



## Data Article

# Data on estimation for sodium absorption ratio: Using artificial neural network and multiple linear regressions



Majid Radfard<sup>d</sup>, Hamed Soleimani<sup>b</sup>, Samira Nabavi<sup>b</sup>,  
Bayram Hashemzadeh<sup>c</sup>, Hesam Akbari<sup>a,\*</sup>, Hamed Akbari<sup>a</sup>,  
Amir Adibzadeh<sup>a,b,c,\*\*</sup>

<sup>a</sup> Health Research Center, Life Style Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran

<sup>b</sup> Department of Environmental Health, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

<sup>c</sup> Department of Environmental Health, School of Public Health, Khoy University of Medical Sciences, Khoy, Iran

<sup>d</sup> Research Center for Health Sciences, Institute of Health, Department of Environmental Health, School of Health, Shiraz University of Medical Sciences, Shiraz, Iran

## ARTICLE INFO

*Article history:*

Received 12 March 2018

Received in revised form

25 August 2018

Accepted 31 August 2018

Available online 5 September 2018

*Keywords:*

Groundwater quality

SAR

Aras

Neural network

Multiple linear regression

## ABSTRACT

In this article the data of the groundwater quality of Aras catchment area were investigated for estimating the sodium absorption ratio (SAR) in the years 2010–2014. The artificial neural network (ANN) is defined as a system of processor elements, called neurons, which create a network by a set of weights. In the present data article, a 3-layer MLP neural network including a hidden layer, an input layer and an output layer had been designed. The number of neurons in the input and output layers of network was considered to be 4 and 1, respectively, due to having four input variables (including: pH, sulfate, chloride and electrical conductivity (EC)) and only one output variable (sodium absorption ratio). The impact of pH, sulfate, chloride and EC were estimated to be 11.34%, 72.22%, 94% and 91%, respectively. ANN and multiple linear regression methods were used to estimate the rate of sodium absorption ratio of groundwater resources of the Aras catchment area. The data of both methods were compared with the model's performance evaluation criteria, namely root mean square error (RMSE), mean absolute error (%) and

\* Corresponding author.

\*\* Corresponding author at: Health Research Center, Life Style Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran.

E-mail addresses: [Radfard.tums.ac.ir@gmail.com](mailto:Radfard.tums.ac.ir@gmail.com) (H. Akbari), [Radfard.tums.ac.ir@gmail.com](mailto:Radfard.tums.ac.ir@gmail.com) (A. Adibzadeh).

correlation coefficient. The data showed that ANN is a helpful and exact tool for predicting the amount SAR in groundwater resources of Aras catchment area and these results are not comparable with the results of multiple linear regressions.

© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

---

## Specifications table

Subject area	Chemistry
More specific subject area	Water quality and monitoring
Type of data	Tables, Figures
How data was acquired	Data on groundwater resources quality in the Aras catchment area was obtained from West Azerbaijan Water and Wastewater Company during the years 2010–2014 and was studied for estimation of sodium absorption ratio (SAR).
Data format	Raw, Analyzed
Experimental factors	The sodium absorption ratio (SAR), were analyzed according to the standards for water and wastewater treatment handbook.
Experimental features	The levels of physical and chemical parameters were determined.
Data source location	Aras, West Azerbaijan province, Iran.
Data accessibility	Data are included in this article
Related research article	A.Takdastana, M. Mirzabeygi (Radfard), M.Yousefi, A. Abbasnia, R. Khodadadia, A.H.Mahvi, D.Jalili Naghan, Neuro-fuzzy inference system Prediction of stability indices and Sodium absorption ratio in Lordegan rural drinking water resources in west Iran, Data in Breif 18(2018) 255–261.

---

## Value of the data

- The data of this article can be used to environmental management and better exploitation of groundwater resources.
  - Considering the present data, many of the sampling drinking water supply reservoirs need to pay attention to achieve Iran national water quality standards.
  - The results clearly indicate that with appropriate selection of input variables, artificial neural network and multiple linear regressions as a soft computing approach can be used to estimate water quality indices properly and reliability.
- 

## 1. Data

Two algorithms, including seven Back-propagation algorithms and Lewenberg-Markow, have been used in this data article. The Comparison of the performance of seven Back-propagation algorithms in estimating the sodium absorption ratio with the number of neurons 10 in the hidden layer has been shown in [Tables 1](#) and [2](#), indicating the comparison of the different neurons performance in the hidden layer in estimating the sodium absorption ratio using the Lewenberg-Markow algorithm. The optimized output of neural network and data performance criteria for has been shown in [Figs. 1](#) and [2](#). Also, [Fig. 3](#) indicates the actual SAR values in groundwater resources and their predicted values via multiple linear regression.

**Table 1**

Comparison of the performance of seven Back-propagation algorithms in estimating the sodium absorption ratio with the number of neurons 10 in the hidden layer.

Back-propagation algorithms	Neural network process	Evaluation of model performance			Repeat number
		R	MAE	MSE	
Trainbfg (BFGS quasi-Newton)	Training	0.883	1.34	5	132
	Test	0.819	1.33	4.54	132
	Training+Test+Validation	0.866	1.36	4.7	132
Traincgp (Polak–Ribie're conjugate gradient)	Training	0.916	1.15	3.07	39
	Test	0.855	1.2	3.16	39
	Training+Test+Validation	0.893	1.2	3.78	39
Traingd (gradient descent)	Training	0.726	13.09	197.07	52
	Test	0.736	13.89	221.88	52
	Training+Test+Validation	0.762	13.17	193.17	52
Traingda (adaptive learning rate back-propagation)	Training	0.806	1.71	7.25	500
	Test	0.85	1.54	4.58	500
	Training+Test+Validation	0.811	1.68	6.55	500
Trainscg (scaled conjugate gradient)	Training	0.898	1.23	4.13	143
	Test	0.86	1.16	2.55	143
	Training+Test+Validation	0.89	1.25	3.85	143
Traincgf (Fletcher–Powell conjugate gradient)	Training	0.868	1.4	4.9	34
	Test	0.0804	1.52	4.49	34
	Training+Test+Validation	0.85	1.43	5.16	34
Trainlm (Levenberg–Marquardt)	Training	0.906	1.05	2.77	29
	Test	0.9	0.92	1.9	29
	Training+Test+Validation	0.0901	1.08	3.52	29

**Table 2**

Comparison of the different neurons performance in the hidden layer in estimating the sodium absorption ratio using the Lewenberg–Markow algorithm.

Number of neurons	R	MAE	MSE	Repeat number
3	0.879	1.23	4.59	25
5	0.86	1.24	4.81	32
7	0.91	1.22	3.61	25
10	0.9	0.923	1.9	29
12	0.89	1.14	4.03	33
17	0.88	1.09	2.37	29
20	0.895	1.08	3.76	30
30	0.892	1.09	3.92	34

## 2. Experimental design, materials and methods

### 2.1. Study area description

Aras Catchment area is a plain located in the northern half of the East Azerbaijan province, West Azerbaijan province and Ardabil province. Extensive precipitation in this region, in addition to its impact on climate moderation, has created numerous rivers [18]. (Fig. 4).

### 2.2. Material and methods

Data on groundwater resources were collected during the years 2010–2014 and water samples were analyzed following the standard methods for examination of water and waste water [1–10] in terms of estimation of sodium absorption ratio (SAR). A two-layer neural network with a tangent-sigmoid transfer function for the hidden layer and a linear transfer function for the output layer was

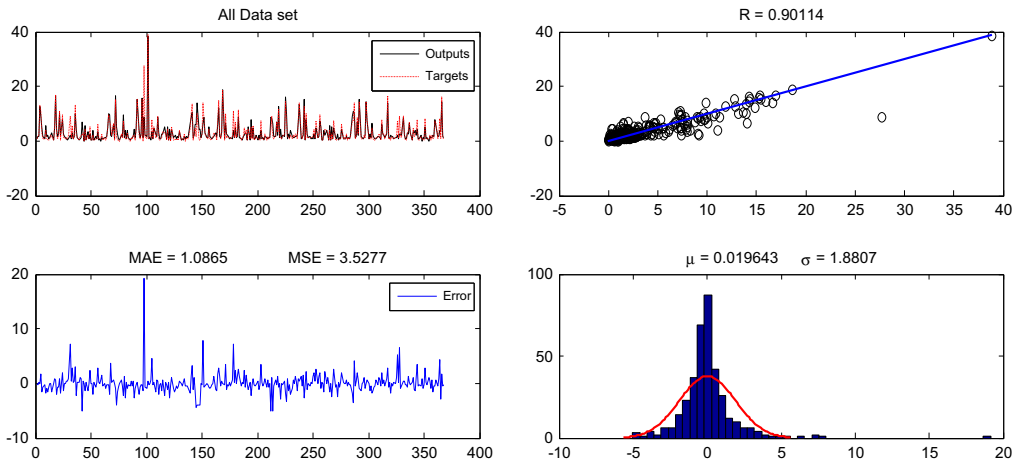


Fig. 1. Optimized neural network output and model performance criteria for all data.

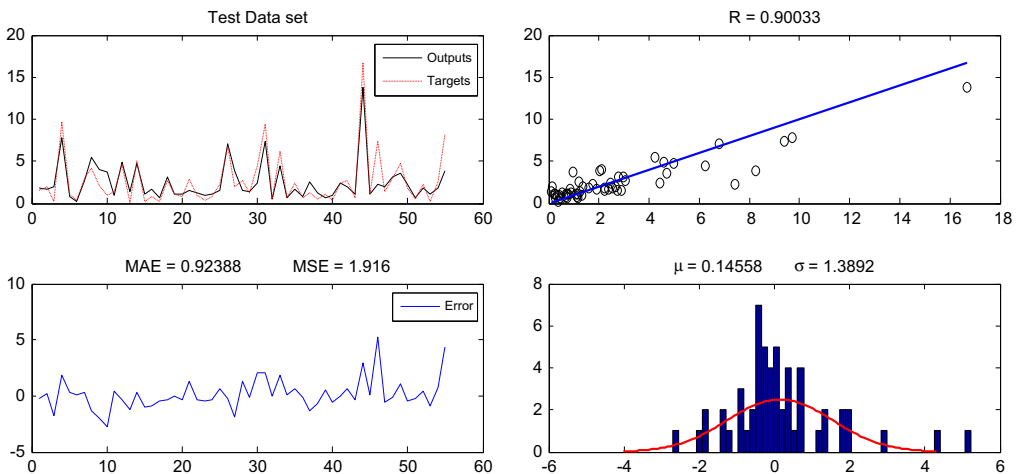


Fig. 2. Optimized neural network output and model performance criteria for test data.

used. The input parameters of the neural network included sulfate, chloride, electrical conductivity (EC) and pH, and the sodium adsorption ratio (SAR) were considered as the network output parameter. The data on these parameters were divided into training, testing and data validation. 70% of these data was used for training, 15% of data for validation and other 15% for testing. Considering that today BP neural networks have become a common tool for modeling environmental systems, so in this study, 8 BP algorithms were selected and their results were tested to obtain the best algorithm. For all algorithms, a dual layer network with a tan-sigmoid transfer function on the hidden layer and a linear transfer function in the output layer was used. In choosing the best BP algorithm, the number of neurons was considered 10. The results of the model's performance with their BP algorithm are presented in Table 1. The performance of the BP algorithm was evaluated with mean squared error (MSE), mean absolute error (MAE) and correlation coefficient ( $R$ ) between the output of the models and the actual data set. The algorithm with the least training error and the maximum correlation coefficient was selected as the most suitable algorithm. The Langenberg-Marquard algorithm (trainlm) was chosen as the best algorithm to predict the sodium adsorption ratio (SAR). To optimize the number of neurons after selecting the best BP algorithm, namely Levenberg-Marquard, the

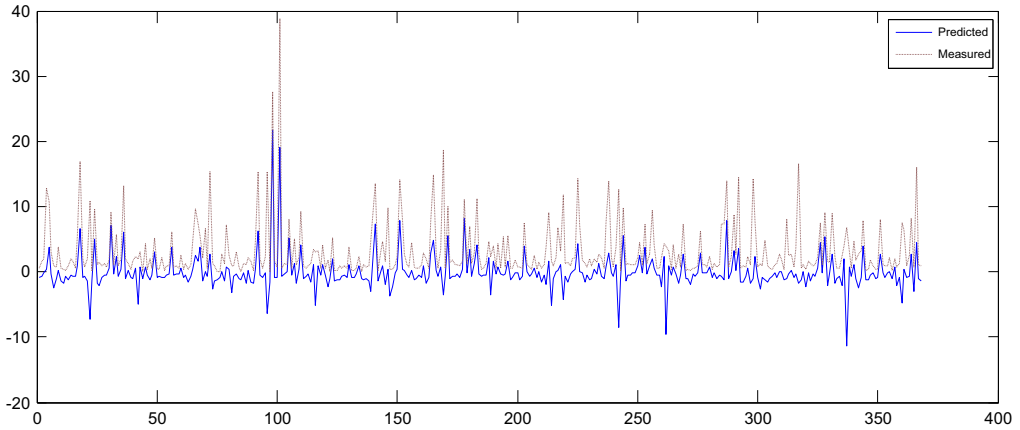


Fig. 3. Actual SAR values in groundwater resources and their predicted values with multiple linear regression.

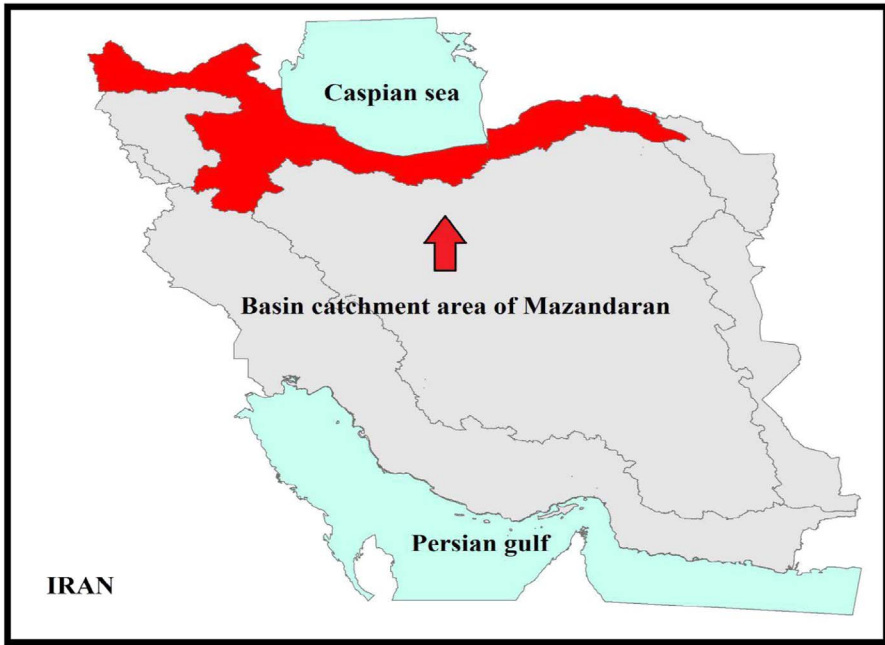


Fig. 4. Study area.

number of neurons was optimized by keeping other parameters intact. As shown in Table 2, in the number of neurons greater than the optimal number of neurons, 10, the mean square error (MSE) was not significantly altered. Therefore, all the modeling steps were done based on the number of neurons 10 and the Lewenberg-Markow algorithm to predict the sodium absorption ratio [1–20].

### Acknowledgements

The authors are grateful to Health Research Center, Life Style Institute, Baqiyatallah University of Medical Sciences, for their support in this study.

## Transparency document. Supplementary material

Transparency document associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.dib.2018.08.205>.

## References

- [1] M. Yousefi, et al., Modification of pumice with HCl and NaOH enhancing its fluoride adsorption capacity: kinetic and isotherm studies, *Hum. Ecol. Risk Assess.: Int. J.* (2018) 1–13.
- [2] A. Babaeia, et al., Data on groundwater quality, scaling potential and corrosiveness of water samples in Torbat-e Heydariyeh rural drinking water resources, Khorasan-e-Razavi province, Iran, *Data Brief* 19 (2018) 2260–2266.
- [3] Z. Heidarinejad, et al., Data on quality indices of groundwater resource for Agricultural Use in the Jolfa, East Azerbaijan, Iran, *Data Brief* (2018).
- [4] D. Jalili, et al., Data on Nitrate-Nitrite pollution in the groundwater resources a Sonqor plain in Iran, *Data Brief* (2018).
- [5] H. Akbari, et al., Data on investigating the Nitrate concentration levels and quality of bottled water in Torbat-e Heydarieh, Khorasan razavi province, Iran, *Data Brief* (2018).
- [6] M. Mirzabeygi, et al., Heavy metal contamination and health risk assessment in drinking water of Sistan and Baluchistan, Southeastern Iran, *Hum. Ecol. Risk Assess.: Int. J.* 23 (8) (2017) 1893–1905.
- [7] A. Abbasnia, N. Yousefi, A.H. Mahvi, R. Nabizadeh, M. Radfard, M. Yousefi, M. Alimohammadi, Evaluation of groundwater quality using water quality index and its suitability for assessing water for drinking and irrigation purposes; case study of Sistan and Baluchistan province (Iran), *Hum. Ecol. Risk Assess.: Int. J.* (2018), <https://doi.org/10.1080/10807039.2018.1458596>.
- [8] M. Radfard, M. Yunesian, R. Nabizadeh Nodehi, H. Biglari, M. Hadi, N. Yosefi, M. Yousefi, A. Abbasnia, A.H. Mahvi, Drinking water quality and Arsenic health risk assessment in Sistan-and-Baluchestan, Southeastern province Iran, *Hum. Ecol. Risk Assess.: Int. J.* (2018), <https://doi.org/10.1080/10807039.2018.1458210>.
- [9] H. Soleimani, et al., Data on drinking water quality using water quality index (WQI) and assessment of groundwater quality for irrigation purposes in Qorveh&Dehgolan, Kurdistan, Iran, *Data Brief* (2018).
- [10] M. Mirzabeygi, N. Yousefi, A. Abbasnia, H. Youzi, M. Alikhani, A.H. Mahvi, Evaluation of groundwater quality and assessment of scaling potential and corrosiveness of water supply networks, Iran, *J. Water Supply: Res. Technol.-AQUA jws2* (2017).
- [11] A. Takdastana, M. Mirzabeygi (Radfard), M. Yousefi, A. Abbasnia, R. Khodadadia, A.H. Mahvi, D. Jalili Naghan, Neuro-fuzzy inference system Prediction of stability indices and Sodium absorption ratio in Lordegan rural drinking water resources in west Iran, *Data Breif* 18 (2018) 255–261.
- [12] M. Mirzabeygi, M. Yousefi, H. Soleimani, A.A. Mohammadi, A.H. Mahvi, A. Abbasnia, The concentration data of fluoride and health risk assessment in drinking water in the Ardakan city of Yazd province, Iran, *Data Brief* 15 (2018) 40–46.
- [13] A. Abbasnia, M. Alimohammadi, A.H. Mahvi, R. Nabizadeh, M. Yousefi, A.A. Mohammadi, H. Pasalari H, M. Mirzabeigi, Assessment of groundwater quality and evaluation of scaling and corrosiveness potential of drinking water samples in villages of Chabahr city, Sistan and Baluchistan province in Iran, *Data Brief* 16 (2018) 182–192.
- [14] A. Abbasnia, M. Radfard, A.H. Mahvi, R. Nabizadeh, M. Yousefi, H. Soleimani, et al., Groundwater quality assessment for irrigation purposes based on irrigation water quality index and its zoning with GIS in the villages of Chabahar, Sistan and Baluchistan, Iran, *Data Brief* 19 (2018) 623–631.
- [15] A. Abbasnia, M. Ghooshani, N. Yousefi, S. Nazmara, M. Radfard, H. Soleimani, et al., Prediction of human exposure and health risk assessment to trihalomethanes in indoor swimming pools and risk reduction strategy, *Hum. Ecol. Risk Assess.: Int. J.* (2018) 1–18. <https://doi.org/10.1080/10807039.2018.1488582>.
- [16] Akazem Neisia, M. Mirzabeygi Radfard, Ghader Zeydunib, Asghar Hamzezadehc, Davoud Jalilid, Abbas Abbasniab, Mahmood Yousefie, Rouhollah Khodadadif, Data on fluoride concentration levels in cold and warm season in City area of Sistan and Baluchistan Province, Iran, *Data Brief* (2018).
- [17] H.R. Maier, G.C. Dandy, The use of artificial neural networks for the prediction of water quality parameters, *Water Resour. Res.* 32 (1996) 1013–1022.
- [18] [https://en.wikipedia.org/wiki/East\\_Azerbaijan\\_Province](https://en.wikipedia.org/wiki/East_Azerbaijan_Province).
- [19] H. Gharibi, et al., A novel approach in water quality assessment based on fuzzy logic, *J. Environ. Manag.* 112 (2012) 87–95.
- [20] M.H. Sowlat, et al., A novel, fuzzy-based air quality index (FAQI) for air quality assessment, *Atmos. Environ.* 45 (12) (2011) 2050–2059.