Simulation System of Lake Eutrophication Evolution based on RS & GIS Technology—A Case Study in Wuhan East Lake

Xin Yang, Mutao Huang* and Kaiyuan Bai
School of Hydropower & Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

*Corresponding author

Abstract. Water is the source of life and monitoring water quality is very important. Traditional water quality monitoring often uses artificial methods, taking water samples on the spot, and then analysing them in the laboratory. This method is time-consuming and labour-intensive, and is affected by time, weather and human activities. It is not only inaccurate, but also difficult to multi-period. It is difficult to monitor large areas of water. This study proposes an intelligent water quality monitoring system—Simulation System of Lake Eutrophication Evolution based on RS & GIS Technology (SSLEE). Based on the .NET Framework, the system interface design is completed through Windows Form platform. C#, IDL and ArcGIS Engine are used to develop the whole system. The system can manage the measured data, remote sensing image data, pre-process the remote sensing image and retrieve the chlorophyll a concentration. Taking Wuhan East Lake as an example, the applicability of SSLEE system is demonstrated. Compared with the measured results, the error is small, indicating that the system is reliable in water quality monitoring of inland lakes.

Keywords: Monitor; Water quality; Lake; RS; GIS; Simulation system.

1. Introduction

Lake is one of the important components of global water resources. There are many lakes in China. In the monitoring and evaluation results of important lakes in China, we can find that the overall water quality of lakes in China is not optimistic, and the level of lake eutrophication is gradually rising [1]. Chl-a is an important components of algae. At the same time, its concentration can be used to estimate the biomass and productivity of phytoplankton, and it is an important parameter to reflect the degree of water eutrophication. To some extent, the content of chlorophyll a can reflect the water quality [2].

In the past, the water quality monitoring often used manual field sampling, and then the water samples were sent to the laboratory for testing. However, due to the time difference between analysis and sampling, the test results are generally inconsistent with the actual situation. Because of the limitation of human activities, it is difficult to monitor the water area of many periods and large areas, which brings a lot of inconvenience for the analysis of water quality in the study area. The characteristics of satellite remote sensing make it possible to carry out real-time monitoring and analysis of lake water quality information. Through remote sensing(RS) monitoring, we can effectively grasp the overall situation of water ecological environment across the constraints of terrain, time and other factors, and provide effective and reliable analysis data for the study of the future situation of water quality. Nowadays, with the continuous deepening of RS image application, the continuous improvement of satellite imaging equipment accuracy, and the continuous promotion of water quality inversion algorithm research, the...
generation of RS image data has become the main data source of water quality analysis, which brings great convenience to water environment protection and monitoring. In recent years, more and more scholars have carried out in-depth analysis and research on the application of RS and GIS technology in water quality monitoring, and achieved some results. However, the mature RS monitoring system for lake water quality is still under development. Most of the existing RS image processing systems and geographic information processing systems are not professional platforms for water pollution monitoring and evaluation, such as ENVI, ArcGIS, etc. These platforms have many problems, such as high requirements for operators, cumbersome operation, high technical threshold, not suitable for fast operation and mismatching of conventional monitoring data. Although the quantitative research of RS monitoring for water quality has developed to a certain extent, the integrated and systematic research of RS dynamic monitoring for water pollution is still lacking. There are still many problems in the application of existing professional RS software system in water pollution monitoring, so there is a lack of comprehensive and mature monitoring system for water pollution. From the perspective of integration of RS and GIS, this paper focuses on the design and implementation of a high automation, high integration, strong business, complete function, simple and easy-to-use visual RS monitoring system for water pollution to achieve dynamic monitoring of water pollution indicators. Based on the .NET Framework, the system interface design is completed through Windows Form platform. C#, IDL and ArcGIS Engine are used to develop the whole system. The main task of the system is to retrieve the chlorophyll a concentration of water, including the radiometric calibration of Landsat 8 RS image and atmospheric correction, and the whole process is completed by combining the lake chlorophyll inversion model. It also provides the image boundary extraction based on vector clipping, which is convenient to display the distribution of water quality in the study area.

2. Model
The core of using RS technology to retrieve the chlorophyll concentration of water is to establish the quantitative mathematical relationship between lake water spectral reflectance and chlorophyll a concentration. Taking Landsat 8 as the RS data source, 35 groups of measured sample data on March 23, 2018 were used as the training data of model, and 25 groups of measured sample data on December 20, 2017 were selected as the model test samples. Comparing the accuracy of the two models (neural network model and support vector machine model), the optimal model is determined to provide data support for the subsequent feature extraction of chlorophyll a changes.

2.1. Neural Network Model
This study uses a three-layer neural network model. It includes input layer, hidden layer and output layer. The input layer is composed of seven sub nodes, each of which represents seven reflectance values corresponding to Landsat 8 image data. The output layer is a single neuron, which represents the chlorophyll a concentration to be retrieved. The number of hidden layers can be determined according to formula (1), and then the optimal number of nodes can be found according to one by one optimization verification method. In the formula below, In the formula below, 6 hidden layer nodes are selected.

\[ m = \sqrt{j + k + a} \]  

(1)

In the formula: m is the number of hidden layer nodes, j is the number of input neurons, k is the number of output layer neurons, and a is a constant of 0-10.

By analyzing the accuracy of the inversion value of chlorophyll a concentration and the measured value, the root mean squared error (RMSE) of the two values is 32.7374, and the coefficient of determination \( R^2 \) is 0.7419, which shows that there is a significant correlation between them. Figure 1 shows the test results of BP neural network model of 25 test samples.
Figure 1. Comparison of BP neural network inversion and measured results. Model validation. Using the RS images of East Lake in December 2017 and March 2018, the concentration distribution of East Lake was obtained through the neural network model and interpolation analysis (Figure 2a, 2b).

Figure 2a. Distribution of chlorophyll a in East Lake in December 2017. Figure 2b. Distribution of chlorophyll a in East Lake in March 2018.

2.2. Support Vector Machine Model (SVM)
SVM is a tool to solve machine learning problems based on the principle of statistics and structural risk minimization. In this simulation, radial basis function (RBF) is selected as the kernel function of SVM, and grid search and cross validation method are used to determine the parameter $C$ of 0.0385 and $\sigma$ of 1.8661. By analyzing the precision of the inversion value of chlorophyll a concentration and the measured value, the RMSE of the two values is 31.2574, and $R^2$ is 0.7566 respectively, indicating that they have significant correlation. Figure 3 shows the test results of 25 samples.
Figure 3. Comparison of SVM inversion and measured results.
Model validation use the RS images of East Lake in December 2017 and March 2018, and the concentration of chlorophyll a distribution in the East Lake was obtained through SVM model and interpolation analysis. (Figure 4a, 4b).

2.3. Model Comparison
The RMSE of the two models are 32.7374 and 31.2574 respectively, and the $R^2$ are 0.7419 and 0.7566 respectively. The accuracy of SVM model is a little higher.

3. System Introduction
The system consists of four functional modules: integrated management of multiple data, RS image processing, RS inversion of lake water quality and comprehensive assessment of lake eutrophication. Lake water quality inversion is the core function module of the system, which can retrieve the water quality distribution of the whole research area through the measured data. This paper only introduces this module. Compared with the RMSE and $R^2$ of the neural network model and SVM model, the accuracy of the SVM model is slightly higher. This system takes chlorophyll a as the water quality evaluation parameter, selects the SVM inversion model, selects the measured data and the algorithm training model, then establishes the relationship between chlorophyll a and the band reflectivity, thus inverts the water quality information of the selected period. The inversion algorithm of this system is based on C#, and the inversion concentration distribution map is produced by the interpolation algorithm provided by ArcGIS Engine and. Net control. The specific flow of system implementation is shown in Figure 5, and the training results of SVM algorithm are shown in Figure 6.
4. Case Study

4.1. Research Area
East Lake is located in the central city of Wuhan, covering an area of 33 square kilometers. In this paper, East Lake is taken as the research area, and the water quality of the lake is simulated by using the system.

4.2. Display Simulation Results
First, enter the system, select the training data and band, set the Gaussian kernel $\sigma$, influence factor and tolerance for model training, and then select the existing RS image data to generate a grid map reflecting the concentration of chlorophyll a in the East Lake through the trained model, as shown in Figure 7.
5. Conclusion
In this study, we illustrated a visual modelling system, named Simulation System of Lake Eutrophication Evolution based on RS & GIS Technology. In view of the difficulties in financial and physical resources in obtaining a large number of measured data simultaneously, it is feasible to use mature satellite RS data of the same kind to retrieve the lake chlorophyll a concentration. The decision coefficient $R^2$ of the model based on Landsat 8 image data is 0.7566, and the root mean square error (RMSE) is 31.2574. Therefore, the support vector machine model can be used for RS inversion of chlorophyll a concentration in East Lake water with good results. However, the accuracy of the model still needs to be improved. The influence of spectral resolution of image data and atmospheric correction effect on the selection of parameters in the inversion model should be considered comprehensively.

Acknowledgments
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.

References
[1] Yumin Yang, Jun Zhao, Xiaoke Wang, Shaodan Zhang. Research Progress of Monitoring Chlorophyll-a Content in Water Using Remote Sensing [J]. Anhui Chemical Industry, 2018, 44(03): 1-3.
[2] Xiaochun Cai, Wu Hu. Retrieval of Chlorophyll-a Concentration Based on Spectral Unmixing Model with Remote Sensing Data—a Case Study in Lake Dianshan [C]. System Simulation Technology & Application, 2009, 11:730-733.
[3] Wanhua Gu. Remote Sensing Based Estimation of Chlorophyll-a Concentration in DianShanHu Lake [D]. Fudan University, 2011.
[4] Biyun Zhao, Bin He, Yunyan Zhu, Guolin Yuan. Study on the Remote Sensing Model for Chlorophyll-a in Dianchi Lake[J]. Yunnan Environmental Science, 2001(03):1-3.
[5] Pulliainen J, Kallio K, Eloheimo K, et al. A semi-operative approach to lake water quality retrieval from remote sensing data [J]. Science of the Total Environment, 2001, 268(1-3): 79-93.
[6] Banglong Pan, Huiyan Shen, Hui Shao, Wei Hua. Combined Inversion of Hyper–Spectral Remote Sensing of Space and spectrum for Lake Chlorophyll[J]. Journal of Atmospheric and Environmental Optics, 2017, 12(06):429-434.
[7] Benesty J, Chen J, Huang Y, et al. Pearson correlation coefficient[M]. Noise reduction in speech processing. Springer, Berlin, Heidelberg, 2009: 1-4.