Integration of Grey with Neural Network Model and Its Application in Data Mining

Changjun Zhu
College of Urban Construction, Hebei University of Engineering, Handan, 056038, China
christorf@126.com

Qinghua Luan
College of Water Conservancy and Hydropower, Hebei University of Engineering, Handan, 056038, China

Zhenchun Hao, Qin Ju
State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, China

***Abstract*—Because of Boundary types and geologic conditions, which possess random and obscure characteristics, groundwater heads vary with the conditions. The prediction of groundwater level is one of the main work of hydraulic government, which is predicted based on the history data and the relative influence factors. Therefore, prediction precision depends on the accuracy of history data. Data mining has provided a new method for analyzing massive, complex and noisy data. According to the complexity and ambiguity of groundwater system, a new integration of grey with neural network model is built to forecast groundwater heads, which were used to judge whether future groundwater heads were extraordinarily over the history range or not. This method overcomes the disadvantages which the grey method only predict the linear trend. The methods were used to analyze the random characteristics of groundwater heads in anyang city. The results indicate that the method is reliable, and reasonable.**

***Index Terms*—grey degree, groundwater level, neural network, Anyang city**

I. I. INTRODUCTION

The Gray systematic theory is proposed by Chinese professor Deng Jurong (1985) in 1982. In the theory, there is not only a large amount of known information called white system, but also much unknown and uncertain information called black system. The system including white system and black system is called gray system. Contaminant transport in an underground system usually occurs in varied flow fields and in anisotropic and heterogenous media. Because the applicability of analytical solutions is extremely limited for such conditions, numerical techniques are essential for underground pollution modeling. Mecarthy(1989) , Basha & El-Habel(1993) , Li *et al.*(2004) ， Li & Wang(2004) made much work about the uncertain issues. Among the numerical techniques, the gray numerical method has become very popular and is recognized as a powerful numerical tool. The distribution and transport of pollutants mentioned by Liu *et al.*(2003) ,Chen & Wagenet(1995) , Xu *et al.*(2002) in groundwater are controlled by physical chemistry and biology functions, which include advection, diffusion, dispersion, sorption, decay and biodegradation. In the courses，there is not only the known information but also uncertain information, therefore, it can be seen as one gray system.

Groundwater trends are related to atmosphere precipitation. There is obvious seasonality variation of atmosphere precipitation. Therefore, the groundwater level is changed by the seasons and periods. In this case, we needed to use the original data of the groundwater in order to predict the groundwater level. In this paper, we calculated the cycle groundwater, trend groundwater, and random groundwater separately after isolating the three items. Traditionally, we analyzed and predicted groundwater using the determinacy and random statistical mathematics model (such as, finite element, finite difference, analysis, harmony wave analysis, time series analysis, probability statistic, etc) . These methods are mainly based on the linear theory. Since the models are simple, the precision is not high (Zhang *et al*., 2002; Luo *et al*., 2003) .

The development and management of groundwater resources should include trends in groundwater head variation. In this study, the statistics estimation method was used to build a wrapping band for groundwater level variation based on groundwater observation data.

II. PREDICTION METHOD

A. Principle of GM (1,1)

The essence of GM (1,1) [5] is to accumulate the original data in order to obtain regular data. By setting up the differential equation model, we obtain the fitted curve in order to predict the system. Assuming the observed original water quality is as follows:

$$\left\{x^{(0)}(t)=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\right\}$$

The data were treated with an accumulated generating operation(AGO) :

$$\left\{x^{(1)}(t)=(x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n)\right\} \quad (1)$$

doi:10.4304/jsw.6.4.716-723

Where: $x^{(1)}(k)=\sum_{t=1}^{k}x^{(0)}(i)$, $k=1,2,\cdots n$;

Mean value of $x^{(1)}$ $x=(x(2),x(3),\cdots x(n))$

Where:

$$x(t)=\frac{1}{2}(x^{(1)}(t)+x^{(1)}(t-1))\ \ t=2,3,\cdots,n$$

Corresponding differential equation:

$$\frac{dx^{(1)}}{dt}+ax^{(1)}=u \tag{2}$$

Utilizing least square method to solve parameters a,u. Where t is the time, a and u are the parameters to be determined.

$$\hat{a}=\begin{bmatrix}a\\u\end{bmatrix}=(B^TB)^{-1}B^TY_N \tag{3}$$

where $B=\begin{bmatrix}-x(2) & 1\\-x(3) & 1\\ \vdots & \vdots\\-x(n) & 1\end{bmatrix}$ $\quad Y_N=\begin{bmatrix}x^{(0)}(2)\\x^{(0)}(3)\\ \vdots\\x^{(0)}(n)\end{bmatrix}$

Grey predicting model of $x^{(1)}$:

$$\hat{x}^{(1)}(t+1)=(x^{(0)}(1)-\frac{u}{a})e^{-at}+\frac{u}{a} \tag{4}$$

Grey predicting model of $x^{(0)}$

$$\hat{x}^{(0)}(t+1)=(1-e^{a})(x^{(0)}(1)-\frac{u}{a})e^{-at}$$

(5)

In order to differentiate models good or bad, the methods are adopted as follows:

$$e_k=x^{(0)}(k)-\hat{x}^{(0)}(k)\ \ k=1,2,\cdots,n)$$

$$\bar{e}=\frac{1}{n}\sum_{k=1}^{n}e_k \qquad \bar{x}^{(0)}=\frac{1}{n}\sum_{t=1}^{n}x^{(0)}(t)$$

$$S_1^2=\frac{1}{n}\sum_{k=1}^{n}[x^{(0)}(k)-\bar{x}]^2 \qquad S_2^2=\frac{1}{n}\sum_{k=1}^{n}[e(k)-\bar{e}]^2$$

While $c=\frac{S_2}{S_1}<0.35$ $\quad P=\{|e_k-\bar{e}|<06745S_1\}>0.95$, the precision of the models is very accurate indeed.

B. Prediction of grey interval

Assuming the mean value , $\bar{x}$,and variance, σ^2 of the groundwater level in every month can be found. Accorinf to the large numbers principles, when the number of the data is enough , the mean of samples $\bar{x}$ can be expressed as a normal distribution, $N(\mu,\frac{\sigma}{\sqrt{n}})$,and $\frac{(\bar{x}-\mu)}{(\frac{\sigma}{\sqrt{n}})}$ satisfies normal distribution, where μ is the mean of population. If the confidence level, $(1-\alpha)$, of μ is given, then the confidence interval of μ is $[\bar{x}-Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\bar{x}+Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}]$, where $Z_{\frac{\alpha}{2}}$ is the accumulated probability corresponding to

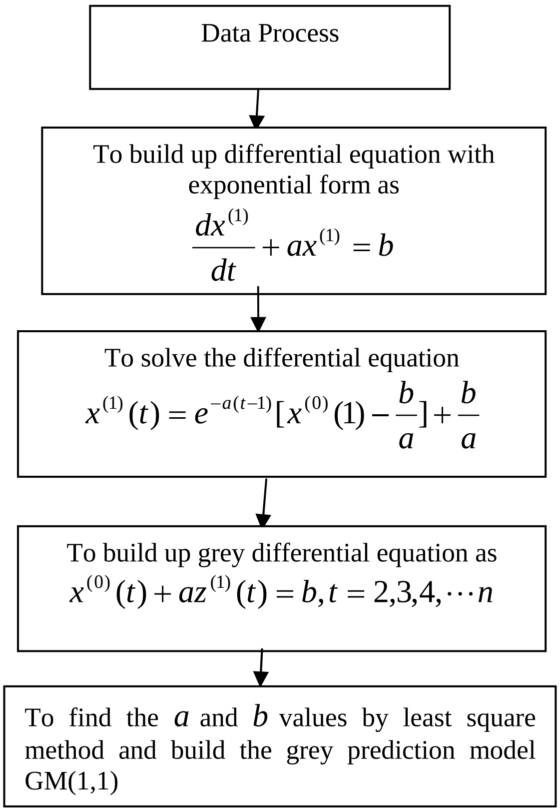

Figure 1. Flow chart of grey method of groundwater level

standard normal value at $\frac{\alpha}{2}$ (Y.-L.Yeh, T.-C.Chen,2004, Hogg RV, Tanis EA 1988)

In this study, the confidence interval of groundwater level can be as the grey interval.. the grey degree can be defined as Eq. (6)

$$G_d=\frac{n_{out}}{n_{all}} \tag{6}$$

C. RBF Neural network

A radial basis function (RBF) is a real-valued function whose value depends only on the distance from the origin, so that $\phi(x)=\phi(\|x\|)$; or alternatively on the distance from some other point *c*, called a *center*, so that $\phi(x,c)=\phi(\|x-c\|)$. Any function φ that satisfies the property $\phi(x)=\phi(\|x\|)$ is a radial function. The norm is usually Euclidean distance.

Radial basis functions are typically used to build up function approximations of the form

$$y(x)=\sum_{i=1}^{N}w_i\phi(\|x-c_i\|) \tag{7}$$

where the approximating function *y*(*x*) is represented as a sum of *N* radial basis functions, each associated with a different center $\boldsymbol{c}_i$, and weighted by an appropriate coefficient w_i. Approximation schemes of this kind have been particularly used in time series prediction and control of nonlinear systems exhibiting sufficiently simple chaotic behaviour, 3D reconstruction in computer graphics (for ex. hierarchical RBF) .

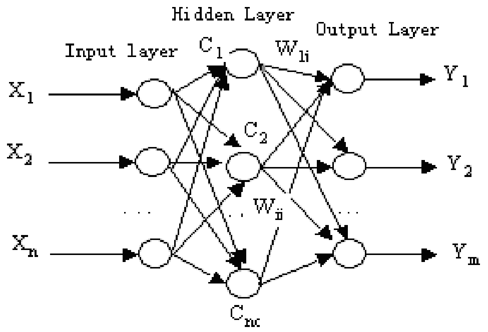

Figure 2. RBF neural network Structure

There are three learning parameters:: center of RBF C_K、σ_K W_K . The process of algorithm as followings:

(1) choosing one series C_K from input matrix;

(2) Calculating variance

$$\sigma = \frac{d_{\max}}{K} \tag{8}$$

Where $d_{\max}$ is the maximum distance, K is the number of C_K ;

(3) Calculating $\hat{y}_i(n)$ according to $x(n)$

$$\hat{y}_i(n) = \sum_{k=1}^{M} W_K \phi[x(n), C_K, \sigma_K] \tag{9}$$

(4) update RBF parameters

$$W(n+1) = W(n) + \mu_w e(n)\phi(n) \tag{10}$$

$$C_K(n+1) = C_K(n) + \mu_C \frac{e(n)W_K(n)}{\sigma_K{}^2(n)} * \phi[x(n), C_K(n), \sigma_K][x(n) - C_K(n)] \tag{11}$$

$$\sigma_K(n+1) = \sigma_K(n) + \mu_\sigma \frac{e(n)W_K(n)}{\sigma_K{}^2(n)} \phi[x(n), C_K(n), \sigma_K] \| [x(n) - C_K(n)] \|^2 \tag{12}$$

$$\phi(n) = \begin{Bmatrix} \phi[x(n), c_1(n), \sigma_1], \phi[x(n), c_2(n), \sigma_2], \\ \ldots, \phi[x(n), c_N(n), \sigma_N] \end{Bmatrix}^T \tag{13}$$

$$e(n) = \hat{y}_i(n) - y_d(n) \tag{14}$$

$y_d(n)$ is the expired output; μ_N, μ_C, μ_σ are the echoes of three parameters.

(5) If network convergence, stop the calculation, otherwise go to process (3) . The detail procedures can be seen in figure2.

D. Optimum model of Combined Gray Neural Network

ANN has many types, according to the characters, three-layer RBF is considered as the basic network for establishment of the model. RBF neural network is one of the most important network in ANN at present, and so far, it is widely used in hydrological science. The GM-ANN coupling model is based on the GM model of groundwater, using the GM model of groundwater to simulate the groundwater level, then the groundwater level of different observation wells is as the input of the neural network to predict the groundwater level.

III. Application

A. Study area

Spring area of small south China sea is located in linzhou city and anyang county in henan province. River water system belongs to Wei river water system in haihe basin, the main river is huan river, and there are a Xiaonanhai revervior and "artificial Tianhe" –red flag canal. The vents is in Huan river valley, the exposed elevation is between 131-135m. Overall look in spring, western mountains and Lin basin aer recharge are, and the central low mountains is runoff area, and eastern Xinan sea is discharge area, whose total area is 934.6Km^2. karst groundwater is buried deeply, the minimum depth is more than 10m, the influence of evaporation of groundwater on level is very small, and the type of karst water discharge is mainly water drainage, mine drainage and artificial mining. The precipitation and evaporation measured in 2003, 2004 ,2005 at 6# station is shown in figure.5.

Spring is located in transition between taihang uplift and north China plain settlement, whose west is Linzhou fault rupture, east to Tangxi fault rupture.

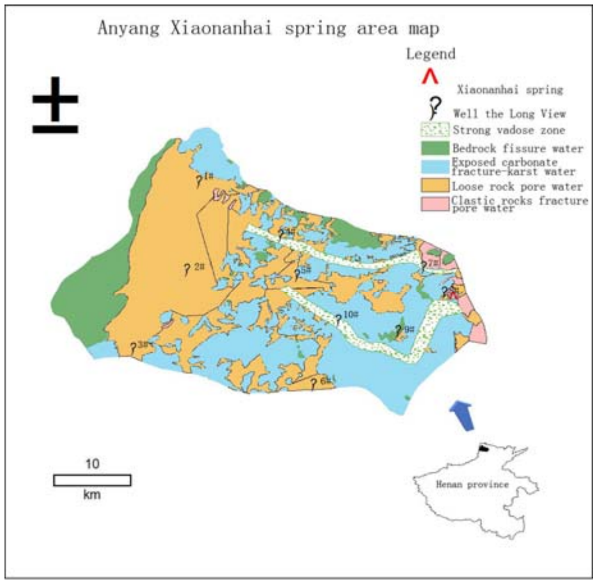

Figure 3. Hydrogeologic map in the Xiaonanhai spring catchment

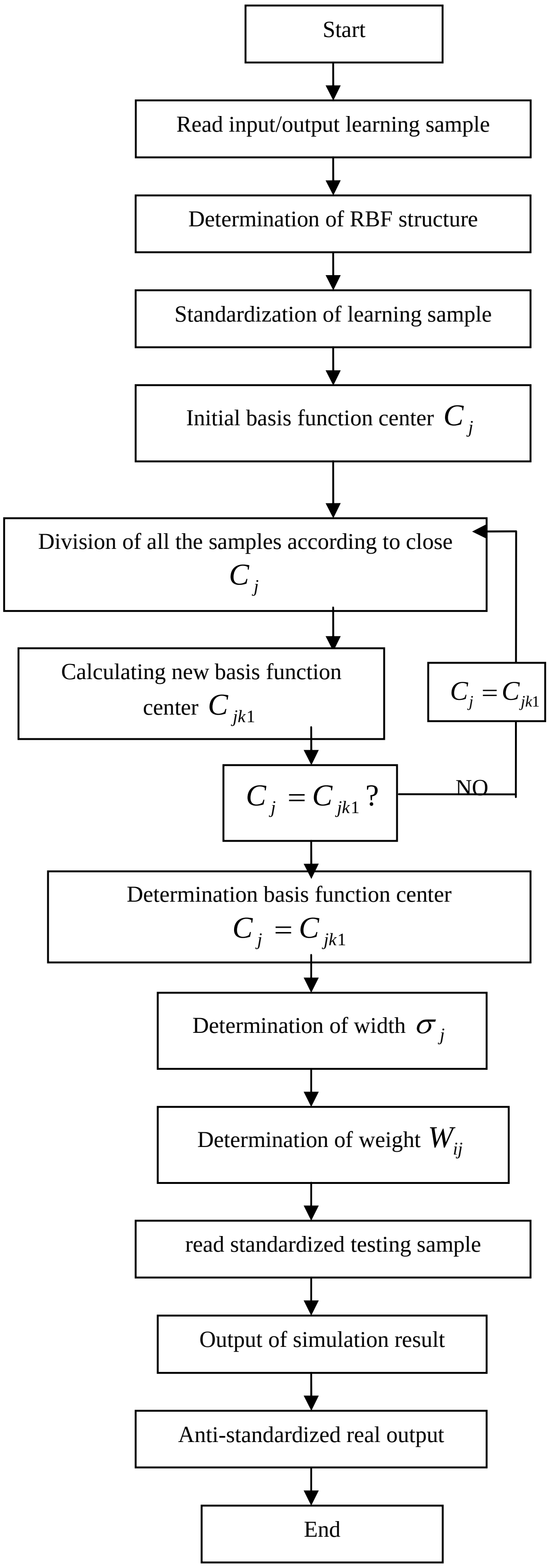

Figure 4. Flow chart of RBF neural network

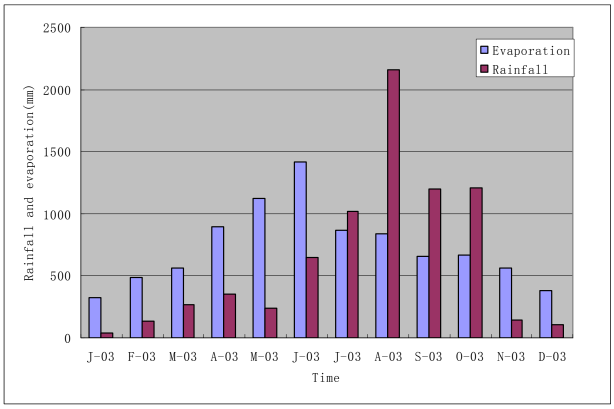

Figure 5. Monthly rainfall and evaporation as measured in 2003 at 6# Station

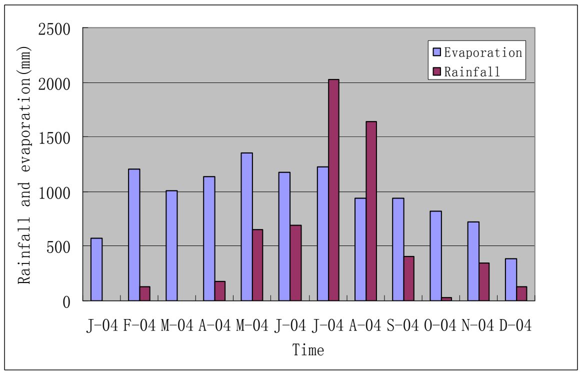

Figure 6. Monthly rainfall and evaporation as measured in 2004 at 6# Station

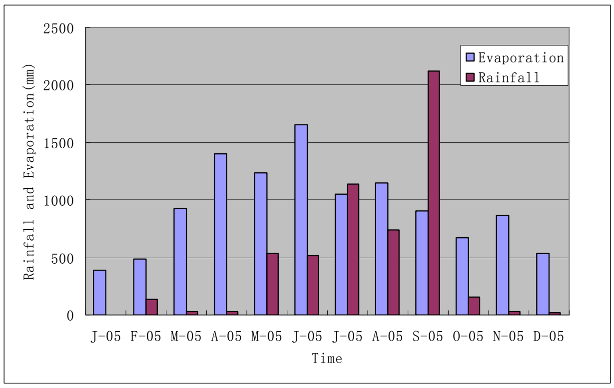

Figure 7. Monthly rainfall and evaporation as measured in 2005 at 6# Station

B. Data collection and analysis

There are 10 automated observation well in Xiaonanhai spring area. Figure 9 shows the observed groundwater level in 6# station. The statistical results of this station are listed in Table.1. The minimum groundwater level was in January and the maximum in December. The greatest variance in groundwater level was in August and the smallest in May.

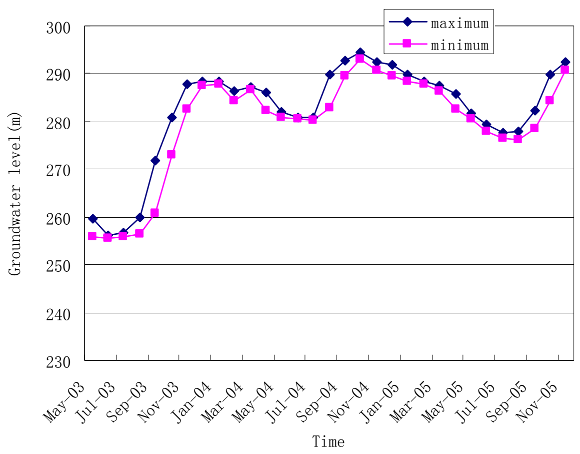

Figure 8. The average groundwater level at 6# station

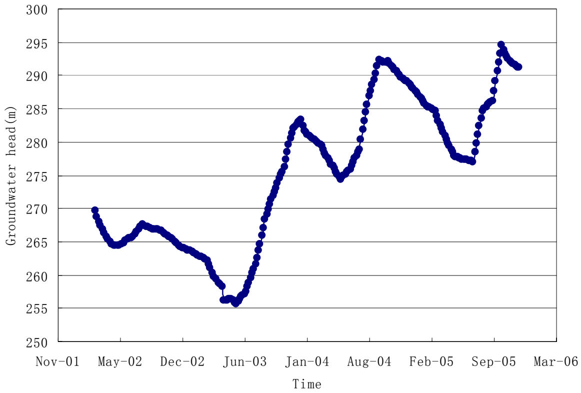

Figure 9. Observation groundwater level in 6# station

TABLE I. STATISTICAL ANALYSIS OF GROUNDWATER LEVEL IN ANYANG STATION IN 2004

Month	Max	Min	Ave	Var
Jan-04	251.56	249.62	250.6583	0.503
Feb-04	252.15	251.93	252.0333	0.007
Mar-04	252.64	252.24	252.5317	0.022
Apr-04	252.63	252.44	252.5533	0.005
May-04	252.34	252.19	252.2433	0.002
Jun-04	252.13	251.89	252.0067	0.008
Jul-04	252.54	251.37	251.79	0.180
Aug-04	255.04	252.44	253.7367	0.946
Sep-04	257.84	255.57	256.755	0.728
Oct-04	259.86	258.21	259.0517	0.392
Nov-04	261.44	260.32	260.9267	0.189
Dec-04	262.47	261.6	262.0783	0.111

C. Results and discussion

In order to analyze the grey properties of groundwater level, several confidence levels were set for the observed groundwater level at 6# station (Fig.8) . the grey degree of monthly groundwater level was obtained by Eq.(6) and the results are listed in table2. and grey degree can be shown in figure11. Similarly, the confidence interval of groundwater level in different confidence level was in figure12 and figure13. The results indicate that the grey degree of groundwater level increased with a decreasing confidence level. This shows that the influence factors could affect the groundwater level, and the groundwater system can be seen as a grey system.

Seen from table1, the largest variance of groundwater level in 2004 was in August. And seen from table2, the maximum grey degree was found in March, and the minimum grey degree was found in September when the confidence level was set to 80%.

In this study, at higher variance the confidence level would be lowered. The relationship between confidence grade and confidence level is listed in table2.

TABLE II. THE RELATIONSHIP BETWEEN CONFIDENCE GRADE AND CONFIDENCE LEVEL

Confidence	Variance
90	<0.1
80	0.1-0.5
70	0.5-1.0
60	>1.0

D. prediction of groundwater level

The prediction of groundwater level is important for groundwater management. The accuracy of groundwater level can influence the management effectively. While, the groundwater can be seen as a grey system, this study built a new model using grey theory and ANN to predict groundwater level at 6# station.

ANN has many types, according to the characters, three-layer RBF is considered as the basic network for establishment of the model. RBF neural network is one of the most important network in ANN at present, and so far, it is widely used in hydrological science. The GM-ANN coupling model is based on the GM model of groundwater, using the GM model of groundwater to simulate the groundwater level, then the groundwater level of different observation wells is as the input of the neural network to predict the groundwater level. The process can be seen in figure9.

Seen from figure 14, the prediction level is well fit for the observation level which indicate that the method is reliable and dependent.

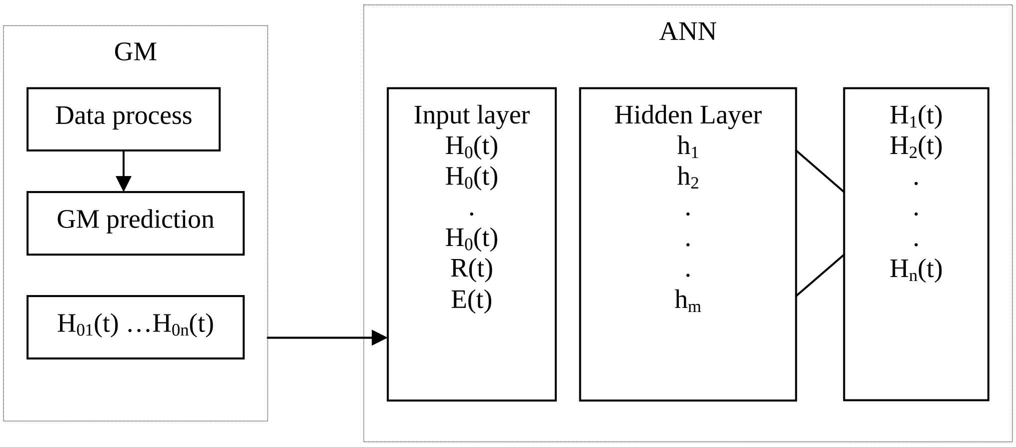

Figure 10. Schematic diagram of the coupled model: integration of ANNs with GM

TABLE III. GREY DEGREE OF GROUNDWATER LEVEL UNDER DIFFERENT CONFIDENCE LEVEL IN ANYANG STATION

Confidence level	J-04	F-04	M-04	A-04	M-04	J-04	J-04	A-04	S-04	O-04	N-04	D-04
0.95	0.17	0.33	0.17	0.17	0.33	0.33	0.50	0.33	0.17	0.33	0.33	0.33
0.9	0.33	0.33	0.67	0.33	0.33	0.67	0.50	0.50	0.17	0.33	0.67	0.50
0.8	0.50	0.67	0.83	0.67	0.50	0.67	0.67	0.67	0.33	0.50	0.67	0.67
0.7	0.67	0.83	0.83	0.83	0.67	0.83	0.67	0.67	0.50	0.67	0.67	0.67
0.6	0.83	0.83	0.83	1.00	0.67	0.83	0.83	0.83	0.67	0.83	0.67	0.67
0.5	0.83	0.83	0.83	1.00	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83
0.4	0.83	1.00	0.83	1.00	0.83	1.00	0.83	0.83	0.83	0.83	0.83	0.83
0.3	0.83	1.00	1.00	1.00	1.00	1.00	1.00	0.83	0.83	1.00	1.00	0.83
0.2	0.83	1.00	1.00	1.00	1.00	1.00	1.00	0.83	0.83	1.00	1.00	0.83
0.1	0.83	1.00	1.00	1.00	1.00	1.00	1.00	0.83	1.00	1.00	1.00	1.00

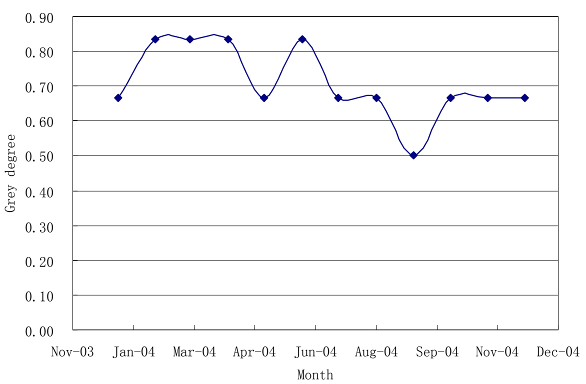

Figure 11. Grey degree of groundwater level in 6# station

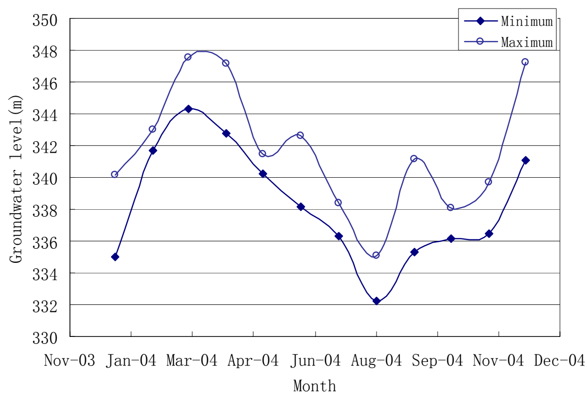

Figure 12. Grey interval of groundwater level in 6# station

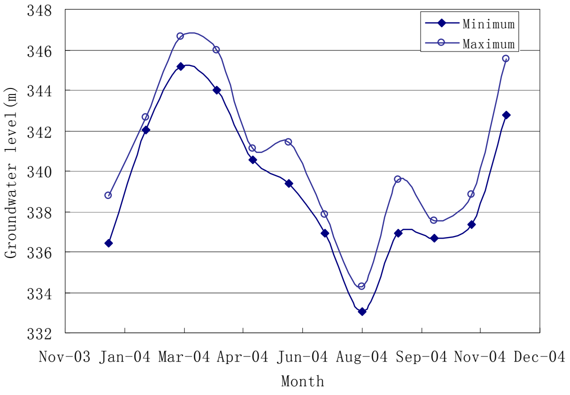

Figure 13. Grey interval of groundwater level in 3# station

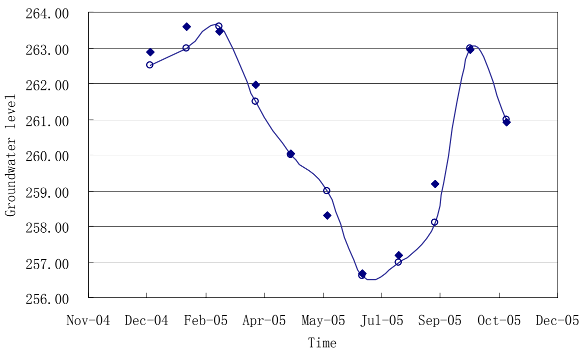

Figure 14. Simulated and measured groundwater level at 6# station in 2005

IV. CONCLUSIONS

Application of combined grey neural network to forecasting of groundwater level is a novel research area. Based on the modeling results obtained in this study, the following conclusions can be drawn:

(1) The model of combined grey neural network is proposed by virtue of the dynamic characteristics of groundwater level, which increased the forecast precision. Therefore, the method is reliable and effective.

(2) The combined grey neural network is the improvement of other groundwater level prediction methods. This method can be used to deal with the nonlinear and periodical issues to improve the precision of the groundwater level prediction.

(3) This method is not only suitable for the dynamic prediction of the groundwater level, but also suitable for other respects (such as surface water quality, prediction of quality in the atmospheric environment, and so on) .

ACKNOWLEDGMENTS

This study was funded by the National Basic Research Program of China (2010CB951101) , the National Natural Science Foundation of China (Grant No. 40830639, 50879016, 50979022 and 50679018) , the program for Changjiang Scholars and Innovative Research Team in University (IRT0717) and the Special Fund of State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering (1069-50985512, 1069-50986312)

REFERENCES

[1] Martin T.Hagan, Howard B.Demuth, Mark H.Beale, Neural network design.Beijing : China Machine Press,2004,239-255(in Chinese)

[2] Linming Zhao, Haoyun Hu, Dehua Wei, Shuqian Wang. Multilayer forward artificial neural network. Zhengzhou:Yellow River Conservancy Press , 1999,24-35(in Chinese)

[3] Bao Liu, Daiping Hu. Studies on applying artificial neural networks to some forecasting problems. Journal of System Engineering.1999,338-343(in Chinese)

[4] Zhijia Li, Xiangguang Kong. Channel flood routing model of artificial neural network. Journal of Hohai University. 1997,25(5) :7-12(in Chinese)

[5] Li, R., Wang, J., Qian, J., 2002. Prediction of river water quality based on grey dynamic model group. Bulletin of Soil and Water Conservation 22(4) , 10-12(in Chinese)

[6] Xiaofang Rui, Linli Wang. A study of flood routing method with forecast period .Advances in Water Science,2000,11(3) :291-295.(in Chinese)

[7] Guoru Huang, Xiaofang Rui. Radial basis function-neural network model for channel flood routing. Journal of Hohai University (Natural Sciences) , 2003,31(6) :621-625(in Chinese)

[8] Xingming Zhu, Changna Lu, Ruyun Wang, Jingyi Bai, Artificial neural network model for flood water level forecasting . Journal of Hydraulic Engineering, 2005,36(7) :806-811(in Chinese)

[9] Yanfang Geng, Zhili Wang, Sheng Jin. River system flood forecasting based on artificial neural network of radial basis function. Journal of Dalian University of Technology, 2006,4(2) :267-271(in Chinese)

[10] Feisi Research and Development Center of Science and Technology, MATAB6.5 auxiliary neural network analysis and design.Beijing, electron industry press, 2003(in Chinese)

[11] Luo, D., Guo, Q., Wang, X., 2003. Simulation and prediction of underground water dynamics based on RBF neural network. *Acta Geoscientia Sinica* **24(5)** , 475-478(in Chinese)

[12] Jun Lang,Xiaohong Su,Xiujie Zhou. Prediction for air pollution index based on combined grey neural network model.Journal of Harbin institute of techonology,2004,36(12) :1598(in Chinese)

[13] Y.-L.Yeh, T.-C.Chen. Grey degree and grey prediction of groundwater head. Stoch Enviro Risk Ass,2004,18:351-363.

[14] Qin Ju, Zhongbo Yu, Zhenchun Hao, Xi Chen, Chao She. Integration of artificial neural networks with a numerical groundwater model for simulating spring discharge [A].2010 4th International Conference on Intelligent Information Technoogy Application[C]. Institute of Electrical and Electronics Engineers, Vol.2, p361-364.

[15] Yeh YL, Su MD, Tsou I (1995) . Forecasting of groundwater level using grey modeling. J Taiwan Water Conservancy 43: 66–73 (in Chinese)

[16] Hogg RV, Tanis EA (1988) Probability and statistical inference. Macmillan Publishing Company, New York

[17] CHEN Xi, LIU Chuan-jie, HU Zhong-ming, et al. Numerical modeling of groundwater in a spring catchment and prediction of variations in the spring discharge[J]. Hydrogeology and Engineering Geology, 2006, 2:36-40.

[18] Ju, Q., Z. Yu, Z. Hao, et al, 2009, Division-based Rainfall-Runoff Simulations with BP Neural Networks and Xinanjiang models [J]. Neurocomputing, 72, 2873-2883.
[19] Parkin, G., S. J., Birkinshaw, P. L., Younger, et al, 2007, A numerical modelling and neural network approach to estimate the impact of groundwater abstractions on river flows, Journal of Hydrology, 339, 15-28.
[20] Birkinshaw, S. J., G. Parkin and Z. Rao, 2008, A hybrid neural networks and numerical models approach for predicting groundwater abstraction impacts, Journal of Hydroinformatics, 10(2) :127-137.
[21] Demissie, Y., K., A. J. Valocchi, B. S. Minsker, et al, Integrating a calibrated groundwater flow model with error-correcting data-driven models to improve predictions, Journal of Hydrology, 364 , 257 – 271.
[22] Garcia, L., A., A. Shigidi, 2006, Using neural networks for parameter estimation in ground water, Journal of Hydrology , 318, 215–231.
[23] McDonaldM, C., and A.W. Harbaugh, 1988. A modular three-dimensional finite difference ground-water flow model: U.S. Geological Survey Techniques of Water Resources Investigations[R]. Book 6, Chapter A1, 586.
[24] Douglas P. Dufresne, and Charles W. Drake, 1999. Regional groundwater flow model construction and wellfield site selection in a karst area, Lake City, Florida [J]. Engineering Geology, 52(1) :129-139.
[25] Chen X., and X. Chen, 2003. Stream Water Infiltration, Bank Storage, and Storage Zone due to Flood Storages in Channels[J]. Journal of Hydrology, 280, 246-264.
[26] CHEN Xi, CHEN Xun-hong. Numerical modeling of groundwater flow and analysis of water budget in Nebraska Sand Hills, USA[J]. Advances in Water Science, 2004, 15:94 -99.
[27] Qin Ju, Zhongbo Yu, Zhenchun Hao, et al, 2007. Hydrologic Simulations with Artificial Neural Networks [A]. Angela Burgess. Third International Conference on Natural Computation [C]. Los Alamitos: IEEE Computer Society Publications, 2:22-27

Changjun Zhu received his B.S. in China University of Petroleum in 1999 and M.S. in the Chinese Academy of science, in 2002. Since then, he is the associate professor of Hebei University of Engineering from 2009. Now he has received Ph.D. in the College of Hydrology and Water Resources, Hohai University in Nanjing, China, His main research interests are in numerical environment modeling

Zhenchun Hao received his B.S. and M.S. in the College of Hydrology and Water Resources, Hohai University in Nanjing, China, in 1981 and 1984, respectively. He continued his education at the same university where he received a Ph.D. in hydrology in 1988. Now he is a professor of hydrology at the College of Hydrology and Water Resources, Hohai University in Nanjing, China. His research interests include large-scale hydrological modeling, watershed hydrological modeling and ecological hydrology.