J. Mater. Environ. Sci. 6 (12) (2015) 3539-3547 *Bijari et al.*
ISSN : 2028-2508
CODEN: JMESCN

Evaluation of artificial neural network models and time series in the estimation of hydrostatic level (Case study: South Khorasan Province-Birjand Aquifer)

Mohammad KazemBijari*, HamidrezaGolkarHamzee Yazd, MojtabaTavousi
Department of Water Engineering, Ferdows Branch, Islamic Azad University, Ferdows, Iran

Received 03 Aug 2015, Revised 27 Nov 2015, Accepted 27 Nov 2015

Corresponding Author: mkbijari@gmail.com

Abstract
Groundwater has always been one of the major sources of drinking and agriculture especially in arid and semi-arid agriculture regions. BirjandAquifer in South Khorasan province, due to its location in the arid region, considers the use of groundwater as the most important and at the same time sole source of freshwater. Modeling and prediction of wells' hydrostatic level is one of the basic tasks to achieve the optimal management of water resources.The current article mainly aims to evaluate the effectiveness of artificial neural network technique and time series in the prediction of hydrostatic level of groundwater. For this purpose, the data on hydrostatic level of 13 piezometer wells present in Birjand plain as one of the sub-basins of Lut Desert was utilized in the 16-year statistical period of 1997-2013 on a monthly basis.In the current study, the discharge parameters including(amount of water extraction per cubic meter from drinking water wells, industry and agriculture), the water entering each polygon in terms of cubic meters (due to precipitationin the area) and water surface level (m) per piezometers in the previous time step were used, and the model output waswater level at the current time step.Wells' hydrostatic level was simulated separately by neural network technique and time series (SARIMA), and finally RMSE, MAE and R^2 were used to determine prediction accuracy of each of two methods. The study results indicated the high precision of both techniques of neural network and time series in the prediction of the hydrostatic level of the wells in the region.

*Keywords***:**Hydrostatic Level; Time Series; Neural Network; Artificial;BirjandAquifer

1.Introduction
The use of artificial intelligence techniques and time series can be of great help in modeling of groundwater. Application of time series in hydrology has been started since four decades ago, and it has reached its peak by presentingBox and Jenkins models.In their research, Mirzaee and colleagues [1] showed that the Box and Jenkins models, based on their own special capabilities, have the ability to predict different time series, especially time series of groundwater data. Daliakopoulos et al.[2] in a research indicated that for the various structures of neural network, Levenberg-Marquardt (LM) training function is more accurate in predicting groundwater level [2].Jothiprakash and Sakhare [3] in their research used neural network model by back-propagation algorithm training, and to evaluate model performance, they utilized three statistical criteria including MSE, RMSE, and R^2. The performance of the models indicated that artificial neural network can be used to predict the groundwater level. Yang et al. [4] showed that the artificial network is able to predict groundwater level more accurate than combined time series. Sreekanth and colleagues [5] showed that the use of neural network by back-prorogation network model and algorithm training (LM) is the most suitable criterion for underground water level prediction. Using neural network model to predict groundwater level, Malekinezhad et al.[6]showed application of feed-forward neural network model with error back-propagation algorithm which is composed of three educational functions (LM), reactionary back-propagation and scaled slope. According to the results obtained, the function (LM) was selected as the best educational function to predict groundwater levels. Pourmohammadi et al. [7] stated in a research that feed-forward neural network

model with error back-propagation algorithm and educational function (LM) can be regarded as the best training function to predict groundwater levels. The aim of this study is to predict the hydrostatic level of groundwater in one of the sub-basins of Lut Desert using artificial neural network and time series and compare them with each other.

2. Materials and Methods

2.1. Location of the study area and data

The study area (Figure 1) is Birjand as a city located in South Khorasan province, with geographic location of 59 degrees longitude and 32 to 33 degrees northern latitude. BirjandAquifer hasa hot and arid climate, average rainfall of 157 mm, minimum and maximum altitude of 1180-2720 meters, the annual average temperature 16.4 ^{o}Cand potential evaporation of 2540 mm per year.BirjandAquifers one the most important agricultural lands of South Khorasan Province. The aquifer of this plain is discharged through the direct infiltration of precipitation, surface flows, return flow from agriculture and green space, drinking water and industry, and through waters flowing from aqueducts outside the scope of feeding balance, and through the extraction of groundwater for various uses, and also groundwater outlet. There are 13 piezometer wells in this sub-basin, which the relevant data of these wells and also a synoptic station in this area were used for the study.

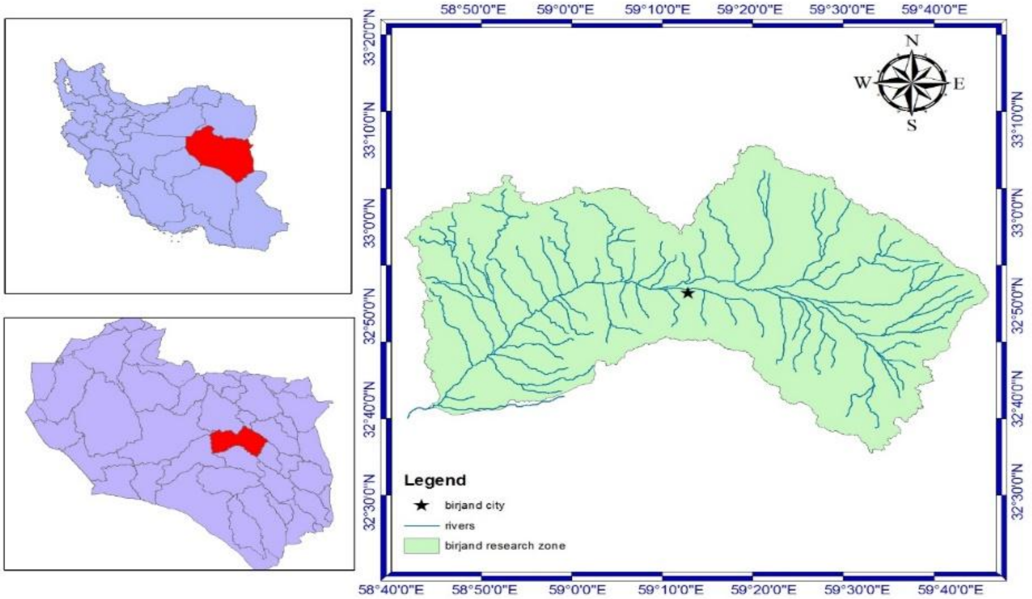

Figure 1: Birjand study sub-basin and propagation (distribution) of rivers in its level

In this study, first,the data relevant to precipitation, evaporation,water flow and hydrostatic level of the wells located in the study area (within a 16-year period since 1997-2013 on a monthly basis) were normalized. Then, using two methods including time series and neural network, water levels in the wells were predicted. In thetime series method, only the data relevant to hydrostatic level of wells were used.In neural network method, the data relevant to precipitation, temperature, wells' dischargedflow, evaporation, and hydrostatic level in previous month were used as the model input to predict thehydrostatic level of the wells. Finally, using criteria to evaluate the accuracy, two methods were compared forthe prediction of hydrostatic level. Each of the stages of the current research is explained as follows.

2.2. Time series

In the current research, ARIMA and SARIMA Models were used to do analyses. ARIMA model as a general model that can represent an extensive class of non-hydrostatic time series is an auto-regressive moving-average integrated process (p, d, and q). Due to the fact that most of time series are practically non-hydrostatic, so this category of processes are used extensively. SARIMA Model is a Seasonal Auto-Regressive Integrated Moving Average Model. In case of the presence of some similarities in a time series after a specified time interval (s), the series has a seasonal behavior with an alternating period (s). The modeling of this model is similar to that of ARIMA, and only alternating period should be taken into account. The extension of SARIMA and ARIMA models is necessary only when time series has both seasonal and non-seasonal behavior. The presence of such behavior makes ARIMA model ineffective. SARIMA model is often named Multiplicative Seasonal Auto-Regressive Integrated Moving Average Model. The overall form of SARIMA for monthly time series is in the form of SARIMA (p, d, q) (P, D, Q) $_S$.

After the normalization of the data relevant to hydrostatic level of wells, the normalized data were inserted into Minitab, and their diagram during statistical period was drawn. To fit the process three process curves were utilized for data fitting including linear, quadratic and exponential curves. Then, given the precision criteria in each curve, the initial appropriate model was chosen. Then, autocorrelation function and partial autocorrelation were plotted, and type of the relevant model was examined and selected. Finally, residual normality assumption, the assumption of constant residual variance, and also independence of data were examined, and the best model was chosen using the results from the properties of statistical test of the predicted data and Akaike information criterion (AIC) extracted from the models. In addition, the data relevant to hydrostatic level was predicted.

2.2.1.Akaike information criterion (AIC)

If there are some acceptable models for a given set of data, better model selection criteria is usually used in terms of statistical parameters or based on prediction errors. One of the selection criteria is AIC which is estimated based on fitted model residuals. Whenever AIC for the rank of p of a model is smaller than other p ranks, then the adequacy of the model parameters is established [8].

3.2.Neural network

Inthisstudy,todetermine the effect ofthemost significant factorsonthe groundwater level, Feed Forward Back Propagation Neural Networks withLevenberg Marquardttrainingfunctionswere used, which are the best method for groundwater level.Back-propagation method is a systematic method for training multilayer networks. The training of a selective network based on the available information includes adjustment of the weight values and biases or initial constant values, in order to minimize the error between observed and calculated output. The algorithm is based on an error correction learning rule. Back-propagation learning rule is used for training of feed-forward multilayer networks, which are commonly called multilayer perception networks. Similar to quasi-Newton methods, Levenberg-Marquardt algorithm seeks to reduce calculations using non-calculation of Hissin Matrix, and it operates very faster than other algorithms. The main disadvantage of Levenberg-Marquardt method is its need to keep voluminous matrices in the memory, which this requires, large space [9]. The network parameters are regulated in a way that the actual response of the network has more and more tendency to favorable response.Artificial neural network model of the data was divided into three parts including test, training and validation, and in 16-year statistical period, 70% of data are used for training (the most important part of the model), 15 percent for the test, and 15 percent for validation.

4.2. Statistical evaluation of the results

Three criteria were used to evaluate the results. RMSE or root mean square error (equation 1), indicates the evaluation of the model accuracy based on difference between the actual values y_{act} and the predicted values y_{es} and naturally whatever it is closer to one, it shows less difference between them, and n is the number of data. Whatever MAE ormean absolute error is smaller, it could be said that estimation error of hydrostatic level is less than that of the obtained models, and it is obtained through the equation 2. In addition, R^3 or coefficient of determination represents the amount of proportion of fitness of linear regression model compared to estimated and observational data pair. The value of this coefficient is always between zero and one, and its proximity to one represents a better performance of the model, and it can be obtained from equation (3):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_{est} - y_{act})^2}{n}} \quad (1)$$

$$MAE = \frac{\sum_{i=1}^{n}|y_{est} - y_{act}|}{n} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{act} - y_{est})}{\left(\sum_{i=1}^{n} y_{act}^2\right) - \left(\frac{\sum_{i=1}^{n} y_{est}^2}{n}\right)} \quad (3)$$

3. Results and Discussion

3.1. The time series

In the curve fitting process, linear model had lower accuracy rates than other models but since the linear process model does not fully cover data changes, it cannot be used as a model for predicting time series, and it is better to use other models such as ARIMA and SARYMA models.

3.1.1. Fit the most suitable model

Model recognition is an experimental knowledge of old data to determine the components of the model. At this stage, autocorrelation methods are applied, and diagrams including autocovariance (ACF) and partial autocovariance (PACF) are used for the diagnosis of series dependence and determining the model's coefficients per unit of time. To determine the value of p in autoregressive model and q in moving average, two diagrams ACF and PACF were used, respectively, and Figure 2 shows ACF and PACF curve in the piezometer (5). In addition, given the time series of data and drop in groundwater levels, piezometer data have a process, which by using simulation, the data process per each piezometer was specified and rank to remove data process was estimated. Figure 4 shows the data process in piezometer 5.
To validate various models in the time series, residual test is used. In this test, in case of the suitability of the selected coefficients, the model has the ability to model all the parameters. Thus, in case of drawing residual correlation model, correlation coefficient does not intersect any confidence line, which this shows the good fit of the model. Given the parameters of the model, to predict groundwater level, Minitab software is used, and to verify the authenticity of modeling, the residual test is performed using software, which their results for piezometer (5) is presented in Figure 3.
Given the model parameters, validation and residual test, the best model based on AIC for each piezometer was chosen, and then groundwater level was predicted for a period of 24 months. In 13 piezometers of Birjand aquifer, different models were chosen for simulation, which the relevant final results are shown in Table 1.
Based on the results derived from this table and the model's parameters, the prediction of groundwater level in the aquifer in two models ARIMA and SARIMA for each piezometer was done. The results of each model and statistical test table in piezometer (5) are presented in Table 2, and the diagram of prediction of each model is presented in Figures 5 and 6. Using the results obtained from the statistical test properties of the predicted data and Akaike information criterion derived from ARIMA and SARIMA models in 13 piezometers of Birjand aquifer, use of SARIMA has more correlation with the observed hydrograph in the region [10].

Table 1: The results from the final models for predicting groundwater level

Piezo meter No.	)q،d،p(ARIMA	s) Q؛D،P)(q،d،p(SARIMA	Piezo meter No	)q،d،p(ARIMA	s) Q؛D،P)(q،d،p(SARIMA
1	(2,1,4)	(2,1,4) $(1,0,1)_{12}$	8	(2,1,2)	(2,1,2) $(1,0,1)_{12}$
2	(1,1,5)	(1,1,5) $(1,0,0)_{24}$	9	(1,2,2)	(1,2,2) $(1,0,1)_{12}$
3	(1,2,1)	(1,2,1) $(1,0,1)_{12}$	10	)(1,1,1	(1,1,1) $(1,0,0)_{12}$
4	(1,1,1)	(1,1,1) $(1,0,0)_{12}$	11	(1,2,1)	(1,2,1) $(1,0,0)_{12}$
5	(2,1,3)	(2,1,3) $(1,0,1)_{12}$	12	(1,1,1)	(1,1,1) $(1,0,0)_{12}$
6	(3,1,3)	(3,1,3) $(1,0,0)_{12}$	13	(1,2,1)	(1,2,1) $(1,0,0)_{12}$
7	(4,1,1)	(4,1,1) $(1,0,0)_{12}$			

Table 2: Specification of the statistical test of the predicted data for piezometer No. 5 with the models including ARIMA (2.1.3) & SARIMA (2.1.3) $(1.0.1)_{12}$

Row	Model	Components	α	SE	T	p-value	SS	MS	AIC
1	ARIMA	2AR	-0.881	0.0475	-18.58	0	4.41	0.0238	-24.25
		3MA	-0.0368	0.047	-0.78	0.435			
		Constant Components	0.013	0.0012	10.51	0			
2	SARIMA	1SAR	0.998	0.0073	136.58	0	3.098	0.017	-34.36
		1SMA	0.944	0.0493	19.15	0			
		Constant Components	0.00007	0.00037	-0.2	0.839			

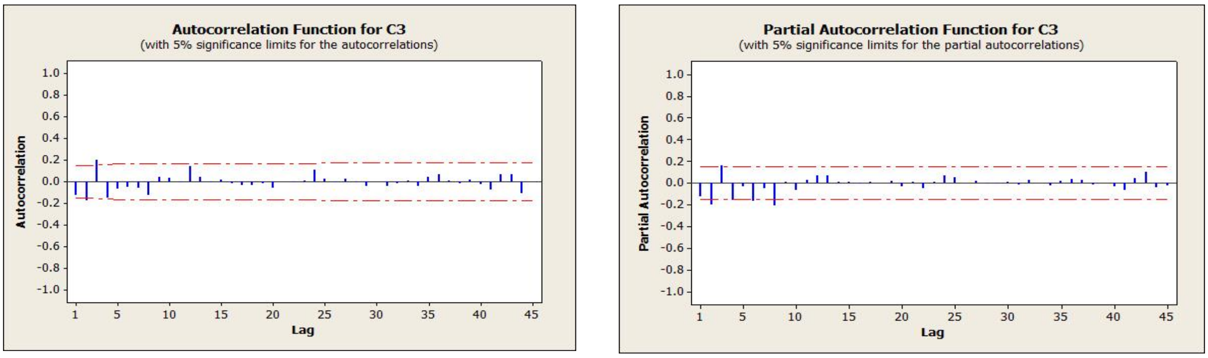

Figure 2: ACFF & PACF Diagram for Piezometer No. 5

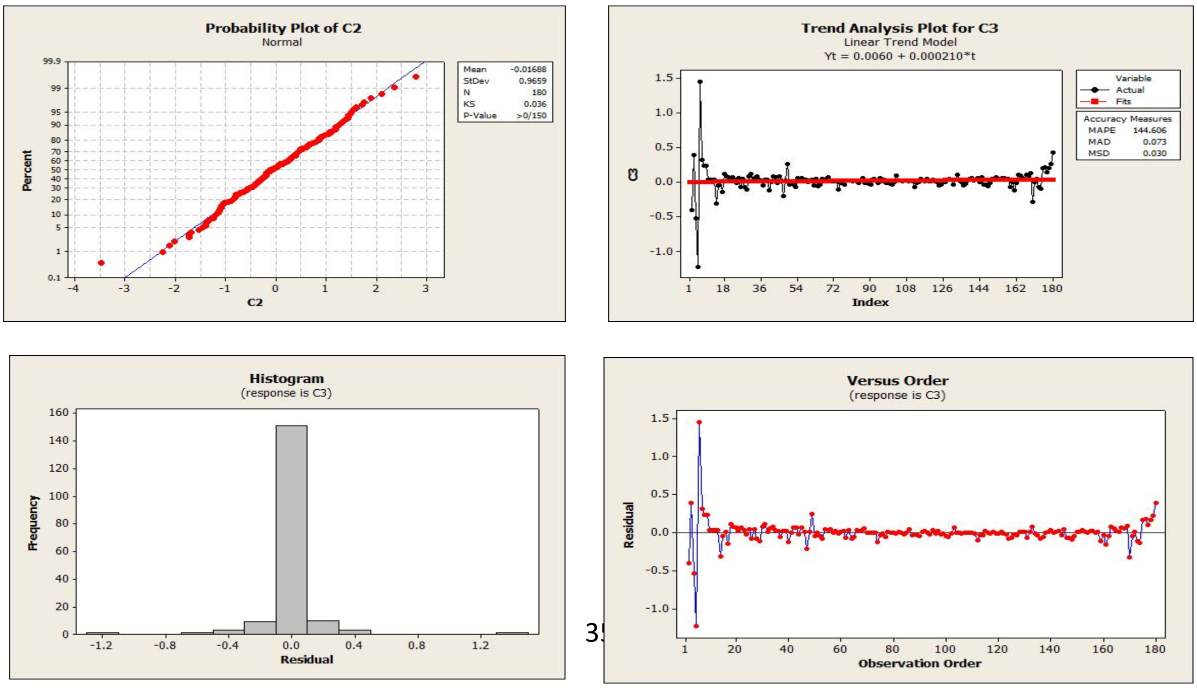

Figure 3: the diagram relevant to the residuals obtained from the fit of ARIMA (2.1.3) Model

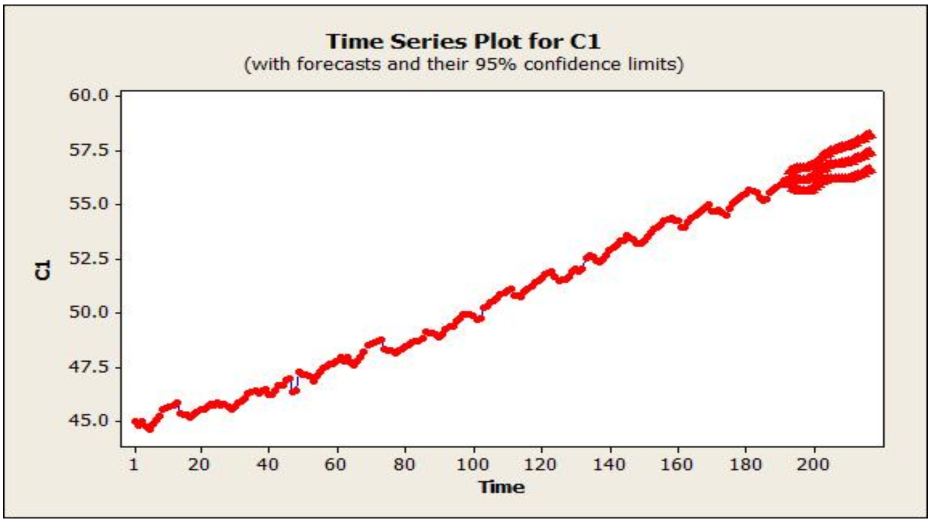

Figure 4: ARIMA Model in predicting Piezometer No. 5

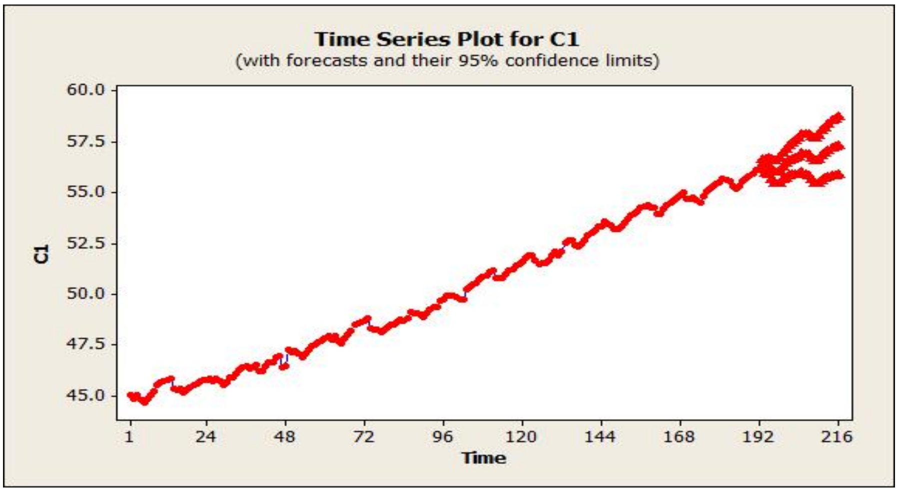

Figure 5: SARIMA Model in predicting Piezometer No.5

3.1.2.Estimatethe aquifer hydrographthrough time series model

Given the volume of groundwater and water exploitation based on regional groundwater balance, paying attention to groundwater hydrograph for decision-making in the management of an aquifer is of critical significance, and given the estimates of the groundwater level in 24 months for doing forecast, it could be said that the hydrographof groundwater in Birjandaquifer using SARIMA model (Figure 6) is closer to the observed results.

3.2. Neural networks

Given the effectiveness of multiple parameters in a regional groundwater level, identify the input parameters of the neural network is very important. Different scenarios for the variables entering artificial neural network were analyzed.

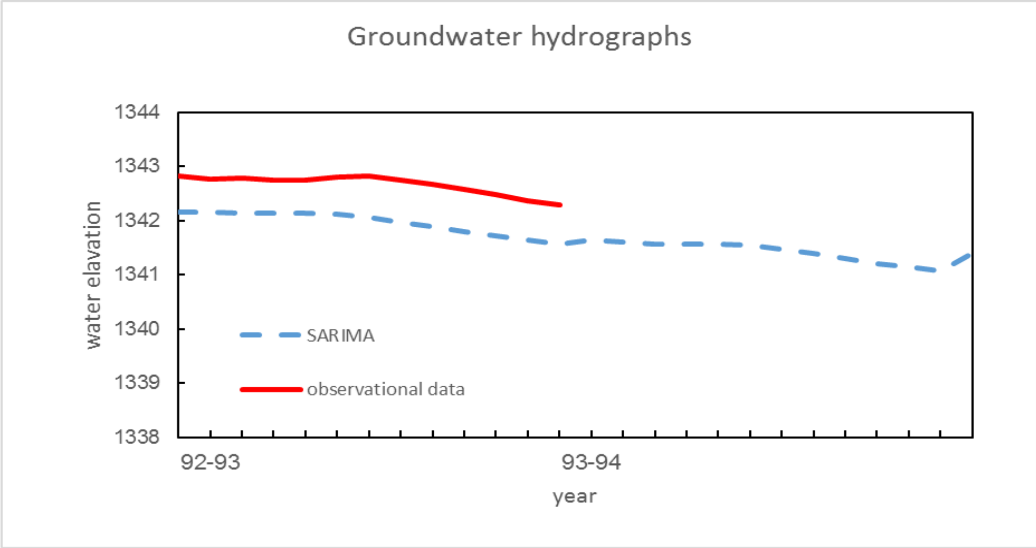

Figure 6: Observational hydrograph and SARIMA Model in predicting Birjand aquifer hydrograph

Some parameters including temperature, evaporation and precipitation are considered to be meteorological parameters, and feeding variables and aquifer discharge (the same month and six-month delay) and water (the same month and with a delay of six months) are regarded as environmental parameters. In this study, 10 scenarios were analyzed, which according to the type of artificial neural network and network evaluation criteria, three factors including the aquifer feeding, groundwater extraction, and groundwater level in previous month, were the most effective scenarios in all piezometers in order to predict groundwater level, and they were used as fundamental variables. The correlation coefficient of this model for the data on training, testing, and validation of piezometers (1) are provided in Figure 7.

3.2.1. Estimation of the aquifer hydrograph by using neural network model

Given the estimation of groundwater level in 23 months in order to do prediction, Birjand aquifer groundwater hydrograph is drawn in the form of Figure 8 using the neural network, and the results indicate the high accuracy of feed-forward neural artificial network.

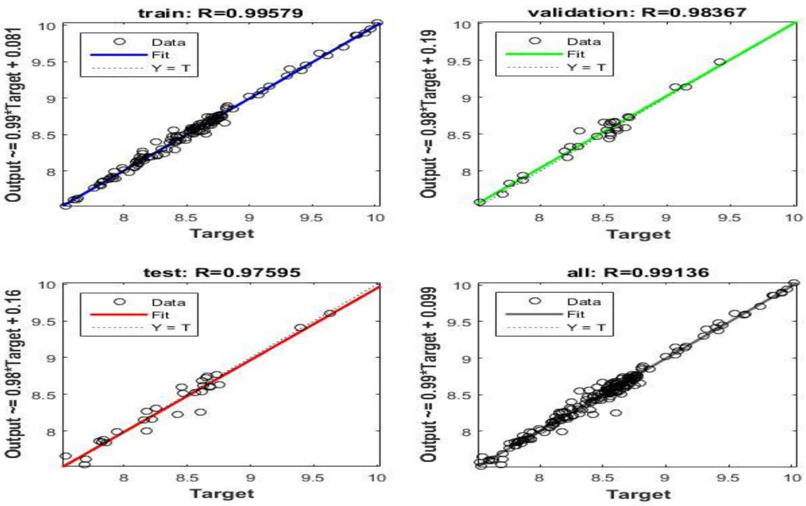

Figure 7: the results from the outlet of groundwater level obtained in the piezometer 1 through neural model

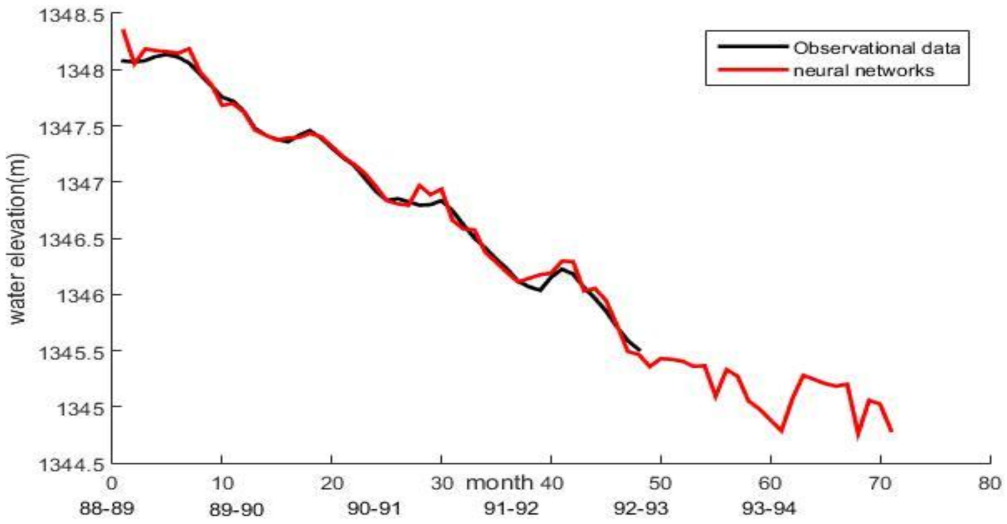

Figure 8: a comparison of observational and predicted groundwater hydrograph in Birjand Aquifer

4.A comparison of the efficiency of two models

A comparative analysis on the efficiency of two models namely neural network and SARIMA in terms of their ability to predict hydrostatic level was done using error evaluation criteria. According to Table 3, R^2 of the data simulated by using feed-forward neural network models is higher than that of time series (SARIMA), and RMSE and MAE in the neural network model is less than that of SARIMA model. In general, both feed-forward neural network model and SARIMA time series have a high correlation with actual data of the hydrostatic level of the wells, and have absolute error and quantitative standard. However, feed-forward neural networks are more efficient in predicting hydrostatic level.In addition, prediction using SARIMA has a good correlation with the observed hydrograph in the region. One of the main reasons for more consistency of SARIMA Model than ARIMA in predicting groundwater level is that the data of groundwater level are highly influenced by meteorological data and the extraction of groundwater.This process is done every year in a cyclical way and has a cycle of 12 months in the year. Thus, the results of simulation using SARIMA model is more similar to the observed results [11].

Table 3: A comparison of two models including neural network and ARIMA in predicting hydrostatic level of groundwater

Type of Model	Feed-forward Neural Network	SARIMA
R^2	0.953	0.819
RMSE	0.094	0.683
MAE	0.077	0.677

Figure 9 shows the values of hydrostatic level predicted by two models including time series and neural network compared with the real data of wells' hydrostatic level. Given the graph of feed-forward neural network, time series SARIMA and real data have a fully compatible and similar process.
The results from the current research showed the efficiency of the two models to predict hydrostatic level of groundwater. This comparison and an evaluation of the efficiency of the two models are also evident in the results from the other research.

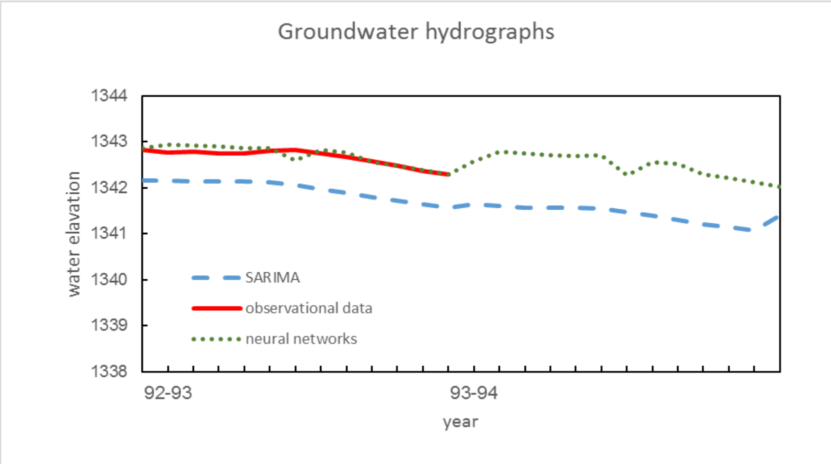

Figure 9: the hydrostatic level values predicted by to models including time series and neural network through real data

Another result of this research is better correlation and lesser error of neural network in the prediction of hydrostatic level in comparison to the time series model. The point that should be noted here is benefits of use of the time series compared to neural network. One of these cases is that, unlike artificial neural model, equations and relationships in time series are clearly specified [12]. Another advantage of using time series in hydrologic sciences, especially the prediction of hydrostatic level is that in time series, there is no need to other sets of data as input, and the model is able to do prediction only by using the same data on hydrostatic level or desired parameter. Nevertheless, the neural network should have input parameters which are able to impact the predicted parameter. This point is especially important in the areas with lack of data and valid statistics.

References

1. Mirzai S.Y., Chitsazan M., ChiniPardaz R., *Geology Conference in Tehran*. (In Persian) (2005).
2. Daliakopoulos I. N., Coulibaly P., TsanisI. K.,*Journal of Hydrology*, 309(2005) 229-240.
3. Jothiprakash V. Sakhare S.*The 12th International Conference of International Association for Computer Methods and Advances in Geomechanics (IACMAG*), Goa, India. (2008).
4. Yang Z.P., LuW.X., Long Y. Q., LiP., *China, Journal of Arid Environments*, 73 (2009) 487-492.
5. Sreekanth P. D., Geethanjali N., Sreedevi P. D., Ahmed S., Ravi, N., KamalaP. D., *Current Science*. 96(2009) 1-7.
6. Malekinejad H., Pourshari'ayti R.,*Journal of Agricultural Science and irrigation*, Volume Sciences. 36 (2012) 81-92.
7. Pourmohammadi S., Malekinejad H., Pourshari'iyati R., *Journal of Soil and Water Conservation and Natural Resources*,Gorgan University of Agricultural Sciences, 12 (2012)201-210.
8. Safavi H. R.,*Isfahan Industrial University*,(2011) 706
9. Kia M., *KianRayanehSabz*(2009) 229
10. Mohtasham D., Akbarpour A., MeftahHalaqi M., E'etebari B. A., *Journal of Irrigation and Drainage;* 4 (2010) 1-10.
11. Chandramouli V., Lingireddy S.,BrionG. M.,*Journal of Computing in Civil Engineering*, 21 (2007) 39- 46.
12. Nayak P., Satyajirao Y.R., Sudheer K.P.,*Water Resources management*, 2(2006) 77-99.