IRRIGATION AND DRAINAGE
Irrig. and Drain. **66**: 428–438 (2017)
Published online 23 January 2017 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/ird.2107

DIAGNOSING DRAINAGE PROBLEMS IN COASTAL AREAS USING MACHINE-LEARNING AND GEOSTATISTICAL MODELS†

ABDULLAH DARZI-NAFTCHALI[1]*, FATEMEH KARANDISH[2] AND AHMAD ASGARI[1]

[1]*Water Engineering Department, Sari Agricultural Sciences and Natural Resources University, Sari, Iran*
[2]*Water Engineering Department, University of Zabol, Zabol, Iran*

ABSTRACT

This study focuses on diagnosing drainage problems in the coastal areas of Iran by using geostatistical methods, support vector machines (SVMs) and the adaptive neuro-fuzzy inference system (ANFIS). Groundwater level (WD) and quality were monitored at 37 shallow wells scattered over a 25 000 ha area at different times. Using prepared raster maps of pH, ESP, EC and WD by the best method, drainage problems were categorized into eight classes. Both SVM and ANFIS models significantly improved predicted data for pH, ESP, EC and WD compared with geostatistical models, while SVMs provided slightly better results which were used for further analysis. More than 60% of the area needs drainage to lower the groundwater table in preplanting and post-harvest periods, while during the growing seasons, more than about 72% of the area requires drainage for salinity control. Based on the results, identifying drainage problems at basin scale is possible with cost-efficient machine learning models with minimum time and data requirements and investment for detailed field surveys. Copyright © 2017 John Wiley & Sons, Ltd.

KEY WORDS: ANFIS; groundwater level; salinity; SVM; waterlogging

Received 1 June 2016; Revised 5 December 2016; Accepted 5 December 2016

RÉSUMÉ

Cette étude porte sur le diagnostic des problèmes de drainage dans les zones côtières de l'Iran en utilisant différents outils, à savoir des méthodes géostatistiques, des machines supports de vecteurs (SVM) et le système adaptatif d'inférence neuro-floue (ANFIS). Le niveau et la qualité des eaux souterraines (DEO) ont été suivis dans le temps sur 37 puits peu profonds disséminés sur une zone de 25 000 ha. Les problèmes de drainage ont été répartis en huit classes en utilisant des cartes raster préparées avec des paramètres sélectionnés avec la meilleure méthode. Les deux modèles SVM et ANFIS ont considérablement amélioré les prédictions pour le pH, ESP, CE et WD par rapport aux modèles géostatistiques, alors SVMs fournissent des résultats légèrement meilleurs, qui ont été utilisés pour une analyse ultérieure. Sauf pour moins de 3% de la zone d'étude, le drainage doit être amélioré pour toutes les périodes. Plus de 60% de la surface a besoin de drainage pour abaisser la nappe phréatique dans les périodes pré-plantation et post-récolte, alors que pendant les saisons de croissance, plus d'environ 72% de la surface nécessite un drainage pour le contrôle de la salinité. D'après les résultats, l'identification des problèmes de drainage à l'échelle du bassin est possible grâce à des modèles d'apprentissage peu couteux, tout en nécessitant un minimum d'investissement en données de terrain et en temps pour les enquêtes détaillées sur le terrain. Copyright © 2017 John Wiley & Sons, Ltd.

MOTS CLÉS: ANFIS; niveau des eaux souterraines; salinité; SVM; engorgement

INTRODUCTION

In recent years, much of the water requirement in the coastal areas of Mazandaran Province, northern Iran, has been supplied through surface water resources such as traditional *abbandans* (ponds) and rivers. Rapid population growth and increased water demand for drinking, domestic,

*Correspondence to: Abdullah Darzi-Naftchali, Water Engineering Department, Sari Agricultural Sciences and Natural Resources University, Sari, Iran. E-mail: abdullahdarzi@yahoo.com

†Diagnostiquer les problèmes de drainage dans les zones côtières en utilisant le machine-learning et des modèles géostatistiques.

Copyright © 2017 John Wiley & Sons, Ltd.

industrial and, especially, agricultural uses have put more pressure on the natural water balance and consequently on the total hydrology of the region, resulting in greater emphasis on the exploitation of groundwater resources. In addition, the area is beset with the problems of waterlogging during rainy seasons and salinity during summer months, mainly due to imbalanced groundwater recharge and extraction. Waterlogging, which occurs due to the relatively flat topography, inadequate natural drainage facilities and excess rainfall, prevents year-round cropping, resulting in considerable areas either going out of production or experiencing reduced yield (Darzi-Naftchali *et al.*, 2013). However, there are increasing concerns about overexploitation from shallow aquifers during spring and summer months. The quality of groundwater determines its usefulness for different applications which, in turn, depends on geology and weathering of parent materials, recharge water quality and inputs from industry, agriculture and other land use practices (Aghazadeh and Asghari-Mogaddam, 2010).

Providing sustainable agricultural production in the areas with the twin problems of waterlogging and salinity is dependent on the provision of some form of drainage (El-Nashar, 2013). The assessment of such improving management strategies requires extensive long-term field surveys related to existing groundwater level and quality. Detailed field surveys are very time-consuming and involve considerable cost and investment. Developing methods with low input data requirements while providing reliable results are of great importance in drainage surveys. Geostatistical and machine-learning models such as adaptive neuro-fuzzy inference (ANFIS) and support vector machines (SVMs) might be good tools for such investigation. There is no need for knowledge of the mathematical relationship between the inputs and the corresponding outputs in these models (Almasri and Kaluarachchi, 2005). Even with limited data, machine-learning models provide quick and flexible estimating methods aimed at achieving a high level of generalization and prediction accuracy (Khalil *et al.*, 2005). Among machine-learning models, SVMs are a new structure which were introduced for classification and regression problems (Liu *et al.*, 2010). Machine-learning models have been applied in hydrologic studies such as predicting monthly stream flows (Wang *et al.*, 2009; Sudheer *et al.*, 2014), groundwater modelling (Yoon *et al.*, 2011), water quality modelling (Singh *et al.*, 2011) and simulating soil water content, temperature and precipitation (Deng *et al.*, 2011). In addition, in previous studies, fluctuation of the groundwater level and its quality in coastal areas were generally investigated to assess the suitability of groundwater quality for drinking, agricultural and industrial purposes. However, the present study focuses on diagnosing drainage problems in the coastal areas of Mazandaran Province by using geostatistical and machine-learning models based on groundwater level and quality. Included in this analysis is identification of the relationship between groundwater level and its quality by using geostatistical methods and ANFIS and SVM models.

MATERIALS AND METHODS

Study area and data collection

The experimental site with 25000 ha area is located within longitudes of 654651 m to 675625 m and latitudes of 4038241 m to 4070903 m in Mazandaran Province, Iran. 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. Long-term annual minimum, maximum and average temperatures are about 12.5, 21.5 and 17°C, respectively.

There are 37 operating wells across the study area which are used for irrigation and drinking. Of the total area of the study region, 43.3, 38.3, 13.2 and 5.2% are attributed to paddy fields, dry lands, orchards and *abbandans*, respectively. Figure 1 shows the location of shallow wells in the study area and the land use pattern. Water samples of the wells were prepared on three occasions, including preplanting (winter, March 2009), cultivation period (spring, May 2010) and post-harvest period (summer, August 2010). The samples were analysed at the Water Quality Laboratory of the Mazandaran Regional Water Company to determine acidity (pH), electrical conductivity (EC) and exchangeable sodium percentage (ESP). Also, groundwater depth (WD) was measured at the time of sampling.

Geostatistical approach

All data with high skewness and kurtosis were normalized with logarithmic transformation since normality is critical for statistical analysis methods. To extract the spatial distribution maps, normalized data of pH, EC, ESP and WD were subjected to four geostatistical methods, including weighting moving average (WMA) with powers of 2 (WMA-2), 3 (WMA-3) and 4 (WMA-4) and ordinary kriging (OK). Geostatistics considers the location of a data point along with its value. The spatial correlation of a given variable is represented by semivariogram which is calculated as follows (Borga and Vizzaccaro, 1997; Issaks and Srivastava, 1989):

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

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 γ

Copyright © 2017 John Wiley & Sons, Ltd.

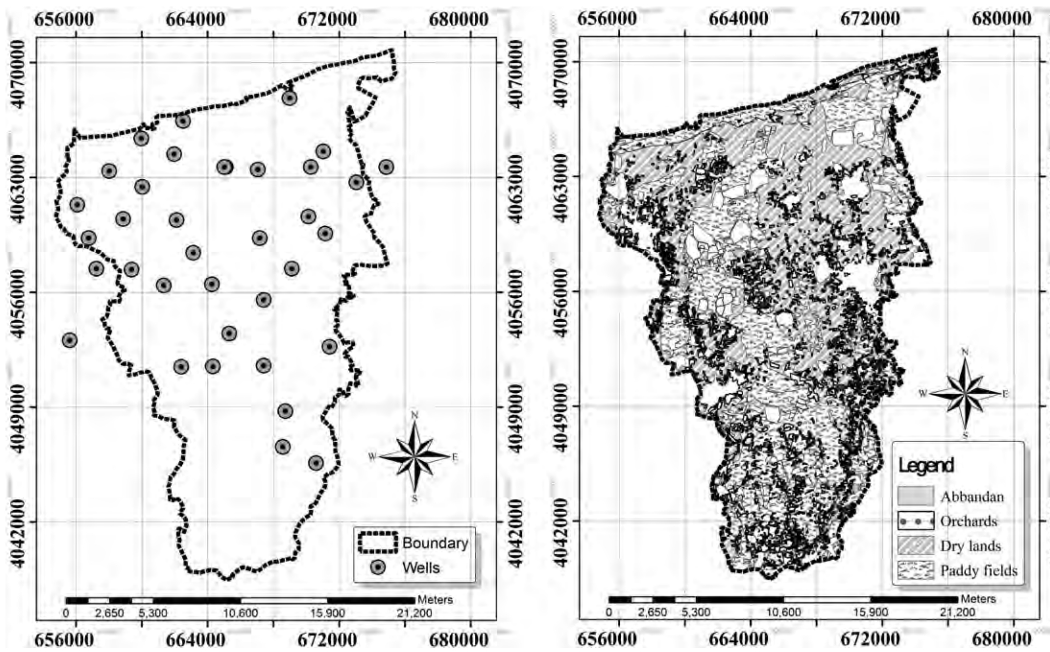

Figure 1. Location of shallow wells in the study area (a) and the land use pattern (b).

value at R is termed as a sill which is theoretically equal to the sample variance under the assumption of second-order stationarity. The value at $h = 0$ is defined as nugget effect C_0. A detailed description of geostatistical theories can be found in Goovaerts (1997), Chiles and Delfiner (1999) and Webster and Oliver (2006). The GS+ software was applied for variogram analysis and cross-validation analysis for all geostatistical methods.

ANFIS

The theory of fuzzy sets has been widely used in different fields of study since its introduction. ANFIS was introduced in 1993 through integrating the capabilities of this theory and neural nets (Tang *et al.*, 2005). Fuzzy system parameters are identified by the hybrid learning algorithms that are also applicable for teaching the model as well in the fuzzy inference system which serves as a teaching method in neuro-fuzzy systems (Rehman and Mohandes, 2008). Detailed description about the structure of ANFIS model characteristics can be found in Tabari *et al.* (2012).

Two sets of parameters including comparative parameters (S1) and consequent parameters (S2) are embedded in ANFIS which are calculated using a least square error algorithm (LSE) and a gradient descent algorithm under forward pass and backward pass steps. Simulation is assumed to be correct when the applied method for estimating both of the parameters in training, examination and validation procedures results in the least amount of error function of the model. The function type used for the model inputs is a major feature of every ANFIS model. In this research, different membership functions (MFs) were employed. The best MF giving the minimum of errors was determined through examining various combinations of MFs. Different ANFIS architectures were evaluated using a MATLAB code that included fuzzy logic. Efficient models were first determined for each combination of input variables. Then, the various ANFIS models were tested and results obtained were evaluated using different criteria indices (Karandish *et al.*, 2016).

SVM

Support vector machines (SVMs) are used for both classification and regression problems. SVMs may be represented as two-layer networks with nonlinear and linear weights in the first and the second layers, respectively (Bray and Han, 2004). Support vector regression (SVR) is used to estimate a function according to a given data set $\{(x_i,y_i)\}_n$, (x_i: input vector; y_i: output value and n: total number of data sets) (Tabari *et al.*, 2012). The following function is used as the linear regression function:

$$f(x) = \omega.\varnothing(x) + b \tag{2}$$

where $\varnothing(x)$ is a nonlinear function by which x is mapped into a feature space, b and ω are, respectively, a weight vector and a coefficient. Linear regression is performed in a high-dimensional feature space via a nonlinear mapping. More details on SVMs can be found in Tabari *et al.* (2012). Different SVM architectures were tried by

Copyright © 2017 John Wiley & Sons, Ltd.

a written program code in MATLAB and the suitable models were found for each input combination. The SVM models were implemented using a MATLAB code. First, different SVM architectures were tried. Once the appropriate model structures were determined for each combination of input variables, the resulting SVM models were tested against the experimental data set, and the results were compared using the performance statistics. In the SVM modelling, an appropriate choice of kernels allows the data to become separable in the feature space despite being non-separable in the original space. This allows one to obtain nonlinear algorithms from algorithms previously restricted to handling linearly separable data sets (Bray and Han, 2004).

Evaluation criteria

Prediction performance was assessed by cross-validation. In addition, three indices including mean bias error (MBE), mean absolute error (MAE) and root mean square error (RMSE) were calculated for comparing different interpolation models as follows:

$$MAE = \frac{\left|\sum_{i=1}^{n} Z(x_i) - Z^*(x_i)\right|}{n} \tag{3}$$

$$MBE = \frac{\sum_{i=1}^{n} \left(Z(x_i) - Z^*(x_i)\right)}{n} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(Z^*(x_i) - Z(x_i)\right)^2}{n}} \tag{5}$$

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

RESULTS AND DISCUSSION

Data quality check

Statistical parameters of measured pH, EC, ESP and WD for all sampling periods are summarized in Table I. For all parameters, the maximum sample variance was in the winter season. The exception was for pH, for which the highest sample variance corresponded to thesummer season. Based on Wilding and Dress (1983), a parameter has low variability if $CV < 0.15$, has medium variability if $0.15 < CV < 0.35$ and has high variability if $CV > 0.35$. High variability was observed for all parameters apart from pH for which a low variability was observed. This result is in line with some other researches (Yost *et al.*, 1982; Zhou *et al.*, 1996). In a wide region, variability could be affected by land use, erosion pattern and fertilization. Non-uniform water and nutrient distribution under surface irrigation as well as topography would increase the variability of a parameter in a small region (Sun *et al.*, 2003).

Being a characteristic of environmental data, non-normal distribution was observed for most data sets in Table I. Non-normality is particularly evident in the winter data set with a

Table I. Statistical parameters of groundwater properties and WD

		Mean	SD	Variance	Min.	Max.	CV (%)	Skewness	Kurtosis
Winter	pH	7.51	0.237	0.06	7.1	8.2	3.16	0.64	0.36
	EC ($\mu S\ cm^{-1}$)	5810	9530	9.07×10^6	802	56 100	164	4.31	19.9
	EC$'$ ($\mu S\ cm^{-1}$)	8.11	0.96	0.92	6.7	10.9	11.8	0.79	0.39
	ESP	12.0	12.4	153	0.30	57.0	103	1.75	3.24
	ESP$'$	2.16	0.94	0.88	0.26	4.07	43.5	−0.03	−0.78
	WD (m)	0.59	0.458	0.21	0.00	2.35	78.0	1.73	4.29
	WD$'$ (m)	0.43	0.26	0.07	0.00	1.21	59.9	0.74	0.82
Spring	pH	7.13	0.23	0.06	6.67	7.55	3.2	−0.13	−0.59
	EC ($\mu S\ cm^{-1}$)	3520	3434	11.8×10^6	635	13 200	97.5	1.93	2.6
	EC$'$ ($\mu S\ cm^{-1}$)	7.84	0.78	0.61	6.45	9.49	9.5	0.56	−0.15
	ESP	11.71	9.71	94.2	0.40	34.9	82.9	0.01	0.28
	ESP$'$	2.22	0.88	0.77	0.34	3.58	39.6	−0.38	−0.73
	WD (m)	1.01	0.237	0.43	0.07	3.17	65.3	0.93	1.76
	WD$'$ (m)	0.65	9530	0.10	0.07	1.43	49.5	−0.11	−0.15
Summer	pH	7.06	0.96	0.116	6.31	7.59	4.8	−0.28	−0.56
	EC ($\mu S\ cm^{-1}$)	8660	12.4	70.8×10^6	1040	27 500	97.2	1.11	−0.23
	EC$'$ ($\mu S\ cm^{-1}$)	83.3	0.94	1780	32.0	166	50.6	0.72	−0.83
	ESP	21.3	0.46	293	1.4	71.6	80.6	1.21	1.12
	ESP$'$	2.80	0.26	0.71	0.88	4.28	30.0	−0.36	−0.41
	WD	1.26	0.23	0.49	0.32	2.8	55.7	0.69	−0.11

Copyright © 2017 John Wiley & Sons, Ltd.

strong positive skew (skewness of 1.73–4.31), and significantly higher kurtosis of normal distribution kurtosis (kurtosis of 3.24–19.9). Spring and summer data sets showed much less deviation from the normal distribution with a skewness of 0.01–1.93 and kurtosis of 0.28–2.6, respectively, for spring, and a skewness of 0.28–1.21 and kurtosis of −0.11 to 1.12 for summer. Since normality is essential in kriging methods, log transformation was applied for data with a skewed distribution and a few very large values to produce bell-shaped histograms. Both sample variance and coefficient of variation (CV) for all parameters decreased after normalization. Thus, further geostatistical analysis was done on the log-transformed data sets. However, to produce a GIS map, the predictions were transformed back to the original values.

Geostatistical approach

Structural analysis. The best semivariogram for different periods is presented in Figure 2 and the characteristics of these semivariograms are summarized in Table II. Different models were fitted to a specific parameter on different sampling dates except for ESP, for which a spherical model was

Table II. Parameters of semivariogram models for groundwater chemical properties and water depth for different measuring periods (OK)

Period		Co	Co + C	*R* (m)	CI
Winter	pH	0.06	0.11	41 100	0.50
	EC	0.15	1.06	6 500	0.15
	ESP	0.00	1.07	5 500	0.00
	WD	0.01	0.07	1 780	0.13
Spring	pH	0.04	0.09	36 100	0.44
	EC	0.49	0.98	41 100	0.50
	ESP	0.00	0.88	910	0.00
	WD	0.08	0.17	35 200	0.46
Summer	pH	0.06	0.11	41 100	0.50
	EC	1690	3390	41 100	0.50
	ESP	0.00	0.84	900	0.00
	WD	0.44	1.76	38 800	0.25

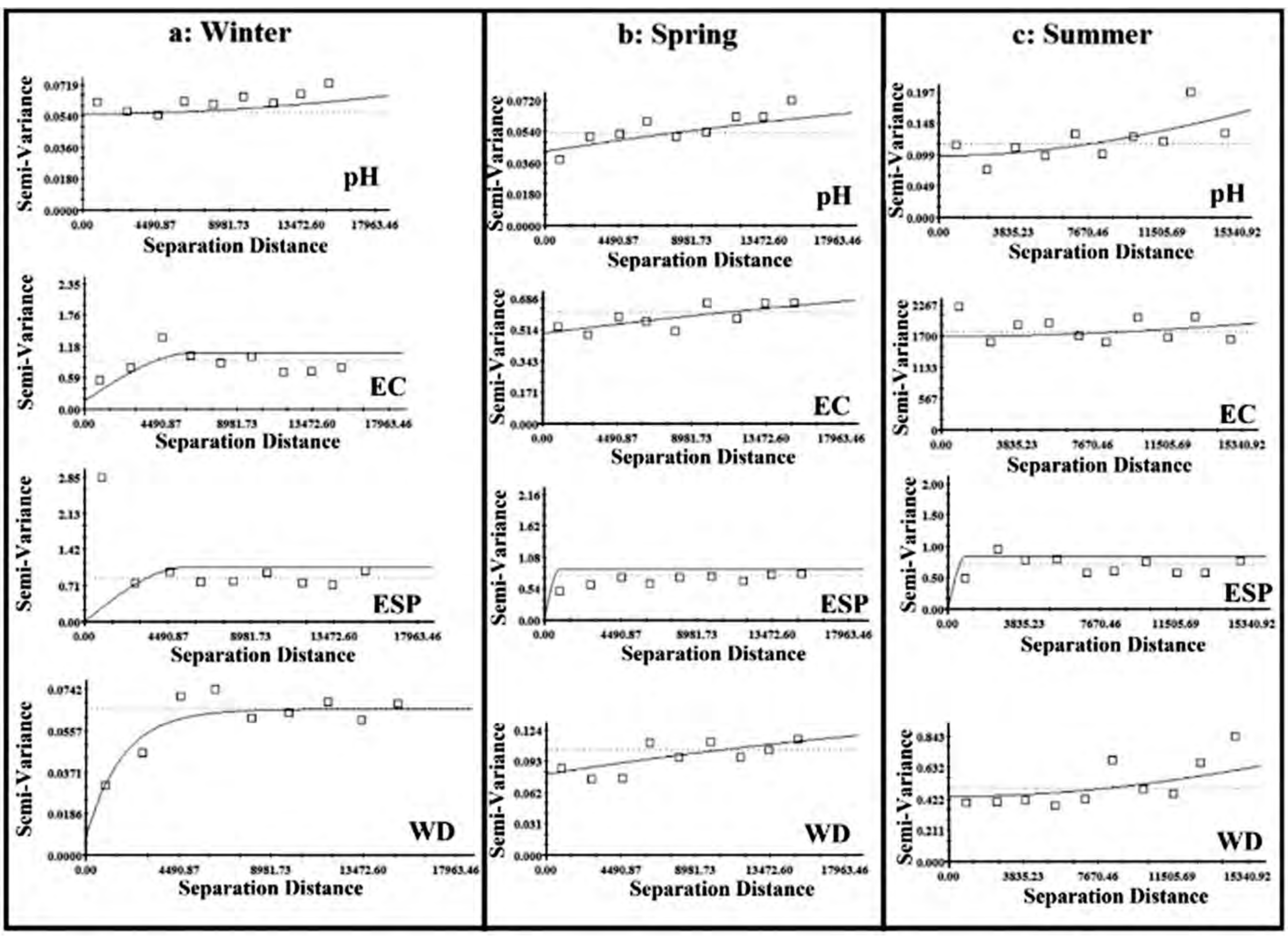

Figure 2. Semivariogram models for groundwater chemical properties and WD for winter (a), spring (b) and summer (c) periods.

Copyright © 2017 John Wiley & Sons, Ltd.

fitted to the observed data on the three sampling dates. A spherical model was advised for investigating groundwater quality parameters in some previous research (Castrignanò *et al*., 2008). However, regional conditions greatly affect the suitability of semivariograms. Spatial dependence of groundwater properties can be classified according to nugget to sill ratio, known as the Cambardella Index (CI) (1994), with CI < 0.25 indicating a strong spatial dependence, between 0.25 and 0.75 indicating moderate spatial dependence, and CI > 0.75 indicating a weak spatial dependence. A medium spatial dependence was observed for all parameters except for EC in winter and WD in winter and summer.

Evaluation of different geostatistical methods. The results of cross-validation between measured and estimated values of pH, EC, ESP and WD for different sampling periods are presented in Table III. Maps produced by OK were smoother than those provided by WMA. It should be noted that OK considers the spatial dependence pattern of observed data, while only the distance between predicted and observed locations is considered in the WMA method. On the other hand, the locations of wells are more obviously shown on WMA maps than on the other maps due to the nature of the WMA which generally provides spikes around the sample points (Lloyd, 2005). Overall, maps generated using OK seemed to be more consistent with observed data. Estimating groundwater quality by OK has been reported by Karandish and Shahnazari (2014). Similar results were found for EC, pH and WD. The different interpolation techniques were visually and quantitatively compared. For all data sets, OK provided the best results, having the least RMSE, MAE and MBE. Other studies also reported that OK would significantly reduce the uncertainty of groundwater chemical properties and water depth and would increase estimation accuracy (Assaf and Saadeh, 2009). A positive MBE index for all interpolation techniques revealed an overall underestimation for all methods. Underestimation may cause false decisions in remediation procedures and may waste investment. Table III shows a relatively high error for all methods which can arise from large variability or sparse sampling or both (Castrignanò *et al*., 2008). However, the OK interpolation method led to better results.

ANFIS and SVM

For both ANFIS and SVM models, geographic coordinates of wells including latitude and longitude were introduced as independent variables and the values of pH, ESP, WD and EC parameters were introduced as dependent variables. The evaluation results of the ANFIS model are tabulated in Table IV. For all MFs, the 5×5 combination for inputs often had the least RMSE and MAE and the highest EF. The gaussmf, gbellmf, gaussmf and gaussmf functions with 5×5 combinations for each of the input variables were the most suitable MFs for pH, ESP, WD and EC, respectively. These functions underestimated all variables except for EC. The selected function for EC estimated electrical conductivity of groundwater slightly more than observed values. The values of RMSE in the training, testing and control stages were, respectively, 0, 1.01 and 0.22 for pH; 0.12, 3.37 and 5.59 for ESP; 0.29, 0.67 and 0.83 m for WD and 0.4, 5.5 and 9.8 dS m^{-1} for EC. Also, for the selected model, the averages of MBE in the training, testing and control stages were, respectively, 0, 0.02 and 0.01 for pH; 1.01, 0.24 and 0.48 for ESP; -0.00, 0.01 and 0.04 m for WD and 0, -1.8 and -3.3 dS m^{-1} for EC. Such values demonstrate the reliability of the selected methods for simulating the variables in the study area.

The results of the testing of SVM models are presented in Table V. For SVM modelling, an appropriate choice of

Table III. Cross-validation between measured and estimated values of pH, EC, ESP and WD

Period	Method	WD (m)			ESP			EC (μS cm^{-1})			pH		
		MBE	MAE	RMSE	MBE	MAE	RMSE	MBE	MAE	RMSE	MBE	MAE	RMSE
Winter	WMA-2	0.06	0.33	0.48	2.80	14.2	19.1	1 030	7 970	15 400	0.01	0.30	0.33
	WMA-4	0.06	0.34	0.49	2.26	14.4	19.2	825	7 860	155 000	0.01	0.31	0.34
	WMA-6	0.05	0.35	0.50	1.80	14.6	19.3	666	7 730	15 500	0.01	0.32	0.35
	Kriging	0.05	0.33	0.48	1.34	14.1	18.9	1 020	6 680	12 500	−0.01	0.28	0.30
Spring	WMA-2	0.11	0.52	0.68	2.20	11.3	13.6	750	3 070	3 940	0.03	0.28	0.28
	WMA-4	0.11	0.54	0.69	1.81	11.7	13.8	660	3 140	3 970	0.02	0.27	0.28
	WMA-6	0.10	0.57	0.71	1.42	12.1	14.1	570	3 220	4 020	0.02	0.27	0.29
	Kriging	0.08	0.50	0.67	1.02	10.6	12.8	118	873	943	0.02	0.27	0.28
Summer	WMA-2	0.09	0.43	0.58	3.93	16.6	23.9	1 230	6 920	9 540	−0.02	0.28	0.33
	WMA-4	0.08	0.44	0.59	3.49	17.1	24.1	1 180	6 890	9 410	−0.02	0.29	0.34
	WMA-6	0.08	0.46	0.60	3.01	17.1	24.1	1 060	6 790	9 280	−0.03	0.30	0.35
	Kriging	0.06	0.42	0.57	2.41	14.8	22.0	1 062	5 140	8 470	0.00	0.26	0.31

Copyright © 2017 John Wiley & Sons, Ltd.

Table IV. Comparison of different MFs with various combinations for pH, EC, ESP and WD

	MF	N. MF	RMSE			MBE			EF		
			Train	Test	Control	Train	Test	Control	Train	Test	Control
WD (m)	gbellmf	2	0.35	0.96	0.93	0.01	0.12	0.02	0.74	0.70	−1.29
		3	0.41	1.24	0.75	0.07	0.24	−0.31	0.65	−2.89	−0.47
		4	0.27	2.28	0.67	0.10	0.70	0.06	0.85	0.51	−0.18
		5	0.31	0.93	0.64	-0.02	−0.11	0.02	0.80	0.77	−0.68
	gaussmf	2	0.40	0.91	0.74	0.01	−0.03	0.23	0.66	0.97	−0.63
		3	0.66	1.67	1.16	0.14	−0.06	−0.28	0.09	−2.26	−3.05
		4	0.40	2.14	0.72	0.12	0.65	0.14	0.66	0.62	−0.41
		5	0.16	0.20	0.19	−0.00	0.02	0.05	0.94	0.97	0.92
	gauss2mf	2	0.38	0.84	0.71	0.02	−0.01	0.16	0.69	0.97	−0.46
		3	0.46	1.66	1.11	−0.01	−0.44	−0.34	0.56	−0.24	−2.79
		4	0.73	1.15	0.85	0.00	−0.12	0.05	−0.14	0.41	−1.23
		5	0.29	0. 67	0.83	−0.00	0.01	0.04	0.83	0.97	0.72
EC (μS cm^{-1})	gbellmf	2	9 630	11 600	16 200	−390	−5 530	−9 730	0.50	1.00	−8.00
		3	60.6	10 800	5 130	0.1	−3 340	−1 890	1.00	−1.10	0.10
		4	3.4	5 100	14 100	−0.3	−1 870	−5 040	1.00	0.80	−5.80
		5	1.0	1 520	29 400	0.0	−181	−1 710	1.00	1.00	0.65
	gaussmf	2	1.0	5 320	29 500	0.0	−2 630	−8 710	1.00	0.90	−26.3
		3	9 690	15 400	10 600	−1030	−6 990	−3 110	0.50	−2.80	−8.70
		4	11.0	735	14 200	−0.5	−190	−5 460	1.00	1.00	−5.20
		5	0.4	5.5	9.8	0.0	−1.8	−3.3	1.00	1.00	1.00
	gauss2mf	2	7 730	23 500	35 700	−393	−11 300	−15 600	0.70	0.90	−38.7
		3	34.5	24 900	19 900	−0.4	−6 030	−6 810	1.00	−10.4	−30.1
		4	8.1	15 900	31 000	0.0	−5 400	−11 900	1.00	1.00	−28.7
		5	0.8	3 880	11 300	0.0	−1 680	−6 490	1.00	0.90	−0.78
pH	gbellmf	2	0.16	0.78	0.69	0.00	−0.07	0.06	0.81	−0.53	−4.68
		3	0.00	0.60	0.94	0.00	−0.32	0.12	1.00	0.99	−9.51
		4	0.00	1.59	2.31	0.00	0.26	0.25	1.00	−1.72	−62.2
		5	0.00	0.48	0.06	0.00	0.01	0.01	1.00	1.00	0.958
	gaussmf	2	0.16	0.90	0.69	0.00	−0.08	0.07	0.82	−1.30	−10.1
		3	0.00	0.65	1.23	0.00	−0.23	-0.02	1.00	0.21	−18.7
		4	0.00	1.56	1.55	0.00	0.58	0.02	1.00	−2.05	−34.7
		5	0.00	1.01	0.22	0.00	0.02	0.01	1.00	0.43	0.45
	gauss2mf	2	0.17	0.47	0.85	0.00	−0.18	−0.14	0.78	0.83	−8.04
		3	0.00	0.51	0.76	0.00	0.04	0.13	1.00	0.47	−7.15
		4	0.00	4.04	2.82	0.00	1.98	0.66	1.00	−4.53	−106
		5	0.00	0.45	0.33	0.00	0.03	0.11	1.00	0.65	0.48
ESP	gbellmf	2	10.1	11.9	12.9	0.60	1.61	3.66	0.56	0.93	−0.52
		3	0.02	3.96	6.14	0.00	−0.50	1.22	1.00	0.99	0.66
		4	5.42	26.0	19.6	1.06	−1.91	4.06	0.87	0.99	−2.50
		5	0.01	2.37	5.86	1.01	0.24	0.48	0.87	0.98	0.82
	gaussmf	2	11.4	13.0	11.9	−0.01	−1.39	1.08	0.43	0.55	−0.59
		3	4.57	30.0	22.7	0.31	14.5	7.44	0.91	0.96	−3.47
		4	4.47	24.4	30.3	0.68	−0.46	8.15	0.91	0.96	−6.93
		5	2.40	12.4	10.7	0.69	−1.27	1.05	0.99	0.98	0.87
	gauss2mf	2	10.4	15.3	20.2	0.63	−6.45	0.74	0.52	0.99	−2.51
		3	4.74	29.1	31.3	0.29	−3.09	8.81	0.90	0.38	−7.92
		4	3.08	19.1	24.1	0.39	−0.49	4.29	0.96	0.98	−3.99
		5	2.46	13.3	16.7	0.63	1.38	0.57	0.91	0.99	0.92

kernels allows the data to become separable in feature space despite being non-separable in the original space. This allows one to obtain nonlinear algorithms from algorithms previously restricted to handling linearity separable data sets (Bray and Han, 2004). Here, the radial basis function was the best kernel for all SVM models. With regard to RMSE (i.e. RMSE in the control phase was 0.57, 0.24, 0.25 and 8.25 for WD, EC, pH and ESP, respectively), SVM models performed slightly better than ANFIS ones for all parameters. Thus, SVM models were selected for preparation of spatial distribution maps of pH, ESP, WD, and EC variables.

Copyright © 2017 John Wiley & Sons, Ltd.

Table V. The final architecture and criteria indices of the SVMs

Parameter	Best kernel	RMSE			MBE			EF		
		Train	Test	Control	Train	Test	Control	Train	Test	Control
WD (m)	Radial basis function	0.33	0.81	0.57	0.03	−0.05	0.06	0.77	0.68	−0.52
EC (µS cm^{-1})	Radial basis function	0.8×10^5	0.01	2.4	0.03×10^5	0.02	0.01	0.83	0.95	0.58
pH	Radial basis function	0.10	0.84	0.25	0.10	0.04	0.06	0.85	0.37	0.46
ESP	Radial basis function	1.90	9.41	8.58	0.61	−0.85	0.87	0.84	0.83	0.73

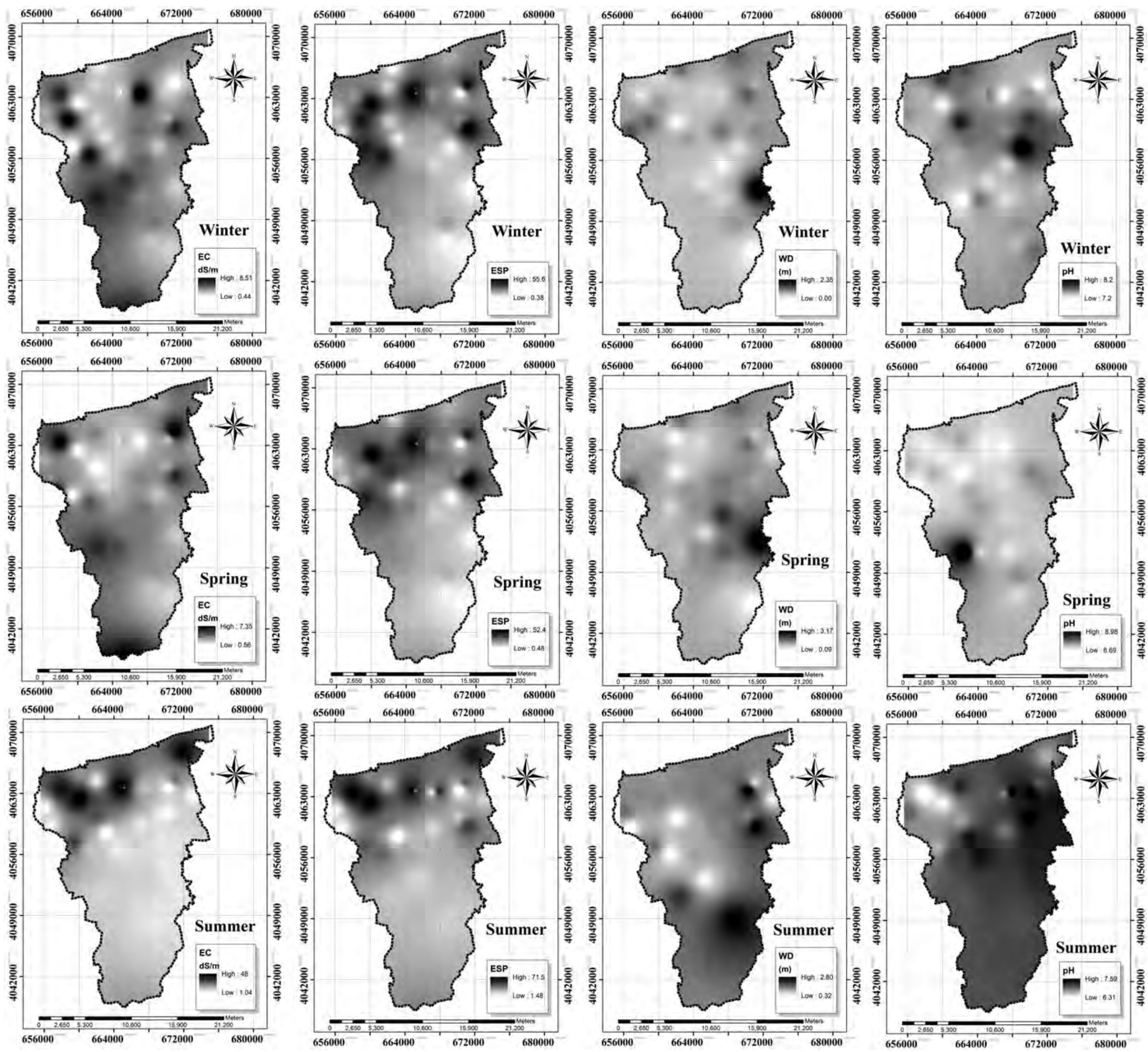

Figure 3. The spatial distribution of EC, ESP, WD and pH for winter, spring and summer periods based on the SVM model.

Copyright © 2017 John Wiley & Sons, Ltd.

Prioritizing drainage requirement

Using selected SVM models, raster maps of predicted values of pH, ESP, WD and EC variables were provided in a regular network in a GIS environment and are presented in Figure 3. Generally, the range of WD fluctuations in spring (0.09–3.17 m) was more than that in summer (0.32–2.8 m) and winter (0–2.35 m). Although the range of WD fluctuations in winter was close to that in summer, WDs were lower in summer in large parts of the region mainly due to water consumption of cultivated crops and low rainfall in this season. Also, frequent rainfall in winter resulted in a lower WD in this season. Changes in ESP and EC were highly correlated with changes in WD, so that the values of these variables in summer and spring were considerably more than those in winter when the WD was low. Due to application of chemical fertilizers and agricultural activities during thegrowing season, salinity and sodicity risks were higher in the study area in summer than the risks in other seasons. Paddy fields as the dominant land use in the study area have a specific hydrology, with the wetland condition prevailing during the growing season as well as in rainy seasons. Paddy farmers generally apply more water to the field to control the weed germination than is needed for meeting crop water requirements. So, a substantial amount of applied water is lost by deep percolation from the root zone. Therefore, agricultural leachates as a result of excessive use of fertilizers and pesticides often contribute significantly to groundwater pollution (Jeyaruba and Thushyanthy, 2009). Previous studies have reported salinity problems of groundwater due to agricultural practices (Mace *et al.* 2006; Chaudhuri and Ale, 2014). EC values were in the range 0.00–7.34 dS m^{-1} in spring and 0.00–8.51 dS m^{-1} in winter, while such values fluctuated between 1.04 and 48 dS m^{-1} in summer. A similar trend was observed for ESP. ESP varied by 0.47–52.4, 1.47–71.6 and 0.38–55.6, respectively, in spring, summer and winter. Fewer changes were observed for pH in the sampling periods. Groundwater samples were almost alkaline in spring and winter, and were almost acidic in summer. Acidity is one of the distinguishing characteristics of saline and sodic soils. The pH of saturated soil pastes of sodic soils is 8.2 or more and in extreme cases may be above 10.5 (Food and Agriculture Organization of the United Nations (FAO), 1988). So, considering the spatial distribution of acidity implies that the study area generally suffers from salinity rather than sodicity. However, in some parts of the region in spring pH was more than 8.2, demonstrating more sodicity problems.

Classification of drainage requirements of the study area was done based on salinity, sodicity and waterlogging problems. Regions with groundwater EC and ESP higher than 4 dS m^{-1} and 15, respectively, were considered to have salinity and sodicity problems. Also, regions with WD lower than 2 m were considered to be waterlogged. Therefore, it is assumed that regions with EC $>$ 4 dS m^{-1}, ESP $>$ 15

Table VI. Seasonal variations in land areas in relation to EC, ESP and WD

Season	Area (%)					
	EC $<$ 4 dS m^{-1}	EC $>$ 4 dS m^{-1}	ESP $<$ 15	ESP $>$ 15	WD $<$ 2 m	WD $>$ 2 m
Winter	80.9	19.1	79.4	20.6	99.4	0.6
Spring	92.1	7.9	75.2	24.8	97.4	2.6
Summer	12.3	87.7	33.1	66.9	92.7	7.3

Table VII. Seasonal changes in drainage requirements in the study area

Drainage class	Description	Area (%)		
		Winter	Spring	Summer
1	No need for drainage	2.6	2	0.61
2	Drainage for lowering groundwater table	68.1	3.9	64.2
3	Drainage for sodic soil reclamation	22.1	3.3	15.8
4	Drainage for lowering groundwater table and saline soil reclamation	0	0.01	0
5	Drainage for saline soil reclamation	4.7	71.8	15
6	Drainage for lowering groundwater table and sodic soil reclamation	0	4.2	0
7	Drainage for saline sodic soil reclamation	2.5	13.7	4.4
8	Drainage for lowering groundwater table and saline sodic soil reclamation	0	1.1	0

Copyright © 2017 John Wiley & Sons, Ltd.

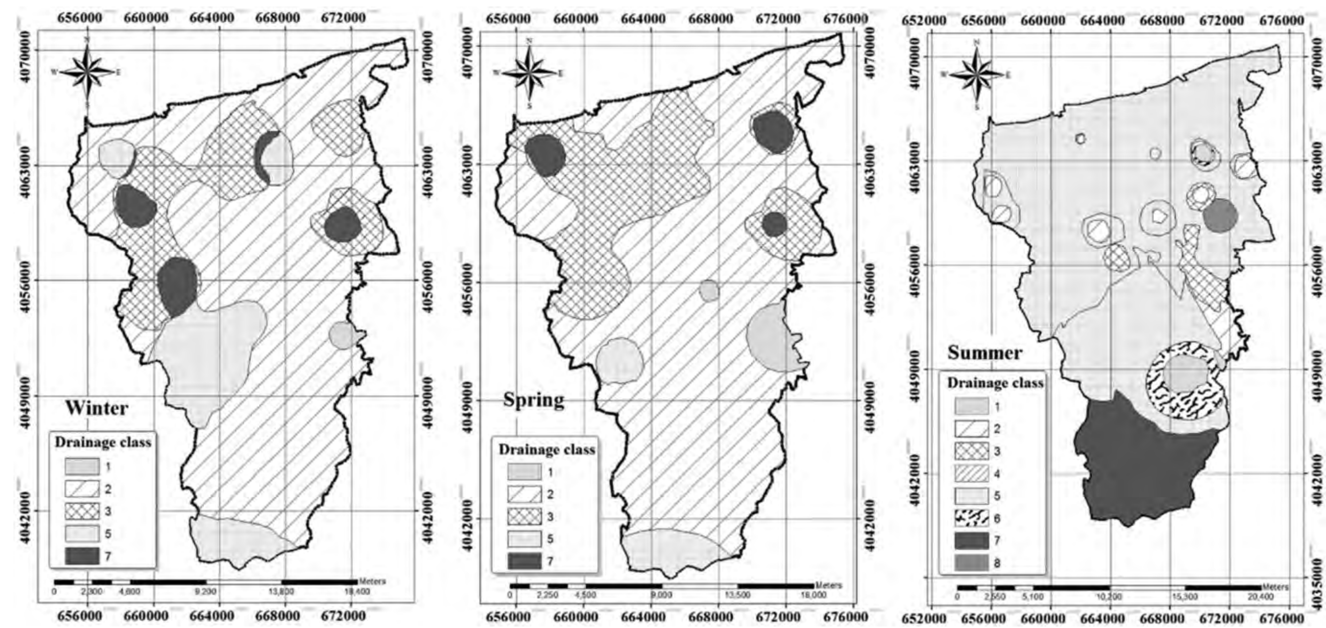

Figure 4. Classification of drainage needs in the study area in winter, spring and summer.

and WD < 2 m require some form of drainage to control salinity, sodicity and waterlogging problems, respectively. Table VI represents seasonal variation of areas with different EC, ESP and WD classes. In summer, more than 87.7 and 66.9% of the study area had salinity and sodicity problems, respectively (Table VI). However, only 19.1 and 7.9% of the study area had salinity problems in winter and spring, respectively. Also, the values of ESP in winter and spring were more than 15, respectively, in 20.6 and 24.8% of the study area, but waterlogging was common problem in winter and spring, occurring in 99.4 and 97.4% of the study area. Salt leaching in winter and increased WD as a result of groundwater pumping exceeding the groundwater recharge rate due to unsuitable management activities in summer, caused salinity problems in winter and summer compared with those in spring. Considering the high risk of salinity and sodicity in summer, improvement of drainage conditions is inevitable if sustainable agriculture in the region is to be achieved. Also, preparing land for on-time cultivation in spring requires implementing drainage systems to lower the groundwater table more rapidly and prevent yield losses due to delayed cropping stress.

Through combination of three indicators (EC, ESP and WD) and overlaying their maps, eight drainage classes were defined as described in Table VII. The spatial distribution of drainage classes is presented in Figure 4. Also, seasonal changes in drainage requirements are presented in Table VII. Drainage was necessary in a large area of the region for various reasons. There was no drainage requirement in only about 2.6, 2 and 0.61% of the total area, respectively, in spring, summer and winter. The maximum drainage requirement in spring, summer and winter was related to, respectively, 'drainage to remove excess water', 'drainage to control salinity' and 'drainage to remove excess water'. Generally, the need for drainage in a region mainly depends on the agricultural activities, rainfall pattern, irrigation, etc. Based on these results, in the wet seasons, drainage should be carried out to remove excess water and to control waterlogging, while at the end of the growing season it is required for controlling salinity and sodicity.

CONCLUSION

Sustainable agriculture in the coastal area of northern Iran is primarily dependent on diagnosing and combating drainage problems. Electrical conductivity (EC), exchangeable sodium percentage (ESP), acidity (pH) and groundwater level (WD) were selected for classifying drainage requirements in this area. These parameters were analysed by using geostatistical methods, SVMs and ANFIS models on three occasions, including pre-planting period (winter), cultivation period (spring) and post-harvest period (summer). Results demonstrated the great ability of SVM models to predict the selected parameters for all periods. Increased agriculture due to rapid population growth in the study area combined with improper usage of fertilizers and overexploitation of groundwater have caused almost all parts of the study area (more than 90%) to need drainage to achieve sustainable agriculture. However, drainage requirements are highly dependent on the measurement period. In fact, in the wet seasons, drainage should be carried out to remove

Copyright © 2017 John Wiley & Sons, Ltd.

excess water and to control waterlogging, while at the end of the growing season, it is required to control salinity and sodicity.

REFERENCES

Aghazadeh N, Asghari-Mogaddam A. 2010. Assessment of groundwater quality and its suitability for drinking and agricultural uses in the Oshnavieh area, northwest of Iran. *Journal of Environmental Protection* **1**: 30–40.

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. *Environmental Modelling and Software* **20**: 851–871.

Assaf H, Saadeh M. 2009. Geostatistical assessment of groundwater nitrate contamination with reflection on DRASTIC vulnerability assessment: the case of the Upper Litani Basin, Lebanon. *Water Resources Management* **23**(4): 775–796.

Borga M, Vizzaccaro A. 1997. On the interpolation of hydrologic variables: formal equivalence of multiquadratic surface fitting and kriging. *Journal of Hydrology* **195**(1–4): 160–171.

Bray M, Han D. 2004. Identification of support vector machines for runoff modeling. *Journal of Hydroinformatics* **6**(4): 265–280.

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 Science Society of America Journal* **58**: 1501–1511.

Castrignanò 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 EuroAgeng 2008 23th –25th June 2008,* Hersonissos, Crete Isle, Greece. CD ROM (in print).

Chaudhuri S, Ale S. 2014. Long term (1960–2010) trends in groundwater contamination and salinization in the Ogallala aquifer in Texas. *Journal of Hydrology* **513**: 376–390.

Chiles JP, Delfiner P. 1999. *Geostatistics: Modelling Spatial Uncertainty*, Wiley: New York; 695 pp.

Darzi-Naftchali A, Mirlatifi SM, Shahnazari A, Ejlali F, Mahdian MH. 2013. Effect of subsurface drainage on water balance and water table in poorly drained paddy fields. *Agricultural Water Management* **130**: 61–68.

Deng J, Chen X, Du A, Zhang Y. 2011. Soil water simulation and predication using stochastic models based on LS–SVM for red soil region of China. *Water Resources Management* **25**: 2823–2836.

El-Nashar WY. 2013. The combined effect of water-logging and salinity on crops yield. *IOSR Journal of Agriculture and Veterinary Science* **6**(4): 40–49.

Food and Agriculture Organization of the United Nations (FAO). 1988. *Salt-affected soils and their management* FAO Soils Bulletin: Rome, Italy; 39.

Goovaerts P. 1997. *Geostatistics for Natural Resources Evaluation*, Oxford University Press: New York; 483 pp.

Issaks E, Srivastava RM. 1989. *Applied Geostatistics*, Oxford University Press: New York.

Jeyaruba T, Thushyanthy M. 2009. The effect of agriculture on quality of groundwater: a case study. *Middle-East Journal of Scientific Research* **4**(2): 110–114.

Karandish F, Shahnazari A. 2014. Appraisal of the geostatistical methods to estimate Mazandaran coastal ground water quality. *Caspian Journal of Environmental Science* **12**(1): 129–146.

Karandish F, Darzi-Natchali A, Asgari A. 2016. Application of machine-learning models for diagnosing health hazard of nitrate toxicity in shallow aquifers. *Paddy and Water Environment*. DOI:10.1007/s10333-016-0542-2.

Khalil A, Almasri MN, McKee M, Kaluarachchi JJ. 2005. Applicability of statistical learning algorithms in groundwater quality modeling, *Water Resources Research* **41**: W05010, DOI:10.1029/2004WR003608.

Liu D, Yu Z, Lü H. 2010. Data assimilation using support vector machines and ensemble Kalman filter for multi-layer soil moisture prediction. *Water Science and Engineering* **3**: 361–377.

Lloyd CD. 2005. Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. *Journal of Hydrology* **308**: 128–150.

Mace RE, Davidson SC, Angle ES, Mullican WF. 2006. *Aquifers of Gulf Coast of Texas*. Texas Water Development Board Report: Texas, USA; 304 pp.

Rehman S, Mohandes M. 2008. Artificial neural network estimation of global solar radiation using air temperature and relative humidity. *Energy Policy* **36**(2): 571–576.

Singh KP, Basant N, Gupta S. 2011. Support vector machines in water quality management. *Analytica Chimica Acta* **703**(2): 152–162.

Sudheer CH, Maheswaran R, Panigrahi BK, Mathur S. 2014. A hybrid SVM-PSO model for forecasting monthly streamflow. *Neural Computing & Applications* **24**(6): 1381–1389.

Sun B, Zhou S, Zhao Q. 2003. Evaluation of spatial and temporal changes of soil quality based on geostatistical analysis in the hill region of subtropical China. *Geoderma* **115**: 85–99.

Tabari H, Nikbakht J, Shifteh S'e B. 2012. Investigation of groundwater level fluctuations in the north of Iran. *Environmental Earth Sciences* **66**: 231–243.

Tang AM, Quek C, Ng GS. 2005. GA-TSKfnn: parameters tuning of fuzzy neural network using genetic algorithms. *Expert Systems with Applications* **29**: 769–781.

Wang WC, Chau KW, Cheng CT, Qiu L. 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology* **374**(3–4): 294–306.

Webster R, Oliver MA. 2006. Optimal interpolation and isarithmic mapping of soil properties. VI. Disjunctive kriging and mapping the conditional probability. *European Journal of Soil Science* **40**: 497–512.

Wilding LP, Dress LR. 1983. Spatial variability and pedology. In: *Pedogenesis and Soil Taxonomy. I. Concepts and Interactions* Wilding LP, Smeckand NE, Hall GF (eds). Elsevier Science: New York, NY; 83–116 pp.

Yoon H, Jun SC, Hyun Y, Bae GO, Lee KK. 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *Journal of Hydrology* **396**(1–2): 128–138.

Yost RS, Uehara G, Fox RL. 1982. Geostatistical analysis of soil chemical properties of large land areas: I. Semivariograms. *Soil Science Society of America Journal* **46**: 1028–1032.

Zhou HZ, Gong ZT, Lamp J. 1996. Study on soil spatial variability. *Acta Pedologica Sinica* **33**: 232–241 (in Chinese with English abstract).

Copyright © 2017 John Wiley & Sons, Ltd.