Data Article

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

Majid Radfard d, Hamed Soleimani b, Samira Nabavi b, Bayram Hashemzadeh c, Hesam Akbari a,*, Hamed Akbari a, Amir Adibzadeh a,b,c,***

a Health Research Center, Life Style Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran
b Department of Environmental Health, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran
c Department of Environmental Health, School of Public Health, Khoy University of Medical Sciences, Khoy, Iran
d Research Center for Health Sciences, Institute of Health, Department of Environmental Health, School of Health, Shiraz University of Medical Sciences, Shiraz, Iran

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...
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 Brief 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.
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
used. The input parameters of the neural network included sulfate, chloride, electrical conductivity (EC) and pH, and the sodium absorption 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
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 Stayle Institute, Baqiyatallah University of Medical Sciences, for their support in this study.
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, Khurasan-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 Jola, 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-Heydari, Khurasan 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. Radfarf, 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 Baluchestan province (Iran), Hum. Ecol. Risk Assess.: Int. J. (2018), https://doi.org/10.1080/10807039.2018.1458596.
[8] M. Radfarf, M. Younesian, R. Nabizadeh Nodehi, H. Biglari, M. Hadi, N. Yousefi, 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 Brief 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 Yazdi 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. Mirzabegi, 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. Radfarf, 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. Ghochani, N. Yousefi, N. Nazmara, M. Radfarf, 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 Zeydumb, Asghar Hamzezadeh, Davoud Jalilid, Abbas Abbsniah, Mahmood Youssef, 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.
[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.