An Optimal Two Bands Ratio Model to Monitor Chlorophyll-a in Urban Lake Using Landsat 8 Data

Qi Chen[1], Mutao Huang[1,*], Kaiyuan Bai[1], Xiaojuan Li[1]

[1] College of Hydropower and Information Engineering, Huazhong University of Science and Technology, NO. 1037, Luoyu Road, Wuhan 430074, China

Abstract. Chlorophyll-a (Chl-a) estimation in inland waters is an essential environmental issue. This study aimed to identify a band ratio model for Chl-a simulation using Landsat 8 OLI data and in situ Chl-a measuring in Lake Donghu. The band B1and B2, respectively at the wavelength of 443 nm and 483 nm, in the band ratio model [B1/B2] performed best in Chl-a estimation with the R^2 of 0.6215. K-means cluster analysis based on water quality indexes (Chl-a, pH, DO, TN, TP, COD, Turbidity) was conducted to further improve the accuracy of inversion model. The MAPE of the optimal [B1/B2] algorithm has decreased by 4.81% and 39.87% respectively for 17 December 2017 (R^2=0.7669, N=42) and 26 March 2018 (R^2=0.9156, N=45).

1 Introduction

Traditional water environment monitoring relies on a large amount of human labors by collecting in-situ water samples and conducting chemical analysis, and then the results of chemical analysis were compared with water quality standards to obtain the water quality status. Because of the high time and labor costs of data collection, the traditional water monitoring is always lag in time and limit in space. Therefore, as a complement to traditional monitoring, remote sensing method was applied for deriving water quality parameters.

Integrating in situ Chl-a concentration data with satellite remote sensing image (Klemas, 2013), and ensuring the continuity of the sampling program in time and space, simultaneously, has already been considered as an simple and feasible method for assessing water quality (Bukata, 2005; Palmer et al., 2015). MODIS Aqua/Terra has been widely applied to monitor Chl-a dynamics in coastal water with a relatively high accuracy, due to its large number of bands (from 412 to 869 nm) and low spatial resolution (250m). But for inland water, especially for some relatively small lakes and reservoirs (Lee et al. 2016), Landsat satellite data series are perfect for monitoring water quality and ecological change because of its relatively high resolution (30m) and long operating history (46 years). Landsat 8 Operational Land Imager (OLI) has optimized the spectral coverage and increased the number of bands (Pahlevan et al. 2014), which can provide higher data quality and richer water information (Concha and Schott 2016). For inland water quality evaluation, Landsat 8 data has always been used to monitor water quality parameters, such as TN and TP, SDD (Lee et al. 2016), Chl-a, colored dissolved organic matter (DOM) (Chen et al. 2017), TSM and dissolved organic carbon (DOC) (Alcântara et al. 2017). W. Tan (2017) have mapped the dynamic of Chl-a distribution in Lake Erhai from 1987 to 2016 using Landsat 5 and Landsat 7 images through an empirical model (log(Chl-a) = $1.5979+10.1431 \times (TM1^{-1}-TM2^{-1}) \times TM4$). S. Deyong (2015) has developed an algorithm to map the distribution of phycocyanin in Lake Dianchi by using Landsat series data (Landsat 4 TM, Landsat 5 TM, Landsat 7 ETM+ and Landsat 8 OLI). I. Ogashawara (2016) proposed a multi-band model to monitor the dynamic of algal blooms using the Landsat 8 OLI data. However, these Landsat 8 OLI data derived Chl-a estimation algorithms are empirical models and are not applicable to lakes in other region (Li et al., 2018).

Based on the background described above, this study aimed to establish a better correlation between the band reflectance and Chl-a concentration and improved the accuracy of the band ratio model. K-means cluster analysis was conducted to distinguish the pollution condition in Lake Donghu, and then to optimize the inversion model according to the classification result using a decision tree analysis, which should provide support for the management of water environment security in urban lake.

2 Study area and dataset

2.1. Study area and in situ experiment

---

* Corresponding author: huangmutao@hust.edu.cn

The average annual water level of the Lake Donghu has not changed much during these years with a value of 0.5 m. Generally, the lake water level begins to rise after March each year, and the relatively high water level period started from May to August. The water level drops after September, and the relatively low water level period started from September to March. The average annual water retention time of Lake Donghu is 0.44a and

(St. Louis, MO, USA). In short, water samples were filtered through 0.45-um Whatman cellulose acetate membranes and then immediately stored in liquid nitrogen. The filters were then soaked with acetone (90%) to extract the Chl-a pigment, and a centrifuge was used to increase the extraction efficiency. After storing at 0 °C for 24 h, the Chl-a was determined by measuring the extracted pigment samples.
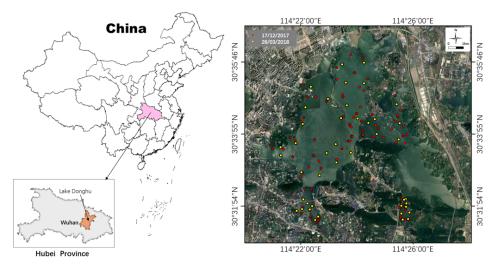

Fig. 1. Study area, sample dates and locations.

the water depth is about 2.2 m. The monthly average temperature distribution shows obvious characteristics of "cold in winter and warm in summer". The average temperature is 28.8°C in July and 4.5°C in January. The maximum water temperature is usually found in July and August, slightly exceeding 30°C; the minimum in February and is approximately 5 °C. The difference in temperature between the surface and the bottom layers of the water column is less than 20 °C. The characteristics of the inter-annual variability of temperature indicate the obvious continental climatic characteristics, that is, the average temperature gradually increases from January to July and decreases from August to December month by month.

Two field measurements was taken in 17 December 2017 (winter) and 26 March 2018 (spring), used for model calibration and validation, respectively. The sampling coordinates were displayed in Figure 1. Each water sample was mixed by three in situ replications and temporarily stored in a 500mL plastic bottle, which was rinsed with in situ water before sampling. The sampling program finished in a full day and after sampling work, all the 42 bottles were transferred to the laboratory, and then stored in a refrigerator at 0-4°C for further analysis. The analysis method is as follows, dissolved oxygen (DO), turbidity and water pH are obtained using a multi-parameter water quality analyzer (HI9829, Hanna, Italy). Secchi disk (SD), total nitrogen (TN), total phosphorus (TP) and Chemical oxygen demand (COD), were analyzed according to the standard Chinese method (A. P. H. 1989). The concentration of Chl-a (in mg·m^{-3}) was measured with an RF-5301 Fluorescent Spectro-photometer (Shimadzu, Kyoto, Japan), calibrated by the Chl-a standards manufactured by Sigma Chemical Co.

2.2 Landsat 8 OLI data processing

Two Landsat 8 scenes of path 123 and row 39, acquired on the same dates of water sampling (17 December 2017, 26 March 2018, and 26 October 2018) were used for comparing and validating. The images were processed using ENVI 5.3 software. For image pre-processing, the original DNs were converted into top-of-atmosphere (TOA) reflectances ρ_λ at wavelength λ by the following equation:

$$\rho_\lambda = \pi \cdot L_\lambda \cdot d^2 / (S_\lambda \cdot \sin\theta) \qquad (1)$$

Where L_λ is radiance (W/m^2 · sr · μm), d is Earth-Sun distance in astronomical units, S_λ is solar irradiance (W/m^2 · sr · μm), and θ is the sun elevation angle (°). The $\sin\theta$ is a correction based on the reflectance gains and offsets of the OLI sensor. An atmospheric correction was applied using 6S method to transform TOA-reflectance into the reflectance at water surface using an atmospheric model for the tropical zone.

Table 1. The band information of Landsat 8 OLI

Band details	abbreviation	Wavelength (nm)
Band 1 Coastal	B1	433–453
Band 2 Blue	B2	450–515
Band 3 Green	B3	525–600
Band 4 Red	B4	630–680
Band 5 NIR	B5	845–885
Band 6 SWIR 1	B6	1560–1651
Band 7 SWIR 2	B7	2100–2300

3 Method

3.1 K-means Clustering Analysis

The K-means clustering is a widely used partitioning clustering algorithm due to its simplicity and efficiency. Firstly, K objects were randomly selected as the initial cluster center, and then the distance between each object and each seed cluster center were calculated, and then each object was assigned to the cluster center closest to it. Once all objects have been assigned, the cluster center of each cluster will be recalculated based on the existing objects in the cluster until the square of the error is minimum and the cluster center is no longer change. The following objective function was used to caculate the minimal error (Qian et al., 2017):

$$J = \sum_{i=1}^{k} \sum_{j=1}^{n} \left\| x_j^i - c_i \right\|^2 \qquad (2)$$

Where x_j^i is the i-th cluster center and the j-th data point, respectively.

The K-means cluster analysis was conducted based on 42 in situ water parameters, including Chl-a, COD, DO, turbidity, pH, TN, TP, collected in 17 December 2017. Figure 2 gives the result of K-means cluster analysis that the whole lake area was divided into two classes, and which also displays water quality parameters of each sample point and its in situ Rrs value in the corresponding class in 17-th December 2017. A comprehensive description of these classes is provided below.

Fig. 2. Boxplot of different water quality parameters and value of reflectance spectra in 2017-12-17, respectively for Class I(a, c) and Class II(b, d)

Class I (Figure. 2a, 2c) is composed of the lower number of samples (N=17). This class displayed a higher level of Chl-a, COD, TP and TN (Figure. 2a). The mean value of Chl-a, COD, TP and TN is 159.23 mg·m⁻³, 36.22 mg/L, 0.30 mg/L, 4.05 mg/L and it was 9.36, 1.83, 2.71, 4.21 times of that in Classes II, respectively. The pH in Class I are above 7.0, which means the water in it is alkaline. In Figure. 2c, the Class I is characterized by a distinct comparable peak at B3 (0.525-0.600 µm). The reflectance trough at B2 (0.450-0.515 µm) is also well marked in this class.

Class II (Figure. 2b, 2d) contained 29 of the 42 samples. The pH level is lower than 7.0, which illustrate the water in Class II is acidic. The turbidity and DO is comparable to that of Class I. Class II has a similar characteristic of reflectance spectra as Class I, which also exhibited a distinct reflectance peak around B3 (0.525-0.600 µm) and a reflectance trough of B2 (0.450-0.515 µm). However, the mean value of reflectance peak around B3 (0.525-0.600 µm) and reflectance trough at B2 (0.450-0.515 µm) in Class II is 2.03 and 2.12 times higher than that of in Class I, respectively.

3.2 Band ratio algorism

As a comparison, Figure 3 shows the performances of these two band ratio models, using the same dataset in Lake Donghu (2017-12-17). When putting all the values of two classes together to simulate the Chl-a, algorithm

(a) perform better than the other three algorithms using the two band ratio (B1/B2). When seprately to simulate log(Chl-a) respectively for Class I and Class II, the algorithm (d) using B2/B3 is best for Class I (R^2=0.5079, n=13) and the algorithm (d) using B1/B3 is best for Class II (R^2=0.0855, n=29). Obviously, the result of separated liner algorithm for Class II is always unsatisfied, which has lower concentration of Chl-a.
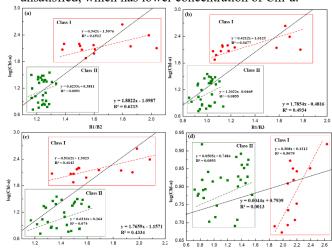

Fig 3. Relationship between different two bands ratio and chlorophyll-a concentration in 2017-12-17
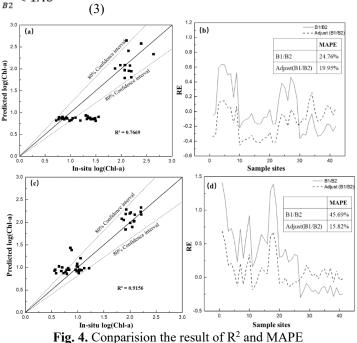
4 Result

4.1 Optimal Band ratio Chl-a model development

In this study, to improve the accuracy of model we introduced the optimization two band ratio algorithm based on (B1/B2) by doing a decision tree method shown in Equation (9). The initial separation was based on the value of Rrs (B1/B2), When Rrs (B1/B2) was greater than 1.45, profiles were well represented by Class I characteristic distribution and will there multiply the correction factor of 1.3. Correspondingly, Class II profiles occurred for the value of Rrs (B1/B2) < 1.45 and will there multiply the correction factor of 0.7.

$$\lg(Chl-a) = \begin{cases} 1.3 * \left(\frac{B1}{B2}\right), \frac{B1}{B2} > 1.45 \\ 0.7 * \left(\frac{B1}{B2}\right), \frac{B1}{B2} < 1.45 \end{cases} \quad (3)$$

4.2 Optimal Band ratio model validation

To test the accuracy of the calibrated model, the dataset of in-situ Chl-a concentration and Rrs of Landsat 8 OLI, respectively in 2017-12-17 and 2018-03-26 were used for algorithms validation, as described in Figure 4. 76.19% and 82.22% of in-situ log(Chl-a) values was in the range of 80% confidence interval of the predicted Chl-a concentration and the MAPE has decreased by 4.81% and 39.87%, respectively for 2017-12-17 and 2018-03-26, showing that optimization algorithm based on the decision tree indeed performs better than the previous two bands ratio Rrs(B1/B2) algorithm. The model is defined in Equation (9):

Fig. 4. Conparision the result of R^2 and MAPE

Conclusion

In this study, an optimal two band ratio algorithm was proposed for Chl-a estimation in Lake Donghu using Landsat 8 OLI data. Through a K-means cluster analysis, the optimized [B1/B2] model after a decision tree analysis performed best in Chl-a inversion. Superiority of the optimized [B1/B2] model was evidenced by its application to the two different stage in lake, and the predicted Chl-a concentration were strongly correlated with the in situ Chl-a concentration both in 17 December 2017 (R^2 = 0.7669; MAPE = 19.95%) and 26 March 2018 (R^2 = 0.9156; MAPE = 15.82%). Because the different optical properties in urban lakes, an important next step is to apply the optimized two band ratio model-[B1(coastal 433 nm)]/[B2(blue 450nm)] to other lake waters to test the stability of this model.

Acknowledge

References

1. V. Klemas, "Remote sensing of emergent and submerged wetlands: an overview," International INT. J. REMOTE SENS. **34** (2013)

2. R. P. Bukata, "Satellite Monitoring of Inland and Coastal Water Quality," Crc Taylor & Francis, (2005)

3. S. C. J. Palmer, T. Kutser, and P. D. Hunter, "Remote sensing of inland waters: Challenges, progress and future directions," Remote Sens. Environ. **157** (2015)

4. Z. Lee, S. Shang, Q. Lin, Y. Jing, and L. Gong, A semi-analytical scheme to estimate Secchi-disk

depth from Landsat-8 measurements. Remote Sens. Environ. **177** (2016)

5. N. Pahlevan, Z. Lee, J. Wei, C. B. Schaaf, J. R. Schott, and A. Berk, "On-orbit radiometric characterization of OLI (Landsat-8) for applications in aquatic remote sensing," Remote Sens. Environ. **154** (2014)

6. J. A. Concha and J. R. Schott, "Retrieval of color producing agents in Case 2 waters using Landsat 8," Remote Sens. Environ. **185** (2016)

7. C. Jiang, W. N. Zhu, Y. Q. Tian, and Y. Qian, "Estimation of Colored Dissolved Organic Matter From Landsat-8 Imagery for Complex Inland Water: Case Study of Lake Huron," IEEE T GEOSCI. REMOTE. **55** (2017)

8. Alcântara et al., "Modeling the spatio-temporal dissolved organic carbon concentration in Barra Bonita reservoir using OLI/Landsat-8 images," Model. Earth Syst. Environ. **3** (2017)

9. W. Tan, P. Liu, L. Yi, Y. Shao, and S. Feng, A 30-Year Assessment of Phytoplankton Blooms in Erhai Lake Using Landsat Imagery: 1987 to 2016, Remote Sens. **9** (2017)

10. S. Deyong, H. Chuanmin, Q. Zhongfeng, and S. Kun, "Estimating phycocyanin pigment concentration in productive inland waters using Landsat measurements: a case study in Lake Dianchi," Optics Express **23** (2015)

11. I. Ogashawara, L. Lin, and M. J. M. Madriñán, Slope algorithm to map algal blooms in inland waters for Landsat 8/Operational Land Imager images, J. APPL. REMOTE SENS. **11** (2016)

12. A. P. H. Association and A. W. W. Association, Standard methods for the examination of water and wastewater. American public health association, (1989)

13. W. Qian, W. Yi, R. Niu, and P. Ling, "Integration of Information Theory, K-Means Cluster Analysis and the Logistic Regression Model for Landslide Susceptibility Mapping in the Three Gorges Area, China," Remote Sens. **9** (2017)