Article (refereed)

Carvalho, Laurence; Miller, Claire A.; Scott, E. Marian; Codd, Geoffrey A.; Davies, P. Sian; Tyler, Andrew N.. 2011 Cyanobacterial blooms: statistical models describing risk factors for national-scale lake assessment and lake management. *Science of the Total Environment*, 409 (24). 5353-5358. 10.1016/j.scitotenv.2011.09.030

Cyanobacterial blooms: statistical models describing risk factors for national-scale lake assessment and lake management

Laurence Carvalho[a*], Claire A. Miller (nee Ferguson)[b], E. Marian Scott[b], Geoffrey A. Codd[c,] , P. Sian Davies[d] and Andrew Tyler[e]

[a]Centre for Ecology & Hydrology, Bush Estate, Penicuik, Midlothian, EH26 0QB, UK
[b]School of Mathematics and Statistics, University of Glasgow, Glasgow, G12 8QW, UK
[c]Division of Molecular Microbiology, College of Life Sciences, University of Dundee, Dundee DD1 4HN, UK
[d]Environment Agency (EA), Waveney Works, Station Road, Haddiscoe, Great Yarmouth. NR31 9JA, UK
[e]School of Biological and Environmental Science, University of Stirling, Stirling, FK9 4LA, UK

***Corresponding author:** Laurence Carvalho, Centre for Ecology & Hydrology, Bush Estate, Penicuik, Midlothian, EH26 0QB, UK Email: laca@ceh.ac.uk. Tel: +44 131 445 4343

Abstract

Cyanobacterial toxins constitute one of the most high risk categories of waterborne toxic biological substances. For this reason there is a clear need to know which freshwater environments are most susceptible to the development of large populations of cyanobacteria. Phytoplankton data from 134 UK lakes were used to develop a series of Generalised Additive Models and Generalised Additive Mixed Models to describe which kinds of lakes may be susceptible to cyanobacterial blooms using widely available explanatory variables. Models were developed for log cyanobacterial biovolume. Water colour and alkalinity are significant explanatory variables and retention time and TP borderline significant (R^2_{adj} = 21.9 %). Surprisingly, the models developed reveal that nutrient concentrations are not the primary explanatory variable; water colour and alkalinity were more important. However, given suitable environments (low colour, neutral-alkaline waters), cyanobacteria do increase with both increasing retention time and increasing TP concentrations, supporting the observations that cyanobacteria are one of the most visible symptoms of eutrophication, particularly in warm, dry summers. The models can contribute to the assessment of risks to public health, at a regional- to national level, helping target lake monitoring and management more cost-effectively at those lakes at highest risk of breaching World Health Organisation guideline levels for cyanobacteria in recreational waters. The models also inform restoration options available for reducing cyanobacterial blooms, indicating that, in the highest risk lakes (alkaline, low colour lakes), risks can generally be lessened through management aimed at reducing nutrient loads and increasing flushing during summer.

Keywords: Algal bloom, blue-green algae, cyanotoxin, phosphorus, restoration, Water Framework Directive

1. Introduction

Cyanobacteria are natural inhabitants of freshwaters, where they fulfil important roles in primary production, nitrogen fixation and the cycling of matter (Howarth et al., 1988). They can, however, present hazards to the health of humans and other animals when large populations flourish to produce blooms and particularly when these accumulate on lake surfaces or along shorelines as scums. They constitute a major health hazard as they frequently produce numerous potent toxins (cyanotoxins) that can result in a range of adverse health effects from mild, e.g. skin irritations and gastrointestinal upsets, to fatal (Codd et al., 1999, 2005). Cyanotoxins constitute one of the most high risk categories of waterborne toxic biological substances. This is because not only are the health hazards which they present significant, but exposure to potentially harmful doses of the cyanotoxins can occur, with blooms and scums being a common annual feature in many lakes or reservoirs worldwide. There is, therefore, a great need for understanding where and when cyanobacterial blooms are likely to occur and to what extent. This knowledge would help target lake monitoring and management more efficiently at those lakes at highest risk of breaching World Health Organisation (WHO) and national guidelines (Chorus & Bartram, 1999; WHO, 2003, 2004).

Research over recent decades has identified a number of physical factors that favour cyanobacterial blooms, with the main focus on seasonal drivers, such as warmer temperatures, windiness and consequently the intensity of thermal stratification of the water column (Foy et al., 1976; Mischke, 2003; Reynolds, 2006). Bloom-forming cyanobacteria have been shown to be favoured by high alkalinities and associated high pH (Shapiro, 1984). It is also a widely held view that the increasing magnitude and frequency of cyanobacterial blooms is primarily related to the nutrient enrichment of freshwaters. Indeed there have been several studies showing empirically that bloom frequency is related to the general nutrient status of a lake (Gorham et al., 1974; Dokulil & Teubner 2000; Downing et al., 2001; Reynolds & Petersen, 2000; Schindler et al., 2008). Supporting evidence of a relationship between nutrient enrichment and cyanobacterial abundance is largely derived from long-term studies of enrichment at a few selected individual sites, usually lowland, alkaline, eutrophic lakes, and often examining individual cyanobacterial species. There have been a few published studies examining the relative % abundance of cyanobacteria across eutrophication gradients in large datasets (Downing et al., 2001; Ptacnik et al., 2008), but a more comprehensive quantitative analysis of what factors affect actual cyanobacterial abundance across a wide range of lake types at a regional or national scale has not been carried out.

A more ‘global’ approach has been adopted to develop empirical statistical models to predict the amount of phytoplankton chlorophyll *a* for given concentrations of phosphorus in lake waters (OECD, 1982; Phillips et al., 2008) and also to predict natural background chlorophyll concentrations for individual lakes (Carvalho et al., 2009). The present study aims to take a similar statistical modelling approach, but to more specifically model the variability in cyanobacteria at a national level. It aims to understand the key environmental drivers which are routinely available that favour cyanobacterial abundance in lakes across the UK. We do not aim to understand the distribution and abundance of individual cyanobacterial species or functional groups as it is the risk from all potentially toxic cyanobacteria that is important. The models can be used to help identify which lakes are most susceptible to developing cyanobacterial blooms, enabling a more proactive, rather than reactive, strategy for monitoring and managing cyanobacterial health risks and other adverse impacts, including guiding

restoration measures. More generally, the analysis examines whether empirical evidence from a large selection of lakes spanning a range of environmental conditions, supports nutrient concentrations as the key driver of cyanobacterial abundance in freshwater lakes.

2. Material and Methods

2.1 Data

Phytoplankton data were available for summer months (July, August, September) from UK lakes during the period 2003 to 2006. Samples were either integrated tube samples from the middle of the lake, or where boat access was not possible, a sub-surface sample taken 0.3 m below the surface, using a weighted bottle and float attached to a rope and thrown from the shore near the outflow. The varying sampling method may add variability to the results (discussed later), although previous studies have shown that open water and outflow samples rarely differ markedly in terms of chlorophyll *a* (e.g. Bailey-Watts, 1978). Samples were preserved with Lugol's iodine solution and stored for counting for no longer than 1 year. Phytoplankton were counted using 5-10 mL Utermöhl sedimentation chambers at a range of magnifications (from x40 to x500) depending on cell size. In general, 400 counting units were measured across magnifications using low magnification full-chamber counts, intermediate magnification transects and high magnification fields of view. Counts and biovolume estimates of cells, colonies and filaments were made following the approach outlined by CEN (2004) and Brierley et al., (2007). A series of training workshops and ring-counts were undertaken to ensure these guidance documents were followed correctly and taxonomic identities were standardised. Sub-samples were also taken for nutrient and alkalinity measurements and analyses were carried out by accredited methods at UK environment agencies analytical chemistry laboratories. Colour was measured using absorbance at 400 nm, with measurement of a 100 Hazen standard solution at 400nm used to convert the colour results to Hazen units.

Usually only one sample was available per month. If more than one sample was available, data were averaged by month. Some lakes were represented in the dataset by samples from different months in the same year. The 262 samples were taken from 134 lakes, with 63 lakes having more than one monthly sample in a particular year. Table 1 highlights that most of the data are for 2004 and 2005, however, there was a similar amount of data available for each month overall.

Table 2 lists the principal cyanobacteria genera considered as potentially toxin-forming in UK lakes. Taxonomy broadly followed John et al. (2002) which uses the genus name *Oscillatoria* for taxa that other floras refer to as *Planktothrix*. Records for *Snowella* and *Gomphosphaeria* were grouped as they were sometimes lumped together by counters. The most common genera known to cause toxin problems in UK freshwaters include *Microcystis*, *Aphanizomenon*, *Snowella*, *Oscillatoria* (*Planktothrix*) and *Anabaena*. The likelihood of an individual bloom of these genera containing potent toxins ranges from about 40% to at least 90% (Codd et al., 1999). The other genera were included in the analysis as they are associated with toxicity, although the toxic components are not so well characterised and they also tend to bloom less frequently.

The natural log cyanobacteria biovolume (μm^3 ml^{-1}) was taken as the response

here with a small arbitrary constant of 0.001 added, before transforming, to eliminate zeros. Cyanobacterial biovolume is a direct measure of actual cyanobacteria abundance and can be, therefore, related to potential toxin concentrations (Codd et al., 2005). No a-priori subjective decisions were made in selecting explanatory variables. Data were obtained for the following widely-available explanatory variables: lake area (km^2), altitude (m above sea level), mean depth (m), alkalinity (mEq L^{-1}), colour (Pt L^{-1}), retention time (years), total phosphorus (TP) (µg L^{-1}), total nitrogen (TN) (mg L^{-1}) and chlorophyll$_a$ (µg L^{-1}). Retention times were taken from the UK lakes database (http://www.uklakes.net/) and were estimates based on lake volume and 30-year average annual total rainfall in the catchment. Table 3 includes summary statistics for all explanatory variables and for the response of cyanobacteria biovolume. It illustrates that there is a small amount of missing data for five of the variables.

2.2 Statistical Methods

Due to the presence of non-linear and non-monotonic relationships between some of the explanatory variables and the cyanobacterial response being investigated, generalised additive models, GAMs (Hastie & Tibshirani, 1990; Wood, 2006) were adopted throughout, assuming normal errors. This approach, with a log cyanobacterial response, was taken since the range of values for cyanobacteria is very large going from 0 to 7.5 x 10^8 and adding a very small constant before transforming reflects the fact that it is unlikely that all of the zeros are truly zero counts. In these models the relationship between the response and the explanatory variables is allowed to be a smooth function instead of restricting relationships to be linear.

Since some of the lakes have measurements recorded for more than one month within the same year, generalised additive mixed models, GAMMs (Pinheiro & Bates, 2000: Wood, 2006) were also fitted. In these models a random intercept is included for lake in order to explore the effect of multiple samples from individual lakes on the results.

Models were developed for the whole lake dataset. In addition, as bloom-forming cyanobacteria are a typical feature of alkaline lakes, models were also developed for a sub-set containing all medium alkalinity (MA: 0.2-1.0 mEq L^{-1}) and high alkalinity (HA: >1.0 mEq L^{-1}) lakes (i.e. excluding low alkalinity lakes <0.2 mEq L^{-1} that are likely to have little or no cyanobacteria).

All of the models were fitted using the `gam` and `gamm` functions in the `mgcv` package (1.7-6) (Wood, 2011) of R version 2.13.1 (R Development Core Team 2011), which is free software available at http://www.r-project.org. Since the aim of the modelling was to identify relationships between potential explanatory variables and the cyanobacterial response, the following modelling strategy was employed. For each of the models fitted, all available data were used and the estimated degrees of freedom (edf), used to determine the amount of smoothing for each explanatory variable, has been automatically selected by the model fitting procedure, using restricted maximum likelihood (REML). REML and maximum likelihood (ML) are less prone to local minima than other criteria and Wood (2011) highlights evidence that smoothness selection based on REML/ML may offer an improvement in terms of mean square error performance over other automatic selection methods. However, models were also re-fitted using generalised cross validation (GCV) for smoothness selection and the *p*-values were very similar, with results in terms of significance of model terms unchanged, see Wood (2004, 2008 and 2011) for a discussion on smoothness selection.

Since the amount of smoothing has been estimated during the model fitting process, the *p*-values (testing the null hypothesis that the smooth term is constant) may be less conservative than stated. Therefore, proportion deviance explained by the fitted model and R^2 –adj were also used to identify important explanatory variables, along with the individual *p*-values for each model term. After fitting the model, plots were produced to highlight the relationship between each of the explanatory variables and cyanobacterial abundance and to indicate the shape of the relationships. A stepwise model-fitting procedure was used, whereby, the first model contains all potential explanatory variables, and subsequent models are fit, where the non-significant terms are removed sequentially. The full and reduced models are discussed here.

Natural log transforms of all the explanatory variables were appropriate to reduce skewness in the data distributions. Two distinct outliers were removed from the biovolume data as they appeared to be large transcription errors. There were also 4 high outliers in the retention time data. These were all reservoirs and were, therefore, removed from the analyses as the actual retention time was likely to be artificially managed and much lower.

3. Results

3.1 Correlations between responses and explanatory variables

A correlation analysis highlighted that log area, log depth and log retention time were highly significantly correlated (Table 4). From this analysis it was decided to remove the former two from the model fitting and retain their influence through retention time. This was to reduce the possibility of concurvity in the model, where concurvity is the nonparametric equivalent of multicollinearity. Of the nutrient factors, both log TN and log TP were highly significantly correlated. As log TP had a stronger correlation with chlorophyll$_a$ and cyanobacterial biovolume it was retained for model fitting. On this basis, the variables considered in the final model fitting were log altitude, log retention time (chosen over depth and area), log alkalinity, log colour and log TP (chosen over TN and chlorophyll$_a$).

3.2 Response of Log Cyanobacterial Biovolume

The results for the smooth terms in the GAMMs and GAMs are presented in Table 5 and it can be seen that the results are very similar. This highlights that the clustering within lake in the data is having little effect and the GAM results will be discussed in detail from this point on.

The full GAM using all possible explanatory variables explained 26.9 % of the deviance in log cyanobacterial biovolume (R^2-adj = 23.1 %, n=207) (Table 5). Removing non-significant terms from the full model resulted in a model incorporating log colour, log alkalinity, and borderline significant log TP and log retention time (deviance explained 25.0%, R^2-adj = 21.9%, n=207) (Table 6). Log retention time and log TP concentrations showed positive linear relationships with the response with coefficients (standard errors) of 0.613 (0.316) and 0.858 (0.457) respectively. Cyanobacteria were not abundant in lakes with a retention time <30 days and TP concentrations <20 µg L^{-1} (Figure 1). Humped relationships were apparent with log colour and log alkalinity with peak biovolumes modelled at 12 Pt L^{-1} and 1.9 mEq L^{-1} (Figure 1). The model for the subset of medium and high alkalinity lakes (Table 5) explained slightly more variability in the response than the whole dataset model (deviance explained = 29.5%, R^2-adj =

23.9 %, n = 151). The reduced model incorporating log colour ($p < 0.001$) and log retention time (0.01) explained 19.6 % of the deviance in log cyanobacterial biovolume (R^2-adj = 17.3 %, n=157). The forms of the relationships were the same as with the all-lake dataset with the highest cyanobacterial biovolumes found in poorly flushed, low humic lakes with a water colour of about 10-20 Pt L^{-1}.

4. Discussion

An improved understanding of which lakes are susceptible to the development of large populations of potentially toxic cyanobacteria can inform regional- and national-scale assessments of risks to public health. In general, colour was the strongest explanatory variable in all models, followed by alkalinity, then retention time and finally TP. Altitude was not selected in any of the models, although this may be because the range of the altitude data was not large, with a high proportion of the lakes of low altitude (<200 m a.s.l.) (Table 3) included in the study dataset. Unsurprisingly, alkalinity was not a significant explanatory variable when only medium and high alkalinity lakes were considered, with retention time becoming much more important in the model.
The forms of the modelled relationships were not unexpected, confirming observations reported from more selective lake studies in the literature. For example, the absence or low abundance of cyanobacteria in lakes of high colour agrees with phytoplankton studies of humic lakes, which describe chrysophyte, cryptophyte and diatom algae as the dominant groups in the community, although cyanobacterial genera such as *Anabaena* and *Woronichinia* can still occur (Arvola et al., 1999). The preference of bloom-forming cyanobacteria for neutral to alkaline waters has also been established in elegant in-lake experimental studies (Reynolds & Allen, 1968; Shapiro 1984). The humped relationship observed, indicating a decline in waters of very high alkalinity, is not reported in the literature, but it may be that these lakes were in fact affected by sea salts, as freshwater cyanobacteria are known to be very sensitive to slight changes in salinity (Paerl, 1988). However, this result should not be over interpreted since the decline occurs towards the end of the smooth function, and such functions can suffer from boundary effects.

The analysis of a large lake dataset has importantly revealed that nutrient concentrations do not appear to be the primary driver of cyanobacteria abundance at a national scale; water colour and alkalinity both appear to be more important. However, given such suitable environments (low colour, neutral-alkaline waters), the analysis did show that cyanobacteria increase with both increasing retention time and increasing TP concentrations. This supports the widely observed increases in bloom incidents in warm, dry summers in individual eutrophic lakes and reservoirs. The impact of retention time on phytoplankton composition has not been so extensively studied, although it has been demonstrated that small phytoplankton with relatively high reproductive rates tend to dominate lakes or seasons when retention times are low (high wash-out) (Dickman, 1969; Bailey-Watts et al., 1990). Large, bloom-forming cyanobacteria are recognised for their slower reproductive rates compared to many diatoms and small green or flagellate algae (Reynolds, 2006) and the positive linear response to retention time observed in these models fits with this knowledge. Elliott (2010) recently highlighted in model simulations at an individual site, that retention time was more important than water temperature in increasing the frequency of algal blooms above WHO thresholds, but that nutrients ultimately controlled the capacity at

an individual site scale. Intensive long-term monitoring at Grasmere in the English Lake District, supports our more broad-scale finding that cyanobacteria were never abundant in lakes with a retention time <30 days. In Grasmere, the filamentous cyanobacterium *Anabaena* was shown to only occur during long, dry summers when retention times are high (Reynolds & Lund, 1988), and *Planktothrix* spp. are generally not recorded from temperate lakes with a retention time less than about 30 days (Reynolds, 2006). Given cyanobacterial replication rates, losses due to wash-out are likely to have some effect in lakes with retention times <100 days (Reynolds, 2006). If the growth of cyanobacterial populations is further limited by additional factors (light, nutrients), then wash-out loss processes may become even more significant.

The positive, linear relationship with TP concentrations conforms with widespread empirical evidence that shows cyanobacteria to increase with nutrient enrichment (Gorham et al., 1974; Reynolds & Petersen, 2000). This response is almost certainly not just a simple population growth response to increased nutrient resources, but may also be because cyanobacteria become competitively more dominant than other algal groups in enriched waters, due to their tolerance of the associated low CO_2 and light environments (Dokulil & Teubner 2000; Reynolds 2006) and associated shifts in the availability of nitrogen (Hyenstrand et al., 1998). Grazer avoidance or mortality associated with large colonies and/or toxic forms have also been reported (Paerl, 1988). TP was correlated with alkalinity in the dataset, however, both TP and alkalinity were still identified as separate explanatory variables in the final model for log cyanobacteria biovolume.

The models provide quantitative support to the more qualitative literature which shows that cyanobacteria tend to be abundant in clear water lakes of neutral to alkaline waters and that their abundance generally increases with increasing retention time and TP concentrations (Reynolds & Lund, 1988; Reynolds & Petersen, 2000). Bloom-forming cyanobacteria are typically slower-growing than other classes of phytoplankton and this analysis supports the fact that they do best where resources (nutrients) are plentiful and loss rates to flushing are minimal.

5. Conclusions

This study identifies higher risk lake environments where more targeted monitoring of cyanobacteria biovolumes should be focused (e.g. water colour 10-20 Pt L^{-1}, alkalinity >1 mEq L^{-1}, retention time >30 days, TP >20 µg L^{-1}).

The R-sq values for the statistical models are low, therefore, their predictive capability at an individual site level is very limited. One reason for the limited explanatory power is that we are attempting to explain the environmental conditions suitable for a mixed community of up to 17 genera of potentially toxic cyanobacteria, all with different environmental preferences, across a large gradient of lake types. Additional reasons for uncertainty may be because of additional explanatory variables not considered in the models, such as grazer densities, water temperature or stratification intensity. Finally, model uncertainty may also be due in part to measurement errors or natural variability in both the response (e.g. estimates of cyanobacteria colony biovolumes) and explanatory variables (e.g. TP and alkalinity). The different type of sampling (e.g. sub-surface outflow vs. integrated epilimnion samples) and different counters will have added additional variability.

The fact that development of blooms and in particular, high-risk surface scums is very uncertain and is also dependent on weather conditions, does highlight that there is a great need for more cost-effective monitoring of the spatial and temporal extent of cyanobacterial blooms. This study identifies higher risk lake environments and looking forward, the next generation of earth observation satellite platforms due to be launched between 2011 and 2013 are set to provide the spatial and temporal resolution and hyperspectral imaging capabilities necessary for effective cyanobacteria monitoring in inland regions where risks are greatest (Hunter et al., 2009; 2010).

Despite uncertainties in the models, the research does highlight possible management measures open to lake managers for reducing cyanobacterial abundance. Risks are greatest in clear water lakes of neutral to alkaline waters. In these lakes, cyanobacterial abundance generally increases with increasing retention time and TP concentrations. The analysis, therefore, supports management aimed at reducing nutrients loads from catchments. In addition to this, in many lakes with controlled outflows, it may also be possible that lake retention times can be manipulated to maximise summer flushing.

Acknowledgments
We thank the Environment Agency of England & Wales and the Scottish Environment Protection Agency for giving access to their data for this work. Funding was provided by the Natural Environment Research Council's Environment and Human Health Programme (Project reference NE/E009360/1).

References

Arvola L, Eloranta P, Järvinen M, Keskitalo J, Holopainen A-L. Food webs of humic waters: Phytoplankton. In Keskitalo J, Eloranta P, editors, Limnology of Humic Waters Backhuys Publishers, Leiden; 1999. pp. 137-171.

Bailey-Watts AE. A nine-year study of the phytoplankton of the eutrophic, and non-stratifying Loch Leven (Kinross, Scotland), Journal of Ecology 1978; 66:741-771.

Bailey-Watts AE, Kirika A, May L, Jones DH. Changes in Phytoplankton over various time scales in a shallow, eutrophic: the Loch Leven experience with special reference to the influence of flushing rate. Freshwater Biology 1990; 23:85-111.

Brierley B, Carvalho L, Davies S, Krokowski J. Guidance on the quantitative analysis of phytoplankton in Freshwater Samples. Report to SNIFFER (Project WFD80), SNIFFER, Edinburgh, December 2007.

Carvalho L, Solimini A, Phillips G, van den Berg M, Lyche Solheim A, Mischke U, Pietiläinen O-P, Poikane S, Tartari G. Predicting site-specific chlorophyll reference conditions for European lakes. Hydrobiologia 2009; 633:59-66.

CEN. Water quality – Guidance standard for the routine analysis of phytoplankton abundance and composition using inverted microscopy (Utermöhl technique) CEN TC 230/WG 2/TG 3/N83, May 2004.

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

Codd GA, Bell SG, Kaya K, Ward CJ, Beattie KA, Metcalf JS. Cyanobacterial toxins, exposure routes and human health. European Journal of Phycology 1999; 34:405-415.

Codd GA, Morrison LF, Metcalf JS. Cyanobacterial toxins: risk management for health protection. Toxicology and Applied Pharmacology 2005; 203:264-272.

Dickman M. Some effects of lake renewal on phytoplankton productivity and species composition. Limnology & Oceanography 1969; 14:660-666.

Dokulil MT, Teubner K. Cyanobacterial dominance in lakes. Hydrobiologia 2000;438: 1-12.

Downing JA, Watson SB, McCauley E.. Predicting cyanobacteria dominance in lakes. Canadian Journal of Fisheries and Aquatic Sciences 2001; 58:1905-1908.

Elliott A. The seasonal sensitivity of Cyanobacteria and other phytoplankton to changes in flushing rate and water temperature, Global Change Biology 2010; 16:864-876.

Foy RH, Gibson CE & Smith RV. The Influence of Daylength, Light Intensity and Temperature on the Growth Rates of Planktonic Blue-Green Algae. British Phycological Journal 1976; 2:151-163.

Gorham E, Lund JWG, Sanger JE, Walter EDJ. Some Relationships Between Algal Standing Crop, Water Chemistry, and Sediment Chemistry in the English Lakes Limnology and Oceanography 1974; 19:601-617.

Hastie TJ, Tibshirani RR. Generalized Additive Models, Chapman & Hall, London; 1990.

Howarth RW, Marino R, Lane J, Cole JJ. Nitrogen-fixation in freshwater, estuarine and marine ecosystems. 1. Rates and importance. Limnology and Oceanography 1988; 33:669-687.

Hunter PD, Tyler AN, Gilvear DJ, Willby NJ. Using remote sensing to aid the assessment of human health risks from blooms of potentially toxic cyanobacteria Environmental Science and Technology 2009; 43:2627–2633.

Hunter PD, Tyler AN, Carvalho L, Codd, GA, Maberly SC. Hyperspectral remote sensing of cyanobacterial pigments as indicators for cell populations and toxins in eutrophic lakes. Remote Sensing of Environment, 2010;114:2705-2718.

Hyenstrand P, Blomqvist P, Petersson A. Factors determining cyanobacterial success in aquatic systems - a literature review. Archiv Hydrobiologie 1998; Special Issue: Lake Erken - 50 Years of Limnological Research; 41-62.

John DM, Whitton BA, Brook AJ. The Freshwater Algal Flora of the British Isles. University Press, Cambridge 2002.

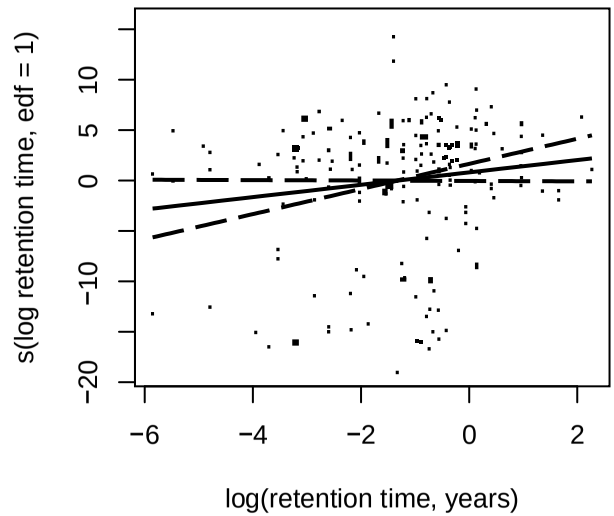
Mischke U. Cyanobacteria associations in shallow polytrophic lakes: influence of environmental factors. Acta Oecologica-International Journal of Ecology 2003; 24: S11-S23.

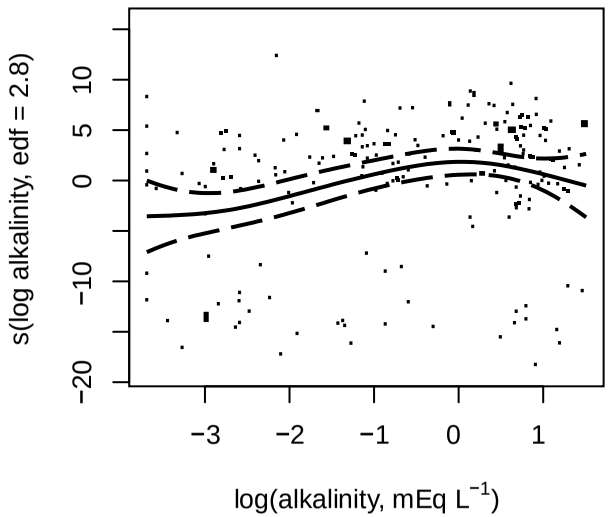
OECD. Eutrophication Of Waters. Monitoring, Assessment And Control, OECD, Paris; 1982. 156 pp.

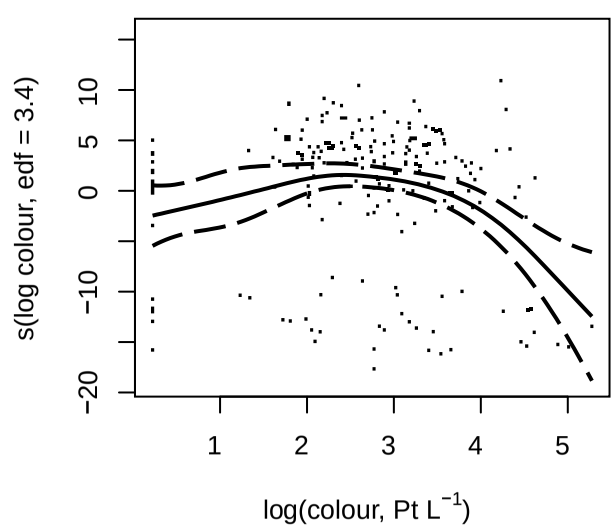
Paerl HW. Growth and reproductive strategies of freshwater blue-green algae (cyanobacteria). Growth and Reproductive Strategies of Freshwater Phytoplankton (ed CD Sandgren), Cambridge University Press, Cambridge; 1988. pp. 261-315.

Phillips G, Pietilainen OP, Carvalho L, Solimini A, Solheim AL, Cardoso AC. Chlorophyll-nutrient relationships of different lake types using a large European dataset. Aquatic Ecology 2008;42:213-226.

Pinheiro J.C. and Bates, D.M. Mixed effects Models in S and S-PLUS. Springer 2000.
Ptacnik R, Lepisto L, Willen E, Brettum P, Andersen T, Rekolainen S, Solheim AL, Carvalho L. Quantitative responses of lake phytoplankton to eutrophication in Northern Europe. Aquatic Ecology 2008;42:227-236.
R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2011. ISBN 3-900051-07-0, URL http://www.R-project.org/.
Reynolds CS. Ecology of Phytoplankton. Cambridge University Press, Cambridge; 2006.
Reynolds CS Allen SE. Changes in the phytoplankton of Oak Mere, following the introduction of base rich water. British Phycological Journal 1968;3:451-462.
Reynolds CS, Lund JWG. The phytoplankton of an enriched, soft-water lake subject to intermittent hydraulic flushing (Grasmere, English Lake District). Freshwater Biology 1988;19:379-404.
Reynolds CS, Petersen AC. The distribution of planktonic Cyanobacteria in Irish lakes in relation to their trophic states. Hydrobiologia 2000;424:91-99.
Schindler DW, Hecky RE, Findlay DL, Stainton MP, Parker BR, Paterson MJ, Beaty KG, Lyng M, Kasian SEM. Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. Proceedings of the National Academy of Sciences of the United States of America 2008;105:11254-11258.
Shapiro J. Blue-green dominance in lakes: The role and management significance of pH and CO_2. Internationale Revue der Gesamten Hydrobiologie 1984;69:765-780.
WHO. Guidelines for Safe Recreational Water Environments. World Health Organisation, Geneva; 2003.
WHO. Guidelines for Drinking-water Quality, 3rd edn. World Health Organisation, Geneva; 2004.
Wood SN. Stable and efficient multiple smoothing parameter estimation for generalized additive models. Journal of the American Statistical Association 2004; 99:673-686.
Wood SN. Generalized Additive Models, An Introduction with R, Chapman & Hall, Boca Raton; 2006.
Wood SN. Fast stable direct fitting and smoothness selection for generalized additive models. J.R.Statist. Soc. B 2008; 70(3):495-518
Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B) 2011; 73(1):3-36


s(log retention time, edf = 1)
10
5
0
−10
−20
−6
−4
−2
0
2
log(retention time, years)


s(log alkalinity, edf = 2.8)
10
5
0
−10
−20
−3
−2
−1
0
1
log(alkalinity, mEq L^{-1})


s(log colour, edf = 3.4)
10
5
0
−10
−20
1
2
3
4
5
log(colour, Pt L^{-1})

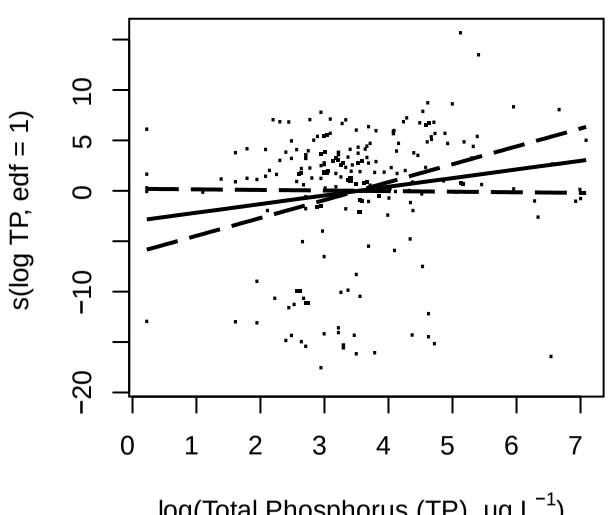

s(log TP, edf = 1)
10
5
0
−10
−20
0
1
2
3
4
5
6
7
log(Total Phosphorus (TP), µg L^{-1})

Table 1. Summary of no. of lake samples available by month and year

Year	July	August	September
2003	0	2	2
2004	56	48	72
2005	15	24	18
2006	8	16	1
All years	79	90	93

Table 2. Cyanobacterial taxa considered as potential toxin-producers, the number of samples present in the dataset and their total biovolume

Taxon	**Samples (n)**	**Total Biovolume in dataset ($\mu m^3 mL^{-1}$)**
Microcystis	30	802498550
Aphanizomenon	38	332888817
Snowella / Gomphosphaeria	39	61838799
Chroococcus	68	59611048
Gloeotrichia	4	57138196
Planktothrix (*Oscillatoria*)	116	45769778
Aphanocapsa	53	33899849
Anabaena	103	24799040
Aphanothece	50	23171273
Woronichinia / Coelosphaerium	37	13709770
Lyngbya	4	2692241
Merismopedia	34	473408
Phormidium	2	14967
Anabaenopsis	1	13485
Pannus	6	3517
Spirulina	1	242
Pseudanabaena	3	67

Table 3. Summary statistics for all explanatory variables and cyanobacteria biovolume response.

	Min.	Mean	Max.	No. missing
Area (km^2)	0.01	1.84	27.98	0
Altitude (m a.s.l)	1	120	455	0
Mean Depth (m)	0.3	8	42	3
Alkalinity (mEq L^{-1})	0.02	1.11	4.53	11
Colour (Pt L^{-1})	1	24	196	37
Retention Time (yr)	0.003	1.4	9.7	3
TN (µg L^{-1})	25	974	7880	38
TP (µg L^{-1})	1	80	1197	0
Chl_a (µg L^{-1})	1	18	292	0
Cyanobacteria biovolume (μm^3 mL^{-1})	0	$5.1x10^6$	$7.5x10^8$	0

Table 4. Pearson correlation coefficients (and p values). Significant relationships highlighted in bold.

	Log Altitude	**Log Area**	**Log Depth**	**Log Retention Time**	**Log Alkalinity**	**Log Colour**	**Log TN**	**Log TP**	**Log chla**
Log Area	0.054								
	0.388								
Log Depth	**0.287**	**0.625**							
	<0.001	**<0.001**							
Log Retention Time	0.005	**0.272**	**0.396**						
	0.933	**<0.001**	**<0.001**						
Log Alkalinity	**-0.440**	**-0.305**	**-0.475**	0.011					
	<0.001	**<0.001**	**<0.001**	0.869					
Log Colour	0.097	**-0.282**	**-0.377**	-0.036	0.083				
	0.145	**<0.001**	**<0.001**	0.595	0.226				
Log TN	-0.122	-0.075	-0.102	0.099	**0.474**	0.090			
	0.068	0.265	0.126	0.147	**<0.001**	0.180			
Log TP	**-0.178**	**-0.205**	**-0.312**	0.095	**0.543**	**0.199**	**0.392**		
	0.004	**0.001**	**<0.001**	0.131	**<0.001**	**0.003**	**<0.001**		
Log chla	**-0.239**	**-0.158**	**-0.262**	0.043	**0.565**	0.059	**0.400**	**0.476**	
	<0.001	**0.010**	**<0.001**	0.498	**<0.001**	0.374	**<0.001**	**<0.001**	
Log cyano biovolume	**-0.168**	0.023	-0.025	**0.174**	**0.269**	-0.039	**0.194**	**0.227**	**0.369**
	0.006	0.711	0.685	**0.005**	**<0.001**	0.565	**0.004**	**<0.001**	**<0.001**

Table 5. GAMM and GAM results for log cyanobacteria biovolume response and log explanatory variables for all lakes and sub-set of medium (MA) and high (HA) lakes. Results are *p*-values to assess the null hypothesis that each smooth term is constant and (edf) = estimated degrees of freedom.

p-values (edf)	**Log Altitude**	**Log Retention Time**	**Log Alkalinity**	**Log Colour**	**Log TP**
All lakes (GAMM)	0.185 (2.192)	0.069 (1)	**0.009 (2.647)**	**<0.001 (3.482)**	0.095 (1)
All lakes (GAM)	0.264 (2.192)	0.069 (1)	**0.024** (2.648)	**<0.001** (3.482)	0.095 (1)
MA & HA lakes (GAMM)	0.342 (1.892)	**0.025 (1)**	0.128 (2.256)	**<0.001 (3.346)**	0.385 (2.509)
MA & HA lakes (GAM)	0.360 (1.892)	**0.025** (1)	0.191 (2.256)	**<0.001** (3.346)	0.276 (2.509)

Table 6. GAM results for log cyanobacteria biovolume response and log explanatory variables after removing non-significant* terms sequentially. Results are *p*-values to assess the null hypothesis that each smooth term is constant and (edf) = estimated degrees of freedom.

All lakes	**Log Altitude**	**Log Retention Time**	**Log Alkalinity**	**Log Colour**	**Log TP**
p-values	*	0.054	**0.007**	**<0.001**	0.062
edf	*	1	2.814	3.363	1