Water Science and Engineering, **Jun. 2008, Vol. 1, No. 2, 1–9**
ISSN 1674–2370, *http://kkb.hhu.edu.cn*, e-mail: *wse@hhu.edu.cn*

Simulation of rainfall-underground outflow responses of a karstic watershed in Southwest China with an artificial neural network

Chen Xi,* Chen Cai, Hao Qingqing, Zhang Zhicai, Shi Peng

State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, P. R. China

Abstract: Karstic aquifers in Southwest China are largely located in mountainous areas and groundwater level observation data are usually absent. Therefore, numerical groundwater models are inappropriate for simulation of groundwater flow and rainfall-underground outflow responses. In this study, an artificial neural network (ANN) model was developed to simulate underground stream discharge. The ANN model was applied to the Houzhai subterranean drainage in Guizhou Province of Southwest China, which is representative of karstic geomorphology in the humid areas of China. Correlation analysis between daily rainfall and the outflow series was used to determine the model inputs and time lags. The ANN model was trained using an error backpropagation algorithm and validated at three hydrological stations with different karstic features. Study results show that the ANN model performs well in the modeling of highly non-linear karstic aquifers.

Key words: *karst*; *underground channel*; *correlation analysis*; *artificial neural network*

DOI: 10.3882/j. issn. 1674-2370.2008.02.001

1 Introduction

Karst terrain occupies 22 million km^2, which is about 15% of the world's land area, inhabited by approximately 1 billion people (Yuan and Cai 1988). Because of karstic landforms, such as poljes, cockpits, sinks and dolines, a limited amount of surface water is available, and groundwater is the primary water resource for domestic, agricultural and industrial utilization. Statistical results from Ford and Williams (1989) demonstrate that karstic water from shallow and deep aquifers supplies about 25% of the world's population. One of the largest karst areas in the world is located on the eastern side of the Yunnan-Guizhou Plateau of Southwest China（Figure 1). Centered in Guizhou Province, the southwestern karst area covers 42.62×10^4 km^2 and is home to a population of over 100 million (Wang et al. 2004). Karstic areas are approximately 95% of the land surface in Guizhou Province, and outcrops of soluble carbonate rocks cover 62% of the karstic areas (Wang et al. 2004). Of the total 1.76×10^5 km^2 of land area in Guizhou Province, 87% is mountainous plateau, 10% is hilly and only 3% is classified as flat (Zeng 1994). Mountainous areas with slopes of more than 15° account for 60% of the area (Li

---

This work was supported by the National Basic Research Program of China (973 Program, Grant No. 2006CB403200), the National Natural Scientific Foundation of China (Grant No. 50679025) and the 111 Project of the Ministry of Education and the State Administration of Foreign Expert Affairs, China (Grant No. B08048).
*Corresponding author (e-mail: *xichen@hhu.edu.cn*)
Received Apr. 24, 2008; accepted Jun. 1, 2008

et al. 2002).

Karstic watershed outflow strongly depends on karstic morphology and geological structure. In developed karstic areas with primarily soluble carbonate rocks, surface flows that run into a swallet or sinkhole may pass directly into a conduit or an underground channel. The major conduit system flows as open channel flow most of the time (Thrailkill et al. 1991). In this case, an underground outflow hydrograph shows a steep rise and decline. In less developed karstic areas, rock fractures are usually small and disconnected. A multi-layer underground conduit system may exist in areas of limestone mixed with dolomite. The major flow may occur in saturated horizontal conduits 10–100 m below the water table during the drought period and rise up to overflow the conduit during the flood period (Jeannin 2001). In this case, the hydrograph fluctuation is gentle due to the strong memory of the karstic system.

Spatially distributed numerical models can be used to simulate karstic hydrological processes (Angelini and Dragoni 1997; Larocque et al. 1999). However, simulating these heterogeneous systems requires large amounts of data for a variety of properties, including porosity, permeability and transmissivity. Parsimonious models, such as conceptual hydrological models, statistical models and ANN models, provide alternatives for karstic hydrological modeling, primarily because very few observations are needed. ANN models have been widely used to evaluate streamflow responses to climate variation in areas with sparse observation data (Chen et al. 2008) as well as very heterogeneous karstic areas (Hu et al. 2008).

The karstic aquifer used as an example in this study is in the Houzhai Watershed, located in Puding County, Guizhou Province, in Southwest China. This watershed has karstic geomorphology, rock fragments and land uses typical of karstic catchments in Southwest China. An ANN model was used to simulate variations of underground outflow through analysis of the memory of the karstic discharge and its response to rainfall.

2 Study area and site description

The study area, the Houzhai subterranean drainage in Puding County, Guizhou Province in Southwest China, with an area of 81 km^2, has typical karstic landforms, cone karst and cockpit karst geomorphology (Figure 1). Its elevation is high in the northwest and low in the southwest, varying from 1 200 m to 1 300 m above the sea level. The cone peaks of the Houzhai Watershed are generally 200–300 m above the adjacent doline depressions, and the cone surface relief and slope are much steeper.

The study site has a subtropical wet monsoon climate. The mean annual temperature is 20.1℃. The highest average monthly temperature is in July, and the lowest is in January. Annual precipitation is 920 mm, with a distinct summer wet season and a winter dry season. Average monthly humidity ranges from 74% to 78%.

The main rock type in the study area is limestone that belongs to the Guanling group (T^2g) of the middle Triassic Period (Wang et al. 2002). Various carbonatites that formed in the middle of the Triassic Period are spread all over the study area, and the karstic topography includes many bared funnels and sink holes and a well developed underground channel network. Buried karst is located in the valleys and poljes, which are surrounded by karstic mountainous peaks.

The Houzhai subterranean drainage system consists of three main surface rivers, the Dengzhan, Haoying and Houzhai rivers, and a main underground channel, the Dayouzhai-Maoshuikeng (MSK) (Figure 1) (Wang et al. 2002). The surface rivers are usually dry. The Qingshan Reservoir only receives surface water from the Dengzhan and Haoying rivers during flood periods. The reservoir outflow discharges into the Houzhai River, part of it entering the Dayouzhai-MSK underground channel and flowing westward to the watershed outlet at Houzhai (HZ) Station. Four hydrological observation stations from the upper watershed to the outlet along the Dayouzhai-MSK underground channel, Muzhudong (MZD), Laoheitan (LHT), Liugu (LG) and Maoshuikeng (MSK), record groundwater discharge and partial surface water discharge.

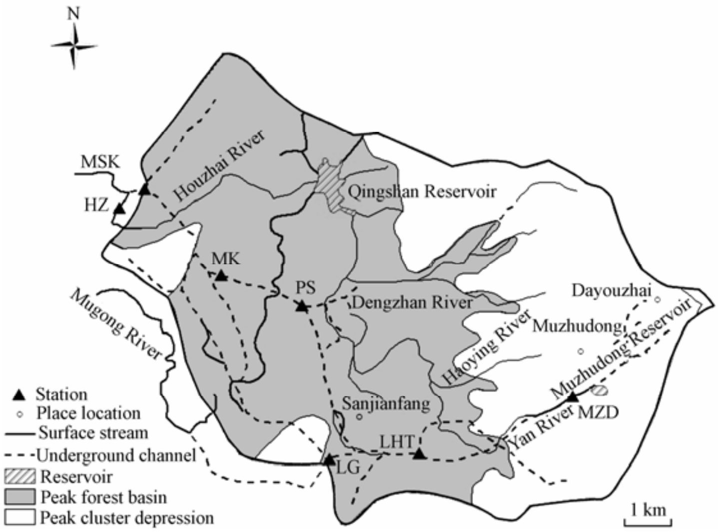

Figure 1 Houzhai karstic system and stream network

There are strong differences in the landform and aquifer characteristics between the upper and lower watersheds. The upper region at MZD Station consists of naked karstic peak clusters with many depressions (Wang et al. 2002). Rich ponors or funnels, which are directly linked with the underground channels, result in a sharp rise and drop of the hydrograph at MZD Station (Figure 2). When the watershed drainage reaches the isolated cones with broad and flat peneplains covered by soil at LHT Station, the dendritic surface and underground water systems overlap or become cross-connected. The enhanced regulating ability of the water channel system leads to a relatively smooth hydrograph at LHT Station. As the watershed drainage system continues downstream to MSK Station, surface water from the Houzhai River joins the underground drainage system. The outflow from MSK Station remains relatively large during flood periods (Figure 2). Characteristics of the hydrologic processes at MZD, LHT and MSK stations reveal the influences of karstic geomorphologic conditions and the drainage system on the rainfall-runoff relationship.

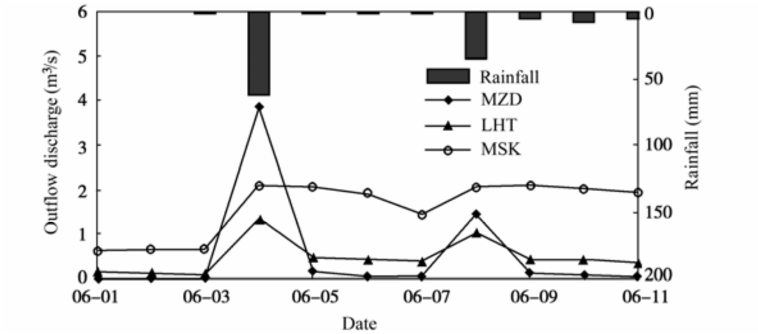

Figure 2 A rainfall event in 1998 and its corresponding outflow discharge at MZD, LHT and MSK stations

3 Artifical neural network (ANN)

An ANN is a network of interconnected neurons or nodes, $u_1,\ldots, u_j,\ldots, u_n$. Each neuron consists of a number of input nodes (in this case, inputs to outflow, $x_1,\ldots, x_i,\ldots, x_m$) and contributes to outflow, y. Thus, a neural network model can be written as

$$y = f(u_j) \tag{1}$$

$$u_j = \sum_{i=1}^{m} w_{ij} x_i - \theta_j \tag{2}$$

In Eq. (2), w_{ij} is the weight vector of neuron j, and θ_j is a critical value. The function f in Eq. (1) is a logistic sigmoid function or a hyperbolic tangent function (Dawson and Wilby 1998, 2001).

$$f(u) = \frac{1}{1+\mathrm{e}^{-u}} \tag{3}$$

ANNs applied in rainfall-runoff modeling are usually multi-layer feed-forward networks (Campolo et al. 1999; Hu et al. 2008). However, obtaining the number of hidden layers and the number of neurons in each hidden layer is not straightforward, and no rules are available to determine the exact numbers. These numbers are usually determined by a trial-and-error procedure. Initial values can be set according to previous study experience. Two hidden layers have been proven to provide the flexibility necessary to model complex-shaped solution surfaces and can be used as a starting point for developing a layered feed-forward network of sigmoidal neurons (Flood and Kartam 1994). The number of neurons in the hidden layer is initially set between $2\sqrt{m}+n$ and $2m+1$, according to Hajela and Berke (1991) and Fletcher and Goss (1993), where m and n represent the numbers of input and output neurons, respectively. Then, neurons are added or removed on a trial-and-error basis in order to find the optimal number of hidden neurons while maintaining a stable solution for the model.

4 Application of ANN to karstic underground outflow modeling

4.1 Determining appropriate inputs and outputs

Karstic groundwater moves from the recharge area to the discharge area, traveling a long distance through the groundwater aquifers. Karstic groundwater always has long residence time, the time lag between rainfall and underground outflow from the watershed. Underground outflow discharge is related not only to the current rainfall but also to prior rainfall and outflow discharge. The form of the relationship between the rainfall and underground outflow discharge can be generally approximated as

$$Q(t) = f\left[P(t), P(t-\Delta t), \cdots, P(t-N\Delta t), Q(t-\Delta t), \cdots, Q(t-M\Delta t)\right] \tag{4}$$

where Q is underground outflow discharge at the outlet of the watershed, P is rainfall, Δt is the data sampling interval, N and M are positive integers reflecting the memory length of the watershed and f is the appropriate model structure (i.e. the mathematical functions) (Chow et al. 1988; Hsu et al. 1995; Hu et al. 2008). In Eq. (4), the function f maps $Q(t-M\Delta t)$ to $Q(t)$ with the knowledge of rainfall at t and $t-N\Delta t$. f is determined by using an artificial neural network model.

The memory length or time lag, N and M, can be determined by the autoregressive moving average (ARMA) model and the correlation test proposed by Maier and Dandy (2000) and Dawson and Wilby (1998). The contribution of daily rainfall $P(t), \cdots, P(t-N\Delta t)$ and outflow discharge $Q(t-\Delta t), \cdots, Q(t-M\Delta t)$ to $Q(t)$ is determined by a correlation test with the significance level of 0.05.

The cross-correlation method permits the comparison of the rainfall time series with the outflow discharge time series. This technique provides two kinds of information: the strength of the relationship between the two series and the lag between them. In this study, daily rainfall and underground outflow during the rainfall periods of 1980–1989 for MSK Station, 1996–2002 for LHT Station and 1990–2002 for MZD Station were used to calculate cross-correlation coefficients. Figure 3 shows that the most recent rainfall strongly influences the karstic system discharge on a given day.

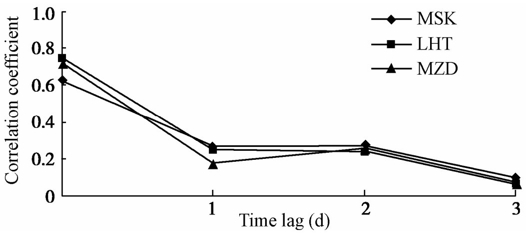

Figure 3 Coefficients of cross-correlation between rainfall and outflow discharge

The autocorrelation function (ACF) measures the amount of linear dependence between observations in a time series of outflow discharge $(Q(t), Q(t-1), \cdots, Q(t-M\Delta t))$. The ACF method is a tool for identifying some overall characteristics of a discharge time series, particularly cyclic variations. These variations may be related to some structural characteristics of karstic aquifers. The ACF can be extended to create a partial autocorrelation function (PACF) in which there is no dependence on intermediate elements within the lag. Both ACF and PACF values reveal dependence between observations in a time series of outflow discharge resulting from karstic aquifer memory.

The time required for the correlogram to drop below 0.2 is called memory effect (Mangin 1971). High memory indicates a poorly developed karstic network with large groundwater flow reserves (storage). In contrast, low memory is believed to reflect low storage in a highly karstified aquifer. Consequently, the shape of the correlogram and the memory effect derived from it depend on the maturity of the karstic system.

The ACF and PACF were calculated on the basis of daily outflow discharge of 1980–1989 for MSK Station, 1996–2002 for LHT Station and 1990–2002 for MZD Station (Figure 4). Based on these graphs, the values of both the ACF and PACF are larger than 0.2 for one day of the time lag at MZD Station, two days of the time lag at LHT Station and three days of the time lag at MSK Station.

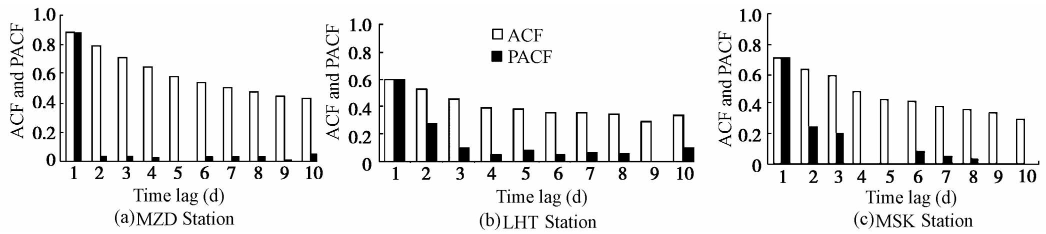

Figure 4 ACF and PACF of daily outflow for three stations

4.2 Model training and validation

Among various types of algorithms used to train the network, the backpropagation method (Rumelhart et al. 1986; Wasserman 1989; Fausett 1994) has been proven to be one of the most successful both for a gradient descent technique that minimizes the network error and for designing multi-layer feed-forward neural networks (Haykin 1994). The backpropagation method has been widely applied in rainfall-streamflow simulation (French et al. 1992; Gautam et al. 2000; Wilby et al. 2003; Chen et al. 2008).

The available data were divided into two subsets: a training period and a validation period (Table 1). In training the model, all input and output data were normalized to an internal representation between 0 and 1 so that they would receive equal attention in the training process (Maier and Dandy 2000).

ANN inputs and outputs were determined on the basis of the ACF and PACF. Two hidden layers were adopted. The number of neurons in the hidden layer was determined in the model training using the trial-and-error procedure, according to Hajela and Berke (1991) and Fletcher and Goss (1993). As the calculated result, there are seven hidden-layer neurons for MZD

Station, seven for LHT Station and nine for MSK Station. Therefore, the optimal network topologies, expressed by the number of inputs, hidden layers, hidden-layer neurons and outputs, are 2-2-7-1, 3-2-7-1, and 4-2-9-1 for MZD, LHT and MSK stations, respectively.

Table 1 Performance indices of ANN model of Houzhai Watershed

Station	Period	C_{NS}	E_{rms} (m^3/s)	E_{ra} (%)
MZD	Training (1990–1998)	0.884	0.335	72.56
	Validation (1999–2002)	0.832	0.398	91.44
LHT	Training (1996–1999)	0.860	0.121	15.37
	Validation (2000–2002)	0.859	0.161	20.24
MSK	Training (1980–1986)	0.935	0.126	10.54
	Validation (1987–1989)	0.873	0.250	15.34

To determine an appropriate set of weights in Eq. (2), the model was trained using the error backpropagation algorithm and momentum was used to speed convergence to an error minimum. Based on the error signal received, connection weights were updated so that the network converged at a stable state that allowed all the training patterns to be encoded. This process was guided using the Nash-Sutcliff efficiency coefficient (C_{NS}), the root-mean-square error (E_{rms}) and the mean of the relative absolute error (E_{ra}) between observed and simulated daily discharge. The calculation was performed using MATLAB software. The error criterion goal for training was 0.0001 and an initial value of outflow was randomly generated.

Model training and validation results are listed in Table 1. The simulations demonstrate that larger errors between observation and simulation result from extremely small outflow discharges. About half of the discharge values at MZD Station are smaller than 0.1 m^3/s. It is usually difficult to accurately measure and simulate these extremely small discharges because a large E_{ra} could result from a small measurement or simulation error. When the discharge at MZD Station is compared with the larger discharge at LHT and MSK stations, the model performance indices demonstrate that an increase in the number of extremely small discharge observations corresponds to a reduction in the accuracy of the ANN simulation. Although the accuracy of the simulated discharge in the validation period is reduced, C_{NS} values for the three stations are still higher than 0.83. Generally, as shown in Figure 5, the ANN model is reliable in describing the rainfall-underground outflow response and simulating flood discharge in the Houzhai Watershed.

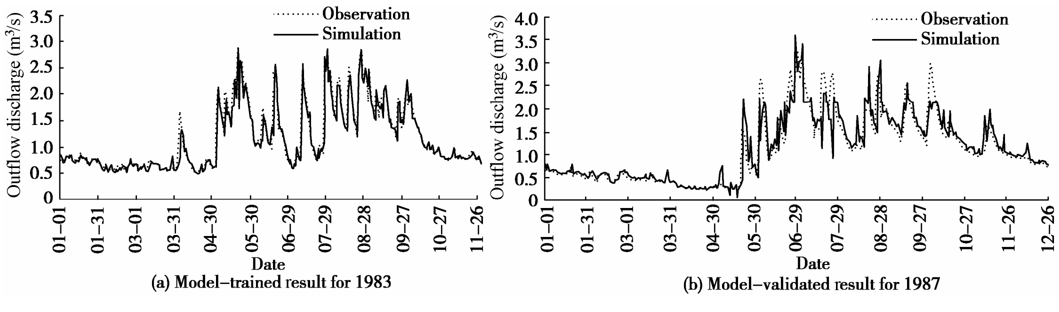

Figure 5 Model-trained results of outflow discharge for 1983 and validated results for 1987 at MSK Station

5 Conclusions

Karstic groundwater is very important to the regional economy as it provides drinking water to residents and water resources for agricultural and industrial utilization. This study developed an artificial neural network model to describe the relationship between rainfall and underground outflow in three regions of the Houzhai Watershed with different geomorphology, surface stream and underground stream channel characteristics.

The correlation tests allow determination of the optimum temporal information to be included in the rainfall and outflow input data in order to simulate the *i*th day discharge. Study results demonstrate that the daily underground outflow throughout the whole Houzhai Watershed is significantly influenced by the corresponding daily rainfall, as well as rainfall in the preceding days. The number of days preceding the *i*th day with significant influence on outflow discharge is one at MZD Station, two at LHT Station and three at MSK Station. The ANN structures for inputs, hidden layers, hidden-layer neurons and outputs are 2-2-7-1, 3-2-7-1 and 4-2-9-1 for MZD, LHT and MSK stations, respectively.

The training of the ANN model is accomplished using the backpropagation Levenberg-Marquardt algorithm. The Nash-Sutcliff efficiency coefficients (C_{NS}) between simulated and observed daily discharges are higher than 0.83 at MZD, LHT and MSK stations. The overall shape of the predicted hydrograph is equivalent to that of the observed hydrograph. Instances of poor performance of the ANN model at three stations, particularly at MZD Station, primarily result from an outflow series that contains many extremely small discharge observations. The developed ANN model, together with correlation analysis, reveals ways in which the karstic aquifer regulates the rainfall-underground outflow relationship. It can be further used for prediction of underground outflow responses to climatic variations, and for estimation of available groundwater resources.

Acknowledgements

The Puding Karst Geological and Ecological Survey Station in Guizhou Province, China, provided necessary observation data.

References

Angelini, P., and Dragoni, W. 1997. The problem of modeling limestone springs: The case of Bagnara (North Apennines, Italy). *Ground Water*, 35(4), 612–618.

Campolo, M., Andreussi, P., and Soldati, A. 1999. River flood forecasting with a neural network model. *Water Resources Research*, 35(4), 1191–1197.

Chen, X., Wu, J., and Hu, Q. 2008. Simulation of climate change impacts on streamflow in the Bosten Lake Basin using an artificial neural network model. *Journal of Hydrologic Engineering*, 13(3), 180–183.

Chow, V. T., Maidment, D. R., and Mays, L. W. 1988. *Applied Hydrology*. New York: McGraw-Hill.

Dawson, C. W., and Wilby, R. L. 1998. An artificial neural network approach to rainfall-runoff modeling. *Hydrological Sciences Journal*, 43(1), 47–66.

Dawson, C. W., and Wilby, R. L. 2001. Hydrological modelling using artificial neural networks. *Progress in Physical Geography*, 25(1), 80–108.

Fausett, L. 1994. *Fundamentals of Neural Networks: Architectures, Algorithms and Applications*. Englewood Cliffs: Prentice Hall.

Fletcher, D., and Goss, E. 1993. Forecasting with neural networks: An application using bankruptcy data. *Information and Management*, 24(3), 159–167.

Flood, I., and Kartam, N. 1994. Neural networks in civil engineering II: Systems and application. *Journal of Computing in Civil Engineering*, 8(2), 131–148.

Ford, D. C., and Williams, P. W. 1989. *Karst Geomorphology and Hydrology*. London: Unwin Hyman.

French, M. N., Krajewski, W. F., and Cuykendall, R. R. 1992. Rainfall forecasting in space and time using a neural network. *Journal of Hydrology*, 137(1–4), 1–31.

Gautam, M. R., Watanabe, K., and Saegusa, H. 2000. Runoff analysis in humid forest catchment with artificial neural network. *Journal of Hydrology*, 235(1–2), 117–136.

Hajela, P., and Berke, L. 1991. Neurobiological computational models in structural analysis and design. *Computers & Structures*, 41(4), 657–667.

Haykin, S. 1994. *Neural Networks: A Comprehensive Foundation*. New York: IEEE Press, McMillan College Publishing.

Hu, C., Hao, Y., Yeh, T.-C. J., Pang, B., and Wu, Z. N. 2008. Simulation of spring flows from a karst aquifer with an artificial neural network. *Hydrological Processes*, 22(5), 596–604.

Hsu, K.-L., Gupta, H. V., and Sorooshian, S. 1995. Artificial neural network modeling of the rainfall-runoff process. *Water Resources Research,* 31(10), 2517–2530.

Jeannin, P.-Y. 2001. Modeling flow in phreatic and epiphreatic karst conduits in the Hölloch Cave (Muotathal, Switzerland). *Water Resources Research*, 37(2), 191–200.

Larocque, M., Banton, O., Ackerer, P., and Razack, M. 1999. Determining karst transmissivities with inverse modeling and an equivalent porous media. *Ground Water*, 37(6), 897–903.

Li, R., Wang, S., and Zhang, D. 2002. The role of man-made factors in eco-environmental detection in Guizhou karst areas. *Bulletin of Mineralogy, Petrology and Geochemistry*, 21(1), 43–47.

Maier, H. R., and Dandy, G. C. 2000. Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environmental Modelling & Software*, 15(1), 101–123.

Mangin, A. 1971. Etude des débits classés d'exutoires karstiques portant sur un cycle hydrologique. *Annales de spéléologie*, 28(1), 21–40. (in French)

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. 1986. Learning internal representations by error propagation. Rumelhart, D. E., and McClelland, J. L., eds. *Parallel Distributed Processing*, 318–362. Cambridge: MIT Press

Thrailkill, J., Sullivan, S. B., and Gouzie, D. R. 1991. Flow parameters in a shallow conduit-flow carbonate aquifer, Inner Bluegrass Karst Region, Kentucky, USA. *Journal of Hydrology*, 129(1–4), 87–108.

Wang, L. C., Zhang, J. X., Zhou, Y. K., Chen, X. L., and Du, J. K. 2002. Approach of deterministic geomorphologic instantaneous unit hydrograph to hydrological processes' simulation in Karst area. *Chinese Geographical Science*, 12(4), 354–358.

Wang, S. J., Liu, Q. M., and Zhang, D. F. 2004. Karst rocky desertification in southwestern China: Geomorphology, land use, impact and rehabilitation. *Land Degradation & Development*, 15(2), 115–121.

Wasserman, P. D. 1989. *Neural Computing: Theory and Practice*. New York: Van Nostrand Reinhold.

Wilby, R. L., Abrahart, R. J., and Dawson, C. W. 2003. Detection of conceptual model rainfall–runoff processes inside an artificial neural network. *Hydrological Sciences Journal*, 48(2), 163–181.

Yuan, D. X., and Cai, G. H. 1988. *The Science of Karst Environments*. Chongqing: Chongqing Publishing House. (in Chinese)

Zeng, Z. 1994. Suggestion on poverty-deviation in the karst mountain areas in South China. Xie, Y. H., and Yang, M. D., eds. *Human Activity and Karst Environments*, 15–19. Beijing: Beijing Science and Technology Press. (in Chinese)