Environmental Science & Technology

Article

pubs.acs.org/est

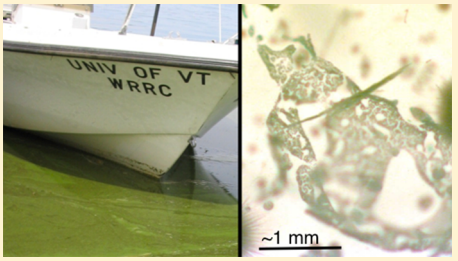
Unraveling Associations between Cyanobacteria Blooms and In-Lake Environmental Conditions in Missisquoi Bay, Lake Champlain, USA, Using a Modified Self-Organizing Map

Andrea R. Pearce,[†,*] Donna M. Rizzo,[†] Mary C. Watzin,[‡,∥] and Gregory K. Druschel[‡,§,⊥]

[†]School of Engineering, [‡]Rubenstein School of Environment and Natural Resources, and [§]Department of Geology, University of Vermont, Burlington, Vermont 05405

Supporting Information

ABSTRACT: Exploratory data analysis on physical, chemical, and biological data from sediments and water in Lake Champlain reveals a strong relationship between cyanobacteria, sediment anoxia, and the ratio of dissolved nitrogen to soluble reactive phosphorus. Physical, chemical, and biological parameters of lake sediment and water were measured between 2007 and 2009. Cluster analysis using a self-organizing artificial neural network, expert opinion, and discriminant analysis separated the data set into no-bloom and bloom groups. Clustering was based on similarities in water and sediment chemistry and non-cyanobacteria phytoplankton abundance. Our analysis focused on the contribution of individual parameters to discriminate between no-bloom and bloom groupings. Application to a second, more spatially diverse data set, revealed similar no-bloom and bloom discrimination, yet a few samples possess all the physicochemical characteristics of a bloom without the high cyanobacteria cell counts, suggesting that while specific environmental conditions can support a bloom, another environmental trigger may be required to initiate the bloom. Results highlight the conditions coincident with cyanobacteria blooms in Missisquoi Bay of Lake Champlain and indicate additional data are needed to identify possible ecological contributors to bloom initiation.



INTRODUCTION

Cyanobacteria are a ubiquitous component of freshwater phytoplankton communities worldwide, but in eutrophic waters, these algae can reach nuisance proportions.[1,2] Cyanobacteria blooms are the result of complex interactions among phytoplankton, zooplankton grazers, nutrients, and other biotic and abiotic factors[3,4] and are problematic due to their density and because some cyanobacteria produce secondary toxic metabolites.[5]

Over the past decade, summer cyanobacteria bloom frequency has increased in Lake Champlain (New York and Vermont, USA, and Quebec, Canada), with regular occurrence in Missisquoi Bay (Figure 1).[6] Intensive quantitative cyanobacteria monitoring shows dominance by three genera: *Microcystis*, *Anabaena*, and *Aphanizomenon*, with relative proportions varying from year to year.[6] Microcystins are the predominant cyanotoxin group within the bay.[7] Despite intensive monitoring, the bloom ecology and links to water quality conditions remain poorly understood.

Phosphorus (P) is believed to be the dominant nutrient tied to increases in Lake Champlain phytoplankton.[8] Under low nitrogen (N) conditions, cyanobacteria can dominate other algal groups, especially when P is abundant.[9] Some cyanobacteria can fix N_2 or have other physiological adaptations enabling efficient uptake of dissolved N and outcompete other phytoplankton for nitrogen.[10] When the N:P ratio is low, cyanobacteria may have an advantage, particularly when the ratio drops below 29:1,[11] although these observations are contradicted.[12] Biological activity can stimulate the seasonal development of sediment anoxia and lead to the release of P and ammonium to the overlying water by reduction (and dissolution) of iron oxide-hydroxide minerals that sorb P and from organic material decomposition by heterotrophic bacteria, respectively.[13,14]

Data-driven computational approaches (e.g., artificial neural networks or ANNs) are ideally suited for assimilating the multiple, intercorrelated data associated with cyanobacteria blooms.[15] These ANN algorithms account for more data variability than linear parametric statistics when applied to geochemical and microbiological data sets.[16] We selected the self-organizing map (SOM), a non-linear and non-parametric clustering method because of its robustness with data that violate the assumptions associated with parametric clustering methods. It outperforms many clustering methods (e.g., hierarchical and K-means) on data sets with high dispersion, outliers, irrelevant variables, and non-uniform cluster densities.[17] In general, clustering methods are attractive for exploratory data analysis because the number of groupings does not need *a priori* specification.[18] The SOM has proved useful for highlighting patterns in cyanobacteria dominance[19] and is more robust than traditional methods for clustering hydrochemical data.[20] Although specific results are not transferrable to another system,

Received: August 6, 2013
Revised: November 18, 2013
Accepted: November 19, 2013
Published: November 19, 2013

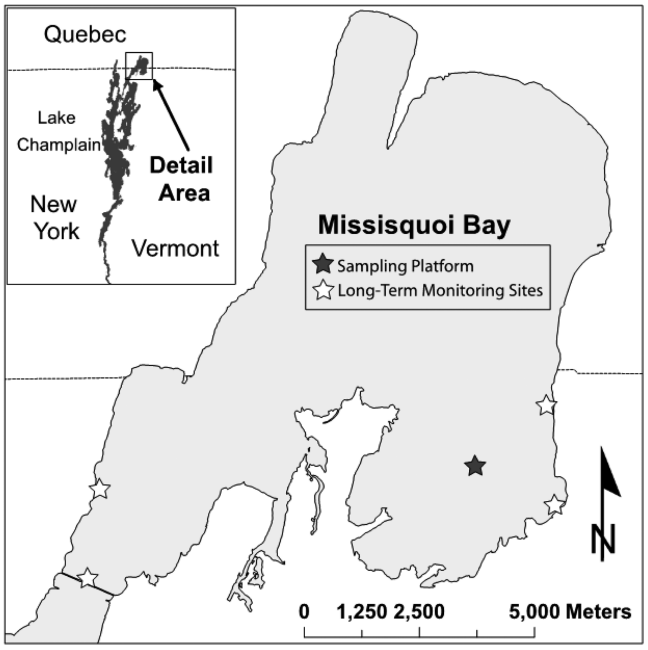


Figure 1. Sampling locations in Missisquoi Bay, Lake Champlain, USA/Canada. Detailed data were collected from the sampling platform. The long-term monitoring data were collected from the platform and at four additional locations in the bay.

the SOM methodology and follow-on 2-D visualizations may be applied to any data set.

In this work, the SOM clusters a unique set of water quality and sediment redox chemistry measurements to examine the in-lake conditions supporting cyanobacteria blooms, especially those of the dominant genus *Microcystis.* Clustering used only *non*-cyanobacteria phytoplankton cell counts, sediment, and water chemistry data. Results were compared to *Microcystis* cell counts with emphasis on individual input parameter contributions, while leveraging the SOM visualization tools, expert opinion, and a modification that allows weighting of input variables.

METHODS

Lake Champlain is the sixth largest lake in the northeastern U.S., draining regions of Vermont and New York, United States and Quebec, Canada into the Richelieu River and ultimately the St. Lawrence River (Figure 1). Missisquoi Bay, straddling the U.S.–Canada border, is hydrodynamically isolated from the rest of Lake Champlain and is one of the most eutrophic segments, exhibiting nearly annual cyanobacteria blooms, at times accompanied by the cyanotoxin, microcystin.[6] The Bay has a maximum depth of ~4 m, a mean depth of 2.8 m, and a watershed to water surface area ratio of 40:1.[21]

The study data set was collected over 12 days in 2007, 2008, and 2009 from a sampling platform in Missisquoi Bay (Figure 1) at multiple times of the day. The intent was to capture conditions (1) after ice-out, prior to the initiation of the bloom and during (2) bloom initiation, (3) peak bloom, and (4) bloom subsidence. During each event, samples were collected three times (mid-day, dusk, and dawn) at multiple depths over a 24-h period, to capture diel patterns in cyanobacteria behavior and chemistry at the sediment–water interface.[22]

Additional samples were collected more frequently from the sampling platform and four additional locations within Missisquoi Bay as part of a long-term monitoring (LTM) partnership between the Vermont Agency of Natural Resources and the University of Vermont (Figure 1). Footnote *a* designation in Table 1 indicates parameters measured as part of this LTM effort.

Data were analyzed using an SOM[23] modified and coded by the first author (MATLAB R2009B, The Mathworks, Natick, MA). The algorithm and its modifications are described in detail

Table 1. Parameters Measured from the Sampling Platform

	parameter	units	minimum	mean	maximum	std dev
nutrients	total phosphorus[a]	μg/L	16.7	55.4	266	45.9
	total nitrogen[a]	mg/L	0.327	0.604	0.997	0.147
	dissolved nitrogen[a]	mg/L	0.29	0.47	0.997	0.157
	soluble reactive phosphorus	μg/L	0.5	6.53	33.26	5.86
	microcystin	μg/L	0.003	3.22	18.5	4.5
phytoplankton	Bacillariophyceae[a]	cells/mL	23.4	292	1694	319.6
	Chlorophyceae[a]	cells/mL	70.2	998	3455	924
	Anabaena[a]	cells/mL	0	3039	17563	4127
	Aphanizomenon[a]	cells/mL	0	1108	9919	2082
	Microcystis[a]	cells/mL	0	31016	319804	63622
physical parameters	temperature[a]	°C	21.1	22.4	24	0.89
	conductivity	μS/cm	82.1	114.6	127.9	12.2
	dissolved oxygen	mg/L	6.52	8.36	10.89	1.12
	PAR (irradiance)	W/m^2	0.057	287.7	1120	421
	fluorescence	mg/m^3	0.98	8.26	18.9	3.78
	turbidity	FTU	3.03	13.3	84	16.9
sediment	dissolved oxygen[b]	μM	5	51.9	150	45.7
	oxic boundary[c]	mm	−1	−0.25	0.5	0.47
	Mn(II) redox boundary[c]	mm	−12	−0.96	10	5.5

[a]Also measured as part of the long-term monitoring data set. [b]Measured at the sediment–water interface by voltametric microelectrode. [c]Relative to the sediment–water interface.

elsewhere.[23,24] Briefly, the SOM teratively self-organizes the input data (i.e., lake water samples) onto two-dimensional maps using a mathematical measure of similarity. Modifications include a mask that enables weighting of the input variables based on their relative importance. Statistical analyses were performed using JMP 8.0.1, SAS Institute, Inc., Cary, NC.

Our choice of the following 10 input data parameters was determined via sensitivity analysis and consultation between coauthors: ln(dissolved nitrogen:soluble reactive phosphorus), cell counts of Bacillariophyceae (diatoms) and Chlorophyceae (green algae) [as ln(cells/mL)], temperature [°C], conductivity [μS/cm], dissolved oxygen (DO) [mg/L], PAR irradiance [W/m^2], chlorophyll fluorescence [ln(mg/m^3)], turbidity [ln-(FTU)], and depth of sediment anoxia measured vertically along the sediment−water interface to the dissolved Mn^{II} [mm] front. Of the measured sediment parameters, depth to the dissolved Mn^{II} front most consistently represented the overall redox condition and potential for nitrate reduction and NH_4 diffusion from the sediment.

Because Missisquoi Bay is shallow, almost always well mixed vertically, and known to have cyanobacteria migrate vertically on a diel basis,[25−27] we assumed sediment redox chemistry impacted cyanobacteria throughout the water column and used the same sediment redox measurement for water samples collected in each vertical profile. The dissolved nitrogen:soluble reactive phosphorus ratio (DN:SRP) was selected instead of total nitrogen:total phosphorus (TN:TP) because a large portion of the total water column nutrients at any time are tied up in organisms or bound to other particles and not available for biological uptake. Natural log transformations were applied when extreme values skewed the analysis. Concentration of microcystins and cyanobacteria cell counts were omitted from the SOM clustering analysis to allow comparison of the SOM-generated clusters to these observed measurements.

Preliminary data analyses revealed that a few samples with low water temperature (in early May and late October) dominated the cluster analysis yet provided no information about what was driving bloom dynamics. *Microcystis* growth is very sensitive to temperature, with growth beginning in earnest when water temperature rises to around 20 °C in the summer. Cells also persist in cooler late-season water by out-competing other algae for space near the water surface allowing blooms to linger with little or no growth.[4,28,29] To eliminate masking of other potentially important associations, samples with water temperatures less than 20 °C were omitted from further analysis (n = 56). Removing the low-temperature samples allows differences related to cyanobacteria bloom dynamics, beyond simple associations with temperature, to emerge during clustering. Removing data that do not capture the phenomenon of interest is a common (and necessary) practice when examining non-stationary, spatially autocorrelated phenomena.[30]

The LTM data set comprises 167 samples collected across all sampling locations (Figure 1) between 2006 and 2009. Although these data are similar to our shorter-term data set, aquatic data are limited, and sediment anoxia measurements do not exist (Table 1). As a result, LTM data for the SOM consists of only 5 input parameters: the natural log of DN:SRP, cell counts of diatoms and green algae [ln(cells/mL)], temperature (>20 °C), and day of year (as a surrogate for sediment anoxia because day of year and the depth to dissolved Mn^{II} are correlated in our shorter-term data set, R^2 = 0.63).

RESULTS

Summary statistics (Table 1) characterize sediment and water quality and assess cyanobacteria bloom conditions. For this work, a 1000 *Microcystis* cells per mL (Figure 2) threshold defines a bloom. Samples with and without a *Microcystis* bloom coincide with lower and higher DN:SRP, respectively. The two samples (Figure 2) with low DN:SRP values and low cyanobacteria cell counts occurred in late July 2008 at ∼13:00 h. Samples collected at ∼20:00 h have similar chemistry but significantly higher cyanobacteria cell counts, likely an indication of the vertical migration of *Microcystis*.

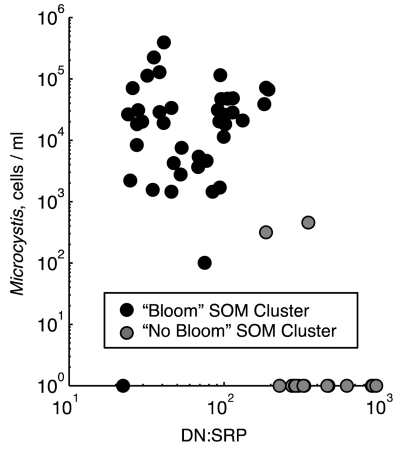


Figure 2. Microcystis cells/mL vs DN:SRP.

An initial SOM organizes the data into groups approximately corresponding to bloom and no-bloom (Figure 3a and b). Circles in Figure 3a identify the final self-organized location of each sample on the 15 × 15 node output map. This 2-dimenionsal map visualizes the non-parametric organization/clustering of the input data and has no physical meaning. Black circles indicate samples associated with a bloom, while gray circles indicate no-bloom as defined by the 1000 cell/mL threshold. A second SOM simulation using the same inputs and settings explicitly forces the

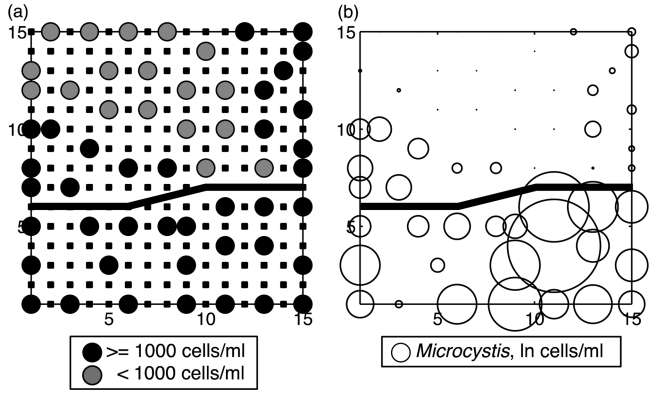


Figure 3. (a) Clustered SOM showing the final location of each sample on the 15 × 15 node output map. Data were directly clustered into 2 groupings, with the separation indicated by the black line. The color of the circles marking map location of each sample is based on a 1000 cell/mL threshold. (b) The final location of each sample on the 15 × 15 node output map is marked with a circle sized proportionally to the count of *Microcystis* cells/mL. Neither *Microcystis* cell counts nor toxin concentrations were included as input data.

data into two clusters using only two output nodes. The black line delineates groups determined by the 2-node SOM. Small squares show unused SOM node locations. Open circles (Figure 3b) identify self-organized samples at the same grid locations as Figure 3a but are now sized by the natural log of the *Microcystis* cell count (cells/mL). All cell counts are plotted as cell count +1 to permit the location of samples with no observed *Microcystis* cells to be visualized on a log scale. Note: neither *Microcystis* cell counts nor toxin concentrations were included as input data when generating these bloom or no-bloom data clusters.

Weighting the SOM inputs using expert opinion refines the bloom/no-bloom groupings (Figure 4a and b). The modified

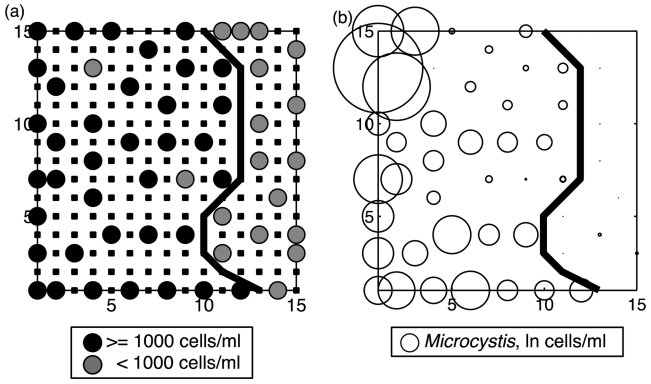


Figure 4. The same information as in Figure 3, but this SOM weights DN:SRP and the depth to the dissolved Mn^{II} front twice as much as the other variables on the recommendation of the subject area experts. Again, neither *Microcystis* cell counts nor toxin concentrations were included as input data. (a) SOM output showing the final location of each sample on the 15 × 15 node output map. Data were explicitly classified into 2 new groupings with the separation between the groupings indicated by the black line. The color of the circles marking map location of each sample is based on a 1000 cell/mL threshold. (b) The final location of each sample on the 15 × 15 node output map is marked with a circle sized proportionally to the count of *Microcystis* cells/mL.

SOM allows users to weight input variables if desired.[24] Our two experts (M.C.W. and G.K.D.) suggested weighting DN:SRP and the depth to dissolved Mn^{II} twice as much as the other variables. Sample clustering (Figure 4a and b) is similar to the unweighted clustering (Figure 3a and b), but the SOM bloom/no-bloom groupings now better correspond to the 1000 *Microcystis* cell/mL threshold. The black division line is superimposed on the 10 two-dimensional SOM component planes (Figure 5), showing the contribution of individual input parameters to the overall organization of the data. The organization of the SOM output data (Figure 4) maps identically to the individual values on the component planes (Figure 5). These component planes enable observations about the characteristics of the groupings. Conductivity and turbidity are distributed relatively uniformly over the map, suggesting little contribution to the clustering. The bloom cluster comprises a large concentration of the highest temperatures. The natural log of DN:SRP and the depth to Mn^{II} (the variables more heavily weighted in the analysis) are lower and higher, respectively in the bloom group.

When the weighted SOM (Figure 4) is applied to the larger, LTM data set, the same division of bloom/no-bloom groupings is observed (Figure 6a–c). The final self-organized sample map locations are labeled with black (bloom) or gray (no-bloom) circles (Figure 6a) as defined by our 1000 cell/mL threshold (Figure 2). In general, samples associated with high *Microcystis* cell counts (open circles sized by the natural log of *Microcystis* cells/mL) cluster to the upper left of the map (Figure 6b). Forty-four samples with less than 1000 *Microcystis* cells/mL are misclassified by the SOM (black circles located to the left of the solid line). Note: Some samples occupy the same map location making it appear that there are fewer misclassified samples. Twenty-one of these misclassified samples were collected in 2007. The black stars (Figure 6c) represent samples collected in 2007, the one year when Missisquoi Bay did not support more than 1000 *Microcystis* cells/mL, though the water quality data are more similar to 'bloom' conditions.

Figure 7 displays a relationship between available nitrogen and the composition of phytoplankton in Missisquoi Bay. At low DN:TN ratios, cyanobacteria at ~20,000 cells/mL are much more abundant relative to green algae and diatoms. At higher DN:TN ratios, green algae and diatoms comprise a larger fraction of the phytoplankton community.

As an alternative to weighting the SOM input by expert opinion, regularized discriminant analysis (DA) was performed ($\gamma = 0.5$, $\rho = 0.5$) to extract the importance (or weight) of input variables using the canonical coefficients generated during sample classification into two categories (bloom and no-bloom). Unlike expert opinion, the DA is a classification tool that requires a predefined threshold (i.e., cell count) to define a bloom; we defined this as 1000 *Microcystis* cells/mL. The DA misclassifies only 2 of the samples associated with high and low *Microcystis* cell counts, while the canonical coefficients show the DN:SRP ratio and the extent of sediment anoxia contribute most

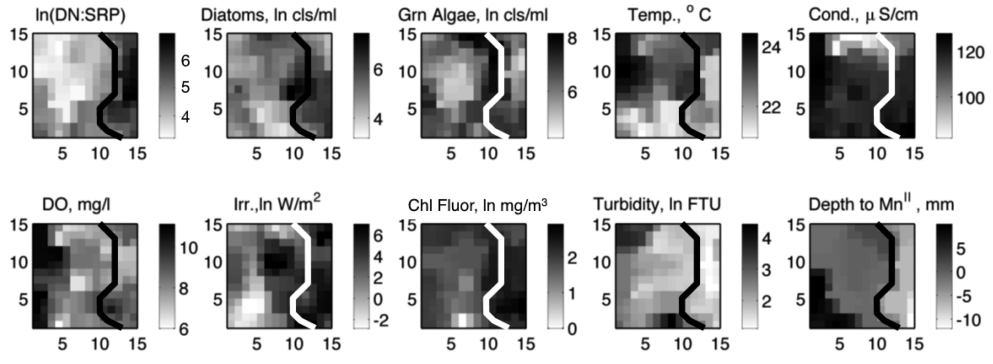


Figure 5. Component planes corresponding to the SOM results of Figure 4. These show the distribution of each input parameter superimposed on the final map and the association with the bloom and no-bloom groupings (i.e., division line Figure 4).

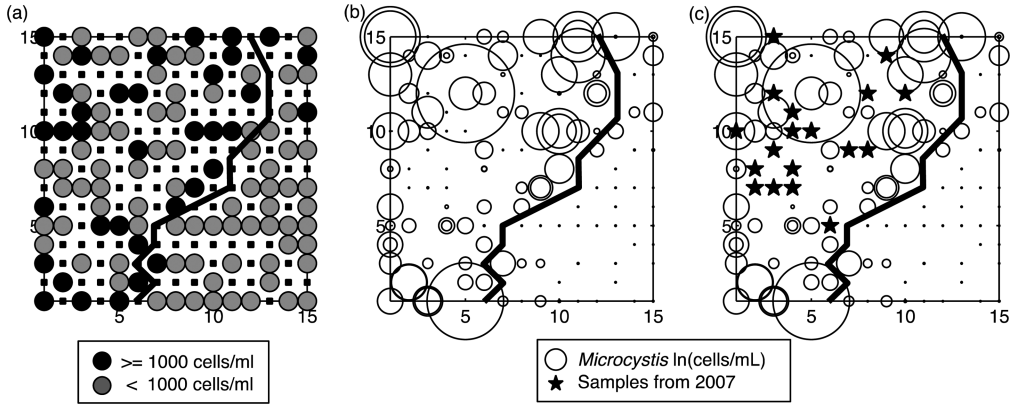


Figure 6. SOM results using the long-term monitoring data. (a) Final SOM showing data clustered into 2 groups, with the groups divided by the black line. The self-organized samples are superimposed by the *Microcystis* cell count. It is important to note that *Microcystis* cell counts were included as input parameters. (b) This is the same map as panel a; however, the open circles mark the final location of each sample on the map and are sized proportionally by the count of *Microcystis* cells/mL associated with the sample. Nearly all of the samples with more than 1000 *Microcystis* cells/mL group together. (c) Similar to panel b, with black stars indicating samples with <1000 *Microcystis* cells/mL collected in 2007. Forty-four samples with less than 1000 *Microcystis* cells/mL were misclassified by the SOM, but 19 of these misclassified samples were collected in 2007, a year that saw no widespread blooms in Missisquoi Bay.

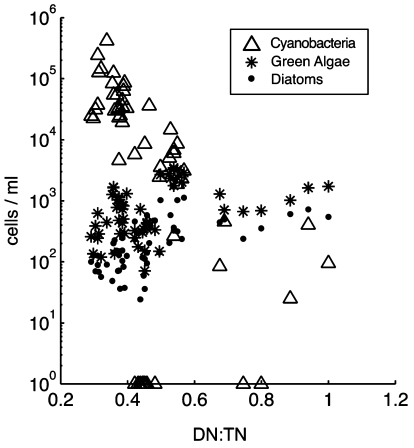


Figure 7. Phytoplankton abundance as a function of the ratio of dissolved nitrogen to total nitrogen (DN:TN). Cyanobacteria (as the sum of *Microcystis*, *Anabaena*, and *Aphanizominon*) are relatively more abundant at low DN:TN.

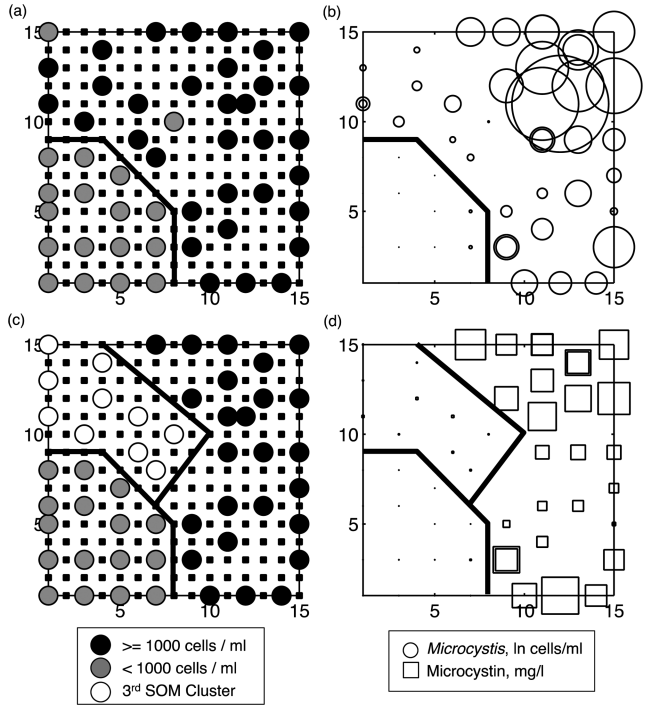


Figure 8. SOM output map using input variables weighted by canonical coefficients generated by discriminant analysis imposing 1000 cells/mL as the threshold between bloom and no-bloom groupings. Neither *Microcystis* cell counts nor toxin concentrations were included as input data. (a) SOM output with 2 clusters; the black line indicates the boundary between the discretely clustered samples, with the circles colored according to the count of *Microcystis* cells/mL. (b) Same map as panel a with each sample now sized proportionally by the *Microcystis* cell count. (c) SOM output map with 3 discrete clusters delineated in black; samples in the third SOM generated cluster are marked with white dots. (d) Squares show the final location of each self-organized sample and are sized proportionally to the measured microcystin sample concentration.

to the discrimination of the bloom and no bloom groupings (Supplementary Figure S2).

SOM clustering results using the canonical coefficient scores to weight the 10 input variables of Table 1 are visualized on a 15 × 15 node map (Figure 8a and b). Solid black and gray circles (Figure 8a) indicate bloom and no-bloom clusters, respectively. Open circles are sized in proportion to the number of *Microcystis* cells (Figure 8b).

The results of using the SOM to cluster data into 3 groups are provided in Figure 8c. The squares (Figure 8d) are sized proportionally to the measured concentration of toxic microcystins. The three clusters are characterized by samples without a bloom, samples with cell counts that indicate a bloom yet have no detectable measured microcystins (i.e., < 0.1 μg/L), and samples with a bloom and measurable microcystins. Neither *Microcystis* cell counts nor toxin concentrations were included as input data.

DISCUSSION

The DN:SRP ratio and the extent of sediment anoxia has a strong influence on sample groupings with and without a cyanobacteria bloom. The N:P ratio has been suggested previously as important in determining conditions that favor cyanobacteria since many cyanobacteria are nitrogen fixers or possess other mechanisms for nitrogen competition.[9] *Microcystis* species, in particular, do

dx.doi.org/10.1021/es403490g | *Environ. Sci. Technol.* 2013, 47, 14267−14274

not fix atmospheric nitrogen but are capable of adjusting their buoyancy to allow for vertical migration within the water column.[25−27,31] In this shallow Missisquoi Bay, sampling does show changes in vertical distribution of *Microcystis* over time,[22] and it is possible that *Microcystis* avoids N limitation by descending to the sediment−water interface to take up ammonium diffusing from anoxic sediments.[32] Of the other dominant cyanobacteria genera observed in Missisquoi Bay, both *Anabaena* and *Aphanizomenon* can fix nitrogen. Figure 7 shows that cyanobacteria in Missisquoi Bay are more successful than green algae and diatoms when DN is a smaller component of TN.

Preliminary analysis indicated that cold-water temperatures (in early May and late October) dominated the clustering, masking important information about bloom dynamics. Since *Microcystis* cells are active primarily above 20 °C,[28] eliminating cold samples allowed patterns specific to cyanobacteria bloom dynamics to be extracted, especially those associated with the intensity and duration of a bloom. Other SOM applications used to investigate cyanobacteria[19,33,34] have also produced strong seasonal trends in data groupings.

Although not necessary for SOM clustering, the ability to weight input variables allows users to fine-tune their importance, when the latter is known. Initially, we did not weight the SOM inputs; however, subsequent weighting using expert opinion significantly refined the groupings. We also weighted the input variables using canonical coefficients from a DA to provide a more systematic approach when expert opinion is not available; however, the DA requires *a priori* distinction of the classes. Initially, our experts advised weighting DN:SRP and the extent of sediment anoxia twice as much as the other variables. The canonical coefficients verified these expert recommendations (Supplementary Figure S2), and as a result, DN:SRP and Mn^{II} are weighted 2−3 times more than other input variables in the subsequent SOM clustering. Weighting the SOM input variables increased our ability to discriminate samples with and without a bloom relative to clustering methods that do not weight the input variables.

The sampling design and schedule (multiple depths at multiple times per day) were intended to capture the diel cycles of *Microcystis* migrating vertically in the water column at multiple points throughout the season, and there is likely multiscale temporal and spatial autocorrelation within the data set, as well correlation between parameters. This inherent autocorrelation within and correlation among the input variables makes the non-parametric SOM an ideal clustering method.

Two biologically meaningfully clusters (i.e., samples with and without a cyanobacteria bloom) were identified by the SOM. Prior work suggests cyanobacteria blooms result from in-lake conditions necessary to sustain a bloom and an environmental event(s) that can initiate a bloom.[35] So, although samples in our "bloom" clusters indicate the conditions (of those measured) necessary to sustain a cyanobacteria bloom, they provide little information about the conditions sufficient to initiate the bloom. The SOM applied to the LTM data creates two groupings: one contains samples without a cyanobacteria bloom, and the other isolates nearly all samples with high counts of *Microcystis* cells. However, this later group also contains many samples from 2007, the season without dominance by cyanobacteria. Since data clustering was based on only a small number of in-lake conditions, we cannot speculate about the combination of events required to initiate a bloom, but the conditions to support a bloom may have existed.

The SOM component planes allow us to examine the magnitude of each variable across the output map. The expert weighted SOM component planes (Figures 4 and 5) show that low DN:SRP ratios correspond with high *Microcystis* cell counts. Also, we have observed a fairly regular seasonal pattern of dominance by different phytoplankton taxa in Missisquoi Bay, with green algae and diatoms dominating the early portion of the warm season and cyanobacteria dominating later in the season. The SOM component planes (Figures 4 and 5) suggest that samples with lower cyanobacteria cell counts clustered with the bloom may represent one snapshot in time of the community succession transition. In this same region of the map, the component planes suggest higher cell counts of green algae and diatoms.

Microcystis are less dominant in summer blooms than 8−10 years ago, providing evidence that the algal community may be shifting due to larger processes such as global climate change or invasive species introduction. Future monitoring should include metrics that more fully capture climate and the entire biological community, including zooplankton, at higher temporal resolution. To understand bloom initiation, meteorological data collection along the shore of Missisquoi Bay and nutrient and sediment loading data associated with river discharge to the Bay will be important. Our own observations suggest that extremely high spring stream flows, the success of cyanobacteria overwintering, intensity and frequency of summer storms, wind and lake circulation patterns, as well as the interactions between phytoplankton and zooplankton grazers affect bloom initiation and duration.[36] These observations are consistent with recent papers highlighting the complexity of bloom dynamics, particularly in the face of global climate change and extreme weather events.[37−40]

This data set combines sediment redox chemistry and water column chemistry. The water column in Missisquoi Bay is typically, but not always, well mixed. The effects of sediment anoxia and its control on the flux of biologically available forms of N and P from the sediment may be felt throughout the water column in large part because of the vertical mobility of *Microcystis*. Although seasonal thermal stratification does not develop, we intermittently observe anoxic layers a few centimeters above the sediment−water interface.[22] This anoxic layer could be a valuable source of reduced N for *Microcystis* when migrating vertically or for the rest of the phytoplankton if the water column is suddenly mixed. Since sediment anoxia data were not available for the LTM data set, the day of year acted as a surrogate for anoxia intensity. This data-driven approach demonstrated the transferability of the methodology; keeping in mind that specific results from statistical methods cannot be applied directly to other locations.

The modified SOM identified patterns of bloom dynamics when used in tandem with expert knowledge of local cyanobacteria behavior and a unique Lake Champlain data set that paired water quality data with precise sediment redox data. Clusters of lake samples with input variables weighted by expert opinion and regularized DA based only on non-cyanobacteria phytoplankton cell counts, sediment, and water chemistry were compared to measured *Microcystis* cell counts. The SOM successfully mined patterns from this highly dimensional data producing groupings in concert with expert feedback. The SOM identified a significant division of samples with and without a cyanobacteria bloom, corroborating a previously suggested hypothesis that cyanobacteria growth is sustained by a set of conditions but requires additional environmental events to

initiate the bloom. In Missisquoi Bay, low DN:SRP and anoxic sediment contribute to the set of bloom-sustaining conditions. Additional environmental drivers are likely required to initiate the rapid growth of cyanobacteria and cyanotoxin production, but we currently lack data at the spatial and temporal frequency necessary to identify these.

Environmental data are often correlated in both space and time and contain multiple correlated variables, rendering traditional parametric statistical analyses inappropriate. Complex systems tools such as the non-parametric clustering algorithm used here are necessary to mine patterns, determine thresholds, and understand non-linear relationships. The SOM, in particular, is useful for organizing data and interpreting the effects of individual input parameters. In addition, the modified SOM allows relative weighting of the input parameters, which likely vary in their effects on cyanobacteria growth. Discerning the input parameters and their weights is improved by iterative collaboration between science experts and complex systems modelers. This research cultivated a positive feedback loop, in which the original expert-derived hypotheses not only necessitated modifications and development of new computational algorithms but also led to more efficient and better-informed hypotheses, which in turn suggested data gaps as well as the need for increased spatial and temporal sampling frequency.

■ ASSOCIATED CONTENT

Supporting Information

Site description, sample collection analysis, and computational details; table of dates, times and depths of sample collection from the sampling platform; self-organizing map network architecture, and canonical coefficients from a regularized discriminant analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: andrea.pearce@uvm.edu.

Present Addresses

[‖]College of Natural Resources, North Carolina State University, Raleigh, North Carolina 27695

[⊥]Department of Earth Science, Indiana University−Purdue University, Indianapolis, Indiana 46202

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Support was provided by Vermont EPSCoR with funds from the National Science Foundation Grant EPS-0701410 and EPS-1101317. Data were collected with support from NOAA grants NA060AR4170231 and NA08OAR4170921. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies. Thanks to L. Stevens for statistical discussion and three anonymous reviewers for their thoughtful review of the manuscript. The authors acknowledge S. Fuller, L. Bronson, Captain R. Furbush, M. Linder, L. Lee, E. Matys, the VT EPSCoR CSYS group, K. Hallock, and A. Pechenik.

■ REFERENCES

(1) Schindler, D. Evolution of phosphorus limitation in lakes. *Science* **1977**, *195*, 260−262.

(2) Xu, H.; Paerl, H. W.; Qin, B.; Zhu, G.; Gao, G. Nitrogen and phosphorus inputs control phytoplankton growth in eutrophic Lake Taihu, China. *Limnol. Oceanogr.* **2010**, *55* (1), 420−432.

(3) Briand, E.; Yepremian, C.; Humbert, J.-F.; Quiblier, C. Competition between microcystin-and non-microcystin-producing *Planktothrix agardhii* (cyanobacteria) strains under different environmental conditions. *Environ. Microbiol.* **2008**, *10*, 3337−3348.

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

(5) World Health Organization *Toxic Cyanobacteria in Water: A Guide to Their Public Health Consequences, Monitoring and Management*; Chorus, I., Bartram, J., Eds.; E & FN Spon: London, 1999.

(6) Watzin, M. C.; Fuller, S.; Bronson, L.; Gorney, R.; Schuster, L. *Monitoring and Evaluation of Cyanobacteria in Lake Champlain*; Lake Champlain Basin Program and Vermont Agency of Natural Resources: Grand Isle, VT, 2010; Lake Champlain Basin Program Technical Report No. 61, p 24.

(7) Boyer, G., et al. The occurrence of cyanobacterial toxins in Lake Champlain. In *Lake Champlain: Partnerships and Research in the New Millennium*; Manley, T., Manley, P., Mihuc, T., Eds.; Kluwer Academic: New York, 2004.

(8) Lake Champlain Basin Program (LCBP). Lake Champlain Basin Atlas. http://www.lcbp.org/Atlas/HTML/intro.htm (accessed 17 September 2010).

(9) Schindler, D.; Hecky, R.; Findlay, D.; Stainton, M.; Parker, B.; Paterson, M.; Beaty, K.; Lyng, M.; Kasian, S. Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105* (32), 11254.

(10) Levine, S.; Schindler, D. Influence of nitrogen to phosphorus supply ratios and physicochemical conditions on cyanobacteria and phytoplankton species composition in the Experimental Lakes Area, Canada. *Can. J. Fish. Aquat. Sci.* **1999**, *56* (3), 451−466.

(11) Smith, V. Low nitrogen to phosphorus ratios favor dominance by blue-green algae in lake phytopankton. *Science* **1983**, *221*, 669−671.

(12) Downing, J.; Watson, S.; McCauley, E. Predicting cyanobacteria dominance in lakes. *Can. J. Fish. Aquat. Sci.* **2001**, *58* (10), 1905−1908.

(13) Bostrom, B.; Anderson, J. M.; Fleischer, S.; Jansson, M. Exchange of phosphorus across the sediment-water interface. *Hydrobiologia* **1988**, *170* (1), 229−244.

(14) Rozan, T. F.; Taillefert, M.; Trowborst, R. E.; Glazier, B. T.; Ma, S. F.; Herszage, J.; Valdes, L. M.; Price, K. S.; Luther, G. W. Iron-sulfur-phosphorus cycling in the sediments of a shallow coastal bay: Implications for sediment nutrient release and benthic macroalgal blooms. *Limnol. Oceanogr.* **2002**, *47* (5), 1346−1354.

(15) Bowden, G.; Dandy, G.; Maier, H. Forecasting cyanobacteria (blue-green algae) using artificial neural networks. In *Artificial Neural Networks in Water Supply Engineering*; Lingireddy, S., Brian, G. M., Eds.; American Society of Civil Engineers: Reston, VA, 2005.

(16) Schryver, J.; Brandt, C.; Pfiffner, S.; Palumbo, A.; Peacock, A.; White, D.; McKinley, J.; Long, P. Application of nonlinear analysis methods for identifying relationships between microbial community structure and groundwater geochemistry. *Microb. Ecol.* **2006**, *51* (2), 177−188.

(17) Mangiameli, P.; Chen, S.; West, D. A comparison of SOM neural network and hierarchical clustering methods. *Eur. J. Oper. Res.* **1996**, *93* (2), 402−417.

(18) Jain, A. K.; Murty, M. N.; Flynn, P. J. Data clustering: a review. *ACM Comput. Surv. (CSUR)* **1999**, *31* (3), 264−323.

(19) Recknagel, F.; Cao, H.; Kim, B.; Takamura, N.; Welk, A. Unravelling and forecasting algal population dynamics in two lakes different in morphometry and eutrophication by neural and evolutionary computation. *Ecol. Inf.* **2006**, *1* (2), 133−151.

(20) Solidoro, C.; Bandelj, V.; Barbieri, P.; Cossarini, G.; Umani, S. Understanding dynamic of biogeochemical properties in the northern Adriatic Sea by using self-organizing maps and k-means clustering. *J. Geophys. Res.: Oceans* **2007**, *112* (C7), C07S90.

(21) Mimeault, M.; Manley, T. O., Missisquoi Bay - an international partnership toward restoration. In *Lake Champlain: Partnerships and research in the new millennium*; Manley, T. O., Manley, P. L., Mihuc, T. B., Eds.; Kluwer Academic/Plenum Publishers: New York, NY, 2004.
(22) Smith, L.; Watzin, M.; Druschel, G. Relating sediment nutrient mobility to seasonal and diel redox fluctuations at the sediment-water interface in a eutrophic freshwater lake. *Limnol. Oceanogr.* **2011**, *56* (6), 2251−2264.
(23) Kohonen, T. The self-organizing map. *Proc. IEEE* **1990**, *78* (9), 1464−1480.
(24) Pearce, A. R.; Rizzo, D. M.; Mouser, P. J. Subsurface characterization of groundwater contaminated by landfill leachate using microbial community profile data and a non-parametric decision-making process. *Water Resour. Res.* **2011**, *47* (6), W06511.
(25) Visser, P.; Passarge, J.; Mur, L. Modeling vertical migration of the cyanobacterium *Microcystis*. *Hydrobiologia* **1997**, *349* (99), 109.
(26) Hyenstrand, P.; Petterson, A. Factors determining cyanobacteria success in aquatic systems - a literature review. *Arch. Hydrobiol. Spec. Iss. Adv. Limnol.* **1998**, *51*, 41−62.
(27) Hunter, P. D.; Tyler, A. N.; Willby, N. J.; Gilvear, D. J. The spatial dynamics of vertical migration by *Microcystis aeruginosa* in a eutrophic shallow lake: a case study using high spectral resolution time-series airborne remote sensing. *Limnol. Oceanogr.* **2008**, *53* (6), 2391−2406.
(28) Coles, J.; Jones, C. Effect of temperature on photosynthesis-light response and growth of four phytoplankton species isolated from a tidal freshwater river. *J. Phycol.* **2000**, *36*, 7−16.
(29) Robarts, R. D.; Zohary, T. Temperature effect on photosynthetic capacity, respiration, and growth rates of bloom-forming cyanobacteria. *N. Z. J. Mar. Freshwater Res.* **1987**, *21* (3), 391−399.
(30) Goovaerts, P. *Geostatistics for Natural Resources Evaluation*; Oxford University Press: New York, 1997.
(31) van Rijn, J.; Shilo, M. Carbohydrate fluctuations, gas vacuolation, and vertical migration of scum-forming cyanobacteria in fishponds. *Limnol. Oceanogr.* **1985**, *30* (6), 1219−1228.
(32) Crawford, K. The effects of nutrient ratios and forms on the growth of *Microcystis aeruginosa* and *anabaena flos-aquae*. M.S. Thesis, University of Vermont, Burlington, VT, 2008.
(33) Chan, W.; Recknagel, F.; Cao, H.; Park, H. Elucidation and short-term forecasting of microcystin concentrations in Lake Suwa (Japan) by means of artificial neural networks and evolutionary algorithms. *Water Res.* **2007**, *41* (10), 2247−2255.
(34) Oh, H.-M.; Ahn, C.-Y.; Lee, J.-W.; Chon, T.-S.; Choi, K.; Park, Y.-S. Community patterning and identification of predominant factors in algal bloom in Daechung Reservoir (Korea) using artificial neural networks. *Ecol. Modell.* **2007**, *204*, 109−118.
(35) Elser, J. The pathway to noxious cyanobacteria blooms in lakes: the food web as the final turn. *Freshwater Biol.* **1999**, *42* (3), 537−543.
(36) Verspagen, J.; Snelder, E.; Visser, P.; Johnk, K.; Ibelings, B.; Mur, L.; Huisman, J. Benthic-pelagic coupling in the population dynamics of the harmful cyanobacterium *Microcystis*. *Freshwater Biol.* **2005**, *50* (5), 854−867.
(37) Paerl, H. W.; Paul, V. J. Climate change: links to global expansion of harmful cyanobacteria. *Water Res.* **2012**, *46*, 1349−1363.
(38) Carey, C. C.; Ibelings, B. W.; Hoffman, E. P.; Hamilton, D. P.; Brooks, J. D. Eco-physiological adaptations that favour freshwater cyanobacteria in a changing climate. *Water Res.* **2012**, *46*, 1394−1407.
(39) El-Shehawy, R.; Gorokhova, E.; Fernandez-Pinas, F.; delCampo, F. Global warming and hepatotoxin production by cyanobacteria: what can we learn from experiments? *Water Res.* **2012**, *46*, 1420−1429.
(40) Reischwaldt, E. S.; Ghadouani, A. Effects of rainfall patterns on toxic cyanobacterial blooms in a changing climate: Between simplistic scenarios and complex dynamics. *Water Res.* **2012**, *46*, 1372−1393.