Exploring the Relationship Between Chlorophyll-A and Other Water Quality Parameters by Using Artificial Neural Network Models:

A Case Study of Lake Erie

By Xue (Ashley) Hu

Supervisors: Les Lavkulich, Hanspeter Schreier

Master of Land and Water Systems

Faculty of Land and Food Systems

University of British Columbia

EXECUTIVE SUMMARY

This study used a machine learning approach to model the water qualities of Lake Erie, Ontario, Canada. The data used in the modelling was obtained from the Environment and Climate Change Canada Agency for Lake Erie between 2000 and 2018 and included chlorophyll-a (CHLA), the dissolved oxygen (DO), total phosphorus (TP), total nitrogen (TN), temperature (T), pH, and depth. Several neural network (NN) models were selected for the data analysis, including the standard Neural Network (NN) model, the Simple Recurrent Neural Network (SRN) model, the Back Propagation Neural Network (BPNN) model and the Jump Connection Neural Network (JCNN) model. CHLA was selected as the key water quality indicators for eutrophication in Lake Erie. The above artificial neural network models were assembled. This study showed that the ANN ensemble model predicted the water quality of Lake Erie in a more timely and accurate manner, which aids in facilitating conventional water quality monitoring and reducing the risk of eutrophication in Lake Erie.

Catalogue

1 Introduction

1.1 Background

Eutrophication of lakes is one of the main factors that affect lake water quality (Canada. Environment and Climate Change Canada, & Canadian Government EBook Collection, 2018). It reduces the transparency of the water, making it difficult for sunlight to penetrate the water layer, negatively affects the photosynthesis of plants in the water and potentially cause an oversaturation of dissolved oxygen. Oversaturation of dissolved oxygen and low dissolved oxygen in water can be harmful to aquatic organisms and cause a large number of fish deaths. At the same time, because of the eutrophication of the water body, large numbers of blue algae and green algae may grow on the surface of the water body and form a layer of green scum over the water body. This will cause the organic matter that accumulates at the bottom to decompose under anaerobic conditions and produce harmful gases and plankton to produce biological toxins (Sayers, M. J. et al., 2019). Also, because the water is eutrophic lakes contains nitrates and nitrites, human beings and animals living in the surrounding areas can be poisoned if they drink the polluted water. In addition, after the green scum is formed, the deepwater algae will utilize the oxygen in the water, because they are not exposed to the sun, and cannot perform photosynthesis. The oxygen in the water will gradually decrease and the organisms in the water will die due to insufficient oxygen. The dead algae and organisms will oxidize in the water. At this time, the water body will become odorous and the water resource will be further contaminated. The blue algae outbreak will also have a negative impact on residents and tourists enjoying boating, swimming, and visiting the landscape around the lake (Basile, S. J. et al., 2017).

Algae contain chlorophyll a (CHLA), which produces nutrients that sustain life through photosynthesis. Therefore, in research, people often use the chlorophyll content in water to reflect the number of algae (Meyer, K. A. et al., 2017). It is necessary to apply more accurate methods to monitor the water quality of Lake Erie because monitoring the chlorophyll content in the lake water to prevent the algae blooms is the key to protecting the water quality of Lake Erie (Fraser & desLibris, 2008).

With the increasingly prominent ecological and environmental problems in lakes, the need for monitoring lake water quality is stricter than before (Millie, D. F. et al., 2014). Traditional water quality sampling is done manually, which is time-consuming and laborious, and it is difficult to consider water quality issues at different depths. When current researchers analyze the water quality of Lake Erie, they cannot quickly analyze the entire lake area. Also, because of the limited number of sampling points and low sampling frequency, continuous monitoring is not possible. Artificial Neural Network (ANN) as a popular machine learning algorithm has a high potential in predicting complex relationships, so it has been increasingly used in environmental modelling (Piasecki, A. et al., 2015). After many experiments, it has been shown that the fitting of the ANN model is significantly better than other models, such as regression models and mechanical models. However, during the running of the ANN model, the data will be divided randomly, which makes the ANN model highly likely to have different performance outputs during repeated simulations (Saber, A. et al., 2020). For example, after a researcher obtains an ideal simulation result when modelling, it is difficult for others to replicate the simulation results.

In order to enable people to obtain similar simulation results when using the ANN model to repeat the simulation and to make it stable and repeatable, the artificial neural network-integrated model needs to be applied to the process of water quality monitoring (Li, Y. L. et al., 2015). Applying integrated simulation to the ANN model can improve its model fitting results in environmental modelling. Firstly, by using multiple ANN models to train the data and combining these ANN models into an ANN integrated model, people can obtain a more stable and accurate model fitting result than a single ANN model. The integrated simulation of an artificial neural network model is highly valued because of its utility (Lu, F. et al., 2016).

1.2 Aims and objective

This study assesses machine learning methods to model the water quality of Lake Erie, Ontario, Canada. Four artificial neural network (ANN) models were selected for the research, including the standard Neural Network (NN) model, the Simple Recurrent Neural Network (SRN) model, the Back Propagation Neural Network (BPNN) model and the Jump Connection Neural Network (JCNN) model. The research used integrated technology to support the implementation of integrated simulation. This method was compared to traditional water quality detection methods and a single NN model, and the ANN model that was most suitable for predicting the chlorophyll-a content of Lake Erie was identified. The artificial neural network (ANN) integrated model quickly and accurately predicts the water quality of Lake Erie (Altunkaynak & Altunkaynak, 2007). For example, when there were enough input variables and chlorophyll-a is used as the output variable, the artificial neural network model will automatically generate codes and formulas that can be used to predict chlorophyll-a after training. People can use this code and enough input variables to predict the chlorophyll-a content in the water. It has a positive effect on promoting traditional water quality monitoring and reducing the risk of eutrophication in Lake Erie. This research proposes a new method to retrieve the optical activity parameters and non-optical activity parameters of water bodies and provide new strategies for water quality monitoring. The results of the study provide local government departments and companies with new methods to monitor water quality, which will have a positive application in preventing lake water pollution. It helps provide education for local residents and tourists.

1.3 Study area

Lake Erie (Figure 1.1, 1.2 & 1.3) is one of the five largest lakes in North America and the thirteenth largest lake in the world (May, C. J. et al., 2020). Lake Erie has an area of 25,700 square kilometres, an average depth of 19 meters, and a water storage capacity of 483 cubic kilometres (Mekonnen & Gorsevski, 2015). Lake Erie straddles the United States and Canada and is oriented from southwest to northeast direction. The southern shore of the lake is in Ohio, Pennsylvania, and New York; the west shore is Michigan; and the north shore is Ontario, Canada. The Detroit River connects Lake Huron with Lake Erie. The Niagara River passes through Niagara Falls to connect the lake into Ontario. The total length of the lakeshore is 1,200 kilometres. There are islands in the lake, concentrated at the western, with Pelee Island in Canada as the largest (Skwor, T. et al., 2014). The lakeshore has industry development and lake water pollution is serious, which has caused many lakeside tourist areas to close.

Figure 1.1 Satellite photo of the Great Lakes (source: NASA)

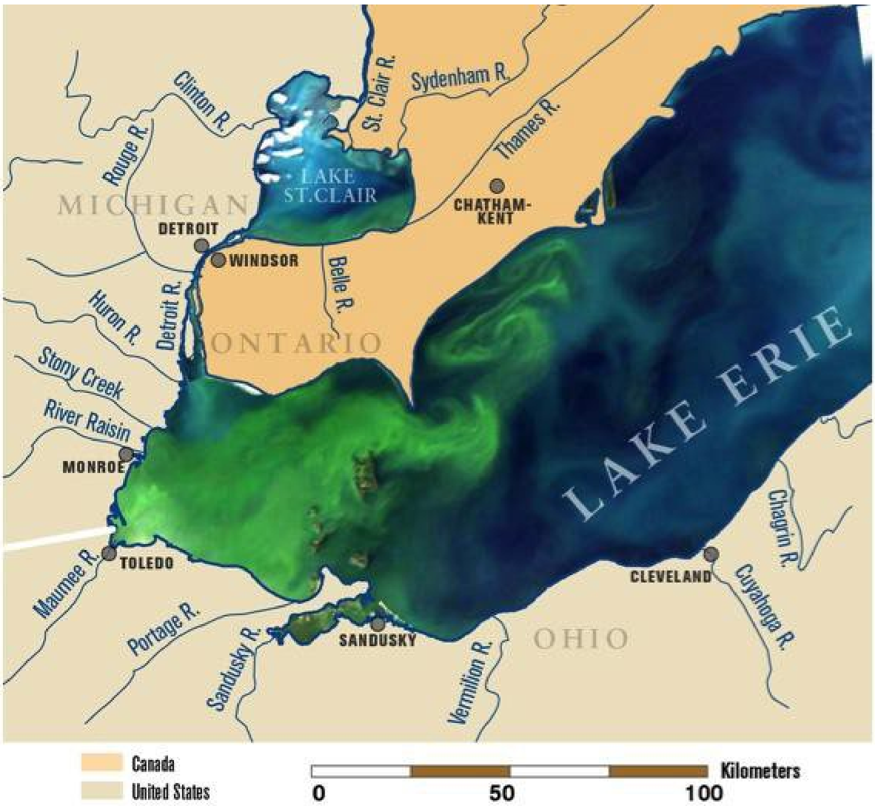


Figure 1.2 map of Lake Eire (U.S. Congress, 1898)

Figure 1.3 map of algae bloom in Lake Erie (U.S. Congress, 1898)

Since the 1950s, the nutrient load of Lake Erie has increased, and the degree of eutrophication has increased year by year (Tian, D. et al., 2017). In the mid to late 1960s, blue algae blooms appeared seasonally in the western bay of Lake Erie, mainly composed of Anabaena, Aphanizomenon, and Microcystis. Cladophora also appeared in large numbers at this time. In the 1970s, water blooms appeared every year, mainly Aphanizomenon flos-aquae. There was no large-scale water bloom in the early 1980s, but the water blooms in the middle and late stages were mainly fascicularia (Zhou, Y. et al., 2013). At this time, the exotic species zebra mussels and quagga mussels began to filter algae. In 1995 and 1998, large-scale water blooms dominated by Microcystis appeared in Lake Erie. Especially in September 1995, green paint-like blooms covered the entire western lake area of Lake Erie. Since 2000, microcystis blooms have frequently appeared. In August 2003, the microcystis bloom lasted for a month in the western lake area of the lake, the worst ever (Mou, X. et al., 2013). From July to August 2014, large-scale microcystis blooms reappeared. The concentration of microcystins in the tap water of Toledo, Ohio exceeded drinking water standards, causing a drinking water crisis for more than 400,000 people. Ohio Governor John Kasich declared a state of emergency in some areas (Wynne & Stumpf, 2015).

The climatic conditions of Lake Erie are conducive to algae reproduction. In July 2019, the wind near Lake Erie was very weak and the airflow was slow. The algae in Lake Erie multiply on the surface of the lake and are not dispersed by the wind. In August, the strong wind mixed the surface algae to a relatively deeper depth and increased growth. As the algae multiplied, the nutrients in the water were consumed. At this time, the precipitation runoff in August brings nutrients from the farmland runoff around Lake Erie to the lake. This accelerates the blue algae bloom. In ten days, the area covered by blue algae can more than double. If these algae were accidentally eaten by humans, liver damage, numbness, dizziness and vomiting would occur. For this reason, many outdoor activities around Lake Erie were temporarily closed. The government also advised people to stay away from waters affected by the dense blue algae (Tian, D. et al., 2017).

Therefore, it is necessary to improve the water quality monitoring methods to detect and predict blue algae blooms in Lake Erie.

1.4 Water quality data

The data used in the modelling was obtained from the Environment and Climate Change Canada Agency for Lake Erie between 2000 and 2018, including chlorophyll-a (CHLA), dissolved oxygen (DO), total phosphorus (TP), total nitrogen (TN), temperature (T), pH, and depth.

2 Methods

This study used ArcGIS to draw a map of the distribution of water quality monitoring points in Lake Erie (Figure 2.1). Next, nearly 200,000 data points were obtained from the Environment and Climate Change Canada Agency.

Excel was used to filter out the points with the same monitoring time, longitude, latitude and depth among 200,000 data. In the end, a total of 252 monitoring points were selected. The time, longitude, latitude, chlorophyll-a, dissolved oxygen, total phosphorus, total nitrogen, pH and depth data were summarized in a table.

Figure 2.1 the distribution of water quality monitoring points in Lake Erie

Subsequently, the time, longitude, and latitude were calculated as shown in Table 2.1 (Krasnopolsky, 2018), and all the filtered data were normalized. Normalization is a way to simplify calculations. It transforms the dimensional expression into a dimensionless expression and becomes a scalar. In this way, indicators of different units or magnitudes can be compared. To facilitate the calculation, the data is normalized, that is, the data becomes a decimal between 0 and 1.

Table 2.1

Variable	**Unit**	**Input**
Day		sin(2πday/366) cos(2πday/366)
Hour		sin(hour*2*π/24) cos(hour*2*π/24)
Latitude		sin(lat)
Longitude		sin(lon) cos(lon)
Dissolved oxygen(DO)	mg/L	(DO-min) / (max-min)
Total phosphorus(TP)	mg/L	(TP-min) / (max-min)
total nitrogen(TN)	mg/L	(TN-min) / (max-min)
ph		(pH-min) / (max-min)
Depth	m	(Depth-min) / (max-min)
Temperature(T)	Degree ℃	(T-min) / (max-min)

Variable	**Unit**	**Output**
Chlorophyll-a (CHLA)	μg/L	(CHLA-min) / (max-min)

Then, the processed data, except chlorophyll-a, were used as the input index, and chlorophyll-a was used as the output index (Figure 2.2). The data were trained with four neural network models. Each training would reduce an index with the least impact until only one index remains.

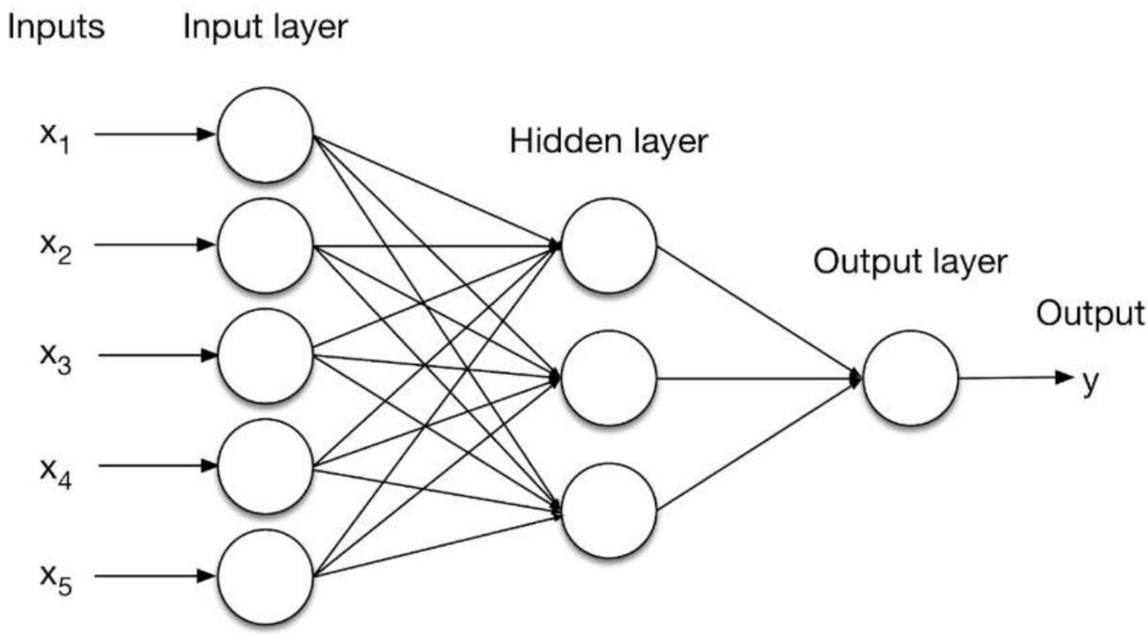


Figure 2.2 Artificial Neural Network

Finally, the data integrated and analyzed the most relevant input parameters and appropriate ANN model with chlorophyll a content (output parameters) based on the results, and draw the final conclusion.

2.1 Neural Network model

2.1.1 Introduction of the neural network model

Artificial neural networks are one of the main tools used in machine learning. Neural networks consist of input and output layers, as well as a hidden layer consisting of units that transform the input into something that the output layer can use (China University of Geosciences, 2019). They are excellent tools for finding patterns.

Learning is an important content of neural network research, and its adaptability is achieved through learning. The process of artificial neural network “learning” was called training (Luo, W. et al., 2019). The artificial neural network has a self-learning function. For example, when implementing image recognition, only input with different image templates and corresponding recognition results into the artificial neural network, and the network will slowly learn to recognize similar images through the self-learning function (Huang & Gao, 2017). The self-learning function is particularly important for prediction. In this project, taking time, longitude, latitude, dissolved oxygen, total phosphorus, total nitrogen, pH, depth and temperature as input variables, and chlorophyll-a as output variables, the artificial neural network can identify them through "learning" the relationship between. After the "learning" is successful, the artificial neural network can derive the code and formula for predicting the output variable (chlorophyll a). When there are enough input variables, this method can predict relatively precisely.

For a basic idea of how a deep learning neural network “learns”, imagine a factory line. Once the raw materials are input, they are passed down a conveyer belt, with each subsequent stop or layer extracts a different set of high-level features (Fantin-Cruz, 2010). If the network is intended to recognize an object, the first layer might analyze the brightness of its pixels. The next layer could then identify any edges in the image, based on lines of similar pixels. After this, another layer may recognize textures and shapes, and so on. By the time the fourth or fifth layer is reached, the deep learning net will have created complex feature detectors. It can determine if these certain image elements are commonly found together, such as a pair of eyes, a nose, and a mouth. Once this is done, the researchers, who have trained the network, can give labels to the output, and then use backpropagation to correct any mistakes which have been made. Along with increasing the number of iteration times, the network can carry out its own classification tasks without needing humans to help every time.

2.1.2 Necessary parameters

R Squared - The coefficient of multiple determination is a statistical indicator usually applied to multiple regression analysis. It compares the accuracy of the model to the accuracy of the benchmark model, wherein the prediction is the mean of all of the samples. A perfect fit would result in an R squared value of 1, a very good fit near 1, and a very poor fit less than 0.

Where

[equations didn't convert over from PDF]

is the actual value

is the predict value of

And is the mean of the values

Correlation Coefficient r - (Pearson's Linear Correlation Coefficient) This is a statistical measure of the strength of the relationship between the actual vs predicted outputs. The *r* coefficient can range from -1 to +1. The closer r is to 1, the stronger the positive linear relationship, and the closer *r* is to -1, the stronger the negative linear relationship. When r is near 0, there is no linear relationship.

r squared - This is the square of the correlation coefficient

where

Where n is the number of patterns, x refers to the set of actual outputs, y refers to the set of predicted outputs

2.2 Input data

Dissolved oxygen (DO), total phosphorus (TP), total nitrogen (TN), temperature (T) and pH

	Unit	Max	Min	Mean	Median	Standard Deviation	Coefficient of Variation
TP	mg/L	0.132	0.005	0.018	0.013	0.016	0.890
Ph		8.740	7.400	8.083	8.080	0.213	0.026
TN	mg/L	0.305	0.006	0.049	0.041	0.033	0.669
DO	mg/L	15.830	9.970	12.992	13.225	1.113	0.086
T	℃	15.300	0.300	6.106	5.900	3.176	0.520
DEPTH	m	20.000	2.000	13.803	14.750	5.994	0.434

2.3 Output data

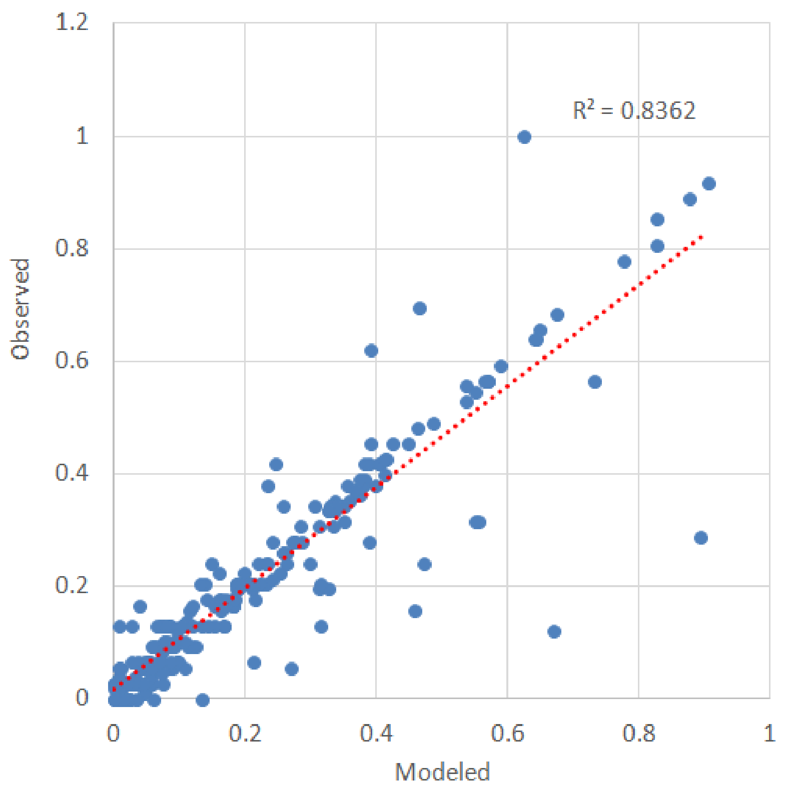
Chlorophyll-a (CHLA)

	Unit	Max	Min	Mean	Median	Standard Deviation	Coefficient of Variation
CHLA	μg/L	10.900	0.100	2.236	1.500	2.069	0.925

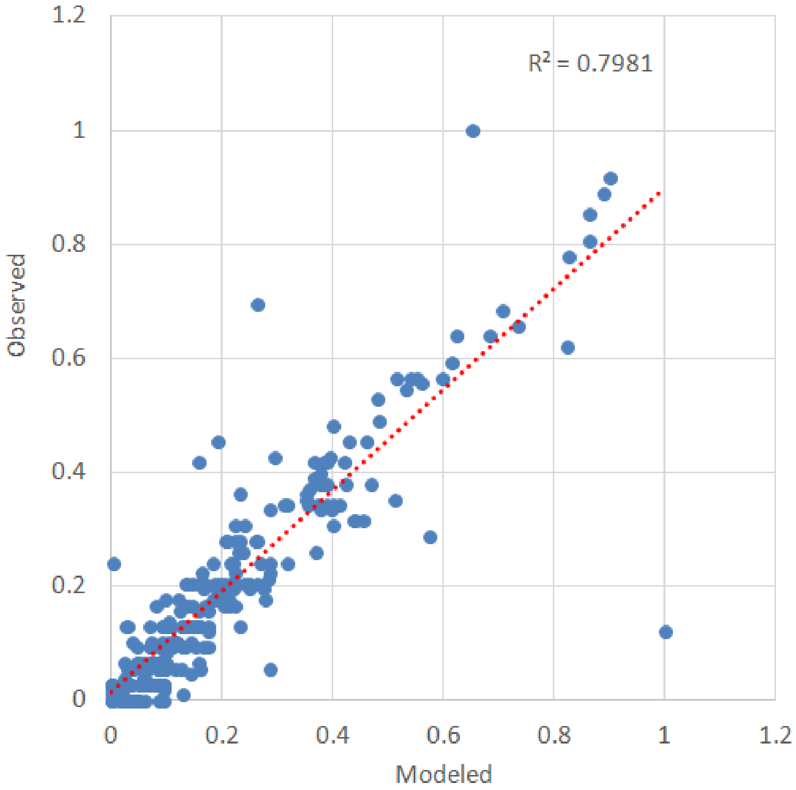
3 THE BODY (Results & Discussion)

In evaluating the simulation effects of the different models, 20% of the data were used as a test set and R Squared and Correlation Coefficient r were used as the evaluation metrics. Several neural network (NN) models were selected for the data analysis, including the standard Neural Network (NN) model, the Simple Recurrent Neural Network (SRN) model, the Back Propagation Neural Network (BPNN) model and the Jump Connection Neural Network (JCNN) model.

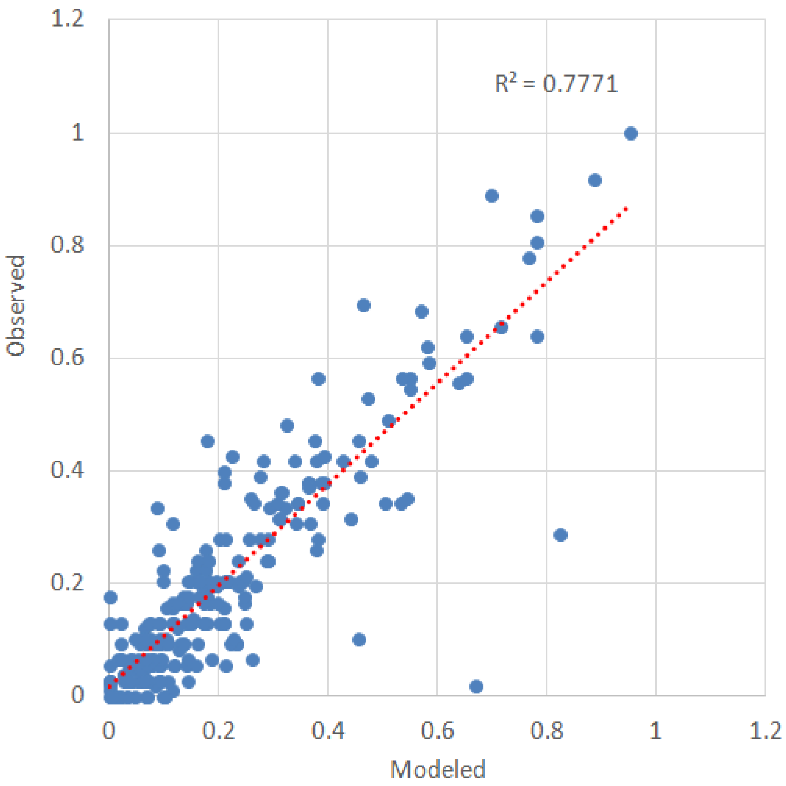
3.1 Standard NN



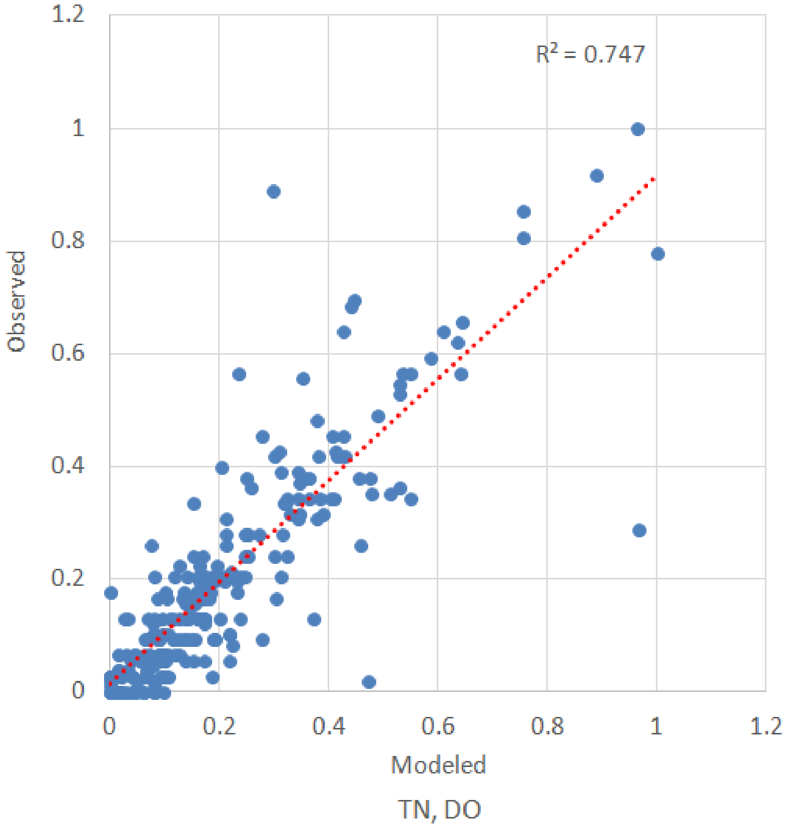
TP, TN, PH, DO, T



TP, TN, PH, DO



TP, TN, DO



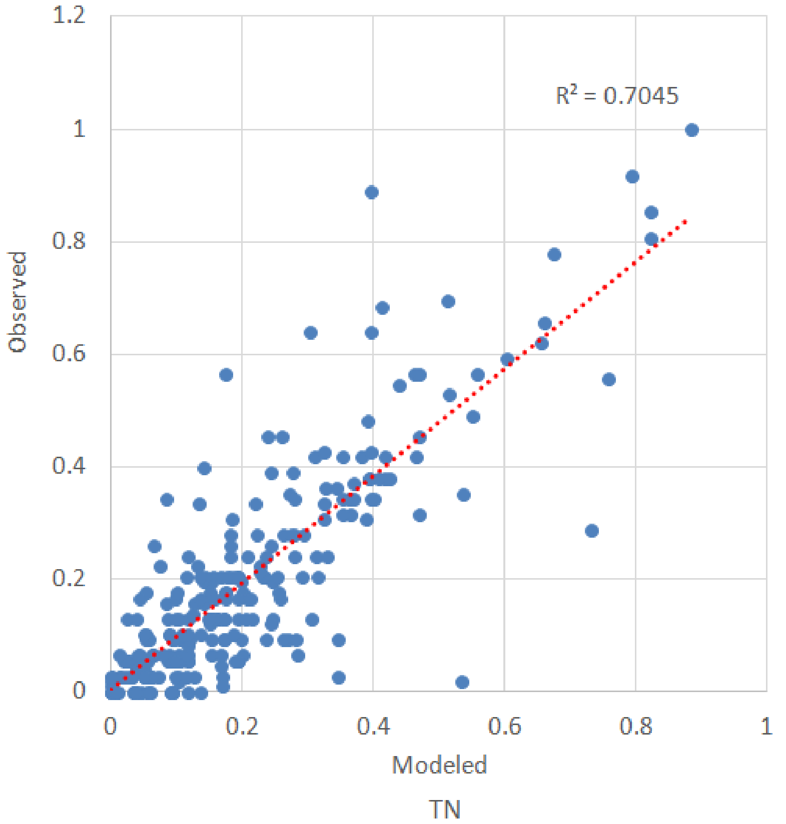


Figure 3.1 Scatter plots of the modelled and observed CHLA values with the Standard NN for Lake Erie

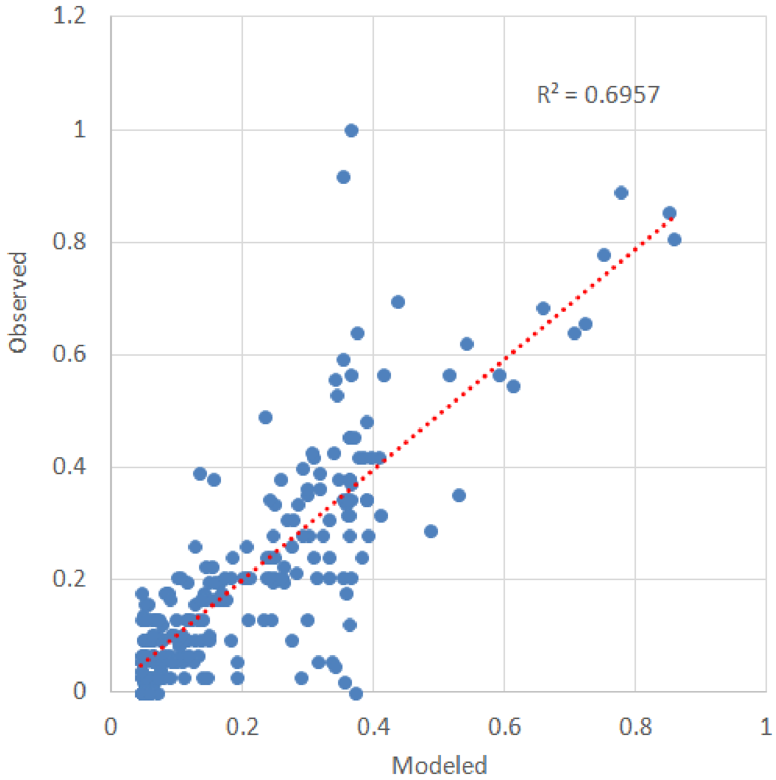
(note: R in the figures represents r in the formula)

Table 3.1 Results of Standard NN

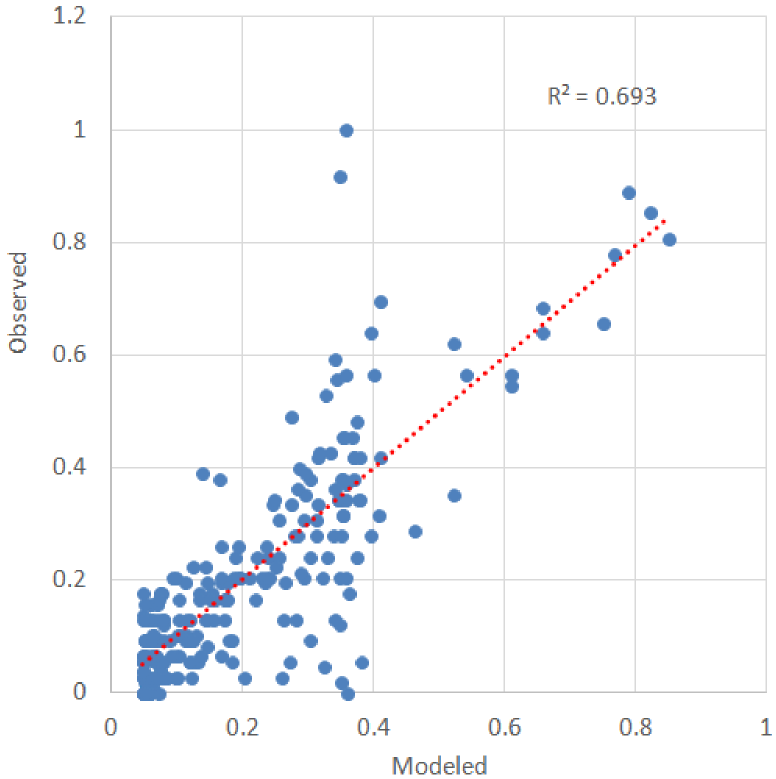
Standard NN		
Input Variables	R squared:	Correlation coefficient r:
TP,TN,PH,DO,T	0.825	0.9144
TP,TN,PH,DO	0.7822	0.8934
TP,TN,DO	0.7663	0.8815
TN,DO	0.7369	0.8643
TN	0.7015	0.8394

The first Figure (3.1) and Table (3.1) are the results of the Standard NN model. When the artificial neural network is used for learning for the first time, the initial R-squared is 0.825. Then, the various water quality parameters are sequentially deleted and the same "learning" is continued with the Standard NN. The results show that the temperature has the smallest effect on R squared. Therefore, it is concluded that among the five water quality parameters, the temperature has the smallest correlation with chlorophyll. Next, the above method was continued, and pH, total phosphorus and dissolved oxygen were deleted in order of relevance. Finally, it is concluded that total nitrogen is the water quality parameter most relevant to chlorophyll.

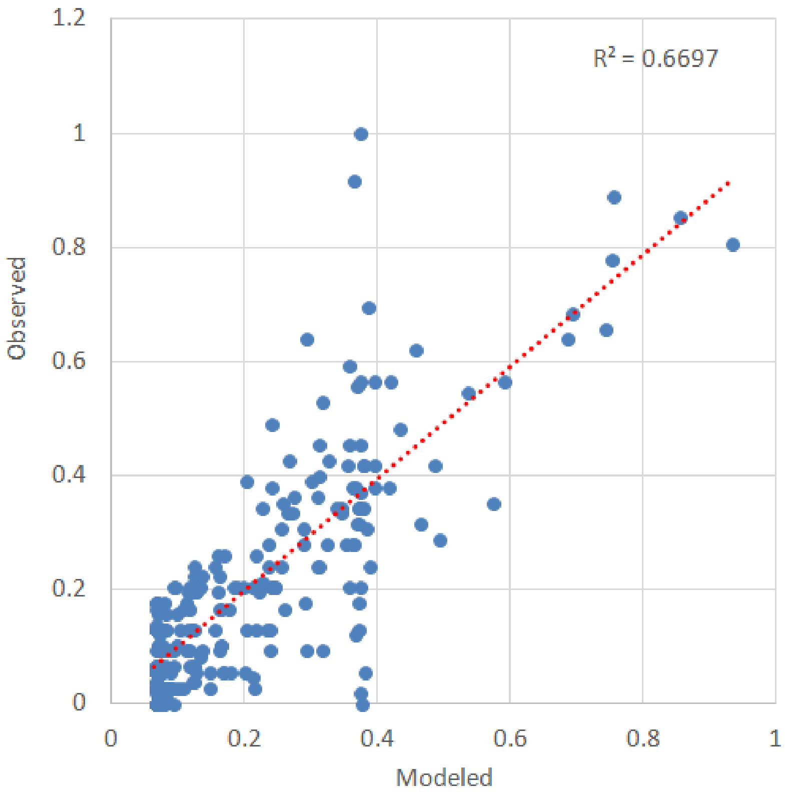
3.2 SRN



TP, TN, PH, DO, T

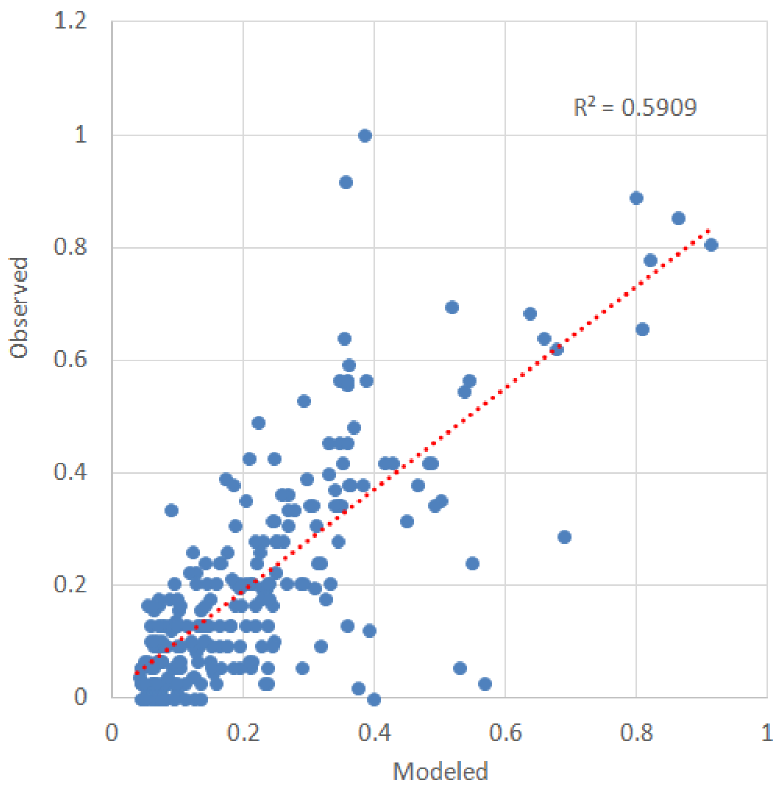


TP, TN, PH, T



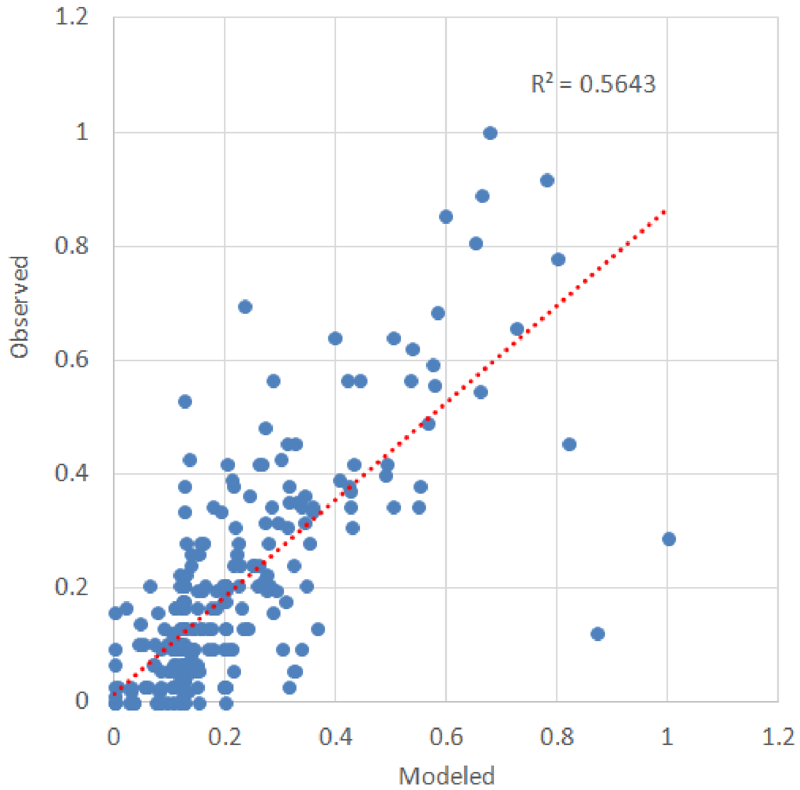
1.2
1
0.8
0.6
0.4
0.2
0
Observed
0
0.2
0.4
0.6
0.8
1
Modeled
R² = 0.6697

TP, TN, PH



1.2
1
0.8
0.6
0.4
0.2
0
Observed
0
0.2
0.4
0.6
0.8
1
Modeled
R² = 0.5909

TN, PH



TN

Figure 3.2 Scatter plots of the modelled and observed CHLA values with the SRN for Lake Erie

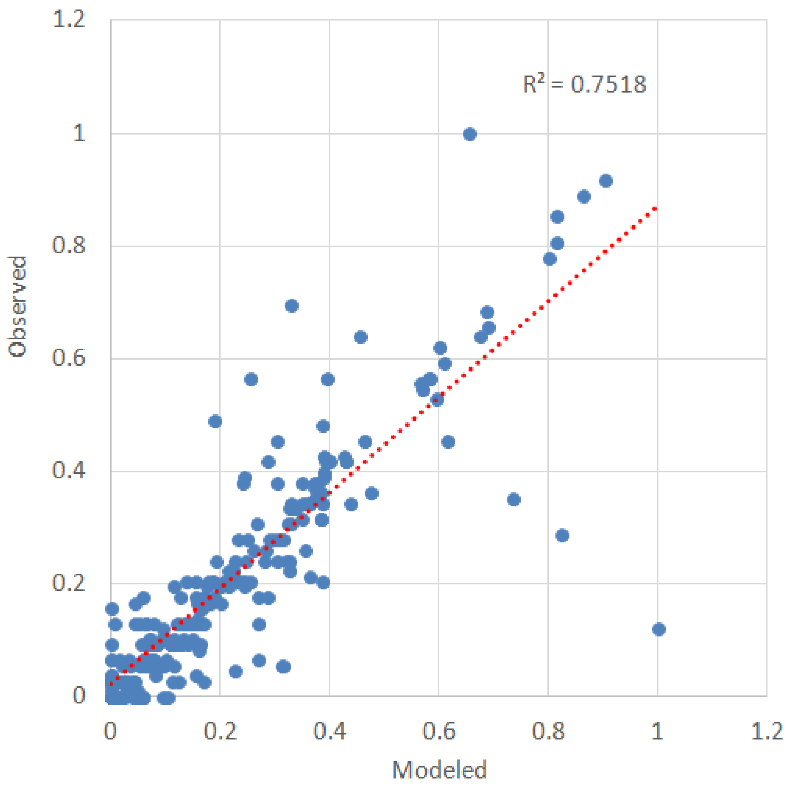
(note: R in the figures represents r in the formula)

Table 3.2 Results of SRN

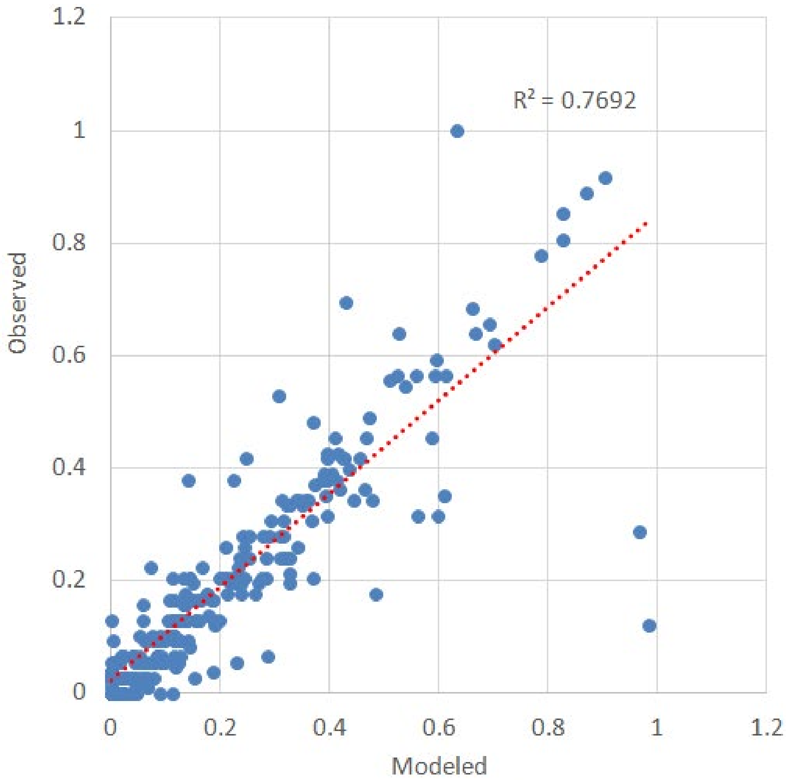
SRN		
Input Variables	R squared:	Correlation coefficient r:
TP,TN,PH,DO,T	0.6955	0.8341
TP,TN,PH,T	0.6928	0.8324
TP,TN,PH	0.6693	0.8184
TN,PH	0.5807	0.7687
TN	0.4541	0.6879

The second is the result of the SRN model, as shown in Figure 3.2 and Table 3.2. When the artificial neural network is used for learning for the first time, the initial R-squared is 0.6955. Then, the various water quality parameters are sequentially deleted and the same "learning" is continued with the SRN. The results show that dissolved oxygen (DO) has the smallest effect on R squared. Therefore, it is concluded that among the five water quality parameters, dissolved oxygen (DO) has the smallest correlation with chlorophyll. Next, the above method was continued, and temperature, total phosphorus and pH were deleted in order of relevance. Finally, it is concluded that total nitrogen is the water quality parameter most relevant to chlorophyll.

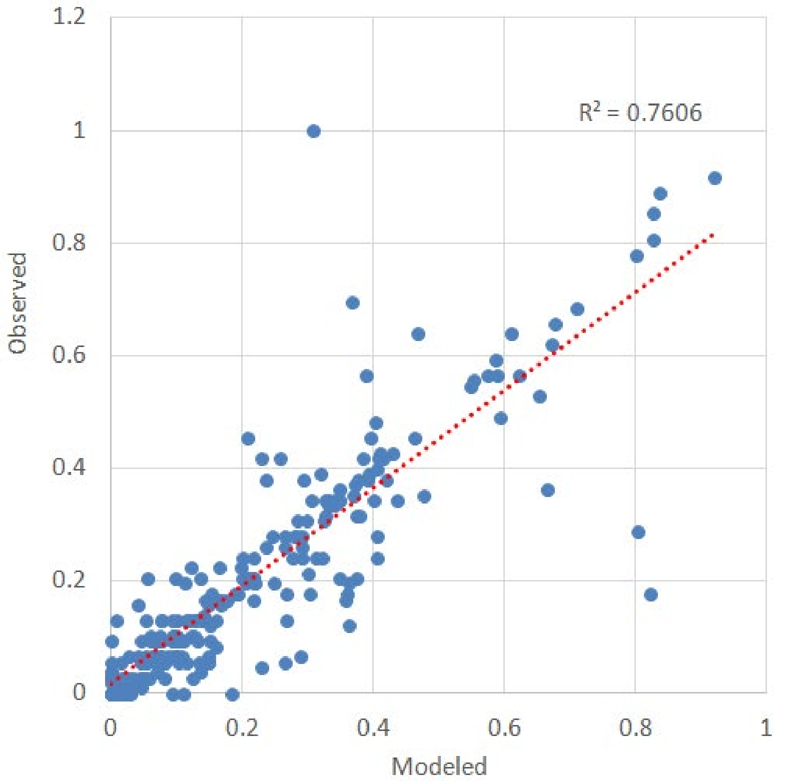
3.3 BPNN



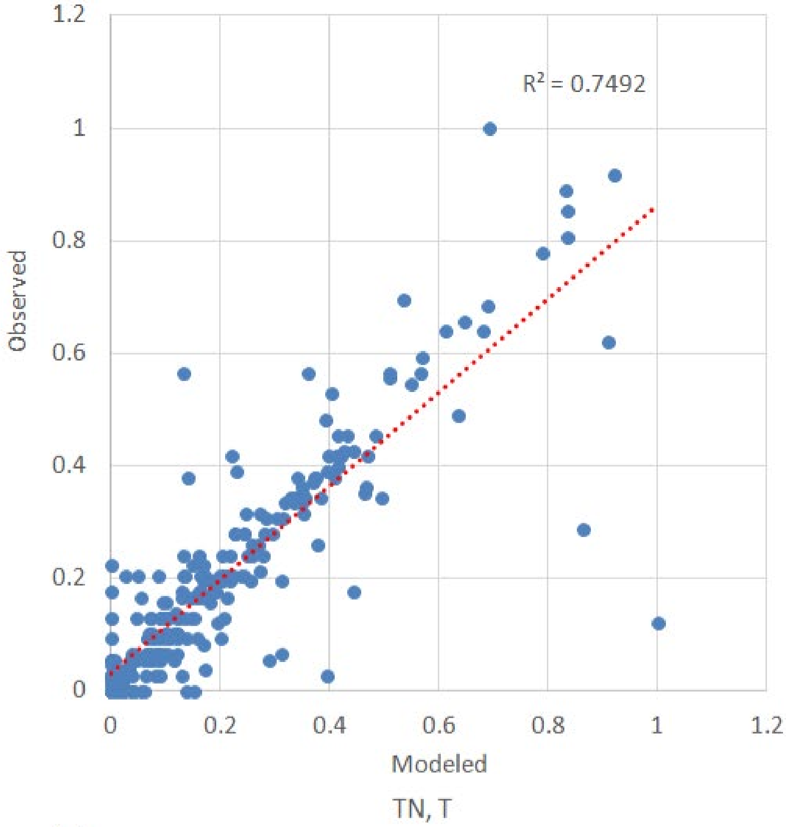
TP, TN, PH, DO, T



TP, TN, DO, T



TP, TN, T



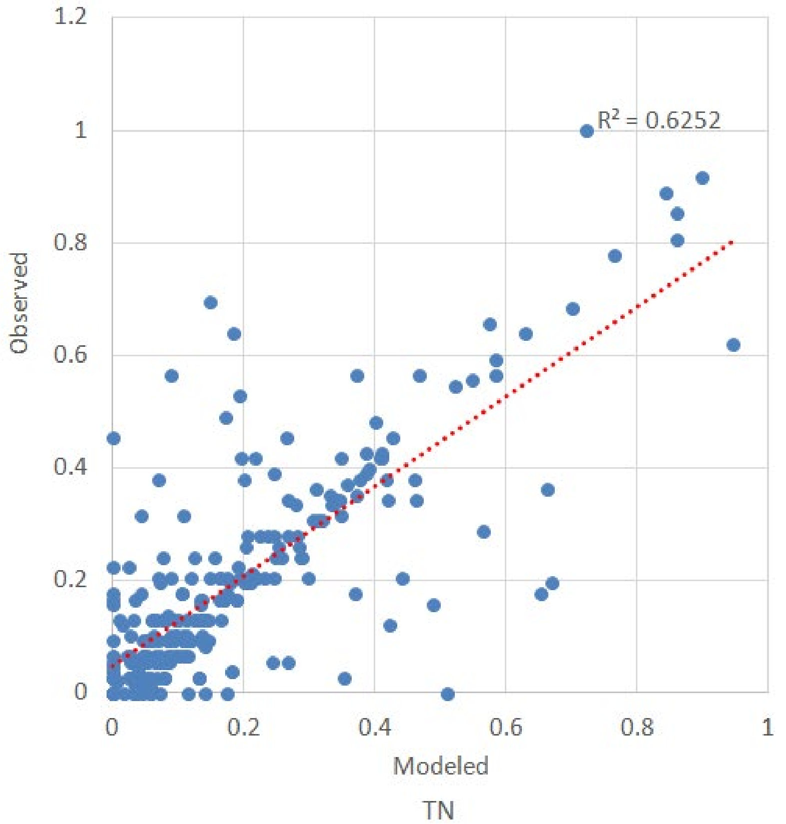


Figure 3.3 Scatter plots of the modelled and observed CHLA values with the BPNN for Lake Erie

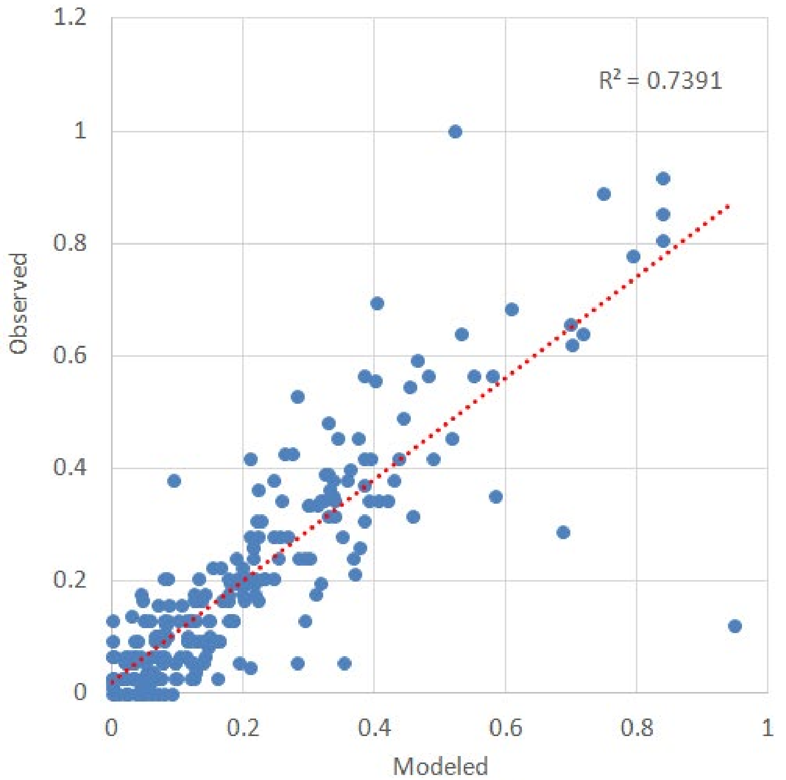
(note: R in the figures represents r in the formula)

Table 3.3 Results of BPNN

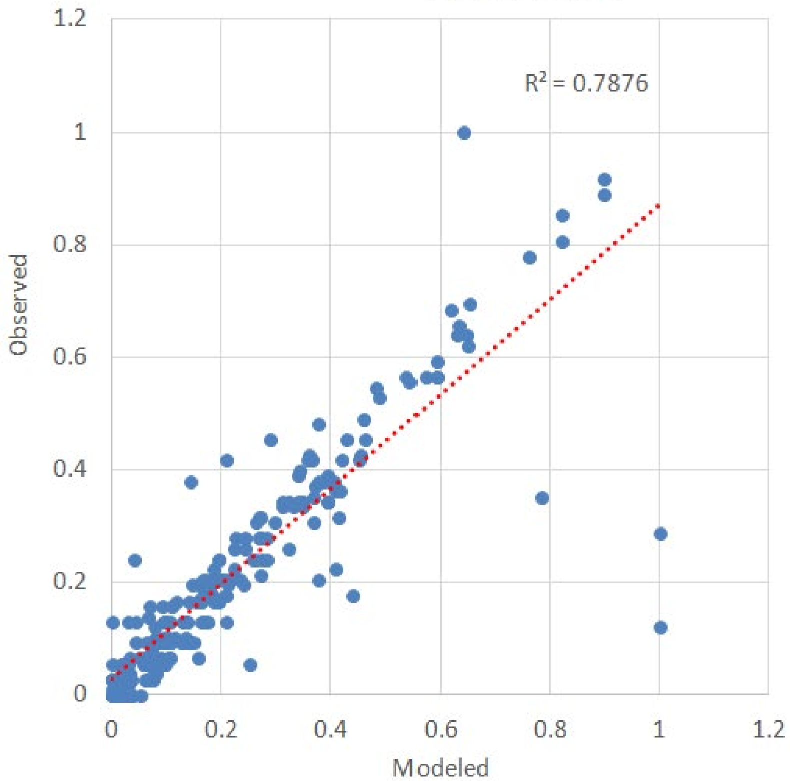
BPNN		
Input Variables	R squared:	Correlation coefficient r:
TP,TN,PH,DO,T	0.7255	0.867
TP,TN,DO,T	0.7307	0.877
TP,TN,T	0.7397	0.8721
TN,T	0.7181	0.8655
TN	0.6066	0.809

The third is the result of the BPNN model, as shown in Figure 3.3 and Table 3.3. When the artificial neural network is used for learning for the first time, the initial R-squared is 0.7255. Then, the various water quality parameters are sequentially deleted and the same "learning" is continued with the BPNN. The results show that pH has the smallest effect on R squared. Therefore, it is concluded that among the five water quality parameters, pH has the smallest correlation with chlorophyll. Next, the above method was continued, and dissolved oxygen, total phosphorus and temperature were deleted in order of relevance. Finally, it is concluded that total nitrogen is the water quality parameter most relevant to chlorophyll.

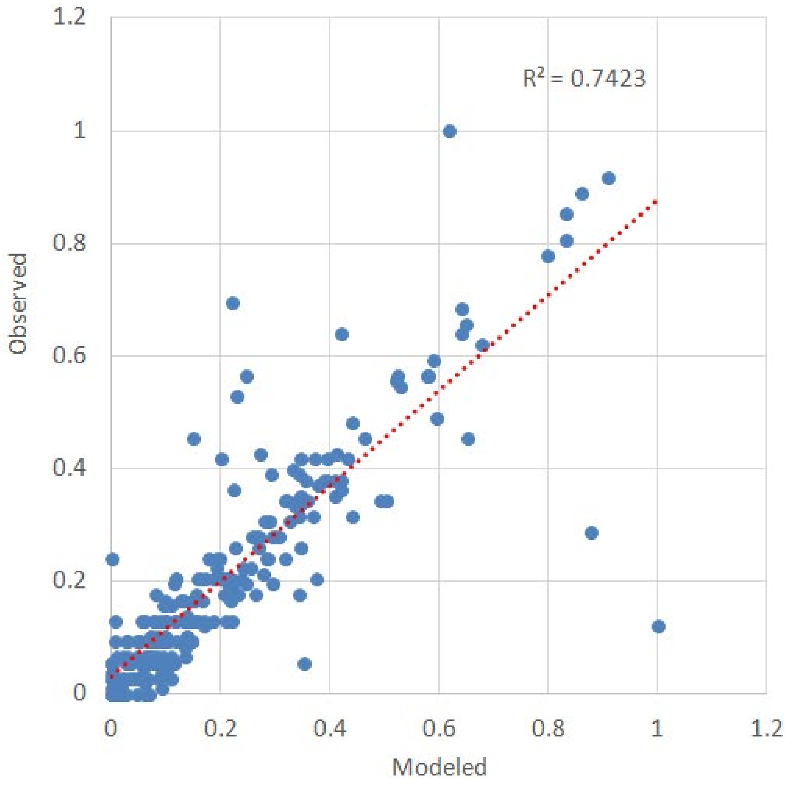
3.4 JCNN



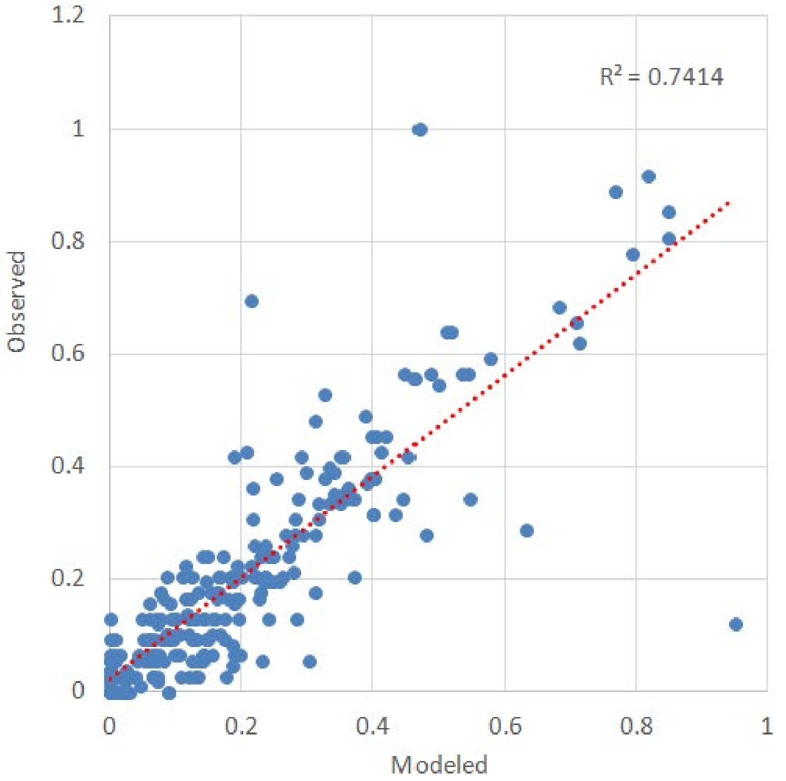
TP, TN, PH, DO, T



TP, TN, DO, T


R² = 0.7423
Observed
Modeled
0
0.2
0.4
0.6
0.8
1
1.2

TP, TN, DO


R² = 0.7414
Observed
Modeled
0
0.2
0.4
0.6
0.8
1
1.2

TN, DO

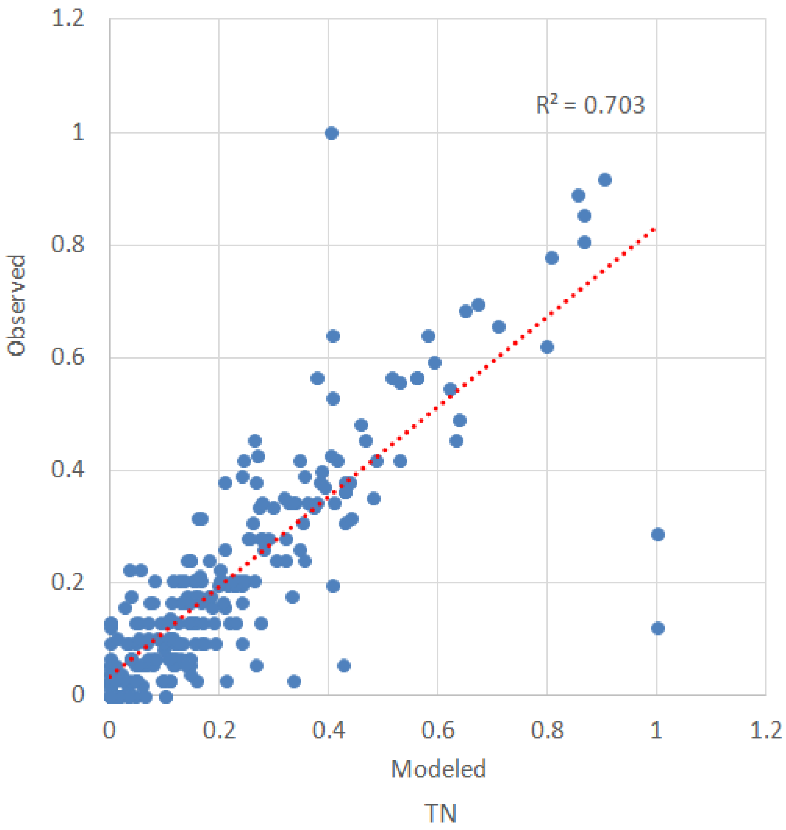


Figure 3.4 Scatter plots of the modelled and observed CHLA values with the JCNN for Lake Erie

(note: R in the figures represents r in the formula)

Table 3.4 Results of JCNN

JCNN		
Input Variables	**R squared:**	**Correlation coefficient r:**
TP,TN,PH,DO,T	0.7303	0.8597
TP,TN,DO,T	0.7601	0.8875
TP,TN,DO	0.7184	0.8616
TN,DO	0.7321	0.861
TN	0.6562	0.8385

Finally, the result of the JCNN model is shown in Figure 3.4 and Table 3.4. When the artificial neural network is used for learning for the first time, the initial R-squared is 0.7303. Then, the various water quality parameters are sequentially deleted and the same "learning" is continued with the JCNN. The results show that pH has the smallest effect on R squared. Therefore, it is concluded that among the five water quality parameters, pH has the smallest correlation with chlorophyll. Next, the above method was continued, and temperature, total phosphorus and dissolved oxygen were deleted in order of relevance. Finally, it is concluded that total nitrogen is the water quality parameter most relevant to chlorophyll.

4 Conclusions

According to the "learning" results of the four models, the first conclusion is found that total nitrogen is the most related to chlorophyll, as well as total phosphorus and dissolved oxygen.

Then, by comparing the results of the four artificial neural network models, it was found that the R-squared of the Standard NN model is closest to 1. Therefore, the fitting result of Standard NN is the best, and it is the most suitable for the chlorophyll in Lake Erie.

In addition to the above two conclusions, the results obtained by each ANN model are different, so it is necessary to consider the results of multiple artificial neural network (ANN) models. The result after comprehensive consideration will be more accurate than the result of considering only a single artificial neural network (ANN) model.

Finally, because the water depth of each monitoring point is different, it is necessary to use the water depth as an important parameter. This can increase the accuracy of the results. Previous studies have not considered the impact of water depth, so considering the water depth is one of the innovations of the project.

5 Limitations and Recommendations

5.1 Limitations

There is not enough input data so only the water quality parameters most relevant to chlorophyll can be found, and the code to predict chlorophyll cannot be generated.

In this project, only four main artificial neural network (ANN) models are selected. In order to make the results more accurate, more artificial neural network (ANN) models need to be applied.

The project lacks model calibration and verification.

5.2 Recommendations

In order to obtain more accurate results and achieve the goal of prediction, more input parameters should be used to predict chlorophyll-a (CHLA). Also, more artificial neural network (ANN) models should be applied in prediction to make results more accurate.

6 Acknowledgments

Thanks to my supervisors Les Lavkulich and Hanspeter Schreier for supervising and guiding me during the project. Thanks to Julie Wilson, Megan Bingham, and all the students in the MLWS programme for their help. In addition, I would like to thank Professor Jeanne Jinhui Huang and Dr. Yu Li of Nankai University for their inspiration and for teaching me to use artificial neural network algorithms. I could not complete this project report without your help. Thank you very much.

References

- Altunkaynak, A., & Altunkaynak, A. (2007). Forecasting surface water level fluctuations of lake van by artificial neural networks. *Water Resources Management, 21*(2), 399-408. doi:10.1007/s11269-006-9022-6
- Artificial neural networks; investigators at china university of geosciences detail findings in artificial neural networks (lake level dynamics exploration using deep learning, artificial neural network, and multiple linear regression techniques). (2019,). *Journal of Robotics & Machine Learning*
- Basile, S. J., Rauscher, S. A., & Steiner, A. L. (2017). Projected precipitation changes within the great lakes and western lake erie basin: A multi-model analysis of intensity and seasonality. *International Journal of Climatology, 37*(14), 4864-4879. doi:10.1002/joc.5128
- Canada. Environment and Climate Change Canada, & Canadian Government EBook Collection. (2018). *Canada-ontario lake erie action plan: Partnering on achieving phosphorus loading reductions to lake erie from canadian sources*. Ottawa: Environment and Climate Change Canada.
- Fantin-Cruz, I., Pedrollo, O., Costa Bonecker, C., da Motta-Marques, D., & Loverde-Oliveira, S. (2010). Zooplankton density prediction in a flood lake (pantanal – brazil) using artificial neural networks. *International Review of Hydrobiology, 95*(4-5), 330-342. doi:10.1002/iroh.201011205
- Fraser, C., & desLibris - Books. (2008). *Lake erie stories: Struggle and survival on a freshwater ocean*. Toronto: Dundurn Press.
- Huang, J., & Gao, J. (2017). An ensemble simulation approach for artificial neural network: An example from chlorophyll a simulation in

lake poyang, china. *Ecological Informatics, 37*, 52-58. doi:10.1016/j.ecoinf.2016.11.012

- Krasnopolsky, V., Nadiga, S., Mehra, A., & Bayler, E. (2018). Adjusting neural network to a particular problem: Neural network-based empirical biological model for chlorophyll concentration in the upper ocean. *Applied Computational Intelligence and Soft Computing, 2018*, 1-10. doi:10.1155/2018/7057363
- Li, Y. L., Zhang, Q., Werner, A. D., & Yao, J. (2015). Investigating a complex lake-catchment-river system using artificial neural networks: Poyang lake (china).*Hydrology Research, 46*(6), 912. doi:10.2166/nh.2015.150
- Lu, F., Chen, Z., Liu, W., & Shao, H. (2016). Modeling chlorophyll-a concentrations using an artificial neural network for precisely eco-restoring lake basin. *Ecological Engineering, 95*, 422-429. doi:10.1016/j.ecoleng.2016.06.072
- Luo, W., Zhu, S., Wu, S., & Dai, J. (2019). Comparing artificial intelligence techniques for chlorophyll-a prediction in US lakes. *Environmental Science and Pollution Research International, 26*(29), 30524-30532. doi:10.1007/s11356-019-06360-y
- May, C. J., Ludsin, S. A., Glover, D. C., & Marschall, E. A. (2020). The influence of larval growth rate on juvenile recruitment in lake erie walleye (sander vitreus).*Canadian Journal of Fisheries and Aquatic Sciences, 77*(3), 548-555. doi:10.1139/cjfas-2019-0059
- Mekonnen, A. D., & Gorsevski, P. V. (2015). A web-based participatory GIS (PGIS) for offshore wind farm suitability within lake erie, ohio. *Renewable and Sustainable Energy Reviews, 41*(C), 162-177. doi:10.1016/j.rser.2014.08.030
- Meyer, K. A., Davis, T. W., Watson, S. B., Denef, V. J., Berry, M. A., & Dick, G. J. (2017). Genome sequences of lower great lakes microcystis sp. reveal strain-specific genes that are present and expressed in

western lake erie blooms. *PloS One, 12*(10), e0183859. doi:10.1371/journal.pone.0183859

- Millie, D. F., Weckman, G. R., Fahnenstiel, G. L., Carrick, H. J., Ardjmand, E., Young, W. A., . . . Shuchman, R. A. (2014). Using artificial intelligence for CyanoHAB niche modeling: Discovery and visualization of microcystis –environmental associations within western lake erie. *Canadian Journal of Fisheries and Aquatic Sciences, 71*(11), 1642-1654. doi:10.1139/cjfas-2013-0654
- Mou, X., Lu, X., Jacob, J., Sun, S., & Heath, R. (2013). Metagenomic identification of bacterioplankton taxa and pathways involved in microcystin degradation in lake erie.*PloS One, 8*(4), e61890. doi:10.1371/journal.pone.0061890
- Piasecki, A., Jurasz, J., & Skowron, R. (2015). Application of artificial neural networks (ANN) in lake drwęckie water level modelling. *Limnological Review, 15*(1), 21-30. doi:10.2478/limre-2015-0003
- Saber, A., James, D. E., & Hayes, D. F. (2019;2020;). Long-term forecast of water temperature and dissolved oxygen profiles in deep lakes using artificial neural networks conjugated with wavelet transform. *Limnology and Oceanography, 65*(6), 1297-1317. doi:10.1002/lno.11390
- Sayers, M. J., Grimm, A. G., Shuchman, R. A., Bosse, K. R., Fahnenstiel, G. L., Ruberg, S. A., & Leshkevich, G. A. (2019). Satellite monitoring of harmful algal blooms in the western basin of lake erie: A 20-year time-series. *Journal of Great Lakes Research, 45*(3), 508-521. doi:10.1016/j.jglr.2019.01.005
- Skwor, T., Shinko, J., Augustyniak, A., Gee, C., & Andraso, G. (2014). Aeromonas hydrophila and aeromonas veronii predominate among potentially pathogenic ciprofloxacin- and tetracycline-resistant

aeromonas isolates from lake erie. *Applied and Environmental Microbiology, 80*(3), 841-848. doi:10.1128/AEM.03645-13

- Tian, D., Xie, G., Tian, J., Tseng, K., Shum, C. K., Lee, J., & Liang, S. (2017). Spatiotemporal variability and environmental factors of harmful algal blooms (HABs) over western lake erie. *PloS One, 12*(6), e0179622. doi:10.1371/journal.pone.0179622
- U.S. Congress. (1898). *Lake erie maps.* ().
- Zhou, Y., Obenour, D. R., Scavia, D., Johengen, T. H., & Michalak, A. M. (2013). Spatial and temporal trends in lake erie hypoxia, 1987–2007. *Environmental Science & Technology, 47*(2), 899-905. doi:10.1021/es303401b
- Wynne, T., & Stumpf, R. (2015). Spatial and temporal patterns in the seasonal distribution of toxic cyanobacteria in western lake erie from 2002–2014. *Toxins, 7*(5), 1649-1663. doi:10.3390/toxins7051649