Groundwater

Susceptibility Assessment of Groundwater Nitrate Contamination Using an Ensemble Machine Learning Approach

by Farzaneh Sajedi Hosseini[1], Bahram Choubin[2], Mehdi Bagheri-Gavkosh[3], Omid Karimi[4], Fereshteh Taromideh[5], and Csaba Mako[6]

Abstract

Groundwater pollution susceptibility mapping using parsimonious approaches with limited data is of utmost importance for water resource and health planning, especially in data-scarce regions. Current research assesses groundwater nitrate susceptibility by considering the various combination of explanatory variables. In this study, the novel machine learning models of weighted subspace random forest (WSRF) and generalized additive model using LOESS (GAMLOESS) are applied, and the results are compared with well-known machine learning models of K-nearest neighbors (KKNN) and random forest (RF). The optimum combination of inputs for groundwater nitrate susceptibility mapping is identified using the *k*-fold cross-validation methodology. Results indicated that the combination of variables of precipitation, groundwater level, and lithology had the best performance among the 16 combinations. Modeling performance using the optimum combination demonstrated that the new ensemble approach, the WSRF model, had superior performance according to the evaluation metrics of accuracy (0.87), kappa (0.73), precision (0.92), false alarm ratio (0.08), and critical success index (0.75). The susceptibility assessment results of this paper can be a useful tool in developing strategies for the prevention and protection of groundwater pollution.

Introduction

Groundwater resource quality plays an essential role in many regions around the world, especially in arid and semi-arid areas such as Iran, where the population has grown significantly. Groundwater is a significant resource for drinking water (Jalili et al. 2018), which provides about 63% of drinking water consumed in Iran (Iranian Ministry of Energy [IMOF] 2014). Iran, as one of the largest countries in the Middle East, is among the world's major groundwater consumers (Dalin et al. 2017). Yet, most of the regions in Iran frequently are challenged with low groundwater quality (Dalin et al. 2017; Pazand et al. 2018; Rawat et al. 2019), chronic groundwater decline (Hashemi 2015; Bagheri et al. 2020), desertification (Jalili et al. 2018) and drying out of Qanats (Abbasnejad et al. 2016).

About 50% of the world's population is hugely dependent on groundwater resources for the provision of drinking water and other consumptions (Oki and Kanae 2006). Clean drinking water resources such as groundwater, are essential for human health (Hoseinzadeh et al. 2015) as more than 80% of human diseases are directly related to contaminated water (Agca 2014). Among groundwater pollutants, nitrate pollution is the most nonpoint source, and widespread chemical contamination in groundwater resources (Spalding and Exner 1993). Groundwater nitrate pollution has significantly increased over the past decade (Alighardashi and Mehrani 2017). It has been identified as a severe environmental problem in many countries (Rutkoviene et al. 2009; Wick et al. 2012; Zhang et al. 2013; Chica-Olmo et al. 2014; Esmaeili et al. 2014; Espejo-Herrera et al. 2015; Han et al. 2015; Matiatos 2016; Ouedraogo et al. 2016). Nitrate pollution is a significant suffering problem in groundwater management in most of the world's agricultural areas.

[1]Corresponding author: Reclamation of Arid and Mountainous Regions Department, Faculty of Natural Resources, University of Tehran, Karaj, Iran; farzsajedi@ut.ac.ir

[2]Soil Conservation and Watershed Management Research Department, West Azarbaijan Agricultural and Natural Resources Research and Education Center, AREEO, Urmia, Iran

[3]Irrigation and Reclamation Engineering Department, University of Tehran, P.O. Box 31587-77871, Karaj, Iran

[4]Department of Irrigation and Drainage, College of Aburaihan, University of Tehran, Tehran, Iran

[5]Department of Irrigation, Sari Agricultural Sciences and Natural Resources University, Sari, Iran

[6]Institute of Information Society, University of Public Service, 1083, Budapest, Hungary

Article impact statement: Machine learning modeling is a useful tool and effective method for accurate groundwater pollution mapping.

Received May 2022, accepted September 2022.

doi: 10.1111/gwat.13258

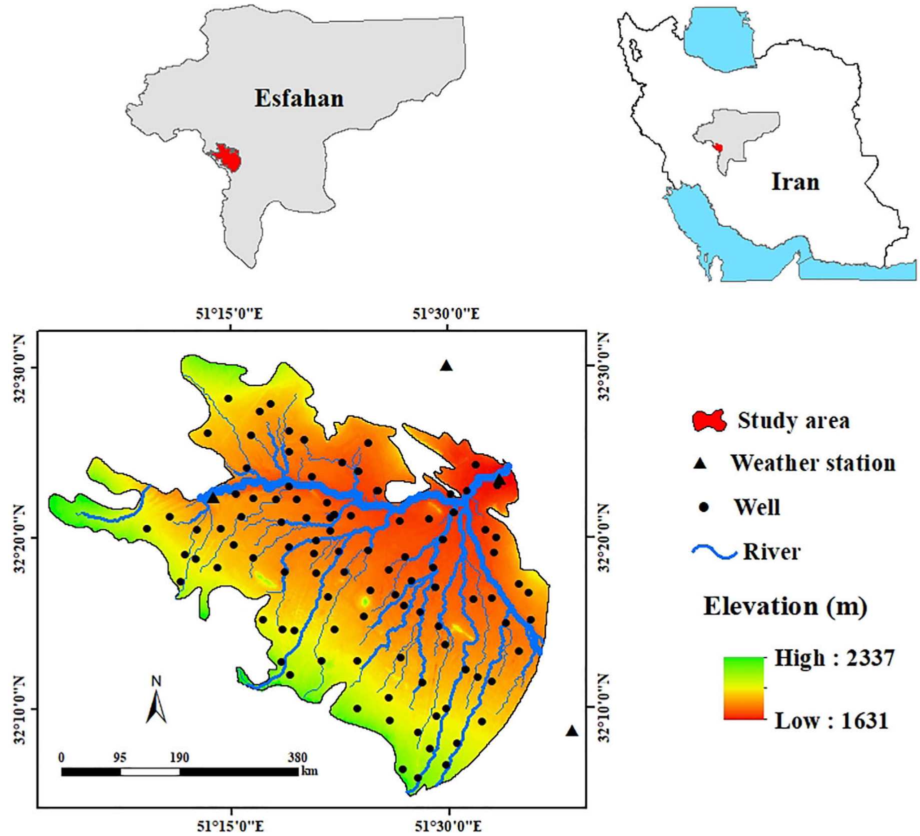

Figure 1. Location of Lenjanat aquifer in Esfahan province, Iran.

The study of groundwater contamination with nitrate in arid and semi-arid regions faces spatial complexity (Nejatijahromi et al. 2019). According to the increasing demands for groundwater resources, strategies such as regular monitoring and identifying the source and behavior of pollutants are essential to retain water quality and supply healthy and safe drinking water for consumers (Alighardashi and Mehrani 2017). Nitrate in groundwater can be initiated by a natural or anthropogenic source. So, currently, contamination of groundwater with nitrate is beyond the scope of some studies.

Machine learning modeling is a useful tool and effective method for complex systems and has recently been applied to predict hazards in environmental science (Choubin et al. 2018; Singh et al. 2018). This method has been especially applied to assess nitrate concentration in aquifers of southwestern United States (Anning et al. 2012), southern Spain (Rodriguez-Galiano et al. 2018), Lowa private wells in the United States (Wheeler et al. 2015), and Andimeshk-Dezful region in Iran (Rahmati et al. 2019). In addition, it has retracted some researchers' attention to assessing groundwater vulnerability quality (Karandish et al. 2017; Ostad-Ali-Askari et al. 2017; Barzegar et al. 2018) by use of different machine learning models, including boosted regression trees (BRT) (Ransom et al., 2017), random forests (RF) (Nolan et al. 2018; Rodriguez-Galiano et al. 2018), artificial neural networks (ANN) (Ostad-Ali-Askari et al. 2017), multivariate discriminant analysis (MDA) (Sajedi-Hosseini et al. 2018), and classification and regression trees (CART) (Rodriguez-Galiano et al. 2018).

Accurate groundwater pollution mapping using parsimonious approaches with a minimum required data is of utmost importance for water resource and health planning, especially in data-scarce regions (Lee and Moon 2007; Kashani et al. 2017; Malekian et al. 2019). Machine learning modeling has some advantages, such as low cost and rapid modeling compared to traditional approaches (Nolan et al. 2018). Therefore, this study applied novel machine learning models such as weighted subspace random forest (WSRF) and generalized additive model using LOESS (GAMLOESS) for groundwater nitrate susceptibility mapping. The results were compared with well-known machine learning models of K-nearest neighbors (KKNN), and random forest (RF). Also, during the modeling process, different input combinations were tested for improving the results. Unlike previous studies (e.g., Sajedi-Hosseini et al. 2018; Rahmati et al. 2019), the main objective was to compare different combinations instead of using all combinations as input. Therefore, besides the comparative analysis of the models, this study determined the best combination across the multitime resampling method in spatial modeling and susceptibility assessment of nitrate. However, the main objectives of this study were: (1) to identify the best combination for groundwater nitrate susceptibility mapping (GNSM), (2) to compare the performance of the different machine learning models in GNSM, and (3) to produce the susceptibility maps based on the best predictive variables and potential models.

Material and Methods

Study Area

The Lenjanat aquifer, with an area of 1180 km^2 is located in Esfahan province, in the center of Iran (Figure 1). Geographically this aquifer is situated between longitudes 51°04′E to 51°41′E and Latitudes 32°04′N to 32°31′N. It is located in an arid region of Esfahan province with annual precipitation of about 250 mm. The elevation in the Lenjanat plain changes from about 1631 to 2337 m above sea level. Zayandehrood River, with a length of 26 km, is the main river of this plain. The mean monthly temperature varies from 4 °C to 27 °C, respectively, in January and August months. Geologically, deposits of the plain are mostly relevant to the Permian to Quaternary periods. Jurassic shale and Cretaceous limestones are the main bedrock of the aquifer (Sajedi-Hosseini et al. 2018). Due to the climate of the region, the main source of drinking and irrigation water is groundwater. Therefore, monitoring, investigating, and modeling the groundwater quality in this region is of utmost importance for water resource management and environmental planning.

Dataset

The dependent variable (predictand) in this study was the location of the polluted and nonpolluted wells. It is worth noting that to reach a balanced number of contaminated and noncontaminated wells, some wells have been excluded. So, the nitrate concentration data from 102 wells (Figure 1) were coded on a binary scale of 0 and 1, which respectively indicates the nonpolluted and polluted wells based on a threshold equal to 50 mg/L for Nitrate values according to the guideline of the World Health Organization (2011) for drinking-water quality. Details of Nitrate data have been presented in our previous work (Sajedi-Hosseini et al. 2018). Predictor variables are including elevation, precipitation (PCP), groundwater level (GWL), distance from industrial (DFI) areas, soil type, land use, and lithology (Figure 2).

Modeling of Groundwater Nitrate Susceptibility

After preparing the input and output variables, the groundwater nitrate pollution was modeled using four classifier machine learning models. Various combinations of input variables were considered and a trial and error method was used to find the best combination in each model. 70% of data was used for the calibration of the models using the 10-fold cross validation (CV) procedure, and the rest of the data (30%) was excluded from the modeling process to validation of the models. In this study the groundwater nitrate susceptibility was modeled using four machine learning models including KKNN, RF, WSRF, and GAMLOESS.

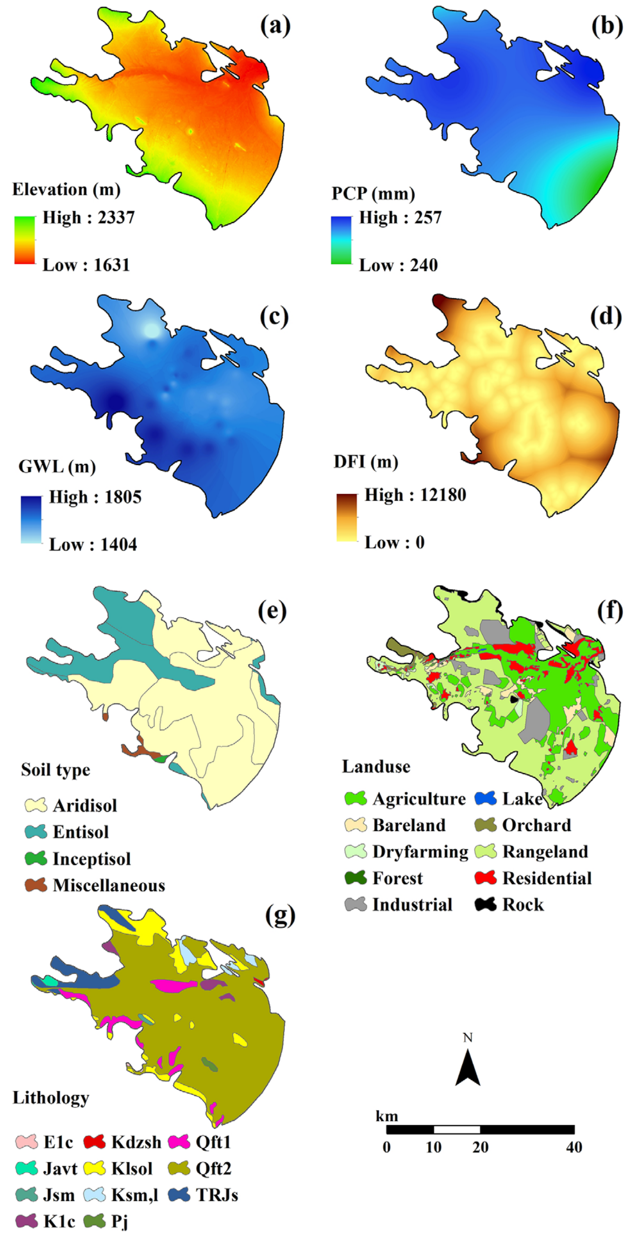

Figure 2. Groundwater nitrate pollution influencing factors: (a) elevation, (b) precipitation (PCP), (c) groundwater level (GWL), (d) distance from industrial (DFI) area, (e) soil type, (f) land use, and (g) lithology.

For evaluating the models, different commonly used metrics are used in this study which is including accuracy, kappa, precision, false alarm ratio (FAR), and critical success index (CSI) (Equations 1 to 5):

$$\text{Accuracy} = \frac{H + CN}{H + FA + M + CN} \tag{1}$$

$$\text{Kappa} = \frac{\text{Accuracy} - P_e}{1 - P_e} \tag{2}$$

$$P_e = \frac{(H + FA)(H + M) + (M + CN)(FA + CN)}{(H + FA + M + CN)^2}$$

Table 1
Accuracy of the Models in Groundwater Nitrate Modeling Using a Different Combination of Inputs by the Testing Dataset

Combination	Variable	RF	KKNN	WSRF	GAMLOESS
Comb1	PCP, GWL	0.80	0.70	0.77	0.57
Comb2	PCP, GWL, DFI	0.67	0.46	0.67	0.60
Comb3	PCP, GWL, lithology	0.80	0.75	0.87	0.76
Comb4	PCP, GWL, land use	0.63	0.53	0.73	0.60
Comb5	PCP, GWL, elevation	0.80	0.63	0.76	0.63
Comb6	PCP, GWL, soil type	0.80	0.67	0.80	0.53
Comb7	PCP, GWL, DFI, lithology	0.70	0.53	0.67	0.66
Comb8	PCP, GWL, DFI, land use	0.67	0.53	0.67	0.60
Comb9	PCP, GWL, DFI, elevation	0.70	0.63	0.63	0.53
Comb10	PCP, GWL, DFI, soil type	0.67	0.50	0.67	0.60
Comb11	PCP, GWL, DFI, lithology, land use	0.67	0.63	0.67	0.63
Comb12	PCP, GWL, DFI, lithology, elevation	0.70	0.70	0.63	0.53
Comb13	PCP, GWL, DFI, lithology, soil type	0.73	0.67	0.63	0.63
Comb14	PCP, GWL, DFI, lithology, land use, elevation	0.67	0.60	0.63	0.50
Comb15	PCP, GWL, DFI, lithology, land use, soil type	0.63	0.63	0.70	0.60
Comb16 (all inputs)	PCP, GWL, DFI, lithology, land use, elevation, soil type	0.70	0.63	0.60	0.57

$$\text{Precision} = \frac{H}{H + FA} \tag{3}$$

$$\text{FAR} = \frac{FA}{H + FA} \tag{4}$$

$$\text{CSI} = \frac{H}{H + M + FA} \tag{5}$$

where H, FA, M, and CN are calculated by a contingency table and respectively denote the number of hits, the number of false alarms, the number of misses, and the number of correct negatives.

Results and Discussion

Best Input Combination

Model calibration was conducted by different combinations (16 combinations) using a 10-fold cross-validation methodology. Validation results calculated by the held-out data (30% of data) for each combination are presented in Table 1. According to the accuracy values, the RF model had 80% accuracy in Comb1 (PCP, GWL), Comb3 (PCP, GWL, lithology), Comb5 (PCP, GWL, elevation), Comb6 (PCP, GWL, soil type). The KKNN model had 75% accuracy with Comb3 (PCP, GWL, lithology). The maximum accuracy (accuracy = 0.87) was occurred by the WSRF model for Comb3 (PCP, GWL, lithology). Also, the GAMLOESS model had the highest accuracy (accuracy = 0.76) in Comb3 (PCP, GWL, lithology) (Table 1).

Therefore, it can be concluded that groundwater modeling has better performance when variables of PCP, GWL, and lithology (i.e., Comb3) are used as input variables. Effects of the PCP on the nitrate concentration in groundwater such as leaching from the soil contents have been proved by scholars (e.g., Schweigert et al. 2004; Rankinen et al. 2007). About GWL, increasing the level of groundwater decreases nitrate concentration (Kraft et al. 1999). Hu et al. (2005) demonstrated that there is a significant correlation between nitrate and groundwater level, which shallow groundwater levels have a higher nitrate concentration. Also, the nitrate removal is dependent on the lithology condition and subsurface variations in soil texture (Haycock and Burt 1993; Gold et al. 1998; Devito et al. 2000; Vidon and Hill 2004).

Table 2
Validation Results of the Groundwater Nitrate Pollution Modeling Using the Best Input Combination (Comb3) by the Testing Dataset

Criterion	KKNN	RF	WSRF	GAMLOESS
Accuracy	0.75	0.80	0.87	0.76
Kappa	0.46	0.58	0.73	0.40
Precision	0.72	0.76	0.92	0.81
FAR	0.30	0.24	0.08	0.22
CSI	0.61	0.68	0.75	0.48

Modeling Results

Evaluation metrics for the best input combination (i.e., Comb3) by the testing dataset are presented in Table 2. According to the accuracy, respectively, models of WSRF, RF, GAMLOESS, and KKNN had better performance. Kappa values indicated that the WSRF and RF models had a good performance ($0.55 <$ kappa < 0.85) (Monserud and Leemans 1992), while the precision metric showed that the WSRF and GAMLOESS models had higher values than the RF and KKNN models. According to the FAR, models of WSRF, GAMLOESS, RF, and KKNN had a lower error, respectively. The

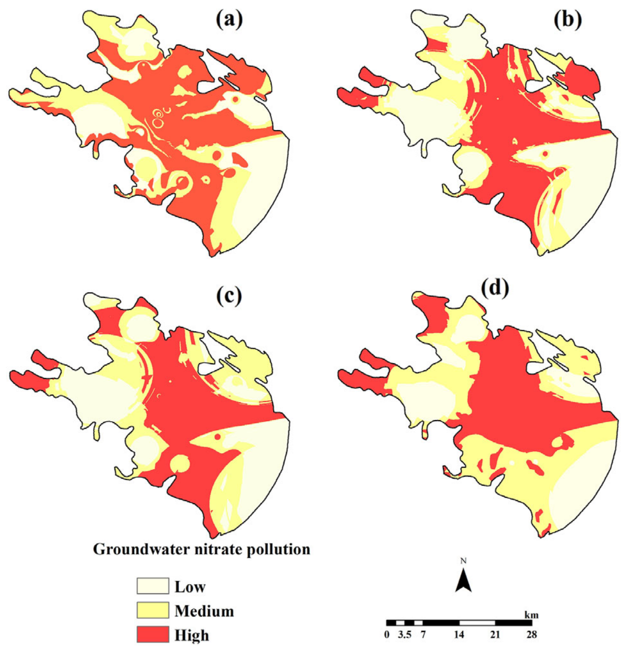

Figure 3. Groundwater nitrate pollution susceptibility: (a) KKNN, (b) RF, (c) WSRF, and (d) GAMLOESS.

CSI values respectively for the WSRF, RF, KKNN, and GAMLOESS were higher (respectively equal to 0.75, 0.68, 0.61, and 0.48) (Table 2).

Therefore, the WSRF model had superior performance due to the evaluation metrics of accuracy (0.87), kappa (0.73), precision (0.92), FAR (0.08), and CSI (0.75). This method can organize and categorize a very sparse dataset with random forests created by small subspaces through a variable weighting method, rather than the traditional approach of random variable sampling (Xu et al. 2012). To the best authors' knowledge, there is not any study that has used the WSRF model for groundwater susceptibility modeling; however, the good performance of this model has been proved in other fields. For example, Singla and Rana (2016) indicated the best performance of the WSRF among the 13 machine learning models in eye state prediction. Also, in another study Wilkes et al. (2018) demonstrated that the WSRF model could accurately predict the biochemical interpretation of urine steroid profiles.

Susceptibility Mapping of Groundwater Nitrate Pollution

After calibration and validation of the machine learning models using the best input combination, the pixel value of inputs for the whole region was used and the groundwater nitrate pollution, spatially, was predicted. Predicted output maps were classified into three susceptibility classes (i.e., low, medium, and high) based on the Natural Breaks (Jenks) classification method (Figure 3). According to the susceptibility maps, the percentage of high susceptibility class was greater respectively for models of GAMLOESS, WSRF, KKNN, and RF with 38.4, 36.4, 35.7, and 34.0% of the whole region. However, the location of the susceptibility classes produced by the different models was approximately matched. The high class is mostly located in the middle regions of the plain, which mostly have land uses for agriculture, residential, and industrial. Inorganic nitrate sources from agricultural areas such as chemical fertilizers, organic nitrate sources from residential areas such as human waste and municipal sewage effluents, and industrial wastewaters can be the main cause of high susceptibility in these areas (Dongol et al. 2005; Amiri et al. 2014; Esmaeili et al. 2014; Matzeu et al. 2017; Sajedi-Hosseini et al. 2018).

Conclusions

This study tried to model groundwater susceptibility using limited and parsimonious parameters. So, the best input combination was identified by the k-fold cross-validation methodology through four machine learning models. The valuable result obtained is that using all variables as input does not guarantee a better model performance. Variables of precipitation, groundwater level, and lithology were identified as the best input combination. Modeling performance using the best combination demonstrated that the new ensemble model, the WSRF model, had superior performance according to the evaluation metrics of accuracy (0.87), kappa (0.73), precision (0.92), FAR (0.08), and CSI (0.75). Therefore, the results of this study indicated a good performance using the limited parameters which are most parsimonious and important for water resources and health planning in a data-scarce region. However, hydrologically, groundwater pollution studies are confronted with some inevitable limitations such as temporal variations of pollution and migration of contaminants in groundwater; which should be considered in groundwater pollution risk management.

Authors' Note

The authors do not have any conflicts of interest or financial disclosures to report.

References

Abbasnejad, A., B. Abbasnejad, R. Derakhshani, and A. Hemmati Sarapardeh. 2016. Qanat hazard in Iranian urban areas: Explanation and remedies. *Environmental Earth Sciences* 75, no. 19: 1–14.

Agca, N. 2014. Spatial variability of groundwater quality and its suitability for drinking and irrigation in the Amik Plain (South Turkey). *Environmental Earth Sciences* 72, no. 10: 4115–4130. https://doi.org/https://doi.org/10.1007/s12665-014-3305-7

Alighardashi, A., and M.J. Mehrani. 2017. Survey and zoning of nitrate-contaminated groundwater in Iran. *Journal of Materials and Environmental Science* 8, no. 12: 4339–4348. https://doi.org/10.26872/jmes.2017.8.12.457

Amiri, V., M. Rezaei, and N. Sohrabi. 2014. Groundwater quality assessment using entropy weighted water quality index (EWQI) in Lenjanat, Iran. *Environmental Earth Sciences* 72, no. 9: 3479–3490. https://doi.org/10.1007/s12665-014-3255-0

Anning, D.W., A.P. Paul, T.S. McKinney, J.M. Huntington, L.M. Bexfield, and S.A. Thiros. 2012. Predicted nitrate and

arsenic concentrations in basin-fill aquifers of the southwestern United States. U.S. Geological Survey. Scientific Investigations Report 2012-5065.
Bagheri, M., M. Kholghi, S.M. Hosseini, F. Amiraslani, and A. Hoorfar. 2020. Participatory approach in aquifer storage and recovery management in arid zones, does it work. *Groundwater for Sustainable Development* 10: 100368. https://doi.org/10.1016/j.gsd.2020.100368
Barzegar, R., A.A. Moghaddam, R. Deo, E. Fijani, and E. Tziritis. 2018. Mapping groundwater contamination risk of multiple aquifers using multi-model ensemble of machine learning algorithms. *Science of the Total Environment* 621, no. 12: 697–712. https://doi.org/10.1016/j.scitotenv.2017.11.185
Chica-Olmo, M., J.A. Luque-Espinar, V. Rodriguez-Galiano, E. Pardo-Igúzquiza, and L. Chica-Rivas. 2014. Categorical indicator kriging for assessing the risk of groundwater nitrate pollution: The case of Vega de Granada aquifer (SE Spain). *Science of the Total Environment* 470: 229–239. https://doi.org/10.1016/j.scitotenv.2013.09.077
Choubin, B., G. Zehtabian, A. Azareh, E. Rafiei-Sardooi, F. Sajedi-Hosseini, and Ö. Kişi. 2018. Precipitation forecasting using classification and regression trees (CART) model: A comparative study of different approaches. *Environmental Earth Sciences* 77, no. 8: 1–13. https://doi.org/10.1007/s12665-018-7498-z
Dalin, C., Y. Wada, T. Kastner, and M.J. Puma. 2017. Groundwater depletion embedded in international food trade. *Nature* 543, no. 7647: 700–704. https://doi.org/10.1038/nature21403
Devito, K.J., D. Fitzgerald, A.R. Hill, and R. Aravena. 2000. Nitrate dynamics in relation to lithology and hydrologic flow path in a river riparian zone. *Journal of Environmental Quality* 29, no. 4: 1075–1084.
Dongol, B.S., J. Merz, M. Schaffner, G. Nakarmi, P.B. Shah, S.K. Shrestha, P.M. Dangol, and M.P. Dhakal. 2005. Shallow groundwater in a middle mountain catchment of Nepal: Quantity and quality issues. *Environmental Geology* 49, no. 2: 219–229. https://doi.org/10.1007/s00254-005-0064-5
Esmaeili, A., F. Moore, and B. Keshavarzi. 2014. Nitrate contamination in irrigation groundwater, Isfahan, Iran. *Environmental Earth Sciences* 72, no. 7: 2511–2522. https://doi.org/10.1007/s12665-014-3159-z
Espejo-Herrera, N., K.P. Cantor, N. Malats, D.T. Silverman, A. Tardón, R. García-Closas, C. Serra, M. Kogevinas, and C.M. Villanueva. 2015. Nitrate in drinking water and bladder cancer risk in Spain. *Environmental Research* 137: 299–307. https://doi.org/10.1016/j.envres.2014.10.034
Gold, A.J., P.A. Jacinthe, P.M. Groffman, W.R. Wright, and R.H. Puffer. 1998. Patchiness in groundwater nitrate removal in a riparian forest. *Journal of Environmental Quality* 27, no. 1: 146–155.
Han, D., G. Cao, J. McCallum, and X. Song. 2015. Residence times of groundwater and nitrate transport in coastal aquifer systems: Daweijia area, northeastern China. *Science of the Total Environment* 538: 539–554. https://doi.org/10.1016/j.scitotenv.2015.08.036
Hashemi, H. 2015. Climate change and the future of water Management in Iran. *Middle East Critique* 24, no. 3: 307–323. https://doi.org/10.1080/19436149.2015.1046706
Haycock, N., and T. Burt. 1993. Role of floodplain sediments in reducing the nitrate concentration of subsurface run-off: A case study in the Cotswolds, UK. *Hydrological Processes* 7, no. 3: 287–295.
Hoseinzadeh, E., H. Khorsandi, C. Wei, and M. Alipour. 2015. Evaluation of Aydughmush River water quality using the National Sanitation Foundation Water Quality Index (NSFWQI), river pollution index (RPI), and forestry water quality index (FWQI). *Desalination and Water Treatment* 54, no. 11: 2994–3002. https://doi.org/10.1080/19443994.2014.913206
Hu, K., Y. Huang, H. Li, B. Li, D. Chen, and R.E. White. 2005. Spatial variability of shallow groundwater level, electrical conductivity and nitrate concentration, and risk assessment of nitrate contamination in North China Plain. *Environment International* 31, no. 6: 896–903.
Iranian Ministry of Energy, 2014. Rehabilitation and balance programfor groundwater resources, 106 pp.
Jalili, D., M. RadFard, H. Soleimani, S. Nabavi, H. Akbari, H. Akbari, A. Kavosi, A. Abasnia, and A. Adibzadeh. 2018. Data on nitrate–nitrite pollution in the groundwater resources a Sonqor plain in Iran. *Data in Brief* 20: 394–401. https://doi.org/10.1016/j.dib.2018.08.023
Karandish, F., A. Darzi-Naftchali, and A. Asgari. 2017. Application of machine-learning models for diagnosing health hazard of nitrate toxicity in shallow aquifers. *Paddy and Water Environment* 15, no. 1: 201–215. https://doi.org/10.1007/s10333-016-0542-2
Kashani, M., M. Ghorbani, Y. Dinpasho, S. Shahmorad, and Z. Kundzewicz. 2017. Comparative study of different wavelets for developing parsimonious Volterra model for rainfall-runoff simulation. *Water Resources* 44: 568–578. https://doi.org/10.1134/S009780781704008X
Kraft, G., W. Sites, and D. Mechanic. 1999. Impact of irrigated vegetable agriculture in a humid North-Central U.S. sand plain aquifer. *Groundwater* 37, no. 13: 572–580.
Lee, H., and Y. Moon. 2007. Analysis and development of conceptual rainfall-runoff model structures for regionalization purposes. *KSCE Journal of Civil Engineering* 11: 57–64. https://doi.org/10.1007/BF02823373
Malekian, A., B. Choubin, J. Liu, and F. Sajedi-Hosseini. 2019. Development of a new integrated framework for improved rainfall-runoff modeling under climate variability and human activities. *Water Resources Management* 33, no. 7: 2501–2515. https://doi.org/10.1007/s11269-019-02281-0
Matiatos, I. 2016. Nitrate source identification in groundwater of multiple land-use areas by combining isotopes and multivariate statistical analysis: A case study of Asopos basin (Central Greece). *Science of the Total Environment* 541: 802–814. https://doi.org/10.1016/j.scitotenv.2015.09.134
Matzeu, A., R. Secci, and G. Uras. 2017. Methodological approach to assessment of groundwater contamination risk in an agricultural area. *Agricultural Water Management* 184: 46–58. https://doi.org/10.1016/j.agwat.2017.01.003
Monserud, R.A., and R. Leemans. 1992. Comparing global vegetation maps with the kappa statistic. *Ecological Modelling* 62, no. 4: 275–293.
Nejatijahromi, Z., H. Nassery, T. Hosono, M. Nakhaei, F. Alijani, and A. Okumura. 2019. Groundwater nitrate contamination in an area using urban wastewaters for agricultural irrigation under arid climate condition, southeast of Tehran, Iran. *Agricultural Water Management* 221: 397–414. https://doi.org/10.1016/j.agwat.2019.04.015
Nolan, B., C. Green, P. Juckem, and J. Liao. 2018. Metamodeling and mapping of nitrate flux in the unsaturated zone and groundwater, Wisconsin, USA. *Journal of Hydrology* 559: 428–441. https://doi.org/10.1016/j.jhydrol.2018.02.029
Oki, T., and S. Kanae. 2006. Global hydrological cycles and world water resources. *Science* 313: 1068–1072.
Ostad-Ali-Askari, K., M. Shayannejad, and H. Ghorbanizadeh-Kharazi. 2017. Artificial neural network for modeling nitrate pollution of groundwater in marginal area of Zayandeh-Rood River, Isfahan, Iran. *KSCE Journal of Civil Engineering* 21, no. 1: 134–140. https://doi.org/10.1007/s12205-016-0572-8
Ouedraogo, I., P. Defourny, and M. Vanclooster. 2016. Mapping the groundwater vulnerability for pollution at the pan

African scale. *Science of the Total Environment* 544: 939–953. https://doi.org/10.1016/j.scitotenv.2015.11.135

Pazand, K., D. Khosravi, M. Ghaderi, and M. Rezvanianzadeh. 2018. Hydrogeochemistry and lead contamination of groundwater in the north part of Esfahan province, Iran. *Journal of Water and Health* 16, no. 4: 622–634. https://doi.org/10.2166/wh.2018.034

Rahmati, O., B. Choubin, A. Fathabadi, F. Coulon, E. Soltani, H. Shahabi, E. Mollaefar, J. Tiefenbacher, S. Cipullo, B. Ahmad, and D. Tien Bui. 2019. Predicting uncertainty of machine learning models for modelling nitrate pollution of groundwater using quantile regression and UNEEC methods. *Science of the Total Environment* 688: 855–866. https://doi.org/10.1016/j.scitotenv.2019.06.320

Rankinen, K., T. Salo, K. Granlund, and H. Rita. 2007. Simulated nitrogen leaching, nitrogen mass field balances and their correlation on four farms in southwestern Finland during the period 2000e2005. *Agricultural and Food Science* 16: 387–406.

Rawat, K.S., L. Jeyakumar, S. Singh, and V. Tripathi. 2019. Appraisal of groundwater with special reference to nitrate using statistical index approach. *Groundwater for Sustainable Development* 8: 49–58. https://doi.org/10.1016/j.gsd.2018.07.006

Rodriguez-Galiano, V., J. Luque-Espinar, M. Chica-Olmo, and M. Mendes. 2018. Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods. *Science of the Total Environment* 624: 661–672. https://doi.org/10.1016/j.scitotenv.2017.12.152

Rutkoviene, V., L. Cesoniene, and S. Kutra. 2009. The influence of water use intensity on nitrate concentration in shallow well water. *Polish Journal of Environmental Studies* 18, no. 3: 435–442.

Sajedi-Hosseini, F., A. Malekian, B. Choubin, O. Rahmati, S. Cipullo, F. Coulon, and B. Pradhan. 2018. A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. *Science of the Total Environment* 644: 954–962. https://doi.org/10.1016/j.scitotenv.2018.07.054

Schweigert, P., N. Pinter, and R. Van Der Ploeg. 2004. Regression analyses of weather effects on the annual concentrations of nitrate in soil and groundwater. *Journal of Plant Nutrition and Soil Science* 167, no. 3: 309–318. https://doi.org/10.1002/jpln.200321291

Singh, S., R. Taylor, M. Rahman, and B. Pradhan. 2018. Developing robust arsenic awareness prediction models using machine learning algorithms. *Journal of Environmental Management* 211: 125–137. https://doi.org/10.1016/j.jenvman.2018.01.044

Singla, D., and P. Rana. 2016. August. Eye state prediction using ensembled machine learning models. In *2016 International Conference on Inventive Computation Technologies*, Vol. 2, IEEE. 1–5.

Spalding, R., and M. Exner. 1993. Occurrence of nitrate in groundwater—A review. *Journal of Environmental Quality* 22, no. 3: 392–402. https://doi.org/10.2134/jeq1993.00472425002200030002x

Vidon, P., and A. Hill. 2004. Landscape controls on nitrate removal in stream riparian zones. *Water Resources Research* 40, no. 3: 1–14. https://doi.org/10.1029/2003WR002473

Wheeler, D., B. Nolan, A. Flory, C. DellaValle, and M. Ward. 2015. Modeling groundwater nitrate concentrations in private wells in Iowa. *Science of the Total Environment* 536: 481–488. https://doi.org/10.1016/j.scitotenv.2015.07.080

Wick, K., C. Heumesser, and E. Schmid. 2012. Groundwater nitrate contamination: Factors and indicators. *Journal of Environmental Management* 111: 178–186. https://doi.org/10.1016/j.jenvman.2012.06.030

Wilkes, E., G. Rumsby, and G. Woodward. 2018. Using machine learning to aid the interpretation of urine steroid profiles. *Clinical Chemistry* 64, no. 11: 1586–1595.

World Health Organization. 2011. *Guidelines for Drinking—Water Quality*, 4th ed., Geneva, Switzerland, Environmental health criteria. http://www.who.int/water_sanitation_health/publications/2011/dwq_guidelines/en/

Xu, B., Z.H. Joshua, W. Graham, W. Qiang, and Y. Yunming. 2012. Classifying very high-dimensional data with random forests built from small subspaces. *International Journal of Data Warehousing and Mining* 8, no. 2: 44–63.

Zhang, X., Z. Xu, X. Sun, W. Dong, and D. Ballantine. 2013. Nitrate in shallow groundwater in typical agricultural and forest ecosystems in China, 2004-2010. *Journal of Environmental Sciences (China)* 25, no. 5: 1007–1014.