ASSESSMENT OF SOIL MAPPING ACCURACY USING RS, GEOSTATISTICS AND GIS TRADITIONAL METHOD CASE STUDY: Um El-HUSE - SIWA OASSIS, EGYPT.

MOHAMED E. A. KHALIFA

Pedology Dept., Water resources and desert land division, Desert Research Center, Egypt.

ABSTRACT

Conventional soil surveys lack to detailed soil spatial information which is required for many land management applications. Soil related modern technologies like remote sensing and geostatistics have the capability to compensate that inadequacy. The present study aimed at combining the knowledge derived from remotely sensed data and geostatistics throughout GIS framework to produce detailed soil maps. Studied area was selected over about 4200 Feddans at Um El-Huse area to the north of Siwa oasis. Semi-detailed survey was carried out using 66 profiles. According to laboratory analysis four soil mapping units have been differentiated, varying in profile depth, salinity, alkalinity, texture and lime content. Representative LANDSAT scene of the area has been classified using isodata unsupervised classification into 15 spectral classes, which have been regrouped according to field investigation and intensive lab. analysis into these predetected at conventional soil map. Results indicated interactions between mapping units have not matched after conventional method, where associated standard mapping errors were 8.1%, 11.6%, 7.3% and 2.8%, respectively for soil units. Remotely sensed soil map has proven an obvious successfulness over 82.2% from check points vs. 65.2% for traditional method. Geostatistically, profile depth, salinity and carbonate were best fitted by Gaussian semivariogram while alkalinity was best fitted by Spherical model. Fitting confidents were 0.90%, 0.77%, 0.84% and 0.90% for previous soil properties, respectively. Punctual kriged maps were
undertaken and overlayed with soil texture to generate the Geostatistical soil map. The Kriging maps were cross-validated and the correlation coefficients were 1.09, 0.98, 0.97 and 1.08, respectively, for soil depth, salinity alkalinity and lime content. Generally, the current study showed the superiority of soil map creation using RS or geostatistics over traditional method due to exploring the inner variations of soil properties and accurate map delineations. Moreover, the integration between modern technologies and traditional measurements may be useful towards more precise reliability.

**Keywords:** Soil mapping, Geostatistics, Kriging, Semivariogram, Geographical Information System, Remote sensing.

**INTRODUCTION**

Detailed spatial information of soil attribute is required for many environmental modeling and land management applications (Beven and Kirkby, 1979; Burrough, 1996). Soil maps based on conventional surveys are the major source of soil spatial information, although, standard soil surveys were not designed to provide the detailed (high-resolution) soil information (Band and Moore, 1995) and crop management applications (Peterson, 1991). Soil maps may be enhanced by other landscape data derived from detailed digital terrain analyses and remote sensing techniques (Band and Moore, 1995). This inadequacy is largely due to the limitations of the discrete data model and polygon-based mapping practice employed in conventional soil surveys. Zhu et al. (1996) developed a soil-land inference model to overcome the limitations in conventional soil surveys by combining the knowledge of soil scientists with GIS techniques under fuzzy logic to map soils.

Remote sensing investigations revealed that the most important distinction between soil type's surfaces is in their reflectance which depend on soil properties (Hoffer, 1978). Morra et al. (1991) used Near-infrared reflectance spectroscopy (NIRS) to determine soil total carbon and nitrogen using regression analysis. Cindy et al. (1994) proposed a complementary method for estimating surface organic
carbon levels of soils using the analysis of digital data from Landsat Thematic Mapper (TM) images. Ben-Dor et al. (2002) used hyperspectral image data to produce prediction equation model for soil organic matter, soil field moisture, soil saturated moisture, and soil salinity, and then produce quantitative maps for these properties. **Unsupervised classification** for satellite images aims to cluster image pixels in a dataset into similar spectral classes, without any prior knowledge of these classes. Thus, a small range of digital numbers (DNs) can fix one cluster that is specified from another cluster (Lillesand and Keifer, 1994). **Revised classification** is much more effectual in terms of accuracy in mapping which used to cluster pixels in a data set into classes corresponding to user-defined classes.

The spatial distribution of soil properties should be monitored for effective management of agricultural field (Yi-Ju et al., 1998). Geostatistical analysis has been considered recently as a solution key of many mapping researches to describe the spatial variability using the semivariogram and kriging modeling to predict soil attributes at un-sampled sites (Warrick et al., 1986; Webster, 1991; Goovaerts, 1999). Soil properties vary in the space according to repeated regular pattern rules and their spatial variability depend mainly on distances apart between observations sites (Lagacherie et al., 1995). Semivariogram model is defined as a plot of variance of paired sample measurements as a function of the distance between samples to expresses the way in which the distribution of a soil property varies. (Oliver and Webster, 1991). The term Kriging is a weighted local average interpolation method used to predict spatial soil properties as a function of modeled semivariogram, locations of samples to each other and to the point or block being estimated. (Oliver and Webster, 1991). The spatial relationship between the values of the attribute is governed by the regionalized variable theory, which states that observations close to each other are more correlated than observations taken at a further distance (Journel and Huijbregts, 1978). This means that points spatially close to the estimation points should be given higher weights than those further away (Cressie, 1993). The spatial distribution of the variables is not random, but follows the regionalized theory, i.e., observations close to each other on the ground tend to be more alike than those further apart (Journel and Huijbregts, 1978). The presence
of such spatial structure is prerequisite to the application of Geostatistics, and represent the first step towards spatial prediction (Burrough and McDonnell, 1998).

The aim of this study is to apply three different methods for mapping the spatial distribution of soil properties; Satellite image classification, Geostatistical procedure and conventional overlaying of measured soil properties by GIS. The resultant soil mapping units have to be compared and evaluated in term of accuracy.

MATERIALS AND METHODS

1- The study site

Studied area covers about 4,200 feds. to the northern west direction of Siwa oasis at Um El-Huse area between latitudes 29° 13’ 04” and 29° 15’ 05” and longitudes 25° 47’ 03” and 25° 50’ 12”. It comprises part of the newly reclaimed soils at Siwa oasis. Map 1 shows location of the study area.

Map 1: Location of the study area at Siwa oasis.

2- Soil survey:

Isodata unsupervised classification was elaborated to classify a represented Landsat 7 (ETM+) scene (2006) of the studied area using ERDAS software (Erdas Imagine, 2001). Multiple pixels in a cluster (class) are corresponding to a ground feature which fixed by a
homogeneous range of digital numbers (DNs) and set apart from other clusters. Fifteen spectral classes were identified based on the spectral signatures of the image pixels (Map 2).

Studied area which extends over 4,200 feddans was surveyed at semi-detailed scale (one profile for each 60 feds.) according to detected spatial extension of unsupervised spectral classes. A total of 66 soil profile were dug, described pedomorphologically according to FAO (1998), and sampled. The samples locations were georeferenced to the UTM coordinate system (ESRI, 2006) (Map 3).

Map 2: Spectral classes of the study area ground surface

Map 3: Representative soil profiles of the studied area.

• Soil Profile
3- Laboratory analysis:
Collected soil samples were analyzed to determine; soil texture using pipette method, Piper (1950), electrical conductivity (EC) in dS/m at 25°C, soluble cations and anions and consequently ESP; soil reaction (pH), according to Page et al. (1982); Total calcium carbonate % according to Page et al. (1982).

4- Semi-supervised classification:
The purpose of that classification is to revise the unsupervised spectral classes by reducing and regrouping them according to field observations and laboratory analysis results. Ground surface classes at processed scene was refined by visiting the predetermined sites of the studied area at field and record field observations (i.e. land forms, vegetation cover, stoniness (gravels, grits and rock fragments), presence of calcareous crust, presence of windblown sand sheets, sand dunes, ..... etc).

5- Geostatistical analysis:
5.1- Semivariogram model: The semivariogram is defined as half of the average squared difference between two attribute values separated by vector h, for one variable (Journel and Huijbregts, 1978, Burrough and McDonnell, 1998):

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

where $N(h)$ is the number of pairs at lag $h$, $Z(x_i)$ is the value of the attribute at location $(x_i)$ and $Z(x_i + h)$ is the value of the attribute at location $(x_i + h)$ separated by distance $h$. The separation vector $h$ is specified with some direction and distance (lag) tolerance. This semivariogram is used to model soil attributes with high CV, and then fitting them to one of the known semivariogram functions (i.e.: Gaussian, Exponential, Spherical….etc), Fitted semivariogram models have main three parameters, namely, nugget variance ($C_0$), sill variance (C), and range (A), which were calculated for each studied soil attribute.

5.2- Kriging: According to fitted variograms, kriged estimates were calculated as linear sum of the data (Journel and Huijbregts, 1978). If there are some measured values of a property (Z) at some sampling point, $X_1, X_2..., X_i$, the estimated value (Z) at non-sampled
point \((X_0)\) is estimated based on the associated weights of sampling points \((\lambda)\) according to the following equation:

\[
(X_0) = \lambda_1(X_1) + \lambda_2(X_2) + \ldots + \lambda_i(X_i)
\]

Kriging was undertaken using punctual kriging method. Estimations and associated standard deviations of soil attributes were calculated. Fitted semivariograms and Kriged estimates were carried out using GS+ (Gamma Design, 2001) and contour maps were plotted using Surfer 8 (Golden software, Ins. 2002).

5.3- Cross Validation: Estimated values \((v')\) were tested at the locations of existing samples by comparing with true values \((v)\) using the information available in the data set (Issaks and Srivastava, 1989). Estimation errors and Mean square error (MSE) were calculated:

\[
Error = r = v' - v \quad MSE = \frac{1}{n} \sum_{i=1}^{n} r^2
\]

RESULTS AND DISCUSSIONS

1- Descriptive statistical analysis

Table 1 shows some summary statistics for soil attributes with high coefficient of variance (CV). Data emphasize on soil salinity has the highest CV (0.92) followed by sodium adsorption ratio (0.88) and calcium carbonate (0.82).

2- Soil mapping units

Soils of the area have moderate soil texture within three classes; sandy loam, silty loam and silty clay loam. Three profile effective depth classes were found: deep, moderately deep and shallow, where depths ranged between 30.0 cm (shallow) and 125 cm (deep). Soil salinity varied in a wide range from 4.0 dS/m (moderately saline) to 23.4 dS/m (highly saline). Soil alkalinity belonged to moderately alkaline and alkaline classes where ESP ranged between 8.0 and 34.1. Lime content was moderate and ranged between 5.2% and 22.9%. Generally, main four soil mapping units have been achieved as processed by traditional method (Table 2 & Map 4)
Table (1) Summarized statistical data for some soil attributes.

<table>
<thead>
<tr>
<th>Statistical Parameter</th>
<th>EC dSm⁻¹</th>
<th>ESP</th>
<th>CaCO₃ %</th>
<th>P.D cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>4.0</td>
<td>8.0</td>
<td>5.2</td>
<td>30.0</td>
</tr>
<tr>
<td>Maximum</td>
<td>23.4</td>
<td>34.1</td>
<td>22.9</td>
<td>125.0</td>
</tr>
<tr>
<td>Mean</td>
<td>10.6</td>
<td>17.4</td>
<td>15.1</td>
<td>73.6</td>
</tr>
<tr>
<td>CV% (coefficient of variation)</td>
<td>0.92</td>
<td>0.88</td>
<td>0.82</td>
<td>0.80</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>3.35</td>
<td>7.0</td>
<td>1.98</td>
<td>1.25</td>
</tr>
<tr>
<td>Variance</td>
<td>55.3</td>
<td>76.2</td>
<td>25.5</td>
<td>18.8</td>
</tr>
</tbody>
</table>

3- Semi-supervised classification:

Field investigation showed that spectral classes of the studied area could be regrouped into four soil mapping units with different characteristics and different soil properties as emphasized by previous traditional soil map (Table 2). Generally, elevation of studied area degraded from +10 m A.S.L towards southern west direction to be -10 m A.S.L. Variations in elevation were associated with high compatibility between land forms and surface cover differences with variations in soil properties (Table 2).

Map (5) differentiates between four soil mapping units where spatial distribution is based on satellite image revised classification. Over the determined reference points at the semi-supervised image classification map, analysis of representative soil samples emphasized on that interference between mapping units. Each soil mapping unit in the traditional map included inner areas belong to one or more other soil units have not been recognized unless image classification took place. Area of unit C, for moderately deep saline soils, obviously has been increased from 46.8% in conventional soil map to 50.5% in RS soil map as seen in maps (4 & 5), with general reduction of the rest of units' areas. Standard errors of mapping associated with traditional soil map were 8.1%, 11.6%, 7.3% and 2.8% for main soil units, respectively. Superiority of image classification mapping over traditional mapping attributed to that successfulness at 82.2% of reference points as seen in the accuracy assessment table (table 3).
Table (2) Description of soil mapping units in the studied area.

<table>
<thead>
<tr>
<th>Unit</th>
<th>Soil properties</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Mod. saline mod. alkaline slightly calcareous deep sandy loam soils</td>
<td>Level surface with common desert pavement over drift sand.</td>
</tr>
<tr>
<td>B</td>
<td>Mod. saline mod. alkaline slightly calcareous mod. deep silty loam soils</td>
<td>Moderately sloped surface with varisized gravels and rock fragments.</td>
</tr>
<tr>
<td>C</td>
<td>Saline alkaline mod. calcareous moderately deep silty clay loam soils</td>
<td>Gently sloped surface with varisized gravels and scattered desert shrubs.</td>
</tr>
<tr>
<td>D</td>
<td>Highly saline alkaline mod. calcareous shallow silty clay loam soils</td>
<td>Level surface with common salt crust and sparsely salt high tolerance plants.</td>
</tr>
</tbody>
</table>

Table (3) Accuracy assessment of classified points in comparison with randomize reference points.

<table>
<thead>
<tr>
<th>Class name</th>
<th>Ref. Points</th>
<th>Cls. Points</th>
<th>Error %</th>
<th>Accuracy%</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>6</td>
<td>5</td>
<td>16.6</td>
<td>83.4</td>
</tr>
<tr>
<td>B</td>
<td>15</td>
<td>17</td>
<td>13.3</td>
<td>86.7</td>
</tr>
<tr>
<td>C</td>
<td>35</td>
<td>31</td>
<td>11.4</td>
<td>88.6</td>
</tr>
<tr>
<td>D</td>
<td>10</td>
<td>13</td>
<td>30.0</td>
<td>70.0</td>
</tr>
<tr>
<td>Total-Average</td>
<td>66</td>
<td>66</td>
<td>17.8</td>
<td>82.2</td>
</tr>
</tbody>
</table>
Map (4) Soil Mapping Units by Traditional

Map (5) Soil Mapping Units through semi-supervised satellite image classification
4- Geostatistical soil map

4.1-Semivariograms: The semivariograms for soil salinity, calcium carbonate and profile depth were best fitted to the Gaussian model as shown in the following equation:

\[ \gamma(h) = C_0 + C \left[ 1 - \exp\left( -\frac{3h^2}{a^2} \right) \right] \]

While soil alkalinity (ESP) has best fitted by Spherical semivariogram model as the following equation:

\[ \gamma(h) = C_0 + C \left[ \frac{3h}{2a} - \frac{1}{2} \left( \frac{h}{a} \right)^3 \right] \quad \text{as } h \leq a \]

\[ \gamma(h) = C_0 + C \left[ \frac{h}{3a} - \frac{1}{2} \left( \frac{h}{3a} \right)^3 \right] \quad \text{as } h \geq a \]

Where \( C_0 \) is the nugget, \( C \) is the sill, \( h \) is the separation distance (lag) in meters, and \( a \) is the range. The formulated equations for these variables are as follows:

\[ \gamma_{EC}(h) = 13.9 + 58.8 \left[ 1 - \exp\left( -\frac{3h^2}{296} \right) \right] \]

\[ \gamma_{CaCO_3}(h) = 15.6 + 62.2 \left[ 1 - \exp\left( -\frac{3h^2}{511} \right) \right] \]

\[ \gamma_{PD}(h) = 373 + 1356 \left[ 1 - \exp\left( -\frac{3h^2}{342} \right) \right] \]

\[ \gamma_{ESP}(h) = 15.4 + 81.8 \left[ \frac{3h}{76.4} - \frac{1}{2} \left( \frac{h}{38.2} \right)^3 \right] \]

Parameters for these fitted semivariograms are shown in table (4) and associated curves are shown in figure (1).

Table 4: Semivariogram types and parameters for soil attributes.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model</th>
<th>Nugget (C0)</th>
<th>Sill (C)</th>
<th>Range (a)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td>Gaussian</td>
<td>13.9</td>
<td>58.8</td>
<td>17.2</td>
<td>0.908</td>
</tr>
<tr>
<td>CaCO₃</td>
<td>Gaussian</td>
<td>15.6</td>
<td>62.2</td>
<td>22.6</td>
<td>0.767</td>
</tr>
<tr>
<td>P. Depth</td>
<td>Gaussian</td>
<td>373</td>
<td>1356</td>
<td>18.5</td>
<td>0.840</td>
</tr>
<tr>
<td>ESP</td>
<td>Spherical</td>
<td>15.4</td>
<td>81.8</td>
<td>38.2</td>
<td>0.906</td>
</tr>
</tbody>
</table>
Figure 1: Best fitted semivariogram models for (a) soil salinity, (b) soil alkalinity, (c) calcium carbonate and (d) profile depth.

It is clear that the coefficient of determination $R^2$, obtained from the fitting process, for both EC and ESP exceeds 0.90, which indicates the goodness of the estimation using Gaussian and spherical models, respectively, while it had been decreased for carbonate and profile depth. Moreover, fitted Gaussian semivariogram model for most soil properties indicates a smoothly spatial varying pattern for these variables (Burrough and McDonnell, 1998).

4.2- Kriged soil map: According to fitted semivariogram models, kriged estimates for studied soil attributes were interpolated as seen in map (6) for (a) profile depth, (b) salinity, (c) alkalinity, (d) carbonate, in addition to spatial variation of (e) soil texture. Kriged soil map was generated by overlaying estimated sheets for different soil properties as seen in map (7). It is clear that spatial distribution of actual data aggregated values of soil properties in contiguous groups due to the lack of information in the area between sampled locations, while interpolation by kriging takes into consideration the spatial variability of the surrounding points. Obviously, the kriged estimates are smoothed out, because estimated values are less variable than actual values (Goovaerts, 1999). The standard deviation of kriged
estimates is much higher for calcium carbonate and profile depth than for salinity and alkalinity (Map 8).

Map (6): Kriged estimates for (a) profile depth, (b) salinity, (c) alkalinity and (d) calcium carbonate; in addition to (e) voronoi soil texture map
Map (7) Soil Mapping Units based on Geostatistics

Map (8): Standard deviations of kriging for (a) soil depth, (b) salinity, (c) alkalinity and (d) calcium carbonate.
4.3- Cross validation of kriging: This process measures the linear regression between kriged and actual values of studied soil attributes (Fig 2).

Figure 2: Linear regression curves between kriged and actual values of (a) profile depth, (b) salinity, (c) alkalinity and (d) carbonate.

Regression analysis showed high correlations between measured and estimated values where regression coefficients were 1.09, 0.98, 0.97 and 1.08 for profile depth, EC, ESP and carbonate, respectively.

However, obvious successful and superiority was achieved for soil map creation throughout either satellite image classification or geostatistical analysis in comparing with conventional method. Modern plotting techniques are powerful in exploring inner soil variations and delineating precisely the boundary between different units.
REFERENCES


ERDAS IMAGINE 8.6. (2001). Geographic imagine package. ERDAS Inc. USA.


مختصر العربية
تقييم دقة خرائط التربة المنشأة بإستخدام الاستشعار عن بعد، الاحصاء الأرضية، والطرق التقليدية لنظم المعلومات الجغرافية. دراسة حالة: أم الهيوس - سيوة، مصر.
محمد عزت عبد الهادى خليفة

الملخص العربي
تقدر خرائط التربة المنشأة بالطرق التقليدية إلى قدر التفاصيل اللازم تطبيق النماذج الحديثة لإدارة الأراضى والمياة والتربة. ويهدف من البحث تقييم عدد إستخدام التقنيات الحديثة وطرق الدراسة المتطورة للأراضى في رسم خرائط التربة، حيث تقارن الطرق التقليدية لإنشاء خرائط التربة بكل من خرائط محاليل صور الأقمار الصناعية وخرائط التحليل الجيوإحصائي. وقد تم اختيار منطقة الدراسة بالقاهرة أم الهيوس – شمال ولاية سيوة - على إمتداد 4.2 كم بعد نظام لاحق من بذل الجهود لدراستها التربية لمساحة 46.4 كم². حيث تم رسم خرائط التربة على المستوى النصفي تفصيلي ببسط عدد (66) قطاع أرضي تعرض كل منها مساحة 0.7 كم² وتم تقسيم الصورة LANDSAT الممثلة للمنطقة إلى عدد (51) وحدة، بناءً على الدراسة الحقلية والتحليل المعتمد. أمكن إختزال الدراسة بضبط النماذج الطبيعية المعتمدة على الاحصائيات والطريقة الأساسية التي تشبه الجغرافية (Gaussian). وتظهر من النتائج أفضلية خرائط التربة المنشأة باستخدام الاستشعار عن بعد، حيث تجاوز نسب التوافق مع خرائط التربة التقليدية (Land use) 65.2% من نقاط التوقيت، وأظهرت رسم خرائط التربة الجيوإحصائية تفوقاً واضحاً في نوعية وبناء خرائط التربة، حيث تجاوزت نسب التوافق مع خرائط التربة التقليدية (Land use) 93.5% من نقاط التوقيت، وأظهرت رسم خرائط التربة الجيوإحصائية تفوقاً واضحاً في نوعية وبناء خرائط التربة، حيث تجاوزت نسب التوافق مع خرائط التربة التقليدية (Land use) 93.5% من نقاط التوقيت، وأظهرت رسم خرائط التربة الجيوإحصائية تفوقاً واضحاً في نوعية وبناء خرائط التربة، حيث تجاوزت نسب التوافق مع خرائط التربة التقليدية (Land use) 93.5% من نقاط التوقيت.
الإحصائي مقارنة بالطريقة التقليدية حيث تفوقت كلاً الطريقتين في استكشاف التباين الداخلي لصفات التربة وتحديد حدود الوحدات الأرضية بدقة. وهكذا تؤكد الدراسة على أهمية التقنيات المتقدمة في رصد الاختلافات الأرضية لرسم وتوقع خرائط التربة حيث تتكامل فيه عناصر الواقعية المصاحبة للطرق التقليدية مع عناصر التقدير المكثفة للطرق الحديثة.