A comparative study of artificial neural networks and multivariate regression for predicting groundwater depths in the Arak aquifer

Soadat Hanifian[1] · Mohammad Reza Khaleghi[2] · Mohsen Najarchi[3] · Reza Jafarnia[1] · Javad Varvani[4]

Abstract

In recent years, the groundwater resources of Arak plain have been under severe stress, so in some areas, due to the drying up of wells, the depth of wells has increased to access water. In some areas, the groundwater depth is high, which will lead to the salinization of those lands in the future. Regional modeling was used to organize and measure the response of the groundwater resources of Arak plain against the implementation of different management and implementation scenarios. This study aims to investigate the effective factors in the groundwater depth to provide a regional model with multiple linear regression (MLR) methods for Arak plain aquifer. For this purpose, the average groundwater potential maps (GPMs) in the Arak plain, as a dependent variable, and the transmissivity of the aquifer formations, groundwater exploitation values, altitude, average precipitation of the region, the amount of evaporation, and the distance from water resources are considered independent variables and regression analysis is done in SPSS software media. It was done to present a linear model. In the next stage, the presented model was evaluated by applying it to places where its statistics and information were not used to present the model, and finally, by applying this model in the GIS environment, the GPMs for the region were created. The study was prepared. Also, an artificial neural network (ANN) was used to simulate the depth of underground water. The performance of the ANN was measured through parameters such as root-mean-square error (RMSE) and correlation coefficient between real and desired outputs (R). The results of both methods indicate that factors such as the transmissivity of aquifer formations, GPMs drawdown, topography (the height of the well site on the level of the watershed), the groundwater exploitation values at the maximum operating radius of the well, and the distance from water resources are the main factors of GPMs drawdown. But the effectiveness of ANN in estimating GPMs drawdown is higher than the MLR method. The implemented methodology could be generalized to other watersheds with water scarcity problems for groundwater management.

Keywords Aquifer formation · GPMs drawdown · Multiple linear regression · Artificial neural network

Edited by Dr. Michael Nones (CO-EDITOR-IN-CHIEF).

✉ Mohammad Reza Khaleghi
drmrkhaleghi@gmail.com

✉ Mohsen Najarchi
m-najarchi@iau-arak.ac.ir

1 Department of Water Engineering, Arak Branch, Islamic Azad University, Arak, Iran

2 Department of Range and Watershed Management, Torbat-e-Jam Branch, Islamic Azad University, Torbat-e-Jam, Iran

3 Department of Water Engineering, Faculty of Civil Engineering, Arak Branch, Islamic Azad University, Arak, Iran

4 Department of Natural Resources and Environmental Sciences, Arak Branch, Islamic Azad University, Arak, Iran

Introduction

In many dry countries of the world, including Iran, the main source of water supply is groundwater resources. The need for water is increasing due to the increase in human population, the development of agriculture, industries, and other human activities (Chowdhury 2016). The limitation of suitable surface water resources and the increasing demand for water consumption has led people to use groundwater resources (Khazaz et al. 2015). On the other hand, in recent years due to successive droughts and climate change in different regions of the globe, the GPMs in these regions have undergone significant changes. Therefore, estimating and simulating the process of GPMs drawdown is very important, and has been studied by various researchers in recent years. For example, Chelsea and Wan (2013),

Nofal et al. (2015), Priyanka and Mahesha (2015), Naghibi et al. (2017a, b), Vaheddoost and Aksoy (2018), Azari et al. (2021), and Zarafshan et al. (2022), conducted field and analytical studies on quantitative and qualitative parameters of groundwater.

To know the status of groundwater resources, it is necessary to make an accurate prediction of the fluctuations in the level of these resources. This prediction and modeling of the behavior of aquifers are not easily possible due to the complexities of these systems. Therefore, due to the many problems of modeling aquifers with mathematical models, computer models have been used by researchers to predict the water level in aquifers. On the other hand, computer models have provided a tool for managing water resources (Gualbert and Essink 2001), and nowadays, the use of software mathematical models for monitoring and managing groundwater has developed significantly (Cho et al. 2011; Wu et al. 2014; Gholami et al. 2018; Nadiri et al. 2018). In the meantime, by using simulation software, conditions similar to what exists in nature can be created. In this context, one of the effective and fast methods of studying how to move, check the balance, and manage the use of groundwater is indirect methods of study, that is, the use of computer models. These models simulate the natural conditions of the aquifers with a series of mathematical relationships. If a model is accurate, it can be used to predict the state of water resources in the future and examine the impact of applied management conditions. It should be noted that artificial intelligence and soft computing methods can model various nonlinear and complex problems, which have recently been widely used to estimate and simulate the flow of salty water into freshwater aquifers (Coppola et al. 2003; Nourani et al. 2008; Karthikeyan et al. 2013; Hamed et al. 2015; Agarwal and Garg 2016; Naghibi et al. 2017a, b; Azari et al. 2021; Gandhi and Patel 2022; Zarafshan et al. 2022). For example, Lallahem et al. (2005) evaluated the GPMs changes using an artificial neural network (ANN) and concluded that ANN has a good performance in GPMs estimation. Ioannis (2005) used the ANN model with the backpropagation (BP) algorithm and Levenberg–Marquardt (LM) algorithm to estimate the GPMs. They concluded that the ANN model can provide an acceptable estimate for future GPMs using limited data. Krishna (2008) used ANN to model groundwater in the coastal city of Kakinada in India and concluded that this model provides the best prediction with the error BP algorithm and LM algorithm. Rakhshandehroo et al. (2017) and Azari et al. (2021) determined the parameters of the unlimited groundwater table using ANN. They used the BP algorithm to train the ANN model. Roshni et al. (2019) estimated the fluctuations of GPMs in composite aquifers by feed-forward ANN and a hybrid model of Violet-ANN. By examining the results of

numerical models, they stated that the combined model simulated the objective function values with better accuracy. Moreover, GIS is one of the most practical tools in decision-making and has caused many people in various fields to use it as a powerful tool (Sahour et al. 2020). Thapa et al. (2017) used multi-influencing factors (MIF) and GIS to analyze potential zones in the Birbhum district, West Bengal.

On the one hand, due to the large extraction of groundwater tables, the modeling of quantitative and qualitative parameters of these sources of drinking water supply is very important. On the other hand, artificial intelligence techniques and numerical models have acceptable accuracy. In addition, the high speed of calculations and reduction of laboratory costs and field studies are other advantages of using numerical methods. Also, recently numerical models have been widely used to simulate various problems and their popularity is increasing day by day.

In the Arak catchment basin, due to the low surface currents, the main source of water supply for the region is the groundwater, which is fed from the alluvial and limestone aquifers of the Arak plain. Observation of the 10-year unit hydrograph of Arak plain showed that the GPMs has a drawdown of about 6 m during this period. Therefore, it is very important to control what is taken from the aquifer. In general, the effective factors in the fluctuations of the GPMs include the type of transmissivity of aquifer formations, GPMs drawdown, topography (the height of the well site on the level of the watershed), the groundwater exploitation values in the maximum operating radius of the well and the distance from water resources (Brunner and Kinzelbach 2005; Zhang 2001).

In recent years, the Arak Plain watershed has become one of the critical plains due to the drawdown in the fluctuations of groundwater level, the increase in population, and also the land use changes (transformation of pastures into agricultural lands). Identification and zoning of GPMs is an important and necessary step in the comprehensive management of the Arak plain watershed. According to the review of the conducted studies, so far no comprehensive study has been conducted in connection with the zoning of GPMs in the Arak plain watershed. Also, performing the modeling process is challenging due to the complexity of the controlling factors and the need for various inputs. Therefore, it is necessary to carry out the present research to evaluate and manage this watershed to prevent and reduce the damages caused by the decrease in the volume of groundwater. This study aims to use the multivariate regression (MLR) method and ANN to estimate GPMs fluctuations and compare the performance and efficiency of these two methods in simulating GPMs fluctuations in the Arak plain. Therefore, the purpose of this research is the quantitative interpretation of the Arak plain aquifer

system. In this study, the recharging and discharge status of the Arak Plain aquifer is investigated. This leads to the long-term management of the aquifer and makes appropriate decisions by the planners.

Methods and materials

Study area

The studied area is located in the central province of Arak, which has a semi-arid climate, an average annual rainfall of 280 mm, and a temperature of 18 °C (Ghadimi 2015). The Arak aquifer is located between the Arak Mountains in the south and Miqan playa in the north of Arak and is fed by the two rivers Qara Kahriz and Aman Abad in the south and east of Arak city. These rivers pass through agricultural lands and various industries on their way. Qara Kahriz River in Arak plain is affected by urban sewage and industrial effluents. The Amanabad River also passes through the Arak garbage dump in the Amanabad plain. Geologically, the area is located in the Sanandaj-Sirjan zone, which contains poorly metamorphosed rocks, including Jurassic shale, sandstone, and slate, as well as Cretaceous crystallized limestone (Ghadimi et al. 2016). The bottom rock of the aquifer is Cretaceous crystallized limestone (Fig. 1). The Arak plain between the Arak Mountains in the south and Miqan playa in the north has drinking water wells that have been affected by the calcareous composition of the highlands and the salty layers of the Miqan playa (Ghadimi et al. 2015).

Methodology

Due to the lack of comprehensive quantitative and qualitative information about water resources is limited in Iran, it important to use different methods and models to estimate the quantitative and qualitative parameters of water resources. In this research, GIS software version 10.3 was used to prepare quantitative groundwater potential maps (GPMs) of the studied areas. Also, the MLR method was used to provide a linear model to estimate the average depth of the groundwater aquifer in the Arak plain. For this purpose, the average GPMs in the Arak plain, as a dependent variable, and the factors of water transmissivity of the aquifer formations, altitude, average precipitation of the region, the amount of evaporation, and the distance from the water resources are considered as independent variables and MLR analysis is done in SPSS software environment. It is done to present a linear model. In the next step, the presented model was evaluated and evaluated by applying it in places where its statistics and information were not used to present the model, and finally, by applying this model in the GIS environment, the GPMs for the study area will be prepared.

Effective factors in the GPMs drawdown

In this part of the study, the values of the annual drawdown of the GPMs were used as the output of the ANN and the effective factors in the fluctuation of the water table depth were used as the inputs of the network in the modeling process. The input factors of the network or the effective factors in the fluctuations of the GPMs include groundwater
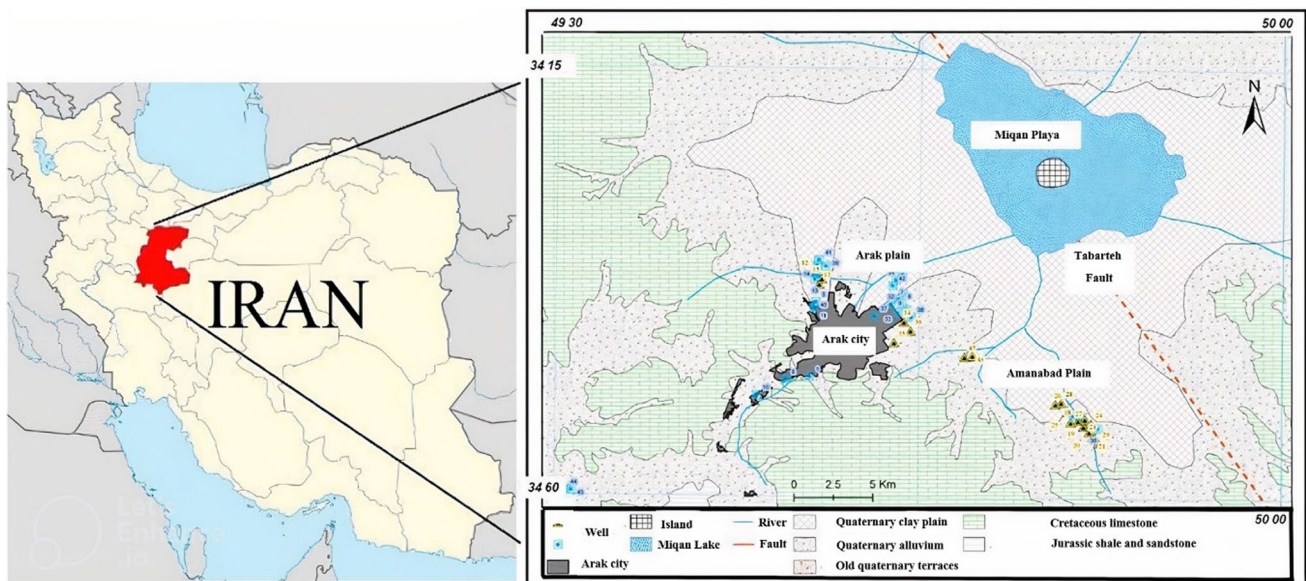

Fig. 1 The studied area in Markazi Province

exploitation values, depth of the water table, transmissivity of aquifer formations, topography (slope and height), location in the watershed, distance from water resources (recharging the water table), annual precipitation, annual evaporation and distance from industries and residential centers. Therefore, it was estimated as follows for the outlet and inlet study wells of the network. Therefore, for the study wells, the outlet and inlets of the network were estimated as follows.

Groundwater depth

Based on the observation statistics of the groundwater depth in the study wells, the values of the average groundwater depth were determined. Then, by using the data of the average groundwater depth and also the interpolation capabilities of the GIS, the spatial distribution of the annual average groundwater depth in the Arak plain was obtained.

Transmissivity of aquifer formation

According to Wei et al. (1990), GPMs is very sensitive to transmissivity (Wei et al. 1990). Therefore, it is one of the most important hydrodynamic parameters of the groundwater. Based on the data from Markazi Regional Water Authority (MRWA) and pumping tests in the study wells, the transmissivity of aquifer formation has been determined in the location of each of the study wells (MRWA). Then, the GPMs of the average transmissivity of aquifer formation in the Arak plain were prepared based on the figures determined in the wells and the determination of homogeneous levels, using geological and hydrogeological GPMs by the MRWA. Transmissivity values ranged from 2013 to 11,675 m^2/day for the studied wells.

Groundwater exploitation values

Determining the amount of groundwater exploitation values is one of the most important and difficult inputs. It is difficult to access accurate information about exploitation values, especially regarding agricultural wells in Iran, because these wells do not have smart contours. For this purpose, the data of all rural and urban drinking water wells and industrial wells were obtained from the MRWA, and their annual groundwater exploitation values were determined.

Annual precipitation and evaporation

To estimate the spatial distribution of precipitation and evaporation at the study level, the statistics of normal values of precipitation (rain gauge, climatology, and synoptic station) and annual evaporation (evapotranspiration station) in thirty years were used in the GIS environment. In this regard, we have used the interpolation technique in the GIS environment and a GPMs or raster layer of average annual precipitation and evaporation was prepared.

Topography

The 10 m contour lines of topographic GPMs have been used to prepare a 10 m digital height model, and slope map and determine the location of the well on the watershed level.

Distances from water resources

All the rivers and lakes of the study area were investigated based on topographical maps and satellite images, and a map of the distance from water resources was prepared in the GIS environment.

Distances from manufacturers and settlement areas

An information bank of the province's industries has been prepared and checked on satellite images. Residential areas were also prepared based on topographical maps with a scale of 25,000 and checked using satellite images. Then, distance maps from manufacturers and settlement areas were prepared in a GIS environment.

Modeling stage

Presenting the model with the MLR method

Statistical analyses were done using SPSS software in a step-by-step manner. The average GPMs drawdown was considered a dependent variable and the factors influencing it were considered independent variables. In the next step, the accuracy and efficiency of the presented model were evaluated in the places of Arak Plain where the statistics and information were not used to present the model, and the estimated values, which included the average GPMs drawdown value for that place, were reported with the average values. They were compared by the TAMAB (80 percent of the data were used to train the model and 20 percent of the data were used to test the model). With this method, a linear model was presented, which requires average values of surface water drawdown, such statistics are available only in the range of hydrometric stations, and on the other hand, by removing this factor, the resulting linear model is not significant. Therefore, a nonlinear model based on the two factors of transmissivity of aquifer formations and distance from water resources was presented to estimate the average GPMs drawdown, and the effectiveness of this nonlinear model was evaluated.

Prediction of GPMs drawdown by using ANNs

Simulating the groundwater system is not easily possible due to the complexities of these systems. This is while ANNs as black box models with their high capabilities are very suitable for modeling complex and nonlinear systems. Therefore, due to the many problems of modeling aquifers with mathematical models, ANNs have been used by researchers to predict the GPMs in aquifers.

For the modeling process, different ANNs were used to reach an optimal network to estimate the annual drawdown in GPMs. In this regard, the drawdown values were considered as output and all effective factors as network input. An example of the input and output data of the ANNs is presented in Table 1. At first, the data were normalized and then randomly generated. In the next step, the data were divided into three categories: training data (65% of data), validation data (10% of data), and test or validation data (25% of data). Also, to evaluate network inputs and determine the main factors of the GPMs drawdown, a correlation analysis between inputs and outputs was done.

Performance evaluation of the models

To evaluate the efficiency of the used methods, error indices and correlation between the estimated and observed values have been used in the training and testing stages. Also, the observed values of GPMs drawdown in study wells were superimposed on the simulated annual GPMs and the accuracy of the simulated and the observed GPMs drawdown values was compared and investigated. The performance of the two methods was evaluated by comparing the predicted GPMs drawdown with the recorded rates on the testing subset using the mean squared error (MSE), the coefficient of determination (R sqr), and the mean absolute error (MAE), and Nash–Sutcliffe coefficient (NSE) in the simulation.

$$MSE = \frac{\sum(Qi - \hat{Qi})}{n} \tag{1}$$

$$R^2 = \left[\frac{\sum_{i=1}^{n} \left(Q_i - \overline{Q}_i \right) \cdot \left(\hat{Q}_i - \widetilde{Q}_i \right)}{\sqrt{\sum_{i=1}^{n} (Q_i - \overline{Q}_i)^2 \cdot \sum_{i=1}^{n} (\hat{Q}_i - \widetilde{Q}_i)^2}} \right]^2 \tag{2}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} \left(\hat{Q}_i - Q_i \right)^2}{\sum_{i=1}^{n} \left(Q_i - \overline{Q} \right)^2}, \tag{3}$$

where Q_i is the observed value, $\hat{Q}_i$ is the simulated value, $\overline{Q}_i$ is the mean of the observed data, $\widetilde{Q}_i$ is the mean of the simulated data, and n is the number of data points. These criteria were used to evaluate the ANN performance in GPMs drawdown modeling because the inputs and output data were parametric (quantitative) variables since these criteria are common in model performance evaluation (Gholami et al. 2018). The optimum model is the model that produces the highest R squared and NSE and the lowest NRMSE values.

Mapping of GPMs drawdown by coupling ANN and GIS

To prepare the average GPMs, the optimal method for modeling and mapping the average GPMs drawdown will be used. The use of GIS along with the optimal method provides a good basis to support the modeling of average GPMs drawdown (Solomatine and Ostfeld 2008). First, the quantitative values of the model parameters, including the transmissivity of the aquifer formation, the groundwater exploitation values in the maximum operational radius of the well, the distance from the water resources, and the topography (site height) are evaluated using data, GPMs and digital layers in the GIS environment. Then, the modeling process is done to simulate the average GPMs drawdown. Then, the tested and optimal model is selected to predict the average GPMs drawdown in the studied plain. Then, the raster layers of model inputs are prepared and combined with a one-kilometer pixel overlap analysis. The model inputs in the GIS environment and the average GPMs drawdown depth are estimated using the best-tested model for the entire plain. Finally, the average GPMs drawdown of the plain is prepared using GIS.

Results

The results of the MLR method

As mentioned, the MLR method was used to provide the model to estimate the average drawdown of groundwater. The results of using the MLR method to present the model were presented in the form of a linear model (Table 1).

Table 1 The results of running the linear model with the MLR method ($R^2 = 0.632$)

Model	Sum of squares error (SSE)	df	Mean absolute error (MAE)	F	Sig.
Residual error (RE)	1.525	3	0.5084		
	1.168	24	0.0486	10.45	0.000
	2.693	27			

In the study area of Arak plain, 800 wells with GPMs records have been identified be investigated. A summary of the characteristics and the study position of the current research is presented in Tables 2 and 3. Table 2 shows a summary of the hydrological characteristics of the Arak plain.

Also, to determine the optimal inputs and to determine the main factors of GPMs drawdown in the study area, the correlation analysis between the inputs and outputs of the network has been used, and their correlation coefficients are presented in Table 4.

As can be seen, several factors affect the changes in the GPMs, such as groundwater exploitation values, the transmissivity of aquifer formations, annual precipitation and evaporation values, altitude of the place, and distance from water resources. Also, based on the correlation coefficient, the most important factors include the transmissivity of aquifer formations (0.74), height (0.56), groundwater exploitation values (0.47), and distance from water resources (0.41). Regarding the transmissivity variable ($m^2.day^{-1}$), there is a direct relationship between the transmissivity of aquifer formations and the depth of the water table. Also, this relationship is a direct relationship regarding the variables of location height (meters), annual evaporation, distance from water resources, and annual exploitation, and an inverse relationship regarding the variable of annual precipitation.

The direct relationship between the transmissivity of aquifer formations and the depth of the water table means that the higher the transmissivity of the aquifer formations, the greater the GPMs due to the higher possibility of exploitation with a higher pumping flow rate (Nordqvist et al. 2008). Precipitation, as one of the main factors of natural nutrition of underground water tables, has a significant and weak inverse relationship with the water table depth values (Doll et al. 2014). Evaporation due to the reduction of surface water resources has a significant and weak direct relationship with the water table depth values in the region (Naghibi et al. 2017a, b). Regarding the altitude variable, Gholami et al. 2018) studies indicate that low-altitude areas, due to their high ability to absorb water, have a high GPMs and less drawdown due to harvesting. According to

the results, lower elevations have more groundwater potential than higher elevations and the water table is higher in them. One of the reasons for this is the creation of more runoff at higher altitudes and the recharging and infiltration of groundwater at lower altitudes, which is based on the results of Davoudi Moghadam et al. (2015), Arabameri et al. (2019) and Rahmati et al. (2018). The slope parameter showed that in lower slopes due to the decrease in the runoff, the amount of underground water potential is higher, which is the result of more water recharge in these areas, and it is consistent with the results of Lee et al. (2012 and 2019) and Golkarian and Rahmati (2018).

Also, by moving away from waterways, the amount of groundwater potential in the studied area has decreased, which shows the connection between groundwater and waterways (Haqizadeh et al. 2017; Naghibi et al. 2017a, b). Concerning the distance from water resources, Sahour et al. (2020) stated that the shorter this distance is, the fluctuations of the GPMs also decrease because the drawdown caused by harvesting is compensated through river recharging.

Table 3 Characteristics of Arak aquifer (Mohammadi Ghaleni et al. 2011)

Characteristic	Value
Aquifer type	Unconfined
The area of the plain (square kilometers)	3392.7
Operating status	Banned
The number of deep wells	1458
The number of semi-deep wells	1554
Total discharge (million cubic meters)	702.35
Drinking water consumption (million cubic meters)	76.424
Agricultural consumption (million cubic meters)	612.390
Industry consumption (million cubic meters)	13.534
Average water depth (meters)	27
Average annual water level drawdown (meters)	0.46
Average reserve ratio (percentage)	0.03
Aquifer area (square kilometer)	1946

Table 2 A summary of the hydrological characteristics of Arak plain

Characteristic	Value
Altitude (meters)	1700
Area (square kilometers)	3400
Slope	8
Average temperature (Celsius)	13.8
Normal annual precipitation (mm)	280
Mean humidity (percent)	46
Mean values of evaporation (mm)	1535.6
Climate (Demartin classification)	Semi-dry cold

Table 4 Correlation coefficients (Pearson) between the GPMs and the parameters affecting the GPMs

Factor	Correlation (Pearson) with the GPMs	Sig. (P value)
Transmissivity	0.74	0.000
Annual exploitation	0.47	0.000
Annual rainfall	−0.005	0.002
Annual evaporation	0.18	0.000
Height	0.56	0.000
Slope	0.06	0.003
Distance from water resources	0.41	0.00

The contour lines of the groundwater of the Arak aquifer in 2015 are drawn in Fig. 2. As shown in Fig. 2, the depth of GPMs in the central part of Arak is lower than in other parts. This is related to the lower height of the center of the plain compared to the surrounding area of the Arak plain.

Spatiotemporal changes of GPMs in the study area

Most observation wells in the Arak aquifer have a drawdown between 0 and 5 m. On average, in 62% of the area of the Arak aquifer, the GPMs are between zero and 3 m, in 22% between 3 and 6 m, in 8% between 6 and 9 m, in 7% between 9 and 12 m and in 0.6% it has a drawdown between 12 and 15 m. Figure 3 presents the time changes of the GPMs of the Arak aquifer in a sample well.
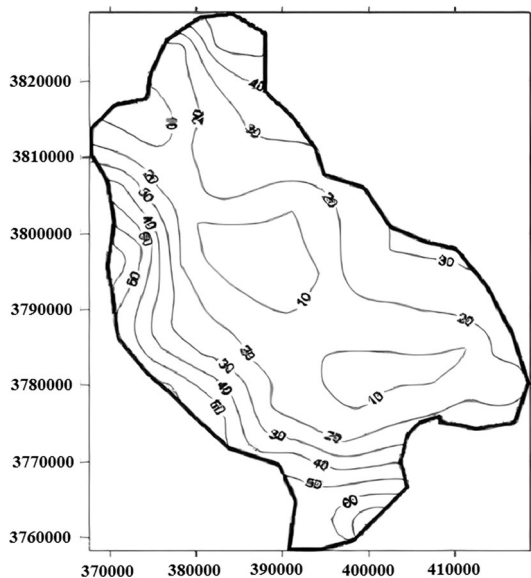
Fig. 2 Map of the depth of underground water in the Arak plain
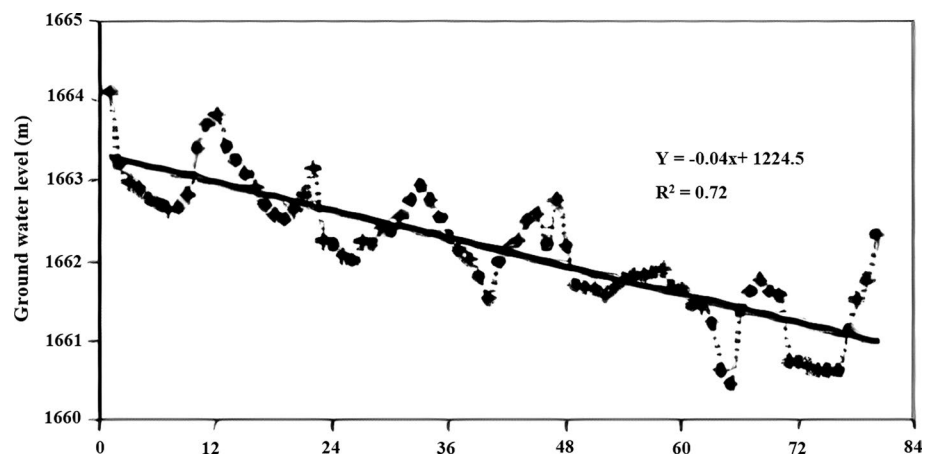
The trend of time changes in the GPMs in the study area

The annual average level of the GPMs in one of the sample wells is plotted in Fig. 2 and a line has been fitted on it to determine the trend of changes in the level of the GPMs. In the equation of the fitted line, y represents the GPMs in meters, x is time in years or months, and R^2 (the coefficient of determination) represents the percentage of total changes that are determined by the linear relationship between the GPMs and time.

The negative slope of the fitted line (a negative sign of the x coefficient) indicates the drawdown in the GPMs in the sample wells. The GPMs of the Arak aquifer have decreased over time, but due to the smaller coefficient of x in these wells, the intensity of the decrease in the wells of the Arak plain has not been so severe. The high coefficient of determination in the fitted lines on the changes of the GPMs in Arak Plain indicates the existence of a linear relationship between the fluctuations of the GPMs with time and its strong temporal structure.

The results of the ANN application

The simulation of groundwater systems is far different from resurfacing water due to the complexities of these systems. This situation is especially evident in the optimization of these systems. In the present research, the GPMs for the studied plain was estimated based on secondary data, observation wells, and the ANN model. It should be noted that 84 months of observation values were used for training artificial intelligence models and 41 months for testing. The results showed that the ANN model had a high ability in simulating the GPMs in the training (R sqr = 0.96) and testing or validation stages (R sqr = 0.85). In this regard, the average GPMs in 800 piezometric wells were estimated and considered as the output of the system. Also, the factors of transmissivity of aquifer formations, GPMs drawdown,

Fig. 3 Time changes of the GPMs of Arak aquifer in a sample well

$$Y = -0.04x + 1224.5$$
$$R^2 = 0.72$$

topography (height of the well site on the level of the watershed), annual evaporation values, groundwater exploitation values in the maximum operating radius of the well, and distance from water resources were considered as system inputs. System training, verification, and testing were done in the environment of Neurosolution software. The ANN evaluation results are presented in Table 5, which shows the error values in the training phase. Based on this, acceptable results were observed in the training phase. Also, the transmissivity of aquifer formations, GPMs drawdown, topography (the height of the well site on the level of the watershed), the groundwater exploitation values at the maximum operating radius of the well, and the distance from water resources are the most important inputs or effective factors in the amount of GPMs drawdown. In the end, using the tested system and the raster layers of the mentioned inputs in the GIS environment, zoning the GPMs drawdown was prepared and its accuracy was evaluated and confirmed by comparing the estimated values and observations of the GPMs (R sqr = 0.96). Digital GPMs of these factors were prepared in the GIS environment, which is presented in Figs. 4 and 5.

The performance evaluation results of the ANN model in simulating the GPMs in the training phase are presented in Table 5. The correlation between observed and simulated data (R) for optimal model input is equal to 0.96.

The results showed that the effective factors in the GPMs and the best inputs in simulating the GPMs include the transmissivity of aquifer formations, GPMs drawdown, topography (the height of the well site on the level of the watershed), the groundwater exploitation values at the maximum operating radius of the well and the distance from water resources. The performance evaluation results of the ANN method in simulating the drawdown of GPMs in the test or validation stage are presented in Table 5 and Fig. 5. Table 5 shows the error of the training phase, and according to them, good results were obtained in the training phase. In such a way, regarding the estimation of the GPMs drawdown index, the ANN model has acceptable efficiency and accuracy (R = 0.96). Such results are consistent with the results of other researchers (Nourani et al. 2008; Naghibi et al. 2017a, b; Sahour et al. 2020). Also, in the training phase, especially in simulating the GPMs, the methods used have acceptable efficiency based on statistical criteria (R squared, NSE, and NRMSD), so that having the least error in similar values

constructed had the highest agreement between observed and simulated values (Table 5). The results of the validation or test phase (Table 5) also showed that based on the NSE, NRMSD, and R-squared indices, the ANN model method is an effective method for simulating the depth of the water table. Figures 5, 6 and 7 show the results of the comparison between the simulated values and the observed values of the table depth in the test phase, which shows the efficiency of the ANN method between the simulated values and the observed values. Then, the estimated depth values of the ANN method for the entire surface were presented in the GIS environment as a GPMs of the annual depth values of the water table (Fig. 8). To evaluate the efficiency of the method used and the accuracy of the obtained results, the annual observation figures of the GPMs in the study wells are overlapped on the mentioned GPMs. The comparison of the GPMs of water table depth values with observed depth values indicates the accuracy of the results and the efficiency of the methodology used (R sqr = 0.96).

The purpose of this study is to evaluate the GPMs in places without data and to make its results available to the public. At this stage, GIS capabilities were used to monitor the results of the ANN model as a raster layer of the GPMs, and the results are presented in Fig. 7. A comparison of observed values with areas of estimated GPMs shows the high performance of the ANN model (Krishna et al. 2008; Agarwal and Garg 2016).

Discussion

Two methods of MLR and ANN were used with the same data to simulate GPMs drawdown in Arak plain. The results obtained from both methods indicate that the transmissivity of aquifer formations, GPMs drawdown, topography (the height of the well site on the level of the watershed), the groundwater exploitation values at the maximum operating radius of the well, and the distance from water resources are the most important inputs or factors. They are effective in lowering the groundwater table of the Arak plain. Therefore, the mentioned parameters were considered as inputs for simulating the drawdown of the GPMs by MLR models and ANN. Kamasi et al. (2016) considered rainfall, temperature, and river discharge as factors affecting the GPMs of

Table 5 Performance of the two models in the training and testing stages of GPMs modeling

	NSE		NRMSD		R squared	
	Training	Testing	Training	Testing	Training	Testing
ANN	0.95	0.8	0.01	0.5	0.96	0.85
MLR	0.82	0.43	0.41	0.83	0.82	0.71

NSE: Nash–Sutcliffe efficiency. *NRMSD*: Normalized root-mean-squared deviation. *R* squared: Coefficient of determination
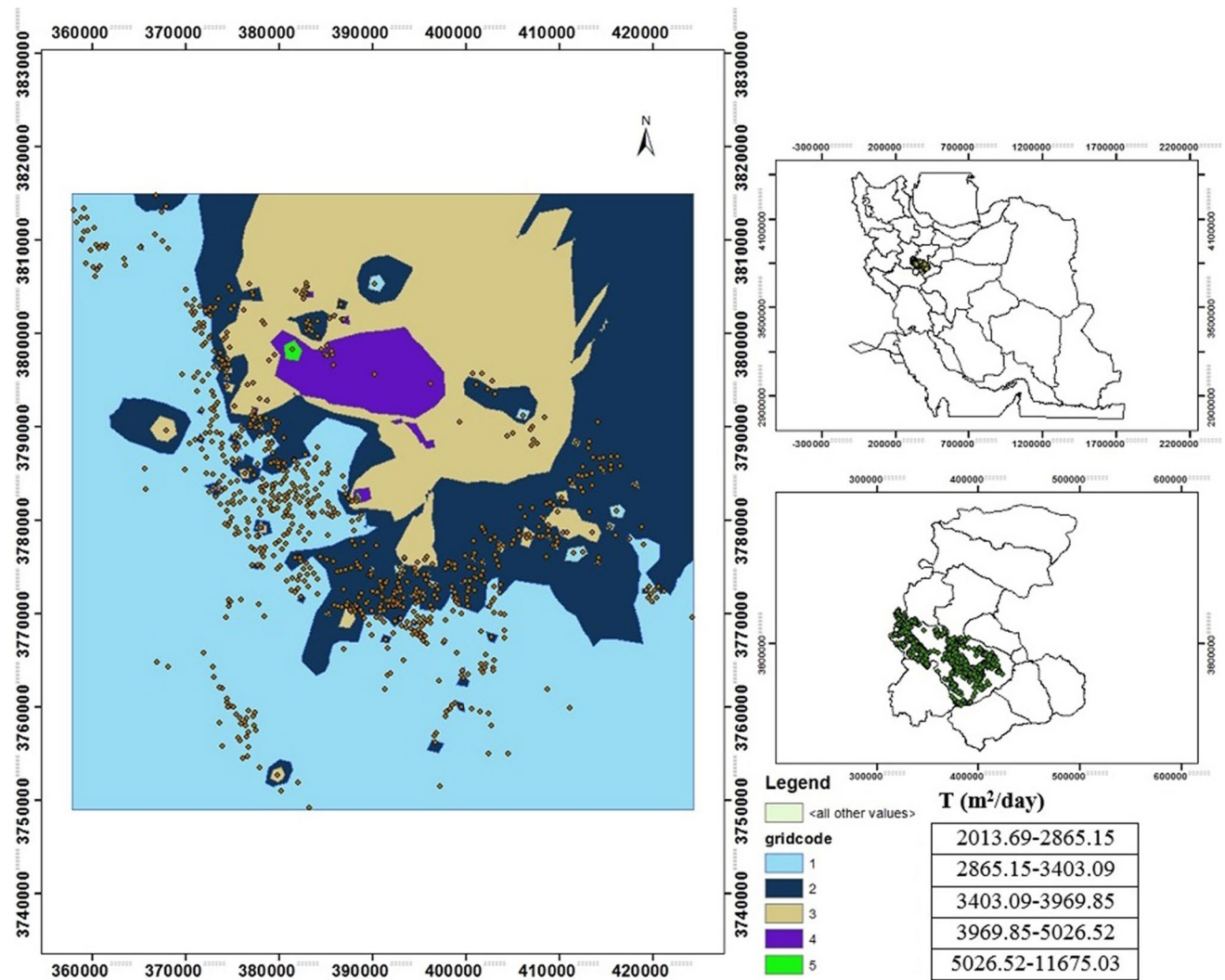
Fig. 4 Transmissivity map of aquifer formations in the study area (square meters per day)

the Silakhor plain aquifer. Naghibi et al. (2017a, b) empha-sized on higher importance of altitude, TWI, slope angle, and slope aspect in groundwater assessment. The review of scientific literature indicates the growing trend of research in the field of various effective factors in the zoning of GPMs (Rahmati et al. 2016; Naghibi et al. 2017a, b; Rakhshan-dehroo et al. 2017; Azari et al. 2021; Gandhi and Patel 2022; Zarafshan et al. 2022). The results of Rahmati et al. (2016) showed that altitude, drainage, density, lithology, and land use had the highest importance in the zoning of GPMs. A comparison of these studies indicates that the importance of each of the effective variables is a function of implemented indicators and methods and hydrological, and geological characteristics of the destination aquifer.

The average drawdown of the GPMs in the wells of Arak is 3.38 m and the highest drawdown is 15.33 m and the annual drawdown is 2.19 m. Akbari et al. (2009) reported

the average drawdown in the GPMs in the Mashhad plain of about 0.6 m, which is less than the drawdown in the Arak plain. This drawdown is due to several important factors. The first and most important reason is the indiscriminate extraction of the groundwater resources of the Arak aquifer and the second reason is the occurrence of recent droughts in the region.

Evaluating the efficiency of linear models with the MLR method showed that the maximum correlation between the estimated values and the observed values does not exceed 0.62. Instead, using the same data to evaluate the efficiency of the ANN, the correlation between the estimated values is 0.68. Therefore, the efficiency and accuracy of the ANN in the simulation of GPMs drawdown are more than the MLR method.

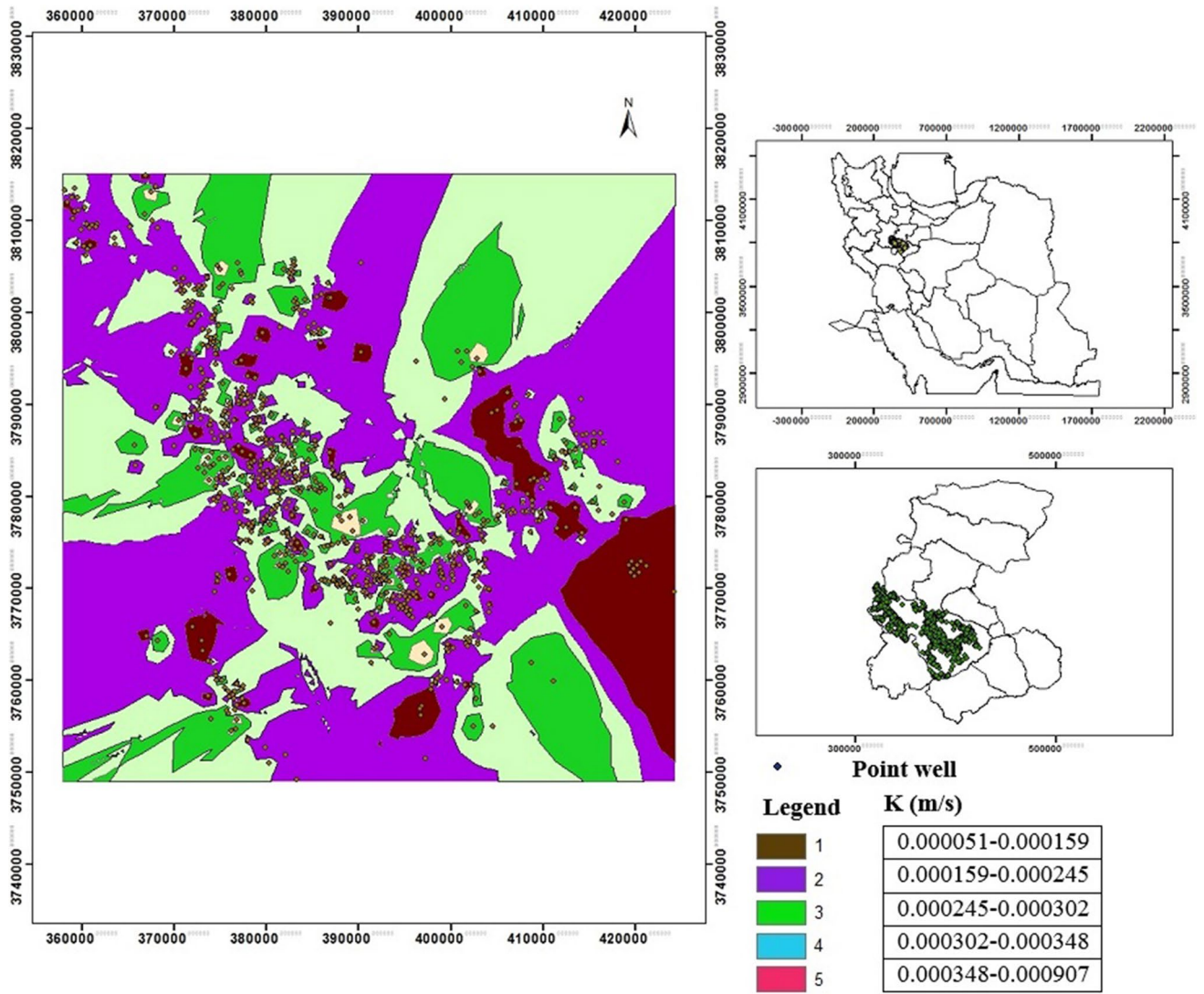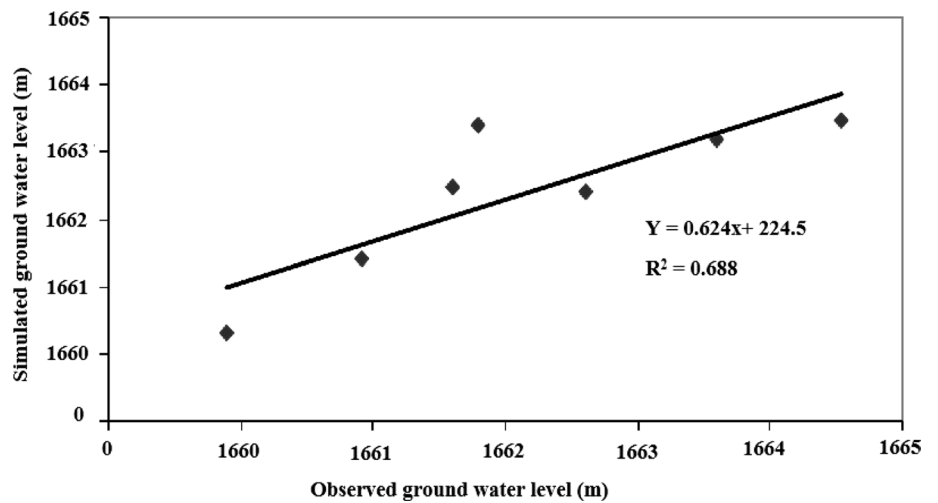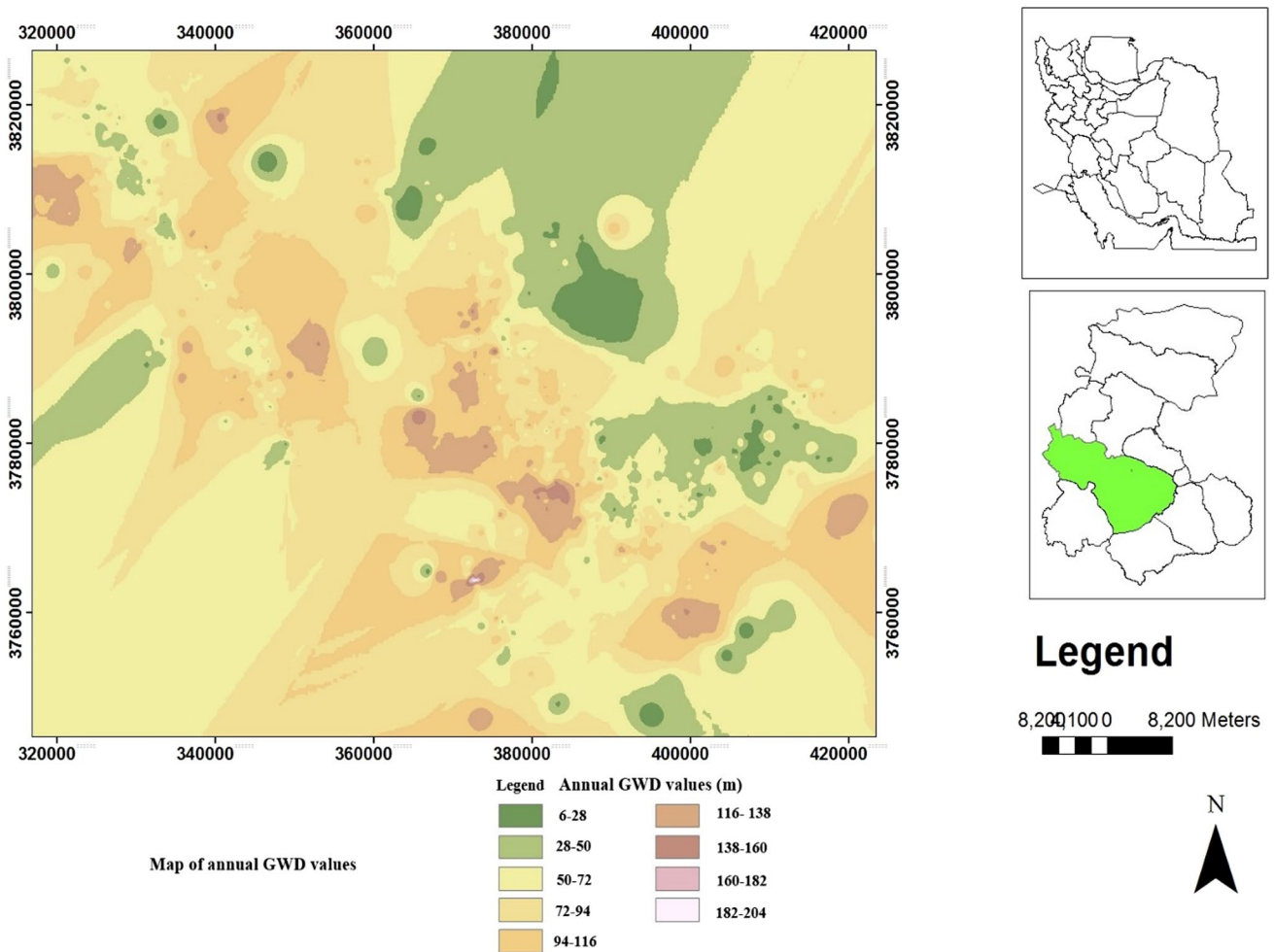According to Table 5, in each of the MLR and ANN mod-els separately, the accuracy of simulating the parameters
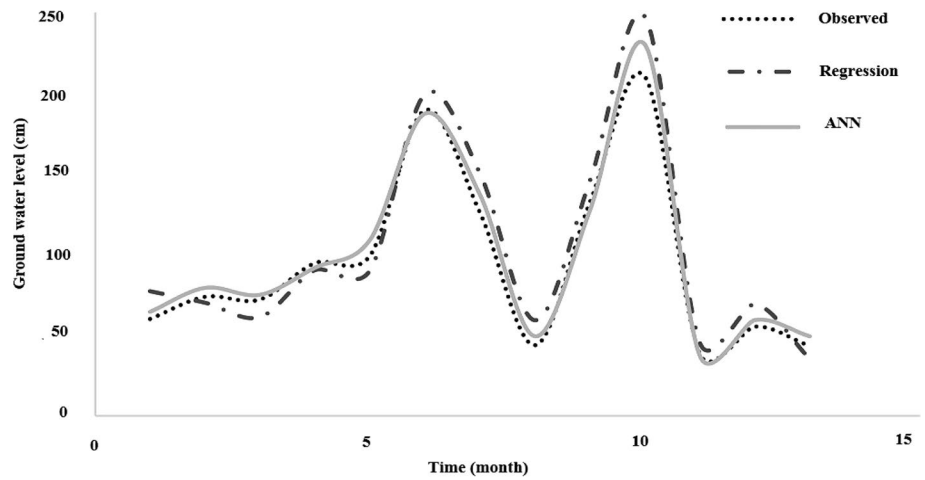
Fig. 5 Hydraulic conductivity map of aquifer formations in the study area (meters per day)

Fig. 6 Scatter diagram of measured and simulated data of GPMs using ANN model

$Y = 0.624x + 224.5$

$R^2 = 0.688$

Fig. 7 Results of GPMs simulation using ANN and MLR

Fig. 8 Map of annual GPMs values

of the GPMs was in the model with a hyperbolic tangent sigmoid transfer function; so that NRMSD and NSE had the lowest value and R-squared index had the highest value. Therefore, the optimal model for estimating GPMs

parameters is ANN with a hyperbolic tangent sigmoid transfer function. Alipour (2012) predicted the GPMs of North Mahyar plain using the ANN model and fuzzy inference system. The optimal learning algorithm was obtained based

on the least mean squared error of Levenberg–Marquardt. The transfer function of the hidden layer and the output of the sigmoid and linear logarithms were obtained. Sreekanth et al. (2009) tested the evaluation of ANN models such as the LM algorithm in predicting the groundwater level in the Maheshwaram watershed in Hyderabad state, India. The results showed that the model is completely consistent with reality and the obtained data have acceptable accuracy and certainty ($R^2 = 0.93$, RMSE $= 4.50$). Keykhosravi et al. (2019) used the perceptron and radial base ANN model to predict the dust of the Sabzevar cement factory. The results indicated a higher accuracy of the perceptron ANN model than the radial basis model in predicting the amount of dust.

Also, as seen in diagrams 4 to 6, the highest overlap with the measured values of the GPMs parameter is related to the ANN model. In the ANN model, the BP algorithm is used for training. This method usually gives relatively good results for prediction, but its problem is getting stuck in the local optimum and slow speed in reaching the answer.

As it is clear in Table 5, a high value of the correlation coefficient was obtained, which indicates the closeness of the predicted values with the measured data and the ability and high accuracy of the relationships between the input and output variables. These results show the acceptable accuracy of the ANN and the good and efficient learning of the network using the desired training algorithm and the data provided to the network. According to the conducted research, the result of using the ANN model in simulating and predicting the average water drawdown can be properly and optimally evaluated and its sufficient ability, in this case, can be confirmed. The results of research by Nourani et al. (2008), Gholami and Ghomi (2013), Agarwal and Garg (2016), Gholami et al. (2018), and Sahour et al. (2020) also indicate the high efficiency of the ANN in estimating and simulating groundwater parameters. The approach of this research was ANN with a multi-layer perceptron structure with the Levenberg–Marquardt algorithm. The results of past research also indicate the high efficiency of the ANN in estimating and simulating groundwater parameters (Gholami and Ghomi 2013).

Conclusion

In recent years, with the increasing population and industrial development, the use of underground water resources has multiplied. With the continuation of this practice, the level of underground water has draw-downed day by day. Therefore, identifying these resources and using them optimally means a sustainable and permanent harvest of this God-given wealth.

Groundwater modeling with the help of the ANN model has a special position due to the ability to learn by example and without the need for equations governing the phenomenon. The present study was conducted to simulate the parameters of the GPMs of Arak plain using MLR models and ANN in the environment of GIS software and finally compare their results with the measured data. At first, based on the results of the calculation of the correlation coefficient, the highest correlation coefficient of the average drawdown of groundwater is related to the parameters of the transmissivity of aquifer formations, GPMs drawdown, topography (height of the well site on the level of the watershed), groundwater exploitation values at the maximum radius of the well and the distance from water resources. Therefore, the mentioned parameters were considered as input for simulating the average drawdown of the GPMs. Next, the results obtained from the ANN model showed that the accuracy of simulating GPMs parameters is the ANN with sigmoid tangent transfer function.

Modeling and simulating GPMs drawdown through ANN has a special place due to its ability to learn by providing examples without the need for equations governing the phenomenon. ANNs are one of the most useful models in predicting and modeling complex hydrogeological and hydrological issues. By choosing the right type and number of input factors and using the right and compatible type of ANN and its proper calibration, it can be said that this technique is a very effective and suitable tool for estimating GPMs drawdown in the Arak plain.

In general, it can be stated that the ANN can extract the law governing the data, even the confusing data. This feature of ANNs can be considered one of the most prominent features of this technique compared to other methods. By choosing the right type of learning technique, the right number of neurons and hidden layer, the right type and the number of input factors, as well as its proper calibration, it can be said that this technique is a very effective and suitable tool for estimating GPMs drawdown in the Arak plain. Also, this model can be useful for facilitating the development and implementation of quantitative and qualitative strategies for groundwater and it is a step in making management decisions to improve the groundwater of watersheds.

The results of both methods indicate that factors such as the transmissivity of aquifer formations, GPMs drawdown, topography (the height of the well site on the level of the watershed), the groundwater exploitation values at the maximum operating radius of the well, and the distance from water resources are the main factors of groundwater loss. But the effectiveness of ANN in estimating groundwater loss is higher than the MLR method.

The results of this research showed that the ANN model with a backpropagation algorithm has the highest network efficiency, the lowest percentage of error, and a high ability in predicting the time of hydrogeological phenomena. As a result, this model can accurately predict the underground

water level in the studied area and avoid spending more time and money to dig more piezometers. From the point of view of methodology, this research can solve many problems in predicting the time of underground water level in aquifers. The results of the ANN model showed the high power of this nonlinear method in estimating the desired parameter. The approach of this research was an ANN with a multi-layered perceptron structure. Therefore, by using ANN, we can estimate the GPMs drawdown or even other quantitative and qualitative water parameters in places without statistics and use them for optimal management of water resources.

Declarations

Conflict of interest There is no conflict of interest among authors.

References

Agarwal R, Garg PK (2016) Remote sensing and GIS-based groundwater potential and recharge zones mapping using multi-criteria decision analysis making technique. Water Resour Manage 30:243–260

Akbari M, Jarge MR, Madani SH (2009) Assessment of decreasing of groundwater table using Geographic Information System (GIS). (Case study: Mashhad Plain Aquifer)

Alipour Z (2012) Evaluation of adaptive fuzzy-neural inference system and neural network in groundwater level prediction (North Mahyar Plain). Msc Thesis University of shahid chamran, Ahvaz, Iran, p 177

Arabameri A, Cerda A, Tiefenbacher JP (2019) Spatial pattern analysis and prediction of gully erosion using novel hybrid model of entropy-weight of evidence. Water 11(6):1129. https://doi.org/10.3390/w11061129

Azari A, Zeynoddin M, Ebtehaj I, Sattar AMA, Gharabaghi B, Bonakdari H (2021) Integrated preprocessing techniques with linear stochastic approaches in groundwater level forecasting. Acta Geophys 69:1395–1411. https://doi.org/10.1007/s11600-021-00617-2

Brunner P, Kinzelbach V (2005) Groundwater Modeling in a remote Chinese basin-how can models be improved in areas where data are scarce? Eur Geosci Union 2005:2p

Chelsea Q, Wan Y (2013) Time series modeling and prediction of salinity in the Caloosahatchee river estuary. Water Resour Res 49(9):5804–5816

Cho KH, Sthiannopkao S, Pachepsky YA, Kim KW (2011) Prediction of contamination potential of groundwater arsenic in Cambodia, Laos, and Thailand using artificial neural network. Water Res 45(17):5535–5544. https://doi.org/10.1016/j.watres.2011.08.010

Chowdhury A (2016) Assessment of spatial groundwater level variations using geo-statistics and GIS in Haringhata block, Nadia district, West Bengal. Int J Res Eng Technol 5(5):276–280

Coppola E, Szidarovszky F, Poulton M, Charles E (2003) Artificial neural network approach for predicting transient water depth in a multilayered groundwater system under variable state, pumping, and climate conditions. J Hydrol Eng 6:348–360

Davoodi Moghaddam D, Rezaei M, Pourghasemi HR, Pourtaghie ZS, Pradhan B (2015) Groundwater spring potential mapping using

bivariate statistical model and GIS in the the Taleghan Watershed Iran. Arabian J Geosci 8(2):913–929

Doll P, Schmied HM, Schuh C, Portmann FT, Eicker N (2014) Global-scale assessment of groundwater depletion and related groundwater abstractions: combining hydrological modeling with information from well observations and GRACE satellites. Water Resour Res 50(7):5698–5720. https://doi.org/10.1002/2014WR015595

Gandhi FR, Patel JN (2022) Groundwater potentiality deciphering and sensitivity study using remote sensing technique and fuzzy approach. Acta Geophys 70:265–282. https://doi.org/10.1007/s11600-021-00711-5

Ghadimi F (2015) Prediction of heavy metals contamination in the groundwater of Arak region using artificial neural network and multiple linear regression. J Tethys 3(3):203–215

Ghadimi F, Ghomi M (2013) Geochemical evaluation of drinking water in Arak City, Iran. Iran J Earth Sci 5(2):54–65. https://journals.iau.ir/article_522868.html

Ghadimi F, Ghomi M, Azimi R (2016) Sources of nitrate and bromide contaminants of groundwater in alluvial aquifer of Arak. Iran J Tethys 4(2):100–115

Gholami V, Booij MJ, Nikzad Tehrani E, Hadian MA (2018) Spatial soil erosion estimation using an artificial neural network (ANN) and field plot data. Catena 163:210–218. https://doi.org/10.1016/j.catena.2017.12.027

Golkarian A, Rahmati O (2018) Use of a maximum entropy model to identify the key factors that influence groundwater availability on the Gonabad Plain. Iran Environ Earth Sci 77:369

Gualbert HP, Essink O (2001) Improving fresh groundwater supply-problems and solutions center of hydrology (ICHU), institute of earth science. Ocean Coast Manag 44(5–6):429–449

Haghizade A, Moghaddam D, Pourghasemi H (2017) GIS-based bivariate statistical techniques for groundwater potential analysis. J Earth Syst Sci 126:109

Hamed Y, Elkiki M, Al Gahtani OS (2015) Prediction of future groundwater level using artificial neural network Southern Riyadh KSA (case study). Eighteenth International Water Technology Conference IWTC18 5(2): 149–169

Ioannis N, Daliakopoulos P, Coulibaly I, Tsanis K (2005) GPMS forecasting using artificial neural networks. J Hydrol 309:229–240

Kamasi M, Sharghi S, Nourani V (2016) Identification of factors affecting groundwater level reduction using wavelet-entropy criterion (Case Study: Silakhor Plain Aquifer). Hydrogeomorph 9 (4):63–86

Karthikeyan L, Kumar ND, Graillot D, Gaur S (2013) Prediction of ground water depth in the uplands of a tropical coastal riparian wetland using artificial neural networks. Water Resour Manage 27:871–883

Keykhosravi SS, Nejadkoorki F, Amintoosi M (2019) Estimation of artificial neural networks accuracy in anticipation of the dust of the Sabzevar cement factory. J Res Environ Health 5(1):43–52

Khazaz L, Oulidi HJ, El Moutaki S, Ghafiri A (2015) Comparing and evaluating probabilistic and deterministic spatial interpolation methods for groundwater level of Haouz in Morocco. Geog Inf Syst 7:631–642

Krishna B, Satyajit Rao YR, Vijaya T (2008) Modeling GPMSs in an urban coastal aquifer using artificial neural networks. Hydrol Processes 22:1180–1188

Lallahem S, Mania J, Hani A, Najjar Y (2005) On the use of neural networks to evaluate GPMSs in fractured media. J Hydrol 307(1–4):92–111

Lee S, Song KY, Kim Y, Park I (2012) Regional groundwater productivity potential mapping using a geographic information system (GIS) based artificial neural network model. Hydrogeol J 20:1511–1527

Lee S, Hyun Y, Lee M (2019) Groundwater potential mapping using data mining models of big data analysis in Goyang-si. South Korea Sustain 11:1678

Mohammadi Ghaleni M, Ebrahimi K, Araghinejad S (2011) Groundwater quantity and quality evaluation: a case study for Saveh and Arak aquifers. Water Soil Sci 21(2):93–108

Nadiri AA, Sadeghfam S, Gharekhani M, Khatibi R, Akbari E (2018) Introducing the risk aggregation problem to aquifers exposed to impacts of anthropogenic and geogenic origins on a modular basis using 'risk cells.' J Environ Manage 217:654–667. https://doi.org/10.1016/j.jenvman.2018.04.011

Naghibi SA, Ahmadi K, Daneshi A (2017a) Application of support vector machine, random forest, and genetic algorithm optimized random forest models in groundwater potential mapping. Water Resour Manage 31:2761–2775. https://doi.org/10.1007/s11269-017-1660-3

Naghibi SA, Moghaddam D, Kalantari B, Pradhan B, Kisi O (2017b) A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater well potential mapping. J Hydrol 548:471–483

Nofal ER, Amer MA, El-Didy SM, Fekry AM (2015) Delineation and modeling of seawater intrusion into the Nile Delta Aquifer: a new perspective. Water Sci 29(2):156–166

Nordqvist R, Gustafsson E, Andersson P, Thur P, AB G (2008) Groundwater flow and hydraulic gradients in fractures and fracture zones at Forsmark and Oskarshamn. SKB Rapport, pp 1–69. https://www.skb.com/publication/1784919

Nourani V, Asgharimoghaddam AA, Nadiri AO, Singh VP (2008) Forecasting spatiotemporal water depth of Tabriz aquifer. Trends Appl Sci Res 3(4):319–329

Priyanka BN, Mahesha A (2015) Parametric studies on saltwater intrusion into coastal aquifers for anticipate sea level rise. Aquat Procedia 4:103–108

Rahmati O, Pourghasemi HR, Melesse AM (2016) Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran region. Iran Catena 137:360–372. https://doi.org/10.1016/j.catena.2015.10.010

Rahmati O, Naghibi SA, Shahabi H, Bui DT, Pradhan B, Azareh A et al (2018) Groundwater spring potential modeling: Comprising the capability and robustness of three different modeling approaches. J Hydrol 565:248–261

Rakhshandehroo G, Akbari H, Afshari Igder M, Ostadzadeh E (2017) Long-term groundwater level forecasting in shallow and deep wells using wavelet neural networks trained by an improved harmony search algorithm. J Hydrol Eng 23(2):04017058

Roshni T, Jha MK, Deo RC, Vandana A (2019) Development and evaluation of hybrid artificial neural network architectures for modeling spatio-temporal groundwater fluctuations in a complex aquifer system. Water Resour Manage. https://doi.org/10.1007/s11269-019-02253-4

Sahour H, Gholami V, Vazifedan M (2020) A comparative analysis of statistical and machine learning techniques for mapping the spatial distribution of groundwater salinity in a coastal aquifer. J Hydrol 591:125321. https://doi.org/10.1016/j.jhydrol.2020.125321

Solomatine DP, Ostfeld A (2008) Data-driven modelling: some past experiences and new approaches. J Hydroinf 10(1):3–22

Sreekanth PD, Geethanjali N, Sreedevi PD, Ahmed S, Ravi Kumar N, Kamala Jayanthi PD (2009) Forecasting groundwater level using artificial neural networks. Curr Sci 96(7):933–939

Thapa R, Gupta S, Guin S, Kaur H (2017) Assessment of groundwater potential zones using multi-influencing factor (MIF) and GIS: a case study from Birbhum district, West Bengal. Appl Water Sci 7:4117–4131

Vaheddoost B, Aksoy H (2018) Interaction of groundwater with Lake Urmia in Iran. Hydrol Processes 32(21):3283–3295

Wei HF, Ledoux E, de Marsily G (1990) Regional modelling of groundwater flow and salt and environmental tracer transport in deep aquifers in the Paris basin. J Hydrol (amsterdam) 120:341–358

Wu W, Dandy GC, Maier HR (2014) Protocol for developing ANN models and their application to the assessment of the quality of the ANN model development process in drinking water quality modeling. Environ Modell Software 54:108–127. https://doi.org/10.1016/j.envsoft.2013.12.016

Zarafshan P, Etezadi H, Javadi S, Roozbahani A, Hashemy SM, Payam Zarafshan P (2022) Comparison of machine learning models for predicting groundwater level, case study: Najafabad region. Acta Geophys. https://doi.org/10.1007/s11600-022-00948-8

Zhang M (2001) Information-statistics evaluation on the effects of ground water buried depth to upper soil and groundwater salinity. China postdoctoral preceding science press, Beijing, China, pp 221–224