Vol. 3, No. 1, 2018

MODIFIED REDFIELD RATIO COMBINED WITH ARTIFICIAL NEURAL NETWORK SIMULATION TO ESTIMATE THE ALGAL BLOOM PATTERN

Tri Nguyen-Quang, Kateryna Hushchyna, Kayla McLellan, Qurat-Ul An Sabir

Biofluids and Biosystems Modeling Lab (BBML)
Department of Engineering, Faculty of Agriculture, Dalhousie University
39, Cox Road, Truro-Bible Hill, B2N 5E3, Canada
Correspondence: tri.nguyen-quang@dal.ca

Received: 30.11.2017

Abstract. In previous work (Hushchyna and Nguyen-Quang, 2017), we have introduced the Modified Redfield Ratio (MRR) to estimate algal blooms occurring in Mattatall Lake, Nova Scotia (Canada). The goals of this paper are to test this modified index based on nutrient level to estimate bloom patterns via field experimental data and by the mathematical simulation with a supervised learning model Artificial Neural Network. Although the original Redfield ratio, a molar index applied in marine water based on the C:N:P ratio, is still discussed, the MRR based on Nitrates/Phosphates ratio we suggested herein seems to have many practical aspects for fresh water to evaluate the onset of bloom patterns.

Key-words: Cyanobacteria, Harmful Algal Blooms (HAB), Modified Redfield Ratio (MRR), Artificial Neural Network (ANN)

1. Introduction

Mattatall Lake (ML) in Nova Scotia (Fig. 1) started to experience severe algal blooms in 2013. During the fall of 2014, massive algal blooms appeared in ML, and persisted until late December. This phenomenon was observed and recorded by our team in the fall-winter of 2014 (Nguyen-Quang, 2015). The duration of this phenomenon was extremely unusual as algal blooms have not been known to last until the winter season or coexist with icy conditions. The dominant species in this bloom was identified to be *Anabaena planctonica* (*Dolicospermum planctonicum*) with a cell count around 250 000 cells/ml, which may produce the neurotoxin *Anatoxin-a.* This cell count is approximately two and a half times higher than the alert level 2 guideline from World Health Organization's drinking water standards (Chorus and Bartram, 1999).

There are certain combinations of multiple factors that trigger HAB specific to each waterbody. However, their coupling effects are not yet understood. No research or data related to ML has been done. Moreover, in Nova Scotia, no systematic investigation has been sketched for cyanobacterial bloom patterns.

Hushchyna and Nguyen-Quang (2017) have presented the Modified Redfield Ratio (hereafter denoted MRR) to estimate the HAB occurrence in Mattatall lake (Nova Scotia, Canada), served as a pilot site for our pioneering systematic study in order to evaluate the recent bloom phenomena in the entire province and in Atlantic Canada.

In this paper, we try to:

- Evaluate the coupling effects of different governing parameters on the recent bloom phenomenon occurring over the last two years. For the first step, we deal with 3 parameters involving in algal growth including Nitrates, Phosphates and Dissolved Oxygen (DO). We consider 2 biological pigments Chlorophyll-a (Chl-a) and Phycocyanin (PC) released by cyanobacterial species as the effects from these causes.
- Sketching again the Modified Redfield Ratio to estimate the onset of algal bloom, i.e the state of instability where there appear some scums of algae.
- Using a mathematical model (Artificial Neural Network – ANN) in simulating coupling cause-effect relationships of five factors (Nitrates, Phosphates, DO, Chl-a, PC). From that, the validation for MRR and prediction for HAB risks in function of environmental factors could be estimated.

2. Experimental processes and methodology

2.1. Experimental processes

Different equipment for field sampling and Lab analysis were used, such as a YSI probe (Professional

Plus, Hoskin scientific LTD, USA) for pH, Dissolved Oxygen (DO), conductivity, and temperature in the water; a Fluorometer for Chlorophyll-a and Phycocyanin measurements; and Photometer for evaluating nutrient components. Samples were taken tri-weekly or every month, depending on the weather conditions, starting from May to November yearly, at the surface and bottom levels. It is also noted that the lake is quite shallow with the maximum depth around 8m. Predetermined sampling locations are represented in Fig. 1 with equivalent coordinates in Table 1. If there were some other places that are not included in the predetermined points that had an algae bloom present, samples were taken at these points as well. A HOBO weather station was installed in one fixed location of the lake. As soon as sampling was finished, they were taken to the lab for micronutrient analyses.

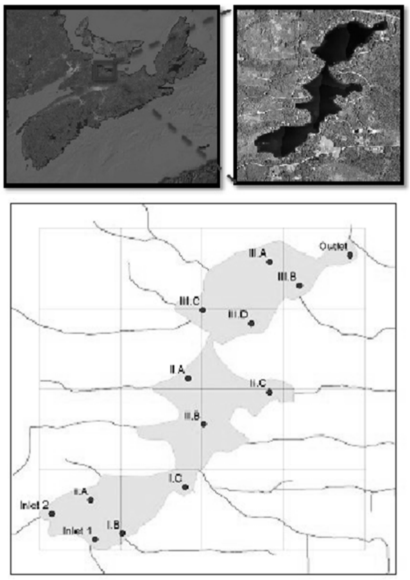

Fig. 1. *Above:* Google Earth view of Nova Scotia highlighting ML area (left) and Government of Nova Scotia aerial photograph of ML (right). *Coordinates of ML are 45° 42' 17" N, 63° 19' 39" W*
Below: Sample stations (locations) predetermined and monitored.

Chlorophyll-a is usually used as a parameter to determine the quantity of primary photosynthetic pigment in cells of aquatic micro plants. Measurement and determination of this parameter are the basic analysis to evaluate the characteristics of algal blooms in many research works in the world. Unfortunately, Chl-a represents just the whole quantity of photosynthesis pigment released from all algae and micro-plants present in water, hence it cannot help to distinguish cyanobacteria existence among all living micro-plants and algae in the waterbody. To be able to define and confirm the presence of Cyanobacteria species in the composition of aquatic microalgae in the waterbody, another pigment form, Phycocyanin (PC), is used. This pigment can determine different cyanobacteria species from another planktonic species.

Table 1

Coordinates of sampling points in Mattatall Lake

Point	X (Longitude/Latitude)	Y (Longitude/Latitude)
IA	-63.483759 W	45.683914 N
IA1	-63.482013 W	45.681338 N
IB	-63.480922 W	45.681845 N
IC	-63.475384 W	45.684767 N
IIA	-63.475148 W	45.691568 N
IIB	-63.47378 W	45.688689 N
IIC	-63.467932 W	45.690711 N
IIIA	-63.467907 W	45.698875 N
IIIB	-63.46526 W	45.697408 N
IIIC	-63.473843 W	45.695859 N
IIID	-63.469584 W	45.695011 N
Inlet 1	-63.483369 W	45.681467 N
Inlet 2	-63.487229 W	45.683072 N
Outlet	-63.460799 W	45.699303 N

2.2. Redfield ratio (RR) and Modified Redfield ratio (MRR)

Redfield (Redfield, 1958) suggested an optimal ratio between Nitrogen and Phosphorus which were contained in living organisms. This ratio, known under the name of *'Redfield ratio'*, can be used to estimate the algal status in a water body. With a mass conversion, Bulgakov and Levich (1999) defined ranges of this ratio (N/P): 1) *from 5 to 10: Cyanophyta* species domination would be considered; 2) from *20 to 50:* green algae growth would be favored.

Inspired from this idea, our suggestion however did use Nitrates and Phosphates (not TP and TN) as we believe that mineral form of these nutrient components triggered the bloom in ML and could be the main cause of it. The mass conversion is based on the ratio of *Nitrates to Phosphates* to define the blooming possibilities as well as their limits for the development of two main potential groups leading to bloom patterns: *green algae* and *blue-green algae*. For a detail of our Modified Redfield Ratio calculation, we refer to Hushchyna and Nguyen-Quang (2017).

2.3. Mathematical model

An Artificial Neural Network (McCulloch and Pitts, 1943) was used as the mathematical model in this research. The model has simple processing elements called neurons or nodes. Direct communication links are what connect the nodes to one another, and each node has a weight function. The weights represent information being used the net to solve the problem. We suggest the back-propagation multi-layer neural network in which, nodes are arranged into layers: input layer (Data observations), hidden layer(s) (intermediate nodes) and output layer (conclusions) (Fig. 2).

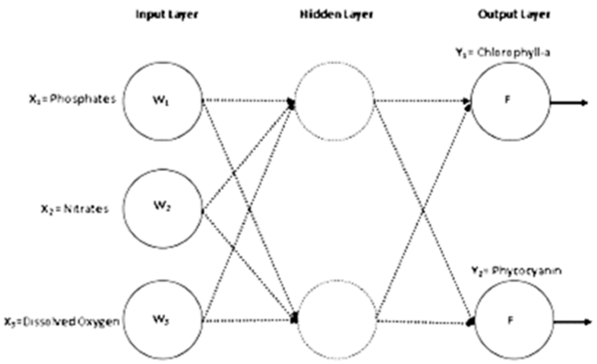

Fig. 2. Multilayer ANN model for Chlorophyll-a and Phycocyanin development

Our study herein is limited on a set of three main parameters for the input sets represented by 3 variables, *X_1, X_2 and X_3 equivalent to*: 1) Phosphates; 2) Nitrates; 3) Dissolved Oxygen (DO). Two outputs in this model are the quantity of Chlorophyll-a and Phycocyanin (Y_1 and Y_2), representing the growth of cyanobacteria or output layer (Fig. 2). The input layer data was collected over the summer months, and over 150 data points was used to train and validate the model.

The weight layer signals form a weight matrix: W_{ij}. The weights undergo a "training step" where they are fed data and adjusted until they can produce results that are similar to the actual outputs. Mathematically, the input can be expressed in a vector form as $\{X\} = \{X_1 \ldots X_n\}$. The i^{th} component of the vector X_i that comes out from the input node i is transferred to anode j on a hidden layer (j=1, 2,.... m) through the weight matrix W_{ij}.

Each hidden node has a summation function operating on the input values, the total input u_j received by the hidden node is:

$$u_j = \sum_{i=1}^{k} W_{ij} X_i \,. \qquad (1)$$

The hidden node j has a transfer function that performs a nonlinear transformation on the total input u_j, and produces an output which becomes the next input fed into a node p of the output layer j (j=1, 2,..., n), which also has a summation function, through another weight V_{jp}.

The total input received by the output node p becomes its output, y_p, expressed as:

$$y_p = \sum_{p=i}^{n} V_{jp} f(u_i) \,. \qquad (2)$$

The outputs can be given in a vector form as $\{Y\} = =\{y_1 \ldots y_n\}$. This neural network performs a nonlinear transformation on $\{X\}$ as expressed in the following equation.

$$\{Y\} = f(\{X\}). \qquad (3)$$

The network is ready to be trained once it has been structured properly. The standard back-propagation algorithm was used for training the network. To train the network, random initial weights are assigned to each node and then known input values and target values are added to the network. Output values are estimated from the input values and then compared to the target values. This is done continuously by the model until the mean squared error R^2 is as low as possible (between the output and target values). Once the model was trained, it was tested with a set of data that was independent from the training data.

3. Results and Discussion

3.1. Risk estimation from field data

The attempt to predict the HAB occurrence and proliferation under complex context of environmental conditions (nutrient, light, meteorological factors, etc.) led to many indexes for the estimation of HAB risks based on chemical components such as micronutrient factors of the waterbody that contribute significantly to algal growth. The simplest index is based only on one factor such as the TP level, Chlorophylla or cell counts.

In the context of this paper, we suggest to use our modified Redfield ratio (MRR) mentioned above (Hushchyna and Nguyen-Quang, 2017) to estimate the HAB risk in a 'simple' way which combine Chlorophyll-a and PC indexes, shown in Table 2. For Chl-a and PC criteria, we refer to Novotny and Olem (1994), Bulgakov and Levich (1999), Brient et al. (2008) and Brylinsky (2009).

Based on the field data obtained during 2015–2016 with our field observations (174 points), we established a table of Index equivalent to PC and Chl-a measured and illustrated by the graph in Fig. 3.

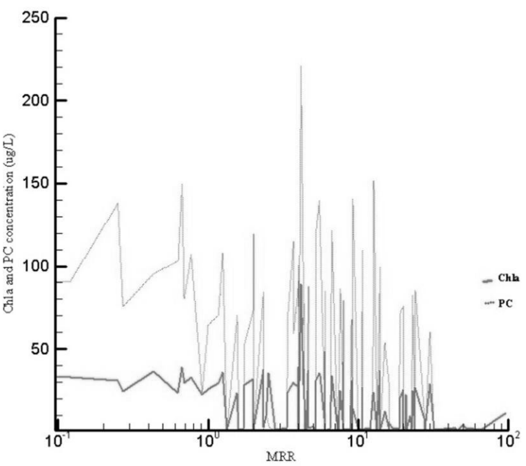

Fig. 3. Chlorophyll-a and Phycocyanin versus MRR (field data 2015 and 2016)

When MRR value of 0-10, the graph contains the most Chl-a values, as well as the highest (120 points, 68.96 %).

From Fig. 3 (with the log scale for MRR), we can see that when MRR between 3.25 and 6.5, all Chl-a greater than 20 µg/L and PC greater than 100 µg/L. In this range, we also have the highest values of Chla (>90 µg/L) and PC (>220 µg/L) at MRR=3.5. In combining Table 2, this show a high risk of HAB. There were in total 27 points that fell into this category of a ratio between 3.25 and 6.5.

With points with MRR less than 3.25, we can observe that Chl-a values vary between 5 and 30 µg/L while PC varies from 20 to 150 µg/L. There were 69 points (39.6 %) with a ratio MRR less than 3.25, and of those points, there was only one point that had the low Chl-a value: 9.91 µg/L. In referring to Table 2, it could be said that there was a medium to medium high risk of HAB.

Table 2

Risk indexes, adapted from Novotny and Olem (1994), Bulgakov and Levich (1999), Brient et al. (2008), Brylinsky M. (2009), Hushchyna and Nguyen-Quang (2017)

Risk	Chl-a index	PC index	Our Modified Redfield Ratio (MRR)
Low risk of HAB	0–5 *µg/L*	≤ *30 ±2 µg/L*	≥ 6.6
Medium risk of HAB	5–10 *µg/L*	Between 30 and 90 *µg/L*	≤ 3.25
High risk of HAB	≥ 10 *µg/L*	≥ *90±2 µg/L*	3.26–6.5

When MRR value of 10-30: This section of the graph has the second highest number of points in the total data distribution (45 points, 25.86 %). Among them, there are 29 points having the Chl-a values between 0 and 5 µg/L, and the rest varied from 5 to 30 µg/L. Regarding PC, the distribution was not homogenous: some points having high values (over 70 µg/L) and some points have less than 30 µg/L. If we consider both PC and Chl-a criteria in the same time as indicated in Table 2, this MRR range should be potentially at low to medium risk.

When MRR value > 30: All points have a very low risk of HAB (9 points, 5.17 %). All of the points have chlorophyll-a values below 5 µg/L. The PC values vary: between 10 and 40 µg/L, except for some outliers with high values (over 80 µg/L). This showed a low risk of HAB.

Actually, our field observations recorded during two summer seasons 2015-2016 showed different bloom scenarios fitting quite well with the above estimation, as illustrated in Figure 4 and Table 3.

Fig. 4. Presence of algal blooms in Mattatall Lake in August 2015

Table 3

Cases of potential blooms based on the modified ratio *nitrates/phosphates*

	2015						2016				
	11-Jun	*30-Jul*	*17-Aug*	*10-Sep*	*04-Oct*	*27-Oct*	*14-Apr*	*10-May*	*17-Jun*	*16-Aug*	*04-Oct*
Cyanobacteria domination	6	4	5	1	6	8	0	4	4	0	0
Green Algae domination	8	9	10	4	1	5	3	0	3	0	0
Potential Cases	12	10	18	26	18	13	0	0	4	22	25

3.2. Simulated results from ANN model versus the risk criteria

The mathematical simulation is highlighted via the relationship curve MRR vs PC and MRR versus Chl-a in Figs. 5 *a, b* (discontinuous lines) which show the regression tendencies generated for both field and simulated data. The discrepancy was due to the fact that the field data was individually measured for each parameter, and the simulated data were combined from mathematical model with coupling parameters (herein with nitrates, phosphates, and DO). These regression curves for both field and simulated data have been chosen so that they are as similar as possible in shape (MRR vs PC or Chl-a). They are both high order polynomial functions.

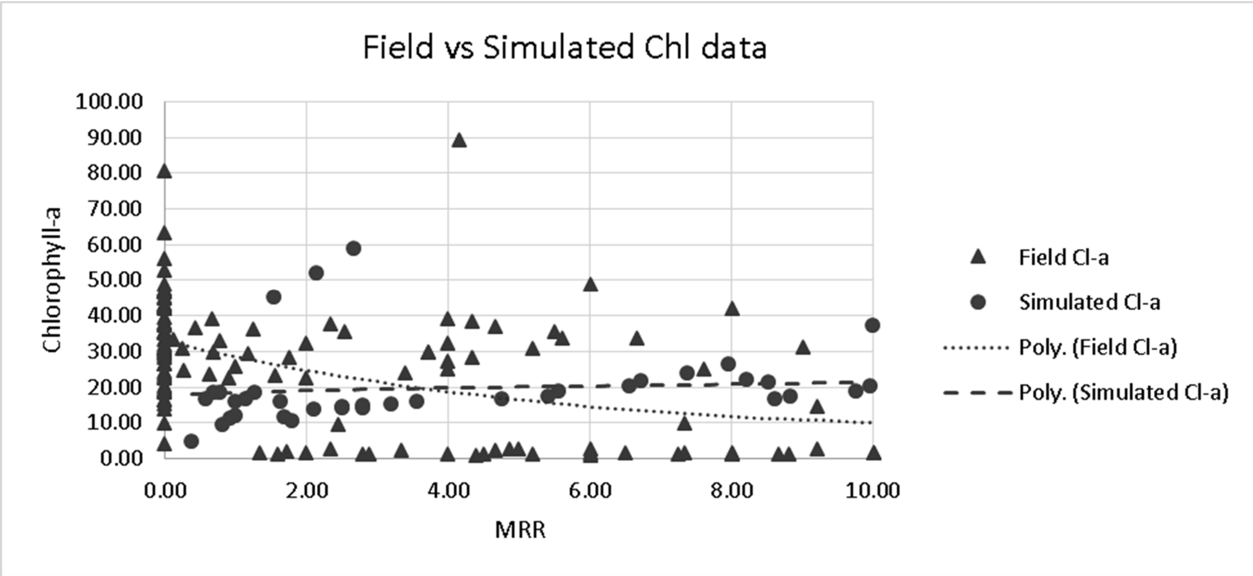

Fig. 5*a*. *Chlorophyll-a versus MRR: Comparison for field data and simulated results*

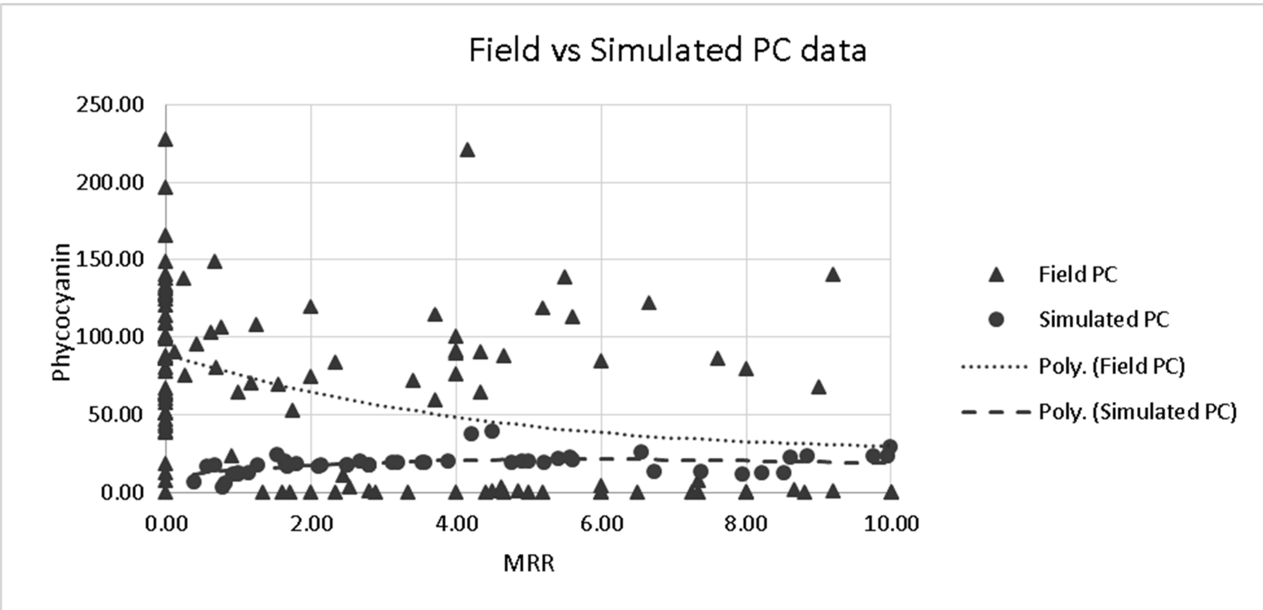

Fig. 5*b*. *Phycocyanin versus MRR: Comparison for field data and simulated results*

For Chl-a, the field data curve had higher values than the simulated curve for a short range (MRR approximately 0 to 3.5). After the MRR value of 3.5 (Fig. 5*a*), the curves have a similar shape, but there is a difference in the concentration between them and simulated data are higher. Two curves cross each other at MRR of 3.5, where the separation two field and simulated curves is easily noticeable. In comparing with Table 2, this value is quite close to the suggested MRR= 3.25.

For PC (Fig. 5*b*), if considering the MRR range from 0 to 10, we can notice the field data is higher than the simulated data. It appears that the simulated data would cross the field data curve and predict higher numbers shortly after the MRR value of 10.

These regression curves do not fit the data very well due to 1) the high nonlinearity of data and 2) the complexity of the coupling effects between parameters.

To consider the combined effects of MRR and DO, for a linear regression, we can see that as the DO increases, so does the concertation of Chl-a (Fig. 6*a*). As the MRR increases, the Chl-a concentration decreases. From Fig. 6*a*, DO increases and MRR remains less than 27.5, the concentration of Chl-a will increase. Beyond this value MRR = 27.5, Chl-a decreases and even disappears.

In eliminating some outliers, we built the nonlinear regression surface for the Chl-a pattern and recognized that the range MRR 0-3 is an interesting range (Fig. 6*b*). This range actually shows a 'chaotic' situation where the Chl-a has, at the same time, increasing and decreasing tendencies. When compared with Table 2, it is noticed that when MRR is < 3.25, we have a medium risk for HAB. The medium risk can be understood in the sense that the *'transition'* situation could lead to bloom or be diminished, depending on other intervening factor(s). This intervening factor herein is DO: obviously in Fig. 6*b*, when DO in between a certain range (0 to 10 mg/L), the risk is higher as Chl-a will increase.

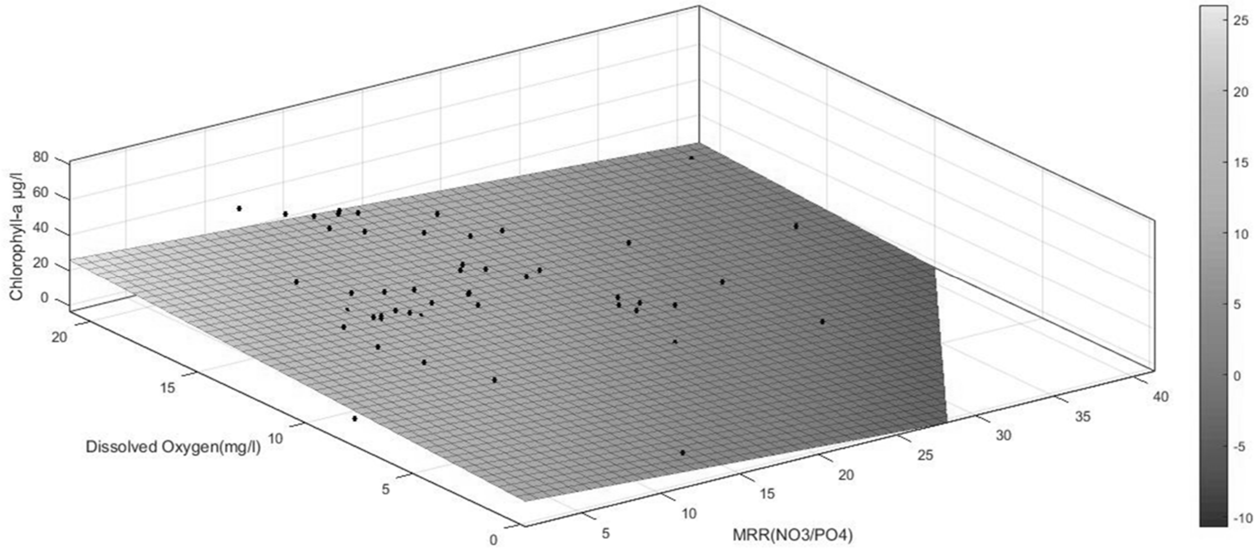

Fig. 6*a*. *Chlorophyll-a vs MRR and DO (Polynomial degree = 1, R^2= 0.9199)*

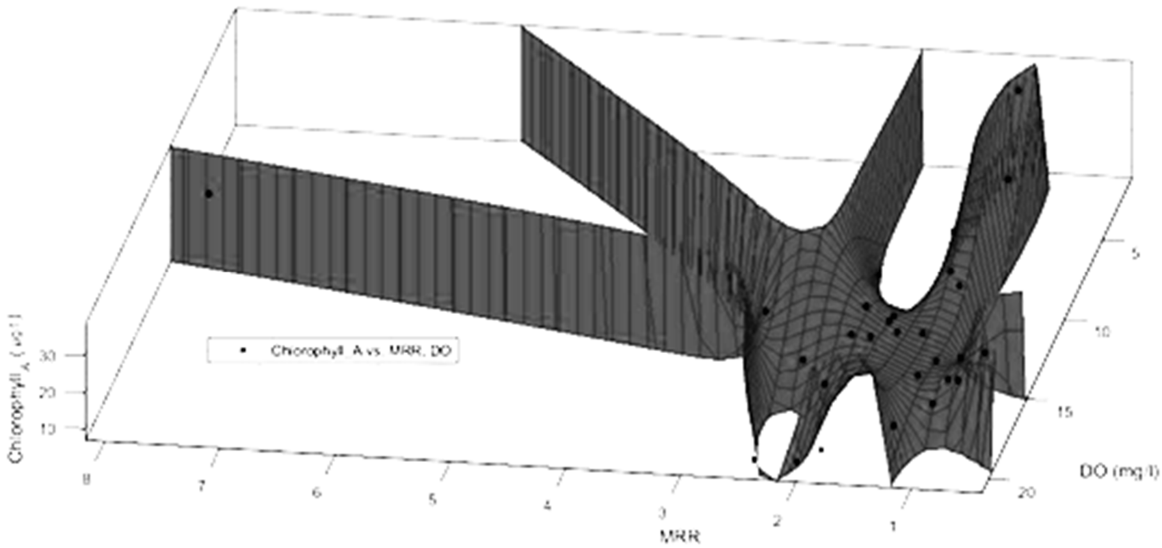

Fig. 6*b*. *Chlorophyll-a vs MRR and DO (Polynomial degree = 5, R^2= 0.6407)*

Also from Fig. 6*b* when MRR went from 3 to 5.5, the combining effects of MRR and DO showed a high value of Chl-a, almost as indicated in Table 2 for the high risk area (MRR 3.25–6.5). Beyond this range, depending on the DO value, the Chl-a value can be diminished.

A low order of regression showed the effect of DO on the PC values in the function of MRR (Fig. 7*a*). The simulated graph reflected some PC patterns obtained from the field trips when Nitrates = 0 (MRR = 0). We also noticed that when DO is in the 0 to 10 mg/l range, PC decreases and then increases at the value of MRR around 5.5. Again, comparing to Table 2, the MRR range belonging to 0 to 6.5 advised a medium to high risk of HAB. As the value of MRR increases beyond than 5.5 as indicated in the graph, the PC increases, depending on DO.

A close look on the nonlinear regression surface as shown in Fig. 7*b* explained the combining effects of DO and MRR. We can see that from the MRR range 0 to near 3.5 a 'chaotic/transitive' situation of PC depends on DO. Beyond 3.5, only with the range of DO from 10 to 13 mg/l PC would continue to increase. All other DO values lead to a PC-free situation no matter what the MRR is (as long as it is greater than 3.5).

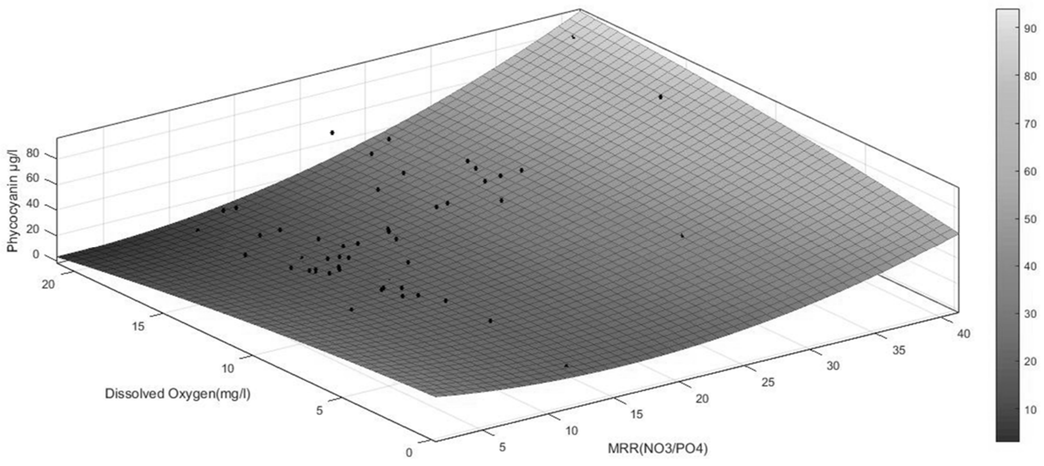

Fig. 7*a*. *Phycocyanin vs MRR and DO (Polynomial degree = 2, R^2= 0.8136)*

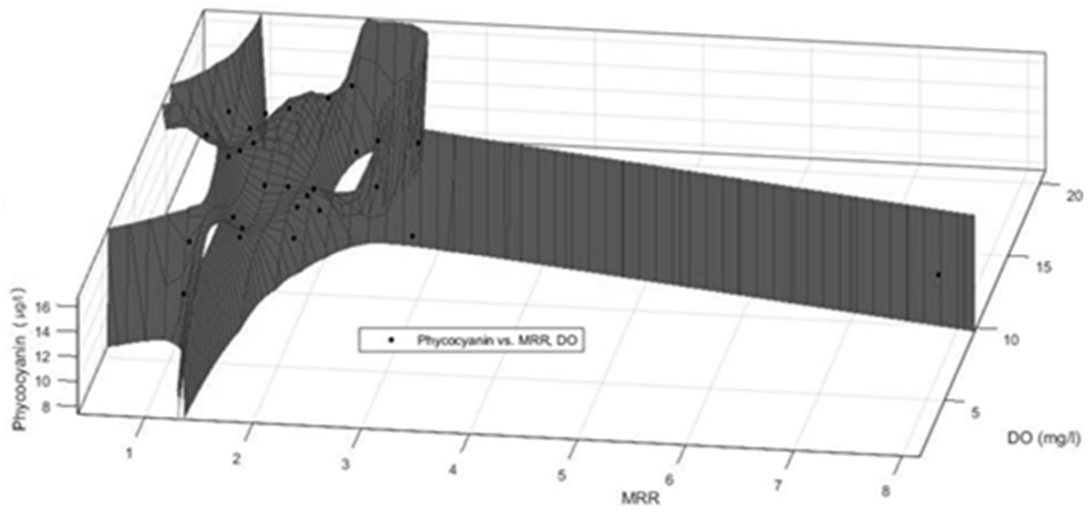

Fig. 7*b*. *Phycocyanin vs MRR and DO (Polynomial degree 5, R^2= 0.9453)*

This finding is interesting as by our field records, the 'optimal' values of DO can be observed when algae growing are around 9 to 14 mg/l. However, DO is a complex factor to be counted in the model and needs to be more studied as it shows in the same time cause and effect, i.e. DO contributes to the environmental conditions for algal development, and once algae grow and bloom, they also increase DO by their photosynthesis process. The change DO values shows us how algae consume or produce oxygen.

4. Conclusions

With the data from two years (2015–2016), ML showed a moderately eutrophic level and contained potential toxic algal species. The formation of blooms can occur only when an optimal combination of favorable conditions including micronutrients, water temperature, etc. Two main categories of blooms were observed in ML seasonally – one normal category of algal bloom from July to September, formed with non-toxic species *Mougeotia sp.* and another one with

Anabaena sp., specifically *A. planctonica* (*Dolicospermum planctonicum*), which started in September and lasted until December, and this species generated blooms in the fall term.

Difference between field data regression and simulated regression are due to:

- High nonlinearity of the problem and hence of data.
- The coupling effects between involving parameters are very visible via the simulated data. We used here the combination of only 3 factors and noticed the discrepancy between the single effect and combined effects between parameters.

The research throughout this paper showed an important insight: bloom patterns can only be explained and predicted by coupling effects of all involved parameters. To combine all effects of all possible parameters, only a mathematical model can help us to deal with this complex issue. The mathematical approach can be a powerful tool to deal with the coupling effects of governing parameters in the bloom occurrence issue. Therefore, the more numerous factors that are combined, the better simulation and hence prediction are. Our main target is to define the declining coupling factor(s) for the pattern HAB which cannot be based on just one or two parameters, but a combination of many involving ones. The threshold of pattern will be hence elucidated only with a mathematical model that we will present in the future.

Although our simulated results herein are not very sophisticated yet, they can show the feasibility and robustness of the mathematical model for the difficult task to predict blooms under coupling effects of all involved physical, chemical and biological factors. And we need more data to process and improve a precise simulation/prediction in the future.The ANN model can be applied to other parameters involving in the HAB pattern, and the simulation and hence prediction will be much more precise and reliable (with the same above structure, there will be more inputs representing involving factors and more outputs standing for effects/ consequences). However, more data also would contribute to the precision of the model.

Finally, the value MRR between 3.25–3.5 can be considered as the threshold for the HAB pattern and Modified Redfield Ratio showed a reasonable reliability to be used for estimating quickly the HAB risks.

Although the original Redfield ratio, the molar index applied in marine water is still discussed, the MRR based on Nitrates/Phosphates ratio we suggested herein seems to have many practical aspects for fresh water to evaluate the algal instability state leading to the onset of bloom patterns. It needs to be more studied and validated by different data sources.

Acknowledgements

TNQ acknowledges the Natural Science and Engineering Research Council of Canada via two grants NSERC RGPIN 03796 and NSERC EGP 484647, Canada Foundation for Innovation as well as Nova Scotia Research and Innovation Trust via Grant No 31188 for equipment. His students are also acknowledged for their contribution in the fieldworks and data treatments (Nadeem, TD, and Van Hai). We acknowledge the ML Steward Committee, especially Bob Maclean and Blaine Odgen for their collaborations and accompaniment during the field trips.

References

[1] Brient, L., Lengronne, M., Bertrand, E., Rolland, D., Sipel, A., Steinmann, D., Baudin, I., Legeas, M. le., Rouzic, B. Le. and Bormans, M. (2008), "A phycocyanin probe as a tool for monitoring cyanobacteria in freshwater bodies", Journal *of Environmental Monitoring,* Vol. 10, pp. 248–255.

[2] Brylinsky, M. (2012), An Assessment of the Sources and Magnitudes of Nutrient Inputs Responsible for Degradation of Water Quality in Seven Lakes Located Within the Carleton River Watershed Area of Digby and Yarmouth Counties. *Nova Scotia Environment, Wolfville Nova Scotia.*

[3] Brylinsky, M. (2009), Lake Utopia Water Quality Assessment. *Commissioned by the New Brunswick Department of the Environment.*

[4] Brylinsky, M. and Sollows, J. (2014), Results of the 2013 Water Quality Survey of Eleven Lakes Located in the Carleton River Watershed Area of Digby and Yarmouth Counties, Nova Scotia.

[5] Bulgakov, N. G. and Levich, A. P. (1999), "The Nitrogen/Phosphorus ratio as a factor regulating phytoplankton community structure". *Archiv für Hydrobiologie*, Vol. 146 No.1, pp. 3–22.

[6] Chorus, I. and Bartram, J. (1999), Toxic Cyanobacteria: A guide to their public health consequences, monitoring and management. *WHO (World Health Organization),* E & FN Spon. London.

[7] Hushchyna, K. and Nguyen-Quang T. (2017), Using the Modified Redfield Ratio to estimate Harmful Algal Blooms. Environmental Problems, 2 (2), pp. 101–108.

[8] McCulloch, W. S. and Pitts, W. (1943), "A logical calculus of the ideas immanent in nervous activity", *The bulletin of mathematical biophysics,* Vol. *5,* No. 4, pp. 115–133.

[9] Nguyen-Quang, T. (2015), "The Mattatall Lake Algal Bloom Study", scientific report. BBML (Biofluids and Biosystems Modeling Lab), Department of Engineering-Dalhousie University Faculty of Agriculture.

[10] Novotny, V. and Olem, H. (1994), "*Water Quality: Prevention, Identification, and Management of Diffuse Pollution*". Wiley, 13 Jan – Technology & Engineering

[11] Redfield, A. (1958), "The biological control of chemical factors in the environment", *Am. Sci.* Vol. 46, pp. 205–220.

[12] WHO. (2003), Guidelines for safe recreational water environments. *World Health Organization.* Vol. 1: Coastal and fresh waters, 253 p.