Wavelet and ANFIS Combination Model for Groundwater Level Forecasting

Qiao Yan
College of Science, Chang'an University, Xi'an 710064, Shaanxi, China

Cong Ma, Yang Song, Weibo Zhou*
College of Environmental Science and Engineering, Chang'an University, Xi'an 710054, Shaanxi, China
*Corresponding author (E-mail: 737108028@qq.com)

Abstract
The reliability and accuracy for groundwater level predicting is significant with regard to water resources management. In the current study, a wavelet transform-adaptive neuro-fuzzy inference system (WT-ANFIS) conjunction model for monthly groundwater level forecasting has been proposed. The model has been applied to forecast the groundwater level fluctuations for the two observation wells in the city of Xi'an, China. The comparison has been made on the performance of groundwater level forecasting for the proposed model (WT-ANFIS) and the performance of regular autoregressive integrated moving average (ARIMA) model and adaptive neuro-fuzzy inference system (ANFIS) model. Average temperature, maximum temperature, total precipitation and average groundwater level of every month within the period 1998-2010 had been applied to develop and validate the models. The data of the first eleven years gained had been adopted for training the applied models and the data of last two years observed for groundwater level series had been reserved for testing. The results indicate that, both during the training and testing period, the WT-ANFIS model provided more precise monthly groundwater level predictions compared to that of the ARIMA and ANFIS model.

Key words: Groundwater Level Forecasting, Wavelet Transform, Adaptive Neuro-fuzzy Inference System

1. INTRODUCTION
Groundwater serves as one of the most important water sources for industrial, municipal and agricultural supplies; however, in many parts of the world, groundwater resources have been overexploited (Konikow and Kendy, 2005). The resources are consumed at excessive rates, and various environmental and geological problems aroused, which seriously hinder the development of national economy (Adamowski and Chan, 2011). For the purpose of assisting water managers and engineers to make effective water resources management, it is necessary to develop a model of high accuracy for groundwater level forecasting.

At the present, there are many groundwater modeling approaches have been applied to forecast groundwater level fluctuations, within which; conceptual and physically based models; are the main types, for depicting hydrological variables and characterizing the complex structures of aquifers. Nevertheless, these modeling approaches do have some limitations in practice; for instance, a large number of accurate data (Mohammadi, 2008) is necessary for modeling. In case of insufficient data, accurate forecasting is more important than understanding actual dynamical behavior of the hydrological system; therefore, empirical models are suitable alternatives, because the models are available to provide accurate and reliable results without costing calibration time. Autoregressive moving average (ARIMA) model is one of empirical models with its particular properties allowing generalizations of the process being analyzed. It is a linear prediction method which assumes that the present data is a function of past data and errors (Faruk, 2010). However, the performance and accuracy of the ARIMA model are not always satisfactory, it is also not adequate to apply ARIMA model to forecast groundwater level as the climate and exploitation changes over time greatly. The variation of groundwater level is highly nonlinear because of interdependencies and uncertainties in the hydro-geological process (Suryanarayani et al., 2014). Artificial intelligence techniques have been proved to be the effective methods in virtually modeling any nonlinear function. To sum up, artificial neural networks (ANNs), adaptive neuro-fuzzy inference systems (ANFIS), support vector machine (SVM), and other artificial intelligence techniques are all widely used for predicting groundwater level at present, and all these techniques have been proved to be effective methods (Coppola et al, 2003; Daliakopoulos et al., 2005; Krishna et al., 2008; Adamowski and Chan, 2011; Chen et al., 2011). ANFIS, a model which combines artificial neural networks and fuzzy systems, is able to depict complex and uncertain modeling (such as groundwater system), and handle blur data. Studies indicate that it is available for ANFIS model to make more accurate predictions compared to the predications acquired via other conventional techniques. In addition, Kisi (2005) estimated the suspended sediments by using ANFIS and ANN method and found that ANFIS performed better than ANN model.
Furthermore, Faruk Dursun (2012) compared the ANFIS model with multiple linear regression (MLR) model and nonlinear regression (NLR) models for estimating discharge coefficient of semi-elliptical side. Moreover, Karimi et al., (2013) developed a neuro-fuzzy network technique to forecast sea level. What’s more, Khoshnevisan et al., (2014) demonstrated that ANFIS has a higher prediction accuracy than ANN model in forecasting greenhouse strawberry yield.

On the contrast, ANFIS model frequently has limitations with non-stationary time series (Cannas et al, 2006), and it may not be able to deal with non-stationary data in case, that input data preprocessing has not been performed (Tiwari and Chatterjee, 2010). Wavelet transform, used for extracting information contained in original time series, gives considerable insight into the physical form of the data. Studies show that wavelet transform performs better compared to the Fourier transform in analyzing non-stationary time series (Adamowski, 2007). Over recent years, wavelet transform method or wavelet transform combined with artificial intelligence techniques have been widely adopted for analyzing variations, periodicities, and trends in time series (Kim and Valdes, 2003; Coulibaly and Burn, 2004; Labat et al., 2005; Zhou et al, 2008; Shiri and Kisi, 2012). In this paper, wavelet transform has been resorted to decompose the time series of original data into various components so that the new time series can be used as inputs for the ANFIS model. The purpose of this research is to build up a combination model, a Wavelet and an ANFIS model to process the history monitoring data, to forecast monthly groundwater level time series data, then to compare its performance with the performance of other existing models like ARIMA and ANFIS.

2. METHODOLOGY

2.1 Wavelet Transform (WT)

The wavelet transform (WT) is a mathematical approach decomposes time series so as to provide both time and frequency information (Deka and Prahlada, 2012). It is a successful tool of capturing the features of target time series and detecting localized phenomena in non-stationary time series. WT is superior to acquire timely information simultaneously, frequency and location of a signal, while the Fourier transform (FT) can only obtain the frequency information. Up till now, WT has been widely used for many engineering purposes, including hydrological forecast, river flow forecast, fault classification and suspended sediment, etc.

The continuous wavelet transform (CWT) \( W(\tau, s) \) of a signal \( x(t) \) with respect to a mother wavelet \( \psi(t) \) is defined as follows (Partal, 2009):

\[
W(\tau, s) = \left| s \right|^{-\frac{1}{2}} \int_{-\infty}^{\infty} x(t) \psi^\ast \left( \frac{t - \tau}{s} \right) dt \quad \tau \in R, \ s \in R, \ s \neq 0
\]

where \( \tau \) is the translation parameter, \( s \) is the wavelet scale parameter, \( t \) is time and \( \ast \) refers to the complex conjugate.

However, the CWT is not often used for forecasting because it consumes a large amount of computation time and data. Instead, the discrete wavelet transform (DWT) is simpler to implement and requires less computation compared to the CWT (Adamowski and Chan, 2011). The DWT is defined as:

\[
\psi_{m,n}(t) = s_0^{-m/2} \psi \left( \frac{t - nt_0 s_0^n}{s_0} \right)
\]

where \( m \) and \( n \) are integers that control the scale and time respectively, \( t_0 \) is the location parameter that must be greater than 0, and \( s_0 \) is a specified fixed dilation step greater than 1.

The discrete wavelet transform (DWT) performs two functions viewed as high-pass and low-pass filters, through which the original time series are passed. And then the original time series are decomposed into one comprising the high frequencies (the detail) and one comprising the trend (the approximation) (Kisi and Cimen, 2011).

There are many wavelets that can be used as mother wavelets, such as Daubechies wavelet, Meyer wavelet, Morlet wavelet, and etc. (Percival and Walden, 2000). Among these wavelets, Daubechies wavelet is the most appropriate one for processing a non-stationary series and has nice performance in time series forecasting (Ramsey, 2002; Kao et al., 2013). As for Daubechies wavelet, its smoothness increases as the order of functions and the support intervals increase simultaneously, which may deteriorate the prediction (Suryanarayana et al., 2014).

2.2. Adaptive Neuro-Fuzzy Inference System (ANFIS)

Adaptive neuro-fuzzy inference system (ANFIS), first proposed by Jang (1993), is a universal approximation tool that can approximate any continuous function on a compact set to any degree of precision (Jang et al., 1997). It is a supervised model that incorporates the advantages of the learning skill of artificial neural network (ANN) and the reasoning capability of fuzzy inference system (FIS). It can employ either a back-propagation gradient method alone or combine with least-squares algorithm to estimate the parameters of membership functions and minimize output errors.
For a first-order Takagi and Sugeno (1985) fuzzy system, a typical rule set with two fuzzy If-then rules can be shown as:

Rule 1: If \( x \) is \( A_1 \) and \( y \) is \( B_1 \), then \( f_1 = p_1 x + q_1 y + r_1 \)

Rule 2: If \( x \) is \( A_2 \) and \( y \) is \( B_2 \), then \( f_2 = p_2 x + q_2 y + r_2 \)

where \( x \) and \( y \) are the input variables, \( A_i \) and \( B_i \) (\( i = 1, 2 \)) are the linguistic labels, \( p_i, q_i \) and \( r_i \) (\( i = 1, 2 \)) refer to the consequent parameters. The general structure of the ANFIS model consists of five layers (Fig. 1), and each one is briefly described subsequently.

Layer 1 (Fuzzy layer): This layer is an input layer and each node \( i \) in this layer is an adaptive node with a node function

\[ O_{1i} = u_A(x), \quad i = 1, 2 \quad (3) \]

where \( O_{1i} \) is the membership function of \( A_i \). Usually the bell-shaped function is chosen as a membership function with a maximum value of 1 and a minimum value of 0. The mathematical detail of bell-shaped function is defined as

\[ u_A(x) = \frac{1}{1 + [(x - c_i) / a_i]^2} \quad (4) \]

where \( a_i, b_i \) and \( c_i \) are the parameters, \( b_i \) is a positive value, and \( c_i \) denotes the center of the curve.

Layer 2 (Product layer): This layer consists of the nodes labeled \( \prod \) which multiply incoming signals and sending the product out. For example,

\[ O_{2i} = u_A(x) \cdot u_B(y) = \omega_i, \quad i = 1, 2 \quad (5) \]

Layer 3 (Normalized layer): In this layer, each node is a stable node and it calculates the ratio of \( i \)th rules firing strength to the sum of all rules’ firing strengths

\[ O_{3i} = \frac{\omega_i}{\omega_1 + \omega_2} = \bar{\omega}_i, \quad i = 1, 2 \quad (6) \]

Layer 4 (Consequent layer): Each node of this layer is an adaptive node, whose node function is given by

\[ O_{4i} = \bar{\omega}_i (p_i x + q_i y + r_i) = \bar{\omega}_i f_i, \quad i = 1, 2 \quad (7) \]

Layer 5 (Output layer): In this layer, the overall output is calculated by combining all incoming signals

\[ O_{5i} = \frac{\sum_{i=1}^{2} \omega_i f_i}{\sum_{i=1}^{2} \omega_i} = \sum_{i=1}^{2} \bar{\omega}_i f_i = \text{overall output} \quad (8) \]

**Figure 1.** The general architecture of ANFIS

### 2.3. Coupled Wavelet and Adaptive Neuro-Fuzzy Inference System (WT-ANFIS)

In this paper, the wavelet transform is linked to the ANFIS model for monthly groundwater level forecasting. The WT-ANFIS model consists of five steps including data preparing, original time series decomposing, training, forecasting and steps evaluating. WT method is used for decomposing step in original time series and ANFIS model is applied in the training and predicting steps. The parameters of the ANFIS are determined by trial and error method (Adamowski and Chan, 2011). As mentioned above, DWT is used in this paper because it requires less computation data and time than that for the CWT. The DWT method converts the input data into an array of constitutive series, which presents better behavior than that of the original data. The proposed method is referred to as WT-ANFIS model herein.
3. STUDY AREA AND DATA

3.1. Coupled Study Area

The model proposed has been applied to forecast groundwater level in the urban area of the city of Xi’an. The location for making this study is an arid and semi-arid region in the central of Shaanxi Province in China. The location of the study is the home to 5.847 million people. It is of semi-arid and semi-humid continental monsoon climate area with an average annual rainfall of 740.4 mm, of which 60% occurs during July to October. And the mean annual temperature is around 13.3 °C and varies between -1.3 °C in January and 26.2 °C in July.

Xi’an city is one of the forty cities of severe water shortage in China; as the result, groundwater is the main water supply for industrial and municipal utilizations. In the year of 2010, groundwater extracted has been estimated to be 94.782 million m³, which represents 60.77% of the total volume of water used in the year, and the remaining 39.23% coming from surface water. Over recent decades, as the rapid urbanization has swept the region, it results in overexploitation of the aquifer system. Together with the fast population growth, groundwater faces the increasing subsequent demand for; finally it causes, gradual decline in groundwater table.

All in all, an accurate groundwater level forecasting model is an important task for the water engineers to make sustainable groundwater resources management.

3.2. Data

The models adopted in this paper were developed by using meteorological and hydrological data. More specifically, the data used consist of average groundwater level (h in m), average temperature (T_{mean} in °C), maximum temperature (T_{max} in °C) and total precipitation (P in mm), all on the monthly basis. Monthly groundwater level recorded by the Station of Geological Environment Monitoring of Shaanxi Province is available for the period from January 1998 to December 2010 for two representative observation wells (291, K83-3) located in the areas of Yuhuazhai and Guanjiacun. The elevations of Yuhuazhai and Guanjiacun observation wells are 402.89m and 388.83m respectively. Additionally, monthly average temperature, maximum temperature and monthly total precipitation from January 1998 to December 2010 are obtained from the Meteorological Administration of Xi’an. The input data series are divided into a training set (from January 1998 to December 2008) and a testing set (from January 2009 to December 2010). Table 1 shows the statistical parameters of the used data, where X_{max}, X_{min}, X_{mean}, S_{X} and C_{V} denote the maximum, minimum, mean, standard deviation and coefficient of variation respectively.

<table>
<thead>
<tr>
<th>Data period</th>
<th>Data set</th>
<th>Statistical parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training period</td>
<td>h1 (m)</td>
<td>X_{max} 393.87, X_{min} 375.47, X_{mean} 384.21, S_{X} 5.12, C_{V} 0.013</td>
</tr>
<tr>
<td></td>
<td>h2 (m)</td>
<td>X_{max} 366.22, X_{min} 355.02, X_{mean} 360.07, S_{X} 3.34, C_{V} 0.009</td>
</tr>
<tr>
<td></td>
<td>T_{mean} °C</td>
<td>29.10, -1.70, 15.10, 9.26, 0.613</td>
</tr>
<tr>
<td></td>
<td>T_{max} °C</td>
<td>42.90, 8.90, 27.71, 9.40, 0.339</td>
</tr>
<tr>
<td></td>
<td>P (mm)</td>
<td>254.40, 0, 47.48, 47.85, 1.008</td>
</tr>
<tr>
<td>Testing period</td>
<td>h1 (m)</td>
<td>X_{max} 388.24, X_{min} 385.64, X_{mean} 387.53, S_{X} 0.67, C_{V} 0.002</td>
</tr>
<tr>
<td></td>
<td>h2 (m)</td>
<td>X_{max} 362.05, X_{min} 359.31, X_{mean} 360.96, S_{X} 0.82, C_{V} 0.002</td>
</tr>
<tr>
<td></td>
<td>T_{mean} °C</td>
<td>28.20, 0.60, 15.15, 9.49, 0.626</td>
</tr>
<tr>
<td></td>
<td>T_{max} °C</td>
<td>39.80, 12.90, 28.59, 8.52, 0.298</td>
</tr>
<tr>
<td></td>
<td>P (mm)</td>
<td>176.00, 0, 48.53, 43.79, 0.902</td>
</tr>
</tbody>
</table>

h1 and h2 denote the groundwater level for 291 and K83-3 observation wells respectively.

4. CRITERIA FOR EVALUATING MODEL PERFORMANCE

For a assessing the applied models and the abilities of making accurate predictions, the root mean squared error (RMSE), the coefficient of efficiency (CE) and the correlation coefficient (R^2) are used. The RMSE is given by

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (h_{\text{est}}(t) - h_{\text{obs}}(t))^2}
\]

where \( h_{\text{est}}(t) \) is the estimated value at time \( t \), \( h_{\text{obs}}(t) \) is the observed value at time \( t \), and \( n \) is the number of time steps. RMSE describes the average magnitude of the errors between the observed values and the calculated results. The lower values of the RMSE, the more precise the prediction is.

The CE is given by
\[ CE = 1 - \frac{\sum_{j=1}^{n} (h_{obs}(t) - \bar{h}_{obs})^2}{\sum_{j=1}^{n} (h_{obs}(t) - \bar{h}_{obs})^2} \] (10)

where \( \bar{h}_{obs} \) is the average value of observed groundwater level. CE ranges from negative infinity to 1; the higher value means the better performance of the model.

The expression for the \( R^2 \) is

\[ R^2 = \left[ \frac{\sum_{j=1}^{n} (h_{obs}(t) - \bar{h}_{obs})(h_{est}(t) - \bar{h}_{est})}{\sqrt{\sum_{j=1}^{n} (h_{obs}(t) - \bar{h}_{obs})^2} \sqrt{\sum_{j=1}^{n} (h_{est}(t) - \bar{h}_{est})^2}} \right]^2 \] (11)

where \( \bar{h}_{est} \) is the average value of estimated groundwater level. The best fit between observed and estimated values would have \( R^2=1 \).

5. MODEL DEVELOPMENT

5.1. ARIMA Modeling

SPSS 19 has been used to develop the ARIMA models for predicting monthly groundwater level for the observation wells in this paper. Since ARIMA model is a univariate time series analysis technique, only one variable that is monthly average groundwater level in this study, can be used. According to the Box-Jenkins methodology (1991), a differencing approach could be used to stable the original time series, and an autocorrelation function (ACF) and a partial autocorrelation function (PACF) could be utilized to decide which (if any) autoregressive or moving average component should be included in the ARIMA model. The groundwater level time series for the two observation wells had been found to be not stationary; therefore, the groundwater level time series are transformed into stationary values through the differencing process. The parameter estimation has been performed by using the generalized least square method.

The best empirical ARIMA model of 291 groundwater level is ARIMA(0,1,0)(1,0,0)12, while the best ARIMA model of K83-3 groundwater level is ARIMA(0,1,0)(0,0,1)12. All the parameter estimators for both ARIMA models are significant.

The ARIMA(0,1,0)(1,0,0)12 model for groundwater level data from 291 observation well can be expressed as follows:

\[ (1 - \Phi B^{12})(1 - B)h(t) = a \] (12)

where \( h(t) \) is the measured groundwater level time series at time \( t \); \( \Phi \) is autoregressive parameter; \( B \) is the backward shift operator \( (\Phi(k)h(t)=h(t-k)) \); and that is the Gaussian white noise. The parameter \( \Phi \) has been estimated to be 0.406 by SPSS software for 291 observation well. Eq. (12) can also be written as

\[ h(t) = h(t-1) + 0.406h(t-12) - 0.406h(t-13) + a, \] (13)

The ARIMA(0,1,0)(0,0,1)12 model for groundwater level data from K83-3 observation well can be expressed as follows:

\[ (1 - B)h(t) = (1 + 0.237B^{12})a, \] (14)

Eq. (14) can also be written as

\[ h(t) = h(t-1) + 0.237a_{t+12} \] (15)

5.2. ANFIS Modeling

The ANFIS models for monthly groundwater level forecasting for both observation wells were developed by using the MATLAB R2010 software program (MATLAB 2010). MATLAB’s ANFIS editor provides many types of membership functions (MFs). Yet, recent studies indicate that the type of MF does not affect the forecasting accuracy in significant way (Vernieuwe et al., 2005). In this Paper, the widely applied bell-shaped function has been chosen for the membership function expressed in Eq. (4). The well-known hybrid learning method has been selected and the training epoch number was set to 85.

Additionally, the input variables consist of various combinations of the following physical variables: average groundwater level, average temperature, the maximum temperature and the total precipitation all on the monthly basis. Different combinations of the input time series could exert influence on the prediction accuracy for groundwater level forecasting. Eq. (13) indicates that the groundwater level at time \( t-j \) \( (j=1,12,13) \) have a strong influence on the groundwater level measured at time \( t \) for 291 observation well. Therefore, \( h(t-1) \), \( h(t-12) \) and \( h(t-13) \) were selected as inputs for 291 observation well. Likewise, the groundwater level at time \( t-1 \) was
selected as input for K83-3 observation well. Various combinations of the remaining variables; average temperature, the maximum temperature and the total precipitation on the monthly basis 1 month before, 2 months before, 3 months before and 4 months before had been tested. The optimal number of member functions (MFs) was identified via using a trial and error procedure (Jain et al, 2001). The criteria for evaluating model performance described earlier could be utilized to find the best ANFIS model. The structure and input variables of the best ANFIS models for 291 and K83-3 observation wells are summarized in Table 2.

<table>
<thead>
<tr>
<th>Well</th>
<th>Number of MFs</th>
<th>Input variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>291</td>
<td>4</td>
<td>(h(t-1), h(t-12), h(t-13), T_{\text{mean}}(t-1), T_{\text{mean}}(t-2), T_{\text{max}}(t-1), P(t-1), P(t-2), P(t-3))</td>
</tr>
<tr>
<td>K83-3</td>
<td>5</td>
<td>(h(t-1), T_{\text{mean}}(t-1), T_{\text{mean}}(t-2), T_{\text{max}}(t-1), P(t-1), P(t-2))</td>
</tr>
</tbody>
</table>

In Table 2, the numbers of MFs of best ANFIS models for 291 and K83-3 observation wells are determined to be 4 and 5, respectively. As for 291 observation well, the best ANFIS model is a function of the average groundwater level 1 month before, 12 months before and 13 months before; the average temperature 1 month before and 2 months before; the maximum temperature 1 month before; and the total precipitation 1 month before, 2 months before and 3 months before. For K83-3 observation well, the best ANFIS model is a function of the average groundwater level 1 month before; the average temperature 1 month before and 2 months before; the maximum temperature 1 month before and 2 months before; and the total precipitation 1 month before and 2 months before.

5.3. WT-ANFIS Modeling

The coupled wavelet-adaptive neuro-fuzzy inference system models have been developed by combining the two methods, namely, WT and ANFIS. The WT-ANFIS is an ANFIS model of sub-series components obtained by applying WT to pre-process original data. MATLAB Wavelet toolbox has been used in this paper to decompose the input data (average groundwater level for the observation wells, average temperature, the maximum temperature and the total precipitation all on the monthly basis) into a certain number of sub-series components (Ds and As denoting details and approximations of the time series, respectively). Fig. 2 gives the input and output structure of the proposed WT-ANFIS model. All sub-series decomposed by DWT approach were used as inputs to the ANFIS model since an averaging or optimizing selection of only certain sub-series would have been a diminutive method. That is, all sub-series coefficients are important and contain useful information about the original time series (Adamowski and Sun, 2010; Suryanarayana et al, 2014; Mohammadi et al, 2015).

![Flow chart of the WT-ANFIS model](image)

As shown in Fig. 2, four resolution levels (2-4-8-16) are employed in this paper. The input data of monthly \(h, T_{\text{mean}}, T_{\text{max}}, P\) in WT-ANFIS models are decomposed into time series of 2-month mode, 4-month mode, 8-month mode, 16-month mode and one approximation mode by using DWT algorithm. Fig. 3 shows the approximation and details components of the groundwater level time series for 291 observation well using Daubechies-2 (db2) wavelet at level 4. In the Fig.3, \(a_t\) represents the original groundwater level series of 291 observation well, \(d_k\) is the low frequency data of the time series, and \(d_1, d_2, d_3,\) and \(d_4\) are the four levels decomposed series in sequence when the data continuously pass four times through the high pass filters, the
time series of 2-month mode, 4-month mode, 8-month mode, 16-month mode, respectively. Other original time series are decomposed and the coefficients obtained from these decompositions are imported as inputs for ANFIS model.

Figure 3. Decomposition of groundwater level series for 291 observation well using db2 mother wavelet at level 4

For the WT-ANFIS models, bell shaped function has been chosen for the membership function. The input data of the WT-ANFIS models consist of various combinations of the following variables: the Ds and As of average groundwater level, average temperature, the maximum temperature and the total precipitation on the monthly basis. The predicted groundwater level is the summation of the outputs from each decomposed series, the comparing the observed groundwater level, by using the trial and error approach, the optimum number of MFs has been determined based on different combinations of variables for both observation wells.

5.4. Results and Discussion

Figs. 4-6 show the comparison of the observed and forecasted groundwater levels for the observation wells during the training period by the best ARIMA, ANFIS and WT-ANFIS models, respectively. It is obvious that the simulated results by the three models used in this paper agree well with the observed groundwater level. That is, the best ARIMA, ANFIS and WT-ANFIS models have a good fitting accuracy in the process of developing monthly groundwater level forecasting model. Further, it can be seen that the best WT-ANFIS models to compute the groundwater levels with less scatter and are closer to the fit line equation \(y=x\) compared to those of other models.

(a) 291 observation well          (b) K83-3 observation well

Figure 4. The observed and forecasted groundwater level by the best ARIMA model during the training period for the observation wells
Figure 5. The observed and forecasted groundwater level by the best ANFIS model during the training period for the observation wells.

Figure 6. The observed and forecasted groundwater level by the best WT-ANFIS model during the training period for the observation wells.

The values of RMSE, CE and $R^2$ of the best ARIMA model, ANFIS model and WT-ANFIS models for the 291 and K83-3 observation wells during the training period are summarized in Tables 3 and 4 respectively. According to the evaluation criteria values in Tables 3 and 4, it is obvious that the best WT-ANFIS model has smaller RMSE and larger CE and $R^2$ comparing with that of the best ARIMA model and ANFIS model for both the observation wells. In other words, the best WT-ANFIS model is more competent in predicting monthly groundwater level as compared to the best ARIMA model and ANFIS model.

Table 3. Comparison of the best WT-ANFIS model with other models for monthly groundwater level forecasting for 291 observation well during the training period

<table>
<thead>
<tr>
<th>Performance criterion</th>
<th>Model</th>
<th>For best ARIMA</th>
<th>For best ANFIS</th>
<th>For best WT-ANFIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.399</td>
<td>0.330</td>
<td>0.231</td>
<td></td>
</tr>
<tr>
<td>CE</td>
<td>0.994</td>
<td>0.996</td>
<td>0.998</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.996</td>
<td>0.997</td>
<td>0.998</td>
<td></td>
</tr>
</tbody>
</table>

Higher fitting accuracy in the training sets does not mean a higher forecasting accuracy in the testing sets (Yang, et al, 2009). Hence, it is necessary to validate the models considered in this study. The monthly groundwater level from January 2009 to December 2010, which has not been used to establish the models, was used to evaluate the forecast accuracy.
Table 4. Comparison of the best WT-ANFIS model with other models for monthly groundwater level forecasting for K83-3 observation well during the training period

<table>
<thead>
<tr>
<th>Performance criterion</th>
<th>For best ARIMA</th>
<th>For best ANFIS</th>
<th>For best WT-ANFIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.057</td>
<td>0.047</td>
<td>0.041</td>
</tr>
<tr>
<td>CE</td>
<td>0.997</td>
<td>0.998</td>
<td>0.999</td>
</tr>
<tr>
<td>R²</td>
<td>0.997</td>
<td>0.998</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Figure 7 and Fig.8 give the comparison of the observed and simulated groundwater level by using the best models of ARIMA, ANFIS and WT-ANFIS during the testing period for 291 and K83-3 observation wells. Although at some data points, the best WT-ANFIS model gives worse predictions than either predictions of the best ARIMA or ANFIS model forecasts, its overall forecasting capability has been improved.

The performance of the best ARIMA model, ANFIS model and WT-ANFIS model for the 291 and K83-3 observation wells during the testing period are given in Table 5 and Table 6. It can be seen from the tables that the best WT-ANFIS model produces better performance than that of the best ARIMA model and the best ANFIS model. In comparison with the best ANFIS model, the RMSE values for the best WT-ANFIS model are lower by 30% for 291 observation well and 21% for K83-3 observation well. Similarly, the best WT-ANFIS model when compared to best ARIMA model shows a decrease in RMSE values of 42% for 291 observation well and 47% for K83-3 observation well. The best WT-ANFIS model gives an increase in CE and R² values over the best ANFIS model respectively by 272% and 51% for 291 observation well and 32% and 30% for K83-3 observation well. The best WT-ANFIS model when compared to the best ARIMA model, shows an increase in CE and R² values of 1140% and 394% for 291 observation well, and 927% and 40% for K83-3 observation well respectively. The CE and R² values of the best WT-ANFIS model are far superior to the best ARIMA model for both observation well as shown in Table 5 and Table 6.
In conclusion, the best ARIMA, ANFIS and WT-ANFIS models can be successfully used to develop the forecasting models which are available to provide reliable and accurate monthly groundwater level predictions. The results indicate that the best performance has been obtained with the best WT-ANFIS model, followed by the best ANFIS model and finally the best ARIMA model. ANFIS is a powerful fuzzy logic neural network and in this paper some factors that influence groundwater level fluctuations have been considered and put in the model, whereas ARIMA is a univariate time series analysis tool and only groundwater level data were used. Hence, ANFIS has a higher precision than the ARIMA model. The WT-ANFIS model is more accurate because discrete wavelet transform decomposes original time series into component series that carry useful information and the wavelet transformed data improves the ability of ANFIS model by capturing useful information on various resolution levels.

5. CONCLUSIONS

In this paper, a new method based on coupling of WT and ANFIS has been developed for forecasting monthly groundwater level at two observation wells in the city of Xi’an, China. To study the accuracy of WT-ANFIS model, regular ARIMA model and ANFIS model have been developed. The ANFIS and WT-ANFIS models are trained by using temperature, groundwater level and rainfall data observed from 1998 to 2008, and then all the data were applied to the test period extending from 2009 to 2010. The best ANFIS and WT-ANFIS models have been obtained by comparing the performances of the models with various input variables. The parameters of the best ARIMA model are determined from the autocorrelation coefficients and partial autocorrelation coefficients.

Results of this study suggest that the best ARIMA, ANFIS and WT-ANFIS models can be applied to forecast reliable and accurate groundwater level. However, comparing the performance of the best ARIMA, ANFIS and WT-ANFIS models, it indicates that RMSE values of the best WT-ANFIS models are lower than those of the best ARIMA and ANFIS models whereas values of CE and $R^2$ for the best WT-ANFIS models are higher than those of the best ARIMA and ANFIS models for both the observation wells during the training and testing period. Thus, it can safely draw a conclusion that the best WT-ANFIS model has better performance in comparison with other models considered in this paper.

This study demonstrates the strong performance of the WT-ANFIS model in monthly groundwater level forecasting. Future studies are necessary to explore the use of WT-ANFIS model in groundwater level prediction for other geographical regions.

Acknowledgements

The authors gratefully acknowledge the Station of Geological Environment Monitoring of Shaanxi Province, China, for providing the groundwater level data. This work was financially supported by the Natural Science Foundation of Shaanxi Province (2014JM1030); the Center of Geological Survey of Xi’an, China (12120113004800) and the Fundamental Research Funds for the Central Universities (310812161011).

REFERENCES


MATLAB. (2010). The MathWorks Inc: Naick, MA, USA.


