



Prediction models for groundwater quality parameters using a multiple linear regression (MLR): a case study of Kermanshah, Iran

Parisa Dargahi¹ · Simin Nasser^{1,2} · Mahdi Hadi² · Ramin Nabizadeh Nodehi¹ · Amir Hossein Mahvi¹

Received: 15 March 2021 / Accepted: 31 August 2022 / Published online: 23 December 2022
© The Author(s), under exclusive licence to Tehran University of Medical Sciences 2022

Abstract

Groundwater is one of the major sources of exploitation in arid and semiarid regions. Spatial and temporal quality distribution is an important factor in groundwater management. Thus for protecting groundwater quality, data production on spatial and temporal distribution is essential. The present study has applied multiple linear regression (MLR) techniques to predict the fitness of groundwater quality in Kermanshah province, west of Iran. The parameters examined were Total dissolved solids (TDS), Total hardness (TH), Sodium adsorption ratio (SAR). The quality variables were modelled by MLR. Finally, the performance of the models was assessed using the coefficient of determination (R^2). The relationship between parameters by MLR showed that TDS and water quality parameters in semi-deep wells and aquifers had a strong positive correlation ($r=0.94$, $r=0.98$) and there was a strong positive significant correlation between SAR and water quality parameters in deep wells and aquifers ($r=0.98$, $r=0.99$). Also, TH and water quality parameters in all water sources had a strong positive correlation ($r=1$). The MLR model could serve as an alternative and cost-effective tool for groundwater quality prediction where there is limitation in laboratory facilities, trained expertise or time. Consequently, the usefulness of these linear regression equations in predicting the groundwater quality is an approach, which can be applied in any other locations.

Keywords Correlation coefficients · Groundwater quality · Multiple linear regressions (MLR) · Modeling

Introduction

Water is one of the most essential and vital resources for human life [1]. According to the WHO in 2017, approximately 2 billion people worldwide were denied access to safe and healthy water [2]. Groundwater is the most important source for various uses, including: agriculture, drinking, and industrial uses [3]. Approximately one-third of the world's population uses groundwater for drinking, and the most important reasons are the nonavailability of potable surface water and a general belief that groundwater is purer

and safer than surface water, due to the protective qualities of the soil cover [4, 5].

With the advent of industrial development in recent years, the large-scale use of synthetic fertilizers for agricultural production and the use of pesticides and insecticides for agriculture have created serious concern about the susceptibility of groundwater contamination. Changes in groundwater quality are related to rock–water contact and oxide-reduction reactions during water percolation across aquifers. In addition to these mechanisms, waterborne contaminants, toxic and non-toxic pollutants are the major factors of water quality that are moved from the recharge area to the discharge area from groundwater aquifers [6, 7].

Assessment of water quality is one of the most major issues in groundwater studies. Evaluation and monitoring of groundwater quality are vital for the sustainable use of these resources [8]. One of the essential approaches to quality control of water resources is to obtain an effective model for predicting groundwater quality [9]. In this study, using statistical methods, we tried to predict the quality characteristics of groundwater in the areas of Kermanshah province. Similar studies have been conducted in this field by Conglian

✉ Simin Nasser
naserise@tums.ac.ir

✉ Mahdi Hadi
m-hadi@sina.tums.ac.ir

¹ Department of Environmental Health Engineering, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

² Center for Water Quality Research (CWQR), Institute for Environmental Research(IER), Tehran University of Medical Sciences, Tehran, Iran

et al. [10], Zouheira et al. [11], Yesilnacar et al. [12], Kisi et al. [13]. Common models in this field are regression models, time series, and new models such as neural networks and genetic algorithms, and etc. The multiple regression model due to extracting all data from the data set at the same time has a better ability to understand the problem and another advantage of this model is its flexibility and proper performance against normal distribution variables.

The purpose of this study was to present multiple regression models to predict the qualitative parameters; Total dissolved solids (TDS), Sodium adsorption ratio (SAR), Total hardness (TH) of Kermanshah groundwater resources in the west of Iran. Investigation of the Sodium adsorption ratio (SAR) parameter is essential in soil management and stability in agricultural lands. Total dissolved solids (TDS) and Total hardness (TH) are also effective parameters in creating the taste. Due to the importance of these three parameters, some suitable models were presented to predict these parameters given the type of water resource (deep wells, semi-deep wells, springs and aquifers).

Materials and methods

Case study

Kermanshah is a province in the west of Iran. It has an area of about $24,640 \text{ km}^2$ and is located at the longitude of $N33^{\circ}37'–35^{\circ}17'$ and latitude of $E 45^{\circ}20'$ to $48^{\circ}1'$ (Fig. 1). The study area altitude ranges from the lowest of 2100 m to the highest of 3357 m. The descriptive statistics of the physicochemical parameters for the groundwater parameters including Calcium (Ca), pH, Chlorine (Cl), Magnesium (Mg), Sodium (Na), Sodium percent (%Na), Electrical conductivity (EC), Sulfate (SO_4), Total hardness (TH), Carbonate (CO_3), Sodium adsorption ratio (SAR), Bicarbonate (HCO_3), Cations, Anions, Total dissolved solids (TDS) are summarized in Table 1. To prepare the spatial distribution map for each parameter, the water quality information of 142 wells (Fig. 2) were analysed.

Multiple linear regressions (MLR)

Multivariate statistical analysis is widely used to test the reliability of different processes that affect the mineralization of the groundwater aquifer system [4]. Multiple regressions have the general purpose of learning more about the relationship between one or more independent or predictor

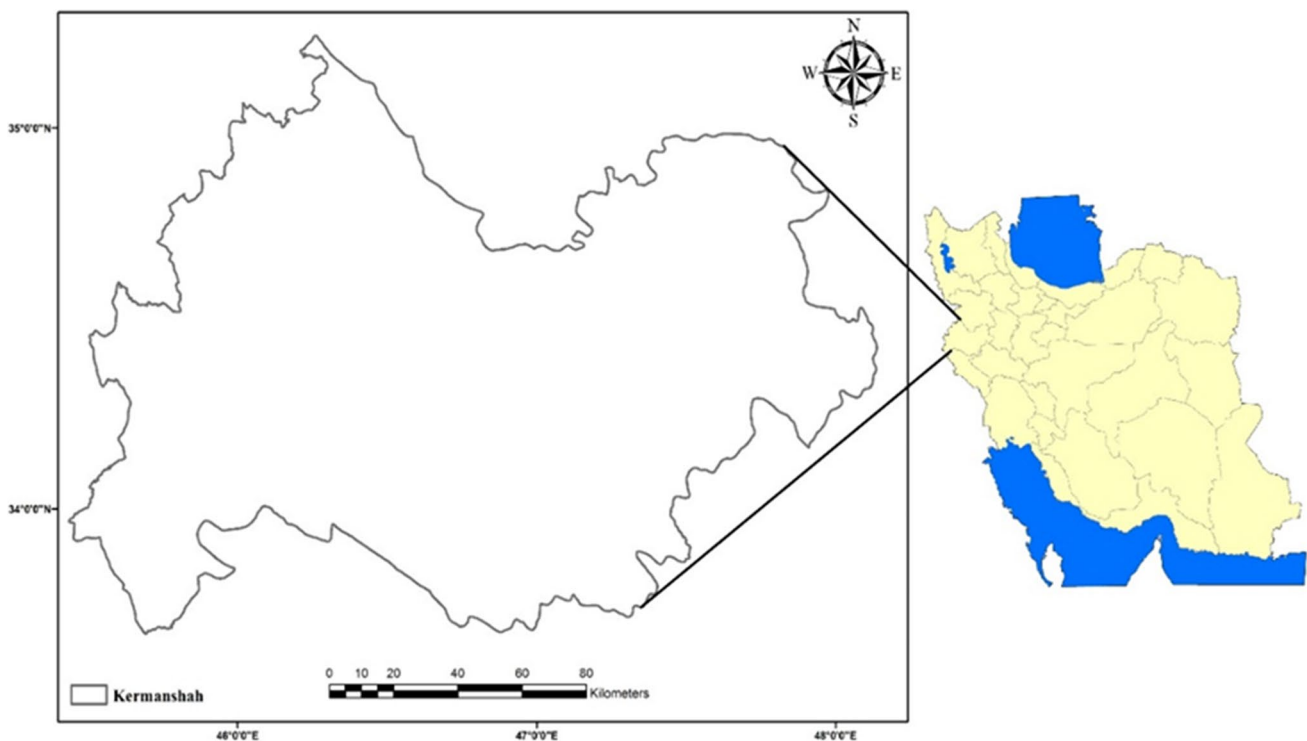
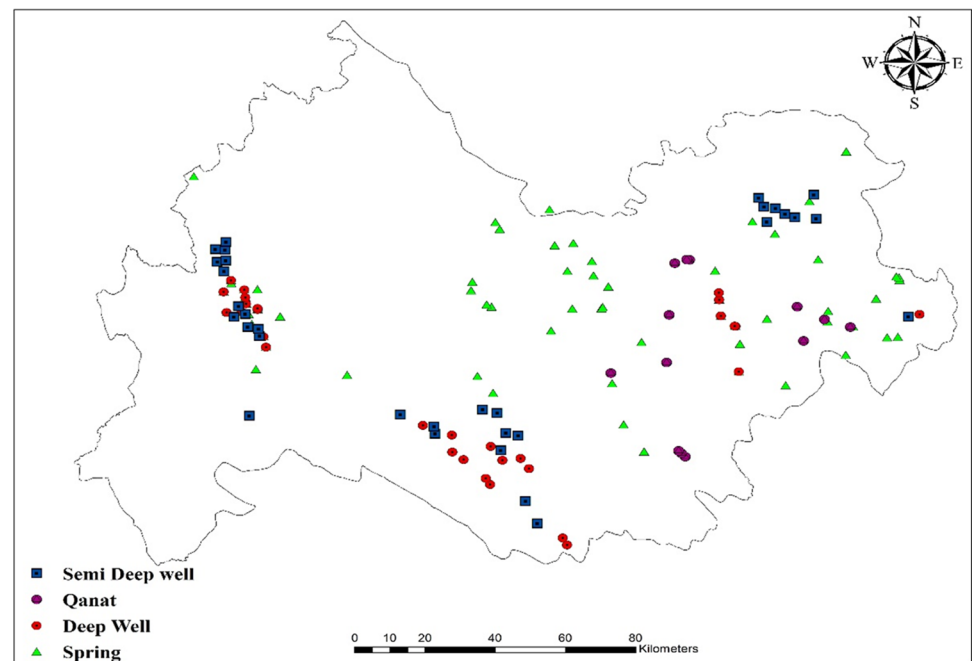


Fig. 1 Position of the study area

Table 1 Summary statistics of chemical compositions of major ions (mg/l) in the groundwater's of the study area

Compositions	Unit	Minimum	Maximum	Average	SD*	SE**
SO ₄	mg/l	3.85	633.11	52.98	59.5	2.69
Na	mg/l	0.69	179.4	11.22	13.49	0.61
Mg	mg/l	0.74	108.22	20.19	11.91	0.54
Ca	mg/l	3.4	196	58.64	18.76	0.84
Cl	mg/l	1.78	5147.5	16.4	104.73	3.39
TH	mg Ca CO ₃ /l	85.5	790	228.88	78.3	3.5
pH	–	0.3	83	7.88	2.78	0.13
TDS	mg/l	4.7	1154.5	323.78	130.31	5.97
EC	umoh/cm	0.62	3210	458.53	176.99	5.72
SAR	–	0	4.06	0.31	0.34	0.02
Cations	meq/l	2	17.4	5.066	1.9	0.09
Anions	meq/l	1.59	445.23	5.31	10.42	0.48
HCO ₃	mg/l	0	27,149.45	257.99	624.24	27.91
CO ₃	mg/l	0	240.4	3.766	13.91	1.24
Na-percent	–	0.76	100	8.59	7.11	0.32

*SD: Standard Deviation, **SE:Standard Error

Fig. 2 Local position of the wells in the study area

variables and a dependent or observed variable [10]. In other words, the multiple linear regression can predict the value of the dependent variable for the given set of predictors. The multiple regression model will match the observed dependent variable with a measured variable by changing the coefficients linearly relating to the predictors [14].

In this study the multiple regression model was used to predict groundwater quality parameters, with a 5% level of significance. The relationship among groundwater quality parameters were examined using mlr function in R software [15].

The dependent variables in this study were the groundwater TDS, SAR, TH, and the independent variables or predictors were SO₄, pH, HCO₃, Mg, Ca, Cl, Na, EC. Modeling was performed for four water resources: deep wells, semi-deep wells, springs and aquifers. The model's characteristics were evaluated using coefficient of determination (R^2) statistic.

Inverse distance weighted (IDW)

The spatial distribution for groundwater quality parameters was done with the help of a spatial analyst module in ArcGIS.10.4.1 software.

Inverse distance weighting (IDW) interpolation obviously means the conclusion that factors closer with each other are the same as those farther apart. IDW can use the calculated values surrounding the prediction position to predict a value for any unmeasured position [16].

Statistical analysis

Analysis of variance (ANOVA) has wide applicability in groundwater quality problems as a versatile diagnostic tool. ANOVA tests significant differences in one or more clusters [15]. The Kruskal–Wallis rank-sum test is, a non-parametric equivalent of an ANOVA test [3]. To find the appropriate parametric or non-parametric test and to check the homogeneity of variance, Flinger test was used. This test showed a significant P-value (<0.05), so the non-parametric Kruskal–Wallis test was used for analysis of variance [17].

Results and discussion

Statistical analysis results

The ANOVA and Kruskal–Wallis test findings are presented in Table 2. The value of df (degrees of freedom) for deep well, semi-deep well, aquifer is 5 and for spring 15. The results showed that there is a significant difference between the quality parameters and the type of water source.

GIS interpolation model

IDW model was used as the interpolation model to produce GIS maps as presented in Fig. 3.

The lowest and the maximum output in IDW was 8.89086 and 901.19 mg/l for TDS; 0.007 and 4.02 mg/l for SAR and 15.80 and 1966.39 mg/l for TH, respectively. All of the parameters had low values in the north and the east of the study area. In contrast, the maximum value of parameters was located in the south and the west of the study area; so the north and the east part of the study area had better quality than the south and the west part of the study area. One of the most important reasons for the decrease in quality in these areas is the use of nitrate fertilizers and the dissolution of calcareous minerals.

Table 2 ANOVA and Kruskal–Wallis results for the TDS, SAR, TH models in groundwater for the study area

Parameters	Variation		
	Taste	df	P-value
TDS			
Deepwell	Kruskal–Wallis	5	$<2.2e-16$
Semi-deep well	ANOVA	5	4.58e-08
Spring	Kruskal–Wallis	15	$<2.2e-16$
Aquifer	Kruskal–Wallis	5	$<2.2e-16$
SAR			
Deepwell	Kruskal–Wallis	5	$<2.2e-16$
Semi-deep well	Kruskal–Wallis	5	2.438e-11
Spring	Kruskal–Wallis	15	$<2.2e-16$
Aquifer	ANOVA	5	4.47e-05
TH			
Deepwell	Kruskal–Wallis	5	$<2.2e-16$
Semi-deep well	ANOVA	5	5.05e-10
Spring	Kruskal–Wallis	15	$<2.2e-16$
Aquifer	Kruskal–Wallis	5	1.104e-05

Multiple linear regression model (MLR)

The MLR model is useful in discovering the association between various independent and dependent variables. MLR (TDS, SAR, TH) models for four water resources: deep wells, semi-deep wells, springs and aquifers were made using R software presented in Fig. 4.

MLR analysis of TDS model

TDS is one of the most important parameters in assessing the suitability of water for irrigation [18] and for overall groundwater quality assessment [19]. TDS is a measure of the combined concentration of cations and anions. In natural water, dissolved solids consist of inorganic salts, small amounts of organic matter, and dissolved materials. Dissolved solids are mainly due to carbonates, chlorides, sulfates, nitrates, phosphates, Ca, Mg, Na, K, Fe, Mn, etc. [5].

TDS of the Groundwater is mainly due to the vegetable decay and the disposal of effluents from industries. TDS value of 500 mg/l is the desirable limit, and water containing more than 500 mg/l TDS causes gastrointestinal irritation [20]. The high value of TDS influences the taste, hardness, and corrosive property of the water [21].

In Fig. 4, the predicted values are graphically showed with the observed data for the models. There are good relationships between predicted values and the observed data for TDS model. The estimated R^2 value and P values of this model are represented in Table 3. Also, the most critical factor in determining the success of the model is, the adjusted R square, in comparison with multiple R or R square. The

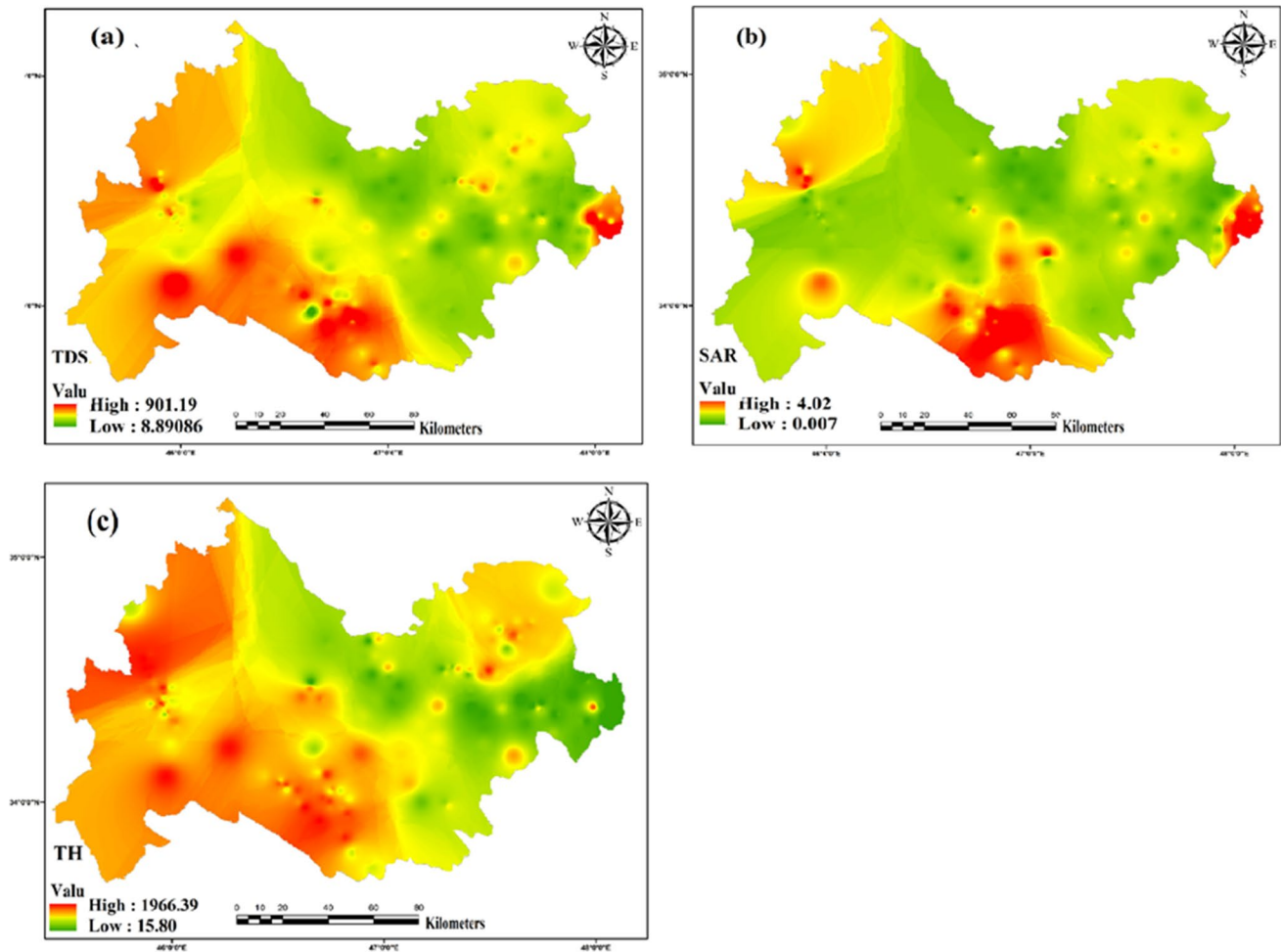


Fig. 3 Interpolated maps of the groundwater quality parameters generated using by IDW. a TDS: b SAR: c TH

adjusted R square is 0.86 for the deep well, 0.94 for the semi-deep wells, 0.88 for the springs and 0.94 for the aquifers.

The P-value is less than $2.2e - 16$. Considering the P-value of the model, it is statistically significant. The independent variables such SO_4 , pH, HCO_3 , Mg, Ca, Cl, Na, EC were significant in predicting TDS value. Independent variables describe the variance of TDS by 86% for the deep well, 94% for the semi-deep wells, 88% for the springs, and 94% for the aquifers.

Adhikari et al. (2009) studied statistical approaches for hydrogeochemical characterization of groundwater in west Delhi, India. The study showed a good correlation between water quality parameters and also showed that multiple regression models can predict quality parameters at 5% level of significance [22].

The results obtained in the present study are Comparison with the previous studies (Pan et al. (2018) and Zouheira et al. (2017)). According to their reports, The high R square shows that about 98% of the total variations in the TDS have been explained by these variables [10, 11]. The value of R

square is 0.98, showing that about 98% of the total variations in the TDS can be accounted for the independent variables. TDS models provide an accurate prediction of quality parameters with considerably high values of R^2 .

Also, This results is in contrast with the ones expressed by Kadam et al. (2019). According to their reports, multiple R^2 is 1 and adjusted R^2 is 1. Also the 'p' value is less than $2.2e - 16$.

The ANN is appropriate compared to the MLR model. ANN models counterpart convincingly fit quality. MLR modelling technique is based on the simple least square method; whereas, the ANN model imitates the functioning of the human being intelligence. According to their report, the ANN model would become more beneficial in the prediction of water quality [23].

Also, Civelekoglu et al. (2007) indicated that ANN modeling appears to be a strong tool in situations where the relations between variables are nonlinear [24].

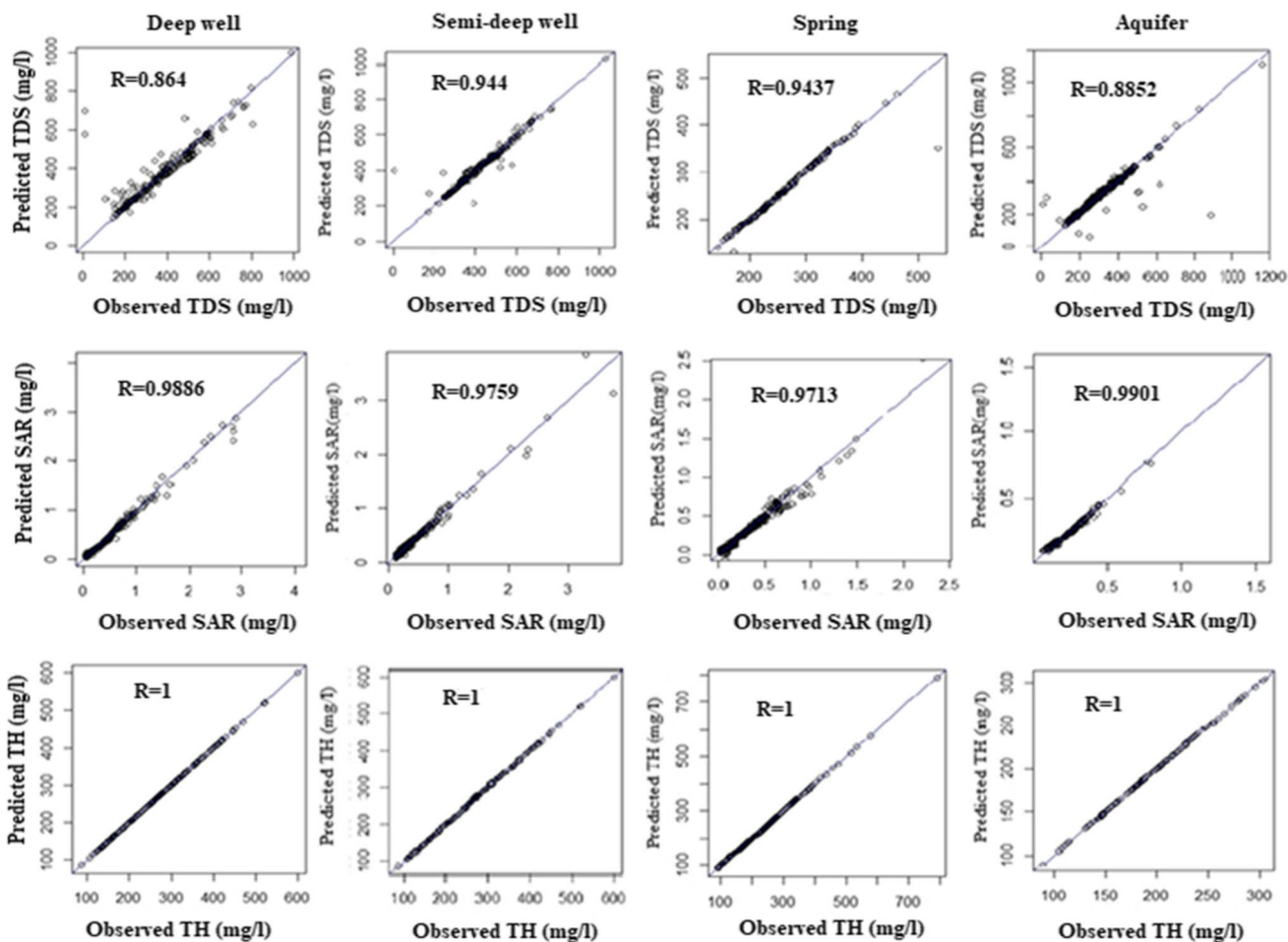


Fig. 4 Multiple linear regressions TDS, SAR, TH models in groundwater for the study area

Table 3 Performance multiple linear regressions TDS models in groundwater for the study area

Water resources	TDS			
	MLR equations	R ²	R ² _{adj}	P-value
Deep well	TDS = 0.16 SO ₄ + 3.44 Mg + 2.49 Na + 2.66Ca + 0.1 EC + 10.20	0.8644	0.8620	< 2.2e-16
Semi-deep well	TDS = 166.6 Mg - 0.3100 Cl + 154.4 Na + 121.7 Ca + 0.4156 EC + 0.7969 year - 1106	0.9440	0.9426	< 2.2e-16
Spring	TDS = 0.17 SO ₄ + 0.2 HCO ₃ + 0.58 Na + 0.51 EC + 3.78	0.8852	0.8840	< 2.2e-16
Aquifer	TDS = 0.66 EC - 5.54	0.9437	0.9403	< 2.2e-16

The results of the multiple regression model can be used as a positive predictive tool for determining the chemistry of groundwater if the dependent variable TDS is measured at every location. The proposed TDS model can be utilized for estimating TDS content in groundwater obtained from such an area. Consequently, the MLR model can serve as an alternative and cost-effective tool for groundwater quality prediction in the circumstances,

where trained expertise and time constraints and the field data are favourable.

MLR analysis of SAR model

Sodium concentration is one of the important parameters in the classification of water irrigation. Soils containing a large proportion of sodium with carbonate as predominant anion are termed alkali soils and those with chloride or sulfate, as

predominant anion, are termed as saline soils, These affect the growth of the plant [21]. For calculating SAR, Na, Mg, and Ca are also needed [13].

The systematic calculation of the correlation coefficient between water quality variables and regression analysis provides an indirect means for rapid monitoring of water quality [21]. The estimated R² value and P values of this model are represented in Table 4. The most critical factor in determining the success of the model is, the adjusted R square, in comparison with multiple R or R square. The adjusted R square is for the deep well and the aquifers 0.98 and for the semi-deep wells and the springs 0.97. The P-value is less than 2.2e – 16.

Based on the explanation coefficient, the independent variables in the deep well and the aquifers were found to affect the SAR by 98%, 97% for the semi-deep well and the spring.

The results obtained in the present study are Comparison with the previous studies Tabari et al. (2012) [25]. According to their study Correlation coefficient and standard error are 0.74 and 1.35 respectively. Measured and predicted values for regression models, fit relatively well, but at a high SAR value, the amount of differences of the measured values and the model is increased. Finally, using comparison between statistical indicators in the artificial neural network and regression model in can be observed superiority of the artificial neural network model in simulated and predicted values of SAR. Therefore the artificial neural network is better performance than the regression model for predicted values of SAR [25].

Also, the results obtained in the present study are Comparison with the previous studies Kisi et al. (2018) [13]. the quality variables were modeled by simple ANFIS and the ANFIS trained by evolutionary algorithms. Finally, the models’ performances were evaluated using determination coefficient (R²). There used the ANFIS model to estimate sodium adsorption ratio SAR. The results indicate that the Na (0.97 and 0.92) and Cl show the highest correlations (0.97 and 0.82) with EC and SAR. Also, Mg, Cl and Ca are the most appropriate variables for TH [13]. Also, the correlation coefficient between the most effective variables and outputs is positive. It is noteworthy that potassium (0.24, 0.23 and 28) and show the lowest correlation with EC, SAR and TH, respectively. the results showed that the ANFIS model could be useful tools to compute and predict the groundwater quality variables [13]. The results of the multiple regression SAR model can be used as a positive predictive tool in determining the groundwater quality parameters.

MLR analysis of TH model

The hardness of water is mainly based on the evaluations of calcium and magnesium. Calcium and magnesium, the two most dominant cations play a major role in determining the hardness of the water. It is noteworthy that there are also some other variables in water such as aluminium, iron, manganese, etc.; but, calcium and magnesium are the most effective variables to the hardness of water [13]. Hardness may be due to the persence of calcium and magnesium salt

Table 4 Performance multiple linear regressions SAR models in groundwater for the study area

Water resources	SAR			
	MLR equations	R ²	R ² _{adj}	P-value
Deep well	SAR = -0.000342 SO ₄ - 0.00134 Mg - 0.00155 Cl + 0.0283 Na - 0.000736 Ca - 0.000117 EC + 0.156	0.9886	0.9862	< 2.2e-16
Semi-deep well	SAR = -0.004 Mg - 0.002 Ca + 0.03 Na - 0.002 Cl + 0.0017 year - 1.99	0.9759	0.9753	< 2.2e-16
Spring	SAR = -0.000212 SO ₄ - 0.000172 Mg + 0.000476 Cl + 0.0291 Na - 0.000155 EC + 0.0998	0.9713	0.9710	< 2.2e-16
Aquifer	SAR = -0.000399 HCO ₃ - 0.00199 Cl + 0.0306 Na + 0.116	0.9901	0.9895	< 2.2e-16

Table 5 Performance multiple linear regressions TH models in groundwater for the study area

Water resources	TH			
	MLR equations	R ²	R ² _{adj}	P-value
Deep well	TH = 409 Mg - 0.00000000000000144 Cl + 0.00000000000000131 Na + 250 Ca + 0.00000000000000303	1	1	< 2.2e-16
Semi-deep well	TH = 410 Mg + 250 Ca + 0.00000000000000159 year - 0.0000000000219	1	1	< 2.2e-16
Spring	TH = 409.8 Mg + 250 Ca + 0.0000000000001682	1	1	< 2.2e-16
Aquifer	TH = 410 Mg + 250 Ca - 0.00000000000000694	1	1	< 2.2e-16

from detergents and soaps used for laundering on the bank of the water body precipitated as calcium carbonate.

The maximum allowable limit of TH for drinking purpose is 500 mg/l, and the most desirable limit is 100 mg/l (per WHO international standard). For total hardness, the most desirable limit is 80–100 mg/l. Groundwater exceeding the limit of 300 mg/l is considered to be very hard [4]. The estimated R^2 value and P values of this model are represented in Table 5. Also, the most critical factor in determining the success of the model is, the adjusted R square, in comparison with multiple R or R square. The adjusted R square is 1 for the deep well, the semi-deep wells, the springs and for the aquifers. The P-value is less than $2.2e - 16$.

According to the results, it can be said that the independent variables in deep wells, semi-deep wells, springs, and aquifer with a coefficient of explanation of 100% are effective on TH. The multiple R^2 value (100%) indicates that 100% of the variability in TH could be ascribed to the combined effect of SO_4 , pH, HCO_3 , Mg, Ca, Cl, Na, EC.

The results obtained in the present study are Comparison with the previous studies Kadam et al. (2019), Kisi et al. (2019); indicated that ANFIS and ANN models could be used as useful tools to predict TH value the groundwater quality variables [13, 23].

Mekparyup et al. (2013) also indicated that all regression coefficients are significant and Highly positive correlation between the response variable and the predictor variables [26].

Therefore the best performances TH model for the groundwater quality parameters respect to the other estimating methods.

Conclusions

This aim of the study was determination of groundwater quality and the relationship between variability of groundwater quality in Kermanshah province, west of Iran. The results of spatial distribution for each parameter by IDW showed that the north and the east part of the study area had better quality in groundwater resources in comparison with the south and the west part the study area. The relationship between parameters by MLR showed that TDS and water quality parameters in semi-deep wells and aquifers had a strong positive correlation ($r = 0.94$, $r = 0.98$) and there was a strong positive significant correlation between SAR and water quality parameters in deep wells and aquifer ($r = 0.98$, $r = 0.99$). Also, TH and water quality parameters in all water sources had a strong positive correlation ($r = 1$). Consequently, the MLR model could serve as an alternative and cost-effective tool for groundwater quality prediction in the circumstances, where trained expertise and time constraints and the field data are favourable.

The present study has been done in annual scales, it is suggested that in future researches in different time scales including daily and monthly study.

Acknowledgements The authors expressed their appreciate to Tehran University of Medical Sciences. The authors are grateful to Urban Water and Wastewater Office, about providing information. The authors also would like to gratefully thank the anonymous reviewers for their invaluable comments to improve the quality of the paper.

Declarations

Conflict of interest The authors declare that They have no conflict of interest.

References

1. Rajankar PN, Gulhane SR, Tambekar DH, Ramteke DS, Wate SR. Water quality assessment of groundwater resources in Nagpur Region (India) based on WQI. *E-J Chem*. 2009;6(3):905–8.
2. Masroor K, Kermani M, Gholami M, Fanaei F, Arfaeinia H, Nemati S, et al. Development and implementation of water safety plans for groundwater resources in the southernmost city of West Azerbaijan Province. *Iran J Environ Health Sci Eng*. 2020;18:629–37.
3. Mogaji KA, Lim HS, Abdullah K. Modeling of groundwater recharge using a multiple linear regression (MLR) recharge model developed from geophysical parameters: a case of groundwater resources management. *Environ Earth Sci*. 2015;73(3):1217–30.
4. Arumugam K, Elangovan K. Hydrochemical characteristics and groundwater quality assessment in Tirupur region, Coimbatore district, Tamil Nadu, India. *Environ Geol*. 2009;58(7):1509.
5. Varol S, Davraz A. Evaluation of the groundwater quality with WQI (Water Quality Index) and multivariate analysis: a case study of the Tefenni plain (Burdur/Turkey). *Environ Earth Sci*. 2015;73(4):1725–44.
6. Annapoorna H, Janardhana MR. Assessment of groundwater quality for drinking purpose in rural areas surrounding a defunct copper mine. *Aquat Procedia*. 2015;1(4):685–92.
7. Sattari MT, Mirabbasi R, Sushab RS, Abraham J. Prediction of groundwater level in Ardebil plain using support vector regression and M5 tree model. *Groundwater*. 2018;56(4):636–46.
8. Taheri K, Taheri M, Parise M. Impact of intensive groundwater exploitation on an unprotected covered karst aquifer: a case study in Kermanshah Province, western Iran. *Environ Earth Sci*. 2016;75(17):1221.
9. Kovačič G, Ravbar N. Analysis of human induced changes in a karst landscape—the filling of dolines in the Kras plateau, Slovenia. *Sci Total Environ*. 2013;447:143–51.
10. Pan C, Ng KTW, Fallah B, Richter A. Evaluation of the bias and precision of regression techniques and machine learning approaches in total dissolved solids modeling of an urban aquifer. *Environ Sci Pollut Res*. 2019;26(2):1821–33.
11. Ali ZI, Frika Y, Ghzel LL, Zairi M. Hydrogeochemical characteristics prediction using multiple linear regression: case study on unconfined aquifer in northeastern Tunisia. *Arab J Geosci*. 2017;10(17):382.
12. Yesilnacar MI, Sahinkaya E. Artificial neural network prediction of sulfate and SAR in an unconfined aquifer in southeastern Turkey. *Environ Earth Sci*. 2012;67(4):1111–9.

13. Kisi O, Azad A, Kashi H, Saeedian A, Hashemi SAA, Ghorbani S. Modeling groundwater quality parameters using hybrid neuro-fuzzy methods. *Water Resour Manag.* 2019;33(2):847–61.
14. Nathan NS, Saravanane R, Sundararajan T. Application of ANN and MLR Models on groundwater quality using CWQI at Lawspet, Puducherry in India. *J Geosci Environ Prot.* 2017;5(03):99.
15. Johnson TD, Belitz K. Assigning land use to supply wells for the statistical characterization of regional groundwater quality: correlating urban land use and VOC occurrence. *J Hydrol.* 2009;370(1–4):100–8.
16. Ikem A, Osibanjo O, Sridhar M, Sobande A. Evaluation of groundwater quality characteristics near two waste sites in Ibadan and Lagos, Nigeria. *Water Air Soil Pollut.* 2002;140(1–4):307–33.
17. Magesh N, Chandrasekar N. Evaluation of spatial variations in groundwater quality by WQI and GIS technique: a case study of Virudunagar District, Tamil Nadu, India. *Arab J Geosci.* 2013;6(6):1883–98.
18. Atta HSAF, Amer AWM, Atta SAF. Hydro-chemical study of groundwater and its suitability for different purposes at Manfalut District, Assuit Governorate. *Water Sci.* 2018;32(1):1–15.
19. Li Z, Wang G, Wang X, Wan L, Shi Z, Wanke H, et al. Groundwater quality and associated hydrogeochemical processes in Northwest Namibia. *J Geochem Explor.* 2018;186:202–14.
20. Adujo AA, Adebawole AE, Uchebulam O. Modeling groundwater total dissolved solid from derived electromagnetic data using multiple linear regression analysis: a case study of groundwater contamination. *Model Earth Syst Environ.* 2020;6(3):1863–75.
21. Saleem A, Dandigi MN, Kumar KV. Correlation-regression model for physico-chemical quality of groundwater in the South Indian city of Gulbarga. *Afr J Environ Sci Technol.* 2012;6(9):353–64.
22. Adhikary PP, Chandrasekharan H, Chakraborty D, Kumar B, Yadav B. Statistical approaches for hydrogeochemical characterization of groundwater in West Delhi, India. *Environ Monit Assess.* 2009;154(1–4):41.
23. Kadam AK, Wagh VM, Muley AA, Umrikar BN, Sankhua RN. Prediction of water quality index using artificial neural network and multiple linear regression modelling approach in Shivganga River basin, India. *Model Earth Syst Environ.* 2019;5(3):951–62.
24. Civelekoglu G, Yigit NO, Diamadopoulos E, Kitis M. Prediction of bromate formation using multi-linear regression and artificial neural networks. *Ozone: Sci Eng.* 2007;29(5):353–62.
25. Tabari H, Sabziparvar AA, Ahmadi M. Comparison of artificial neural network and multivariate linear regression methods for estimation of daily soil temperature in an arid region. *Meteorol Atmos Phys.* 2011;110(3):135–42.
26. Mekparyup J, Saithanu K, Dujjanut J. Multiple linear regression equation for estimation of daily averages solar radiation in Chonburi, Thailand. *Appl Math Sci.* 2013;7(73):3629–39.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.