*Article*

# Monitoring Coastal Chlorophyll-a Concentrations in Coastal Areas Using Machine Learning Models

**Yong Sung Kwon [1], Seung Ho Baek [2], Young Kyun Lim [2,3], JongCheol Pyo [1], Mayzonee Ligaray [1], Yongeun Park [4],\* and Kyung Hwa Cho [1],\***

[1]    School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Korea; wizkys@naver.com (Y.S.K.); pjsddet1@naver.com (J.P.); mayzonee@gmail.com (M.L.)
[2]    South Sea Research Institute, Korea Institute of Ocean Science and Technology, Geoje 53201, Korea; baeksh@kiost.ac.kr (S.H.B.); limyk0913@kiost.ac.kr (Y.K.L.)
[3]    Department of Oceanography and Ocean Environmental Sciences, Chungnam National University, Daejon 34134, Korea
[4]    School of Civil and Environmental Engineering, Konkuk University, Seoul 05029, Korea
\*    Correspondence: yepark@konkuk.ac.kr (Y.P.); khcho@unist.ac.kr (K.H.C.);
       Tel.: +82-2-450-6106 (Y.P.); +82-52-217-2829 (K.H.C.)

**Abstract:** Harmful algal blooms have negatively affected the aquaculture industry and aquatic ecosystems globally. Remote sensing using satellite sensor systems has been applied on large spatial scales with high temporal resolutions for effective monitoring of harmful algal blooms in coastal waters. However, oceanic color satellites have limitations, such as low spatial resolution of sensor systems and the optical complexity of coastal waters. In this study, bands 1 to 4, obtained from Landsat-8 Operational Land Imager satellite images, were used to evaluate the performance of empirical ocean chlorophyll algorithms using machine learning techniques. Artificial neural network and support vector machine techniques were used to develop an optimal chlorophyll-a model. Four-band, four-band-ratio, and mixed reflectance datasets were tested to select the appropriate input dataset for estimating chlorophyll-a concentration using the two machine learning models. While the ocean chlorophyll algorithm application on Landsat-8 Operational Land Imager showed relatively low performance, the machine learning methods showed improved performance during both the training and validation steps. The artificial neural network and support vector machine demonstrated a similar level of prediction accuracy. Overall, the support vector machine showed slightly superior performance to that of the artificial neural network during the validation step. This study provides practical information about effective monitoring systems for coastal algal blooms.

**Keywords:** harmful algal blooms; remote sensing; Landsat-8 Operational Land Imager; machine learning

## 1. Introduction

Currently, harmful algal blooms (HABs) are among the problematic environmental issues worldwide [1,2]. The frequency and intensity of HAB events has dramatically increased globally since the 1970s [3,4]. Since that time, fish-killing events by red tides have been observed in Korea; a red tide event of *C. polykrikoides* was first recorded during the 1980s and the frequency of red tide events has gradually increased up to the 2000s [5]. HABs lead to severe damage to the aquaculture industry, resulting in shellfish and fish kills, and may even threaten human health [6–9]. Huge economic losses are caused by HABs amounting to approximately $1 billion per year in Europe and $100 million per year in the USA [10]. Economic losses suffered by Korea resulting from HABs since the 1980s amount to $121 million [11].

Ecological and biological studies have been conducted to monitor HABs and to address their critical effects [12–14]. HAB monitoring is essential for effective decision making and to develop management strategies [15]. In particular, remote-sensing techniques using multispectral optical satellite sensors have been widely used in the oceanic research field because of their extensive spatial coverage and frequent temporal resolution [16,17]. For example, ocean chlorophyll (OC) algorithms have been applied to monitor the chlorophyll-a (chl-a) concentration [18]. For the Korean coastal area, two-stage filtering using sea-surface temperature and the 667-nm band of Moderate Resolution Imaging Spectroradiometer (MODIS) images have been used to detect a *C. polykrikoides* bloom as presented in Kim, et al. [19]. Son, et al. [20] applied the normalized water leaving radiance to classify non-bloom and bloom water.

The optical complexity of coastal waters, however, might result in difficulty using satellite images because coastal and estuarine waters contain a significantly high concentration of terrigenous material and inorganic particles [21,22]. In addition, the resolution of satellite sensors has limitations in covering complex coastlines because one pixel may include both coastal water and the land surface [23]. Satellite systems such as MODIS and the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) have relatively coarse spatial resolution; thus, they may not be applicable to monitor coastal water quality. The only possible means is monitoring via shipboard field sampling [24,25].

For this study, an intensive two-year field survey in the South Sea of Korea was conducted from 2016 to 2017. We collected chl-a samples from 62 sampling stations near coastal areas. The main objective of this study was to investigate a proper method to retrieve the chl-a concentration in coastal waters. First, OC algorithms were applied to Landsat-8 Operational Land Imager (OLI) to evaluate land imager efficiency in the retrieval of the chl-a concentration. Then, an artificial neural network (ANN) and a support vector regression (SVR) model were developed using intensive field observation and remote sensing reflectance (Rrs) data to estimate the chl-a concentration. Via this study, we propose the most desirable method for monitoring chl-a in coastal areas.

## 2. Materials and Methods

### 2.1. Site Description

The research area in this study is the middle of the South Sea of Korea (Figure 1). The area extends from $34°00'$ N–$35°30'$ N and $127°00'$ E–$129°30'$ E. This area could be characterized as a complex coastline with many aquaculture farms, influenced by an oceanic current; i.e., the Tsushima Current. The Tsushima Current is a branch of the Kuroshio Current and transports high levels of nutritious warm water to the South Sea of Korea [26,27]. The first bloom event of *C. polykrikoides* was recorded during the early 1980s at a semi-enclosed bay; i.e., Jinhae Bay. Since the 1980s, many coastal areas in the South Sea of Korea have been affected by *C. polykrikoides* blooms [28–30].

### 2.2. Satellite and Field Sampling Data

We used Landsat-8 OLI, which has 11 bands in total. Bands 1 to 4 are in the visible light zone, band 5 is near infrared, bands 6 and 7 are in the shortwave infrared zone, bands 8 and 9 are the panchromatic and cirrus band, respectively, and band 10 and band 11 are in the thermal infrared zone [31]. Their wavelength widths and spatial resolutions are represented in Table 1. The research area covers two orbit pathways of Landsat-8; the path/row numbers of the research area are 115/36 for the west side and 114/36 for the east side. The Landsat-8 level 1B image data were acquired from the EarthExplorer website (https://earthexplorer.usgs.gov/) of the U. S. Geological Survey (USGS). To collect *in situ* chl-a data; biweekly field sampling was conducted from June 2016 to September 2017. Surface water samples were collected using 15-L buckets and then filtered using Whatman GF/F glass fiber filters (47 mm in diameter with a pore size of 0.45 µm) for chl-a analysis. The filter papers were stored frozen at $-20$ °C for 24 h, in a dark environment and chl-a was extracted with 90% acetone from the filter paper. The chl-a concentration was analyzed using a Turner BioSystems fluorescence analyzer

(Sunnyvale, CA, USA). Satellite images, which had the closest acquisition date to that of the field survey, were selected and processed. Reflectance values were extracted from the satellite images by each sampling station; a total of 147 data were extracted for model construction. Table 2 shows information on the general statistics of the chl-a, field sampling period, and Landsat-8 image counterparts.
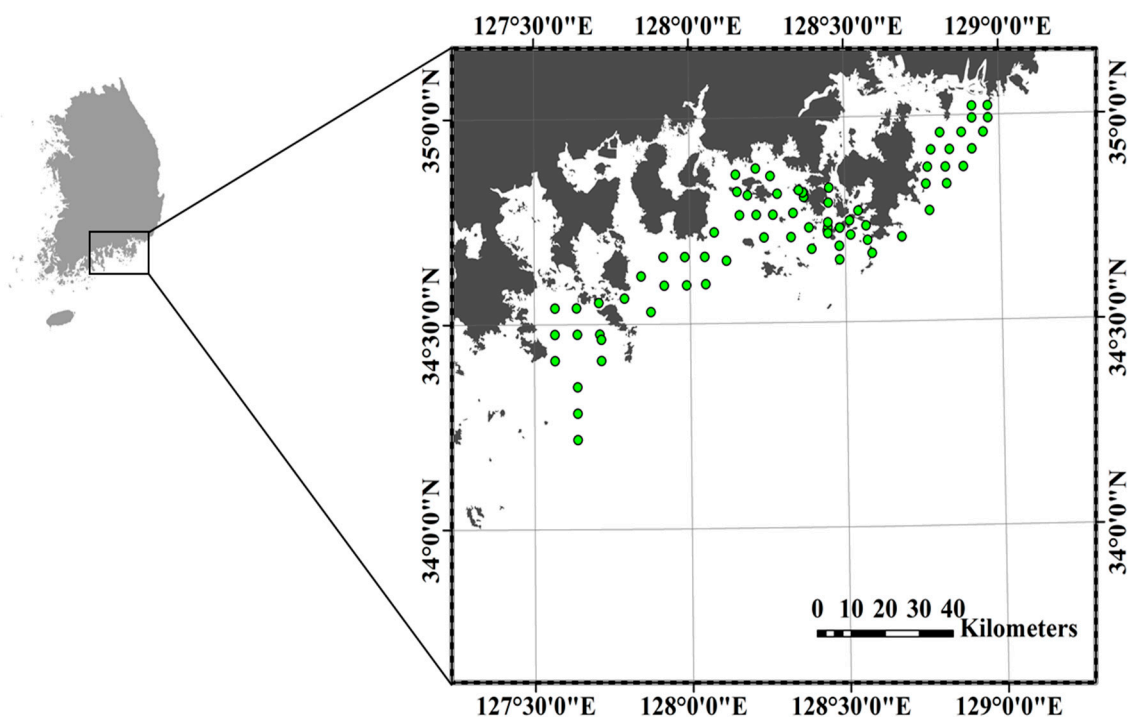


**Figure 1.** Map of research site including field sampling stations (green circles).

**Table 1.** Landsat-8 Operational Land Imager (OLI) image information.

| Band | Wavelength (nm) | | | | Spatial Resolution (m) |
|---|---|---|---|---|---|
| | **Minimum** | **Center** | **Maximum** | **Range** | |
| 1 | 433 | 443 | 453 | 20 | 30 |
| 2 | 450 | 482.5 | 515 | 65 | 30 |
| 3 | 525 | 562.5 | 600 | 75 | 30 |
| 4 | 630 | 655 | 680 | 50 | 30 |
| 5 | 845 | 865 | 885 | 40 | 30 |
| 6 | 1560 | 1610 | 1660 | 100 | 30 |
| 7 | 2100 | 2200 | 2300 | 200 | 30 |
| 8 | 500 | 590 | 680 | 180 | 15 |
| 9 | 1360 | 1375 | 1390 | 30 | 30 |

**Table 2.** Information regarding the satellite images, field sampling, and extracted band data.

| Satellite Image Data | | | Field Sampling Data | | |
|---|---|---|---|---|---|
| **Image Name** | **Date** | **Path/Low** | **Survey Period** | **Number of Data** | **Chl-a Concentration (mg m$^{-3}$)** |
| LC08_L1TP_114036_20170422_20170501_01_T1 | 22 April 2017 | 114/36 | 25–26 April 2017 | 18 | Min: 0.380 Max: 2.392 Median: 0.957 |
| LC08_L1TP_115036_20170616_20170629_01_T1 | 16 June 2017 | 115/36 | 19–21 June 2017 | 34 | Min: 0.281 Max: 3.426 Median: 1.087 |
| LC08_L1TP_114036_20170812_20170824_01_T1 | 12 August 2017 | 114/36 | 14–16 August 2017 | 18 | Min: 0.584 Max: 4.087 Median: 1.695 |
| LC08_L1TP_114036_20170913_20170928_01_T1 | 13 September 2017 | 114/36 | 12–13 September 2017 | 23 | Min: 1.442 Max: 6.191 Median: 3.931 |

### 2.3. Atmospheric Correction of Satellite Data

The satellite images were first radiometrically corrected to convert digital numbers (DNs) to top-of-atmosphere (TOA) radiance, after which Rrs was calculated. The atmospheric correction for the Landsat-8 OLI images was performed using the case-2 regional coast color (C2RCC) algorithm with the Sentinel Application Platform (SNAP) software, which is accessible via the European Space Agency (ESA) website (http://step.esa.int/main/toolboxes/snap/) [32]. C2RCC is an atmospheric correction algorithm, based on a neural network, for coastal and turbid inland water bodies with complex optical properties. C2RCC uses a database of simulated water leaving reflectance and TOA radiance. Training of networks aimed toward inversion of spectra for the atmospheric correction [33].

### 2.4. Model for Predicting Chl-a Concentration

In this study, we analyzed the performance of two types of models (i.e., OC algorithms and machine learning algorithms) for quantifying chl-a concentration. The OC and machine learning algorithms were constructed using Landsat-8 images to estimate the chl-a concentration. The ANN and SVR were constructed using MATLAB version R2017a (The MathWorks Inc., Natick, MA, USA). We used bands 1 to 4 and their ratios as input data. Three types of input datasets were used for the ANN and SVR models; surface Rrs values from bands 1 to band 4, a four-band-ratio (band 1/band 3 or 4 and band 2/band 3 or 4), and a mixed dataset containing both the four bands and four-band-ratio.

#### 2.4.1. OC Algorithm

The OC algorithms were developed by O'Reilly, Maritorena, Mitchell, Siegel, Carder, Garver, Kahru and McClain [18] using the data collected during the SeaWiFS Bio-optical Algorithm Mini-Workshop (SeaBAM), which had finalizing the operational SeaWiFS chl-a algorithm as a goal. In this study, we used the five OC algorithms to quantify the chl-a concentration using the monitoring data collected from the South Sea of Korea. The OC algorithm equations are represented as follows:

$$\text{OC1a, Power, Chl-a } = 10^{a0 + a1 \times R} \tag{1}$$

$$\text{OC1b, Geometric, Chl-a } = 10^{a0 + a1 \times R} + a2 \tag{2}$$

$$\text{OC1c, Quadratic polynomial, Chl-a } = 10^{a0 + a1 \times R + a2 \times R^2} \tag{3}$$

$$\text{OC1d, Cubic polynomial, Chl-a } = 10^{a0 + a1 \times R + a2 \times R^2 + a3 \times R^3} \tag{4}$$

$$\text{OC2b, OC2 and OC3d, Modified cubic polynomial (MCP), Chl-a } = 10^{a0 + a1 \times R + a2 \times R^2 + a3 \times R^3} + a4 \tag{5}$$

where R represents the log of the band ratio and a0, a1, a2, a3, and a4 are the coefficients for each algorithm. OC1a, OC1b, OC1c, and OC1d commonly use R as the log(Rrs490/Rrs555). OC2b, OC2, and OC3d use Equation (5) where the R values are log(Rrs443/Rrs555), log(Rrs490/Rrs555), and log (Rrs443 > Rrs490/Rrs555), respectively. In this study, 483-nm and 563-nm bands were used after tuning with vicarious calibration gains instead of 490-nm and 555-nm bands [34].

#### 2.4.2. ANN Model

The ANN model is among the widely applied tools for modeling complex environmental processes and water quality modeling [35–41]. The ANN model is constructed via three layers (i.e., the input, hidden, and output layers) that are composed of nodes. All layers in ANN are connected by weight and bias. The number of hidden nodes and hidden layers are very important model parameters because too many layers and nodes may result in over-fitting of the ANN model [42]. The input layer consists of input band datasets, and the output layer has a target variable, the chl-a concentration. A hidden layer forms an internal structure of the neural network as the input data pass through [43]. The data, starting from an input layer, are transferred to a hidden layer, finally reaching an output layer. Before the nodal data are transferred to the next layer, the data are multiplied and added by

weight and bias. The structure of ANN can be represented by a generalized mathematical expression as in Equations (6) and (7) [44] as follows:

$$H_q^i = f_1\left(\sum_{p=1}^{N} w_{pq}^h x_p^i + b_1^q\right) \tag{6}$$

$$g_1^i = f_2\left(\sum_{q=1}^{L} w_{q1}^0 H_q^i + b_1^2\right) \tag{7}$$

where $H_q^i$ is the hidden layer output, $g_1^i$ is the neural network output, $x_p^i$ is the pth element of the ith input variable, $w_{pq}^h$ is the weight of the connection between the pth node of the input layer and the qth node of the hidden layer, and $w_{q1}^0$ is the weight between the qth node of the hidden layer and the output layer node. Term b is the bias term, and p and q represent the node number in the input and hidden layers. $f_1$ is the activation function for the input vector and $f_2$ is the output function calculating the scalar output. During a training process, the initial weights are randomly provided. The Mean squared error (MSE) is calculated between the final output and the observed target variable after the signal passes through the structure. After training, backpropagation is applied to update the weights via backward signal transferring in which the Levenberg–Marquardt algorithm was used [45]. This algorithm is fast and stable in convergence to minimize a non-linear function [46]. There are three activation functions for developing ANN in MATLAB. The activation functions include tangent-sigmoid, log-sigmoid, and linear functions. A trial-and-error method was applied to select the proper activation functions; the tangent-sigmoid function was finally chosen for this study. Other parameters in ANN include learning rate and momentum constant. Parameters such as learning rate, momentum constant, and the number of layer and nodes in a layer affects ANN performance and should be optimized [47]. The learning rate, momentum constant, and the number of nodes in the hidden layer were optimized using a pattern-search algorithm in the MATLAB toolbox.

### 2.4.3. SVR Model

The support vector model (SVM) is a useful tool for pattern recognition and non-linear regression; major applications include classification, regression, and time series prediction [39,48,49]. SVR is the regression version of an SVM and was used for this study [50]. SVR is generally used for regression of continuous variables. The basic theory of SVR can be represented by a mathematical equation on the network output ($s_i$) as follows:

$$s(X_i) = \left(\sum_{i=1}^{T} w_i \varphi(X_i) + b\right) \tag{8}$$

where $w_i$ and b are the coefficients that are determined by minimizing the error between the network output and the target variable. $\varphi(X_i)$ is a nonlinear mapping function. To simplify calculation of the nonlinear mapping function, a kernel function, $\kappa(X_i, X)$, is applied. Before developing the SVR model, three kernel functions were compared to maximize the performance of the SVR models. The kernel functions include linear, polynomial, and Gaussian radius bias kernel functions. All three kernel functions were pre-applied to choose an optimal kernel for the SVR model. Finally, the Gaussian radius bias kernel was selected for the study. Additionally, three parameters, box constraint, epsilon, and sigma, were used to construct the SVR model in MATLAB. A box constraint leads to a strict separation of the data by applying a cost to the error. Epsilon is a complexity factor, adjusting the number of support vectors. Sigma is a scale parameter that is relevant to model stability. For the SVR model, these three parameters were determined via pattern-search optimization.

### *2.5. Evaluation of Model Performance*

Cross-validation was used to evaluate the overall performance of the models. 20% of the data were used for validation and the remaining 80% represents the training data. The accuracy of each

model was mainly evaluated via the coefficient of determination ($R^2$) and the root mean squared error (RMSE). The RMSE value can be calculated by the following equation:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \tag{9}$$

where, n is the number of data, i shows $i^{th}$ chl-a observation, and $e_i$ is the residual between the observed and estimated chl-a [51].

## 3. Results and Discussion

### 3.1. Atmospherically Corrected Rrs Spectra

In Figure 2 atmospherically corrected Rrs spectra are described. For a high chl-a concentration (>2.5 mg m$^{-3}$), low Rrs was observed to be dominant in bands 1 and 2. For medium concentrations (1 to 2.5 mg m$^{-3}$), mid-range values of Rrs spectra with large variations were observed. The spectral shapes were similar to those of previous studies that used ocean color satellites [21,52–55]. The spectral features of the oceanic chl-a concentration have also been reported in other studies. Ahn and Shanmugam [21] reported low water-leaving radiance in the 412–510-nm wavelength range for red tide waters. A study of the bio-optical properties of the Antarctic Peninsula waters showed that a high concentration of chlorophyll retrieves low blue band Rrs and lowers the blue-green ratio [55]. Bricaud, Morel, Babin, Allali and Claustre [54] also showed that a higher chl-a concentration resulted in a lower reflectance in the blue wavelength.
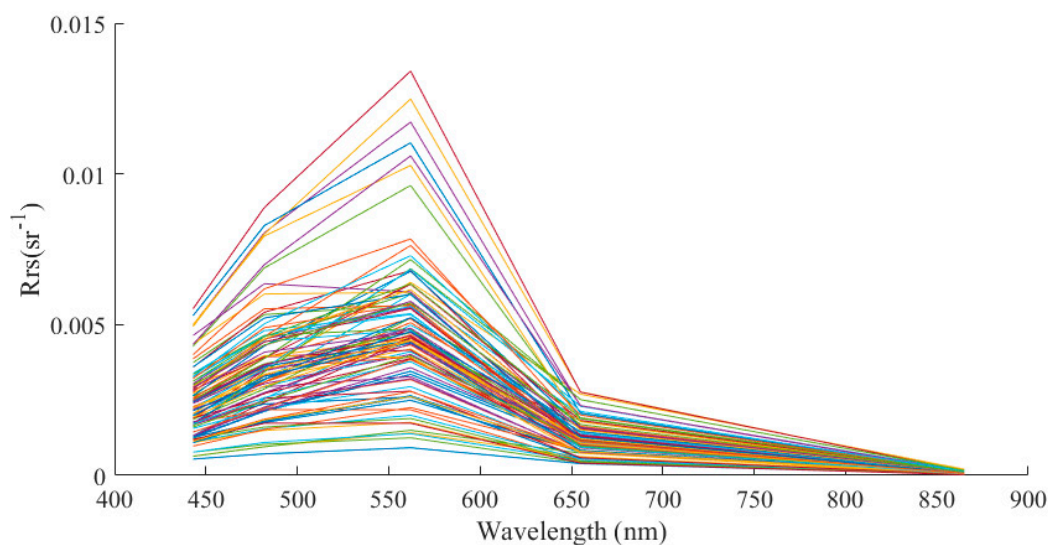


**Figure 2.** Rrs spectra. Each line represents Rrs spectrum from each station.

### 3.2. Retrieval Results Using the OC Algorithms

The OC algorithms were applied using Landsat-8 reflectance data to quantify chl-a concentrations. The performance parameters of the OC algorithms are listed in Table 3. OC1a and OC1b showed similar $R^2$ values of 0.2972 and 0.2957, respectively. OC1c and OC1d also showed similar performance, with $R^2$ values of 0.2992 and 0.2930, respectively. Among the OC algorithms, OC1c showed the best performance. The performances of OC2b and OC2 were poor as their respective $R^2$ values were 0.0194 and 0.0620. OC3d yielded an $R^2$ of 0.2960, similar to that of the OC1 group. Compared to the other OC algorithms, the OC2 group showed poor estimation performance. The scatter plots of in situ and estimated chl-a concentrations for the OC algorithms are shown in Figure 3. For all ranges, the OC algorithm tended to underestimate chl-a concentrations, except for the OC2 group. As shown in

Figure 3e,f, the OC2 group showed overestimation at low chl-a concentrations (<3 mg m$^{-3}$). The OC1 group and OC3d showed comparable estimations of chl-a, while OC2 and OC2b made it difficult to ascertain significant correlations between the estimated and observed chl-a; all of the estimated chl-a concentrations were less than 2.5 mg m$^{-3}$.
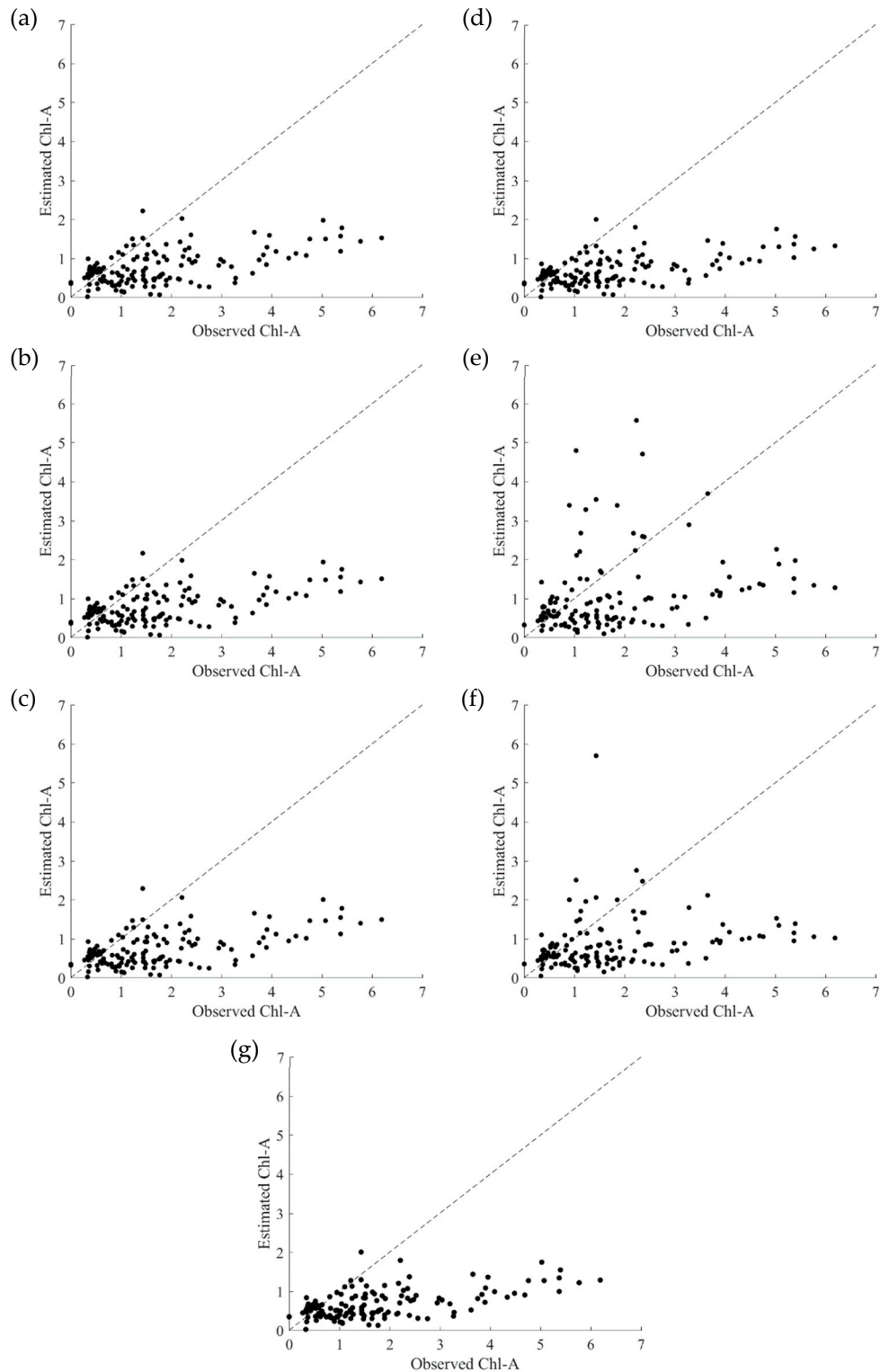


**Figure 3.** Scatter plots of in situ and estimated chl-a for OC algorithms. Used algorithms: (**a**) OC1a, (**b**) OC1b, (**c**) OC1c, (**d**) OC1d, (**e**) OC2, (**f**) OC2b, and (**g**) OC3d.

**Table 3.** Ocean Chlorophyll (OC) algorithms and performance of the Landsat-8 images.

| OC Algorithm | Coefficient | | | | | $R^2$ |
| --- | --- | --- | --- | --- | --- | --- |
| | a0 | a1 | a2 | a3 | a4 | |
| OC1a | 0.3734 | −2.4529 | - | - | - | 0.2972 |
| OC1b | 0.3636 | −2.3500 | −0.0100 | - | - | 0.2957 |
| OC1c | 0.3920 | −2.8500 | 0.6580 | - | - | 0.2992 |
| OC1d | 0.3335 | −2.9164 | 2.4686 | −2.5195 | - | 0.2930 |
| OC2b | 0.1909 | −1.9961 | 1.3020 | −0.5091 | −0.0815 | 0.0194 |
| OC2 | 0.3410 | −3.0010 | 2.8110 | −2.0410 | −0.0400 | 0.0620 |
| OC3d | 0.3483 | −2.9959 | 2.9873 | −1.4813 | −0.0597 | 0.2960 |

OC algorithms have shown high performance in ocean chl-a studies using oceanic color sensors such as SeaWiFS, MODIS-Aqua, etc. [56–58]. In this study, OC algorithm applications using Landsat-8 images did not show good performance as compared to previous studies using oceanic color sensors [56–58]. This result can be attributed to atmospheric correction using ACOLITE. Overestimation of the green wavelength (563 nm) can be observed by calculating the low aerosol contribution [59]. Because the OC algorithms use the 563-nm band as a denominator of the band ratio, a relatively high magnitude of green band can cause underestimation of the chl-a concentration. The spectral resolution of Landsat-8 is lower than that of other satellite sensors typically used in oceanic studies. While bands 2, 3, and 4 of the Landsat-8 images have 65-nm, 75-nm, and 50-nm ranges of band width, respectively, MODIS has a 10- to 20-nm range in width through the 443 nm to 555 nm bands. Visible Infrared Imaging Radiometer Suite (VIIRS) also has an approximately 20-nm bandwidth range [60]. SeaWiFS, commonly used for developing OC algorithms, has a 20-nm range [61]. A broader bandwidth might decrease the sensitivity of surface reflectance in terms of the chl-a concentration [62].

*3.3. Determination of Optimized Model Parameters*

The parameters optimized for the ANN and SVR models are shown in Table 4. The learning rate and momentum constant for the four-band dataset was 0.5000 and 0.5625, respectively, and the number of hidden layer nodes was 6. For the four-band-ratio dataset, the optimized learning rate was 0.1250 and the momentum constant was 0 and the hidden layer node number was 7. The ANN model, constructed using a mixed dataset, had a learning rate of 0.4980, a momentum rate of 0.9990, and 4 hidden layers, which are similar to that of the four-band dataset. The epsilon, kernel scale, and box constraint were optimized, as described in Table 4. The epsilon values were 0.0583, 0.0505, and 0.0999 for the four-band, four-band-ratio, and mixed datasets, respectively.

The kernel scale was 2.0005 for the four-band dataset, 500.0005 for the four-band-ratio dataset, and 2.9848 for the mixed dataset. The four-band-ratio showed the highest kernel scale value. The box constraint for the four-band dataset was 206.5474 and was the lowest. The four-band-ratio and mixed datasets showed similar values, 533.9840 and 511.2826, respectively.

**Table 4.** Best optimized parameters for the artificial neural network (ANN) and support vector model (SVR) models.

| Models | Parameter | Four-Band | Four-Band-Ratio | Mixed Dataset |
| --- | --- | --- | --- | --- |
| ANN | Learning rate | 0.5000 | 0.1250 | 0.4980 |
| | Momentum constant | 0.5625 | 0 | 0.9990 |
| | Number of hidden layer nodes | 6 | 7 | 4 |
| SVR | Epsilon | 0.0583 | 0.0505 | 0.0999 |
| | Kernel scale | 2.0005 | 500.0005 | 2.9848 |
| | Box constraint | 206.5474 | 533.9840 | 511.2826 |

### 3.4. Retrieval Results Using the Machine Learning Algorithms

The performance of the optimized ANN and SVR models is shown in Table 5. The $R^2$ and RMSE values for both the training and validation sets were used to compare the performances of the ANN and SVR models. Both the ANN and SVR models had three different models in terms of the three different inputs: four-band, four-band-ratio, and mixed datasets. The overall performances of the SVR models were better than those of the ANN models. In the case of the ANN model, the $R^2$ values for the training and validation steps of the four-band dataset were 0.4368 and 0.6322 and the RMSE values were 1.0444 mg m$^{-3}$ and 1.2187 mg m$^{-3}$, respectively. The $R^2$ values of the four-band-ratio dataset were 0.6663 and 0.3886 while the RMSE values were 0.8626 mg m$^{-3}$ and 1.3619 mg m$^{-3}$ for the training and validation steps, respectively. For the mixed dataset, the $R^2$ and RMSE values for the training and validation steps were 0.6621 and 0.2199, and 0.8713 mg m$^{-3}$ and 1.5943 mg m$^{-3}$, respectively. The four-band-ratio dataset showed the highest $R^2$ values for the training step and the four-band dataset was the highest for the validation step. However, the training $R^2$ value of the four-band dataset was less than the validation. In terms of RMSE, the four-band and four-band-ratio datasets showed similar overall performance but the difference between the training and validation was lower in the four-band dataset.

**Table 5.** Model performance of the ANN and SVR models.

| Dataset | Performance | Four-Band | | Four-Band-Ratio | | Mixed Dataset | |
|---|---|---|---|---|---|---|---|
| | | ANN | SVR | ANN | SVR | ANN | SVR |
| Training | $R^2$ | 0.4368 | 0.7119 | 0.6663 | 0.0082 | 0.6621 | 0.6948 |
| | RMSE | 1.0444 | 0.7442 | 0.8626 | 1.5337 | 0.8713 | 0.8294 |
| Validation | $R^2$ | 0.6322 | 0.7648 | 0.3886 | 0.0056 | 0.2199 | 0.6263 |
| | RMSE | 1.2187 | 0.9633 | 1.3619 | 1.5849 | 1.5943 | 0.9933 |

The scatter plots for the ANN models are shown in Figure 4. The four-band model overestimated the chl-a concentration in the low chl-a region (<2 mg m$^{-3}$) and underestimated mid-to-high concentrations (>3 mg m$^{-3}$) during both the training and validation steps. In the case of the four-band-ratio model, the under- and over-estimation trend was weak during the training step and showed a relatively high $R^2$ value. However, during the validation step a strong overestimation trend was apparent at low concentrations (<2 mg m$^{-3}$). In the mixed-dataset model, a high concentration of 6 mg m$^{-3}$ was well estimated compared to that of the other datasets.

Overall, the optimized SVR models developed using the four-band and mixed datasets showed better performance than the optimized ANN models. The estimation trends were similar to the ANN models, underestimation at high concentrations (>3 mg m$^{-3}$) and overestimation at low concentrations (<2 mg m$^{-3}$). For the four-band model, the $R^2$ values were 0.7119 and 0.7648 for the training and validation, respectively. However, the $R^2$ values during the validation step were higher than those during the training step for both models. The RMSE values were also the best for training and validation at 0.7442 mg m$^{-3}$ and 0.9633 mg m$^{-3}$, respectively. The four-band-ratio dataset showed the worst performance in terms of both $R^2$ and RMSE; 0.0082 and 0.0056 for $R^2$ and 1.5337 mg m$^{-3}$ and 1.5849 mg m$^{-3}$ for RMSE for the training and validation steps, respectively. Repeated training was conducted, however, the SVR model using the four-band-ratio dataset was not well-trained. The $R^2$ and RMSE values of the mixed dataset were 0.6948 and 0.8294 mg m$^{-3}$ for the training step and 0.6948 and 0.9933 mg m$^{-3}$ for the validation step, respectively. Except for the four-band ratio, the performances of the SVR model in both training and validation were superior to those of the ANN model. Figure 5 shows the scatter plots of the SVR results using each dataset. Overestimation was dominant at low concentrations (<2 mg m$^{-3}$) and underestimation was prevalent at medium and high concentrations (>3 mg m$^{-3}$). Table 5 shows that the SVR model is superior to the ANN model in terms of the $R^2$ and RMSE values. In addition, the SVR model results in a lesser performance difference between the training and validation steps. Dzwonkowski and Yan [63] used five reflectance bands between 443 nm and 670 nm obtained from SeaWiFS and a neural network to estimate chl-a

in coastal waters. Vilas, et al. [64] developed three different ANN models using Medium Resolution Imaging Spectrometer (MERIS) images. For SVRs, there are few research studies of oceanic chl-a. Zhan, et al. [65] used a SeaBAM dataset, used for developing OC algorithms, to estimate chl-a using an SVR model. The performance range was approximately from 0.5 to 0.9. As described in Section 3.1, Landsat-8 OLI, which has a broad spectral resolution, has poor ability in discriminating pigments. This implies that machine learning techniques can overcome the weak points of the empirical equations. In particular, the underestimation that results from the broadband width was efficiently fixed. In terms of a simple performance of chl-a estimation, the model developed in this study can be considered to have low accuracy. For coastal and estuary waters, however, a high spatial resolution is powerful. The latest studies noted that a land imager such as Sentinel-2 and Landsat-8 can retrieve reliable results from these bodies of water [52,66,67]. In recent, Automatic Model Selection Algorithm (AMSA) to determine the best model was developed for several datasets from MERIS and MODIS-Aqua by Blix and Eltoft [68]. SVR and Gaussian Process Regression (GPR) was selected as the best model and the r-squared performance was between 0.75 and 0.96. The determining of the best model structure and dataset is similar to our study. However, the characteristic of the dataset is different. In this study, four band Rrs dataset had been used to create four-band-ratio and mixed dataset. This can be used for determining the best dataset from a raw Rrs dataset.
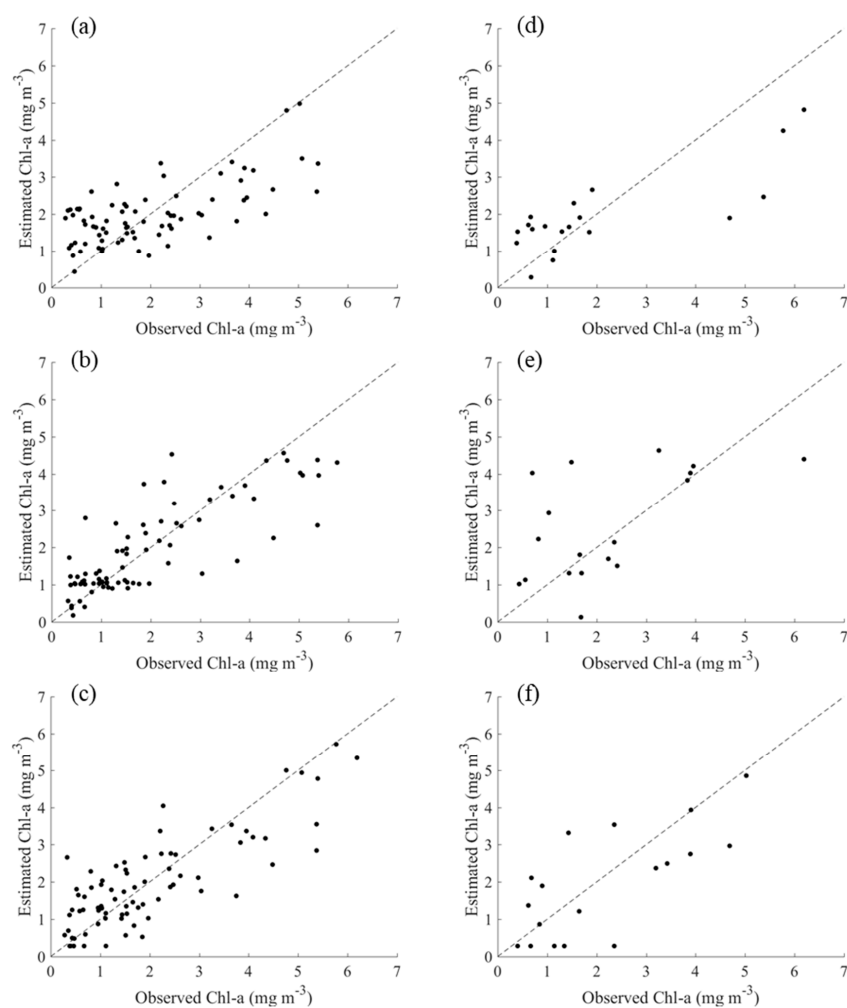


**Figure 4.** Scatter plots of in situ and estimated chl-a for the ANN model. (**a**) the training results for the four-band, (**d**) the validation results for the four-band, (**b**) the training results for the four-band-ratio, (**e**) the validation results for the four-band-ratio, (**c**) the training results for mixed dataset (**f**) the validation results for mixed dataset.
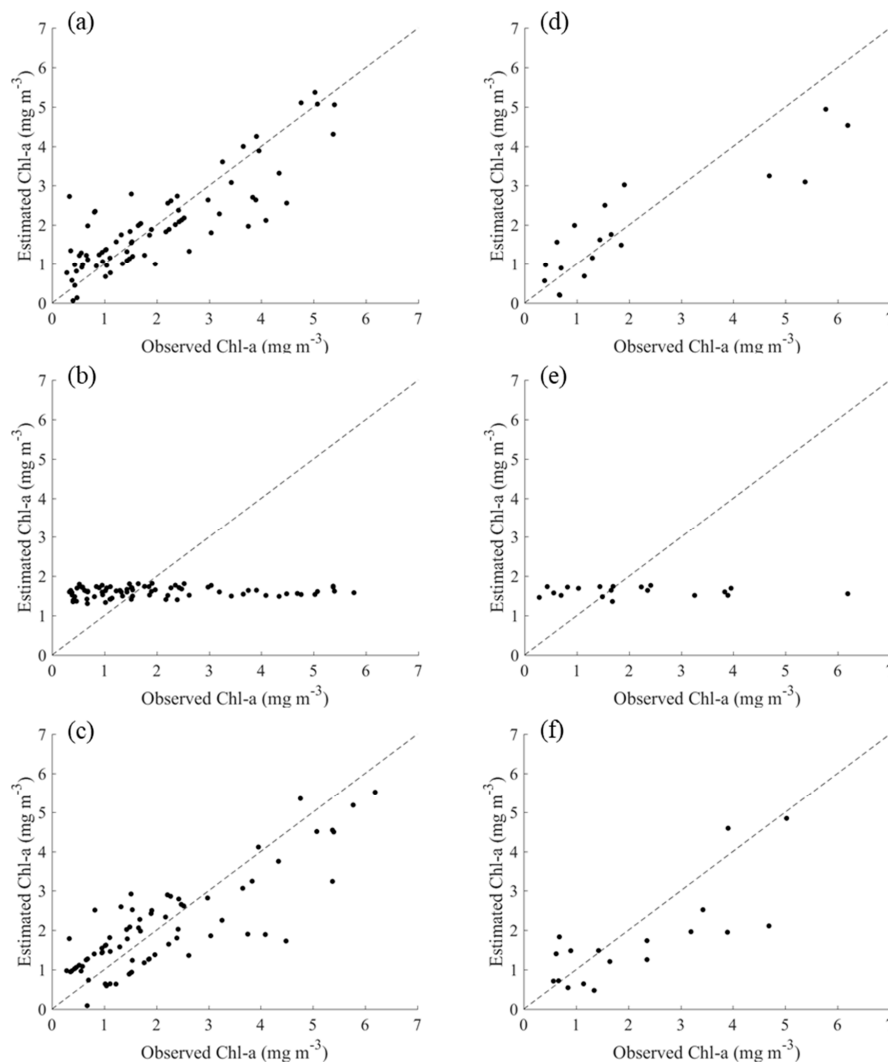
**Figure 5.** Scatter plots of in situ and estimated chl-a for the SVR model using the Gaussian radius bias kernel. (**a**) the training results for the four-band, (**d**) the validation results for the four-band, (**b**) the training results for the four-band-ratio, (**e**) the validation results for the four-band-ratio, (**c**) the training results for mixed dataset (**f**) the validation results for mixed dataset.

In addition, distribution maps of chl-a, created using machine learning models, are shown in Figure 6. From the raw Landsat-8 image, the ANN and SVR models were applied and chl-a maps were developed. For the raw images, a total of three images, 3 August, 12 August, and 13 September 2017, were used. To avoid an overfitting problem, the models that had the lowest difference between the training and validation performance was selected. The four-band-ratio and four-band datasets were chosen for the ANN and SVR models, respectively. In the ANN model case, for (a), (b) and (c), the trends in the coastal areas have a relatively high chl-a concentration and open ocean has under 2 mg m$^{-3}$ appeared, generally. However, the four-band SVR model showed poor distribution performance. There was no remarkable difference between the coastal and open ocean areas. In the case of 13 September 2017, a high concentration distribution appeared but there was also more than 4 mg m$^{-3}$ in the open ocean area.
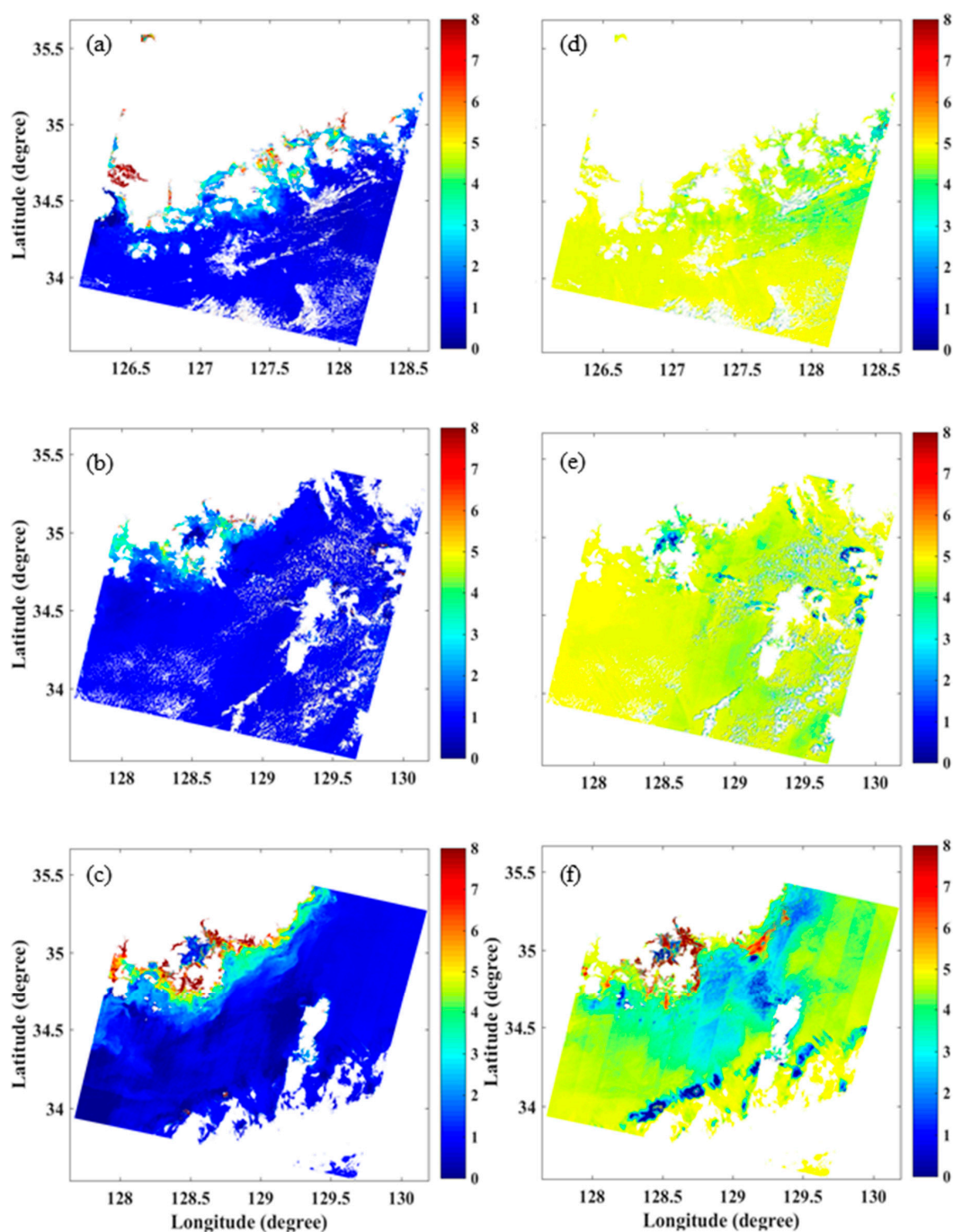
**Figure 6.** Chl-a distribution map developed using machine learning models and the Landsat-8 OLI images of 3 August, 12 August, and 13 September 2017. (**a**–**c**) show the four-band-ratio ANN results and (**d**–**f**) show the four-band SVR model maps.

## 4. Conclusions

The purpose of this research was to evaluate the performance of OC algorithms and develop optimal ANN and SVR models for chl-a estimation in coastal waters. The major findings are as follows:

1.　Seven OC algorithms were evaluated after applying various calibration gains. All OC algorithms showed poor performance using Landsat-8 satellite data. OC1c showed the best performance among the OC algorithms with an $R^2$ of 0.2992. The $R^2$ values of OC2 and OC2b were less than 0.1.

2. The ANN and SVR models showed better estimation performance than that of the OC algorithms. Compared to previous studies using oceanic color sensors, the machine-learning techniques using Landsat-8 images showed satisfactory performance.

3. The SVR model showed slightly better results than those of the ANN model during the training and validation steps. The four-band-ratio dataset SVR is not appropriate for chl-a estimation. However, the ANN model generated a more reasonable and reliable distribution of chl-a as compared to the raw image.

This study demonstrates that Landsat-8 OLI satellite data can potentially be used in coastal and oceanic research for remote sensing of HABs and that machine-learning techniques are efficient and useful tools to estimate chl-a concentration using reflectance data obtained by Landsat-8 OLI satellites from complex coastal waters.

**Author Contributions:** All authors contributed in conceiving and designing of the approach and realizing this manuscript. Conceptualization, K.H.C., Y.P. and Y.S.K.; Methodology, Y.S.K., J.P., S.H.B., Y.K.L. and Y.P.; Validation, K.H.C., Y.P., M.L., J.P. and Y.S.K.; Formal Analysis, K.H.C. and Y.S.K.; Data Curation, S.H.B. and Y.K.L.; Writing-Review & Editing, M.L., K.H.C. and Y.P.; Supervision, Kyong Hwa Cho and Y.P.

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

## References

1. Anderson, D.M. Toxic algal blooms and red tides: A global perspective. In *Red Tides: Biology, Environmental Science and Toxicology*; Elsevier: Amsterdam, The Netherlands, 1989; pp. 11–16.

2. Berdalet, E.; Fleming, L.E.; Gowen, R.; Davidson, K.; Hess, P.; Backer, L.C.; Moore, S.K.; Hoagland, P.; Enevoldsen, H. Marine harmful algal blooms, human health and wellbeing: Challenges and opportunities in the 21st century. *J. Mar. Biol. Assoc. UK* **2016**, *96*, 61–91. [CrossRef] [PubMed]

3. Glibert, P.M.; Anderson, D.M.; Gentien, P.; Granéli, E.; Sellner, K.G. The global, complex phenomena of harmful algal blooms. *Oceanography* **2005**, *18*, 136–147. [CrossRef]

4. Hallegraeff, G.M. A review of harmful algal blooms and their apparent global increase. *Phycologia* **1993**, *32*, 79–99. [CrossRef]

5. Lee, C.K.; Park, T.G.; Park, Y.T.; Lim, W.A. Monitoring and trends in harmful algal blooms and red tides in korean coastal waters, with emphasis on cochlodinium polykrikoides. *Harmful Algae* **2013**, *30*, S3–S14. [CrossRef]

6. Moore, S.K.; Trainer, V.L.; Mantua, N.J.; Parker, M.S.; Laws, E.A.; Backer, L.C.; Fleming, L.E. Impacts of climate variability and future climate change on harmful algal blooms and human health. *Environ. Health* **2008**, *7*, S4. [CrossRef] [PubMed]

7. Heil, C.A.; Glibert, P.M.; Al-Sarawi, M.A.; Faraj, M.; Behbehani, M.; Husain, M. First record of a fish-killing gymnodinium sp bloom in kuwait bay, arabian sea: Chronology and potential causes. *Mar. Ecol. Prog. Ser.* **2001**, *214*, 15–23. [CrossRef]

8. Glibert, P.M.; Landsberg, J.H.; Evans, J.J.; Al-Sarawi, M.A.; Faraj, M.; Al-Jarallah, M.A.; Haywood, A.; Ibrahem, S.; Klesius, P.; Powell, C. A fish kill of massive proportion in kuwait bay, arabian gulf, 2001: The roles of bacterial disease, harmful algae, and eutrophication. *Harmful Algae* **2002**, *1*, 215–231. [CrossRef]

9. Shumway, S.E. A review of the effects of algal blooms on shellfish and aquaculture. *J. World Aquac. Soc.* **1990**, *21*, 65–104. [CrossRef]

10. Hoagland, P.; Anderson, D.M.; Kaoru, Y.; White, A.W. The economic effects of harmful algal blooms in the united states: Estimates, assessment issues, and information needs. *Estuaries* **2002**, *25*, 819–837. [CrossRef]

11. Park, T.G.; Lim, W.A.; Park, Y.T.; Lee, C.K.; Jeong, H.J. Economic impact, management and mitigation of red tides in Korea. *Harmful Algae* **2013**, *30*, S131–S143. [CrossRef]

12. Jeong, H.J.; Yoo, Y.D.; Kim, J.S.; Seong, K.A.; Kang, N.S.; Kim, T.H. Growth, feeding and ecological roles of the mixotrophic and heterotrophic dinoflagellates in marine planktonic food webs. *Ocean Sci. J.* **2010**, *45*, 65–91. [CrossRef]

13. Jeong, H.J. The ecological roles of heterotrophic dinoflagellates in marine planktonic community. *J. Eukaryot. Microbiol.* **1999**, *46*, 390–396. [CrossRef]

14. Qi, Y.; Hong, Y.; Lu, S.; Chu, J.; Zhang, J.; Zhu, C.; Li, Y.; Liang, S.; Li, J. Outbreak of chattonella marina red tide and its relations to ecological parameters in dapeng bay, South China Sea. *Oceanol. Limnol. Sin.* **1994**, *25*, 132–137. (In Chinese)

15. Hu, C.; Murch, B.; Corcoran, A.A.; Zheng, L.; Barnes, B.B.; Weisberg, R.H.; Atwood, K.; Lenes, J.M. Developing a smart semantic web with linked data and models for near-real-time monitoring of red tides in the eastern Gulf of Mexico. *IEEE Syst. J.* **2016**, *10*, 1282–1290. [CrossRef]

16. Stumpf, R.P. Applications of satellite ocean color sensors for monitoring and predicting harmful algal blooms. *Hum. Ecol. Risk Assess. Int. J.* **2001**, *7*, 1363–1368. [CrossRef]

17. Carvalho, G.A.; Minnett, P.J.; Fleming, L.E.; Banzon, V.F.; Baringer, W. Satellite remote sensing of harmful algal blooms: A new multi-algorithm method for detecting the florida red tide (*Karenia brevis*). *Harmful Algae* **2010**, *9*, 440–448. [CrossRef] [PubMed]

18. O'Reilly, J.E.; Maritorena, S.; Mitchell, B.G.; Siegel, D.A.; Carder, K.L.; Garver, S.A.; Kahru, M.; McClain, C. Ocean color chlorophyll algorithms for seawifs. *J. Geophys. Res.* **1998**, *103*, 24937–24953. [CrossRef]

19. Kim, Y.; Byun, Y.; Kim, Y.; Eo, Y. Detection of cochlodinium polykrikoides red tide based on two-stage filtering using modis data. *Desalination* **2009**, *249*, 1171–1179. [CrossRef]

20. Son, Y.B.; Ishizaka, J.; Jeong, J.-C.; Kim, H.-C.; Lee, T. Cochlodinium polykrikoides red tide detection in the south sea of korea using spectral classification of modis data. *Ocean Sci. J.* **2011**, *46*, 239–263. [CrossRef]

21. Ahn, Y.-H.; Shanmugam, P. Detecting the red tide algal blooms from satellite ocean color observations in optically complex Northeast-Asia coastal waters. *Remote Sens. Environ.* **2006**, *103*, 419–437. [CrossRef]

22. Woodruff, D.L.; Stumpf, R.P.; Scope, J.A.; Paerl, H.W. Remote estimation of water clarity in optically complex estuarine waters. *Remote Sens. Environ.* **1999**, *68*, 41–52. [CrossRef]

23. Wynne, T.T.; Stumpf, R.P.; Tomlinson, M.C.; Ransibrahmanakul, V.; Villareal, T.A. Detecting karenia brevis blooms and algal resuspension in the western gulf of mexico with satellite ocean color imagery. *Harmful Algae* **2005**, *4*, 992–1003. [CrossRef]

24. Gregg, W.W.; Conkright, M.E. Global seasonal climatologies of ocean chlorophyll: Blending in situ and satellite data for the coastal zone color scanner era. *J. Geophys. Res.* **2001**, *106*, 2499–2515. [CrossRef]

25. Boss, E.; Swift, D.; Taylor, L.; Brickley, P.; Zaneveld, R.; Riser, S.; Perry, M.; Strutton, P. Observations of pigment and particle distributions in the western north atlantic from an autonomous float and ocean color satellite. *Limnol. Oceanogr.* **2008**, *53*, 2112–2122. [CrossRef]

26. Onitsuka, G.; Yanagi, T.; Yoon, J.H. A numerical study on nutrient sources in the surface layer of the Japan Sea using a coupled physical-ecosystem model. *J. Geophys. Res.* **2007**, *112*, C05042. [CrossRef]

27. Morimoto, A.; Takikawa, T.; Onitsuka, G.; Watanabe, A.; Moku, M.; Yanagi, T. Seasonal variation of horizontal material transport through the eastern channel of the Tsushima Straits. *J. Oceanogr.* **2009**, *65*, 61–71. [CrossRef]

28. Kim, H.G. Recent harmful algal blooms and mitigation strategies in Korea. *Oceanogr. Lit. Rev.* **1998**, *7*, 1163–1164.

29. Jeong, H.J.; Du Yoo, Y.; Lee, K.H.; Kim, T.H.; Seong, K.A.; Kang, N.S.; Lee, S.Y.; Kim, J.S.; Kim, S.; Yih, W.H. Red tides in Masan Bay, Korea in 2004–2005: I. Daily variations in the abundance of red-tide organisms and environmental factors. *Harmful Algae* **2013**, *30*, S75–S88. [CrossRef]

30. Yang, J.-S.; Choi, H.-Y.; Jeong, H.-J.; Jeong, J.-Y.; Park, J.-K. The outbreak of red tides in the coastal waters off Kohung, Chonnam, Korea: 1. Physical and chemical characteristics in 1997. *Sea* **2000**, *5*, 16–26.

31. Roy, D.P.; Wulder, M.A.; Loveland, T.R.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Helder, D.; Irons, J.R.; Johnson, D.M.; Kennedy, R.; et al. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.* **2014**, *145*, 154–172. [CrossRef]

32. Fan, Y.; Li, W.; Gatebe, C.K.; Jamet, C.; Zibordi, G.; Schroeder, T.; Stamnes, K. Atmospheric correction over coastal waters using multilayer neural networks. *Remote Sens. Environ.* **2017**, *199*, 218–240. [CrossRef]

33. Brockmann, C.; Doerffer, R.; Peters, M.; Kerstin, S.; Embacher, S.; Ruescas, A. Evolution of the c2rcc neural network for sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters. In Proceedings of the Living Planet Symposium, Pregue, Czech Republic, 9–13 May 2016; Ouwehand, L., Ed.; ESA: Paris, France, 2016; p. 54.

34. Franz, B.A.; Bailey, S.W.; Kuring, N.; Werdell, P.J. Ocean color measurements with the operational land imager on Landsat-8: Implementation and evaluation in seadas. *J. Appl. Remote Sens.* **2015**, *9*, 096070. [CrossRef]

35. Singh, K.P.; Basant, A.; Malik, A.; Jain, G. Artificial neural network modeling of the river water quality—A case study. *Ecol. Model.* **2009**, *220*, 888–895. [CrossRef]

36. Maier, H.R.; Dandy, G.C. The use of artificial neural networks for the prediction of water quality parameters. *Water Resour. Res.* **1996**, *32*, 1013–1022. [CrossRef]

37. Maier, H.R.; Morgan, N.; Chow, C.W.K. Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. *Environ. Model. Softw.* **2004**, *19*, 485–494. [CrossRef]

38. Cho, K.H.; Sthiannopkao, S.; Pachepsky, Y.A.; Kim, K.W.; Kim, J.H. Prediction of contamination potential of groundwater arsenic in Cambodia, Laos, and Thailand using artificial neural network. *Water Res.* **2011**, *45*, 5535–5544. [CrossRef] [PubMed]

39. Park, Y.; Kim, M.; Pachepsky, Y.; Choi, S.-H.; Cho, J.-G.; Jeon, J.; Cho, K.H. Development of a nowcasting system using machine learning approaches to predict fecal contamination levels at recreational beaches in Korea. *J. Environ. Qual.* **2018**. [CrossRef]

40. Kim, M.; Baek, S.; Ligaray, M.; Pyo, J.; Park, M.; Cho, K.H. Comparative studies of different imputation methods for recovering streamflow observation. *Water* **2015**, *7*, 6847–6860. [CrossRef]

41. Park, Y.; Pyo, J.; Kwon, Y.S.; Cha, Y.; 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] [PubMed]

42. Farifteh, J.; Van der Meer, F.D.; Atzberger, C.; Carranza, E.J.M. Quantitative analysis of salt-affected soil reflectance spectra: A comparison of two adaptive methods (plsr and ann). *Remote Sens. Environ.* **2007**, *110*, 59–78. [CrossRef]

43. Moody, A.; Gopal, S.; Strahler, A.H. Artificial neural network response to mixed pixels in coarse-resolution satellite data. *Remote Sens. Environ.* **1996**, *58*, 329–343. [CrossRef]

44. Norgaard, M.; Ravn, O.; Poulsen, N.; Hansen, L. *Neural Networks for Modelling and Control of Dynamic Systems: A Practitioner's Handbook. Advanced Textbooks in Control and Signal Processing*; Springer: Berlin, Germany, 2000.

45. Demuth, H.; Beale, M.; Hagan, M. *Neural Network Toolbox™ 6 User's Guide*; The MathWorks, Inc.: Natick, MA, USA, 2008; pp. 37–55. Available online: http://kashanu.ac.ir/Files/Content/neural_network_toolbox_6.pdf (accessed on 15 September 2015).

46. Yu, H.; Wilamowski, B.M. Levenberg–marquardt training. In *Industrial Electronics Handbook*; CRC: Boca Raton, FL, USA, 2011; Volume 5.

47. Park, Y.; Cho, K.H.; Park, J.; Cha, S.M.; Kim, J.H. Development of early-warning protocol for predicting chlorophyll-a concentration using machine learning models in freshwater and estuarine reservoirs, Korea. *Sci. Total Environ.* **2015**, *502*, 31–41. [CrossRef] [PubMed]

48. Park, Y.; Ligaray, M.; Kim, Y.M.; Kim, J.H.; Cho, K.H.; Sthiannopkao, S. Development of enhanced groundwater arsenic prediction model using machine learning approaches in Southeast Asian countries. *Desalination Water Treat.* **2016**, *57*, 12227–12236. [CrossRef]

49. Guo, H.; Jeong, K.; Lim, J.; Jo, J.; Kim, Y.M.; Park, J.-P.; Kim, J.H.; Cho, K.H. Prediction of effluent concentration in a wastewater treatment plant using machine learning models. *J. Environ. Sci.* **2015**, *32*, 90–101. [CrossRef] [PubMed]

50. Xie, X.S.; Liu, W.T.; Tang, B.Y. Spacebased estimation of moisture transport in marine atmosphere using support vector regression. *Remote Sens. Environ.* **2008**, *112*, 1846–1855. [CrossRef]

51. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [CrossRef]

52. Concha, J.A.; Schott, J.R. Retrieval of color producing agents in case 2 waters using Landsat 8. *Remote Sens. Environ.* **2016**, *185*, 95–107. [CrossRef]

53. Vanhellemont, Q.; Ruddick, K. Advantages of high quality swir bands for ocean colour processing: Examples from Landsat-8. *Remote Sens. Environ.* **2015**, *161*, 89–106. [CrossRef]

54. Bricaud, A.; Morel, A.; Babin, M.; Allali, K.; Claustre, H. Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models. *J. Geophys. Res.* **1998**, *103*, 31033–31044. [CrossRef]

55. Dierssen, H.; Smith, R. Bio-optical properties and remote sensing ocean color algorithms for antarctic peninsula waters. *J. Geophys. Res.* **2000**, *105*, 26301–26312. [CrossRef]

56. Kahru, M.; Mitchell, B.G. Empirical chlorophyll algorithm and preliminary seawifs validation for the california current. *Int. J. Remote Sens.* **1999**, *20*, 3423–3429. [CrossRef]

57. Sancak, S.; Besiktepe, S.T.; Yilmaz, A.; Lee, M.; Frouin, R. Evaluation of seawifs chlorophyll-a in the Black and Mediterranean Seas. *Int. J. Remote Sens.* **2005**, *26*, 2045–2060. [CrossRef]

58. Tassan, S. Local algorithms using seawifs data for the retrieval of phytoplankton, pigments, suspended sediment, and yellow substance in coastal waters. *Appl. Opt.* **1994**, *33*, 2369–2378. [CrossRef] [PubMed]

59. Bernardo, N.; Watanabe, F.; Rodrigues, T.; Alcantara, E. Atmospheric correction issues for retrieving total suspended matter concentrations in inland waters using OLI/Landsat-8 image. *Adv. Space Res.* **2017**, *59*, 2335–2348. [CrossRef]

60. Murphy, R.E.; Barnes, W.L.; Lyapustin, A.I.; Privette, J.; Welsch, C.; DeLuccia, F.; Swenson, H.; Schueler, C.F.; Ardanuy, P.E.; Kealy, P.S. Using viirs to provide data continuity with modis. In Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Symposium, Sydney, Australia, 9–13 July 2001; pp. 1212–1214.

61. Hooker, S.B.; Firestone, E.R.; Esaias, W.E.; Feldman, G.C.; Gregg, W.W.; Mcclain, C.R. *SeaWiFS Technical Report Series. Volume 1: An overview of SeaWiFS and Ocean Color*; NASA Goddard Space Flight Center: Greenbelt, MD, USA, 1992; Volume 1, pp. 1–24.

62. Decker, A.; Malthus, T.; Wijnen, M.; Seyhan, E. The effect of spectral bandwidth and positioning on the spectral signature analysis of inland waters. *Remote Sens. Environ.* **1992**, *41*, 211–225. [CrossRef]

63. Dzwonkowski, B.; Yan, X.H. Development and application of a neural network based ocean colour algorithm in coastal waters. *Int. J. Remote Sens.* **2005**, *26*, 1175–1200. [CrossRef]

64. Vilas, L.G.; Spyrakos, E.; Palenzuela, J.M.T. Neural network estimation of chlorophyll a from meris full resolution data for the coastal waters of galician rias (NW Spain). *Remote Sens. Environ.* **2011**, *115*, 524–535. [CrossRef]

65. Zhan, H.G.; Shi, P.; Chen, C.Q. Retrieval of oceanic chlorophyll concentration using support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 2947–2951. [CrossRef]

66. Dörnhöfer, K.; Göritz, A.; Gege, P.; Pflug, B.; Oppelt, N. Water constituents and water depth retrieval from sentinel-2a—A first evaluation in an oligotrophic lake. *Remote Sens.* **2016**, *8*, 941. [CrossRef]

67. Pahlevan, N.; Lee, Z.; Wei, J.; Schaaf, C.B.; Schott, J.R.; Berk, A. On-orbit radiometric characterization of OLI (Landsat-8) for applications in aquatic remote sensing. *Remote Sens. Environ.* **2014**, *154*, 272–284. [CrossRef]

68. Blix, K.; Eltoft, T. Machine learning automatic model selection algorithm for oceanic chlorophyll-a content retrieval. *Remote Sens.* **2018**, *10*, 775. [CrossRef]