

Examination of Multiple Linear Regression (MLR) and Neural Network (NN) Models to Predict Eutrophication Levels in Lake Champlain

L. E. Farra¹, K. Wang², Z. Chen^{1*}, and Y. Y. Zhu¹

¹ Faculty of Engineering and Computer Sciences, Concordia University, Montreal, Quebec H3G 1M8, Canada

² Department of Engineering Mechanics, Dalian University of Technology, Dalian 116024, China

Received 6 March 2019; revised 30 April 2019; accepted 8 June 2019; published online 30 June 2019

ABSTRACT. Eutrophication is one of the main causes of the degradation of lake ecosystems. In this paper, multiple linear regression (MLR) and neural network (NN) methods were developed as empirical models to predict chlorophyll-a (Chl-a) concentrations in Lake Champlain. The models were developed using a large dataset collected from Lake Champlain over a 24-year period from 1992 to 2016. The dataset consisted of monitoring depth (*Depth*), total phosphorus (*TP*), total nitrogen (*TN*), alkalinity (*RegAlk*), temperature (*TempC*), chloride (*Cl*) and secchi depth (*Secchi*). Statistical analyses showed that *TP*, *Secchi*, *TN* and *Depth* demonstrated strong relationships with Chl-a concentrations. The simulation results revealed that both the MLR and NN models performed well in predicting Chl-a concentrations, especially for low to moderate concentrations of Chl-a ($< 7.5 \mu\text{g/L}$). The NN model showed better accuracy and generalization performance in comparison with the MLR model for both the training and verification processes. In addition, both the developed MLR and NN models produce good results when used to predict Chl-a concentrations from 2017 to 2021. However, neither the MLR nor NN models can accurately predict high Chl-a concentrations ($> 7.5 \mu\text{g/L}$). These models can be useful for improving lake management and providing early warnings regarding the problem of eutrophication.

Keywords: neural network (NN), multiple linear regression (MLR), eutrophication, chlorophyll-a (Chl-a) prediction, Lake Champlain, empirical models

1. Introduction

Eutrophication has been one of the main water-based environmental problems in lakes and reservoirs since the 1960s (Smith et al., 2006). It can promote the excessive growth and decay of plants and bacteria, including cyanobacteria blooms. During photosynthesis, cyanobacteria blooms consume nutrients for lake biome survival and produce toxins which are poisonous to humans and lake wildlife. It may take decades before enough nutrients are naturally eliminated from a lake for it to recover from eutrophication (Chambers et al., 2001; Hiscock et al., 2003). Chlorophyll-a (Chl-a) can be used as a biomarker for the presence of cyanobacteria, as there is a direct relationship between the mass of the cyanobacterial algal bloom and the concentration of Chl-a in fresh water (Elliott et al., 2007; Burger et al., 2008; Freeman et al., 2009).

The German agricultural chemist Justus von Liebig conducted the first eutrophication study in 1950. Prior to this, Weber (1907) and Johnstone (1908) found a link between nutrients and aquatic productivity (Smith et al., 1999). In the following years, many eutrophication studies were conducted. Phosphorus is generally regarded as the principal cause of lake eutrophication. In

addition, the increasing frequency of cyanobacteria blooms might be linked to climate change (Anderson et al., 2012; Pomati et al., 2017; Yao et al., 2018; Le Moal et al., 2019). Since the 1970s, the modeling approach has been considered to be an interesting tool due to the rapid development of computer techniques (Imboden, 1974; Vollenweider, 1975). There are two main classes of eutrophication modeling: empirical models and mechanistic models (Fornarelli et al., 2013).

Empirical models mainly analyze relationships between predictor variables (e.g., Chl-a) and interest variables (e.g., nutrients and temperature) based on available data sets. In addition, they are more generic than mechanistic models and can provide good predictions for lakes (Van Huet, 1992). On the other hand, empirical models have large uncertainty levels due to inaccuracies in the input variables and modeling structures. The main modeling techniques of empirical models include the multiple linear regression (MLR), neural network (NN), data mining (DM) and tree-based methods. The first empirical model was the regression model, which assumed linear nutrients-Chl-a relationships between variables (Dillon and Rigler, 1974; Reckhow, 1993). Brown et al. (2000) developed MLR models representing nutrients Chl-a relationships using a large data set collected over 10 years from 360 lakes in Florida. However, their precision is low. Huszar et al. (2006) revealed that a multiple regression with log *TP* and log *TN* separating systems with high *TN:TP* (> 17 by weight) improved the predictive power

* Corresponding author. Tel.: +1 514-848-2424 ext. 8775;
E-mail address: zhi.chen@concordia.ca (Z. Chen).

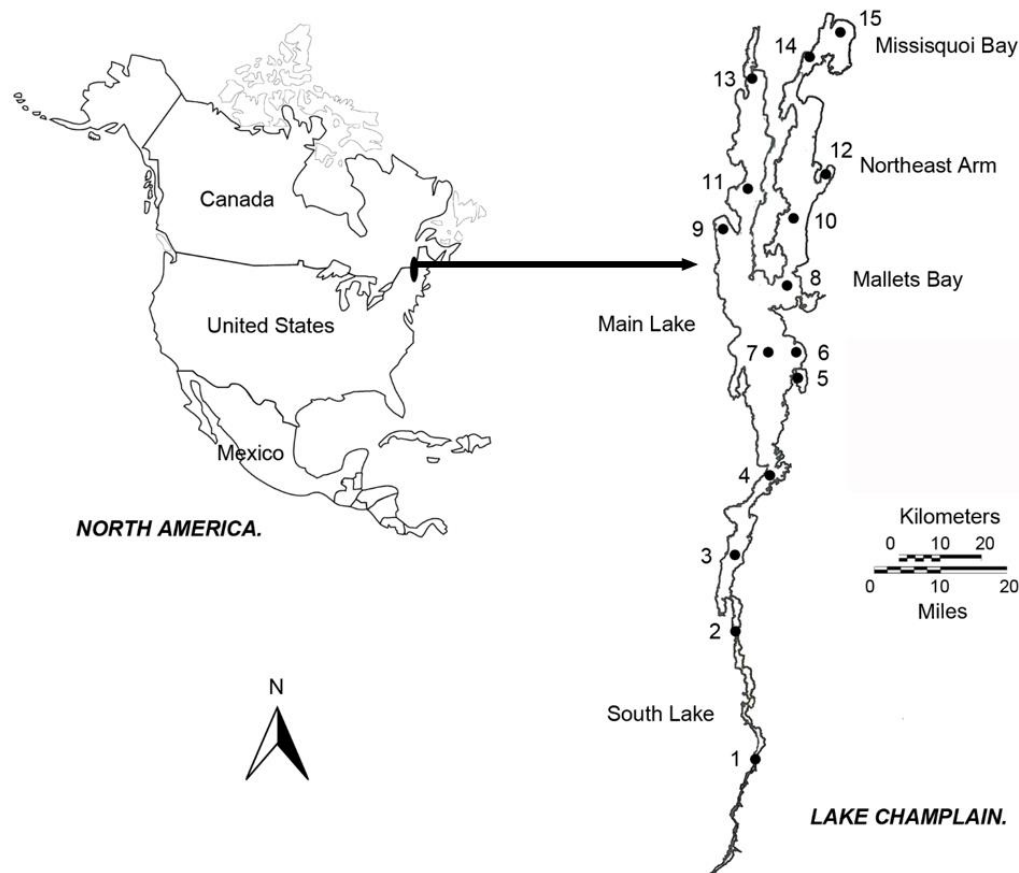


Figure 1. Location of study area in North America and sampling stations of Lake Champlain watershed.

of the log *TP*-log Chl-*a* relationship based on a data set from 192 lakes in tropical and subtropical regions. A new monthly regression model was developed and accurately predicted monthly Chl-*a* medians based on the median summer *TP* concentrations and latitude (Dimberg et al., 2013). Matus-Hernández et al. (2018) evaluated the predictive performance of the simple linear regression (SLR), multiple linear regression (MLR) and generalized additive models (GAMs) based on Landsat imagery. The result showed that MLR performed best in predicting log-transformed Chl-*a*.

The statistical models of the 1970s and 1980s have evolved significantly over the last decade towards “data-driven” models. Such models need large data sets which were unavailable in the past, but which lake monitoring systems are now able to collect. In addition, the neural network (NN) has been applied in areas as diverse as business, physics, geology, medicine, engineering and, in particular, environmental engineering (Bazregar, and Moghaddam, 2016). NN is utilized as an efficient modeling approach to simulating, predicting, classifying or controlling a system and can be applied in every situation in which a relationship between the predictor variables and interest variables exists regardless of the relationship’s complexity. However, NN may not produce a specific equation. It is possible that running with the same data set would result in a different

NN model because that NN is a learning algorithm that works by minimizing the error successively from the previous equation. NN is now commonly used to study lake eutrophication because of its recognized ability to predict highly nonlinear and complex relationships (Ranković et al., 2010; Fornarelli et al., 2013; Chen and Liu, 2014; Jeong et al., 2015). For example, the principle component regression (PCR), neural network (NN) and hybrid models were applied to predict and forecast the phytoplankton dynamics in the Macau Main Storage Reservoir. The simulation result showed that the NN model has greater accuracy and generalization performance compared with the PCR models (Jeong et al., 2015). This paper is intended to examine data-driven statistical model and neural network methods through a regional scale lake eutrophication case study in the North America. First, the contributions of lake water quality parameters will be analyzed (i.e., monitoring depth (*Depth*), total phosphorus (*TP*), total nitrogen (*TN*), alkalinity (RegAlk), temperature (*TempC*), chloride (*Cl*) and secchi depth (*Secchi*)) to algae bloom using correlation analysis; then Chl-*a* models using the multiple linear regression (MLR) and neural network (NN) methodologies will be developed for the Lake Champlain based on 24-year data from 1992 to 2016 and to predict the Chl-*a* concentrations of Lake Champlain from 2017 to 2021 based on the validated MLR and NN models.

2. Data and Methodology

2.1. Case Study Area and Data Collection

Lake Champlain is one of the largest glacially formed lakes in North America, existing partially in the US states of Vermont and New York, and partially in Quebec, Canada. It is approximately 94 km in length, 19 km in maximum width, 122 m in maximum depth (though most of it is shallow water of about 1.5 m in depth), and 19 m in mean depth (Smeltzer et al., 2012). The lake has 5 different environmental zones, which are shown in Figure 1 (New York State DEC and Vermont DEC, 2012): the South Lake, which is long, narrow and shallow; the Main Lake, which is the deepest and widest section of the lake; Malletts Bay, which is circumscribed by historical railroad and road cause-ways; the Inland Sea, which lies to the east of the Hero Islands; the Missisquoi Bay, which is a large and discrete bay rich with wildlife. The eutrophication problem in Lake Champlain started in the early 1980s due to high phosphorus levels from agricultural runoff and municipal sewage treatment plants, which caused excessive cyanobacteria algal blooms (CABs). As a result, the drinking water was contaminated by trihalomethanes (THMs) produced by CABs (Gopal, 2005).

The 24-year data set (1992 ~ 2016) of water quality parameters used in the modeling process was taken from the website of the state of Vermont, which has operated under the Long-Term Water Quality and Biological Monitoring Program on Lake Champlain (LTMP) since 1992 (New York State DEC and Vermont DEC, 2012). The state of Vermont decided to share Lake Champlain's environmental data with the public to help researchers conduct studies that could provide solutions for the lake's problems. Each year, 15 sampling locations throughout Lake Champlain were monitored approximately every two weeks from May to early November (Figure 1). These lake stations were sampled consistently throughout the entire monitoring period beginning in 1992, with the exception of lake stations 4 and 5, which were added in 2001, and station 15, which was added in 2006 (New York State DEC and Vermont DEC, 2012).

The data includes information on the total maximum daily loads of pollutants, available to the public and researchers via the Lake Champlain watershed management web site. In this modeling study, the data regarding 8 water quality parameters (i.e., chlorophyll-a (Chl-a, µg/L), total phosphorus (TP, µg/L), chloride (Cl, µg/L), secchi depth (Secchi, m), total nitrogen (TN, µg/L), temperature (TempC, °C), alkalinity (RegAlk) and monitoring depth (Depth, m)) from the 15 stations from 1992 ~ 2016 were downloaded.

2.2. Multiple Linear Regression (MLR)

Some published Chl-a models use the statistical method of multiple linear regression (MLR) to investigate multiple scalar independent variables (Z = water quality parameters) that are hypothesized to be linearly related to the dependent variable (Y = Chl-a), which causes eutrophication in lakes. This is described below in the general matrix format for the MLR Equation (Khuri, 2013):

$$Y = Z\beta + \varepsilon \quad (1)$$

$$Y = \begin{bmatrix} CHLa_1 \\ CHLa_2 \\ \dots \\ CHLa_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_n \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{bmatrix}, \quad (2)$$

$$Z = \begin{bmatrix} 1 & TP_{11} & TN_{12} & Cl_{13} & Secchi_{14} & TempC_{15} & Depth_{16} & RegAlk_{17} \\ 1 & TP_{21} & TN_{22} & Cl_{23} & Secchi_{24} & TempC_{25} & Depth_{26} & RegAlk_{27} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & TP_{n1} & TN_{n2} & Cl_{n3} & Secchi_{n4} & TempC_{n5} & Depth_{n6} & RegAlk_{n7} \end{bmatrix}$$

In this type of MLR model, Y is an n -by-1 vector of responses and represents the dependent or output variable (chlorophyll-a, Chl-a). Z is an n -by-8 design matrix for the model and represents the independent or input variables: total phosphorus (TP), total nitrogen (TN), chloride (Cl), secchi depth ($Secchi$), temperature ($TempC$), monitoring depth ($Depth$) and alkalinity ($RegAlk$). β is an 8-by-1 vector of coefficients for the independent or input variables. ε is an n -by-1 vector of error terms. The short version of the general MLR format is written as Equation (3):

$$CHLa = \beta_0 + \beta_1 \times TP + \beta_2 \times TN + \beta_3 \times Cl + \beta_4 \times Secchi + \beta_5 \times TempC + \beta_6 \times Depth + \beta_7 \times RegAlk + \varepsilon \quad (3)$$

The MLR equation is solved using the least squares method. The unknown vector of coefficients β of the linear equation is estimated by minimizing the sum of squares of the residual errors between the observed data and the data predicted by the linear equation. The coefficients β that produce the best solution are found when the error between the linear equation model and the observed data is zero (Kariya, 2004). By setting $\varepsilon = 0$ and rearranging the equation, we get $\beta = S(\beta)/Z$, which gives the coefficients of matrix β and the predicted values. By comparing the predicted values with the observed values, we can judge the model's accuracy.

2.3. Neural Network (NN)

Unlike other modeling techniques, NN does not generate a model with coefficients but instead produces multilayer interconnected neural processing units that imitate human brain activity, in which each neuron in a layer is connected to every neuron in the next layer. Figure 2 shows the structure of neural networks. The outputs of the NN model are weights which are saved as an xml file. The file can then be used in the forecasting and verification of a different data set.

In NN, the input data goes through the NN model, where it is multiplied by the weights (brain) in a forward direction. The information is then processed, and the output is compared with the observed value. The error resulting from the comparison is backpropagated and becomes an input for the next prediction while the model weight (brain) is adjusted to minimize the error via several iterations. The data are plotted as variable importance charts which show the impacts of variables on the model, and also as synaptic weight charts which show the in-

fluences of neurons (Chen and Liu, 2014).

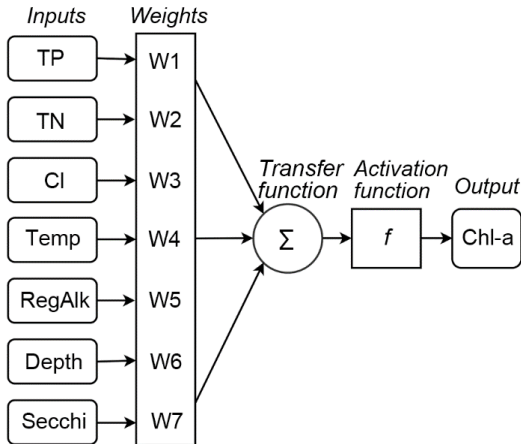


Figure 2. The structure of neural networks (Nasrabadi, 2007).

2.4. Examination of Modeling Methods

To find the best Chl-a model for the Lake Champlain case study, three published methods, including the determination of coefficient (R^2), standard error and confidence interval (CI) were used to evaluate the Chl-a models. The determination of coefficient (R^2) was used to evaluate Chl-a models used in lake eutrophication studies (Çamdevýren et al., 2005; Rui et al., 2012). The Pearson correlation coefficient (R) measures the linear correlation between two variables (value between +1 and -1). The two variables show a strong positive or negative linear relationship when the R value is close to +1 or -1. Normally, there is a strong positive or negative linear relationship with two variables with higher R values. R^2 (the square of the Pearson R) indicates how closely the regression model fits the observed data (value between 0 and 1). The closer the R^2 value is to 1, the better the model fit. The standard error (SE) is an estimate of the average squared error (Kenney, 2013). The model is more accurate with the lower SE value. The standard error was used in the verification analysis of Lake Ontario (Thomann et al., 1979). Confidence interval (CI) thresholds are used to maintain small and evenly distributed errors, and error values outside the threshold (also called the limit) values are ignored. The lower the CI threshold value, the better the model. The critical values are the boundaries of the CI, found by using the z score table (the lower critical value = $-Z(\alpha/2)$; the upper critical value = $Z(\alpha/2)$). Most commonly, the 95% confidence level is used with $Z = 1.960$. However, other confidence levels can be used, for example, 90% ($Z = 1.645$) and 99% ($Z = 2.576$). The critical values in most data analysis software packages are user-defined inputs which are set manually before data processing. Figure 3 shows the flow chart of the methods of modeling Chl-a concentrations in Lake Champlain.

3. Results and Discussion

3.1. Correlation Analysis

Correlation analysis was performed on the water parameters to rank their relative significance to Chl-a concentrations, as well as to describe their relationship. Table 1 and Figure 4 show the result of the cross-correlation analysis, which examines the relationship and impact of seven independent water quality parameters (i.e., TP, TN, CI, Secchi, TempC, RegAlk and Depth) on the dependent variable (Chl-a) in Lake Champlain.

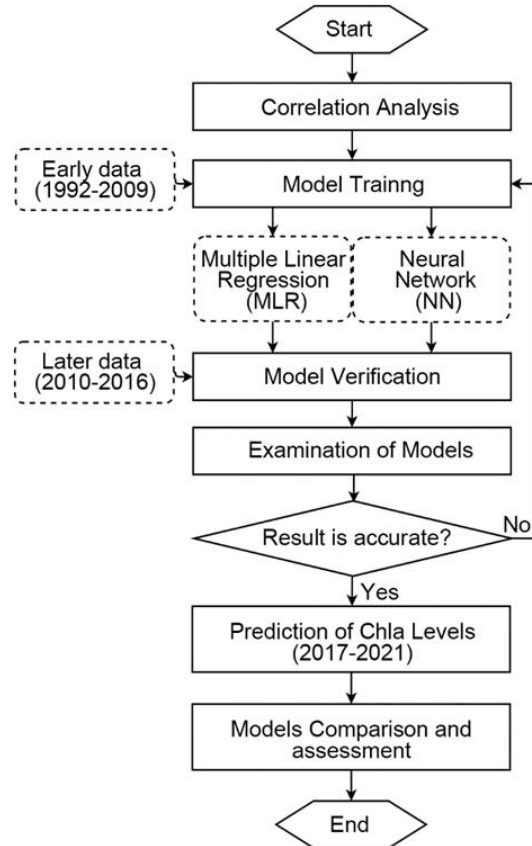


Figure 3. The modeling approach.

As expected, the cross-correlation analysis indicated that there was a strong positive linear relationship between the independent variable TP and the dependent variable Chl-a ($R = 0.817$). Moreover, the independent variable TN and the dependent variable Chl-a showed at medium positive linear interrelationship ($R = 0.682$). However, there was a medium negative linear relationship between the independent variable *Secchi*, a measure of water clarity which is reduced by the growth of CABs, and the dependent variable Chl-a ($R = -0.722$) (Libes, 2011).

In addition, the analysis confirmed that there was a medium positive linear relationship between two of the independent variables, TP and TN ($R = 0.678$), due to the fact that they are both nutrients used in CAB growth (Downing and McCauley, 1992). The two independent variables *Secchi* and TP also showed a strong negative linear interrelationship ($R =$

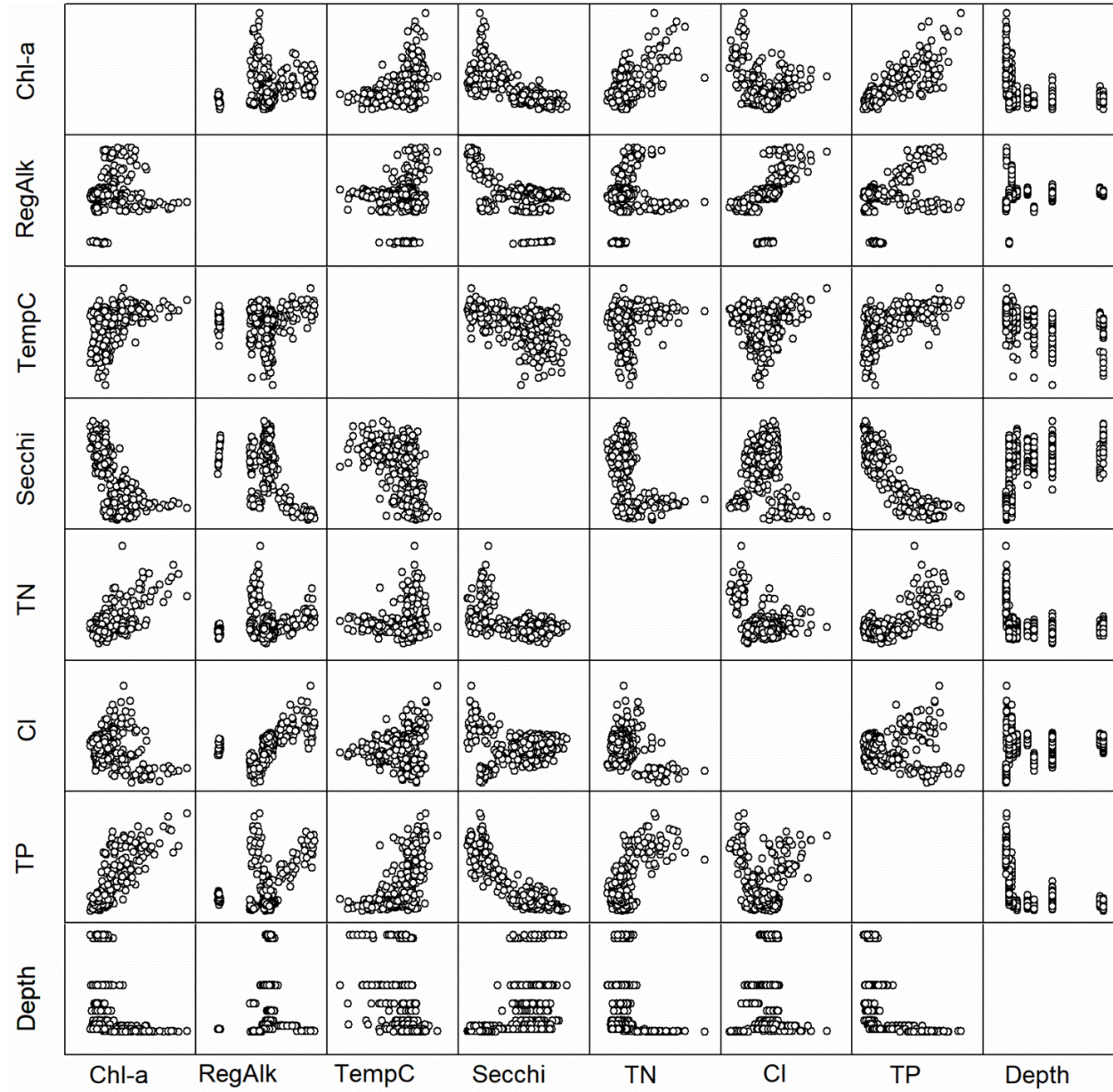


Figure 4. Cross-correlation scatter plots showing the relationship between the 7 independent water quality variables and dependent variable Chl-a (1992 ~ 2016).

Table 1. Pearson's Correlation Matrix

Parameters	Chl-a ($\mu\text{g/L}$)	RegAlk ($\mu\text{g/L}$)	TempC ($^{\circ}\text{C}$)	Secchi (m)	TN ($\mu\text{g/L}$)	Cl ($\mu\text{g/L}$)	TP ($\mu\text{g/L}$)	Depth (m)
Chl-a($\mu\text{g/L}$)	1.000	0.159	0.373	-0.722	0.682	-0.282	0.817	-0.402
RegAlk($\mu\text{g/L}$)	0.159	1.000	0.144	-0.453	0.065	0.534	0.347	0.042
TempC($^{\circ}\text{C}$)	0.373	0.144	1.000	-0.486	0.107	0.161	0.482	-0.441
Secchi(m)	-0.722	-0.453	-0.486	1.000	-0.544	-0.059	-0.874	0.534
TN($\mu\text{g/L}$)	0.682	0.065	0.107	-0.544	1.000	-0.380	0.678	-0.286
Cl($\mu\text{g/L}$)	-0.282	0.534	0.161	-0.059	-0.380	1.000	-0.078	0.083
TP($\mu\text{g/L}$)	0.817	0.347	0.482	-0.874	0.678	-0.078	1.000	-0.515
Depth(m)	-0.402	0.042	-0.441	0.534	-0.286	0.083	-0.515	1.000

Table 2. Lake Champlain MLR Modeling Results for 1992 ~ 2009

MLR model#	R	R ²	Adjusted R ²	Std. Error	Predictors
1	0.766	0.587	0.586	2.076	Constant, TP
2	0.808	0.654	0.651	1.906	Constant, TP, Cl
3	0.814	0.663	0.659	1.884	Constant, TP, Cl, Secchi
4	0.818	0.669	0.663	1.872	Constant, TP, TP, Cl, Secchi, TN

Table 3. Chlorophyll-a (Chl-a) MLR Coefficients of the Four Lake Champlain MLR Models (1992 ~ 2009)

MLR model#		Unstandardized Coefficients		Standardized Coefficients		t	Sig.	95% Confidence Interval	
		B	Std. Error	Beta				Lower Bound	Upper Bound
1	(Constant)	1.800	0.250			7.211	0.000	1.308	2.292
	TP(μg/L)	0.173	0.010	0.766		18.172	0.000	0.154	0.192
2	(Constant)	5.484	0.600			9.137	0.000	4.301	6.666
	TP(μg/L)	0.172	0.009	0.760		19.607	0.000	0.154	0.189
	Cl(μg/L)	0.002	0.000	-0.257		-6.641	0.000	0.000	0.000
	(Constant)	8.417	1.294			6.503	0.000	5.867	10.967
3	TP(μg/L)	0.128	0.019	0.566		6.650	0.000	0.090	0.166
	Cl(μg/L)	0.003	0.000	-0.279		-7.112	0.000	0.000	0.000
	Secchi(m)	-0.431	0.169	-0.218		-2.550	0.011	-0.765	-0.098
	(Constant)	6.646	1.567			4.241	0.000	3.558	9.734
	TP(μg/L)	0.111	0.021	0.490		5.277	0.000	0.069	0.152
4	Cl(μg/L)	0.003	0.000	-0.243		-5.635	0.000	0.000	0.000
	Secchi(m)	-0.436	0.168	-0.220		-2.593	0.010	-0.767	-0.105
	TN(μg/L)	0.004	0.002	0.112		1.978	0.049	0.000	0.007

Table 4. ANOVA Analysis of the Four Lake Champlain MLR Models (1992 ~ 2009)

MLR model#		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1422.661	1.000	1422.661	330.227	0.000
	Residual	999.486	232.000	4.308		
	Total	2422.147	233.000			
2	Regression	1582.888	2.000	791.444	217.839	0.000
	Residual	839.259	231.000	3.633		
	Total	2422.147	233.000			
3	Regression	1605.958	3.000	535.319	150.852	0.000
	Residual	816.189	230.000	3.549		
	Total	2422.147	233.000			
4	Regression	1619.669	4.000	404.917	115.550	0.000
	Residual	802.478	229.000	3.504		
	Total	2422.147	233.000			

-0.874). The analysis revealed the less significant trend towards a medium negative linear relationship between *Secchi* and *TN* ($R = -0.544$). These results suggested that the following MLR and NN modeling approaches should be successful if the levels of *TP* and *TN* are assumed to indeed be positively linearly related to Chl-a, and, conversely, *Secchi* is negatively linearly related to Chl-a.

3.2. Multiple Linear Regression (MLR) Models

3.2.1. Training Processes of MLR Models

The IBM SPSS multiple linear regression (MLR) tool was used to explore the MLR models using Lake Champlain monitoring station data from 1992 to 2016. Early yearly data (1992 ~ 2009) were used to create the models, while the later yearly data (2010 ~ 2016) were used to verify the models. The statistical software package was used to analyze eutrophication levels using the later yearly Lake Champlain dataset as the model training dataset, with the critical value for the models being 95% and

the error range being from -1.96 to 1.96 (models with errors outside this range were rejected). The stepwise selection feature in the IBM SPSS software was used to determine whether each independent variable (i.e., *TP*, *TN*, *Cl*, *Secchi*, *TempC*, *RegAlk* and *Depth*) was significant for the dependent variable (Chl-a) in the model. In the stepwise method, insignificant variables were excluded from the models. The results of this modeling analysis are shown in Table 2 and the Chl-a MLR coefficients of the four Lake Champlain MLR models are presented in Table 3.

The analysis of variance (ANOVA) provided a method to compare the different models. Table 4 shows the results of the ANOVA, which indicated that although all four MLR models found were significant ($\text{Sig.} = 0$), the last model (MLR model #4) had the minimum error mean square, meaning that of the all models, it provided the most accurate prediction. The maximum Pearson product-moment correlation coefficient of MLR model #4 ($R = 0.818$) indicated the strongest linear relationship between the variables. In addition, the maximum determination

coefficient of MLR model #4 ($R^2 = 0.669$) indicated the linear model had a good fit. The minimum standard error of MLR model #4 (error = 1.872) indicated that the error of the sample mean with respect to true mean was slight low, which showed the linear regression model provided accurate estimation. The minimum confidence interval (CI) for the coefficients of MLR model #4 indicated that the error limit boundary for accepting or rejecting a model was small for all the variables, which further demonstrated that MLR model #4 was the most accurate MLR model.

The results of MLR model #4 showed that four parameters (i.e., *TP*, *Cl*, *Secchi* and *TN*) were significant predictors of Chl-a concentrations over the early period (1992 ~ 2009), and thus of CAB growth. The parameters *TP* and *TN* were positively correlated with the Chl-a concentrations, while *Secchi* was negatively correlated with the Chl-a concentrations. These results supported the results of the previous correlation analysis and are also in agreement with other scientific literature (Dimberg et al., 2013; Cha and Stow, 2014; Bhagowati and Ahamad, 2018). Moreover, the results of MLR model #4 revealed that *Cl* was the fourth new significant predictor, showing a negative relationship with Chl-a concentrations. Although this result was not obtained by the statistical analysis above, it was in agreement with other scientific literature (Shillito and De Marle, 1992).

MLR models #1 to #4 indicated that *TP* was the most significant predictor of Chl-a concentrations and therefore CAB growth, which is supported by most of the scientific literature

(Correll, 1998; Brown et al., 2000; Vinçon-Leite and Casenave, 2019). Four parameters (*TP*, *TN*, *Secchi*, *Cl*) were significant predictors of Chl-a concentrations over the entire time period of 24 years. These useful predictors necessitate testing with historical data. Therefore, MLR model #4 was chosen for further verification studies and its equation is shown below:

$$\text{Chl-a} = 6.646 + 0.111 \times TP + 0.004 \times TN - 0.436 \times \text{Secchi} + 0.0003 \times Cl \quad (4)$$

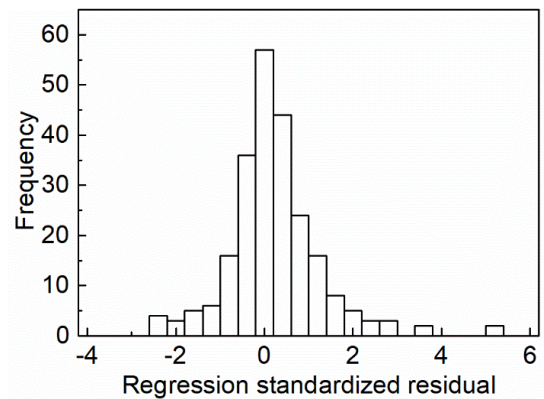


Figure 5. MLR-Model #4 standard error distribution (Mean = -1.65×10^{-15} , standard deviation = 0.991, and sample number = 234).

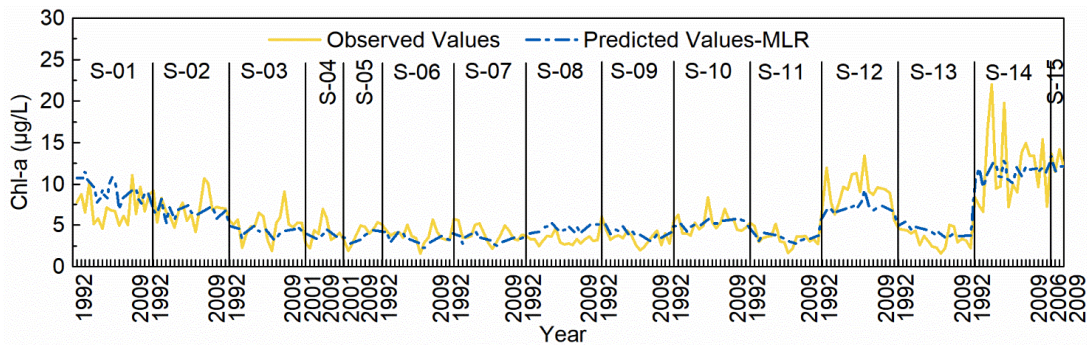


Figure 6. Lake Champlain chlorophyll-a levels by year (1992 ~ 2009), MLR model $R^2 = 0.669$, Observed Chl-a values (solid yellow line) compared to predicted Chl-a values using the MLR model (dotted blue line).

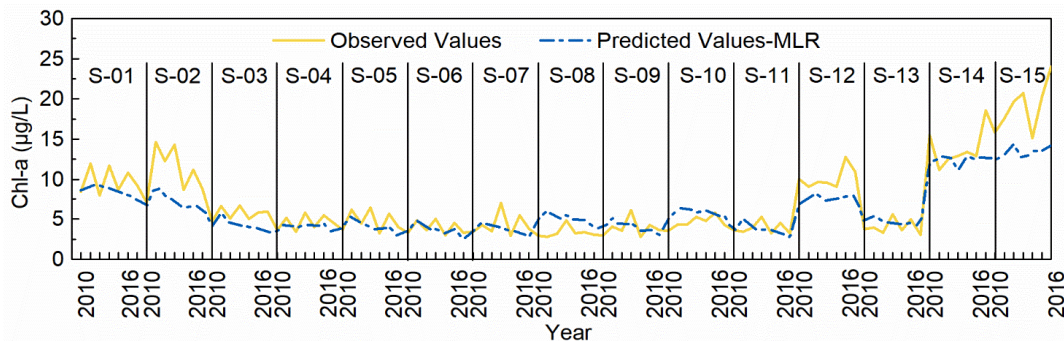


Figure 7. Lake Champlain chlorophyll-a levels by year (2010 ~ 2016), MLR model $R^2 = 0.830$, Observed Chl-a values (solid yellow line) compared to predicted Chl-a values using the MLR model (dotted blue line).

3.2.2. Verification Processes of MLR model #4

As explained above, yearly Lake Champlain water quality monitoring data for the period from 2010 to 2016 were used for the verification of the most accurate model, MLR model #4. The examination of the standard error distribution histogram in Figure 5 revealed that the model errors had a uniform distribution with few errors outside of the standard distribution range, indicating good verification results. The verification data set produced an $R^2 = 0.830$, which was higher than the $R^2 = 0.669$ obtained from the MLR training data set. The predicted Chl-a values of MLR model #4 aligned well with the observed values from both the training and verification processes (Figures 6 and 7), especially when concentrations were lower than 7.5 $\mu\text{g/L}$. However, most of MLR model #4's prediction errors were obtained from Lake Champlain monitoring stations 01, 02, 14 and 15, where most concentrations were higher than 7.5 $\mu\text{g/L}$. As shown in Figures 6 and 7, these errors were not random. Marshall and Peters (1989) indicated that eutrophic lakes have higher Chl-a variability than oligotrophic lakes. The nutrients levels in these four stations are higher than other stations, which may affect the accuracy of the modeling results.

3.3 Neural Network (NN) Model

3.3.1 Training Processes of NN Models

Several studies suggested using neural network (NN) methods to provide effective Chl-a prediction (Coad et al., 2014; Jeong et al., 2015; Xu et al., 2015). The IBM SPSS multilayer perceptron (MLP) NN tool was used to explore NN models using Lake Champlain monitoring station data. Approximately 70% of the early period's (1992 ~ 2009) yearly data were used to create the NN models, while the remaining 30% were used to simultaneously verify each model as it was created. Later yearly data (2010 ~ 2016) were used to verify the NN model. It is noted that running the same data set using NN model could result in a different NN model with different number of runs. Improving the NN results is an iterative and time-consuming process. Previous studies suggested manipulating the number of layers and the ratio between the training and prediction data. After implementing many of the recommended techniques to improve the prediction of the NN model, the best NN model found was employed in this study for the further NN analysis.

Figure 8 displays the results of the NN synaptic weight chart, which represents the feedforward NN Chl-a model architecture as connections flowing forward from the input layer to the output layer without any feedback loops.

The NN importance chart (Figure 9) shows the degree of importance of each independent variable to the network's model-predicted Chl-a value. The results of the analysis showed that the independent variable *TP* (Normalized Importance $r_i = 100\%$) had the greatest effect on how the network classifies Chl-a models, followed by the variables *TN* ($r_i = 57.41\%$), *CI* ($r_i = 47.45\%$) and *Secchi* ($r_i = 36.44\%$), which showed the same result as the analysis of correlation and MLR model #4. Interestingly, the NN analysis suggested that *RegAlk* ($r_i = 34.69\%$) and *TempC* ($r_i = 30.03\%$) were also important. The remaining tested variable (*Depth*) was considered relatively unimportant. These

results are also in agreement with those of the published literature (Karul et al., 1999).

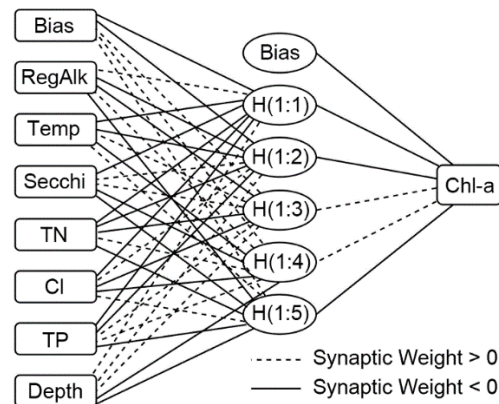


Figure 8. Synaptic weight chart of the NN model for Lake Champlain using early period data (1992 ~ 2009).

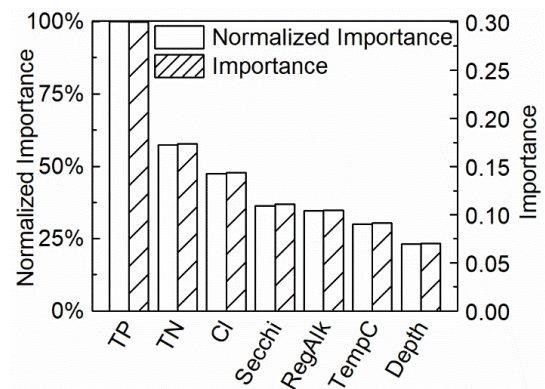


Figure 9. NN analysis water quality variables importance chart.

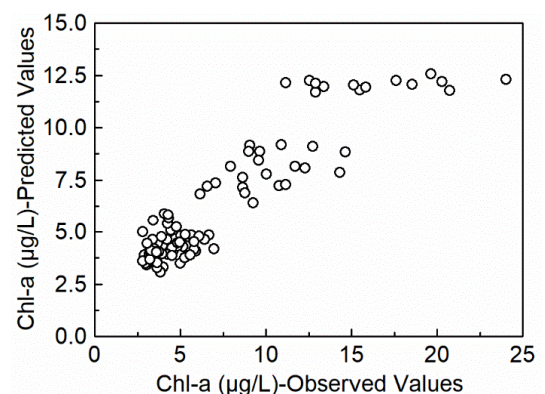


Figure 10. NN model predicted Chl-a values versus observed Chl-a values.

Figure 10 shows the observed Chl-a values compared with the NN model's predicted Chl-a values. It can be seen that the errors between the predicted values of the NN model and the observed values were low when the Chl-a concentration was

lower than 7.5 µg/L. However, the predicted values became less accurate at higher Chl-a concentrations (> 7.5 µg/L). This is in good agreement with the MLR modeling study results.

3.3.2. Verification Processes of NN Models

As explained above, yearly Lake Champlain water quality monitoring data for the early period (2010 ~ 2016) were used for the verification of the NN model. The predicted Chl-a values of the NN model aligned well with the observed values both in the training and verification processes (Figures 11 and 12), especially when the concentration was lower than 7.5 µg/L. The verification dataset produced the R^2 value of 0.845, which is greater than the R^2 value of 0.719 obtained from the NN training data set. In addition, the higher R^2 showed that the NN model has better simulation performance than MLR model #4 in both the training and verification processes. However, at stations 01, 02, 14 and 15, when Chl-a concentrations were higher than 7.5 µg/L, the predicted values of the NN model were relatively less accurate, showing the same results as the MLR model #4. The nutrients levels in these four stations are higher than other stations which may lead to the relatively lower modeling accuracy. Marshall and Peters (1989) also indicated that eutrophic lakes have higher Chl-a variability than oligotrophic lakes.

3.4. MLR and NN Models for Chl-a Prediction

Both the developed #4 MLR model and the NN model were applied to predict the Chl-a levels of Lake Champlain over the next five years (2017 ~ 2021). Firstly, the dataset of seven independent parameters (i.e., *TP*, *TN*, *Cl*, *Secchi*, *TempC*, *RegAlk* and *Depth*) over the period from 2018 to 2021 were predicted using the EXCEL prediction tool based on the parameters' respective changing trends. In particular, the 2017 temperature parameter dataset was lacking, while the datasets of the six other independent parameters and Chl-a were able to be collected. Only the temperature data in 2017 were predicted; the Chl-a concentration data in 2017 were used here as validation indicators for the prediction processes.

The predicted values of the Chl-a levels in Lake Champlain by year (2017 ~ 2021) are shown in Figure 13. The predicted values and observed values align well with each other in stations where Chl-a concentrations were lower than 7.5 µg/L, while the prediction values were less accurate for stations with higher concentrations (> 7.5 µg/L). In addition, the prediction results of the developed MLR #4 and NN models showed similar trends and magnitudes in most stations. The predicted values of MLR model #4 showed more obvious fluctuations than the predicted values of the NN model in stations 01, 14 and 15. Moreover, MLR model #4 produced a better prediction result ($R^2 = 0.845$) than the NN model ($R^2 = 0.837$) regarding the Chl-a levels in Lake Champlain from 2017 to 2021, although the differences were small.

3.5. Discussion

The algal conditions caused by excessive phosphorus levels in Missisquoi Bay and the South Lake of Lake Champlain

normally last for much of the year, especially in summer. Due to the detrimental effects of eutrophication problems in Lake Champlain, related research has increased in recent years. Parts of Lake Champlain are eutrophic, resulting in the excessive growth of aquatic plants and unpleasant and potentially toxic seasonal algal blooms (Jokela et al., 2004). Several studies have been published concerning Lake Champlain. For example, in 1989, a group of scientists from the Vermont Department of Environmental Conservation published a comprehensive study of Lake Champlain and concluded that it would be unrealistic to use daily data for Lake Champlain to detect emerging lake eutrophication problems (Smeltzer et al., 1989). In 1997, satellite images of the watershed were used to estimate the proportions of attributed nonpoint source loads (Millette, 1997), and in 2009 a Danish study suggested that eutrophication in Lake Champlain was also affected by changes in climate (Jeppesen et al., 2009). In addition, in 2012, sewer installation, population expansion and agricultural and urban development were found to be the main causes of the eutrophication problems in Lake Champlain (Levine et al., 2012).

According to the correlation analysis, *TP* has the most linear relationship with Chl-a (positive), followed by *Secchi* (negative), *TN* (positive) and *Depth* (negative). *TP* and *TN* were needed for CAB growth, while the parameters *Depth* and *Secchi* were reduced by the growth of CABs, which could explain the results of the correlation analysis.

The determination coefficients R^2 of MLR model #4 for the training and verification processes were respectively 0.669 and 0.830. The determination coefficients R^2 of the NN model for both the training (0.719) and verification (0.845) processes were greater than those of MLR model #4, which showed a better ability to predict algal growth in Lake Champlain. In a similar study, Maier et al. (2010) also reported that NN provided better results than the MLR Models, especially for simulating both linear and nonlinear systems. First the correlations between the lake water quality parameters, such as *TP* and *TN*, may be associated with multicollinearity issue. Additionally, input parameters for MLR may have high complex nonlinear relationship with Chl-a, expecting that MLR alone may have less prediction accuracy.

The predicted results of both MLR and the NN models for the Chl-a levels in Lake Champlain from 2017 to 2021 were closer to the observed values in stations with low to moderate levels of Chl-a (< 7.5 µg/L). The determination coefficients R^2 of MLR model #4 and the NN model were respectively 0.845 and 0.837, indicating that MLR model #4 had relatively better prediction results than the NN model. The future prediction for the period from 2018 to 2021 can be further improved when more data including more field data are available for the seven input parameters (i.e., *TP*, *TN*, *Cl*, *Secchi*, *TempC*, *RegAlk* and *Depth*).

For stations 01, 02, 14 and 15 in the narrow north and south ends of Lake Champlain, high levels of Chl-a (> 7.5 µg/L) were recorded. For these south and north ends of the lake, the results of training, validation and prediction processes of both the MLR and NN models were not close to the observed values compared to other monitoring stations. This could be based on the follows: (1) the southern monitoring stations 01 and 02 are in

shallow water at the mouth of the Poultney river, which contributes a large amount of nutrients, particularly *Cl*, to the lake, (2) farms and agricultural land may contribute fertilizers with high *TV* concentrations to the northern area of the lake including monitoring stations 14 and 15, (3) Marshall and Peters (1989) indicated that eutrophic lakes had a higher variability of Chl-a than oligotrophic lakes. Most stations in the Lake Champlain are with low to moderate levels of Chl-a ($< 7.5 \mu\text{g/L}$), where both MLR and NN have much better accuracy except the southern and

northern end areas with high water nutrient concentrations ($> 7.5 \mu\text{g/L}$), and (4) the complex geometry and hydraulic conditions in the narrow northern and southern end areas may further affect the MLR and NN modeling results, better physical models considering complex near shore mixing can be used for such small mixing zones; or hybrid PCR-NN methods that have both linear and nonlinear modeling capabilities could be other alternative methods for these small complex near shore mixing zones (Jeong et al., 2015).

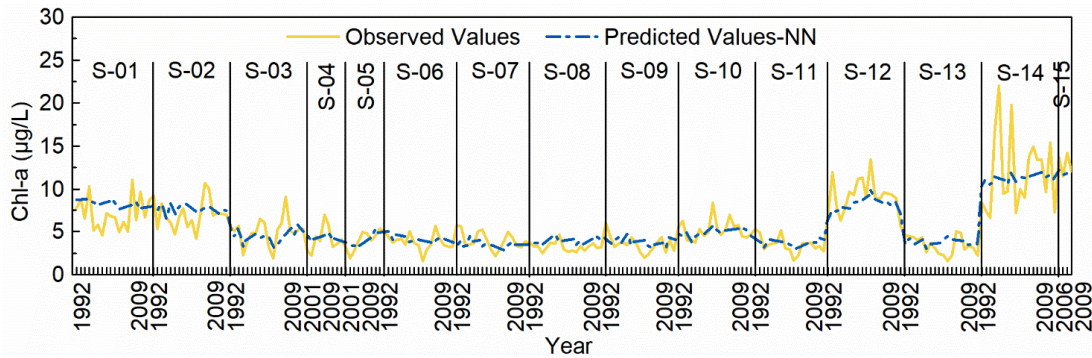


Figure 11. Lake Champlain chlorophyll-a levels by year (1992 ~ 2009), NN model $R^2 = 0.719$, Observed Chl-a values (solid yellow line) compared to predicted Chl-a values using the MLR model (dotted blue line).

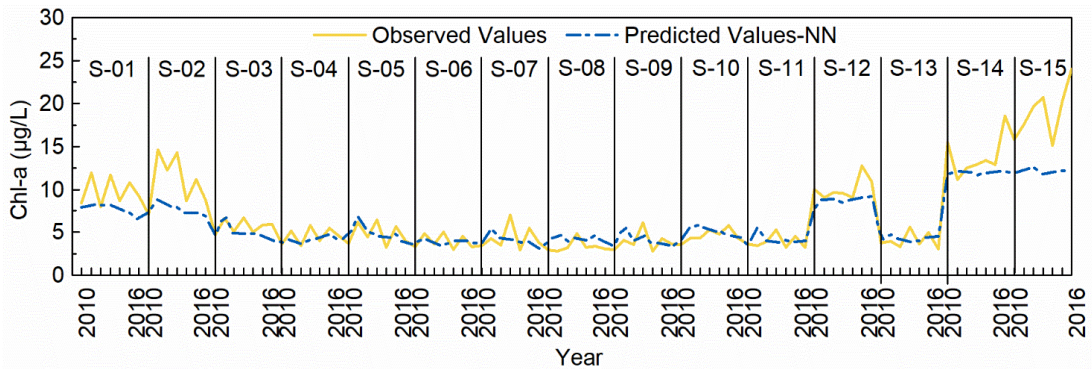


Figure 12. Lake Champlain chlorophyll-a levels by year (2010 ~ 2016), NN model $R^2 = 0.845$, Observed Chl-a values (solid yellow line) compared to predicted Chl-a values using the MLR model (dotted blue line).

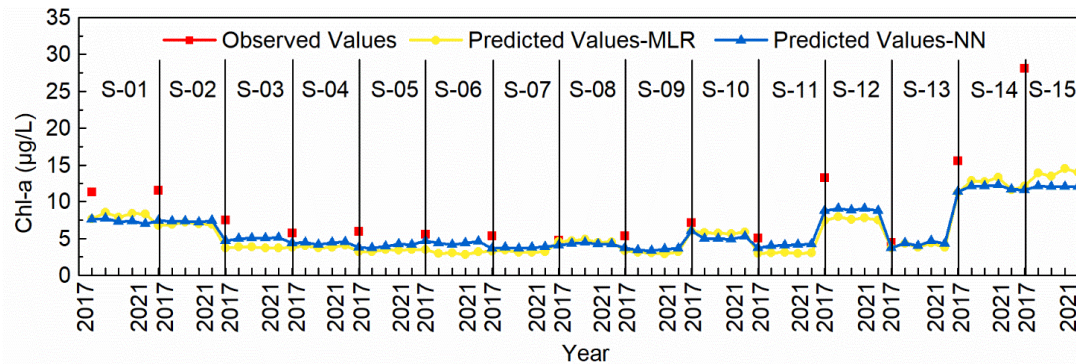


Figure 13. Lake Champlain chlorophyll-a levels by year (2017 ~ 2021), MLR model $R^2 = 0.845$, NN model $R^2 = 0.837$, observed Chl-a values (red point) compared to predicted Chl-a values using the MLR model (yellow line) and NN model (blue line).

4. Conclusions

The parameters of *TP*, *Secchi*, *TN* and *Depth* showed great impact on the concentrations of Chl-a. For most parts of Lake Champlain, where low to moderate levels of Chl-a ($< 7.5 \mu\text{g/L}$) were recorded at the majority of water quality monitoring stations, both MLR and NN models can accurately simulate algal growth. The determination coefficients R^2 of the NN model for both the training (0.719) and verification (0.845) processes were greater than those of MLR model #4, which showed a better ability to predict algal growth in Lake Champlain. Furthermore, the determination coefficients R^2 of MLR model #4 and the NN model were respectively 0.845 and 0.837, which shows that the results predicted by MLR model #4 were superior to those predicted by the NN model for the period from 2017 to 2021 in Lake Champlain. In general, there was no significant difference between MLR model #4 and the NN model in predicting Chl-a concentrations in Lake Champlain. However, for the narrow northern and southern end areas of Lake Champlain with high levels of Chl-a ($> 7.5 \mu\text{g/L}$), including southend water quality monitoring stations 01, 02 and northend stations 14, 15, more monitoring stations with more field data or other modeling methods than MLR and NN are preferable for these small near-shore mixing zones.

Acknowledgments. The authors would like to thank Mr. Eric Smeltzer from the Vermont Department of Environmental Conservation, who provided insightful information on Lake Champlain.

References

- Anderson, D.M., Cembella, A.D., and Hallegraeff, G.M. (2012). Progress in understanding harmful algal blooms: paradigm shifts and new technologies for research, monitoring, and management. *Annual Review of Marine Science*, 4, 143-176. <https://doi.org/10.1146/annurev-marine-120308-081121>
- Barzegar, R., and Moghaddam, A.A. (2016). Combining the advantages of neural networks using the concept of committee machine in the groundwater salinity prediction. *Modeling Earth Systems and Environment*, 2(1), 26. <https://doi.org/10.1007/s408-015-0072-8>
- Bhagowati, B., and Ahamad, K.U. (2018). A review on lake eutrophication dynamics and recent developments in lake modeling. *Ecology & Hydrobiology*, <https://doi.org/10.1016/j.ecohyd.2018.03.002>
- Brown, C.D., Hoyer, M.V., Bachmann, R.W., and Canfield Jr, D.E. (2000). Nutrient-chlorophyll relationships: an evaluation of empirical nutrient-chlorophyll models using Florida and north-temperate lake data. *Canadian Journal of Fisheries and Aquatic Sciences*, 57(8), 1574-1583. <https://doi.org/10.1139/cjfas-57-8-1574>
- Burger, D.F., Hamilton, D.P., and Pilditch, C.A. (2008). Modeling the relative importance of internal and external nutrient loads on water column nutrient concentrations and phytoplankton biomass in a shallow polymictic lake. *Ecological Modelling*, 211(3-4), 411-423. <https://doi.org/10.1016/j.ecolmodel.2007.09.028>
- Çamdevýren, H., Demýr, N., Kanik, A., and Keskýn, S. (2005). Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs. *Ecological Modelling*, 181(4), 581-589. <https://doi.org/10.1016/j.ecolmodel.2004.06.043>
- Cha, Y., and Stow, C.A. (2014). A Bayesian network incorporating observation error to predict phosphorus and chlorophyll a in Saginaw Bay. *Environmental Modelling & Software*, 57, 90-100. <https://doi.org/10.1016/j.envsoft.2014.02.010>
- Chambers, P., Guy, M., Roberts, E., Charlton, M., Kent, R., Gagnon C., Grove G., and Foster N. (2001). *Nutrients and Their Impact on the Canadian Environment*, Agriculture and Agri-Food Canada, Environment Canada, Fisheries and Oceans Canada, Health Canada and Natural Resources Canada, 241.
- Chen, W.B., and Liu, W.C. (2014). Artificial neural network modeling of dissolved oxygen in reservoir. *Environmental Monitoring and Assessment*, 186(2), 1203-1217. <https://doi.org/10.1007/s10661-013-3450-6>
- Coad, P., Cathers, B., Ball, J.E., and Kadluczka, R. (2014). Proactive management of estuarine algal blooms using an automated monitoring buoy coupled with an artificial neural network. *Environmental Modelling & Software*, 61, 393-409. <https://doi.org/10.1016/j.envsoft.2014.07.011>
- Correll, D.L. (1998). The role of phosphorus in the eutrophication of receiving waters: A review. *Journal of Environmental Quality*, 27(2), 261-266. <https://doi.org/10.2134/jeq1998.00472425002700020004x>
- Dillon, P.J., and Rigler, F. (1974). The phosphorus - chlorophyll relationship in lakes. *Limnology and Oceanography*, 19(5), 767-773. <https://doi.org/10.4319/lo.1974.19.5.0767>
- Dimberg, P.H., Hytteborn, J.K., and Bryhn, A.C. (2013). Predicting median monthly chlorophyll-a concentrations. *Limnologia*, 43(3), 169-176. <https://doi.org/10.1016/j.limno.2012.08.011>
- Downing, J.A., and McCauley, E. (1992). The nitrogen: phosphorus relationship in lakes. *Limnology and Oceanography* 37(5), 936-945. <https://doi.org/10.4319/lo.1992.37.5.0936>
- Elliott, J., Persson, I., Thackeray, S., and Blenckner, T. (2007). Phytoplankton modelling of Lake Erken, Sweden by linking the models PROBE and PROTECH. *Ecological Modelling*, 202(3-4), 421-426. <https://doi.org/10.1016/j.ecolmodel.2006.11.004>
- Fornarelli, R., Galelli, S., Castelletti, A., Antenucci, J.P., and Marti, C.L. (2013). An empirical modeling approach to predict and understand phytoplankton dynamics in a reservoir affected by interbasin water transfers. *Water Resources Research*, 49(6), 3626-3641. <https://doi.org/10.1002/wrcr.20268>
- Freeman, A.M., Lamon III, E.C., and Stow, C.A. (2009). Nutrient criteria for lakes, ponds, and reservoirs: a Bayesian TREED model approach. *Ecological Modelling*, 220(5), 630-639. <https://doi.org/10.1016/j.ecolmodel.2008.12.009>
- Gopal, B. (2005). Does inland aquatic biodiversity have a future in Asian developing countries? *Aquatic Biodiversity II*, 69-75. https://doi.org/10.1007/1-4020-4111-X_10
- Hiscock, J.G., Thourot, C.S., and Zhang, J. (2003). Phosphorus budget-land use relationships for the northern Lake Okeechobee watershed, Florida. *Ecological Engineering*, 21(1), 63-74. <https://doi.org/10.1016/j.ecoleng.2003.09.005>
- Huszar, V.L., Caraco, N.F., Roland, F., and Cole, J. (2006). Nutrient-chlorophyll relationships in tropical-subtropical lakes: do temperate models fit? *Biogeochemistry*, 239-250. <https://doi.org/10.1007/s10533-006-9007-9>
- Jeong, I.I., Lou, I., Ung, W.K., and Mok, K.M. (2015). Using principle component regression, artificial neural network, and hybrid models for predicting phytoplankton abundance in Macau storage reservoir. *Environmental Modeling & Assessment*, 20(4), 355-365. <https://doi.org/10.1007/s10666-014-9433-3>
- Imboden, D.M. (1974). Phosphorus model of lake eutrophication. *Limnology and Oceanography* 19(2), 297-304. <https://doi.org/10.4319/lo.1974.19.2.0297>
- Jeppesen, E., Kronvang, B., Meerhoff, M., Søndergaard, M., Hansen, K.M. (2009). Climate change effects on runoff, catchment phosphorus loading and lake ecological state, and potential adaptations. *Journal of Environmental Quality*, 38(5), 1930-1941. <https://doi.org/10.2134/jeq2008.0113>
- Jokela, W.E., Clausen, J.C., Meals, D.W., and Sharpley, A.N. (2004).

- Effectiveness of agricultural best management practices in reducing phosphorous loading to Lake Champlain, *Lake Champlain: Partnerships and Research in the New Millennium*, 39-52. https://doi.org/10.1007/978-1-4757-4080-6_4
- Kariya, T. and Kurata, H. (2004). Generalized Least Squares. John Wiley & Sons. <https://doi.org/10.1002/0470866993>
- Karul, C., Soyupak, S., and Yurteri, C. (1999). Neural network models as a management tool in lakes. *Hydrobiologia*, 139-144. <https://doi.org/10.1023/A:1017007313690>
- Kenney, J.F. (2013). *Mathematics of Statistics*. D. Van Nostrand Company Inc, Toronto; Princeton; New Jersey; London; New York; Affiliated East-West Press Pvt-Ltd; New Delhi.
- Khuri, A.I. (2013). Introduction to Linear Regression Analysis, Fifth Edition by Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining. *International Statistical Review*, 81(2), 318-319. https://doi.org/10.1111/insr.12020_10
- Le Moal, M., Gascuel-Oudoux, C., Ménesguen, A., Souchon, Y., Étrillard, C. (2019). Eutrophication: A new wine in an old bottle? *Science of The Total Environment*, 651, 1-11. <https://doi.org/10.1016/j.scitotenv.2018.09.139>
- Levine, S.N., Lini, A., Ostrofsky, M.L., Bunting, L., Burgess, H. (2012). The eutrophication of Lake Champlain's northeastern arm: Insights from paleolimnological analyses. *Journal of Great Lakes Research*, 38, 35-48. <https://doi.org/10.1016/j.jglr.2011.07.007>
- Libes, S. (2011). *Introduction to Marine Biogeochemistry*, Academic Press.
- Maier, H.R., Jain, A., Dandy, G.C., and Sudheer, K.P. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *Environmental Modelling & Software*, 25(8), 891-909. <https://doi.org/10.1016/j.envsoft.2010.02.003>
- Marshall, C.T., and Peters, R.H. (1989). General patterns in the seasonal development of chlorophyll a for temperate lakes. *Limnology and Oceanography*, 34(5), 856-867. <https://doi.org/10.4319/lo.1989.34.5.0856>
- Matus-Hernández, M.Á., Hernández-Saavedra, N.Y., and Martínez-Rincón, R.O. (2018). Predictive performance of regression models to estimate Chlorophyll-a concentration based on Landsat imagery. *PLoS One*, 13(10), e0205682. <https://doi.org/10.1371/journal.pone.0205682>
- Millette, T.L. (1997). *Development of Land Cover/Land Use Geographic Information System Data Layer for the Lake Champlain Basin and Vermont Northern Forest Lands Project Areas*, LCBP Technical Report #24, Lake Champlain Basin Program.
- Nasrabadi, N.M. (2007). Pattern Recognition and Machine Learning. *Journal of Electronic Imaging*, 16(4), 049901. <https://doi.org/10.1117/1.2819119>
- New York State Department of Environmental Conservation (DEC) and Vermont Department of Environmental Conservation (DEC). (2012). *Long-Term Water Quality and Biological Monitoring Project for Lake Champlain*, Grand Isle, VT., Lake Champlain Basin Program.
- Pomati, F., Matthews, B., Seehausen, O., and Ibelings, B.W. (2017). Eutrophication and climate warming alter spatial (depth) co-occurrence patterns of lake phytoplankton assemblages. *Hydrobiologia*, 787(1), 375-385. <https://doi.org/10.1007/s10750-016-2981-6>
- Ranković, V., Radulović, J., Radojević, I., Ostojić, A., and Čomić, L. (2010). Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia. *Ecological Modelling*, 221(8), 1239-1244. <https://doi.org/10.1016/j.ecolmodel.2009.12.023>
- Reckhow, K.H. (1993). A random coefficient model for chlorophyll-nutrient relationships in lakes. *Ecological Modelling*, 70(1-2), 35-50. [https://doi.org/10.1016/0304-3800\(93\)90071-Y](https://doi.org/10.1016/0304-3800(93)90071-Y)
- Rui, X., Zhi, C., and Yun, Z. (2012). Impact Assessment of Climate Change on Algal Blooms by a Parametric Modeling Study in Han River. *Journal of Resources and Ecology*, 3(3), 209-219. <https://doi.org/10.5814/j.issn.1674-764x.2012.03.003>
- Shillito, M.L., and De Marle, D.J. (1992). *Value: Its Measurement, Design, and Management*, John Wiley & Sons.
- Smeltzer, E., Shambaugh, A.d., and Stangel, P. (2012). Environmental change in Lake Champlain revealed by long-term monitoring. *Journal of Great Lakes Research*, 38, 6-18. <https://doi.org/10.1016/j.jglr.2012.01.002>
- Smeltzer, E., Walker, W.W., and Garrison, V. (1989). Eleven years of lake eutrophication monitoring in Vermont: a critical evaluation, *Enhancing States' Lake Management Programs*, US Environmental Protection Agency and North American Lake Management Society, 53-62.
- Smith, V.H., Joye, S.B., and Howarth, R.W. (2006). Eutrophication of freshwater and marine ecosystems. *Limnology and Oceanography*, 51(1part2), 351-355. https://doi.org/10.4319/lo.2006.51.1_part_2.0351
- Smith, V.H., Tilman, G.D., and Nekola, J.C. (1999). Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environmental Pollution*, 100(1-3), 179-196. [https://doi.org/10.1016/S0269-7491\(99\)00091-3](https://doi.org/10.1016/S0269-7491(99)00091-3)
- Thomann, R.V., Winfield, R.P., Segna, J.J., Laboratory, E.R., and Station, L.L.R. (1979). *Verification analysis of Lake Ontario and Rochester Embayment Three Dimensional Eutrophication Models*, Environmental Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency.
- Van Huet, H.J.W.J. (1992). Phosphorus eutrophication in the SW Frisian lake district I. Monitoring and assessment of a dynamic mass balance model. *Hydrobiologia*, 233(1), 259-270. <https://doi.org/10.1007/BF00016114>
- Vinçon-Leite, B., and Casenave, C. (2019). Modelling eutrophication in lake ecosystems: A review. *Science of The Total Environment*, 651, 2985-3001. <https://doi.org/10.1016/j.scitotenv.2018.09.320>
- Vollenweider, R.A. (1975). Input-output models. *Schweizerische Zeitschrift für Hydrologie*, 37(1), 53-84. <https://doi.org/10.1007/BF02505178>
- Xu, Y., Ma, C., Liu, Q., Xi, B., Qian, G., Zhang, D., Huo, S. (2015). Method to predict key factors affecting lake eutrophication - A new approach based on Support Vector Regression model. *International Biodeterioration & Biodegradation*, 102, 308-315. <https://doi.org/10.1016/j.ibiod.2015.02.013>
- Yao, X., Zhang, Y., Zhang, L., and Zhou, Y. (2018). A bibliometric review of nitrogen research in eutrophic lakes and reservoirs. *Journal of Environmental Sciences*, 66, 274-285. <https://doi.org/10.1016/j.jes.2016.10.022>