DESERT

Spatial prediction of soil electrical conductivity using soil axillary data, soft data derived from general linear model and error measurement

N. Hamzehpour[*], M. Rahmati, B. Roohzad

Department of Soil Science and Engineering, University of Maragheh, Maragheh, East Azerbaijan, Iran

Abstract

Indirect measurement of soil electrical conductivity (EC) has become a major data source in spatial/temporal monitoring of soil salinity. However, in many cases, the weak correlation between direct and indirect measurement of EC has reduced the accuracy and performance of the predicted maps. The objective of this research was to estimate soil EC based on a general linear model via using several soil properties. Through calibration equations, the error involved in such model-based data was calculated and employed in mapping soil EC using kriging with measurement errors (KME) method. The results were then compared with those of ordinary kriging (OK) and co-kriging (CK). Soil samples were taken from the depth of 0-20 cm in 78 points with spatial intervals of 500 m from an area of 40 km², and they were analyzed for their electrical conductivity (EC) and certain other soil properties. Measured soil EC data (hard data) and auxiliary soil data were further used to develop the semi-variance and cross-semi-variance functions; moreover, soil salinity prediction was done on a grid of 100 m with OK and CK methods. Afterwards, the most optimal EC estimation model was developed using auxiliary soil data and GLM. As predicted values always involve uncertainty, the error involved with the predicted values was calculated and then the calibration equations were adjusted. Lastly, soil salinity was predicted using KME method. Results showed that the OK method had the lowest MSE and RMSE values, 0.65 and 0.8 dS m⁻¹, respectively. Furthermore, among the auxiliary data, pH and silt content resulted in some of the best cross-semi-variance functions, among which, silt had a better performance regarding the spatial prediction of soil EC. The GLM model developed with the calculated error and KME resulted in predictions close to those of OK method (with MSE and RMSE of 0.74 and 0.86 dS m⁻¹, respectively). KME method provided the possibility of merging error resulting from the use of soft data, derived from prediction equations; therefore, it successfully improved the spatial prediction of soil electrical conductivity.

Keywords: Co-kriging; Kriging with measurement errors; Soil salinity; Spatial dependency

1. Introduction

Over the past decades, spatial prediction methods such as kriging have played a significant role in reducing both the number of samples necessary to monitor a large area and the data gathering expenses, through predicting soil variables in some un-sampled locations. Recent improvements in the field of geostatistics and the advances in calculating complex problems have made it possible to analyze variables with spatial correlation. Kriging methods have always had a

widespread use in geostatistical modeling, which has been discussed in detail in several studies (Krivoruchko and Gribov, 2019; Shi *et al.,* 2019; Keskin and Grunwald, 2018; Li and Heap, 2011). There have also been numerous attempts at mapping the spatial variability of soil electrical conductivity (EC) using kriging methods (Triantafilis *et al.,* 2004 ; Malins and Metternicht, 2006; de Clercq *et al.,* 2009; Giordano *et al.,* 2010; Acosta *et al.,* 2011; Hamzehpour and Bogaert, 2017; Zare *et al.,* 2019; Yang *et al.,* 2019).

However, during the past few years, there have been several studies that have essentially focused on using soil properties and remote sensing derived indices as covariates in modeling

---
[*] Corresponding author. Tel.: +98 41 37276060
   Fax: +98 41 37276060
   E-mail address: nhamzehpour@maragheh.ac.ir

and predicting soil EC (Rahmati and Hamzehpour, 2017; Zovko *et al.,* 2018; Hammam and Mohamed, 2018; Abdullah *et al.,* 2019; Zare *et al.,* 2019). For instance, Abdullah *et al*. (2019) modeled soil salinity via the use of five salinity indices and eleven environmental variables. Based on their results, the most optimal performance was obtained from the combination of both salinity indices and indirect variables. Zovko *et al.* (2018) defined and used a spectral index which synthesized most of the salt-affected soil properties in order to predict soil electrical conductivity. They further compared it with the results of ordinary kriging and a multivariate predictor utilizing certain soil chemical properties. Juan *et al.* (2011) made use of a spatial Gaussian linear mixed model to calculate soil salinity using soil EC and Na content. Clercq *et al.* (2009) employed a first-order polynomial equation to map the spatial and temporal variations of soil salinity. However, none of these studies have accounted for the uncertainty and error involved in the use of model-derived estimations of soil electrical conductivity, which are not the actual measurements of soil EC.

Several researchers have tried to calculate the errors attached to uncertain data in order to improve the prediction accuracy (D'Or *et al.,* 2001; Bogaert and D·Or, 2002; Douaik *et al.,* 2005; Fazekas and Kukush, 2005; Brus *et al.,* 2008; Hamzehpour *et al.,* 2013; Hamzehpour and Bogaert, 2017). Hamzehpour and Bogaert (2017) used calibration equations between field-measured EC and laboratory-measured EC. Then, they calculated the error in one-time interval and used it in spatio-temporal prediction of soil salinity at some other time intervals. Hamzehpour *et al.* (2013) employed field-measured EC to predict the top soil salinity using kriging with measurement errors method. They further made use of the calibration equations between field and laboratory measured soil salinity and defined probability density functions to calculate the error of soft data used in prediction. Nonetheless, there are not many studies focusing on the spatial prediction of soil properties using auxiliary soil data and their uncertainty.

In almost all studies where EC values (or any other soil variable) were calculated based on any modeling type (linear/nonlinear) or any kind of data (soil/remote sensing), the estimated values were treated as error-free and used in spatial prediction of target variables without considering their uncertainty. In this paper, we attempted to estimate soil EC using several soil properties and the general linear model (GLM). Because estimated EC values always involve uncertainty, through calibration equations, their error was calculated and used in the spatial prediction of soil EC using kriging with measurement errors method (KME). Therefore, the results were compared with those of ordinary kriging (OK) and co-kriging (CK). The present work provided a framework for increasing the effectiveness of auxiliary soil data in the spatial prediction of any soil variable where there are limited available data.

2. Materials and Methods

2.1. Study site

The study area is located southeast of Urmia Lake, north-western Iran. The corresponding area is located between 45º 58' 41" to 46º 02' 35" E and 37º 16' 18" to 37º 20' 21" N (Figure 1), including 40 km^2 of lands with varying degrees of soil salinity. The elevation variations over the whole study area range from 1270 to 1283 m. The mean annual precipitation and temperature are 264.73 mm and 15 ºC, respectively. The potential evaporation in the area is between 900 and 1170 mm. Geologically, the study area is composed of playa and alluvial deposits.

2.2. Data description

In total, during autumn 2014, 78 soil samples were taken on a grid of 500 m, covering all of study area. After passing soil samples through 2 mm sieve, soil EC, pH, Na$^+$, and Ca^{2+}+Mg^{2+} were determined in the saturated extracts of the soil samples (Rhoades, 1996). Calcium carbonate equivalent (CCE) was measured using back titration of the remaining HCl (Page *et al*., 1982). Soil texture was measured through the use of a hydrometer method (Gee and Or, 2002). Soil organic carbon (OC) was determined with acid digestion (Nelson and Sommers, 1996). Sodium adsorption ratio (SAR) was calculated using equation 1 (soil survey staff, 2014):

$$SAR = \frac{soulable\ Na\ (mmol\ l^{-1})}{\sqrt{\frac{(soulable\ (Ca+Mg)(mmol\ l^{-1}))}{2}}} \qquad (1)$$
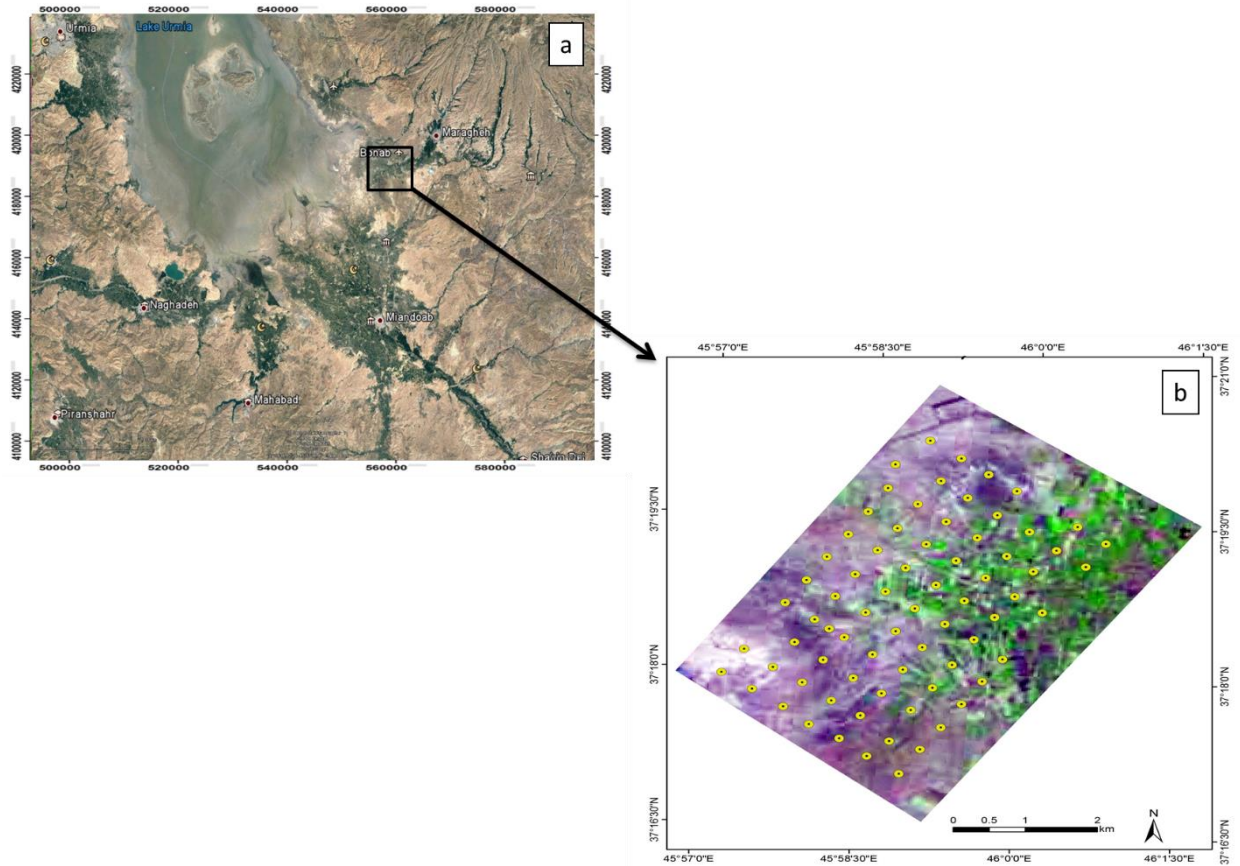
Fig. 1. (a) The location of the study area in southeastern Urmia Lake in Bonab plain; (b) the location of soil sampling points in Bonab plain

2.3. Geostatistical analysis

In this study, ordinary kriging (OK), co-kriging (CK), and kriging with measurement errors (KME) were used to predict the soil salinity using soil properties and soft data derived from general linear model (GLM).

2.3.1. Semivariance and cross-semivariance functions

The basic Kriging estimator is defined as follows (Li and Heap, 2011):

$$Z(x_0) - \mu = \sum_{i=1}^{n} \lambda_i [Z(x_i) - \mu(x_0)] \qquad (2)$$

which can be extended to incorporate additional information, as in equation 3:

$$\hat{Z}_1(x_0) - \mu_1 = \sum_{i_1=1}^{n_1} \lambda_i [Z_1(x_{i_1}) - \mu_1(x_{i_1})] + \sum_{j=2}^{n_v} \sum_{i_j=1}^{n_j} \lambda_{i_j} [Z_j(x_{i_j}) - \mu_j(x_{i_j})] \qquad (3)$$

where μ_1 is the stationary mean of the initial variable, $Z_1(x_{i_1})$ is the data at point i_1, $\mu_1(x_{i_1})$ is the mean of samples within the search window, n_1 is the number of sample points within the

search window for point x_0 used for estimation, (λ_{i_1}) is the weight selected to reduce the estimation variance of the initial variable, n_v is the number of secondary variables, n_j is the number of jth secondary variables within the search window, λ_{i_j} is the weight assigned to i_j^{th} point of jth secondary variable, $Z_j(x_{i_j})$ is the data at i_j^{th} point of jth secondary variable, and $\mu_j(x_{i_j})$ is the mean of the samples of jth secondary variable within the search window.

Cross-semivariance can be estimated using the following equation:

$$\hat{\gamma}_{12}(h) = \frac{1}{2n} \sum_{i=1}^{n} [z_1(x_i) - z_1(x_i) - z_1(x_i + h)][z_2(x_i) - z_2(x_i + h)] \qquad (4)$$

In the present research, Z_1 refers to EC, and Z_2 refers to the soil covariates.

2.4. Kriging with measurement errors

In an ordinary kriging, all of the measurements that are considered as error-free are called hard data. However, when the data used in predictions are not the result of direct

measurements (for example, calculated from a model developed from other soil properties), they are involved with uncertainty; moreover, although they might be highly correlated with direct measurements, they are not error-free (called soft data). In our case, the residual variation can be considered as a measurement error which can be smoothed out through the interpolation process. Ordinary kriging is briefly modified to account for measurement errors (Savelyeva *et al.*, 2010).

It is assumed that $Z(x_i)$ is a random field where Z_i is the measured value at the sampled location x_i for this random field, and ε_i is the corresponding error. Instead of the exact measurement of Z_i, we can measure $Z_i + \varepsilon_i$. The assumptions for taking ε_i into account are: (1) errors are not spatially correlated (i.e., $S_{ij}=E\{\varepsilon_i\varepsilon_j\}=0$), (2) errors are not correlated with the true value ($E\{Z_i\varepsilon_i\}=0$), and (3) they have a variance σ_i^2.

Considering the conventional kriging, the predicted value, $Z^*(x_i)$, at an unsampled location x_0 is presented in equation 5. Here we consider the estimation variance (σ_i^2) to infer the amount of error in variance equation:

$$Var(Z^*(x_0) - Z(x_0)) = E(\sum_{i=1}^{N(x_0)} \lambda_i(x_0)Z(x_i) - Z(x_0)^2 = \tag{5}$$

$$\sum_{i=1}^{N(x_0)} \lambda_i^2(x_0)(\sigma_i^2 + C_{00}) + 2\sum_{i=1}^{N(x_0)} \sum_{\substack{j=1 \\ j\neq i}}^{N(x_0)} \lambda_i(x_0)\lambda_j(x_0)C_{ij}$$

$$-2\sum_{i=1}^{N(x_0)} \lambda_i(x_0)C_{i0} + C_{00}$$

, where

$$C_{00} = Var(ZZ)\forall Z(x), C_{ij} = Var(Z_iZ_j) = Var(Z(x_i)Z(x_j)), \text{ and } C_{i.} = var(Z(x_i)Z(x.))0$$

Thus, the kriging system of equations is:

$$\lambda_i(x_0)(C_{00} + \sigma_i^2) + \sum_{\substack{j=1 \\ j\neq i}}^{N(x_0)} \lambda_j(x_0)C_{ij} = C_{i0} + \mu, i = 1 \dots N(x_0) \tag{6}$$

$$\sum_{i=1}^{N(x_0)} \lambda_i(x_0) = 1$$

Where σ_i^2 is the measurement error variance, which may differ from one location to another, and μ is the known stationary mean. The corresponding kriging variance is given by:

$$\sigma_k^2 = C_{00} - \sum_{i=1}^{N(x_0)} \lambda_i(x_0)C_{i0} + \mu. \tag{7}$$

For more detail on the formulas and theoretical framework of kriging with measurement errors, refer to Fazekas and Kukush (2005).

2.5. Validation and comparison criteria

In order to compare the results using cross-validation procedure (Davis, 1987), soil salinity was predicted at each observation location by deleting the measured values. This resulted in pairs of estimated-observed soil salinity. Among these pairs of values, three global performance criteria were computed, namely mean error (ME), mean squared error (MSE), and root mean square error (RMSE). They are defined as follows:

$$ME = \frac{\sum_{i=1}^{n} [z(u_i) - z^*(u_i)]}{n} \tag{8}$$

$$RME = \frac{\sum_{i=1}^{n} [z(u_i) - z^*(u_i)]2}{n} \tag{9}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} [z(u_i) - z^*(u_i)]^2}{n}} \tag{10}$$

Where $Z(u_i)$ is the measured value of z at location u_i, $z^*(u_i)$ is the predicted value at the same location, and *n* is the number of samples. Therefore, accurate predictions are characterized by ME, MSE, and RMSE values close to zero. In addition, cross-validation is used as another approach to assessing the predictive capabilities of interpolators.

All the analyses were done using the BMElib toolbox (Christakos and Serre, 2000) in Matlab (MathWorks, 1999).

2.6. GLM model performance assessment

To perform the regression analysis, the measured soil properties significantly correlated with soil EC at 1% probability level were imported to SPSS. The following criteria were applied to evaluate the regressions both in training and testing data:

a) Coefficient of determination (R^2)

Coefficient of determination depicts the correlation coefficient between the measured and predicted values, where zero indicates the worst performance and 1 means the most optimal one. The following equation was used to determine R^2:

$$R^2 = 1 - \frac{SSE}{SSTO} \tag{11}$$

Where R^2, SSE, and SSTO are coefficient of determination error, sum of squares, and total sum of squares, respectively.

b) Adjusted coefficient of determination (R^2_{adj})

The magnitude of R^2 typically increases with the increase in the number of independent variables. Adjusted coefficient of determination (R^2_{adj}) was employed to compare several methods with different numbers of independent variables. The following equation specified R^2_{adj}:

$$R^2_{adj} = 1 - \left(1 - R^2\right) \frac{N-1}{N-P-1} \qquad (12)$$

Where R^2_{adj}, N, and P are the adjusted coefficient of determination, number of observations, and independent variables, respectively.

3. Results and Discussion

3.1. Soil salinity prediction with ordinary kriging

The summarized statistics of the studied soil properties are presented in Table 1. According to Table 1, soil EC varied from 0.33 to 107.5 dS m^{-1} with a mean value of 10.95 dS m^{-1}. Due to the high skewness and kurtosis of soil EC data, their logarithmic form was used for further studies.

Table 1. Summary statistics of measured soil properties used as axillary data in soil EC prediction

Soil property	Unit	Mean	SD*	Max	Min	Skewness	Kurtosis
EC	dS m^{-1}	10.95	20.48	107.50	0.33	2.94	9.01
pH	-	8.16	0.44	8.97	7.10	-0.50	-0.24
Na	meq l^{-1}	29.58	67.60	436.92	0.92	4.04	18.32
Ca+Mg	meq l^{-1}	24.58	16.19	108.00	4.00	11.28	2.88
SAR	-	4.92	8.82	48.25	0.23	16.02	3.67
OM	(%)	0.60	0.32	1.92	0.00	1.09	2.56
CCE	(%)	18.57	3.91	30.58	12.64	0.72	0.95
Sand	(%)	61.55	18.19	90.00	14.00	2.41	1.39
Silt	(%)	26.70	13.03	56.00	6.00	-0.65	0.39
Clay	(%)	11.75	7.12	36.00	2.00	-0.34	-0.66

*Standard deviation

The best fitted semi-variogram to EC data set and its parameters are displayed in Figure 2a. Semi-variogram models with the smallest residual sum of squares were selected as the best fitting model. Figure 2 (b and c) shows the EC prediction map with ordinary kriging method along with cross-validation points. The fitted semi-variogram model had two parts: a nugget effect equal to 0.16 (dS m^{-1})^2 and a spherical part, with a range of 4580m and a sill of 5.25 (dS m^{-1})^2. The resulted semi-variogram was used in the spatial prediction of soil salinity on a grid of 100 meters. Cross-validation results showed that spatial prediction of EC with kriging method had an acceptable performance with MSE and RMSE of 0.65 and 0.8 dS m^{-1}, respectively (Table 2).

3.2. Soil salinity prediction with co-kriging

In order to use a soil property as covariate in the spatial prediction of a specific soil parameter such as EC, the covariate should have a number of significant correlations with the targeted parameter. To investigate the role of the studied soil properties in improving the spatial prediction of EC, the Pearson correlation coefficient was calculated between EC and the other studied soil properties. Figure 3 presents the correlation between the selected covariates and EC and the linear models fitted between them. As observed,

at 1% probability level, EC had a significant correlation with soil pH, soluble Na, soluble Ca+Mg, SAR, CCE, clay, silt, and sand. EC was negatively correlated with soil sand and pH. Aini *et al.* (2014) also reported significant negative correlation at 1% probability level between soil EC and pH. Accordingly, these eight soil parameters were selected as covariates in the spatial prediction of EC.

Later, the selected soil properties were used as covariates to calculate cross-semivariance functions. Among all the properties, only pH and silt resulted in the development of cross-semivariograms (Figure 4a and b) because only these two properties showed some elements of strong spatial dependency. Therefore, these two parameters were applied to the spatial prediction of soil EC with the co-kriging method. Studies have shown that there is a significant correlation between soil texture and soil EC (Jung *et al.*, 2005). Among soil texture fractions, sand content is negatively correlated with soil EC while silt and clay contents of soils have a positive correlation. This could be due to the positive role of silt and clay in increasing soil porosity in comparison to sand content (Bernoux *et al.*, 1998; Corwin and Lesch, 2005). Increased soil porosity can augment soil water holding capacity, thereby raising soil salt content and electrical conductivity (Inman *et al.,* 2002;

Brevik *et al*., 2006). The significant negative correlation between soil pH and soil EC is associated with the higher salt dissolution as soil pH decreases, resulting in increased soil EC (Mohd-Aizat *et al.,* 2014; Aini *et al.,* 2014).

Table 2 shows the cross-validation results of the co-kriging of soil EC with pH and silt as covariates, and Figure 4 (c and d) illustrates the corresponding maps. Based on the results, using soil pH, as covariate, reduced the accuracy of soil EC predictions through smoothing out the effect of points with significant differences with neighboring sampling points.
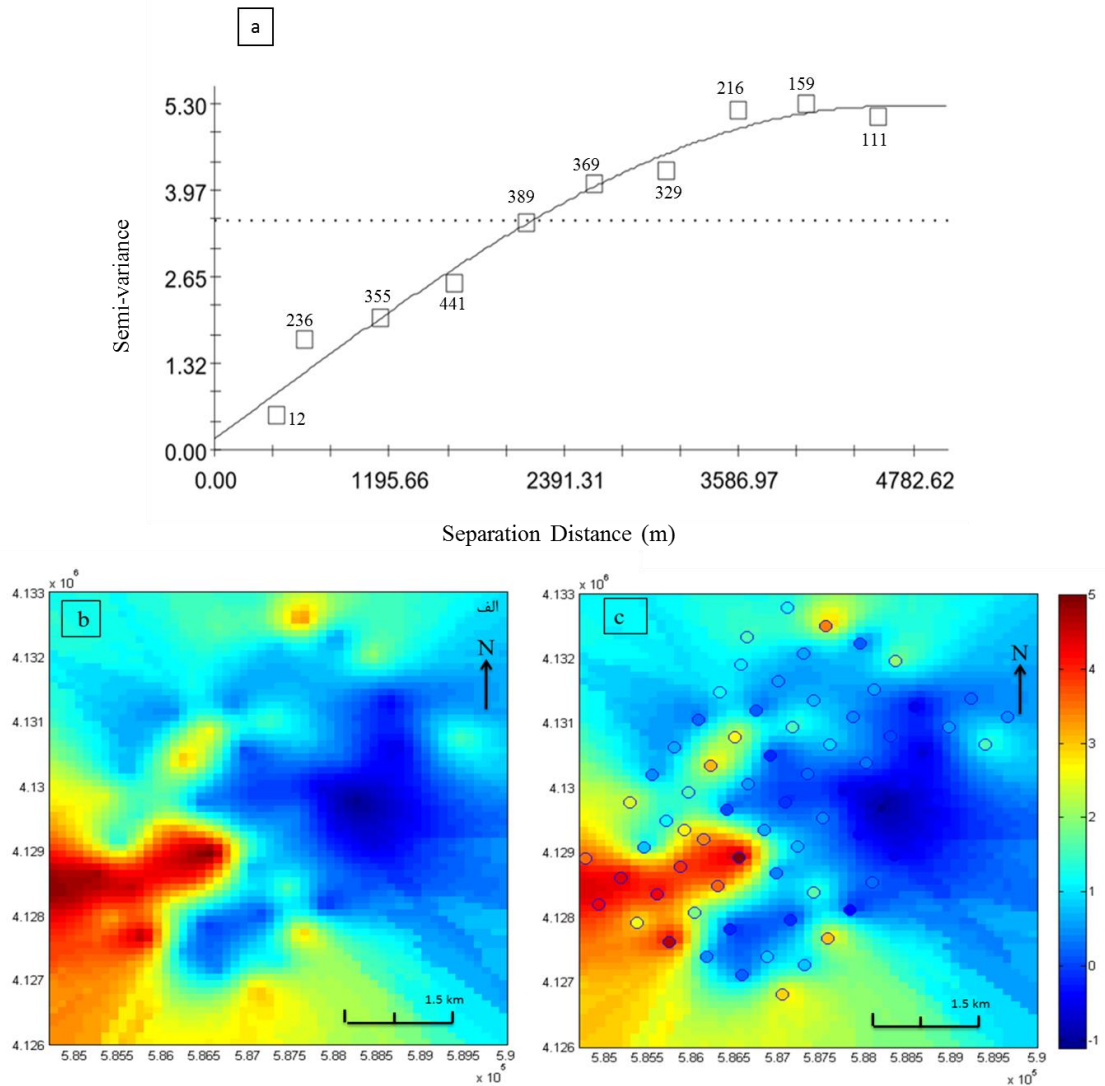

Fig. 1. a) Spatial semi-variance function for laboratory measured soil EC. Dots correspond to the calculated values; solid line is the corresponding fitted model. The number of pairs for variogram derivation is also provided next to the points; b: spatial prediction of top soil salinity using fitted semi-variance function; and c: predicted soil EC map along with the validation points. The level of colour reflects the top soil measured EC values in dS.m^{-1} in log scale. The similarity between points and background maps regarding colour indicates the prediction accuracy
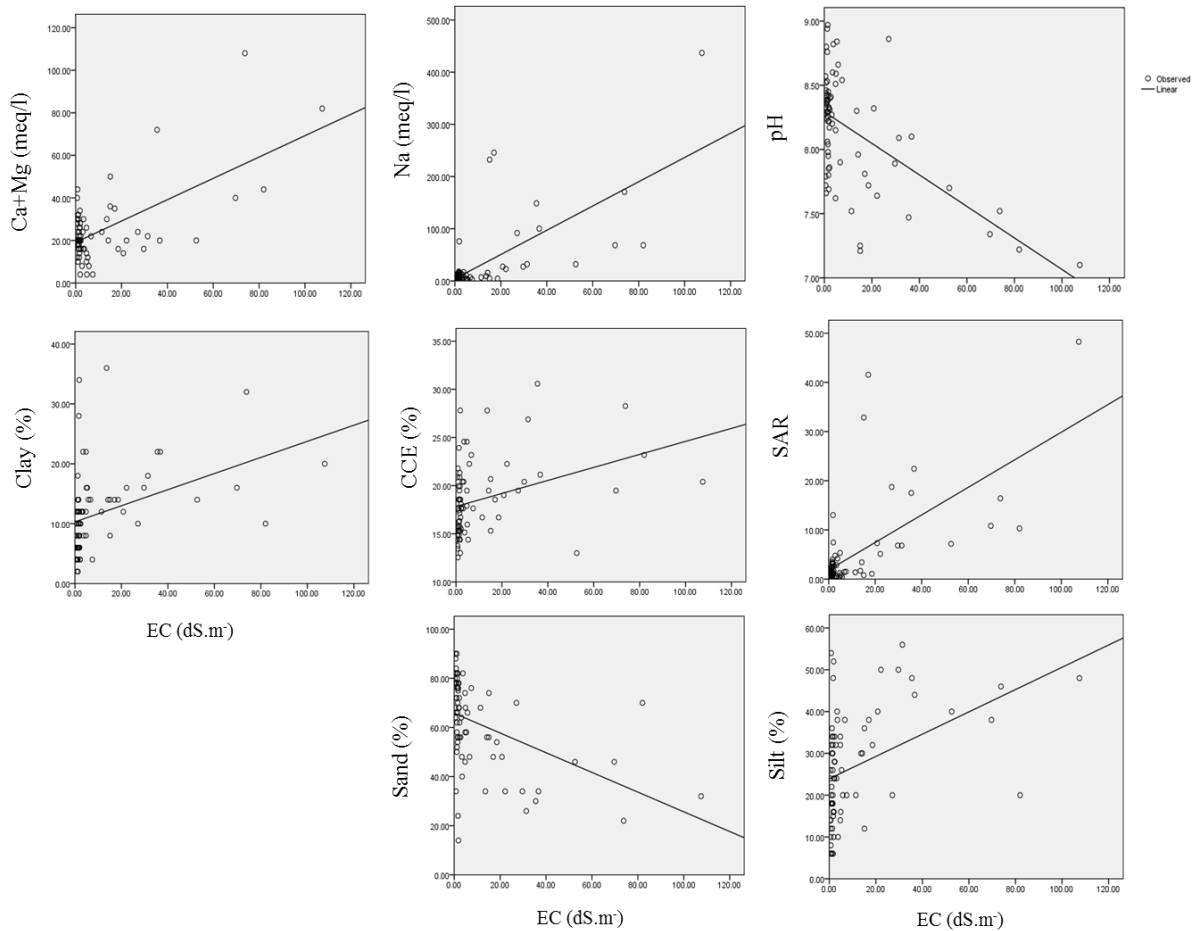
Fig. 3. Plots of the correlations between EC and the selected soil properties as covariates. The correlations were significant at 1% probability. OM: soil organic matter; SAR: sodium adsorption ratio; CCE: calcium carbonate equivalent

Table 2. Quantitative criteria to compare EC prediction with ordinary kriging and co-kriging methods with different covariates

Comparison criterion	OK	CK pH as covariate	CK silt as covariate
ME	-0.33	-0.69	-0.38
MSE	0.65	1.36	0.71
RMSE	0.80	1.17	0.84

Nazmul *et al.* (2015) also showed that despite the high level of correlation between soil pH and EC, soil pH, as covariate, was not able to improve the spatial prediction of soil EC. However, soil silt percentage resulted in a prediction close to that of OK with MSE and RMSE of 0.71 and 0.84 dS m^{-1}, respectively (Table 2). This could be attributed to its high correlation with soil EC, its strong spatial dependency, and the development of strong spatially-dependent cross-semivariogram (Figure 4a).

Different studies have demonstrated the superiority of co-kriging over ordinary kriging (Stein and Coresten, 1991; Zhang *et al.*, 1997; Wu *et al.*, 2009); as with co-kriging, additional covariates can be used to predict the soil variables. However, Alemi *et al.* (1988)

compared kriging and co-kriging methods in the spatial prediction of soil salinity with clay as covariate. They showed that the kriging method had a better performance in comparison with co-kriging. Juang *et al.* (2001) also reported that concerning soil EC data normality, kriging method also had a better performance. Meanwhile, other studies have reported that when auxiliary data are not highly correlated with target variable, co-kriging might not perform better than co-kriging (Martínez, 1996; Triantafilis *et al.*, 2001), highlighting the importance of selecting the appropriate auxiliary variables. It has further been shown that when the correlation between auxiliary data and target variable is higher than 0.5, co-kriging can perform better than kriging (Yates and Warrick, 1987). However, in the present study, although

silt had a lower correlation with EC than pH, its higher spatial dependency dominated its lower

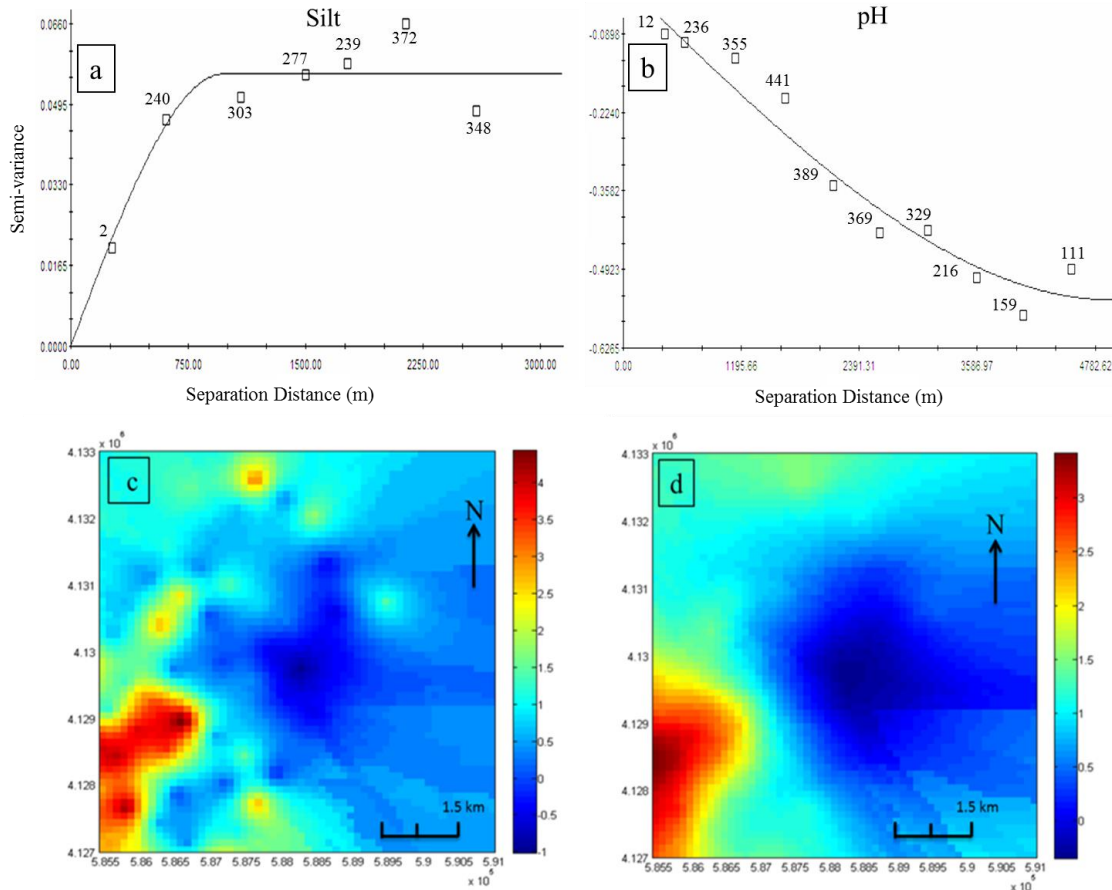correlation coefficient, resulting in better EC predictions.

Fig. 4. Cross-semivariance functions and spatial prediction of soil EC with covariates. a and b represent cross-semivariance functions with pH and silt, respectively. The number of pairs for variogram derivation is also provided next to the points. c and d present the spatial prediction maps of soil EC with pH and silt as covariates, respectively

3.3. Spatial prediction of soil EC using GLM-derived soft data

3.3.1. Development of general linear model (GLM) using auxiliary soil data

In order to develop the best soil EC prediction model using all available soil parameters, stepwise GLM was developed using soil parameters having significant correlations with soil EC at 1% probability level (Figure 3). The developed regression model was defined as follows:

$$EC\ (dS/m) = 95.71 - (12.517*pH) + (0.99*Na) + (0.239*(Ca+Mg)) + (0.147*SAR) - (0.264*CCE) + (0.132*clay) + (0.421*silt) - (0.289*sand) \tag{13}$$

Further stepwise including/excluding of soil parameters from GLM led to the equation

number 14 where CCE, sand, and clay parameters were also excluded. This equation had R and adjusted R^2 of 0.81 and 0.65, respectively, which is higher than that of equation number 13 (R and adjusted R^2 of 0.74 and 0.55, respectively) (Table 3). Although there were no statistically significant differences between these two equations, equation number 14 was selected for future use owing to the fewer soil properties involved in its development, hence the ease of use.

$$EC\ (dS/m) = 91.314 - (12.34*pH) + (0.103*Na) + (0.231*\ Ca+Mg) + (0.142*\ SAR) + (0.421*silt) \tag{14}$$

Several studies have reported low R^2 values between covariates and predicted soil EC (smaller than 0.5) (; Douaoui *et al.,* 2006; Bannari *et al.,* 2008; Bouaziz *et al.,* 2011).

However, Hammam and Mohamed (2018) obtained an R^2 of 0.58 between normalized vegetation index (NDVI) and soil EC. Rahmati and Hamzehpour (2017) further reported R^2 of 0.807 for their constructed regression relations for soil EC prediction. Nevertheless, one of these studies neither calculated the error nor attached it to the estimated EC values while using them in the spatial prediction of soil salinity. In the present work, as the next step of mapping soil EC, error was calculated and incorporated in the predictions; in this regard, equation 14 was robust enough for future use.

Table 3. Evaluation results for different constructed regression analysis between measured EC and different soil auxiliary data

R	Equation	R^2_{adj}	R^2
0.74	13	0.55	0.55
0.81	14	0.65	0.65

3.3.2. Semi-variance function development using GLM derived soft data

Figure 5 presents the semi-variogram best fitted to soil EC data derived from equation 14. The fitted semi-variogram model had two distinctive parts: a nugget effect equal to 0.338 $(dSm^{-1})^2$, a spherical part with a range of 4477 m, and a sill of 1.97 $(dS\ m^{-1})^2$. The semi-variance function developed using the exact measurement of soil EC (Figure 2a) was compared with the one developed from EC-predicted values (Figure 5). Based on the comparison, nugget/sill ratio (NE), which indicates the spatial dependency of a semi-variogram, was stronger in the former than in the latter semi-variance function (3% versus 17%). However, NE <25% is considered as a strong spatial dependence and a reliable spatial predictor of the target variable (Duffera *et al.,* 2007). Therefore, both developed semi-variance functions were robust enough to be used in predictions.
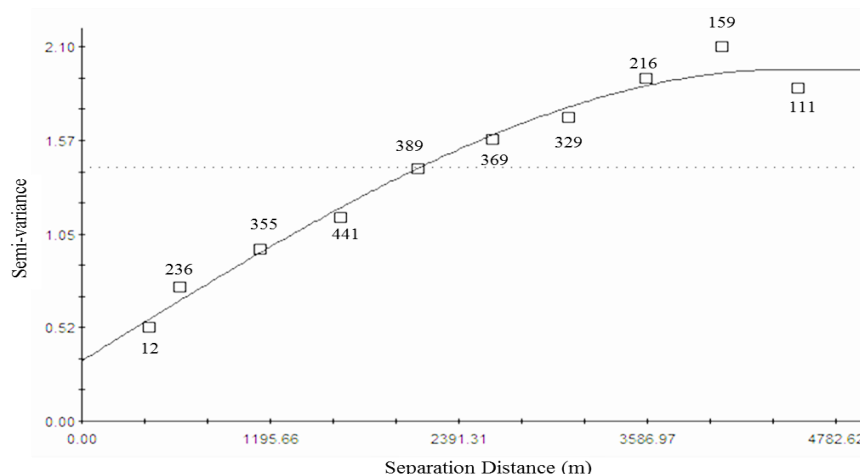

Fig. 5. Spatial semi-variance function for soil EC soft data derived from GLM. Dots correspond to the calculated values; the solid line is the corresponding fitted model. The number of variogram derivation pairs is also provided next to the points

3.3.3. Definition of probabilistic type soft data and error measurement

To go forward with soft data derived from equation 14 and its use in the spatial prediction of soil EC with KME, first, there should be significant difference between direct EC measurements (hard data) and those data gained indirectly (soft data). Next, soft data should be calibrated with hard data through calibration equations. Results showed a significant difference between hard and soft data (p-value <0.05). The calibration plot between hard and soft data and the fitted prediction line are shown in Figure 6a. As seen, R^2 of 0.65 for calibration line was not strong enough; thus, it was expected that the resulting calibrated data would involve unneglectable error. Therefore, despite the calibration of soft data, their use in the spatial prediction of soil EC without accounting for error, would involve large sources of uncertainty, leading to the reduced accuracy of the predicted maps.

Accordingly, as the next step in defining the probabilistic soft data, the histogram of the differences between calibrated data and actual measurements was plotted (residuals), Figure 6b. The normal distribution of residuals (the

Kolmogorov-Smirnov correlation coefficient was significantly higher than 0.05) showed that the residuals of EC predictions with GLM and

auxiliary data had no trend; therefore, the variance of the residuals was considered as error in further calculations.
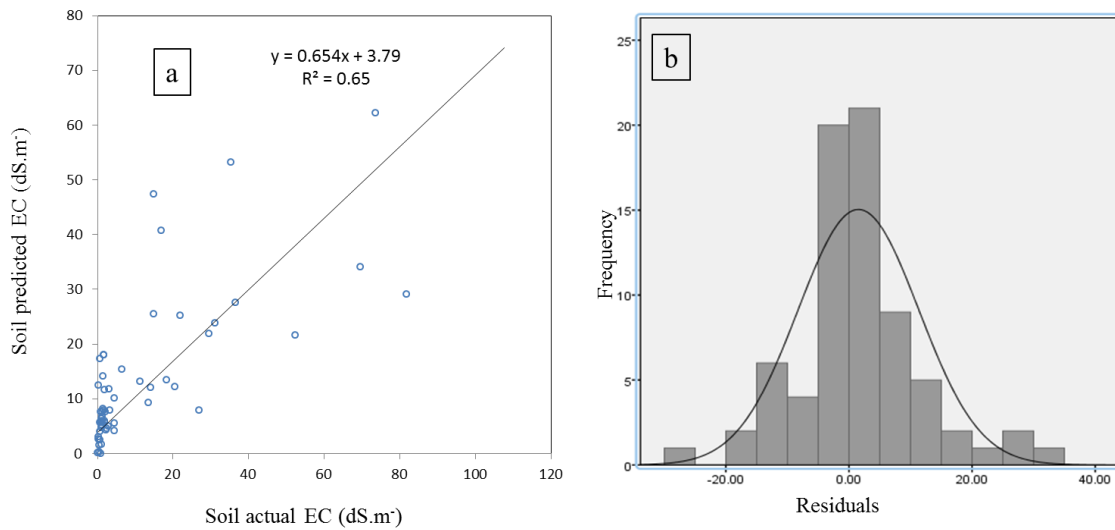

Fig. 6. (a) The correlation between EC measurements (hard data) and EC values form GLM (soft data). Dots correspond to the measured values while the solid line is the corresponding fitted model; (b) histogram of the residuals between hard and soft data and fitted probability density function (pdf)

3.3.4. Spatial prediction of soil EC with soft data

In the following step, the spatial prediction of soil EC was done on a scale of 100 m using semi-variance function presented in Figure 5. Predictions were done using the soft data defined in the previous section, once by taking the error into account and once without error involvement. Equations 7 and 2 were used in predictions with and without error involvement, respectively. These results were then compared to EC map which was generated using hard data (Figure 2). Results of spatial prediction maps are presented in Figure 7.

As observed in Figure 7, there were a couple of distinctive differences between soil EC predictions using only soft data (Figure 7a) and that of soft data with error involvement (Figure 7b). However, the error calculated from the use of soft data and KME method enhanced the predictions (Figure 7b versus 7c). The validation results revealed that KME method had an acceptable prediction accuracy with MSE and RMSE of 0.74 and 0.86 dS m^{-1}, respectively, which is close to that of silt as covariate. Among the studied methods, however, EC predictions with only soft data resulted in the weakest results with MSE and RMSE of 2.47 and 1.6 dS m^{-1}, respectively.

Table 4. Quantitative criteria to compare three different methods used for soil EC prediction.

Comparison criterion	Soft data	Soft data with error	Hard data
ME	-1.49	-0.45	-0.33
MSE	2.47	0.74	0.65
RMSE	1.57	0.86	0.80

These results showed that although GLM and auxiliary soil data resulted in untrusworthy predictions of soil EC values (Figure 7a), the error calculation and their use in predictions with KME method, could highly improve the spatial prediction of soil EC.

Some researchers have reported the passive role of error measurment and its envolvement in predictions (Hamzehpour and Bogaert, 2017; Hamzehpour *et al*., 2013; Douaik *et al*., 2005), which is in line

with the present study. However, in almost all these studies, the soft data used in predictions was a simplified method for measuring target soil variables such as EC. In the present research, the soft data used for predictions was estimated based on other soil properities, hence involved with more error and higher uncertainty compared with previous works. Therefore, although neglecting error during prediction process led to the weak predictions, error measurment and its involvement in the

framework of KME method played a more significant role in improving the spatial prediciton of soil EC.

Therefore, through proposed method, remote sensing data are also of high capability of being used as soft data in spatial prediction of soil properties.

Several researches (e.g., Bilgili *et al.,* 2011; Gorji *et al.,* 2015; Scudiero *et al.,* 2016; Rahmati and Hamzehpour, 2017) have developed polynomial fuctions between soil EC and remote sensing data which are capable to be used as saft data.
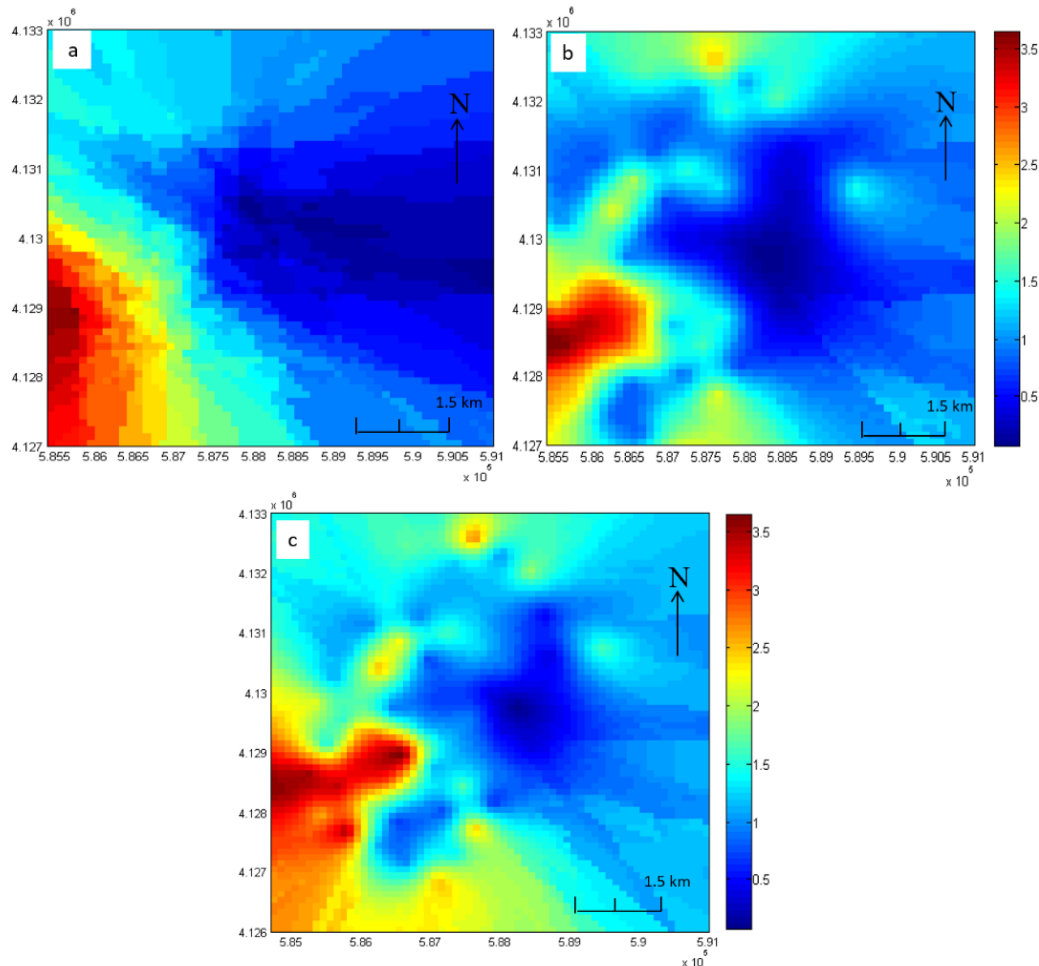

Fig. 7. Soil EC prediction maps with: a) using soft data without error; b) soft data with error; and c) hard data

4. Conclusion

There has been an increase in the degradation of agricultural lands and soil quality in Iran due to issues, such as drought, secondary salinization caused by saline irrigation water, and miss-cultivation practices. Therefore, it is necessary to constantly monitor soil properties such as soil salinity in order to keep the agricultural production yields at their optimum levels with the least water usage and minimum damage to the environment. However, lack of information concerning soil properties in target regions, have made

soil monitoring expensive and time-consuming. In the present research, an attempt was made to extract specific soil property such as soil electrical conductivity from other soil properties using general linear model (GLM). However, with such model derived data, there is still need for their calibration with hard data and the consideration of error involved in the data of such ilk. In this regard, through calibration equations, the error was calculated, and using kriging with measurement errors method (KME), the spatial prediction of soil electrical conductivity was successfully performed.

This work is a step forward in the use of soil auxiliary data where there is little information available on the parameter under question. However, as of today, the un-denaible error involved with such model based data, has restricted their use as auxillary data source. KME method seems to be a stong tool for using auxiliary soil data in spatial predictions of soil variables.

References

Abdullah, A.Y.M., R.K. Biswas, A.I. Chowdhury, S.M. Billah, 2019. Modeling soil salinity using direct and indirect measurement techniques: A comparative analysis. Environmental Development, 29; 67-80.

Aini, I.N., M.H. Ezrin, W. Aimrun, 2014. Relationship between soil apparent electrical conductivity and pH value of Jawa series in oil palm plantation. Agriculture and Agricultural Science Procedia, 2; 199-206.

Acosta, J.A., A. Faz, B. Jansen, K. Kalbitz, S. Martínez-Martínez, 2011. Assessment of salinity status in intensively cultivated soils under semiarid climate, Murcia, SE Spain. Journal of Arid Environments, 75; 1056-1066.

Alemi, M.H., A.S. Azari, D.R. Nielsen, 1988. Kriging and univariate modeling of a spatially correlated data. Soil Technology, 1; 133-147.

Bannari, A., A. Guedon, A. El-Harti, F. Cherkaoui, A. El-Ghmari, 2008. Characterization of Slightly and Moderately Saline and Sodic Soils in Irrigated Agricultural Land using Simulated Data of Advanced Land Imaging (EO-1) Sensor. Communications in Soil Science and Plant Analysis, 39; 2795–2811.

Bernoux, M., C., Cerri, D., Arrouays, C., Jolivet, B., Volkoff, 1998. Bulk densities of Brazilian Amazon soils related to other soil properties. Soil Science Society of America Journal, 62; 743-749.

Bilgili, A.V., M.A. Cullu, H. van Es, A. Aydemir, S. Aydemir, 2011. The Use of Hyperspectral Visible and near Infrared Reflectance Spectroscopy for the Characterization of Salt-Affected Soils in the Harran Plain, Turkey. Arid Land Research and Management, 25; 19–37.

Bouaziz, M., J. Matschullat, R. Gloaguen, 2011. Improved Remote Sensing Detection of Soil Salinity from a Semi-Arid Climate in Northeast Brazil. Comptes Rendus Geoscience, 343; 795–803.

Bogaert, P., D. D'Or, 2002. Estimating soil properties from thematic soil maps. Soil Science Society of America Journal, 66; 1492-1500.

Brevik, E.C., T.E. Fenton, A. Lazari, 2006. Soil electrical conductivity as a function of soil water content and implications for soil mapping. Precision Agriculture, 7; 393-404.

Brus, D.J., P. Bogaert, G.B.M. Heuvelink, 2008. Bayesian maximum entropy prediction of soil categories using a traditional soil map as soft information. European Journal of Soil Science, 59; 166-177.

Christakos, G., M.L. Serre, 2000. BME analysis of spatiotemporal particulate matter distributions in North Carolina. Atmospheric Environment, 34; 3393-3406.

Corwin, D.L., S.M. Lesch, 2005. Apparent soil electrical conductivity measurements in agriculture. Computers and electronics in agriculture, 46; 11-43.

Davis, B.M., 1987. Uses and abuses of cross-validation in geostatistics. Mathematical geology, 19; 241-248.

De Clercq, W.P., M. Van Meirvenne, M.V. Fey, 2009. Prediction of the soil-depth salinity-trend in a vineyard after sustained irrigation with saline water. Agricultural Water Management, 96; 395-404.

Douaoui, A.E.K., H. Nicolas, C. Walter, 2006. Detecting Salinity Hazards within a Semiarid Context by Means of Combining Soil and Remote-Sensing Data. Geoderma, 134; 217–230.

Douaik, A., M. Van Meirvenne, T. Tóth, 2005. Soil salinity mapping using spatio-temporal kriging and Bayesian maximum entropy with interval soft data. Geoderma, 128; 234-248.

D'Or, D., P. Bogaert, G. Christakos, 2001. Application of the BME approach to soil texture mapping. Stochastic Environmental Research and Risk Assessment, 15; 87-100.

Duffera, M., J.G. White, R. Weisz, 2007. Spatial variability of Southeastern US Coastal Plain soil physical properties: Implications for site-specific management. Geoderma, 137; 327-339.

Fazekas, I., A.G. Kukush, 2005. Kriging and measurement errors. Discussions Mathematicae. Journal of Probability and Statistic, 25; 139–159.

Gee, G.W., D. Or, 2002. 2.4 Particle-size analysis. Methods of soil analysis: Part, 4. 5; 255-293.

Giordano, R., S. Liersch, M. Vurro, D. Hirsch, 2010. Integrating local and technical knowledge to support soil salinity monitoring in the Amudarya river basin. Journal of Environmental Management, 91; 1718-1729.

Gorji, T., A. Tanik, E. Sertel, 2015. Soil Salinity Prediction, Monitoring and Mapping Using Modern Technologies. Procedia Earth and Planetary Science, 15; 507–512.

Hammam, A.A., E.S. Mohamed, 2018. Mapping soil salinity in the East Nile Delta using several methodological approaches of salinity assessment. The Egyptian Journal of Remote Sensing and Space Science, 23(2); 125-131.

Hamzehpour, N., P. Bogaert, 2017. Improved spatiotemporal monitoring of soil salinity using filtered kriging with measurement errors: An application to the West Urmia Lake, Iran. Geoderma, 295; 22-33.

Hamzehpour, N., M.K. Eghbal, P. Bogaert, N. Toomanian, R.S. Sokouti, 2013. Spatial prediction of soil salinity using kriging with measurement errors and probabilistic soft data. Arid land research and management, 27; 128-139.

Inman, D.J., R.S., Freeland, J.T., Ammons, R.E., Yoder, 2002. Soil investigations using electromagnetic induction and ground-penetrating radar in southwest Tennessee. Soil Science Society of America Journal, 66; 206-211.

Juan, P., J. Mateu, M.M. Jordan, J. Mataix-Solera, I. Meléndez-Pastor, J. Navarro-Pedreño, 2011. Geostatistical methods to identify and map spatial variations of soil salinity. Journal of Geochemical Exploration, 108; 62-72.

Juang, K.W., D.Y. Lee, T.R. Ellsworth, 2001. Using rank-order geostatistics for spatial interpolation of

highly skewed data in a heavy-metal contaminated site. Journal of Environmental Quality, 30; 894-903.

Jung, W.K., N.R., Kitchen, K.A., Sudduth, R.J., Kremer, P.P., Motavalli, 2005. Relationship of apparent soil electrical conductivity to claypan soil properties. Soil Science Society of America Journal, 69; 883-892.

Keskin, H., S. Grunwald, 2018. Regression kriging as a workhorse in the digital soil mapper's toolbox. Geoderma, 326; 22-41.

Krivoruchko, K., A. Gribov, 2019. Evaluation of empirical Bayesian Kriging. Spatial Statistics. 32; 100368.

Li, J., A.D. Heap, 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. Ecological Informatics, 6; 228-241.

Malins, D., G. Metternicht, 2006. Assessing the spatial extent of dryland salinity through fuzzy modeling. Ecological Modelling, 193; 387-411.

Martínez-Cob, A., 1996. Multivariate geostatistical analysis of evapotranspiration and precipitation in mountainous terrain. Journal of Hydrology, 174; 19-35.

MathWorks Inc., 1999. MatLab, the language of technical computing, using MATLAB version 5. the Mathwork Inc. http://www.mathworks.com, Natick.

Mohd-Aizat, A., M.K. Mohamad-Roslan, W.N.A. Sulaiman, D.S., Karam, 2014. The relationship between soil pH and selected soil properties in 48 years logged-over forest. International journal of environmental sciences, 4(6); 1129-1140.

Nelson, D., L.E. Sommers, 1996. Total carbon, organic carbon, and organic matter, In: Methods of soil analysis, Part 3, Chemical methods, p. 539-579.Page, A.L., R.H. Miller, D.R. Jeeney, 1982. Methods of soil analysis, part 2, Chemical and mineralogical properties. American Society of Agronomy.

Rahmati, M., N. Hamzehpour, 2017. Quantitative remote sensing of soil electrical conductivity using ETM+ and ground measured data, International Journal of Remote Sensing, 38; 123-140.

Rhoades, J.D., 1996. Salinity: electrical conductivity and total dissolved solids. In: Sparks, DL, editor. Methods of Soil Analysis, Part 3. Chemical Methods. SSSA Book Ser. 5. SSSA and ASA, Madison, WI, p. 417-435.

Savelyeva, E., S. Utkin, S. Kazakov, V. Demyanov, 2010. Modeling spatial uncertainty for locally uncertain data. In: geoENV VII–Geostatistics for Environmental Applications. Springer, Dordrecht. p. 295-306.

Scudiero, E., T.H. Skaggs, D.L. Corwin, 2016. Comparative Regional-Scale Soil Salinity Assessment with Near-Ground Apparent Electrical Conductivity

and Remote Sensing Canopy Reflectance. Ecological Indicators, 70; 276–284.

Serre, M.L. G. Christakos, 1999. Modern geostatistics: computational BME analysis in the light of uncertain physical knowledge–the Equus Beds study. Stochastic Environmental Research and Risk Assessment, 13; 1-26.

Shi, Y., Z. Lu, L. Xu, S. Chen, 2019. An adaptive multiple-Kriging-surrogate method for time-dependent reliability analysis. Applied Mathematical Modelling, 70; 545-571.

Soil Survey Staff, 2014. Keys to Soil Taxonomy, 12th ed. USDA. NRCS

Stein, A., L.C.A. Corsten, 1991. Universal kriging and cokriging as a regression procedure. Biometrics, 575-587.

Triantafilis, J., I.O.A. Odeh, A.L. Jarman, M.G. Short, E. Kokkoris, 2004. Estimating and mapping deep drainage risk at the district level in the lower Gwydir and Macquarie valleys, Australia. Australian Journal of Experimental Agriculture, 44; 893-912.

Triantafilis, J., I.O.A. Odeh, A.B. McBratney, 2001. Five geostatistical models to predict soil salinity from electromagnetic induction data across irrigated cotton. Soil Science Society of America Journal, 65, 869-878.

Wu, C., J. Wu, Y. Luo, L. Zhang, S.D. DeGloria, 2009. Spatial prediction of soil organic matter content using cokriging with remotely sensed data. Soil Science Society of America Journal, 73; 1202-1208.

Yang, S.H., F. Liu, X.D. Song, Y.Y. Lu, D.C. Li, Y.G. Zhao, G.L. Zhang, 2019. Mapping topsoil electrical conductivity by a mixed geographically weighted regression kriging: A case study in the Heihe River Basin, northwest China. Ecological Indicators, 102; 252-264.

Yates, S.R., A.W. Warrick, 1987. Estimating Soil Water Content Using Cokriging 1. Soil Science Society of America Journal, 51; 23-30.

Zare, S., S.R.F. Shamsi, S.A. Abtahi, 2019. Weakly-coupled geo-statistical mapping of soil salinity to Stepwise Multiple Linear Regression of MODIS spectral image products. Journal of African Earth Sciences, 152; 101-114.

Zhang, R., P. Shouse, S. Yates, 1997. Use of pseudo-crossvariograms and cokriging to improve estimates of soil solute concentrations. Soil Science Society of America Journal, 61; 1342-1347.

Zovko, M., D. Romić, C. Colombo, E. Di Iorio, M. Romić, G. Buttafuoco, A. Castrignanò, 2018. A geostatistical Vis-NIR spectroscopy index to assess the incipient soil salinization in the Neretva River valley, Croatia. Geoderma, 332; 60-72.