ORIGINAL ARTICLE

CrossMark

A risk evaluation model for karst groundwater pollution based on geographic information system and artificial neural network applications

Li Bo[1,2,3] · Zeng Yi-Fan[4] · Zhang Bei-Bei[4] · Wang Xian-Qing[2]

Abstract

The risk analysis on karst groundwater pollution is a research hotspot in current international hydrogeological field as well as the premise of preventing and controlling groundwater pollution. According to the characteristics of groundwater pollution in the typical study area, the study selected main-control factors of risk evaluation on karst groundwater pollution in mountainous areas at first. Based on this, the research determines the method for quantifying the factors and established a risk evaluation index system for karst groundwater pollution. To overcome drawbacks of the method for determining weights of factors in traditional evaluation method, the study determines the structure of the artificial neural network model by combining the selected evaluation factors. And also, the weight coefficients of evaluation factors on each layer are calculated. On this basis, the model for evaluating the risk of karst groundwater pollution is established. Moreover, the risk zoning evaluation map of groundwater pollution in the typical study area is prepared after conducting the weighted stacking of various sub-layers using the geographic information system. The method applied in the study can comprehensively and objectively reflect that the groundwater pollution is controlled by multiple factors and reveal the nonlinear characteristic of the pollution process. Additionally, the evaluation result is institutive and visible, which can provide a certain basis and reference for relevant researches.

Keywords Karst groundwater · Pollution · Risk evaluation · Artificial neural network (ANN) · Geographic information system (GIS)

Introduction

Generally, the soil layer of karst areas in southwest China is thin with a surface-ground bilayer structure, so that it is easy for pollutants to enter into aquifers through weak overlying

✉ Li Bo
libo1512@163.com

1 Key Laboratory of Karst Environment and Geohazard, Ministry of Land and Resources, Guizhou University, Guiyang 550000, Guizhou, China

2 College of Resource and Environmental Engineering, Guizhou University, Guiyang 550000, Guizhou, China

3 Key Laboratory of Groundwater Contamination and Remediation, China Geological Survey (CGS) and Hebei Province, Shijiazhuang 050061, Hebei, China

4 National Engineering Research Center of Coal Mine Water Hazard Controlling, China University of Mining and Technology (Beijing), Beijing 100083, China

strata and sinkholes. Compared with other non-karst aquifers, the karst groundwater pollution is increasingly significant and, therefore, how to reasonably and effectively protect karst groundwater has been an urgent problem to solve (Kaçaroğlu 1999; Shi et al. 1999; Tiwari et al. 2012; Zhang et al. 2017). The risk prediction of groundwater pollution is an important mean to prevent the groundwater pollution. Carrying out the risk identification on groundwater pollution not only can comprehensively explore the relationship between groundwater pollution and human social practices but also can timely determine the key areas of groundwater pollution, which provides a scientific basis for the management and protection of regional groundwater resources (Li et al. 2015, 2016; Masciopinto et al. 2017; Kourakos et al. 2012).

As an open system, the groundwater system shows a close and complex relation with the external factors and, therefore, the risk evaluation on groundwater pollution is a complex problem (Secunda and Collin 1998; Tiwari

et al. 2015; Sangam et al. 2015). It is found through literature review that hydrogeologists predicted groundwater pollution risk using different methods from different perspectives according to actual conditions of various areas due to complexities of evaluation factors and limitations of research levels. For example, the United States Environmental Protection Agency put forward the DRASTIC evaluation model in 1987, which is mainly designed for investigating the vulnerability evaluation of pores and fractured aquifers (Aller et al. 1987); Martin and Abraham of Israel comprehensively considered the influences of crucial environmental factors and land use intensity in risk evaluation on groundwater pollution. The four basic environmental factors involve hydrology, landform, soil and vegetation (Martin and Abraham 2001); Sappa and Vitale of Italy has established a pollution impact map for different agrochemicals through water quality simulation mode base on geographic information system (GIS), and chose different key parameters to evaluate groundwater pollution risk (Sappa and Vitale 2001); The UK uses the GOD method designed by Foster and Skinner in the risk assessment of groundwater source pollution (G is the groundwater outlet, O is the overall lithology, D is the depth of the groundwater) (Foster and Skinner 1995); Liu et al. constructed a risk assessment index system for groundwater pollution including influencing factors such as the exposure of the evaluation subject, anti-pollution of unsaturated zone, aquifer vulnerability and so on. In addition, the weight of each evaluation index was determined by analytic hierarchy process (AHP) method (Liu et al. 2012). Current researches have realized significant achievements about risk prediction on non-karst groundwater pollution while it is difficult to establish a uniform evaluation index system owing to the motion laws of karst groundwater in different karst areas show great disparities. Therefore, the subjective (such as Delphi, AHP) and objective (such as entropy, grey relational analysis) weighting methods are generally employed to determine weights of factors in current risk prediction models of groundwater pollution. However, these methods show shortcomings: subjectivity is strong and weights cannot reflect the actual significance of factors. To solve the aforementioned problems, the study established a risk evaluation system for karst groundwater pollution applicable to karst areas based on systematically investigating the typical study area. The study also introduced artificial neural network (ANN) method with learning mechanism containing weights. On this basis, the risk of karst groundwater pollution in Huaxi district of Guiyang city, Guizhou province, China, was evaluated using the data processing and display functions of the GIS, attaining a favorable effect. The research can provide bases and references for similar researches.

General situation of the study area

Geographical location

Huaxi district is located in southwest of Guiyang city in Guzihou province within the geographic coordinate of 106°26′40″–106°41′56″ (east longitude) and 26°15′28″–26°31′03″ (north latitude). The overall research area is 359.79 km². The geographic location is displayed in Fig. 1.
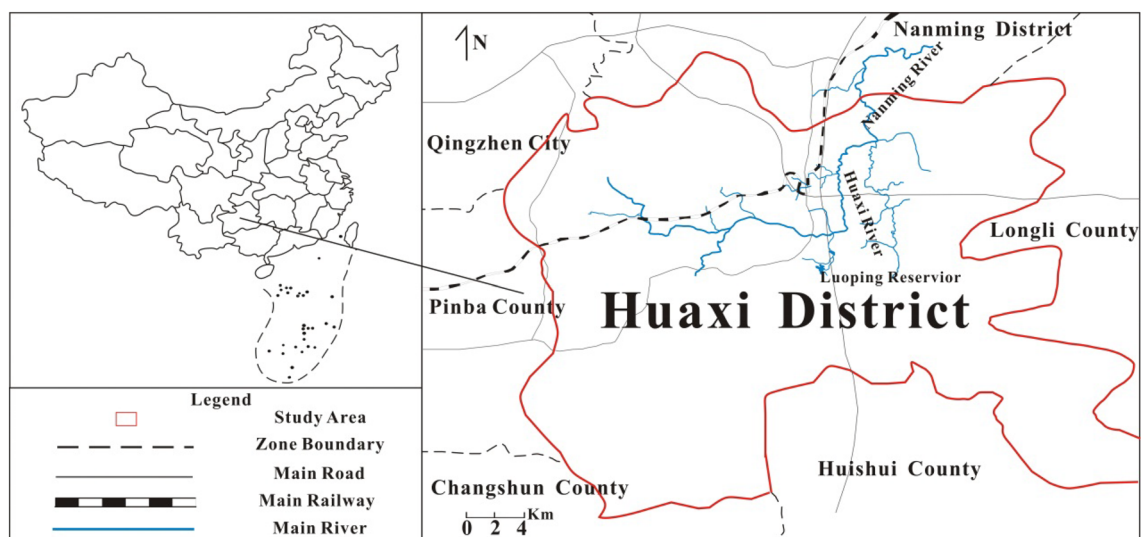

Fig. 1 Geographic location

Meteorological and hydrological conditions

The study area is located in the central Guizhou plateau, which has the subtropical plateau monsoon moist climate in the northern hemisphere with various characteristics of plateau, humid and monsoon climates. According to statistical observation data of Huaxi meteorological station since 1956, the average precipitation for years, the annual maximum and minimum precipitations in the study area were 1099.4, 1450.9 (in 2000) and 765.6 mm (in 1986), respectively. The precipitation mainly takes place in the summer half year, accounting for about 77.6% of the total.

The study area is located in the upstream zone of Qingshui River, a major tributary in Wujiang River of Yangtze River basin, which mainly includes Chetian, Lengfan, Gan and Kailun rivers, with the total runoff in the length of 60.1 km and the overall area of 225.91 km^2.

Topography

Located in the middle section of Miaoling Mountain, the study area gradually declines from southwest to northeast in terrain, which mainly appears as karst landform. The combination forms mainly include slopes, depressions, and clusters and valleys as well as eroded middle–low mountains and denudated hills with relative smaller areas.

Aquifer rock formations

The primary exposed strata in the study area from old to new are separately displayed as Permian, Triassic and Quaternary systems. According to the lithology of exposed strata and characteristics of aquifer media, the groundwater in the area is divided into carbonate karst water and pore water in loose Quaternary rocks, where the former can be further divided into fracture–cavern water, karst cave–fracture water, and dissolved pore–fissure water. The aquifer rock formations and the water-abundance characteristics are shown in Table 1.

Chemical properties of groundwater

The pH value of groundwater in the study area generally ranges from 7.2 to 8.04, appeared as neutral–weakly alkaline water. The total hardness (as $CaCO_3$) is in the range of 286–611 mg/L generally, shown as slightly–extremely hard water. As to the water quality, the groundwater in the study area mainly consists of various cations (such as Ca^{2+}, Mg^{2+}, K^+ and Na^+) and anions (including HCO^{3-}, $SO4^{2-}$ and Cl^-) in chemical compositions. The chemical types of groundwater are mainly shown as HCO_3^-Ca, HCO_3^-Ca·Mg and $HCO_3 \cdot SO_4^-$Ca.

The risk evaluation indexes of groundwater pollution

Compared with other groundwater systems, the karst groundwater system in Huaxi district shows significant characteristics: non-uniform aquifer media, diverse landforms, large topographic slope and poor contamination resistance of unsaturated zones. Consequently, the optimal evaluation effect cannot be acquired by simply applying current index systems. Thus, by fully investigating the characteristics of groundwater environment, the study established a risk evaluation index system for karst groundwater pollution in the study area from inherent and special factors of groundwater. The cause of the groundwater pollution is the conjunction of two kinds of factors, the first is the inherent factor which refers to the sensitivity to pollution of groundwater in a natural state and it is static, immutable and uncontrollable; the second is the special factor which refers to the sensitivity

Table 1 Water abundance of aquifer rock formations

Types of aquifer rock formation		Strata	Symbol	Lithology
Karst aquifer rock formation	Fracture–cavern karst water	Qixia and Maokou formations	P_2q+m	Limestone, cherty limestone, silicalite and mudstone
	Karst cave–fracture water	Guiyang formation	T_2gy	Argillaceous dolomite and limestone
		Huaxi formation	T_2h	Dolomite, argillaceous dolomite and limestone
		Daye formation	T_1d	Limestone
		Changxing formation	P_3c	Argillaceous limestone and bioclastics
		Wujiaping formation	P_3w	Shale with limestone, silicalite, sandstone and coal
	Dissolved pore–fissure water	Anshun formation	T_1a	Dolomite and limestone
Aquifer rock formation of pore water in loose rocks		Quaternary	Q	Clay, sand, conglomerate, etc.

to specific pollutants or human activities, and it is related to human activities and is dynamic and can be controlled. Then each evaluation factor was marked with a certain score (1–10) in different ranges. The larger the score is, the higher the risk degree (Tables 2, 3). The details are displayed as follows:

Inherent evaluation factors of groundwater

(1) Types of aquifers

The types of aquifers control the runoff path and length of groundwater, determine the contact condition of pollutant-bearing groundwater and aquifers and influence the migration, adsorption, dilution and dispersion of pollutants in the groundwater. Generally, the larger the scales and numbers of fractures and karst caves contained in an aquifer are, the stronger the permeability, the weaker the filtering capacity for pollutants and the poorer the degradation and dilution degrees. According to the drilling and field investigation, the types of aquifers in the study area can be divided into carbonate, clastics and carbonate–clastic aquifers.

(2) Topographic slope

The topographic slope influences the rainfall infiltration. Under a low slope, water gently flows and plenty of pollutants are infiltrated owing to the rainfall has sufficient contact areas with the ground surface so as to pollute the groundwater. On the contrary, the rainfall flows out as surface runoff so that there is a small quantity of infiltrations of rainfall with pollutants under a large topographic slope.

Table 2 Division of karst development degrees

Symbol	Description of lithology	Karst corrosion ratio (%)	Karst development degree
J_1zl	Sandy shale with limestone	0–2	Weak
T_3e+s	Sandstone and siltstone	0–1	Weak
T_2gy	Argillaceous dolomite and limestone	3–10	Weak–moderate
T_2h	Dolomite, argillaceous dolomite and limestone	10–15	Moderate
T_1a	Dolomite and limestone	15–30	Moderate–strong
T_1d	Limestone	15–30	Moderate–strong
P_3c	Argillaceous limestone and bioclastics	0–3	Weak
P_3w	Shale with limestone, silicalite, sandstone and coal	0–3	Weak
P_2m	Limestone, cherty limestone, silicalite and mudstone	> 30	Strong
P_2q	Cherty limestone and intercalation of muddy limestone in the lower part	> 30	Strong

Table 3 Scoring criteria of inherent evaluation indexes of groundwater

Topographic slope (°)	Scores	Thickness of unsaturated zone (m)	Scores	Precipitation (mm)	Scores	Types of aquifers	Scores	Average hydraulic conductivity (m/d)	Scores
< 5	10	2.00–3.00	9	990–1020	2	Carbonate rocks	4	0.05–0.10	1
6–15	9	3.01–5.00	7	1020–1050	4	Carbonate rocks with clastics	3	0.11–0.15	3
16–25	5	5.01–7.00	5	1050–1080	6	Clastics	2	0.16–0.20	5
26–35	3	7.01–9.00	3	1080–1110	8			0.21–0.25	7
36–62	1	9.01–10.5	1	1110–1140	10			0.26–0.40	9
Karst landform types	Scores	Types of media of unsaturated zones		Scores		Types of media of unsaturated zones	Scores	Karst development degree	Scores
Karst depression	9	Sand shale with limestone		5		Dolomite and limestone	8	Strong	10
Karst valley	6	Sandstone and siltstone		5		Argillaceous dolomite and limestone	8	Moderate–strong	8
Slope	5	Limestone		6		Dolomite and argillaceous dolomite	8	Moderate	6
Middle–low mountains	3	Shale with argillaceous limestone		6		Karst limestone	10	Weak–moderate	4
Eroded hills	1							Weak	2

Majorities of the study area show gentle terrain with the slope of 1°–15° while local areas are steep with the slope of 36°–62°.

(3) Thickness of unsaturated zones

The thickness of unsaturated zones determines the path and time of pollutants from the ground surface to the aquifer as well as the degradation and adsorption degrees. In general, the thicker an unsaturated zone is, the longer and the more complex the infiltration path of pollutants. Additionally, the probability that pollutants are degraded and adsorbed increases with the growing contact time between infiltration pollutants and unsaturated zones. Therefore, the thickness of the unsaturated zone exerts a protection effect on the groundwater system and the thicker an unsaturated zone is, the more favorable for preventing the groundwater from pollution.

(4) Precipitation

Precipitation is the primary carrier for liquid and solid pollutants to infiltrate and recharge to the groundwater system. Pollutants diffuse into aquifer through horizontal runoff after vertically migrating to aquifers through precipitation. The more the precipitations are, the more the infiltrating pollutants carried by precipitation and the larger the possibility of groundwater to be polluted. The average precipitation of the study area for years is found to be lowest in the southwest (990 mm) while most in the northeast (1140 mm), that is, the precipitations gradually rise from the southwest to the northeast.

(5) Karst development degree

The karst development degree is an important index for evaluating the risk of karst groundwater pollution. The higher the karst development degree is, the larger the development scales of karst caves and fractures in aquifers. Under the condition, it is easier for pollutants to flow in aquifers and more difficult to degrade and dilute the pollutants, which thereby leads to a higher possibility of groundwater pollution. The karst development degree in the study area is divided into five grades (Table 2): strong, moderate–strong, moderate, weak–moderate and weak by taking lithology and dissolved rates of boreholes as inference indexes. Different grades correspond to different scores.

(6) Hydraulic conductivity

The hydraulic conductivity depends on the property of rocks. The fractures inside rocks and size and connectivity of karst conduits influence the flow rate of groundwater and migration rate of pollutants to some extent and further affect the contact time between groundwater and aquifer media. The larger the hydraulic conductivity is, the stronger the water permeability of an aquifer, the faster the pollutants flow into the aquifer, the smaller the probability of the pollutants to be diluted and degraded during migration and the wider the pollution scope. According to the data obtained through pumping test, the hydraulic conductivity in the study area gradually declines from the southeast to the northwest.

(7) Types of karst landforms

Karst landforms are diverse and groundwater infiltrates in different approaches and modes under different landforms. For example, in karst depressions without soil layer or with thin soil layer, precipitations directly flow into underground rivers through depressions and sinkholes. In this way, groundwater is likely polluted if precipitations contain pollutants. However, in karst valleys, precipitations slowly flow out along the surface and mainly filtrate through small holes and karst caves. The filtration mode takes a long time. The combination types of karst landforms in the study area are divided into five types: karst landform types, karst depression, karst valley, slope, middle–low mountains, and eroded hills.

(8) Types of media of unsaturated zones

The media of unsaturated zones determine the filtration paths and approaches. Generally, the larger the particles of media of unsaturated zones are, the lower the clay contents and, therefore, the poorer the adsorptive capacity for pollutants while the larger the possibility of pollutants flowing into groundwater.

Special evaluation factors of groundwater

(1) Land use types

Human activities greatly influence groundwater system with the social progress. The most significant human behavior is shown as land use and land use types affect the capacity of pollutants to flow into groundwater. By analyzing satellite images, the study divides the land use types into land for construction, arable land, forest land and waters. And also, the scoring criteria of land use types in the study area are established according to the producing condition and capacity to flow into groundwater of pollutants under different land use types.

(2) Pollution sources of groundwater

The pollution sources of groundwater refer to the sources of various substances causing groundwater pollution. The pollution sources in the study area are divided into agricultural, domestic and industrial and mineral pollution sources according to data. The first two pollution sources separately refer to agricultural livestock farms and domestic sewage while the industrial and mineral pollution source denotes waste coal fields, sewage treatment plants and polluting enterprises. According to the migration rate and mode of groundwater, the types and influence scopes of pollution sources are scored. The corresponding scores are added up when random two or three scopes of pollution sources in the research study are superposed.

Weight determination

Back propagation neural network

In the comprehensive evaluation of groundwater pollution risk, the rationality of weights of factors directly influences the reliability of the comprehensive evaluation results. Currently, the weights of evaluation factors are generally determined using subjective and objective weightings. Subjective weight method reflects the empirical judgment of decision-makers and the weights determined using the method generally conform to the practices. However, it is easy to exaggerate or reduce the effects of certain factors owing to different decision-makers having diverse empirical judgments (Wang and Li 2006; Fan et al. 2002). In contrast, the weights obtained by applying the objective weighting method are not the true importance of factors but measurements of useful information provided by data of various factors although the method effectively utilizes the data information of evaluation factors. Moreover, the more the information is, the larger the corresponding weights of factors and viceversa (Lu et al. 2008). Additionally, as a comprehensive evaluation system is a complex nonlinear system, the influence degrees of various factors on the evaluated problems vary with time and space. It means that weights of majorities of evaluation factors change with the development of evaluated objectives and progress of human awareness. Therefore, it is necessary to establish a learning mechanism for weights to adapt to the ever-changing evaluation requirements. Artificial neural network (ANN) has special advantages in solving the problem (Dawson and Wilby 1999; Snell et al. 2000; Du and Zhao 2006). Back propagation (BP) neural network is a multilayer feedforward neural network established based on BP algorithm whose topology structure is composed of input, hidden and output layers. The completeness theorem of mapping capacity of BP neural network indicates that a three-layer BP neural network can approximate to any continuous function at random precision (Yang and Ma 2016; Ni et al. 2017). The study attempts to systematically identify the unknown relations between various evaluation factors using BP neural network to establish the neural network model and determine weights of various evaluation factors. The method for determining weights of factors using BP neural network is displayed as follows (Wang 2013; Deng et al. 2016):

(1) Establishment of BP neural network

Data are chosen as training samples. The training samples contain m factors, each of which has n samples. Thereby, an input data matrix $X = (x_{ij})_{n \times m}$, ($i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$) can be established in which x_{ij} refers to ith factor of ith sample.

The network model contains three layers: input, hidden and output layers. There are m nodes in the input layer, namely m evaluation factors, while the input vector is expressed as $X = (x_1, x_2, \ldots, x_i, \ldots, x_m)^T$. There is one layer in the hidden layer with k nodes (the value of k can be calculated according to Formula (1) or (2)). The output vector is shown as $Y = (y_1, y_2, \ldots, y_j, \ldots, y_k)^T$ and the output layer contains one node. The output value is O while the desirable output is expressed as $D(d_1, d_2, \ldots, d_j, \ldots, d_k)$. The connection weight matrix between the input and the hidden layers is displayed as $V = (v_{ij})_{m \times k}$ in which v_{ij} denotes the connection weight between ith node in the input layer and jth node in the hidden layer. The connection weight vector between the hidden and the output layers is expressed as $Z = (z_1, z_2, \ldots, z_j, \ldots, z_k)$ in which z_j refers to the connection weight between the output layer and jth node in the hidden layer.

$$k = (\text{nodes in input layer} + \text{nodes in output layer})/2 \qquad (1)$$

$$k = (\text{nodes in input layer} \times \text{nodes in output layer})^{1/2} \qquad (2)$$

There is the following relation between the input and the hidden layers:

$$y_j = f\left(\sum_{i=1}^{m} v_{ij}x_i\right) \quad j = 1, 2, \ldots, k \qquad (3)$$

The following relation exists between the hidden and the output layers:

$$O = f\left(\sum_{j=1}^{k} z_j y_j\right) \qquad (4)$$

$f(x)$ in Formulae (3) and (4) is an activation function and the nonlinear transfer function sigmoid is generally applied, namely

$$f(x) = \frac{1}{1 + e^{-x}}. \qquad (5)$$

Samples are chosen to train the neural network, namely the factor of the first sample is input after randomly taking a group of small data. The output error E_i (the output error of ith sample) occurs when the output result does not conform to the desirable value. The formula is expressed as follows:

$$E_i = \frac{1}{2(d_i - o_i)} \qquad i = 1, 2, \ldots, n \qquad (6)$$

The total error is expressed as follows:

$$E = \frac{1}{2} \sum_{i=1}^{n} E_i. \qquad (7)$$

To improve the generalization ability of the network, it is necessary to train the network by employing regularized adjustment method. The mean square error (MSE) of the network is employed as the performance function of the BP network. The formula is shown as follows:

$$\text{mse} = \frac{1}{N} \sum_{i=1}^{N} \left(e_i \right)^2 \tag{8}$$

where e_i refers to the training error of ith training sample ($e_i = t_i - a_i$) while t_i and a_i denote the target output and network output of ith training sample, respectively. The network performance function undergoing regularized adjustment is expressed as follows:

$$\text{msereg} = \gamma \text{mse} + (1 - \gamma) \text{msw} \tag{9}$$

where γ refers to the performance function while $\text{msw} = \frac{1}{n} \sum_{j=1}^{n} \left(w_j \right)^2$.

The error signal δ^o is obtained by calculating output desirable and actual values to adjust the connection weight between hidden and output layers. The specific method refers to Formulae (10)–(14). All samples are trained and the training ends when $E < E_{\min}$ (the training precision is generally a positive decimal fraction). Otherwise, the connection weight is modified and the calculation is repeated until satisfying the requirement.

$$\delta^o = (d - o)o(1 - 0), \tag{10}$$

$$\delta_j^y = \delta^o z_j y_j \left(1 - y_j \right), \tag{11}$$

$$z_j(t) = z_j(t - 1) + \eta \delta^o y_j + \mu \Delta z_i(t - 1), \tag{12}$$

$$v_{ij}(t) = v_{ij}(t - 1) + \eta \delta_j^y x_i + \mu \Delta v_{ij}(t - 1), \tag{13}$$

where the learning rate $\eta \in (0,1)$ and the momentum $\mu \in (0,1)$.

After the training ends and reaches required precision, the sum of absolute values of all connection weights (V) is calculated according to Formula (14), followed by normalization, to further obtain the weights of all factors.

$$w_j = \frac{\sum_{l=1}^{k} \left| v_{ji} \right|}{\sum_{i=1}^{m} \sum_{l=1}^{k} \left| v_{il} \right|} \quad j = 1, 2, \ldots, m \tag{14}$$

Weight calculation

If too many input samples are chosen, the calculation result may not meet the requirement for high precision. However, if only few are chosen, it would be hard to satisfy the accuracy requirement for the evaluation. After many attempts, 10 sets of data have been chosen as input samples. The locations of these data are distributed evenly in the study area, and then the quantization of input data is realized according to Tables 3 and 4. Meanwhile, to achieve contrast between input and output samples, we have further normalized the

Table 4 Scoring criteria of special evaluation indexes of groundwater
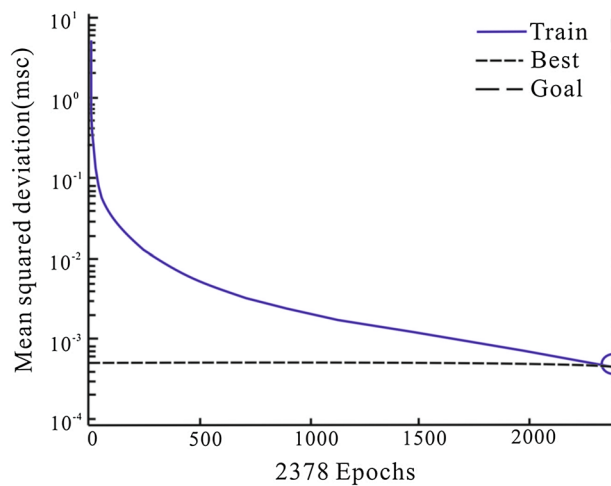
Land use types	Scores	Types of pollution sources	Influence scope	Scores	Influence scope	Scores	Influence scope	Scores
Land for construction	9	Agricultural pollution source	0–500 m	3	500–1500 m	2	1500–3000 m	1
Arable land	5	Domestic pollution source	0–500 m	5	500–1500 m	3	1500–3000 m	1
Forest land	2	Industrial and mineral pollution source	0–500 m	8	500–1500 m	5	1500–3000 m	2
Waters	1							

Table 5 Input samples for training the neural network

Input sample X	1	2	3	4	5	6	7	8	9	10
Topographic slope	1	1	1	0.889	0.889	0.889	1	1	0.444	0
Types of aquifers	0.5	0.5	0.5	0.5	0.5	1	1	0.5	0.5	1
Precipitation	0.25	0.25	0.75	0.25	0.5	0.75	1	0	0	0.5
Thickness of unsaturated zones	0.25	0.25	0	0.75	0.75	0.5	0.25	0.75	0.75	0.25
Types of media of unsaturated zones	0.6	0.6	0.2	0.2	0.2	0.6	0.6	0.2	0.2	0.6
Karst landforms	0.5	0.5	1	0	1	1	1	0.625	0.625	0.5
Average hydraulic conductivity	0	0.25	0.5	1	1	0.75	0.75	0.75	0.75	0.25
Karst development degree	0.5	0.5	0.5	0.25	0.25	0.75	0.75	0.5	0.5	0.75
Land use types	0.5	0.125	0.5	0	0.5	0.5	1	0.5	0.5	0.125
Types of pollution sources	0	0.054	0.324	0.676	0.216	0.108	0.054	0.162	0.324	0

Table 6 Desirable output samples for the training of the neural network

Desirable output Y	0.2078	0.1524	0.5071	0	0.3733	0.7852	1	0.1268	0.1114	0.3433

Fig. 2 Network training precision
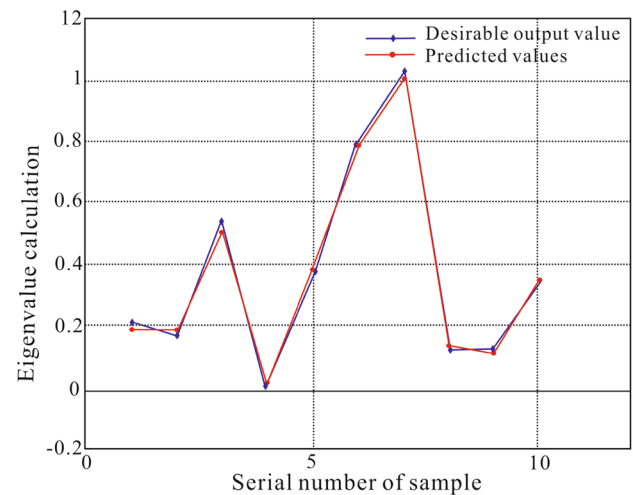

Fig. 3 Prediction of BP neural network

quantified data and converted into the numbers between 0 and 1, and Table 5 is the input sample used for network training. The three-layer structure composed of an input layer, a hidden layer and an output layer is employed. The hidden layer contains 9 nodes and the network training precision is 0.001. The MATLAB software is used in calculation.

The desirable output sample is the risk level of groundwater pollution of each input sample. Values between 1 and 5 were selected to quantify the output samples, in which 1 represents very low risk, while 5 represents very high risk. Similarly, we have normalized the quantified data and Table 6 is the desirable output samples according to artificial neural network training.

The blue line and the dotted line in Fig. 2 denote the precision condition and target of training, respectively. After training for 2378 times, the training ends owing to the network satisfies the requirement, namely, MSE < 0.001. At this time, the predicted values calculated by artificial neural network are consistent with the desirable output samples (Fig. 3). By the connection weigh matrix V between the input layer and the hidden one, we can get the weight of each evaluation factor shown in Table 7 according to formula 14.

Maps of various evaluation factors based on GIS

The thickness of unsaturated zone, the lithologic combination of unsaturated zone and aquifer, karst corrosion ratio, hydraulic conductivity and so on are calculated using

the collected data of geological drilling and pumping test about maps and records in the study area. To compare different factor data, it is necessary to remove the influences of different dimensions and units of various factors (Wu et al. 2016). Therefore, the dimensionless treatment is conducted on raw data according to the scoring criteria in Tables 3 and 4 to convert the data in the range of 1–10. Finally, thematic maps can be drawn by the interpolation method combined with GIS software which include the types of aquifers, the thickness of unsaturated zones, the media type of unsaturated zones, the hydraulic conductivity and the degree of karst development. In the same way, the distribution map of atmospheric precipitation has been made according to the years of monitoring data of the rain station in the study area. According to the existing topographic map of the study area, the topographic slope map has been drawn. According to the remote sensing image of the study area, the land use situation was divided into four types including construction land, arable land, woodland and water area, and the land use type map was made on this basis. Thematic map of Karst landform and influence scopes of pollution sources map have been drawn according to field investigation data, the influence scopes of pollution sources are harder to unify, the influence scopes of the representative pollution sources in Huaxi district was used as a basis in this study. According to the observation, the further the pollution distance is, the lighter the degree of pollution is. The evaluation factor thematic maps are as follows, based on which the spatial distribution law of

Table 7 Weights of various evaluation factors

Evaluation factors	Topographic slope	Aquifer types	Precipitation	Thickness of unsaturated zones	Types of media of unsaturated zones	Karst landform	Hydraulic conductivity	Karst development degree	Land use types	Types of pollution sources
Weight W_i	0.0626	0.116	0.0955	0.0952	0.1305	0.1066	0.0933	0.1174	0.1081	0.0748
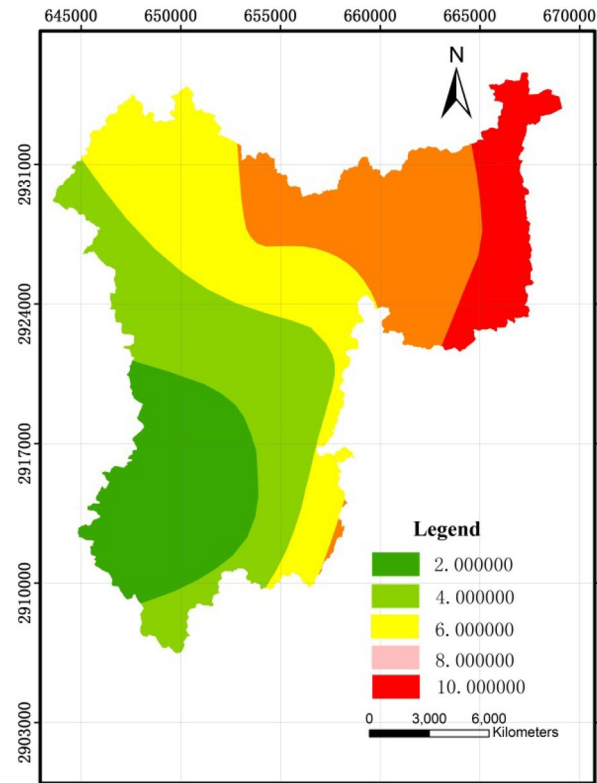
Fig. 4 Precipitation

the quantized values of various factors can be institutively mastered (Figs. 4, 5, 6, 7, 8, 9, 10, 11, 12, 13).

Risk prediction of karst groundwater pollution in Huaxi district

The risk evaluation model of karst groundwater pollution

Establishing a risk evaluation model for karst groundwater pollution is to establish a mathematical model which shows the influences of various evaluation factors on groundwater pollution. Thereby, the risk index *VI* is introduced to evaluate the risk of karst groundwater pollution. The risk evaluation is summarized to the sum of superposed influences of difference evaluation factors on groundwater pollution in an element of a section in an area, which can be expressed using the following model formula:

$$VI = \sum_{k=1}^{n} W_k \cdot f_k(x, y) \tag{15}$$

where *VI*, W_k, $f_k(x,y)$, *x, y* and *n* refer to risk index, weights of diverse evaluation factors, the score of *k*th evaluation factor,
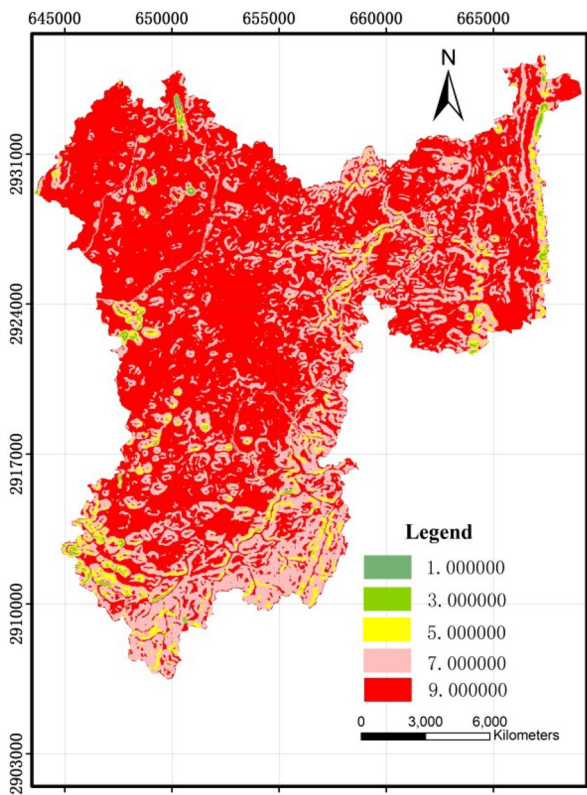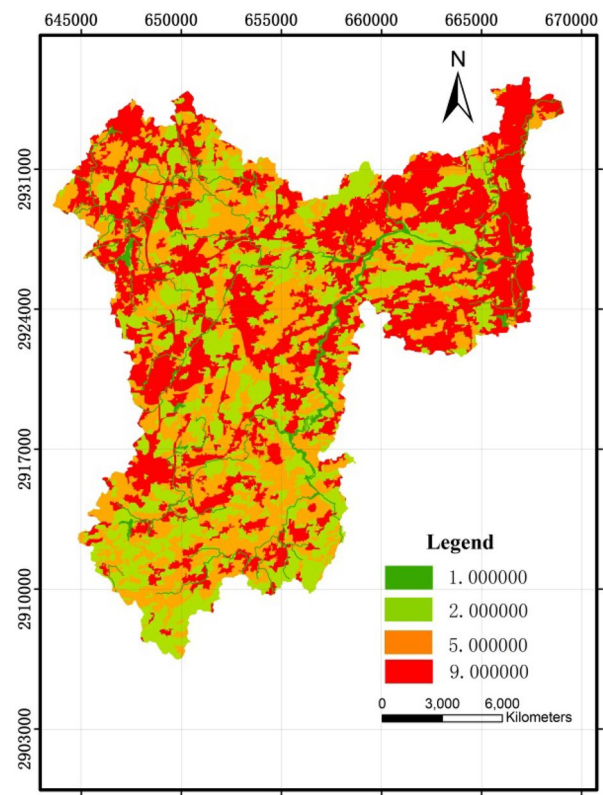
Fig. 5 Topographic slope
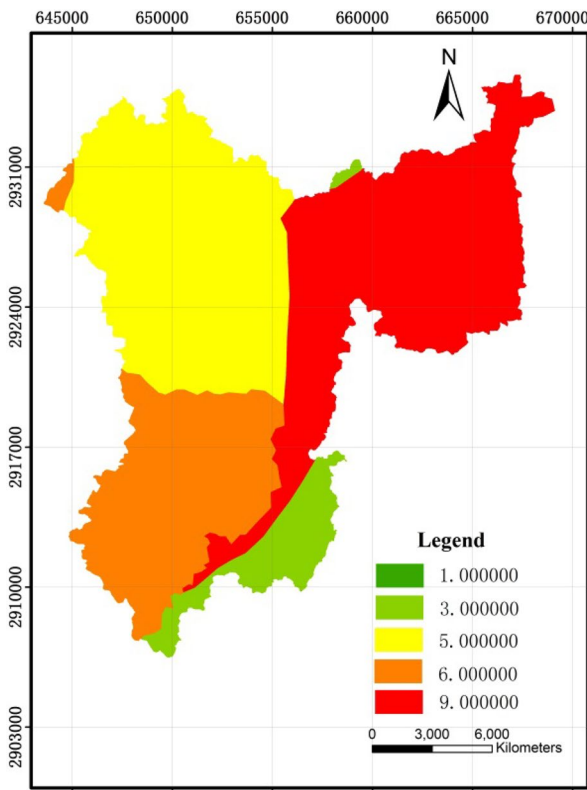
Fig. 7 Land use types
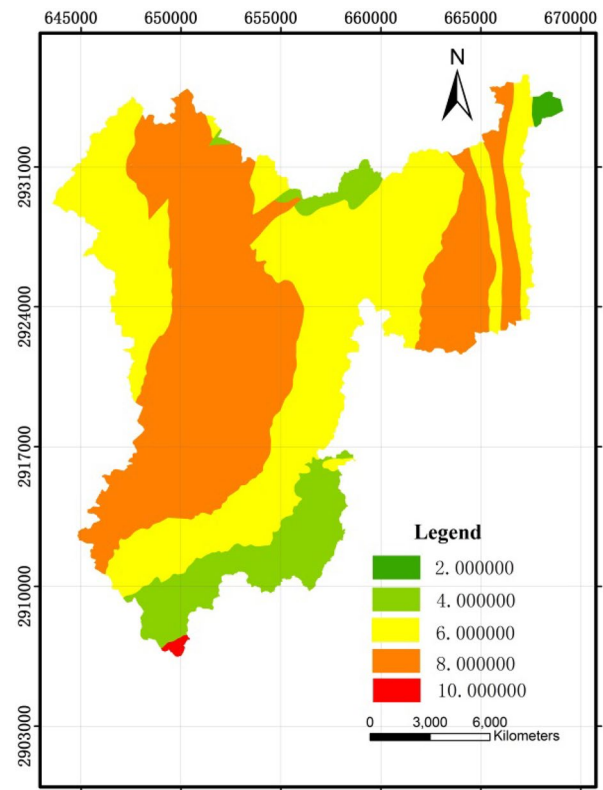
Fig. 6 Types of landforms
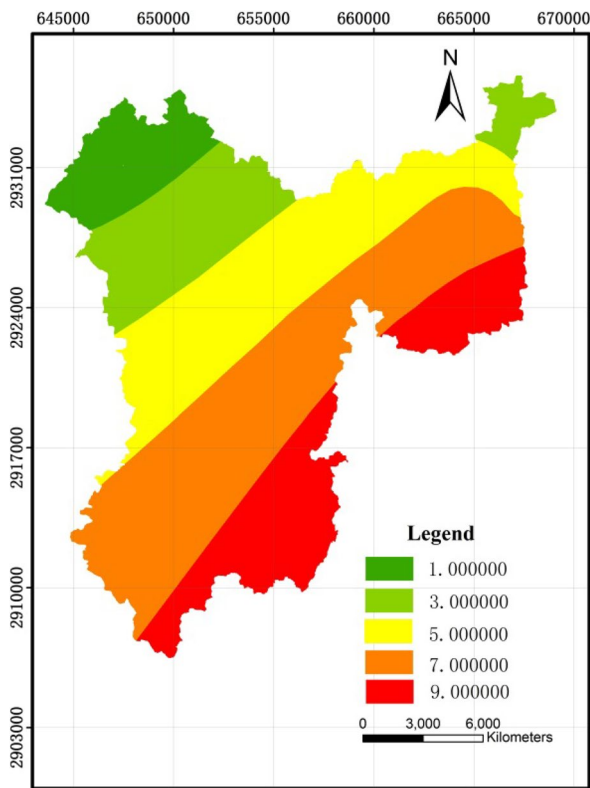
Fig. 8 Karst development degree

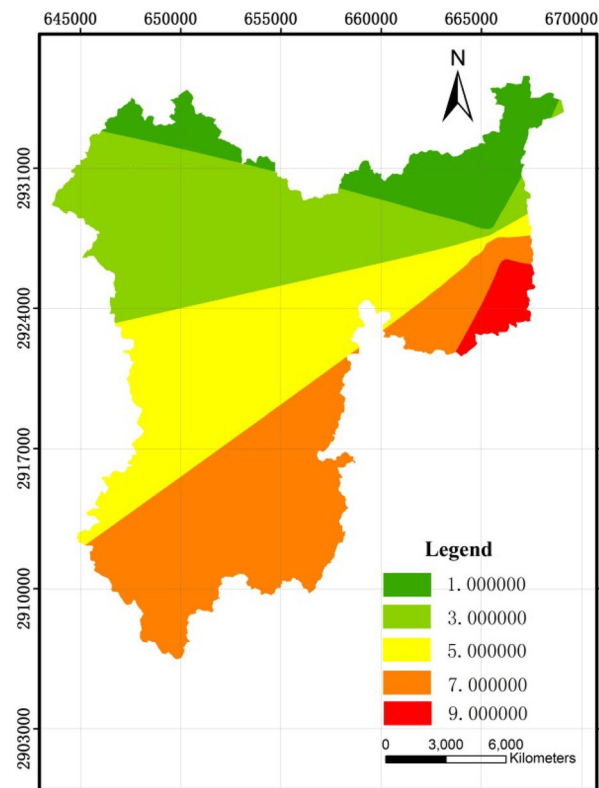**Fig. 9** Hydraulic conductivity

Fig. 10 Thickness of unsaturated zones

geographic coordinates, and the number of evaluation factors, respectively.

Risk evaluation of karst groundwater pollution in Huaxi district

On the basis of drawing maps of multiple evaluation factors, the maps of different evaluation factors are superposed by employing the information overlap technology of GIS to form the superposed cells combining multi-source information. Afterwards, the comprehensive risk evaluation indexes in different superposed cells are calculated using the established risk evaluation model, and then graded using Natural Breaks of GIS software to get the optimal five-level grading result. Eventually, the zoning map of comprehensive evaluation about the risk of karst groundwater pollution in Huaxi district combining multiple factors is prepared (Fig. 14).

The high-risk and relatively high-risk zones are mainly distributed in the northeastern and central regions of the study area with an area about 150 km², which takes up 43.9% of the total study area. The zone with low slope is abundant with precipitation and mainly shown as karst hills and depressions in landform. Moreover, the zone mainly has carbonate aquifers in terms of types of aquifers,

showing high karst development degree, large-scale karst caves and sinkholes, and thin unsaturated zones in local areas. Thus, it is easy for the surface pollutants to infiltrate into groundwater with precipitation. Especially, owing to intensive urban constructions and distributions of waste coal mines in this zone, karst groundwater is likely to be seriously polluted once the pollution is improperly prevented.

Low-risk and relatively low-risk zones are mainly distributed in the northwest, north, and southwest of the study area with an area about 100 km², taking up 30.56% of the total area. The precipitation is at a low–moderate level while the zones have large–moderate topographic slopes. These zones are mainly shown as monadnocks, slopes, peak clusters and valleys while carbonate rock with clastics and clastic aquifers are types of aquifers most widely seen. The karst poorly develops mainly appearing as karst fissures and holes and the thickness of unsaturated zones is at a moderate–thick level. Moreover, the lands are mainly used for forest and waters.

The moderate–risk zone is mainly distributed in the area between relatively low-risk and relatively high-risk zones covering an area about 92.06 km², which accounts for 25.54% of the total study area. In this zone, the karst favorably develops and there is a gentle topographic slope.
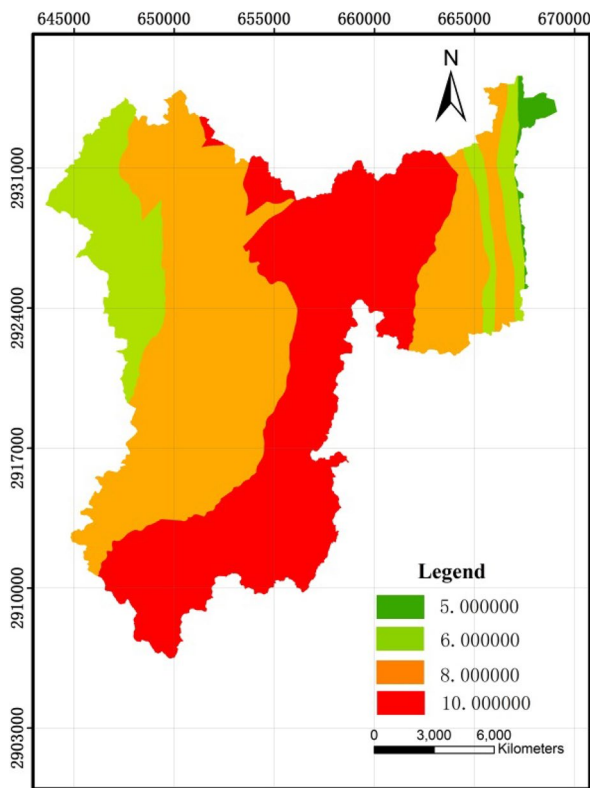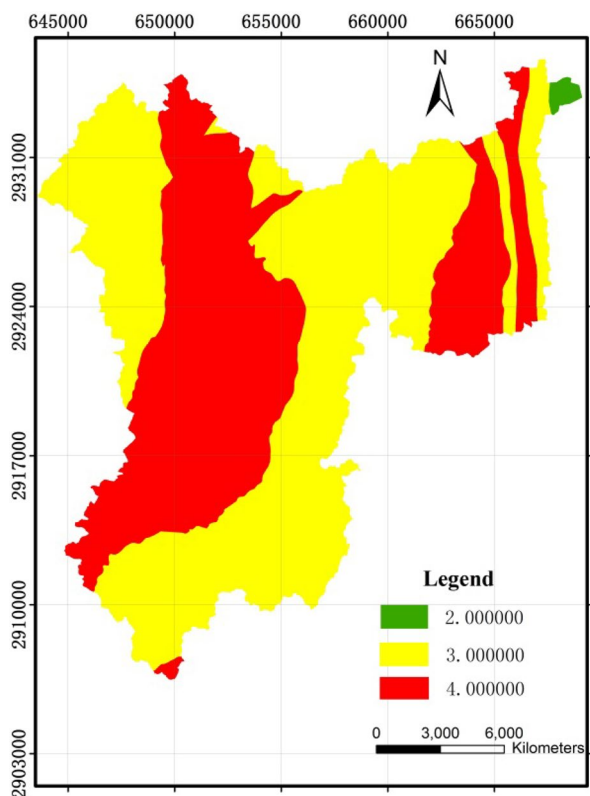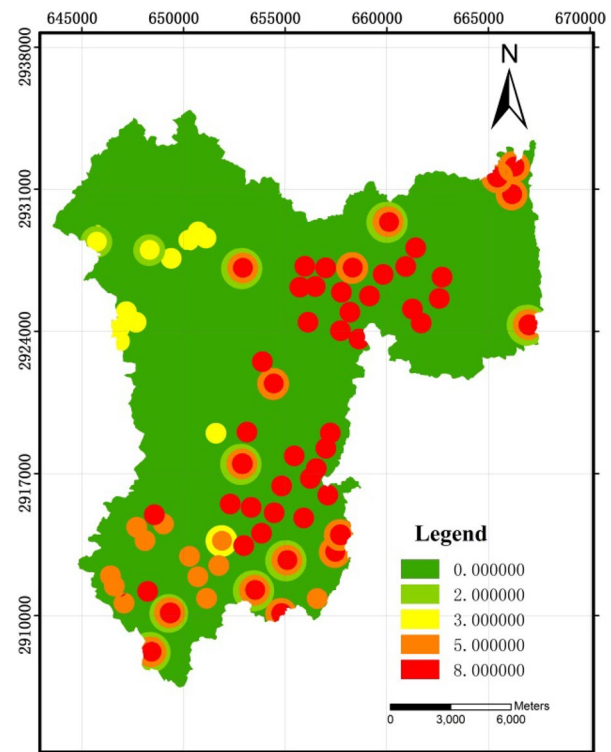
Fig. 11 Media types of unsaturated zones

Fig. 13 Influence scopes of pollution sources

The precipitation retains for a long time while the media of the unsaturated zones in the zone are not in favor of infiltration of groundwater.

Conclusions

Currently, the determination method of factor weights in the risk evaluation model of groundwater pollution is usually determined by subjective or objective weight; however, all of these methods are either too subjective or the weight value can not reflect the actual meaning of the factor. The artificial neural network method has the characteristics of obtaining the approximate relationship of system from the limited data set, which can better reflect the complicated relationship among groundwater pollution evaluation factors and, therefore, this study constructed the risk prediction model of groundwater pollution based on artificial neural network. Factor values of ten different locations in the study area were chosen as the training samples, the training sample's groundwater pollution risk as the output samples and the BP neural network was used to systematically identify unknown relationship among various factors, and then the factor weights were calculated. On this basis, combined with GIS technology, the risk evaluation of groundwater pollution in Huaxi District of Guiyang

Fig. 12 Types of aquifers
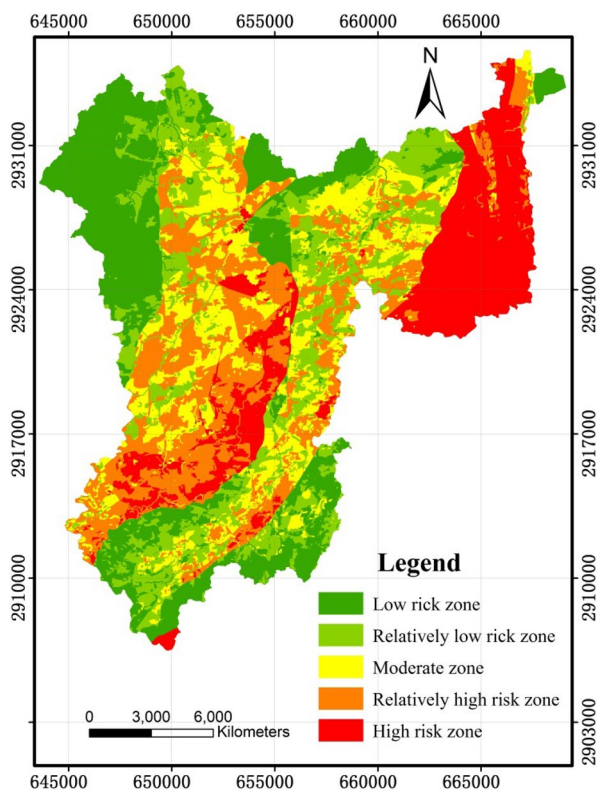
Fig. 14 The vulnerability zoning of karst groundwater

City was completed, and the method proved to be feasible through the analysis of the evaluation results. The groundwater pollution risk evaluation based on GIS-ANN technology can comprehensively and objectively reflect the characteristics of the groundwater pollution which is controlled by many factors and has a complicated nonlinear formation mechanism. At the same time, the evaluation results are intuitive, visual, and can provide a certain basis and reference for related research.

In southwest China, karst groundwater pollution is caused by the natural factors and human activities. Compared with other regions, the thickness of unsaturated zone is extremely uneven and corrosion effect is stronger, the unique karst development structure can connect groundwater with surface water. Furthermore, the unsaturated zone water retention capacity is very weak, which makes pollutants easier to wind its way to the groundwater, so in the process of prevention and control of groundwater pollution in this area, we must fully consider the particularity of the pollution approach and takes special prediction and control measures.

References

Aller L, Bennet T, Lehr JH, Petty RJ (1987) DRASTIC: a standardized system for evaluating groundwater pollution potential using hydrogeologic settings, US EPA Report 600/2-85/018. EPA, Washington

Dawson CW, Wilby RL (1999) A comparison of artificial neural networks used for river flow forecasting. Hydrol Earth Syst Sci 3(4):529–540

Deng K, Xiao L, Xu L, Song HY (2016) Prediction model of sports performance based on grey BP neural network. Int J u- e- Serv Sci Technol 9(8):87–96

Du HY, Zhao YL (2006) Comparison and research of artificial neural networks. Comput Technol Dev 16(5):97–99

Fan ZP, Ma J, Zhang Q (2002) An approach to multiple attribute decision making based on fuzzy preference information on alternatives. Fuzzy Sets Syst 131:101–106

Foster SSD, Skinner AC (1995) Groundwater protection: the science and practice of land surface zoning in groundwater quality. Proc Remed Protect Prague IAH Publ 225:471–482

Kaçaroğlu F (1999) Review of groundwater pollution and protection in karst areas. Water Air Soil Pollut 113(1–4):337–356

Kourakos G, Klein F, Cortis A, Harter T (2012) A groundwater non-point source pollution modeling framework to evaluate long-term dynamics of pollutant exceedance probabilities in wells and other discharge locations. Water Resour Res 48:W00L13

Li FW, Zhao Y, Feng P, Zhang W, Qiao JL (2015) Risk assessment of groundwater and its application. Part I: risk grading based on the functional zoning of groundwater. Water Resour Manag 29(8):2697–2714

Li B, Wu Q, Chen LX (2016) An analytical method of regional water resources carrying capacity in karst area—a case study in Guizhou province, China. Water Pract Technol 11(4):796–805

Liu ZC, He LS, Dong J, Meng R, Song BY, Xi BD (2012) Risk assessment of groundwater pollution for simple waste landfill. Res Environ Sci 25(7):833–839

Lu WX, Liang CY, Ding Y (2008) A method determining the objective weights of experts based on evidence distance. Chin J Manag Sci 16(6):95–99

Martin LC, Abraham JM (2001) Combined land –use and environmental factors for sustainable groundwater management. Urban Water 3:229–237

Masciopinto C, Vurro M, Palmisano VN, Liso IS (2017) A suitable tool for sustainable groundwater management. Waste Water Resour Manag 31(13):4133–4147

Ni XW, Liu DR, Ai L, Feng JM, Liang X, Ao XF (2017) Forecast water production of CBM reservoir from well logging data based on BP neural network algorithm. Coal Technol 36(9):105–107

Sangam S, Ranjana K, Vishnu PP (2015) Evaluation of index-overlay methods for groundwater vulnerability and risk assessment in Kathmandu Valley, Nepal. Sci Total Environ 575: 779–790

Sappa GS, Vitale (2001) Groundwater protection: contribution from Italian experience. Ministry of the Environment, Polish

Secunda S, Collin ML (1998) Groundwater vulnerability assessment using a composite model combining DRASTIC with extensive agricultural land use in Israel' s Sharon region. J Environ Manag 54:39–57

Shi ZT, Liu XY, Liu Y, Huang Y, Peng HY (1999) Catastrophic groundwater pollution in a karst environment: a study of phosphorus sludge waste liquid pollution at the Penshuidong Cave in Yunnan, China. Environ Earth Sci 59(4):757–763

Snell SS, Gopal S, Kaufman RK (2000) Spatial interpolation of surface air temperatures using artificial neural networks: evaluating their use for downscaling GCMs. J Clim 13(5):886–895

Tiwari RN, Mishra A, Dubey DP, Mishra UK (2012) Evaluation of pollution potential of groundwater, Rampur Baghelan area, Satna district, Madhya Prade. Int J Earth Sci Eng 5(4):703–709

Tiwari AK, De MM, Amanzio G (2015) Evaluation of metal contamination in the groundwater of the Aosta Valley Region, Italy. Int J Environ Res 11(3):291–300

Wang GS (2013) An improved BP neural network algorithm and it's application in customer classification. Int J Appl Math Stat 46(16):111–118

Wang ZX, Li QX (2006) An approach to integrate the final weights based on the subjective and objective weightsm. Commun Appl Math Comput 20(1):87–92

Wu Q, Li B, Chen YL (2016) Vulnerability assessment of groundwater inrush from underlying aquifers based on variable weight model and its application. Water Resour Manag 30(10):3331–3345

Yang Z, Ma C (2016) Risk prediction of water Inrush of Karst tunnels based on BP neural network. Tunnel Constr 36(11):1337–1348

Zhang LK, Qin XQ, Tang JS, Liu W, Yang H (2017) Review of arsenic geochemical characteristics and its significance on arsenic pollution studies in karst groundwater, Southwest China. Appl Geochem 77:80–88