



Article

Optimal Band Selection for Airborne Hyperspectral Imagery to Retrieve a Wide Range of Cyanobacterial Pigment Concentration Using a Data-Driven Approach

Wonjin Jang ¹, Yongeun Park ², JongCheol Pyo ³, Sanghyun Park ⁴, Jinuk Kim ¹, Jin Hwi Kim ¹, Kyung Hwa Cho ⁵, Jae-Ki Shin ⁶ and Seongjoon Kim ^{2,*}

- ¹ Graduate School of Civil, Environmental and Plant Engineering, Konkuk University, Seoul 05029, Korea; jangwj0511@konkuk.ac.kr (W.J.); saertt@konkuk.ac.kr (J.K.); jinhwi@konkuk.ac.kr (J.H.K.)
- ² Division of Civil and Environmental and Plant Engineering, Konkuk University, Seoul 05029, Korea; yepark@konkuk.ac.kr
- ³ Center for Environmental Data Strategy, Korea Environment Institute, Sejong 30147, Korea; jcpyo@kei.re.kr
- ⁴ Water Quality Assessment Research Division, National Institute of Environmental Research, Incheon 22689, Korea; pbaby75@korea.kr
- ⁵ School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Korea; khcho@unist.ac.kr
- ⁶ Office for Busan Region Management of the Nakdong River, Korea Water Resources Corporation (K-Water), Busan 49300, Korea; jaekishin@kwwater.or.kr
- * Correspondence: kimsj@konkuk.ac.kr; Tel.: +82-2-444-0186



Citation: Jang, W.; Park, Y.; Pyo, J.; Park, S.; Kim, J.; Kim, J.H.; Cho, K.H.; Shin, J.-K.; Kim, S. Optimal Band Selection for Airborne Hyperspectral Imagery to Retrieve a Wide Range of Cyanobacterial Pigment Concentration Using a Data-Driven Approach. *Remote Sens.* **2022**, *14*, 1754. <https://doi.org/10.3390/rs14071754>

Academic Editor: Mathew Grant Allan

Received: 23 February 2022

Accepted: 4 April 2022

Published: 6 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Understanding the concentration and distribution of cyanobacteria blooms is an important aspect of managing water quality problems and protecting aquatic ecosystems. Airborne hyperspectral imagery (HSI)—which has high temporal, spatial, and spectral resolutions—is widely used to remotely sense cyanobacteria bloom, and it provides the distribution of the bloom over a wide area. In this study, we determined the input spectral bands that were relevant in effectively estimating the main two pigments (PC, Phycocyanin; Chl-a, Chlorophyll-a) of cyanobacteria by applying data-driven algorithms to HSI and then evaluating the change in the spatio-temporal distribution of cyanobacteria. The input variables for the algorithms consisted of reflectance band ratios associated with the optical properties of PC and Chl-a, which were calculated by the selected hyperspectral bands using a feature selection method. The selected input variable was composed of six reflectance bands (465.7–589.6, 603.6–631.8, 641.2–655.35, 664.8–679.0, 698.0–712.3, and 731.4–784.1 nm). The artificial neural network showed the best results for the estimation of the two pigments with average coefficients of determination 0.80 and 0.74. This study proposes relevant input spectral information and an algorithm that can effectively detect the occurrence of cyanobacteria in the weir pool along the Geum river, South Korea. The algorithm is expected to help establish a preemptive response to the formation of cyanobacterial blooms, and to contribute to the preparation of suitable water quality management plans for freshwater environments.

Keywords: phycocyanin; chlorophyll-a; hyperspectral image; data-driven model; band selection

1. Introduction

Cyanobacterial blooms, also named algal blooms, are massive growths of cyanobacteria in eutrophic lakes or rivers with low flow rates which change the color of the water surface to green. Recently, the duration, intensity, and frequency of algal blooms have gradually been increasing due to global warming, urbanization, and changes in precipitation patterns [1]. Excessive occurrence of algae poses environmental, societal, and economic threats due to the resulting damage to the aquatic ecosystems and the health of the aquatic animals and plants therein, as well as increased water purification costs due to harmful algal blooms (HABs)-associated toxins [2,3]. Understanding the patterns of distribution of

algal blooms is very important for reducing the negative effects of the blooms on aquatic environments [4]. The study area, Baekje weir (BJW) pool along the Geum river, South Korea, has problems due to the periodic occurrence of algal blooms and is in need of improved monitoring and management tools.

Conventional water quality monitoring in rivers and lakes is implemented at a specific time and location to measure concentrations. Single-point sampling methods are limited in their ability to identify the spatial and temporal distributions of HABs. Fixed sensors have also been widely used as real-time and automatic systems for monitoring HABs. In situ sensor networks are limited in terms of their spatial and temporal resolutions and reliability, which is due to the maintenance problem involving the water quality observation sensor. To address these monitoring issues, many researchers have applied remote sensing techniques. Remote sensing can achieve regular-interval monitoring and detection of a wide area by measuring the reflectance or emission of objects.

Various types of multispectral and hyperspectral imagery have been used to remotely detect harmful algal blooms in coastal and inland waters [5–9]. In particular, hyperspectral imageries provide fine spectral data, thereby leading us to investigate the detailed statistical relations. Many empirical models, such as a band ratio algorithm that considers the relationship between spectral reflectance and algal pigments, were constructed to estimate algal pigments using spectral reflectance from remotely sensed data [10]. The band ratios in these models have typically been calculated using a small number of bands (less than four bands, e.g., 443, 665, 704, and 740 nm) due to the limited number of reflectance bands available from 400 nm to 900 nm. Eight different bands from satellite data were applied to machine learning models to estimate Chlorophyll-a (Chl-a) in coastal water [11,12]. A few of these studies have used airborne hyperspectral imagery with a fine spectral resolution (4 to 7 nm). Kutser et al. [13] used AISA hyperspectral spectrometer to obtain fine spectral data (7 nm resolution), but they applied an inverse model using a specific spectral band data to estimate Chl-a. Lunetta et al. [14] and Li et al. [15] retrieved Chl-a using a band ratio algorithm based on 673, 675, 693, and 704 nm, which were values obtained from AVIRIS and AISA hyperspectral imagery. Pyo et al. [16,17] obtained 86 reflectance bands from 400 to 800 nm taken by the AISA eagle sensor (4–5 nm resolution) and they applied all band data to construct machine learning models to estimate Chl-a and phycocyanin (PC) in a reservoir. Keller et al. [18] used 111 spectral bands from the hyperspectral snapshot sensor Cubert UHD 285, and the band data was used to develop linear and nonlinear bio-optical models for Chl-a in a river. Airborne hyperspectral imagery may provide fine spectral data, which can be used to investigate the statistical relations in detail. However, the existing literature studies have rarely evaluated the statistical relationships between fine-resolution spectral reflectance and each algal pigment. Some studies have used conventional spectral reflectance data to construct band ratio algorithms based on a small number of bands even though they observed fine-resolution hyperspectral [13–15]. Some studies have applied numerous spectral reflectance (more than 86 bands) data to a machine learning model [16–18]. Applying numerous reflectance bands may lead to increases in complexity, uncertainty, and input noise in the model and cause overfitting, low accuracy, and difficult interpretation [19]. That is, there have been few studies examining relevant feature selection for inputs of empirical models among fine-resolution hyperspectral data.

Therefore, to construct accurate and reliable empirical models, it is important to select relevant input reflectance bands. PC estimation mainly uses the wavelengths from 600 to 625 nm and from 700 to 720 nm in fresh water, and the 625 to 650 nm wavelength is used in some cases as well [20–22]. In the case of Chl-a, the wavelength range from 670 to 700 nm was used [23,24]. In general, the reflectivity of the eutrophic water surface shows low reflectance at 600 to 620 nm and 670 to 675 nm, and reflectance at around 650 nm. It also shows a peak at 700 to 724 nm, which is attributed to the strong absorption and fluorescence of PC and Chl-a [25,26]. The spectral wavelengths used to estimate PC or Chl-a vary according to the sites and concentration levels. In a high Chl-a concentration condition, the peak at 550 to 580 nm is slightly shifted to higher wavelengths, and the magnitude

of the peak near 700 nm is significantly higher than it is for the low condition. Moreover, the magnitude of peaks is shifted depending on the spatial distribution of cyanobacterial density [25,27].

Therefore, the objectives of this study are to (a) determine relevant spectral bands within the range of 400–800 nm for use in empirical algorithms to effectively retrieve two pigments, (b) apply the selected input spectral reflectance data to empirical algorithms such as data-driven models, (c) evaluate the performance of the algorithms, and (d) suggest the best input spectral band combination and an algorithm that shows good performance for understanding the characteristics of the spatial distribution in a weir pool.

2. Materials and Methods

2.1. Study Area

BJW in South Korea was constructed in October 2011 to secure water storage in Geum river and reduce flood and drought damage (Figure 1). The weir is located 60 km upstream from the river estuary, and the basin area of the weir is 7976 km² while its width and height are 311 m (movable 120 m, fixed 191 m), and 5.5 m, respectively. The manageable water level and storage capacity of water are 4.2 m and 24.2 × 10⁶ m³, respectively. Since the construction of the weir, cyanobacteria blooms have become a major concern due to continued drought, increases in water temperature, accumulation of nutrients load, and increased residence time. The algae alert based on an increase in the density of cyanobacteria has been steadily issued from 2012 to 2018.

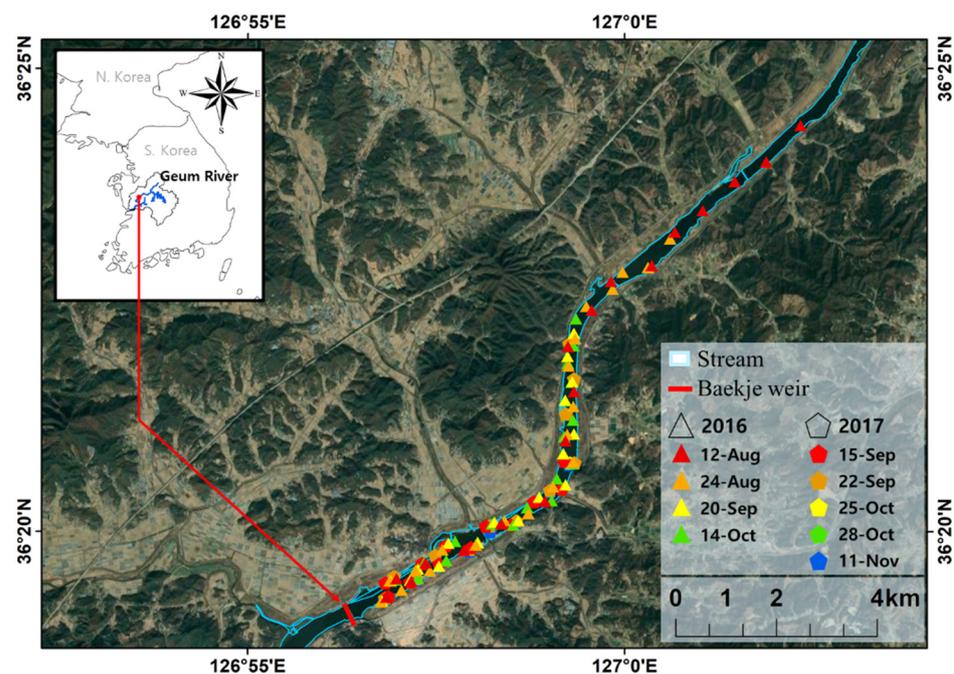


Figure 1. Map of airborne hyperspectral scanning area, Location of Baekje weir and monitoring points indicated, respectively, by red line and marks (2016: triangles; 2017: pentagons).

2.2. Data Acquisition

A total of nine monitoring events were conducted to collect water samples, field hyperspectral reflectance, and airborne hyperspectral imagery during the 2-year period from 2016 to 2017 in the BJW (Figure 1 and Table 1). Water sampling and field hyperspectral measurement were performed concurrently with image acquisition by an aircraft equipped hyperspectral sensor. Water samples were analyzed for PC, Chl-a, and suspended solid.

Table 1. Descriptive two-year water quality sampling statistics (mg/m³) of Baekje weir.

Date	Number of Samples	PC				Chlorophyll-a			
		* Avg.	* Std.	Min	Max	Avg.	Std.	Min	Max
12 August 2016	18	35.46	36.10	6.04	146.99	40.65	23.38	14.19	111.40
24 August 2016	20	38.07	23.58	12.25	100.00	37.24	8.02	25.95	61.44
20 September 2016	17	1.23	0.27	0.83	1.64	25.51	11.32	11.85	60.88
14 October 2016	20	0.33	0.17	0.19	0.88	28.21	9.38	13.74	46.17
15 September 2017	12	8.34	0.66	7.41	9.66	47.28	8.54	30.24	61.52
22 September 2017	12	12.63	3.96	7.64	21.69	17.57	3.80	14.08	27.89
25 October 2017	12	3.51	0.67	2.64	4.56	13.18	2.85	10.56	20.92
28 October 2017	12	4.35	4.52	1.18	14.77	10.54	2.28	8.45	16.73
11 November 2017	11	0.35	0.14	0.23	0.71	22.00	6.76	12.76	38.43

* Avg.: Average, * Std.: Standard deviation.

2.2.1. Hyperspectral Imagery

Hyperspectral sensors measure reflectivity over hundreds of spectral bands for inland water. The high spectral resolution data of water surface reflectance are ideal data for the development and validation of PC algorithms [28,29]. The hyperspectral image sensor (AISA Eagle, SPECIM Inc., Oulu, Finland) has a spectral range of 400 to 970 nm, and a (4–5) nm spectral resolution. The spatial resolution of the hyperspectral imageries taken from the aircraft is 2 m, and the average river width in the 23 km upstream, where the observation was conducted, is 310 m (150 cells). Field hyperspectral reflectance data were collected using hand-held spectroradiometer (FieldSpec HandHeld2, ASD, Inc., Longmont, CO, USA). Field hyperspectral reflectance includes water surface reflectance, sky radiance, and irradiance [30]. Geometry correction and atmospheric correction using field hyperspectral data and MODTRAN 6 software were performed to improve the surface reflectance of the captured hyperspectral image [31].

2.2.2. Chl-a and PC Sampling

Concentrations of chlorophyll-a and phycocyanin were measured for a total of 134 samples collected through in situ monitoring (Table 1). The concentration of Chl-a was measured using the following equation [32]:

$$\text{Chlorophyll - a (mg/m}^3\text{)} = \frac{\left(11.64 \frac{a_{663}}{a_{750}} - 2.16 \frac{a_{645}}{a_{750}} + 0.10 \frac{a_{630}}{a_{750}}\right) \times V_1}{V_2}$$

where V_1 is the volume of supernatant, V_2 is the total volume of filtered sample, and a is the absorbance by each wavelength of supernatant. The supernatant was extracted using the solvent-extraction method and centrifugation. PC was measured using the following equation with a laboratory experiment:

$$\text{Phycocyanin (mg/m}^3\text{)} = \frac{a_{620} - (0.474 \times a_{652})}{5.34}$$

where a is the absorbance by each wavelength of supernatant. The supernatant was extracted by physical force using the freezing and thawing method and centrifugation [33,34]. The absorbance of the sample was measured using a Cary 5000 UV-vis-NIR Spectrophotometer (Agilent Inc., Santa Clara, CA, USA) that provided a wavelength range from 200 to 3300 nm.

2.3. Selection of Input Bands

Input data of the PC and Chl-a optical models were extracted from the hyperspectral imagery (HSI) of the same location at which water sampling had been conducted (Figure 2). The band selection was performed using random forest feature selection (RFFS) for more

efficient and improved Chl-a and PC estimations with varying concentrations according to time and space from his (theoretical background for random forest is described in Section 2.4.2). After calculating the importance for each variable in a regression problem, RFFS removes insignificant variables. It is particularly effective for the dimensionality reduction in data with hundreds of consecutive data volumes, such as hyperspectral images [35,36]. Further, the reflectivity of 400 to 450 nm and above 900 nm were removed due to the influence of atmospheric scattering, absorption of colored dissolved organic matter, and noise [37–39]. In addition, RFFS was performed for three concentration sections for each of PC and Chl-a. The input variables consisted of a total of 9 band ratio from cyanobacteria two pigments (PC, Chl-a) key feature reflectance (R_{pp}/R_{pa} , R_{cp}/R_{pa} , R_{gp}/R_{pa} , R_{pp}/R_{ca} , R_{cp}/R_{ca} , R_{gp}/R_{ca} , R_{pp}/R_{wa} , R_{cp}/R_{wa} , and R_{gp}/R_{wa}). Band ratio data were used to reduce the effects of atmospheric and irradiance [20,40–42].

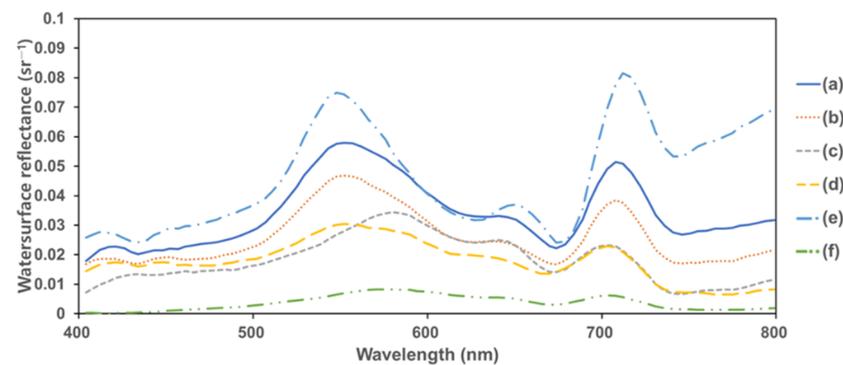


Figure 2. A hyperspectral reflectance curve extracted from a sampling point. The positions of peak and absorption were shifted according to the concentration of the water quality factors. (a) PC, 146.99; Chl-a, 111.4; and SS, 40.14. (b) PC, 54.81; Chl-a, 66.18; and SS, 16.93. (c) PC, 1.54; Chl-a, 60.88; and SS, 16.67. (d) PC, 15.89; Chl-a, 51.54; and SS, 9.47. (e) PC, 100; Chl-a, 36.78; and SS, 13.2. (f) PC, 8.93; Chl-a, 17.52; and SS, 7.33.

2.4. Development of Optical Models to Retrieve Pigments

Six regression models were used to predict the concentrations of the two pigments (i.e., PC and Chl-a) for cyanobacteria detection. The models include Partial Least Squares (PLS), tree-based ensemble regressions (Random Forest (RF), Gradient Boosting (GB), Support Vector Machine (SVM), K-Nearest Neighbor regression (KNN), and Artificial Neural Network (ANN). All models are implemented in Python open-source library scikit-learn and Keras.

2.4.1. Partial Least Squares

PLS regression is a method of generalizing and combining the features of principal components analysis and multiple regression to predict dependent variables by extracting a set of orthogonal factors—called latent variables—with the best predictive power from the predictors [43,44]. The PLS regression predictor is:

$$Y = \beta \cdot X + \varepsilon$$

where Y is the vector of dependent (predictable) variables, X is the matrix of independent (predictors) variables, β is the matrix of regression coefficient, and ε is the matrix of error.

2.4.2. Tree-Based Ensemble Regression

Random Forest (RF) regression is an ensemble learning method that combines a set of regression trees. Ensemble models aim to reduce bias and/or variance of such weak learners by combining several of them together to create a strong learner that achieves better performance. RF can handle many data sets in a short time, and it has both high

prediction accuracy and the capability of determining variable importance [45,46]. The RF regression predictor is:

$$Y_{pr} = \frac{1}{N} \sum_{n=1}^N T(x)$$

where Y_{pr} is the predictable variable, N is the number of regression trees, $T(x)$ is the result of each regression tree, and x is the input variables. RF increases tree diversity by growing trees from different training data sets that are generated through a procedure called bagging; this is to avoid correlation between trees. Bagging is used to generate training data by randomly resampling the original data set. Gradient Boosting regression tree (GB) is a model that is improved by applying a boosting statistical technique to the RF, and it sequentially creates a new regression tree that minimizes the residual of the existing tree. The sequential tree creation process is the form of gradient descent; i.e., a new tree is added to optimize the model at each step to minimize the loss function [47].

2.4.3. K-Nearest Neighbors Regression

K-nearest neighbors (KNN) regression is a nonparametric regression method that is a kind of instance-based lazy learning algorithm. Nonparametric regression is characterized by the fast learning of complex target functions without loss of information, because no assumptions are made about the distribution of the data during the training phase. KNN works based on the weighted average of the k-nearest neighbors, which applies the inverse of their distance [48]. The learning method of KNN calculates the Euclidean distance of the input data, then rearranges the existing data by increasing or decreasing the distance. Next, the inverse distance weighted average is calculated by considering the K-nearest neighbors, and the number K of nearest neighbors is optimized to minimize the loss function.

2.4.4. Support Vector Machine

Support Vector Machine (SVM) is machine learning that is specialized in pattern recognition for classification, but it can be employed in regression analysis for nonlinear and high-dimensional data using kernel functions [49]. The principle of SVM is to (1) create a linear non-probabilistic decision boundary that can classify data sets, (2) calculate the vertical distance (margin) between a support vector (a vector that determines a decision boundary) and the decision boundary, and (3) optimize to maximize the margin by adjusting the boundary. Support Vector Regression (SVR) applies an ϵ -insensitive loss function to the SVM [50,51]. As a result, data are located closer than ϵ from the regression function, and data located at a distance greater than ϵ are penalized. In the same way as SVM, the optimization process determines a regression function that maximizes margin while minimizing penalty.

2.4.5. Artificial Neural Network

Artificial Neural Network (ANN) is a machine learning model based on the information processing system in the biological brain [52]. It is mainly composed of an input layer, a hidden layer, and an output layer; here, a layer is a collection of nodes called artificial neurons. ANN is widely applied to solve the nonlinear function approximation problem, and it has the features of creating large-scale parallel networks to explore the characteristics of data and learning relationships directly from the data [53]. Further, it is not very sensitive to noise in data. In this study, single layer perceptron was constructed to predict PC and Chl-a concentrations.

2.4.6. Regression Model Optimization

The following parameters are used for optimization for each regression technique: Tree ensemble (min sample leaf, depth, number of estimators, and minimum ratio of sample split), SVM (kernel function, C, and gamma), KNN (number of neighbors, weight, and leaf size), and ANN—single layer perceptron structure (activation function of the input

and output layers, and number of nodes). The input data of the models were divided into training and test sets in a 7:3 ratio; 94 were used for training data and 40 were used for test data. The input data (water surface reflectance) were normalized with StandardScaler using the mean and standard deviation, while the output data (PC, Chl-a concentration) were normalized with MinMaxScaler to have a range of (0–1). The training loss function of the models was set to be Mean Square Error (MSE), and parameter optimization was performed for each of the six models using 100 randomly selected data sets. For more detailed descriptions of the parameters and libraries used in this study, refer to Pedregosa et al. [54].

2.5. Performance Evaluation Parameters

The model performances were evaluated using the four statistics of the coefficient of determination (R^2), Nash–Sutcliffe efficiency (NSE), and root mean square error (RMSE). The statistics are computed as shown in the following equations:

$$R^2 = \left(\frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (P_i - \bar{P})^2}} \right)^2$$

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2}$$

where O_i is the observed value for algae concentration (Chl-a, PC); P_i is the estimated value for algae concentration; \bar{O} and \bar{P} are the averages of the observed and estimated values, respectively; and n is the total number of data. R^2 expresses how well the estimated values of the model describe the observed values, and it ranges from 0 to 1; a value close to 1 indicates the best-fit model, and values greater than 0.5 are typically considered acceptable [55,56]. NSE is used to evaluate the efficiency of the model. NSE has a range of $(-\infty$ to 1), and when the value of NSE is 1, this means that the observed value and the estimated value are perfectly matched. Further, for NSE values greater than 0, the values calculated from the model are acceptable, but values less than 0 are judged to be unacceptable [57]. RMSE is a statistical value that judges the error between the observed value and the simulated value. The closer RMSE is to 0, the less error there is in the model result. If the value of RMSE is less than half of the standard deviation of the observed value, then the performance of the model is determined to be suitable [58].

3. Results

3.1. Band Selection

The ranges of concentrations obtained from 134 samples were (0.19–146.99) for PC and (2.28–111.40) mg/m^3 for Chl-a (Table 1). High concentrations of PC were observed in August 2016 (36.77 mg/m^3) and September 2017 (10.49 mg/m^3), and the concentrations decreased dramatically on subsequent dates. A similar concentration decrease pattern was observed for Chl-a (in 2016 and 2017: 40.65 to 25.51 and 47.28 to 10.54 mg/m^3 , respectively).

Based on the RFFS result, six band ranges were selected whose reflectivity was shown to change sensitively according to the concentration conditions (PC, Chl-a) (Tables 2 and 3). For each band, the maximum or minimum value within the range was used to consider the shift in the position of the peak or absorption according to the cyanobacterial biomass, suspended solid (SS) concentration (Figure 2).

Table 2. Result of band selection according to PC and Chl-a concentration using random forest feature importance.

Pigment	Concentration (mg/L)	No. of Bands Selected	Band (nm)
PC	0–3	14	452, 470, 484, 604, 674, 684, 708, 712, 717, 727, 784, 789, 794, 799
	3–15	14	466, 525, 679, 741, 746, 751, 755, 760, 765, 770, 775, 779, 784, 789
	15–147	20	457, 461, 470, 475, 488, 497, 502, 507, 511, 516, 520, 525, 529, 543, 665, 670, 674, 693, 717, 784
Chl-a	0–20	21	566, 571, 580, 585, 590, 604, 646, 651, 655, 660, 665, 670, 674, 698, 703, 708, 717, 722, 727, 784, 789
	20–35	25	452, 466, 470, 488, 507, 511, 516, 520, 525, 529, 539, 543, 552, 674, 689, 722, 736, 751, 755, 760, 770, 775, 784, 789, 794
	35–111	22	488, 548, 590, 599, 604, 627, 646, 651, 655, 689, 698, 703, 708, 712, 717, 731, 736, 741, 755, 779, 789, 794

Table 3. Cyanobacteria key features' reflectance bands (nm).

Pigment	Type	MIN		MAX	
PC	peak	641.18	~	655.35	R _{pp}
	* abs.	603.60	~	631.76	R _{pa}
Chl-a	peak	698.04	~	712.33	R _{cp}
	abs.	664.81	~	679.03	R _{ca}
Green	peak	465.74	~	589.58	R _{gp}
Water	abs.	731.42	~	784.11	R _{wa}

* abs.: absorbance.

3.2. Model Development

To develop optical models to retrieve pigments, six regression models (PLS, RF, GB, SVM, KNN, and ANN) and three input variable cases were compared. Band ratio data were used to reduce the effects of atmosphere and irradiance. Table 4 summarizes the training and validation statistics results of each case. In Cases 1 and 2, the Chl-a and PC concentrations were estimated using the spectral reflectance bands that were sensitive to absorption by each pigment, and in Case 3, the two pigments were estimated by all spectral reflectance bands that were sensitive to absorption by both Chl-a and PC. When the bands of Chl-a and PC were considered together (Case 3), the resulting performance was higher than those of the other cases and, in particular, the Chl-a estimation performance was significantly improved in all models. In the comparison between PC estimation models in Case 3, each model showed good regression results, particularly in tree ensemble (RF, GB) and ANN. The tree ensemble model showed the best training performance as R^2 0.77, 0.85, NSE 0.76, 0.84, and RMSE 9.63, 7.78 mg/m³, and validation performance as R^2 0.71, 0.74, NSE 0.74, 0.74, and RMSE 15.38, 15.32 mg/m³. In the Chl-a regression results, most models showed that the regression performance was lower than that of PC, and PLS did not show a proper regression result. ANN indicated the best results in the two pigments for both training (R^2 (0.80, 0.81), NSE (0.65, 0.79), and RMSE (6.09, 11.72) mg/m³) and validation (R^2 (0.67, 0.79), NSE (0.84, 0.92), and RMSE (11.38, 11.92) mg/m³).

Table 4. Performances of six data-driven models on estimation of two pigments in terms of different input spectral band combinations.

Case *	Pigment	Training			Validation			
		Method	R ²	NSE	RMSE	R ²	NSE	RMSE
* Case 1	PC	PLS	0.60	0.60	14.69	0.34	0.33	9.46
		RF	0.73	0.70	11.55	0.51	0.43	12.84
		GB	0.79	0.73	10.55	0.59	0.54	12.13
		SVM	0.68	0.81	12.50	0.71	0.43	19.58
		KNN	0.69	0.69	13.40	0.51	0.34	8.06
		ANN	0.80	0.76	11.41	0.68	0.49	9.99
		Avg.	0.71	0.72	12.35	0.56	0.43	12.01
* Case 2	Chl-a	PLS	0.29	−1.44	13.73	0.28	−6.77	10.98
		RF	0.46	−1.02	11.37	0.35	−4.10	15.97
		GB	0.67	0.18	9.62	0.30	0.11	14.00
		SVM	0.48	−0.46	11.77	0.34	0.06	11.14
		KNN	0.38	−1.18	11.09	0.30	−4.75	18.05
		ANN	0.52	0.48	11.60	0.43	0.05	12.09
		Avg.	0.47	−0.57	11.53	0.33	−2.57	13.70
* Case 3	PC	PLS	0.69	0.69	10.98	0.56	0.64	18.09
		RF	0.77	0.76	9.63	0.71	0.74	15.38
		GB	0.85	0.84	7.78	0.74	0.74	15.32
		SVM	0.70	0.69	11.14	0.68	0.70	16.35
		KNN	0.74	0.73	10.39	0.67	0.73	15.61
		ANN	0.81	0.65	11.72	0.79	0.84	11.92
		Avg.	0.76	0.73	10.27	0.69	0.73	15.45
	Chl-a	PLS	0.35	0.35	10.81	0.29	0.79	17.79
		RF	0.59	0.58	8.71	0.42	0.83	15.97
		GB	0.58	0.53	9.16	0.43	0.82	16.63
		SVM	0.47	0.45	9.94	0.46	0.85	15.08
		KNN	0.53	0.51	9.39	0.46	0.83	16.12
		ANN	0.80	0.79	6.09	0.67	0.92	11.38
		Avg.	0.55	0.54	9.02	0.46	0.84	15.50

* Case 1: PC estimation with PC feature spectral reflectance band, * Case 2: Chl-a estimation with Chl-a feature spectral reflectance band, and * Case 3: PC and Chl-a estimation with combined feature reflectance of two pigments.

According to the comparison between observed and estimated pigment concentrations in Case 3, RF and GB tended to be underestimated in the high concentration section, despite the high statistical value in PC estimation (Figure 3b,c). On the other hand, ANN showed good overall performance for all concentration ranges (Figure 3f). In the Chl-a estimation of PLS, RF, and GB, the distributions of the points showed weak linear relationships between (Figure 3a–c). In the cases of RF, GB, and KNN, the estimated values did not exceed more than a certain value (see the red-dotted circle in Figure 3b,c,e). Likewise, ANN had good regression performance for Chl-a estimation, but it was lower than the corresponding performance for PC estimation. From Table 4 and Figure 3, ANN was the most suitable regression model for cyanobacteria estimation using the nine wavelength combination ratios considered in this work.

Additional analysis was performed to confirm the effects of the two bands that are rarely used for inland cyanobacteria estimation on the optical model estimation performance. To determine the effects of R_{gp} and R_{wa} on the estimation performance of the model, the ANN algorithm results shown in Table 4 Case 3, applying all bands (ANN_O) and the ANN algorithm for three cases were compared. Table 5: (A) removed R_{wa} , (B) removed R_{gp} , and (C) removed R_{wa} and R_{gp} from ANN_O . In PC estimation, the R^2 of each case in the training step was almost the same as that of ANN_O , but in the validation step, it decreased by 0.11 on average, compared to the original value of 0.79. The average performance of the three cases of estimating Chl-a was decreased to 0.56 in both training and validation.

Comparing the results between the three cases, both the R_{gp} and R_{wa} band reflections have effects that are almost similar to that of the ANN algorithm, and the two bands have a great influence on the algorithm, particularly in terms of validation.

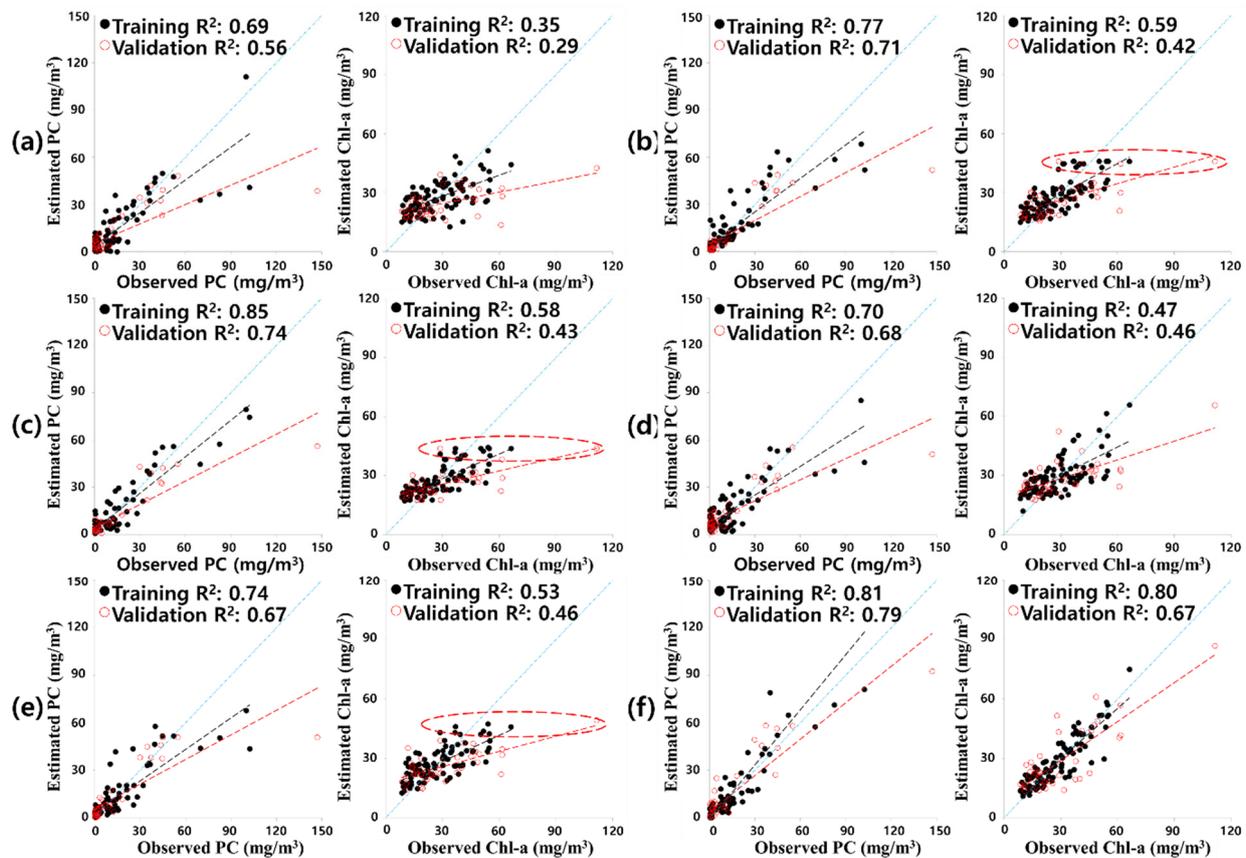


Figure 3. Comparison of concentration of two pigments (PC, Phycocyanin; Chl-a, Chlorophyll-a) between observed and estimated values from each of the following optimized regression methods: (a) partial least square, (b) random forest, (c) gradient boosting regression tree, (d) support vector machine, (e) K-nearest neighbors, and (f) artificial neural network. Training is denoted by a black closed circle. Validation is denoted by a red open circle.

Table 5. Comparison of change in artificial neural network (ANN) performance according to input band change.

Pigment	Case	Training			Validation		
		R ²	NSE	RMSE	R ²	NSE	RMSE
PC	A	0.805	0.799	9.447	0.719	0.188	14.881
	B	0.869	0.848	7.176	0.717	0.328	14.438
	C	0.795	0.666	9.974	0.597	0.216	16.813
Chl-a	A	0.544	0.288	9.360	0.604	0.030	12.898
	B	0.564	0.220	8.934	0.587	−0.992	14.075
	C	0.565	0.272	9.094	0.480	−1.763	14.995

A: Reflection of water absorption of 731.42–784.11 nm removed from Origin ANN input data. B: Reflection of green peak of 465.74–589.58 nm removed from Origin ANN input data. C: Reflection of water absorption and green peak removed from Origin ANN input data.

3.3. Baekje Weir Algae Spatial Distribution Generation

The concentrations of the two pigments were estimated by dividing the study area into seven parts for a more granular study. Z1 is the area adjacent to the weir dam, Z2 and Z3 are tributary-affected areas, Z4 and Z6 are river bends, and Z5 and Z7 are locations

to check for the effects of any changes in flow velocity according to cross-sectional area (Figure 4).

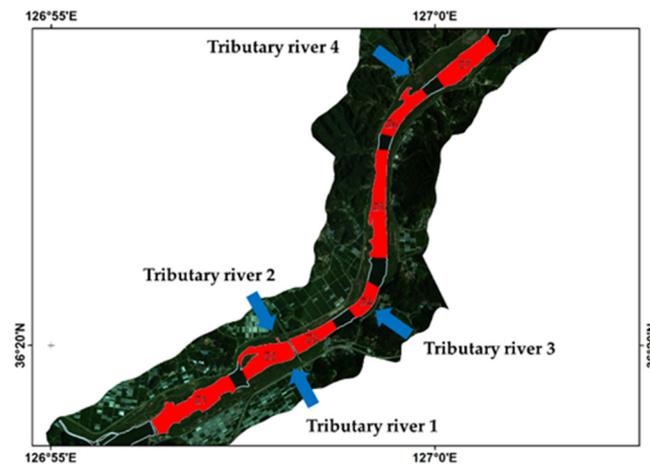


Figure 4. Map of the cyanobacteria distribution divided into seven sections according to the hydrodynamic conditions and tributary rivers.

The spatial distribution characteristics of algae (cyanobacteria) in the BJW were analyzed through the six selected models for the hyperspectral image of 12 August 2016 (Z1), at which time cyanobacteria were predominant (Figure 5). In terms of PC spatial distribution, ANN and PLS more precisely estimated the distribution of high PC concentration near the weir, while the other four models underestimated the PC concentration in the same region. Similarly, ANN, PLS, and KNN performed well in Chl-a spatial distribution; however, in the case of PLS, the Chl-a distribution near the bridge 1 km away from the weir (Z2, Z3) was underestimated.

The spatio-temporal distribution changes of PC and Chl-a over nine sampling days were estimated using the ANN model, which showed good performance in the statistical values and spatial distribution estimation. Figures 6 and 7 show the results, and Appendix A shows the detailed concentration values. The maximum PC and Chl-a concentrations were observed on 12 August 2016, and high concentrations of the two pigments could be seen to be distributed until 24 August 2016. The Chl-a was observed to be above the concentration of 18.05 mg/m^3 over the entire monitoring period, but PC was only observed on a specific date (August 2016; September–October 2017) or in a specific section. The concentration of two pigments tended to decrease as it went upstream, and the change in concentration was greater for PC than it is for Chl-a. The overall PC concentration adjacent to the weir dam was high (12 August 2016), and the PC concentration of Z1 (43.80 mg/m^3) appeared to be higher than those of Z2 and Z3, which were 34.11 and 30.88 mg/m^3 , respectively. On the other hand, in the low PC concentration (15 September 2017), Z2 and Z3 of 8.49 and 9.28 mg/m^3 , respectively, were both similar or higher than Z1 (1.22 mg/m^3). Chl-a was generally distributed throughout the entire upstream with an average concentration of 29.42 mg/m^3 . Unlike Chl-a, high PC concentration were partially observed in Z2, Z3, Z5, and Z6 on August 2016.

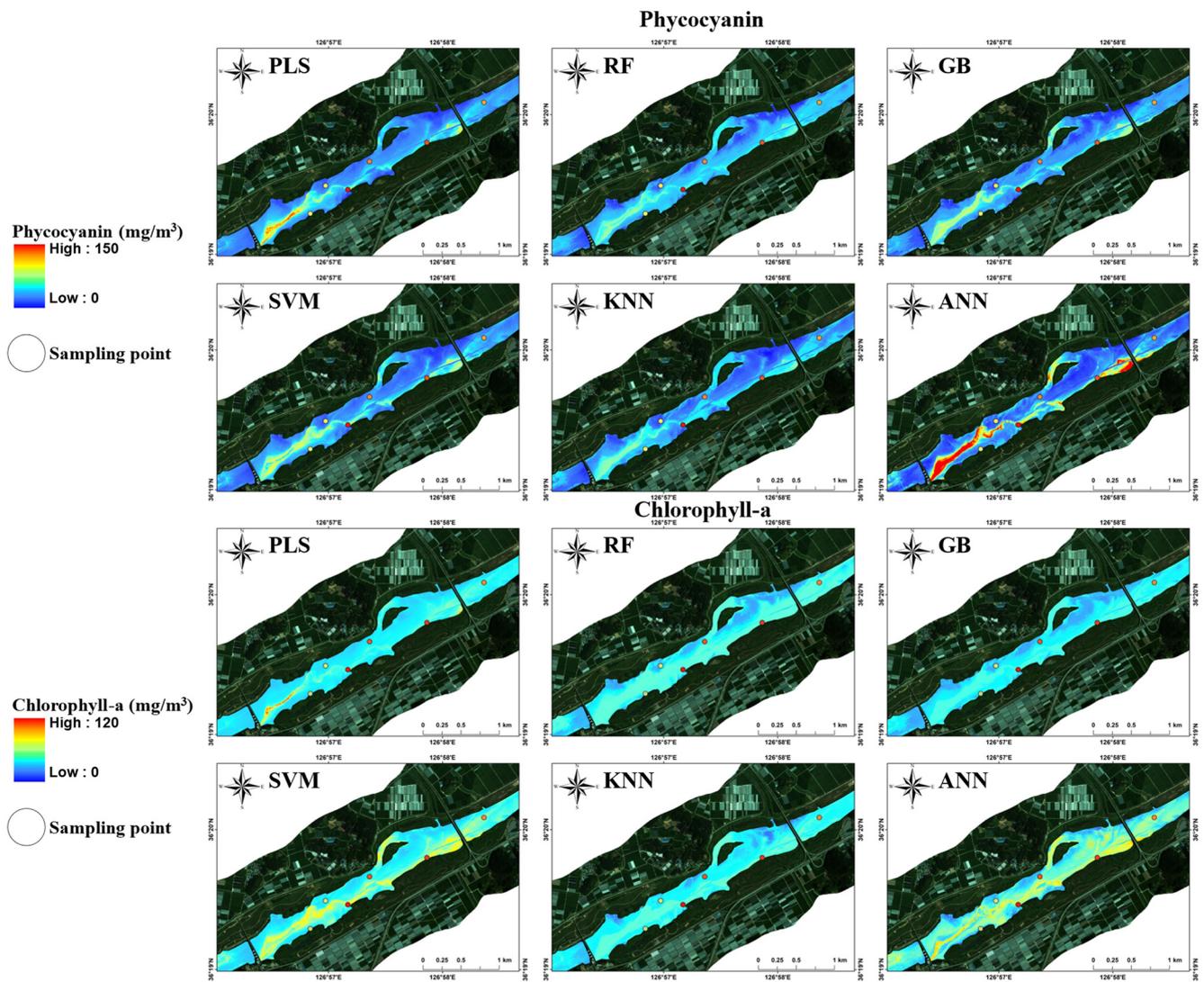


Figure 5. Comparison map of phycocyanin (PC) and chlorophyll-a (Chl-a) concentration distribution by six optimized regression models from hyperspectral imagery near Baekje weir on 12 August 2016, when cyanobacteria were predominantly distributed; sampling point is marked with PC and Chl-a concentration.

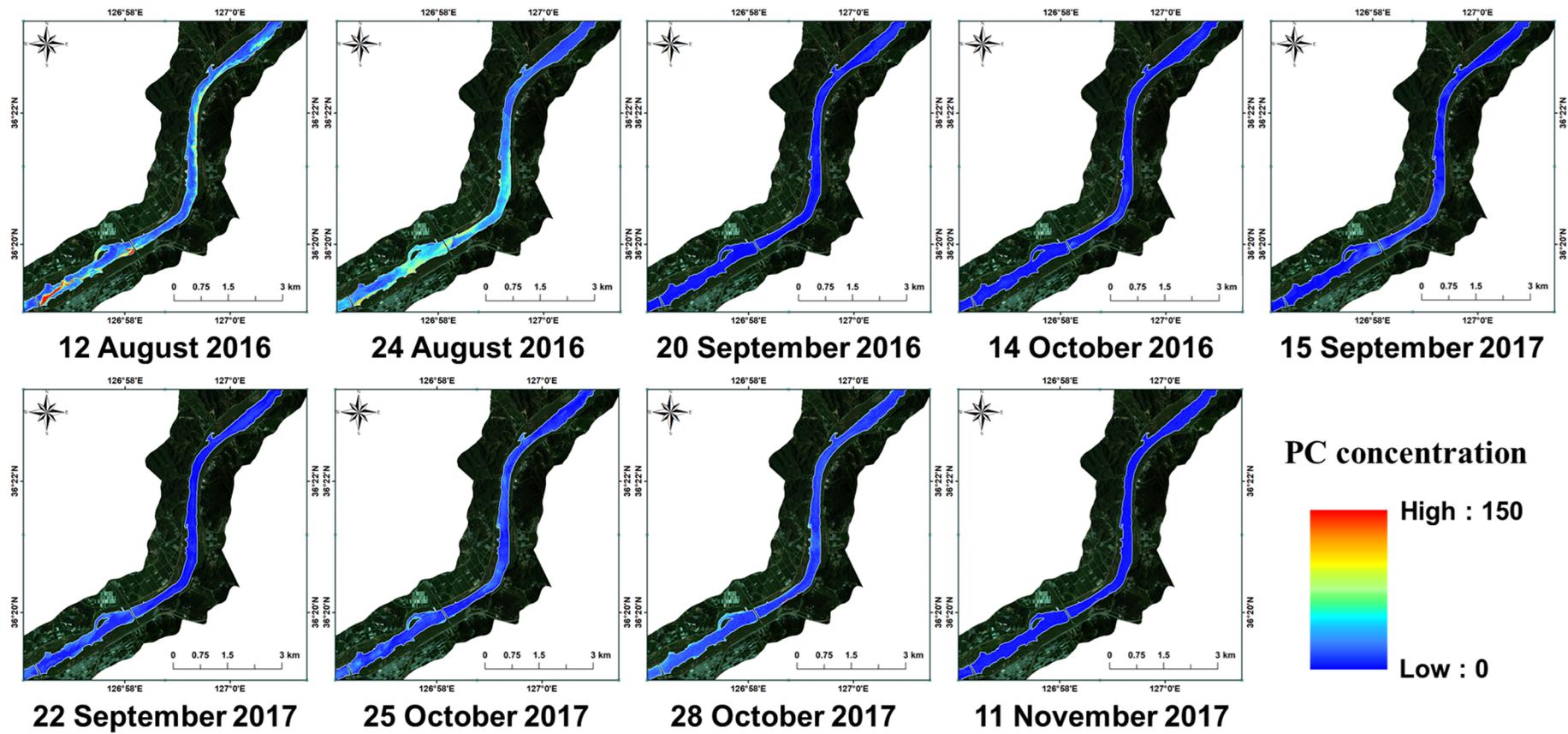


Figure 6. Spatio-temporal change estimation of phycocyanin concentration by applying artificial neural network to the hyperspectral imagery of the nine sampling dates.

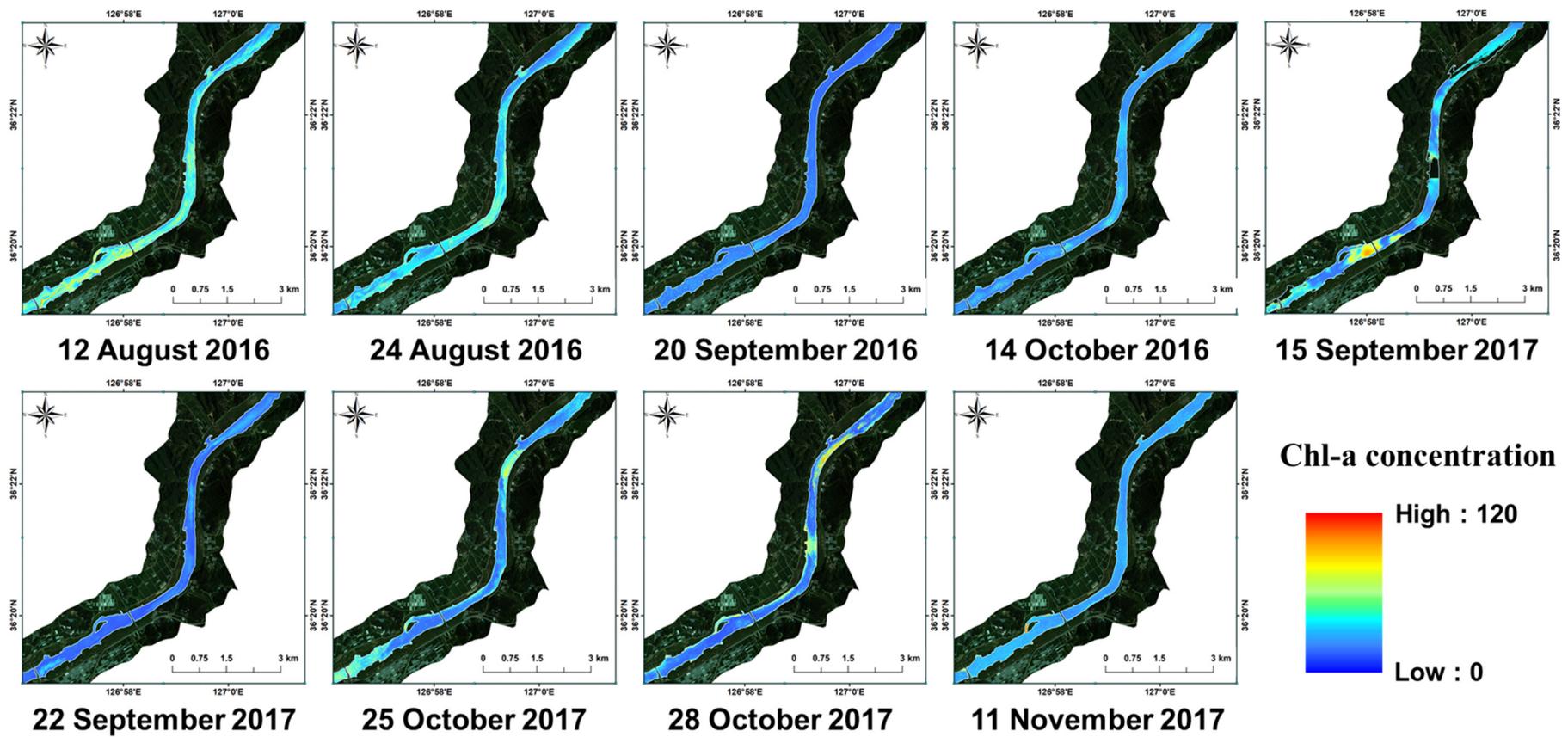


Figure 7. Spatio-temporal change estimation of chlorophyll-a concentration by applying artificial neural network to the hyperspectral imagery of the nine sampling dates.

4. Discussion

4.1. Band Selection for Inland Cyanobacteria Pigments

Band selection is known to critically affect the performance of the optical algorithm for estimation [59]. The selected six spectral bands included the feature reflectance of cyanobacteria and the water characteristic reflectance (algae, inorganic particles, and suspended solids) (Table 3). The first band, called the green peak, appears in R_{gp} at around 460–590 nm due to the light reflection by algal cells and the low absorption of Chl-a [25,60]. The bands at 600–630 and 660–680 nm showed minimum reflectance, which was attributed to the strong absorption of both PC (absorption reflectance of PC: R_{pa}) and Chl-a (absorption reflectance of Chl-a: R_{ca}). The reflectance peak at 640–660 nm appears to be due to the fluorescence characteristics of PC (peak reflectance of PC: R_{pp}), while the peak at 700–712 nm is a relative peak caused by Chl-a absorption (peak reflectance of Chl-a: R_{cp}). The last band at 730–784 nm was affected by scattering stemming from the presence of inorganic particles in water (R_{wa}) [61–63]. The four bands in the range from 600–712 nm contain most of the reflectance used in previous studies that estimated the concentrations of PC and Chl-a inland [20–26]. The two reflectance bands (R_{gp} and R_{wa}) improved the performance of the optical algorithm by minimizing the influence of inorganic particles in inland water and removing interference from suspended particles in the green peak reflectance (Table 5) [61,64]. The empirical method using machine learning is generally the most powerful method for estimating a wide range of variables, as it does not require prior understanding of the complex interactions between water and light reflection. However, the performance of this method varies substantially depending on the water quality conditions, locations, and variable ranges from which the data were obtained [65]. Therefore, the optical algorithms in this study obtained high regression performance with the six-reflection band ranges that can correct for the reflection interference effect.

4.2. Cyanobacteria Optical Algorithm Specialized for BJW

The ANN model showed the highest performance in estimating the two pigments. Previous studies that have involved the estimation of two pigments made comparisons according to the use of input spectral bands and the performance of the model. Song et al. [66] estimated the PC concentrations of Central Indiana USA and South Australia using reflection bands of 620 to 630, 685 to 700, and around 555 to 625 nm obtained from Sentinel-3/OLCI and Hyperion satellite data. Partial least squares–ANN and three-band model were applied for PC estimation, and the results were found to be R^2 (0.84–0.98). Chang and Vannah [67] estimated the microcystin concentration in Lake Erie using Landsat and MODIS (Landsat: 1–5, 7; MODIS: 1–4, 6, 7). Six reflectance bands obtained from the two satellites were used to construct ANN and genetic programming; the resulting R^2 values for the two machine learning models were 0.53 and 0.60, respectively. He et al. [68] estimated the Chl-a concentration in the Gulf of St. Lawrence using 10 bands of MODIS reflectance: 412, 443, 469, 488, 531, 547, 555, 645, 667, and 678 nm. The performances of the five models SVM, ANN, GB, RF, and MLR were compared using R^2 values. Among the models, SVM showed the highest R^2 values of 0.71 and 0.91 in the training and validation steps, respectively. Zhou et al. [69] estimated Chl-a concentration in Dianshan lake using principal component analysis–ANN to obtain in situ hyperspectral surface reflectance. The R^2 values were 0.85 and 0.64 in the training and validation steps, respectively. The ANN model to estimate PC and Chl-a developed in this study showed significantly improved performance over the regression models and offered similar or better performance than the algorithms described in previous studies.

4.3. Spatio-Temporal Distribution Characteristics of Cyanobacteria in BJW

In South Korea, cyanobacterial blooms have intermittently occurred in rivers and reservoirs. However, after the Four Major Rivers Project (2010–2011), many weirs were installed in the major rivers, such as the Nakdong, Geum, and Yeongsan Rivers, which subsequently led to the frequent formation of cyanobacteria blooms, which caused various

water quality problems in the weir pools [70]. In the study area, i.e., the BJW of Geum river, algal bloom caused by cyanobacteria has frequently occurred every summer. The maximum air temperatures on the sampling dates (12 August 2016, 24 August 2016, 22 September 2017, and 28 October 2017) when the two cyanobacteria pigments were dominantly distributed in BJW were 36.2, 34.4, 27.0, and 23.8 °C, respectively (Figure A1). The algae growth rate is sensitive to temperature conditions: cyanobacteria in fresh water shows an optimal growth rate of about 30 °C, while other chlorophytes or diatoms have optimal growth rates of around 20–30 °C [71–74]. In the spatio-temporal change of the two pigments, PC was more sensitive to temperature than Chl-a. In August 2016 (the air and water temperatures were 34.4–36.2 and 30.6–30.9 °C), high PC concentration was widely observed in the BJW under continuous water discharge due to relatively high cyanobacterial growth rates. However, considerably low PC concentrations were collected due to a decrease in the growth rates with decreasing water temperature (23.0 °C) and continuous wash-out by the discharge. Berg and Sutula [75] also reported that if sufficient N and P are provided to allow cyanobacteria to grow, the temperature has the greatest effect on the growth rate.

Chl-a concentration was maintained with relatively high values regardless of sampling dates (e.g., 14 October 2016 vs. 24 August 2016), compared to PC. Based on changes in cyanobacterial and diatom cell density in the BJW [70], relatively high Chl-a and low PC concentrations in September may result from the transition of phytoplankton community [76]. Therefore, it is assumed that there is no noteworthy concentration change compared to PC. The spatial distribution characteristics of PC and Chl-a showed similar patterns when massive cyanobacteria dominated in the upstream of the BJW. In addition, the two pigments are distributed in various patterns in different sections due to the influence of hydrodynamic and tributary river factors. Due to the velocity difference, high concentration is found at the insides of the river bends (Z4, Z6), in large river cross-sections, and at the river boundary, rather than at the center of the river (Z5, Z7). Cyanobacteria grow actively in low velocity and they have a long residence time [77–79]. Comparing Z2 and Z3, Z2 distributed a higher concentration due to the tributary river, advective transport and accumulation of biomass at the weir.

5. Conclusions

Based on nine hyperspectral and water quality monitoring campaigns, this study selected major spectral bands related to the effective retrieval of cyanobacteria pigments (i.e., PC, Chl-a) from hyperspectral reflectance data and developed data-driven algorithms for the remote sensing of cyanobacteria. Six reflection band ranges were selected by the random forest feature selection (RFFS) while considering peak and absorption reflectance as affected by PC, Chl-a, SS, and water characteristics. The sensitive reflectance of each band ratio model using both PC and Chl-a showed better performance in the estimation of each pigment than the individual model using each pigment's sensitive reflectance. This result shows that the two pigments-specific reflectance may be applied simultaneously to construct retrieval models. That is, the selection of relevant reflectance may be critical to the retrieval models, and reflectance bands should be investigated in terms of the pigment sensitivity and improvement of model performance. Overall, this study identified a reflection band that can consider the interference effects of various water characteristics in hyperspectral reflectance imagery, and it therefore provides a useful method for constructing a retrieval model with which to estimate the spatio-temporal concentrations of the main two cyanobacteria pigments (PC and Chl-a). It is expected that future reliable models derived from this study can support the development of efficient management practices for mitigating algal blooms.

Author Contributions: Methodology, W.J. and Y.P.; Writing—Original draft preparation, W.J.; Conceptualization, Y.P.; Data curation, J.P.; Investigation, J.K. and J.H.K.; Validation, K.H.C. and S.P.; Formal analysis, J.-K.S.; and Supervision and Editing, S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was supported by Konkuk University in 2018.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Statistics of Phycocyanin and Chlorophyll-a Spatial Concentration

Table A1. Summary of ANN-estimated phycocyanin and chlorophyll-a concentrations of the seven sections.

Pigment	Date	Zone														Overall
		1		2		3		4		5		6		7		
		MEAN	STD													
PC	12 August 2016	43.80	34.70	34.11	30.16	30.88	20.30	22.10	12.85	37.20	26.53	33.24	23.46	25.79	18.40	32.45
	24 August 2016	36.13	24.68	49.92	15.46	49.21	24.11	39.82	13.87	33.08	16.23	20.78	5.98	11.84	3.52	34.40
	20 September 2016	4.54	4.57	3.82	3.93	3.67	3.34	3.28	3.74	4.57	3.69	5.96	4.15	6.49	4.83	4.62
	14 October 2016	4.51	4.43	4.26	5.35	5.46	7.19	7.45	7.34	3.61	2.81	3.76	2.54	4.48	4.32	4.79
	15 September 2017	1.22	2.01	8.49	7.78	9.28	5.89	10.13	4.52	3.09	3.13	3.74	3.70	1.32	1.76	5.32
	22 September 2017	11.99	9.25	11.45	8.82	8.63	7.22	2.51	2.46	2.49	3.05	3.63	3.80	5.93	4.69	6.66
	25 October 2017	11.07	9.51	12.73	10.73	6.21	8.54	14.91	8.65	11.16	8.10	12.66	8.33	5.37	5.14	10.59
	28 October 2017	19.87	10.72	15.26	13.80	12.32	14.42	10.12	9.82	12.81	9.59	12.97	7.41	10.08	7.87	13.35
	11 November 2017	3.03	4.92	2.78	4.18	2.27	3.15	2.42	2.78	1.59	2.45	2.04	2.86	2.91	3.86	2.43
Avg.	15.13	11.64	15.87	11.13	14.21	10.46	12.53	7.34	12.18	8.40	10.97	6.92	8.25	6.04	12.73	
Chl-a	12 August 2016	48.35	15.29	50.45	15.05	47.14	11.81	45.59	13.02	41.87	11.15	37.71	12.13	33.92	9.77	43.57
	24 August 2016	36.34	10.29	35.12	8.18	39.68	9.01	47.14	9.11	37.10	10.55	30.82	11.60	28.51	9.47	36.39
	20 September 2016	22.36	9.43	21.99	7.01	23.05	6.81	20.44	4.65	19.09	3.99	17.31	2.94	17.02	3.22	20.18
	14 October 2016	24.03	8.01	23.78	6.99	25.06	8.33	29.55	8.29	27.09	4.69	24.28	3.51	26.61	5.26	25.77
	15 September 2017	33.09	12.83	52.56	24.65	40.87	24.36	27.29	8.68	30.42	12.14	30.03	9.68	35.70	5.20	35.71
	22 September 2017	18.69	6.15	16.16	3.36	15.67	4.75	18.04	3.87	17.81	5.35	18.41	5.91	21.55	5.08	18.05
	25 October 2017	37.85	13.87	23.54	12.51	24.65	9.95	26.32	11.76	25.42	10.43	37.51	16.99	26.86	7.73	28.88
	28 October 2017	23.61	13.43	23.45	16.44	24.88	15.78	24.15	12.98	32.52	21.71	47.05	29.11	26.27	18.20	28.85
	11 November 2017	30.03	7.10	29.93	13.58	25.67	5.78	25.27	3.68	26.19	4.52	27.80	7.37	26.65	4.82	27.36
Avg.	30.48	10.71	30.77	11.98	29.63	10.73	29.31	8.45	28.61	9.39	30.10	11.03	27.01	7.64	29.42	

Appendix B. July to November Weather and Water Quality Features in Baekje Weir

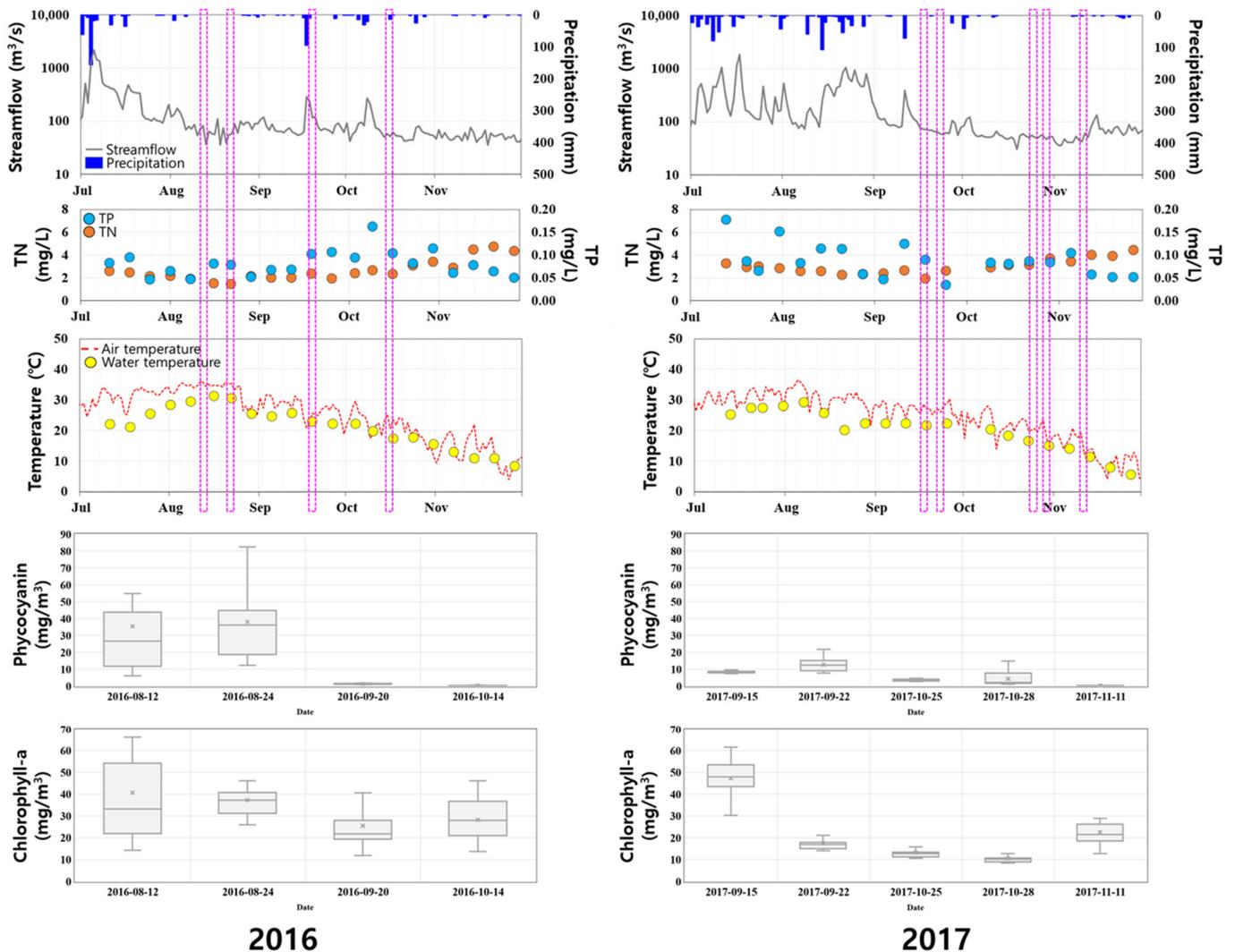


Figure A1. Temporal variations in streamflow; Total nitrogen (TN): Total phosphorus ratio (TP). TN and TP concentrations in air, and water temperatures during the 2016 to 2017 period; the nine sampling dates are indicated by magenta rectangles; PC and Chl-a concentration box plot of each sampling date.

References

- O'Neil, J.M.; Davis, T.W.; Burford, M.A.; Gobler, C.J. The rise of harmful cyanobacteria blooms: The potential roles of eutrophication and climate change. *Harmful Algae* **2012**, *14*, 313–334. [[CrossRef](#)]
- Brooks, B.W.; Lazorchak, J.M.; Howard, M.D.A.; Johnson, M.V.V.; Morton, S.L.; Perkins, D.A.K.; Reavie, E.D.; Scott, G.I.; Smith, S.A.; Steevens, J.A. Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems? *Environ. Toxicol. Chem.* **2016**, *35*, 6–13. [[CrossRef](#)] [[PubMed](#)]
- Hudnell, H.K. The state of U.S. freshwater harmful algal blooms assessments, policy and legislation. *Toxicol.* **2010**, *55*, 1024–1034. [[CrossRef](#)] [[PubMed](#)]
- Chapra, S.C.; Canale, R.P.; Amy, G.L. Empirical Models for Disinfection By-Products in Lakes and Reservoirs. *J. Environ. Eng.* **1997**, *123*, 1–12. [[CrossRef](#)]
- Matthews, M.W.; Bernard, S.; Robertson, L. An algorithm for detecting trophic status (chlorophyll-a), cyanobacterial-dominance, surface scums and floating vegetation in inland and coastal waters. *Remote Sens. Environ.* **2012**, *124*, 637–652. [[CrossRef](#)]
- Matthews, M.W.; Odermatt, D. Improved algorithm for routine monitoring of cyanobacteria and eutrophication in inland and near-coastal waters. *Remote Sens. Environ.* **2015**, *156*, 374–382. [[CrossRef](#)]
- Su, T.C.; Chou, H.T. Application of multispectral sensors carried on unmanned aerial vehicle (UAV) to trophic state mapping of small reservoirs: A case study of Tain-Pu reservoir in Kinmen, Taiwan. *Remote Sens.* **2015**, *7*, 10078–10097. [[CrossRef](#)]

8. Binding, C.E.; Greenberg, T.A.; Jerome, J.H.; Bukata, R.P.; Letourneau, G. An assessment of MERIS algal products during an intense bloom in Lake of the Woods. *J. Plankton Res.* **2011**, *33*, 793–806. [[CrossRef](#)]
9. Duan, H.; Ma, R.; Zhang, Y.; Loisel, S.A.; Xu, J.; Zhao, C.; Zhou, L.; Shang, L. A new three-band algorithm for estimating chlorophyll concentrations in turbid inland lakes. *Environ. Res. Lett.* **2010**, *5*, 044009. [[CrossRef](#)]
10. Shi, K.; Zhang, Y.; Qin, B.; Zhou, B. Remote sensing of cyanobacterial blooms in inland waters: Present knowledge and future challenges. *Sci. Bull.* **2019**, *64*, 1540–1556. [[CrossRef](#)]
11. Kim, Y.H.; Im, J.; Ha, H.K.; Choi, J.K.; Ha, S. Machine learning approaches to coastal water quality monitoring using GOCI satellite data. *GISci. Remote Sens.* **2014**, *51*, 158–174. [[CrossRef](#)]
12. Zhang, Y.; Pulliainen, J.; Koponen, S.; Hallikainen, M. Application of an empirical neural network to surface water quality estimation in the Gulf of Finland using combined optical data and microwave data. *Remote Sens. Environ.* **2002**, *81*, 327–336. [[CrossRef](#)]
13. Kutser, T.; Herlevi, A.; Kallio, K.; Arst, H. A hyperspectral model for interpretation of passive optical remote sensing data from turbid lakes. *Sci. Total Environ.* **2001**, *268*, 47–58. [[CrossRef](#)]
14. Lunetta, R.S.; Knight, J.F.; Paerl, H.W.; Streicher, J.J.; Peierls, B.L.; Gallo, T.; Lyon, J.G.; Mace, T.H.; Buzzelli, C.P. Measurement of water colour using AVIRIS imagery to assess the potential for an operational monitoring capability in the Pamlico Sound Estuary, USA. *Int. J. Remote Sens.* **2009**, *30*, 3291–3314. [[CrossRef](#)] [[PubMed](#)]
15. Li, L.; Sengpiel, R.E.; Pascual, D.L.; Tedesco, L.P.; Wilson, J.S.; Soyeux, A. Using hyperspectral remote sensing to estimate chlorophyll-a and phycocyanin in a mesotrophic reservoir. *Int. J. Remote Sens.* **2010**, *31*, 4147–4162. [[CrossRef](#)]
16. Pyo, J.C.; Duan, H.; Baek, S.; Kim, M.S.; Jeon, T.; Kwon, Y.S.; Lee, H.; Cho, K.H. A convolutional neural network regression for quantifying cyanobacteria using hyperspectral imagery. *Remote Sens. Environ.* **2019**, *233*, 111350. [[CrossRef](#)]
17. Pyo, J.C.; Duan, H.; Ligaray, M.; Kim, M.; Baek, S.; Kwon, Y.S.; Lee, H.; Kang, T.; Kim, K.; Cha, Y.K.; et al. An integrative remote sensing application of stacked autoencoder for atmospheric correction and cyanobacteria estimation using hyperspectral imagery. *Remote Sens.* **2020**, *12*, 1073. [[CrossRef](#)]
18. Keller, S.; Maier, P.M.; Riese, F.M.; Norra, S.; Holbach, A.; Börsig, N.; Wilhelms, A.; Moldaenke, C.; Zaake, A.; Hinz, S. Hyperspectral data and machine learning for estimating CDOM, chlorophyll a, diatoms, green algae and turbidity. *Int. J. Environ. Res. Public Health* **2018**, *15*, 1881. [[CrossRef](#)]
19. Šindelář, R.; Babuška, R. Input selection for nonlinear regression models. *IEEE Trans. Fuzzy Syst.* **2004**, *12*, 688–696. [[CrossRef](#)]
20. Dekker, A.G. *Detection of Optical Water Quality Parameters for Eutrophic Waters by High Resolution Remote Sensing*; Institute for Environmental Studies: Amsterdam, The Netherlands, 1993.
21. Mishra, S.; Mishra, D.R.; Schluchter, W.M. A novel algorithm for predicting phycocyanin concentrations in cyanobacteria: A proximal hyperspectral remote sensing approach. *Remote Sens.* **2009**, *1*, 758–775. [[CrossRef](#)]
22. Simis, S.G.H.; Peters, S.W.M.; Gons, H.J. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. *Limnol. Oceanogr.* **2005**, *50*, 237–245. [[CrossRef](#)]
23. Kallio, K.; Kutser, T.; Hannonen, T.; Koponen, S.; Pulliainen, J.; Vepsäläinen, J.; Pyhälähti, T. Retrieval of water quality from airborne imaging spectrometry of various lake types in different seasons. *Sci. Total Environ.* **2001**, *268*, 59–77. [[CrossRef](#)]
24. Shafique, N.A.; Fulk, F.; Autrey, B.C.; Flotemersch, J. Water area extraction using geocoded high resolution imagery of TerraSAR-X radar satellite in cloud prone Brahmaputra River valley. *J. Geomat.* **2009**, *3*, 9–12.
25. Ogashawara, I.; Mishra, D.R.; Mishra, S.; Curtarelli, M.P.; Stech, J.L. A performance review of reflectance based algorithms for predicting phycocyanin concentrations in inland waters. *Remote Sens.* **2013**, *5*, 4774–4798. [[CrossRef](#)]
26. Schalles, J.F.; Yacobi, Y.Z. Remote detection and seasonal patterns of phycocyanin, carotenoid and chlorophyll pigments in eutrophic waters. *Ergeb. Limnol.* **2000**, *55*, 153–168.
27. Soja-Woźniak, M.; Darecki, M.; Wojtasiewicz, B.; Bradtke, K. Laboratory measurements of remote sensing reflectance of selected phytoplankton species from the Baltic Sea. *Oceanologia* **2018**, *60*, 86–96. [[CrossRef](#)]
28. Hestir, E.L.; Brando, V.E.; Bresciani, M.; Giardino, C.; Matta, E.; Villa, P.; Dekker, A.G. Measuring freshwater aquatic ecosystems: The need for a hyperspectral global mapping satellite mission. *Remote Sens. Environ.* **2015**, *167*, 181–195. [[CrossRef](#)]
29. Yan, Y.; Bao, Z.; Shao, J. Phycocyanin concentration retrieval in inland waters: A comparative review of the remote sensing techniques and algorithms. *J. Great Lakes Res.* **2018**, *44*, 748–755. [[CrossRef](#)]
30. Pyo, J.C.; Ligaray, M.; Kwon, Y.S.; Ahn, M.H.; Kim, K.; Lee, H.; Kang, T.; Cho, S.B.; Park, Y.; Cho, K.H. High-spatial resolution monitoring of phycocyanin and chlorophyll-a using airborne hyperspectral imagery. *Remote Sens.* **2018**, *10*, 1180. [[CrossRef](#)]
31. Berk, A.; Conforti, P.; Kennett, R.; Perkins, T.; Hawes, F.; Van Den Bosch, J. MODTRAN[®] 6: A major upgrade of the MODTRAN[®] radiative transfer code. In Proceedings of the Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing, Lausanne, Switzerland, 24–27 June 2014; pp. 1–4.
32. Pyo, J.; Ha, S.; Pachepsky, Y.A.; Lee, H.; Ha, R.; Nam, G.; Kim, M.S.; Im, J.; Cho, K.H. Chlorophyll-a concentration estimation using three difference bio-optical algorithms, including a correction for the low-concentration range: The case of the Yiam reservoir, Korea. *Remote Sens. Lett.* **2016**, *7*, 407–416. [[CrossRef](#)]
33. Pyo, J.C.; Pachepsky, Y.; Baek, S.S.; Kwon, Y.S.; Kim, M.J.; Lee, H.; Park, S.; Cha, Y.K.; Ha, R.; Nam, G.; et al. Optimizing semi-analytical algorithms for estimating chlorophyll-a and phycocyanin concentrations in inland waters in Korea. *Remote Sens.* **2017**, *9*, 542. [[CrossRef](#)]

34. Sarada, R.; Pillai, M.G.; Ravishankar, G.A. Phycocyanin from *Spirulina* sp.: Influence of processing of biomass on phycocyanin yield, analysis of efficacy of extraction methods and stability studies on phycocyanin. *Process Biochem.* **1999**, *34*, 795–801. [[CrossRef](#)]
35. Le Bris, A.; Chehata, N.; Briottet, X.; Pappadimitis, N. A random forest class memberships based wrapper band selection criterion: Application to hyperspectral. *Int. Geosci. Remote Sens. Symp.* **2015**, *2015*, 1112–1115. [[CrossRef](#)]
36. Jaiswal, J.K.; Samikannu, R. Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression. In Proceedings of the 2017 World Congress on Computing and Communication Technologies (WCCCT), Tiruchirappalli, India, 2–4 February 2017; pp. 65–68. [[CrossRef](#)]
37. Reinart, A.; Kutser, T. Comparison of different satellite sensors in detecting cyanobacterial bloom events in the Baltic Sea. *Remote Sens. Environ.* **2006**, *102*, 74–85. [[CrossRef](#)]
38. Menken, K.D.; Brezonik, P.L.; Bauer, M.E. Influence of chlorophyll and colored dissolved organic matter (CDOM) on lake reflectance spectra: Implications for measuring lake properties by remote sensing. *Lake Reserv. Manag.* **2006**, *22*, 179–190. [[CrossRef](#)]
39. Brezonik, P.L.; Olmanson, L.G.; Finlay, J.C.; Bauer, M.E. Factors affecting the measurement of CDOM by remote sensing of optically complex inland waters. *Remote Sens. Environ.* **2015**, *157*, 199–215. [[CrossRef](#)]
40. Ha, N.T.T.; Thao, N.T.P.; Koike, K.; Nhuan, M.T. Selecting the best band ratio to estimate chlorophyll-a concentration in a tropical freshwater lake using sentinel 2A images from a case study of Lake Ba Be (Northern Vietnam). *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 290. [[CrossRef](#)]
41. Gitelson, A.A.; Schalles, J.F.; Rundquist, D.C.; Schiebe, F.R.; Yacobi, Y.Z. Comparative reflectance properties of algal cultures with manipulated densities. *J. Appl. Phycol.* **1999**, *11*, 345–354. [[CrossRef](#)]
42. Quibell, G. The effect of suspended sediment on reflectance from freshwater algae. *Int. J. Remote Sens.* **1991**, *12*, 177–182. [[CrossRef](#)]
43. Geladi, P.; Kowalski, B.R. Partial least-squares regression: A tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17. [[CrossRef](#)]
44. Abdi, H. Partial least square regression (PLS regression). *Encycl. Res. Methods Soc. Sci.* **2003**, *6*, 792–795. [[CrossRef](#)]
45. Guo, L.; Chehata, N.; Mallet, C.; Boukir, S. Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 56–66. [[CrossRef](#)]
46. Rodriguez-Galiano, V.; Sanchez-Castillo, M.; Chica-Olmo, M.; Chica-Rivas, M. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* **2015**, *71*, 804–818. [[CrossRef](#)]
47. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [[CrossRef](#)] [[PubMed](#)]
48. Khazaee Poul, A.; Shourian, M.; Ebrahimi, H. A Comparative Study of MLR, KNN, ANN and ANFIS Models with Wavelet Transform in Monthly Stream Flow Prediction. *Water Resour. Manag.* **2019**, *33*, 2907–2923. [[CrossRef](#)]
49. Vapnik, V.; Guyon, I.; Hastie, T.; Rosset, S.; Zhu, J.; Tibshirani, R. Support vector machines. *Mach. Learn.* **1995**, *20*, 273–297.
50. Coulibaly, M.S.K.P. Application of support vector machine in power system. *Study Dyn. Syst.* **2006**, *11*, 199–205. [[CrossRef](#)]
51. Mohammadpour, R.; Shaharuddin, S.; Chang, C.K.; Zakaria, N.A.; Ghani, A.A.; Chan, N.W. Prediction of water quality index in constructed wetlands using support vector machine. *Environ. Sci. Pollut. Res.* **2015**, *22*, 6208–6219. [[CrossRef](#)]
52. Zou, J.; Han, Y.; So, S.S. Overview of artificial neural networks. *Methods Mol. Biol.* **2008**, *458*, 15–23. [[CrossRef](#)]
53. Palani, S.; Liong, S.Y.; Tkalic, P. An ANN application for water quality forecasting. *Mar. Pollut. Bull.* **2008**, *56*, 1586–1597. [[CrossRef](#)]
54. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Duchesnay, E. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
55. Santhi, C.; Arnold, J.G.; Williams, J.R.; Dugas, W.A.; Srinivasan, R.; Hauck, L.M. Validation of the swat model on a large river basin with point and nonpoint sources 1. *JAWRA J. Am. Water Resour. Assoc.* **2001**, *37*, 1169–1188. [[CrossRef](#)]
56. Moriasi, D.N.; Arnold, J.G.; Van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **2007**, *50*, 885–900. [[CrossRef](#)]
57. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [[CrossRef](#)]
58. Singh, J.; Knapp, H.V.; Arnold, J.G.; Demissie, M. Hydrological modeling of the Iroquois River watershed using HSPF and SWAT. *J. Am. Water Resour. Assoc.* **2005**, *41*, 343–360. [[CrossRef](#)]
59. Oki, K. Why is the ratio of reflectivity effective for chlorophyll estimation in the lake water? *Remote Sens.* **2010**, *2*, 1722–1730. [[CrossRef](#)]
60. Vincent, R.K.; Qin, X.; McKay, R.M.L.; Miner, J.; Czajkowski, K.; Savino, J.; Bridgeman, T. Phycocyanin detection from LANDSAT TM data for mapping cyanobacterial blooms in Lake Erie. *Remote Sens. Environ.* **2004**, *89*, 381–392. [[CrossRef](#)]
61. Yacobi, Y.Z.; Moses, W.J.; Kaganovsky, S.; Sulimani, B.; Leavitt, B.C.; Gitelson, A.A. NIR-red reflectance-based algorithms for chlorophyll-a estimation in mesotrophic inland and coastal waters: Lake Kinneret case study. *Water Res.* **2011**, *45*, 2428–2436. [[CrossRef](#)]

62. Augusto-Silva, P.B.; Ogashawara, I.; Barbosa, C.C.F.; de Carvalho, L.A.S.; Jorge, D.S.F.; Fornari, C.I.; Stech, J.L. Analysis of MERIS reflectance algorithms for estimating chlorophyll-a concentration in a Brazilian reservoir. *Remote Sens.* **2014**, *6*, 11689–11707. [[CrossRef](#)]
63. Rundquist, D.C.; Han, L.; Schalles, J.F.; Peake, J.S. Remote measurement of algal chlorophyll in surface waters: The case for the first derivative of reflectance near 690 nm. *Photogramm. Eng. Remote Sens.* **1996**, *62*, 195–200.
64. Schalles, J.F.; Rundquist, D.C.; Schiebe, F.R. The influence of suspended clays on phytoplankton reflectance signatures and the remote estimation of chlorophyll. *SIL Proc.* **2001**, *27*, 3619–3625. [[CrossRef](#)]
65. Sagan, V.; Peterson, K.T.; Maimaitijiang, M.; Sidike, P.; Sloan, J.; Greeling, B.A.; Maalouf, S.; Adams, C. Monitoring inland water quality using remote sensing: Potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. *Earth-Sci. Rev.* **2020**, *205*, 103187. [[CrossRef](#)]
66. Song, K.; Li, L.; Tedesco, L.P.; Li, S.; Hall, B.E.; Du, J. Remote quantification of phycocyanin in potable water sources through an adaptive model. *ISPRS J. Photogramm. Remote Sens.* **2014**, *95*, 68–80. [[CrossRef](#)]
67. Chang, N.-B.; Vannah, B. Intercomparisons between empirical models with data fusion techniques for monitoring water quality in a large lake. In Proceedings of the 2013 10th IEEE International Conference on Networking, Sensing and Control (ICNSC 2013), Evry, France, 10–12 April 2013; pp. 258–263. [[CrossRef](#)]
68. He, J.; Chen, Y.; Wu, J.; Stow, D.A.; Christakos, G. Space-time chlorophyll-a retrieval in optically complex waters that accounts for remote sensing and modeling uncertainties and improves remote estimation accuracy. *Water Res.* **2020**, *171*, 115403. [[CrossRef](#)]
69. Zhou, L.; Ma, W.; Zhang, H.; Li, L.; Tang, L. Developing a PCA-ANN model for predicting chlorophyll a concentration from field hyperspectral measurements in dianshan lake, China. *Water Qual. Expo. Health* **2015**, *7*, 591–602. [[CrossRef](#)]
70. Kim, S.; Chung, S.; Park, H.; Cho, Y.; Lee, H. Analysis of environmental factors associated with cyanobacterial dominance after river weir installation. *Water* **2019**, *11*, 1163. [[CrossRef](#)]
71. Konopka, A.; Brock, T.D. Effect of temperature on blue-green algae (*Cyanobacteria*) in Lake Mendota. *Appl. Environ. Microbiol.* **1978**, *36*, 572–576. [[CrossRef](#)]
72. Lüring, M.; Eshetu, F.; Faassen, E.J.; Kosten, S.; Huszar, V.L.M. Comparison of cyanobacterial and green algal growth rates at different temperatures. *Freshw. Biol.* **2013**, *58*, 552–559. [[CrossRef](#)]
73. Nalley, J.O.; O'Donnell, D.R.; Litchman, E. Temperature effects on growth rates and fatty acid content in freshwater algae and cyanobacteria. *Algal Res.* **2018**, *35*, 500–507. [[CrossRef](#)]
74. Paerl, H.W.; Paul, V.J. Climate change: Links to global expansion of harmful cyanobacteria. *Water Res.* **2012**, *46*, 1349–1363. [[CrossRef](#)]
75. Berg, M.; Sutula, M. Factors Affecting Growth of Cyanobacteria. *Monaldi Arch. Chest Dis. Pulm. Ser.* **2015**, *59*, 103–107.
76. Dev, P.J.; Sukenik, A.; Mishra, D.R.; Ostrovsky, I. Cyanobacterial pigment concentrations in inland waters: Novel semi-analytical algorithms for multi- and hyperspectral remote sensing data. *Sci. Total Environ.* **2022**, *805*, 150423. [[CrossRef](#)]
77. Park, Y.; Pyo, J.C.; Kwon, Y.S.; Cha, Y.K.; Lee, H.; Kang, T.; Cho, K.H. Evaluating physico-chemical influences on cyanobacterial blooms using hyperspectral images in inland water, Korea. *Water Res.* **2017**, *126*, 319–328. [[CrossRef](#)]
78. Ha, K.; Cho, E.A.; Kim, H.W.; Joo, G.J. Microcystis bloom formation in the lower Nakdong River, South Korea: Importance of hydrodynamics and nutrient loading. *Mar. Freshw. Res.* **1999**, *50*, 89–94. [[CrossRef](#)]
79. Oliver, R.L.; Ganf, G.G. Freshwater blooms. In *The Ecology of Cyanobacteria*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 149–194.