Desalination and Water Treatment
www.deswater.com

doi: 10.1080/19443994.2016.1190106

57 (2016) 26773–26782
November

Taylor & Francis
Taylor & Francis Group

Determination of the forecasting-model parameters by statistical analysis for development of algae warning system

Gooyong Lee[a], Faridah Othman[b], Shaliza Ibrahim[c], Min Jang[c,*]

[a]*Water Research Center, Institute of Research Management and Monitoring, University of Malaya (UM), Kuala Lumpur, Malaysia*
[b]*Faculty of Engineering, Department of Civil Engineering, University of Malaya (UM), Kuala Lumpur, Malaysia*
[c]*Department of Environmental Engineering, Kwangwoon University, Seoul, Republic of Korea, Tel. +82 2 940 5125; Fax: +82 2 911 2033; emails: minjang@kw.ac.kr, heejaejang@gmail.com*

Received 15 December 2015; Accepted 29 December 2015

ABSTRACT

The aim of this study is to determinate optimal model parameters for prediction of long-term forward (>1 month) chlorophyll-a (Chl-a) concentration in lakes. To optimize model parameters, water quality data from 93 lakes in South Korea were collected and analyzed. Among the 93 lakes, 30 problematic lakes were selected as study sites. Correlation analysis using Chl-a and other water quality data were conducted, and the results indicated that electrical conductivity (EC) and turbidity are important key parameters, which are less considerable than in previous research. To verify effectiveness of the selected parameters, one-month forward prediction of Chl-a concentration was performed using water quality data from the most problematic lakes in South Korea. Artificial neural networks were used as a prediction model. The results of Chl-a prediction using selected parameters showed higher accuracy compare to using general parameters based on the literature reviews. EC and turbidity are important parameters, showing high correlation with Chl-a. This study will corroborate effective model parameters to predict long-term Chl-a concentration in lakes.

Keywords: Chlorophyll-a forecasting; Long-term forecasting; Correlation analysis; Artificial neural networks; Algae early warning system

1. Introduction

Algal bloom problems in lakes have been widely reported, becoming a serious environmental issue because of their harmfulness to ecosystems, causing huge economical loss. To prevent algal bloom, an algae early warning system is one of effective strategies in terms of proactive management concept. An algae-alert system, similar to an algae early warning system, is a forecasting and warning system to prevent damages caused by the massive algal bloom by prediction of Chl-a concentration and proactive strategies. The system has been applied in European, North American, and Asian countries. In South Korea, an alert system was implemented in 1998 at four major lakes used as a drinking water resource. After that, the system was installed to 22 lakes until 2013, and will soon be applied to almost all lakes in South Korea.

*Corresponding author.

Presented at the 8th International Desalination Workshop (IDW) 2015 November 18–21, 2015, Jeju Island, Korea

Algae alert system must have three functions; real-time analysis of water quality, forecasting of major water quality parameter, and warning function. Currently, real-time analysis and warning function have been developed rapidly. For example, more than 70 auto-analysis systems already exist in South Korea. However, forecasting system for water quality is badly applied because of lack of effective forecasting models. Several studies have been conducted to simulate algae concentration, optimize model parameters, develop prediction models, and finally support the algae alert system. Selection of model parameters is a key challenge for high prediction accuracy of models [1–6]. Generally, phosphorous and nitrogen compounds are considered the main parameters related to lake eutrophication, so they are used in many studies [1–3]. The authors absolutely agree with previous studies because there are plenty of references to verify the relationships between nitrogen, phosphorous, and lake eutrophication. However, this study started with a different view point, positing that environment and land usage differ from country to country, so specific eco-condition should be considered based on water quality data. To identify correlation between variables, correlation analysis, factor analysis, and regression analysis are commonly used. Among these methods, correlation analysis is a bivariate analysis that measures the strengths of association between two variables, and has showed excellent results in several studies [4–7].

After selecting effective model parameters, forecasting models should be developed with consideration of field situation. To develop forecasting models, prediction period is also an important factor. Prediction period such as daily or weekly is not helpful for real operations because of the fact that algae prevention technologies such as hazardous control, algae fence, and aeration are impossible to be effective within one week. Obviously, it is possible to predict one-month or one-year forward Chl-a concentration using weekly prediction data, but model uncertainty must be increased in this case. For making proactive technologies effective in reducing eutrophication, the authors believe that one-month (or two-week) forward prediction is needed.

For a direct prediction of one-month forward Chl-a concentration, a deterministic model such as black-box model has got attention in terms of accuracy and fast calculation [8]. Among several black box models, artificial neural networks (ANN) have showed high performance to predict water quality in several studies [1,3,6,9–14]. However, the model parameters and prediction period still need to be developed.

This study was therefore conducted to figure out an appropriate statistic analysis method to select optimal parameters for long-term prediction of Chl-a, and verify effectiveness of the parameters using ANN. Moreover, common parameters in other studies have been reviewed, and compared with results of this study. Finally, proactive strategies using a forecasting model developed in this study are discussed to minimize lake eutrophication.

2. Methodology

2.1. Correlation analysis of water quality data

Correlation analysis methods are generally classified into Pearson correlation analysis, Spearman rank correlation test, and Kendall rank correlation. Pearson correlation analysis is widely used in statistics to measure the degree of the relationship between linear related variables, while spearman rank correlation is a non-parametric test that is used to measure the degree of association between two variables. Spearman rank correlation test does have any assumptions about the distribution of the data and is the appropriate correlation analysis when the variables are measured on a scale that is at least ordinal. Simply saying, Pearson correlation benchmarks linear relationship, and Spearman correlation benchmarks monotonic relationship between parameters. Kendall rank correlation is also a non-parametric test that measures the strength of dependence between two variables. The main difference between Spearman correlation and Ketal correlation is that Spearman method tends to match with power function while Kendall method is rather intended to other nonlinear correlations. Considering the usage of each methods and literature reviews [6,15], Pearson correlation analysis was used in this study. The general formula of Pearson's correlation coefficient "r" is provided in Eq. (1):

$$r = r_{xy} = \frac{\sum_{i=1}^{n}\{(x_{\mathrm{i}} - \bar{x})(y_{\mathrm{i}} - \bar{y})\}}{\sqrt{\sum_{i=1}^{n}(x_{\mathrm{i}} - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_{\mathrm{i}} - \bar{y})^2}} \quad (1)$$

where x and y are data-sets containing n values and $\bar{x}$ and $\bar{y}$ are mean value of x and y, respectively.

2.2. Artificial neural networks

ANN have got attention in various engineering communities when models need deterministic prediction and fast calculation in real-time prediction [16–18]. In studies on water resource management, many experts [17,19,20] agreed with the benefits of ANN, such as low data requirements (ex. geological data), and better consideration of non-linear parameters.

Multi-layer perceptron (MLP) is a kind of ANN and have showed good performance in both short- and long-term prediction [2,21–23]. Furthermore, a double hidden layer was matched properly with MLP in several studies [24,25]. Therefore, in this study, MLP with double hidden layers has been tested with various number of nodes (5, 10, 15, and 20) (Table 1), and an optimal node number was selected for prediction of Chl-a at each study site.

For activated function at hidden- and output-layer, sigmoid and linear function were used, respectively [26–28]. To adjust weight, the back-propagation algorithm as first suggested by Rumelhart and McClelland [29] was used because it is an adequate method to train the MLP, and particularly it is less sensitive to the noises or errors inherent of parameters. For validating model accuracy, three criteria were used in this study (Table 2). In Table 2, P is the result of prediction, O is the measured values from sites, $\bar{O}$ is the average of O values, and n is the number of samples. Data were adjusted on a scale ranging from 0 to 1, and then used for input and output values.

2.3. Study sites

The parameters for prediction of Chl-a are various and show different aspect depending on geological and environmental conditions. In this study, to optimize the parameters related with Chl-a, survey and data analysis were conducted for 93 lakes, which are almost all of the lakes in South Korea. Among the lakes, 30 lakes were selected as shown in Fig. 1 in terms of amount of accessible data. Water quality data were measured monthly by Korea Nation Institute of Environmental Research, and opened to public since 2005. Fifteen types of water quality data such as temperature (temp.), pH, dissolved oxygen (DO), chemical oxygen demand, biological oxygen demand (BOD), suspended solid (SS), nitrate (NO_3-N), dissolved total nitrogen, ammoniacal nitrogen (NH_3-N), total nitrogen (TN), total phosphorus (TP), phosphate phosphorus (PO_3-P), dissolved total phosphorus, electrical conductivity (EC), and Chl-a were collected. Six years of data (2006–2011) were used. Flow and rainfall data were not used in this study because hydraulic retention time changed rapidly, and also these obviously showed low correlation with Chl-a in preliminary tests. Similar results that hydraulic retention time is not the main parameter for prediction of Chl-a was obtained in other researches [6,30]. All data were subdivided into training-validation (4 years) and testing period (2 years). The four-year training-validation period (2006–2009) was divided into training and validation parts. For correlation analysis, water quality data from upstream and water intake point were used. For some intake point where upstream exists more than two, water quality data of upstream were averaged and used.

3. Results and discussions

3.1. Key parameters for Chl-a prediction

Based on the literature review, 13 of effective parameters for prediction of Chl-a concentration are summarized in Table 3. SS and turbidity were counted together, and expressed as "Tun.SS." In the literature, 96% of articles agree that temperature (Temp.) of water is the most important parameter, while pH, DO, TP, and PO_4-P were also considered as key parameters in that order. It is noteworthy that the meaning of total (%) is numbers of use in the literature. It does not mean its importance, but the parameters with high percentage have big potential to be valuable parameters for prediction. Results of correlation analysis in this study are shown in Table 4, while Fig. 2 illustrates comparisons between literature reviews and results of correlation analysis.

Results of correlation analysis showed same and different aspects compared to previous studies

Table 1
ANN structure for one-month forward prediction of Chl-a

Composition of ANN	Description
Structure type	Multi-layer perceptrons
Number of hidden layer	2
Number of nodes at first hidden layer	5, 10, 15, 20
Number of nodes at second hidden layer	5, 10, 15, 20
Activated function at hidden layer	Sigmoid function, $f(x) = 1/(1 + e^{-x})$
Activated function at output layer	Linear function ($f(x) = x$)
Weight adjust algorithm	Back-propagation algorithm

Table 2
Criteria for testing validity of the model

Criteria	Purpose	Estimation
Coefficient of determination (R^2)	To evaluate the goodness-of-fit of models	$R^2 = \frac{\sum_{i=1}^{n}(P_i - \bar{O})^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2}$
Mean square error (MSE)	To quantify the error of models	$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2$
Mean percentage error (MAE)	To compare the error between populations	$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\frac{P_i - O_i}{O_i}$

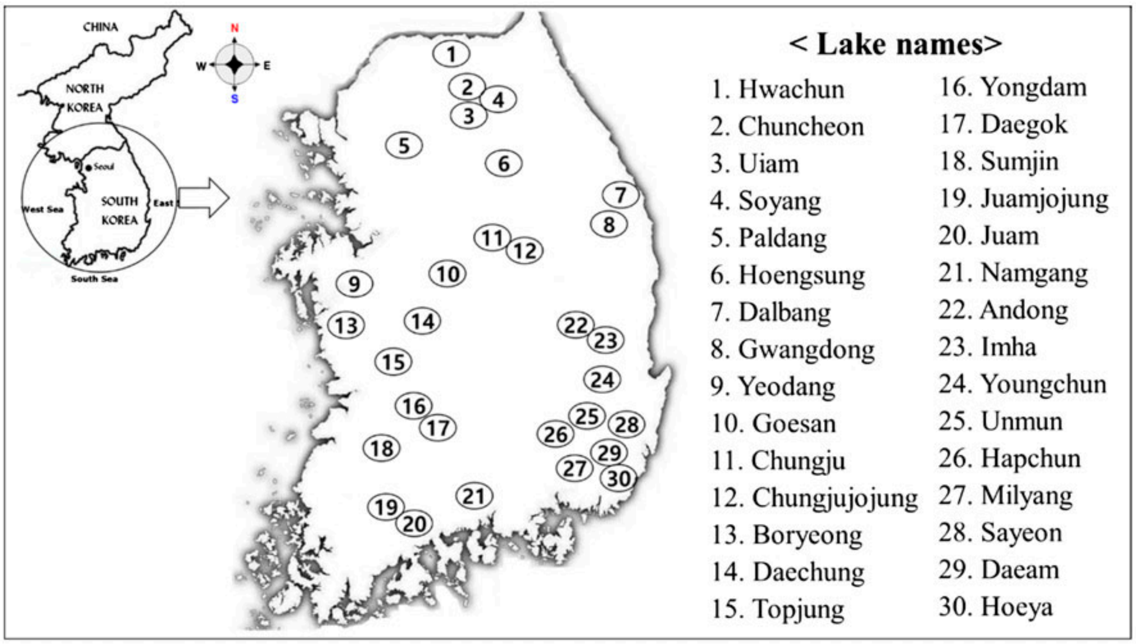


Fig. 1. Study sites in this study.

(Fig. 2). First of all, temperature of water was the most important parameter in both the literature review and correlation analysis. Secondly, DO ranked third in literature review, while holding the second rank in correlation analysis. Many studies agree that temperature and DO show high relationship with Chl-a [3,9,11,13,31]. Lake eutrophication generally increases with temperature because algae grow rate is higher at higher temperatures. The increase in DO concentration can increase algae concentration [13,32]. Specially, Wei et al. [13] found increasing DO concentration will stimulate the growth of Phormidium and Synedra. Therefore, temperature and DO are regarded as parameters which are highly correlated with Chl-a in most environments.

One distinctive finding from correlation analysis concerns pH. Correlation analysis results in this study showed a low correlation of pH with Chl-a, though pH is high influential and the second important parameter in literature reviews. The correlation between pH and Chl-a is high in only 5 lakes among 30 study sites. Generally, pH is closely related to Chl-a because algae use carbon dioxide as a carbon source during photosynthesis, and carbon dioxide in water affects pH [33]. The possible reason for low correlation of pH and Chl-a in this study is the lag-time which means time difference between parameters. A Chl-a concentration at the study sites was used for one-month before pH at upstream for correlation analysis. This finding has two meanings: one thing is that pH is not prior factor for one-month Chl-a prediction. The other factor is that the effect of carbon dioxide to pH is not worth to consider for long-term prediction.

The other remarkable thing from correlation analysis results in this study is that the correlation with Chl-a to EC and turbidity is higher than phosphorous- and nitrogen-compounds in many study sites. Phosphorous and nitrogen compounds are regarded as very important factors in the literature reviews because these two parameters are essential nutrients for algal growth [1,5,6,14]. However, several studies also agree with the importance of EC [3,6,21,34], while turbidity has been

Table 3
Literature reviews of related parameters with chlorophyll a concentration

Refs.	Temp.	pH	DO	TP	PO_4-P	TN	NO_3-N	NH_3-N	Clean.	EC	Tub.SS	COD	BOD
[3]	O	O	O	–	O	–	O	–	O	O	O	–	–
[4]	O	–	–	O	–	O	–	–	–	–	–	–	–
[6]	O	O											
[7]	O	–	–	O	–	O	–	–	O	–	–	–	–
[9]	O	O	O	–	–	–	O	O	O	–	–	–	–
[10]	O	O	O		O		O	O		O	O	–	–
[11]	O	O	O	–	O	–	O	–	O	–	O	–	–
[12]	O	–	O	–	O	–	–	–	O	–	–	–	–
[13]	O	O	O	O	–	O	–	–	–	–	–	O	–
[14]	O	–	–	–	–	–	–	–	O	–	–	–	–
[31]	O	O	O	O	–	O	–	O	O	–	–	O	O
[36]	O	O	–	–	O	–	O	O	–	–	O	–	–
[37]	O	O	O	O	O	–	O	O	O	O	–	–	–
[38]	O	–	–	O[a]	–	O[a]	–	–	–	–	–	–	–
[39]	O	O	O	–	–	–	–	–	–	–	–	–	–
[40]	O	–	–	O	–	–	–	–	–	–	–	–	–
[41]	O	–	–	O	–	–	–	O	O	–	–	–	–
[34]	O		O				O	O				O	O
[42]	O	O	O	O	O	–	O	O	–	–	–	–	O
[43]	O	O			O		O	O	O		O		
[44]		O		O		O	O	O		O	O	O	
[45]	O		O	O	O	O	O	O					
[46]	O	O	O		O		O				O		O
[35]	O	O	O	O	O	O	O	O	O		O	O	O
[b]Total (%)	96	63	58	50	46	29	54	50	46	17	33	21	21

[a]Meaning of TP and TN in Chen et al. [38] is total inorganic phosphorus and inorganic nitrate, respectively.
[b]Meaning of "Total" is the percentage value of "numbers of use per total numbers of literature".

used as an input parameter in many studies [13,21,35]. A special meaning of correlation analysis results is that phosphorous and nitrogen are high correlated with Chl-a, but not the essential factors which should be considered for all sites. Furthermore, EC and turbidity are important factors related to Chl-a, and it would be preferable to consider EC and turbidly for simulation of Chl-a concentration.

Therefore, similar to the literature review, it has been confirmed that temperature and DO are highly correlated with Chl-a in the results of correlation analysis in this study. A singularity is EC and turbidity, with higher correlation than phosphorous and nitrogen, and pH which is regarded as an important parameter in literature showed a low correlation with Chl-a in South Korea.

3.2. Prediction of 1-month forward Chl-a concentration

Based on the results of literature reviews and correlation analysis, prediction of one-month forward Chl-a concentration was conducted using ANN. There are three reasons why one-month prediction was performed in this study. Firstly, water quality data are measured every month in study sites. Secondly, a short forecasting period (less than two weeks) is not enough to prevent algae bloom by using proactive strategies such as algae fence and aeration. Lastly, the current algae alert system in South Korea has been operated based on monthly water quality. Therefore, prediction of one-month forward Chl-a concentration was carried out in this study.

Generally, it is hard to predict one-month after Chl-a concentration directly using governing equations (ex. Darcy's Law, Lagrange's equations, and etc.) or regression analysis (ex. linear regression, multiple regression analysis, ordinary least squares regression, etc.). The main reason is the difficulty in making a formula including all parameters due to various environmental conditions related to eutrophication. Obviously, it is possible to predict one-month to one-year using governing equations or regression analysis. However, this also requires assumptions of nature or stepwise prediction, which can increase uncertainty of models. Therefore, box-box models which do not require governing equations are preferred to predict Chl-a concentration.

Table 4
Correlation analysis results between Chl-a and other parameters

Lake	Temp.	pH	DO	TP	PO_4-P	TN	NO_3-N	NH_3-N	Clean.	EC	Tub.SS	COD	BOD
1	O												
2	O		O										
3	O							O		O			
4	O		O										
5		O	O			O	O				O	O	O
6	O		O	O	O			O			O	O	
7	O		O										
8	O		O	O	O								
9	O	O	O			O				O			
10	O		O					O			O		
11	O					O							
12	O						O				O	O	O
13	O			O		O		O				O	O
14	O	O	O			O	O			O	O		
15	O	O				O						O	O
16	O		O										
17	O		O			O	O						
18	O												
19	O		O							O			
20	O												
21	O		O	O	O		O		O	O	O	O	
22	O		O	O	O					O	O		
23	O		O		O					O	O		
24	O		O	O	O					O	O		
25		O						O					
26	O		O							O			
27	O		O	O	O		O					O	O
28	O		O			O	O	O			O		
29	O		O	O	O					O			
30	O		O	O	O	O	O			O			O
Total (%)	93	17	70	30	30	30	27	17	3	33	33	23	20

The lake where there were the most serious problems with Chl-a in the last decade was selected as a study site. In the algae alert system in Korea, the first warning starts when Chl-a concentration is more than 15 mg/m^3 and the second warning starts when Chl-a concentration is more than 25 mg/m^3, while the final warning level is Chl-a concentration at more than 100 mg/m^3.

Water quality of Hoeya Lake is a big problem because of high Chl-a concentration, which was over the second warning level every year during the study period (2006–2011), and 60% of the second warning levels reached the final warning level. Therefore, Hoeya Lake was selected as the case study site in this study. Hoeya Lake is used as a drinking water source and has an upstream named Hoeya river. The catchment area is 125.5 km^2 while lake surface area is 2.37 km^2. Effective storage volume is 17,700,000 m^3.

Correlation analysis results of Hoeya Lake and Hoeya river are shown in Table 5. Current data in Hoeya river and one-month after data in Hoeya Lake were compared by correlation analysis results. Within 0.01 of significance probability, temperature, DO, BOD, EC showed high correlation with one-month after Chl-a. Additionally, NO_3-N, TN, PO_4-P, and TP showed high correlation within 0.05 of significance probability. The remarkable thing is that EC and BOD are more strongly correlated with Chl-a than nitrogen and phosphorous.

Based on results of correlation analysis, three types of ANN were designed. Differences between ANN models are input parameters. Based on the following rules, input parameters are listed in Table 6.

C-ANN: Water quality data (eight data are selected in order of frequent usage) in Table 3 are used as input parameters.

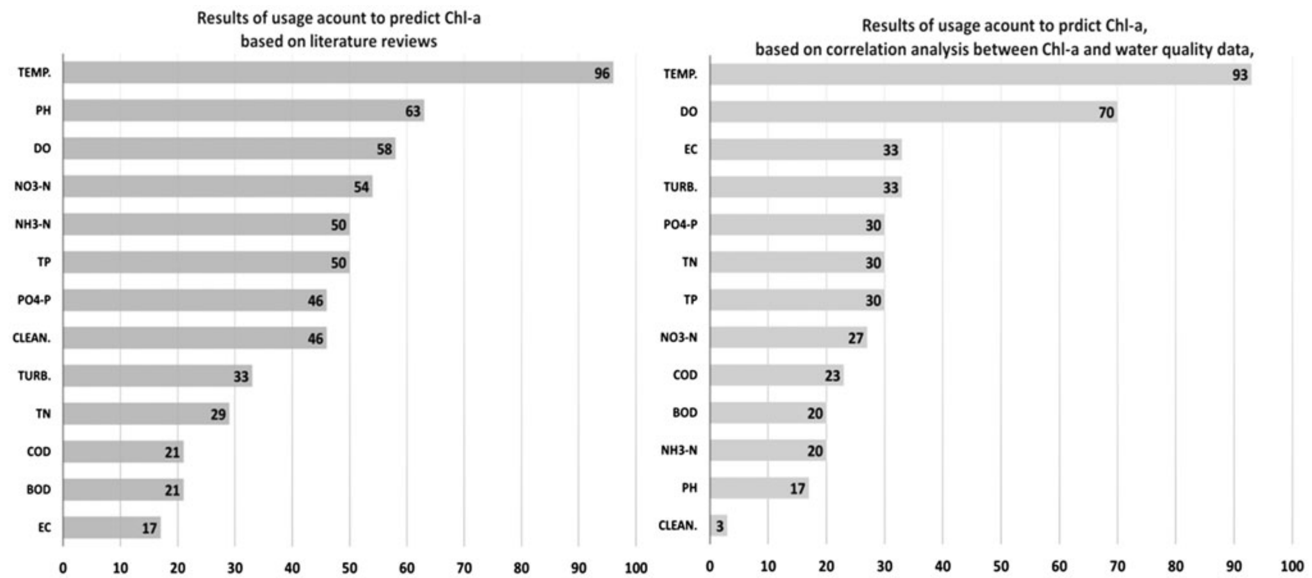


Fig. 2. Comparison between literature reviews and results from correlation analysis.

Table 5
Correlation analysis results of water quality data in Hoeya Lake

Water-quality data	Pearson correlation coefficient (r)	Significance probability ($=\rho$)
Temp.	−0.556	0.000
DO	0.396	0.001
pH	−0.156	0.190
BOD	−0.249	0.005
COD	−0.133	0.265
SS	−0.136	0.256
NH_3-N	−0.027	0.825
NO_3-N	0.395	0.041
TN	0.365	0.022
PO_4-P	0.375	0.021
TP	0.324	0.015
EC	0.326	0.003

Table 6
ANN input and output parameters in Hoeya Lake

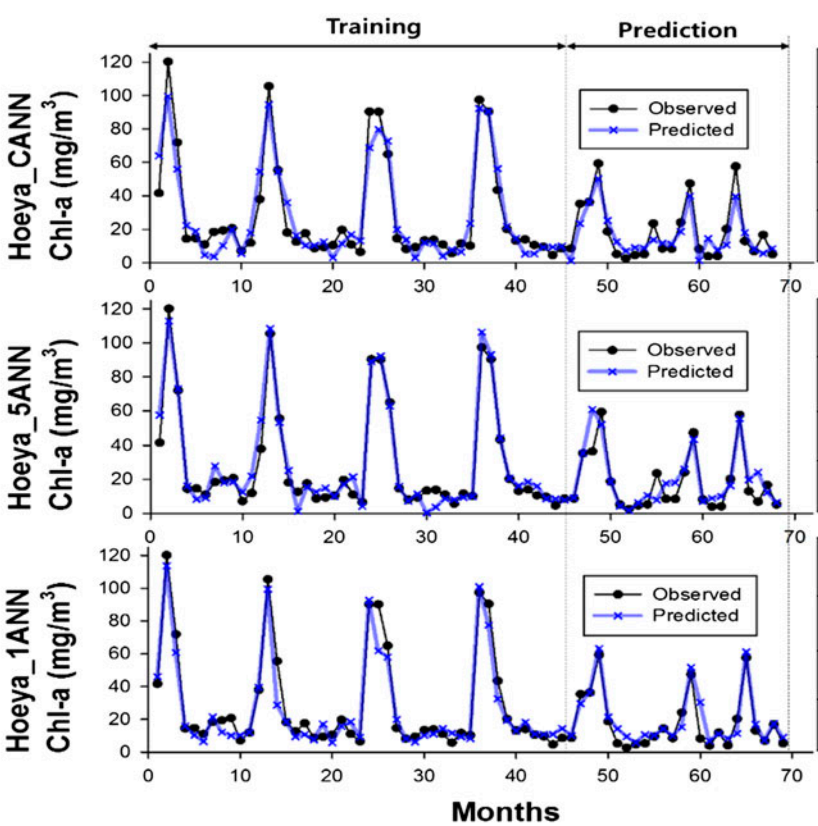
Parameters		Description
Input water quality data	C-ANN	Temp., pH, DO, NO_3-N, NH_3-N, TP, PO_4-P, Clean.
	5-ANN	Temp., DO, BOD, EC, NO_3-N, TN, PO_4-P, TP
	1-ANN	Temp., DO, BOD, EC
Output data	1-month forward Chl-a concentration	

5-ANN: Water quality data showing lower than 0.05 of significance probability in Table 5 are used as input parameters.

1-ANN: Water quality data showing lower than 0.01 of significance probability in Table 5 are used as input parameters.

Results of the one-month forward prediction of Chl-a in Hoeya Lake using ANN models are shown in Fig. 3. Overall, prediction results showed high accuracy in terms of high coefficient of determination (=R^2) recorded greater than 0.91 and 0.80 in training and prediction periods, respectively. 5-ANN showed the highest prediction accuracy. Parameters within significance probability 0.05 are used as input parameters for 5-ANN. It has been shown that high correlation parameters strongly effect Chl-a prediction. 1-ANN showed the second highest prediction accuracy. The reason why the prediction accuracy is lower than 5-ANN despite higher correlation is because of small numbers of input parameter. In other words, it is possible to simulate Chl-a concentration with only temperature, DO, BOD, and EC in a study site, but adding parameters of NO_3-N, TN, PO_4-P, and TP makes for a more accurate simulation. However, as C-ANN shows in Fig. 3, many parameters without statistical analysis can cause low prediction accuracy. In addition, MSE is one of the important criteria which increases much more for errors at high peak point. In Fig. 3, MSE of C-ANN in the training part is 2–3 times higher than 1-ANN or 5 ANN. One of the most important things in order to practically apply prediction models is exactly predicting peak time and damage of event. In this aspect, pH, which has been used in other studies, is a hindrance to prediction in the study sites of this study. So, it is confirmed that using input data based on personal knowledge or experiences without statistical analysis rather reduce prediction accuracy. Correlation analysis, factor analysis, and regression analysis are used to analyze connectivity between parameters and it has been proven that correlation analysis shows good efficiency to select parameters for the prediction of Chl-a, while ANN shows high accuracy of one-month forward prediction of Chl-a.

The results of this study can provide one-month Chl-a concentration, which is necessary for effective operation of an algae early warning system. Using one-month after Chl-a concentration, proactive water quality management is possible. An operator of lake management systems should consider various conditions of lakes and prepare several proactive strategies such as algae removal boats, algae fences, aeration systems, and others. After preparing prevention plans, application order of the prevention technologies



Training		Prediction	
R^2	0.91	R^2	0.80
MSE	91.7	MSE	55.8
MAE	7.53	MAE	6.35

Training		Prediction	
R^2	0.96	R^2	0.87
MSE	37.6	MSE	43.2
MAE	4.50	MAE	4.73

Training		Prediction	
R^2	0.93	R^2	0.86
MSE	59.6	MSE	36.8
MAE	5.25	MAE	3.86

Fig. 3. Results of one-month forward prediction of Chl-a in Hoeya Lake using C-ANN, 5-ANN, and 1-ANN.

should be decided. Using forecasting models such as ANN, once a risky level of Chl-a concentration is predicted, the operator has to suggest possible strategies to decision-makers, and they must select effective actions for algae reduction. Finally, the authors believe that successful long-term forecasting may support an algae early warning system, and that the statistical analysis and ANN developed in this study show strong potential to develop Chl-a forecasting models.

4. Conclusion

Selecting model parameters is considered a very essential part to design forecasting models. Optimization of the model parameters is therefore important. To determinate optimal model parameters to predict long-term forward chlorophyll-a concentration in lakes, 30 problematic lakes have been selected as study sites among 93 lakes in South Korea. Correlation analysis between Chl-a and other water quality data was carried out, and the results represented that temperature and DO have high correlation with Chl-a. A remarkable finding of this study is that EC and turbidity showed higher correlation with Chl-a compared to phosphorous and nitrogen. Furthermore, pH, which is regarded as an important parameter in the literature, showed a low correlation with Chl-a in South Korea. Although this parameter, which has a high correlation with Chl-a, does not guarantee good performance for Chl-a simulation, it has high potential to increase model accuracy as an input parameter. To verify the effectiveness of selected parameters as forecasting models inputs, ANN was applied to simulate one-month forward Chl-a concentrations, and the parameters were used as input valuables for ANN. The results of Chl-a prediction verified that high correlated parameters showed better performance than general parameters selected based on literature or modeler's knowledge. Finally, correlation analysis is a preferable statistical method which may be used to select optimal parameters for Chl-a simulation, while ANN showed high accuracy for simulation of Chl-a. Furthermore, EC and turbidity were highlighted as important parameters in this study.

Acknowledgments

This research was supported by Fundamental Research Grants (FRGS) funded by the Ministry of Higher Education in Malaysia (Grant No. FP016-2014A).

References

[1] S. Huo, Z. He, J. Su, B. Xi, C. Zhu, Using artificial neural network models for eutrophication prediction, Procedia Environ. Sci. 18 (2013) 310–316.
[2] I.E. Mulia, H. Tay, K. Roopsekhar, P. Tkalich, Hybrid ANN–GA model for predicting turbidity and chlorophyll-a concentrations, J. Hydro-environ. Res. 7 (2013) 279–299.
[3] C. Karul, S. Soyupak, A.F. Çilesiz, N. Akbay, E. Germen, Case studies on the use of neural networks in eutrophication modeling, Ecol. Modell. 134 (2000) 145–152.
[4] Q. Chen, A.E. Mynett, Integration of data mining techniques and heuristic knowledge in fuzzy logic modelling of eutrophication in Taihu Lake, Ecol. Modell. 162 (2003) 55–67.
[5] Y. Park, K.H. Cho, J. Park, S.M. Cha, J.H. Kim, Development of early-warning protocol for predicting chlorophyll-a concentration using machine learning models in freshwater and estuarine reservoirs, Korea, Sci. Total Environ. 502 (2015) 31–41.
[6] S. Cho, B. Lim, J. Jung, S. Kim, H. Chae, J. Park, S. Park, J.K. Park, Factors affecting algal blooms in a man-made lake and prediction using an artificial neural network, Measurement 53 (2014) 224–233.
[7] J. Jiang, P. Wang, Z. Tian, L. Guo, Y. Wang, A comparative study of statistical learning methods to predict eutriphication tendency in a reservoir, northeast China, in: Mechanic Automation and Control Engineering (MACE), 2011 Second International Conference on, IEEE, 2011, pp. 1883–1886.
[8] H.R. Maier, A. Jain, G.C. Dandy, K.P. Sudheer, Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions, Environ. Modell. Software 25 (2010) 891–909.
[9] K.-S. Jeong, G.-J. Joo, H.-W. Kim, K. Ha, F. Recknagel, Prediction and elucidation of phytoplankton dynamics in the Nakdong River (Korea) by means of a recurrent artificial neural network, Ecol. Model. 146 (2001) 115–129.
[10] K.-S. Jeong, D.-K. Kim, G.-J. Joo, River phytoplankton prediction model by Artificial Neural Network: Model performance and selection of input variables to predict time-series phytoplankton proliferations in a regulated river system, Ecol. Inf. 1 (2006) 235–245.
[11] J.-T. Kuo, M.-H. Hsieh, W.-S. Lung, N. She, Using artificial neural network for reservoir eutrophication prediction, Ecol. Modell. 200 (2007) 171–177.
[12] J.H. Lee, Y. Huang, M. Dickman, A. Jayawardena, Neural network modelling of coastal algal blooms, Ecol. Modell. 159 (2003) 179–201.
[13] B. Wei, N. Sugiura, T. Maekawa, Use of artificial neural network in the prediction of algal blooms, Water Res. 35 (2001) 2022–2028.
[14] H. Wilson, F. Recknagel, Towards a generic artificial neural network model for dynamic predictions of algal abundance in freshwater lakes, Ecol. Modell. 146 (2001) 69–84.
[15] L. Velo-Suárez, J. Gutiérrez-Estrada, Artificial neural network approaches to one-step weekly prediction of Dinophysis acuminata blooms in Huelva (Western Andalucía, Spain), Harmful Algae 6 (2007) 361–371.

[16] D.-S. Joo, D.-J. Choi, H. Park, The effects of data preprocessing in the determination of coagulant dosing rate, Water Res. 34 (2000) 3295–3302.
[17] L. Gooyong, L. Sangeun, P. Heekyung, Improving applicability of neuro-genetic algorithm to predict short-term water level: A case study, J. Hydroinf. 16 (2014) 218–230.
[18] D.-J. Choi, H. Park, A hybrid artificial neural network as a software sensor for optimal control of a wastewater treatment process, Water Res. 35 (2001) 3959–3967.
[19] Ö. Kişi, River flow modeling using artificial neural networks, J. Hydrol. Eng. 9 (2004) 60–63.
[20] D. Cameron, P. Kneale, L. See, An evaluation of a traditional and a neural net modelling approach to flood forecasting for an upland catchment, Hydrol. Processes 16 (2002) 1033–1046.
[21] W.A. Young, D.F. Millie, G.R. Weckman, J.S. Anderson, D.M. Klarer, G.L. Fahnenstiel, Modeling net ecosystem metabolism with an artificial neural network and Bayesian belief network, Environ. Modell. Software 26 (2011) 1199–1210.
[22] M. Awad, Sea water chlorophyll-a estimation using hyperspectral images and supervised Artificial Neural Network, Ecol. Inf. 24 (2014) 60–68.
[23] D. Seo, M. Kim, J.H. Ahn, Prediction of chlorophyll-a changes due to weir constructions in the nakdong river using EFDC-WASP modelling, Environ. Eng. Res. 17 (2012) 95–102.
[24] G. Sahoo, C. Ray, E. De Carlo, Use of neural network to predict flash flood and attendant water qualities of a mountainous stream on Oahu, Hawaii, J. Hydrol. 327 (2006) 525–538.
[25] W. Wang, P.H. Gelder, J. Vrijling, J. Ma, Forecasting daily streamflow using hybrid ANN models, J. Hydrol. 324 (2006) 383–399.
[26] A. Abebe, R. Price, Managing uncertainty in hydrological models using complementary models, Hydrol. Sci. J. 48 (2003) 679–692.
[27] M. Aqil, I. Kita, A. Yano, S. Nishiyama, A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behaviour of runoff, J. Hydrol. 337 (2007) 22–34.
[28] V. Chandramouli, P. Deka, Neural network based decision support model for optimal reservoir operation, Water Resour. Manage. 19 (2005) 447–464.
[29] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning internal representations by error propagation, in: 40 DTIC Document, 1985, pp. 318–362.
[30] S. Malek, S.M.S. Ahmad, S.K.K. Singh, P. Milow, A. Salleh, Assessment of predictive models for chlorophyll-a concentration of a tropical lake, BMC Bioinf. 12(Suppl. 13) (2011) S12.
[31] X.-L. Wang, Y.-L. Lu, G.-Z. He, J.-Y. Han, T.-Y. Wang, Exploration of relationships between phytoplankton biomass and related environmental variables using multivariate statistic analysis in a eutrophic shallow lake: A 5-year study, J. Environ. Sci. 19 (2007) 920–927.
[32] C. Sivapragasam, N. Muttil, S. Muthukumar, V. Arun, Prediction of algal blooms using genetic programming, Mar. Pollut. Bull. 60 (2010) 1849–1855.
[33] C.Y. Chen, Effects of pH on the growth and carbon uptake of marine phytoplankton, Mar. Ecol. Prog. Ser. 109 (1994) 83–94.
[34] K.H. Cho, J.-H. Kang, S.J. Ki, Y. Park, S.M. Cha, J.H. Kim, Determination of the optimal parameters in regression models for the prediction of chlorophyll-a: A case study of the Yeongsan Reservoir, Korea, Sci. Total Environ. 407 (2009) 2536–2545.
[35] G. Wu, Z. Xu, Prediction of algal blooming using EFDC model: Case study in the Daoxiang Lake, Ecol. Modell. 222 (2011) 1245–1252.
[36] Q. Chen, A.E. Mynett, Modelling algal blooms in the Dutch coastal waters by integrated numerical and fuzzy cellular automata approaches, Ecol. Modell. 199 (2006) 73–81.
[37] H. Çamdevýren, N. Demýr, A. Kanik, S. Keskýn, Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs, Ecol. Modell. 181 (2005) 581–589.
[38] Q. Chen, A.E. Mynett, A robust fuzzy logic approach to modelling algae biomass, J. Hydraul. Res. 42 (2004) 303–309.
[39] S. Marsili-Libelli, Fuzzy prediction of the algal blooms in the Orbetello lagoon, Environ. Modell. Software 19 (2004) 799–808.
[40] H.R. Maier, T. Sayed, B.J. Lence, Forecasting cyanobacterium Anabaena spp. in the River Murray, South Australia, using B-spline neurofuzzy models, Ecol. Modell. 146 (2001) 85–96.
[41] Y. Liu, H. Guo, P. Yang, Exploring the influence of lake water chemistry on chlorophyll a: A multivariate statistical model analysis, Ecol. Modell. 221 (2010) 681–688.
[42] B. de Ceballos, A. König, J. de Oliveira, Dam reservoir eutrophication: A simplified technique for a fast diagnosis of environmental degradation, Water Res. 32 (1998) 3477–3483.
[43] H. Cao, F. Recknagel, G.-J. Joo, D.-K. Kim, Discovery of predictive rule sets for chlorophyll-a dynamics in the Nakdong River (Korea) by means of the hybrid evolutionary algorithm HEA, Ecol. Inf. 1 (2006) 43–53.
[44] M.A.E. Forio, D. Landuyt, E. Bennetsen, K. Lock, T.H.T. Nguyen, M.N.D. Ambarita, P.L.S. Musonge, P. Boets, G. Everaert, L. Dominguez-Granda, P.L.M. Goethals, Bayesian belief network models to analyse and predict ecological water quality in rivers, Ecol. Modell. 312 (2015) 222–238.
[45] K. Kim, M. Park, J.-H. Min, I. Ryu, M.-R. Kang, L.J. Park, Simulation of algal bloom dynamics in a river with the ensemble Kalman filter, J. Hydrol. 519, Part D (2014) 2810–2821.
[46] A. Fonseca, C. Botelho, R.A.R. Boaventura, V.J.P. Vilar, Integrated hydrological and water quality model for river management: A case study on Lena River, Sci. Total Environ. 485–486 (2014) 474–489.