Spatial Variability Analysis of Selected Soil Properties at Musayab, Babil, Iraq

Ahmad S. Muhaimeed, Amal M. Saleh
Agriculture college/Baghdad University

Abstract: Great Musayab project was chosen to assess spatial variability of some soil properties, and furthermore to investigate its implications for sampling design. Two hundred and forty composted soil samples were collected across the project and the surrounding areas. Soil properties including electrolytic conductivity (ECe), calcium carbonate (CaCO3), cation exchange capacity (CEC), as well as sand, silt, and clay were analyzed for each sample. Classic statistical analysis showed that ECe had the highest CV which was caused by some unusually high measurements. Semivariograms of all properties were constructed and compared to estimate the spatial variability of the soil properties in the area. Semivariograms of soil properties were best described by an exponential model. Geo-statistical analysis showed that all the soil properties had a moderate or strong spatial dependency. Ordinary kriged maps indicated soils with high ECe, CEC, CaCO3, sand, silt, and clay in the surface horizons were found in the southern parts of the project. Water flows may be the dominant driving force for the spatial variability of chemical properties and texture parameters, implying more samples or analysis are required to achieve a similar level of precision.

1. Introduction:

Spatial dependence - the tendency for observations close together in space to be more highly correlated than those that are further apart. Also called spatial autocorrelation. Spatial dependence imputes that up to some distance apart from each other, two observations at different locations are not statistically independent (Chiles and Delfiner, 1999).

Semi-variance is a measure of the spatial dependence between two observations as a function of the distance between them. Semivariogram - a graph of how semivariance changes as the distance between observations changes. Semivariograms are used for measuring the degree of dissimilarity between observations as a function of distance. Based on the “first rule of geography” that things close together are more similar than things far apart, semi-variance is generally low when two locations are close to each other (i.e. observations at each point are likely to be similar to each other). Typically, semi-variance increases as the distance between the locations grows until at some point the locations are considered independent of each other and semi-variance no longer increases (Karl and Maurer, 2010).

Geostatistics, as a rapidly evolving branch of applied statistics and mathematics that offers a collection of tools, has been utilized extensively to illustrate the spatial variability of a variety of natural phenomena as well as spatial characteristics of soil attributes (Webster and Oliver, 2001; Hoover and Wolman, 2005; Jackson et al., 2007). Geostatistics takes into account both the structured and random characteristics of spatially distributed variables to provide optimal and unbiased estimations. This enables spatial relationships among sample values to be quantified and used for interpolation of values at unsampled locations (Zuo et al. 2008). Huang et al. (2001) showed that knowledge of soil spatial variability and relationships among soil properties is important for the evaluation of agricultural land management practices. His study was to characterize the spatial variability of selected soil properties along transect across a field that was partially grassed Conservation Reserve Program land for 10 years (CRP) and partially continuously cropped land (CCL). Soil chemical properties including pH, available phosphorus (P), and soil total carbon content (STC) were compared and geostatistically analyzed to construct semivariograms.

Iqbal et al. (2005) indicated that analysis and interpretation of spatial variability of soils is a keystone in site-specific farming. The objectives of his study were to determine the degree of spatial variability of soil physical properties and variance structure, and to model the sampling interval of alluvial floodplain soils. Geostatistical analyses illustrated that the spatially dependent stochastic component was predominant over the nugget effect. Structured semivariogram functions of each variable were used in generating fine-scale kriged contour maps. The magnitude and spatial patterns soil physical property variability have implications for variable rate applications and design of soil sampling strategies in alluvial floodplain soils.
Weindorf and Zhu (2010) explained that Non-agricultural lands are surveyed sparsely in general. Meanwhile, soils in these areas usually exhibit strong spatial variability which requires more samples for producing acceptable estimates. Semivariograms of all properties were constructed, standardized, and compared to estimate the spatial variability of the soil properties in the area. Based on the similarity among standardized semivariograms, they found that the semivariograms could be generalized for physical and chemical properties, respectively. Optimal sampling density (OSD), which is derived from the generalized semivariogram and defines the relationship between sampling density and expected error percentage, was proposed to represent, interpret, and compare soil spatial variability and to provide guidance for sample scheme design. OSDs showed that chemical properties exhibit a stronger local spatial variability than soil texture parameters.

The purposes of this study was to describe and interpret the spatial distribution patterns of some soil properties in an area of Great Musayab, central of Iraq project based on geostatistics.

II. Materials and Methods:

Description of the study site:

The project is located within the lands of the governorate of Babil between the Tigris and Euphrates rivers on the left bank of the Euphrates River, just ten kilometers from the Hindiyah dam and the boundaries of the project end about 80 kilometers east of the Euphrates river (Fig. 1). The land sloping of the project rises towards the south 35 m above sea level, and has a hot arid climate with subtropical influence. Summer temperatures frequently exceed 48 °C. Winter temperatures infrequently exceed 21 °C. Typically precipitation is low. Because of very high rates of evaporation, soil and plants rapidly lose the little moisture obtained from the rain, and vegetation could not survive without extensive irrigation. The land of the project is naturally vegetated with Agoobs (Alhagia marorum), but most of area is cultivated barley. The major soil families of the study area are (fine, Smectitic, superactive, calcareous, hyperthermic, Vertic Torrifluvents) and (fine, Smectitic, active, calcareous, hyperthermic, Typic Torrifluvents) (Soil Survey Staff, 2010).

Soil sampling and laboratory analysis

Two hundred and forty soil samples were randomly selected, from 0 to 25 cm depth for chemical and physical property analyses. Soil properties including electrolytic conductivity (ECe), calcium carbonate (CaCO3), cation exchange capacity (CEC), as well as sand, silt, and clay were analyzed for each sample by Department of Soil Investigations Laboratory/Ministry of Irrigation (Muhammad et al., 2001).

Statistical analysis

Means, standard deviations, standard error, coefficients of variation (CV), skewness and kurtosis for each variable were analyzed using classical statistical methods. Data distributions were tested for normality. If data were not normally distributed, they were transformed using natural logarithm to a nearly normal distribution.

![Figure 1. Location of the study site at Musayab, Babil, Iraq.](image-url)
Skewness and kurtosis Measurements

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point.

Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak. A uniform distribution would be the extreme case.

Definition of skewness

For univariate data \(Y_1, Y_2, \ldots, Y_N\), the formula for skewness is:

\[
\text{skewness} = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^3}{(N - 1)s^3}
\]

Where:
- \(\bar{Y}\) is the mean;
- \(s\) is the standard deviation;
- and \(N\) is the number of data points.

The skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero. Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicate data that are skewed right (Hosking, 2006).

Definition of kurtosis

For univariate data \(Y_1, Y_2, \ldots, Y_N\), the formula for kurtosis is:

\[
kurtosis = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})^4}{(N - 1)s^4}
\]

Where:
- \(\bar{Y}\) is the mean;
- \(s\) is the standard deviation;
- and \(N\) is the number of data points.

The kurtosis for a standard normal distribution is three. In addition, positive kurtosis indicates a "peaked" distribution and negative kurtosis indicates a "flat" distribution (Hosking, 2006).

Geostatistical Analyses

Semivariance analysis using ArcGIS (v 9.3 – ESRI Inc.) was used to quantify spatial autocorrelation between neighboring observations, and to facilitate subsequent mapping of soil properties (Boerner et al, 1998). This analysis calculates an index of autocorrelation among groups of paired samples separated by increasing distances.

In order to interpolate surface maps of measured soil properties, the data was fitted to theoretical models. Data was fit to Exponential semivariogram models for the data that was ordinary kriged (Kriging is a geostatistical estimator that infers the value of a random field at an unobserved location) (Strano, 2008).

Characteristics of the Semivariogram

A number of parameters were extracted from the fitted models including the nugget (the semivariance at distance zero), the sill (the y-value at which the semivariance reaches an asymptote), and the range (the distance [x-value] at which this leveling occurs)(Fig.2). We used a system proposed by Cambardella et al. (1994) to define different classes of spatial dependence for the soil properties measured in this study that are based on the ratio of the nugget to the sill. If the nugget to sill ratio was \(\leq 25\%\), the soil property was considered to be strongly spatially dependent, or distributed in patches; if the ratio was between \(26\%\) and \(75\%\), the soil property was considered to be moderately spatially dependent; and if the ratio was \(>75\%\) the soil property was considered to be weakly spatially dependent (Cambardella et al. 1994).
Semivariance

The geostatistical measure of semivariance for interpolation of unsampled locations was determined using the general equation for semivariograms as presented below:

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

Where:
- \( \hat{\gamma}(h) \) is the semivariance at lag distance \( h \);
- \( n(h) \) is the number of observation pairs separated by \( h \);
- \( z(x_i) \) is a measured variable at spatial location \( i \);
- \( z(x_i + h) \) is a measured variable at spatial location \( i + h \) (Bachmaier and Backes, 2008).

III. Results and Discussion

Explanatory statistics

Descriptive statistics of measured soil properties were presented in Table 1. As the sampling scheme adopted in this study is almost evenly distributed, classic statistics could be utilized to reveal the spatial variability of the soil properties.

Soil ECe ranged from 1.10 to 210.00 dS m\(^{-1}\). Distribution of ECe was positively skewed, indicating that there were some extreme high values in this area of Great Musayab. ECe had the highest positive kurtosis value indicating a "peaked" distribution. The CV is the ratio of the standard deviation (SD) to the mean values times 100. ECe had the highest CV (158.12) which was the only one over 100. The extremely high CV of ECe in this study was caused by some unusually high measurements. The reason for such high measurements may be geological, climatic trends, or human activities.

Cation exchange capacity (CEC) ranged from 6.50 to 29.50 cmol kg\(^{-1}\). Distribution of CEC was negatively skewed, indicating that there were some extreme low values of CEC in this area. On the other hand, distribution of CEC was kurtotic. Soil calcium carbonate (CaCO\(_3\)) ranged from 194.00 to 340.00 g kg\(^{-1}\). Distribution of CaCO\(_3\) was negatively skewed but was positively kurtotic.

Descriptive statistics of soil texture parameters were: Sand varied from 1.00 to 85.00 g kg\(^{-1}\). Distribution of sand was positively skewed and also was kurtotic. Silt varied about 7 times from 11.00 to 70.00 g kg\(^{-1}\). Distribution of silt was negatively skewed but was positively kurtotic. Clay varied about 4 times from 4.00 to 46.00 g kg\(^{-1}\). Positive kurtosis values of clay and silt were similar.

Mean values of the soil properties except electrolytic conductivity (ECe) were similar with median values. This similarity was also noted by Emadi et al. (2008). Soil properties are often distributed normally in space. Only two of soil properties studied had a high skewness value greater than one (Table 1), implying that the frequency distributions were highly skewed. Special care should therefore be taken in applying the natural-logarithmic transformation to stabilize the variance (Grunwald et al., 2007).

Correlation coefficients between the soil properties are given in Table 2. Correlations were found to be significantly high between all variables as generally reported, e.g., sand and silt (\( r^2 = 0.725^{**} \)), sand and clay (\( r^2 = 0.819^{**} \)), silt and clay (\( r^2 = 0.946^{**} \)). High significant correlations can also be identified between soil chemical properties, i.e., ECe and CEC (\( r^2 = 0.788^{**} \)), ECe and CaCO\(_3\) (\( r^2 = 0.708^{**} \)), CEC and CaCO\(_3\) (\( r^2 = 0.960^{**} \)).
Spatial Variability Analysis of Selected Soil Properties at Musayab, Babil, Iraq

Table 1. Descriptive statistics of selected soil properties at Musayab, Iraq.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Min.</th>
<th>Max.</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Variance</th>
<th>SE (%)</th>
<th>SD (%)</th>
<th>CV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECe m³</td>
<td>24.967</td>
<td>5.200</td>
<td>1.100</td>
<td>210.00</td>
<td>0</td>
<td>2.989</td>
<td>12.935</td>
<td>2</td>
<td>1558.51</td>
<td>2.548</td>
</tr>
<tr>
<td>CEC cmol kg⁻¹</td>
<td>19.088</td>
<td>18.600</td>
<td>6.500</td>
<td>29.500</td>
<td>-0.226</td>
<td>4.124</td>
<td>16.483</td>
<td>0.262</td>
<td>4.060</td>
<td>21.270</td>
</tr>
<tr>
<td>CaCO₃ g kg⁻¹</td>
<td>280.75</td>
<td>279.00</td>
<td>194.00</td>
<td>340.00</td>
<td>-0.440</td>
<td>2.969</td>
<td>1211.9</td>
<td>2.254</td>
<td>34.91</td>
<td>7</td>
</tr>
<tr>
<td>Sand g kg⁻¹</td>
<td>21.125</td>
<td>17.000</td>
<td>1.000</td>
<td>85.000</td>
<td>2.036</td>
<td>7.911</td>
<td>300.780</td>
<td>1.119</td>
<td>17.34</td>
<td>3</td>
</tr>
<tr>
<td>Silt g kg⁻¹</td>
<td>49.250</td>
<td>53.500</td>
<td>11.000</td>
<td>70.000</td>
<td>-1.088</td>
<td>3.582</td>
<td>179.747</td>
<td>0.865</td>
<td>13.40</td>
<td>7</td>
</tr>
<tr>
<td>Clay g kg⁻¹</td>
<td>29.625</td>
<td>30.000</td>
<td>4.000</td>
<td>46.000</td>
<td>-0.635</td>
<td>3.853</td>
<td>83.229</td>
<td>0.588</td>
<td>9.123</td>
<td>30.795</td>
</tr>
</tbody>
</table>

*a) Standard error; b) Standard deviation; c) Coefficient of variation.

Geostatistics:
The geostatistical parameters describing soil properties from the dataset were listed in Table 3. Regression coefficients (R²) suggested that all models were best fitting to the R² value (greater than 0.5) of the best-fitted model (Duffera et al., 2007).

The semivariograms of soil properties were best described by an exponential model (Fig. 3). Except for soil texture parameters of sand, nuggets for all models were equal to zero. Smaller nugget indicates the sampling intervals proper to reflect the variance. The sill value for soil ECe (917.982) was approximately twice as high than the sill value of soil CaCO₃ (460.251), this implies that ECe had greater variation.

Table 3 shows that all the soil properties have a moderate or strong spatial dependency (Cambardella et al. 1994). The effective ranges of CEC, CaCO₃, silt, and clay are greater than 2000 m, indicating a large-patched distribution pattern (Fig. 4, 5). Given variables with similar nugget/sill ratios, related effective ranges may differ substantially. For instance, Soil EC and silt in this study have similar ratios (0.00) but they have effective ranges of 1753.487 and 2539.616 m, respectively. Apparently, ECe reached its maximum variance level within a shorter lag distance, implying a stronger local variability than silt.

The cross-validation value is the determination coefficient (r²) of the correlation between the measured values and the cross-validation values, which were predicted based on the semivariogram and neighbor values (Robertson, 2008). Despite strong spatial dependency for soil ECe, the prediction efficiency (r²) was slow, and for all other variables the efficiency of spatial prediction ranged from 0.346 to 0.640.

Table 2. Correlation coefficients between selected soil properties at Musayab, Iraq.

<table>
<thead>
<tr>
<th>Variable</th>
<th>ECe</th>
<th>CEC</th>
<th>CaCO₃</th>
<th>Sand</th>
<th>Silt</th>
<th>Clay</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECe</td>
<td>0.788 **</td>
<td>0.708 **</td>
<td>0.960 **</td>
<td>0.945 **</td>
<td>0.640 **</td>
<td>0.705 **</td>
</tr>
<tr>
<td>CEC</td>
<td>0.708 **</td>
<td>0.708 **</td>
<td>0.960 **</td>
<td>0.945 **</td>
<td>0.640 **</td>
<td>0.705 **</td>
</tr>
<tr>
<td>CaCO₃</td>
<td>0.945 **</td>
<td>0.848 **</td>
<td>0.813 **</td>
<td>0.725 **</td>
<td>0.953 **</td>
<td>0.986 **</td>
</tr>
<tr>
<td>Sand</td>
<td>0.599 **</td>
<td>0.926 **</td>
<td>0.957 **</td>
<td>0.725 **</td>
<td>0.953 **</td>
<td>0.819 **</td>
</tr>
<tr>
<td>Silt</td>
<td>0.705 **</td>
<td>0.953 **</td>
<td>0.986 **</td>
<td>0.819 **</td>
<td>0.946 **</td>
<td>0.946 **</td>
</tr>
<tr>
<td>Clay</td>
<td>0.705 **</td>
<td>0.953 **</td>
<td>0.986 **</td>
<td>0.819 **</td>
<td>0.946 **</td>
<td>0.946 **</td>
</tr>
</tbody>
</table>

**Significant at P = 0.01.

Table 3. Semivariogram models and model parameters for selected soil properties at Musayab, Iraq.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Transf.</th>
<th>Model</th>
<th>Model</th>
<th>Nugget</th>
<th>Sill</th>
<th>Nugget/Sill</th>
<th>Spatial dependency</th>
<th>Effective Range</th>
<th>Cross validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECe</td>
<td>Yes</td>
<td>Exp</td>
<td>0.589</td>
<td>0.000</td>
<td>917.982</td>
<td>0.000</td>
<td>Strong</td>
<td>1753.487</td>
<td>0.176</td>
</tr>
<tr>
<td>CEC</td>
<td>No</td>
<td>Exp</td>
<td>0.500</td>
<td>0.000</td>
<td>6.712</td>
<td>0.000</td>
<td>Strong</td>
<td>2082.485</td>
<td>0.522</td>
</tr>
<tr>
<td>CaCO₃</td>
<td>No</td>
<td>Exp</td>
<td>0.527</td>
<td>0.000</td>
<td>460.251</td>
<td>0.000</td>
<td>Strong</td>
<td>2718.658</td>
<td>0.604</td>
</tr>
<tr>
<td>Sand</td>
<td>Yes</td>
<td>Exp</td>
<td>0.605</td>
<td>91.233</td>
<td>183.212</td>
<td>49.796</td>
<td>Moderate</td>
<td>5987.276</td>
<td>0.346</td>
</tr>
<tr>
<td>Silt</td>
<td>No</td>
<td>Exp</td>
<td>0.500</td>
<td>0.000</td>
<td>69.912</td>
<td>0.000</td>
<td>Strong</td>
<td>2539.616</td>
<td>0.640</td>
</tr>
<tr>
<td>Clay</td>
<td>No</td>
<td>Exp</td>
<td>0.611</td>
<td>0.000</td>
<td>42.320</td>
<td>0.000</td>
<td>Strong</td>
<td>2406.116</td>
<td>0.479</td>
</tr>
</tbody>
</table>
a) Transformation of original measurements is applied using natural logarithm if the coefficient of skewness is greater than one; 

b) Semivariogram model: 

\[ \text{Exp} \] (exponential); 

c) Nugget/sill(%) = (nugget/sill) x 100; 

d) Spatial dependency was defined as strong, moderate, weak or pure nugget based on nugget to sill ratios; 

e) The effective range is the model range; 

f) The cross-validation values for a given variable are coefficients of correlation between observed value and values cross-validated by GS + 9.3.

Figure 3. Generalized semivariogram models for chemical properties and soil texture parameters at Musayab, Iraq.

Figure 4. Interpolation maps of selected soil chemical properties using ordinary kriging at Musayab, Iraq.
Generalized semivariogram models

General patterns can be identified for soil chemical properties and soil texture parameters which can be fitted by exponential models (Fig. 3). The soil properties correlograms were reflected in a positive spatial autocorrelation structure. The autocorrelation for soil texture parameter of sand at zero lag was 0.88, and for all the other variables was 0.00. It begins to increase as the lag distance increases, when the autocorrelation does not change significantly with increasing lag distance, the plateau reached, called the sill, reflects the magnitude of random variation (Nielsen, 1998).

Soils in the Great Musayab project, especially along the Tigris and Euphrates rivers, minimally developed Entisols showing little evidence of pedogenesis, therefore differences in spatial autocorrelation extent are not likely related to pedogenic processes, such as eluviation and illuviation. These alluvial floodplain soils have different stratification extents for the soil properties; this suggests that the degree of accumulation and the extent of stratification during deposition of the alluvial materials is the most important factor in explaining the significant extent of spatial autocorrelation.

Ordinary kriged maps indicated soils with high ECe, CEC, CaCO3 in the surface horizons were found in the southern parts of the project (Fig. 4). Similarly, high sand, silt, and clay contents were found in the same spatial pattern (Fig. 5).

The distinctness between the generalized semivariograms of chemical properties and texture parameters may be attributed to the different driving forces during soil formation. The waters of the Tigris and Euphrates are heavily silt laden, irrigation and fairly frequent flooding deposit large quantities of silty loam in much of the project area. Windborne silt contributes to the total deposit of sediments. By the time, the flow of the rivers is substantially reduced, and the surface area of the resulting sediment volume increases. The Tigris and Euphrates also carry large quantities of salts. These, too, are spread on the land by sometimes excessive irrigation and flooding. A high water table and poor surface and subsurface drainage tend to concentrate the salts near the surface of the soil. Most soils of Iraq are located in arid and semi-arid regions with high amount of calcium carbonate which results in higher calcification rate. Extensive leaching may have removed the CaCO3 from soil of the northern parts of the project area, but often the amount of CaCO3 in soils derived from calcareous parent material is considerable.

Water flows may be the dominant driving force for the spatial variability of texture parameters, soil particles can move with water and tend to deposit and accumulate on the areas where water flows slow down.

IV. Conclusions

Classic statistical analysis showed that ECe had the highest CV which was the only one over 100. Mean values of the soil properties except electrolytic conductivity (ECe) were similar with median values. However, soil properties are often distributed normally in space.

Geo-statistical analysis showed that all the soil properties had a moderate or strong spatial dependency. General patterns can be identified for soil chemical properties and soil texture parameters which can be fitted by exponential models. Ordinary kriged maps indicated soils with high ECe, CEC, CaCO3 as well as sand, silt, and clay in the surface horizons were found in the southern parts of the project. Except soil texture parameter of...
sand, nuggets for all models were equal to zero. Smaller nugget indicates the sampling interval is proper to reflect the variance.

It should be noted that the generalized semivariogram models enables soil scientists to use measured soil chemical and physical data over greater distances to estimate attributes in the unsampled locations.

References