Mapping spatial variability of soil salinity in a coastal area located in an arid environment using geostatistical and correlation methods based on the satellite data

M. Samiee\textsuperscript{a}, R. Ghazavi\textsuperscript{a*}, M. Pakparvar\textsuperscript{b}, A.A. Vali\textsuperscript{a}

\textsuperscript{a} Department of Watershed Management, Faculty of Natural Resources and Earth Sciences, University of Kashan, Kashan, Iran
\textsuperscript{b} Fars Agricultural and Natural Resources Research and Education Center

Received: 29 October 2017; Received in revised form: 6 June 2018; Accepted: 18 June 2018

Abstract

Saline lakes can increase the soil and water salinity of the coastal areas. The main aim of this study is to distinguish the characteristics of the spectral reflectance of saline soil, analyze the statistical relationship between soil EC and characteristics of the spectral reflectance of saline soil, and to map soil salinity east of the Maharloo Lake. The correlation between field measurements of electrical conductivity and remote sensing spectral indices was evaluated using multiple regression analysis. In this study, Kriging, CoKriging, and multiple regressions were applied for soil salinity mapping and classification using 100 soil samples. After radiometric, geometric, and atmospheric corrections of Landsat OLI images, the statistical correlation between the electrical conductivity of field measurements and spectral reflectance was investigated. According to obtained results, the modified salinity index (MSI) with the highest correlation ($R^2 = 0.78$) was used as an auxiliary variable for the coKriging method. Kriging with a spherical model was selected for soil salinity mapping (RMSE = 50.5 and $R^2 = 0.18$). The RMSE and $R^2$ values for CoKriging were (43.2 and 0.42), respectively. Because of their acceptable $R^2 (=0.65)$ and low standard deviation (33.8) for salinity analysis, MSI and difference vegetation index (DVI) were used to estimate and zonate soil salinity in the study area. The results showed that soil salinity could be estimated via spectral indices with acceptable accuracy, $R^2$ and RMSE. Overall, this method leads to a decrease in the costs involved in the soil mapping of saline soil areas.

Keywords: Soil salinity; Maharlo lake; Geostatistical methods; Regression; Arid environment

1. Introduction

Soil salinity is a dynamic process with intense effects on soil, and possessing hydrological, climatic, geochemical, agricultural, social, and economic aspects (Juan \textit{et al.}, 2010; Allbed \& Kumar, 2013). In arid or semi-arid regions, poor natural drainage may cause a serious salt accumulation hazard for soils. The accumulation of ions such as Na\textsuperscript{+}, K\textsuperscript{+}, Mg\textsuperscript{2+}, and Ca\textsuperscript{2+} in soil affects the chemical and physical properties of the soil (Juan \textit{et al.}, 2010). Mapping spatial variations of soil salinity and studying the effectual factors of salinization has an important and key role in the management of saline lands. Due to the vastness of the saline lands, their studies are time-consuming and expensive (Shresta, 2006, Weng \textit{et al}, 2008). The study of soil salinity using remote sensing techniques was proposed to reduce fieldwork and data generation costs (Allbed and Kumar, 2013). For detecting and mapping salt-affected soils, several spectral indices such as the bare soil index (BI), Normalized Difference Salinity Index NDSI, and Salinity Index (SI) have been proposed (Mehrjerdi \textit{et al}., 2008, Khan \textit{et al}., 2001; Taghizadeh Nematolahi \textit{et al}., 2012). According to Judkins and Myint (2012), NDVI, PCA 1, Tasseled Cap 3, and Tasseled Cap 5 presented the most promising correlations with soil salinity in the Mexicali Valley, Mexico. In
order to detect saline soils in the Najmabad region of Savojbolagh in Iran, satellite IRS-P6’s LISS-III sensor data was used (Shirazi et al. 2012). The results showed that the INT2 and PVI indices have the best indicators for the detection of salty soils. Pakparvar et al. (2012) investigated soil salinity changes in the Darab plain in Iran in 1990 and 2002. The results showed that saline soil area increased by 42% in 2002 compared to 1990.

Jafari et al. (2007) studied salinity variations (EC and SAR) in different layers of soil in the agricultural lands of Kermanshah Province, Iran. The results show that both EC and SAR were increased with an increase in depth in the steep slopes of the rainfed lands of Paveh. Fernandez-Buces et al. (2006) mapped soil salinity by a combined spectral response index for bare soil and vegetation, called COSRI, in the former lake Texcoco, Mexico. Soil salinity in Cuddalore was monitored and evaluated by way of spectral bands using the NDVI, SI, and SAVI indices (Narmada et al., 2015). The SAVI was presented as the best index for separating plants under severe and moderate conditions of soil salinity. The soil salinity in Turkey’s Seyhan plate was mapped using the multi-temporal Landsat 7 and several salinity indices such as NDSI, BI, SI, RVI, SAVI, and EVI (Azabdaftari and Sunar 2016).

Reviewing the previous studies shows that the best indices and original bands for soil salinity mapping are different based on the imagery types and the extent of the salinity. Therefore, an assessment of the best band combination is necessary to generate a reliable salinity map. Several researchers have investigated the application of image processing techniques for soil salinity classification, but there are a limited number of published studies on the combination of image processing and geostatistics, which should increase the accuracy of the classification. In this study, a combination of image processing and geostatistics were applied to: (1) investigate the spectral reflectance behavior of saline soils in Maharloo Lake, (2) analyze the statistical correlation between the electrical conductivity of field measurements and remotely sensed spectral indices, and (3) select the best method for generating the soil salinity map of the study area using image processing and geostatistics.

2. Materials and Methods

2.1. Study area

The study area is located in the eastern part of Maharloo Lake with an area of 1740 ha. The area is 10 km southeast of Shiraz city within latitudes 3240000 to 3250000 and longitudes 692000 to 685500 UTM zone 39 (Fig. 1). The Maharloo Lake has a high annual variability of water level ranging from 5000 to 12500 ha. From the winter to mid-spring, the water level rises gradually to the maximum level and then drops through evaporation until late summer, whereupon it is normally dried out. Consequently, water salinity and its density gradually increase in the summer. The main soil texture is clay loam and the salinity level ranges from 2.3 to 232 dS/m.

2.2. Methodology

In this study, images acquired in September 2016 by Landsat OLI were used to delineate the soil salinity in the eastern area of Maharloo Lake, an area with a bare land surface or with sparse vegetation around waterways. During the summer period, when soil moisture is very low, the light-colored appearance of saline soils makes it more noticeable and salt-affected soils are commonly found there (Douaoui et al., 2006).

The geometric correction was made based on maps (scale: 1/25000) published by the Iranian national cartographic center. Dark objects and quick methods were used for radiometric and atmospheric corrections, respectively. Sampling sites were selected using the Fishnet method in GIS at intervals of 400 m. Next, 100 soil samples were systematically collected from a depth of 1-5 cm (Fig. 1). The satellite sensor recorded a reflection of a short-wave spectrum related to the surface layers; this recording, however, may not provide details of the underlying layers. Therefore, sampling should be limited to this section (1-5 cm in this study).

All soil samples were air-dried and passed through a 2 mm sieve prior to analysis. Electrical conductivity was measured in saturation pastes after a 4h equilibration using the conductivity meter.

A variety of original bands (2-7), principal components (the first three components), and 25 spectral indices (Table 1) were used to examine the relationship between the electrical conductivity of field measurements and remotely sensed spectral indices. These indices
have been employed for salinity detection by several researchers (Azabdaftari and Sunar 2016; Narmada et al. 2015; Judkins and Myint 2012). The SAVI presented the best index for separating plants under severe and moderate conditions of soil salinity (Azabdaftari and Sunar 2016). The most commonly used technique to identify salinity is the calculation of different indicators from infrared and visible bands (Matinfar and Zandieh, 2016). Soil salinity could be estimated by using spectral indices as a good auxiliary variable in spatial estimation and mapping salinity in irrigated land. For this purpose, the indices based on visible spectral bands are more sensitive to the soil salinity and the SI index has a better correlation with soil salinity in our region than others (Lhissou et al., 2014).

Salt-affected soils have high spectral values in red and near-infrared bands (131.48 and 182.12). This observation is extremely useful as it helps distinguish salt-affected, waterlogged, and normal soils. Severely salt-affected soils had the highest reflectance in all three bands (green, red, and infrared), followed by moderately salt-affected soil. In comparison, normal soils have a low reflectance. Kumar Koshal (2012) used the red band to distinguish the image characteristics of normal soils from salt-affected lands.

Most salt-affected soils have a higher visible and NIR reflectance (Rao et al., 1995). These indices are always located in visible and infrared bands. The field points were crossed to individual bands and indices and the corresponding values were extracted. The indicators that showed the highest correlation between reflectance data and soil EC were selected via multiple regression analysis using stepwise regression. A high coefficient of determination ($R^2$) and a low root mean square error (RMSE) were used to select the best equation. The salinity index that had a higher coefficient of determination and the lower standard deviation was used to generate an enhanced image for soil salinity. The resulting salinity map was further classified into diverse numbers and intervals of classes in order to obtain the best-classified salinity map.

2.3. Evaluation of classification

The best regression model exhibiting the input salinity index with highest accuracy was used to generate the salinity map. This map was then classified based on a variety of classes, numbers, and intervals using a trial-and-error approach to examine the highest possible number of classes possessing the highest accuracy. Each salinity and their interval class were considered as a classification system. A set of 66 randomly chosen points was
considered to verify the accuracy of these systems (Table 1).

Classification accuracy was tested using Kappa coefficient. The Kappa coefficient computes the overall agreement of a matrix. In contrast to the overall accuracy, i.e., the ratio of the sum of diagonal values to a total number of cell counts in the matrix, the Kappa coefficient also takes non-diagonal elements into account (Banko, 1998).

Overall accuracy between remotely-sensed classification and the observed data can be calculated via Eq. 1 (Congalton and Green, 2009).

\[
O.A. = \frac{\sum_{i=1}^{k} n_{ij}}{N}
\]

where \( n_{ij} \) refer to the number of samples classified into groups \( i \) (\( i = 1, 2, ..., k \)) in the map and groups \( j \) (\( j = 1, 2, ..., k \)) in the reference dataset.

The kappa coefficient is taken into account to assess the quality of each classification (Wu et al., 2008).

\[
Kappa = \frac{P_o - P_e}{1 - P_e}
\]

where \( P_o \) is the accuracy of the observed agreement and \( P_e \) is an estimate of chance agreement.

### Table 1. Spectral indices used to assess the relationship between electrical conductivity and spectral indices

<table>
<thead>
<tr>
<th>Salinity indices</th>
<th>Equation</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salinity Index(1)</td>
<td>( SI_1 = (B/R) )</td>
<td>Abbas(2007)</td>
</tr>
<tr>
<td>Salinity Index(2)</td>
<td>( SI_2 = (B - R)/(B + R) )</td>
<td>Abbas(2007)</td>
</tr>
<tr>
<td>Salinity Index(3)</td>
<td>( SI_3 = (G* R)/B )</td>
<td>Abbas(2007)</td>
</tr>
<tr>
<td>Salinity Index(4)</td>
<td>( SI_4 = \sqrt{B* R} )</td>
<td>Khan et al.(2001)</td>
</tr>
<tr>
<td>Salinity Index(5)</td>
<td>( SI_5 = (B*R)/G )</td>
<td>Abbas(2007)</td>
</tr>
<tr>
<td>Salinity Index(6)</td>
<td>( SI_6 = (R*NIR)/G )</td>
<td>Abbas(2007)</td>
</tr>
<tr>
<td>Salinity Index(7)</td>
<td>( SI_7 = \sqrt{G*R} )</td>
<td>Douaoui et al.(2006)</td>
</tr>
<tr>
<td>Salinity Index(8)</td>
<td>( SI_8 = \sqrt{G^2 + R^2} - NIR^2 )</td>
<td>Khaier(2003)</td>
</tr>
<tr>
<td>Salinity Index(9)</td>
<td>( SI_9 = \sqrt{G^2 + R^2} )</td>
<td>Douaoui et al.(2006)</td>
</tr>
<tr>
<td>Salinity Index(10)</td>
<td>( SI_{10} = (NIR - SWIR)/(NIR + SWIR) )</td>
<td>Khaier(2003)</td>
</tr>
<tr>
<td>Simple Ratio</td>
<td>( SR = NIR/R )</td>
<td>Birth and McVey(1968)</td>
</tr>
<tr>
<td>Normalized Difference Salinity Index</td>
<td>( NDSI = (R - NIR)/(R + NIR) )</td>
<td>Khan et al.(2001)</td>
</tr>
<tr>
<td>Salinity Index(A)</td>
<td>( SI_A = (R/B)*100 )</td>
<td>Khan et al.(2005)</td>
</tr>
<tr>
<td>Salinity Index(T)</td>
<td>( SI_T = (R/NIR)*100 )</td>
<td>Khan et al.(2005)</td>
</tr>
<tr>
<td>Brightness Index</td>
<td>( BI = \sqrt{R^2 + NIR^2} )</td>
<td>Dehni &amp; Lounis(2012)</td>
</tr>
<tr>
<td>Infrared Percentage Vegetation Index</td>
<td>( IPVI = NIR/(NIR + R) )</td>
<td>Crippen(1990)</td>
</tr>
<tr>
<td>Difference Vegetation Index</td>
<td>( DVI = NIR - R )</td>
<td>Tucker(1979)</td>
</tr>
<tr>
<td>Normalized Difference Vegetation Index</td>
<td>( NDVI = (NIR - R)/(NIR + R) )</td>
<td>Rouse et al.(1973)</td>
</tr>
<tr>
<td>Combined Spectral Response Index</td>
<td>( Cosri = (B + G)/(R + NIR)*NDVI )</td>
<td>Fernandez-Buces et al.(2006)</td>
</tr>
<tr>
<td>Soil adjusted Vegetation Index</td>
<td>( SAVI = 1.5*(NIR - R)/(NIR + R + 0.5) )</td>
<td>Huiete(1998)</td>
</tr>
<tr>
<td>Enhanced Vegetation Index</td>
<td>( EVI = 2.5*(NIR - R)/(NIR + 6* R - 7.5* B +1) )</td>
<td>Huiete et al.(2002)</td>
</tr>
<tr>
<td>Generalized Difference Vegetation Index</td>
<td>( GDVI = NIR^2 - R^2 )</td>
<td>WU(2012)</td>
</tr>
<tr>
<td>Modified Salinity Index</td>
<td>( MSI = SWIR/NIR )</td>
<td>Tajgardan et al.(2009)</td>
</tr>
<tr>
<td>Vegetation Soil Salinity Index</td>
<td>( VSSI = 2<em>G - 5</em>(R + NIR) )</td>
<td>Dehni &amp; Lounis(2012)</td>
</tr>
</tbody>
</table>

B, G, R and NIR correspond to the blue (band 2), green (band 3), red (band 4) and near infrared (band 5) bands in Landsat imagery, respectively

### 2.4. Soil salinity mapping

Kriging and CoKriging methods were investigated to map the spatial variation of soil salinity. Using geostatistical methods provides us with a spatial correlation between neighboring observations to forecast values at non-sampled locations (Goovaerts, 2000). Basic components of geostatistics include Semivariogram analysis (characterization of spatial correlation) and Kriging (optimal interpolation technique).
3. Results and Discussion

3.1. Geostatistics

Ordinary Kriging and CoKriging were used in electrical conductivity mapping based on measured electrical conductivity. The most suitable interpolation method was selected based on the variogram analysis using cross-validation criteria and error evaluation methods (RMSE and $R^2$). According to obtained results, the modified Salinity Index (MSI) with the highest correlation ($R^2$=0.78) was used as an auxiliary variable for the CoKriging method.

Kriging with a spherical model was selected for soil salinity mapping (RMSE = 50.5 and $R^2 = 0.18$). The RMSE and $R^2$ values for the CoKriging method were 43.2 and 0.42, respectively. Table 2 presents the characteristics of the variogram. Experimental variograms and fitted variograms for EC using Kriging and CoKriging are shown in Fig. 2. Presentation errors map and variance reduction in weighting for estimating is the benefits of Kriging against others interpolation methods. Errors in this method are independency from variable and dependent to spatial location and it cause to predict the best location sampling is possible. Variogram relationship based on the measured points is as follows:

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [(z(x + h) - (z(x)))]^2$$  \hspace{1cm} (3)$$

$\gamma$ is the variogram for a distance (lag) $h$ between observations, $z(x)$ and $z(x+h)$, $n(h)$ is the number of pairs of observations which are at distance $h$. $z(x)$ is the observed variable and $z(x+h)$ is the observed variable in the $h$ distance from $z(x)$ and variogram

Varioogram is similar to variance in classic statistic but against variance that is around average, variogram show two samples differences. The main purpose of variogram calculating is investigating variability of variable ratio distance of place or time. For this reason, is necessary to draw a graph with summary of mean square differences of pair points that located at $h$ as $x$ axis.

Figure 3 shows the electrical conductivity map of the study area using Kriging and CoKriging.

| Table 2. Characteristics of semi-variogram in the study area |
|---------------------------------|----------|----------|----------|----------|----------|----------|
| Method  | Model type | Number of lags | Lag Size | Range | Nugget | $R^2$ | RMSE |
| CoKriging  | Spherical | 12 | 500 | 1776 | 2024 | 0.42 | 43.2 |
| Kriging  | Spherical | 13 | 285 | 2575 | 2151 | 0.18 | 50.5 |

![Fig. 2. Experimental variograms and fitted variograms for EC using Kriging (right) and CoKriging (left)](image)

![Fig. 3. Electrical conductivity maps using Kriging (right) and CoKriging (left)](image)
3.2. Multiple linear regression analysis

Residual plots were used to evaluate the quality of a regression. According to results, the distribution of the field measurement data of dependent and independent variables was normal (Figs. 4a and 4b). In addition, no trend was observed in the distribution of standardized residuals (Fig. 5). The standardized residual is the residual divided by its standard deviation. Plot the standardized residual of the simple linear regression model of the data set faithful against the independent variable waiting. The scatter plot of the residuals should be disordered when regression is respectable. When the histogram plot of the residuals shows a symmetric bell-shaped distribution, the normality assumption should be true.

Several multiple linear regression models in the study area were examined to achieve different regression equations (Table 3). Related statistics, such as tolerance value or variance inflation factor were used to examine collinearity in the equations. The Variance Inflation Factor (VIF) indicates the degree to which each independent variable is explained by the other independent variable. Tolerance is the amount of variability of the selected independent variable not explained by other independent variables. Therefore, very small tolerance values (and thus large VIF value because VIF=1/Tolerance) denote a high collinearity (Hair, 1998). According to obtained results, both the tolerance and variance inflation factor were close to 1, suggesting the equation’s low multi-collinearity (Table 3). The Durbin Watson statistic test was used to investigate the autocorrelation in the residuals. A value of 2 indicates that there is no autocorrelation in the sample. Values close to 0 indicate a positive autocorrelation and a value near 4 shows a negative autocorrelation (Habibi, 2016). According to these results, there is no autocorrelation in the samples (Table 3). Besides, no statistically significant correlation coefficient was observed between MSI and DVI (P<0.01).
3.3. Soil salinity zoning

Due to the high coefficient of determination and low standard deviation of the salinity, we applied MSI and DVI index in order to estimate and zonate soil salinity in the study area. The scatter plot (Fig. 6) shows an acceptable relationship between predicted and measured EC when applying the developed regression model ($R^2=0.65$).

Based on the results of this model, the prediction of high soil salinity values was underestimated, while the prediction of the low soil salinity values was overestimated. EC map generated by model 9 in the study area is shown in Figure 7.

The salinity map taken from the stepwise regression model shows that the estimated value is higher than the observed value. The results show that the electrical conductivity in the south and southwest of the Maharloo Lake is low and irregular, but soil EC was increased as distance from the lake decreased.

3.4. Classification of soil salinity map

There are some restrictions on the use of remote sensing data for mapping salt-affected areas related to the spectral behavior of salt types, the spatial distribution of salts on the ground surface, temporal changes on salinity, vegetation diversity, and spectral confusions with other ground surfaces (Metternicht and Zink, 2003).

For soil salinity mapping, 7 soil salinity classes were categorized with intervals of 64, 100, 120, 140, 160, 210, and $>210$ (Pakparvar et al., 2009) The overall accuracy and kappa coefficient for this classification were 87% and 79% respectively. It is worth mentioning that the 7-class system resulted in a higher accuracy than the six-class system, which is normally unexpected. This result might be due to a high level of soil salinity in the study area. As the 6-class system focuses more on lower salinity classes and integrates the high salinity areas in one class of $>64$ ds/m$^2$, it neglects the classification of higher salinity levels, leading to the lower accuracy of the six-class system compared to the seven-class system (Table 4). A classified map of the study area is shown in Fig. 8.

![Fig. 6. Comparison of the observed EC (dS m$^{-1}$) vs. the predicted EC using a stepwise regression model](image-url)
Table 4. Number of classes and interval of salinity and related overall accuracy

<table>
<thead>
<tr>
<th>Number of classes</th>
<th>Maximum EC (dsm$^{-1}$) for each salinity class</th>
<th>Overall accuracy (%)</th>
<th>Kappa Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>24 64 80 100 120 140 180 210 &gt;210</td>
<td>32.69</td>
<td>0.23</td>
</tr>
<tr>
<td>8</td>
<td>64 100 120 140 160 180 210 &gt;210</td>
<td>38.12</td>
<td>0.28</td>
</tr>
<tr>
<td>7</td>
<td>64 100 120 140 160 210 &gt;210</td>
<td>87</td>
<td>0.79</td>
</tr>
<tr>
<td>6</td>
<td>4 8 16 32 64 &gt;64</td>
<td>82.71</td>
<td>0.72</td>
</tr>
</tbody>
</table>

4. Conclusion

In this study soil salinity zoning were investigated using several methods. According to results, the classification system incorporating higher salinity classes showed a better accuracy; hence, it might be stated that the accuracy of salinity detection by remote sensing is increased with the level of salinity. In other words, the spectral response increases with an increase in salt content at the terrain surface.

In this study, Kriging, CoKriging, and multiple regression techniques were used to zone the electrical conductivity changes of the soil. Among investigated methods, MSI with DVI salinity indices were used for the regression, mapping of soil salinity because of their high $R^2$, and low standard error. Near-infrared (NIR), red (RED), and mid-infrared band (SWIR) ratios were applied in this study for salinity detection. Rao et al (1995) cited that salt-affected soils tend to have a higher visibility and NIR reflectance. Matinifar and Zandie (2016) stated that the most common method of recognizing salinity is to calculate the different indicators using infrared and visible bands. Menenti et al. (1986), Shrestha (2006), and Judkind and Mint (2012) showed that Landsat bands, particularly SWIR bands
have an adequate strength to detect soil salinity. Shrestha (2006) showed that the mid-infrared band (Landsat band 7) and near-infrared band (band 4) have an acceptable correlation with the observed EC values of the soil surface layer in the north of Thailand.

Evaluating and monitoring the spatial-temporal change of soil salinity is useful to identify the progression of salinity hazards and to evaluate the effectiveness of remediation strategies (Douaouia et al., 2006). The results of this study show that soil salinity could be estimated via spectral indices with an acceptable accuracy. This method leads to decreases in the costs of the soil mapping in saline soil areas. In aid to future studies, the seasonal variation of the EC level can be investigated via multi-temporal soil and remote sensing data within a year. As many researchers found, to improve the mapping accuracy of soil salinity, the images with better spectral/spatial resolutions will cause better accuracy. Nevertheless, the cost of soil salinity mapping is increased by choosing higher resolution images. Accordingly, an optimal approach should be followed, based on the overall objective, to obtain a more reliable salinity map with the least possible costs.

References


Matinfar, M., A. Shawan, S. Erasmi, 2005. Remote sensing based classification of salt affected soils as
an indicator for landscape degradation in the south of Aleppo, Syria, in UN convention to combat desertification, Trier, Germany.


