remote sensing

MDPI

Article

A 30-Year Assessment of Phytoplankton Blooms in Erhai Lake Using Landsat Imagery: 1987 to 2016

Wenxia Tan [1,2,*], Pengcheng Liu [1,2], Yi Liu [3], Shao Yang [3] and Shunan Feng [1,2]

1 Key Laboratory for Geographical Process Analysis & Simulation of Hubei Province, Central China Normal University, Wuhan 430079, China; liupc@mail.ccnu.edu.cn (P.L.); fsn.1995@foxmail.com (S.F.)
2 College of Urban and Environmental Sciences, Central China Normal University, Wuhan 430079, China
3 School of Life Sciences, Central China Normal University, Wuhan 430079, China; liuyiabcd@126.com (Y.L.); mryangshao@126.com (S.Y.)
* Correspondence: tanwenxia@mail.ccnu.edu.cn

Received: 11 October 2017; Accepted: 5 December 2017; Published: 6 December 2017

Abstract: Long-term information of phytoplankton bloom is critical for assessing the processes driving blooms in lakes. A three-decade survey of the phytoplankton blooms was completed for Erhai Lake from 1987 to 2016 with Landsat imagery. A modified three-band model using Landsat broad bands is developed by comparing reflectance data from Landsat imagery to two field datasets. The model is applied to the archived imagery (1987–2016) to predict chlorophyll-a (Chl-a). Predicted ln(Chl-a) and observed ln(Chl-a) measurements are significantly correlated (R^2 = 0.70; RMSE = 0.13 μg/L). Bloom maps are generated by identifying Landsat pixels that have Chl-a concentrations larger than 20 μg/L as bloom area. Bloom extent and magnitude are estimated. Our study reveals that algal blooms first occurred in 1996 with a bloom area of 150 km^2. Bloom occurred frequently from 2002 to 2016, with extreme blooms in 2003, 2013 and 2016. Algal blooms were mostly distributed in the northern and southern part of the lake. The proposed method uses one model for all Landsat images for Erhai Lake and can predict past blooms and extend the record to early years when field data is not available. The bloom extent and magnitude produced in this study can be used as the basis for the understanding of the processes that control the bloom outbreak.

Keywords: Landsat; chlorophyll-a; phytoplankton bloom; algal bloom; surface reflectance

1. Introduction

Due to increasing anthropogenic nutrients input and global climate change, eutrophication and phytoplankton blooms have affected inland water like rivers and lakes worldwide in recent years [1]. Understanding the factors that influence the initiation and growth of the blooms are of vital importance to the ecosystem [2]. Temperature, light and nutrient status are the most common factors for the growth of algae or cyanobacteria. Long term water quality monitoring programs are critical for providing data to find the driving factors of blooms. However, field observations are usually only available for a few stations and often cover short periods of time due to the high cost of maintaining personnel and equipment to collect field data [3]. Moreover, the extent of the bloom can hardly be provided as measurements are only available for limited locations. Thus, historic bloom information such as the occurrence, spatial extent, magnitude and timing of blooms is scarce for a given system. However, observational information of historical blooms would help understand the cause of the bloom and improve current management efforts, and provide new insight into processes driving bloom spatial and temporal variability.

Chlorophyll-a (Chl-a) concentrations are indicators of algal biomass, and can be easily related to the trophic status of lakes. The level of eutrophication has a positive correlation with the concentration of Chl-a because Chl-a is a major pigment that absorbs solar radiation energy as part of the optical response of phytoplankton biomass [4]. Remote sensing technology can provide synoptic observations with short revisit time at low cost, and have been widely used for monitoring the Chl-a concentrations for inland waters. Many studies have developed models from both broad band and narrow band multispectral satellite sensors to predict Chl-a concentrations for inland waters, but the models and the corresponding parameters can only be applied to the very few images under study and cannot be used on other images [5–8]. A few studies have produced historical Chl-a concentration values and bloom maps using Landsat [9–12], MERIS [13,14], MODIS [15], suggesting satellite remote sensing is promising for the retrieval of historic lake environmental parameters. Among them, Landsat series images provide the longest observation record of the earth surface free of charge with a spatial resolution of 30 meters [16,17]. Thus, the continuity of images from Landsat archive allow for long term monitoring of environmental parameters at regional scale especially for small lakes which are too small to be seen from MODIS.

Erhai Lake is a famous tourism destination in China. Concerns have been drawn to Erhai Lake in recent years as the worsening of the ecological environment of the lake [18]. The research status and problems on ecology and environment of Erhai Lake were summarized in [18], but detailed phytoplankton bloom history data of the lake is still unknown. This study is part of a larger research effort aimed at better characterizing Erhai Lake for effective future management. The aims of this study are to (1) develop the best algorithm for Chl-a retrieval for Erhai Lake using Landsat imagery; (2) produce historic chlorophyll-a concentration maps for 1987 to 2016; (3) hindcast bloom location, extent and magnitude for 1987 to 2016.

2. Study Area and Data

2.1. Study Site and In Situ Data

Erhai Lake, located in Dali, the southwest of China in Yunnan Province (25°36′~25°58′N, 100°06′~100°18′E), is the second largest fresh water body in that province. The average water depth is 10.5 m with the deepest point more than 20 m. The Erhai Lake region has a humid climate with annual mean temperature of 15 to 16 °C. Precipitation of the Erhai Lake region shows significant seasonality, with the wet seasons (May–October) accounting for 70–90% of the total annual precipitation, while the dry seasons (November–April) only representing 10–30%. The lake is mainly fed by precipitation and runoff. Two main inflows are in the northern tip of the lake and a third inflow is in the southeast. The only outlet river of the lake is located at the southwest corner. Blooms in Erhai Lake usually occur during summer and autumn and are dominated by Microcystis that are harmful [19].

Regular water quality observation data for Erhai Lake are available for a relatively short period since 2009. Most information on blooms for Erhai Lake prior to 2009 is based on types and/or locations of observations that are not entirely intercomparable. Thus, little is known about the severity of historical blooms. Erhai Lake was sampled once per month for Chl-a analysis from March 2009 to January 2011 by the Institute of Hydrobiology, Chinese Academy of Sciences, including 12 sampling points that were distributed in four parallel line from the northern to the southern part of lake with three points in each line from west to east (Figure 1). From August 2014 to October 2016, Erhai Lake was sampled monthly (but with a few month gaps) by the School of Life Sciences, Central China Normal University, including five sampling points that were in northern, southern and central part of the lake. Thus, the total number of in situ samples is more than three hundred. As shown in Figure 2, the measured Chl-a show a strong seasonal variation, with the minimum in March to April and the maximum in October to November. A few Chl-a measurements in Fall 2009 have values larger than 30 μg/L, which are much higher than the Chl-a concentrations in other months, indicating severe

algae blooms occurred in Fall 2009. Chl-a data from these two sources are used for model development and validation.

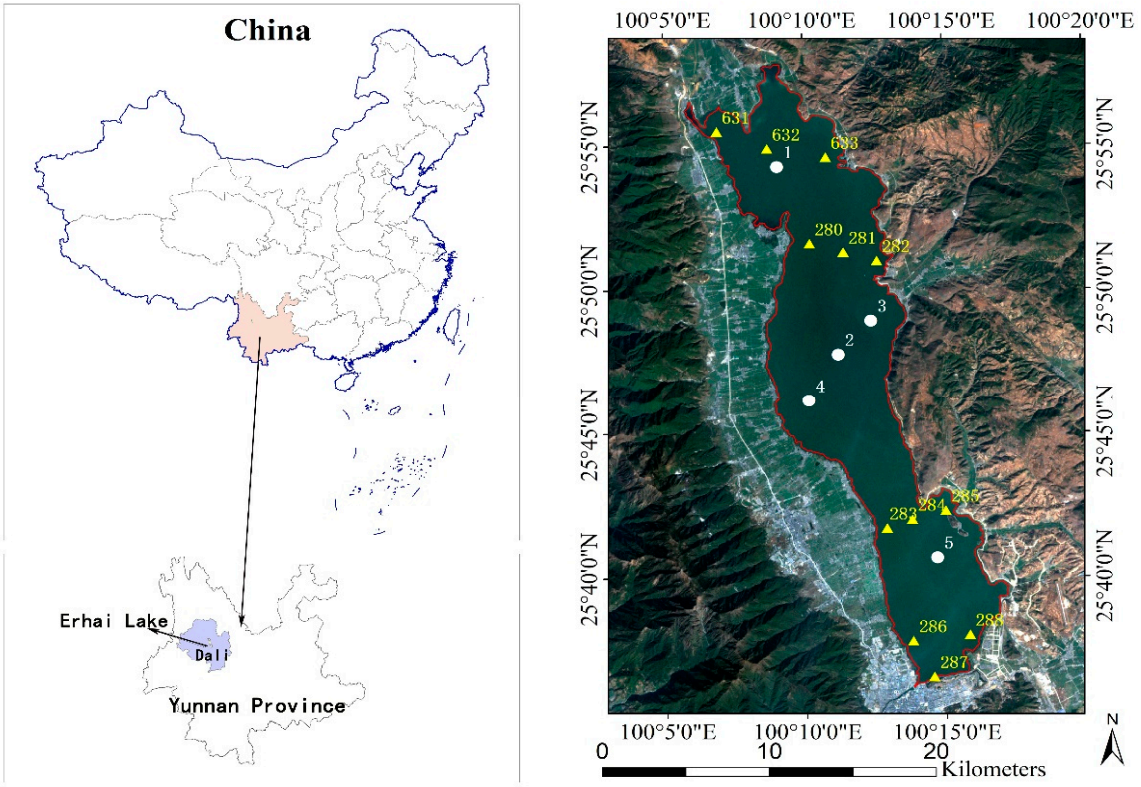

Figure 1. Study area and in situ chlorophyll-a (Chl-a) sample locations. The background is a true-color composite Landsat 7 image acquired on 11 December 2009.

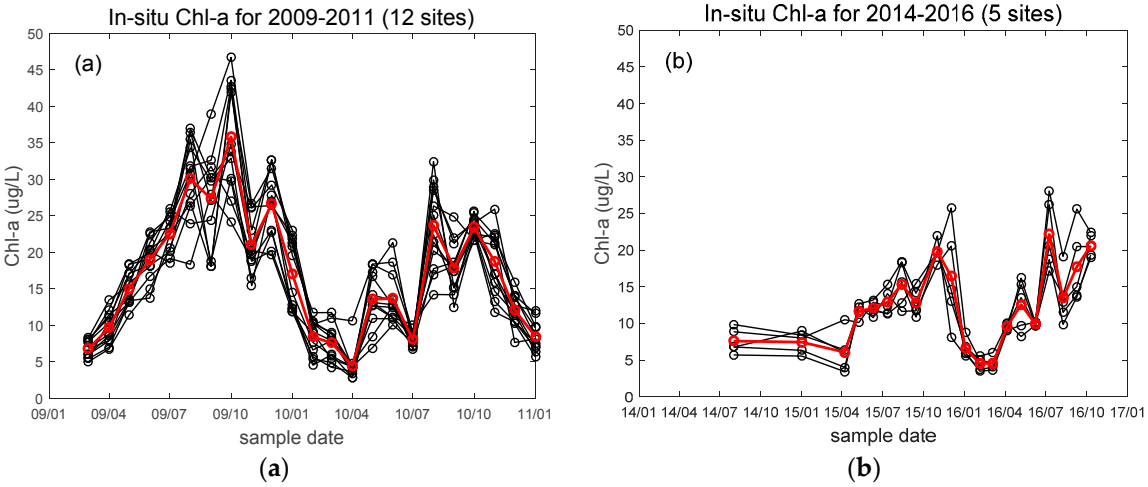

Figure 2. (**a**) In situ Chl-a measurements for each of the sampling site shown in black for 2009–2010, imposed red line is the mean for all sites (**b**) The same as (**a**), but for 2014–2016.

Our historical model is also tested by comparing predicted data versus actual Chl-a data. This actual data with 21 sampling point was collected between 10 to 12 September 2011 and has been used in previous studies [20,21]. Landsat 7 image collected on 28 September 2011 was the only image available for that month. As five sampling points are located on the scan line error, 16 sampling points are used in the validation.

2.2. Landsat Data

All available Landsat Surface Reflectance Level-2 science data products of Landsat 5 and 7 are obtained from the USGS EarthExplorer website (earthexplorer.usgs.gov) (Table 1). Landsat 5 was launched in 1982, but continuous data collection for Erhai Lake did not begin until the end of 1986. Erhai Lake lies in Path 131 and Row 42 of the WRS-2 orbit. Cloud-free images over Erhai Lake during the wet season are quite rare, thus through manual interpretation images with a few cloud covers are also included, but only data outside the cloud cover or cloud shadow are utilized. More than three hundred images covering the lake area are used in the analysis for a 30-year period (1987–2016). For the period of 1987 to 1999, only Landsat 5 was operational. From 1999 to 2011, Landsat 5 and 7 were both operational. From 2012 to 2016, only Landsat 7 collected imagery. Although Landsat 7 has been hampered by the failure of the Scan Line Corrector (SLC-off) since 2003, the image areas that have not been affected are included in this analysis as the scan line error areas were masked and excluded from analysis using USGS's reflectance products [22]. Landsat 5 and 7 sensors have a revisit time of 16 days, but 8-days revisit time can be achieved when Landsat 5 and 7 were both operational from 1999 to 2011. Table 2 shows the number of Landsat images used in this study. Though Chl-a usually increases from May to October, very limited satellite data is available for May to August because of the high cloud cover during the wet season in Erhai Lake region.

Table 1. Comparison of Landsat 7 Enhanced Thematic Mapper Plus (ETM+) and Landsat 5 Thematic Mapper (TM) bands. The bands used in this study are shown in bold letters.

Landsat5 (TM)		**Landsat7 (ETM+)**	
Bands	**Wavelength (μm)**	**Bands**	**Wavelength (μm)**
Band 1-Blue	**0.45–0.52**	**Band 1-Blue**	**0.45–0.52**
Band 2-Green	**0.52–0.60**	**Band 2-Green**	**0.52–0.60**
Band 3-Red	**0.63–0.69**	**Band 3-Red**	**0.63–0.69**
Band 4-Near Infrared (NIR)	**0.76–0.90**	**Band 4-Near Infrared (NIR)**	**0.77–0.90**
Band 5 -Shortwave Infrared (SWIR) 1	**1.55–1.75**	**Band 5-Shortwave Infrared (SWIR) 1**	**1.55–1.75**
Band 6-Thermal	10.40–12.50	Band 6-Thermal	10.40–12.50
Band 7-Shortwave Infrared (SWIR) 2	**2.08–2.35**	**Band 7-Shortwave Infrared (SWIR) 2**	**2.09–2.35**
		Band 8-Panchromatic	0.52–0.90

Table 2. The number of Landsat images used in our study for different months.

Year	Jan.–Apr.	May–Aug.	Sept.–Dec.	Year	Jan.–Apr.	May–Aug.	Sept.–Dec.
1987	4	2	1	2002	9	4	8
1988	4	2	3	2003	13	2	8
1989	5	4	6	2004	10	3	7
1990	6	0	4	2005	7	3	5
1991	6	2	5	2006	6	6	7
1992	5	3	4	2007	4	4	3
1993	5	3	3	2008	8	2	5
1994	6	0	4	2009	8	0	7
1995	5	0	3	2010	9	1	4
1996	6	2	5	2011	6	2	6
1997	2	1	2	2012	1	2	4
1998	3	2	5	2013	5	2	5
1999	6	2	6	2014	6	1	6
2000	4	2	6	2015	1	2	2
2001	7	6	4	2016	5	2	5

The Landsat images are geometrically corrected by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) before the images are downloaded [23]. Pixels with clouds, cloud shadows, and snow are removed using the pixel quality assessment data along with the surface reflectance products [24]. Water-only images are created by clipping the original images using the lake

boundary. A 3 by 3 mean filter is used to smooth data and reduce random high frequency noise [25]. All the reflectance values are converted to a 0–1 value range by a scale factor of 0.0001.

3. Method

3.1. Selection of Landsat Spectral Bands and Ground Data

In situ Chl-a concentrations are temporally matched with the closest available Landsat images. Kloiber et al. [26] found that the closeness of fit (r^2 values) of regression models between satellite reflectance data and measured water clarity decreases with increasing length of the time window between image collection and ground observation due to constant water flow. Thus, Kloiber et al. [26] concluded that a window of up to ±7 days yielded reasonable results between satellite data and lake water clarity when observation data were limited. As the wide time window allows the inclusion of more observation data to conduct retrospective analyses with archived satellite imagery, all Landsat 5 and 7 images captured within eight days of the sampling dates are investigated in this study (Table 3).

Table 3. Image dates for Landsat 5 and 7 and Chl-a field sampling dates that are within a window of ±8 days of the satellite overpass.

Landsat Date	Sample Date	lag	Landsat Date	Sample Date	lag
7/4/2009	1/4/2009	6	25/12/2014	1/1/2015	−7
30/9/2009	1/10/2009	−1	2/5/2015	9/5/2015	−7
24/10/2009	1/11/2009	−8	25/10/2015	1/11/2015	−7
12/27/2009	1/1/2010	−5	26/11/2015	2/12/2015	−6
28/1/2010	1/2/2010	−4	29/1/2016	4/2/2016	−6
1/3/2010	1/3/2010	0	1/3/2016	5/3/2016	−4
9/9/2010	1/9/2010	8	4/5/2016	7/5/2016	−3
30/12/2010	1/1/2011	−2	9/9/2016	9/9/2016	0
			11/10/2016	10/10/2016	1

3.2. Remote Sensing Reflectance for Different Chl-a Concentrations

The histogram of all in situ Chl-a concentrations that have matched Landsat imagery within five days are shown in Figure 3a. The Chl-a concentrations range from 3.6 to 46.7 μg/L with a mean of 14.6 μg/L. The Chl-a are grouped into three categories to explore the relationship between Chl-a concentrations and Landsat surface reflectance at different bands based on the distribution of Chl-a concentrations, with concentration values less than 20 μg/L (category 1), between 20 and 30 μg/L (category 2), and larger than 30 μg/L (category 3).

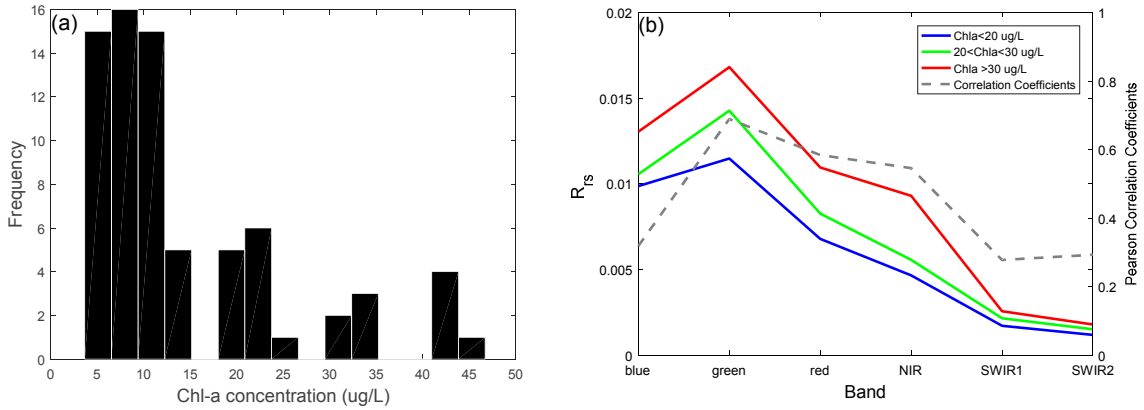

Figure 3. (**a**) The histogram of all in situ Chl-a measurements that have matched Landsat imagery within five days; (**b**) The mean remote sensing reflectance for different Chl-a concentration categories and the Pearson correlation coefficients between ln(Chl-a) and remote sensing reflectance.

Remote sensing reflectance (R_{rs}) are calculated by dividing the Landsat surface reflectance by π [27]. R_{rs} curves for the three Chl-a categories are shown in Figure 3b. All R_{rs} curves are similar in general outline (Figure 3b), regardless of the image acquisition dates or the locations of the sampling points or the Chl-a concentrations. These R_{rs} curves are also similar in magnitude and shape to the typical R_{rs} curves collected for Erhai Lake using a spectroradiometer during two field surveys in 2011 and 2012 [20,21]. The reflectance curves show maximum reflectances at Green band, and exhibit a downturn at the Red band. This is because Red band overlaps a region with strong absorbance by Chl-a, but reflectance from algal cells offsets the absorbance to some extent [28]. It is worth noting that reflectances in the NIR band are much smaller than that in the Red band for Erhai Lake, compared to many investigations of other lakes which reported that reflectances in NIR band are very high for bloom areas due to surface scum [25,29]. Thus, pixels with reflectance values in NIR band larger than in the Red band are masked in our study, as other components of the lake water may contribute to the increase of reflectance of NIR band. Due to the high absorption coefficient of pure water, reflectance values decrease to less than one percent and the reflectance curves get tangled at SWIR bands (Figure 3b). Thus, pixels with reflectance values in SWIR1 band larger than two percent are suspected to be influenced by fog and are masked.

The values of R_{rs} for all the six bands are positively correlated to the Chl-a concentrations as shown in Figure 3b, and these correlations are significant for Green, Red and NIR bands at $p < 0.01$ level. The correlation coefficient for Green band is 0.69 which is the largest for all the bands investigated, followed by 0.58 and 0.55 for blue and NIR band. The increase of remote sensing reflectances can be attributed to the increased biomass at high Chl-a concentrations and the biomass increase the scattering at these wavelengths. However, the magnitudes of the increase vary due to the differences in absorbance and reflectance of algal pigments in these bands [10].

3.3. Chl-a Algorithm Development and Validation

Comprehensive reviews of empirical and semi-analytical procedures for Chl-a retrieval in inland and near-coastal waters using passive optical remote sensing data can be found in [3,25,30]. Semi-analytical models, such as the three-band model and the floating algae index (FAI) [31], are developed based on solutions to the radiative transfer equation and can be applied to different water types that have similar optical properties. However, semi-analytical models are designed and tuned for narrow band sensors such as MODIS and Medium Resolution Imaging Spectrometer (MERIS), as the absorbing feature of phycocyanin at ~620 nm are often used. Empirical models, such as linear regression and multi-linear regressions based on bands, band ratios and band arithmetic algorithms, are easy to implement, but they usually lack physical foundations and are geographically specific [32].

The first objective of this study is to develop a single equation using Landsat reflectance that have matched ground data by linear regression and use this equation on all archived images for Erhai Lake. Thus, empirical models are used, while complex algorithms like machine learning, relevance vector machines are avoided [33]. Various band combinations are explored and the estimated Chl-a are compared with in situ Chl-a data. The model with the highest closeness of fit is selected. Natural log-transformed Chl-a observations is used because the relationship between band combinations values and log(Chl-a) were shown to be approximately linear over a broad range of Chl-a [34]. The five most effective algorithms for the study area are listed in Table 4. They all use Landsat band 1 to 4. Algorithm #1 to #4 are based on band arithmetic and band ratios. According to [3], (TM1 − TM3)/TM2 and TM2/TM1 are recommended for low Chl-a concentration situations (<20 μg/L), while TM3/TM1 is recommended for high Chl-a concentration situations (>20 μg/L). Algorithms #4 $(TM1^{-1} - TM2^{-1}) \times TM4$ is inspired by the classical three-band model, which is originally designed to estimate pigment contents in terrestrial vegetation, and has been proved to be very effective for Chl-a concentration retrieval [32]. The three-band model uses two reciprocal reflectance where λ1 is the wavelength of the band maximally sensitive to Chl-a absorption while λ2 is the wavelength of the band minimally sensitive to Chl-a absorption. Additionally, λ3 is the wavelength of the band

minimally affected by absorption that accounts for scattering. Multivariate linear regression (MLR) has also been considered [12,35]. Several MLR combination were explored through stepwise procedures for our data, and the combination of TM1, TM2 and TM4 is the most effective.

Table 4. Evaluation of algorithm. MLR is short for multi-linear regression. RRMSE is the relative root mean square error (RMSE), which is RMSE divided by the mean.

#	Algorithm	±8 Days & Samples:126			±5 Days & Samples: 73		
		R^2	RMSE	RRMSE (%)	R^2	RMSE	RRMSE (%)
1	(TM1 − TM3)/TM2	0.48	0.20	8.2	0.69	0.14	5.7
2	TM2/TM1	0.47	0.20	8.2	0.44	0.25	10.2
3	TM3/TM1	0.39	0.23	9.4	0.64	0.16	6.5
4	$(TM1^{-1} - TM2^{-1}) \times TM4$	0.59	0.16	6.5	0.70	0.13	5.3
5	MLR: TM1, TM2, TM4	0.59	0.16	6.5	0.78	0.10	4.1

The selection of the optimal model is a key step to obtaining a model with good predictive capability. To assess model performance, validation of the models is undertake using a leave one out cross-validation approach. A calibration model is built using N − 1 samples and the sample left out is predicted. The root mean sum of squares of prediction error of cross validation (MSECV) (Equation (1)) is used as a measure to select the optimal model because of its effectiveness in indicating how well the model predicts new samples. The corresponding regression equation for the validated model is applied to all Landsat images to generate Chl-a concentration maps, obtaining a time-series of Chl-a for Erhai Lake.

$$\text{RMSECV} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y' - y)^2}, \quad (1)$$

where y' and y are measured and predicted log(Chl-a) concentrations and n is the total number of samples. Moreover, linear regression analysis and root mean square error (RMSE) is used to assess model performance by comparing the predicted Chl-a with actual Chl-a collected in September 2011.

3.4. Bloom Occurrence, Extent and Magnitude Extraction

Under conditions of cyanobacterial dominance, World Health Organization (WHO) uses a Chl-a threshold of 10 μg/L for increased odds of irritative or allergenic effects, and 50 μg/L for increased probability of irritative symptoms and toxic impacts [36]. As Erhai Lake is in the early stage of eutrophication, 20 μg/L is selected as the threshold for bloom. The size of a bloom, defined as bloom extent, is the sum of area where Chl-a concentration is larger than 20 μg/L. As Chl-a data are not available for areas with cloud cover or scan line error, the extent reported is a minimum estimation for the day under consideration. Based on visual examination of the bloom maps, pixels along the shoreline are occasionally detected as bloom pixels in a few images. These pixels are suspected to be under the influence of water plants growing along the shoreline. To eliminate these errors, a bloom occurrence event is identified only when the bloom extent is larger than one tenth of the total area of Erhai Lake, which is larger than 25 km^2. The severity of a bloom, defined as bloom magnitude, is inferred from the mean Chl-a concentrations for those pixels that have Chl-a concentrations larger than 20 μg/L. Thresholds of 20, 25, 30 μg/L are used to define the severity of the blooms as moderate, high and very high.

4. Results

4.1. Chl-a Model Calibration and Validation

As shown in Table 4, algorithm #4 and #5 perform better than the other three algorithms when using samples of eight days and five days. The relationship between the estimated log(Chl-a) and

in situ log(Chl-a) using samples that are within five days of the corresponding Landsat images are shown in Figure 4. A model will have the advantage of covering a wider variety of image observation conditions and the advantage of covering various lake conditions if a greater number of separation days are included. Thus, the effect of separation date on the performance of the algorithms is explored by successively increasing the time window between satellite overpass and ground observation from within one day to within eight days (Figure 5). For both model #4 and #5, the highest R^2 and the lowest standard error of estimate are achieved when time window is within five days. When the time window increases to ±6 days, R^2 decreases for both algorithms. When further increase the time window to ±7 and ±8 days, R^2 and RMSE keep almost stable with the gradual increase in the number of observations. Thus, the optimum time window is set as ±5 days to ensure the best fit of the model.

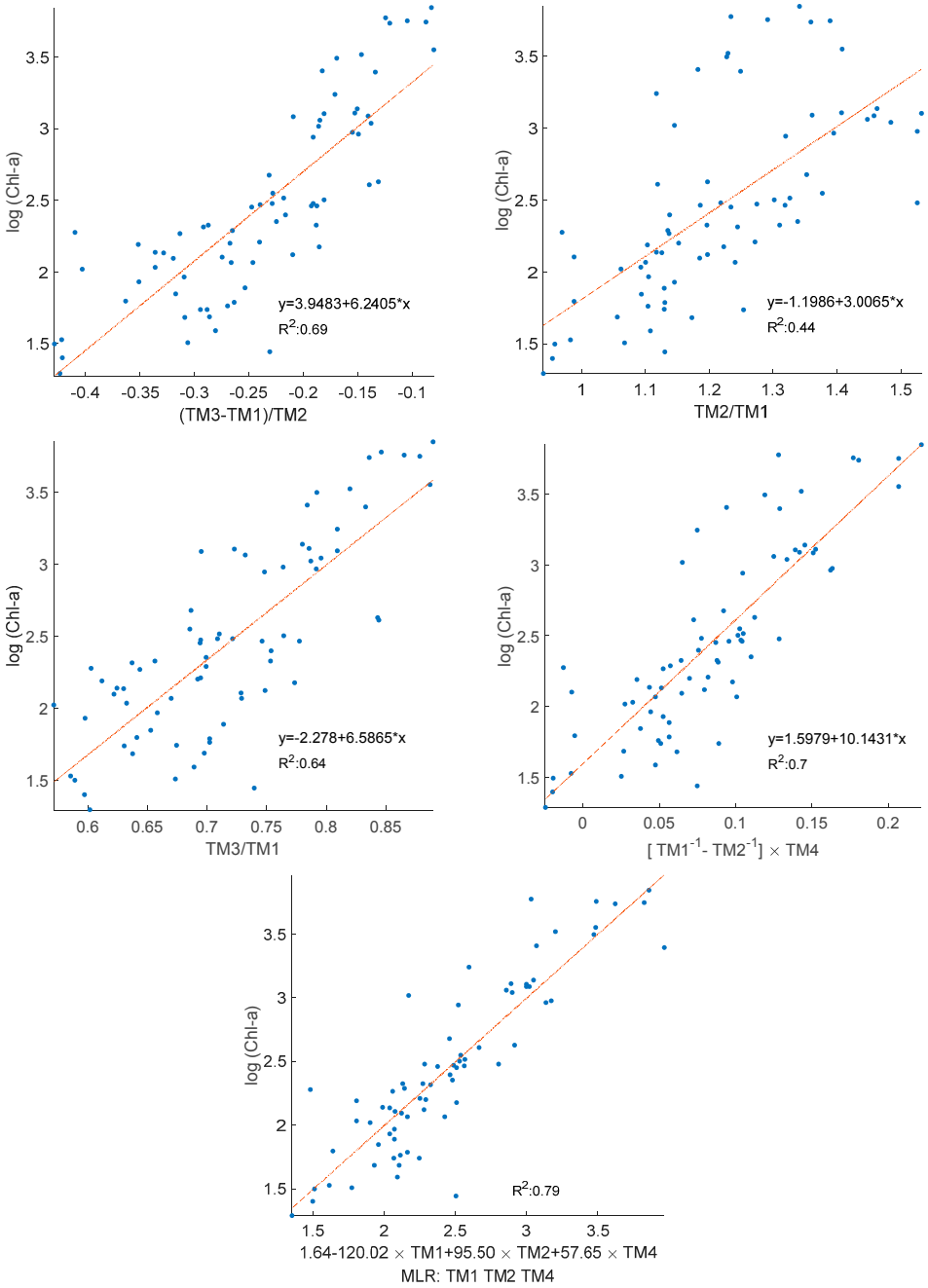

Figure 4. Comparison between estimated log(Chl-a) and in situ log(Chl-a) using different algorithms using in situ observations that have matched Landsat images within five days.

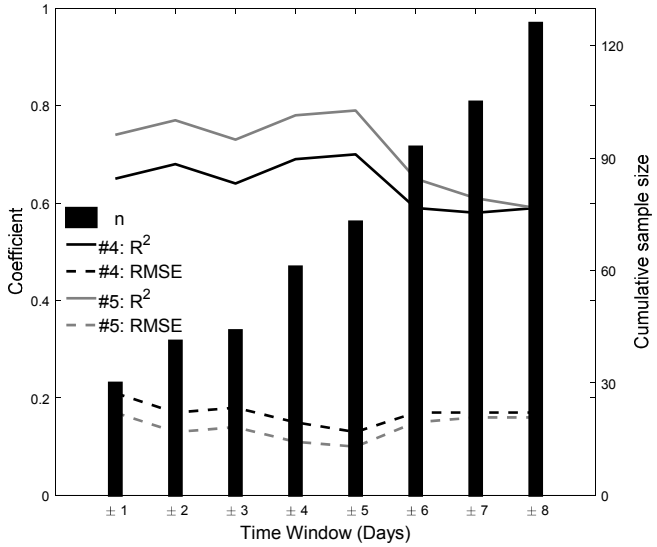

Figure 5. Relationship between the width of the time window between satellite and ground observations and the number of samples (n), R^2, and RMSE for algorithm #4 and #5.

Calibration and validation of algorithms #4 and #5 are conducted as described in Section 3.3. The model predicted values for each sample point produced by the leave one out cross validation procedure is used in a linear regression with the in situ measurements of Chl-a concentrations (Figure 6). The RMSECV for algorithms #4 and #5 are 0.37 and 0.64, while the R^2 for algorithms #4 and #5 are 0.69 and 0.06, showing that algorithm #4 performs better than algorithm #5. As MLR suffers from the problem of overfitting, algorithm #4 is selected as the final model for modeling Chl-a concentrations. Using a total sample number of 73 for a time window of ±5 day, the model is defined in Equation (2):

$$\log(\text{Chl} - \text{a}) = 1.5979 + 10.1431 \times \left(TM1^{-1} - TM2^{-1}\right) \times TM4, \quad (2)$$

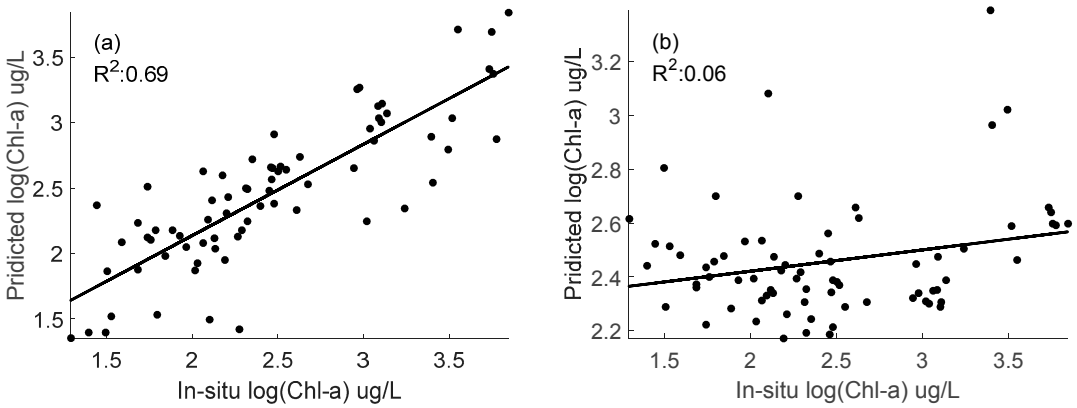

Figure 6. Linear regression of in situ measurements of Chl-a and model predicted for algorithm #4 (**a**) and #5 (**b**).

The chosen model is tested on actual data (Figure 7). There is significant relationship between the predicted and actual log(Chl-a) (n = 16, R^2 = 0.6, $p < 0.001$, RMSE = 0.44 µg/L, RRMSE = 16%) as shown in Figure 8. The Chl-a model tend to underestimate actual Chl-a concentrations. Considering the fact that the imaging dates are more than 16 days apart from the sampling date, the predicted Chl-a concentrations are acceptable.

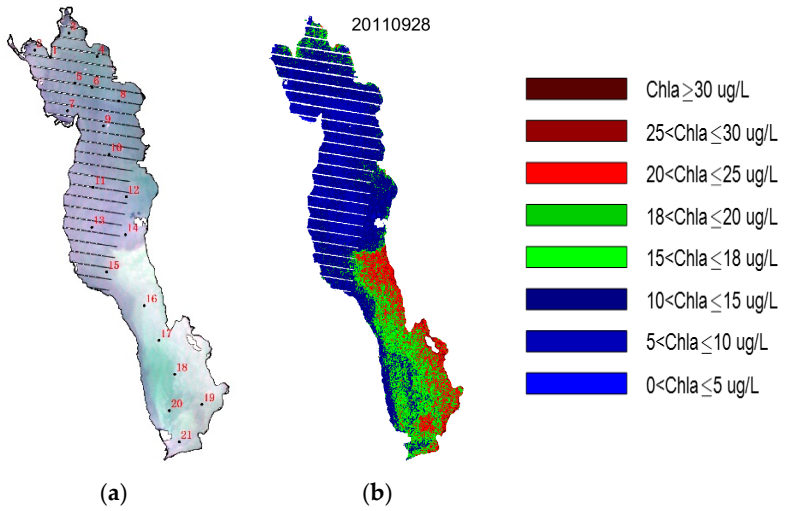

Figure 7. (**a**) The 21 actual data points collected during 10–12 September 2011 are shown with the true-color composite Landsat 7 image acquired on 28 September 2011; (**b**) The Chl-a concentration map for 28 September 2011 derived using our selected model. The parallel lines on the maps are areas affected by the Landsat 7 scan line error.

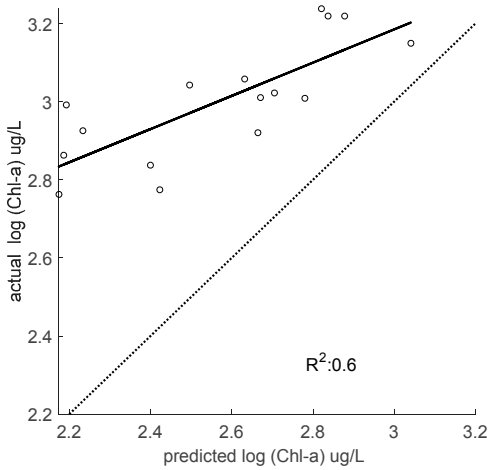

Figure 8. Predicted and actual log (Chl-a). The 1:1 line is represented by the dashed line, while the solid line represents the observed relationship between the satellite estimated values and actual measured values (n = 16, R^2 = 0.60, RMSE = 0.44 μg/L, RRMSE = 16%).

4.2. Spatial and Temporal Analysis of Blooms

Based on peer-reviewed literature, qualitative depictions of Erhai blooms have been reported in the embayment and some parts of the lake in years 1996, 1998, 2003, 2006, 2009 and 2013 [37–41]. As shown in Figure 9, the bloom extents for Erhai Lake were small in the late 1980s to 2001 except 1996 and 1997. During that period, year 1996 was the first time a large bloom occurred in Erhai Lake for the 30-year study period. Bloom in 1996 was record setting, even for the 1986–2016 study period, with a mean Chl-a concentration of more than 50 μg/L. Since 2002, an increasing number of algae blooms have occurred based on our results. Bloom extent was larger than 25 km^2 in year 2003, 2005, 2006, 2008, 2009, 2010, 2013 and 2016 (Figure 9), which generally consistent with other researches that Erhai Lake as developed to moderately eutrophication since 2002 [18]. A Mann–Kendall statistical analysis performed on the yearly maximum bloom extents show that there is significant trend in the bloom extent for Erhai Lake. The bloom magnitude was very high in years 1996, 2006 and 2013 with mean Chl-a larger than 30 μg/L (Table 5). We note that the 1998 bloom was not detected in our data.

All days captured in 1998 had a very low Chl-a concentration. One possible reason could be that the bloom lasted only a short period and the bloom period were not imaged by Landsat.

As shown in Figure 10, high Chl-a concentrations most frequently occur on the southern and northern part of the lake, as 60% of the water of Erhai Lake comes from the two main inflows at the northern tip [38]. These two inflows are the sources of pollution for Erhai Lake. The nutrients input from runoff causes the high Chl-a concentration in the northern part of the lake. On the other hand, waste water with nutrients is discharged into Erhai Lake from Dali city located on the southwest shore of Erhai Lake. Thus, Chl-a concentrations are high in the northern and southern part of the lake.

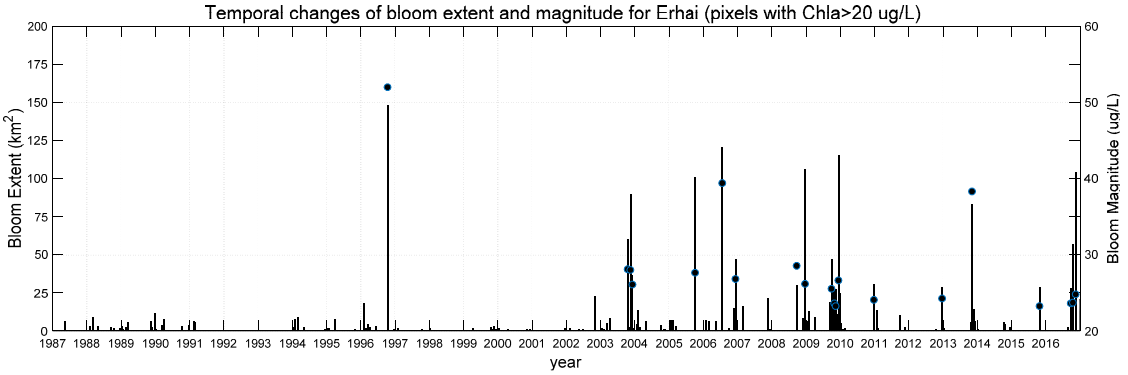

Figure 9. Temporal changes of bloom extent, defined as the area with Chl-a larger than 20 μg/L from 1987 to 2016. The bloom magnitude, which is the mean Chl-a for days with a bloom extent larger than 25 km^2, are shown as dots.

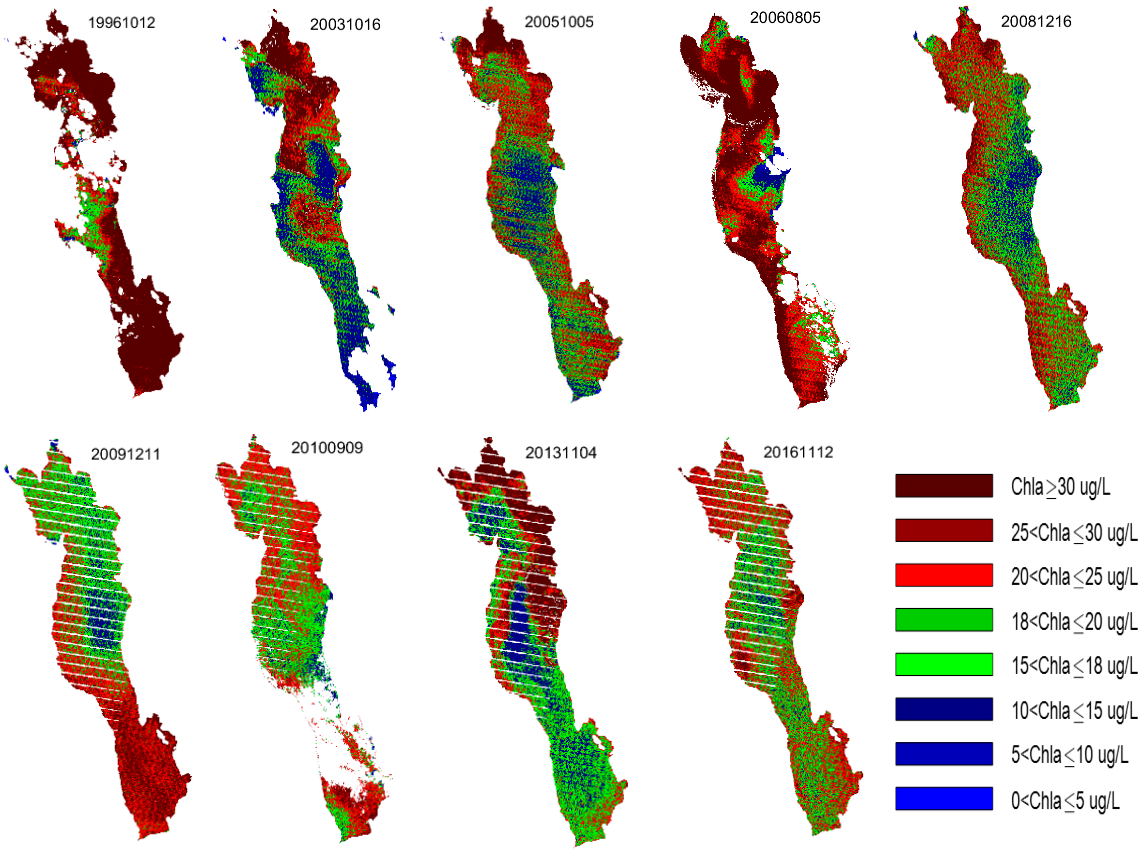

Figure 10. The Chl-a concentration map for selected years with large areas of high Chl-a concentration. The areas affected by the scan line errors are shown in white color on the images.

Table 5. Bloom magnitude definition using the mean Chl-a concentrations for those pixels that have Chl-a concentrations larger than 20 μg/L.

Bloom Magnitude	Moderate	High	Very High
Mean Chl-a concentration (μg/L)	20–25	25–30	30
year	2010, 2012, 2015	2003, 2005, 2009, 2016	1996, 2006, 2013

5. Discussion

From the bloom extent data, large bloom rarely occurred in Erhai Lake before 2002, other than the outbreak of algae bloom in 1996. Nevertheless, bloom occurred almost every year since 2002. In 1996, the estimated bloom extent reached 148 km^2, accounting for half of the lake water area. One of the most important reasons for this disaster is the proliferation of silver fish. It is reported that the silver fish was successfully introduced into the lake as alien biota in 1991, and the harvest was about 253 t in 1993, and 650 t in September 1996 [41]. Silver fish mainly feed on zooplankton, which is a key part of the grazing food chain and thus has a strong control for phytoplankton density. The excessive reproduction of silver fish in Erhai Lake destroyed the balance of the food chain, resulting in a dramatic increase in phytoplankton. Meanwhile, the rainfall was low in 1996. The reduced water storage capacity lead to the increase in nutrient concentration, which is conducive to algae reproduction [34]. The interference of human activities was another important reason. The number of Dali tourists increases from the 0.89 million in 1995 to 3.24 million in 1996 (Dali Yearbook). The increase in the number of tourists led to the increase of wastewater and nutrients discharge into the lake and intensified the bloom in 1996 [42].

Since 2002, both the bloom extent and bloom frequency increased, the bloom extent reached more than 25 km^2 almost every year. The driving force for this comes from both natural and human factors. From the mid-1990s, the trend of temperature growth in Dali is more obvious. The average temperature in 2000–2009 was 0.41 °C higher than that in 1990–1999 (Dali state weather station), which is beneficial to algal blooms. In the aspect of human disturbance, the number of tourists in Dali, the area of crops sown, the number of pig stocks and the output of Erhai fish have entered a period of rapid growth. The industrial and agricultural output value has increased significantly (Dali Yearbook). The rapid development of agriculture, animal husbandry and fishery increase a comprehensive load in Erhai Lake and create conditions for algae breeding rapidly. Our study is consistent with Wang et al. [43] that 2001 is a critical transition point for Erhai Lake.

Because bloom is a mixture of water and algae, the surface reflectance generally decreases from green band to red and infrared bands for all Chl-a concentrations. It is noteworthy to mention that almost all data in 1992 and 1993 have NIR reflectance larger than that in red band. These pixels are suspected to be floating vegetation and are not included for bloom extent estimation, as their surface reflectance features are closer to that of green vegetation. Consequently, almost no data is available for years 1992 and 1993 as shown in Figure 9. The reasons why surface reflectance in 1992 and 1993 show characteristics of green vegetation need further investigation.

It is noticeable that some of the estimated Chl-a concentrations are not well-correlated with the in situ data. This can be attribute to several reasons. First and foremost, the Landsat reflectance is influenced by the changes in the lake water components, including the proportion of total suspended solids and colored dissolved organic matter, in different seasons and through the years. Secondly, the sampling dates do not lie well with the imaging dates, mostly within five days of image acquisition. Concomitant sampling with image acquisition is difficult to obtain due to factors including weather condition, man power, etc. Another reason is the influence of wind and light. During windy conditions on the lake, water quality variables change quickly as phytoplankton floating upward was accommodated by sunlight. For days with relatively high wind speed, the estimated Chl-a concentrations are low and not reliable due to complete mixing of Chla concentrations throughout the water column by wind. Besides, this paper used in situ Chl-a data from different years and seasons, and the water surface's reflectance may vary sharply as the season or year varied. The two groups of

Chl-a observations used also have systematic difference due to different sampling techniques used, as the first group used surface water while the second group used water samples that were mixtures of water from the surface, middle and bottom of the lake. For future modeling of water parameters using remote sensing, it is important to take field observations concurrent with satellite data acquisition.

Although Landsat has a revisit time of 16 days, data gaps can last several months especially during the wet season due to high cloud cover for Erhai Lake. Moreover, several images were screened out by visual interpretation of the true color image composition due to high wind or fog. Thus, the analysis of the phenology of the bloom growth is impossible. MODIS data which have a shorter revisit time but 250 m spatial resolution may be integrated with Landsat in the future to see if bloom phenology could be extracted.

6. Conclusions

The effectiveness of Landsat archive observations for identifying historic phytoplankton blooms in Erhai Lake is demonstrated in this study. Using field measurements, a three-band model adapted to Landsat broad bands was selected as the best algorithm for Chl-a estimation. The proposed model was considered more robust than previous studies of Chl-a retrieval, due to the much larger number of images and in situ sampling days used for model development and validation, as well as the different times of year and the wide range of Chl-a concentrations. The proposed model was then applied to all historical images to extract the Chl-a concentration maps over the 30-year period from 1987 to 2016. The bloom extent and bloom magnitude were estimated quantitatively based on the Chl-a maps using a threshold of 20 μg/L. The timing, spatial distribution, extent, magnitude and the of historic blooms are all generated for the 30-year period. Overall, Erhai Lake did not burst blooms before 2002 except the outbreak of algae bloom in 1996. Meanwhile, both the bloom extent and the outbreak frequency increased greatly since 2002 and bloom can be found almost every year. Year 1996, 2006 and 2016 were identified as the years of most severe blooms. The bloom extent of 1996 was close to 150 km^2 with bloom magnitude of more than 30 μg/L, meaning that more than half of the lake had a Chl-a concentration larger than 30 μg/L on the imaging date of 12 October 1996. Historic quantities of bloom extent, bloom magnitude and the spatial distribution maps clearly provide a reliable and ecological useful indicator of bloom extent and occurrence in Erhai Lake. The results are consistent with blooms reported in other literature, indicating that the developed method is capable of routine monitoring of algal blooms. Our results suggest that the use of a consistent equation for Chl-a is applicable to all historic Landsat 5 and 7 images.

This research used one model for all historic Landsat images of Erhai Lake, instead of one model per image. Since the model has been tested with independent data, it can be used on future images captured in Erhai for water monitoring. The results from this investigation suggest that remote sensing can serve a valuable tool to assess the historic blooms. To fully understand the cause of the blooms for Erhai Lake, a study of variability in river discharge patterns should be considered. Weather conditions such as lake water temperature, cumulative sunny days and precipitation before the large bloom should also be studied using concurrent meteorological data. The data produced in this study can be used as the basis for the understanding of the processes that control the bloom outbreak, and consequently used for bloom early warning.

Acknowledgments: Authors would like to acknowledge Genbao Li and Gongliang Yu from the Institute of Hydrobiology (IHB), Chinese Academy of Sciences for providing field data for model development and validation. Gongliang Yu also provided helpful discussions during the revision of this work. This work was supported by the National Natural Science Foundation of China under grants 41506208 and 31670464. Landsat 5 and 7 images are available from the U.S. Geological Survey.

Author Contributions: Wenxia Tan and Shao Yang conceived and designed the experiments; Shunan Feng and Yi Liu performed the experiments; Wenxia Tan and Pengcheng Liu analyzed the data and wrote the paper.

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

References

1. Callisto, M.; Molozzi, J.; Barbosa, J.L.E. Eutrophication of lakes. In *Eutrophication: Causes, Consequences and Control: Volume 2*; Ansari, A.A., Gill, S.S., Eds.; Springer: Dordrecht, The Netherlands, 2014; pp. 55–71.
2. Ho, J.C.; Michalak, A.M. Challenges in tracking harmful algal blooms: A synthesis of evidence from lake erie. *J. Great Lakes Res.* **2015**, *41*, 317–325. [CrossRef]
3. Matthews, M.W. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. *Int. J. Remote Sens.* **2011**, *32*, 6855–6899. [CrossRef]
4. Carlson, R.E. A trophic state index for lakes. *Limnol. Oceanogr.* **1977**, *22*, 361–369. [CrossRef]
5. Brivio, P.A.; Giardino, C.; Zilioli, E. Validation of satellite data for quality assurance in lake monitoring applications. *Sci. Total Environ.* **2001**, *268*, 3–18. [CrossRef]
6. Ma, R.; Dai, J. Investigation of chlorophyll-a and total suspended matter concentrations using Landsat ETM and field spectral measurement in taihu lake, China. *Int. J. Remote Sens.* **2005**, *26*, 2779–2795. [CrossRef]
7. Mayo, M.; Gitelson, A.; Yacobi, Y.Z.; Ben-Avraham, Z. Chlorophyll distribution in lake Kinneret determined from Landsat thematic mapper data. *Int. J. Remote Sens.* **1995**, *16*, 175–182. [CrossRef]
8. Wheeler, S.M.; Morrissey, L.A.; Levine, S.N.; Livingston, G.P.; Vincent, W.F. Mapping cyanobacterial blooms in lake Champlain's missisquoi bay using Quickbird and MERIS satellite data. *J. Great Lakes Res.* **2012**, *38*, 68–75. [CrossRef]
9. Allan, M.G.; Hamilton, D.P.; Hicks, B.; Brabyn, L. Empirical and semi-analytical chlorophyll a algorithms for multi-temporal monitoring of new zealand lakes using Landsat. *Environ. Monit. Assess.* **2015**, *187*, 364. [CrossRef] [PubMed]
10. Bonansea, M.; Rodriguez, M.C.; Pinotti, L.; Ferrero, S. Using multi-temporal Landsat imagery and linear mixed models for assessing water quality parameters in río tercero reservoir (argentina). *Remote Sens. Environ.* **2015**, *158*, 28–41. [CrossRef]
11. Ho, J.C.; Stumpf, R.P.; Bridgeman, T.B.; Michalak, A.M. Using Landsat to extend the historical record of lacustrine phytoplankton blooms: A lake Erie case study. *Remote Sens. Environ.* **2017**, *191*, 273–285. [CrossRef]
12. Yip, H.D.; Johansson, J.; Hudson, J.J. A 29-year assessment of the water clarity and chlorophyll-a concentration of a large reservoir: Investigating spatial and temporal changes using Landsat imagery. *J. Great Lakes Res.* **2015**, *41*, 34–44. [CrossRef]
13. Han, X.; Feng, L.; Chen, X.; Yesou, H. Meris observations of chlorophyll-adynamics in Erhai Lake between 2003 and 2009. *Int. J. Remote Sens.* **2014**, *35*, 8309–8322. [CrossRef]
14. Stumpf, R.P.; Wynne, T.T.; Baker, D.B.; Fahnenstiel, G.L. Interannual variability of cyanobacterial blooms in lake Erie. *PLoS ONE* **2012**, *7*, e42444. [CrossRef] [PubMed]
15. Zhang, Y.; Ma, R.; Zhang, M.; Duan, H.; Loiselle, S.; Xu, J. Fourteen-year record (2000–2013) of the spatial and temporal dynamics of floating algae blooms in lake Chaohu, observed from time series of MODIS images. *Remote Sens.* **2015**, *7*, 10523. [CrossRef]
16. Wulder, M.A.; Masek, J.G.; Cohen, W.B.; Loveland, T.R.; Woodcock, C.E. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sens. Environ.* **2012**, *122*, 2–10. [CrossRef]
17. Wulder, M.A.; White, J.C.; Loveland, T.R.; Woodcock, C.E.; Belward, A.S.; Cohen, W.B.; Fosnight, E.A.; Shaw, J.; Masek, J.G.; Roy, D.P. The global Landsat archive: Status, consolidation, and direction. *Remote Sens. Environ.* **2016**, *185*, 271–283. [CrossRef]
18. Cai, Y.; Zhang, H.; Chen, G.; Duan, L.; Zhang, Z.; Wang, S.; Li, H. The research status and problems on ecology and environment of lake Erhai. *Adv. Geosci.* **2013**, 241–252.
19. Yang, W.; Deng, D.; Zhang, S.; Hu, C. Seasonal dynamics of crustacean zooplankton community structure in Erhai Lake, a plateau lake, with reference to phytoplankton and environmental factors. *Chin. J. Oceanol. Limnol.* **2014**, *32*, 1074–1082. [CrossRef]
20. Matsushita, B.; Yang, W.; Yu, G.; Oyama, Y.; Yoshimura, K.; Fukushima, T. A hybrid algorithm for estimating the chlorophyll-a concentration across different trophic states in asian inland waters. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 28–37. [CrossRef]
21. Yu, G.; Yang, W.; Matsushita, B.; Li, R.; Oyama, Y.; Fukushima, T. Remote estimation of chlorophyll-a in inland waters by a NIR-red-based algorithm: Validation in asian lakes. *Remote Sens.* **2014**, *6*, 3492–3510. [CrossRef]

22. Zhu, Z. Change detection using Landsat time series: A review of frequencies, preprocessing, algorithms, and applications. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 370–384. [CrossRef]
23. Schmidt, G.; Jenkerson, C.; Masek, J.; Vermote, E.; Gao, F. *Landsat Ecosystem Disturbance Adaptive Processing System (Ledaps) Algorithm Description*; Open-File Report 2013–1057; U.S. Geological Survey: Reston, VA, USA, 2013; p. 27.
24. Zhu, Z.; Wang, S.; Woodcock, C.E. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and sentinel 2 images. *Remote Sens. Environ.* **2015**, *159*, 269–277. [CrossRef]
25. Kutser, T. Passive optical remote sensing of cyanobacteria and other intense phytoplankton blooms in coastal and inland waters. *Int. J. Remote Sens.* **2009**, *30*, 4401–4425. [CrossRef]
26. Kloiber, S.M.; Brezonik, P.L.; Olmanson, L.G.; Bauer, M.E. A procedure for regional lake water clarity assessment using Landsat multispectral data. *Remote Sens. Environ.* **2002**, *82*, 38–47. [CrossRef]
27. Moses, W.; Bowles, J.; Corson, M. Expected improvements in the quantitative remote sensing of optically complex waters with the use of an optically fast hyperspectral spectrometer—A modeling study. *Sensors* **2015**, *15*, 6152. [CrossRef] [PubMed]
28. Gitelson, A.A.; Yacobi, Y.Z.; Schalles, J.F.; Rundquist, D.C.; Han, L.; Stark, R.; Etzion, D. Remote estimation of phytoplankton density in productive waters. *Arch. Hydrobiol. Spec. Issues Adv. Limnol.* **2000**, *55*, 121–136.
29. Kutser, T. Quantitative detection of chlorophyll in cyanobacterial blooms by the satellite remote sensing. *Limnol. Oceanogr.* **2004**, *49*, 2179–2189. [CrossRef]
30. Odermatt, D.; Gitelson, A.; Brando, V.E.; Schaepman, M. Review of constituent retrieval in optically deep and complex waters from satellite imagery. *Remote Sens. Environ.* **2012**, *118*, 116–126. [CrossRef]
31. Hu, C. A novel ocean color index to detect floating algae in the global oceans. *Remote Sens. Environ.* **2009**, *113*, 2118–2129. [CrossRef]
32. Gitelson, A.A.; Dall'Olmo, G.; Moses, W.; Rundquist, D.C.; Barrow, T.; Fisher, T.R.; Gurlin, D.; Holz, J. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. *Remote Sens. Environ.* **2008**, *112*, 3582–3593. [CrossRef]
33. Camps-Valls, G.; Gómez-Chova, L.; Muñoz-Marí, J.; Vila-Francés, J.; Amorós-López, J.; Calpe-Maravilla, J. Retrieval of oceanic chlorophyll concentration with relevance vector machines. *Remote Sens. Environ.* **2006**, *105*, 23–33. [CrossRef]
34. Bricaud, A.; Claustre, H.; Ras, J.; Oubelkheir, K. Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations. *J. Geophys. Res.* **2004**, *109*, C11010. [CrossRef]
35. 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]
36. Falconer, I.; Bartram, J.; Chorus, I.; Kuiper-Goodman, T.; Utkilen, H.; Burch, M.; Codd, G.A. Safe levels and safe practices. In *Toxic Cyanobacteria in Water*; Chorus, I., Bartram, J., Eds.; E & FN Spon: London, UK, 1999; pp. 155–178.
37. Zhu, R.; Wang, H.; Yu, D.; Zeng, C.; Shen, H.; Chen, J. Dynamic changes of microcystins and phytoplankton during the cyanobacterial bloom in lake Erhai in 2013. *J. Lake Sci.* **2015**, *27*. [CrossRef]
38. Du, B. Study on entrophication of Erhai Lake. *J. Lake Sci.* **1992**, *4*, 86–92.
39. Han, T.; Peng, W.; Li, H.; Mao, Z. Evolution of eutrophication in the Erhai Lake and its relevant research progress. *J. China Inst. Water Resour. Hydropower Res.* **2005**, *3*, 71–78.
40. Yang, W.; Deng, D.; Zhang, S.; Xie, P.; Guo, L.; Wang, S. Seasonal dynamic and spatial distribution of chlorophyll-a concentration in lake Erhai. *J. Lake Sci.* **2012**, *24*, 858–864.
41. Du, B. Loss of biodiversity results in Erhai Lake eutrophication. *Yunnan Environ. Sci.* **1997**, *16*, 30–34.
42. Li, J. Research and countermeasures for Erhai Lake eutrophication. *J. Lake Sci.* **2001**, *2*. [CrossRef]
43. Wang, R.; Dearing, J.A.; Langdon, P.G.; Zhang, E.; Yang, X.; Dakos, V.; Scheffer, M. Flickering gives early warning signals of a critical transition to a eutrophic lake state. *Nature* **2012**, *492*, 419–422. [CrossRef] [PubMed]