Environ Sci Pollut Res (2017) 24:5335–5348
DOI 10.1007/s11356-016-8155-2

RESEARCH ARTICLE

Dynamic monitoring and prediction of Dianchi Lake cyanobacteria outbreaks in the context of rapid urbanization

Yi Luo[1,2,3] · Kun Yang[1,2] · Zhenyu Yu[1] · Junyi Chen[3] · Yufei Xu[1] · Xiaolu Zhou[4] · Yang Yang[1]

Received: 4 August 2016 / Accepted: 25 November 2016 / Published online: 24 December 2016

Abstract Water crises have been among the most serious environmental problems worldwide since the twenty-first century. A water crisis is marked by a severe shortage of water resources and deteriorating water quality. As an important component of water resources, lake water quality has deteriorated rapidly in the context of fast urbanization and climate change. This deterioration has altered the water ecosystem structure and influenced lake functionality. To curb these trends, various strategies and procedures have been used in many urban lakes. Among these procedures, accurate and responsive water environment monitoring is the basis of the forecasting and prevention of large-scale cyanobacteria outbreaks and improvement of water quality. To dynamically monitor and predict the outbreak of cyanobacteria in Dianchi Lake, in this study, wireless sensors networks (WSNs) and the geographic information system (GIS) are used to monitor water quality at the macro-scale and meso-scale. Historical, real-time water quality and weather condition data were collected, and a combination prediction model (adaptive grey model (AGM) and back propagation artificial neural network (BPANN)) was proposed. The correlation coefficient (R) of the simulation experiment reached 0.995. Moreover, we conducted an empirical experiment in Dianchi Lake, Yunnan, China using the proposed method. R was 0.93, and the predicting error was 4.77. The results of the experiment suggest that our model has good performance for water quality prediction and can forecast cyanobacteria outbreaks. This system provides responsive forewarning and data support for lake protection and pollution control.

Keywords Urbanization · Cyanobacteria · Wireless sensor networks · GIS · Grey model · Artificial neural network

Responsible editor: Suresh Pillai

✉ Kun Yang
kmdcynu@163.com

[1] School of Information Science and Technology, Yunnan Normal University, Yunnan 650500, China

[2] GIS Technology Research Center of Resource and Environment in Western China, Ministry of Education, Yunnan Normal University, Yunnan 650500, China

[3] School of Tourism and Geographical Science, Yunnan Normal University, Yunnan 650500, China

[4] Department of Geology and Geography, Georgia Southern University, Statesboro, GA, USA

Introduction

With the rapid development of urbanization, continuous expansion of the human footprint, and negative influence of global warming, environmental problems of inland lakes have become more serious and complex. For instance, many lakes face eutrophication, suffer from both point and non-point pollution, and are exposed to both endogenous and exogenous pollutions. The deterioration of water quality has severely hindered the sustainable social and economic development of nearby cities (Xin et al. 2014; Xun et al. 2012; Ren et al. 2014).

As the "eyes" of water protection and management, water quality monitoring is the basis for forecasting cyanobacteria outbreak and the assessment of bloom intensity. As technology advances, the monitoring tools and indicators being measured become increasingly diverse, and the measurement accuracy is also constantly improving (Matthews and Odermatt 2015; Lunetta et al. 2015). The current methods used to monitor inland water quality can be divided into two categories, indirect and direct monitoring. Representative indirect monitoring methods are based on remote sensing. Remote sensing-based approaches

covering large study areas, are cost-effective, and have become a powerful research tool in lake studies. Studies using remote sensing have evolved from traditional water body detection to water quality monitoring and prediction (Stumpf R P et al. 2016). With the development of sensors, the types of substances being monitored and the overall accuracy have also increased, such as the chlorophyll concentration, concentration of suspended solids, total nitrogen, and total phosphorus (Wu Shengli et al. 2009; Sun LY et al. 2010; Klemas V 2012; Olmos and Birch 2010). However, there are still several challenges.

1. The update cycle for remotely sensed imageries is too long to reflect cyanobacteria outbreak in a timely manner.
2. Reliable image retrieval is subject to weather conditions (e.g., clouds and rain). The derived water quality indices are not consistent because of meteorological factors.
3. A large number of research results show that the water quality parameters (such as chlorophyll a, dissolved oxygen, and turbidity) and meteorological conditions (such as air temperature, illumination intensity, and wind speed) are the primary causes of cyanobacteria outbreak (Zou Zhiqiang et al. 2012; Lunetta et al. 2015; Qin Boqiang et al. 2014). Although water quality parameters and distribution can be acquired using remote sensing, the meteorological conditions around the lake cannot be determined in this way.

With the development of sensor design, wireless communication, and microelectromechanical systems (MEMS), direct monitoring using wireless sensors networks (WSNs) has become increasingly popular. Monitoring nodes in WSNs have multiple features, such as low cost, low power, and self-organization (Zhu Haiyang et al. 2014; Yu and Naiming 2013; Zhu Yonghong et al. 2015; Yi et al. 2014a, b). These features make wide-area, long-term, and unattended monitoring possible (Gong Peng 2007). WSNs, as part of "earth observation systems", have attracted increasing attention. WSNs are important components of the future development of environmental monitoring around the world. They have been applied in many areas, such as atmospheric environmental monitoring, soil environmental monitoring, mine environmental monitoring, and geological disaster monitoring (Evans J et al. 2008; Jin Rui et al. 2012). In the domain of water quality monitoring, the combination of position sensors and WSNs enables high-density, high-precision, and continuous observations. In addition, such methods can simultaneously collect both water quality data and surrounding environmental data, which can be used to model the interaction between the internal and external factors of water pollution and the mechanism of the pollution process.

Because of the advantages of sensor networks, this study uses multi-source heterogeneous water environment sensors to achieve high-precision and continuous water quality monitoring. Based on the collected data, we used the adaptive grey model (AGM) and back propagation artificial neural network (BPANN) to predict the outbreak of cyanobacteria. The study introduces an innovative approach to monitor water quality and to predict cyanobacteria outbreak.

Study area

Yunnan Province is located in the southwest part of China (21° to 29°′N, 97° to 106 ′E); the total land mass covers approximately 394,000 km^2 and accounts for 4.11% of the country's land mass. Dianchi Lake is located in Kunming City, which is the capital city of Yunnan Province. It is the largest freshwater lake in Yunnan Province. Because of the environmental and aesthetic importance, Dianchi Lake is known as the "pearl of the plateau". Figure 1 shows the research area location. Rapid urbanization, global warming, and other natural and human factors have largely deteriorated the water quality in Dianchi Lake. It has been exposed to multiple types of pollution, including point and non-point source pollutions, as well as endogenous and exogenous pollutions. The water quality has changed from Grade II to worse than Grade V (based on Chinese water quality standards) in just 50 years. Plagued by severe eutrophication and frequent outbreaks of cyanobacteria bloom, Dianchi Lake is currently one of the most polluted lakes in China (Genbao L I et al. 2014; Yong L et al. 2012).

The main cause of watershed environment deterioration is that the formation mechanism of urban lake pollution has not been efficiently addressed. Effective methods for controlling pollution have not been fully developed. The focus has been on industrial point pollution and agricultural non-point pollution, ignoring urban non-point pollution due to rapid urbanization. A full understanding of the changes to the natural environment and human environment is the basis for controlling lake pollution and improving the lake environment. In this study, we discuss the natural environment and human environment change in the Dianchi watershed using data from the China Statistical Yearbook and the Yunnan Province Statistical Yearbook.

Natural environment analysis

The landforms of the research area

Dianchi Lake, an urban lake, is downstream of Kunming City and is the lowest place in the Dianchi Basin. The basin area below 2–35° accounts for 87.6% of the total area. The minimum distance between Dianchi Lake and the Dianchi watershed boundary is approximately 600 m. Eventually, these features will lead to Dianchi Lake being the ultimate receiving water body of all the basin water containing pollution (Yao Xu et al. 2010; Guolin and Bin 2008; Yu and Naiming 2013). Figure 2 shows the slope and the aspect of the Dianchi watershed.

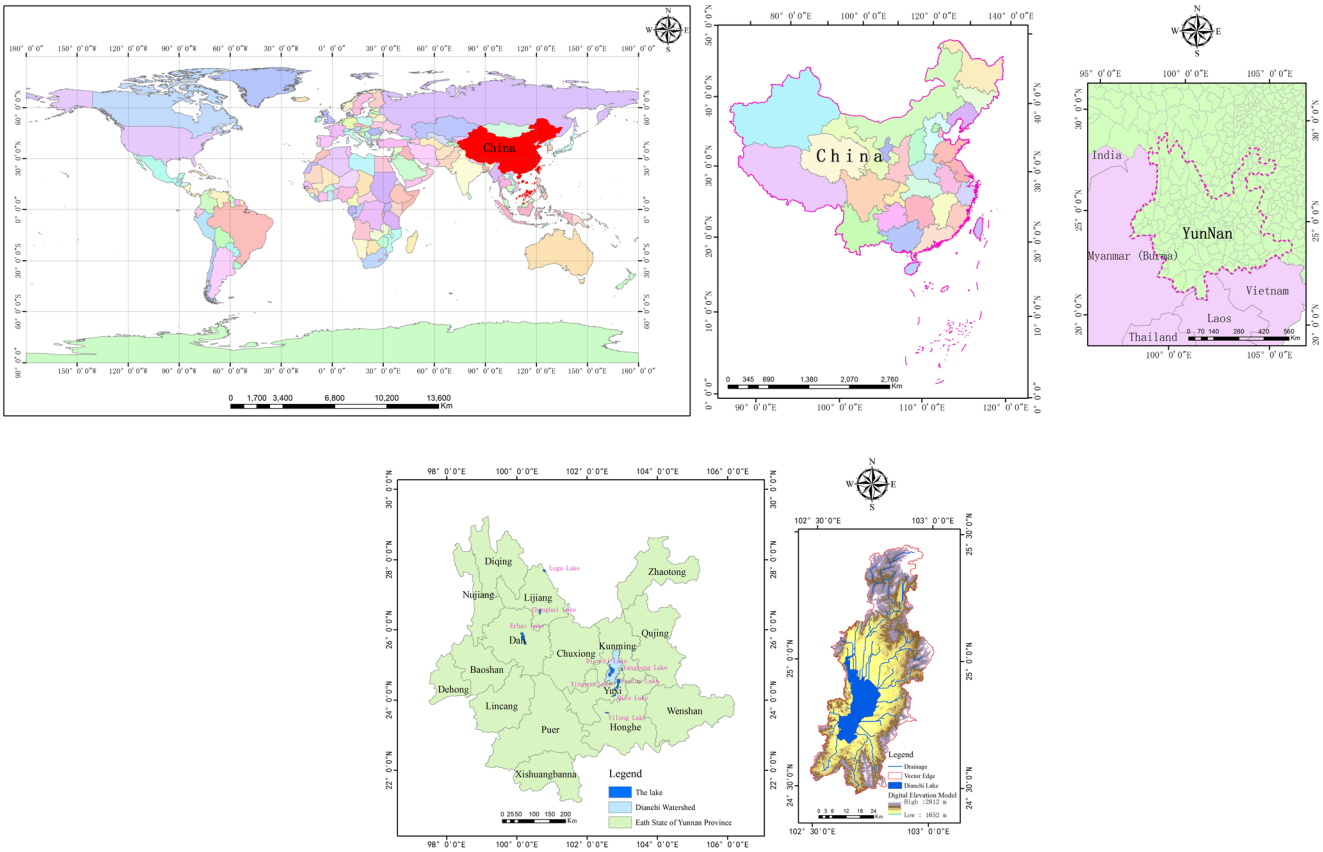


Fig. 1 The research area locations

Air temperature change analysis

In the past 60 years, the surface air temperature around the world increased in a linear trend by approximately 0.74 °C (IPCC 2007; Karl T R et al. 2015). Under these circumstances, the amount and pattern of rainfall has changed. Figure 3a shows the surface air temperature trend of Kunming during the last 25 years.

The results show that the average surface air temperature was approximately 15.9 °C during 1990–2014, and the last 10 years rank among the ten warmest years for Kunming since 2005. The overall trend in the average air temperature change is increasing at a rate of 0.391 °C/10, and the trend is significant ($p < 0.01$). The variation tendency of air temperature in Kunming is consistent with that around the world. In addition, oscillation in the air temperature trend can be observed. The temperature oscillates between ±2 °C, but we do not know the cause and rule.

Relative humidity change analysis

The average relative humidity during 1990–2014 was 68.9% RH in Kunming. Figure 3b shows the relative humidity change trend of Kunming over the last 25 years. The overall trend in the average relative humidity change is a significant decrease ($p < 0.01$). The phenomenon matches the objective law of atmospheric sciences. Similarly, oscillation in the relative humidity trend is obvious (±13% RH). The air temperature and relative humidity change rates are strongly related, with an obvious lag.

Rainfall change analysis

The average total rainfall during 1990–2014 was approximately 977.4 mm. The overall trend in the total rainfall is a significant decrease ($p < 0.01$). Figure 3c shows the analysis results. The rainfall in the last 10 years was the lowest for Kunming since 1990, impacting the on water supply to Dianchi Lake and resulting in a longer water renewal cycle. Oscillation in the rainfall trend is obvious.

Wind speed change analysis

The average wind speed during 1990–2014 was approximately 9.43 km/h. Wind speed significantly ($p < 0.01$) increased since 1990 (Fig. 3d). Wind plays an important and extensive role in the lake water quality and cyanobacteria outbreak, directly affecting the water environment. Figure 3e clearly illustrates that the southwest wind is dominant nearly all year, and the city of Kunming is in the east and south area, which

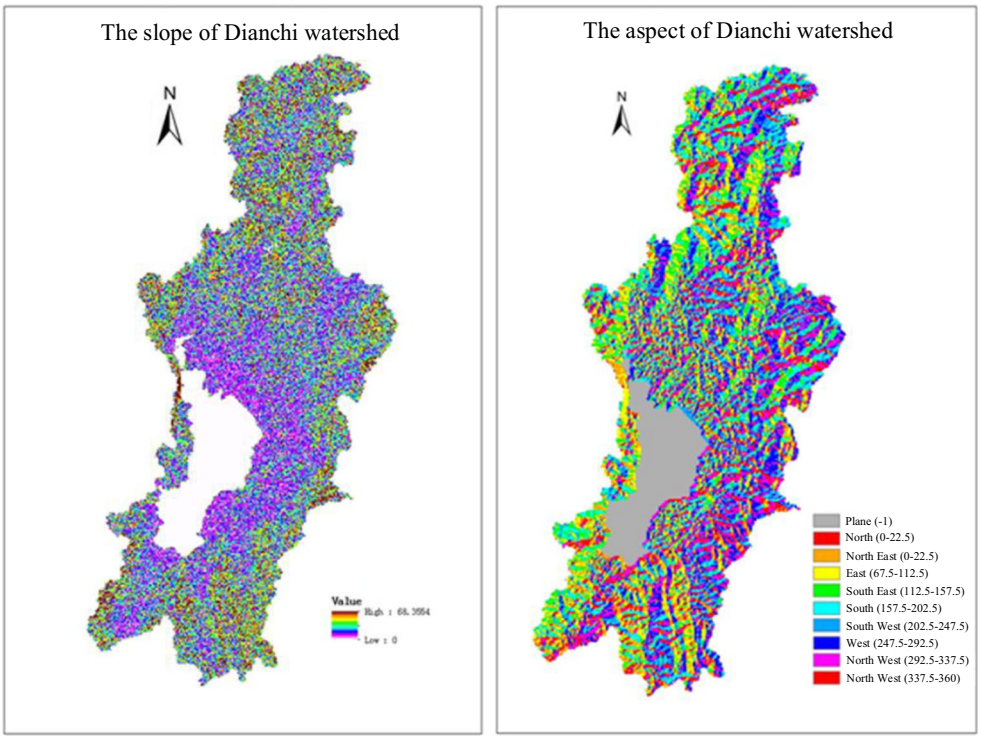


Fig. 2. The slope and the aspect of the Dianchi watershed. **a** The slope of Dianchi watershed; **b** the aspect of Dianchi watershed

has resulted in pollutant accumulation and environment deterioration.

Human environment change analysis

The urbanization level of Kunming reached large-scale enhancement in the strategy context of "One Belt and One Road", "Bridgehead Strategy", "New Kunming Construction of One Lake with Four City", and "Protecting Farmland in Flatland Areas and Constructing Mountainous Cities" in recent years. These development situations have increased the intensity of Dianchi Lake pollution. One of the most important aspect is that the impervious surface area (ISA) is increasing due to the urbanization process. An increase in ISA leads to soil erosion, the urban heat island effect, and urban waterlogging (Pham S V et al. 2008; Yuewen Ze et al. 2006; Anderson J R 1976; Peng Jian et al. 2006; Wu and Murray 2003; Xiao R B 2011; Voorde T V D et al. 2011).

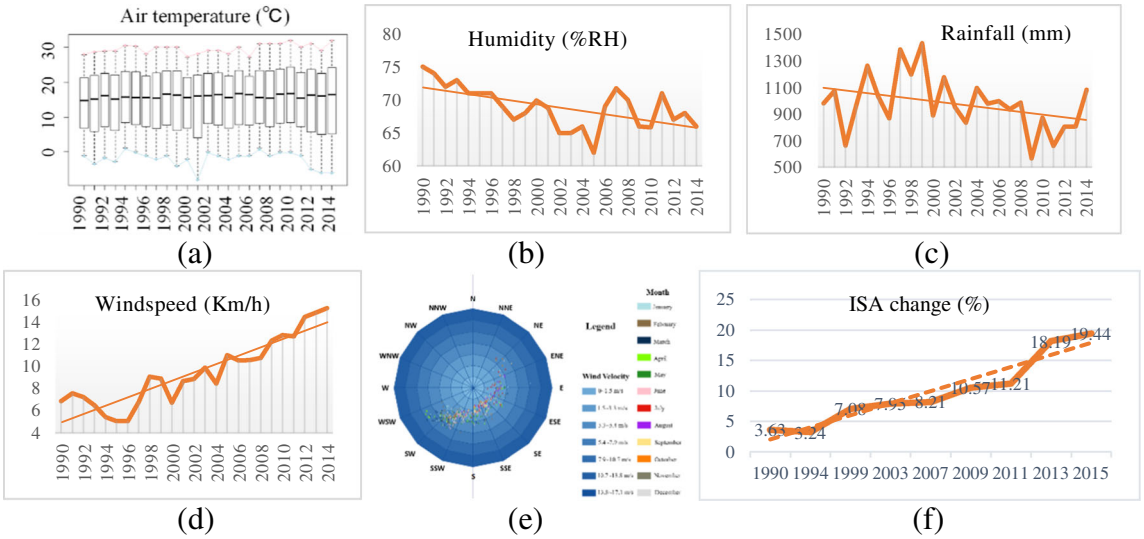


Fig. 3 The natural and human parameters change trend of Dianchi watershed over the last 25 years (1990–2014). **a** The surface air temperature change trend; **b** relative humidity change trend; **c** rainfall change trend; **d** wind speed change trend; **e** wind location analysis result; **f** the impervious surface area change trend

ISA change analysis

ISA is an important parameter in the study of the impact of urbanization on the water environment. In 1994, Schueler T R (1994) proposed that the impact of ISA on the water environment can be ignored when ISA less than 10%; the deterioration of the water environment is obvious when ISA is 10–25%; when ISA is greater than 25%, the deterioration of the water environment is hard to reverse. After analysing the changing trends of the 17 sub-watersheds of the Dianchi watershed during 1990–2015, we found that all the sub-watershed ISA values were significantly increasing ($p < 0.001$). Figure 3f shows the changing trend of the Dianchi watershed during 1990–2015, and the spatial-temporal distribution of each sub-watershed ISA is shown in Fig. 4.

The results show that the ISA changing trend was linear ($R^2 = 0.93$). In 1993, the ISA of the Dianchi watershed was approximately 3.63, and the impact of ISA on the water environment could be ignored. In 2009, the ISA of the Dianchi watershed was approximately 10.57%, indicating that the deterioration of the water environment was obvious due to the increasing ISA. In 2015, the ISA of the Dianchi watershed reached 19.44%, and the impact of ISA on the water environment would be disastrous if the ISA continued to expand.

Figure 5 shows the changing trend analysis of the 17 sub-watersheds of the Dianchi watershed during 1990–2015. In 1999, four sub-watersheds had ISA of 10–25%, and the remaining sub-watershed ISA values were less than 10%. From 1999 to 2009, ten sub-watersheds had ISA of 10–25%, and three sub-watersheds had ISA greater than 25%. In 2015, the number of sub-watersheds with ISA greater than 25% was ten. In addition, the ISA values of the Jinye River sub-watershed, the Gucheng River sub-watershed, Chuanfangzi River sub-watershed, and Dongbaisha River were 32.09, 34.41, 36.91, and 40.75%, respectively. The analysis shows that the urbanization speed of Kunming is extremely fast, and the effect on the water environment is severe.

House construction area change analysis

House construction areas are the places where people live. In the last 30 years, Kunming has been in a stage of rapid development. As a significant characteristic of urbanization, increasing house construction area accompanied this development. The increasing house construction has subjected the lake to increasing environmental pressure, resulting in decreasing wood land, farmland, and bare land. Figure 6a shows the Kunming house construction area change trend during 1993–2014.

There is a significant increasing trend of house construction area ($p < 0.01$). This changing trend can be broken into two phases. In the first phase of 1993–2009, the changing trend was linear ($R^2 = 0.83$); in the second phase during 2009–2014, Kunming experienced rapid growth following a linear relationship ($R^2 = 0.95$).

Population change analysis

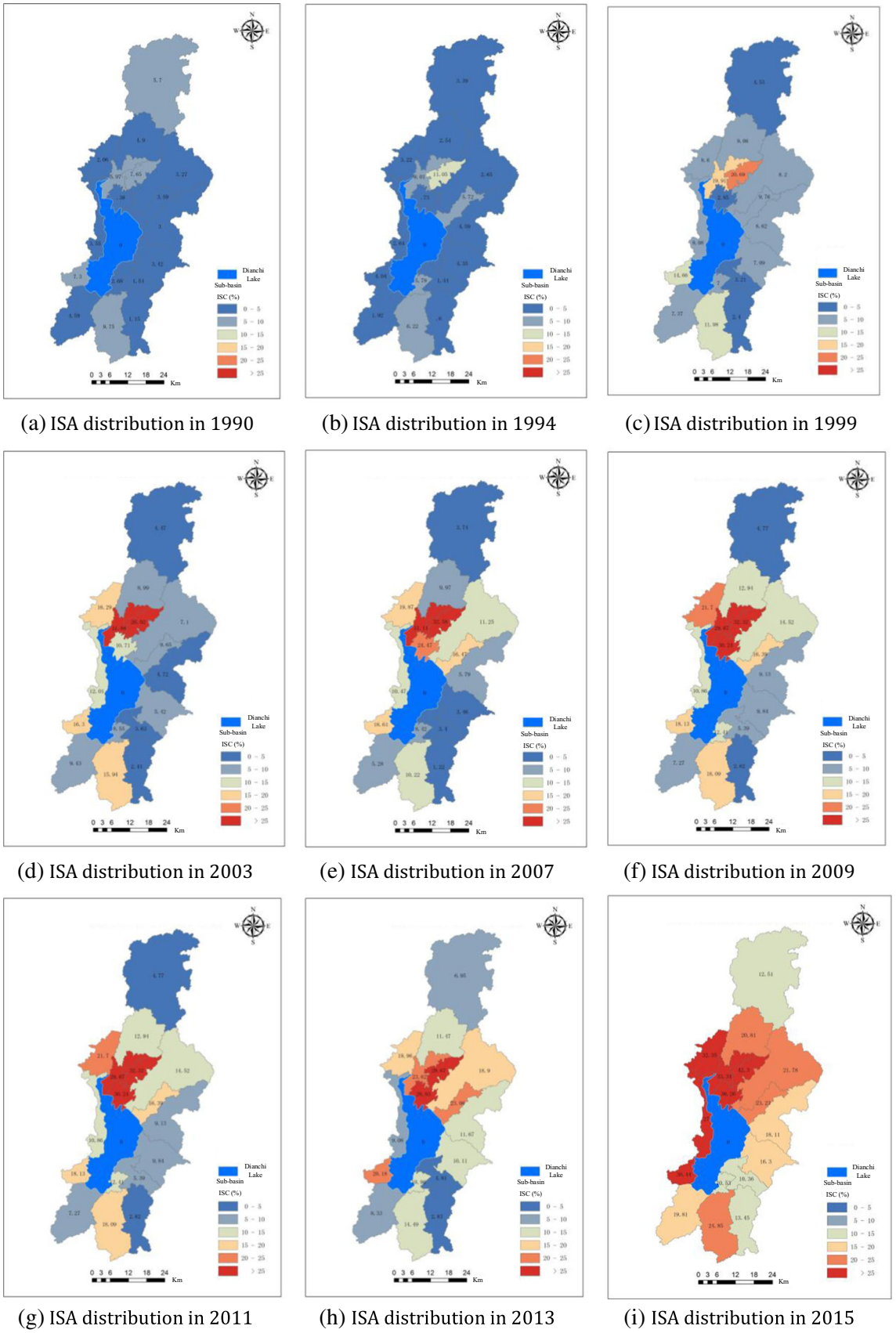
A fast urbanization process is accompanied by large-scale migration, resulting in a rapidly increasing city population and a dramatically declining rural population. Figure 6b shows the population change analysis during 1990–2014. The population of Kunming rose to 6.626 million in 2014, twice the population of 1990. There was a significantly increasing population trend ($p < 0.001$). Furthermore, Kunming is a population inflow city and is the main centralization area. The eco-carrying capacity of the Dianchi watershed is being challenged.

It is necessary to perform long-term, dynamic, and real-time monitoring to provide data and a theoretical basis to control the Dianchi Lake pollution to protect and control the ecology of the Dianchi watershed.

Model parameter selection and evaluation criteria

Model parameter selection

The selection of indicators is fundamental to predicting the trends and areas of outbreak. There are wide ranges of indicators associated with eutrophication. In general, they can be divided into physical, chemical, and biological. Since outbreaks of cyanobacteria result from the interaction of a variety of factors, we cannot predict the degree of eutrophication or cyanobacteria outbreak based on only one or a few indicators. However, treating all related indicators as inputs in the model is not only unrealistic but may result in model over-fitting. Too many input variables slow the convergence rate and generate substantial noise. Some indicators cannot be measured by in situ sensors. These indicators can only be measured through water sample collection and analysis in the laboratory. The complex process restricts their application in real-time systems. The water quality of Dianchi Lake is Grade V, the average chemical oxygen demand (COD) is approximately 58 mg/L, the average total phosphorus (TP) is approximately 0.22 mg/L, the average total nitrogen (TN) is approximately 0.85 mg/L, the average of permanganate index (PI) is approximately 9.8 mg/L and the average trophic status index is approximately 72 based on the Environment Condition Gazette, which was published by the Yunnan Province Environment Monitoring Canter in 2014. According to the Environmental Quality Standard of Surface Water published by the National Environmental Protection Bureau, Dianchi Lake is at a high eutrophication level, and this situation will not change



(a) ISA distribution in 1990 (b) ISA distribution in 1994 (c) ISA distribution in 1999

(d) ISA distribution in 2003 (e) ISA distribution in 2007 (f) ISA distribution in 2009

(g) ISA distribution in 2011 (h) ISA distribution in 2013 (i) ISA distribution in 2015

Fig. 4 Spatial-temporal distribution of each sub-watershed ISA

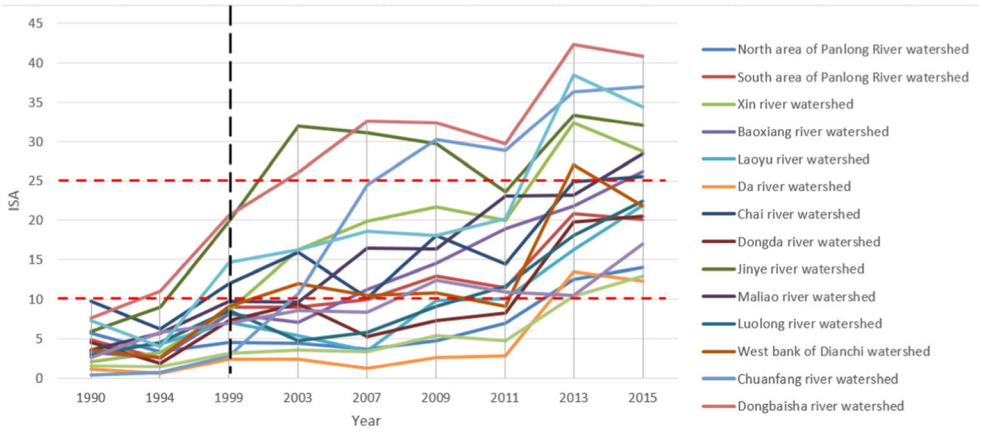


Fig. 5 The changing trend analysis of the sub-watersheds of the Dianchi watershed during 1990–2015

quickly. Some of the water quality parameters will remain within a narrow range (such as COD, TP, TN, and PI). Because Dianchi Lake is at a high eutrophication level, algal blooms will occur when the environment parameters (such as air temperature, rainfall, and Chla) meet the appropriate conditions.

This paper selects 13 parameters that are highly related to eutrophication, TN, TP, COD, PI, Chla, dissolved oxygen (DO), turbidity (TUR), pH, water temperature, light, wind speed, wind direction, and precipitation, for the model.

We obtained TN, TP, COD, and PI from the data published by Yunnan Environmental Monitoring Center. Chla, DO, TUR, pH, and temperature were measured using in situ sensors. Light, wind speed, wind direction, and precipitation were retrieved from meteorological data provided by the Yunnan Meteorological Bureau.

Evaluation criteria for outbreak

Li et al. used the density index of cyanobacteria and classified outbreaks into five categories (2014). Liu et al. used floating areas of cyanobacteria in Taihu Lake and divided the bloom degree into small, medium, large, serious, and catastrophic (2011). By monitoring the movement patterns of fish in the lake, Chen et al. selected the curvature and proximity features to evaluate water quality (2015). Kong et al. used ten meteorological and hydrological parameters as evaluation criteria, with Chla used as an indicator to monitor the degree of bloom (2009).

When a lake is in a eutrophic state, the number of dominant species grows exponentially. Hence, we can evaluate the degree of eutrophication based on the population of the dominant planktonic algae. The concentration of Chla can be used to represent the quantity of algae. Analysing and predicting the change in Chla helps to estimate the biomass status and trends of lake phytoplankton.

In this paper, we used 13 indicators as model inputs to predict the Chla concentration in the next 24 to 72 h. This indicator can be used as a signal for cyanobacteria outbreak.

Algorithm analysis

The lake system is a multi-level, multi-factor, and multi-target complex system. The change process of the algal blooms reveals a significant level of complexity. Data generated from this process show obvious patterns of randomness, incompleteness, and uncertainty. Such characteristics suggest that cyanobacteria bloom prediction is a typical grey system. In addition, cyanobacteria outbreaks show significant regional differences due to interactions between external environmental factors and internal water substances. This feature suggests that outbreak prediction is a nonlinear problem. Considering these

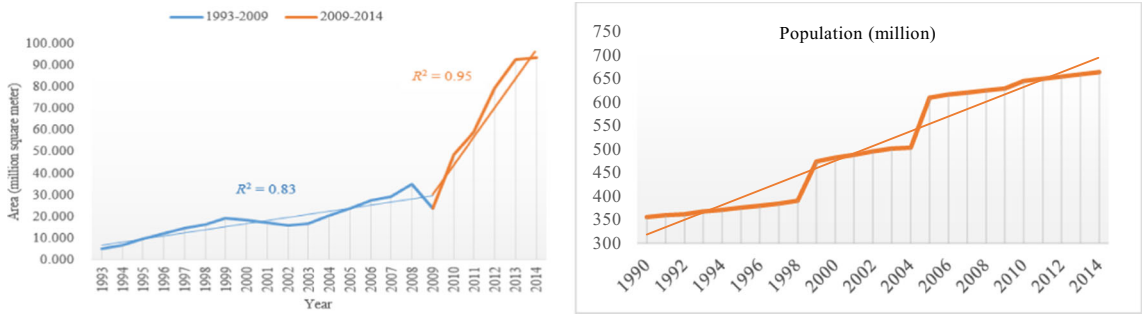


Fig. 6 House construction uses area and population change during 1990–2014. **a** house construction uses area change trend; **b** population change trend

characteristics, we use an artificial neural network (ANN) in this study to model the process of cyanobacteria outbreak.

We combined the adaptive grey model and BP ANNs in an AGM-BPANN model to predict the process of cyanobacteria outbreak. This model incorporated both environmental factors (precipitation, wind speed, wind direction) and internal water environment factors (nitrogen, phosphorus, COD, permanganate, chlorophyll, DO, pH, temperature). We used the model to predict the variation of cyanobacteria blooms on a specific time scale. Figure 7 shows the workflow of the AGM-BPANN model.

The AGM algorithm

AGM (1, 1) is the core of the grey model and evolves from GM (1, 1). GM (1, 1) is a first-order differential equation of a single variable. Its discrete-time response function is exponential. The modelling process is as follows, assuming that the sequence of original data can be represented as $X^{(0)} = \{X^{(0)}(1), X^{(0)}(2),\ldots, X^{(0)}(n)\}$ and the accumulated sequence can be obtained after accumulating them.

$$X^{(1)}(t) = \sum_{m=1}^{i} X^{(0)}(m), \mathrm{i} = 1, 2, \ldots, \mathrm{t} \tag{1}$$

Therefore, the new sequence can be acquired, $X^{(1)} = \{X^{(1)}(1), X^{(1)}(2),\ldots, X^{(1)}(n)\}$, and the differential equation of the new sequence is as follows:

$$\frac{dx^{(1)}}{dt} + kx^{(1)} = m \tag{2}$$

Equation (3) shows the discrete form of Eq. (4).

$$x^{(0)}(t+1) + \frac{1}{2}k\left[x^{(1)}(t+1) + x^{1}(t)\right] = m \tag{3}$$

K stands for the parameter to be identified; m stands for an endogenous variable. Assume that the estimating parameter is as follows:

$$A = \begin{bmatrix} \hat{k} \\ \hat{m} \end{bmatrix} = \left(X^T X\right)^{-1}\left(X^T Y\right) = \begin{bmatrix} \frac{2\left(1-e^{k}\right)}{1+e^{k}} \\ \frac{2A}{1+e^{a}} \end{bmatrix} \tag{4}$$

$$Y = \begin{bmatrix} x^0(2) \\ x^0(3) \\ \ldots\ldots \\ x^0(n) \end{bmatrix}$$

$$X = \begin{bmatrix} -\frac{1}{2}\left[x^{(1)}(1) + x^{(1)}(2)\right] & 1 \\ -\frac{1}{2}\left[x^{(1)}(2) + x^{(1)}(3)\right] & 1 \\ \ldots\ldots & \\ -\frac{1}{2}\left[x^{(1)}(n-1) + x^{(1)}(n)\right] & 1 \end{bmatrix} \tag{5}$$

Because X and Y are confirmed parameters, A is an unknown parameter. The number of equations is $n - 1$, so the equation has a solution. A can be calculated using the least squares method. The cumulative time series of the GM (1,1) model can be acquired by inserting A into Eq. (2).

$$\hat{x}^{(1)}(t+1) = \left[x^{(0)}(1) - \frac{\hat{m}}{\hat{k}}\right]e^{-\hat{k}t} + \frac{\hat{m}}{\hat{k}} \tag{6}$$

$x(1)(t + 1)$ stands for the accumulative prediction value. The prediction model of the original sequence can be acquired after subtracting the series.

$$\hat{X}^{(0)}(t+1) = \hat{X}^{(1)}(t+1) - \hat{X}^{(1)}(t), (t = 1, 2, \ldots, \mathrm{n}) \tag{7}$$

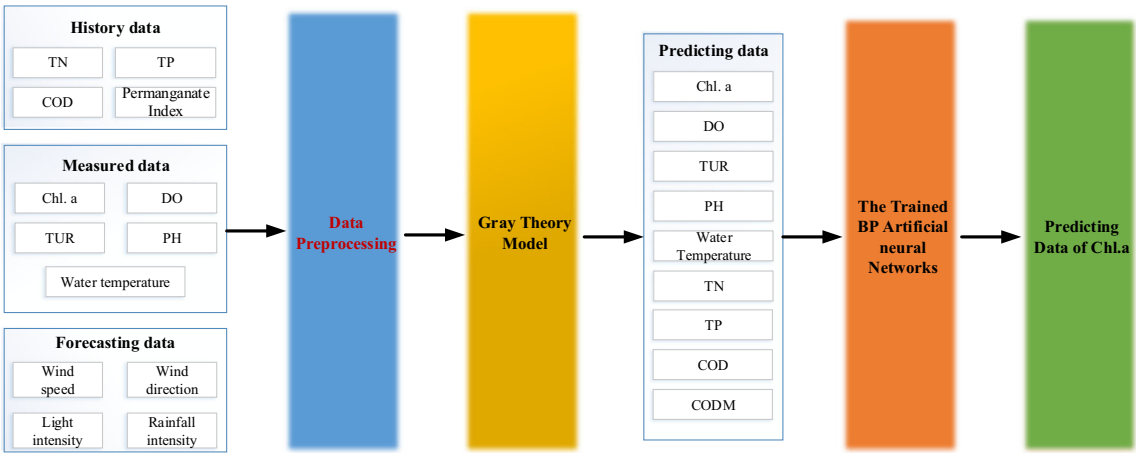


Fig. 7 The structure of GM-BPANN

There are two accumulation and restore steps in which error occurs when the randomness of the raw data is weakened in the procedure of constructing the model for GM (1,1). Therefore, GM (1,1) is a biased exponential model. The effort is uneven in practical application.

We can acquire Eq. (8) by replacing k and m with $\hat{k}$ and $\hat{m}$.

$$k = \ln\frac{2-\hat{k}}{2+\hat{k}}, \mathrm{A} = \frac{2\hat{m}}{2+\hat{k}} \tag{8}$$

Equation (9) shows the new, improved AGM (1,1) model.

$$\hat{x}^{(0)}(t+1) = Ae^{-kt} \quad t = 1, 2, \ldots, n \tag{9}$$

BP ANN algorithm

The BP ANN consists of three layers, an input layer, a hidden layer, and an output layer. Figure 8 shows the network architecture. The input layer consists of water quality variables. Based on the network structure, information is passed to the next layer in certain ways. The hidden layer typically includes one or more network layers, with each layer containing a certain number of network nodes.

The most commonly used activation function $f(x)$ (f_1) is shown in Fig. 8a), and is a sigmoid function or hyperbolic tangent function.

$$\text{Sigmoid}: \mathrm{f}(x) = \frac{1}{1+e^{-x}} \ (0 < f(x) < 1) \tag{10}$$

$$\text{Hyperbolic tangent function}: \mathrm{f}(x) = \frac{1-e^{-x}}{1+e^{-x}} \ (-1 < f(x) < 1) \tag{11}$$

We usually select a linear function (such as $f(x) = \mathrm{a}x + \mathrm{b}$) as the activation function when the model is used to approximate the function (f_2 is shown in Fig. 8a).

Assuming that O_l stands for the output node, H_n stands for the hidden node, I_m stands for the input node, W_{nm} stands for the weight of layer m to layer n, V_{ln} stands for the weight of layer n to the output layer, β stands for the threshold of node n, and Eq. (12) stands for each hidden node output.

$$H_n = f\left(\sum\nolimits_m W_{mn} I_m - \beta_n\right) \tag{12}$$

Therefore O_l can be expressed by Eq. (13).

$$O_l = f\left(\sum_m V_{ln} H_n - \beta_n\right) \tag{13}$$

The BP ANN learning process is divided into two phases, forward propagation and backward propagation. The outputs are generated through forward propagation. If there is a significant deviation between the actual output and expected value, the errors are calculated and passed backward based on the path connections. Based on this process, the weights and thresholds of the network are updated to reduce the prediction error. When the error is reduced to the required value or the number of iterations reaches a predetermined value, the learning process is stopped.

Cyanobacteria bloom prediction based on AGM-BPANN

The physical, chemical, and biological factors in the water body, as well as the interaction between natural and human factors, jointly influence eutrophication in the lake. Due to the uncertainty of the lake system, it is difficult to accurately describe the boundary and structure. The grey model can use approximate differential equations to describe future tendencies based on previously known or inaccurate information. By using time series data to determine the differential equation parameters, the grey model performs well in many uncertain systems. There is randomness, uncertainty, and a complex nonlinear relationship between various elements in the lake system. The ANN is good at transforming incomplete, unreliable and uncertain information into complete, reliable, and definite information and is suitable for high-dimensional, nonlinear problems when the mechanism is not clear.

Experimental results and discussion

Design of the system architecture

Because the goals of the system are to obtain, transfer, fuse, predict, and visualize water quality information, based on previous theoretical and algorithm preparation (Luo Yi et al. 2015; Luo Yi et al. 2014; Luo Yi et al. 2014), this project constructs a 4-layer architecture, including a monitoring layer, data storage layer, model layer, and application layer (Fig. 9).

The application layer collects real-time water quality information from meteorological and environmental monitoring nodes installed in Dianchi Lake. Managers can query and update the parameters of the monitoring nodes through the application interface. Similarly, technicians can monitor and maintain the nodes

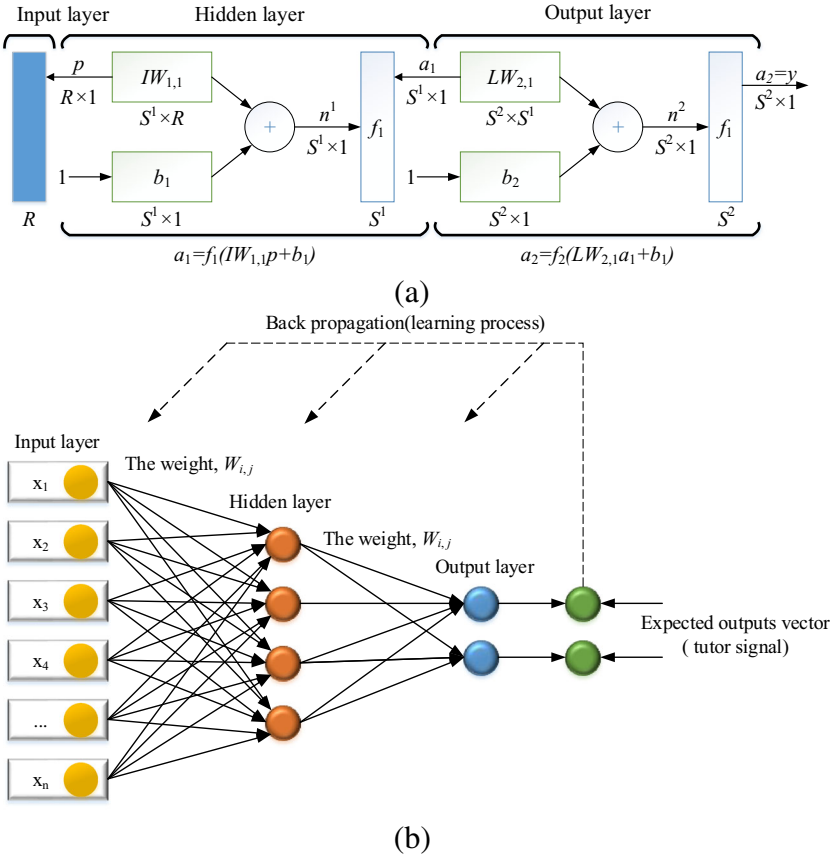


Fig. 8 The structure of GM-BPANN. **a** Matrix description of the algorithm; **b** node description of the algorithm

through the application interface. Users can obtain information about the water environment through the application interface. The interface also provides data visualization functionality based on GIS procedures. Because the system contains a series of structured/unstructured and relational/non-relational data, we used the non-relational database MongoDB to build the data storage layer to achieve multiple and heterogeneous data management. Before using the model for prediction, we pre-trained and optimized the model. The model layer was designed to conduct the training and optimization steps. When using the system, the model can be called directly after the training phase is completed.

Design of the monitoring node

Figure 10 shows the hardware design architecture for the monitoring node. It includes environment perception, sensor signal conversion, data processing, data fusion, and data exchange components. The environmental perception component is composed of a series of water quality sensors and weather sensors to measure water quality and collect meteorological data.

Figure 11a, b shows a photograph of a Chla sensor and water quality sensor. The signal conversion component processes the sensor array electrical signals through amplification, shaping,

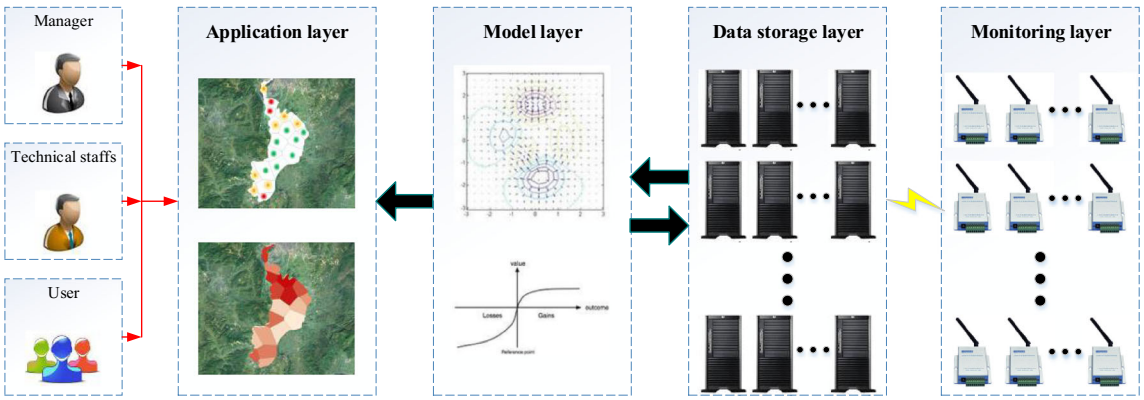


Fig. 9 The structure of the monitoring system

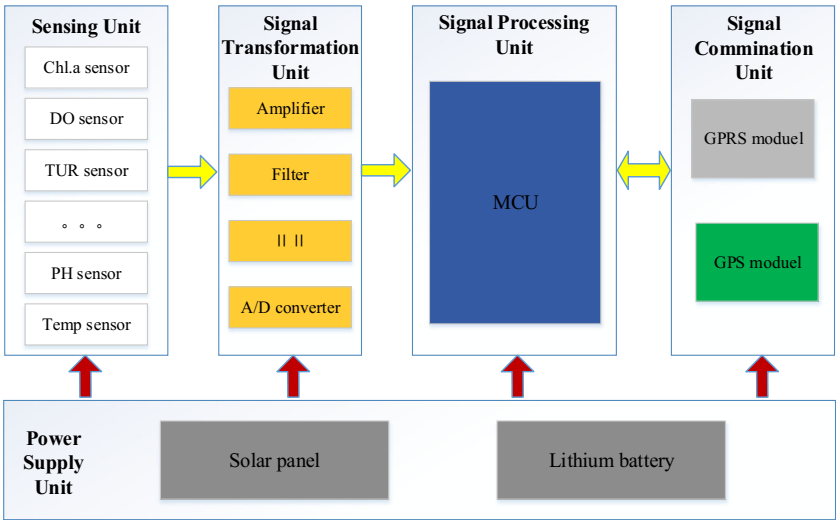


Fig. 10 The structure of the monitoring node

filtering, and A/D conversion and passes the data to the next layer. The data processing and fusion components use an LPC1788 as the main chip to conduct data collection, pre-processing, and other functions. The data exchange component is composed of a GPS module and a GPRS module; the former is used to obtain spatiotemporal information in the study area, while the latter implements data or command exchange between the monitoring node and monitoring centre.

Since the monitoring nodes are placed in the lake, the sensors are left unattended for long periods. Hence, we use solar panels to convert light energy into electrical energy and use lithium batteries to store the electricity to power the monitoring nodes. The power of a solar panel (Q) and the capacity of a lithium battery (P) can be calculated using e (14) and (15).

$$Q = K \times Q_d \times \alpha \times D \div C \tag{14}$$

$$P = 5618 \times K \times \frac{Q_d}{(K_{OP} \times H_L)} \tag{15}$$

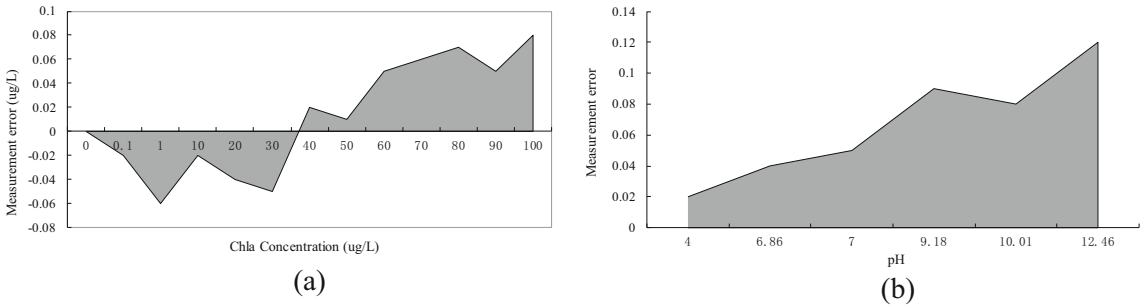
K stands for the safety factor, Q_d stands for the daily average power consumption of the monitoring node, α stands for the temperature correction factor, D stands for the longest successive number of rainy days, C stands for the discharge capacity coefficient, and H_L stands for the average solar radiation in a year. According to the argument list and historical meteorological data, the power of a solar panel must be greater than 35 W, and the capacity of the lithium batteries must be greater than 31 Ah. After considering the system margin, a 12 V/35 Ah lithium battery and a 40 W solar panel were selected. Figure 11c shows a photograph of the monitoring system.

Sensor calibration

Chlorophyll a sensor calibration

Figure 12a shows the measurement error distribution for chlorophyll a. Based on sampling principles for selecting the test point and sample size, we prepared 13 chlorophyll a standard solutions with different concentrations within the range of the sensor spectrum. The concentrations were 0, 0.1, 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 μg/L We used the chlorophyll a sensor to measure the concentration of each of the 13 solutions. We also used the monitoring nodes to conduct the measurements. Based on the testing results, the maximum

Fig. 11 Photographs of monitoring node. **a** Photograph of Chla sensor and water quality sensor array; **b** photograph of Chla sensor; **c** photograph of monitoring system



(a) (b)

Fig. 12 Measurement error distribution of Chla and pH. **a** Measurement error distribution of chlorophyll a; **b** measurement error distribution of pH

deviation of the chlorophyll a value was approximately 0.08 μg/L. We used the measured error rate to calibrate the sensor.

pH sensor calibration

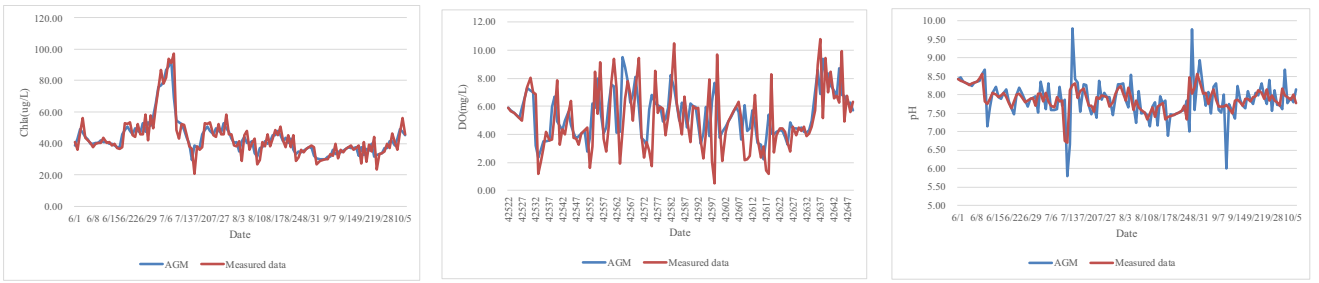
Because the pH sensor uses an electrode method to measure the pH, we calibrated the sensor before using it for monitoring. Based on sampling principles for selecting the test point and sample size, we selected six pH buffer solutions to calibrate the sensor: pH 4.00, pH 6.86, pH 7.00, pH 9.18, pH 10.01, and pH 12.46. Figure 12b shows the error distribution. The maximum pH measurement error was approximately 0.12, which was used to calibrate the sensor.

DO sensor calibration

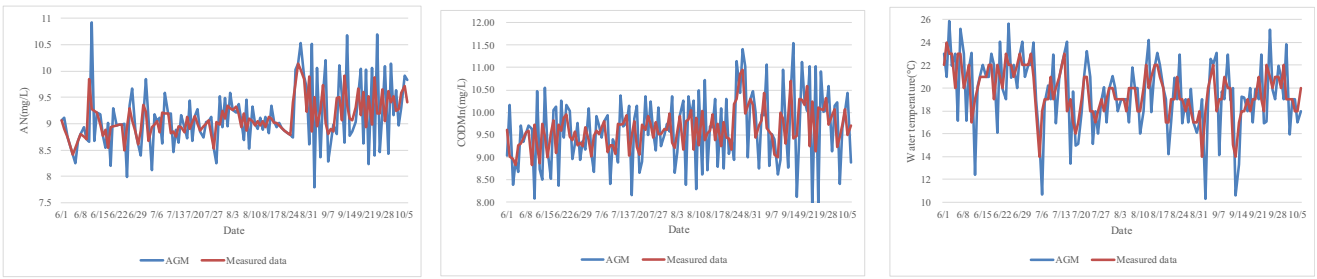
The common practice used to calibrate DO sensors is a two-point linear correction. We placed the sensor in a 0% DO solution and recorded the output at the monitoring nodes. Then, we placed the DO sensor in saturated moist air and recorded the output value. We calibrated the sensor based on the observed zero point error and range error.

Analysis of the experimental results

We started the monitoring experiment and continuously collected data since June 2015 at Dianchi Haigeng Park. The monitored parameters include chlorophyll, DO, pH, ammonia, and temperature. The data were updated every 6 h. We conducted data pre-processing before importing the data into the model. We averaged the daily water quality data before using them in the model. After repeated experiments and model calibration, and given full consideration of the accuracy of weather forecast data, we used the proposed AGM-BPANN model to predict the water quality conditions and outbreak probability 1 day in the future.



(a) Chla (b) Dissolved oxygen (c) pH



(a) Ammonia nitrogen (b) Permanganate index (c) Water temperature

Fig. 13 The curves of the measured and predicted water quality values

(a) (b)

Fig. 14 Field experiment and map of the cyanobacteria bloom. **a** Photograph of the monitoring site; **b** map of the cyanobacteria outbreak

AGM (1,1)-based prediction

We imported water quality data, including chlorophyll, DO, pH, ammonia nitrogen, PI, and temperature, into the AGM model. Figure 13 shows the results of the predicted values compared with the measured values. The red curve represents the measured values, while the white curve represents the predicted values.

We used the coefficient of determination (R) to evaluate the model performance. $R \leq 0.6$ indicates bad forecasting ability; $0.7 \leq R < 0.8$ indicates moderate forecasting ability; $0.8 \leq R < 0.9$ indicates good forecasting ability; and $R \geq 0.9$ indicates excellent forecasting ability. R and the error of the parameters were $R_{\mathrm{Chla}} = 0.86$, $\delta_{\mathrm{Chla}} = 9.74$, $R_{\mathrm{DO}} = 0.81$, $\delta_{\mathrm{DO}} = 1.15$, $R_{\mathrm{PH}} = 0.88$,$\delta_{\mathrm{PH}} = 0.26$, $R_{A\text{-}N} = 0.88$, $\delta_{A\text{-}N} = 0.40$, $R_{\mathrm{COD}Mn} = 0.86$, $\delta_{\mathrm{COD}Mn} = 0.19$, $R_T = 0.84$, and $\delta_T = 2.03$. The experiment results show that the water quality prediction performance of the model is satisfactory.

Chlorophyll a concentration forecast based on BPNN analysis

Chlorophyll a (Chla) is a comprehensive indicator of phytoplankton biomass and is one of the most important indicators of the degree of eutrophication. A change in the Chla value results from the combined effects of multiple factors. In addition, the Chla is transitive, i.e., the Chla concentration is temporally correlated. Hence, we used DO, pH, ammonia nitrogen, PI, water temperature, Chla, TN, TP, COD, CODMn, wind speed, wind direction, and precipitation as model inputs. The Chla value is used as the BPANN model output. From 4.4.1, we know that the AGM (1,1) model can accurately predict the water quality indicators 1 day in the future. Hence, we used the outputs from AGM (1,1) to train and validate the BPANN model. We trained the network 1000 times and found the model convergence. In addition, based on Mirchadani theory, we used three hidden layers (each layer contained 8 nodes). The training accuracy reached 10^{-2}. We used the sim function to test the network performance. R reached 0.995, which indicates that the model has good modelling performance. We also used the measured water quality data to validate the network and used the residuals to evaluate the learning performance. Because there was a serious outbreak of cyanobacteria between July 6 and July 9 near Dianchi Lake Park, we used the dataset between June 24, 2015 and July 23, 2015 to validate the model. Figure 14a shows the monitoring site. Figure 14b shows the map of the cyanobacteria outbreak.

Figure 15 shows the measured and predicted curves for the Chla concentration at one site. Based on the computation results, $R_{\mathrm{Chla}} = 0.93$ and $\delta_{\mathrm{Chla}} = 4.77$.

Conclusion

In recent years, the rapid development of urbanization and climate change have contributed to the deterioration of the ecological environment of Dianchi Lake. Therefore, it is necessary to conduct long-term, dynamic, and real-time monitoring of the water quality. Hence, this study used a wireless sensor network and GIS techniques to monitor the water quality in Dianchi Lake. Based on real-time data, forecast data, and the historical data, this study proposed that the AGM-BPANN model is to predict the water quality and concentration of Chla. The AGM-BPANN model could explain the

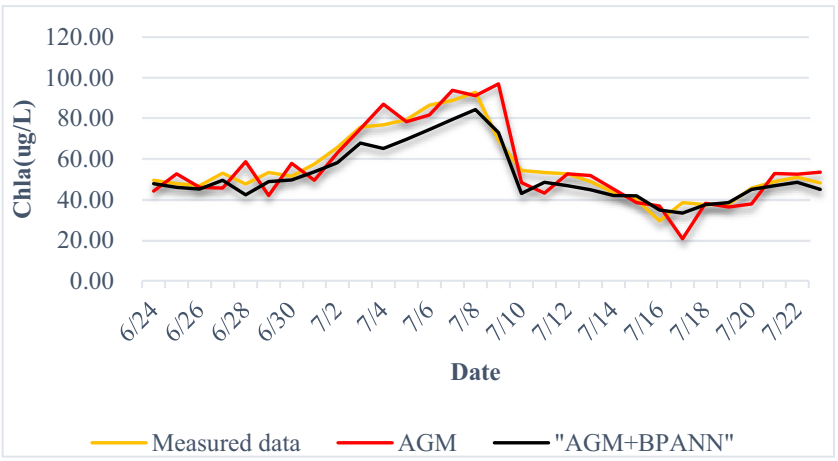


Fig. 15 Curves of the measured and predicted Chla

process of eutrophication and simulate the nonlinear trend. The successful prediction of the Chla concentration provides a scientific basis for early-warning of the cyanobacteria outbreaks. In future work, we will increase the number of monitoring sites in Dianchi Lake and conduct continuous and long-term monitoring. The results of the spatiotemporal modelling will be discussed in future papers.

References

Anderson JR (1976) A land use and land cover classification system for use with remote sensor data[J]. Usgs Prof Pap 964:964

Boqiang Q, Yunlin Z, Guang G et al (2014) The key factor for lake ecological restoration analysis [J]. Prog Geogr 33(7):918–924

Cheng S, Liu J, Li L (2015) Study on anomaly water quality assessment factor based on fish movement behavior[J]. Chin J Sci Instrum 36(8):1759–1766

Evans J, Janek JF, Hunter BL (2008) Wireless sensor network design for flexible enuironmental monitoring[J]. J Eng Technol 25(1):46–52

Genbao LI, Lin LI, Pan M et al (2014) The degradation cause and pattern characteristics of Lake Dianchi ecosystem and new restoration strategy of ecoregion and step-by-step implementation[J]. J Lake Sci 26(4):485–496

Guolin Y, Bin HE (2008) Influence of geographical features in Dianchi Lake Basin water pollution [J]. Environ Sci Surv 27(5):21–23

Haiyang Z, He Z, Shaojie M (2014) Design of low power acoustic-ultrasonic compound sensor node[J]. Chin J Sci Instrum 35(10): 2223–2230

IPCC (2007) Climate Change 2007: The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC. Cambridge University Press

Jian P, Lin WY, Yuan Z et al (2006) Effect on landscape pattern classification index [J] [25]. Land Use Geogr Sci 61(2):157–168

Karl TR, Arguez A, Huang B et al (2015) Possible artifacts of data biases in the recent global surface warming hiatus[J]. Science 348(6242): 1469–1472

Klemas V (2012) Remote sensing of algal blooms: an overview with case studies[J]. J Coast Res 28(1A):34–43

Lunetta RS, Schaeffer BA, Stumpf RP et al (2015) Evaluation of cyanobacteria cell count detection derived from MERIS imagery across the eastern USA[J]. Remote Sens Environ 157:24–34

Luo Y, Yang K, Yunbo S et al (2014) Research of radiosonde humidity sensor with temperature compensation function and experimental verification[J]. Sensors Actuators A Phys 218:49–59

Matthews MW, Odermatt D (2015) Improved algorithm for routine monitoring of cyanobacteria and eutrophication in inland and near-coastal waters[J]. Remote Sens Environ 156:374–382

Olmos MA, Birch GF (2010) A novel method using sedimentary metals and GIS for measuring anthropogenic change in coastal lake environments[J]. Environ Sci Pollut Res Int 17(2):270–287

Peng G (2007) Environmental monitoring wireless sensor networks for remote sensing technology [J]. J Remote Sens 11(4):545–551

Pham SV, Leavitt PR, McGowan S et al (2008) Spatial variability of climate and land-use effects on lakes of the northern Great Plains[J]. Limnol Oceanogr 53(2):728–742

Ren L, Cui E, Sun H (2014) Temporal and spatial variations in the relationship between urbanization and water quality[J]. Environ Sci Pollut Res 21(23):13646–13655

Rui J, Xin L, Baoping Y et al (2012) Heihe River watershed hydrological sensor networks [J]. Adv Earth Sci 27(9):993–1005

Schueler TR (1994) The importance of imperviousness[J]. Watershed Protection Techniques 1(3):100–111

Shengli W, Cheng L, Jun S et al (2009) Satellite remote sensing in Taihu Lake cyanobacteria bloom meteorological factors affecting the distribution and analysis [J]. Meteorological 35(1):18–23

Stumpf RP, Davis TW, Wynne TT et al (2016) Challenges for mapping cyanotoxin patterns from remote sensing of cyanobacteria[J]. Harmful Algae 54:160–173

Sun LY, Chen YZ, Wang XQ et al (2010) Comparison of multi-sensor data application in algal bloom detection[J]. Int Congr Image Signal Proc 3:2144–2148

Voorde TVD, Jacquet W, Canters F (2011) Mapping form and function in urban areas: an approach based on urban metrics and continuous impervious surface data[J]. Landsc Urban Plan 102(3):143–155

Wu C, Murray AT (2003) Estimating impervious surface distribution by spectral mixture analysis[J]. Remote Sens Environ 84(4):493–505

Xiao R B. (2011) Modeling spatial Ddistribution of population density for urban planning[J]. China Population Resources & Environment

Xin Z, Xi C, Wanqing L et al (2014) Non-point source pollution watershed landscape pattern remote sensing sources and sinks resolve[J]. Agric Eng 2(32):191–197

Xun L, Xiaomei X, Jia H et al (2012) Point source pollution control in Dianchi Lake Basin and resolve problems [J]. Lake Sci 22(5):633–639

Yao X, Wang G, Yang W et al (2010) Dynamics of the water bloom-forming Microcystis and its relationship with physicochemical factors in Lake Xuanwu (China)[J]. Environ Sci Pollut Res 17(9):1581–1590

Yi L, Yunbo S, Kun Y et al (2014a) Humidity sensor for radiosonde and its measuring circuit[J]. Opt Precis Eng 22(11):3050–3060

Yi L, Yunbo S, Liliang Q et al (2014b) Research on atmosphere aloft temperature probing based on harmonic analysis and AC comparison methods[J]. Chin J Sci Instrum 35(4):721–729

Yi L, Kun Y, Lin D et al (2015) Research on dynamic measurement method for button contact resistance[J]. Chin J Sci Instrum 36(1):49–55

Yong L, Yang P, Hu S et al (2012) Watershed pollution prevention planning and eutrophication control strategy for Lake Dianchi[J]. Huanjing Kexue Xuebao 32(8):1962–1972

Yonghong Z, Enjie D, Yanjun H (2015) Energy balance routing algorithm for WSNs optimized with PSO[J]. Chin J Sci Instrum 36(1):78–86

Yu G, Naiming Z (2013) Kunming city urban surface runoff pollution characteristics [J]. J Environ Eng 7(7):2587–2595

Ze Y, Jianhua X, Lihua X (2006) Based on the environmental effects of ecological urban land use remote sensing images - Ecology of urban thermal environment and vegetation index as an example [J]. 26 (5): 1450–1460

Zhiqiang Z, Yue W, Chao S et al (2012) Underwater wireless sensor monitoring network for sampling and sparse approximate reconstruction[J]. Sci Instrum 33(12):2728–2734