

The Comparison of Four Different Groundwater Level Prediction Models in Baoding City

Qi Wang^{1, 3}, Tingshan Tian^{2, 3} and Changqing Li²

¹University of Waterloo, ON, Canada

²China Geological Environmental Monitoring Institute, Beijing, China

³China University of Geosciences, Beijing, China

q244wang@uwaterloo.ca, {taints, changqli}@mail.cigem.gov.cn

Keywords: BP neural network prediction model; regression analysis prediction model; time series prediction model; Markov prediction model; Baoding City; groundwater level.

Abstract. This paper has quantitatively predicted the groundwater level variation in Baoding City based on BP neural network model, regression analysis model, time series model, and Markov model. It can be concluded that the prediction accuracy of BP neural network model and time series model is the highest, and their average relative prediction deviations are 3.5% and 2.3%. The prediction accuracy of regression analysis model is lowest, and the average relative prediction deviation is 14.69%. The range of average relative prediction deviation of Markov model is from 1.8% to 4.3%. The prediction accuracy of Markov model is high. Markov model can only predict the specific state of groundwater level, which is an interval value, rather than a specific value. Therefore, the reliability of Markov model can be improved by expanding the scope of the forecast on the premise of definitely meeting the requirements of actual work. This study provides scientific support towards groundwater level prediction of Baoding City, and has practical application value for urban water planning of Baoding City.

Introduction

Baoding City is located in central Hebei Province at longitude 113°40'~116°20' east and latitude 38°15'~39°57' north. It is north adjacent to Beijing and Zhangjiakou, south to Hengshui and Shijiazhuang, west to Shanxi Province and east to Cangzhou and Langfang. The average annual rainfall of the city is 540mm and the average number of precipitation days per year is 68 days. Its maximum evaporation capacity is 1977.9 mm and minimum is 1749 mm. The hydrographic net in Baoding is comparatively developed, with main rivers including Tang River, Fu River, Cao River, North Juma River, South Juma River and Daqing River.

There are two hydrogeological units in Baoding: mountain and plain. Groundwater recharge resources in mountain areas are mainly from atmospheric precipitation and surface water infiltration. Groundwater flows from the Northwest to the Southeast, but the direction has changed in the tectonic zone. The anticlinorium in Fuping-Baijian is distributed in the central region of the city and becomes a watershed between the area of surface water and groundwater area, resulting in the groundwater in the Northwest pooling in Laiyuan basin, and overflowing the surface formed spring group and then discharging in the plains, while the groundwater in Southeast directly discharging in the plains. The main ways of discharge are runoff and springing. The recharge resources of shallow groundwater aquifer in plain areas are mainly from atmospheric precipitation and lateral runoff supply, and recharge resources of deep groundwater aquifer are mainly from lateral runoff supply. The flow direction of groundwater is the same as the terrain inclined direction and surface water runoff direction, and the main ways of discharge are exploitation and downstream flow.

With the rapid development of industry and agriculture and the improvement of people's living standards, the exploitation quantity of groundwater is increasing annually, leading to a downward trend of groundwater level. In recent years, along with the protection and rational exploitation of groundwater, the groundwater level decline trend is alleviated.

Application of Four Models in the Prediction of Groundwater Level Variation

BP neural network model, Markov model, time series model and regression analysis model are usually used in the study of prediction of groundwater levels. Therefore, this paper will discuss the application of these four prediction models to the study of groundwater levels in Baoding city and analyze the relative prediction accuracy of the four models.

BP Neural Network Prediction Model

Artificial neural network technology is a technology to study by simulating the structure and function of human brain neural network, of which BP neural network technology is most widely used and with the best prediction accuracy [1], and has a wide range of application in the prediction research of groundwater levels [2,3,4]. Therefore, the artificial neural network model established in this paper is BP neural network model. It consists of three layers: input layer, hidden layer and output layer. The number of hidden layers is determined by the complexity of the specific problems and the decrease of the deviation. The connection weight in BP neural network model reflects the specific mapping relationship between input variables and output variables in the learning sample, and it is modified in the network learning.

We train the network model by using the monthly precipitation, evaporation, exploitation and groundwater level data from 2009 to 2012. The number of the hidden layer units is determined by the deviation decline of actual calculation. The practice result shows that the network model can obtain the reasonable experimental structure when the number of hide layers is four. The four input units correspond to the average precipitation, evaporation, exploitation and groundwater level data and one output unit corresponds to the average value of the predicted groundwater level. The comparison of parameter selection experiments is shown in table 1.

Table 1. The comparison of parameter selection experiments

No. of hidden layers	No. of hidden nodes	Learning rate	Times of learning	Results (Accuracy rate within 5% deviation)
2	10	0.01	5000	0.6
	20	0.01	5000	0.8
	40	0.01	5000	0.8
3	10	0.01	5000	0.8
	20	0.01	5000	1
	40	0.01	5000	1
4	10	0.01	5000	0.8
	20	0.01	5000	0.6
	40	0.01	5000	0.4

We use the trained network model to predict the groundwater level of Baoding City from January to April in 2013. The comparative analysis of the result is shown in Table 2.

Table 2. The comparison of BP neural network prediction and actual measurement

Time	Actual measurement (m)	BP network prediction (m)	Relative deviation (%)
Jan. 2013	19.89	20.15	1
Feb. 2013	20.45	21.50	5
Mar. 2013	20.02	20.93	5
Apr. 2013	21.01	21.66	3

From the table, it can be seen that the relative deviations of the average groundwater level value predicted by BP neural network to measured value are rather small, which are all not more than 5%. And the mean relative deviation is 3.5%, so the prediction result is quite satisfactory.

Markov Prediction Model

Another commonly used model in the prediction of groundwater level is Markov prediction model [5, 6]. Markov Process is a theory that studies the state of things and state transition. It's a stochastic process, also known as Markovian process, referring to what state a time series will be and what

value to assign, and only relying on present state and value and independent of its past state and value.

This study makes use of 48 measured data of groundwater levels in each month from 2009 to 2012 in Baoding City, calculating the mean value $\bar{x}=21.11\text{m}$ and standard deviations=1.48. Divide the groundwater levels variation into five states: very small, relatively small, general, relatively large and very large, as shown in Table 3.

Table 3. Grading and internal division of groundwater levels variation

State	Grade	Grading standard	Growth interval
One	Very small	$x < \bar{x} - 1.1s$	$x < 19.59$
Two	Relatively small	$\bar{x} - 1.1s \leq x < \bar{x} - 0.5s$	$19.59 \leq x < 20.48$
Three	General	$\bar{x} - 0.5s \leq x < \bar{x} + 0.5s$	$20.48 \leq x < 21.96$
Four	Relatively large	$\bar{x} + 0.5s \leq x < \bar{x} + 1.1s$	$21.96 \leq x < 22.85$
Five	Very large	$x \geq \bar{x} + 1.1s$	$x \geq 22.85$

Statistics of groundwater levels variation transition times in each month from 2009 to 2012 in Baoding is shown in Table 4.

Table 4. Statistics of groundwater levels variation transition times

Initial state	Transition state grade					Total
	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	
Grade 1	6	1	0	0	0	7
Grade 2	0	12	0	1	0	13
Grade 3	1	0	1	0	0	2
Grade 4	0	1	1	15	1	18
Grade 5	0	0	0	1	6	7
Total	7	14	2	17	7	47

We use the groundwater level series in each month from 2009 to 2012 of Baoding to test whether it has Markov property. From Table 4, the transfer probability frequency and transition probability matrix can be obtained:

$$f_{ij} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 12 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 15 & 1 \\ 0 & 0 & 0 & 1 & 6 \end{bmatrix} \quad (1)$$

$$P_{ij} = \begin{bmatrix} 0.857 & 0.143 & 0 & 0 & 0 \\ 0 & 0.923 & 0 & 0.077 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.056 & 0.056 & 0.832 & 0.056 \\ 0 & 0 & 0 & 0.143 & 0.857 \end{bmatrix} \quad (2)$$

Thus, the value of X^2 is 115.402. Given the significance level $\alpha=0.05$, the value of quantile X_{α}^2 is 26.296 from the quantile table. Because $X^2 > X_{\alpha}^2(5-12)$, so the series have Markov property.

According to the statistics of Table 4, a state transition probability matrix of groundwater levels variation can be obtained.

$$P = \begin{bmatrix} 0.857 & 0.143 & 0 & 0 & 0 \\ 0 & 0.923 & 0 & 0.077 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.056 & 0.056 & 0.833 & 0.56 \\ 0 & 0 & 0 & 0.143 & 0.857 \end{bmatrix} \quad (3)$$

According to the statistics of groundwater levels variation transition times in Table 3, the groundwater level variation state is in grade 2, given initial state probability vector $A(0) = [0 \ 1 \ 0 \ 0 \ 0]$, there is,

$$A(1) = PA(0) = [0 \quad 0.92 \quad 0 \quad 0.08 \quad 0]. \quad (4)$$

$$A(2) = PA(1) = [0 \quad 0.861 \quad 0.004 \quad 0.131 \quad 0.004]. \quad (5)$$

$$A(3) = PA(2) = [0.012 \quad 0.787 \quad 0.009 \quad 0.174 \quad 0.018]. \quad (6)$$

$$A(4) = PA(3) = [0.016 \quad 0.737 \quad 0.014 \quad 0.207 \quad 0.026]. \quad (7)$$

According to above analysis, the groundwater level variation from January to April in 2013 has the maximum probability in grade 2, that is, the average groundwater level of these four months will be within 19.59~20.48m. The actual measured groundwater levels of the four months are 20.45m, 20.02m, 21.01m and 19.89m, and only the predicted data in April is smaller than actual measured. The comparison between predicted value range and measured value is shown in Table 5.

Table 5. The comparison between predicted value range by Markov prediction model and measured value

Time	Predicted range (m)	Measured value (m)	Relative deviation range (%)
Jan. 2013	19.59~20.48	19.89	2~3
Feb. 2013	19.59~20.48	20.45	0.1~4
Mar. 2013	19.59~20.48	20.02	2~3
Apr. 2013	19.59~20.48	21.01	3~7

According to the data in the table, the relative deviation of groundwater level value from January to March predicted by Markov prediction model to measured value is rather small, not more than 5%, and the relative deviation in April is probably greater than 5%. The mean relative deviation is within 1.8%~4.3%, so the prediction result is quite satisfactory.

Time Series Prediction Model

Time series model is also widely used in the study of the variation of groundwater levels [7,8]. Time series model generally includes trend term, periodic term and random noise term. So the observation series model of groundwater levels is:

$$H_t = T_t + S_t + R_t. \quad (8)$$

Of which, H_t refers to observation series of groundwater level state, T_t refers to trend term series, S_t refers to periodic term series, R_t refers to random term series.

According to the optimized frequency analysis, the best observation time interval is 5 days. Making use of SPSS and according to AIC Standard, the time series model will be ARIMA (0,1,1). The comparison of predicted value by time series prediction model and measured value is shown in Table 6.

Table 6. Comparison of predicted value by time series prediction model and measured value

Time	Predicted value (m)	Measured value (m)	Relative deviation (%)
Jan. 2013	19.80	19.89	0.5
Feb. 2013	19.91	20.45	2.7
Mar. 2013	19.84	20.02	0.9
Apr. 2013	19.99	21.01	5.0

From Table 6, the relative deviation of groundwater level values predicted by time series model to measured value is rather small, not more than 5%, and the mean relative deviation is 2.3%, so the prediction result is quite satisfactory.

Regression Analysis Prediction Model

Regression analysis model is a kind of stochastic model, which has a great advantage in predicting groundwater level under a very complex hydrogeological condition. Regression analysis method is to build a regressive function relationship between independent variables and dependent variables

with mathematical statistics on a basis of a long-time mastering of series data. Multivariate linear regression analysis is widely used in the areas with complex hydrogeological conditions [9,10]. Its mathematical model formula is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_P X_P + \epsilon. \quad (9)$$

Of which, Y refers to dependent variable; X_i ($i=1,2,\dots, P$) is independent variable; $\beta_0, \beta_1, \beta_2, \dots, \beta_P$ are unknown regression parameters and ϵ is random variable;

Use the actual measurement data in Baoding City to build a multivariate linear regression model between groundwater levels and precipitation, evaporation, exploitation:

$$H = 7.83 + 0.009P - 0.003E + 0.00044Q. \quad (10)$$

Of which, H refers to groundwater level in month i; P is precipitation capacity in month i; E is evaporation in month i and Q is manual exploitation in month i.

According to above formula, it can be obtained that the negative correlation coefficient $|r|$ is 0.88, which is within the range of [0.8-1]. It means that there is a strong linear relationship between independent variables and dependent variables in the regression model. Likewise, $P=0.02 \leq 0.05$ also shows a significant linear relationship between independent variables and dependent variables. In this study, the value of F is 6.89. From F distribution table, we can get $F_{0.95}(3,6)=4.76$. $F > F_{0.95}$, so H_0 is rejected. Finally, according to F test criterion, it can be assumed that there is a strong linear relationship between independent variables and dependent variables.

We use already built regression model to predict the groundwater levels from January to April in 2013 in Baoding, and the comparison between predicted value and measured value is shown in Table 7.

Table 7. Comparison between predicted value by regression analysis and measured value

Time	Predicted value (m)	Measured value (m)	Relative deviation (%)
Jan. 2013	16.45	19.89	17.29
Feb. 2013	18.00	20.45	11.98
Mar. 2013	17.37	20.02	13.23
Apr. 2013	17.59	21.01	16.28

From Table 7, it can be concluded that the relative deviation of groundwater level value predicted by regression analysis prediction model to measured value is rather large, more than 5%, and the mean relative deviation is 14.69%, so the prediction result is not satisfactory, which is not suitable for the study on groundwater level prediction in Baoding City.

Comparative Study on These Four Prediction Models

Comparative Study on Prediction Methods

Visualizability

Regression equation concluded from regression analysis can visually describe the regression relationship between variables, while the neural network model has a topological structure which cannot directly present the regression equation. It can get close to the objective function by learning and record the information on relevant connection weights, forming a corresponding relationship between input variable and output variable when the deviation up to a certain standard. But the neural network model can show all the regression results in a certain range, and present the relationship between input and output by diagram. Time series model is to find the function conforming to the change of system based on past observed value series, and use the observed value as independent variable and predicted value as dependent variable to conduct prediction analysis.

However, due to the influence of many factors, the development of the system is not linear, so it is impossible to find an accurate function to describe the law of the development of the system. Markov prediction model can only predict an interval rather than an exact function relationship.

Operation Difficulties

Multivariate regression analysis requires lots of calculation, and the amount of calculation gets larger with the increase of the numbers of variables, also does the numbers of curves, therefore, it is difficult to select the optimal regression curve. In addition, different regression analysis methods have different restrictions on variable values. So we need a segmented regression analysis, even a stepwise regression analysis if a model cannot reflect the whole relationship between the variables. But the calculation of neural network model can be operated with a unified practicing method without being distinguished, and it's simple and easy to be carried out. The calculation of time series model is complicated because it needs a large amount of time series data to identify whether the time series is stationary, and transfer the nonstationary time series into stationary time series by using difference. The calculation of Markov prediction model is more complex since it needs to implement data states division and state transition probability calculation and so on.

Objectivity

When using the method of regression analysis, it needs to draw a scatter diagram to judge the type of regression curve, no matter to implement multivariate regression analysis, linear or nonlinear regression analysis. However, regression curve can only roughly describe the relationship between the variables rather than the details of the regression curve. In addition, there are a few types of basic regression curves so some of the curves are treated as linear relationship even if they are not, so it will be difficult to describe the real relationship between variables, leading to big deviation. Therefore, the selection of models decides the result of regression effect. Neural network is to approach to the objective function by learning, so the regression deviation will be smaller if using reasonable network structure and practice method. The prediction result of time series prediction model sometimes cannot be so good by using long-time series, for it cannot eliminate the non-significant factors as regression analysis with the use of F test. The prediction result of Markov prediction model is a range or an interval rather than a certain value. Therefore, there is a direct relationship between prediction accuracy and the number of historical data, the reasonability of intervals division, the more the data is and the more reasonable the intervals are divided, the more accurate the prediction result will be.

Comparative study of Prediction Results

According to the above comparative analysis of the prediction accuracy of four prediction models, the relative deviations of predicted value by BP neural network to measured value is rather small, which are all not more than 5%, so the prediction result is quite satisfactory. A prediction interval can be obtained from Markov prediction model and 80% of the predicted values are within this interval. The relative deviation of predicted value by time series model to measured value is very small, which are all less than 5%. However, the relative deviation of predicted value by regression analysis to measured value is very large, which are all more than 5%, so the prediction result is not satisfactory.

It can be concluded that regression analysis is not suitable for the prediction study on the variation of groundwater levels and time series prediction model and BP network prediction model can quite accurately predict the values of groundwater levels. And Markov prediction model can be used to predict an interval if there is no necessary to predict certain values.

Conclusions

With regarding the groundwater level in Baoding City as the research objective, this paper analyzes and studies the hydrogeological conditions and groundwater level data of each month in nearly five years of Baoding City, and elaborates the variation law of groundwater levels, and also discusses the prediction accuracy of four groundwater level prediction models. Through the research, it can be concluded that:

The annual groundwater level from 1996 to 2012 in study area is about 20m with an amplitude of 10m. The groundwater level keeps decreasing due to past excessive exploitation in Baoding City. However, with the implementation of relevant restriction policies, the groundwater is exploited and used reasonably. Therefore, the groundwater level has restored back to previous average level in recent years.

The Markov prediction model can predict the state of groundwater level, and the prediction result is an interval rather than a certain value. Therefore, on the premise of meeting the requirements of research, the reliability of the prediction result of Markov model can be improved by expanding prediction range. The relative deviation of predicted value by BP neural network to measured value is rather small, so the prediction result is satisfactory. The relative deviation of predicted value by time series model to measured value is very small, which are all less than 5%, so the result is also satisfactory. But the relative deviation of predicted value by regression analysis to measured value is very large, which are all more than 5%. The prediction result of regression analysis is not satisfactory, so it's not suitable for the prediction study of groundwater levels in Baoding City. Through the research of four prediction models, this study can provide scientific support towards groundwater level prediction in Baoding City, and have practical application value for urban water planning in Baoding City.

References

- [1] B. Widrow: Neural Networks, Application in Industry, Business and Science. Communication of the ACM, 37, 1994: 93-105.
- [2] Shengli Zhao, Yan Li, Quanshu Li: BP Neural Network Prediction in Groundwater State, Agricultural University of Hebei Journal, 2002, 25(4):205-207.
- [3] Yantao Zhao: Urban Groundwater Levels Prediction Based on BP Neural Network, Zhengzhou University, 2011.
- [4] Liangping Yang, Zhenjiao Jiang, Yiting Zhao: Prediction Study on Groundwater Levels Variation and Tendency in Hohhot City, Hydrogeological Working Medium, 2009,36(4):46-49.
- [5] Yinsheng Song: Application of Markov Prediction Model in Groundwater Level Prediction, Shangdong Geology, 1998(1)34~40.
- [6] Miaolin Wang: Application of Grey Markov Prediction Model in Annual Maximum Groundwater Levels Prediction in Cuntan Station, 2004(2)10—11.
- [7] Yinfeng Chen: Application of Chaotic Time Series in Hydrologic Forecasting, Changan University, 2005
- [8] Shujing Zhang, Lixin Qi,: Concise Textbook of Time Serie. Tsinghua University Press, North Jiaotong University Press, 2003.
- [9] Yani Wang, Gongqiang Gao, Mingke Cai: Groundwater Levels State Forecasting in Baoji City Based on Multivariate Regression Analysis [J]. Hydraulic and Architectural Engineering Journal, 2010(05)
- [10]Shangpeng Gao, Qinglu Zhao, Zhiheng Shao: Groundwater Levels Prediction Based on Regression Analysis. Groundwater, 2014(03).