Groundwater Level Prediction based on Neural Networks: A case study in Linze, Northwestern China

Hui Zhang, Jixuan Zhao, Chong Chen

College of Information Science And Engineering, China University of Petroleum-Beijing, Beijing, China

Abstract: Groundwater level is an important factor in evaluating groundwater resources. Due to numerous non-linear factors, establishing theoretical models is difficult.. Therefore, this paper proposesthe BP (Back Propagation) neural network and the Radial Basis Function (RBF) neural network. The study area is divided into two zones. The R^2 (coefficient of determination) and RMSE (Root Mean Squared Error) are used to evaluate the performance. The BP neural network is used to predict groundwater level in the two zones with the R^2of0.57 and 0.54, with the RMSE of 0.0804 meters and 0.1864 meters respectively. The RBF neural network is implemented with $R^2$of 0.65 and 0.61, with RMSE of 0.0720 meters and 0.1519 meters, respectively. The results show the RBF neural network performs better than the BP neural network in the accuracy of predicting groundwater level. This study shows the feasibility and superiority of groundwater simulation using neural network.

1 Introduction

The development of Artificial Neural Networks (ANN) provides more ideas and methods for groundwater research. The geographical environment of each region is different, and the groundwater conditions are also different. Therefore, choosing a more suitable neural network to predict the groundwater level in this region has become a major research issue. In 1991, Mr. Ding Hailiang [1] established a prediction model of GM (1,3) in the karst groundwater mining area with the extraction volume and rainfall as the influencing factors. In 2002, Fu Qiang et al. [2] optimized the BP neural network model with the momentum term learning rules, and conducted simulation experiments on the depth of groundwater in the rice irrigation area. In 2007, Lu Wenxi et al. [3] used the learning rate adaptive momentum BP algorithm to predict the groundwater depth in western Jilin. In 2020, Yan Baizhong et al. [4] constructed a multivariate Long Short Term Memory(LSTM) network, taking 13-yeardata from monitoring well J1, Daiyue District, Tai'an City as the research object, and exploring the application of LSTM network in groundwater level prediction,and Multivariate LSTM network is superior to the univariate LSTM network and the BP neural network.The neural network proposed in these studies for the corresponding study area is of great significance for predicting the depth of groundwater in similar environments in the future. Neural network has important application value in the study of groundwater in oil

exploitation and development stage. Cheng Yishanet al. [5] used BP neural network model to forecast the grade of inorganic scale produced in the process of oil field water injection, which provides a method with high precision, less data demand and less calculation time. Kong Xiangchao [6] used the GA-BP model to study the groundwater pollution in the process of oil development in Northern Shanxi Province of China, and established a prediction system between oil development pollution and groundwater environmental impact, which reduced the workload while providing effective data results.

The spatial characteristics of water resources in China are as follows: scarce water resources in the north and the west, rich in the South and East, scarce in the inland and more in the coastal areas [7]. Due to the large population of China, in the case of water resources are not dominant, the per capita share of water resources is even more fragmented, which is equivalent to feeding 21% of the world's population with 6% of the world's water resources. In the 21st century, with the rapid development of society and the substantial improvement of living conditions, China's population shows a trend of growth, followed by a large increase in water demand for production and living, which aggravates the shortage of water resources. Linze area of Gansu Province is a typical arid area in Northwest China [8]. Soil erosion and desertification are serious, and the ecological environment is fragile. Effective management of agricultural water resources is the key to solve the ecological problems in this area and an important way to improve the quality of life of people in Linze area. This paper collects the groundwater level recharge value and the actual evapotranspiration in the experimental area.Two modelsare developed to predict and analyze the groundwater level, compares the fitting results of the model with the actual value, and selects the better model, which can provide convenient basis and technical support for us to improve the utilization rate of water resources.

The maincontributions of this work are:

(a) the prediction of thegroundwater level by considering the initial groundwater level and actual evapotranspiration as input factors;

(b) the suitable BP neural network and the RBF neural network models with appropriate hyperparameters for predicting groundwater level.

2 Methodology

2.1 BP neural network

The BP neural network [9] is a multi-layer feedforward neural network trained according to the error back propagation algorithm. It is composed of input layers, hidden layers, and output layers (shown in Figure 1).
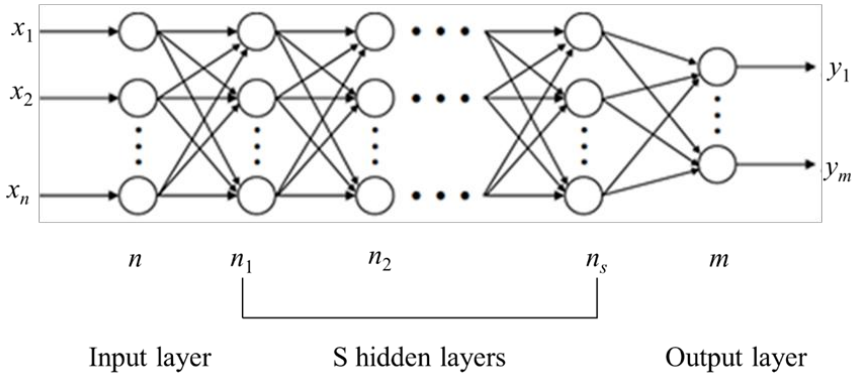
Fig. 1. Structure of the BP neural network.

The BP neural network has strong nonlinear mapping ability and self-learning ability. The basic idea of BP algorithm is to use the difference between the target value and the results generated by the forward propagation of the network. The error is back propagated to make the network continuously learn and optimize the weight and threshold value of each layer by layer under the conditions that (1) the difference value does not meet the requirements; or (2) the number of iterations does not reach the predefined value. The learning rule of the BP neural network is the gradient descent method which is also used in this study [10].

Firstly, calculate the partial derivative of the loss function for each neuron in the output layer.

$$\frac{\partial e}{\partial wh_o} = \frac{\partial e}{\partial yi_o}\frac{\partial e}{\partial wh_o} = \delta_o(k)ho_h(k) \tag{1}$$

$$\frac{\partial e}{\partial yi_o} = \frac{\delta(\frac{1}{2}\sum_{o=1}^{q}(d_o(k)-yo_o(k)))^2}{\delta yi_o}$$

$$= -(d_o(k)-yo_o(k))yo'_o(k) = -(d_o(k)-yo_o(k))f(yi_o(k)) = \delta_o(k)$$

(2)

$$\frac{\partial yi_o(k)}{\delta wh_o} = \frac{\delta(\sum_{h}^{p} w_{ho}ho_k(k)-b_o)}{\delta wh_o} = ho_h(k) \tag{3}$$

where e = Loss function; x=input vector; hi=input vector of the hidden layer; ho=output vector of the hidden layer; yi=input vector of output layer;yo=output vector of output layer; d= output vector; w_{ih}=the connection weight of the input layer and the hidden layer; w_{ho}=the connection weight between the hidden layer and the output layer.

Secondly, calculate the partial derivative of the loss function to each neuron in the hidden layer.

$$\frac{\delta e}{\delta hi_h(k)} = -(\sum_{o=1}^{q} \delta_o(k)w_{ho})f'(hi_h(k)) = \delta_h(k) \quad (4)$$

$$\frac{\delta hi_h(k)}{\delta w_{ih}} = x_i(k) \quad (5)$$

$$\frac{\delta e}{\delta w_{ih}} = \delta_h(k)x_i(k) \quad (6)$$

Thirdly, use the partial derivative of the loss function to each neuron in the output layer to modify the connection weight of the output layer.

$$\Delta w_{ho}(k) = -\frac{\delta e}{\delta w_{ho}} = \delta_o(k)ho_h(k) \quad (7)$$

$$w_{ho}^{N+1} = w_{ho}^{N} + \eta\delta_o(k)ho_h(k) \quad (8)$$

where η =learning rate.

Fourthly, the partial derivative of the loss function to the neurons in the hidden layer is used to modify the hidden layer connection weights.

$$\Delta w_{ih}(k) = -\frac{\delta e}{\delta w_{ih}} = -\frac{\delta e}{\delta hi_h(k)}\frac{\delta hi_h(k)}{\delta w_{ih}} = \delta_h(k)x_i(k) \quad (9)$$

$$w_{ih}^{N+1} = w_{ih}^{N} + \eta\delta_h(k)x_i(k) \quad (10)$$

Finally, calculate the global error (m samples).

$$E = \frac{1}{n}\sum_{k=1}^{m}\sum_{o=1}^{q}(d_o(k) - yo_o(k))^2 \quad (11)$$

Repeat the above steps until the result meets the requirements.

2.2 RBF neural network

The RBFneural network [9] is a non-linear layered feed forward network with strong non-linear fitting ability and simple learning rules, which is convenient for computer implementation. In the context of neural networks, the hidden unit provides a "function" set, which constructs an arbitrary "base" when the input pattern (vector) is extended to the hidden space. The functions in this function set are called Radial basis function. Briefly, a neural network with a radial basis function as an activation function is called an RBF neural

network. The RBF neural network can approximate any non-linear functionwith strong generalization ability, solve the regularity that cannot be analyzed in the system software. It has excellent results in aspects such as image processing and data classification. Figure 2 shows the structure of the RBF neural network.
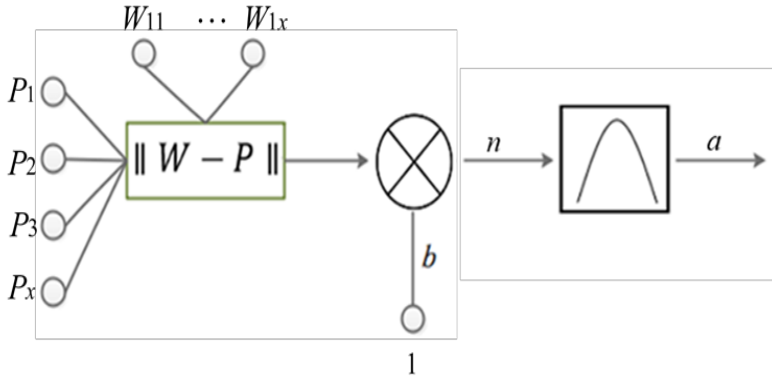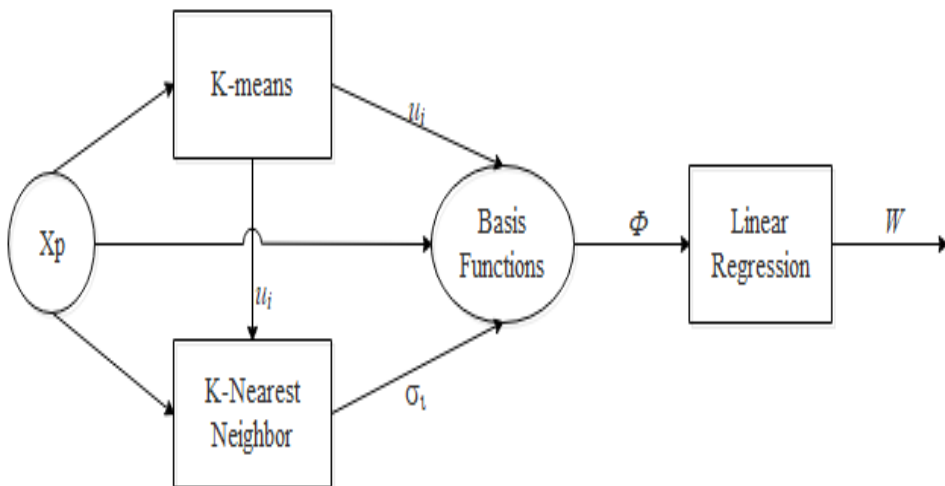

Fig. 2. General RBF learning process.

The basic idea of RBF is: transform the data into a high-dimensional space so that it is linearly separable in the high-dimensional space [11]. Therefore, the output layer is linear. The common activation function of RBF hidden layer is Gaussian function, and the function is as follows.

$$\phi(||x-u||) = e^{-\frac{|x-y|^2}{\sigma^2}} \tag{12}$$

The learning algorithm flow of the RBF neural network is shown in Figure 3.

Fig. 3. General RBF learning process.

According to the structure above, the RBF learning process can generally be described as the following steps.

(a) Use the k-Means algorithm to find the center vector u^i.

(b) To calculate the variance σ_k, the calculation formula is as follows.

$$\sigma_i = \sqrt{\frac{1}{k}\sum_{k=1}^{K}||u_k - u_i||^2} \qquad (13)$$

(c) Use the least square method to get the weight *W*.

3 Case study

3.1 Problem statement

A synthetic study area is conducted by considering two different sub-areas. The distribution of the two sub-areas are different according to the data observed at Linze Agricultural Ecosystem Comprehensive Observation Field. Corns are planted in the study area. The crop parameters adopt the crop data file "MAIZ.W41" in the WOFOST database. The synthetic system is shown in Figure 4. According to the different groundwater recharge and actual evapotranspiration in regions 1 and 2, respectively, the groundwater levels of the two regions are predicted.
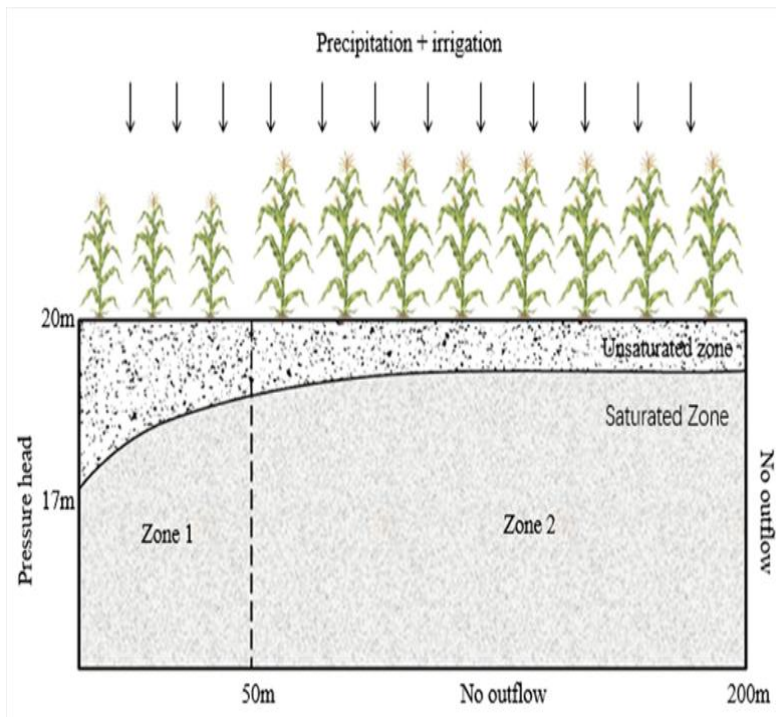

Fig. 4. Illustrative 2D case.

3.2 Experiments setup

3.2.1 The parameters of the BP neural network

First, the dataset is divided into training dataset by randomly select 70% of the data and verification dataset by randomly select 30% of the data using the *train_test_split* method of *sklearn.model_selection*. Because of the different units of groundwater level, groundwater recharge and actual evapotranspiration, *sklearn* is used to normalize the training data to avoid the large error and low accuracy of the model results. The normalization method used in this study is described as follows.

$$X = \frac{(x - \mu)}{\sigma} \tag{14}$$

where X=the processed dataset; μ =the average value; σ =variance.The pre-processing makes the variance of the new dataset of 1 and the mean value of 0.

In this experiment, we choose to build a neural network with two hidden layers. The activation function uses "*ReLU*"; the loss function uses "*mean_squared_error*", which is the mean square error (MSE); the learning algorithm uses "Adam optimization algorithm".

The hyper-parameters of the BP neural network are set according to feature of the dataset, i.e., the learning rate is set to 0.001, the maximum number of iterations (epoch) is set to 1000, the total number of experimental samples is 130, and the number of samples selected for one training is set to 32.

### 3.2.2 The parameters of the RBF neural network

First, the same data normalization method as the BP neural network is used to process the data. Then, the *fit()* method in the *sklearn.cluster.KMeans* is used to construct the k-means cluster.

The number of neurons in the hidden layer of the RBF neural network is set to 17. The multiplication of the Gaussian kernel matrix and the weight matrix is used to calculate the predicted value.

4 Results and discussion

4.1 Comparison chart

Through the training and prediction of the above two neural network models, we can see that they all have the ability to predict the groundwater level dynamically. In order to facilitate the comparison and select the better model, we compare the experimental results of the above two models separately according to the regions and draw the comparison charts of the training set results and test set results of the two neural networks in the two regions, as well as the comparison diagram between the prediction results and the real value.
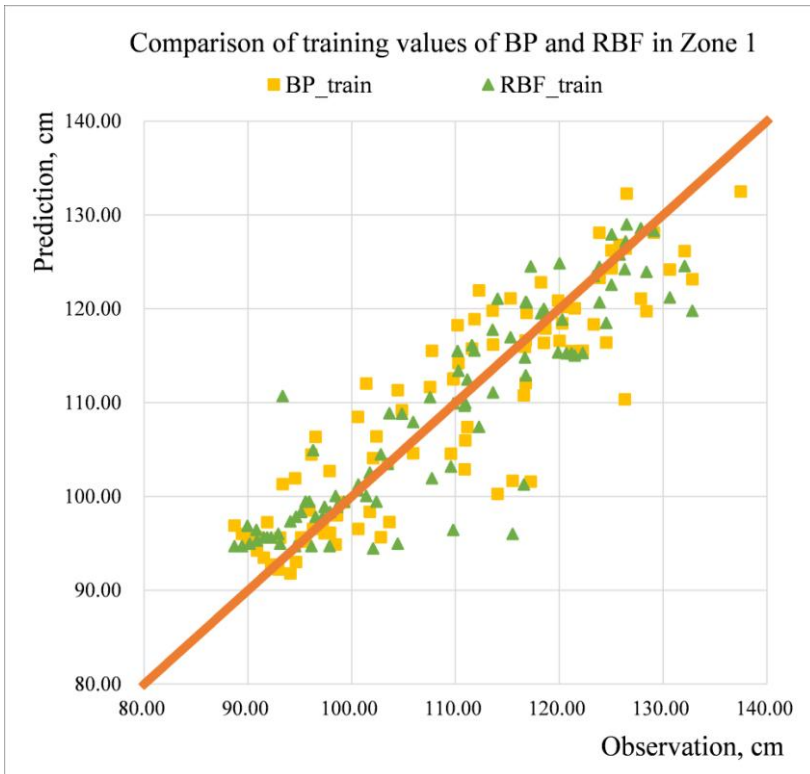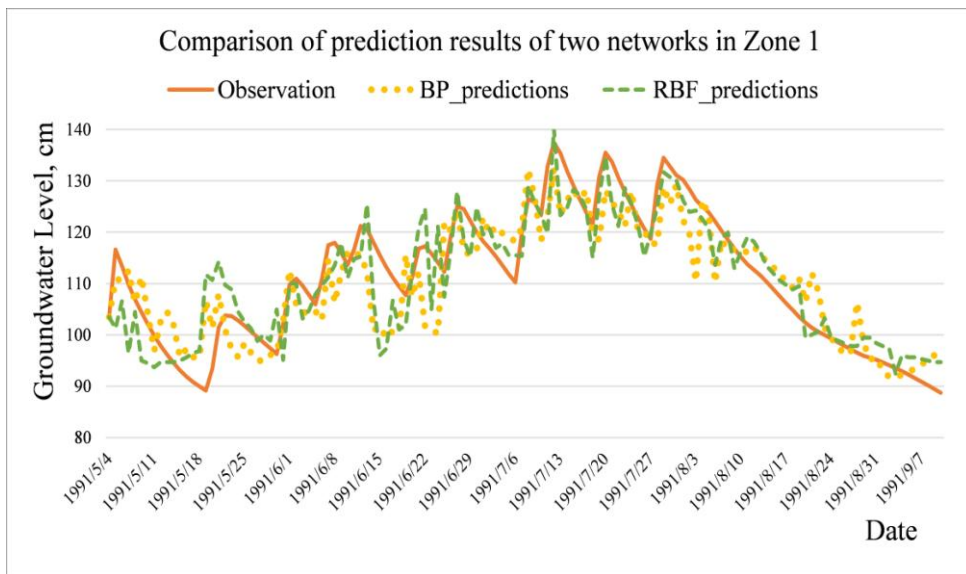
Fig. 5. Comparison of training values of BP and RBF in Zone 1.
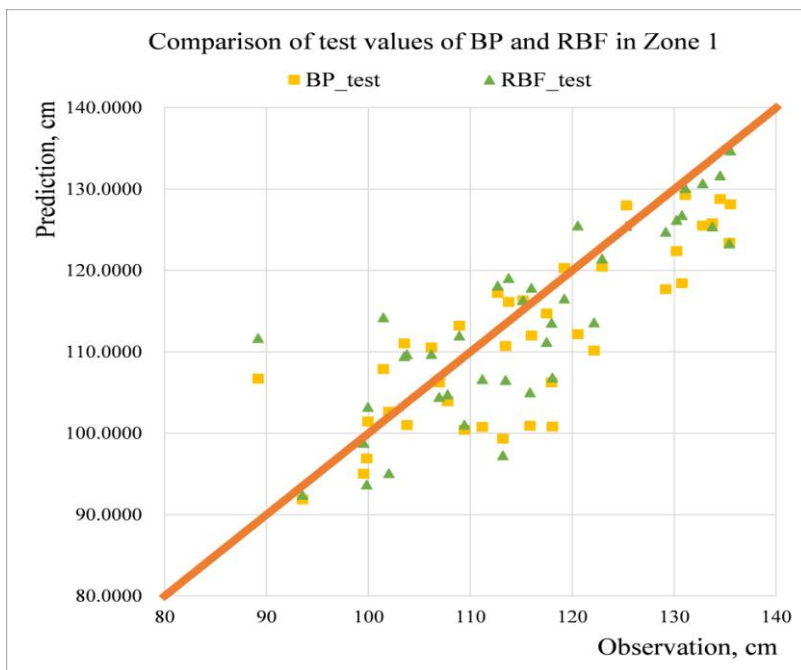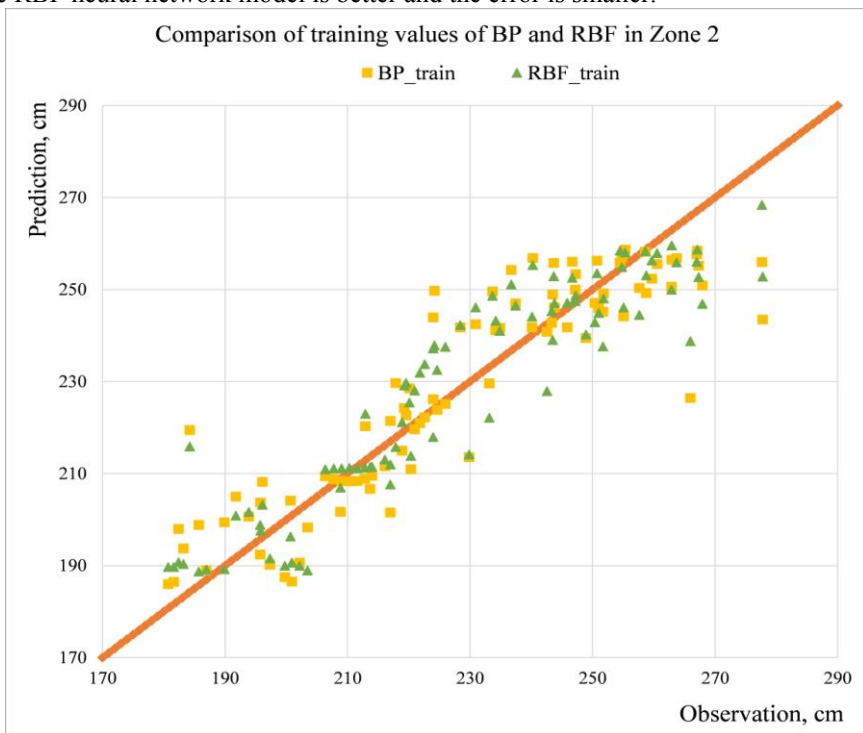

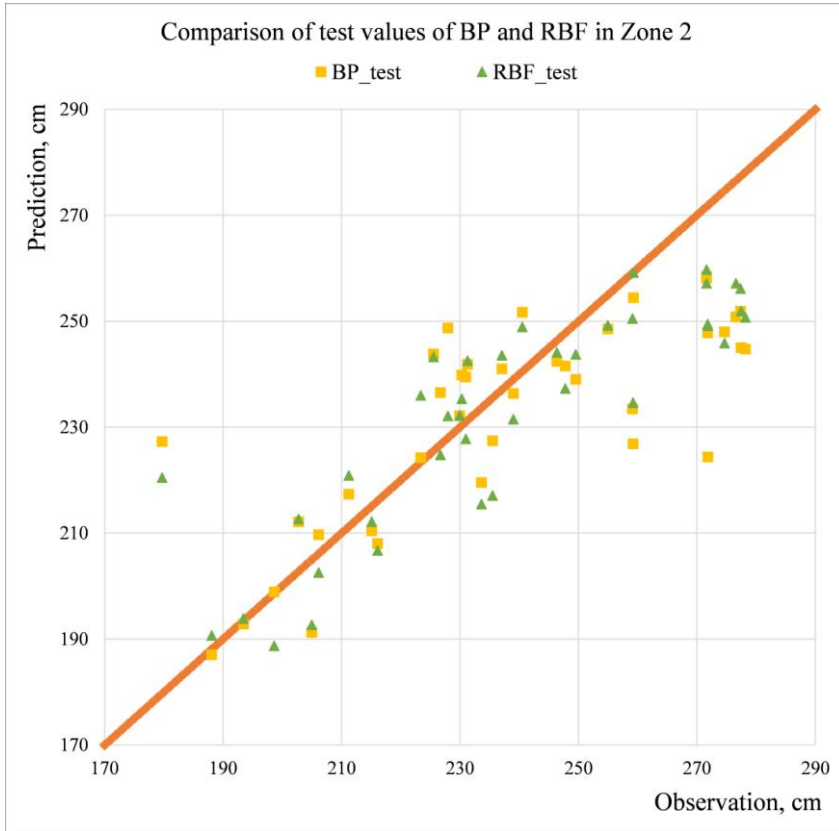**Fig. 6.** Comparison of test values of BP and RBF in Zone 1.

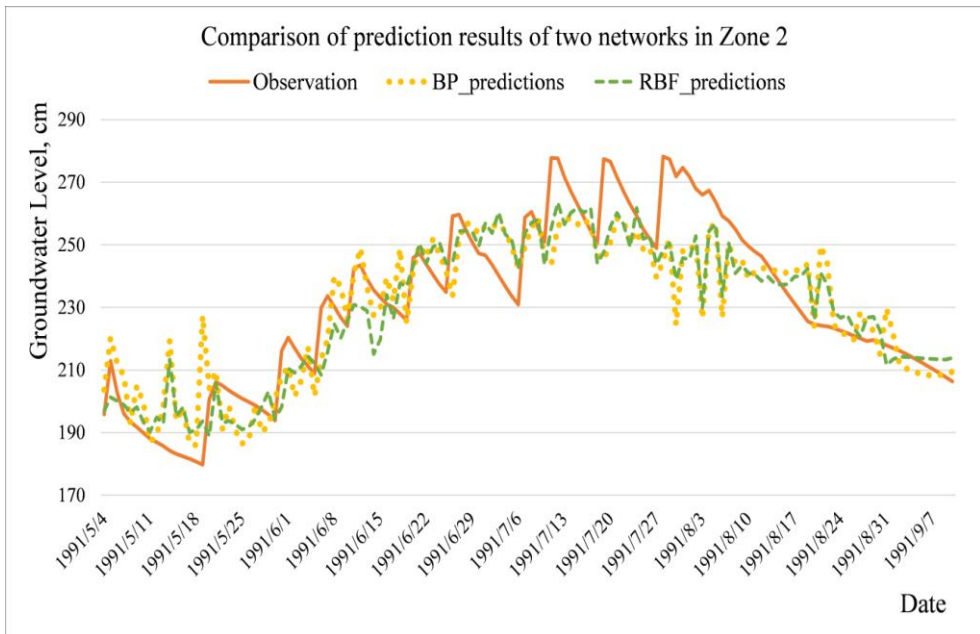**Fig. 7.** Comparison of prediction results of two networks in Zone 1.

It can be seen from the above three comparison charts that the results of the two neural network prediction processing can reflect the change trend of the original groundwater level and roughly predict the groundwater level value, but on the whole, the latter prediction value is closer to the actual groundwater level value, which shows that the prediction result of the RBF neural network model is better and the error is smaller.

Fig. 8. Comparison of training values of BP and RBF in Zone 2.

Fig. 9. Comparison of test values of BP and RBF in Zone 2.

Fig. 10. Comparison of prediction results of two networks in Zone 2.

From Figure 8 ~ Figure 10, one can conclude that the neural networks are able to capture the variation of groundwater levels in Zone 2. The test results indicate that the neural network are trained with sufficient accuracy. The results calculated from the RBF neural network perform better than those from the BP neural network.

4.2 Coefficient of determination (R^2) and Root Mean Square Error(RMSE)

The coefficient of determination (R^2) [12] indicates that in the regression relationship, the explained independent variable can be used to reflect the reliability of the dependent variable in the regression model. The closer to 1 the better the effect. The Root Mean Square Error (RMSE) is the square root of the average of the sum of the squares of the D-value between the predicted value and the actual value. It is more accurate and effective to measure the deviation. In this experiment, they all indicate the fitness between the predicted value of groundwater level and the true value. The R^2 and the RMSE data of the BP neural network and the RBF neural network in Zone 1 and 2 is shown in Tables 1 and 2.

Table 1. Comparable coefficient (R^2) data comparison table.

R^2	BP neural network	RBF neural network
Zone 1	0.57	0.65
Zone 2	0.54	0.61

Table 2. Root Mean Square Error (RMSE) data comparison table.

RMSE/m	BP neural network	RBF neural network
Zone 1	0.0804	0.0720
Zone 2	0.1864	0.1519

Through the comparison of the data in Table 1 and 2, we conclude that the results of the RBF neural network prediction are higher than the BP prediction results with a higher coefficient of determination and a smaller root mean square error. The RBF neural network model reflects the changes of groundwater level more accurately and more stable, although sometimes the prediction value of BP network is better, the prediction of the BP neural network is not as stable as that of RBF network.

5 Conclusion

From the prediction results, the D-value between the predicted value and the true value of the BP neural network model is larger than that of the RBF neural network model. Their performance indicators are relatively close, but the former is not as stable as the latter. Because the BP neural network is easy to fall into the local minimum and the convergence speed is slow, all of these affect its accuracy. When the center point of the RBF neural network model is far away from the query point, the output result is close to 0. What really matters is some points away from the query point. The characteristic of "local mapping" is obvious. Moreover, due to the small number of layers, the accuracy is higher and the effect is relatively stable.

From the perspective of the complexity of the experimental method, the BP neural network model has more hidden layers, which can have one or more layers. The adjustment of the number of neurons in each layer will affect the final result, and there is no authoritative calculation method. Predecessor experience and a large number of experiments make it more complicated to adjust the parameters, however, the RBF network only needs a hidden layer, so it is easier to adjust.

Because the influencing factors of the groundwater level are extremely numerous and complex, and in this experiment, we only used 2 influencing factors to roughly simulate the dynamic changes of the groundwater level, which can be used to study the changes of the groundwater level in complex environments in the future Provide some basis. This method of predicting the difficult-to-measure groundwater level through some well-measured factors reduces some human, material and financial resources. Ithas important implications in the scientific management of groundwater and agricultural technology in Linze, and brings great value to society. The growth process of crops is affected by various factors such as groundwater depth, precipitation, runoff, radiation and so on. Therefore, a variety of models are integrated to describe the crop growth process more accurately, and different types of models are used for different processes, so as to integrate the advantages of machine learning model and numerical model. This method can improve the simulation accuracy without increasing the amount of calculation.

References

1. D. Hailiang A preliminary study of grey model (1,3) applied to groundwater level forecast in a karstified area of groundwater exploitation, *Carsologica Sinica*, **02**, 13-20 (1991).

2. F. Qiang, L. Jianyu, W. Likun, et al Based on ANN model to forecast the groundwater level on the area of well irrigation rice, *Journal of Northeast Agricultural University*, **02**, 152-159 (2002).

3. L. Wenxi, Y. Zhongping, L. Ping, et al Dynamic prediction model of groundwater level based on improved BP algorithm **03**, 5-8 (2007).

4. Y. Baizhong, S. Jian, W. Xinzhou, et al Multivariable LSTM Neural Network Model for Groundwater Levels Prediction, *Journal of Jilin University (Earth Science Edition)*, **50(01)**, 208-216 (2020).

5. C. Yishan, L. Zhiping, X. Longfei, et al BP neural network method for predicting the inorganic scaling in the reservoir, *Petroleum Geology & Oilfield Development In Daqing*,1-10 (2020).

6. K. Xiangchao *Research on predicting the impact of oil development pollution on groundwater environment*. (Xi'an Shiyou University, 2020).

7. C. Wang, Y. Zhang, T. Shen Risk Assessment and Spatial Analysis of Water Shortage in China, *Journal of Capital Normal University (Natural Sciences Edition)*, **36(1)**, 81-87 (2015).

8. Y. Zhang, L. Wang Research on Coordinated Development among Agricultural Water Resources Utilization, Environment and Economy in Arid Area - A Case Study of Construction Pilot of Water-saving Society in Zhangye City, *Resource Development & Market*, **32(2)**, 142-145 (2016).

9. S. Haykin *Neural Networks: A Comprehensive Foundation, 2nd Edition*. (China Machine Press 2004).

10. L. Goodfellow, Y. Bengio, A. Courville *Deep Learning*. (Posts and Telecommunications Press 2017).

11. F. Chollet *Deep learing with Python*. (Posts and Telecommunications Press 2017).

12. H. Jia *Applied Statistics (5th Edition)*. (University of International Business and Economics Press, 1998).