CrossMark

Mapping Aquifer Vulnerability Indices Using Artificial Intelligence-running Multiple Frameworks (AIMF) with Supervised and Unsupervised Learning

Ata Allah Nadiri[1] · Maryam Gharekhani[1] · Rahman Khatibi[2]

Abstract DRASTIC-based vulnerability indices and their variations for an aquifer are investigated in this paper, each of which is regarded as a framework since their rationale of using seven DRASTIC data layers is consensual and lacks empirical or theoretical formulations. The Basic DRASTIC framework (BDF) is implemented by a set of prescribed rules; whereas its three variations involve unsupervised learning from the data, which comprise: (i) learning the rates by the Wilcoxon test (WDF) but using BDF weights; (ii) using BDF rates but learning the weights by Genetic Algorithm (BDF-GA); and (iii) learning rates as in WDF and the weights as in BDF-GA (WDF-GA). These four frameworks are not supervised, but the novelty of the paper is to introduce supervised learning at the second stage by Artificial Intelligence to run Multiple Frameworks (AIMF), for which the paper uses Support Vector Machine (SVM). AIMF uses the outputs of the four frameworks as its input data and a function of observed nitrate-N values as its target data. The AIMF strategy is evaluated in the aquifer of Ardabil plain, which is exposed to anthropogenic contamination such as nitrate-N. The coefficient of correlation (r-values) between the results and nitrate-N values for the above frameworks are: 0.2, 0.37, 0.38 and 0.45; whereas AIMF enhances it to 0.84; attributable to the supervised learning.

✉ Ata Allah Nadiri
   nadiri@tabrizu.ac.ir

   Maryam Gharekhani
   m_gharekhani@tabrizu.ac.ir

   Rahman Khatibi
   gtev.rex@gmail.com

[1] Department of Earth Sciences, Faculty of Natural Sciences, University of Tabriz, 29 Bahman
     Boulevard, Tabriz, East Azerbaijan, Iran

[2] GTEV-ReX Limited, Swindon, UK

⌂ Springer

1 Introduction

Application of Artificial Intelligence (AI) is topical in research for improving professional practices on assessing aquifer vulnerability indices (VI) using the DRASTIC framework. VI maps are generated by overlaying a set of data layers prepared by prescribed rates at the grid scale for a set of data layers and prescribed weights for the relative importance of each data layer. The choice of the data layers and their rates and weights are not based on any empirical or theoretical rationale but are consensual and represent intrinsic aquifer vulnerability. The problems include: (i) rates and weights are prescriptive; and (ii) intrinsic vulnerability cannot be measured directly and therefore supervised learning is not feasible; and (iii) measured nitrate-N values are used to serve as a measure of vulnerability and thereby to estimate their correlation with VI values. Supervised learning becomes feasible by using AI models and this paper investigates AI-running Multiple Frameworks (AIMF). AIMF is a strategy offered by the paper for supervised learning (Techniques at Level 2) from unsupervised multiple frameworks (Techniques at Level 1) and the goal of the paper is to study AIMF performances.

The methods for assessing aquifer vulnerability may be categorised as follow: (i) process-based models, which use simulation models to estimate subsurface flow contaminant and transport (Kauffman and Chapelle 2010); (ii) statistical methods, which use descriptive statistics for incorporating effects of several variables; (iii) overlay-index methods, which use a combination of different regional maps by assigning a numerical index (Tesoriero et al. 1998); and (iv) framework approaches, which include AVI (Van Stemproot et al. 1993), DRASTIC (Aller et al. 1987); GALDIT (Najib et al. 2012); GOD (Foster 1987); IRISH (Daly and Drew 1999); SI (Ribeiro 2000) and SINTACS (Civita 1994). The rationale in the above frameworks is the use of prescribed rates at the grid scale for the input data and prescribed weights for the relative importance of each data layer but they differ for using different data layer and different prescribed rates and weights.

Among the various frameworks, the paper uses the DRASTIC framework for its wide usage to mapping groundwater vulnerability to contamination (Sorichetta et al. 2012; Wu et al. 2014; Sadeghfam et al. 2016a, 2016b; Jafari and Nikoo 2016). DRASTIC is the acronym of seven geological and hydrogeological data layers, which are: Depth to water Table (D), Recharge (R), Aquifer media (A), Soil media (S), Topography or slope (T), Impacts due to thickness (I), hydraulic Conductivity (C). These data layers are widely available at the grid scale, which collectively provide an insight into intrinsic vulnerability. Measured nitrate-N is used as a proxy of vulnerability to reflect on aquifer responses to anthropogenic contaminants. DRASTIC frameworks rely on expert judgment by using prescribed rates and weights. Hence excessive uncertainties are likely without learning from the correlation between the data and nitrate-N values. This makes a research case to investigate possible improvements through the strategy of learning from local data instead of using prescriptive rules.

Existing AI applications to aquifer vulnerability problems involve modifying DRASTIC weights, which include: (i) Analytic Hierarchy Process (AHP), as studied by Bello-Dambatta et al. (2009), Bai et al. (2012); (ii) single-parameter sensitivity analysis, which

reduces subjectivity in weights, e.g. Rupert (1999), Mclay et al. (2001); and (iii) correlation analysis between the rated data and nitrate concentrations as basis to modify weights, e.g. Huan et al. (2012).

Modification of rates through AI is another possibility to improve aquifer vulnerability frameworks (Huan et al. 2012). The latter modifies the rates by nonparametric statistical tests using the Wilcoxon rank-sum technique by assigning the lowest and highest rates to the lowest and highest mean nitrate concentrations of Aller's rates, respectively and using prescribed rate by Aller et al. (1987) to the locations without measured nitrate concentration. As the paper uses the Wilcoxon test, its description is given later.

Identifying gaps in the ongoing research is not an easy task, as the ongoing research is diversifying without an inherent vision. However, the authors are following a program of research to develop the next generation for mapping aquifer vulnerability indices with a focus on learning from convergence and divergence in different base models/results. The authors build on their published works of Nadiri et al. (2017a, 2017b, 2017c), which show that remarkable improvements are likely by supervised AI learning (at the second level) from lower-level results of supervised models of vulnerability indices. This paper investigates another strategy, in which supervised AI models learn directly from four unsupervised frameworks, as follows. (i) BDF: (Basic DRASTIC frameworks) uses prescribed rates and weights with no learning; (ii) WDF (Wilcoxon-based DRASTIC Framework): learns rates from the data by Wilcoxon test but weights are as per BDF; (iii) BDF-GA (Genetic Algorithms): rates are as per BDF but weights are learned by GA without supervised learning; and (iv) WDF-GA: rates are as per WDF and weights as per GA. The paper uses Support Vector Machine (SVM) to run these four frameworks (BDF and its three variations) to map the groundwater vulnerability.

Ardabil plain, in the Ardabil province, Northwest Iran, is the study area. Water resources for domestic, agricultural and industrial demands are met from groundwater withdrawal of the aquifer in Ardabil plain. The nitrate-N levels in the plain is exposed to the exceedance of the permissible 10 mg/L, as per United State Environmental Protection Agency (USEPA 2012). The paper presents an outcome of a study to map the vulnerability of the aquifer, which may be used for its management to reduce impacts of the polluting agricultural and industrial activities giving rise to nitrate contamination and other anthropogenic contaminants.

2 Methodology

The methodology is detailed in this section for Artificial Intelligence (AI) running Multiple Frameworks (AIMF), in which SVM is the AI supervised learning model and runs four frameworks: Basic DRASTIC Framework (BDF) and three variations of BDF (and hence four frameworks), see Fig. 1.

2.1 Basic DRASTIC Framework (BDF)

The Basic DRASTIC Framework (BDF), introduced by the USEPA in 1987, is used widely as a tool to assess vulnerability of groundwater to contaminations, see the illustration in Fig. 1. BDF uses seven data layers, which are: water table Depth (D), net Recharge (R), Aquifer media (A), Soil media (S), Topography or slope (T), Impact of the vadose zone (I), and the aquifer hydraulic Conductivity (C). The prescribed rates have the values are between 1 and 10
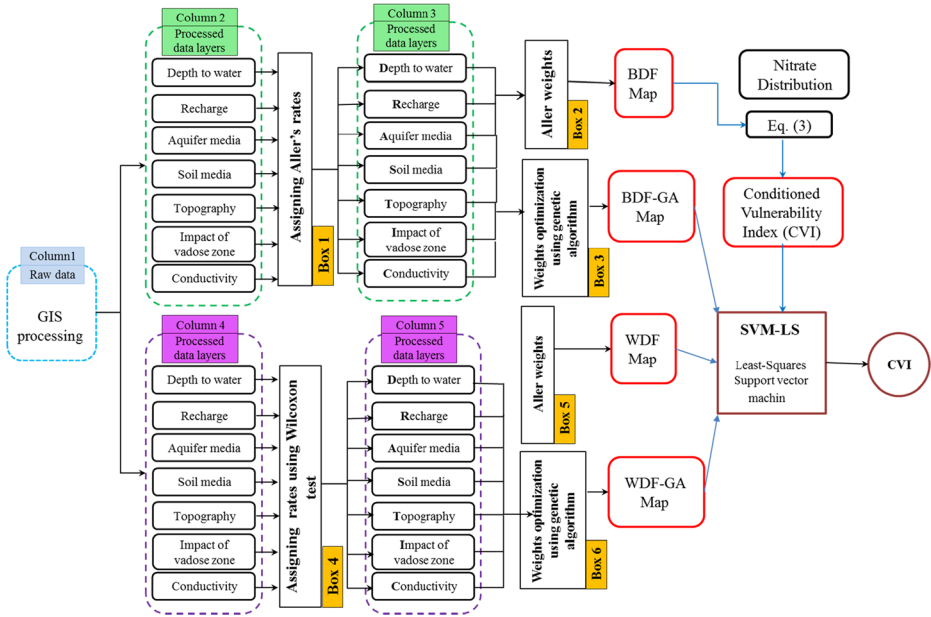
Fig. 1 Flowchart of the methodology

and are reproduced in Table 1, which also gives the prescribed weight values, in the range of 1 to 5, both as per Aller et al. (1987). The DRASTIC vulnerability index is calculated as:

$$DRASTIC\ Index = D_r D_w + R_r R_w + A_r A_w + S_r S_w + T_r T_w + I_r I_w + C_r C_w \qquad (1)$$

where D, R, A, S, T, I, and C are the seven data layers and the subscripts "r" and "w" refer to rates and weights, respectively. Table 1, also shows the interpolation method for each data layer. All DRASTIC data layers and groundwater vulnerability indices are mapped using a commercial GIS software package. The authors' implementation of the framework is detailed by Nadiri et al. (2017a, 2017b, 2017c).

2.2 Using Wilcoxon Test to Formulate BDF Variation – WDF

The algorithm for Wilcoxon test (Wilcoxon 1945) is used to develop a variation for BDF by learning the values of the rates from the data instead of using prescribed values. Wilcoxon test is a paired difference test and the underlying capability is used to learn the rates from the dependencies between the rates and hydrogeological characteristics. The procedure uses the mean of each parameter defined in the initial model based on the statistical test and compares two samples to identify their matching sample values or repeated measurements. It is a non-parametric statistical hypothesis test and assesses if the ranks of their population mean differ.

The concept of the Wilcoxon test is straightforward, as follows: (i) arrange the data in order and produce two rank totals, with each total representing a case; (ii) study the difference between the two cases such that: most of the low ranks belongs to one case and most of the high ranks to the other case; (iii) assess the rank totals, they should be quite different and one of them will be quite small; (iv) there will be conditions when high and low ranks are

Table 1 Prescribed BDF rates and weights and learning their values by variant frameworks

DRASTIC* Data Layers	Description	Interpolation method	Classification range	BDF rates	Mean NO$_3$-N (mg/L)	WDF- rates	BDF weight	BDF-GA weight	WDF-GA weight
D: Depth to water (m)	Measure the depth from ground surface to the water table. The lower the distance, the greater the chance for aquifer pollution	Ordinary Kriging	0.87–1.5	10	1.5	1.12	5.0	5.0 [Note 1]	5.0 [Note 1]
			1.5–4.6	9	4.4	3.28			
			4.6–9.1	7	10.95	8.17			
			9.1–15.2	5	7.76	5.79			
			15.2–22.9	3	13.4	10			
			22.9–30.5	2	5.15	3.84			
			30.5–49.5	1	5.07	3.78			
R: Recharge (cm/year)	Measures quantity of water infiltrates the ground and percolates water table: The greater the recharge the higher the susceptibility to contamination.	IDW	0.18–5.08	1	7.2	8.2	4.0	1.0	1.1
			5.08–10.16	3	7.92	9			
			10.16–17.78	6	7.48	8.5			
			17.78–25.4	8	5.2	5.9			
			25.4–43.5	9	6.84	7.7			
A: Aquifer media (rate)	Measures capacity to attenuate pollutants: The lower permeability, the stronger the attenuation capacity of the aquifer media.	Ordinary Kriging	Clay, silt and sand	3	4.52	4.33	3.0	5.0	4.7
			Sand and clay	5	8.34	8			
			Sand& gravel	8	6.89	6.6			
S: Soil media (rate)	Soil type controls recharge to infiltrate pollutants in vadose zones: the higher the holding capacity, the longer the travel time	Thematic map	Clay	1	4.1	3.82	2.0	1.0	1.78
			Silty loam	4	8.57	8			
			Sandy loam	6	7.7	7.18			
			Gravel/sand	8	4.86	4.53			
T: Topography (percent)	Measures the slope of an area: The greater the slope, the lower the rate of Infiltration and vice versa.	Thematic map	0–2	10	6.7	7.38	1.0	1.1	1.1
			2–6	9	8.46	9.33			
			6–12	5	9.06	10			
			12–18	3	0.1	0.1			
			18–19.3	1	0.01	0.01			

Table 1 (continued)

DRASTIC* Data Layers	Description	Interpolation method	Classification range	BDF rates	Mean NO$_3$-N (mg/L)	WDF- rates	BDF weight	BDF-GA weight	WDF-GA weight
I: Impact of vadose zone media (rate)	The zone above the water table: It generally measures the buffer zone as the unsaturated layer to protect the aquifer.	Ordinary Kriging	Silt and clay Silt, clay & sand Sand and silt Gravel and sand	2 4 6 8	10.4 6.68 6.82 7.38	8 5.13 5.24 5.67	5.0	1.0	1.0
C: Conductivity (m/day)	Measures the ability of the aquifer material ability to transmit water: The higher conductivity the highervulnerability to contamination	IDW	<4 4–12 12–28	1 2 4	6.91 7.16 2.7	3.85 4 1.5	3.0	1.6	3.1

Note 1: The weight identified by GA is coincidentally equal to the prescribed weight for the D data layer

Note 2: The default values of parameters used in GA implementations are: Initial population size = 100; Probability of Crossover Operation = 0.8. Mutation Function is Gaussian. Scale Value = 1; Shrink value = 1

*DRASTIC Data Layers and description reproduced from Nadiri et al. (2017a, 2017b)

distributed fairly evenly between the two cases, at which time their rank totals will be fairly similar and quite large; (v) assign the maximum rates to the maximum mean nitrate-N and minimum rates to minimum mean nitrate-N concentrations; and (vi) derive modified rate values of BDF by the mean of nitrate-N concentrations of each class reducing to a 10-grade scale (Huan et al. 2012).

2.3 Genetic Algorithm (GA)

The two BDF-GA and WDF-GA variations use Genetic Algorithm (GA) to optimise the weights of the 7 DRASTIC data layers, treated as the decision variables by minimising an objective function using GA. GA, introduced in 1975 by John Holland (Holland 1975), is in a wide use, which generates solutions to optimisation and search problems by emulating biological operators, e.g. mutation crossover and selection (Mitchell 1996). GA carries out the search process by using four evolutionary processes: initialisation, selection, crossover, and mutation (Davis 1991).

The whole procedure is as follows: The initialisation process generates a population of chromosomes as a starting point; their fitness function is selected in terms of objective function, decision variable and constraints. This initial population is then allowed to evolve toward better solutions; although a better solution is a mathematical concept and not natural. The population in each iteration is referred to as a generation associated with an assessed fitness for every individual in the population. Each individual's genome is recombined while allowing for random mutations to form a new generation for the sage in the next iteration. The process is terminated at a set criteria.

The objective function is formulated to maximise correlation coefficient between vulnerability indices and a measure of nitrate-N values by identifying weights (w_j) in BDF-GA and WDF-GA, which is expressed as:

$$F = \frac{\sum_{i=1}^{n}\left(V_i - \overline{V}\right)\left(N_i - \overline{N}\right)}{\sqrt{\sum_{i=1}^{n}\left(V_i - \overline{V}\right)^2}\sqrt{\sum_{i=1}^{n}\left(N_i - \overline{N}\right)^2}} \tag{2}$$

Constrain: $1 < w_j < 5$, j = 1,2, …, 7.

where F is objective function; n is the number of data; N_i is NO_3-N concentration; $\overline{N}$ is the mean NO_3-N concentration; V_i vulnerability index; and $\overline{V}$ is mean vulnerability index.

2.4 AI-running Multiple Frameworks (AIMF) using SVM-LS

The paper uses Support Vector Machine (SVM), which is a supervising model and a machine learning method to run the four frameworks described above. The predictor mode of SVM is implemented in various ways but the paper uses the modification by Suykens (2000) for the predictor mode, and the Least Squares technique for the estimator mode, and hence SVM-LS. The supervision in SVM-LS is based on: (i) using outputs of the for framework as inputs to SVM; and (ii) Conditioned Vulnerability Index (CVI) values as target values, expressed as:

$$CVI = \frac{VI_{max}}{(NO_3-N)_{max}} \times (NO_3-N)_i \tag{3}$$

where i, is index and counts the number of datapoints in an observation well; $(NO_3 - N)_{max}$ is the maximum nitrate concentration; VI_{max} is the maximum vulnerability calculated from BDF.

A training set $\{x_i, y_i\}_{i=1}^N$ is considered, where x is a vector containing the four arrays of outputs of BDF, WDF, BDF-GA and WDF-GA, $x_t \epsilon R^n$, and R^n is n-dimensional vector space; y is the array of Conditioned Vulnerability Index (CVI) and serves as target values, $y_t \epsilon R$ and R is one-dimensional vector space. SVM-LS is formulated as follows:

$$y(x) = w^T \circ \varphi(x) + b \tag{4}$$

where b is bias values of regression function; $\varphi(x)$ maps nonlinearly input data onto a feature space of a higher dimension; w is weight used in the regression function. Note that Eqs. (1) and (4) are express the same effect in different ways. The estimator problem of SVM-LS minimises the regular function expressed by Eq. (5), see Suykens et al. (2002) for more detail. The subsequent SVM-LS model for function estimation is Eq. (6):

$$\min R(w, e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^n e_i^2 \tag{5}$$

$$y(x) = \sum_{i=1}^n \alpha_i K(x_i, x) + b \tag{6}$$

where, α_i is part of the solution for the linear system; b is the other part of the solution to the linear system; $K(x_i, x)$ is kernel function; γ is a positive real constant to avoid overfitting and for noisy data a smaller value is used. The Radial Basis Function (RBF) is used to express the kernel function as:

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{\sigma^2}\right) \tag{7}$$

where, σ is the parameter of kernel function. The values of the parameters (γ and σ) are selected by the least square technique.

3 Study Area

The paper investigates vulnerability of the aquifer in the eastern part of Ardabil plain, the Ardabil province, northwest Iran, over an approximately 990 km^2, where the historic city of Ardabil is located. Relevant features of the study area are discussed in detail by Nadiri et al. (2017a, 2017b) but their salient information is outlined below and shown in Figs. 2. The composition of the aquifer in Ardabil plain include: clay, sand and gravel and may be divided into upper and lower aquifers. The upper aquifer is unconfined and multi-layered, where extraction wells are largely drilled and the piezometers are installed for monitoring groundwater levels. Nonetheless, few of the well have penetrated to the lower confined aquifer. In this study area, groundwater withdrawals are from the upper layer through 2622 active pumping wells, 36 qanats and 77 springs.

Dataset: The processing of each of the data layers from their raw values to rated datasets is a tedious process and these have been described by Nadiri et al. (2017a, 2017b). Their processed values are presented in Fig. 3a-3g.
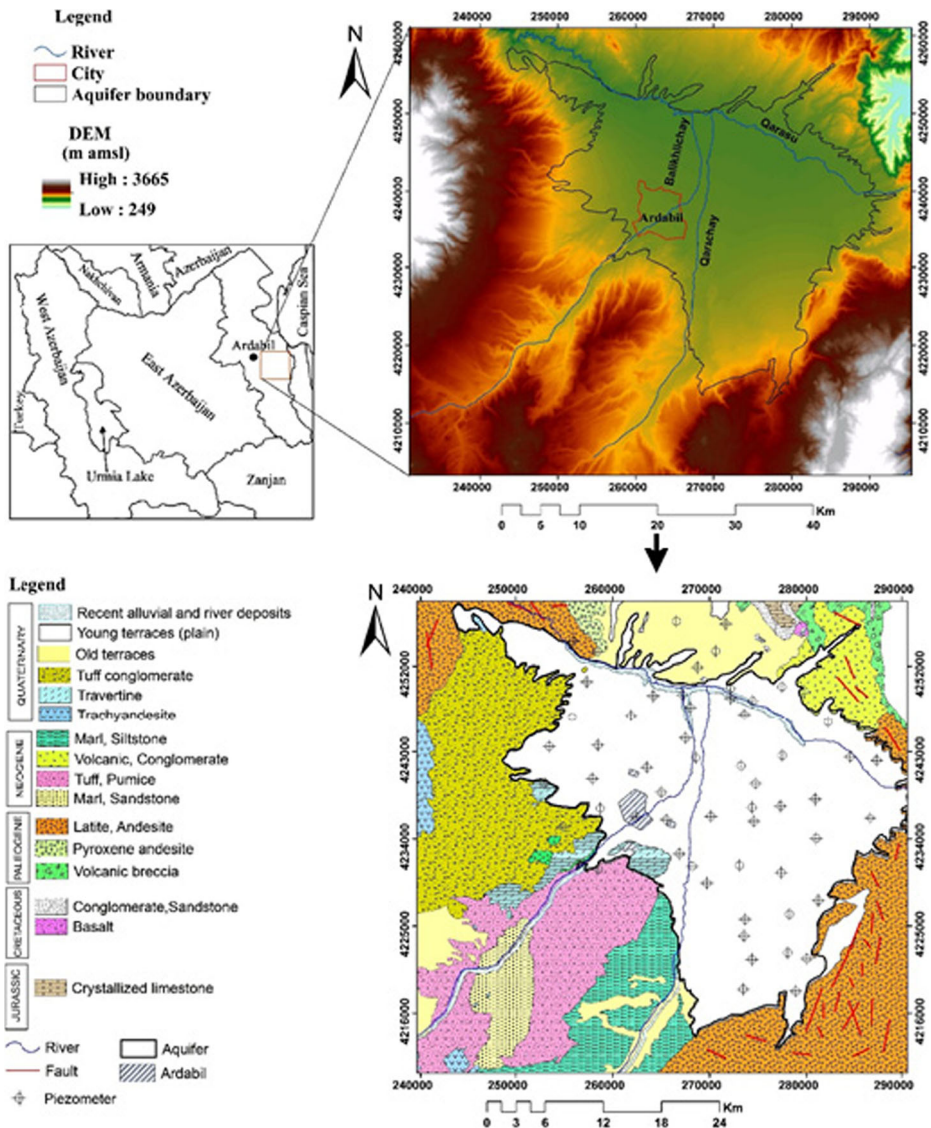
Fig. 2 Location map and geological map of the study area – see also Nadiri et al. (2017a, 2017b)

Nitrate-N: Observed nitrate-N values are measured in mg/L at 65 numbers of wells within the study area (Nadiri et al. 2017a, 2017b) and distributed spatially by the interpolation method of Inverse Distance Weighting, (IDW), although different interpolation techniques were tested to lower Root Mean Square Errors (RMSE). The output is displayed is Fig. 3h. EPA (the US Environmental Protection Agency) specifies the maximum nitrate-N to be 10 mg/L.

The paper uses $NO_3 - N$ concentration values in three ways: (i) absolute concentration values are used to calculate the performance metric of Correlation Index (CI) as given in Section 5; (ii) Conditioned Vulnerability Index (CVI) values as defined by Eq. (3); (iii) CVI-values are used as the target values in SVM-LS.
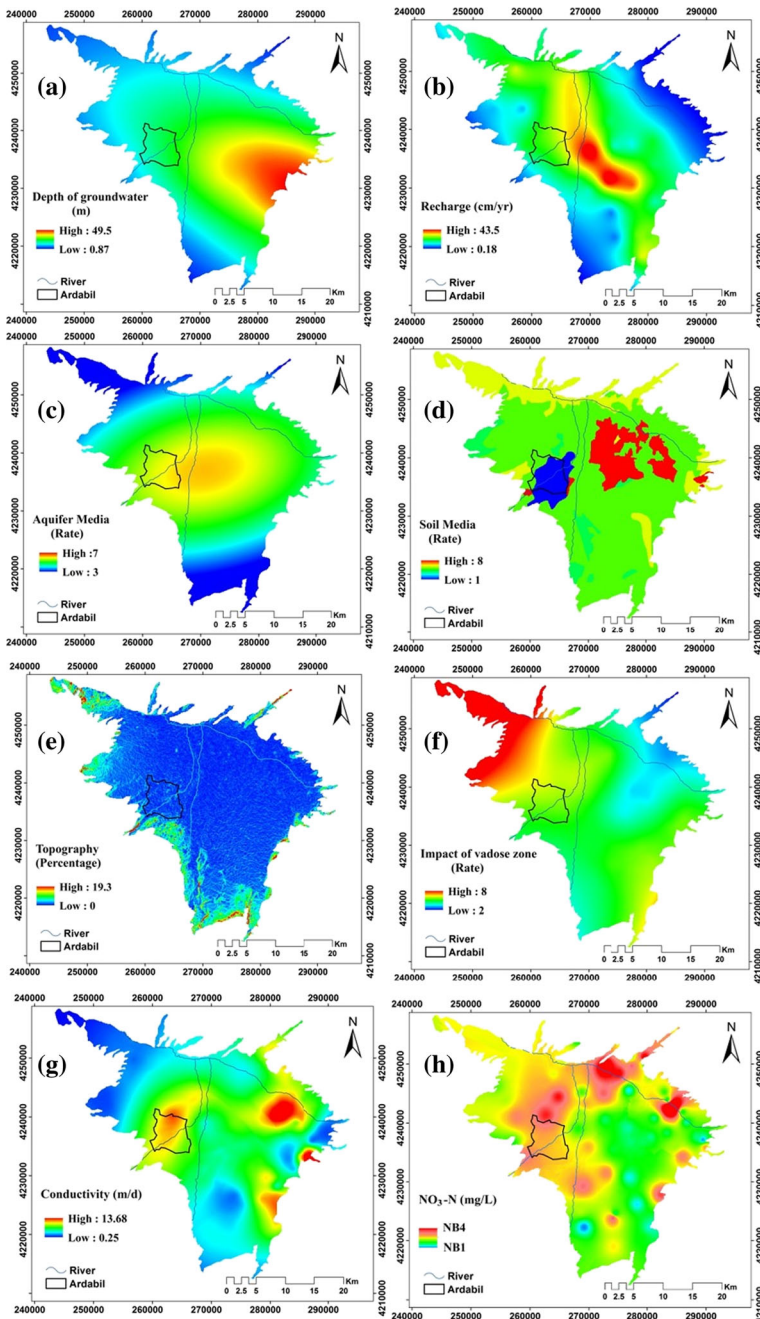
Fig. 3 The DRASTIC data layers: (a) Depth to water table, (b) net Recharge, (c) Aquifer media, (d) Soil media, (e) Topography, (f) Impact of the vadose zone, (g) hydraulic Conductivity, and (h) spatial distribution of NO$_3$-N concentrations

VIs by BDF have the minimum value of 23 and the maximum value of 230. There is no recommended classification for the ranges of the VI values but the paper divides the range

between the minimum and maximum into four bands of Vulnerability Band 1 (VB1: 23–75); Vulnerability Band 2 (VB2: 75–127) and Vulnerability Band 3 (VB3: 127–179) and Vulnerability Band 4 (VB4: 179–230). Measured nitrate-N values, have generally neither a minimum value nor a maximum value but its minimum value for the study area is 0.6 and its maximum value is 42.8 and these have been subdivided into four bands for convenience as follows: Nitrate-N Band 1 (NB1: 0.6–5); Nitrate-N Band 2 (NB2: 5–10); Nitrate-N Band 3: (NB3: 10–30); and Nitrate-N Band 4 (NB4: 30–42.8).

4 Result

The basic spatial unit is grid cell or pixel and the study area is divided into 4000 cells, each 500 m × 500 m. Each grid cell has appropriate data fields for each data layer of the BDF, WDF, BDF-GA, WDF-GA and AIMF. The processing of BDF at each grid cell is carried out through data fields for the rate and weight values, as well as for the calculated intrinsic vulnerability index and target values. These data fields are also the same for WDF, BDF-GA, WDF-GA but each grid cell for AIMF has different data fields identifies its association with training or testing phases, as well as input data fields (i.e. outputs of BDF, WDF, BDF-GA, WDF-GA) and CVIs.

## 4.1 Groundwater Vulnerability by BDF

BDF uses the prescribed rates and weights as specified in Table 1 for each data layers, as presented in Fig. 3 a-g. These were processed using a commercial GIS software to calculate the VI values as per Eq. (1), which are a reproduction of those reported by the authors in a different study (Nadiri et al. 2017b). The values range from 82 to 151 for the whole plain (Fig. 4 a) and are categorised to: VB1, VB2, VB3 and VB4, as specified above.

VI values for the study area by BDF are displayed in Fig. 4 a and their comparative performance is discussed in Section 5. The correlation coefficient between the intrinsic and CVIs is as low as 0.2 at the significance level of <5% (so the null hypothesis is rejected), see also Table 2. This poor correlation is attributed to uncertainties inherent in the values of rates or weights or both and this makes a research case for improving their correlation by seeking variations in the frameworks or by introducing AI processes, as presented below.

4.2 Groundwater Vulnerability by WDF

WDF uses the prescribed weights as in BDF but learns the rates from the data. The procedure is as follows: (i) the rates at each and all grid cells for each data layer are assigned as per BDF; (ii) CVIs are worked for each and all grid cells; (iii) the nitrate-N values corresponding to each data layer are subdivide into the same number of bands; (iv) mean nitrate-N values are estimated for each data layer; (v) the WDF band with the highest nitrate-N mean value is assigned the highest rate as per BDF but the remaining WDF rate values for the lower bands are worked out by the proportionality rule; and (vi) VI values by WDF are calculated by multiplying modified rates at each grid cell and Aller's weights.

The WDF results show that its VI values range from 113 to 177 and they fall to Band 1, Band 2 and Band 3 and Table 2 gives its correlation coefficient to be 0.37 at the significance
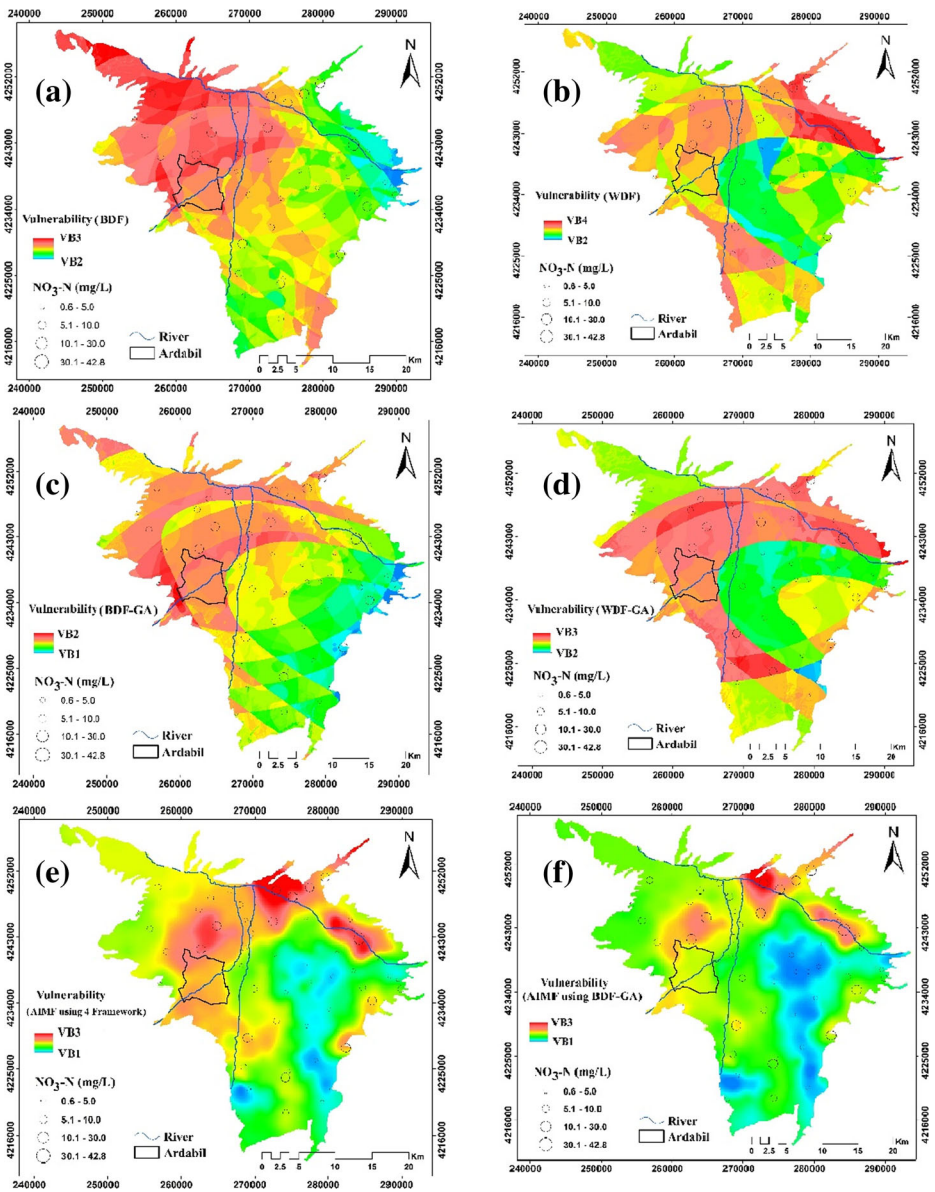
Fig. 4 Vulnerability maps using different methods: (**a**) BDF, (**b**) WDF, (**c**) BDF-GA, (**d**) WDF-GA and (**e**) AIMF using 4 frameworks, (**f**) AIMF using BDF-GA

level of <5% (so the null hypothesis is rejected). VIs are mapped in Fig. 4 b and Section 5 discusses their comparative performance.

4.3 Groundwater Vulnerability by BDF– GA

The framework variation of BDF-GA optimises the weight values using the genetic algorithm (GA) but retains the prescribed rate values based on BDF. A GA optimiser in MATLAB® is

Table 2 Correlation index between nitrate and vulnerability index for BDF, WDF, BDF-GA and WDF-GA, and AIMF model

Method	Vulnerability Bands	Nitrate-N Concentration Bands				CI	r [Note 1]	RMSE		R^2 (see Note 1)	
		NB1	NB2	NB3	NB4			Training	Test	Training	Test
BDF	VB1	0	0	0	0	155	0.2	–	–	–	–
	VB2	10	2	2	1						
	VB3	17	13	9	0						
	VB4	0	0	0	0						
WDF	VB1	0	0	0	0	139	0.37	–	–	–	–
	VB2	1	0	0	0						
	VB3	21	14	8	0						
	VB4	5	1	3	1						
BDF-GA	VB1	12	4	3	0	181	0.38	–	–	–	–
	VB2	15	11	8	1						
	VB3	0	0	0	0						
	VB4	0	0	0	0						
WDF-GA	VB1	0	0	0	0	171	0.45	–	–	–	–
	VB2	21	9	5	0						
	VB3	6	6	6	1						
	VB4	0	0	0	0						
AIMF using 4 Framework	VB1	13	0	0	0	194	0.84	4.09	5.1	0.8	0.72
	VB2	12	15	5	0						
	VB3	2	0	6	1						
	VB4	0	0	0	0						
AIMF [Note 2] using BDF-GA	VB1	11	0	0	0	193	0.83	4.1	5.2	0.76	0.7
	VB2	14	15	4	0						
	VB3	2	0	7	1						
	VB4	0	0	0	0						
Nadiri et al. 2017a	Supervised Intelligence Committee Machine (SICM) using classification of the study area					201	0.96	0.99	1.4	0.99	0.94
Nadiri et al. 2017b	Supervised Committee to Combine Fuzzy Logic (SCFL); no classification of the study area					192	0.93	2.9	4.1	0.9	0.86

- For BDF, the CI value is calculated as:
- 11 wells have a difference of 0 and score 4;
- 25 wells have a difference of 1 in the categories and score 3;
- 18 wells, have a difference of 2 and score 2; and
- 0 well, a difference of 3 and score 1. • CI is calculated as $11 \times 4 + 25 \times 3 + 18 \times 2 + 0 \times 1 = 155$.

Note 1: r is coefficient of correlation and refers to using the values at observation wells; whereas R^2 is coefficient of determination and refers to using the values of all of the grid cells

Note 2: the values of γ and σ are optimised to be 6.14 and 0.022, respectively

Note 3: AIMF runs BDF, WDF, BDF-GA and WDF-GA involve one full level of supervised learning and one level of unsupervised learning

Note 4: SICM and SCFL are models, in which AI runs Multiple Models (AIMM); there are two levels of supervised learning in AIMM

used, which identifies the values of the weights by maximising correlation coefficient between intrinsic VIs and CVI (Eq. 2). The default values and optimised weight values are given in Table 1, which shows they evidently differ from those of BDF.

The calculated VI values for BDF-GA range from 59 to 106, as mapped in Fig. 4c and these fall into two bands, Band 1 and Band 2. The correlation coefficient between intrinsic VI values

by BDF-GA and CVIs is 0.38 at the significance level of <5% (so the null hypothesis is rejected). This is given in Table 2, which reflects a notable improvement sufficient to warrant a variation.

4.4 Groundwater Vulnerability by WDF–GA

WDF-GA uses the rate values learned by WDF but also uses Genetic Algorithm (GA) to optimise the weight values. The optimised weight values for the 7 data layers of WDF-GA are given in Table 1.

The VIs for WDF–GA use rates optimised by GA for each and all the data layers at each and all the grid cells, which are mapped in Fig. 4d. They range from 81 to 133 but the results in the figure are further discussed in Section 5. The VIs fall into the range of VB1 and VB2. The correlation coefficient between VIs and CVIs are calculated to be 0.45 at the significance level of <5% (so the null hypothesis is rejected). This is given in Table 2, which shows a significant improvement compared with the other three sets of results but still not good enough for a defensible decision-making.

4.5 AI-running Multiple Frameworks (AIMF)

The AIMF model based on SVM-LS, which runs the inputs from BDF, WDF, BDF-GA, WDF-GA and produces Vis with respect to matching with the target values of CVIs. The dataset was divided into two sets, comprising training sets using 80% of the dataset selected at random, and testing sets using the remaining 20%. The training phase identifies the values of the two parameters: σ and γ by the least square error method and they are given in Table 2.

The RMSE and R^2 performance criteria are used to estimate the performance metrics of the AIMF model both during the training and testing phases. Table 2 shows the performance results for the training and testing phases. The table shows considerable improvements by SVM over the individual framework results and any decision-making based on AIMF is deemed defensible. Fig. 4 e maps the finalised results but are discussed further in Section 5.

5 Discussion

5.1 Inter-comparison of Results

One way of assessing the comparative performance of the results by four frameworks and by AIMF is to Correlation Index (CI). The calculation of CI makes use of the bands as introduced in Section 3, through the following procedure: (i) calculate the difference between the bands in VI and NO_3-N values for a given observation well; (ii) assign a score of 4 if the difference in between the respective bands is 0 but assign score of 3, 2 and 1 if the differences are of 1, 2 and 3, respectively; (iii) multiply the scores with the band ranks and add the product to produce the CI value for each framework or AIMF. The results are presented in Table 2, which includes an example for the illustration of the calculations.

As per CI results in Table 1, the performances are improved in the order of: WDF, BDF, WDF-GA, BDF-GA and AIMF; whereas the *r-values* suggest the following ranking: BDF, WDF, BDF-GA, WDF-GA and AIMF. However, the authors are not keen on ranking the techniques as such a task renders results of local significance and not applicable elsewhere.

The authors recommend reusing the results through appropriate modelling strategies for further learning, see Nadiri et al. (2013), Tayfur et al. (2014), Khatibi et al. (2017).

5.2 Contextual Overview of the Results

Attention is given to the performance of the four frameworks to identify the possible role of their underlying strategy on the results. BDF is a prescriptive strategy and their inherent learning is limited to a one-time expert opinion without the benefit of any numerical learning from local data or of any periodic review. Not surprisingly, its performance given in Table 2 in terms of correlation coefficient of BDF with nitrate-N values at the observation wells is the poorest and at the border of being rejected. A visual display of the performance of BDF in Fig. 4 a indicates sharp interfaces, patchiness in the behaviour and discontinuity in the vulnerability of the study area but there is no reason to support for sharp interfaces. Notably, published works on DRASTIC do not often quote their performance measures and therefore further comparisons is not possible.

The learning from the data by WDF suggests some improvement on the rating values using the Wilcoxon test procedure despite the conflicts between their CI and r values; see the performance measures in Table 2 and the display of the results in Fig. 4 b. The results are suggestive that the limited learning from data is helpful in terms of improving performance measures and reducing patchiness but the sharp interfaces remain unchanged. Huan et al. (2012) used the geometrical interval method for optimising the rate values in the DRASTIC framework. Their results showed improvements on the correlation between vulnerability index and nitrate concentration, which increased to 0.66 from 0.26 as in the DRASTIC framework.

The improvement by WDF is not sufficient and this makes the research case for a further research and hence the strategy of BDF-GA, i.e. retaining prescriptive rates but learning from the data to improve the weights. The results in Table 2 show some improvements in terms of performance measures and a further reduction of patchiness as displayed in Fig. 4 c. Although WDF-GA aims to learn from data for both rates and weights, the results in Table 2 show that the strategy is not successful as the sharp contrasts in the interfaces are not reduced and the patchiness is amplified. However, Jafari and Nikoo (2016) used the Wilcoxon test with GA and optimised the rate and weight in the DRASTIC framework, in which their results showed the correlation coefficient between nitrate concentration and vulnerability indices increased from 0.57 to 0.82. Notably, the model by Jafari and Nikoo (2016) uses directly supervised learning at Level 1; whereas the results reported by the authors in this study is an unsupervised learning at Level 1 and a supervised learning at Level 2.

The AIMF model takes these four framework results on board and learns from them under supervision by using measured vulnerability as the target values. The results are given in Table 2, which indicates that the improvements are remarkable. Also, Fig. 4 e presents modelled vulnerability values, in which AIMF reduces patchiness and sharp contrasting interfaces.

The success of AIMF displayed in Fig. 4 e gives rise to the question that: is the AIMF strategy dependent on the choice of input framework results? Therefore, a further investigation was carried out by selecting BDF-GA as the only input data. The results are given in Table 2 and displayed in Fig. 4 f, which provide a revealing evidence on the way AIMF is working. A comparison of Figs. 4 e, f with Fig. 3 h shows that AIMF mimics successfully target values but each framework feed as input data simply exerts influence

on the degree of vulnerability. Therefore, as long as the noise in a given framework results does not undermine the information content for VI values, it could and should be used as input data in the AIMF model.

5.3 Contributions towards AI-based Best Practice

Knowledge integration is intended for model performance across published works on the subject. However, the scope for this is limited to the fruitless task of seeking citations that a number of other researchers have arrived at similar conclusions. The problem is that different researchers make some common choices but differ from one another by a wide range of other choices. This leads to stalemate as no algorithm performs persistently the best. Khatibi et al. (2017) address this stalemate and suggest that model/framework reuse may resolve the problem, as discussed below.

A comparison of the results of this research with the authors' recently published research papers is also outlined, as follows. These comprise: (i) Supervised Intelligence Committee Machine (SICM) model (Nadiri et al. 2017a), and (ii) by Supervised Committee to Combine Fuzzy Logic (SCFL) model (Nadiri et al. 2017c). Both of these investigations do not use any framework but their modelling strategy (SICM and SCFL) may be referred to as *AI-running Multiple Models* (AIMM), in which they employ two levels of supervised learning. In contrast AIMF in the paper runs multiple frameworks, which is a supervised learning building on an unsupervised learning. As per Table 2, AIMM (SICM and SCFL) underpins the efficacy of two levels of supervised learning for a further improvement up to %15 and this is attributable to: (i) the levels of supervision; and (ii) SICM performs better than SCFL and AIMF for dividing the study area into three parts to achieve more homogeneous properties in its parts.

The authors are engaged in a program of research with the aim of transforming framework-based aquifer vulnerability mapping practices to next generation practices with maximised information using AI. As outlined in the introduction section, application of AI to VI mapping is already topical but the authors' approach is systematic and with a clear aim of maximising the extracted information from the data layers. The focus so far has been on studying the role of AI and frameworks in a single aquifer (e.g. Nadiri et al. 2017a, 2017c) and multiple aquifers (Nadiri et al. 2017b). Further investigations are currently being carried out to assess the contribution of different techniques of conditioning the values of nitrate-N and VI mapping by learning from impacts of multiple contaminants or from a mix of models and frameworks at the first level. The authors are also investigating the application of the principle modelling developments in other water and environmental problems, e.g. Nadiri et al. (2013), Asadi et al. (2014), and the emerging finding is that remarkable improvements are likely.

6 Conclusion

The paper investigates the role of different strategies of ascertaining the rating and weighting values in DRASTIC-based aquifer Vulnerability Index (VI) mapping. This research produced four frameworks, which are: (i) Basic DRASTIC Framework, (BDF) with prescribed rates and weights; (ii) Wilcoxon-based DRASTIC Framework (WDF), which learns the rates by Wilcoxon test but uses weights as per BDF; (iii) BDF-GA, which uses the rates as per BDF but optimises the values of the weights using Genetic Algorithm; (iii) WDF-GA, which uses

WDF rates and GA-based weights. The primary novelty of the paper is that it formulates a modelling strategy, which uses AI (Artificial Intelligence) to run Multiple Framework (AIMF). AIMF uses Support Vector Machine with Least Squares (SVM-LS) and serves as a supervised learning model for learning from unsupervised multiple frameworks. Learning at two levels is another novel feature of the authors approach to modelling practices, which is open to the formalising and diversifying learning strategies.

The contribution of the paper is that it investigates the application of a supervised AI model to run frameworks with unsupervised learning including one with no learning. The results show that the improvement in the performance metrics is remarkable and can be as much as 40% in this study by the virtue of a further supervised learning. A comparison of this finding with those of similar works is not quite possible, but involving two levels of supervised learning, underpins the efficacy of supervised learning over unsupervised learning.

The paper also contributes towards increased knowledge that frameworks with no learning or with unsupervised learning provide a good-start but their results are unlikely to be defensible for decision-making. However, using supervised Artificial Intelligence (AI) techniques are likely to make the results defensible in decision-making by a further learning of the values of rates and weights from site-specific data.

Compliance with ethical standards

Conflict of Interest None.

References

Aller L, Bennett T, Lehr JH, Petty RJ, Hackett G (1987) DRASTIC: A Standardized System for Evaluating Ground Water Pollution Potential Using Hydrogeologic Settings. EPA 600/2–87-035. U.S. Environmental Protection Agency, Ada, Oklahoma

Asadi S, Hassan M, Nadiri AA, Dylla H (2014) Artificial intelligence modeling to evaluate field performance of photocatalytic asphalt pavement for ambient air purification. Environ Sci Pollut Res 21(14):8847–8857. https://doi.org/10.1007/s11356-014-2821-z

Bai L, Wang Y, Meng F (2012) Application of DRASTIC and extension theory in the groundwater vulnerability evaluation. Water and Environment journal 26(3):381–391

Bello-Dambatta A, Farmani R, Javadi AA, Evans BM (2009) The Analytical Hierarchy Process for contaminated land management. Adv Eng Inform 23(4):433–441. https://doi.org/10.1016/j.aei.2009.06.006

Civita M (1994) Le carte della vulnerabilita` degli acquiferi all'inquinamento Teoria & practica (Aquifer vulnerability maps to pollution) (in Italian). Pitagora Ed, Bologna, p 325

Daly D, Drew D (1999) Irish methodologies for karst aquifer protection. In: Beek B (ed) Hydrogeology and engineering geology of sinkholes and karst. Balkema, Rotterdam, pp 267–327

Davis L (1991) Handbook of Genetic Algorithms. Van Nostrand Reinhold, New York

Foster S (1987) Fundamental concepts in aquifer vulnerability, pollution risk and protection strategy. international conference Noordwijk a Zee, Netherlands, 1–30 Apri 1987

Holland JH (1975) Adaptation in natural and artificial systems. University of Michigan, Ann Arbor

Huan H, Wang J, Teng Y (2012) Assessment and validation of groundwater vulnerability to nitrate based on a modified DRASTIC model: A case study in Jilin City of northeast China. Sci Total Environ 440:14–23. https://doi.org/10.1016/j.scitotenv.2012.08.037

Jafari SM, Nikoo MR (2016) Groundwater risk assessment based on optimization framework using DRASTIC method. Arab J Geosci 9:742. https://doi.org/10.1007/s12517-016-2756-4

Kauffman LJ, Chapelle FH (2010) Relative vulnerability of public supply wells to VOC contamination in hydrologically distinct regional aquifers. Ground Water Monit Remediat 30:54–63. https://doi.org/10.1111/j.1745-6592.2010.01308.x

Khatibi R, Ghorbani MA, Akhoni Pourhosseini F (2017) Stream flow predictions using nature-inspired Firefly Algorithms and a Multiple Model strategy – Directions of innovation towards next generation practices. Journal of Advanced Engineering Informatics. http://www.sciencedirect.com/science/article/pii/S1474034617301271

McLay CDA, Dragden R, Sparling G, Selvarajah N (2001) Predicting groundwater nitrate concentrations in a region of mixed agricultural land use: a comparison of three approaches. Environ Pollut 115:191–204

Mitchell M (1996) An Introduction to Genetic Algorithms. Massachusetts Institute of Technology

Nadiri AA, Fijani E, Tsai FT-C, Moghaddam AA (2013) Supervised committee machine with artificial intelligence for prediction of fluoride concentration. J Hydroinf 15(4):1474–1490. https://doi.org/10.2166/hydro.2013.008

Nadiri AA, Gharekhani M, Khatibi R, Asghari Moghaddam A (2017b) Assessment of groundwater vulnerability using supervised committee to combine fuzzy logic models. Environ Sci Pollut Res 24(9):8562–8577. https://doi.org/10.1007/s11356-017-8489-4

Nadiri AA, Gharekhani M, Khatibi R, Sadeghfam S, Asghari Moghaddam A (2017a) Groundwater vulnerability indices conditioned by Supervised Intelligence Committee Machine (SICM). Sci Total Environ 574:691–706. https://doi.org/10.1016/j.scitotenv.2016.09.093

Nadiri AA, Sedghi Z, Khatibi R, Gharekhani M (2017c) Mapping vulnerability of multiple aquifers using multiple models and fuzzy logic to objectively derive model structures. Sci Total Environ 593-594:75–90. https://doi.org/10.1016/j.scitotenv.2017.03.109

Najib S, Grozavu A, Khalid Mehdi KM, Breabăn IG, Guessir H, Boutayeb K (2012). Application of the method GALDIT for the cartography of groundwaters vulnerability: aquifer of Chaouia coast (Morocco). Analele Stiintifice ale Universitatii "Al. I. Cuza" din Iasi. Serie Noua. Geografie, 2012, vol. 58, No 2, p. 77

Ribeiro L (2000) Desenvolvimento de um ıʼndice para avaliar a susceptibilidade. ERSHA-CVRM, 8

Rupert MG (1999) Improvements to the DRASTIC groundwater vulnerability mapping method. US Geol Surv Fact Sheet FS-066-99

Sadeghfam S, Hassanzadeh Y, Nadiri AA, Zarghami M (2016a) Localization of groundwater vulnerability assessment using catastrophe theory. Water Resour Manag 30(13):4585–4601. https://doi.org/10.1007/s11269-016-1440-5

Sadeghfam S, Hassanzadeh Y, Nadiri AA, Khatibi R (2016b) Mapping groundwater potential field using catastrophe fuzzy membership functions and Jenks optimization method: a case study of Maragheh-Bonab plain, Iran. Environmental Earth Sciences 75(7):545. https://doi.org/10.1007/s12665-015-5107-y

Sorichetta A, Masetti M, Ballabio C, Sterlacchini S (2012) Aquifer nitrate vulnerability assessment using positive and negative weights of evidence methods, Milan, Italy. Comput Geosci 48:199–210

Suykens JAK (2000) Least squares support vector machines for classification and nonlinear modelling. Neural Network World. Special Issue on PASE 2000, 10 (1–2), 29–48

Suykens JAK, Van GT, Brabanter JD, De MB, Vandewalle J (2002) Least Squares Support Vector Machines. World Scientific Publishing, Singapore

Tayfur G, Nadiri AA, Moghaddam AA (2014) Supervised intelligent committee machine method for hydraulic conductivity estimation. Water Resour Manag 28(4):1173–1184. https://doi.org/10.1007/s11269-014-0553-y

Tesoriero AJ, Inkepan EL, Voss FD (1998) Assessing groundwater vulnerability using logistic regression. In: proceedings for the source water assessment and protection 98 conference. Dallas:157–165

USEPA (2012) National Primary Drinking Water Regulations. US Environmental Protection Agency, EPA816-F-09-004

Van Stemproot D, Evert L, Wassenaar L (1993) Aquifer vulnerability index: a GIS compatible method for groundwater vulnerability mapping. Canadian Water Resources Journal 18:25–37. https://doi.org/10.4296/cwrj1801025

Wilcoxon F (1945) Individual comparisons by ranking methods. Biom Bull 1:80–83. https://doi.org/10.2307/3001968

Wu W, Yin S, Liu H, Chen H (2014) Groundwater Vulnerability Assessment and Feasibility Mapping Under Reclaimed Water Irrigation by a Modified DRASTIC Model. Water Resour Manag 28(5):1219–1234. https://doi.org/10.1007/s11269-014-0536-z