Ecological Modelling 398 (2019) 67–76

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Determination of sensitive variables regardless of hydrological alteration in artificial neural network model of chlorophyll *a*: Case study of Nakdong River

Hyo Gyeom Kim[a], Sungwon Hong[a], Kwang-Seuk Jeong[b], Dong-Kyun Kim[c], Gea-Jae Joo[a,*]

[a] Department of Biological Sciences, Pusan National University, Republic of Korea
[b] Department of Nursing Science, School of Public Health, Dongju College, Republic of Korea
[c] Fisheries Science Institute, Chonnam National University, Republic of Korea

ARTICLE INFO

Keywords:
Artificial neural network
Chlorophyll *a*
Sensitivity analysis
River modification
Algal blooms
Regulated river

ABSTRACT

The Nakdong River has suffered from hydrological alterations in the river channel and riverine area during the Four Major Rivers Restoration Project (FMRRP). As these anthropogenic modifications have induced intensive algal blooms, the prediction of algal abundances has become an important issue for securing a source of drinking water and ecosystem stability. This study aimed to assess the changed river system in terms of chlorophyll *a* concentrations using artificial neural network (ANN) models trained for the pre-FMRRP period and tested for the post-FMRRP period in the middle reaches of such a river-reservoir system, and identify the descriptors that consistently affect algal dynamics. A total of 19 variables representing biweekly water-quality and meteo-hydrological data over 10 years were used to develop models based on different ANN algorithms. To identify the major descriptor to the algal dynamics, sensitivity analyses were performed. The best and most feasible model incorporating five parameters (wind velocity, conductivity, alkalinity, total nitrogen, and dam discharge) based on the topology of a probabilistic neural network with a smoothing parameter of 0.028 showed satisfactory results ($R = 0.752$, $p < 0.01$). Some mismatches were found in the post-FMRRP period, which may be due to a discrete event with a newly adapted over-wintering species and different causes of the summer growth of cyanobacteria owing to the river alteration. Based on the lowest sensitivity of dam discharge and the combination results of environmental management with total nitrogen, ANN modelling indicated that short-term water quality variables are persistent factors shaping algal dynamics.

1. Introduction

River damming has modified the characteristics of water body from those of a natural riverine ecosystem to those of a man-made lentic ecosystem. Because these physical alterations have the potential to induce eutrophication or dramatic changes in the hydrology and ecology, it is difficult to obtain a precise and detailed prediction of their ecological impact (Friedl and Wüest, 2002). Physical processes of water transport and mixing not only determine the spatial and temporal distributions of nutrients and solids, but also regulate the ecological conditions for biological processes (Rueda et al., 2006). In the river-reservoir system, changes in the water types can alter the pattern of phytoplankton proliferation in a species-specific manner. For example, increased water residence time due to extensive river fragmentation and flow regulation allows some algal species, which have the ability to adapt to stagnant conditions, to dominate, leading to algal blooms (Tomás et al., 2010).

Algal blooms are a prevalent ecological issue resulting in the deterioration of the water quality and disturbance to the ecosystem (Hallegraeff, 2003; Conroy et al., 2005). Although numerous freshwater phytoplankton genera are capable of forming blooms, harmful algal blooms (i.e., cyanobacteria) are cumbersome because of their toxicity and adaptability to extreme environmental conditions (Paerl et al., 2001). Therefore, by monitoring and predicting the dynamics of algal communities, it is necessary to understand the impact of fundamental processes, including growth characteristics, toxicity, and the bloom-forming response to nutrients, light, and other environmental conditions, as well as the factors affecting the management of algal blooms (Glibert and Burkholder, 2006). Several studies on the prediction of the growth and abundance of algal communities have illustrated that light and nutrient resources are key factors for determining algal dynamics (Reynolds, 2006). While the relationships between abiotic and biotic

* Corresponding author at: Department of Biological Sciences, Pusan National University, Jangjeon-dong, Gumjeong-gu, Busan, 46241, Republic of Korea.
E-mail address: gjjoo@pusan.ac.kr (G.-J. Joo).

https://doi.org/10.1016/j.ecolmodel.2019.02.003
Received 31 May 2018; Received in revised form 9 February 2019; Accepted 10 February 2019

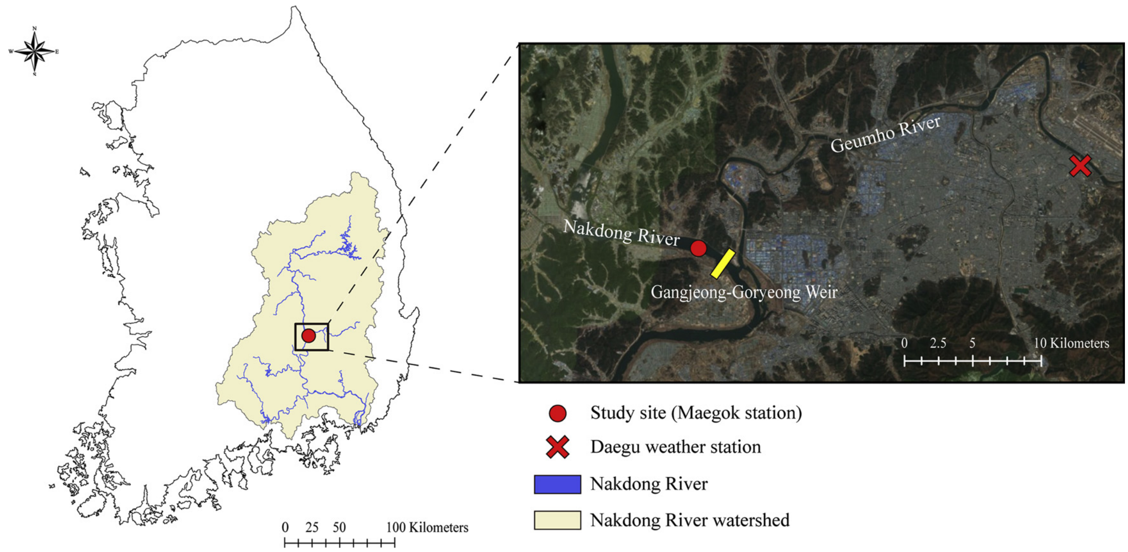

Fig. 1. Catchment of the Nakdong River and a description of the sampling site.

factors and algal biomass have been largely discovered, the complexity of the inter-relationships causes difficulties for identifying the factors responsible for the processes associated with algal structures (Oh et al., 2007).

One promising method for understanding complex water-quality interactions and evaluating the main contributors to algal dynamics is employing a prediction model. Because of the strength of their nonlinearity, ANNs have been used to model chl*a* using various types of algorithms, such as back-propagation, probabilistic neural networks (PNNs), and self-organizing maps (Jeong et al., 2006; Oh et al., 2007; Liu et al., 2015). Although ANNs have a severe limitation as a black-box model in terms of providing little explanatory insight into the contribution of the independent parameters in the prediction, they are undoubtedly a powerful tool for delineating nonlinearity. Furthermore, several methods (e.g., Garson's algorithm, sensitivity analysis, and a randomization approach) have been suggested for interpreting the relative influence of the descriptors, and this criticism has been challenged as the inner workings of neural networks have been illuminated (Olden and Jackson, 2002; Beck, 2015).

The Nakdong River, which is the second longest river in South Korea, is a eutrophic and flow-regulated river system. The river has been characterized by having two annual algal blooms—one in summer season by blue-green algae causing harmful toxins and odorous compounds in high-temperature water, and the other in winter by diatom (Ha et al., 1998). Various deterministic models have been developed for predicting the chl*a* concentration and biomass of *Microcystis aeruginosa*, and *Stephanodiscus hantzschii* in the lower Nakdong River, leading to the successful delineation of the driving factors and behavior of algal blooms (Jeong et al., 2006; Kim et al., 2007). These studies delineate the relationships between abiotic factors and chl*a* and the blooming species under the ecological conditions of the upstream weirs and estuarine barrages. However, further assessment of changes in the water type is required because eight weirs were additionally constructed during the implementation of the Four Major Rivers Restoration Project (FMRRP) from 2009 to 2011 (Shin and Chung, 2011).

Our hypothesis is that the data, pertaining to the abundance of algae in various conditions in the time periods before hydrologic alterations, could predict the algal dynamics after the alteration and especially short-term water quality variables are likely persistent factors shaping algal dynamics. Because algae respond rapidly and predictably to wide range of water quality conditions rather than hydrological alteration which occur over a relatively long period of time (Holling, 1973). Therefore, the specific aims of this study were (i) to assess the changed river system in terms of algal dynamics using ANN models trained for the pre-FMRRP period; and (ii) to identify the consistently important descriptors of algal blooms, excluding the impact of anthropogenic riverine alterations.

2. Materials and methods

2.1. Sampling site and data acquisition

The Nakdong River is one of the main river channels in South Korea. It is located in the southeastern part and is the second-longest river in the country, with a length of approximately 530 km. The study site—Maegok—lies near the middle of the Nakdong River and plays a key role in the supply of drinking water to the Daegu metropolitan city, which has a population of 2.5 million and contains six water-purification plants. Because of the variable rainfall and monsoon climate, to protect against flood and drought damage, the FMRRP was conducted from November 2009 to November 2011. Through this national project, the channels of four major rivers were dredged and deepened, and 3 multi-purpose dams and 16 weirs were built to increase the water-storage capacity. As a total of eight weirs were newly constructed in the Nakdong River, this river has been highly regulated by four multi-purpose dams at the origin, an estuarine barrage at the river mouth, and weirs at mean intervals of 29.6 ± 19.4 km. Our study site is situated above the Gangjeong–Goryeong weir, with four weirs located upstream and four weirs downstream (Fig. 1).

Water samplings were conducted biweekly at a floating dock approximately 20 m off the shore of Maegok. The site has a continuous mixing regime through the entire water column because of the relatively shallow depth of the site as well as the wind (unpublished data). Samples were collected within 0.5 m of the river surface over 10 years, from 2006 to 2015. The water-quality data include 13 parameters: water temperature (WT), dissolved oxygen (DO), conductivity (CO), pH, turbidity (Tu), transparency (Tr), alkalinity (AL), total nitrogen (TN), total phosphorus (TP), nitrate (NO), phosphate (PO), silicate (SI), and chl*a* concentration. The WT and DO were measured using a YSI model 58 dissolved oxygen probe, and the CO was measured using a YSI model 30 salinity meter (YSI, USA). The pH, Tu, and Tr were measured using an Orion model 407 A Ionanalyzer (Orion, USA), Micro100 Turbidity meter (HF Scientific Inc., USA), and a 20-cm-diameter Secchi disc. The AL was determined by titrating H_2SO_4 against a water sample. For the measurement of nutrients, unfiltered water for TN and TP and filtered water for NO, PO, and SI were analyzed via spectrophotometry.

The chl*a* concentrations were determined using a Shimadzu UV-1601 spectrophotometer (Shimadzu Corp., Japan), following a previously published extraction method (Wetzel and Likens, 1991).

Other environmental data were also obtained for the model development. The data pertaining to five meteorological parameters—sunshine hours, radiation, wind velocity, precipitation, and 7-day-totaled precipitation—were collected from the Daegu Local Meteorological Station, which is the station nearest to the study site. Because precipitation was intermittent, we added the parameter of 7-day-totaled precipitation. Daily hydrological data of dam and river discharge were supported by the National Water Resource Management Information System. Two hydrological parameters were included. Dam discharge was represented by the sum of discharge from two major upstream dams (Andong and Imha) and river discharge was represented by the discharge from the nearest upstream station, Waegwan. To prepare the dataset with water quality variables, we used the values of the seven parameters corresponding to our survey date.

To test whether the water quality, meteorological, and hydrological variables differed between the pre- and post-period FMRRP, *t*-test was performed at the 5% level of significance (Zar, 1999)

2.2. Theoretical background of ANN and model selection

The ANN is a model based on the information processing of synaptic neurons in the human brain. It resembles the brain in that knowledge is acquired through a learning process, and interneuron connection strengths (known as synaptic weights) are used to store that knowledge. The weights of the network are modified to attain a desired design objective by performing iterative learning processes (Haykin, 1994). The structure of the ANN consists of two or more layers composed of processing nodes, including input, hidden, and output layers that are connected by links with varying weights (Fig. 2).

To be used as training patterns for neural networks, the datasets were transformed for being treated differently by dependent and independent variables. As neural networks use the sigmoid function in their development, the raw values of dependent variable, chl*a*, were scaled into a [0, 1] interval. To standardize the measurement scales of the input descriptors, z-score normalization (i.e. mean = 0, standard deviation = 1) was applied (Olden and Jackson, 2002). In this study, 19 explanatory descriptors and the dependent variable of chl*a* were converted using the above methods (Table 1).

NeuroSolutions Infinity was used to develop the ANN model. This program provides an efficient approach for selecting a set of input vectors, topology, and configuration that is otherwise costly and time consuming. This program automatically configures the architectures of the ANNs, adds or removes input descriptors, optimizes the number of neurons or PNN smoothing parameters, or varies hidden layers to determine the optimum forecast model for each dataset. To promote the 'best' model, "the experiment score" was evaluated using a pre-set formula incorporating the root-mean-square error (RMSE), mean absolute error (MAE), and correlation coefficient (R). To select potential input variables, they were automatically scored in a combination of ways: (i) according to the correlation of the input to the desired output and (ii) according to the performance in a neural network model. For the objective of prediction, the desired output could be set manually on rows in advance (NeuroDimension, Inc; Gainesville, FL, USA). In addition, collinearity diagnostics were run to test for possible multicollinearity among the explanatory variables using the Durbin-Watson test. As some of descriptors may have been correlated, we used the variance inflation factor (VIF) to estimate the model stability. Values ≥ 10 represent highly correlated descriptors (Ter Braak, 1986).

To identify descriptors common to both pre- and post-period FMRRP, the datasets for training were assigned to the pre-FMRRP period and those for testing were assigned to the post-FMRRP period. Of a total of 257 exemplars, 105 (40%) from 2006 to 2009 were repeatedly trained to determine the appropriate weights and develop a reasonable model for predicting chl*a*. To provide an unbiased estimation of the predictive success of the network concurrent with optimal training and ensure a feasible model design, the RMSE was also computed for "cross-validation" (accounting for 30% of the exemplars from 2010 to 2012). The network training was set to terminate immediately if the RMSE within the training or cross-validation datasets fell below 0.01 or began to increase, which indicates that the network is "over-trained." To assess the predictive performance of the models, we used 20-fold cross-validation. In this procedure, data were segmented into 20 subsamples to apply the leave-one-out algorithm; A single subsample was retained as the validation data to test the model, and the remaining 19 subsamples were used as training data. The testing process involved applying a trained network with frozen weights to a data subset (remaining 30% of exemplars from 2013 to 2015).

2.3. Sensitivity analyses

Sensitivity analyses were performed to examine how chl*a* responded to each descriptor. To shed light on the inner workings of neural networks, a number of methods that can evaluate the influence of the independent variable have been proposed in the ecological literature (Olden and Jackson, 2002; Gevrey et al., 2003). From the results of the

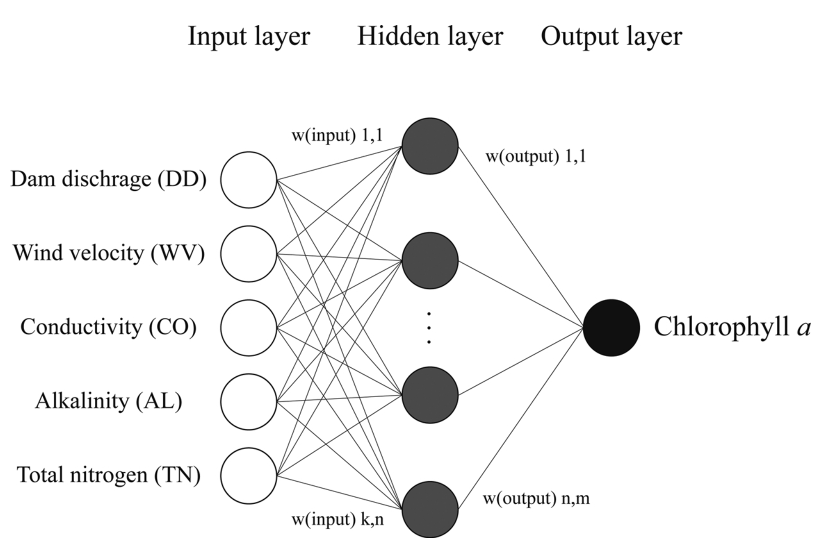

Fig. 2. Basic structure of an ANN model.

Table 1
Descriptive limnological variables at Maegok station from 2006 to 2015 (n = 259, bold type; the selected variables for developing the predictive model).

Variables	Acronym	Input Score	Training (n = 105)		Cross-validation (n = 77)		Evaluation (n = 77)	
			Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
River discharge (CMS)	RD	47.7	162.5	325.7	182.5	232.6	91.6	57.8
Dam discharge (CMS)	DD	49.9	49.7	78.2	54.2	45.2	33.9	17.7
Sunshine hours (h)	SU	49.7	5.7	3.9	5.6	4.1	6.7	3.3
Radiation (MJ m^{-2})	RD	54.5	13.4	6.6	13.5	6.9	15.2	5.9
Wind velocity (m/s)	WV	69.6	2.3	1.0	2.1	0.9	2.1	0.9
Precipitation (mm d^{-1})	Ra	60.0	2.1	6.5	3.7	7.9	1.0	3.6
Accumulated Rainfall for 7 days (mm d^{-1})	ARa	58.2	22.0	40.9	23.3	35.0	14.5	19.6
Water temperature (°C)	WT	47.9	16.5	8.8	16.1	9.2	16.9	9.4
Dissolved oxygen (mg L^{-1})	DO	46.2	11.7	3.0	10.9	2.9	12.2	2.8
Dissolved oxygen (%)	D%	48.5	115.0	20.5	107.9	19.0	122.3	20.9
pH	pH	52.4	8.2	0.7	7.9	0.6	8.4	0.6
Turbidity (NTU)	Tu	52.4	14.7	12.3	38.2	50.3	25.1	58.0
Conductivity (μS cm^{-1})	CO	69.1	261.0	66.2	243.9	63.6	280.6	63.3
Alkalinity (mg L^{-1})	Al	67.0	63.5	14.4	64.6	19.7	63.8	10.8
Total nitrogen (mg L^{-1})	TN	59.8	3.5	1.5	3.4	1.7	3.5	1.3
Total phosphorus (μg L^{-1})	TP	59.0	155.5	99.0	132.4	127.3	72.4	88.8
Nitrate-N (mg L^{-1})	NO	52.0	2.5	1.2	2.3	1.0	2.1	0.7
Phosphate-P (μg L^{-1})	PO	52.4	83.6	85.1	43.0	39.9	20.4	23.6
Silica (mg L^{-1})	SI	44.6	6.7	3.5	6.6	4.1	1.9	1.4
Chlorophyll *a* (μg L^{-1})	Chl*a*	–	18.6	15.7	11.1	10.9	24.3	22.6

relative contribution and/or the contribution profile of the input descriptors, it is possible to determine the feasibility of neural networks for understanding the causal relationship between the input and output (Choi and Choi, 1992). When the weights of the neural network are fixed and one of the input dithers, the effect of the input can be revealed at the output. NeuroSolution Infinity software shows this corresponding change in output as the percentage effect. Additionally, in this study, to delineate the individual contributions of each descriptor and the response of chl*a*, the "Profile" method was additionally used. We constructed contribution plots with 101 scales from the minimum to the maximum of each input descriptor, while the others were blocked at their minimum, first quartile (Q1), median, third quartile (Q3), and maximum, with the output repeatedly obtained through the execution of ANN. The relative contributions of each descriptor were expressed by the range values (maximum–minimum) of their contributions.

2.4. Optimal combination of environmental management

Scenario analysis is traditionally considered the domain of process-based models, which execute process equations based on scenario-specific parameters and input settings. The simulated scenarios affect the output and other input descriptors through the resulting state trajectories. Because ANNs are the connection between descriptors and outputs with significantly non-linear components, they are not appropriate for the indication of mechanistic or causal relationships (Hornik, 1991; Cheng et al., 2018). Nevertheless, in this study, to provide information for the management of the proliferation of phytoplankton using the developed neural network, we observed the response of chl*a* to the orthogonal combinations of two descriptors: TN and dam discharge. For TN criteria, we referred to the suggestion of water quality standard in Kim et al. (2015): the 'very good' level is below 1.5 mg L^{-1}, the 'fair' level is below 4.0 mg L^{-1}, and the 'very poor' level is above 8.0 mg L^{-1}. DD regimes were created from historical data and by referenced the scenario of river regulation mechanism in Jeong et al. (2007), being divided into a rainy season (June–September) and non-rainy season (October–May). The differences in the chl*a* responses between the combinations were analyzed by Kruskal-Wallis test and a posteriori test. All statistical analyses were performed using the software IBM SPSS 23 (Zar, 1999; IBM Inc., Chicago, IL, USA).

3. Results

3.1. Limnological features of pre- and post-period FMRRP

There was a drastic change in the physical structure and the hydrology of the Nakdong River after the Four Major Rivers Restoration Project. The cross-sectional area of the Maegok Station increased from 1104.2 ± 66.0 m^2 to 2006.3 ± 44.4 m^2 (SE), and the flow velocity decreased from 0.6 ± 0.1 m s^{-1} to 0.2 to 0.0 m s^{-1} (SE, $p < 0.05$). In contrast, no meteorological parameter—radiation time, radiation amount, rainfall, and accumulated rainfall—showed significant difference between the pre- and post-periods ($p > 0.05$). Accumulated rainfall values were highest in July, with the maximum value of 277.6 mm, confirming the temperate monsoon climate of the river.

Contrary to hydrologic changes, water quality variables did not show significant differences between the pre- and post-periods of the project, except for turbidity, TP, and inorganic nutrients. In the pre-FMRRP period, the mean values of TP, NO, SI, and PO were lower than those during the post-FMRRP period ($p < 0.05$). Despite the decrease in TP concentration, the river was classified as eutrophic because of the mean value of TP (26.0 μg L^{-1}). While chl*a* did not show a significant difference between the pre- and post-period FMRRP, the mean values of chl*a* slightly increased after showing lower values during the project.

3.2. Development and selection of ANN model

The prediction results, which were based on the post-FMRRP period of January 2013 to December 2015, showed satisfying accuracy in timing, but a little underestimated in peak concentration (R = 0.752, $p < 0.01$), even though the model was trained to fit the pre-FMRRP period (Fig. 3). Among the total of 51,518 trials with 80 experiments, the network based on the PNN topology with a smoothing parameter of 0.028 exhibited the highest score. The selected set of inputs comprised five variables: wind velocity (WV), CO, AL, TN, and dam discharge (DD). There was no significant collinearity neither among descriptors (all values were below 10) nor between each descriptor and output variable.

There were models that used other descriptors, such as WT, RD, TP, and inorganic nutrients with high input scores, but they showed worse correspondence between the measured and forecasted chl*a* than the selected model. These models, including either WT or RD as high-scored input, needed 16 other descriptors to present their best performance. It

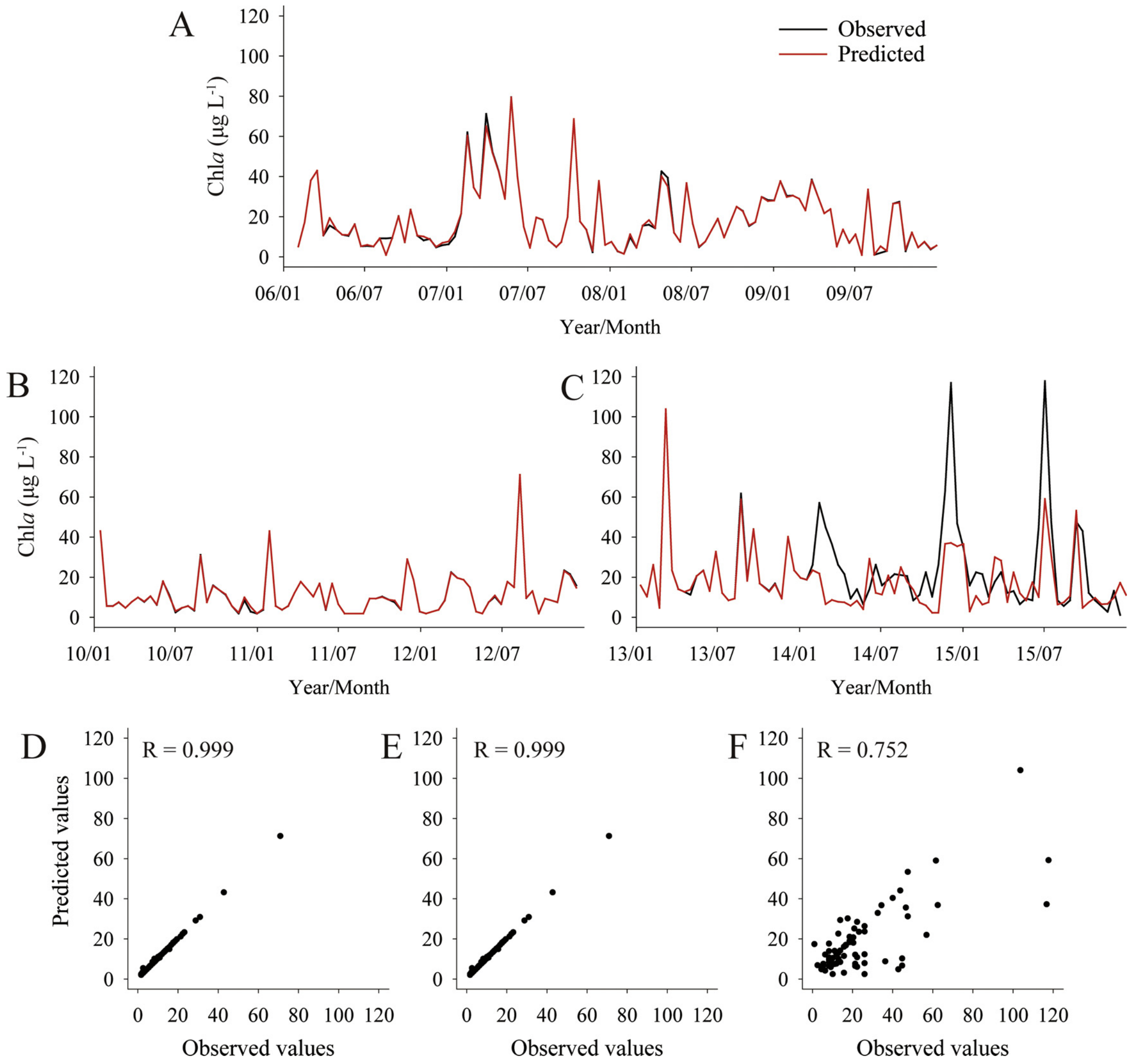

Fig. 3. Prediction of the chlrophyll *a* (chl*a*) using the ANNs developed. (A) Results for training, (B) cross-validation, and (C) testing. Relationships between observed and predicted values in the (D) training, (E) cross-validation, and (F) testing.

was possible to develop a model using TP rather than TN with a few other descriptors, but the predictability was lower than the one using TN. In terms of inorganic nutrients, it was also possible to construct an ANN model with a few descriptors; in this case too, the model performance was not as good as that of the selected final model.

3.3. Sensitivity analyses

To assess the relative importance and validity of descriptors in the developed models, sensitivity analyses were performed. According to the results of the sensitivity studies in Fig. 4A, the importance of the descriptors in recognizing and predicting chl*a* dynamics can be clearly seen. WV had the highest impact on the output (27.9%), followed by CO (27.5%), AL (18.6%), and TN (16.0%), and DD was lowest as 10.0%.

The profile of chl*a*, according to the increase of each descriptor provided two elements of information: (1) the order of contribution as the variation range and (2) the mode of action (Fig. 4B). Among the five descriptors, CO had the greatest effect on chl*a* over the large range, followed by AL, TN, and WV. DD had the lowest effect on chl*a* variation. The contribution plots of CO and DD showed higher values of chl*a* at descriptors' lower values, while that of WV had higher values of chl*a* at its higher values. AL contributed the greatest at intermediate values, and had a greater effect in the range of 56.8–121.2 mg L^{-1}. The influence of TN on chl*a* was observed differently by the values of other descriptors. When other descriptors were low at the minimum and 1st quartile levels, TN had its greatest impact at high values; on the contrary, when the other descriptors were at medium and 3rd quartile levels, TN had its greatest impact at intermediate values.

3.4. Response of chlorophyll a according to environmental changes

Using orthogonal design, a Latin table of TN at three levels and DD at two regimes was designed, and six simulated outputs for these new inputs are listed in Table 2. The average change in TN, which was sensitive to anthropogenic impacts and could be prevented in practice, was empirically changed at three levels, 1.5, 4.0, and 8.0 mg L^{-1}. In terms of DD, with the total amount fixed, two flow regimes were defined: 15% in rainy season with -15% in non-rainy season, and -15% in rainy season with 15% in non-rainy season.

At the TN level of "very good" and "fair" regardless of flow regimes, the chl*a* concentration decreased by 17.6 ± 0.8, 17.7 ± 0.7, 17.1 ± 1.0, and 17.2 ± 1.0 μg L^{-1} (mean ± SE). In addition, when DD increased in rainy season and decreased in non-rainy season, chl*a* showed lower values than in the reverse condition. However, there is no statistically significant difference with its original condition. Contrarily, an increase in TN up to the "very poor" level would induce a significant increase of chl*a*. For the "poor" level of TN, controlling through DD timing and magnitude did not result in a change in chl*a*

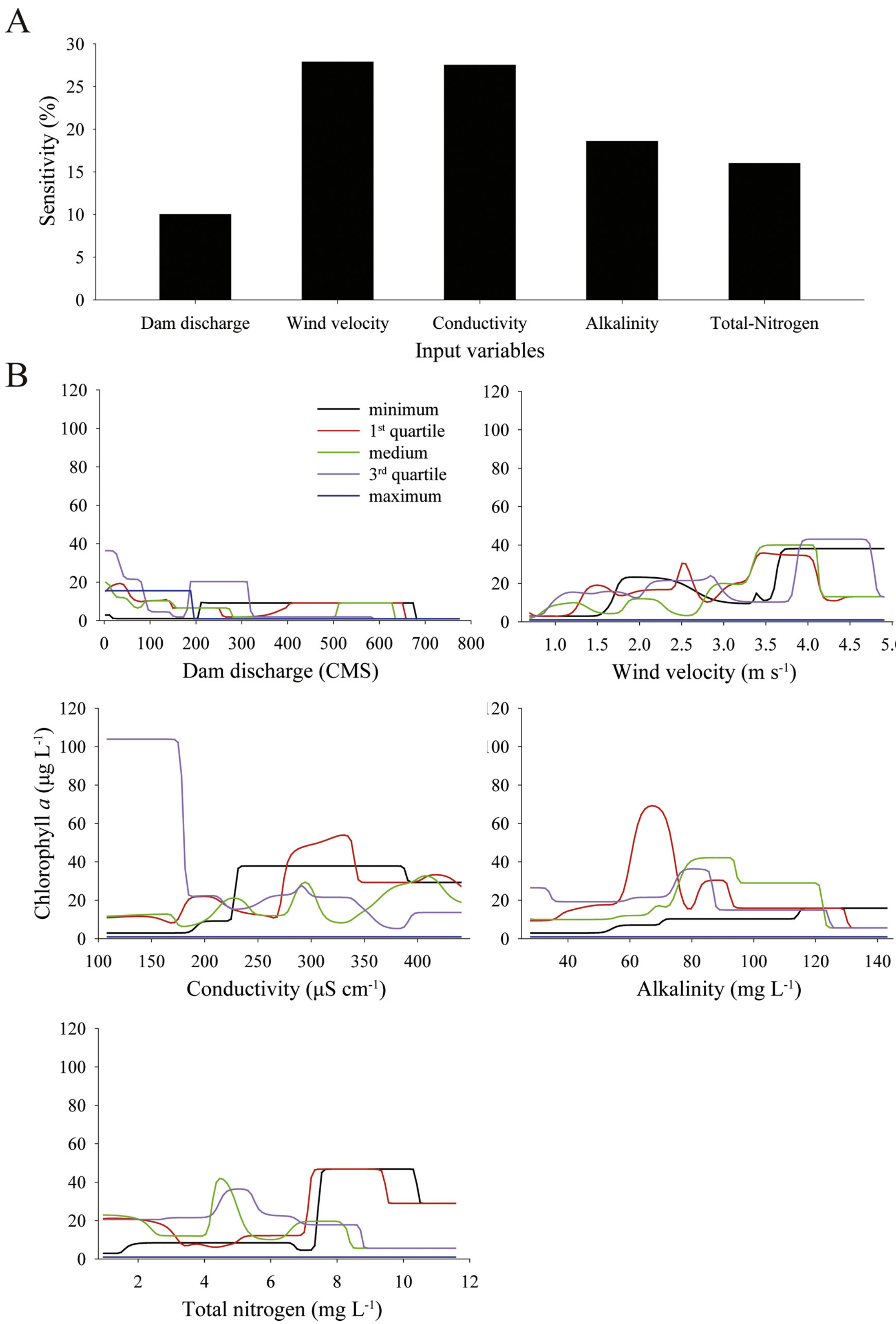

Fig. 4. Evaluation of the relative importance of environmental variables for predicting the chlorophyll *a* using (A) the dithering and (B) the "Profile" method.

Table 2
Changes of chlorophyll *a* (chl*a*) concentration with different combination of the environmental changes according to scenarios using factorial orthogonal design. Variable abbreviation: TN, total nitrogen; DD, dam discharge.

No.	Changes of environmental factors		Response of chl*a* ($\mu g\ L^{-1}$)
	TN	DD	
1	–	–	18.0 ± 17.0^{ab}
2	Very good	Increase in monsoon,decrease in non-monsoon	17.6 ± 12.4^{ab}
3	Very good	Decrease in monsoon,increase in non-monsoon	17.7 ± 11.9^{b}
4	Fair	Increase in monsoon,decrease in non-monsoon	17.1 ± 16.7^{a}
5	Fair	Decrease in monsoon,increase in non-monsoon	17.2 ± 16.5^{ab}
6	Very poor	Increase in monsoon,decrease in non-monsoon	20.5 ± 11.9^{c}
7	Very poor	Decrease in monsoon,increase in non-monsoon	20.5 ± 12.0^{c}

($20.5 \pm 0.7\ mg\ L^{-1}$ vs. $20.5 \pm 0.7\ mg\ L^{-1}$).

4. Discussion

4.1. Limnological features and the maintenance of chlorophyll a concentration

Our study aimed to understand the dominant forcing factors in the pre- and post-period FMRRP in the middle reach of the Nakdong River, which is used for water supply. The results highlighted the importance of persistent descriptors in explaining the phytoplankton biomass despite hydrological alterations. Descriptors, such as dam discharge, conductivity, alkalinity, and wind velocity, related to the physical condition of the water column had a higher explanatory power to predict algal dynamics in the river. In contrast, descriptors such as light sources and water temperature, which are essential in the process-based models for the estimation of algal dynamics, were not incorporated in our model, suggesting that a direct relation of the descriptors with chl*a* was not detected in this study. An ANN has been found to have tremendous potential as a forecasting tool and is faster even with minimal inputs (Palani et al., 2008). Thus, our modelling approach with ANN enables the selection of the most effective descriptors to predict algal dynamics in both pre- and post-period FMRRP, but cannot delineate the direct relationship between descriptors and chl*a*.

Among five nutrient-associated factors, the developed model only included TN as an explanatory descriptor. Phosphorus-family nutrients were not included in the best-performing model because they exhibited a lower score than selected the input descriptors, except DD of 49.9 (Table 1). According to the mean values of the TN: TP ratio (molar) of 123.1 for the Nakdong River, which were observed at low values (< 20:1) only eight times throughout the 10 years in the study period, phosphorus appears to be the primary limiting nutrient for phytoplankton growth (Downing and McCauley, 1992). Moreover, with newly implemented total maximum daily load and installation of water quality improvement facilities, the concentrations of TP and PO were significantly decreased (Shin and Chung et al., 2011). This suggested either the existence of nonlinear relationships between TN and P-limited adapted algae or the necessity for N reduction. Although P load reduction is the prime strategy to secure water quality, N load reduction is also essential for controlling the magnitude and duration of algal blooms (Xu et al., 2010). While the importance of P-associated nutrients could be increased in the post-period FMRRP, availability of N is a key factor for the proliferation and maintenance of cyanobacterial blooms if both pre- and post-periods of the project are considered.

4.2. Modelling of chla in post-FMRRP period

Although the ecosystem of the Nakdong River has suffered from drastic changes in hydrological factors mainly caused by the FMRRP over two years, the ANN model yielded accurate results for the test period of post-FMRRP. Considering that these anthropogenic changes to the river occurred abruptly and their consequences were unpredictable, the data analysis by ANN modelling was successful. Similarly, in the Yuqiao reservoir, significant changes in the ecosystem due to the natural evolution of the submerged vegetation over decades causes difficulties for the model long-term series chl*a* data obtained using ANNs, but the level of precision for the testing data was satisfactory (Liu et al., 2015). Compared with other studies involving the prediction of the chl*a* concentration in relatively consistent aquatic environments, such as the Juam reservoir of Korea (Park et al., 2015), Baiyangdian Lake of China (Wang et al., 2013), and the Kielstau catchment of Germany (Wu et al., 2014), our model showed a good correspondence. This indicated that the patterns of algal communities might not be immediately sensitive to slowly evolving changes in the physical characteristics of the river and be affected after a long period of time (Friedl and Wüest, 2002).

Interestingly, the model did predict the first great peak ($> 100\ \mu g\ L^{-1}$) in the test data, but did not predict the second smaller peak (about $50\ \mu g\ L^{-1}$), and the second and third great peak. This may be because the first peak from May 8, 2013 represents a similar type of event as the algal growth event in the pre-FMRRP period. In contrast, the discrepancy observed at the second smaller peak and the second great peak of late fall-winter season could be a different phenomenon caused by over-wintering cyanobacteria, *Aphanizomenon* spp.. Yu et al. (2014) and Ryu et al. (2016) reported that *Aphanizomenon flos-aquae* abruptly showed development in the fall-winter season and had a great tolerance for water temperature in the range of 5–15 ℃ in the Nakdong River. The other underestimated value at the third great peak observed on July 3, 2015 can be attributed to a decrease in the dam discharge effect due to the newly constructed upstream weirs. The significance of these is that the network has not trained on extensive data with this phenomenon included (i.e. a peak of over-wintering cyanobacteria) and it would not necessarily then be able to predict a discrete event with a newly adapted species and different causes for the summer growth of cyanobacteria.

In this study, the topology of the PNN for a smoothing factor of 0.028 exhibited the best performance among the nine topologies, i.e., five non-temporal networks and four temporal networks. The PNN is an ANN algorithm that is mainly employed to classify patterns in multidimensional datasets having non-linear relationships with the input, pattern, summation, and output layers. (Specht, 1990). The algorithm has become an effective tool for solving many problems because of its ease in training and sound statistical foundation (Strobl et al., 2007). However, in some studies on forecasting the chl*a* concentration using the PNN topology, the precision was lower than that for temporal ANNs (Liu et al., 2015). Two factors that could affect the efficiency of the classification system are the smoothing factor, which represents the width of the calculated Gaussian curve for each probability density function, and the training dataset. Yu and Chen (2007) examined the effect of the smoothing factors and noted that when the value of the factor increases, the specificity, sensitivities, and the overall classification accuracy decrease. Thus, this might be because of the differences in the data-allocation and training sequences or the lack of consideration of the optimal smoothing parameter for the combination of input vectors.

An n-day-ahead input vector can be used to give any developed model the capability to make short-term predictions (Recknagel et al., 1997). With the objective of controlling the algal blooms, various time-lag effects for the input parameter at the intervals of a day, a week, or a year on predicting the chl*a* concentration were calculated based on successful ANN models (Yabunaka et al., 1997; Lee et al., 2003). In our study, an ANN model with 28-d-ahead data as the input was selected. That is, matching the “28-day-ahead” environmental data to “today’s” dataset is reasonable for forecasting the level of chl*a* in the middle reach of the Nakdong River.

The applicability domain of the ANN model would be different between the pre-and post-period FMRRP. ANN modelling is a promising and useful tool that optimizes monitoring networks by identifying essential monitoring parameters (thereby permitting cost reduction). In the pre-FMRRP period, summer cyanobacterial blooms and winter diatom bloom, which would be interpreted as increased chl*a* concentrations, could be forecast without monitoring other related variables such as water temperature, radiation, and P-associated nutrient concentrations. Although there was a lack of fit between the observed and estimated data in the post-period FMRRP, it is possible to infer new patterns caused by river damming and the factors influencing those events. For example, the recruitment of *Aphanizomenon* akinetes from the sediments because of dredging of the riverbed allowed them to germinate in the late fall-winter season (Karlsson-Elfgren and Brunberg, 2004; Kim et al., 2005).

4.3. Sensitivity analyses

Through sensitivity analyses, we would know the effect that each of the descriptors have on the chl*a* concentration. This provides feedback regarding which input channels are the most significant, allowing the simplification of the interpretation of neural networks by reducing the number of axon pathways (Olden and Jackson, 2002). With the automated selection of input parameters, wind velocity (WV) and conductivity (CO) were observed as important descriptors to pattern chl*a* dynamics.

CO was selected as the most influential descriptors and may be selected to reflect the water mass characteristics. It has been an essential descriptor in profiling the chemical and physical characteristics through the entire water column with respect to the temperature and depth to represent water mass fluctuations (Berry et al., 2017). In a previous modeling study on the lower Nakdong River, to optimize the equations enabling the prediction of winter bloom diatoms, CO was selected as an input parameter with a high frequency (Kim et al., 2007). For the same site, Recknagel et al. (2017) predicted the response of two major blooming species, *Microcystis aeruginosa* and *Stephanodiscus hantzschii*, with the CO as a flow-dependent variable, by using inferential models through model ensembles. Although the middle reach in the Nakdong River were less affected by river regulation than the lower region in the pre-FMRRP period, CO seems to be an important descriptor in the prediction of chl*a* both before and after alteration.

Wind-induced resuspension and entrainment of algal species would result in the increase of chl*a*, which is consistent with the WV sensitivity profile of our network. In shallow lakes, the stabilization of the retention zone may cause meroplankton to dominate and increase in community size (Carrick et al., 1993). In terms of highly buoyant cyanobacteria, increase in wind velocity would decrease their biomass in surface water (Ha et al., 2000). However, large colonies of *Microcystis aeruginosa* from 36 to 120 μm seem to have an ability to overcome the entraining forces of turbulence and access the surface region; therefore, they are mainly concentrated in the surface layer (Wu and Kong, 2009). Moreover, intense mixing with wind-induced hydrodynamic disturbance can increase the chance of the aggregation of cyanobacteria (Wu et al., 2013). These positive relationships between wind velocity and algal biomass were well presented in our network, which attempted to predict chl*a* in a shallow, mixing, and frequently bloom-forming river.

According to the contribution plot, AL has a smaller effect on chl*a* at lower and higher values, and a large effect at middle values. Many studies regarding water quality using multivariate statistics in freshwater ecosystems have suggested that AL exhibits seasonality with CO (Interlandi and Crockett, 2003; Singh et al., 2004). In this study, the highest correlation between AL and CO ($R = 0.349$) with a small variation in the mean of AL from pre-FMRRP to post-FMRRP, which indicates that AL was an influential descriptor in chl*a* dynamics regardless of the hydrological alteration.

TN was only one nutrient descriptor in the network and the sensitivity differed in the function of the values other descriptors. Eutrophication with anthropogenic nutrient inputs to aquatic ecosystems is a major cause of the increase in algal blooms (Hallegraeff, 2003). However, in particular, phytoplankton nutrient deficiency would be differed depending on river flow; when there are the low and intermediate discharge, nitrogen is limiting to phytoplankton growth. In contrast, phytoplankton were nutrient-replete in high discharge condition, thus the response of chl*a* to TN showed high values at the intermediate values (Chaffin et al., 2014).

In both sensitivity results, DD had the lowest importance but clearly exhibited the greatest influence at its lower values. The increase in river flow with the dam operation has been widely accepted as a target to suppress algal blooms (Maier et al., 2001, 2004; Salmaso and Braioni, 2008). Similarly with the sensitivity of DD, chl*a* had a significantly negative relationship with flow and flood pulses compared with the isolation condition in river-floodplain ecosystems (Palijan, 2012). Although Li et al. (2013) asserted that a universal critical velocity for the management of algal communities does not exist, the parameter of DD could be an important role in determining the extent and timing of phytoplankton proliferation.

4.4. Management on algal blooms

According to the output of the combinations of environmental control, the increase in TN would be greatly associated with the increase in chl*a* from 18.0 to 20.5 $mg\,L^{-1}$, and the change in the dam discharge regime would have a small impact on the remediation of chl*a*. Controlling the nutrient loadings and securing the water quality of rivers may be effective in restoring their former conditions (Williamson et al., 1999). The decline of TN to the level of “very good” slightly lowered chl*a*, which highlights the necessity of the management on N to prevent further increases. Furthermore, in the regulated river system, dam should be operated to recover the natural flow regime as well as to secure the water requirement (Nilsson et al., 2005). On this account, Recknagel et al. (2017) asserted that adaptive management of seasonal water release from dams and weirs is necessary to remediate algal blooms, but in this study, the effect was not significant. This result would suggest that: (i) it is only the consequence from the lowest weight of dam discharge in the network, and (ii) alternatives in discharge management are required such as pulsed-flow discharge.

5. Conclusions

Our modelling approach with ANN indicates that the dataset collected during the pre-FMRRP period reflects the pattern of chl*a* response for the post-FMRRP period, which emphasizes that the dominant forcing factors were similar in those periods relative to hydrologic changes occurring on longer time scales. Although there was a lack of fit between the observed and estimated values in the post-period FMRRP, this provided an inference for new patterns caused by river damming and the factors influencing those events. Moreover, the sensitivity results clearly delineate the validity of the network and selected components driving the algal dynamics, excluding the impact of anthropogenic riverine alterations. Desirable combinations of TN and DD, which would effectively remediate algal blooms, are obtained. Furthermore, we observed that the constructed upstream weirs would attenuate the effectiveness of adaptive management on dam discharge. Therefore, for the management or remediation of increasing algal blooms due to river regulations, a consistent descriptor of the algal abundance that considers time periods before the hydrological alterations must be formulated.

Acknowledgments

This work was supported by the both grants from the “Basic Research (NRF-2016R1D1A1B01009492)” supported by an NRF (National Research Foundation of Korea).

References

Beck, M., 2015. NeuralNetTools: Visualization and Analysis Tools for Neural Networks. Version 1.3.2.
Berry, M.A., Davis, T.W., Cory, R.M., Duhaime, M.B., Johengen, T.H., Kling, G.W., Marino, J.A., Den Uyl, P.A., Gossiaux, D., Dick, G.J., 2017. Cyanobacterial harmful algal blooms are a biological disturbance to western Lake Erie bacterial communities. Environ. Microbiol. 19, 1149–1162.
Carrick, H.J., Aldridge, F.J., Schelske, C.L., 1993. Wind influences phytoplankton biomass and composition in a shallow, productive lake. Limnol. Oceanogr. 38, 1179–1192.
Chaffin, J.D., Bridgeman, T.B., Bade, D.L., Mobilian, C.N., 2014. Summer phytoplankton nutrient limitation in Maumee Bay of Lake Erie during high-flow and low-flow years. J. Great Lakes Res. 40, 524–531.
Cheng, X., Khomtchouk, B., Matloff, N., Mohanty, P., 2018. Polynomial regression as an alternative to neural nets. arXiv 1806.06850.
Choi, J.Y., Choi, C.-H., 1992. Sensitivity analysis of multilayer perceptron with differentiable activation functions. IEEE Trans. Neural Netw. 3, 101–107.
Conroy, J.D., Edwards, W.J., Pontius, R.A., Kane, D.D., Zhang, H., Shea, J.F., Richey, J.N., Culver, D.A., 2005. Soluble nitrogen and phosphorus excretion of exotic freshwater mussels (*Dreissena* spp.): potential impacts for nutrient remineralisation in western Lake Erie. Freshw. Biol. 50, 1146–1162.
Downing, J.A., McCauley, E., 1992. The nitrogen: phosphorus relationship in lakes. Limnol. Oceanogr. 37, 936–945.
Friedl, G., Wüest, A., 2002. Disrupting biogeochemical cycles-Consequences of damming. Aquat. Sci. 64, 55–65.
Gevrey, M., Dimopoulos, I., Lek, S., 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Modell. 160, 249–264.
Glibert, P.M., Burkholder, J.M., 2006. The complex relationships between increases in fertilization of the Earth, coastal eutrophication and proliferation of harmful algal blooms. In: Granéli, E., Turner, J. (Eds.), The Ecology of Harmful Algae. Springer Verlag, New York, pp. 341–354.
Ha, K., Kim, H.-W., Joo, G.-J., 1998. The phytoplankton succession in the lower part of hypertrophic Nakdong River (Mulgum), South Korea. Phytoplankton and Trophic Gradients. Springer Verlag, New York, pp. 217–227.
Ha, K., Kim, H.-W., Jeong, K.-S., Joo, G.-J., 2000. Vertical distribution of *Microcystis* population in the regulated Nakdong River, Korea. Limnology 1, 225–230.
Hallegraeff, G.M., 2003. Harmful algal blooms: a global overview. In: Hallegraeff, G.M. (Ed.), Manual on Harmful Marine Microalgae. IOC Manuals and Guides No. 33. UNESCO, pp. 1–22.
Haykin, S., 1994. Neural Networks, A Comprehensive Foundation. Macmillan, New York, NY, USA.
Holling, C.S., 1973. Resilience and stability of ecological systems. Annu. Rev. Ecol. Syst. 4, 1–23.
Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. Neural Netw. 4, 251–257.
Interlandi, S.J., Crockett, C.S., 2003. Recent water quality trends in the Schuylkill River, Pennsylvania, USA: a preliminary assessment of the relative influences of climate, river discharge and suburban development. Water Res. 37, 1737–1748.
Jeong, K.-S., Recknagel, F., Joo, G.-J., 2006. 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. Inform. 1, 235–245.
Jeong, K.-S., Kim, D.-K., Joo, G.-J., 2007. Delayed influence of dam storage and discharge on the determination of seasonal proliferations of *Microcystis aeruginosa* and *Stephanodiscus hantzschii* in a regulated river system of the lower Nakdong River (South Korea). Water Res. 41, 1269–1279.
Karlsson-Elfgren, I., Brunberg, A.-K., 2004. The Importance of shallow sediments in the recruitment of *Anabaena* and *Aphanizomenon* (Cyanophyceae)[1]. J. Phycol. 40, 831–836.
Kim, B.H., Lee, W.S., Kim, Y.-O., Lee, H.-O., Han, M.-S., 2005. Relationship between akinete germination and vegetative population of *Anabaena flos-aquae* (Nostocales, Cyanobacteria) in Seokchon reservoir (Seoul, Korea). Arch. Hydrobiol. 163, 49–64.
Kim, D.-K., Jeong, K.-S., Whigham, P.A., Joo, G.-J., 2007. Winter diatom blooms in a regulated river in South Korea: explanations based on evolutionary computation. Freshw. Biol. 52, 2021–2041.
Kim, H.K., Jeong, H., Bae, S.J., 2015. Deriving water quality criteria of total nitrogen for nutrient management in the stream. J. Korean Soc. Agric. Eng. 57, 121–127.
Lee, J.H., Huang, Y., Dickman, M., Jayawardena, A.W., 2003. Neural network modelling of coastal algal blooms. Ecol. Modell. 159, 179–201.
Li, F., Zhang, H., Zhu, Y., Xiao, Y., Chen, L., 2013. Effect of flow velocity on phytoplankton biomass and composition in a freshwater lake. Sci. Total Environ. 447, 64–71.
Liu, Y., Xi, D.-G., Li, Z.-L., 2015. Determination of the optimal training principle and input variables in artificial neural network model for the biweekly chlorophyll-*a*prediction: A case study of the Yuqiao Reservoir, China. PLoS One 10, e0119082.
Maier, H.R., Burch, M.D., Bormans, M., 2001. Flow management strategies to control blooms of the cyanobacterium, *Anabaena circinalis*, in the River Murray at Morgan, South Australia. Regul. Rivers: Res. Manage. 17, 637–650.
Maier, H.R., Kingston, G.B., Clark, T., Frazer, A., Sanderson, A., 2004. Risk-based approach for assessing the effectiveness of flow management in controlling cyanobacterial blooms in rivers. River Res. Appl. 20, 459–471.
Nilsson, C., Reidy, C.A., Dynesius, M., Revenga, C., 2005. Fragmentation and flow regulation of the world's large river systems. Science 308, 405–408.
Oh, H.-M., Ahn, C.-Y., Lee, J.-W., Chon, T.-S., Choi, K.H., Park, Y.-S., 2007. Community patterning and identification of predominant factors in algal bloom in Daechung Reservoir (Korea) using artificial neural networks. Ecol. Modell. 203, 109–118.
Olden, J.D., Jackson, D.A., 2002. Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. Ecol. Modell. 154, 135–150.
Paerl, H.W., Fulton, R.S., Moisander, P.H., Dyble, J., 2001. Harmful freshwater algal blooms, with an emphasis on cyanobacteria. Sci. World J. 1, 76–113.
Palani, S., Liong, S.-Y., Tkalich, P., 2008. An ANN application for water quality forecasting. Mar. Pollut. Bull. 56, 1586–1597.
Palijan, G., 2012. Abundance and biomass responses of microbial food web components to hydrology and environmental gradients within a floodplain of the River Danube. Microb. Ecol. 64, 39–53.
Park, Y., Cho, K.H., Park, J., Cha, S.M., Kim, J.H., 2015. Development of early-warning protocol for predicting chlorophyll-*a* concentration using machine learning models in freshwater and estuarine reservoirs, Korea. Sci. Total Environ. 502, 31–41.
Recknagel, F., French, M., Harkonen, P., Yabunaka, K.-I., 1997. Artificial neural network approach for modelling and prediction of algal blooms. Ecol. Modell. 96, 11–28.
Recknagel, F., Kim, D.-K., Joo, G.-J., Cao, H., 2017. Response of *Microcystis* and *Stephanodiscus* to alternative flow regimes of the regulated river Nakdong (South Korea) quantified by model ensembles based on the hybrid evolutionary algorithm (HEA). River Res. Appl. 33, 949–958.
Reynolds, C.S., 2006. The Ecology of Phytoplankton. Cambridge University Press, New York, NY, USA.
Rueda, F., Moreno-Ostos, E., Armengol, J., 2006. The residence time of river water in reservoirs. Ecol. Modell. 191, 260–274.
Ryu, H.-S., Park, H.-K., Lee, H.-J., Shin, R.-Y., Cheon, S.-U., 2016. Occurrence and succession pattern of cyanobacteria in the upper region of the Nakdong River: factors influencing *Aphanizomenon* bloom. J. Korean Soc. Water Environ. 32, 52–59.
Salmaso, N., Braioni, M.G., 2008. Factors controlling the seasonal development and distribution of the phytoplankton community in the lowland course of a large river in Northern Italy (River Adige). Aquat. Ecol. 42, 533–545.
Shin, J.-H., Chung, J.-Y., 2011. The four major rivers restoration project in South Korea. P. I. Civil Eng.–Civ. En. 164, 19–26.
Singh, K.P., Malik, A., Mohan, D., Sinha, S., 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. Water Res. 38, 3980–3992.
Specht, D.F., 1990. Probabilistic neural networks. Neural Netw. 3, 109–118.
Strobl, R.O., Forte, F., Pennetta, L., 2007. Application of artificial neural networks for classifying lake eutrophication status. Lakes Reserv. Res. Manag. 12, 15–25.
Tomás, P., Oscoz, J., Durán, C., Fernández, D., Marín, J.P., Miranda, R., Leunda, P.M., et al., 2010. Distribution of the bloom-forming diatom *Didymosphenia geminata* in the Ebro River basin (North-East Spain) in the period 2006-2009. Aquat. Invasions 5, 285–289.
Wang, F., Wang, X., Chen, B., Zhao, Y., Yang, Z., 2013. Chlorophyll *a* simulation in a lake ecosystem using a model with wavelet analysis and artificial neural network. Environ. Manage. 51, 1044–1054.
Wetzel, R.G., Likens, G.E., 1991. Limnological Analyses, 2nd. ed. Springer Verlag, New York, NY, USA.
Williamson, C.E., Morris, D.P., Pace, M.L., Olson, O.G., 1999. Dissolved organic carbon and nutrients as regulators of lake ecosystems: resurrection of a more integrated paradigm. Limnol. Oceanogr. 44, 795–803.
Wu, X., Kong, F., 2009. Effects of light and wind speed on the vertical distribution of *Microcystis aeruginosa* colonies of different sizes during a summer bloom. Int. Rev. Hydrobiol. 94, 258–266.
Wu, T., Qin, B., Zhu, G., Luo, L., Ding, Y., Bian, G., 2013. Dynamics of cyanobacterial bloom formation during short-term hydrodynamic fluctuation in a large shallow, eutrophic, and wind-exposed Lake Taihu, China. Environ. Sci. Pollut. Res. Int. 20, 8546–8556.
Wu, N., Huang, J., Schmalz, B., Fohrer, N., 2014. Modeling daily chlorophyll *a* dynamics in a German lowland river using artificial neural networks and multiple linear regression approaches. Limnology 15, 47–56.
Xu, H., Paerl, H.W., Qin, B., Zhu, G., Gaoa, G., 2010. Nitrogen and phosphorus inputs control phytoplankton growth in eutrophic Lake Taihu, China. Limnol. Oceanogr. 55, 420–432.
Yabunaka, K., Hosomi, M., Murakami, A., 1997. Novel application of a back-propagation artificial neural network model formulated to predict algal bloom. Water Sci.

Technol. 36, 89–97.
Yu, S.-N., Chen, Y.-H., 2007. Electrocardiogram beat classification based on wavelet transformation and probabilistic neural network. Pattern Recognit. Lett. 28, 1142–1150.
Yu, J.J., Lee, H.J., Lee, K.L., Lyu, H.S., Whang, J.W., Shin, L.Y., Chen, S.U., 2014. Relationship between distribution of the dominant phytoplankton species and water temperature in the Nakdong River, Korea. Korean J. Ecol. Environ. 47, 247–257.
Zar, J.H., 1999. Biostatistical Analysis, third ed. Prentice Hall, Upper Saddle River, USA.