Environ. Eng. Res. 2019; 24(3): 404-411
https://doi.org/10.4491/eer.2018.245

pISSN 1226-1025
eISSN 2005-968X

Prediction of short-term algal bloom using the M5P model-tree and extreme learning machine

Hye-Suk Yi[1,2], Bomi Lee[2], Sangyoung Park[2], Keun-Chang Kwak[3], Kwang-Guk An[1†]

[1]*Department of Bioscience and Biotechnology, Chungnam National University, Daejeon 34134, Republic of Korea*
[2]*K-water Research Institute, Korea Water Resources Corporation, Daejeon 34350, Republic of Korea*
[3]*Department of Control and Instrumentation Engineering, Chosun University, Gwangju 61452, Republic of Korea*

ABSTRACT

In this study, we designed a data-driven model to predict chlorophyll-*a* using M5P model tree and extreme learning machine (ELM). The Juksan weir in the Youngsan River has high chlorophyll-*a*, which is the primary indicator of algal bloom every year. Short-term algal bloom prediction is important for environmental management and ecological assessment. Two models were developed and evaluated for short-term algal bloom prediction. M5P is a classification and regression-analysis-based method, and ELM is a feed-forward neural network with fast learning using the least square estimate for regression. The dataset used in this study includes water temperature, rainfall, solar radiation, total nitrogen, total phosphorus, N/P ratio, and chlorophyll-*a*, which were collected on a daily basis from January 2013 to December 2016. The M5P model showed that the prediction model after one day had the highest performance power and dropped off rapidly starting with predictions after three days. Comparing the performance power of the ELM model with the M5P model, it was found that the performance power of the 1-7 d chlorophyll-*a* prediction model was higher. Moreover, in a period of rapidly increasing algal blooms, the ELM model showed higher accuracy than the M5P model.

Keywords: Algal bloom, Chlorophyll-*a*, Extreme learning machine, Juksan weir, M5P model tree, Water quality

1. Introduction

Recently, there have been frequent occurrences of algal blooms owing to climate change and rising water temperatures (WTs), leading to a rising interest in water quality management of rivers and reservoirs owing to physical environmental changes. Such algal blooms not only threaten the aquatic ecosystem but also have direct and indirect influences on human life; therefore, solving this problem is extremely important [1-2]. Algal blooms are influenced by a variety of factors including those influencing water quality, such as increasing nutrients and changing WT, as well as climate factors such as air temperature, solar radiation (SR), and precipitation [3-4]; there are ongoing studies to discover the factors influencing algal blooms. However, there are limitations to presenting clear causal relationships between influential factors pertaining to algal blooms, given a complex web of influence from environmental factors as well as the differences in their characteristics by region and period.

One of algal bloom management methods is to predict through numerical modeling considering various influential factors such as WT and nutrients. However, there are some difficulties associated with numerical modeling such as the requirement of large amounts of data, time to construct the inputs required for modeling, as well as large uncertainties associated with model parameters [5-6]. To resolve these issues, there are ongoing studies on algal bloom prediction using machine learning methods, such as the data-based artificial neural network (ANN) and model trees (MTs). Park et al. [7] used an ANN and a support vector machine (SVM) to predict chlorophyll-*a* concentration for providing early warning in the Juam reservoir and Yeongsan reservoir, which are located in an upstream region (freshwater reservoir) and a downstream region (estuarine reservoir), respectively. Jung et al. [8] tested and proposed M5 MTs using partial least-squares regression (PLSR) on a particular dataset and then compared the results to those obtained using M5 MTs, MLF- and RBF-ANN, and k nearest neighbors (kNN). Ye et al. [9] presented an integrated system for real-time observation,

Received July 23, 2018 **Accepted** October 1, 2018

† Corresponding author
Email: kgan@cnu.ac.kr
Tel: +82-42-821-6408 Fax: +82-42-822-9690

early warning, and forecasting of phytoplankton blooms by integrating automated online sondes and an ecological model. Lee et al. [10] suggested that an ANN model with a small number of input variables can capture the trends of algal dynamics, but data with a minimum sampling interval of one week are necessary. Kim et al. [11] proposed an effective method for establishing algal bloom forecasting models using ANNs. Recently, among the various machine learning methods, the extreme learning machine (ELM) method has been proven to have quick learning speeds and high performance; it is being used for predictive modeling in various areas such as predicting power supply stability, river flooding, and algal blooms. Xu et al. [12] developed an ELM-based predictor for real-time frequency stability assessment to enhance the dynamic security of power systems. Yadav et al. [13] studied a new technique, online sequential extreme learning machine (OS-ELM) that is capable of updating the model equation based on new data entry without much increase in computational cost for flood forecasting; the performance of the OS-ELM was comparable to those of other widely used artificial intelligence (AI) techniques like SVMs, ANNs and genetic programming (GP). Lou et al. [14] attempted to develop an ELM-based predictive model to simulate the dynamic change in phytoplankton abundance in Macau reservoir, given a variety of water variables. Boyer et al. [15] assessed the chlorophyll-*a* indicator as being relevant and reflecting the state of the Florida Bay ecosystem; this indicator is sensitive to ecosystem drivers (stressors, especially nutrient loading), feasible to monitor, and scientifically defensible.

Algal blooms tend to change over short terms given climatic conditions, polluting matter, and hydraulic characteristics; as such, these characteristics have led to various policies or studies that observe algal blooms through real-time monitoring. The real-time monitoring of chlorophyll-*a* concentration, an indicator of algal blooms, was presented as an effective method to predict algal blooms [15]; the monitoring data were applied in machine learning to develop short-term algal bloom prediction models.

In this study, we developed two short-term algal bloom prediction models for the Juksan weir, located downstream from the Yongsan River. Real-time water quality measurement data were applied in M5P and ELM for developing algal bloom prediction models. Furthermore the optimal data structure for input-output was determined. We compared the performance of M5P and ELM for developing short-term chlorophyll-*a* concentration prediction models (1-7 d). It is expected that the data-based chlorophyll-*a* concentration prediction model developed in this study would be useful in predicting the influential factors and size of algal blooms.

2. Materials and Methods

2.1. Study Area

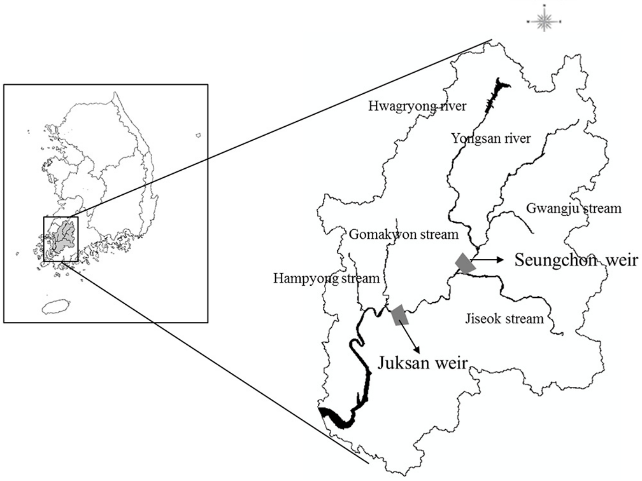

Fig. 1. Study area and main streams.

The Juksan weir in the Youngsan River, situated in South Korea, was selected as the study area with length and watershed area being 129.5 km and 3,455 km^2, respectively. The representative tributaries in the watershed include the Hwangryong River with a basin area of 564.3 km^2 and the Jiseok stream with a basin area of 657.2 km^2. Fig. 1 shows the Juksan weir locations of the Youngsan River in South Korea and the watershed area. The Youngsan River has two weirs (Seungchon weir and Juksan weir), which were built in sequence from 2012. Particularly, the Juksan weir of the lower Youngsan River has harmful algal bloom every summer. These algal blooms can cause water treatment problems for agricultural water supply, residential drinking water, and industrial water supply. Furthermore, the chlorophyll-*a* concentration in winter is higher than other rivers in Korea. Harmful algal bloom means toxic, hypoxia-generating cyanobacterial bloom genera, and it is controlled by the synergistic effects of nutrient (nitrogen and phosphorus) supplies, light, temperature, water residence, and biotic interactions [4].

Water quality data in this study were obtained from a real-time water quality monitoring station called Naju. This station, operated by Korean Ministry of Environment, was monitored on a daily basis from January 2013 to December 2016, and the database is managed by Real-Time Water Quality Information System. The chlorophyll-*a* concentration at the Naju real-time water quality station, which is located upstream 1 km from the Juksan weir was averaged 39.8 μg/L from 2013 to 2016, and the maximum concentration was 206.2 μg/L in August 2016. Table 1 shows the statistical values of water qualities such as chlorophyll-*a* concentration, WT, total nitrogen (T-N), and total phosphorus (T-P). Ecosystems can be classified into trophic categories using nutrients and algal biomass through various methods. The boundaries placed between these categories by aquatic scientists are similar but not universal. The US EPA suggested a eutrophic state based on values of annual average chlorophyll-*a* concentration exceeding 35 μg/L [16]. Accordingly, the Juksan weir can be considered as the eutrophic state, where the yearly average chlorophyll-*a* concentrations were 40.7, 36.0, 38.4, 44.1 μg/L in 2013, 2014, 2015, and 2016, respectively.

The T-P concentration averaged 0.101 mg/L, which exceeded the OECD eutrophication standards of T-P concentration of 0.035 mg/L [17]. This study collected climate data separately from the water quality data to analyze their correlation with chlorophyll-*a* concentration, an indicator of algal blooms; the correlation between each item and chlorophyll-*a* was found to be low. The T-N and N/P ratios had positive correlations, whereas T-P, WT, rainfall (RF), and SR had negative correlations with low degrees of correlations.

Table 1. Variables for Water Qualities in Juksan Weir (2013-2016)

Variables	Average	Max.	Min.	Std.
Chlorophyll-*a* (μg/L)	39.8	206.2	4.2	29.3
Water temperature (℃)	17.4	32.5	2.3	8.4
Total nitrogen (mg/L)	4.126	7.747	1.482	1.241
Total phosphorus (mg/L)	0.101	0.526	0.016	0.048

2.2. Algorithms

2.2.1. M5P model tree

MTs, although simple, are efficient and accurate tools for modeling the patterns and relationships for large datasets [18]. Quinlan et al. [19] developed a new type of tree called the M5 tree to predict continuous variables. An over-fitting problem can occur during MT construction based on training data. Predictably, the accuracy of the tree for training examples increases monotonically as the tree grows. However, this increases over-fitting; thus, the accuracy measured over the independent test examples first increases, then decreases. A method for reducing this problem is called "pruning." The final stage is to use a smoothing process to compensate for sharp discontinuities that inevitably occur between adjacent linear models at the leaves of the pruned tree, particularly for some models constructed from a small number of training instances. The smoothing procedure described by Quinlan et al. [19] uses the leaf model to compute the predicted value. The value is then filtered along the path back to the root, smoothing it at each node by combining it with the value predicted by the linear model for that node. In summary, the three major steps for M5 tree development are (1) tree construction; (2) tree pruning; and (3) tree smoothing. The M5 tree construction process attempts to maximize a measure called the standard deviation reduction (SDR). SDR is defined as

$$SDR = sd(T) - \sum_{i} \frac{|T_i|}{|T|} \times sd(T_i) \qquad (1)$$

where T is the set of cases, T_i is the *i*th subset of cases that result from the tree splitting based on a set of variables (attributes), $sd(T)$ is the standard deviation of T, and $sd(T_i)$ is the standard deviation of T_i as a measure of error [20].

Wang et al. [21] modified the original M5 tree algorithm to handle enumerated attributes and attribute missing values; they called the new tree algorithm "the M5P algorithm". In the M5P tree algorithm, all enumerated attributes are transformed into binary variables before tree construction. This algorithm can effectively deal with missing values and enumerated attributes. The M5P tree algorithm has three main steps, namely building the tree, pruning the tree, and smoothing. The basic tree is formed using the splitting criterion, which treats the standard deviation of the class values that reach a node as a measure of the error at that node, and calculates the expected reduction in error as a result of testing each attribute at that node. The attribute that maximizes the expected error reduction is then selected. The M5P MT has only recently been introduced in the water sector and has not yet been widely applied [8, 22].

2.2.2. Extreme learning machine

ELM is a new type of single-hidden-layer feed-forward network (SLFN) [23-25]. The Moore-Penrose generalized inverse and the minimum norm least-squares solution of a general linear system play important roles in the ELM learning algorithm. A general linear system Ax = y in Euclidean space, where A ∈ $R^{m \times n}$ and y ∈ R^m. Given a set of N samples (x_i, t_i), i = 1, 2, . . . , N, where x_i = [x_{i1}, x_{i2}, . . . , x_{in}]^T ∈ R^n and t_i = [t_{i1}, t_{i2}, . . . , t_{im}]^T ∈ R^m, standard SLFNs with $\tilde{N}$ hidden neurons and activation function *g*(*x*) are mathematically modeled as

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i \cdot x_j + b_i) = o_j, \quad j = 1, ..., N, \qquad (2)$$

where w_i = [w_{i1}, w_{i2}, · · · , w_{in}]^T is the weight vector connecting the *i*th hidden neuron and the input neurons, β_i = [β_{i1}, β_{i2}, · · · , β_{im}]^T is the weight vector connecting the *i*th hidden neuron and the output neurons, and b_i is the threshold of the *i*th hidden neuron. $w_i \cdot x_j$ denotes the inner product of w_i and x_j.

The standard SLFNs with hidden neurons with activation function *g*(*x*) can approximate these N samples with zero error, i.e., there exist β_i, w_i, and b_i such that

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i \cdot x_j + b_i) = t_j, \quad j = 1, ..., N, \qquad (3)$$

The above N equations can be written compactly as

$$H\beta = T \qquad (4)$$

where

$$H(w_1, \cdots, w_{\tilde{N}}, b_1, \cdots, b_{\tilde{N}}, x_1, \cdots, x_N) \qquad (5)$$
$$= \begin{bmatrix} g(w_1 \cdot w_1 + b_1) & \cdots & g(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \cdots & \vdots \\ g(w_1 \cdot w_N + b_1) & \cdots & g(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m} \text{ and } T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \qquad (6)$$

As named in Huang et al. [26-27], H is the hidden layer output matrix of the neural network; the *i*th column of H is the *i*th hidden neuron's output vector with respect to inputs x_1, x_2, · · · , x_N.

2.3. Model Development

Chlorophyll-*a* was used as the primary indicator for algal blooms. Other water qualities monitored were WT, pH, dissolved oxygen, electrical conductivity, turbidity, T-N, T-P, and total organic carbon. Weather data in this study were obtained from the Gwangju station monitored by the Korea Meteorological Administration. Table 2 shows the input variations, periods, and sources.

To analyze the inter-period correlation within the time-series

Table 2. Variables and Sources for the Chlorophyll-*a* Prediction Model Development

Items	Model inputs	Source
Weather	Rainfall (RF), Solar radiation (SR)	Korea Meteorological Administration (http://kma.go.kr)
Water qualities	Water temperature, Total nitrogen, Total phosphorus, N/P ratio, Chlorophyll-*a*	Korean Ministry of Environment, National Institute of Environmental Research (http://water.nier.go.kr)

data with chlorophyll-*a*, a serial correlation analysis (SCA) was conducted [11]. The analysis results indicated that the chlorophyll-*a* concentration of the target day had a correlation coefficient of 0.90 with 1-day-ahead chlorophyll-*a* concentration (CHL_1), 0.70 with 3-days-ahead chlorophyll-*a* concentration (CHL_3), 0.61 with 5-days-ahead chlorophyll-*a* concentration (CHL_5) and 0.52 with 7-days-ahead chlorophyll-*a* concentration (CHL_7). It was expected that the previous chlorophyll-*a* observation data would be variable in order to raise the predictability of the model, and using chlorophyll-*a* concentration 1-7 d before the model was utilized was expected to help construct the short-term algal bloom predictive model (Table S1).

In order to develop the algal bloom prediction model, independent parameters in the dataset include WT, RF, SR, T-N, T-P, N/P ratio, and chlorophyll-*a* as M5P MT and ELM model input. The nutrients N and P are the most important limiting factors influencing productivity. Moreover, the chlorophyll-*a* concentration increased with high P or N and low N/P ratio that supports some previous studies in a lotic environment [28-29]. Daily chlorophyll-*a* concentration was used as the model output that was the primary indicator of algal blooms. Parameters were selected considering input minimization and optimization [5-6, 30]. Both M5P MT and ELM were designed to predict the chlorophyll-*a* concentration after 1, 3, 5, and 7 d in terms of short-term algal bloom prediction. Table S2 shows the model input variables for short-term algal bloom prediction.

A total of 50% of the dataset were applied for model training and the remaining 50% were applied for model testing to develop algal bloom prediction model by each weir. The performance of the models for the Juksan weir was evaluated using the following indicators: square of correlation coefficient (R^2) that provides the variability measure for the data reproduced in the model; root-mean-square error (RMSE) that measures residual errors, providing a global idea of the difference between observation and modeling. The indicators were defined as shown below through Eq. (7) and (8).

$$R^2 = 1 - \sum \frac{(Y_i - \widehat{Y}_i)^2}{(Y_i - \overline{Y}_i)^2} \qquad (7)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{Y}_i - Y_i)^2} \qquad (8)$$

A total of 50% of the dataset were applied for model training and the remaining 50% were applied for model testing to develop algal bloom prediction model by each weir. In two equations, n is the number of data; Y_i are $\overline{Y}_i$ are observation data and the mean of observation data, respectively; and $\widehat{Y}_i$ denotes the modelling results.

3. Results

3.1. M5P Model Results

The algal bloom prediction model using the M5P MT was constructed based on the daily real-time water quality and weather data collected from 2013 to 2016. Of the collected data set orders, odd-numbered orders were utilized for training, and even-numbered orders were used for testing. The M5P MT was applied to develop the algal bloom prediction model to predict chlorophyll-*a* concentration after 1, 3, 5, and 7 d. In the present paper, the M5P algorithm implemented in Weka software (version 3.9.1 1999-2016) was used. This algorithm results were then evaluated using R^2 and RMSE. The performance of the M5P models is shown in Fig. 2. The performance of the model to predict chlorophyll-*a* concentration after 1 d has R^2 values of 0.79 for training and 0.83 for testing data sets and RMSE values of 14.0 μg/L for training and 12.2 μg/L for testing data sets. The performance of the model to predict chlorophyll-*a* concentration after 3 d has R^2 values of 0.55 for training and 0.46 for testing data sets and RMSE values of 20.0 μg/L for training and 22.1 μg/L for testing data sets, indicating a lower predictive power compared with that of the prediction model after 1 d. The performance of the model to predict chlorophyll-*a* concentration after 5 d has R^2 values of 0.49 for training and 0.44 for testing data sets, and RMSE values of 21.3 μg/L for training and 22.6 μg/L for testing data sets. The performance of the model to predict chlorophyll-*a* concentration after 7 d has R^2 values of 0.40 for training and 0.39 for testing data sets and RMSE values of 23.5 μg/L for training and 23.5 μg/L for testing data sets.

We compared the equation's independent variables developed by M5P for each model to predict chlorophyll-*a* after 1, 3, 5, 7 d. For the prediction model after 1 d, only chlorophyll-*a* was selected as the independent variable. Furthermore, the WT, SR, RF, and chlorophyll-*a* were selected as independent variables to predict chlorophyll-*a* after 3 d. The WT, SR, chlorophyll-*a* were selected as independent variables to predict chlorophyll-*a* after 5 d. Furthermore, the WT, SR, T-N, N/P ratio, and chlorophyll-*a* were selected as independent variables to predict chlorophyll-*a* after 7 d. The short-term algal bloom prediction model using M5P indicated that the performance of the 1 d prediction model was the highest. As the predictive periods increased to 3, 5 and 7 d, the number of independent variables included in the predictive equation increased. Moreover, prediction of 3 and 5 d periods included weather data, and the 7 d period included nutrient data (Table S3). The M5P method is based on classification and regression analysis. Comparing the measured and predicted values of the concentration of chlorophyll-*a*, the independent variables can help in calculating the concentration of chlorophyll-*a* within a certain range.

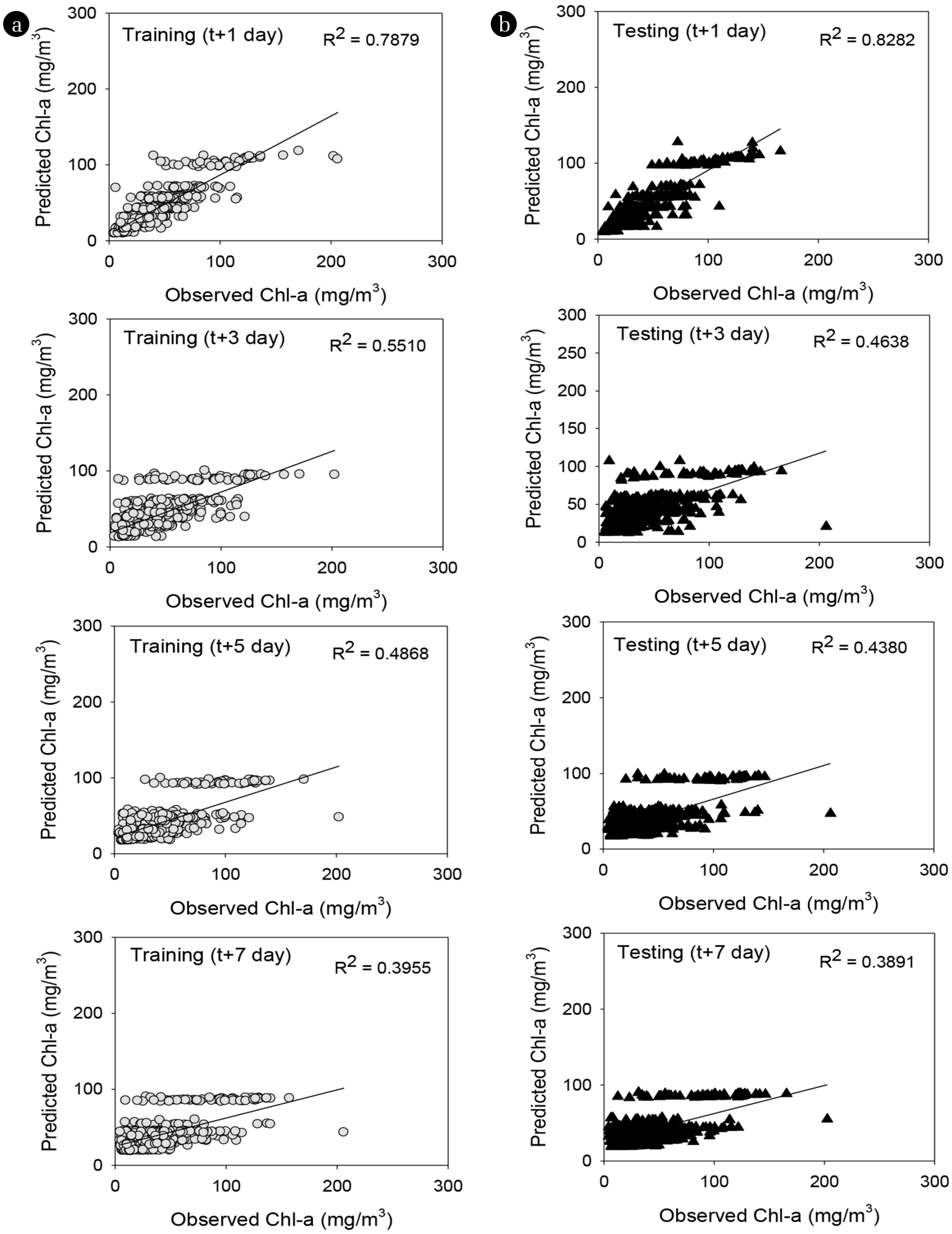

Fig. 2. Results of algal bloom prediction models using M5P; (a) Training results; (b) Testing results.

3.2. ELM Model Results

In this study, we also developed a short-term algal bloom prediction model for Juksan weir using ELM with the same water quality and weather data from 2013 to 2016 used in M5P. In the present study, the ELM algorithm was implemented in MATLAB software. Similar to M5P, among the collected data set orders, the odd number orders were utilized for training, while the even numbers were used for testing and constructing the models to predict chlorophyll-*a* concentration after 1, 3, 5, 7 d. ELM models were constructed in order to determine the optimum number of nodes in the hidden layer. The number of hidden nodes is determined when the performance of the test set for model validation reaches a minimum while the number of hidden node increases from 2 to 20. The training and testing performance of the ELM are shown in Fig. S1.

The performance power of the model to predict chlorophyll-*a* after 1 d has R^2 of 0.82 for training and 0.87 for testing data sets, RMSE of 13.0 μg/L for training and 10.7 μg/L for testing data sets. The performance power of the model to predict chlorophyll-*a* after 3 d has R^2 of 0.62 for training and 0.59 for testing data sets, RMSE

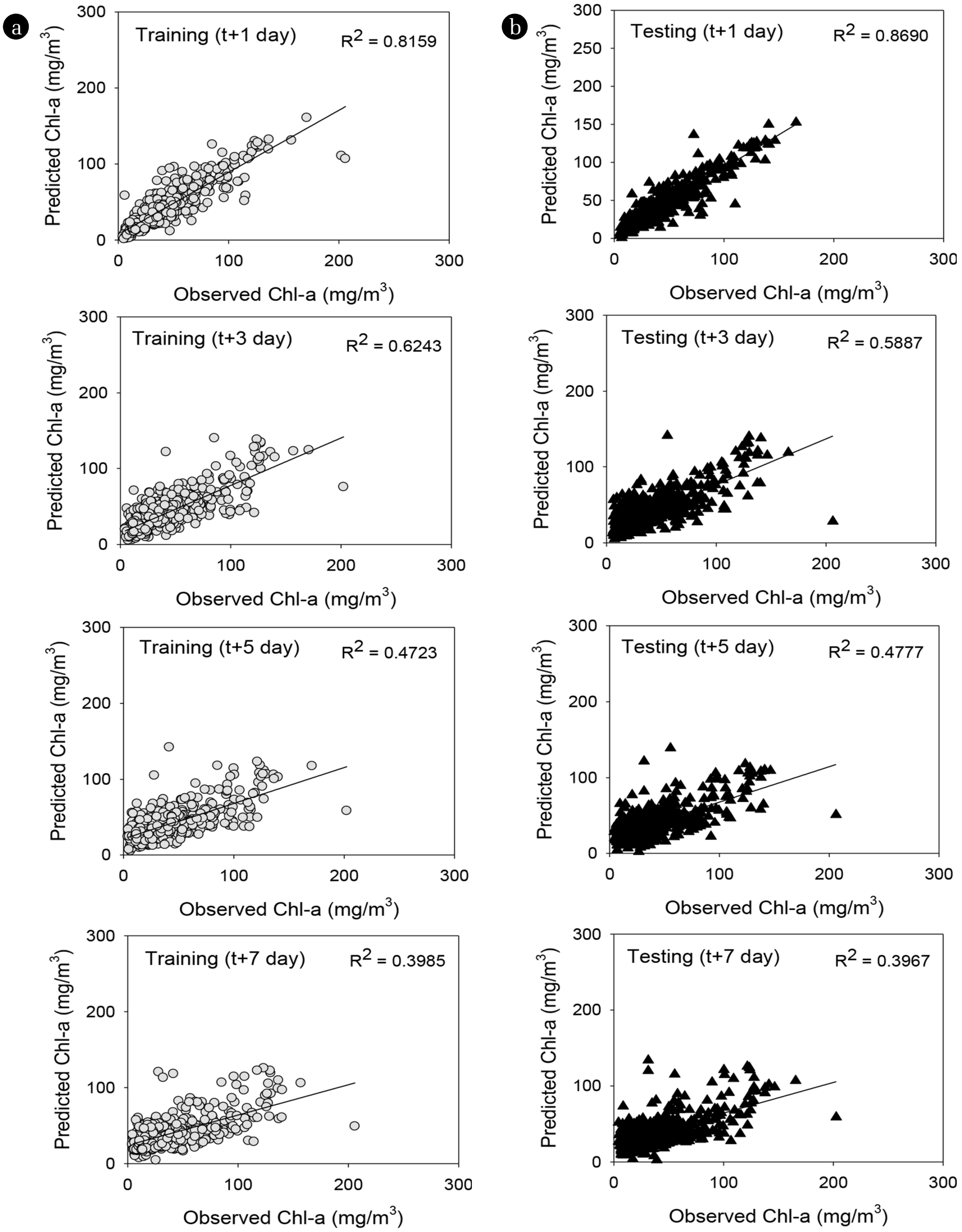

Fig. 3. Results of algal bloom prediction models using ELM; (a) Training results; (b) Testing results.

of 18.3 μg/L for training and 19.4 μg/L for testing data sets, indicating lower performance power compared to model to predict chlorophyll-*a* after 1 d. The performance power of the model to predict chlorophyll-*a* after 5 d has R^2 of 0.47 for training and 0.48 for testing data sets, RMSE of 21.6 μg/L for training and 21.8 μg/L for testing data sets. And performance power of the model to predict chlorophyll-*a* after 7 d has R^2 of 0.40 for training and 0.40 for testing data sets, RMSE of 23.4 μg/L for training and 23.3 μg/L for testing data sets. The performance of the ELM model is shown in Table 3 and Fig. 3.

The performance power of the ELM model was higher than that of M5P model. In addition, during the period when the chlorophyll-*a* concentration increased rapidly, the ELM model had higher accuracy than did the M5P model; this difference was more pronounced for longer prediction periods (Figs. 4, S2. S3, 5). As the M5P model presents independent variable that influence variations in chlorophyll-*a*, it is useful for determining influencing factors for algal blooms. However, the accuracy of this model decreases with an increase in the prediction periods for algal bloom prediction models. As such, the ELM model appears to be more appropriate for predicting the chlorophyll-*a* concentration within 7 d.

Table 3. Statistical Analysis Results of M5P and ELM Models

Prediction model			After 1 d	After 3 d	After 5 d	After 7 d
M5P	R^2	Training	0.79	0.55	0.49	0.40
		Testing	0.83	0.46	0.44	0.39
	RMSE	Training	14.0	20.0	21.3	23.5
		Testing	12.2	22.1	22.6	23.5
ELM	R^2	Training	0.82	0.62	0.47	0.40
		Testing	0.87	0.59	0.48	0.40
	RMSE	Training	10.3	18.3	21.6	23.4
		Testing	10.7	19.4	21.8	23.3

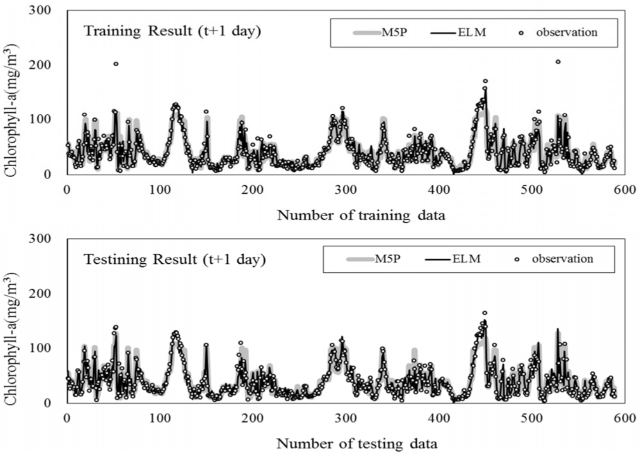

Fig. 4. Comparison of the observed and simulated chlorophyll-*a* of prediction models after 1 d.

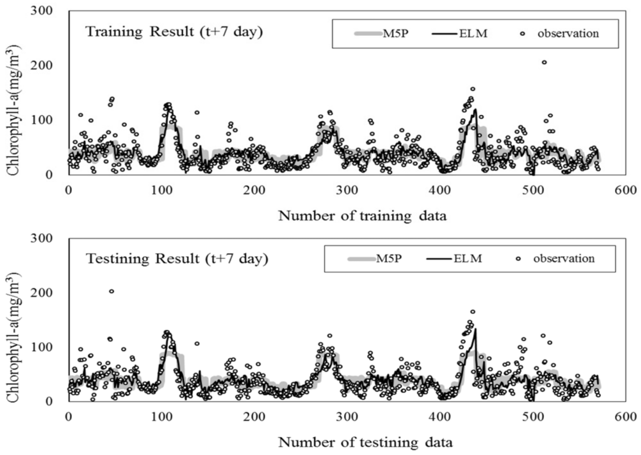

Fig. 5. Comparison of the observed and simulated chlorophyll-*a* of prediction models after 7 d.

4. Conclusions

We applied the M5P MT and ELM to develop a data-based model for short-term algal bloom prediction at Juksan weir, and developed 1, 3, 5, 7 d short-term prediction models. Chlorophyll-*a* concentration, which was the primary indicator of the algal bloom prediction model, was used to develop and compare the algal bloom models. The input variables included T-N, T-P, N/P ratio, WT, chlorophyll-*a*, RF, and SR. The results of the autocorrelation analysis for chlorophyll-*a* indicated that previous measurements of chlorophyll-*a* would serve as a good variable to increase the performance power of the prediction model. 50% of the dataset was applied for model training, while the remaining 50% was applied for model testing to develop the algal bloom prediction model. M5P model showed that the prediction model after 1 d had the highest performance power and dropped off rapidly starting with prediction after 3 d. Comparing the variables used in M5P model equations depending, for the prediction after 1 d chlorophyll-*a* concentration, value yielded was the chlorophyll-*a* concentration as measured 1 d ago; for the prediction model after 3 and 5 d, they included weather data, and the prediction model after 7 d also added nutrients. The present study has analyzed the performance power of ELM model; the prediction model after 1 d had the highest performance power. Comparing the performance power of the ELM model with the M5P model, it was found that the predictive power of the 1-7 d chlorophyll-*a* concentration prediction model was higher. Moreover, in a period of rapid algal blooms increases, the ELM model had higher accuracy than MT; this difference was more pronounced with longer prediction periods. As the M5P model presents the independent variable that influences changes in the chlorophyll-*a* concentration, it is useful for determining affecting factors for algal blooms. However, its accuracy drops with longer prediction periods; as such, the ELM model appears to be more appropriate for chlorophyll-*a* concentration prediction within 7 d.

The present study has utilized and compared M5P and ELM models to construct data-based chlorophyll-*a* concentration prediction model, providing foundations for proactive algae management through accurate predictions of occurrence periods and sizes. These results showed ELM can handle more the nonlinearity of algal bloom than M5P. Furthermore, these results lead us to the conclusion that ELM is effective for short-term algal bloom prediction. In future research, we will develop algal blooms prediction model using recurrent neural network and deep neural network.

References

1. Anderson DM, Cembella AD, Hallegraeff GM. Progress in understanding harmful algal blooms: Paradigm shifts and new tech-

nologies for research, monitoring, and management. *Annu. Rev. Mar. Sci.* 2012;4:143-176.

2. Glasgow HB, Burkholder JM, Reed RE, Lewitus AJ, Kleinman JE. Real-time remote monitoring of water quality: A review of current applications, and advancements in sensor, telemetry, and computing technologies. *J. Exp. Mar. Biol. Ecol.* 2004;300: 409-448.
3. Conley DJ, Paerl HW, Howarth RW, et al. Controlling eutrophication: Nitrogen and phosphorus. *Science* 2009;323:1014-1015.
4. Paerl HW. Controlling cyanobacterial harmful blooms in freshwater ecosystems, microbial biotechnology. *Microb. Biotechnol.* 2017;10:1106-1110.
5. Zhang X, Recknagel F, Chen Q, Cao H, Li R. Spatially-explicit modelling and forecasting of cyanobacteria growth in Lake Taihu by evolutionary computation. *Ecol. Modell.* 2014;306:216-225.
6. Xie Z, Lou I, Ung WK, Mok KM. Freshwater algal bloom prediction by support vector machine in Macau Storage Reservoirs. *Math. Probl. Eng.* 2012;27:1-12.
7. Park Y, Cho KH, Park J, Cha SM, Kim JH. Development of early-warning protocol for predicting *chlorophyll-a* concentration using machine learning models in freshwater and estuarine reservoirs, Korea. *Sci. Total Environ.* 2015;502:31-41.
8. Jung NC, Popescu I, Kelderman P, Solomatine DP, Price RK. Application of model trees and other machine learning techniques for algal growth prediction in Yongdam reservoir, Republic of Korea. *J. Hydroinform.* 2010;12:262-274.
9. Ye L, Cai Q, Zhang M, Tan L. Real-time observation, early warning and forecasting phytoplankton blooms by integrating *in situ* automated online sondes and hybrid evolutionary algorithm. *Ecol. Inform.* 2014;22:44-51.
10. Lee JHW, Huang Y, Dickman M, Jayawardena AW. Neural network modelling of coastal algal blooms. *Ecol. Modell.* 2003;159: 179-201.
11. Kim ME, Shin HS. Study on establishing algal bloom forecasting models using the artificial neural network. *J. Korea Water Resour. Assoc.* 2013;46:697-706.
12. Xu Y, Dai Y, Dong ZY, Zhang R, Meng K. Extreme learning machine-based predictor for real-time frequency stability assessment of electric power systems. *Neural Comput. Applic.* 2013;22:501-508.
13. Yadav B, Ch S, Mathur S, Adamowski J. Discharge forecasting using an online sequential extreme learning machine (OS-ELM) model: A case study in Neckar River, Germany. *Measurement* 2016;92:433-445.
14. Lou I, Xie Z, Ung WK, Mok KM. Freshwater algal bloom prediction by extreme learning machine in Macau Storage Reservoirs. *Neural Comput. Applic.* 2016;27:19-26.
15. Boyer JN, Kelble CR, Ortner PB, Rudnick DT. Phytoplankton bloom status: Chlorophyll *a* biomass as an indicator of water quality condition in the southern estuaries of Florida, USA. *Ecol. Indic.* 2009;9:S56-S67.
16. U.S. E.P.A. Water quality criteria research of the U.S. Environmental Protection Agency. Proceedings of an EPA Sponsored Symposium; 1976.
17. OECD. OECD Eutrophication programme-regional project alpine lakes. Swiss Federal Board for Environmental Protection OECD; 1980.
18. Nikoo MR, Karimi A, Kerachian R, Poorsepahy-Samian H, Daneshmand F. Rules for optimal operation of reservoir-river-groundwater systems considering water quality targets: Application of M5P model. *Water Resour. Manage.* 2013;27:2771-2784.
19. Quinlan JR. Learning with continuous classes. In: Proceedings AI'92, Adams & Sterling, eds. World Scientific. 1992. p. 343-348.
20. Zhan C, Gan A, Hadi M. Prediction of lane clearance time of freeway incidents using the M5P tree algorithm. *IEEE Trans. Intell. Transp. Syst.* 2011;12:1549-1557.
21. Wang Y, Witten IH. Inducing model trees for continuous classes. In: Proceedings of the Poster Papers of the 9th European Conference on Machine Learning (ECML 97). van Someren M, Widmer G, eds. 1997. p. 128-137.
22. Almasi SN, Bagherpour R, Mikaeil R, Ozcelik Y, Kalhori H. Predicting the building stone cutting rate based on rock properties and device pullback amperage in quarries using M5P model tree. *Geotech. Geol. Eng.* 2017;35:1311-1326.
23. Huang GB, Zhu QY, Siew CK. Extreme learning machine: A new learning scheme of feedforward neural networks. In: Proceedings of the IEEE International Joint Conference on Neural Networks; 25-29 July 2004; Budapest, Hungary. 2004. p. 985-990.
24. Zhou J, Peng T, Zhang C, Sun N. Data pre-analysis and ensemble of various artificial neural networks for monthly streamflow forecasting. *Water* 2018;10:628.
25. Huang GB, Zhu QY, Siew CK. Extreme learning machine: Theory and applications. *Neurocomputing* 2006;70:489-501.
26. Huang GB, Babri HA. Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions. *IEEE Trans. Neural Netw.* 1998;9: 224-229.
27. Huang GB. Learning capability and storage capacity of two-hidden-layer feedforward networks. *IEEE Trans. Neural Netw.* 2003;14:274-281.
28. Van Nieuwenhuyse EE, Jones RJ. Phosphorus-chlorophyll relationship in temperature strems and its variation with stream catchment area. *Can. J. Fish. Aquat. Sci.* 1996;53:53-99.
29. Mamun M, Lee SJ, An KG. Temperature and spatial variation of nutrients, suspended solids, and chlorophyll in Yeongsan watershed. *J. Asia Pac. Biodivers.* 2018;11:206-216.
30. Reckagel F, French M, Harkonen P, Yabunaka KI. Artificial neural network approach for modelling and prediction of algal blooms. *Ecol. Modell.* 1997;96:11-28.