MODELING OF GROUNDWATER LEVEL USING ARTIFICIAL INTELLIGENCE TECHNIQUES: A CASE STUDY OF REYHANLI REGION IN TURKEY

DEMIRCI, M.* – ÜNEŞ, F. – KÖRLÜ, S.

Hydraulics Division, Civil Engineering Department, Iskenderun Technical University
31040 Iskenderun, Hatay, Turkey

*Corresponding author
e-mail: mustafa.demirci@iste.edu.tr; phone: +90-326-613-5600 (2810)

Abstract. Determination of the change in groundwater level in terms of planning and managing resources is important. In this study, the groundwater level of Reyhanlı region in Turkey was predicted using multi-linear regression (MLR), adaptive neural fuzzy inference system (ANFIS), Radial basis neural network (RBNN), support vector machines with radial basis functions (SVM-RBF) and support vector machines with poly kernels (SVM-PK) methods. Models were carried out using 192 data of monthly ground water level, monthly total precipitation and monthly average temperature values measured for 16 years between 2000 and 2015. Comparisons revealed that the SVM-RBF and SVM-PK models had the most accuracy in the groundwater level prediction.

Keywords: groundwater level prediction, multi-linear regression, support vector machines, adaptive neural fuzzy inference system, radial basis neural network

Introduction

Groundwater level (GWL) drops due to excessive water withdrawal are serious threats to the sustainability of drinking water basins. Due to global warming and the growing population, the increase in agricultural activities negatively affects the groundwater especially in rural areas. In Turkey, many basins are experiencing problems in terms of the groundwater potential. Increasing water demand, adverse conditions created by climate change and lack of planning related to underground water management in the basin have increased these problems. There are a number of studies that have examined groundwater level reductions due to various reasons. In the field of hydrology, models of groundwater level require knowledge or estimation of the hydrologic parameters of the basin. Prediction of groundwater level is important for effective planning and sustainable groundwater management.

In recent years, artificial intelligence approaches have been widely used in water resource management and hydrological projects (e.g., Kaya et al., 2011, 2016; Mohanty et al., 2013; Li et al., 2012; Shiri et al., 2013; Demirci and Baltaci, 2013; Üneş et al., 2015a, b, 2017; Demirci et al., 2015a, b; Tasar et al., 2017, 2018; Feng and Tian, 2018; Yılmaz et al., 2018; Kankan et al., 2018; Nacar et al., 2018; Wunsch et al., 2018).

Many investigations have been also conducted to predict groundwater level fluctuations with using artificial intelligence techniques. Emamgholizadeh et al. (2014) studied the potential of artificial neural network (ANN) and adaptive neural fuzzy inference (ANFIS) for the groundwater levels (GWL) predictions. Heesung et al. (2011) studied two nonlinear time-series models for predicting the groundwater level (GWL) fluctuations using ANNs and SVM. Nourani et al. (2015) used feed-forward neural network (FFNN), Auto Regressive Moving Average (ARIMAX) models for GWL forecasting of the plain of...
Ardabil, Northwestern Iran. Guzman et al. (2015) used Artificial Neural Networks (ANN) and Support Vector Regression (SVR) in daily groundwater levels predictions. According to their results ANN and SVR have good accuracy to predict daily groundwater levels. Yoon et al. (2016) used a weighted error function approach to improve the performance of artificial neural network (ANN)- and support vector machine (SVM)-based recursive prediction models for the long-term prediction of groundwater levels in response to rainfall. Zhou et al. (2017) proposed a data-driven prediction model combining discrete wavelet transform (DWT) preprocess and support vector machine (SVM) for groundwater depth forecasting. Ebrahimi and Rajaee (2017) investigated the effect of wavelet analysis of ANN, MLR and SVR approaches using GWL data which recorded over in the Qom plain, Iran. Zare and Koch (2018) used new hybrid Wavelet-ANFIS model with several combinations of inputs and mother wavelets to simulate and predict GWL-fluctuations in the Miandarband plain, Iran. The results showed that all model approaches could be used with acceptable accuracy. Other authors also found that artificial neural network in general are exceptionally well suited to predict groundwater levels for different geological settings (Coulibaly et al., 2001; Nayak et al., 2006; Mohammadi, 2008; Sreekanth et al., 2009; Adamowski and Chan, 2011; Trichakis et al., 2011; Chitsazan et al., 2015; Khalil et al., 2015; Chang et al., 2016; Bargezar et al., 2017; Yu et al., 2018).

The main purpose of the present study is to analyze the performance of the MLR, ANFIS, RBNN, SVM-RBF and SVM-PK techniques in monthly groundwater level forecasting using the Monthly Mean Precipitation (MP), Monthly Average Temperature (MT), Monthly Ground Water Level (GWL+1) data. The predictions of the models were compared with those of observed values using te statistical analysis. The monthly groundwater level data measured from DSI (Turkish General of State Hydraulic Works) between 2000 and 2015 and the monthly total precipitation and monthly average temperature data recorded by Antakya Meteorological Station were used to determine groundwater level.

Material and methods

Study area

In this study, groundwater level of Reyhanlı region on Amik plain is studied (Fig. 1). Amik plain is the most efficient plain of Turkey in terms of productivity. It is covered with thick alluvial soil layer and is known as the most efficient basin in terms of agricultural potential. It forms the largest flat land of Hatay Province. Amik Plain nourishes its soil primarily with Asi River, Afrin and Karasu Stream. The plain which locates in the basin of Asi has an area of approximately 65 000 ha. The groundwater level in Amik Plain reduces by approximately 5-6 m each year. Due to drying of Amik Lake, global climate changes and unconscious irrigation, there is a very serious water shortage in the plain which lead to several problems especially for farmers. For these reasons, using existing water resources more effectively has become inevitable.

The DSI observation-well with number 15768 used in the present study is in Reyhanlı region and located at 36.2685 latitude and 36.5676 longitude. It has a depth of 50 m. The monthly groundwater level data taken from DSI (Turkish General of State Hydraulic Works) data and the monthly total precipitation and monthly average temperature data recorded by Antakya Meteorological Station were used to determine groundwater level. Models were carried out using 192 data of monthly ground water level, monthly total precipitation and monthly average temperature values measured for 16 years.
The monthly groundwater level fluctuation between the years 2000 and 2015 is shown in Figure 2. In the Reyhanlı region, generally, high precipitations occur in early and end of spring. However, drought is observed in the summer months. It is observed that consequent groundwater level fluctuations are entirely dependent on meteorological conditions, which affect the drainage basin. It can be seen from this figure that the vertical axis of Figure 2 shows well gauge readings located in Amik region and the level fluctuations range in the order of 27 m.
In all of the models, Monthly Mean Precipitation (MP), Monthly Average Temperature (MT), Monthly Groundwater Level (GWL+1) were used for estimating of the Ground Water Level fluctuations. The models are first trained by using the first set of data, namely training data comprising the first 144 data of 192 months’ observations. Once the training stage was completed, the models were applied to the testing data which comprise the last remaining 48 months’ observations. For applied each model, the mean square error (MSE), mean absolute error (MAE), and the determination coefficient between the model estimations and the observed values are computed. The results are used to compare the performance between model prediction and observed well data. The MSE and MAE are given as in Equations 1 and 2:

\[
MSE = \frac{1}{N} \left( \sum_{i=1}^{N} (Y_{\text{observed}} - Y_{\text{forecast}})^2 \right)
\]

(Eq.1)

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |Y_{\text{observed}} - Y_{\text{forecast}}|
\]

(Eq.2)

where N and Yi denote the number of data sets, groundwater level data respectively. Different number of input combinations were tried for the best performance of using models were compared with each other.

**Multiple-linear regression (MLR)**

This method is used to find out how much a dependent variable is affected and the value of the independent variables in which it is affected and associated.

In multiple regression analysis, dependent variable y indicates the relationship between independent variables x₁, x₂, ..., xₚ and it can be written as in Equation 3:

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ...... + \beta_j x_j + ... + \beta_p x_p + \varepsilon
\]

(Eq.3)

Here, β₀, β₁, β₂, β₃, ..., βₚ are called the regression coefficients. Any βₗ regression coefficient gives the expected change in the y variable versus a unit change in the xₗ variable when other variables are kept constant (when the effect of the other variables is eliminated).

**Adaptive neural fuzzy inference system (ANFIS)**

Fuzzy rules are expressed in form of if-then statements between system inputs and outputs depending on the linguistic variables. ANFIS was first represented by Jang (1993). ANFIS works by a learning algorithm derived from neural network functional rules. The parameters of the fuzzy inference system are determined by the neural network learning algorithms in a fuzzy rule-based system. Fuzzy Inferences System can be classified into two types; Mamdani’s system and Sugeno’s system. ANFIS techniques are based on the Sugeno’s system. ANFIS with Sugeno type works according to “If-Then” rule and the ANFIS structure uses the Sugeno-Fuzzy rules.

The Sugeno NF system, generated by two rules using three inputs, is shown in Figure 3a. Where, w₁ or w₂ is obtained by weighted mean of individual rule outputs. NF
structure is shown in Figure 3b. NF is connected via directional links and contains several nodes. Every node has a node function that can be constant or adjustable parameters. In the figures, $A_1$, $B_1$, $C_1$ and $A_2$, $B_2$, $C_2$ are linguistic labels (such as “low”, “medium” or “high”), $f_1$ and $f_2$ denote, respectively, output functions of 1. and 2. rule, $\{p_i, q_i, r_i, s\}$ specified as result parameters. Sugeno’s system is more compact and computationally more efficient (Jang, 1993). For more information, it can be looked at Jang (1993).

Figure 3. a Sugeno type fuzzy model. b Neuro-fuzzy (NF) system structure for three inputs. In this presented fuzzy model, input data were taken as monthly mean precipitation (MP), monthly average temperature (MT), monthly ground water level (GWL+1), and ground water level (GWL) were estimated as output data.
Radial basis neural network (RBNN)

Radial basis neural network was first presented into the ANN literature by Broomhead and Lowe (1988). The RBNN models consist of two layers whose output nodes form a linear combination of the basis functions. The learning scheme of RBNN is fundamentally different from that of the feed-forward ANN. The RBNN assumes a radially symmetric function, typically the Gaussian function, for its hidden units. The relation between inputs and outputs is demonstrated in Figure 4. Detailed information about RBNN theory can be found in the works of Haykin (1998).

In Figure 4, the net input to the radbas transfer function is the vector distance between its weight vector \( w_i \) and the input vector \( p_i \), multiplied by the bias \( b \). The \( \| \text{dist} \| \) box in this figure accepts the input vector \( p \) and the single row input weight matrix, and produces the dot product of the two. The radial basis function has a maximum of 1 when its input is 0. As the distance between \( w \) and \( p \) decreases, the output increases. Thus, a radial basis neuron acts as a detector that produces 1 whenever the input \( p \) is identical to its weight vector \( w \). The transfer function for a radial basis neuron is given as Equation 4:

\[
\text{radbas}(n) = e^{-\|p\|^2}
\]  

(Eq.4)

Support vector machines (SVM)

The SVM has become a relatively novel and promising estimator in data-driven research fields, of which basic concept and theory have been introduced by Vapnik (1998). The generalization ability of the SVM is considered to be better than ANN, in the sense that it is based on the structural risk minimization rather than the empirical risk minimization of ANN. The main process of SVM model building consists of selecting support vectors which support the model structure and determining their weights. Figure 5 shows the SVM models schematic representations. The idea behind SVM is to find a hyperplane that separates two classes in the transformed feature (input) space with a maximum distance. SVM aims to find the optimal regression hyperplane, that all training samples lay within an \( \varepsilon \)-margin around it and is also as flat as possible (Schölkopf and Smola, 2002). SVM models are closely related to artificial neural
networks and use a sigmoid kernel function while having a two-layer, forward-feed artificial neural network (Haykin, 1998).

![Diagram of SVM](image)

**Figure 5. Schematic representations of SVM**

An interesting feature of SVM is that by minimizing the average error rate on the data set from the empirical risk minimization principle derived, statistical learning can be achieved.

SVM defines how to draw this boundary between variable groups. SVM studies according to statistical learning theory. A set of training data \([p_1, y_1], (p_2, y_2), \ldots, (p_n, y_n)\], where “\(p_i\)” value indicates the input space of the sample and has a corresponding target “\(y_i\)” value.

The SVM function is expressed as in *Equation 5*:

\[
y = (K_{pi}W_{jk}) + b
\]

(Eq.5)

where the Kernel function is \(K_{pi}\), \(b\) is bias term of SVM network and \(W_{jk}\) is called as the Lagrange multipliers that obtain the significance of the training data sets for the output data. The kernel function of non-linear radial basis (Hsu et al., 2003) is in *Equation 6*:

\[
K_{xi} = e^{-\gamma \| x_i - y_i \|^2}
\]

\(\gamma > 0\) and \(i = 1, 2, 3, \ldots n\)  

(Eq.6)

where \(\gamma\) is a user-defined parameter. The kernel function of polynomial (Hsu et al., 2003) is in *Equation 7*:

\[
K_{xi} = (p_i y + c)^d
\]

\(i = 1, 2, 3, \ldots n\)  

(Eq.7)

SVM differs from the other classification methods significantly. Its intent is to create an optimal separating hyperplane between two classes to minimize the generalization error and thereby maximize the margin. SVM is an approximate implementation of structural risk minimization approach. Structural risk minimization method described that the error rate of learning machine on test data is bounded by the sum of training error rate and a term that based on Vapnik–Chervonenkis dimension (Haykin, 1998).
Results and discussion

In order to demonstrate the models capability for predicting groundwater level variations, this section will discuss the validation process of the optimal MLR, RBNN, RBNN, SVM-PK and SVM-RBF models. The correlation coefficient (R), mean square error (MSE) and absolute mean error (MAE) for the performance evaluation of all models are calculated. Comparison parameters of MSE, MAE and R obtained from testing data are shown in Table 1. Results are used to compare the performance of model prediction and the observation data.

In Figure 6a, results of MLR model shows that the correlation coefficient is high and the groundwater level estimate is closer to the actual values. MLR models performed less efficiently than the other models based on MSE, MAE and R (5.44-1.73-0.84) criteria shown in Table 1.

For RBNN model, the distribution and scatter graphs are shown in Figure 6b. When scatter graphs for testing data are analyzed, RBNN model results also showed a good performance. RBNN estimated values are close to the actual values. Correlation coefficient was obtained as R = 0.88. According to the MSE, MAE and R (4.64, 1.62, 0.88) criteria, compared to MLR model, the RBNN model results showed better performance.

Figure 6. a Measurement and MLR model scatter and distribution. b Measurement and RBNN model scatter and distribution
16 years data were evaluated for ANFIS model, and the results are defined as follows. In ANFIS models, distribution and scatter graphs for testing data are shown in Figure 7a. Correlation coefficient were obtained as R = 0.87. When distribution and scatter charts are analyzed, estimated values are close to the actual values and ANFIS results showed similar results according to RBNN results. When we looked MSE, MAE and R (4.82-1.63-0.87) criteria values shown in Table 1, The ANFIS estimated values gave better results than the MLR values.

Figure 7. a Measurement and ANFIS model scatter and distribution. b Measurement and SVM-PK model scatter and distribution. c Measurement and SVM-RBF model scatter and distribution
Table 1. Model results according to statistical criteria

<table>
<thead>
<tr>
<th>MODELS</th>
<th>INPUTS</th>
<th>MSE</th>
<th>MAE</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLR</td>
<td>MP, MT, GWL(L+1)</td>
<td>5.44</td>
<td>1.73</td>
<td>0.84</td>
</tr>
<tr>
<td>RBNN</td>
<td>MP, MT, GWL(L+1)</td>
<td>4.64</td>
<td>1.62</td>
<td>0.88</td>
</tr>
<tr>
<td>ANFIS</td>
<td>MP, MT, GWL(L+1)</td>
<td>4.82</td>
<td>1.63</td>
<td>0.87</td>
</tr>
<tr>
<td>SVM-PK</td>
<td>MP, MT, GWL(L+1)</td>
<td>3.81</td>
<td>1.41</td>
<td>0.90</td>
</tr>
<tr>
<td>SVM-RBF</td>
<td>MP, MT, GWL(L+1)</td>
<td>3.78</td>
<td>1.42</td>
<td>0.90</td>
</tr>
</tbody>
</table>

MSE: mean square error, MAE: mean absolute error, R: correlation coefficient, MP: monthly mean precipitation, MT: monthly average temperature, GWL+1: monthly groundwater level

The SVM Poly Kernels (SVM-PK) and SVM Radial Bases Function (SVM-RBF) models were also applied to estimate the groundwater level. Estimated testing results for SVM models are shown in Figure 7b and c. The SVM Poly Kernels (SVM-PK) and SVM Radial Basis Function (SVM-RBF) estimated values gave closer results according to the RBNN values. The correlation coefficient was obtained for both model to be R = 0.90 as seen in Figure 7b and c. SVM model results data are close to the actual values shown in figures. As seen from the scatter plots, the SVM model estimates are less scattered in relation to the other models.

According to MSE, MAE and R, the MLR (5.44-1.73-0.84) model has the lowest success rate. ANFIS approach (4.82-1.63-0.87) and RBNN approach (4.64-1.62-0.88) have the higher success rate than the MLR model. SVM-PK (3.81-1.41-0.90) and SVM-RBF (3.78-1.42-0.890) models were found to perform better than the other models at all performance evaluations.

Conclusions

In this study, groundwater level fluctuations of Reyhanlı region in Turkey were predicted with using MLR, ANFIS, RBNN, SVM-RBF and SVM-PK approaches.

The accuracy of all models in groundwater level estimation was also investigated, and the results were compared with each other model. Comparisons revealed that the SVM -RBF and SVM-PK had the closer accuracy in the groundwater level prediction. In general, MLR model has the smaller lowest success rate. ANFIS and RBNN models gave better results than the MLR model. According to the MSE, MAE and R criteria, the best results were obtained in SVM-RBF and SVM-PK models. The results of the case study are satisfactory and demonstrate that SVM-RBF and SVM-PK models can be a useful prediction tool in the area of groundwater hydrology.

It is recommended for future studies that authors should study the use of new methods of artificial intelligence techniques in daily, weekly, or yearly groundwater level forecasting for different geographical regions.

REFERENCES


