

Application of machine-learning models for diagnosing health hazard of nitrate toxicity in shallow aquifers

Fatemeh Karandish¹ · Abdullah Darzi-Naftchali² · Ahmad Asgari²

Received: 23 January 2016/Revised: 10 July 2016/Accepted: 21 July 2016/Published online: 8 August 2016
© The International Society of Paddy and Water Environment Engineering and Springer Japan 2016

Abstract There is a growing concern about health hazards linked to nitrate (NO_3) toxicity in groundwater due to overuse of nitrogen fertilizers in rice production systems of northern Iran. Simple-cost-effective methods for quick and reliable prediction of NO_3 contamination in groundwater of such agricultural systems can ensure sustainable rural development. Using 10-year time series data, the capability of adaptive neuro-fuzzy inference system (ANFIS) and support vector machine (SVM) models as well as six geostatistical models was assessed for predicting NO_3 concentration in groundwater and its noncarcinogenic health risk. The dataset comprised 9360 water samples representing 26 different wells monitored for 10 years. The best predictions were found by SVM models which decreased prediction errors by 42–73 % compared with other models. However, using well locations and sampling date as input parameters led to the best performance of SVM model for predicting NO_3 with $\text{RMSE} = 4.75\text{--}8.19 \text{ mg l}^{-1}$ and $\text{MBE} = 3.3\text{--}5.2 \text{ mg l}^{-1}$. ANFIS models ranked next with $\text{RMSE} = 8.19\text{--}25.1 \text{ mg l}^{-1}$ and $\text{MBE} = 5.2\text{--}13.2 \text{ mg l}^{-1}$ while geostatistical models led to the worst results. The created raster maps with SVM models showed that NO_3 concentration in 38–97 % of the study area usually exceeded the human-affected limit of 13 mg l^{-1} during different seasons. Generally, risk probability went beyond 90 % except for winter when groundwater quality was safe from nitrate viewpoint.

Noncarcinogenic risk exceeded the unity in about 1.13 and 6.82 % of the study area in spring and summer, respectively, indicating that long-term use of groundwater poses a significant health risk to local resident. Based on the results, SVM models were suitable tools to identify nitrate-polluted regions in the study area. Also, paddy fields were the principal source of nitrate contamination of groundwater mainly due to unmanaged agricultural activities emphasizing the importance of proper management of paddy fields since a considerable land in the world is devoted to rice cultivation.

Keywords ANFIS · Paddy fields · Noncarcinogenic risk · Ordinary kriging · SVM · Risk probability

Introduction

In many parts of the world, agriculture mostly relies on groundwater resources which are generally exposed to higher risk of contamination due to excessive depletion of low water stocks and rapidly enhanced human activities such as excessively unmanaged application of fertilizers and pesticides, producing urban sewage and industrial wastes. These contamination might render groundwater unsuitable for consumption and put human and animal life as well as the whole environment at great risk (Babiker et al. 2004).

Having high solubility and weak adsorption to soil particles, nitrate is the most frequently introduced pollutant into groundwater systems (Spalding and Exner 1993) since it accompanies the largest life/death cycle in the biosphere. Intensive agriculture is usually the main source of nitrate contamination due to unmanaged and excessive application of chemical fertilizers which alter the balance of nitrogen

✉ Fatemeh Karandish
f.karandish@uoz.ac.ir; karandish_h@yahoo.com

¹ Water Engineering Department, University of Zabol, Zabol, Iran

² Water Engineering Department, Sari Agricultural Sciences and Natural Resources University, Sari, Iran

compounds in soil and introduce the excessive nitrogen to the soil and consequently to the groundwater. High levels of nitrate in groundwater originating mainly from agricultural activities were encountered in almost all regions of the world (e.g., Spalding and Exner 1993; Dudley et al. 2008; Burow et al. 2010; Melo et al. 2012; Dahan et al. 2014). The adverse health effects of high nitrate levels in drinking water are well documented (Mousavifazl et al. 2013; Sundaraiah and Sudarshan 2014). Groundwater with nitrate concentration exceeding the threshold of $3 \text{ mg l}^{-1} \text{ NO}_3^- \text{-N}$ or $13 \text{ mg l}^{-1} \text{ NO}_3^-$ is considered due to human activities (the so-called human-affected value; Burkart and Kolpin 1993; Eckhardt and Stackelberg 1995). However, the maximum acceptable concentration of nitrate for potable water according to the World Health Organization (WHO) is $11.3 \text{ mg l}^{-1} \text{ NO}_3^- \text{-N}$ or $50 \text{ mg l}^{-1} \text{ NO}_3^-$. Thus, exploring the contaminated area and proposing suitable solution for sustaining the groundwater resources are essential.

The first step in planning and management of groundwater resources is assessing and predicting water quality. Such assessment requires reliable quantitative models which led to developing a lot of physical and mathematical models. Conceptual or physical models have been widely used to simulate water and solute transport in subsurface environment during past decades. Nevertheless, the lack of precise data diminishes the reliability of these models (Almasri and Kaluarachchi 2005). Machine-learning models such as adaptive neuro-fuzzy interference (ANFIS) and support vector machines (SVM) might be better alternatives since these models acquire implicit knowledge among dataset without requiring the knowledge of mathematical relationship between the inputs and the corresponded outputs as well as explicit characterization and quantification of physical properties and conditions (Almasir et al. 2005). Even under limited data, machine-learning models provide methods for quick and flexible estimation aimed at achieving high level of generalization and prediction accuracy (Khalil et al. 2005). Among machine-learning models, SVM are a new structure which were introduced as robust and significant learning tool for classification and regression problems. It is a novel inductive rule for learning from a finite dataset and has shown good performance with small samples (Liu et al. 2010). Although machine-learning models have been widely used for simulating hydrological process of surface water resources during past decades (Jiang and Cotton 2004; Ahmad and Simonovic 2005; Elshorbagy and Parasuraman 2008; Zou et al. 2010; Dai et al. 2011; Asefa et al. 2006; Yu and Liong 2007; Lin et al. 2009; Liu et al. 2010; Deng et al. 2011), they have rarely been used for assessing groundwater quality (Khalil et al. 2005).

Geostatistical modeling is another method for assessing ground water quality. This method is based on the theory of regionalized variables (Goovaerts 1997), and is generally preferred for such assessment, because it allows to take into account spatial correlation between neighboring observations to predict attribute values at un-sampled locations. Since the pioneering work of Delhomme (1976), geostatistics has been widely employed in hydrogeology to obtain maps of peizometric surface, to estimate transmissivity or hydraulic conductivity, to create iso-concentration maps of groundwater contaminations, to produce iso-probability maps that the concentration of a specific contaminant exceeds a threshold and as tool for inverse problem solutions (Assaf and Saadeh 2006). Kriging method has been reported as a suitable method for creating the spatial distribution of nitrate concentration in contaminated groundwater (Babiker et al. 2004; Castrignano et al. 2008; Assaf and Saadeh 2006). However, the other geostatistical interpolation methods such as weighting moving average and cokriging methods have less been considered.

Literature review indicates a lack of a comprehensive study dealing with the evaluation of different fast-cost-effective methods for assessing nitrate toxicity in shallow aquifers under point- and non-point-sources of pollution. Therefore, the present research was designed to clarify the performance of both machine-learning and geostatistical modeling for predicting seasonal variations of nitrate toxicity and its associated noncancerous health risk in shallow aquifers under both point- and non-point-sources pollution. Based on long-term time series data, ANFIS and SVM models as well as six geostatistical models including ordinary Kriging (OK), Ordinary Cokriging with the covariates of Ca^{2+} concentration (OCK1) and total hardness (OCK2) and weighting moving average with powers of 2 (WMA2), 4 (WMA4) and 6 (WMA6) were employed for diagnosing the negative impacts of human activities on polluting groundwater resources.

Materials and methods

Site description and sampling

The experimental site located within 653794–679951 m longitudes and 4030314–4064326 m latitudes in Mazandaran province, Northern Iran (Fig. 1a, b). In this part of the province, the long-term average annual rainfall is 700 mm about 70 % of which occurs over the October–March period (Karandish 2016). Long-term annual minimum, maximum, and average temperatures of the study area are about 12.5, 21.5, and 17 °C, respectively. Paddy

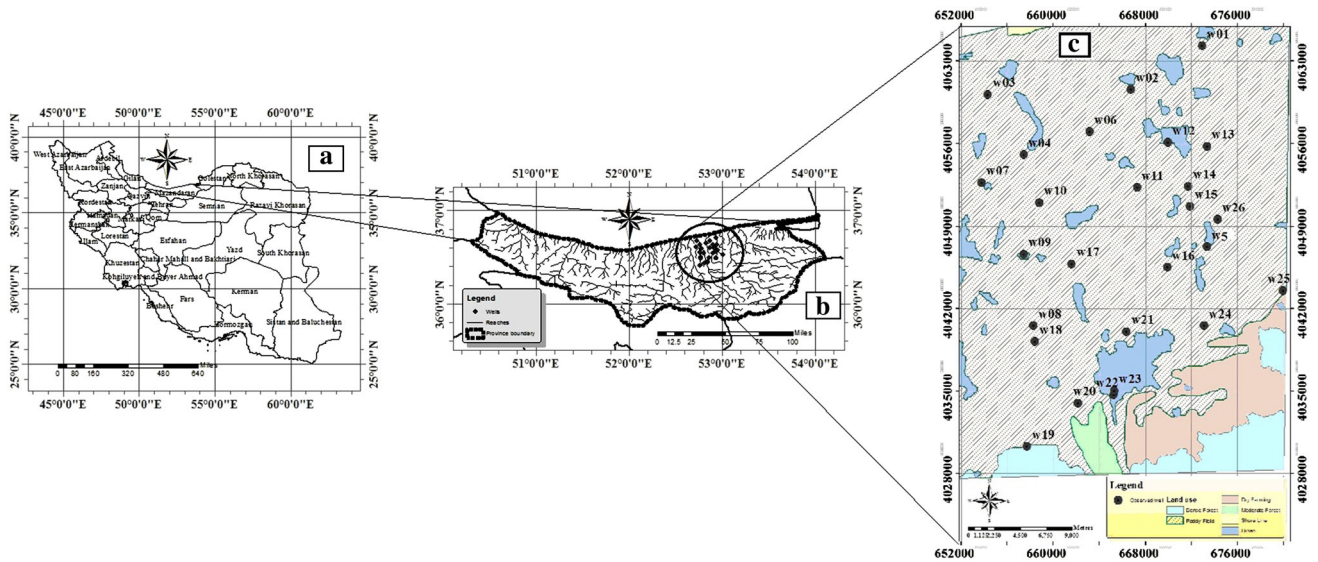


Fig. 1 The location of the study area in Iran (a), in Mazandaran province (b), land use classes and location of monitoring wells (w01–w 26) (c)

fields are dominant in crop sown area in the study area (Fig. 1c).

Groundwater in 26 monitoring wells distributed across the study area (Fig. 1c) were sampled three times a month during 10 years. Water samples were analyzed at the Water Quality Laboratory of Mazandaran Regional Water Company for Nitrate concentration (NO₃) and the 11 physical and chemical properties of water were also analyzed. Pearson correlation analysis was performed between the measured chemical parameters to identify possible relationships and to define covariates for predicting methods. The log normalized data were subjected to six geostatistical methods including: weighting moving average (WMA) with powers of two (WMA-2), four (WMA-4) and six (WMA-6), ordinary kriging (OK) and Co-kriging with two covariates of Ca²⁺ concentration (OCK1) and TH (OCK2). In addition, the capability of two machine-learning methods including adaptive neuro-fuzzy inference systems (ANFIS) and SVM for simulating NO₃ concentration in groundwater was assessed. The simulated and observed data were analyzed by considering the cropping season of major crop (rice) grown in the study area so that a year was divided into four periods including preplanting period (winter), cropping season (spring and summer), and the postharvest period (autumn). Based on the national standards of Iran, seasons were defined as spring (April, May, and June), summer (July, August, and September), autumn (October, November, and December), and winter (January, February, and March).

Geostatistical analysis approach

In classic statistical analysis, samples are treated as if they are stripped of spatial dimension. In geostatistics, however,

the location of a data point is considered in conjunction to its value. Semivariogram is a measure of spatial correlation of a given variable. The experimental semi variogram (γ) is calculated by (Issaks and Srivastava 1989):

$$\gamma(h) = \frac{1}{2 \cdot n(h)} \sum_{i=1}^{n(h)} [Z(x_i + h) - Z(x_i)]^2 \quad (1)$$

where n (h) is the number of sample pairs separated by distance h, Z(x_i) is the measured value at location x_i, and Z(x_i + h) is the measured value at distance h from x_i. Range (R) is the distance where γ reaches a constant value. The value of semivariogram at R is defined as sill. In theory, sill value is equal to the sample variance under the assumption of second-order stationary. The value of γ at h = 0 in the semivariogram is called the nugget effect (C₀). γ is used in kriging methods for estimating the weight parameters and deriving the spatial distribution of estimation variance. A detailed presentation of geostatistical theories can be found in Goovaerts (1997), Chiles and Delfiner (1999), and Webster and Oliver (2006).

ANFIS

In 1993, Jung introduced ANFIS considering the capabilities of the fuzzy theory and the neural nets (Tang et al. 2005). A popular teaching method in neuro-fuzzy systems is the fuzzy inference system, which uses the hybrid learning algorithms to identify the fuzzy system parameters and teach the model as well (Rehman and Mohandes 2008). ANFIS model is a five-layer structure, which is the result of adding the fuzzy logical models to the artificial neural net:

In the layer 1 or input layer, the membership degree of the nodes entering different fuzzy periods is determined by using the membership functions (MFs). The shape of the membership function and the amount of their overlapping are optional, and are determined by Eq. (2):

$$\mu_A(x) = \frac{1}{1 + \left| \frac{x-c_i}{a_i} \right|^{2b_i}} \tag{2}$$

where x is input, and a , b , and c are the comparative parameters and the nonlinear coefficients of this equation, which determine the shape of the membership function. The set of the fuzzy variable coefficients is called S1 set or the left-handed set (LHS). The output amounts of the first layer show the membership amount for each input regarding the different membership functions of the inputs.

Layer 2 is the result of multiplying the input amounts by the nodes, and finally the firing strength. For example, for the first node, we have

$$w_i = \mu_{A_1}(x_1)\mu_{B_1}(x_2). \tag{3}$$

Nodes of layer 3 normalize the firing strength as follows:

$$\bar{w} = \frac{w_i}{\sum_{i=1}^n w_i}, \quad i = 1, 2, \dots, n \tag{4}$$

where n is the number of nodes in each layer.

Layer 4 is the terms layer in which terms are achieved. These terms are the results of operation on the input signals into this layer:

$$Z_1 = \bar{w}_1 f_1 = \bar{w}_1 (p_1 x_1 + q_1 x_2 + r_1) \tag{5}$$

where r_1 , q_1 , and p_1 are the consequent parameters.

Layer 5 is the last layer of the net which includes only one node and is calculated via adding all input amounts into its total output as Eq. (6):

$$\sum_{i=1}^n \bar{w}_i f_i \tag{6}$$

Detailed description of the structure of a ANFIS model and the calculation procedures were discussed in Dastirani et al. (2010). One of the features of every ANFIS model is the type of function considered for the model inputs. In this research, different membership functions were employed. In addition, different numbers of MFs were tried in each application, and the best one giving the minimum errors was selected.

SVM

Being introduced by Vapnik (1995), SVM are a classifier derived from statistical learning theory. The SVM can be used both for classification and regression problems and can

be represented as two-layer networks where the weights are nonlinear in the first layer and linear in the second layer (Bray and Han 2004). Support vector regression (SVR) is used to describe regression with SVM in the literature. The regression estimation with SVR is to estimate a function according to a given dataset $\{(x_i, y_i)\}_n$, where x_i denotes the input vector; y_i denotes the output value; and n is the total number of datasets (Tabari et al. 2012). Herein, the input vectors (x_i) refer independent variables whereas the target values (y_i) refer to nitrate concentration. The linear regression function uses the following function:

$$f(x) = \omega \cdot \phi(x) + b \tag{7}$$

where $\phi(x)$ is a nonlinear function by which x is mapped into a feature space, b and ω are a weight vector and a coefficient that should be estimated from the data. Linear regression is performed in a high dimensional feature space via a nonlinear mapping. The coefficients b and x are estimated by minimizing the sum of the empirical risk and a complexity term.

$$R = c \sum_{i=1}^n L_\epsilon(f(x_i), y_i) \frac{1}{2} \|\omega\|^2 \tag{8}$$

$$L_\epsilon(f(x_i), y_i) = \begin{cases} 0 & \text{for } |f(x_i) - y_i| < \epsilon \\ f(x) - y - \epsilon & \text{otherwise} \end{cases} \tag{9}$$

where c is a positive constant (implied as additional capacity control parameter) that determines the trade-off between the model complexity and the amount up to which error larger than ϵ are tolerated, $\|\omega\|^2$ is the regularization term which denotes the Euclidean norm, and L_ϵ is called ϵ -insensitive loss function that measuring the empirical risk and has the advantage that we will not need all the input data for describing the regression vector x .

As shown in Eq. (12), the loss function is equal to 0 if the difference between the predicted $f(x)$ and the measured value y_i is less than ϵ . The choice of ϵ is easier than the choice of c and it is often given as a desired percentage of the output values y_i . So, a nonlinear regression function is given by a function that minimizes Eq. (11) subject to Eq. (12), as in the following expression:

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^n (\alpha_i - \alpha^*) k(x_i, x) + b \tag{10}$$

With $\alpha_i \alpha^* = 0$, $\alpha_i \alpha^* \geq 0, i = 1, \dots, N$ and the kernel function $k(x_i, x)$ describes the inner products in the D -dimensional feature space.

$$k(x, y) = \sum_{i=1}^D \phi_j(x) \phi_i(y) \tag{11}$$

It should be mentioned that the features ϕ_j need not be computed; rather what is needed is the kernel function that

is very simple and has a known analytical form. In this study, linear, polynomial, radial basis function, and sigmoid kernels were used. The best kernel was determined by a trial and error process. The coefficients $\alpha_i\alpha^*$ are obtained by maximizing the following form:

$$\begin{aligned}
 R(\alpha_i\alpha^*) &= -\epsilon \sum_{i=1}^N (\alpha_i + \alpha^*) + \sum y_i (\alpha^* - \alpha_i) \\
 &\quad - \frac{1}{2} \sum_{j=1}^N (\alpha^* + \alpha_i) \times (\alpha^* + \alpha_i) k(x_i, y_i) \\
 &\quad \times \text{subject to } \sum_{i=1}^N (\alpha^* - \alpha_i) = 0, \quad 0 \leq \alpha_i\alpha^* \leq C
 \end{aligned}
 \tag{12}$$

Only a number of coefficients $\alpha_i\alpha^*$ will be different from zero, and the data points associated to them are called support vectors (Mohandes et al. 2004; Kisi and Cimen 2009; Zhou et al. 2009).

Criteria indices

Prediction performances were assessed by cross-validation. Also, mean bias error (MBE) and root mean square error (RMSE) were calculated for comparing different interpolation models as follows:

$$\text{MBE} = \frac{\sum_{i=1}^n (Z(x_i) - Z^*(x_i))}{n}
 \tag{13}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (Z^*(x_i) - Z(x_i))^2}{n}}
 \tag{14}$$

where $Z(x_i)$ and $Z^*(x_i)$ are the observed and estimated values of NO_3 concentration, respectively, and n is the number of data points.

Health risk assessment

Risk assessment is defined as the process of estimating the probability of occurrence of an event and the probable magnitude of adverse health effects over a specified time period (Gao et al. 2012). This research used the HRA model suggested by the United States Environmental Protection Agency (EPA, 1989), including mathematical models of noncarcinogenic risks caused by a single factor (Ni et al. 2010), to calculate the health risk of groundwater for drinking water supply with respect to nitrate concentration. The noncarcinogenic chronic toxic property of chemical pollutants in the human body takes the reference dose as a yardstick: people whose exposure level is higher than the reference dose are probably at risk; those whose exposure level is equal to or lower than the reference dose are less likely to be at risk. Potential noncarcinogenic risk for exposure to contaminant was evaluated as follows (EPA, 1989):

$$\text{HQ} = \text{CDI}/\text{Rfd}
 \tag{15}$$

where HQ is hazard quotient (unitless), Rfd is the reference dose ($\text{mg kg}^{-1} \text{day}^{-1}$) and CDI is the exposure dose rate ($\text{mg kg}^{-1} \text{day}^{-1}$) representing the daily intake of assessed pollutant per kg weight of a person. Rfd for nitrate was $1.6 \text{ mg kg}^{-1} \text{day}^{-1}$. CDI was calculated as follows:

$$\text{CDI} = (C \times \text{IR} \times \text{ED} \times \text{EF})/(\text{BW} \times \text{AT})
 \tag{16}$$

where C is the nitrate concentration in groundwater (mg l^{-1}); RI is the drinking rate (L day^{-1}), representing a person’s daily amount of drinking water; EF is the exposure frequency, representing the days of assessed pollutants intake per year in the evaluation period (day year^{-1}); ED is the exposure duration (year, the value recommended by the U.S. EPA is 30 years), representing the years of lifelong intake of assessed pollutants; BW is the average weight of human bodies (kg); and AT is the average time (d, the average noncarcinogenic time is $\text{ED} \times 365$ days). According to Gao et al. (2012), $\text{IR} = 2 \text{ L day}^{-1}$, $\text{ED} = 30$ years, $\text{EF} = 365 \text{ day year}^{-1}$, $\text{BW} = 70 \text{ kg}$

Results and discussion

Data quality check and exploration

The descriptive statistics of NO_3 , total hardness (TH), and calcium concentrations (Ca^{2+}) data are presented in Table 1. Nitrate concentration in 9360 samples ranged from 0.1 to 120.05 mg l^{-1} . On average, the concentration of NO_3 in groundwater was generally below the permissible limit of 50 mg l^{-1} as drinking health standard while it was exceeded the human-affected value of $13 \text{ mg l}^{-1} \text{NO}_3$ most of the study period. Average of NO_3 concentrations of groundwater samples collected during winter seasons was 17.5 mg l^{-1} , consistently lower than their for the other seasons (23.7, 36.8, and 29.6 mg l^{-1} related to spring, summer, and autumn seasons, respectively). Nutrient and water management and cultural practices, timing and amount of precipitation, water table depth, and differences in soil types and nitrate uptakes by plants are major factors affecting nitrate concentration in groundwater. The maximum and minimum sample variances belonged to summer (736.85) and spring (422.9) period, respectively.

All datasets exhibit non-normal distribution; particularly in the spring (Skewness = 2.36, Kurtosis = 6.63) and winter (Skewness = 2.88, Kurtosis = 9.09) datasets. Previous researches indicate that nitrate concentrations follow a log-normal distribution. So, the data were log-transformed prior to the calculation of semivariogram in the geostatistical modeling to make the data normally

Table 1 Descriptive statistics of TH, NO₃, and Ca²⁺ in groundwater

Parameter	Month		Statistical parameter						
			Mean (mg l ⁻¹)	SD (mg l ⁻¹)	Variance (mg l ⁻¹) ²	Min (mg l ⁻¹)	Max (mg l ⁻¹)	Skewness	Kurtosis
NO ₃	Spring	–	23.68	20.56	422.9	4.87	103.22	2.36	6.63
		Log	2.887	0.752	0.566	1.58	4.64	0.15	–0.38
	Summer	–	36.83	27.145	736.85	6.65	120.05	1.26	1.56
		Log	3.341	0.77	0.5927	1.89	4.79	–0.14	–0.89
	Autumn	–	29.615	26.49	701.6	0.1	108.54	1.13	1.14
		Log	3.035	1.033	1.067	0.74	4.71	–0.45	–0.81
	Winter	–	17.465	21.09	444.6	1.97	104.11	2.88	9.09
		Log	2.412	0.922	0.849	0.57	4.65	0.43	–0.15
Ca ²⁺	Spring	–	125.38	47.6	2265.53	46	316	2.38	8.16
		Log	4.776	0.325	0.1124	3.83	5.76	0.07	2.96
	Summer	–	138.1538	48.53	2355.18	76	296	1.82	3.11
		Log	4.881	0.298	0.0889	4.33	5.69	0.97	1.09
	Autumn	–	135	47.51	2257.36	78	310	1.99	5.16
		Log	4.857	0.303	0.092	4.36	5.74	0.76	1.07
	Winter	–	150.46	53.69	2883.3	66	332	1.65	3.43
		Log	4.962	0.322	0.1036	4.19	5.81	0.36	1.07
TH	Spring	–	521.34	232.48	54047.11	305	1260	2.03	3.76
		Log	6.186	0.358	0.128	5.72	7.14	1.15	1.01
	Summer	–	567.12	216.59	46914.35	345	1125	1.05	0.18
		Log	6.278	0.35	0.1226	5.84	7.03	0.56	–0.78
	Autumn	–	537.88	216.42	46836.3	325	1225	1.46	2.04
		Log	6.222	0.355	0.1261	5.78	7.11	0.68	–0.26
	Winter	–	571.35	257.2	66153.12	325	1315	1.58	1.57
		Log	6.272	0.377	0.1419	5.78	7.18	0.98	0.11

distributed and satisfy assumptions of constant variability. The minimum and maximum SD corresponded to spring and autumn, respectively, after normalization.

Geostatistical modeling

Structural analysis

By superimposing trials of various models with different combinations of model parameters including nugget, sill and range of influence, the best approximation of the semivariogram was selected based on root sum square (RSS) index. A spherical model was selected for summer and autumn dataset while the spring and winter semivariogram was fitted with an exponential model (Fig. 2). Spherical model was advised in the results of other investigation for groundwater nitrate concentrations (Babiker et al. 2004; Castrignano et al. 2008; Assaf and Saadeh 2006).

The selected semivariogram for different periods and their components are presented in Table 2. The

semivariogram for autumn had the highest sill which could be attributed to the highest sample variance after normalization and relatively large amount of maximum NO₃ concentration (Table 1). On the contrary, the sill of the semivariogram model for spring was the smallest which is in consistent with lower sample variance and lower amount of maximum nitrate concentration for spring. Spatial dependence of groundwater nitrate concentration can be classified according to nugget to sill ratio, known as Cambardella Index (CI) (1994) with a ratio of CI < 0.25, CI 0.25–0.75 and CI > 0.75 indicating a strong, moderate, and weak spatial dependence, respectively. Nitrate concentration in groundwater had a moderate spatial structure according to CI index (CI 0.26 for spring; 0.54 for Autumn and 0.5 for winter) with *R* ranging from 2310 m (spring) to 43470 m (winter). Exception was for summer when NO₃ had strong spatial dependency with *R* of 12540 m (Table 2). The relatively well spatial continuity is a reflection of the high stability and mobility of nitrates in groundwater, which facilitate the migration of nitrates untransformed well beyond their source of input given the

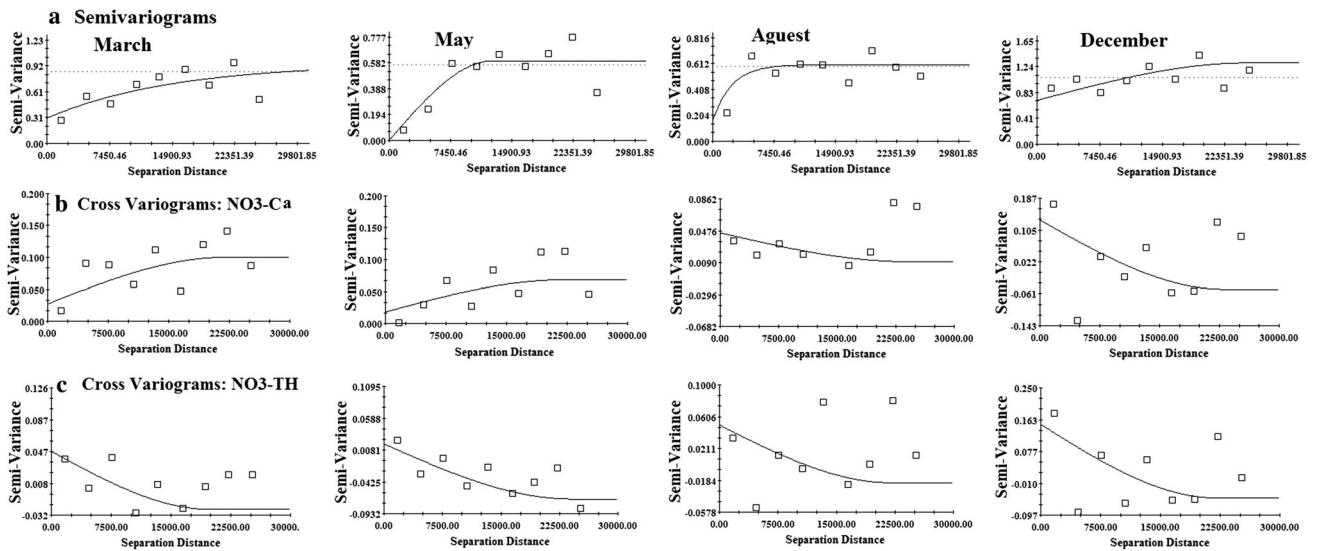


Fig. 2 Semivariograms (a) and Cross-Variograms for $\text{NO}_3\text{-Ca}^{2+}$ (b) and $\text{NO}_3\text{-TH}$ (c) for different periods

Table 2 Parameters of semivariogram and cross-variogram models for different measuring periods

Parameter	Period	Model	C_0	$C_0 + C$	R	$CI = C_0/(C_0 + C)$
NO_3^-	Spring	Exponential	0.16	0.605	2310	0.26
	Summer	Spherical	0.001	0.598	12,540	0.002
	Autumn	Spherical	0.7	1.3	25,000	0.54
	Winter	Exponential	0.57	1.141	43,470	0.50
$\text{NO}_3^- \text{-Ca}^{2+}$	Spring	Spherical	0.018	0.0682	22,000	0.26
	Summer	Exponential	0.045	0.01	24,000	4.5
	Autumn	Spherical	0.13	-0.049	23,000	-2.65
	Winter	Spherical	0.027	0.1	22,000	0.27
$\text{NO}_3^- \text{-TH}$	Spring	Spherical	0.0176	-0.07	24,000	0.25
	Summer	Spherical	0.05	-0.0214	22,000	-2.34
	Autumn	Spherical	0.15	-0.05	22,000	-3.0
	Winter	Spherical	0.047	-0.025	20,000	-1.88

presence of highly permeable subsurface materials with adequate dissolved oxygen (Canter 1997).

Pearson-correlation analysis was carried out between NO_3^- concentration and different effective parameters including water table depth (WD), total amount of precipitation (Rain), water temperature (T), and the soil chemical properties including electrical conductivity (EC), total dissolved solids (TDS), acidity (pH), CO_3 concentration, HCO_3^- concentration, Cl concentration, SO_4 concentration, Ca^{2+} concentration, Mg concentration, Na concentration, and K concentration. These parameters well represent the nutrient and water management and cultural practices in the study area which affect the NO_3^- concentration in groundwater. Figure 3 shows that, amongst all parameters, TH and Ca^{2+} had the highest correlation coefficient with NO_3^- concentration, and therefore, they were incorporated for NO_3^- estimation using cokriging method as secondary variables (covariates). Structural

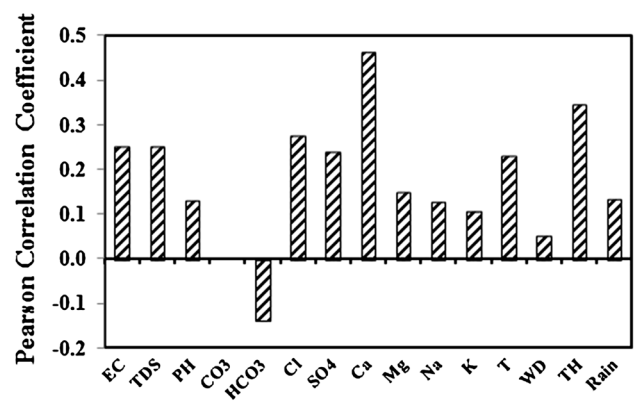


Fig. 3 Pearson correlation coefficients of the different groundwater parameters with NO_3^- concentration

analysis showed that Except for cross-variogram on NO_3^- and Ca^{2+} in summer for which an exponential model had the least RSS, a spherical model was the best fitted to the

dataset for the other periods (Table 2; Fig. 2). According to CI, for OCK1 and OCK2 semi-variograms, groundwater nitrate concentration had generally a weak spatial structure with some exceptions (Table 2). No significant improvement was observed in R with respect to Ca^{2+} or TH.

Evaluating the geostatistical models

Different geostatistical techniques were used to estimate NO_3^- concentration in groundwater for all sampling periods. Produced maps by OK were smoother and were substantially different with WMA since OK considers the pattern of spatial dependence of observed data, while the WMA method considers only the distance between estimated and observed locations (data not shown). Also, the locations of wells are more obviously shown in WMA maps since WMA generally produces spikes around the sample points (Lloyd 2005). Unlike WMA, produced maps by OCK represented more similarities with the OK map especially when incorporating Ca^{2+} as covariate.

Having the least RMSE, MAE, and MBE (Table 3), OK provided the best results for all datasets among different geostatistical methods, which is in consistent with other studies (Babiker et al. 2004; Assaf and Saadeh 2006). Less accurate results under OCK could be attributed to insignificant correlation between the primary and secondary variables and weak spatial correlation (Goovaerts 1997, 2000). Despite a higher correlation between NO_3^- and Ca^{2+} compared with that between NO_3^- and TH, Ca^{2+} may not still improve the prediction accuracy through OCK1 if the spatial continuity of Ca^{2+} is weaker than of NO_3^- , as was observed here (Tables 3). However, the better performance of OCK than WMA shows that OCK may increase the estimation accuracy by using the secondary data available at un-sampled locations even in the presence of a moderate cross-correlation between the primary and secondary variables (Table 2).

Positive MBE index for all interpolation techniques revealed an overall underestimation for all methods (Table 3). Higher bias was observed in summer and

autumn which could be attributed to higher variation of NO_3^- concentration due to both agricultural activities and intensive precipitation. Overall, OK provided more acceptable bias for all observed wells than the other methods. Also, created error maps for OK showed that the average values of standard error of the estimations were 1.11, 1.79, 1.63, and 1.34 mg/l for spring, summer, autumn, and winter, respectively, which were less than sample variance on the corresponded sampling dates (data not shown). This result indicates that OK interpolation is a reliable method for estimating NO_3^- concentration. However, the visual inspection of the observed NO_3^- and the NO_3^- values estimated by the OK model by means of scatter plots in Fig. 4 demonstrates that even applying Ok method did not led to a high agreement between observed and simulated NO_3^- in the study area.

Evaluating ANFIS and SVM models

For both ANFIS and SVM models, NO_3^- concentration was defined as a dependent output variable in ANFIS and SVM models while the Julian day number (DOY) of the sampling date, well locations as X- and Y-coordinates, Ca^{2+} and TH concentrations were defined as independent input variables in these models. DOY was defined based on the sampling dates. For example, if sampling is carried out on February 5, 15, and 25, the DOY will be set to 36, 46, and 56, respectively, and so on for the other months. Three input combinations evaluated in this research were as follows: (i) DOY, X-coordinate, and Y-coordinate (ANFIS1 and SVM1); (ii) DOY, X-coordinate, Y-coordinate, and Ca^{2+} (ANFIS2 and SVM2); and (iii) DOY, X-coordinate, Y-coordinate, and TH (ANFIS3 and SVM3). Out of the total 10-year data, 70 % of data which (data for 7 years) were incorporated in the train phase while the other 30 % of data related to the other 3 years were incorporated in the test phase.

Different ANFIS architectures were tried using a written code in MATLAB language including fuzzy logic. The efficient models were determined for each input

Table 3 Cross-validation between measured and estimated values of NO_3^- concentration

Period	Method	Evaluating index (mg/lit)					
		WMA-2	WMA-4	WMA-6	Kriging	CoKrig-1	CoKrig-2
Spring	MBE	6.55	6.71	6.76	5.22	5.25	5.02
	RMSE	20.35	20.51	20.52	20.01	20.55	20.49
Summer	MBE	31.52	31.52	31.52	9.69	7.70	7.75
	RMSE	44.353	44.353	44.353	29.706	33.188	33.49
Autumn	MBE	31.52	31.52	31.52	9.69	7.70	7.75
	RMSE	31.01	30.91	31.09	29.06	30.50	29.97
Winter	MBE	6.48	6.15	5.99	6.61	5.10	5.10
	RMSE	22.00	22.74	23.33	22.13	23.86	23.90

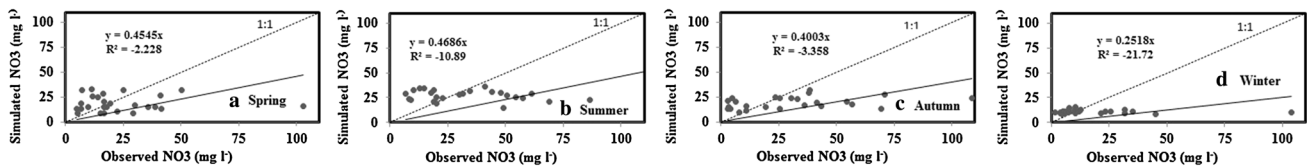


Fig. 4 Scatter plot of observed and simulated NO₃⁻ for OK models

combinations. Then, ANFIS models were tested and given results were compared using criteria indices. Table 4 represents the final architectures of ANFIS models for the test phase where the performance statistics are included. Two MFs were found to be sufficient for simulating NO₃⁻ with ANFIS models. The increase in the number MFs not only does not provide any significant improvement in the results, but also increases the number of parameters to be optimized. Because of that, the number of MFs should be kept in the minimum number (Ozger and Yildirim 2008). Gaussian combination (Gus2) membership functions were the best MFs for all ANFIS models. With respect to RMSE, ANFIS1 model with inputs of DOY, X- and Y-coordinates had the best performance (RMSE = 8.19–17.3 mg l⁻¹ and MBE = 5.2–9.7 mg l⁻¹) for all seasons followed by ANFIS2 with inputs of DOY, X- and Y-coordinates and Ca²⁺ concentration (RMSE = 9.73–25.1 mg l⁻¹ and MBE = 6–13.2 mg l⁻¹). Generally, better results were obtained for spring periods while the worst one was obtained for autumn period.

As shown in Table 4, all ANFIS models underestimated NO₃ concentrations. However, a visual inspection of the observed NO₃ and the NO₃ values estimated by the ANFIS models by means of scatter plots in Fig. 5 clearly indicate the high potential of ANFIS modeling. Despite high values of R² in Fig. 5, it should be noted that the R² term provides

information for the linear dependence between observed and simulated NO₃⁻; and R² values equal to 1 does not guarantee that a model captures the behavior of the investigated parameters (Tabari et al. 2012).

SVM models were implemented by a written program code in MATLAB language by which different SVM architectures were tried. Finally, the appropriate model structures were determined for each input combination. Then, SVM models were tested and the results were compared with performance statistics. The results of SVM models are presented in Table 5. When SVM models are applied, an appropriate choice of kernels allows the data became separable in feature space although being non-separable in the original space. This allows one to obtain nonlinear algorithms from algorithm previously restricted to handling linearity separable datasets (Bray and Han 2004). Here, radial basis function was the best kernel for all SVM models. SVM1 with RMSE = 4.75–8.19 mg l⁻¹ and MBE = 3.3–5.2 mg l⁻¹ had the best performance for simulating NO₃⁻ concentration. SVM2 and SVM3 models had nearly the same performance, and the selection of one of these models over the other should be dependent upon the available data. Scatterplots of observed and simulated NO₃ concentrations with SVM models demonstrate the high potential of SVM modeling with high R² index (Fig. 6).

Table 4 The final architecture and performance statistics of the ANFIS models for the test phases

Periods	Models	Best MF ^a	NMF	RMSE	MBE
Spring	ANFIS1	Gus2	2	8.19	5.2
	ANFIS2	Gus2	2	9.73	6
	ANFIS3	Gus2	2	8.96	5.6
Summer	ANFIS1	Gus2	2	17.3	9.7
	ANFIS2	Gus2	2	21.7	11.7
	ANFIS3	Gus2	2	22.4	12
Autumn	ANFIS1	Gus2	2	16.9	9.5
	ANFIS2	Gus2	2	25.1	13.2
	ANFIS3	Gus2	2	20.4	11.1
Winter	ANFIS1	Gus2	2	9.93	6.1
	ANFIS2	Gus2	2	10.5	6.4
	ANFIS3	Gus2	2	10.4	6.35

^a *Gb:Gus2* Gaussian combination MF, *MF* membership function

Comparison of the models

The global evaluation ranking of the approaches used was done based on RMSE and MBE indices and the results are summarized in Table 6. For all periods, SVM1 ranked first (RMSE = 4.75–8.19 mg l⁻¹) followed by SVM2 (RMSE = 6.34–12.1 mg l⁻¹) and SVM3 (RMSE = 5.1–11.5 mg l⁻¹). ANFIS models could be considered as 4th to 6th best models (RMSE = 8.19–25.1 mg l⁻¹) while geostatistical methods occupied that the last places for estimating NO₃⁻ concentration with RMSE = 22.1–44.6 mg l⁻¹.

ANFIS models combine the transparent, linguistic representation of fuzzy system with the learning ability of ANN. Therefore, they can be trained to perform an input/output mapping just as with an ANN, but with the additional benefit of being able to provide the set of rules on which the model is based. This given further insight into the process being modeled (Sayed et al. 2003). The main

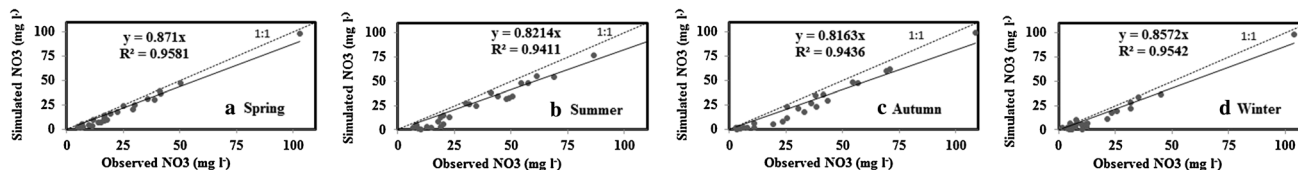


Fig. 5 Scatter plot of observed and simulated NO_3^- for ANFIS models

Table 5 The final architecture and performance statistics of the SVM models for the test phase

Period	Model	Best Kernel	RMSE	MBE
Spring	SVM1	Radial Basis Function	4.75	3.3
	SVM2	Radial Basis Function	6.34	4.2
	SVM3	Radial Basis Function	5.10	3.5
Summer	SVM1	Radial Basis Function	8.19	5.2
	SVM2	Radial Basis Function	10.7	6.5
	SVM3	Radial Basis Function	11.5	6.9
Autumn	SVM1	Radial Basis Function	7.44	4.8
	SVM2	Radial Basis Function	12.1	7.2
	SVM3	Radial Basis Function	10.1	6.2
Winter	SVM1	Radial Basis Function	5.98	4.0
	SVM2	Radial Basis Function	7.26	4.7
	SVM3	Radial Basis Function	8.89	4.5

advantage of using SVM is their flexibility and ability to model nonlinear relationship. Furthermore, the SVM training process always seeks a global optimized solution and avoid over-fitting that eventually leads to better generalization performance than ANN models. SVM is able to select the key vectors in the training process as it supports vectors and remove the nonsupport vectors automatically from the model. This makes the model to cope well with the noisy condition. The main disadvantage of the SVR technique is that it has no physical basis and belongs to a class of data-box approaches. In addition, the SVR can only be used when the training data are available (Bray and Han 2004; Zhou et al. 2009; Kisi and Cimen 2009).

NO_3^- spatial variability pattern and quality assessment

Spatial variability of nitrate concentration was prepared by SVM1 model in five thematic classes indicating “safe”

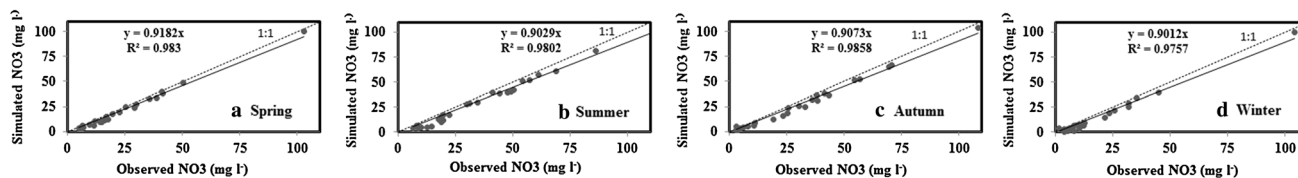


Fig. 6 Scatter plot of observed and simulated NO_3^- for SVM models

class ($\text{NO}_3 < 13 \text{ mg l}^{-1}$), “moderately human affected” class ($13 < \text{NO}_3 < 20 \text{ mg l}^{-1}$), “extremely human affected” class ($20 < \text{NO}_3 < 50 \text{ mg l}^{-1}$, i.e., 50 is the maximum acceptable level for drinking), “unacceptable for drinking” class ($50 < \text{NO}_3 < 75 \text{ mg l}^{-1}$) and “extremely unacceptable for drinking” class ($\text{NO}_3 > 75 \text{ mg l}^{-1}$) (Fig. 7). The pattern of spatial distribution was nearly erratic while it was more spatially structured in autumn when NO_3 concentration gradually decreased southward and westward. This pattern has been suggested to result from the disappearances of nitrate by natural de-nitrification in the direction of groundwater flow (Babiker et al. 2004).

Nearly, most parts of the study area displayed a relatively weak association with maximum acceptable concentration of nitrate for potable water (50 mg l^{-1}) according to the World Health Organization (WHO). Exception was for a few spots in spring and summer, especially centered around W01 observation well, indicating a possible pollution by point sources. Nevertheless, a large part of the study area was found to be suffering from nitrate concentration exceeding the human-affected threshold of 13 mg l^{-1} (Fig. 7). Of all the period tested, winter had the smallest area affected by excessive nitrate pollution (40.3 % of the study area) whereas just about 9.8–23 % of the area felt in the range of $0\text{--}13 \text{ mg l}^{-1}$ for the other periods. Nitrate concentration commonly ranged from 20 to 50 mg l^{-1} during spring, summer, and autumn period, indicating an ongoing potential risk from groundwater pollution. Thus, recognizing the most vulnerable area could help with a better understanding of the pollution source in the study area.

The probability maps of $\text{NO}_3 > 13 \text{ mg l}^{-1}$ ($P_{\text{N}13}$) were generated for all periods (Fig. 8) which were classified with 5 equal classes of $P_{\text{N}13}$ including $0\text{--}0.2$, $0.2\text{--}0.4$, $0.4\text{--}0.6$, $0.6\text{--}0.8$, and $0.8\text{--}1$. Despite a decreasing trend of NO_3 from summer to winter, about 56–97 % of the area

Table 6 Summary of the ranking of the selected models

Period	Method	Rank	Period	Method	Rank	Period	Method	Rank	Period	Method	Rank
Spring	SVM1	1	Summer	SVM1	1	Autumn	SVM1	1	Winter	SVM1	1
	SVM3	2		SVM2	2		SVM3	2		SVM2	2
	SVM2	3		SVM3	3		SVM2	3		SVM3	3
	ANFIS1	4		ANFIS1	4		ANFIS1	4		ANFIS1	4
	ANFIS3	5		ANFIS2	5		ANFIS3	5		ANFIS3	5
	ANFIS2	6		ANFIS3	6		ANFIS2	6		ANFIS2	6
	OK	7		OK	7		OK	7		WMA-2	7
	WMA-2	8		OCK1	8		OCK2	8		OK	8
	OCK2	9		OCK2	9		OCK1	9		WMA-4	9
	WMA-4	10		WMA-2	10		WMA-4	10		WMA-6	10
	WMA-6	11		WMA-4	11		WMA-2	11		OCK1	11
	OCK1	12		WMA-6	12		WMA-6	12		OCK2	12

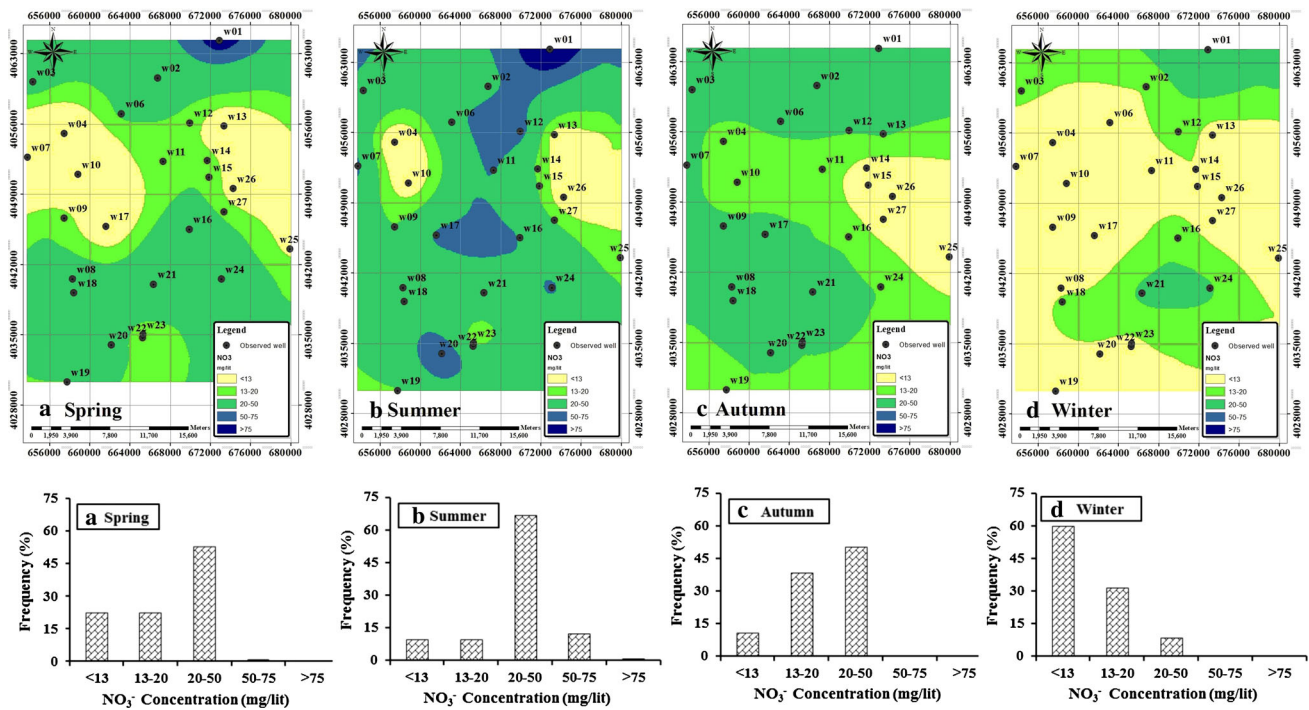


Fig. 7 Classified spatial variability maps of groundwater nitrate level combined with related histograms for spring (a), summer (b), autumn (c), and winter (d)

showed the highest probability ($0.8 < P_{N13} < 1$) of $NO_3 > 13 \text{ mg l}^{-1}$ with the minimum and maximum P_{N13} for spring and autumn, respectively. Exception was for winter when less than 10 % of the study area was at high risk of groundwater contamination ($0.8 < P_{N13} < 1$). Also, a considerable parts of the study area (0.6–20.1 %) showed a strong probability ($0.6 < P_{N13} < 0.8$) of groundwater pollution induced by human activities. From nitrate contamination viewpoint, water of acceptable quality ($P_{N13} < 10 \%$) may be found in a few part of the study

area. Overall, having $P_{N13} > 50 \%$, a large part of the study area is at high risk of nitrate contamination, quite likely due to human-related pollution, accounted for 78.5, 97.7, 97.8 and 42.7 % in spring, summer, autumn, and winter, respectively. Thus, groundwater is more vulnerable to nitrate contamination in summer and autumn than the other periods.

Results revealed a considerable difference in groundwater NO_3 concentration attributed to land use (Fig. 1b) which is well documented in previous researches (Burow

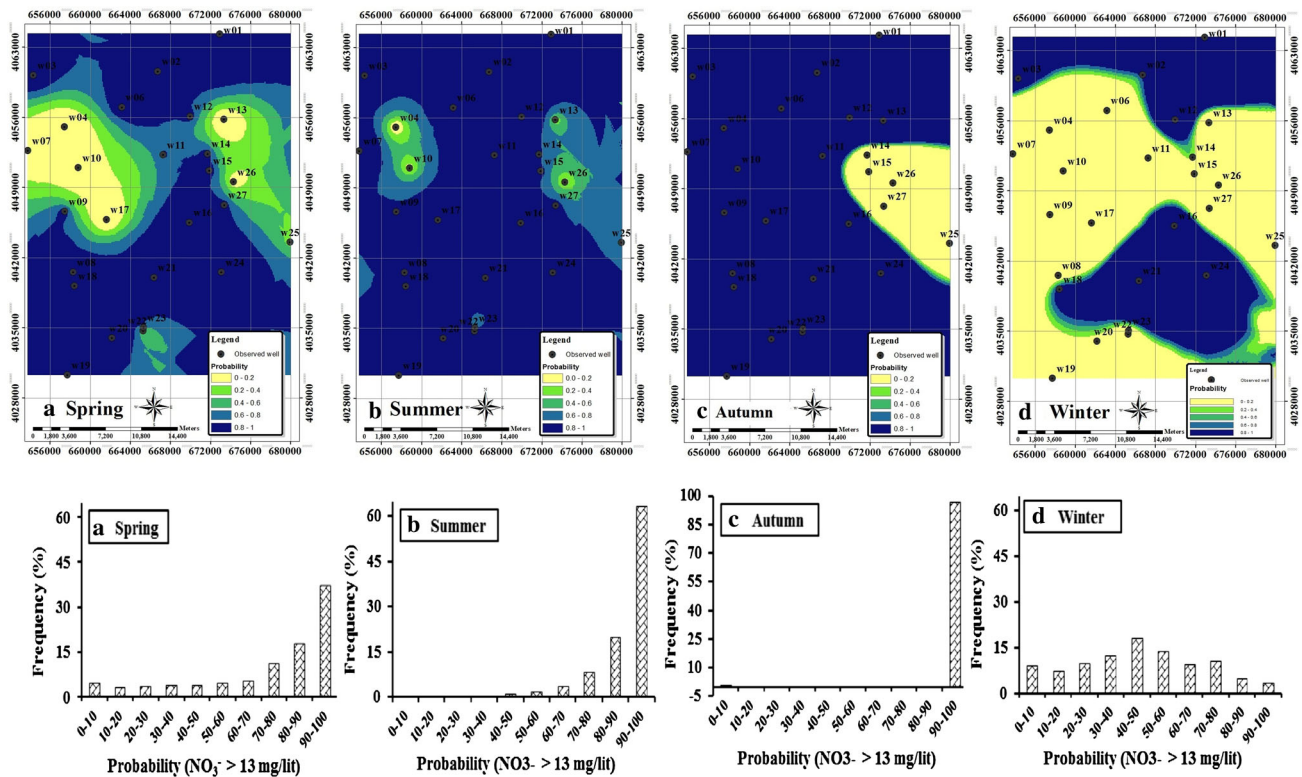


Fig. 8 Spatial distribution of excessive groundwater nitrate probabilities in the study area combined with related histograms for spring (a), summer (b), autumn (c), and winter (d)

et al. 2010; Melo et al. 2012; Dahan et al. 2014). Commonly, monitoring wells located in “paddy fields” class had highest average of nitrate concentration followed by those located in “urban” class. The least NO_3^- concentration was attributed to “dense forest” class with an average of 7.97–19.49 mg l^{-1} during the study period. Non-point-source pollution is evidence in paddy fields, where more than 90 % of it was associated with $\text{NO}_3^- > 13 \text{ mg/lit}$. Being an essential for productive agriculture, fertilizer application is one of the most important factors affecting groundwater contamination under paddy fields as is documented in the previous researches (Spalding and Exner 1993; Dudley et al. 2008; Burow et al. 2010; Melo et al. 2012; Dahan et al. 2014). Since 80 % of the study area devoted to rice cultivation, temporal variations in NO_3^- concentration could be explained based on three periods of preplanting (winter), transplanting (spring and summer), and postharvest (autumn). In paddy fields, fertilizers are commonly applied in spring. Thus, higher nitrate concentration in summer well reflects the lag time between the applications of fertilizers during the rainless season to the time required to nitrates leach into the groundwater. During postharvest period (autumn), nitrate concentration in groundwater was higher than that of preplanting (winter) period mainly related to solute transport to the groundwater

due to heavy rainfall in autumn as what does heavy irrigation followed by fertilization for the summer crops sown in the spring. However, higher nitrate concentration in planting and postharvest periods indicates the considerably negative effect of agricultural activities in the study area. Irrigation practice and management may also influence nutrient transport to groundwater in the paddy fields (Gao et al. 2012). Flooding irrigation, especially after fertilizer application, increases nitrate losses into groundwater compared with dry farming. Also, irrigation water quality can have a profound impact on groundwater quality especially under flooding irrigation due to having dissolved mineral salts (Grattan 2002). A considerable depletion in groundwater level in summer (data not shown) due to irrigation may deteriorate groundwater quality during rice growing season due to a strong relationship between the increase of salt contamination and the lowering of piezometric levels (Lashkaripour and Ghafouri 2011; Chaudhuri and Ale 2014). In addition, the shallow aquifer in paddy fields which is relatively close to the surface may receive direct input of NO_3^- -rich leachate from the agricultural soils (Gao et al. 2012).

Figures 1b, 7, and 8 show that regions surrounded by both paddy fields and urban classes had the highest NO_3^- concentration which reflects the additive negative effect of

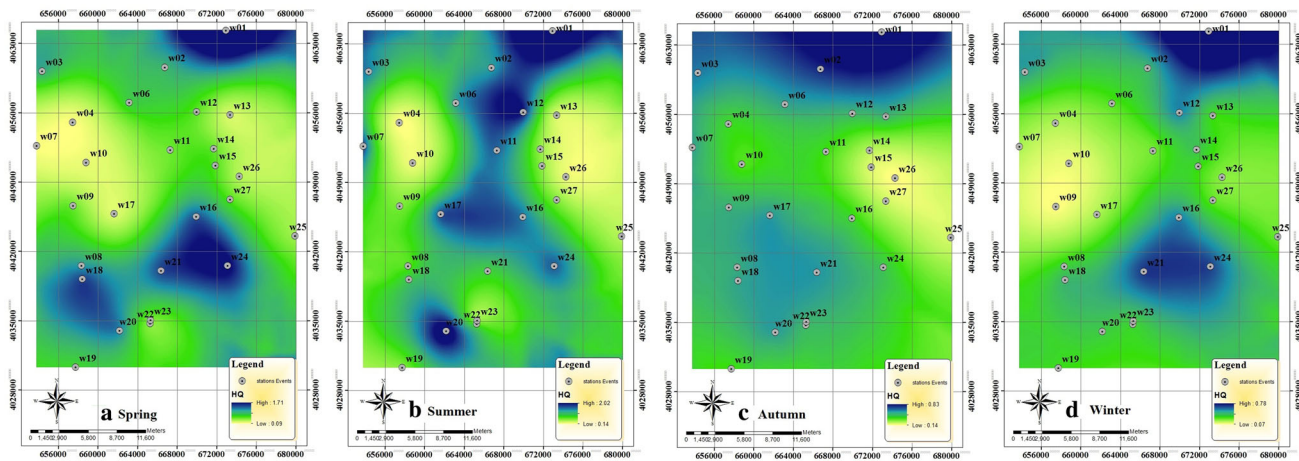


Fig. 9 Spatial distribution of HQ due to nitrate in spring (a), summer (b), autumn (c), and winter (d)

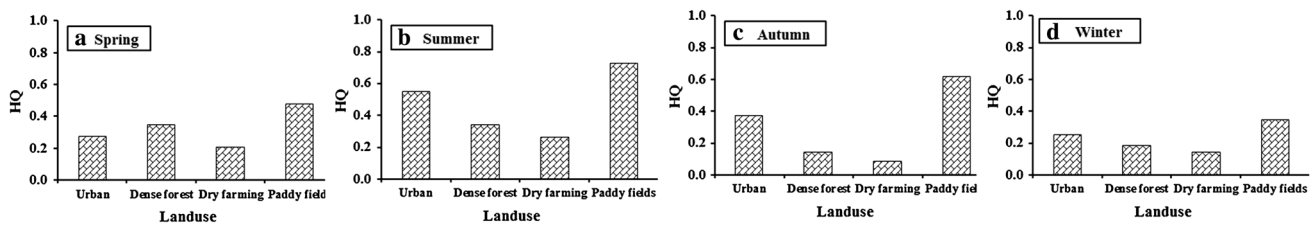


Fig. 10 HQ due to nitrate for different land uses in spring (a), summer (b), autumn (c), and winter (d)

both classes for contaminating groundwater. Point source pollution could be found around W01 borehole with consistent $\text{NO}_3 > 50 \text{ mg/lit}$, presumably due to releases of untreated wastewater to open areas, ditches, and septic tanks and the common practice of dumping solid waste.

Health risk assessment

The spatial distribution of HQ of nitrate is illustrated in Fig. 9. The HQ assumes that there is a level of exposure (i.e., Rdf) below which it is unlikely for even sensitive pollutions to experience adverse health effects; there may be a concern arising for the potential noncarcinogenic effect of the HQ exceeds the unity. HQ due to nitrate showed seasonal changes of 0.09–1.71 in spring, 0.14–2.02 in summer, 0.14–0.83 in autumn, and 0.07–0.78 in winter. HQ exceeded the unity in about 1.13 and 6.82 % of the study area in spring and summer, respectively, which indicates that long-term drinking of groundwater or irrigation using groundwater poses a significant health risk to local resident (Gao et al. 2010).

Figure 10 shows that paddy fields were commonly associated with the highest risk value. Totally, noncarcinogenic risk values of nitrate decreased in the following order: paddy fields > urban > dense forest > dry farming. Exception was for spring when it had higher HQ compared

with urban land use. Relatively higher risk value under agricultural lands has been reported by other researchers (Gao et al. 2010). Higher values of HQ under paddy fields could be attributed to unmanaged use of fertilizers during the growing season.

Conclusion

In this research, the spatial and temporal variation of nitrate concentration in groundwater of Northern Iran’s coastal areas as well as its noncarcinogenic health risk was assessed using different machine-learning and geostatistical models. The comparative analysis demonstrated the high capability of simple SVM models for predicting nitrate concentration in groundwater. Created raster maps with SVM models revealed a high risk of nitrate pollution in a large part of the study area which is mainly under rice production. However, nitrate contamination from urban sources is also possible. Higher risk of groundwater contamination during the rice cultivation as well as postharvest periods compared with preplanting period well reflects the negative effect of unmanaged use of chemical fertilizers for contaminating the groundwater. In addition, over-abstraction of groundwater resources for irrigation in rice growing season and leaching of nitrate through heavy precipitation

during postharvest periods causes further contamination of shallow groundwater aquifers. Health risk analysis revealed a necessity of devising policy guidelines for using groundwater due to high levels of HQ which poses a significant health risk for Mazandaran's residents. Overall, sustaining the environment and protecting the human health requires a serious attention to manage the agricultural activities in paddy fields since a considerable part of the agricultural land in Mazandaran province is allotted to rice cultivation in which more than 40 % of the total Iran's rice production is produced. Controlled use of fertilizers, optimal irrigation based on the plant water demand, preventing overexploitation of the shallow groundwater, and removing point source pollution would help with providing sustainable agriculture in the study area.

References

- Ahmad S, Simonovic SP (2005) An artificial neural network model for generating hydrograph from hydro-meteorological parameters. *J Hydrol* 315:236–251
- Almasri MN, Kaluarachchi JJ (2005) Modular neural networks to predict the nitrate distribution in ground water using the on ground nitrogen loading and recharge data. *Environ Model Softw* 20:851–871
- Assaf H, Saadeh M (2006) Development of an integrated decision support system for water quality control in the Upper Litani Basin, Lebanon. In: Voinov A, Jakeman AJ, Rizzoli AE (eds) Proceedings of the iEMSs third Biennial meeting: "summit on environmental modelling and software", international environmental modelling and software society. Burlington, USA
- Asefa T, Kembrowski M, McKee M, Khalil AF (2006) Multi-time scale stream flow predictions: the support vector machines approach. *J Hydrol* 318:7–16
- Babiker IS, Mohamed AAM, Terao H, Kato K, Ohta K (2004) Assessment of groundwater contamination by nitrate leaching from intensive vegetable cultivation using geographical information system. *Environ Int* 29(8):1009–1017
- Bray M, Han D (2004) Identification of support vector machines for runoff modeling. *J Hydroinf* 6(4):265–280
- Burkart MR, Kolpin DW (1993) Hydrologic and land use factors associated with herbicides and nitrates in near-surface aquifers. *J Environ Qual* 22:646–656
- Burow KR, Nolan BT, Rupert MG, Dubrovsky NM (2010) Nitrate in groundwater of the United States, 1991–2003. *Environ Sci Technol* 44:4988–4997
- Cambardella CA, Moorman TB, Novak JM, Parkin TB, Karlen DL, Turco RF, Konopka AE (1994) Field-scale variability of soil properties in central Iowa soils. *Soil Sci Soc Am J* 58:1501–1511
- Canter LW (1997) Nitrates in groundwater. Lewis, Boca Raton
- Castrignano A, Buttafuoco G, Troccoli A, Colecchia SA, Di Bitetto V, Pisante M, Basso F, Cafiero G, Cammarano D, Basso B (2008) Multivariate geostatistical analysis for delineation of management zones using crop index. In: Proceedings of the international conference on agricultural engineering, Hersonissos (Ageng2008), Crete Isle, Greece, 23–25 June 2008
- Chaudhuri S, Ale S (2014) Long term (1960–2010) trends in groundwater contamination and salinization in the Ogallala aquifer in Texas. *J Hydrol* 513:376–390
- Chiles JP, Delfiner P (1999) Geostatistics: modelling spatial uncertainty. Wiley, New York 695 p
- Dahan O, Babad A, Lazarovitch N, Russak EE, Kurtzman D (2014) Nitrate leaching from intensive organic farms to groundwater. *Hydrol Earth Syst Sci* 18:333–341
- Dai X, Huo Z, Wang H (2011) Simulation for response of crop yield to soil moisture and salinity with artificial neural network. *Field Crops Res* 121:441–449
- Dastirani MT, Moghadamnia A, Piri J, Rico-Ramirez M (2010) Application of ANN and ANFIS models for reconstructing missing flow data. *Environ Monit Assess* 166:421–434
- Delhomme JP (1976) Kriging in the hydrosociences. *Adv Water Resour* 1:251–266
- Deng J, Chen X, Du Z, Zhang Y (2011) Soil water simulation and predication using stochastic models based on LS-SVM for red soil region of China. *Water Resour Manage* 25:2823–2836
- Dudley LM, Ben-Gal A, Lazarovitch N (2008) Drainage water reuse: biological, physical, and technological considerations for system management. *J Environ Qual* 37:25–35
- Eckhardt DAV, Stackelberg PE (1995) Relation of ground-water quality to land use on Long Island, New York. *Groundwater* 33:1019–1033
- Elshorbagy A, Parasuraman K (2008) On the relevance of using artificial neural networks for estimating soil moisture content. *J Hydrol* 362:1–18
- Gao Y, Mao L, Miao C, Zhou P, Gao J, Zhi Y, Shi W (2010) Spatial characteristics of soil enzyme and microbial community structure under different land use in Chongming Island, China: Geostatistical modeling and PCR-RAPD method. *Sci Total Environ* 408:3251–3260
- Gao Y, Yu G, Luo C, Zhou P (2012) Groundwater nitrogen pollution and assessment of its health risk: a case study of a typical village in rural-urban continuum, china. *PLoS One* 7(4):1–8
- Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York, p 483
- Goovaerts P (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J Hydrol* 228:113–129
- Grattan SR (2002) Irrigation water salinity and crop production. University of California, Division of Agriculture and National Resources. Publication 8066
- Issaks E, Srivastava RM (1989) Applied geostatistics. Oxford University Press, New York
- Jiang H, Cotton WR (2004) Soil moisture estimation using an artificial neural network: a feasibility study. *Can J Rem Sens* 30:827–839
- Karandish F (2016) Improved soil-plant water dynamics and economic water use efficiency in a maize field under locally water stress. *Arch Agron Soil Sci*. doi:10.1080/03650340.2015.1135326
- Khalil AF, McKee M, Kembrowski M, Asefa T (2005) Basin scale water management and forecasting using artificial neural network. *J Am Water Resour Assoc* 41:195–208
- Kisi O, Cimen M (2009) Evapotranspiration modelling using support vector machines. *Hydrol Sci J* 54(5):918–928
- Lashkaripour GR, Ghafoori M (2011) The effects of water table decline on the groundwater quality in aquifer of torbat jam plain, Northeast Iran. *Int J Emerg Sci* 1(2):153–163
- Lin GF, Chen GR, Wu MC, Chou YC (2009) Effective forecasting of hourly typhoon rainfall using support vector machines. *Water Resour Res* 45:W08440
- Liu D, Yu Z, Lü H (2010) Data assimilation using support vector machines and ensemble Kalman filter for multi-layer soil moisture prediction. *Water Sci Eng* 3:361–377
- Lloyd CD (2005) Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. *J Hydrol* 308:128–150

- Melo A, Pinto E, Aguiar A, Mansilha C, Pinho O, Ferreira I (2012) Impact of intensive horticulture practices on groundwater content of nitrates, sodium, potassium, and pesticides. *Environ Monit Assess* 184:4539–4551
- Mohandes MA, Halawani TO, Rehman S, Hussain AA (2004) Support vector machines for wind speed prediction. *Renew Energy* 29:939–947
- Mousavifazl H, Alizadeh A, Ghahraman B (2013) Application of geostatistical methods for determining nitrate concentration in groundwater (Case study of Mashhad plain, Iran). *IJACS* 5(4):318–324
- Ni F, Liu G, Tan Y, Deng Y (2010) Spatial variation of health risk of groundwater for drinking water supply in Mingshan County, Ya'an City, China. *Water Sci Engin* 3(4):454–466
- Ozger M, Yildirim G (2008) Determining turbulent flow friction coefficient using adaptive neuro-fuzzy computing technique. *Adv Eng Soft.* doi:10.1016/j.advengsoft.2008.04.006
- Rehman S, Mohandes M (2008) Artificial neural network estimation of global solar radiation using air temperature and relative humidity. *J Energy Policy* 63(2):571–576
- Sayed T, Tavakolie A, Razavi A (2003) Comparison of adaptive network based fuzzy inference systems and B-spline neuro-fuzzy mode choice models. *Water Resour Res* 17(2):123–130
- Spalding RF, Exner ME (1993) Occurrence of nitrate in groundwater—a review. *J Environ Qual* 22:392–402
- Sundaraiah R, Sudarshan V (2014) Nitrate contamination in groundwater of Kalwakurthy area, Mahabubnagar District, Andhra Pradesh, India. *Indian J Appl Res* 4(9):240–243
- Tabari H, Kisi O, Ezani A, Talaei PH (2012) SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climate data in a semi-arid highland environment. *J Hydrol* 444–445:78–89
- Tang LS, Li Y, Zhang J (2005) Physiological and yield responses of cotton under partial root-zone irrigation. *Field Crop Res* 94:214–223
- U.S. Epa (United States Environmental Protection Agency) (1989) Supplement risk assessment (Part 1). U.S. EPA, Washington, D.C.
- Vapnik V (1995) *The nature of statistical learning theory*. Springer, New York
- Webster R, Oliver MA (2006) Optimal interpolation and isarithmic mapping of soil properties. VI. Disjunctive kriging and mapping the conditional probability. *Eur J Soil Sci* 40:497–512
- Yu X, Liang SY (2007) Forecasting of hydrologic time series with ridge regression in feature space. *J Hydrol* 332:290–302
- Zhou H, Li W, Zhang C, Liu J (2009) Ice breakup forecast in the reach of the Yellow River: the support vector machines approach. *Hydrol Earth Syst Sci Discuss* 6:3175–3198
- Zou P, Yang J, Fu J, Liu G, Li D (2010) Artificial neural network and time series models for predicting soil salt and water content. *Agric Water Manag* 97:2009–2019