Sadat-Noori, Mahmood, Glamore, William, and Khojasteh, Danial (2020) ***Groundwater level prediction using genetic programming: the importance of precipitation data and weather station location on model accuracy.*** **Environmental Earth Sciences, 79 (1) .**

Groundwater level prediction using Genetic Programming: The importance of precipitation data and weather station location on model accuracy

Mahmood Sadat-Noori*, William Glamore, Danial Khojasteh

Water Research Laboratory, School of Civil and Environmental Engineering, UNSW Sydney, NSW, Australia

*Corresponding author:

M. Sadat-Noori, Water Research Laboratory, School of Civil & Environmental Engineering, University of New South Wales, Sydney, 110 King St., Manly Vale, NSW, 2093, Australia

Email: m.sadat-noori@unsw.edu.au

Abstract

Groundwater (GW) level prediction is important for effective GW resources management. It is hypothesized that using precipitation data in GW level modelling will increase the overall accuracy of the results and that the distance of the observation well to the weather station (where precipitation data are obtained) will affect the model outcome. Here, Genetic Programming (GP) was used to predict GW level fluctuation in multiple observation wells under three scenarios to test these hypotheses. In Scenario 1, GW level and precipitation data were used as input data. Scenarios 2 only had GW level data as inputs to the model and in Scenarios 3, only precipitation data were used as inputs. Long term GW level time series data covering a period of eight years were used to train and test the GP model. Further, to examine the effect of data from previous time periods on the accuracy of GW level prediction, 12 models with input data up to 12 months prior to the current period, were investigated. Model performance was evaluated using two criteria, Coefficient of Determination (R^2) and Root Mean Square Error (RMSE). Results show that when predicting GW levels through GP, using GW level and precipitation data together (Scenario 1) produces results with higher accuracy compared to only using GW level (Scenario 2) or precipitation data (Scenario 3). Additionally, it was found that model accuracy was highest for the well located closest to the weather station (where precipitation data were collected), demonstrating the importance of weather station location in GW level prediction. It was also found that using data from up to six previous time periods (months) can be the most efficient combination of data for accurate predictions. The findings from this study are useful for increasing prediction accuracy of GW level variations in unconfined aquifers for sustainable GW resources management.

Keywords: Unconfined aquifer, coefficient of determination, root mean square error, Tabriz plain.

1. Introduction

Groundwater (GW) is known as the largest liquid freshwater resource on earth and stores almost 90% of the total non-frozen freshwater worldwide [1]. It is reported that 50% of global megacities are dependent on GW for supplying potable water [1]. Additionally, excessive exploitation of GW is leading to swift depletion of aquifers, posing threats to the sustainability of food/water production worldwide. [2]. In addition, GW acts as a natural storage of water protected from surface evaporation, is distributed spatially and can be utilized with limited capital expenditure [3]. GW is also a more promising and reliable source of fresh water, in comparison to surface water, during droughts that regularly affect the quantity of surface water resources globally [4].

Increasing environmental pressures, such as agriculture development, urbanisation and climate change, can intensify GW resources stresses [5]. In such conditions, the rate of GW recharge may become lower than the withdrawal rate, resulting in environmental repercussions, GW storage decline, degradation of water quality, and increasing extraction costs [6]. Despite the considerable importance of GW resource worldwide, studies and management strategies on GW resources are often less available than surface waters resources, as GW data can be challenging to collect and time consuming [5,7].

To manage GW resources effectively and sustainably, accurate GW level determination is required as poor management may lead to water quality deterioration, declining GW levels, and decreasing aquifer storage [5, 8]. For example, GW level prediction has significant importance in the management of seawater intrusion on fresh GW resources in coastal regions [9], and in developing effective irrigation schemes to prevent GW contamination in locations with agricultural activities [10, 11]. Additionally, determining sustainable GW extraction policies, facilitating environmental protection, and developing water price policies are all dependent on accurate, efficient and reliable forecasting of GW level variations [8].

In the past decade, Artificial Intelligence (AI) and machine learning techniques have received increasing interest in the water resources literature due to their high accuracy and low computational efforts compared to conventional modelling techniques such as regression, statistical, probabilistic parametric, semiparametric and nonparametric models [12]. Models such as Adaptive Neural Fuzzy Inference System (ANFIS) and Genetic Programming (GP) have been described as effective tools to predict GW level elevation [13-16].

It is hypothesized that using precipitation data in GW level modelling will increase the overall accuracy of the results and that the distance of the observation well to the weather station (where precipitation data are obtained) will affect the model outcome. This paper provides a review of the current knowledge of using GP to predict GW levels and develops a GP model to predict GW level fluctuation in multiple observation wells under three scenarios to test these hypotheses.

A GP based model was used to forecast GW level variations near the Amaravathi River, India [17]. It was found that the model could precisely capture the non-linearity nature of GW level fluctuations without requiring explicit knowledge of the physical characteristics of the system. Shiri and Kisi [14] examined the ability of Gene Expression Programming (GEP - a multi-branch GP with the ability of creating expression trees) and ANFIS data driven models to predict GW level time series (one-, two- and three-day forecasts) and found that GEP performed slightly better than ANFIS based on error criteria.

In another investigation, the same authors analyzed the ability of Artificial Neural Network (ANN), GEP, and ANFIS techniques to forecast daily GW levels for 1 to 7-days ahead, using several input combinations, such as GW levels, rainfall and evapotranspiration data in South Korea

[18]. The study was based on GW levels from a single well located 500 meters away from the weather station. Shiri and Kisi [18] reported that the GEP method can be satisfactorily applied to predict GW level fluctuations up to 7 days beyond data records using the mentioned data as inputs.

In two separate studies by Fallah-Mehdipour et al. [19, 20], ANFIS and GP methods were used to obtain governing GW flow equations in Ghaen and Karaj aquifers in Iran, using various recharges and discharges situations as input data sets. They found that GW level predictions are more accurate when using GP compared to ANFIS. Additionally, GP was used for GW budget forecasting by Gorgij et al. [6] to predict a 0.12 m reduction in GW levels for the Azarshahr plain aquifers, Iran, by validating the accuracy of the GP model and using current period GW levels as input data. In another study, Sivapragasam et al. [21] used monthly GW level data as input and GP was applied to predict spatial variations of GW levels in Arjuna Nadhi region, India. They reported that in forecasting GW level for a specific well, information from neighboring wells should incorporate GW level predictions as it significantly improves the prediction accuracy. Amaranto et al. [22] provided GW level prediction for an unconfined aquifer in the northern high plains of Nebraska, USA, by employing different inputs such as crop water demand, ice melting, GW level, precipitation and evapotranspiration. Their study demonstrated that ANN and GP can produce similar prediction results depending on input data.

Overall, when using AI models to predict GW levels, obtaining the highest accuracy in predictions is the main goal. Based on our literature review (Table 1), it was found that although precipitation data is the most widely used parameter (after GW level data) in GW level prediction, the effects of the weather station distance to the observation well on model accuracy has yet to be assessed. Therefore, the main objective of this study is to build on the literature by (1) assessing whether precipitation data is a significant input for GW level prediction under GP application and

(2) to determine how the location of the weather station (where precipitation data are collected) may affect the accuracy of the GW level predictions.

2. Material and Methods

2. 1. Study site

Tabriz plain is located in North-west Iran and accommodates Iran's fifth most densely populated city. The plain has an area of approximately 700 km^2 and is located between latitudes 45°30′ and 46°15′ N, and altitudes 37°56′ and 38°17′ E [23,24]. It is bordered to the north by the Mishow, Moro and west Garadug mountains; to the east by Tabriz city; to the south by the northern slopes of Sahand Mountain; and to the west by Urmia lake [23]. The area experiences a cold climate in winter, mild in spring, semi-hot in summer, and a mild rainy weather in autumn. Its elevation ranges between 1,350 to 1,600 m above sea level. Tabriz plain is classified as arid with a cold and dry climate and has a mean annual precipitation of 280-290 mm (63% lower than world's average rainfall, 800 mm) and a mean annual temperature of 12.6 °C [25].

The geology of the plain is comprised of a 120 m deep alluvial layer that sits atop an impermeable bedrock with the alluvial layer containing sand, gravel, silt and clay material. The significant formations of this area include an upper red formation (in the north-eastern part), volcanic tuffs (in the southern part), and quaternary deposits (in the western part) [26]. Three rivers including the Aji-Chay, Gomanab-Chay and Sinekh-Chay cross this area and transport suspended sediments and saline water during the high discharge and low discharge periods, respectively [24].

The area has a shallow unconfined and a deeper confined aquifer system separated by a low permeable clay-silt layer. The central part of the plain comprises unconfined and confined aquifers, while the highlands solely contain unconfined aquifers [24]. The unconfined aquifer contains saline water up to 60 m below the surface in some parts, mainly due to the saline water of Aji Chay River in the region, which discharges into Urmia Lake (saline lake), with an average annual discharge rate of 10 m^3/s. Fresh GW can be sourced from 60-120 m depths in the unconfined aquifer. The direction of the GW flow generally follows the topography of the region and is mainly from northeast to southwest [27]. The source of recharge for the confined and semi-confined aquifer is precipitation (in winter and autumn) localized to small areas, while the unconfined aquifer recharges from the river and irrigation return water.

Tabriz city has a large drinking water demand (roughly 4500 litres per second in 2017), with GW extraction supplying 33% of demand and the Zarineh-rood river supplying the remaining drinking water demand. However, GW levels are potentially at risk having decreased by 5 m (on average) over the last 30 years in parts of the plain [28]. Figure 1 shows the location of the study site, observation wells and location of the weather station.

2. 2. Genetic Programming

GP is an artificial intelligence model that detects data patterns and approximate functions to best define relationships between inputs and outputs [17]. The primary advantage of GP is the flexibility to simulate a complex phenomenon using mathematical and logical relations with significantly lower computational costs in comparison to conventional methods [19]. This is the main reason GP is increasingly being used in water resources engineering problems [29].

In a traditional tree-based structure of GP, various parameters, operators and functions are placed in the nodes that are connected by several branches. Each tree node consists of two different sets: 1) a function set, and 2) a terminal set. Functions are nodes with children arguments and contain mathematical functions (e.g., square, sin, tan), arithmetic operators (+, -, /), boolean operators (e.g., and, or), and other user-defined expressions. Terminals include numerical constants and variables [20]. Two tree structures of a GP are shown in Figure 2(a) which can be interpreted as $(7-X_1)*(X_2+5)$ (left) and $sqrt(Y_1)+(X_1/8)$ (right) where X and Y represent random variables.

The algorithm initializes by randomly selecting a combination of functions and terminals to form a population of equations that are represented by a tree. Each tree (potential solution) is evaluated through an evolutionary process called fitness. The fitness function is considered an error criterion between the actual and predicted output [16]. Based on the values of fitness function of each tree, selection techniques ranking method are applied to determine trees that can survive in the next generation, while trees that have the least fit with the data are discarded [19]. The elements of these selected trees are then combined to create the next generation of algorithm with some of the characteristics of each parent. In order to serve this purpose, two genetic operators are employed that mimic the natural world reproduction system: crossover and mutation. In the crossover process, two trees are randomly selected, and two or more branches of those trees are randomly swapped (Figure 2(b)). In the mutation process the functions, operators and variables in the nodes are randomly chosen and exchanged (Figure 2(c)) [20]. This evolution process is repeated over successive generations until a termination condition (e.g. a user-defined threshold error) is satisfied.

2. 3. Groundwater level modeling - Input data

In this study, three observation wells (A, B and C) were chosen based on their relative distance to the weather station where precipitation data were obtained. Well A was the furthest (43 km) from the weather station while Well C was the closest (3 km) from the weather station (Figure 1) and Well B was located in-between, at 18 km from the weather station. Monthly GW level data for an eight-year (96-months) period was used to predict monthly GW level variations within each well. Average monthly precipitation data for the same eight-year time period were obtained from the Iranian Meteorological Organization and used as input data. From the total data set, seven years were used for model training and a one-year period was used for testing. A detailed statistical description of the data is provided in Table 2, indicating that there is more skewness in the rainfall data compared to GW level data. Based on the statistics presented in Table 2, the data are assumed to be stationary, meaning its probability distribution does not change when shifted in time or space [18]. Here, we opted not to normalise the input data as distance between the data was considered important. However, data normalisation should not affect the performance of the neural networks tested [30].

In the prediction process for each time period (t), which was one month, inputs to the model were GW level (h) and precipitation (P) for current (t) and previous (t-1, t-2…) time periods. GW level prediction was carried out under three different scenarios to investigate the importance of both GW level and precipitation data as inputs for modelling. The three applied scenarios are described below:

I) Predicting GW level (h) using the 96-month GW level time series data from two previous time periods (h_{t-1} and h_{t-2}), plus precipitation data for current (P_t) and two previous time periods (P_{t-1} and P_{t-2}) (a total of five inputs). In other words, GW level in the current time period (h_t) was assumed to be a function of GW levels from two previous time periods and precipitation during the current and two previous time periods (with time periods being on a monthly scale).

II) Predicting GW level using only GW level data from two previous time periods (h_{t-1} and h_{t-2}) without including precipitation data.

III) Predicting GW level using only precipitation data from the current (P_t) and two previous time periods (P_{t-1} and P_{t-2}). This scenario has a total of three inputs with no GW level data as input.

The following equations show each scenario in mathematical terms:

$$\text{Scenario 1: } h_t = f[h(_{t-1}) + h(_{t-2}) + P_t + P(_{t-1}) + P(_{t-2})] \quad (1)$$

$$\text{Scenario 2: } h_t = f[h(_{t-1}) + h(_{t-2})] \quad (2)$$

$$\text{Scenario 3: } h_t = f[P(_{t-1}) + P(_{t-2})] \quad (3)$$

where h_t is the predicated GW level for each well in the period t; f is the prediction function for each well using the corresponding data set; $h(_{t-1})$ and $h(_{t-2})$ are the GW level in the t-1 and t-2 time periods; P_t is the precipitation in the current time period (t); and $P(_{t-1})$ and $P(_{t-2})$ are precipitation in the t-1 and t-2 time periods, respectively, with time periods being months. The number of lags for the data was chosen based on the Partial Auto-Correlation Function (PACF) of monthly GW levels. As shown in Figure 3, the GW levels from the first two lags have a significant effect on h_t.

Table 3 shows the parameters used in setting up the GP model. The cross over and mutation parameter values were obtained through extensive trails of different combination sets [16]. Generally, after the model parameters are defined complex equations are formed. The results from these equations are evaluated using a fitness function (mean square error) and subsequently the model performance is evaluated based on the fitness criteria. Models with acceptable performance are maintained through cross over and mutation processes. This process continued until the defined number of generations (1000) was reached or there was no improvement after 300 generations (Table 3) [17]. In future studies, an uncertainty analysis of the input data (inherent errors associated with the data) may help improve the confidence of the developed model [31].

2. 4. Performance measures

Two statistical evaluation criteria were used to assess the model performance: Coefficient of determination (R^2); which is defined as the proportion of the alteration in the dependent variable that is predictable from the independent variable, and Root Mean Square Error (RMSE). The Coefficent of determination varies between 0 to 1, with higher values (close to unity) indicating that the predictions fit the data [32,33], following Equation (4). RMSE always has a non-negative value and values closer to zero are representative of a perfect relationship between observed and estimated values. This was calculated using Equation (5).

$$R^2 = \frac{\left[\sum_{i=1}^{t}(h_o - \bar{h}_o)(h_e - \bar{h}_e)\right]^2}{\sqrt{\sum_{i=1}^{t}(h_o - \bar{h}_o)^2 \times \sum_{i=1}^{n}(h_e - \bar{h}_e)^2}} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{t}\sum_{i=1}^{t}(h_e - h_o)^2} \quad (5)$$

where t is the number of time periods, h_o and h_e are observed and estimated values at the ith time period, and $\bar{h}_o$ and $\bar{h}_e$ are the mean of the observed and estimated values, respectively [34].

3. Results and Discussion

3. 1. Is precipitation a significant input for GW level prediction using GP?

Table 4 shows the model outcome for Well A (furthest from the weather station) based on the statistical error criteria. Results show that the highest level of fitting between observed and predicted data was through Scenario 1, which had the lowest root mean square error and the highest correlation coefficient (R^2) value for training and test data (Table 4). Figure 4 shows the observed and predicted values of GW level for both training and test data under Scenario 1. Scenario 2 produced an R^2 of 0.87 for training data and 0.78 for test data and ranked second, whereas Scenario

3 ranked last with the lowest R^2 for the training and test data (Table 4). The results indicated that including both GW level and precipitation data (Scenario 1) produces better results compared to only using GW level or precipitation data (Scenarios 2 and 3), however, not including precipitation data (Scenario 2) does not significantly affect the prediction accuracy (Table 4).

To further assess the effect of GW level data as an input to the model, in comparison with precipitation data, and to investigate the effect of the distance of the weather station on the accuracy of GW level prediction, the two additional Wells (B and C) located closer to the weather station were analysed. The GW level prediction results for wells B and C are shown in Table 5. It was found that the best results for R^2 and RMSE were under Scenario 1, when using GW level and precipitation data as inputs to the model. A comparison between the three wells showed that Well C, which was located closest to the weather station (3 km), produced the best fit of data among the three wells with the highest R^2 values for both training and test data under all scenarios. Well A, located furthest from the weather station produced the least accurate results, demonstrating the importance of the weather station location on the model outcome.

Overall, it was found that in predicting GW levels using GP in unconfined aquifers with sufficient infiltration to recharge the aquifer; (1) the effect of not including precipitation data on the results, is much lower than not including GW level data; and (2) if precipitation data are used, the most accurate predictions will be obtained for wells nearest to the weather station, where precipitation data are collected. However, if precipitation data are not available or not included, using only GW level data will produce reasonable results. However, not including GW level data in the modeling process will produce the least accurate results. Here, the observation wells from the unconfined aquifer were used and only GW level and precipitation data were considered in the modelling [17]. However, other parameters such as temperature, water abstraction, river flow can also affect the GW level in Tabriz plain, as it is a complex aquifer system. A comprehensive data set on the recharge and discharge components of the aquifer may increase modelling accuracy further.

3. 2. The effect of preceding data on GW level prediction

Based on the literature, the maximum previous time periods used for GW level prediction under GP application was two months (using GW level data from two months prior to the current time period, t) (Table 1). Therefore, to investigate the effect of preceding data on the accuracy of GW level prediction, 12 prediction models using GW level data from up to 12 previous time periods were constructed. Analysis were performed on Well C (nearest well to the weather station) as the most accurate predictions were obtained for this well. The modeling procedure was such that the level of the GW in the current time period (t) was a function of the GW level from previous time periods (t-1, t-2, . . ., t-12). Equations 6 to 17 show the various functions used:

$$\mathrm{h}(t) = f(\mathrm{h}_{t-1}) \tag{6}$$

$$\mathrm{h}(t) = f(\mathrm{h}_{t-i}) \quad i = 1, 2 \tag{7}$$

$$\mathrm{h}(t) = f(\mathrm{h}_{t-i}) \quad i = 1, 2, 3 \tag{8}$$

$$\mathrm{h}(t) = f(\mathrm{h}_{t-i}) \quad i = 1, 2, 3, 4 \tag{9}$$

$$\mathrm{h}(t) = f(\mathrm{h}_{t-i}) \quad i = 1, 2, 3, 4, 5 \tag{10}$$

$$\mathrm{h}(t) = f(\mathrm{h}_{t-i}) \quad i = 1, 2, 3, 4, 5 ,6 \tag{11}$$

$$\mathrm{h}(t) = f(\mathrm{h}_{t-i}) \quad i = 1, 2, 3, 4, 5 ,6, 7 \tag{12}$$

$$\mathrm{h}(t) = f(\mathrm{h}_{t-i}) \quad i = 1, 2, 3, 4, 5 ,6, 7, 8 \tag{13}$$

$$\mathrm{h}(t) = f(\mathrm{h}_{t-i}) \quad i = 1, 2, 3, 4, 5 ,6, 7, 8, 9 \tag{14}$$

$$\mathrm{h}(t) = f(\mathrm{h}_{t-i}) \quad i = 1, 2, 3, 4, 5 ,6, 7, 8, 9, 10 \tag{15}$$

$$h(t) = f(h_{t-i}) \quad i = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 \tag{16}$$

$$h(t) = f(h_{t-i}) \quad i = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 \tag{17}$$

where h is GW level, and t is the current time period for which predictions are made, while t-1 to t-12 indicate GW levels at previous time periods (in reference to the current time period). Based on the two model performance criteria (R^2 and RMSE), the results indicated that for the training data, the best R^2 was obtained when using data from up to 3 previous time periods (months). In contrast, using data from one previous time period (month) had the lowest RMSE (Table 6). For test data, the highest R^2 was observed when using data from up to 6 months, which also corresponded to the lowest RMSE. It was observed that by including more data beyond 6 months prior to the currents period (up to 12 months) in the modelling process, R^2 values decreased, indicating the model accuracy can decline as older data are used. Overall, results suggested that when predicting monthly GW levels, using data from up to 6 months prior to the current time period, may produce the most accurate results based on the highest R^2 and lowest RMSE value.

4. Conclusion

Accurate groundwater level prediction is a crucial factor for sustainable GW resources management worldwide. Using data-driven models such as artificial intelligence techniques to accurately predict GW level in aquifers can provide a robust tool for decision makers to monitor, manage and protect GW resources. In this study, genetic programming modelling techniques were used to predict GW levels under various scenarios and prediction accuracy was assessed based on statistical error criteria. Modelling results indicated that GW level predictions are most accurate for wells closest to the meteorological weather station when precipitation data are included as inputs. This illustrates the important role of weather station location on the modelling outcome accuracy in GP. In addition, when only using GW level data the prediction accuracy was maximized by including data from up to six prior time periods.

Acknowledgment

The authors would like to thank the University of New South Wales and the University of Tehran for providing funding and resources for this work.

Data Availability

All data used in the study are available from the authors upon request.

References

[1] Bricker, S. H., Banks, V. J., Galik, G., Tapete, D., & Jones, R. (2017). Accounting for groundwater in future city visions. Land Use Policy, 69, 618-630.

[2] Dalin, C., Wada, Y., Kastner, T., & Puma, M. J. (2017). Groundwater depletion embedded in international food trade. Nature, 543(7647), 700.

[3] Wheater, H. S., Mathias, S. A., & Li, X. (Eds.). (2010). Groundwater modelling in arid and semi-arid areas. Cambridge University Press.

[4] Wada, Y. (2016). Modeling groundwater depletion at regional and global scales: Present state and future prospects. Surveys in Geophysics, 37(2), 419-451.

[5] Famiglietti, J. S. (2014). The global groundwater crisis. Nature Climate Change, 4(11), 945.

[6] Gorgij, A. D., Kisi, O., & Moghaddam, A. A. (2017). Groundwater budget forecasting, using hybrid wavelet-ANN-GP modelling: a case study of Azarshahr Plain, East Azerbaijan, Iran. Hydrology Research, 48(2), 455-467.

[7] Ziolkowska J. R., Reyes R. (2017). Groundwater Level Changes Due to Extreme Weather—An Evaluation Tool for Sustainable Water Management. Water, 9, 117; doi:10.3390/w9020117.

[8] Adamowski, J., & Chan, H. F. (2011). A wavelet neural network conjunction model for groundwater level forecasting. Journal of Hydrology, 407(1-4), 28-40.

[9] Krishna, B., Satyaji Rao, Y. R., & Vijaya, T. (2008). Modelling groundwater levels in an urban coastal aquifer using artificial neural networks. Hydrological Processes: An International Journal, 22(8), 1180-1188.

[10] Yadav, B., Ch, S., Mathur, S., & Adamowski, J. (2017). Assessing the suitability of extreme learning machines (ELM) for groundwater level prediction. Journal of Water and Land Development, 32(1), 103-112.

[11] Djurovic, N., Domazet, M., Stricevic, R., Pocuca, V., Spalevic, V., Pivic, R., Gregoric, E., & Domazet, U. (2015). Comparison of groundwater level models based on artificial neural networks and ANFIS. The Scientific World Journal, 2015:742138.

[12] Goetz, J. N., Brenning, A., Petschko, H., & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Computers & geosciences, 81, 1-11.

[13] Taormina, R., Chau, K. W., & Sethi, R. (2012). Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. Engineering Applications of Artificial Intelligence, 25(8), 1670-1676.

[14] Shiri, J., & Kişi, Ö. (2011). Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. Computers & Geosciences, 37(10), 1692-1701.

[15] Yu, H., Wen, X., Feng, Q., Deo, R. C., Si, J., & Wu, M. (2018). Comparative study of hybrid-wavelet artificial intelligence models for monthly groundwater depth forecasting in extreme arid regions, Northwest China. Water Resources Management, 32(1), 301-323.

[16] Cobaner, M., Babayigit, B., & Dogan, A. (2016). Estimation of groundwater levels with surface observations via genetic programming. Journal of the American Water Works Association, 108(6), E335-E348.

[17] Kasiviswanathan, K. S., Saravanan, S., Balamurugan, M., & Saravanan, K. (2016). Genetic programming based monthly groundwater level forecast models with uncertainty quantification. Modeling Earth Systems and Environment, 2(1), 27.

[18] Shiri, J., Kisi, O., Yoon, H., Lee, K. K., & Nazemi, A. H. (2013). Predicting groundwater level fluctuations with meteorological effect implications—A comparative study among soft computing techniques. Computers & Geosciences, 56, 32-44.

[19] Fallah-Mehdipour, E., Haddad, O. B., & Marino, M. A. (2014). Genetic programming in groundwater modeling. Journal of Hydrologic Engineering, 19(12), 04014031.

[20] Fallah-Mehdipour, E., Haddad, O. B., & Mariño, M. A. (2013). Prediction and simulation of monthly groundwater levels by genetic programming. Journal of Hydro-environment Research, 7(4), 253-260.

[21] Sivapragasam, C., Kannabiran, K., Karthik, G., & Raja, S. (2015). Assessing suitability of GP modeling for groundwater level. Aquatic Procedia, 4, 693-699.

[22] Amaranto, A., Munoz-Arriola, F., Corzo, G., Solomatine, D. P., & Meyer, G. (1753). Semi-seasonal groundwater forecast using multiple data-driven models in an irrigated cropland. Journal of Hydroinformatics, 20(6), 1227-1246.

[23] Barzegar, R., Moghaddam, A. A., & Kazemian, N. (2015). Assessment of heavy metals concentrations with emphasis on arsenic in the Tabriz plain aquifers, Iran. Environmental earth sciences, 74(1), 297-313.

[24] Barzegar, R., Moghaddam, A. A., & Tziritis, E. (2017). Hydrogeochemical features of groundwater resources in Tabriz plain, northwest of Iran. Applied Water Science, 7(7), 3997-4011.

[25] Zarghami, M., & Akbariyeh, S. (2012). System dynamics modeling for complex urban water systems: Application to the city of Tabriz, Iran. Resources, Conservation and Recycling, 60, 99-106.

[26] Barzegar, R., Moghaddam, A. A., Najib, M., Kazemian, N., & Adamowski, J. (2016). Characterization of hydrogeologic properties of the Tabriz plain multilayer aquifer system, NW Iran. Arabian journal of Geosciences, 9(2), 147.

[27] Barzegar, R., Moghaddam, A. A., & Baghban, H. (2016). A supervised committee machine artificial intelligent for improving DRASTIC method to assess groundwater contamination risk: a case study from Tabriz plain aquifer, Iran. Stochastic Environmental Research and Risk Assessment, 30(3), 883-899.

[28] Karami, F. (2011). Evaluation of Meteorological Drought Effects in the Reduction of Ground Water table (Case study: Tabriz Plain). Journal of Geography and Planning, 37, 111-131. (In Persian)

[29] Mehr, A. D., Nourani, V., Kahya, E., Hrnjica, B., Sattar, A. M., & Yaseen, Z. M. (2018). Genetic programming in water resources engineering: A state-of-the-art review. Journal of Hydrology, 566, 643-667.

[30] Shanker, M., Hu, M. Y., & Hung, M. S. (1996). Effect of data standardization on neural network training. Omega, 24(4), 385-397.

[31] Shrestha, D. L., & Solomatine, D. P. (2006). Machine learning approaches for estimation of prediction interval for the model output. Neural Networks, 19(2), 225-235.

[32] Khojasteh, D., Bordbar, A., Kamali, R., & Marengo, M. (2017). Curvature effect on droplet impacting onto hydrophobic and superhydrophobic spheres. International Journal of Computational Fluid Dynamics, 31(6-8), 310-323.

[33] Bordbar, A., Taassob, A., Khojasteh, D., Marengo, M., & Kamali, R. (2018). Maximum Spreading and Rebound of a Droplet Impacting onto a Spherical Surface at low Weber numbers. Langmuir, 34(17), 5149-5158.

[34] Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377(1-2), 80-91.

Table 1. A comparison of previous studies in the literature using GP to predict groundwater level using various data as model inputs.

Location	**Method**	**Input data**				**Time scale**	**R^2 (maximum value based on test data)**	**RMSE (minimum value based on test data)**	**Time period**	**Authors**
		Groundwater level	Precipitation	Evaporation	Additional					
Bondville & Perry wells, Illinois State, US	GEP		✓			Daily	0.99	0.10	Up to 1 week	Shiri & Kişi [14]
Karaj plain, Iran	GP	✓	✓	✓		Monthly	0.81	0.33	1 month	Fallah-Mehdipour et al. [20]
Hongcehon well, Korea	GEP	✓	✓	✓		Daily		0.06	Up to 1 week	Shiri et al. [18]
Ghaen & Karaj aquifers, Iran	GP	✓			Aquifer recharge & discharge rates	Monthly	0.90	0.15	1 month	Fallah-Mehdipour et al. [19]
Arjuna Nadhi, India	GP	✓				Monthly	0.72		1 month	Sivapragasam et al. [21]
Amarawathi basin, India	GP	✓	✓			Monthly		1.23	2 months for groundwater level & 14 months for Precipitation	Kasiviswanathan et al. [17]
North Central Florida, US	MGGP		✓	✓	Surface water level	Monthly	0.90		1 month	Cobaner et al. [16]
Azarshahr plain, Iran	GP	✓				Monthly	0.97	0.07	1 month	Gorgij et al. [6]
Nebraska, US	GP	✓	✓		Crop water demand, snowmelt, evapotranspiration	Monthly		0.10	1 month	Amaranto et al. [22]
Tabriz plain, Iran	GP	✓	✓		Distance to weather station	Monthly	0.98	0.09	Up to 12 months	Current study

409 **Table 2.** Statistical parameters of the input dataset.

Data period	Data set	Statistical parameters						
		Observation	Avg.	Min.	Max.	Std Dev.	Skewness	Coefficient of variation
Training period	Groundwater level (h_t) - (m)	84	1286.0	1285.0	1286.9	0.48	0.31	0.0004
	Precipitation (P_t) - (mm)	84	19.7	0.0	114.80	21.55	2.05	1.09
Testing period	Groundwater level (h_t) - (m)	12	1286.1	1285.8	1286.5	0.22	0.22	0.0002
	Precipitation (P_t) - (mm)	12	20.9	0.3	68.0	22.96	1.23	1.10
Whole period	Groundwater level (h_t) - (m)	96	1286.0	1285.0	1286.9	0.45	0.40	0.0004
	Precipitation (P_t) - (mm)	96	19.9	0.0	114.8	21.60	1.91	1.09

410

Table 3. Parameters used in setting up the GP model.

Parameter	Value
Population size	128
Generation	1000
Crossover rate	0.93
Mutation rate	0.65
Fitness function	Mean square error
Termination	300 Generations without improvement

Table 4. Model outcomes for training and test data for Well (A) under Scenarios 1, 2, and 3.

	Scenario	Train		Test	
		R^2	RMSE	R^2	RMSE
Well (A)	Scenario 1	0.91	0.112	0.85	0.171
	Scenario 2	0.87	0.158	0.78	0.244
	Scenario. 3	0.45	0.994	0.36	0.362

Table 5. Model outcomes for training and test data for Wells B and C under Scenarios 1, 2, and 3.

	Scenario	**Train**		**Test**	
		R^2	RMSE	R^2	RMSE
Well (B)	Scenario 1	0.92	0.262	0.89	0.228
	Scenario 2	0.89	0.238	0.82	0.258
	Scenario 3	0.54	0.501	0.37	0.591
Well (C)	Scenario 1	0.94	0.202	0.92	0.232
	Scenario 2	0.92	0.230	0.89	0.347
	Scenario 3	0.63	0.459	0.47	0.572

Table 6. Statistical performance metrics for train and test data for the 12-time periods.

Time period (month)	Train		Test	
	R^2	RMSE	R^2	RMSE
1	0.900	0.083	0.826	0.264
2	0.906	0.223	0.940	0.173
3	0.950	0.173	0.910	0.223
4	0.939	0.201	0.881	0.223
5	0.945	0.141	0.934	0.141
6	0.919	0.201	0.980	0.101
7	0.865	0.223	0.885	0.387
8	0.834	0.141	0.827	0.141
9	0.825	0.173	0.788	0.173
10	0.848	0.141	0.808	0.201
11	0.865	0.101	0.817	0.189
12	0.847	0.223	0.836	0.101

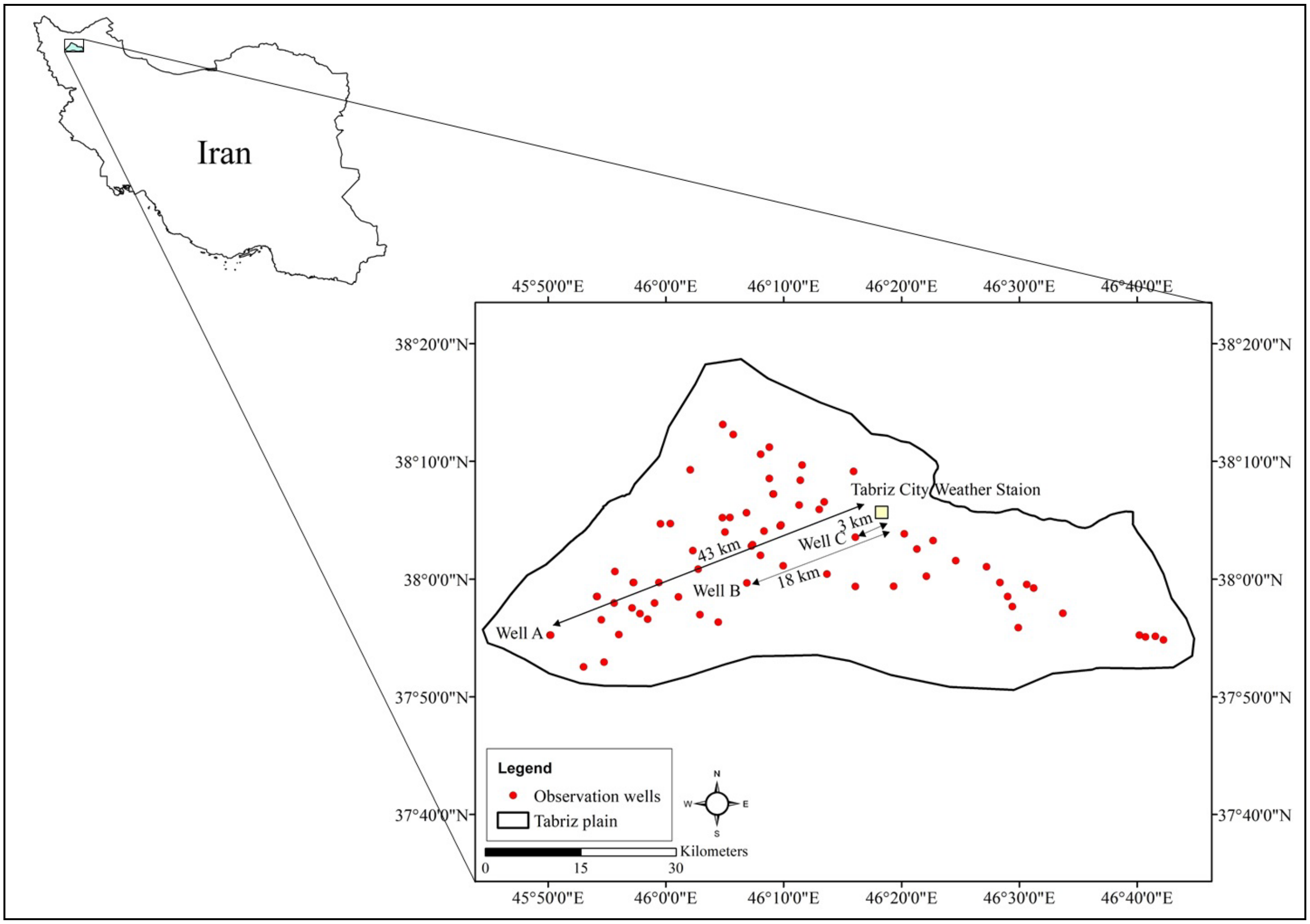

Figure 1. The location of the study site, Well A, B and C and the meteorgogical weather sation location.

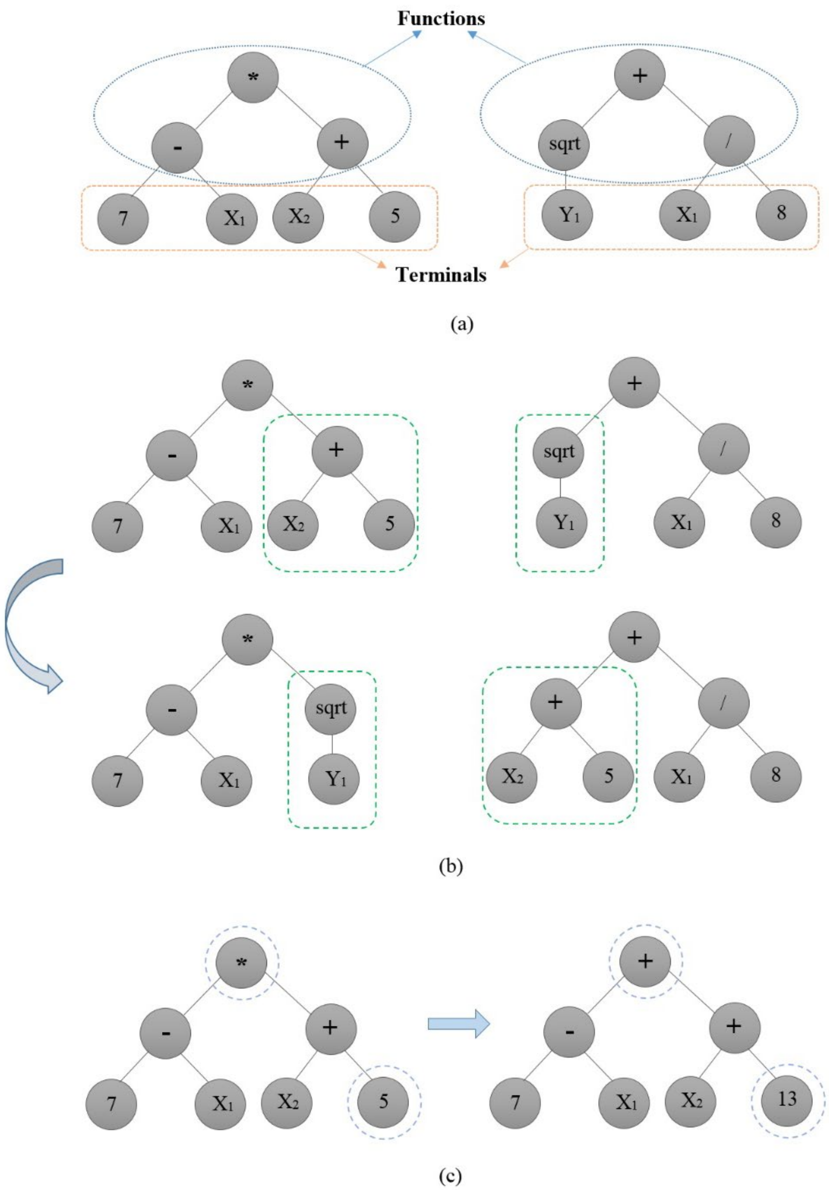

Figure 2. GP structures: (a) tree structure, (b) cross over process, (c) mutation process.

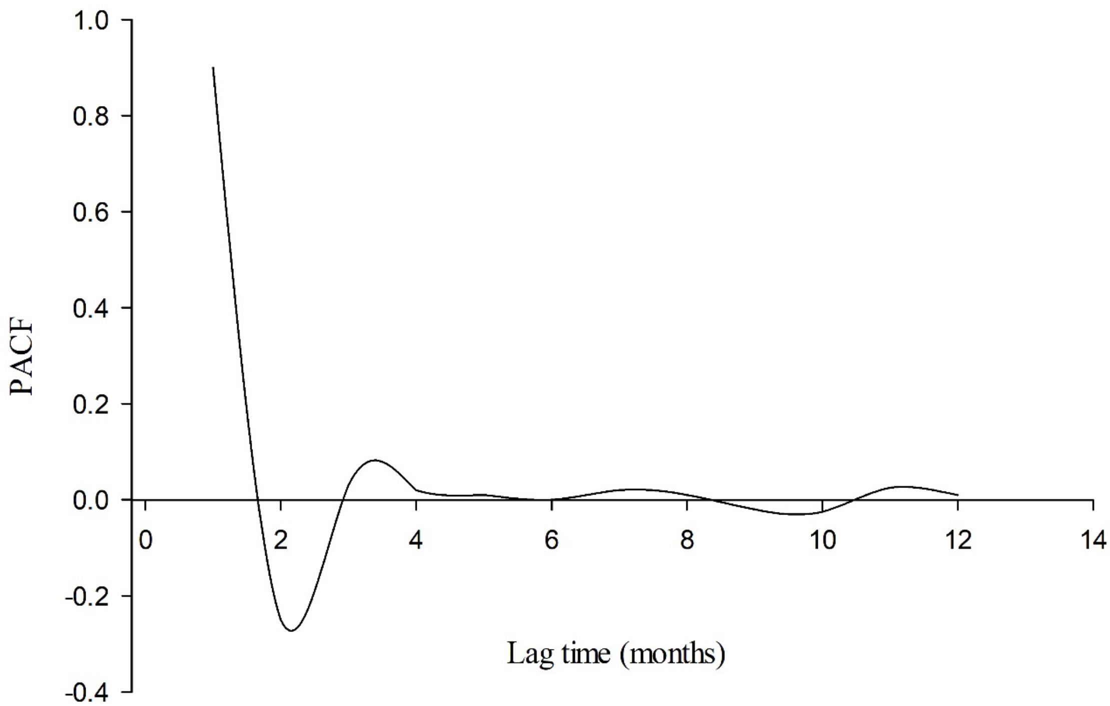

Figure 3. Partial auto-correlation function (PACF) of groundwater level data.

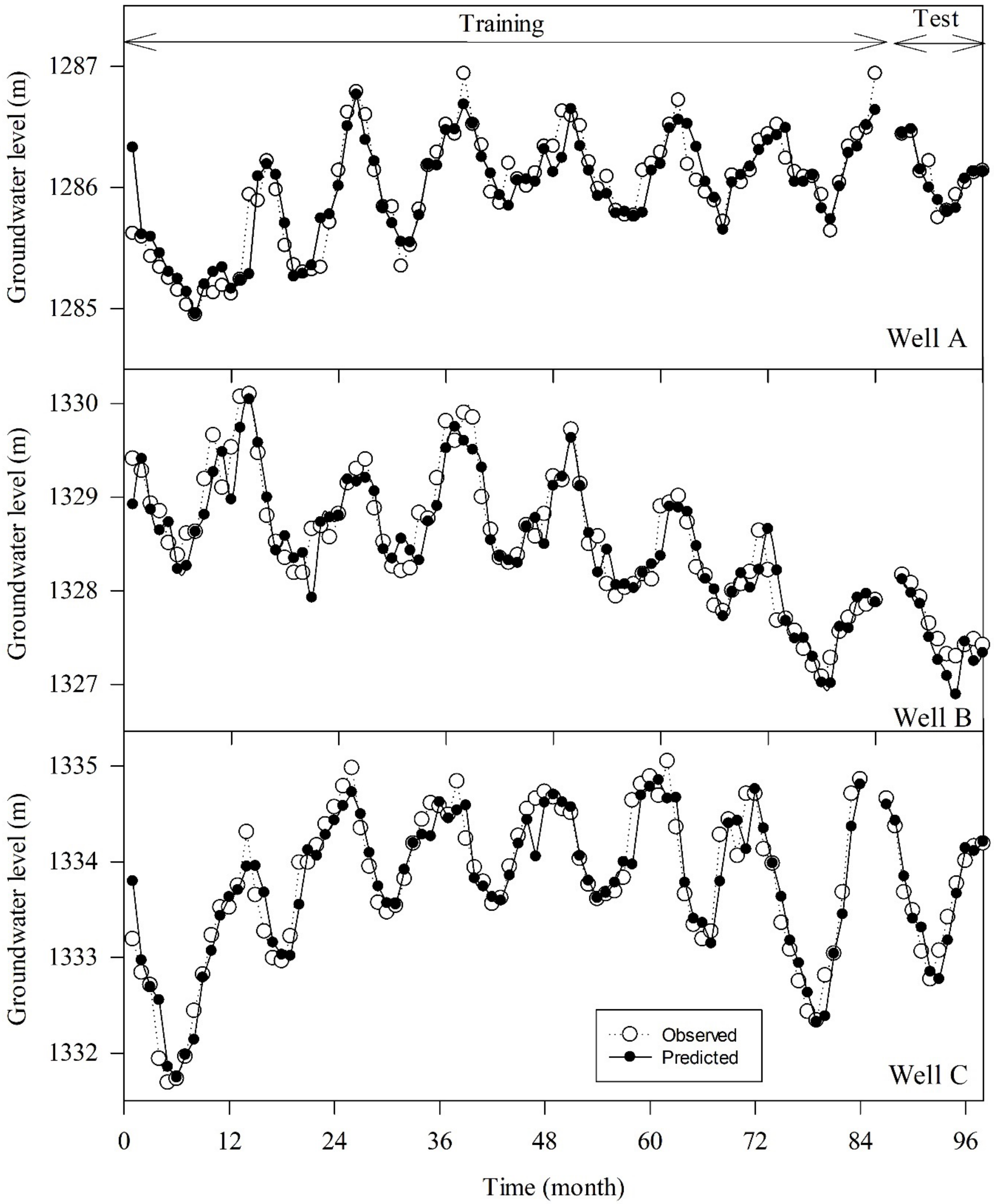

Figure 4. Observed and predicted values of groundwater level for scenario 1 in Well A, B and C.