

Genetic Algorithm – Back Propagation (GA-BP) Neural Network for Chlorophyll-a Concentration Inversion Using Landsat 8 OLI Data

Qi Chen¹, Mutao Huang^{1,*}, Ronghui Wang¹

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) accurate inversion in inland water is important for water environmental protection. In this study, we tested the Genetic Algorithm optimized Back Propagation (GA-BP) neural network model to precisely simulated the Chl-a in an inland lake using Landsat 8 OLI images. The result show that the R^2 of GA-BP neural network model has increased 28.17% compared to traditional BP neural network model. Then this GA-BP model was applied to another two scenes of Landsat 8 OLI image with the R^2 of 0.961, 0.954 respectively for March 26 2018, October 26 2018. And the spatial distribution have shown a reasonable result of Chl-a variation in Lake Donghu. This study can provide a new method for Chl-a concentration inversion in urban lakes and support water environment protection on a large scale.

1 Introduction

Water environment quality is increasingly a concerning problem in recent years (Hanson et al. 2016). Due to the long-term and frequent influence of human activities, many urban lakes are seriously troubled with eutrophication all the year round. The environment of eutrophication is suitable for algae grows, thus leading to water bloom and causing serious damage to the lake ecosystem. Many measurements have been taken in response to anthropogenic eutrophication by reducing the nutrient input and planting aquatic plants to improve the trophic status of freshwater ecosystems (Dörnhöfer and Oppelt 2016). Thus, the purpose of water environment monitoring programs are to predict phytoplankton population dynamic, and then, to prevent these problematic situations by taking early measures. 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 technology was applied for deriving water quality parameters.

Operational Land Imager (OLI) aboard Landsat 8 launched in February 2013 significantly improved in data quality and spectral coverage (Pahlevan et al. 2014) and can provide more accurate water quality information (Concha and Schott 2016). For inland water quality evaluation, Landsat 8 data has been widely used to estimate water quality parameters, such as SDD (Lee et al. 2016), Chl-a (Tan et al. 2017), TSM, colored dissolved organic matter (DOM), dissolved organic carbon concentration (Alcântara et al. 2017), and TN and TP (Li et al. 2017).

In recent years, data-driven technology is taking the emerging computing intelligence as the main development direction, such as artificial neural network, genetic computing and fuzzy systems. They simulate human intelligent activities from different perspectives and provide new insights into the evolution mechanism of ecosystems.

However, the traditional neural network uses the gradient method to iteratively optimize the initial weights and thresholds, and is easy to fall into the local optimal solution. In this study, the genetic algorithm's global search ability is used to optimize the initial weight and threshold to improve this shortcoming. On the other hand, in order to improve the prediction accuracy, the fuzzy pattern recognition technology is used to select the training samples that match the test samples for training to avoid interference with unmatched samples. Based on the measured water quality data of Lake Donghu, the Genetic Algorithm – Back Propagation (GA-BP) neural network Chl-a prediction model was constructed and the results were compared and analyzed in this study. This can provide a theoretical basis for the further use of water informatics of aquatic environment in inland waters.

2 Study area and dataset

2.1. Study area and in situ experiment

* Corresponding author: huangmutao@hust.edu.cn

Lake Donghu (30° 13' 30" N, 114° 12' 30" E) is located in the northeast part of Wuhan City, the capital of Hubei Province, China. It has a surface area of 32 km², a volume of 62 million m³. Since the late 1960s, the Lake Donghu was artificially divided into several sub-lakes, including Shuiguo Lake, Tangling Lake, Guozhen Lake, and Houhu Lake, Niuchaohu Lake, Miaohu Lake, Yujia Lake, etc., while Guozheng Hu (with a surface area of 10.0 km²) is the main study area. At present, only the Shuiguo Lake and Guozheng Lake water surface are connected together, and the rest of the lake area only has bridge holes, gate holes or culverts. This kind of artificial lake segmentation directly leads to poor water flow blockage of water mass exchange, and formation of local "dead water" in the Lake Donghu.

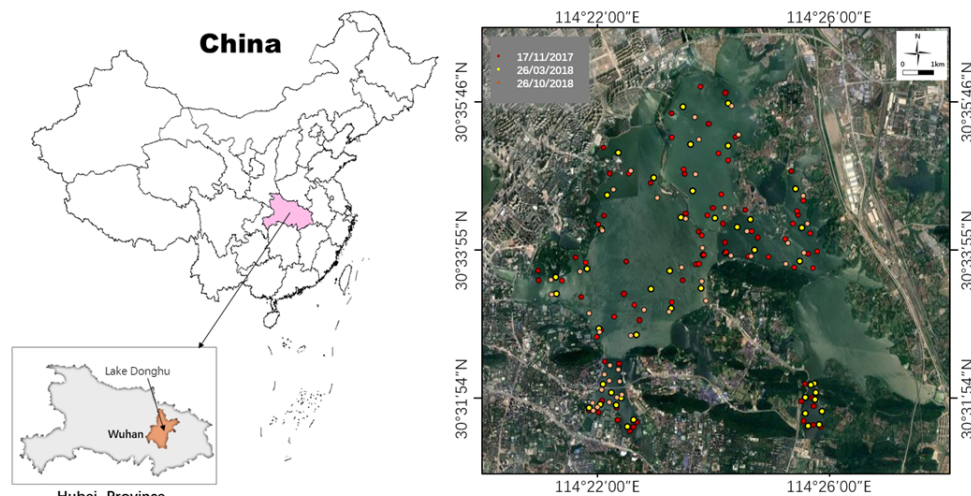


Fig. 1. Study area and sampling sites

Three field surveys were conducted in 17 December 2017, 6 March 2018 and 26 October 2018, representing the different season conditions of Lake Donghu, respectively. The in situ stations were annotated in Figure 1. The mixed water samples from three replications in each sampling site were collected in 500 mL plastic bottles, which were rinsed with surface water before sampling. The concentration of Chl-a (in mg·m⁻³) was measured using an RF-5301 Fluorescent Spectrophotometer (Shimadzu, Kyoto, Japan), calibrated by the Chl-a standards manufactured by Sigma Chemical Co. (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, Chl-a was determined by measuring the extracted pigment samples.

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 November 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 DN_s were converted into top-of-atmosphere (TOA) reflectances ρ_λ at wavelength λ by the following equation:

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

Where L_λ is radiance (W/m² · sr · μm), d is Earth-Sun distance in astronomical units, S_λ is solar irradiance (W/m² · 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. An aerosol

model for the tropospheric layer, and over-water retrieval standard for identifying dark water pixels by the reflectance ratio of NIR band and SWIR band 7 at 2200 nm proven to be a highly accurate method for lakes in tropical regions (Ha et al. 2017).

3 Model method

3.1 BP neural network

To capture nonlinear relationships among Chl-a and Landsat 8 OLI reflectances in a specific water system remains a technical challenge due to the complex physical, chemical, and biological processes involved. Moreover, to accurately simulate the upwelling process requires a significant amount of field data to support the analysis. Because of its strong learning ability and self-adaption, BP neural network is capable of simulating complex nonlinear systems, which is superior to linear regression models. Therefore, BP neural network is proposed to model the correlation between chl-a concentration and the Landsat 8 OLI reflectances. BP neural network architecture consists of two or more layers of neurons connected by weights. The information is captured by the network when input data pass through the hidden layer of neurons to the output layer. In our

research, we set the number of hidden layers and hidden nodes of every layer by the repeated tests. In order to ensure that the prediction results are stable and reliable, overmuch hidden nodes were avoided. This setting principle also helps to generate quick prediction results.

3.2 GA-BP neural network

The GA-optimized BP neural network mainly consists of three parts: (1) BP neural network topology determination, genetic algorithm optimization homing, and BP neural network training prediction. The BP neural network topology is determined according to the number of input and output parameters in actual cases, and then the genetic algorithm individual (chromosome) length is determined; (2) initialize BP neural network parameters, use GA optimization to obtain weights and thresholds. And at this time, each chromosome in the population contains all the weights and thresholds of the network, and the individual calculates its adaptation through fitness function, and the genetic algorithm finds the optimal fitness value through selection, intersection and mutation operation; (3) BP neural network uses the optimal individual to assign the network weight and threshold, trains the network, and finally outputs the prediction result. The corresponding algorithm flow diagram is shown in Figure 2.

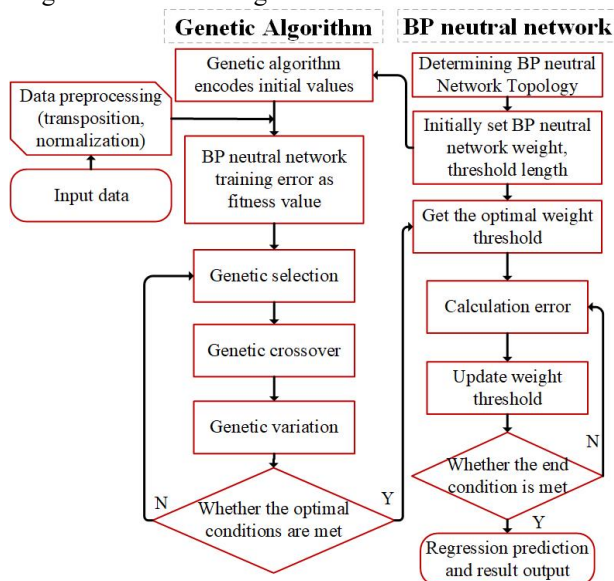


Fig. 2. Algorithm structure diagram

4 Result

4.1 GA-BP model validation

Table 1 shows the statistical results of the correlation coefficient of training, testing, verification during the simulation of the concentration of Chl-a in the Lake Donghu. The GA algorithm ends in the 50th generation,

and the fitness error value gradually decreases, while the corresponding optimal individual fitness value is 2.47. The model achieves the best fitting effect at the 15th iteration, and the corresponding minimum mean square error is 0.055, while the model total correlation coefficient R value reaches 0.869, and the simulation effect is better.

4.2 Comparison with non-optimal BP model

In order to verify the improvement effect of the optimization model, the traditional BP neural network model was used to simulate the Chl-a concentration in the Lake Donghu under the same parameter setting and the same sample setting conditions, so as to form a vertical comparison with the optimized model. From the statistical results shown in Table 1., the total correlation coefficient R^2 of the traditional BP model is 0.678, and the correlation coefficient of the model training samples is only 0.594, and the simulation results have strong uncertainty. Comprehensively, the BP neural network model with GA optimization has higher precision, which reduces the uncertainty of the results of artificial participation in model parameter adjustment. The resulting simulation results are more reliable.

Table 1. The Comparison of model Correlation coefficient R^2

Dataset	model	Training	Test	All
2017/11/17	BP	0.594	0.901	0.678
2017/11/17	GA-BP	0.938	0.769	0.869
2018/03/26	GA-BP	0.998	0.821	0.961
2018/10/26	GA-BP	0.991	0.999	0.954

4.3 GA-BP model application to Landsat 8 OLI images

From the statistical results, the correlation coefficient R between the simulated and measured values of the inversion model in March and October is 0.961 and 0.954, respectively. From the results of inversion distribution 2017-2018, the overall chlorophyll-a concentration has a downward trend, and the lake eutrophication condition has improved, which may attributed to the decreasing of sewage outlet numbers around Donghu.

In summary, the BP neural network regression model based on genetic algorithm optimization can be used to invert the Chl-a concentration in the Lake Donghu. After the verification of the training statistics and the theoretical verification of the actual working conditions, the simulation results have higher precision and the simulation result spatial region. The distribution of Chl-a in Lake Donghu is reasonable, with strong time and space adaptability.

5 Conclusion

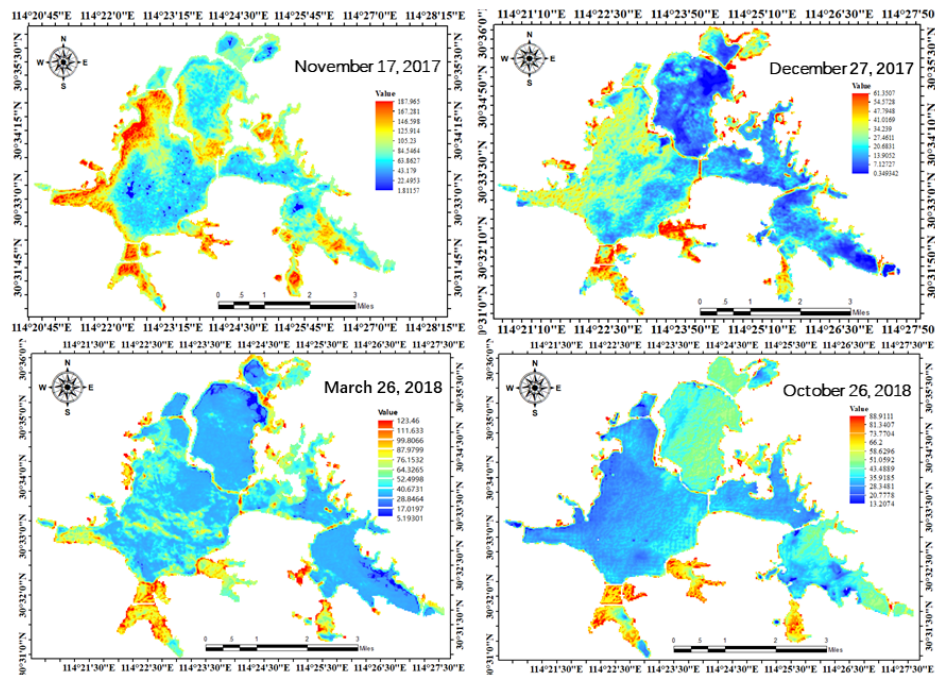


Fig. 3. The distribution of GA-BP model simulated Chl-a in November 17 2017, December 27 2017, March 26 2018 and October 26 2018

In this study, we pre-processed the water remote sensing images and obtained the spectral reflectance values of the seven bands of Landsat 8 OLI, then established the correspondence relationship between in situ measured Chl-a concentration values and spectral values.

The theoretical application and implementation process of BP neural network regression algorithm based on genetic algorithm optimization are introduced. On this basis of this GA-BP algorithm, the remote sensing inversion model of Donghu is established, and the measured data of Chl-a and the spectral value of water body are used as training and test samples for inversion. After the model is trained and validated, four times in situ measured Chl-a data are divided into three phases to verify the statistical results of the model and the adaptive verification in different time ranges. The results show the correlation coefficient between the simulated and actual values of the three-phase model. R is 0.869, 0.961 and 0.954, respectively.

Based on the actual environment of Lake Donghu, the spatial region analysis of the GIS thematic map generated by the inversion model is carried out. The results show that the simulated distribution of Chl-a concentration is consistent with the actual situation, and the rationality of the GA-BP inversion model is demonstrated from the perspective of Chl-a spatial region distribution.

Acknowledge

This work was financially supported by the National Natural Science Foundation of China under Grant No. 51579108, the National Key R&D Program of China under Grant No.2017YFC0405900, the Fundamental Research Funds for the Central Universities under Grant No. 2017KFYXJJ203, and the Graduates' Innovation

Fund of Huazhong University of Science and Technology under Grant No. 2019YGCSCXY004.

References

1. Alcântara, Enner, Bernardo, Nariane, Rodrigues, Thanan, Watanabe, & Fernanda. Modeling the spatio-temporal dissolved organic carbon concentration in Barra Bonita reservoir using OLI/Landsat-8 images. *Model. Earth Syst. Environ.* **3** (2017)
2. Concha, J.A., & Schott, J.R. Retrieval of color producing agents in Case 2 waters using Landsat 8. *Remote Sens. Environ.* **185** (2016)
3. Dörnhöfer, K., & Oppelt, N. Remote sensing for lake research and monitoring – Recent advances. *Ecol. Indic.* **64** (2016)
4. Hanson, P.C., Weathers, K.C., & Kratz, T.K. Networked lake science: how the Global Lake Ecological Observatory Network (GLEON) works to understand, predict, and communicate lake ecosystem response to global change. *Inland Waters* **6** (2016)
5. Lee, Z., Shang, S., Lin, Q., Jing, Y., & Gong, L. A semi-analytical scheme to estimate Secchi-disk depth from Landsat-8 measurements. *Remote Sens. Environ.* **177** (2016)
6. Li, Y., Zhang, Y., Shi, K., Zhou, Y., Zhang, Y., Liu, X., & Guo, Y.. Spatiotemporal dynamics of chlorophyll-a in a large reservoir as derived from Landsat 8 OLI data: understanding its driving and restrictive factors. *Environ Sci Pollut Res Int.* **25** (2018)

7. 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.* **154** (2014)
8. Tan, W., Liu, P., Yi, L., Shao, Y., & Feng, S. A 30-Year Assessment of Phytoplankton Blooms in Erhai Lake Using Landsat Imagery: 1987 to 2016. *Remote Sens.* **9** (2017)