CrossMark

ORIGINAL ARTICLE

Combined gamma and M-test-based ANN and ARIMA models for groundwater fluctuation forecasting in semiarid regions

Bahram Choubin[1] · Arash Malekian[2]

Abstract The shortage of surface water in arid and semi-arid regions has led to the more use of the groundwater resources. In these areas, the groundwater is essential for activities such as water supply and irrigation. One of the most important stages in sustainable yield of groundwater resources is awareness of groundwater level. In this study, we have applied artificial neural networks (ANN) and autoregressive integrated moving average (ARIMA) models for groundwater level forecasting to 4 months ahead in Shiraz basin, southwestern Iran. Time series analysis was conducted according to the Box–Jenkins method. Meanwhile, gamma and M-test were considered for determining the optimal input combination and length of training and testing data in the ANN model. The results indicated that performance of multilayer perceptron neural network (4, 14, 1) and ARIMA (2, 1, 2) is satisfactory in the groundwater level forecasting for one month ahead. The performance comparison shows that the ARIMA model performs appreciably better than the ANN.

Keywords Groundwater level · Artificial neural network · ARIMA · Forecasting · Gamma test · M-test

✉ Arash Malekian
malekian@ut.ac.ir

Bahram Choubin
bahram.choubin@ut.ac.ir

[1] Sari University of Agricultural Sciences and Natural Resources, P.O. Box 737, Sari, Iran

[2] University of Tehran, Karaj 31585-3314, Iran

Introduction

One of the most important factors in wise management of water resources is a proper attitude and a vision of future events which may happen. This has not been exempted in water resources management. Awareness of the status of water resources in a region, especially in arid and semiarid regions, where groundwater is scarce and vital plays an important role in the planning process for different sectors such as domestic, industry and agriculture. Due to the stochastic nature of hydrologic parameters such as groundwater level, its status in the future can be predicted using statistical analysis, mathematical models, etc. Evaluation and forecast of groundwater level through specific models help in groundwater resources management. Hence, we can use time series modeling to predict groundwater level fluctuations during the following months for optimal and proper management of groundwater resources.

Since groundwater resources are mostly related to many factors and have complex fluctuations, it is necessary to decompose the complexity and their variations by mathematical methods (Lu et al. 2013). Among the different available robust tools, the artificial networks (ANNs) and ARIMA models are commonly used to hydro-climatological variables forecasting (Choubin et al. 2014; Sigaroodi et al. 2014; Choubin et al. 2016a, b, 2017a).

ARIMA models are a mathematical approach capable to simulating the both stationary and non-stationary time series. However, these models are lesser studied in the field of groundwater resources. In recent years, ARIMA model has been used for predicting hydro-meteorological parameters (e.g., Boochabun et al. 2004; Abghari et al. 2010; Chattopadhyay and Chattopadhyay 2010; Chattopadhyay et al. 2011; Zakaria et al. 2012). Also, Lee et al. (2009)

Fig. 1 Location of Shiraz basin, Iran
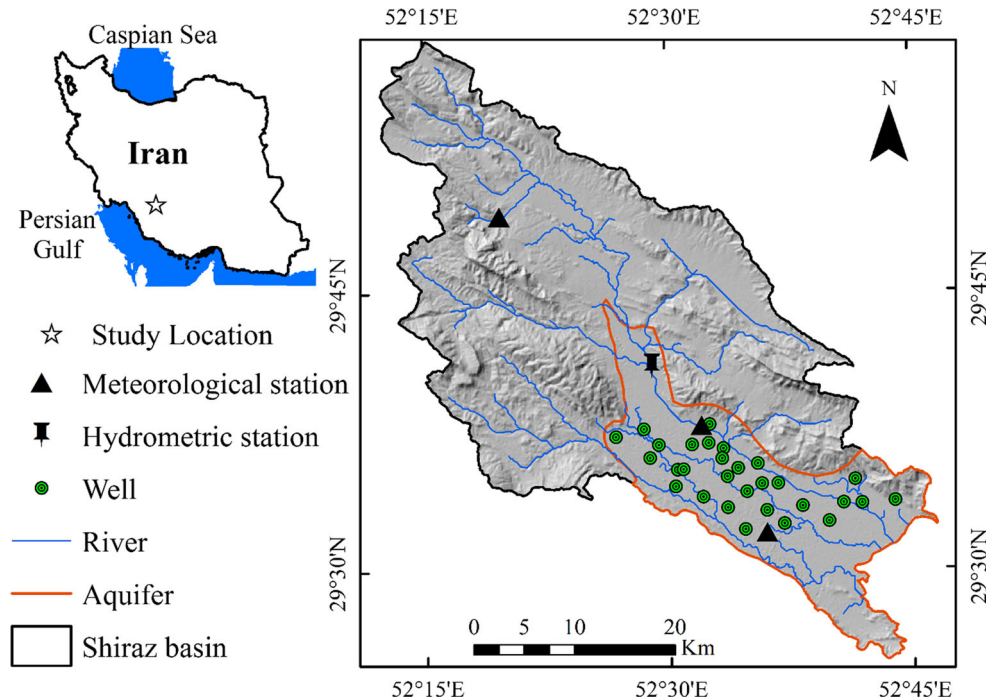
used ARIMA model according to the Box–Jenkins method to groundwater level forecasting in Changwon, Korea.

The intelligence knowledge methods such as neural networks as have been applied for groundwater level forecasting (Coulibaly et al. 2001; Lin and Chen 2005; Daliakopoulos et al. 2005; Bidwell 2005; Nayak et al. 2006; Tsanis et al. 2008; Trichakis et al. 2009; Banerjee et al. 2009; Sethi et al. 2010; Dash et al. 2010; Behzad et al. 2010; Nourani et al. 2008; 2011; Shirmohammadi et al. 2013). However, in the previous studies, determining the optimal input variables for nonlinear models (such as ANN) in groundwater modeling is less considered. In this regard, Rashidi et al. (2016) mentioned that determination of optimal parameters in nonlinear modeling is important. They used gamma test to selecting the best input to simulate the suspended sediment. Also, Jajarmizadeh et al. (2015) applied gamma test to identifying the best combination of the input variables for support vector machines (SVM) to predict the stream flow in a semiarid basin in Iran.

Therefore, the objectives of this research are (1) determining the optimal input combination for ANN modeling approach; (2) selecting the best length of data during training and testing periods in the ANN model; and (3) comparing the performance of linear (ARIMA) and nonlinear (ANN) mathematical models in monthly groundwater level forecasting at a semiarid region of Iran. Besides the time series model considered for groundwater level forecasting, another advantage of this study is determining the optimal input combination and best length of training and testing data in the ANN model based on the gamma and M-tests.

Materials and methods

Study area and data

The study area is located in Shiraz basin, Fars province, southwestern Iran. Shiraz basin extends between $52°12'$ and $52°45'$ E longitude and $29°25'$ to $29°58'$ N latitude and 1450 km^2. Location of Shiraz aquifer and piezometric monitoring wells, hydrometric and meteorological stations is shown in Fig. 1. The long-term average annual precipitation of Shiraz plain is 350 mm. The time period considered in this study is 18 years (1993–2010), and the data used are including monthly total precipitation, monthly average stream flow, temperature, evaporation and groundwater level.

ARIMA models

Box and Jenkins (1970) introduced autoregressive integrated moving average (ARIMA) models which are a class of linear models representing stationary and non-stationary time series. If non-stationarity (d) is combined to a mixed ARMA (p, q) model, then the general ARIMA (p, d, q) is obtained. Equation for non-seasonal ARIMA model of order (p, d, q) for a standard normal variable (Z_t) is as follows (Box and Jenkins, 1970):

$$\varphi(B)(1 - B)^d Z_t = \theta(B)\,\varepsilon_t \tag{1}$$

In Eq. 1, $\phi(B)$ and $\theta(B)$ polynomial of degree p and q, respectively, are:

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p \qquad (2)$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \qquad (3)$$

where p is the number of autoregressive terms, d is the number of differences and q is the number of moving average terms.

The time series modeling with Box–Jenkins approach is consisting three steps namely identification, estimation and diagnostic check (Box and Jenkins 1970). In this study, the time series were tested for normality and then Augmented Dickey–Fuller (ADF) and Phillips–Perron (PP) tests were used to analyze groundwater level time series stationarity. Non-stationary series converted to stationary ones through the method of differencing (Yurekli et al. 2007) that number of differencing determined the value of d. ADF or unit root test by Dickey and Fuller (1979) and PP method by Phillip and Perron (1988) were conducted. Then, the graphical properties of the autocorrelation function and the partial autocorrelation function were used in the estimation step, to determine the value of p and q. To select the best fitted model, we used the minimum amount of Akaike Information Criterion (AIC) and Schwarz Bayesian Criterion (SBC). In the general case, the *AIC* is (Akaike 1974):

$$AIC = -2\log(L) + 2m \qquad (4)$$

where m is the number of parameters in the statistical model and L is the maximized value of the likelihood function for the estimated model. *SBC* criterion (Schwarz 1978) is similar in use to Akaike's index which is defined as:

$$SBC = -2\log(L) + m\ln(n) \qquad (5)$$

where n is denotes the number of observations.

In the diagnostic checking step, the models must be checked for adequacy. In this study, we used Kolmogorov–Smirnov (K–S) test and P–P plot to check the normality of residuals, while Portmanteau test was considered as the criterion to determine the independence of the residuals.

Artificial neural networks

An artificial neural network retrieved from natural nerve cells in order to transform the inputs into meaningful outputs. In this study, we used a feedforward artificial neural network called multilayer perceptron (MLP) for groundwater level forecasting. According to Kim and Valdés (2003), MLP is able to simulate 90% of the processes related with the climate. The Levenberg–Marquardt algorithm is one of the fastest methods implemented with high performance for neural network training (Huang et al. 2006). So, we have used it as the training algorithm in the MLP, also the Logsig and Purelin transfer function in the hidden and output layers. The time lags of $t-1$, $t-2$, $t-3$ and $t-4$ for input layers were chosen to forecasting of the monthly groundwater level from one to 4 months ahead ($t + 1$, $t + 2$, $t + 3$, $t + 4$), while hidden neurons was determined by trial-and-error process.

Gamma test

Koncar (1997) and Agalbjörn et al. (1997) reported the gamma statistic (Γ) which can provide the best mean square error in any nonlinear smooth models (Han et al. 2010). The gamma test is based on $N[k,i]$, which are the kth ($1 \le k \le p$) nearest neighbors $x_{N[k,i]}$ for each vector $x_i (1 \le k \le p)$. Particularly, the gamma test is taken from the Delta function of the input vectors (Moghaddamnia 2009c),

$$\delta_m(k) = 1/M \sum_{i=1}^{M} |x_{N[k,i]} - x_i|^2 \ldots (1 \le k \le p) \qquad (6)$$

where$|\ldots|$ gives the meaning Euclidean distance, and the corresponding gamma function of the output values,

$$\gamma_m(k) = 1/2M \sum_{i=1}^{M} |y_{N[k,i]} - y_i|^2 \ldots (1 \le k \le p) \qquad (7)$$

where $y_{N[k,i]}$ is the corresponding y value for the kth nearest neighbor of xi in Eq. 6. In order to calculate Γ, a least squares regression line is constructed for the p points $(\delta_m(k), \gamma_m(k))$.

$$\gamma = A\delta + \Gamma \qquad (8)$$

The intercept on the y axis ($\delta = 0$) is the Γ value, as can be shown, $\gamma_m(k) \to Var(r)$ in probability as $\delta_m(k) \to 0$.

The graphical output of Eq. 7 provides valuable information. First, the intercept (Γ) on the y axis (or gamma) represents an estimate of the best MSE attainable utilizing a modeling method for unclear smooth functions of continuous variables (Evans and Jones 2002). Second, the gradient gives the complexity of model (whatever slope be steeper indicates that model have greater complexity), (Moghaddamnia 2009c). V-ratio returns a scale invariant noise estimate between 0 and 1. A V-ratio close to zero shows a high degree of predictability (by a smooth model) of the specific output. The V-ratio is obtained by dividing the gamma to the output data variance, (Durrant 2001). Smaller values of the gamma and V-ratio indicate the optimal combination of the used input data (Agalbjörn et al. 1997; Končar 1997).

M-test

Determining the proper length for the training data is important to improve the prediction (Choubin et al.

2014). Wingamma M-test curve is a method for determining the number of data required to produce a stable asymptote. Here, we used M-test based on the V-ratio and gamma value to select the best length of training and testing data in the neural networks method similar to some other works (e.g., Evans and Jones 2002; Remesan et al. 2008; Moghaddamnia et al. 2008; 2009a, b; Piri et al. 2009; Tsui et al. 2002; Piri et al. 2009; Singh 2005; Stefansson et al. 1997; Noori et al. 2010; Han et al. 2010). The values of V-ratio and gamma statistics are determined with increasing number of data points. Data length is determined based on M-test curve stabilized for a specific value of V-ratio and gamma statistics. This test reduces overfitting in the nonlinear modeling (Shamim et al. 2016).

Data normalization

Data normalization is the best way to ensuring data integrity and eliminating redundancy (Choubin et al. 2017b). Thus, the hydrologic data must be normalized, and the best range recommended for normalization is between 0.05 and 0.95 (Hsu et al. 1955). Thus, the series was normalized to the range [0.05, 0.95] as follows:

$$X_{norm} = 0.05 + 0.95 \frac{X - X_{min}}{X_{max} - X_{min}} \tag{9}$$

where X_{norm} and X_r are the normalized and the original inputs, and X_{min} and X_{max} are the minimum and maximum of input ranges, respectively.

Performance criteria

The performance criteria used in the current research are RMSE, MAE and R (Eqs. 10, 11 and 12). Also, Violin plot (Hintze and Nelson 1998) was used to visual diagnostic analysis.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (O_i - P_i)^2} \tag{10}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |(O_i - P_i)| \tag{11}$$

$$R = \frac{\sum_{i=1}^{N} (O_i - \overline{O})(P_i - \overline{P})}{\sqrt{\sum_{i=1}^{N} (O_i - \overline{O})^2 \sum_{i=1}^{N} (P_i - \overline{P})^2}} \tag{12}$$

where N is the number of data points, O_i and P_i are the observed and predicted value, $\overline{O}$ and $\overline{P}$ are the mean of the observed and predicted values, respectively.

Results

Selection of the ARIMA model structure

At this step, the stationary and normality status of the GL time series were investigated. Table 1 shows the result of ADF and PP test before and after differencing. The null hypothesis of the ADF and PP test is H_0: $\theta = 0$ (i.e., the data are non-stationary and need to be differenced to make it stationary). When the opposite hypothesis is true that P value is lower than confidence level ($\alpha = 0.01$). Table 1 indicates unit root test for assessing the stationary status of the GL time series. First, unit root test was conducted for groundwater level time series (i.e., the observed data without any differencing). According to Table 1 and the significance level of ADF and PP test statistic (P value > 0.01), GL data are non-stationary and need to be converted to stationary ones for time series modeling. Then, stationarity of data was evaluated through first differencing of the time series. The results indicate that the GL time series is stationary after the first differencing (P value < 0.01; Table 1). Afterward, using the Box–Jenkins method in the estimation step, the orders of p and q ($p \leq 2$ and $q \leq 4$) were determined through the graphical properties of the autocorrelation function and the partial autocorrelation function. The best fitted model among the different models was identified based on the orders of p and q and evaluation of AIC and SBC criteria through trial-and-error method. The best model was ARIMA (2, 1, 2) with lowest AIC and SBC than other candidate models (114.2 and 129.9, respectively).

The result of Portmanteau test showed that the residuals are independent, since Ljung–Box–Pierce statistic, i.e., Q statistic is less than χ^2 value ($Q = 32.399 < \chi^2 = 33.4$) with degrees of freedom equal to 17 in the one percent confidence level. The normality of the residuals was confirmed through probability–probability (P–P) plot and Kolmogorov–Smirnov Test (Z value of K–S test is equal to 0.45 with the P value of 0.98 which is greater than 0.05, so the residuals distribution is normal). The result of Portmanteau test and K–S test showed that ARIMA (2, 1, 2)

Table 1 ADF and P–P test for evaluation of stationary status of the GL time series

Time series	Unit root test	Statistic	P value
Before differencing (level)	ADF test	1.20	0.55
	PP test	1.12	0.57
After differencing (one differencing)	ADF test	11.34	0.003
	PP test	54.14	0.001

Table 2 Coefficients of the ARIMA (2, 1, 2) model

Model parameters	Coefficients	Standard error	t	P value
φ_1	1.7281	0.004	468.85	0.00
φ_2	−0.9988	0.004	−254.75	0.00
θ_1	1.7073	0.039	43.55	0.00
θ_2	−0.958	0.045	−21.16	0.00
Constant	−0.006	0.023	−0.24	0.81

can be adequately used for prediction purposes. Table 2 shows the coefficients for the ARIMA (2, 1, 2) model. As regards to $\phi_1 + \phi_2 < 1$ and $\theta_1 + \theta_2 < 1$, the values obtained are allowable. Finally, the forecast of ARIMA (2, 1, 2) is generated by the following equation for the next month (t + 1).

$$Y_{t+1} = Y_t + \varphi_1 Y_t - \varphi_1 Y_{t-1} + \varphi_2 Y_{t-1} - \varphi_2 Y_{t-2} - \theta_1 e_t - \theta_2 e_{t-1} + c \tag{13}$$

$$Y_{t+1} = Y_t + 1.7281 Y_t - 1.7281 Y_{t-1} - 0.9988 Y_{t-1} + 0.9988 Y_{t-2} - 1.7073 e_t + 0.958 e_{t-1} - 0.006 \tag{14}$$

where Y is the groundwater level and e is the white noise (the difference between observed and predicted groundwater level).

Model input selection and training data length

Mostly, the limiting factor on the predictive accuracy of the model will be measurement noise or insufficient data. Wingamma software package estimates the least mean squared error that any smooth data model can achieve on the given data without over-training. In this study, we have determined the best combination of input data, length of training and testing data with gamma test and M-test, respectively. To determine the best combination of input data, the different combinations were applied to assess their influence on the groundwater level modeling. We used genetic algorithms (GA) for finding the best combinations that the optimal combination has minimum of gamma (Γ). The goal of model identification for a particular output is to choose a selection of inputs that minimizes the asymptotic value of the modulus of the gamma statistic. At each time step ahead (up to 4 ahead steps), we choose the suitable combination of the inputs including precipitation (P), stream flow (SF), temperature (T), evaporation (E) and groundwater level (GL). Table 3 shows the different combination for 1 month ahead. The optimal combination was selected on the basis of the least amount of V-ratio and gamma statistic. Table 3 clearly shows that V-ratio and gamma statistic in the 10111 mask are less than others. Therefore, the combination of precipitation (P), temperature (T), evaporation (E) and groundwater level (GL) can

Table 3 Determining the best combination for GL forecasting in 1 month ahead

Input	Gamma	Standard error	V-Ratio	Mask
E, GL	0.0020	0.0	0.045	00011
T, GL	0.0020	0.0	0.048	00101
P, GL	0.0020	0.0	0.05	10001
P, T, GL	0.0010	0.0	0.029	10101
SF, E, GL	0.0010	0.0	0.033	01011
SF, T, GL	0.0010	0.0	0.042	01101
P, SF, GL	0.0020	0.0	0.045	11001
P, T, E, GL	0.0009	0.0	0.021	10111
SF, T, E, GL	0.0010	0.0	0.036	01111
P, SF, E, GL	0.0011	0.0	0.026	11011
P, SF, T, E, GL	0.0013	0.0	0.028	11111

P precipitation; SF stream flow; T temperature; E evaporation; GL groundwater level
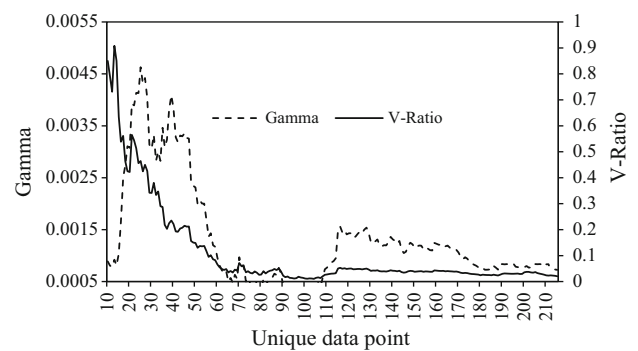

Fig. 2 M-test curve: the variation of gamma statistic and V-ratio with unique data points to determining the proper length for training data

make a good model compared to the other inputs combination (for 1 month ahead).

After achieving the optimal input combination, M-test was used to determine the proper length of training and testing data (Fig. 2) for the best combination of 10111 model (i.e., with P, T, E, GL) in the 1 time step ahead. M-test curve stabilized around 180 data points with the gamma statistic equal to 0.00083. The value of V-ratio is close to zero in the 180 data points which indicate a high degree of predictability of the output data by a smooth model. Therefore, the best length of training data is about

Table 4 Optimal input combination and data length obtained through gamma and M-test

Forecasting step	Optimal input combination	Mask	Training length		Testing length	
			N	%	N	%
1 month ahead	P, T, E, GL	10111	180	83	35	17
2 months ahead	P, D, E, GL	11011	198	92	16	8
3 months ahead	P, D, T, E, GL	11111	183	85	30	15
4 months ahead	P, D, T, E, GL	11111	182	85	30	15

*P* precipitation; *SF* stream flow; *T* temperature; *E* evaporation; *GL* groundwater level

Table 5 Performance of MLP and ARIMA models for GL forecasting

Model	Structure	Forecasting step (month)	Training period			Testing period		
			RMSE	MAE	R	RMSE	MAE	R
MLP	(4, 14, 1)[a]	1 ahead	0.305	0.234	0.974	0.537	0.446	0.874
	(4, 10, 1)[a]	2 ahead	0.901	0.682	0.861	1.403	1.254	0.691
	(5, 8, 1)[a]	3 ahead	0.876	0.695	0.860	1.247	1.092	0.718
ARIMA	(5, 12, 1)[a]	4 ahead	1.405	1.137	0.608	1.455	1.297	0.459
	(2, 1, 2)[b]	1–4 ahead[c]	0.325	0.241	0.968	0.209	0.171	0.980

[a] Number of neurons in input, hidden and output layers

[b] Orders of *p*, *d*, and *q*

[c] ARIMA model simulates same data for training and testing period in each time ahead, so performance results are similar in each time step forecasting

Fig. 3 Scatter plots between observed and forecasted data by MLP and ARIMA in 1 month ahead for testing data sets
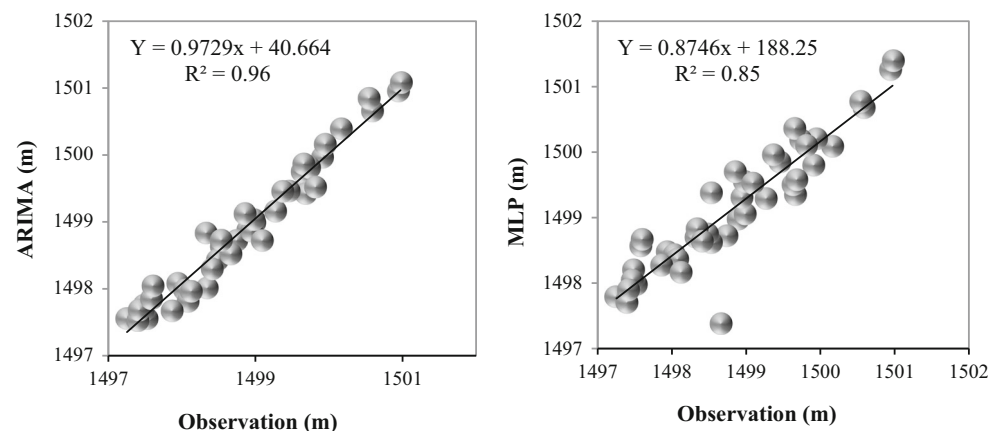

180 data (i.e., 83% of the total data). The result of gamma and M-test in the model input selection and training and testing data length for 1–4 time steps ahead is shown in Table 4.

Results of forecasting groundwater level by ARIMA and MLP network

The multilayer perceptron (MLP) neural network and ARIMA modeling were done for forecasting groundwater level. We used the results of gamma test (the optimal input combination) and M-test (training and testing data length) to forecasting of groundwater level by MLP neural network (Table 4). The root-mean-square error (RMSE), mean

absolute error (MAE) and correlation coefficient (R) were calculated to check the accuracy of the models performance (Table 5). The results indicate that ARIMA (2, 1, 2) has a better performance than the MLP. So that in the ARIMA model, RMSE and MAE values are less whiles the value of R is more than the MLP. It is noticeable that ARIMA model predicts based on the historical data, so the model performance is not different in the months ahead. The result of MLP (4, 14, 1) neural network shows that model has better performance in the 1 month ahead forecasting. MLP (4, 14, 1), i.e., multilayer perceptron network with 4 input neurons (obtained by the gamma test), 14 neurons in the hidden layer (obtained by trial and error) and has one output neuron. Figure 3 shows the scatter plot of

Fig. 4 Observed versus forecasted data by MLP and ARIMA in 1 month ahead
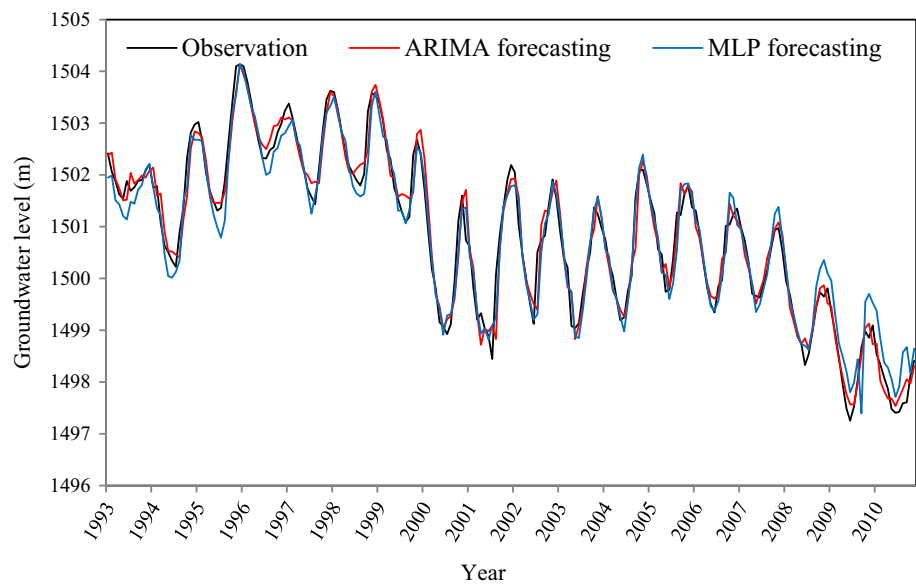
testing data sets between the observed and forecasted by MLP (4, 14, 1) and ARIMA (2, 1, 2) for 1 month ahead. As shown, ARIMA have better fit with the observation ($R^2 = 0.96$) than MLP model ($R^2 = 0.85$). Figure 3 confirms higher accuracy of the results obtained from ANN and ARIMA in the forecasting GL and the observed versus forecasted data results by MLP and ARIMA in 1 month ahead are presented in Fig. 4.

In addition to the performance criteria (Table 5) and scatter plot (Fig. 3), we applied the Violin plot (Hintze and Nelson 1998) to evaluate the model performance. This plot is a boxplot combined with kernel density plots, to show the probability distribution of the data (Choubin et al. 2017a). The Violin plot (Fig. 5) indicates the visual performance of models in forecasting the GL in 1 month ahead, where the ARIMA model has better fit with the observation compared with the MLP model. As, the median of the observed data is well predicted by ARIMA (white points in the graphs), also the 25th and 75th percentiles (thick lines in plots) in ARIMA have better fit than the MLP model. Although, ARIMA overestimated the 5th percentile (thin lower line in violin plots) of GL data than MLP but have closer fit with the observation in 95th percentile (thin upper line in the violin plots).

Discussions

One of the most important stages in sustainable utilization of groundwater resources is understand of groundwater level fluctuations. Exploitation and utilization of groundwater resources in the Shiraz aquifer and persistently drought periods in recent years are caused a dramatic reduction in groundwater table. As a result, forecasting the
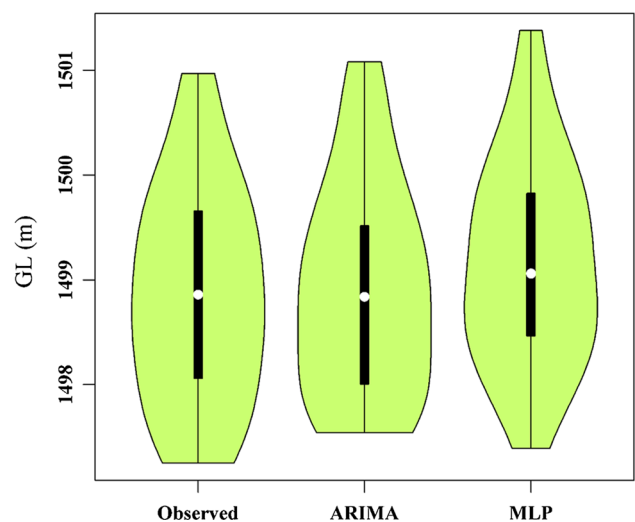

Fig. 5 Violin plots for comparison of the models performance in 1 month ahead

groundwater level as a tool for better and proper management is very crucial and important issue in this the plains.

In this study, we tried to forecast the groundwater level (GL) from one to four months ahead in Shiraz plain, Iran. The result of the ANN indicated that the model has better performance in the 1 months ahead forecasting. Similarly, Shirmohammadi et al. (2013) reported that prediction of groundwater level for 1 and 2 months ahead is better than 3 months ahead.

We also evaluated various performance criteria to examine the abilities of ANN and ARIMA models in forecasting the GL. Although the result of ANN model was satisfactory in the one month ahead forecasting (RMSE = 0.537, MAE = 0.446 and $R = 0.874$), the

evaluation results showed that the ARIMA model performs better than the ANN (RMSE = 0.209, MAE = 0.171 and R = 0.980). Lee et al. (2009) have obtained satisfactory results for groundwater level forecasting by ARIMA model according to the Box–Jenkins method. Also, some other studies (Voudouris 2002; Aflatooni and Mardaneh 2011; Adhikary et al. 2012; Lu et al. 2013) have successfully demonstrated the performance of ARIMA model in the groundwater level forecasting. Lu et al. (2013) suggested that ARIMA model has less accuracy in groundwater level forecasting compared with the decomposition method in China. The scatter and violin plots of current study reveal that the predicted values have suitable fit with the observed data, both in ANN and in ARIMA models, although the ARIMA model performance is better than MLP neural network. Narayanan et al. (2013) suggested that ARIMA modeling is capable to forecast of premonsoon rainfall over the northwest part of India. Yang et al. (2009) indicated that the backpropagation ANN (BPANN) model is superior to the integrated time series (ITS) in forecasting the groundwater level time series.

Selection of the proper input variables and the training data length in the neural network method using gamma and M-test is one of the advantages of this study. Jajarmizadeh et al. (2015) and Rashidi et al. (2016) suggested that preprocessing the input variables in forecasting process by nonlinear models is important as confirmed by the current study results. Also, Kakaei Lafdani et al. (2013) indicated that ANN models based on gamma test can estimate accurately during training and testing periods.

According to the Moghaddamnia et al. (2009c), gamma test reduces huge workload of the trial-and-error process prior to the actual model development. One reason for efficiency of the gamma test is that it can immediately tell us directly from the data whether or not we have sufficient data form a smooth nonlinear model and how a model can present good results.

Conclusions

The results show that both of ANN and ARIMA have good forecasting accuracy, and they are suitable for the forecasting the groundwater level in semiarid regions. This study presented how the gamma test and M-test can be applied together to reduce the huge workload of the trial and error in nonlinear modeling process. In general, the potential of identifying the input parameters and best length of training data may turn gamma test and M-test as an efficient technique for preprocessing the data to predict the groundwater level. It might be helpful for future researches to use these methods as a time-consuming

approach for swiftly attaining the appropriate results. We, in this study, indicated that both performance of MLP (4, 14, 1) and ARIMA (2, 1, 2) are satisfactory in the groundwater level forecasting for 1 month ahead. Some works have suggested that ANNs can be a promising alternative to the traditional ARMA structure; however, this study demonstrates that ARIMA model can be useful to predict the groundwater level.

References

Abghari H, Rezaeianzadeh M, Singh VP, Moradzadeh Azar F (2010) Comparison of monthly rainfall prediction using linear stochastic-base models in Gharalar rain gauge station Iran. Geophys Res Abstr 12, EGU2010-2652-1

Adhikary SK, Rahman MM, Gupta AD (2012) A stochastic modelling technique for predicting groundwater table fluctuations with time series analysis. Int J Appl Sci Eng Res 1(2):238–249

Aflatooni M, Mardaneh M (2011) Time series analysis of ground water table fluctuations due to temperature and rainfall change in Shiraz plain. Int J Water Resour Environ Eng 3(9):176–188

Agalbjörn S, Končar N, Jones AJ (1997) A note on the gamma test. Neural Comput Appl 53:131–133

Akaike H (1974) A look at the statistical model identification. IEEE Trans Autom Control AC 19:716–723

Banerjee P, Prasad RK, Singh VS (2009) Forecasting of groundwater level in hard rock region using artificial neural network. Environ Geol 58:1239–1246

Behzad M, Asghari K, Coppola EA (2010) Comparative study of SVMs and ANNs in aquifer water level prediction. J Comput Civ Eng 24(5):408–413

Bidwell VJ (2005) Realistic forecasting of groundwater level, based on the Eigenstructure of aquifer dynamics. Math Comput Simul 69:12–20

Boochabun K, Tych W, Chappell NA, Carling PA, Lorsirirat K, PaObsaeng S (2004) Statistical modelling of rainfall and river flow in Thailand. J Geol Soc India 64:503–515

Box GEP, Jenkins GM (1970) Time series analysis: forecasting and control. Holden-Day, San Francisco

Chattopadhyay S, Chattopadhyay G (2010) Univariate modeling of summer-monsoon rainfall time series: comparison between ARIMA and ARNN. CR Geosci 342:100–107

Chattopadhyay S, Jhajharia D, Chattopadhyay G (2011) Univariate modelling of monthly maximum temperature time series over northeast India: neural network versus Yule–Walker equation based approach. Meteorol Appl 18:70–82

Choubin B, Khalighi-Sigaroodi S, Malekian A, Ahmad S, Attarod P (2014) Drought forecasting in a semi-arid watershed using climate signals: a neuro-fuzzy modeling approach. J Mt Sci 11(6):1593–1605. doi:10.1007/s11629-014-3020-6

Choubin B, Khalighi-Sigaroodi S, Malekian A, Kişi Ö (2016a) Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. Hydrol Sci J 61(6):1001–1009. doi:10.1080/02626667.2014.966721

Choubin B, Malekian A, Golshan M (2016b) Application of several data-driven techniques to predict a standardized precipitation index. Atmósfera 29(2):121–128. doi:10.20937/ATM.2016.29.02.02

Choubin B, Malekian A, Samadi S, Khalighi-Sigaroodi S, Sajedi-Hosseini F (2017a) An ensemble forecast of semi-arid rainfall

using large-scale climate predictors. Meteorol Appl. doi:10. 1002/met.1635

Choubin B, Solaimani K, Habibnejad Roshan M, Malekian A (2017b) Watershed classification using remote sensing indices: a fuzzy c–means clustering approach. J Mt Sci. doi:10.1007/s11629-017-4357-4

Coulibaly P, Anctil F, Aravena R, Bobee B (2001) Artificial neural network modeling of water table depth fluctuations. Water Resour Res 37(4):885–896

Daliakopoulos IN, Coulibaly P, Tsanis IK (2005) Groundwater level forecasting using artificial neural networks. J Hydrol 309(1–4):229–240

Dash NB, Panda SN, Ramesan R, Sahoo N (2010) Hybrid neural modeling for groundwater level prediction. Neural Comput Appl 19(8):1253–1261

Dickey DA, Fuller WA (1979) Estimators for autoregressive time series with a unit root. J Am Stat Assoc 74:427–431

Durrant PJ (2001) WinGamma: a non-linear data analysis and modeling tool with applications to flood prediction. Ph.D. thesis, Department of Computer Science, Cardiff University, Wales, UK

Evans D, Jones AJ (2002) A proof of the gamma test. Proc R Soc Ser A 458:2759–2799

Han D, Yan W, Moghaddamnia A (2010) Uncertainty with the gamma test for model input data selection. In: WCCI 2010 IEEE world congress on computational intelligence. July 18–23, 2010—CCIB, Barcelona, Spain, pp 1–5

Hintze JL, Nelson RD (1998) Violin plots: a box plot-density trace synergism. Am Stat 52(2):181–184

Hsu KL, Gupta HV, Sorooshian S (1955) Artificial neural network modeling of rainfall-runoff process. Water Resour Res 31(10):2517–2530

Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70:489–501

Jajarmizadeh M, Kakaei Lafdani E, Harun S, Ahmadi A (2015) Application of SVM and SWAT models for monthly streamflow prediction, a case study in South of Iran. KSCE J Civ Eng 19(1):345–357

Kim T, Valdés J (2003) Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. J Hydrol Eng 8(6):319–328

Končar N (1997) Optimisation methodologies for direct inverse neurocontrol. Ph.D. thesis, Deptartment of Computing, Imperial College of Science, Technology and Medicine, University of London, London

Kakaei Lafdani E, Nia AM, Ahmadi A (2013) Daily suspended sediment load prediction using artificial neural networks and support vector machines. J Hydrol 478:50–62

Lee SI, Lee SK, Hamm SY (2009) A model for groundwater time-series from the well field of riverbank filtration. J Korea Water Resour Assoc 42:673–680

Lin GF, Chen LH (2005) Time series forecasting by combining the radial basis function network and the self-organizing map. Hydrol Process 19(10):1925–1937

Lu WX, Zhao Y, Chu HB, Yang LL (2013) The analysis of groundwater levels influenced by dual factors in western Jilin Province by using time series analysis method. Appl Water Sci 4(3):251–260

Moghaddamnia A, Ghafari Gousheh M, Piri J, Amin S, Han D (2009a) Evaporation estimation using artificial neural networks and adaptive neurofuzzy inference system techniques. Adv Water Resour 32(1):88–97

Moghaddamnia A, Ghafari M, Piri J, Han D (2008) Evaporation estimation using support vector machines technique. Proc World Acad Sci Eng Technol 33:14–22

Moghaddamnia A, Ghafari M, Piri J, Han D (2009b) Evaporation estimation using support vector machines. Int J Eng Phys Sci 3(3):134–142

Moghaddamnia A, Remesan R, Hassanpour Kashani M, Mohammadi M, Han D, Piri J (2009c) Comparison of LLR, MLP, Elman, NNARX and ANFIS models with a case study in solar radiation estimation. J Atmos Solar Terr Phys 71:975–982

Narayanan P, Basistha A, Sarkar S, Kamna S (2013) Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India. CR Geosci 345(1):22–27

Nayak PC, Rao YRS, Sudheer KP (2006) Groundwater level forecasting in a shallow aquifer using artificial neural network approach. Water Resour Manag 20(1):77–90

Noori R, Karbassi AR, Sabahi MS (2010) Evaluation of PCA and Gamma test techniques on ANN operation for weekly solid waste predicting. J Environ Manag 91:767–771

Nourani V, Ejlali RG, Alami MT (2011) Spatiotemporal groundwater level forecasting in coastal aquifers by hybrid artificial neural network-geostatistics model: a case study. Environ Eng Sci 28(3):217–228

Nourani V, Mogaddam AA, Nadiri AO (2008) An ANN-based model for spatiotemporal groundwater level forecasting. Hydrol Process 22(26):5054–5066

Phillips PCB, Perron P (1988) Testing for a unit root in time series regression. Biometrika 75:335–346

Piri J, Amin S, Moghaddamnia A, Keshavarz A, Han D, Remesan R (2009) Daily pan evaporation modeling in a hot and dry climate. J Hydrol Eng 14(8):803–811

Rashidi S, Vafakhah M, Kakaei Lafdani E, Javadi MR (2016) Evaluating the support vector machine for suspended sediment load forecasting based on gamma test. Arab J Geosci 9(11):583

Remesan R, Shamim MA, Han D (2008) Model data selection using Gamma test for daily solar radiation estimation. Hydrol Process 22(21):4301–4309

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6(2):461–464

Sethi RR, Kumar A, Sharma SP, Verma HC (2010) Prediction of water table depth in a hard rock basin by using artificial neural network. Int J Water Resour Environ Eng 2(4):95–102

Shamim MA, Hassan M, Ahmad S, Zeeshan M (2016) A comparison of artificial neural networks (ANN) and local linear regression (LLR) techniques for predicting monthly reservoir levels. KSCE J Civ Eng 20(2):971–977

Shirmohammadi B, Vafakhah M, Moosavi V, Moghaddamnia A (2013) Application of several data-driven techniques for predicting groundwater level. Water Resour Manag 27(2):419–432

Sigaroodi SK, Chen Q, Ebrahimi S, Nazari A, Choobin B (2014) Long-term precipitation forecast for drought relief using atmospheric circulation factors: a study on the Maharloo Basin in Iran. Hydrol Earth Syst Sci 18(5):1995–2006. doi:10.5194/hess-18-1995-2014

Singh S (2005) Implementation of the Gamma test in MATLAB using a fast near algorithm in C++. Dissertation, Department of Computer Science, Cardiff University

Stefansson A, Koncar N, Jones AJ (1997) A note on the gamma test. Neural Comput Appl 5:131–133

Trichakis IC, Nikolos IK, Karatzas GP (2009) Optimal selection of artificial neural network parameters for the prediction of a karstic aquifer's response. Hydrol Process 23(20):2956–2969

Tsanis IK, Coulibaly P, Daliakopoulos IN (2008) Improving groundwater level forecasting with a feedforward neural network and linearly regressed projected precipitation. J Hydroinf 10(4):317–330

Tsui APM, Jones AJ, De Oliveira AG (2002) The construction of smooth models using irregular embeddings determined by a gamma test analysis. Neural Comput Appl 10(4):318–329

Voudouris K (2002) Time series analysis using ARIMA models of the groundwater table in Patras industrial area aquifer system NW Peloponnese, Greece. Terra Nostra (Bonn) 3:335–340

Yang ZP, Lu WX, Long YQ, Li P (2009) Application and comparison of two prediction models for groundwater levels: a case study in Western Jilin Province, China. J Arid Environ 73:487–492

Yurekli K, Simsek H, Cemek B, Karaman S (2007) Simulating climatic variables by using stochastic approach. Build Environ 42:3493–3499

Zakaria S, Al-Ansari N, Knutsson S, Al-Badrany T (2012) ARIMA Models for weekly rainfall in the semi-arid Sinjar District at Iraq. J Earth Sci Geotech Eng 2(3):25–55