An overview of pedometric techniques for use in soil survey.
Une revue des techniques pédométriques utilisables en cartographie des sols

MCBRATNEY Alex., ODEH Inakwu

Department of Agricultural Chemistry & Soil Science, The University of Sydney Ross St Building A03 NSW 2006 Australia

I - Introduction

Soil survey methods had in the past been criticised, perhaps justifiably, for being too qualitative in character. In response to these criticisms, quantitative models have been developed within the last twenty years or so, which are being used to describe, classify and study the spatial distribution patterns of soil as it occurs in the field. These quantitative methods should enable precise statements about the soil to be made. The methods are collectively categorised in the emerging field of soil science known as pedometrics (Webster, 1994). This paper briefly describes the pedometric techniques being used for soil survey, summarising their appropriateness in various instances and their limitations. As space does not allow us to present worked examples here, more details will be presented during the oral Symposium at the Congress. Before proceeding further, however, we briefly examine the various sources of environmental information that provide links with the pedometric analysis for the enrichment of soil data.

II - Sources of exogenous information for soil data enrichment

The task of a soil survey is to provide soil information for either general purposes or for a specific use. In the past, surveyors based their approach on the qualitative analysis of the landscape either by physiographic analysis or by aerial photographic interpretation or both. These were all attempts to enrich the soil information through the use of exogenous data. Due to increasing awareness of environmental pollution and associated problems by the general community, quantitative soil information is now required to enable precise statements on the status of the environment to be made. Pedometric techniques have been developed to meet these requirements. Basically, these techniques stem from the classical approach when the soil surveyor generally would study the climate, geology, geomorphology, vegetation, land use and land use history prior to any soil survey. Lately technical advances have provided us with a wealth of new environmental data sources. These are summarised in Table 1.
Air-borne and space sensors are now generating streams of terabytes of data almost always on a daily basis. The air-borne sensors vary from the traditional aerial photography to air-borne videography (Table 1). The space sensors generally consists of series of satellites, many of which were originally for general studies of the various earth resources. Increasingly, proximal sensors are now being used for the purpose of research and development for soil-specific crop management (McBratney and Pringle, 1997). We now examine pedometric techniques that are used for the analysis of the sensed data for more efficient and inexpensive, quantitative soil inventories.

Table 1. Examples of sources of exogenous data for soil inventory

<table>
<thead>
<tr>
<th>Carrier</th>
<th>Sensor/Scanner</th>
<th>Sensed Data</th>
<th>Land/Soil Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air-borne: (Aeroplanes and Balloons)</td>
<td>Photogrammetric and videographic cameras SLAR NIR LIDAR Gamma-radiometer</td>
<td>spectral imageries (albedo), radiance energy radiance energy gamma radiation</td>
<td>DEM, landscape, crop growth, vegetation moisture, landscape moisture, clay, etc. K, U and Th isotopes</td>
</tr>
<tr>
<td>Space-borne: Landsat</td>
<td>Multi-spectral thematic mapper, etc High resolution visible (panchromatic and multispectral)</td>
<td>radiance energy and albedo radiance energy and visible spectral imageries (albedo) surface reflectance, surface temperature</td>
<td>vegetation, moisture</td>
</tr>
<tr>
<td>SPOT Satellites</td>
<td>Advanced very high resolution radiometers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOAA Satellites</td>
<td>Time domain reflectrometer Electromagnetic induction instrument Gamma-radiometer Soil chemical sensors Yield monitors</td>
<td>apparent dielectric constant apparent conductivity gamma radiation soil chemical composition crop yield (ha⁻¹) and quality</td>
<td>moisture, salinity, clay and moisture K, U and Th isotopes soil fertility, organic carbon, etc. yield data associated with soil variability</td>
</tr>
<tr>
<td>Proximal Sensors and Scanners: (humans, vehicular, etc.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humans</td>
<td>Various, human</td>
<td>existing ground-truth data and maps, etc</td>
<td>e.g., topographic, geologic, vegetation</td>
</tr>
</tbody>
</table>

III - Pedometric techniques

There are basically two generic groups of pedometric techniques for soil survey in general and the enrichment of soil information specifically: 1) the classical methods collectively referred to here as the CLORPT methods, and 2) the geostatistical methods. To these we add the third group as 3) the hybrid methods. The hybrid methods are some combinations of techniques from the two generic groups to optimise prediction of soil properties (Fig. 1).
1 - Generic techniques

A) - The CLORPT methods are based on the empirical-deterministic models that originated from Jenny’s (1941) *Factors of Soil Formation* (Fig. 1). Jenny (1941) mathematical function of soil formation is expressed as:

\[ S = f(CL, O, R, P, T) \]

where \( S \) is some soil properties as a function of the factors: \( CL \) as climate, \( O \) as organisms, \( R \) as relief, \( P \) as parent material, and \( T \) as time. Soil spatial variability is therefore considered as being causative realisations of the complex combinations of soil-forming processes as influenced by the soil forming factors. The CLORPT function stimulated numerous studies, especially on the effects of relief and time factors. Much of the earlier studies, and indeed some recent ones, were based on general and bivariate-simple linear regression (e.g., Ruhe and Walker, 1968; Furley, 1971; Moore et al., 1993). The realisation that these studies do not accommodate the non-linearity in the relations has led to recent application of the more robust methods such as generalised linear models, generalised additive models and regression trees (Odeh et al., 1994). But the question to ask is: while the classical methods may take care of the deterministic relations, do they account for spatial autocorrelations of the soil properties, especially at the local level? To answer this question, the pioneer pedometricians initiated the application of geostatistics (which was primarily developed for the mining industry), the compendium of which is given below.

![Diagram of CLORPT techniques, Hybrid techniques, and Geostatistical techniques](image)

Figure 1. The main pedometric techniques used for soil survey

B) - Geostatistical methods are based on the theory of regionalised variables (Matheron, 1965), which allows us to consider spatial variability of a soil property as a realisation of a random function that can be represented by a stochastic model. The geostatistical method of spatial interpolation is termed kriging. The first major applications of ordinary kriging in soil studies emerged in the early 1980’s (e.g., Burgess and Webster, 1980). Since then there has been massive explosion of its use in various sub-fields of soil science: in soil reclamation (Samara and Singh, 1990), in soil classification (e.g., Odeh et al., 1992), in soil pollution studies (e.g., Hendricks Frassen et al., 1997), etc. Major limitations of the univariate geostatistics technique of kriging are due to the assumptions of stationarity which are not often met by the field-sampled data sets and, of course, the often cited requirement of large amount of data to define the spatial autocorrelation. The univariate usage of the techniques is also limiting in situations of complex terrain where the soil-forming processes are themselves complex.
In such situations there is the need to model both the structured and the spatially
dependent components of the soil variable. Also there are economic and logistic reasons
for including the factors influencing soil variability, especially if the latter are more
readily and cheaply available. As both the soil and the exogenous factors are multivariate,
the most obvious choices are appropriate combinations of multivariate analysis using the
CLORPT factors and the geostatistical methods. These combinations constitute the
hybrid techniques (Fig. 1).

2 - Hybrid techniques
The hybrid techniques for soil survey are based on combinations of the geostatistical and
multivariate or univariate CLORPT methods. Let us suppose that a data vector
descriving a soil property is a random variable \( Z \), determined at locations in a region, \( X = x_1, ..., x_N \), and consisting of three components as

\[
Z = m + Z_1 + e
\]

where \( m \) is the local mean for the region, \( Z_1 \) is the spatially dependent component and \( e \)
the residual error term, spatially independent. Now there may be situations where \( m \) is
varying and dependent on some exogenous factors such as the CLORPT factors. In other
words it is deterministically related to some causative factors (in geostatistics parlance,
the variable is said to exhibit trend). Wherever trend exists, the ordinary univariate
kriging is inappropriate. Several methods have been designed to accommodate the trend.

A) - Universal kriging (Matheron, 1969) has been the commonly used method
to accommodate the trend or the “changing drift”, as it is sometimes known, in a soil
variable. The universal kriging is a combination of the standard model of multiple-linear
regression and the geostatistical method of ordinary kriging (Webster, 1994), which is
also analogous to combining CLORPT methods with the univariate kriging using the
geographical coordinates as the exogenous factors. More recently, a more advanced
approach, the Intrinsic Random Function of Order k- IRF-k, has been used to
accommodate the varying nature of the trend in a regionalised soil variable (McBratney
et al., 1991). The term k represents the order of polynomial trends- k = 0 means constant
drift, and the IRF-k is equivalent to ordinary kriging system of equations; if k = 1, we
have linear drift; k = 2 yields quadratic drift. But when the deterministic relationships are
with some known or readily-available and inexpensive covariates (CLORPT factors) or
other easy-to-measure soil variables, cokriging has played a major role in efficiently
predicting the target soil variable (McBratney and Webster, 1983; Stein et al., 1988;
Odeh et al., 1995).

B) - Cokriging is the multivariate extension of kriging that allows the inclusion
of more readily available and inexpensive attributes in the prediction process. There are
many instances in soil survey where the CLORPT factors such as topography, time,
variable parent material, etc, are easily discernible or are either readily available and/or
are cheap to obtain. The most efficient way to predict the expensive-to-measure target
soil variable that is highly correlated with the CLORPT factors is to use the factors in
cokriging the target soil variable- sampled at few locations, into denser grid nodes. This
is termed as heterotopic cokriging (Wackernagel, 1994) in comparison with isotopic
cokriging which requires that data be available at all sample locations. A variant of both
is the generalised cokriging that involves simultaneously prediction of all the correlated
variables into more dense locations. Heterotopy can either be complete or partial
(Wackernagel, 1995). The complete case is the case where the covariates and the target
variable do not share any common locations. This is often the case with using exogenous variables especially landform attributes derived from DEM (Odeh et al., 1995) and satellite imageries for predicting the target soil variable. The partial heterotopy involves cases where there are some coincidence of the locations of the target variable and the ancillary variables. The latter is often the case when other soil covariates are used (McBratney and Webster, 1983).

C) - **Regression-kriging** is another hybrid method that combines either a simple or multiple-linear regression model with ordinary kriging of the regression residuals (Odeh et al., 1995). The assumption here is that the deterministic component \( m \) in equation 1 of the target (soil) variable is accounted for by the regression model, while the model residuals represent the spatially varying but dependent component \( Z_1 \) in equation 1. If the exogenous variables used in the regression equation are available at more dense locations than the target variable, the equation can then be used to predict the \( m \) onto those locations. The \( Z_1 \) can also be predicted to the same locations by ordinary kriging system of equations, and then added to the \( m \) to obtain \( Z^* \). Odeh et al. (1995) and Odeh and McBratney (1998) have demonstrated the superiority of regression-kriging to other prediction methods such as ordinary kriging, universal kriging, multiple-linear regression and cokriging. A variant of regression-kriging is **kriging with uncertainty** (Ahmed and DeMarsily, 1987). The kriging with uncertainty, introduces the regression residuals (as representing model uncertainty) into the kriging system which is then used to predict the target soil variable. This reduces the extrema of the target variable and therefore produces smoother function of the predicted values. Odeh et al. (1995), while found kriging with uncertainty as not good as regression-kriging, nevertheless reported it to be better than ordinary kriging or cokriging alone.

D) - **Kriging with external drift** is a somewhat different hybrid technique which integrates the universality conditions into the kriging system using one or more of the ancillary drift variables (Wackernagel, 1995). These variables could be digitised covariate derived from DEM, or rainfall data or scanned images. The universality conditions need to be known not only at the sampled locations but also at the prediction locations. This hybrid approach to spatial prediction has not been widely used in soil science, but as remotely sensed data become more readily available, it may well be the method to be used along with regression-kriging (Odeh and McBratney, 1998), especially for soil inventory at the regional/catchment (10 m - 1 km) scale.

E) - **Factorial kriging** was recently introduced to soil science (Wackernagel, 1988; Goovaerts, 1992). The method involves a combination of classical multivariate analysis and geostatistics in which multivariate variogram modelling, **principal component analysis** and cokriging are carried out on a multitude of soil variables. The assumption behind factorial kriging is that many of the soil variables have the same communality (as defined by soil forming processes influenced by the CLORPT factors) that enables principal component (or its variants) analysis of the variance-covariance matrices of the variables which are themselves associated with spatial scales (Goovaerts, 1992). Prior to this, McBratney and Webster (1981) transformed their soil data by principal component analysis before embarking on spatial analysis- a form of factorial kriging. Also closely related to factorial kriging is the application of **fuzzy set theory** for classification of the soil into continuous classes. Various combinations of fuzzy logic and classification with kriging have been adopted for soil mapping (see examples in de Gruijter et al., 1997). New methods involving **fuzzy-kriging integral** and fuzzy
inference (e.g. Pham, 1997) are now emerging that could prove useful for optimal spatial prediction of the soil variables. A major problem with factorial kriging is the linearity assumption which is often not met by many soil variables, but the problem can largely be solved by some transformations or by using correspondence analysis or even by using fuzzy integral as mentioned above. In general, factorial kriging and indeed many of the multivariate statistical techniques, are mainly exploratory tools for revealing the correlation structure of the soil variables but are not spatial prediction techniques per se.

Table 2. A comparison of pedometric surveys at different scales and pixel resolutions

<table>
<thead>
<tr>
<th>Areal extent of survey</th>
<th>National or Continental</th>
<th>Regional or Catchment</th>
<th>Field or Farm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal linear scale</td>
<td>1: 5 x 10^6 to 1: 250,000</td>
<td>1:250,000 to 1:10,000</td>
<td>1: &lt;10,000</td>
</tr>
<tr>
<td>Pixel resolution</td>
<td>&gt; 1 km</td>
<td>10 m to 1 km</td>
<td>&lt; 10 m</td>
</tr>
<tr>
<td>Purpose</td>
<td>National planning</td>
<td>Regional or catchment management, etc.</td>
<td>Precision farming, soil contamination</td>
</tr>
<tr>
<td>Ancillary data source</td>
<td>NOAA AVHRR</td>
<td>Landsat, SPOT, Gamma radiometer, Air-borne EM Land resources maps</td>
<td>Air-borne videography, Proximal sensors</td>
</tr>
<tr>
<td>Deterministicity/</td>
<td>Deterministic</td>
<td>Deterministic or deterministic-stochastic</td>
<td>Mainly stochastic, could be both</td>
</tr>
<tr>
<td>Stochasticity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best pedometric</td>
<td>Mostly CLORPT</td>
<td>Mostly hybrid</td>
<td>Mainly geostatistics, may be hybrid</td>
</tr>
<tr>
<td>techniques</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Map precision (% of range value of the target variable)</td>
<td>15-20</td>
<td>10-15</td>
<td>&lt; 10</td>
</tr>
</tbody>
</table>

IV - Soil data interpolation at different scales and pixel resolutions

Application of each pedometric techniques depends on the purpose and the scale of the survey as the ultimate use of soil survey information determines the accuracy required. Different techniques produce different error of interpolation. For example, the accuracy of soil information required for total catchment management is clearly different from accuracy required for national planning (Table 2). In most cases as the purposes are different, the risk of taken wrong decisions due to survey error are also different. Therefore the pedometric techniques described above cannot just be applied to any situations without consideration of the specific needs and appropriateness of the inherent assumptions of the techniques. Table 2 presents a summary of examples as a guide for selecting the best pedometric technique, given the purpose of the soil survey, the precision required, the scale of the survey and the final pixel resolutions of the resulting thematic maps. These techniques need to be incorporated into the mainstream GIS packages for appropriate (geo)statistical analysis prior to GIS operations for land/soil
quality analysis, soil contamination and pollution studies, various decision maps (including precision agriculture) and total environmental management.

V - References


Keywords: Soil geostatistics, spatial prediction, generalised additive models, regression trees, neural networks.

Mots clés: géostatistique, prédiction spatiale, modèles additifs, arbres de régression, réseau de neurones.