

# Using artificial intelligence for CyanoHAB niche modeling: discovery and visualization of *Microcystis*–environmental associations within western Lake Erie

David F. Millie, Gary R. Weckman, Gary L. Fahnenstiel, Hunter J. Carrick, Ehsan Ardjmand, William A. Young II, Michael J. Sayers, and Robert A. Shuchman

Abstract: Cyanobacterial harmful algal blooms (CyanoHABs), mainly composed of the genus *Microcystis*, occur frequently throughout the Laurentian Great Lakes. We used artificial neural networks (ANNs) involving 31 hydrological and meteorological predictors to model total phytoplankton (as chlorophyll *a*) and *Microcystis* biomass from 2009 to 2011 in western Lake Erie. Continuous ANNs provided modeled-measured correspondences (and modeling efficiencies) ranging from 0.87 to 0.97 (0.75 to 0.94) and 0.71 to 0.90 (0.45 to 0.88) for training–cross-validation and test data subsets of chlorophyll *a* concentrations and *Microcystis* biovolumes, respectively. Classification ANNs correctly assigned up to 94% of instances for *Microcystis* presence–absence. The influences of select predictors on phytoplankton and CyanoHAB niches were visualized using biplots and three-dimensional response surfaces. These then were used to generate mathematical expressions for the relationships between modeled CyanoHAB outcomes and the direct and interactive influences of environmental factors. Based on identified conditions (~40 to 50  $\mu$ g total phosphorus (TP)·L<sup>-1</sup>, 22 to 26 °C, and prolonged wind speeds less than ~19 km·h<sup>-1</sup>) underlying the likelihood of occurrence and accumulation of phytoplankton and *Microcystis*, a "target" concentration of 30  $\mu$ g TP·L<sup>-1</sup> appears appropriate for alleviating blooms. ANNs generated robust ecological niche models for *Microcystis*, providing a predictive framework for quantitative visualization of nonlinear CyanoHAB–environmental interactions.

Résumé : Les fleurs d'eau de cyanobactéries nuisibles (CyanoHAB), constituées principalement du genre Microcystis, sont fréquentes dans tous les Grands Lacs laurentiens. Nous avons utilisé des réseaux neuronaux artificiels (« ANNs ») incluant 31 prédicteurs hydrologiques et météorologiques pour modéliser la biomasse de phytoplancton totale (chlorophylle a) et de Microcystis, de 2009 à 2011, dans la partie occidentale du lac Érié. Des ANNs continus ont fourni des correspondances modèle-mesures (et des valeurs d'efficacité de la modélisation) allant de 0,87 à 0,97 (0,75 à 0,94) et de 0,71 à 0,90 (0,45 à 0,88) pour des sous-ensembles de données d'entraînement-de contrevalidation et d'essai des concentrations de la chlorophylle a et des biovolumes de Microcystis, respectivement. Les ANNs de classification ont affecté correctement jusqu'à 94 % des cas d'absence ou de présence de Microcystis. Les influences de prédicteurs sélectionnés sur les niches de phytoplancton et de CyanoHAB ont été visualisées à l'aide de diagrammes de double projection et de surfaces de réponse tridimensionnelles, qui ont ensuite été utilisés pour produire des expressions mathématiques pour les relations entre les CyanoHAB modélisées et les influences directes et interactives de facteurs environnementaux. À la lumière des conditions cernées ( $\sim$ 40 à 50  $\mu$ g phosphore total (PT)·L<sup>-1</sup>, de 22 à 26 °C et des vitesses du vent soutenues inférieures à ~19 km·h<sup>-1</sup>) sous-tendant la probabilité de présence et d'accumulation de phytoplancton et de Microcystis, une concentration « cible » de 30 µg PT-L-1 semble adéquate pour réduire les fleurs d'eau. Les ANNs ont produit des modèles de niche écologique robustes pour Microcystis, fournissant ainsi un cadre prévisionnel pour la visualisation quantitative d'interactions non linéaires entre le milieu ambiant et les CyanoHAB. [Traduit par la Rédaction]

## Introduction

The characterization of biotic responses to environmental controls, along with the dynamics of those responses, fuels ecological modeling on local scales and forms the foundation for regional forecasting efforts. Ecological niche models (inclusive of species distribution models; see Peterson 2006) are empirical tools relating biotic occurrence or abundance to a locale's environmental or geographical characteristics and typically are based upon statistically or theoretically derived response surfaces (Guisan and Thuiller 2005). Response surfaces are utilized by researchers to (*i*) assess

Received 24 December 2013. Accepted 15 July 2014.

D.F. Millie. Michigan Technological University, Michigan Tech Research Institute, Ann Arbor, MI 48105, USA; Michigan Technological University, Great Lakes Research Center, Houghton, MI 49931, USA; Palm Island Enviro-Informatics LLC, Sarasota, FL 34232, USA.

**G.R. Weckman.** Ohio University, Russ College of Engineering and Technology, Department of Industrial and Systems Engineering, Athens, OH 45701, USA; Palm Island Enviro-Informatics LLC, Sarasota, FL 34232, USA.

**G.L. Fahnenstiel.** Michigan Technological University, Michigan Tech Research Institute, Ann Arbor, MI 48105, USA; Michigan Technological University, Great Lakes Research Center, Houghton, MI 49931, USA; University of Michigan, Water Center, Graham Sustainability Institute, Ann Arbor, MI 48104, USA. **H.J. Carrick.** Central Michigan University, Department of Biology and Institute for Great Lakes Research, Mount Pleasant, MI 48859, USA.

E. Ardjmand. Ohio University, Russ College of Engineering and Technology, Department of Industrial and Systems Engineering, Athens, OH 45701, USA. W.A. Young II. Ohio University, College of Business, Management Department, Athens, OH 45701, USA.

M.J. Sayers. Michigan Technological University, Michigan Tech Research Institute, Ann Arbor, MI 48105, USA.

R.A. Shuchman. Michigan Technological University, Michigan Tech Research Institute, Ann Arbor, MI 48105, USA; Michigan Technological University, Great Lakes Research Center, Houghton, MI 49931, USA.

Corresponding author: David F. Millie (e-mail: dfmillie@mtu.edu; enviroinformatics@gmail.com).

predictive capability or uncertainty of a conceptual model for biotic distributions, (*ii*) visualize functional environmental, climatic, and (or) spatial influences upon biota operating on dissimilar scales, and (*iii*) support conservation prioritization and resource management (Austin 2002; Guisan and Thuiller 2005; Peterson 2006). Such an "informatics" approach to data-driven modeling (see Millie et al. 2011, 2013; Michener and Jones 2012) affords the means to collectively mine, model, and visualize biota–environmental associations within aquatic survey data. Yet, predictive success for and knowledge derivation arising from such models are "driven" by technical innovation, contingent upon the suitability and robustness of modeling software (Elith and Leathwick 2009; Thuiller et al. 2009).

Expansive accumulations of the cyanobacterium *Microcystis aeruginosa* have become commonplace within productive waters of the Laurentian Great Lakes and are professed to threaten aquatic resources and public health. Such cyanobacterial harmful algal blooms (CyanoHABs) are fueled by excessive loading of anthropogenic nutrients (particularly phosphorus, P), warm temperatures, and water column stratification (Davis et al. 2009, 2010; Paerl et al. 2011*a*; O'Neil et al. 2012). Although point-source loading has been reduced within the Lakes' nearshore waters (Dolan and Chapra 2012), P remains a primary nonpoint source contaminant (DeMarchi et al. 2010; Bridgeman et al. 2012; Michalak et al. 2013) and is considered the most important nutrient regulating phytoplankton growth (Fahnenstiel et al. 1998; International Joint Commission 2014; c.f. Chaffin et al. 2011, 2013).

Empirical modeling of extrinsic environmental influences upon intrinsic phytoplankton patterns typically has utilized classical multivariate analytics, with their parametric assumptions and (or) limitations (i.e., linearity between predictor-response variables, independence and homoscedasticity of errors, normality of error distributions). Yet, nonlinearity in population biomass-growth and environmental interactions dominate aquatic time series (e.g., Pascual and Ellner 2000; Austin 2002). Great Lakes' waters impacted by CyanoHABs exhibit spatially discontinuous hydrological and ephemeral meteorological variances (Millie et al. 2006a, 2009; De Stasio et al. 2008). As a result, predictable linear associations among quasiperiodic, often chaotic environmental conditions and the Lakes' phytoplankton typically do not occur, and one's use of conventional parametric analytics to model CyanHABenvironmental associations appears ill-posed and formidable in effort (see Downing et al. 2001; Millie et al. 2011).

To this end, artificial neural networks (ANNs) identify and reproduce recurring nonlinear patterns within large databases displaying sizable complexity and variance. These adaptive artificial intelligence agents do not require a known probability distribution of variables and provide impartial prediction, whereby patterns between predictors (e.g., environmental conditions) and responses (e.g., species biomass) are discovered and mapped (Jørgensen et al. 2009; Quetglas et al. 2011). However, unlike parametric analytics that provide users with comfortable degrees of transparency and comprehension, predictor–response relationships within ANNs are encoded within architectures that (superficially) appear incomprehensible in regard to user-friendly knowledge (c.f. Gevrey et al. 2003, 2006).

Proactive assessments for and forecasting of aquatic health and environmental status is the goal of many federal, state, and private partnerships. To incorporate prediction of local CyanoHAB distributions (both realized and potential) into forecast simulations, a modeling technology capable of identifying and reproducing nonlinear associations among interacting environmental conditions and the localized, time-specific patterns of HAB biomass accumulations is required. Here, we (*i*) formulate ANNs to model total phytoplankton and *Microcystis* within western Lake Erie using select hydrological and meteorological variables as predictors, (*ii*) extract, from trained networks, the functional influence of environmental predictors upon modeled responses, (*iii*) evaluate the adequacy of ANNs to derive qualitative or quantitative knowledge concerning environmental influences of importance to CyanoHAB potentials, and (*iv*) discuss the relevancy of results in regard to resource stewardship of the Great Lakes.

## Methods

#### Study site

The western basin of Lake Erie (USA and Canada; Fig. 1) comprises  $\sim$ 13% of the Lake's surface area (3300 of 25 690 km<sup>2</sup>; Hartman 1973) and has a water retention time of  $\sim$  51 days (2.6 years for the entire lake). Major tributaries impacting the basin include the Detroit River (draining the oligotrophic "upper" Great Lakes and contributing  $\sim$ 80% to 90% of total inflows) and the nutrient- and sediment-laden Maumee River (draining an agricultural-industrialurban watershed and adding  $\sim$ 10% of the basin's total inflows). The basin is shallow (mean and maximum depths of 7.4 and 19 m) and resembles a sublittoral zone with respect to hydrologic conditions; waters warm and cool rapidly with water column stratification only occurring during wind-free periods (Chandler 1940; Bolsenga and Herdendorf 1993). Over the last decade, CyanoHABs have become a dominant component of the mid- and late-summer phytoplankton with toxic populations of Microcystis commonplace (Millie et al. 2009; Chaffin et al. 2011, 2013; Michalak et al. 2013).

#### Data acquisition

As a component of the Center of Excellence for Great Lakes and Human Health initiative by the National Oceanic and Atmospheric Administration, surveillance monitoring was conducted throughout western Lake Erie during 2009 to 2011. Master sampling stations were established within vicinity of the confluence of the Maumee River and the lake proper (Fig. 1), with whole-water sampling (~1 m depth) conducted at these sites on a biweekly basis from July through September (2009) and June through October (2010 and 2011). Owing to the expansive distribution of the 2011 bloom event (see Stumpf et al. 2012; Michalak et al. 2013), supplemental sampling was extended throughout the entire basin and into the western region of the Lake's central basin during late summer of the year (Fig. 1). Water for abiotic and biotic characterization was collected using clean Niskin bottles (Millie et al. 2009).

Select abiotic variables derived from whole-water collections included the following (as unit; abbreviation): water temperature (°C; TEMP); water clarity, as Secchi depth ( $K_D$ ; Secchi); total, totaldissolved, and soluble-reactive phosphorus (µg·L<sup>-1</sup>; TP, TDP, and SRP, respectively); nitrate- and ammonia-nitrogen (mg·L<sup>-1</sup>; NO<sub>3</sub>-N and  $\mu g \cdot L^{-1}$ ; NH<sub>4</sub>-N); soluble silica (mg·L<sup>-1</sup>; SolSi); and chloride (mg·L<sup>-1</sup>; CL). Analytical procedures for abiotic hydrological variables were presented in Millie et al. (2006a, 2009). Meteorological variables, including wind direction (compass direction; WndDir), wind speed (m·s<sup>-1</sup>; WndSpd), ambient temperature (°C; ATEMP), and total daily irradiance (watts (W)·m-2; TotIR) were acquired from NOAA's National Buoy Data Center (Station THLO1, Toledo Light No. 2; http://www.ndbc.noaa.gov) and the National Solar Radiation database (http://rredc.nrel.gov/solar/old\_dasta/nsrdb). Because surface and near-surface phytoplankton accumulations reflect contemporary and preceeding conditions resulting from short-term meteorological events, a "lagged" time series (as -1- to -10-day running maxima, totals, and means of each variable; hereafter, variable<sub>Max-#Days</sub>, variable<sub>Tot-#Days</sub>, and variable<sub>Ave-#Days</sub>, respectively) was added to the dataset. From these contemporary hydrological and contemporary and time-lagged meteorological predictors, ~13 600 data cells were generated for analysis and modeling.

Biotic variables derived from whole-water collections included chlorophyll *a* ( $\mu$ g·L<sup>-1</sup>, CHL *a*) — a proxy for total phytoplankton biomass, and total *Microcystis* biovolume ( $\mu$ m<sup>3</sup>·L<sup>-1</sup>). Concentrations of CHL *a* were determined via in vitro fluorometry, subsequent



**Fig. 1.** Sampling stations throughout the western basin. Inset: Study area relative to the Laurentian Great Lakes and Lake Erie. Symbols for master stations are labeled ( $WE_1$  to  $WE_9$ , solid squares), whereas supplemental stations only are delineated as open circles.

to photopigment extraction with DCMU (3-(3,4-dichlorophenyl)-1, 1-dimethylurea; after Millie et al. 2010). *Microcystis* was enumerated microscopically via Utermöhl's (1958) sedimentation technique, with cell densities converted to biovolumes by means of geometric figures best approximating the shape of individual taxa. In western Lake Erie, the type species of *M. aeruginosa* can account for up to ~95% of the total *Microcystis* abundance; accordingly, the type species was combined with its morphospecies (i.e., *Microcystis novacekii, Microcystis wesenbergii*, and *Microcystis botrys*) to produce a unified species complex (Otsuka et al. 2001; Millie et al. 2009).

### Data characterization and modeling

Data distributions and relationships were characterized using basic statistics and pairwise scatterplots. A principal component analysis (PRIMER version 6.1 software; Primer-E Ltd., Plymouth, UK) utilized Euclidean distances to order master sampling dates and sites with respect to environmental variables. The uncertainties of hydrological and meteorological predictors for model development were assessed via a coefficient of variation (Håkanson 2000). Annual and monthly differences among CHL *a* concentrations and *Microcystis* biovolumes were assessed using a nonparametric Kruskal–Wallis analysis of variance (ANOVA) on ranks (see Results), with pairwise comparison of means completed via a Dunn's test (SigmaPlot software; Systat Software Inc., Chicago, Illinois, USA). Owing to the large data range, biovolumes were log-transformed prior to analyses to improve computational analytics and enhance data visualization.

With ANNs incorporating supervised learning, CHL a concentrations were modeled from environmental predictors as a con-

tinuous problem, whereas *Microcystis* biovolumes were modeled as classification and continuous problems. Multilayer perceptrons (MLPs) utilizing a backpropagation learning algorithm were originated (NeuroSolutions version 6.0 software; NeuroDimension, Inc., Gainesville, Florida USA), as follows:

$$\begin{aligned} \text{[CHL a] or biovolumes} &= f\{W_{P_1,P_3}[f(W_{X_1,P_1} \cdot X_1 + W_{X_2,P_1} \cdot X_2 \dots W_{X_i,P_1} \cdot X_i + \varepsilon_1)]\} \\ &+ f\{W_{P_2,P_3}[f(W_{X_1,P_2} \cdot X_1 + W_{X_2,P_2} \cdot X_2 \dots W_{X_i,P_2} \cdot X_i + \varepsilon_2)]\} \\ &+ f\{W_{P_1,P_3}[f(W_{X_1,P_1} \cdot X_1 + W_{X_2,P_i} \cdot X_2 \dots W_{X_i,P_i} \cdot X_i + \varepsilon_j)]\} \end{aligned}$$

where  $X_{1,2,...,i}$  are predictor variables,  $P_{1,2,3,...,j}$  are processing elements,  $W_{X_{1,2,...,i}}$  are scalar weights, and  $\varepsilon_{1,2,...,j}$  is the error (after Principe et al. 2000). Topologies were optimized for the number of processing elements within hidden layers and the types of transfer functions (e.g., sigmoid, hyperbolic tangent) and learning rules (e.g., conjugate gradient, momentum; see Millie et al. 2012, 2013). Data vectors were assigned randomly to subsets for network training (to "fit" the data), cross-validation (to provide unbiased estimation of prediction), and testing (to assess performance) of 60%, 15%, and 25% of data, respectively. During training, learning and momentum rates and step sizes were allowed to vary, thereby accelerating learning and ensuring convergence to a global minimum. For modeling continuous concentrations or biovolumes, performance statistics included correspondence ( $\rho$ ), root mean square error (RMSE), and modeling efficiency (ME). For categorical networks, performance metrics included receiver operating characteristics (i.e., accuracy in case ordering, true positive or negative

rates, and false positive or negative rates in classification for the confusion matrices; Brown and Davis 2006).

## Assessing and visualizing predictor-response relationships

The architecture, weights, biases, and transfer-threshold functions for trained networks were incorporated into a data spreadsheet. Connected weights and global sensitivity analyses identified hydrological and meteorological variables having the greatest influence upon network prediction. Connected weights analysis incorporated final synaptic weight values to identify the relative importance of input variables (Olden and Jackson 2002). For global sensitivity (c.f. local sensitivity analysis; Sun et al. 2012), a state-based approach was utilized. Briefly, input data vectors associated with the values of a predictor variable were assigned to a bin (state) along a presumed normal distribution where every state was equivalent to a predetermined variance (0.2 to 0.5 sigma,  $\sigma$ ) of that distribution. Quantitative uncertainties for predictors upon the modeled output were measured as a function of dithered alterations for each variable (up to ±6 standard deviations from the mean) and across its data range.

To extend model prediction-uncertainty to more than a "one input - one output" relationship (like that provided by connected weights and sensitivity analyses), variables reported to interactively influence Microcystis occurrence and (or) accumulation (e.g., nutrients, TEMP, WndSpd; see Davis et al. 2009, 2010; Paerl et al. 2011a; Paerl and Otten 2013; O'Neil et al. 2012) were chosen from candidate predictors for detailed analysis and model visualization. Based on sensitivity analysis, select (pairs of) predictors were varied across their data distributions, with three-dimensional (3D) response surfaces for CHL a concentrations and Microcystis biovolumes generated via reproduced network computations (after Millie et al. 2012, 2013). Two-dimensional (2D) response plots for select predictors and concentrations-biovolumes were derived (via averaging across contrasting variables within 3D surfaces), with predictor valuations relating to half-maximal concentrationsbiovolumes (akin to  $EC_{50}/IC_{50}$  metrics) calculated via four-parameter logistic equations; SigmaPlot software). Simplistic curvilinear equations were fit to the response surfaces-biplots (TableCurve 2D and 3D softwares; Systat, Inc., Chicago, Illinois, USA) to algorithmically characterize the relationship(s) between the modeled responses and singular and interacting predictors.

## Results

With respect to abiotic variables, spatial and temporal gradients of master sampling sites were evident within the principal components ordination. Values of Secchi and nutrient concentrations (TP, TDP, NO<sub>3</sub>-N) were lesser and greater, respectively, at sites closest to the confluence Maumee River and the Lake (Fig. 2A) than at sites offshore, predictably reflecting the decreasing impacts of tributary-derived nutrient and sediment "loads" with increased distance from the River mouth. Temperature (TEMP, ATEMP) and wind (WndSpd, WndDir) were variables most responsible for differentiating temporal groups (Fig. 2B). Although differences among inter-annual station groups were not readily apparent, annual means for TEMP and TP concentrations were greater for 2010–2011 ( $\sim$ 24 °C and 56  $\mu$ g·L<sup>-1</sup>) than for 2009 (22 °C and 42  $\mu$ g·L<sup>-1</sup>;  $p \le 0.001$  and  $p \le 0.05$ , respectively). Intra-annually, mean TEMPs were greatest for June, July, and August (24 to 26 °C;  $p \leq 0.001$ ), with maximal mean TP concentrations occurring in June ( $\sim$ 68  $\mu$ g·L<sup>-1</sup>) and to a lesser degree, August and September (53 to 59  $\mu$ g·L<sup>-1</sup>;  $p \le 0.001$ ). However, the first and second principal components explained only 51.3% of the total variance, whereas the cumulative variance explained by successively adding the third, fourth, and fifth component axes was 66.5%, 75.7%, and 81.6%, respectively. The lesser and consistent coefficients of variation for meteorological (compared with hydrological) variables and to a lesser degree, Secchi and CL, identified these predictors to **Fig. 2.** (A and B) Two-dimensional principal components (PCs) ordination of master stations based on water column physical-chemical and meteorological variables. Percentages of total variability explained by physical or chemical variable for each principal component are indicated. Stations are denoted as a function of (A) distance from Maumee River mouth (see Fig. 1 for station locations and label identifications) and (B) sampling month. (C) Box-and-whisker plots for coefficients of variation (standard deviation of the mean) derived from monthly means of hydrological, meteorological, and biotic variables. For each box, the dotted line indicates the mean value, boundaries closest or farthest to the solid (median) line signify the respective 25th and 75th data percentiles, the whiskers (error bars) denote the 10th and 90th data percentiles, and the dots represent the 5th and 95th percentiles of data outliers. See Methods for variable abbreviations.



the have the least dispersion across their data ranges and as such were the most conservative predictors (Fig. 2C). In contrast, the nutrients TDP, SRP,  $NH_4$ -N, and  $NO_3$ -N displayed the greatest dispersion among candidate hydrological predictors.

CHL a concentrations and Microcystis biovolumes displayed asymmetrical, non-normal distributions (as determined by a Shapiro-Wilks test, p < 0.001 for both variables, and skewness values of 2.17 and -1.47 for concentrations and biovolumes, respectively), thereby necessitating nonparametric characterization. Concentrations of CHL a were moderately dispersed across their data range (from 0.35 to 194 µg·L<sup>-1</sup>; Fig. 2C). Mean annual concentrations (~31 to 35  $\mu$ g·L<sup>-1</sup>) at all stations in 2010 and 2011 were approximately twofold greater than concentrations in 2009 (p < 0.001; Fig. 3A). The greatest mean concentrations occurred in August and September (~37  $\mu$ g·L<sup>-1</sup>), with the least mean concentration in October (Fig. 3B; p < 0.001). The distribution for Microcystis biovolumes was zero-inflated (~20% of samples having no Microcystis present) and displayed the greatest dispersion of any variable across its data range (from 0 to  $1.17 \times 10^{12} \mu m^3 \cdot L^{-1}$ ; Figs. 2C, 3B). Annual mean biovolumes were greater in 2011 (~10<sup>7.55</sup> µm<sup>3</sup>·L<sup>-1</sup> at master and surveillance stations) than those in 2009 and 2010 (Fig. 3A;  $p \le 0.05$ ). The greatest monthly mean biovolumes occurred in August and September ( ${\sim}10^{8.41}$  and  $10^{7.37}~\mu m^3 {\cdot} L^{-1},$  respectively), with the least mean biovolumes in June and October (Fig. 3B;  $p \le 0.001$ ).

#### Continuous CHL a modeling

MLPs, utilizing distinct supervised learning topologies, distinct transfer functions, and learning algorithms, were surveyed as models for CHL a concentrations. The best-performing network (31 candidate inputs, 15, 14, and 10 processing elements (PEs) within hidden layers (HLs) one, two, and three with hyperbolic tangent transfer functions and implementing a momentum learning algorithm) produced an enormously complex topology with 65 100 synaptic influences upon the modeled response. This MLP, developed with training and cross-validation data subsets prior to its application to test data, provided an adequate estimate of prediction (Figs. 4A, 4B), with the best holistic performance metrics across all evaluated models. The network performed better in estimating concentrations within the training and cross-validation data subsets (modeled-measured  $\rho$  and MEs ranging from 0.89 to 0.97 and 0.78 to 0.94, respectively) than those of the test data (values for modeled-measured values for  $\rho$  and ME of 0.87 and 0.74, respectively; see Discussion). Phosphorus, both TP and TDP, was identified via connected weights analysis (Fig. 4C) to have a consistent principal impact upon modeled concentrations, with TP confirmed via sensitivity analysis to have the greatest predictive influence (Fig. 4D).

The 3D surfaces and 2D plots arising from the optimal model depicted the interacting influences and singular significances, respectively, of select predictors (TP, TEMP, WndSpdAve-3, and NO<sub>3</sub>-N; Figs. 5A, 5B) upon CHL a concentrations; variables displayed nonlinear relationships with CHL a, with modeled concentrations greatest at maximal TP concentrations and TEMPs and minimal NO3-N concentrations and prolonged WndSpds. From the 2D plots, 52 µg TP·L<sup>-1</sup>, 23 °C, 2.7 mg NO<sub>3</sub>-N·L<sup>-1</sup>, and WndSpd<sub>Ave-3</sub> of 18 km·h<sup>-1</sup> were identified (via logistic equations) as conditions at which the response concentrations were one-half of maximal modeled estimates. Nonetheless, biplots portrayed only the "average" influence of predictor variables upon modeled CHL a. For example, parallel "slices" taken across 3D surfaces identified the relationship between modeled CHL *a* and TP at TEMPs  $\leq \sim 20$  °C to be relatively uniform and only slightly curvilinear, having minimal slope (Fig. 5C). Thereafter, the response slope increased substantially and became progressively nonlinear with increasing TEMPs. Equations fit to the response surfaces and 2D plots quantitatively described modeled CHL a conditional to both predictor interaction and singular variable effects (see online Supplementary Material equations<sup>1</sup>).

**Fig. 3.** Total chlorophyll *a* concentrations and *Microcystis* biovolumes at master and supplemental (-Suppl) stations as a function of (A) sampling year and (B) month. Data are means + standard error; totals that follow show data for chlorophyll *a* and biovolumes, respectively: 2009, n = 127 (for both); 2010, n = 147 and 157; 2011, n = 31 and 33; 2011-Suppl, n = 62 and 74; June, n = 63 and 60; July, n = 101 and 100; August, n = 106 and 103; September, n = 72 and 83; October, n = 35). Group means (designated by letters with no superscript for concentrations and letters with the superscript (') for biovolumes) with distinct characters are different at the 0.05 probability level.



## Continuous Microcystis modeling

An MLP (consisting of 30 720 synaptic weights arising from 32 hydrological–meteorological predictors, 12, 10, and 8 PEs within HLs one, two, and three with hyperbolic tangent transfer functions and implementing a momentum learning algorithm) provided the best modeling performance for continuous *Microcystis* biovolumes. This network performed better for training data than for its application to test data (compare Figs. 6A, 6B). Modeled test data mirrored measured data well at biovolumes greater than  $\sim 10^8 \ \mu m^3 \cdot L^{-1}$ , but generally overestimated measured data at lesser biovoumes. Neither connected weights nor global sensitivity analysis identified any one (or few) predictor(s) to greatly influence prediction of *Microcystis* to the exclusion of other variables (Figs. 6C, 6D).

Assorted 3D surfaces and corresponding 2D plots depicted the interacting influences and singular significances, respectively, of a select predictor upon modeled *Microcystis* biovolumes (e.g., Figs. 7A, 7B). Selected predictors displayed unique, nonlinear relationships with *Microcystis*, with ~50  $\mu$ g TP·L<sup>-1</sup>, 26 °C, 1.5 mg NO<sub>3</sub>·N·L<sup>-1</sup>, and WndSpd<sub>Ave-3</sub> of 23 km·h<sup>-1</sup> identified (via logistic equations) as conditions at which response biovolumes were one-half of maximal modeled estimates. Maximal biovolumes occurred at conditions of

<sup>&#</sup>x27;Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjfas-2013-0654.

**Fig. 4.** Modeled chlorophyll *a* concentrations as a function of measured concentrations in (A) training and cross-validation and (B) test data subsets for the optimal continuous network. The dashed line represents a 1:1 relationship. (C) The relative share of prediction associated with model inputs, as determined via the connected weights analysis. (D) Results of state-based sensitivity analyses across a variation of predictors. The initial 12 predictors having the greatest influence in connected weights–sensitivity analyses are depicted. See Methods for variable abbreviations.



greater than ~100  $\mu$ g TP·L<sup>-1</sup> and 28 °C, and WndSpd<sub>Ave-3</sub> less than ~19 km·h<sup>-1</sup>. Similar to that observed for CHL *a*, 2D plots depicted only the "average" influence of a predictor upon modeled biovolumes; parallel "slices" taken across 3D surface identified the modeled biovolume–TP association at TEMPs greater than 24 °C to approximate those at lesser TEMPs, but at a greater biomass capacity and with a lesser influence of TEMP at elevated TP concentrations (Fig. 7C). Equations fit to the 3D surface and 2D plots quantitatively described modeled biovolumes conditional to predictor interaction and (or) singular variable effects (see online Supplementary Material equations<sup>1</sup>).

## Categorical Microcystis modeling

To balance the number of instances within the zero-inflated and continuous portions of the data distribution and thereby improve the development of classification networks for *Microcystis* presence–absence, a Synthetic Minority Over-sampling Technique (SMOTE; Waikato Environment for Knowledge Analysis (WEKA) software, version 3.7.10; http://www.cs.waikato.ac.nz/~ml/weka/) was implemented. As a resampling tool for imbalanced data in classification models (Chawla et al. 2002), SMOTE generated data

vectors through interpolation between neighboring minority values and selection of new values for all variables between the minority pair(s). Following SMOTE, a classification MLP for Microcystis presence-absence (consisting of 53 760 synaptic weights arising from 32 predictors, 14, 12, and 10 PEs within HLs one, two, and three with sigmoidal transfer functions and implementing a momentum learning algorithm) ordered dualistic outcomes into contingency matrices (Table 1). Networks effectively classified training and cross-validation data, with class accuracies and errors of 93.98 and 6.02% and 89.00 and 11.00%, respectively. Application of the trained and cross-validated networks to test data provided a slightly lesser model "fit" (class accuracy and error of 87.21 and 12.79%, respectively). A greater true presence rate (of 93.20%) than the true absence rate (of 58.82%) for test data likely arose from the paucity of representative data vectors within the holistic database (irrespective of using SMOTE, see Discussion).

Biotic (CHL *a*) and abiotic (WndSpd (both lagged and maximum), TotIR, ATEMP, TP) predictors were identified via sensitivity analysis to have the greatest influences upon modeled classification (Fig. 8A). Assorted 3D surfaces were constructed to depict the

**Fig. 5.** (A) Three-dimensional (3D) response surface for chlorophyll *a* concentrations as a function of total phosphorus (TP) and temperature (TEMP), as derived from the optimal artificial neural network. (B) Two-dimensional (2D) response plots for concentrations as a function of TP, TEMP, mean 3-day wind speed (WndSpd<sub>Ave-3</sub>), and nitrate-nitrogen (NO<sub>3</sub>-N), as averaged across 3D surfaces of interacting variable pairs. Predictor valuations relating to one-half of the maximum, modeled concentrations are indicated. (C) 2D depictions of parallel, vertical "slices" taken across the modeled 3D surface as a function TEMP and TP. The changing shape of the "slices" illustrates the nonlinear interaction among variables.



interacting influences of the (aforementioned) abiotic predictors upon modeled *Microcystis* presence or absence (data not shown). From the contrasting, nonlinear 2D plots arising from averaging the singular variable influences across 3D response surfaces, the likelihood of *Microcystis* occurrence was greater with increasing TP concentrations and ATEMPs and less with increasing WndSpd<sub>Ave-3</sub> (Fig. 8B). Specifically, the regularity in *Microcystis* occurrence was generally assured (i.e., probabilities greater than 0.5) during concomitant conditions of TP and ATEMPs greater than 45 mg·L<sup>-1</sup> and 21 °C, respectively, and WndSpds<sub>Ave-3</sub> less than 20 km·h<sup>-1</sup>. Although the sensitivity analysis identified TotIR to influence modeled classifications, variability in this predictor did not appear to dramatically alter *Microcystis* occurrence (as likelihood probabilities were ~0.6 or greater).

## Discussion

In summarizing the theoretical aspects of species modeling as it relates to phytoplankton blooms, Carrick (2011) affirmed that HAB "niche space" arises from defined, albeit complex, environmental and geospatial conditions and from which the parameterization and prediction of species occurrence–distribution patterns can emerge. In this context, important questions that modelers repeatedly address for dynamic and productive waters include the following: (i) What is (are) the influence(s) of singular or interacting suites of predictor (environmental) variables upon a response (biotic) variable? and (ii) How does one visualize and quantify the predictive uncertainties for this (these) environmental-biotic relationship(s)? Here, an "informatic" approach affording the identification, visualization, and quantification of environmental conditions relating to bloom-forming phytoplankton within western Lake Erie was presented. In characterizing CHL a concentrations and Microcystis biovolumes and presence-absence as a function of select physical-chemical predictors, ANNs proved to be robust ecological niche models: modeled-measured values of o for continuous ANNs ranged from 0.87 to 97 and 0.70 to 0.94 for CHL a and Microcystis biovolumes, respectively, and classification networks correctly assigned up to 94% of instances for Microcystis presence-absence. Arising from the data-assimilative models, 3D response surfaces and 2D plots portrayed the nonlinear uncertainties associated with phytoplankton and Microcystis "niche space" within the distribution boundaries of select hydrological-meteorological predictors. In addition, the chosen pedagogical approach allowed generation of simplistic equations for the singular or interacting relationship(s) between modeled outcomes and select predictor variables.

Local (i.e., hydrological) and regional (i.e., meteorological) factors were interactively associated with modeled phytoplankton **Fig. 6.** Modeled *Microcystis* biovolumes as a function of measured biovolumes in (A) training and cross-validation and (B) test data subsets for the optimal continuous network. The dashed line represents a 1:1 relationship. (C) The relative share of prediction associated with model inputs, as determined via the connected weights analysis. (D) Results of state-based sensitivity analyses across variations of predictors. The initial 12 predictors having the greatest influence in connected weights–sensitivity analyses are depicted. See Methods for variable abbreviations.



and Microcystis biomass (c.f. Paerl et al. 2011a). Classification networks identified "baseline" conditions signifying an occurrence likelihood (i.e., probability ≥0.50) for Microcystis presence within samples. During concurrent instances in which water column TP was  ${\sim}40$  to 45  $\mu g$  TP·L^-1, ATEMP was  ${\sim}21$  °C, and WndSPd\_{Ave-3} was less than ~20 km·h<sup>-1</sup>, Microcystis was certain (probabilistically) to be present. In models for continuous data distributions, the TP concentrations for half-maximal CHL a concentration and Microcystis biovolumes were nearly exact ( $\sim$ 52 µg·L<sup>-1</sup>). Such a relatively large TP concentration may have arisen as a consequence of adaptions by Lake Erie phytoplankton to the nutrient-enriched environment in which they occur (Reynolds 2006). Alternatively, the influence of TEMP upon CHL a and Microcystis accumulations differed; half-maximal biovolumes occurred at TEMPs greater (~24 to 26 °C) that that for CHL *a* concentrations (~22 °C). Cyanobacteria are known to exhibit optimal growth rates at water column temperatures typically in excess of 25 °C (Paerl and Huisman 2008). The lesser TEMP identified for CHL a likely arose from the presence or dominance of influential taxa other than Microcystis and possessing distinct autecological requirements (e.g., chlorophytes and diatoms) within the Lake throughout the year (Nicholls 1997; Smol and Stoermer 2010; Allinger and Reavie 2013).

Predictive uncertainties for CHL a concentrations and Microcystis biovolumes largely were associated with contemporary or timelagged variables indicative of water column nutrient sufficiency (TP-TDP-SRP and NO3-N), and mixing (WndSpd-WndDir) and TEMP, ambient meteorological conditions (TotIR, ATEMP), and (or) proxy measurements of phytoplankton biomass (Secchi) and tributary influences (CL; refer to Figs. 4 and 6). Dependence upon such variables for empirical modeling of near-surface phytoplankton accumulations was not surprising; nutrient enrichment, hydrological stability, increased irradiance, and warm temperature are conditions that collectively favor (localized) Microcystis prevalence within lentic systems affected by tributary inflows (such as western Lake Erie). Nevertheless, distinctions between causality and correlative processes (in regards to phytoplankton-CyanoHAB biomass and interactive physical-chemical associations) are difficult to formulate. The empirical linkages among phytoplankton-CyanoHABs and identified environmental conditions arose from an observational database, with no experimental corroboration. The temporal inconsistencies for CHL a concentrations and Microcystis biovolumes in western Lake Erie were considerable; inter-annual mean concentrations and biovolumes varied up to twofold and 1.25 orders of **Fig. 7.** (A) Three-dimensional (3D) response surface for *Microcystis* biovolumes as a function of total phosphorus (TP) and temperature (TEMP), as derived from the optimal artificial neural network. (B) Two-dimensional (2D) response plots for biovolumes as a function of TP, TEMP, mean 3-day wind speed (WndSpd<sub>Ave-3</sub>), and nitrate-nitrogen (NO<sub>3</sub>-N), as averaged across 3D surfaces of interacting variable pairs. Predictor valuations relating to one-half of the maximum, modeled concentrations are indicated. (C) 2D depictions of parallel, vertical "slices" taken across the modeled 3D surface as a function TEMP and TP. The changing shape of the "slices" illustrates the nonlinear interaction among variables.



**Table 1.** Contingency matrices for training, cross-validation, and test data subsets arising from the optimal categorical (presence–absence) artificial neural network for *Microcystis*.

Data subset (n, overall % correct)	Classification	
	Absence	Presence
Training ( <i>n</i> = 299, 93.98%)		
Absence	130	10
Presence	8	151
% Correct	94.20	93.79
Cross-validation ( <i>n</i> = 100, 89.00%)		
Absent	43	8
Present	3	46
% Correct	93.48	85.19
Test ( <i>n</i> = 86, 87.21%)		
Absent	10	7
Present	4	65
% Correct	71.43	90.28

magnitude, whereas intra-annual mean concentrations varied 2.7fold and 2.3 orders of magnitude, respectively. Meteorological variables were the most conservative predictors, while hydrological predictors (particularly nutrients) displayed considerable dispersion across their data ranges. Almost one-fifth of the total environmental variability could not be accounted for (by the PCA), illustrating the extreme dynamics and heterogeneity in the Lake's hydrological-meteorological conditions. In any event, network models provided an effective assessment of environmentalbiota complexity and interaction and yielded realistic, albeit temporal-specific, portrayals for relationships of P, TEMP-ATEMP, and WndSpd with the Lake's bloom-forming phytoplankton.

Networks effectively modeled continuous concentrationsbiovolumes and presence-absence classification within training data. Predictive performances for test data were slightly less; modeled values generally underestimated concentrations greater than 45 µg CHL a·L<sup>-1</sup> and over- or underestimated biovolumes less than  $10^8 \ \mu m^3 \cdot L^{-1}$ , whereas the correct assignment for instances in which biovolumes were absent declined substantially. These decreases in prediction performances likely arose from inadequate data representation within training subsets, thereby precluding optimal model development (e.g., Millie et al. 2006b). Instances of < $\sim$ 45 µg CHL a·L<sup>-1</sup> and >10<sup>8</sup> µm<sup>3</sup> Microcystis·L<sup>-1</sup> were 4.3- and 1.7-fold greater, respectively, in the database than instances of greater and lesser concentrations or biovolumes, respectively. Such insufficient data representation in a HAB monitoring database, including that for Lake Erie, is not uncommon; because of costs associated with sample acquisition-processing, data routinely is

**Fig. 8.** (A) Results of state-based sensitivity analyses for the optimal, categorical network modeling *Microcystis* presence–absence. The initial 12 predictors having the greatest influence are depicted. Because categorical networks utilized entropy error functions, modeled values represent the probability ( $\geq 0$  to  $\leq 1$ ) for the classification state. See Methods for variable abbreviations. (B) Two-dimensional plots for modeled probabilities as a function of total phosphorus (TP), ambient temperature (ATEMP), mean 3-day wind speed (WndSpd<sub>Ave-3</sub>), and total irradiance (TotIR) as averaged across 3D surfaces of interacting variable pairs (not shown).

acquired during HAB presence or perceived event initiation (but not so much when blooms are absent or environmental conditions perceived unfavorable). For realistic data-driven models, it is equally important to acquire data during periods when Cyano-HAB biomass is minimal or absent (e.g., during early summer or late fall, when TEMPs-ATEMPS and periods of shortened water column stratification are unfavorable to bloom formation), as it is during periods when blooms typically are present.

Nevertheless, biotic-abiotic associations within dynamic, often chaotic aquatic systems (such as western Lake Erie) are intrinsically difficult to understand because associations are nonlinear. distributed nonrandomly, and changing continuously in response to the system state (Quetglas et al. 2011). The complex, highdimensional databases that characteristically arise from multiyear surveillance programs require robust empirical analytics incorporating rigorous, even unconventional, approaches for data manipulation-minimization, identifying variable interactions and trends and the derivation of user-friendly information (Wood 2010; Michener and Jones 2012). Because of their capability for detecting and reproducing recurring patterns within data volumes exhibiting substantial "noise", redundancy, autocorrelation, nonlinearity, and outliers, ANNs afford powerful modeling capabilities for water resources. Yet, ANNs require little expert knowledge for their application and, aside from their "pure" prediction capabilities, typically only provide qualitative information regarding predictor influences (e.g., Figs. 4C, 4D and 6C, 6D). Accordingly, many ecologists perceive ANNs to be numerical enigmas having little relevance to ecological applications. This (perceived) deficiency was overcome, in part, via the presented pedagogical approach; ANN-modeled estimates for CHL a concentrations and Microcystis biovolumes and presence-absence were depicted via multidimensional surfaces and biplots. Discrete mathematical expressions then were generated to quantitatively define physical-chemical associations and interactions with modeled outcomes (see Millie et al. 2012, 2013)

Regression analysis often is used to predict algal biomass based on an assumed correspondence with a single (or multiple) abioticbiotic variable(s) (e.g., Bachmann et al. 2001; Heffernan et al. 2010; Stumpf et al. 2012). To compare network results with that of parametric models, multiple linear regression was conducted to estimate CHL a concentrations and Microcystis biovolumes using identical independent variables as the ANNs. Measured:modeled correspondences for regression models were less than that for comparable ANNs, indicating that networks outperformed linear models (data not shown). This greater performance was anticipated; in theory, an ANN encompasses linear regression and because of a model architecture suited for identifying the nonlinear complexities of a biotic response to environmental forcing, should perform as well or better than regression models (Gonzalez 2000; Millie et al. 2006b, 2012). However, multicollinearity existed among the independent variables, and the predictor-response surfaces and biplots arising from the ANNs depicted interacting, nonlinear predictor influences. Although modeled residuals for CHL a and biovolume regressions were independent (determined by Durbin-Watson statistics), the errors of the predictor coefficients were not normally distributed around the regression estimate (determined by Shapiro-Wilks tests), and response variables displayed non-normal distributions and heteroscedasticity (see Results). Such conditions invalidate the underlying assumptions for linear regression (Reckhow et al. 1990; Osborne and Waters 2002), thereby resulting in parametric models that lacked merit.

#### **Relevancy to Great Lakes stewardship**

To mitigate eutrophication of the Great Lakes, federal and state or provincial agencies within the United States and Canada have made it a priority to evaluate relationships among anthropogenic impairments, environmental variability, and phytoplankton production–abundance (e.g., International Joint Commission 2014). To this end, knowledge of the empirical relationships between CyanoHABs and associated physical–chemical factors are integral to developing threshold criteria for decision-making in environmental resource stewardship (e.g., shifts in ecosystem state, establishment of nutrient "critical load" limits, extrinsic factor identification; Groffman et al. 2006). Importantly, the considerable amount of ecological redundancy displayed within phytoplankton provides an assured confidence in one's usage of data-driven metrics (such as those presented here) for assessments pertaining to anthropogenic stressors and (or) natural disturbances (Carrick 2011).



1652

The increased awareness concerning the timing for and amount of nonpoint source contaminants (e.g., over the last two decades, 90% of the P loads delivered from the Maumee River occurs during storm events; Ohio Environmental Protection Agency 2010) underscores the necessity for knowledge concerning contemporary nutrient dynamics and their impact upon Great Lakes phytoplankton (Great Lakes Commission 2012; Michalak et al. 2013). The range in TP concentrations (~40 to 55  $\mu$ g·L<sup>-1</sup>) underlying the occurrence likelihood for and (or) significant accumulation of Microcystis biovolumes and CHL a concentrations provided a realistic portrayal for nutrient association(s) with Lake Erie CyanoHABs and were similar to those previously identified for mitigating CyanoHABs in northern temperate waters. Millie et al. (2011) noted maximal Microcystis biomass within Saginaw Bay, Lake Huron (1991–1996), concurrent with 40 to 45  $\mu$ g TP·L<sup>-1</sup>, whereas Downing et al. (2001) identified the "risk" of cyanobacterial dominance to abruptly increase between 30 and 70 µg TP·L<sup>-1</sup>. As such, a water column "target" concentration of 30  $\mu g \mbox{ TP} \cdot L^{-1}$  appears appropriate for alleviating contemporary bloom potentials in western Lake Erie. Yet, this concentration is twofold greater than the concentration (of 15 µg TP·L-1) originally established for western Lake Erie in the 1978 USA-Canada Great Lakes Water Quality Agreement (Environment Canada and US Environmental Protection Agency 2003) and upheld in the 2012 agreement (http://www. ec.gc.ca/grandslacs-greatlakes/default.asp?lang=En&n=A1C62826-

1&offset=5&toc=show; last accessed 13 July 2014). Although phytoplankton growth within productive waters of the Great Lakes is presumed to be dependent upon P availability, assemblages episodically can be regulated by contrasting N-limitation arising from nutrient-laden tributary inflows and the ingestion-excretory activities of dreissenid mussels (see Conroy et al. 2005; Millie et al. 2009, 2011). Here, N-containing compounds displayed distinct dynamics with Microcystis biovolumes and TP concentrations (on average across the database); minimal NO<sub>3</sub>-N concentrations occurred during instances of maximal Microcystis biovolumes (see Figs. 7B, 8B), and the likelihood for Microcystis presence slightly increased with increasing NH<sub>4</sub>-N concentrations, similar to that for TP availability (data not shown). Chaffin et al. (2013) noted that Microcystis blooms in the western basin may remain N-replete during instances of minimal NO<sub>3</sub>-N availability (i.e., during mid- to late summer when cyanobacterial biomass is greatest), whereas Paerl et al. (2011b) stated that the non-N<sub>2</sub>-fixing Microcystis effectively competes for reduced forms of N (particularly NH<sub>4</sub>-N), suggesting that uptake of water column regenerated NH<sub>4</sub>-N (arising from lake sediments) might assist or support growth during instances of N-limitation (c.f. Davis et al. 2010). Unfortunately, the 2009–2011 database did not contain the necessary chemical variables for definitive assessments of nutrient limitation (via derivation of total C:N:P metrics and (or) the proportion of reduced-oxidized forms of N to organic forms) in regard to CyanoHAB presence-absence or biovolumes

Empirical delineations of environmental conditions influencing CyanoHABs are restricted to the time-dependent, physicalchemical makeup of the water body in question and the prevalent ecotypes present within the assemblage (Chorus and Mur 1999). It appears most realistic to interpret the aforementioned target concentration as a generalized threshold for Microcystis proliferation within western Lake Erie during a specific period (2009 to 2011), rather than a definitive regulatory target. Expansive Microcystis blooms have been prevalent throughout the basin for over a decade, during which time hydrological regimes, tributary nutrient loadings, and meteorological conditions have changed annually. Although the 2009 to 2011 database displayed extreme ranges for CHL concentrations and Microcystis biovolumes, modeling efforts likely encompassed only a "snapshot" of the Lake's ecological continuum and might not be representative of past or future bloom events of greater magnitude and (or) variability. For example, the CyanoHABs observed during 2010 and 2011 were not the greatest bloom events (in terms of biomass magnitude) to occur within western Lake Erie during the last decade. Millie et al. (2009) reported a mixed, localized bloom composed of the chlorophyte *Pandorina morum* (up to 727 µg chlorophyte Chl *a*·L<sup>-1</sup>) and the cyanobacteria *Microcystis* spp. and *Planktothrix* spp. (up to 100 µg cyanobacterial Chl *a*·L<sup>-1</sup>) within waters immediately offshore the mouth of Maumee River during August 2003 (in the immediate vicinity of stations WE<sub>4</sub> and WE<sub>5</sub> of the present study). Although *Pandorina* is not known to produce toxins (like cells of *Microcystis*; Dyble et al. 2008), the taxon is capable of producing allelopathic chemicals inhibiting plant photosynthetic and mitochondrial electron transport (Patterson et al. 1979) and may be influential in the competitive dynamics of the Lake's phytoplankton, including *Microcystis*.

CyanoHAB forecasting across the Great Lakes requires modeling efforts having adequate fidelity to resolve localized, timespecific bloom occurrences. Yet, the modeling of phytoplankton displaying non-normal, heteroscedastic, and zero-inflated distributions is onerous and its interpretation is obscure. The legitimacy of data-driven analytics is constrained by the amount of system variability omitted from the sample distributions, particularly as it applies towards data "extremes" and the occurrence of ecological "surprises" (Peek et al. 2003; Doak et al. 2008). Simplistic, parametric models may not describe and (or) be appropriate for the mosaic of interactive, nonlinear, and dynamic (even stochastic) associations arising between assemblages and environmental variables. Rather, adaptive nonlinear protocols will be required. ANNs modeled the intrinsic variance and magnitude of CHL a and Microcystis within western Lake Erie, affording a framework from which environmental-CyanoHAB associations could be visualized and quantifiable knowledge derived. Ideally, such data-driven approaches can be used to parameterize conceptual models projecting CyanoHAB mechanisms and causal relationships across extended temporal-spatial coverage (e.g., Zhang et al. 2013).

### Acknowledgements

This work is contribution Nos. 11 and 45 of the Great Lakes Research Center, Michigan Technological University, and the Institute for Great Lakes Research, Central Michigan University, respectively. Funding for this work was provided, in part, by the Michigan Tech Research Institute and by the National Oceanic and Atmospheric Administration's Office of Global Programs and Center for Sponsored Coastal Ocean Research via the Center of Excellence for Great Lakes and Human Health. Reference to proprietary names are necessary to report factually on available data; however, the authors' institutions neither guarantee nor warrant the standard of a product and imply no approval of a product to the exclusion of others that may be suitable.

#### References

- Allinger, L.E., and Reavie, E.D. 2013. The ecological history of Lake Erie as recorded by the phytoplankton community. J. Gt. Lakes Res. 39(3): 365–382. doi:10.1016/j.jglr.2013.06.014.
- Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol. Modell. 157(2–3): 101–118. doi:10.1016/S0304-3800(02)00205-3.
- Bachmann, R.W., Hoyer, M.V., and Canfield, D.E., Jr. 2001. Evaluation of recent limnological changes at Lake Apopka. Hydrobiologia, 448(1–3): 19–26. doi:10. 1023/A:1017564911355.
- Bolsenga, S.J., and Herdendorf, C.E. 1993. Lake Erie and Lake St. Clair handbook. Wayne State University Press, Detroit, Mich.
- Bridgeman, T.B., Chaffin, J.D., Kane, D.D., Conroy, J.D., Panek, S.E., and Armenio, P.M. 2012. From river to lake: phosphorus partitioning and algal community compositional changes in Western Lake Erie. J. Gt. Lakes Res. 38(1): 90–97. doi:10.1016/j.jglr.2011.09.010.
- Brown, C.D., and Davis, H.T. 2006. Receiver operating characteristics curves and related decision measures: a tutorial. Chemom. Intell. Lab. Syst. 80(1): 24–38. doi:10.1016/j.chemolab.2005.05.004.
- Carrick, H.J. 2011. Niche modeling and predictions of algal blooms in aquatic ecosystems. J. Phycol. 47(4): 709–713. doi:10.1111/j.1529-8817.2011.01042.x.

1653

- Chaffin, J.D., Bridgeman, T.B., Heckathorn, S.A., and Mishra, S. 2011. Assessment of *Microcystis* growth rate potential and nutrient status across a trophic gradient in western Lake Erie. J. Gt. Lakes Res. **37**(1): 92–100. doi:10.1016/j.jglr. 2010.11.016.
- Chaffin, J.D., Bridgeman, T.B., and Bade, D.L. 2013. Nitrogen constrains the growth of late summer cyanobacterial blooms in Lake Erie. Adv. Microbiol. 3(6A): 16–26. doi:10.4236/aim.2013.36A003.
- Chandler, D.C. 1940. Limnological studies of western Lake Erie. 1. Plankton and certain physical–chemical data of the Bass Islands region, from September, 1938 to November, 1939. Ohio J. Sci. 40(6): 291–336.
- Chawla, N., Bowyer, K., Hall, L., and Kegelmeyer, W. 2002. SMOTE: synthetic minority oversampling technique. J. Artific. Intell. Res. 16(January–June): 321–357.
- Chorus, I., and Mur, L. 1999. Preventative measures. *In* Toxic cyanobacteria in water: a guide to their public health consequences, monitoring and management. *Edited by* L. Chorus and J. Bartram. E & FN Spon on behalf of the World Health Organization, London, England. pp. 235–274.
  Conroy, J.D., Edwards, W.J., Pontius, R.A., Kane, D.D., Zhang, H., Shea, J.F., Richey, J.N., and Culver, D.A. 2005. Soluble nitrogen and phosphorus excre-
- Conroy, J.D., Edwards, W.J., Pontius, R.A., Kane, D.D., Zhang, H., Shea, J.F., Richey, J.N., and Culver, D.A. 2005. Soluble nitrogen and phosphorus excretion of exotic freshwater mussels (*Dreissena* spp.): potential impacts for nutrient remineralisation in western Lake Erie. Freshw. Biol. 50(7): 1146–1162. doi:10.1111/j.1365-2427.2005.01392.x.
- Davis, T.W., Berry, D.L., Boyer, G.L., and Gobler, C.J. 2009. The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of Microcystis during cyanobacteria blooms. Harmful Algae, 8(5): 715–725. doi:10.1016/j.hal.2009.02.004.
- Davis, T.W., Harke, M.J., Marcoval, M.A., Goleski, J., Orano-Dawson, C., Berry, D.L., and Gobler, C.J. 2010. Effects of nitrogenous compounds and phosphorus on the growth of toxic and non-toxic strains of *Microcystis* during cyanobacterial blooms. Aquat. Microbiol. Ecol. **61**(2): 149–162. doi:10.3354/ ame01445.
- DeMarchi, C., Tao, W., Johengen, T.H., and Stow, C.A. 2010. Uncertainty in phosphorus load estimation from a large watershed in the Great Lakes basin. *In* SimHydro 2010: hydraulic modeling and uncertainty. Sophia–Antipolis, 2–4 June. Hydraulic Engineering Company of France, International Association of Hydraulic Research and Engineering. pp. 1–13.
- De Stasio, B.T., Schrimpf, M.B., Beranek, A.E., and Daniels, W.C. 2008. Increased chlorophyll *a*, phytoplankton abundance, and cyanobacteria occurrence following invasion of Green Bay, Lake Michigan by dreissenid mussels. Aquat. Invasions, 3(1): 21–27. doi:10.3391/ai.2008.3.1.5.
- Doak, D.F., Estes, J.A., Halpern, B.S., Jacob, U., Lindberg, D.R., Lovvorn, J., Monson, D.H., Tinker, M.T., Williams, T.M., Wootton, J.T., Carroll, I., Emmerson, M., Micheli, F., and Novak, M. 2008. Understanding and predicting ecological dynamics: are major surprises inevitable? Ecology, 89(4): 952– 961. doi:10.1890/07-0965.1. PMID:18481520.
- Dolan, D.M., and Chapra, S.C. 2012. Great Lakes total phosphorus revisited: 1. Loading analysis and update (1994–2008). J. Gt. Lakes Res. 38(4): 730–740. doi:10.1016/j.jglr.2012.10.001.
- Downing, J.A., Watson, S.B., and McCauley, E. 2001. Predicting Cyanobacteria dominance in lakes. Can. J. Fish. Aquat. Sci. 58(10): 1905–1908. doi:10.1139/f01-143.
- Dyble, J., Fahnenstiel, G.L., Litaker, R.W., Millie, D.F., and Tester, P.A. 2008. Microcystin concentrations and genetic diversity of *Microcystis* in the lower Great Lakes. Environ. Toxicol. 23(4): 507–516. doi:10.1002/tox.20370. PMID: 18247416.
- Elith, J., and Leathwick, J.R. 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40: 677–697. doi:10.1146/annurev.ecolsys.110308.120159.
- Environment Canada and US Environmental Protection Agency. 2003. Implementing Indicators 2003 — A Technical Report. Environment Canada, Downsview, Ont.
- Fahnenstiel, G.L., Krause, A.E., McCormick, M.J., Carrick, H.J., and Schelske, C.L. 1998. The structure of the planktonic food-web in the St. Lawrence Great Lakes. J. Gt. Lakes Res. 24(3): 531–554. doi:10.1016/S0380-1330(98)70843-3.
- Gevrey, M., Dimopoulos, I., and Lek, S. 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Modell. **160**(3): 249–264. doi:10.1016/S0304-3800(02)00257-0.
- Gevrey, M., Dimopoulos, I., and Lek, S. 2006. Two-way interaction of input variables in the sensitivity analysis of neural network models. Ecol. Modell. 195(1–2): 43–50. doi:10.1016/j.ecolmodel.2005.11.008.
- Gonzalez, S. 2000. Neural networks for macroeconomic forecasting: a complimentary approach to linear regression models. Canadian Department of Finance Working Paper No. 2000-07, Ottawa, Ont., Canada.
- Great Lakes Commission. 2012. Priorities for reducing phosphorus loadings and abating algal blooms in the Great Lakes-St. Lawrence River basin: opportunities and challenges for improving Great Lakes aquatic ecosystems. Phosphorus Reduction Task Force Technical Report, Great Lakes Commission, Ann Arbor, Mich.
- Groffman, P.M., Baron, J.S., Blett, T., Gold, A.J., Goodman, I., Gunderson, L.H., Levinson, B.M., Palmer, M.A., Paerl, H.W., Peterson, G.D., Poff, N.L., Rejeski, D.W., Reynolds, J.F., Turner, M.G., Weathers, K.C., and Wiens, J. 2006. Ecological thresholds: the key to successful environmental management or an important concept with no practical application? Ecosystems, 9(1): 1–13. doi:10.1007/s10021-003-0142-z.

- Guisan, A., and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. Ecol. Lett. 8(9): 993–1009. doi:10.1111/j.1461-0248. 2005.00792.x.
- Håkanson, L. 2000. The role of characteristic coefficients of variation in uncertainty and sensitivity analyses, with examples related to the structuring of lake eutrophication models. Ecol. Modell. **131**(1): 1–20. doi:10.1016/S0304-3800(00)00219-2.
- Hartman, W. 1973. Effects of exploitation, environmental changes, and new species on the fish habitats and resources of Lake Erie. Great Lakes Fish Commission, Tech. Rep. 22, Ann Arbor, Mich.
- Heffernan, J.B., Liebowitz, D.M., Frazer, T.K., Evans, J.M., and Cohen, M.J. 2010. Algal blooms and the nitrogen-enrichment hypothesis in Florida springs: evidence, alternatives, and adaptive management. Ecol. Appl. 20(3): 816–829. doi:10.1890/08-1362.1. PMID:20437966.
- International Joint Commission. 2014. A balanced diet for Lake Erie: reducing phosphorus loadings and harmful algal blooms. Report of the Lake Erie Ecosystem Priority, Washington, D.C., and Ottawa, Ont.
- Jørgensen, S.E., Chon, T.-S., and Recknagel, F.A. (*Editors*). 2009. Handbook of ecological modelling and informatics. WIT Press, Southhampton, UK.
- Michalak, A.M., Anderson, E.J., Beletsky, D., Boland, S., Bosch, N.S., Bridgeman, T.B., Chaffin, J.D., Cho, K., Confesor, R., Daloğlu, I., DePinto, J.V., Evans, M.A., Fahnenstiel, G.L., He, L., Ho, J.C., Jenkins, L., Johengen, T.H., Kuo, K.C., LaPorte, E., Liu, X., McWilliams, M.R., Moore, M.R., Posselt, D.J., Richards, R.P., Scavia, D., Steiner, A.L., Verhamme, E., Wright, D.M., and Zagorski, M.A. 2013. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. Proc. Nat. Acad. Sci. 110(16): 6448–6452. doi:10.1073/pnas.1216006110. PMID:23576718.
- Michener, W.K., and Jones, M.B. 2012. Ecoinformatics: supporting ecology as a data-intensive science. Trends Ecol. Evol. **27**(2): 85–93. doi:10.1016/j.tree.2011. 11.016. PMID:22240191.
- Millie, D.F., Pigg, R., Tester, P.A., Dyble, J., Litaker, R.W., Carrick, H.J., and Fahnenstiel, G.L. 2006a. Modeling phytoplankton abundance in Saginaw Bay, Lake Huron: using artificial neural networks to discern functional influence of environmental variables and relevance to a Great Lakes Observing System. J. Phycol. 42(2): 336–349. doi:10.1111/j.1529-8817.2006.00209.x.
- Millie, D.F., Weckman, G.R., Paerl, H.W., Pinckney, J.L., Bendis, B.J., Pigg, R.J., and Fahnenstiel, G.L. 2006b. Neural net modeling of estuarine indicators: hindcasting phytoplankton biomass and net ecosystem production in the Neuse (North Carolina) and Trout (Florida) Rivers, USA. Ecol. Indic. 6(3): 589– 608. doi:10.1016/j.ecolind.2005.08.021.
- Millie, D.F., Fahnenstiel, G.L., Bressie, J.D., Pigg, R.J., Rediske, R.R., Klarer, D.M., Tester, P.A., and Litaker, R.W. 2009. Late-summer phytoplankton in western Lake Erie (Laurentian Great Lakes): bloom distributions, toxicity, and environmental influences. Aquat. Ecol. 43(4): 915–934. doi:10.1007/s10452-009-9238-7.
- Millie, D.F., Pigg, R.J., Fahnenstiel, G.L., and Carrick, H.J. 2010. Algal chlorophylls: a synopsis of analytical methodologies. *In* American Water Works Association, Manual M57, Algae. American Water Works Association, Denver, Colo., USA. pp. 93–122.
- Millie, D.F., Fahnenstiel, G.L., Weckman, G.R., Klarer, D.M., Dyble, J., Vanderploeg, H.A., and Fishman, D.B. 2011. An 'enviro-informatic' assessment of Saginaw Bay (Lake Huron, USA) phytoplankton: data-driven characterization and modeling of Microcystis (Cyanophyta). J. Phycol. 47(4): 714–730. doi:10.1111/ j.1529-8817.2011.01022.x.
- Millie, D.F., Weckman, G.R., Young, W.A., Ivey, J.E., Carrick, H.J., and Fahnenstiel, G.L. 2012. Modeling microalgal abundance with artificial neural networks: demonstration of a heuristic 'Grey-Box'to deconvolve and quantify environmental influences. Environ. Modell. Soft. 38(December): 27–39. doi:10.1016/j.envsoft.2012.04.009.
- Millie, D.F., Weckman, G.R., Young, W.A., II, Ivey, J.E., Fries, D.P., Ardjmand, E., and Fahnenstiel, G.L. 2013. Coastal 'big data' and nature-inspired computation: prediction potentials, uncertainties, and knowledge derivation of neural networks for an algal metric. Estuar. Coast. Shelf Sci. 125(July): 57–67. doi:10.1016/j.ecss.2013.04.001.
- Nicholls, K.H. 1997. Planktonic green algae in western Lake Erie: the importance of temporal scale in the interpretation of change. Freshw. Biol. 38(2): 419– 425. doi:10.1046/j.1365-2427.1997.00247.x.
- Ohio Environmental Protection Agency. 2010. Ohio Lake Erie Phosphorus Task Force Final Report Executive Summary. Division of Surface Water, Columbus, Ohio.
- Olden, J.D., and Jackson, D.A. 2002. Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecol. Modell. 154(1–2): 135–150. doi:10.1016/S0304-3800(02)00064-9.
- O'Neil, J.M., Davis, T.W., Burford, M.A., and Gobler, C.J. 2012. The rise of harmful cyanobacteria blooms: the potential roles of eutrophication and climate change. Harmful Algae, 14: 313–334. doi:10.1016/j.hal.2011.10.027.
- Osborne, J., and Waters, E. 2002. Four assumptions of multiple regression that researchers should always test [online]. Pract. Assess. Res. Eval. 8(2). Available from http://pareonline.net/getvn.asp?n=2&v=8.
- Otsuka, S., Suda, S., Shibata, S., Oyaizu, H., Matsumoto, S., and Watanabe, M.M. 2001. A proposal for the unification of five species of the cyanobacterial genus *Microcystis* Kützing ex Lemmermann 1907 under the rules of the bacte-

riological code. Int. J. Syst. Evol. Microbiol. **51**(3): 873–879. doi:10.1099/00207713-51-3-873. PMID:11411709.

- Paerl, H.W., and Huisman, J. 2008. Blooms like it hot. Science, **320**(5872): 57–58. doi:10.1126/science.1155398. PMID:18388279.
- Paerl, H.W., and Otten, T.G. 2013. Blooms bite the hand that feeds them. Science, 342(6157): 433–434. doi:10.1126/science.1245276. PMID:24159037.
- Paerl, H.W., Hall, N.S., and Calandrino, E.S. 2011a. Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climaticinduced change. Sci. Tot. Environ. 409(10): 1739–1745. doi:10.1016/j.scitotenv. 2011.02.001.
- Paerl, H.W., Xu, H., McCarthy, M.J., Zhu, G., Qin, B., Li, Y., and Gardner, W.S. 2011b. Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): The need for a dual nutrient (N & P) management strategy. Wat. Res. 45(5): 1973–1983. doi:10.1016/j.watres.2010.09.018.
- Pascual, M., and Ellner, S.P. 2000. Linking ecological patterns to environmental forcing via nonlinear time series models. Ecology, 81(10): 2767–2780. doi:10. 1890/0012-9658(2000)081[2767:LEPTEF]2.0.CO;2.
- Patterson, G.M.L., Harris, D.O., and Cohen, W.S. 1979. Inhibition of photosynthetic and mitochondrial electron transport by a toxic substance isolated from the alga *Pandorina morum*. Plant Sci. Lett. **15**: 293–300. doi:10.1016/0304-4211(79)90133-0.
- Peek, M.S., Leffler, A.J., Flint, S.D., and Ryel, R.J. 2003. How much variance is explained by ecologists? Additional perspectives. Oecologia, 137(2): 161–170. doi:10.1007/s00442-003-1328-y.
- Peterson, A.T. 2006. Uses and requirements of ecological niche models and related distribution models. Biodiv. Inform. **3**: 59–72.
- Principe, J.C., Euliano, N.R., and Lefebrve, W.C. 2000. Neural and adaptive systems: fundamentals through simulation. John Wiley and Sons Inc., New York.

- Quetglas, A., Ordines, F., and Guijarro, B. 2011. The use of artificial neural networks (ANNs) in aquatic ecology [online]. In Artificial neural networks application. Edited by C.L.P. Hui. ISBN: 978-953-307-188-6. InTech. Available from http://www.intechopen.com/books/artificial-neural-networksapplication/the-use-of-artificial-neural-networks-anns-in-aquatic-ecology [accessed 13 July 2014]. pp. 576–586.
- Reckhow, K.H., Clements, J.T., and Dodd, R.C. 1990. Statistical evaluation of mechanistic water-quality models. J. Environ. Eng. 116(2): 250–268. doi:10. 1061/(ASCE)0733-9372(1990)116:2(250).
- Reynolds, C.S. 2006. The ecology of phytoplankton. Cambridge University Press, Cambridge, UK.
- Smol, J.P., and Stoermer, E.F. (Editors). 2010. The diatoms: applications for the environmental and earth Sciences. 2nd ed. Cambridge University Press, UK.
- Stumpf, R.P., Wynne, T.T., Baker, D.B., and Fahnenstiel, G.L. 2012. Interannual variability of cyanobacterial blooms in Lake Erie. PLoS ONE, 7(8): e42444. doi:10.1371/journal.pone.0042444. PMID:22870327.
- Sun, X.Y., Newham, L.T.H., Croke, B.F.W., and Norton, J.P. 2012. Three complementary methods for sensitivity analysis of a water quality model. Environ. Modell. Soft. **37**(November): 19–29. doi:10.1016/j.envsoft.2012.04.010.
- Thuiller, W., Lafourcade, B., Engler, R., and Araújo, M.B. 2009. BIOMOD a platform for ensemble forecasting of species distributions. Ecography, 32(3): 369–373. doi:10.1111/j.1600-0587.2008.05742.x.
- Utermöhl, H. 1958. Zur vervollkommung der quantitativen phytoplanktonmethodik. Mitt. Int. Ver. Limnol. 9: 1–38.
- Wood, S.N. 2010. Statistical inference for noisy nonlinear ecological dynamic systems. Nature, 466(7310): 1102–1104. doi:10.1038/nature09319. PMID:20703226.
- Zhang, H., Hu, W., Gu, K., Li, Q., Zheng, D., and Zhai, S. 2013. An improved ecological model and software for short-term algal bloom forecasting. Environ. Modell. Soft. 48(October): 152–162. doi:10.1016/j.envsoft.2013.07.001.

# This article has been cited by:

- 1. Carla Kruk, Angel M. Segura, Lucía Nogueira, Ignacio Alcántara, Danilo Calliari, Gabriela Martínez de la Escalera, Carmela Carballo, Carolina Cabrera, Florencia Sarthou, Paola Scavone, Claudia Piccini. 2017. A multilevel trait-based approach to the ecological performance of Microcystis aeruginosa complex from headwaters to the ocean. *Harmful Algae* **70**, 23-36. [Crossref]
- 2. A.M. Segura, C. Piccini, L. Nogueira, I. Alcántara, D. Calliari, C. Kruk. 2017. Increased sampled volume improves Microcystis aeruginosa complex (MAC) colonies detection and prediction using Random Forests. *Ecological Indicators* **79**, 347-354. [Crossref]
- 3. Iman Ghalehkhondabi, Ehsan Ardjmand, William A. Young, Gary R. Weckman. 2017. Water demand forecasting: review of soft computing methods. *Environmental Monitoring and Assessment* 189:7. . [Crossref]
- 4. Mohammad Jahedi, Ehsan Ardjmand, Marko Knezevic. 2017. Microstructure metrics for quantitative assessment of particle size and dispersion: Application to metal-matrix composites. *Powder Technology* **311**, 226-238. [Crossref]
- 5. Emon Butts, Hunter J. Carrick. 2017. Phytoplankton Seasonality along a Trophic Gradient of Temperate Lakes: Convergence in Taxonomic Compostion during Winter Ice-Cover. *Northeastern Naturalist* 24:sp7, B167-B187. [Crossref]
- 6. Isabella Bertani, Cara E. Steger, Daniel R. Obenour, Gary L. Fahnenstiel, Thomas B. Bridgeman, Thomas H. Johengen, Michael J. Sayers, Robert A. Shuchman, Donald Scavia. 2017. Tracking cyanobacteria blooms: Do different monitoring approaches tell the same story?. *Science of The Total Environment* 575, 294-308. [Crossref]
- Siyun Wang, Daniel Weller, Justin Falardeau, Laura K. Strawn, Fernando O. Mardones, Aiko D. Adell, Andrea I. Moreno Switt. 2016. Food safety trends: From globalization of whole genome sequencing to application of new tools to prevent foodborne diseases. *Trends in Food Science & Technology* 57, 188-198. [Crossref]
- 8. Euan D. Reavie, Meijun Cai, Michael R. Twiss, Hunter J. Carrick, Timothy W. Davis, Thomas H. Johengen, Duane Gossiaux, Derek E. Smith, Danna Palladino, Ashley Burtner, Gerald V. Sgro. 2016. Winter–spring diatom production in Lake Erie is an important driver of summer hypoxia. *Journal of Great Lakes Research* **42**:3, 608-618. [Crossref]
- 9. F. García-Camacho, L. López-Rosales, A. Sánchez-Mirón, E.H. Belarbi, Yusuf Chisti, E. Molina-Grima. 2016. Artificial neural network modeling for predicting the growth of the microalga Karlodinium veneficum. *Algal Research* 14, 58-64. [Crossref]
- 10. Wen-Ping Tsai, Fi-John Chang, Edwin E. Herricks. 2016. Exploring the ecological response of fish to flow regime by soft computing techniques. *Ecological Engineering* 87, 9-19. [Crossref]
- Ehsan Ardjmand, David F. Millie, Iman Ghalehkhondabi, William A. Young II, Gary R. Weckman. 2016. A State-Based Sensitivity Analysis for Distinguishing the Global Importance of Predictor Variables in Artificial Neural Networks. *Advances in Artificial Neural Systems* 2016, 1-11. [Crossref]