EVALUATION OF SPATIAL INTERPOLATION TECHNIQUES
FOR MAPPING AGRICULTURAL TOPSOIL PROPERTIES IN CRETE

Christos G. Karydas[1], Ioannis Z. Gitas[2], Eirini Koutsogiannaki[3], Nikolaos Lydakis-Simantiris[4] and Georgios N. Silleos[5]

1. Aristotle University of Thessaloniki, Laboratory of Remote Sensing and GIS, Thessaloniki, Greece; xkarydas(at)agro.auth.gr
2. Aristotle University of Thessaloniki, Laboratory of Forest Management and Remote Sensing, Thessaloniki, Greece; igitas(at)for.auth.gr
3. Technical University of Crete, Laboratory of Water Resources Management and Coastal Engineering, Chania, Greece; irini(at)enveng.tuc.gr
4. Technological Education Institution of Crete, Department of Natural Resources and the Environment, Chania, Greece; lydakis(at)chania.teicrete.gr
5. University of Macedonia, Department of Applied Informatics, Thessaloniki, Greece; gsilleos(at)tellas.gr

ABSTRACT

The aim of this work was to evaluate prediction maps created with interpolation for five common topsoil properties, namely organic matter, total $CaCO_3$, electric conductivity, Fe content, and clay content in a Mediterranean agricultural system. Knowledge about spatial variance of soil properties is of significant importance when implementing "good farming rules" towards sustainable rural development. A 25 ha block of fields cultivated with grape vines and olive trees in the island of Crete was targeted as a study site. 106 topsoil samples were collected on a 50×50 m^2 grid and then analysed in the laboratory. Three well-known spatial interpolation techniques, namely Ordinary Kriging (OK), Inverse Distance Weighting (IDW), and Radial Basis Functions (RBF) were applied for generating the prediction maps, which then were assessed for their accuracy and effectiveness with a new, independent set of samples. This time the samples were collected randomly. The results indicated that there was not a method clearly more accurate than the other methods for any of the tested properties. The Goodness of prediction (G) had positive values for Fe content and total $CaCO_3$ for all interpolation techniques and positive values for Organic Matter only for IDW. Among G values, only the value for total $CaCO_3$ was big enough for all techniques, indicating that only this specific property counted for mapping. It is believed that the topography, crop alteration and farming practices significantly affected the results in the specific landscape. Future research will focus on taking account of between-field variations that are due to different crops or farming practices, incorporating segmentation techniques on cadastral data and satellite imagery.

INTRODUCTION

Set in power since 2006, the New Common Agricultural Policy (CAP) of the European Union has been promoting environment-friendly agriculture through the implementation of "good farming" rules. Among these rules, minimisation of fertilisation inputs to the agricultural fields is a priority target towards sustainable rural development (http://europa.eu). In this direction, knowledge on spatial variance of topsoil properties becomes a prerequisite for estimating fertilising needs throughout an area of interest. A standard method for creating maps of topsoil properties is to sample the targeted area using a grid sampling scheme, the density of which depends on the heterogeneity of the area. Then, a prediction map can be made by interpolating the measured property values of the samples. However, variance of soil properties is not only controlled by the global spatial variability throughout the targeted area, but also by between-field variance due to different crops or management practices applied by the farmers. Moreover, agricultural environments are numerous throughout the world and moreover, the spatial distribution of different soil properties depends on many factors, thus making extraction of generic rules for interpolation impossible.

Mediterranean agricultural areas are generally characterised by large spatial variation in soil properties and small field size, while farmers cannot afford expensive solutions. Therefore, choosing an optimal interpolation method for making prediction maps from a limited number of soil samples is a subject of significant importance.

Brief literature review in spatial interpolation

Interpolation is the procedure of predicting the value of attributes at un-sampled sites from measurements made at point locations within the same area. Interpolation is used to convert data from point observations to continuous fields so that the spatial patterns sampled by these measurements can be compared with spatial patterns of other spatial entities. The rationale behind spatial interpolation is the very common observation that, on average, values at points close together in space are more likely to be similar than points further apart. Among spatial interpolation methods, one can find Inverse Distance Weighting (IDW), Radial Basis Functions (RBF) and Kriging techniques (1). The IDW method is based on the assumption that the value of an attribute z at some unvisited point is a distance-weighted average of data points occurring within a neighbourhood or window surrounding the unvisited point. The unknown value is estimated by eq. (1):

$$\hat{z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i) \tag{1}$$

where $\hat{z}(s_0)$ is the estimated value for an un-sampled location s_o, n is the number of measured sample points surrounding the prediction location used for the prediction, λ_i is the weight for each measured point, and $Z(s_i)$ is the observed value at location s_i. The weights λ_i are calculated according to eq. (2):

$$\lambda_i = \frac{d_{io}^{-p}}{\sum_{i=1}^{n} d_{io}^{-p}} \tag{2}$$

where

$$\sum_{i=1}^{n} \lambda_i = 1$$

The weight is reduced by a factor of p, as the distance increases. Finally, d_{io} is the distance between the prediction location s_o and each of the measured locations s_i. Weighting of the sampled locations highly depends on the power parameter p, meaning that when distance increases the weight decreases exponentially. IDW belongs to the category of local deterministic methods of interpolation (1).

Radial Basis Functions (RBF) is a family of five deterministic exact interpolation techniques: thin-plate spline, spline with tension, completely regularised spline, multi-quadratic function and inverse multi-quadratic function. An exact interpolator is that one which predicts values identical with those measured at the same point. RBF methods predict values that can vary above the maximum or below the minimum of the measured values. For all RBF methods, there is a parameter that controls the smoothness of the resulting surface. The estimated values of the methods are based on a mathematical function that minimises overall surface curvature, generating quite smooth surfaces. The differences among them are small, so the generated surfaces are almost similar (1). A formula f, which minimises the following factor (eq. (3)) is an example of RBF technique and more specifically of the exact spline method:

$$A(f) + \sum_{i=1}^{n} w_i^2 [f(x_i) - y(x_i)]^2 \tag{3}$$

where $y(x_i)=z(x_i)+\varepsilon(x_i)$ is the source of random error, where z is the measured value of an attribute at point x_i and epsilon is the associated random error. The term $A(f)$ represents the smoothness of the function f and the second term represents its proximity to the data (1).

Starting with the recognition that the spatial variation of any continuous attribute is often too irregular to be modelled by a simple, smooth mathematical function, Kriging (after D. G. Krige) is a wide

family of interpolation methods using geostatistics. Geostatistical methods for interpolation are based on the assumption of spatial autocorrelation, which states that the distance and direction between sample points are the major factors governing the estimated values at unknown points. The function used for fitting the data generates better estimates when spatial autocorrelation exists. Among several Kriging methods there are: Ordinary, Simple, Universal, Indicator, Probability, Disjunctive Kriging and Co-Kriging and they all rely on the concept of autocorrelation. The Kriging equation (4) represents a weighted sum of the data, which is:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i Z(s_i) \tag{4}$$

where N is the number of samples used for the estimation, $Z(s_i)$ is the measured value at location i, λ_i is the unknown weight for the measured value at i^{th} location determined using the fitted variogram and $Z(s_0)$ is the predicted value at location s_0 (1).

Gotway *et al*. (1994) found that the IDW method generated more accurate results than Kriging for mapping soil organic matter and soil NO_3 levels. Wollenhaupt *et al*. (1994) compared these two interpolation techniques and concluded that IDW was more accurate for mapping P and K levels of soil, too. Mueller *et al*. (2004) observed that for the optimal parameters of the method, the accuracy of IDW interpolation generally equalled or exceeded the accuracy of Kriging at all scales of measurement (2,3). On the other hand, other researchers observed that Kriging was more accurate for the interpolation of several soil attributes. Leenaers *et al*. (1990) assessed the Kriging interpolation method as more accurate compared to IDW, for mapping soil Zn content. Kravchenko (2003) evaluated the effect of data variability and the strength of spatial correlation in the data on the performance of the grid soil sampling of different sampling density and two interpolation procedures, ordinary point Kriging and optimal Inverse Distance Weighting (IDW). Kriging with known variogram parameters performed significantly better than the IDW for most of the studied cases (4,5). Many other studies also compared Kriging, IDW, and RBF in soil science. Indicatively, Schloeder *et al*. (2001) observed that Ordinary Kriging (OK) and IDW were similarly accurate and effective methods, while thin-plate smoothing spline with tensions was not. Weller *et al*. (2002) concluded that even in cases where the assumptions for Kriging were not fulfilled the Kriging approach was as good as any other radial base function interpolation. Summarising the above, the accuracy of each method depends on the assumptions and subjective judgments that are made, such as the application of smoothness of the results or not and linearity of interpolation functions or not (6). Apart from research studies to judge the quality of a technique's performance, validation or accuracy assessment of a method is rarely done in operational applications because of its high cost (1).

Aim and objectives

The main aim of this work was to evaluate the reliability of maps of topsoil properties in a Mediterranean agricultural system created by Kriging, IDW, and RBF interpolation techniques. More specific objectives were:

- to produce maps of topsoil properties and more specifically of clay content, organic matter, total $CaCO_3$, electric conductivity and Fe content, based on a 50×50 m^2 sampling point-grid,

- to assess the accuracy and effectiveness of the topsoil maps produced with Kriging, IDW and RBF interpolation methods,

- to suggest the most appropriate technique (if any) for the examined topsoil properties in the specific site,

- to examine the sources of uncertainty in the specific landscape.

Study site

An agricultural area close to the town of Archanes and more specifically a part of the Kampos valley, located 13 km south of Heraklion city in the island of Crete (Greece) was selected as the study site. The landscape of Archanes is semimountainous; small valleys are surrounded by steep hills, while westwards mount Giouchtas dominates the area (811 m high; 3.5 km long). The climate of Archanes is typical Mediterranean, characterised by mild winters and dry summers. Seventy percent of Archanes' inhabitants are farmers, while 2,000 ha (2/3 of the total municipality land) are dedicated to agriculture (Figure 1). The two main cultivations in Archanes are olive trees and grape vines. Both are high value plants, their cultivation dating back to the Minoan civilisation era. The agricultural environment of Archanes is particularly heterogeneous, due to the facts that the mean size of the fields is small (0.65 ha), they have long and irregular shapes, the relief is quite undulating and small natural land patches interfere between cultivated zones. Access to the fields is not always easy. Only few fields are currently under good farming monitoring through agri-environmental measures. The study site in Kampos valley is covered mainly by vineyards, while the olive plantations are relatively new. The mean altitude of the study site is 380 m (Std: 13.5 m) (Figure 1). In terms of soil background the study site is covered by Rendzinas over marls, constituting mostly silty, shallow soils with a high percentage of gravels.
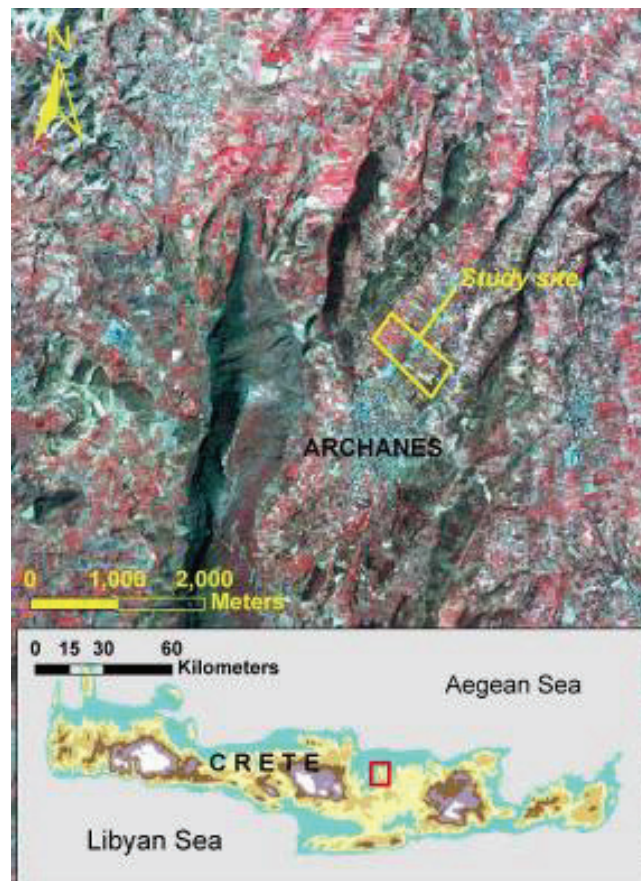

Figure 1: Archanes is located in central Crete, while the study site lies in the valley of Archanes and is mainly covered by grape vines and olive trees.

METHODS

Dataset description and methodology overview

Interpolation of soil samples over a point grid scheme was implemented, using three different interpolators for five surface soil properties. The dataset comprised soil samples collected for training the interpolator (106 samples) and for testing the interpolator (70 samples). An IKONOS satellite image acquired on 24th August 2000 available as a 4a-level Pro Orthorectified product (pan-

sharpened with a 1 m nominal pixel size) was used for identifying the agricultural system, characterising the landscape, and for locating the soil samples in the field during validation survey (collection of testing samples) (7,8). It was also, used for presentation purposes in the mapping outputs.

Soil sampling and analysis

Ideally for mapping, samples should be taken evenly over the study site. A completely regular sampling network, however, can be biased if it coincides with a regular pattern in the landscape. On the other hand, any kind of randomisation can lead to uneven distribution of the samples (1). The study site in Archanes is under the same agricultural system and moreover is smooth in terms of transition of landscape features, such as elevation, slope, aspect, soil categories, crop alteration, field size and orientation, etc. For these reasons, it was not found necessary to conduct preliminary sampling for possible stratification or to select random samples in order to record hidden heterogeneity. Consequently, the selected scheme was the grid sampling scheme, oriented to the N-S direction (Figure 2). The grid method is the most common in the literature for assessing soil properties variability and appropriate when no other information is available about soil variability prior to sampling. The selected 50×50 m^2 grid provides equally spaced observations and therefore reveals systematic variations, which were the main target of the research.
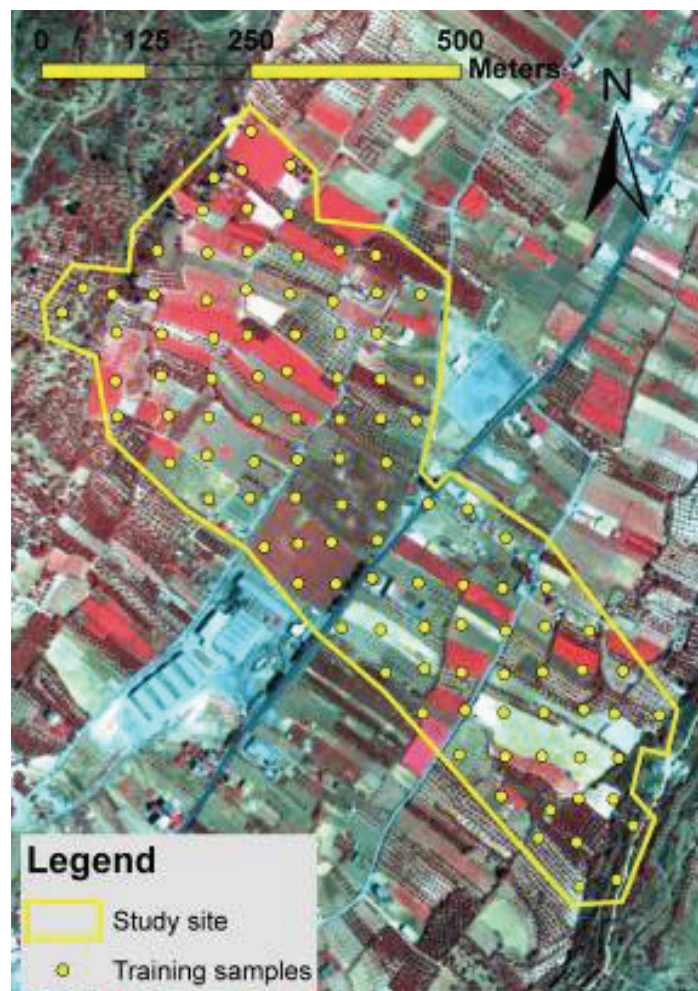

Figure 2: The grid sampling scheme selected for the study site.

The data was collected in two different time frames: The first training dataset was collected in April 2002 (82 samples) and the second training dataset in October 2002 (24 samples; total 106 samples). The sampling points (nodes of the grid) and ancillary features were overlaid the IKONOS image and an ortho-photomap was printed out in the lab. Then, a differential GPS receiver was used for locating these sampling points in the field. In a small number of cases, the sampling points deviated from the scheduled grid due to difficulties in accessing the point. Sample extraction

was implemented using a tube auger from the upper stratum (0 - 0.30 m depth, i.e. topsoil sampling). Roughly 1 - 1.5 kilograms of each sample was taken from each sampling-bag, which then was coded and exsiccated for 24 hours at 35°C. Each soil sample was pulverised and placed in closed plastic dark containers at room temperature until use. Each individual sample was analysed separately and each measurement was repeated three times for the same extract. Therefore, the final values of the measured attributes are represented by the mean value of three different measurements. Any wrong or extreme measurements were repeated and corrected.

The methods followed for soil analysis are described below (9):

- Soil texture (in order to isolate the clay content): Vouyoucos method. Fifty (50) g of sample was incubated for 16-20 h in sodium polyphosphate solution and then transferred to a 1 L volumetric cylinder. The mixture was diluted to 1 L with distilled water and its density was measured with a Vouyoucos densitometer after 40 seconds, and 2 h after the final shaking.

- Electric conductivity (mS/cm): The electric conductivity was measured with an electrode in the soil solution which was obtained by centrifugation of the soil paste.

- Organic matter (%): About 0.5 g of soil sample was oxidised by 10 ml potassium dichromate 1N. The excess of the oxidant was titrated with an iron sulfate standard solution and the percentage of organic matter was calculated.

- Total $CaCO_3$ (%): The volume of CO_2 produced by the reaction of excess HCl solution with soil was measured. The percentage of $CaCO_3$ was calculated using the relevant chemical formula.

The above soil parameters were selected among ten parameters originally measured. The other soil parameters were excluded from further processing, because exploratory data analysis indicated strong seasonal influence which could affect the final results.

Exploratory data analysis

Exploratory analysis including descriptive statistics (Table 1) and global trend of all soil properties was implemented using the 'Geostatistical Analyst' of ArcGIS 8.1 software package. Global trend is an overriding process that affects all measurements in a deterministic way. This is achieved by projecting the sample locations on an *x*, *y* plane. The value of the attribute of each sample is given in the *z* dimension. Moreover, the values of the attributes are projected on *x*, *z* and *y*, *z* planes as scatter plots. Global trend exists if a curve that is not flat (i.e., a polynomial line) can fit through the data. For the soil samples of the study area, trend analysis showed that most of the studied soil attributes had a trend except from electric conductivity and Fe content, which did not have a trend or a trend too weak to be identified (Figure 3). More specifically, for total $CaCO_3$ and clay content a directional trend was identified from NW to SE, a fact that can be attributed to the geographic characteristics of the study area (small valley surrounded by hills on the SE-NW direction) that mainly affected the measured attributes. For organic matter some trend was identified in the direction SW to NE.

Table 1: Descriptive statistics of the sampled topsoil properties.

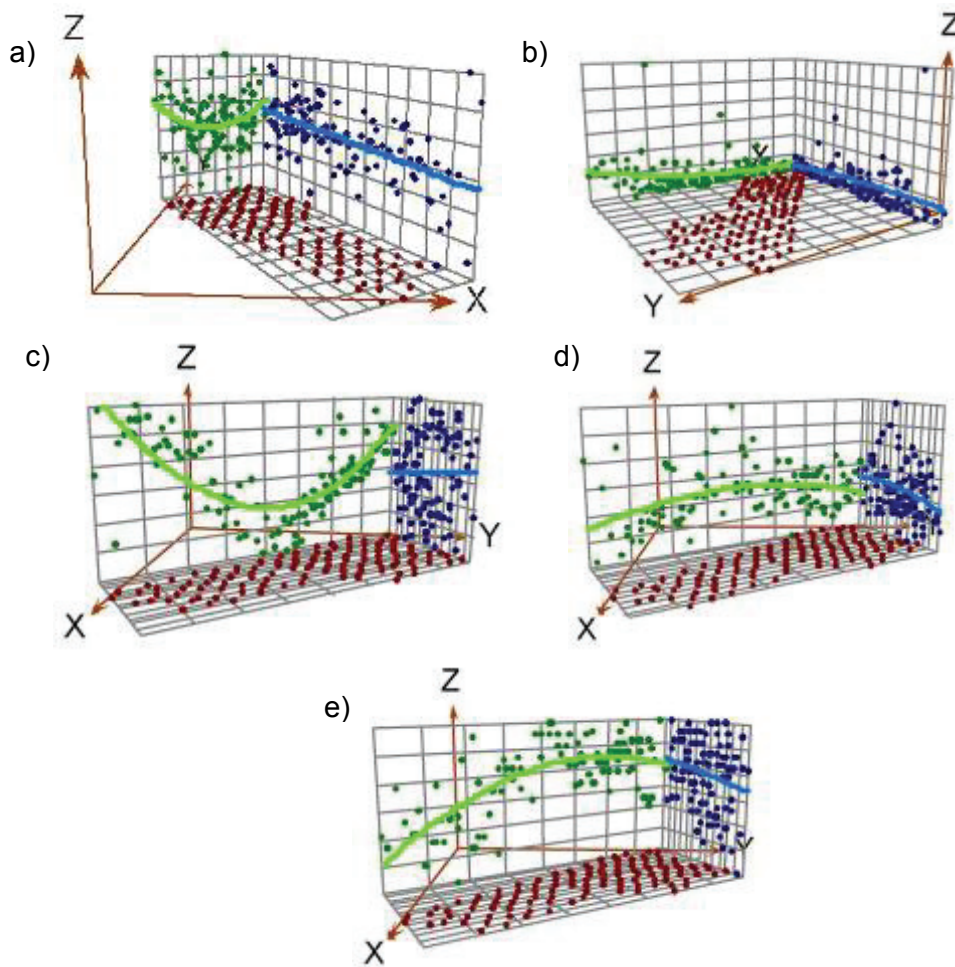| Soil parameter	Min	Max	mean	std. dev.	skewness	kurtosis	median	coeff. of var.
Organic Matter (%)	0.7	4.7	2.79	0.69	0.23	3.69	2.8	0.25
Electric conductivity (mS/cm)	0.24	2.65	0.49	0.30	4.15	27.4	0.405	0.61
Total CaCO3 (%)	9.5	67.8	38.9	15.29	-0.018	1.98	38.05	0.39
Fe content (ppm)	1.27	11.1	4.86	1.75	0.71	3.79	4.68	0.36
Clay content (%)	17	45	34.9	6.45	-0.52	2.56	37	0.18

Figure 3: Trends of soil properties a) OM, b) EC, c) CaCO3, d) Fe and e) Clay.

Interpolation

The parameters used in interpolation procedures for creating the prediction maps of Organic Matter, Electric Conductivity, CaCO₃, Fe content, and Clay content, are presented in Tables 2 - 6, respectively.

Table 2: Parameters of the interpolation methods for prediction map of Organic Matter.

Ordinary Kriging	Inverse Distance Weighting	Radial Basis Functions

Number of Points: 106
 	
Order of Trend Removal: Second (Estimated by Global Polynomial Interpolation)
 	
Semivariogram/Covariance:
  Model:
  0.18959·Spherical(888.44)
              +0.3259·Nugget
  Error modelling:
    Microstructure: 0.3259 (100%)
    Measurement error: 0 (0%)
	Method Parameter(s):
	  Power: 2
		Method Parameter(s):
		  Kernel Function: Completely Regularised Spline
		  Anisotropy:
		    Angle: 90.00
		    Ratio: 1.00
		  Parameter: 0.26047

Searching Neighbourhood:
  Neighbours to Include: 5 or at least 2 for each angular sector
  Searching Ellipse:
    Angle: 0
    Major Semiaxis: 888.44
    Minor Semiaxis: 888.44
    Angular Sectors: 4
	Searching Neighbourhood:
	  Neighbours to Include: 15 (include at least 10)
	  Searching Ellipse:
	    Angle: 0
	    Major Semiaxis: 286.92
	    Minor Semiaxis: 286.92
	    Sector Mode: 0
		Searching Neighbourhood:
		  Neighbours to Include: 15 (include at least 10)
		  Searching Ellipse:
		    Angle: 0
		    Major Semiaxis: 286.92
		    Minor Semiaxis: 286.92
		    Sector Mode: 0

Table 3: Parameters of the interpolation methods for Electric Conductivity.

Ordinary Kriging	Inverse Distance Weighting	Radial Basis Functions

Number of Points: 106
 	
Order of Trend Removal: Second (Estimated by Global Polynomial Interpolation)
 	
Semivariogram/Covariance:
  Model:
  0.002768·Spherical(888.44)
              +0.07918·Nugget
  Error modelling:
    Microstructure: 0.07918 (100%)
    Measurement error: 0 (0%)
	Method Parameter(s):
	  Power: 2
		Selected Method: Radial Basis Functions
		
		Method Parameter(s):
		  Kernel Function: Completely Regularised Spline
		  Anisotropy:
		    Angle: 90.00
		    Ratio: 1.00
		  Parameter: 0.26047

Searching Neighbourhood:
  Neighbours to Include: 5 or at least 2 for each angular sector
  Searching Ellipse:
    Angle: 0
    Major Semiaxis: 888.44
    Minor Semiaxis: 888.44
    Angular Sectors: 4
	Searching Neighbourhood:
	  Neighbours to Include: 15 (include at least 10)
	  Searching Ellipse:
	    Angle: 0
	    Major Semiaxis: 286.92
	    Minor Semiaxis: 286.92
	    Sector Mode: 0
		Searching Neighbourhood:
		  Neighbours to Include: 15 (include at least 10)
		  Searching Ellipse:
		    Angle: 0
		    Major Semiaxis: 286.92
		    Minor Semiaxis: 286.92
		    Sector Mode: 0

Table 4: Parameters of the interpolation methods for total CaCO3.

Ordinary Kriging	Inverse Distance Weighting	Radial Basis Functions

Number of Points: 106
 	
Order of Trend Removal: Second (Estimated by Global Polynomial Interpolation)
 	
Semivariogram/Covariance:
  Model:
  37.541·Spherical(888.44)
          +86.559·Nugget
  Error modelling:
    Microstructure: 86.559 (100%)
    Measurement error: 0 (0%)
	Method Parameter(s):
	  Power: 2
		Method Parameter(s):
		  Kernel Function: Completely Regularised Spline
		  Anisotropy:
		    Angle: 90.00
		    Ratio: 1.00
		  Parameter: 0.26047

Searching Neighbourhood:
  Neighbours to Include: 5 or
  at least 2 for each angular sector
  Searching Ellipse:
    Angle: 0
    Major Semiaxis: 888.44
    Minor Semiaxis: 888.44
    Angular Sectors: 4
	Searching Neighbourhood:
	  Neighbours to Include: 15
	  (include at least 10)
	  Searching Ellipse:
	    Angle: 0
	    Major Semiaxis: 286.92
	    Minor Semiaxis: 286.92
	    Sector Mode: 0
		Searching Neighbourhood:
		  Neighbours to Include: 15
		  (include at least 10)
		  Searching Ellipse:
		    Angle: 0
		    Major Semiaxis: 286.92
		    Minor Semiaxis: 286.92
		    Sector Mode: 0

Table 5: Parameters of the interpolation methods for Fe content.

Ordinary Kriging	Inverse Distance Weighting	Radial Basis Functions

Number of Points: 106
 	
Order of Trend Removal: Second (Estimated by Global Polynomial Interpolation)
 	
Semivariogram/Covariance:
  Model:
  0.39928·Spherical(888.44)
          +2.422·Nugget
  Error modelling:
    Microstructure: 2.422 (100%)
    Measurement error: 0 (0%)
	Method Parameter(s):
	  Power: 2
		Method Parameter(s):
		  Kernel Function: Completely Regularised Spline
		  Anisotropy:
		    Angle: 90.00
		    Ratio: 1.00
		  Parameter: 0.26047

Searching Neighbourhood:
  Neighbours to Include: 5 or
  at least 2 for each angular sector
  Searching Ellipse:
    Angle: 0
    Major Semiaxis: 888.44
    Minor Semiaxis: 888.44
    Angular Sectors: 4
	Searching Neighbourhood:
	  Neighbours to Include: 15
	  (include at least 10)
	  Searching Ellipse:
	    Angle: 0
	    Major Semiaxis: 286.92
	    Minor Semiaxis: 286.92
	    Sector Mode: 0
		Searching Neighbourhood:
		  Neighbours to Include: 15
		  (include at least 10)
		  Searching Ellipse:
		    Angle: 0
		    Major Semiaxis: 286.92
		    Minor Semiaxis: 286.92
		    Sector Mode: 0

Table 6: Parameters of the interpolation methods for Clay content.

Ordinary Kriging	Inverse Distance Weighting	Radial Basis Functions

Number of Points: 106
 	
Order of Trend Removal: Second (Estimated by Global Polynomial Interpolation)
 	
Semivariogram/Covariance:
  Model:
  4.5816·Spherical(888.44)
           +20.612·Nugget
  Error modelling:
    Microstructure: 20.612 (100%)
    Measurement error: 0 (0%)
	Method Parameter(s):
	  Power: 2
		Method Parameter(s):
		  Kernel Function: Completely Regularised Spline
		  Anisotropy:
		    Angle: 90.00
		    Ratio: 1.00
		  Parameter: 0.26047.

Searching Neighbourhood:
  Neighbours to Include: 5 or at least 2 for each angular sector
  Searching Ellipse:
    Angle: 0
    Major Semiaxis: 888.44
    Minor Semiaxis: 888.44
    Angular Sectors: 4
	Searching Neighbourhood:
	  Neighbours to Include: 15 (include at least 10)
	  Searching Ellipse:
	    Angle: 0
	    Major Semiaxis: 286.92
	    Minor Semiaxis: 286.92
	    Sector Mode: 0
		Searching Neighbourhood:
		  Neighbours to Include: 15 (include at least 10)
		  Searching Ellipse:
		    Angle: 0
		    Major Semiaxis: 286.92
		    Minor Semiaxis: 286.92
		    Sector Mode: 0

Measures of accuracy and effectiveness

The test of the prediction maps was based on a measure of accuracy, namely the Mean-Squared Error (MSE) and on one measure of effectiveness, namely the Goodness-of-Prediction Estimate (*G*) (3). The MSE is given by the following equation:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(\left| z(x_i) - \hat{z}(x_i) \right| \right)^2 \qquad (5)$$

where $z(x_i)$ is the observed value at location *i*, $\hat{z}(x_i)$ is the predicted value at location *i*, and *n* is the sample size. Squaring the difference at any point gives an indication of the magnitude of differences, in such a way that small MSE values indicate more accurate point-by-point predictions. The *G* measure (eq. (6)) gives an indication of how effective a prediction might be, relative to that derived from using the sample mean alone (3):

$$G = \left(1 - \frac{\sum_{i=1}^{n} \left(z(x_i) - \hat{z}(x_i) \right)^2}{\sum_{i=1}^{n} \left(z(x_i) - \overline{z} \right)^2} \right) \cdot 100 \qquad (6)$$

where $z(x_i)$ is the observed value at location *i*, $\hat{z}(x_i)$ is the predicted value at location *i*, *n* is the sample size and $\overline{z}$ is the sample mean. A G value equal to 100% indicates a perfect prediction, positive values (i.e. from 0 to 100%) indicate that the predictions are more reliable than the use of the sample mean, and negative values indicate that the predictions are less reliable than the use of the sample mean instead.

RESULTS

Prediction maps of topsoil properties

Fifteen prediction maps were created for the topsoil properties previously mentioned, using the Ordinary Kriging (OK), the IDW, and the RBF methods. As an example, the prediction maps of Organic Matter for the tested interpolation methods are presented in Figure 4.
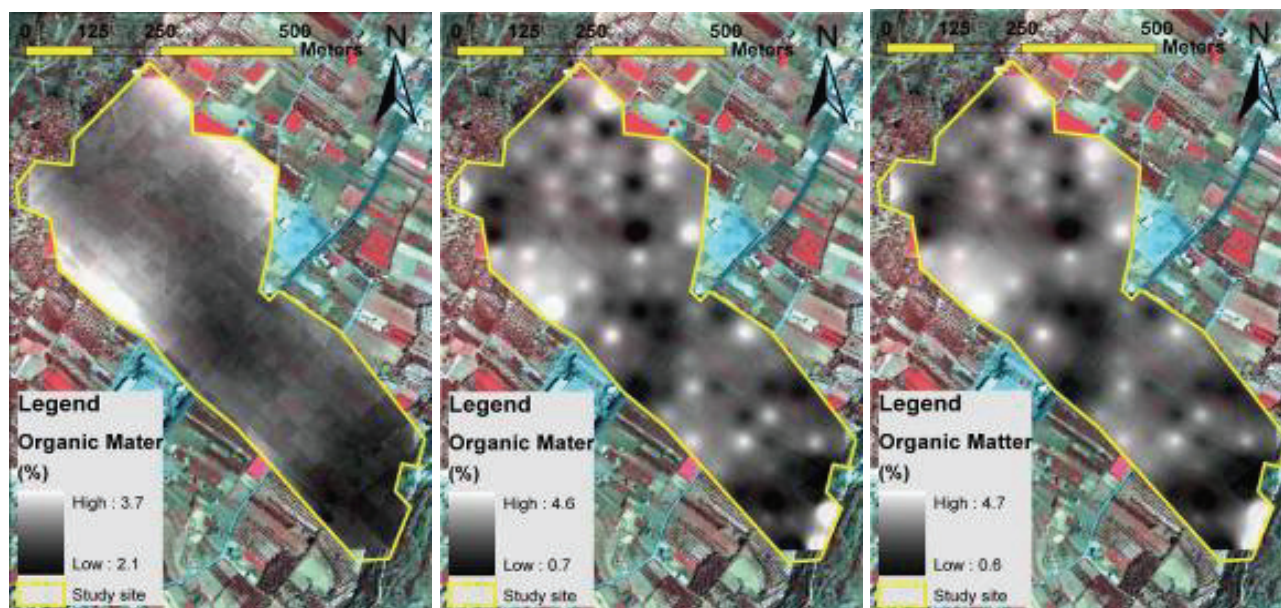
Figure 4: Organic Matter prediction map by the Ordinary Kriging (left side), IDW (centre), and RBF method (right side), respectively.

Evaluation of the prediction maps

Accuracy and effectiveness assessment of the prediction maps was conducted using an independent set of 70 soil samples, i.e. the maps were validated and not cross-validated. This set was collected in October 2002 together with the complementary (second) training dataset. Sampling for creating the testing set was random and scheduled during the field survey. Soil analysis of the testing set followed exactly the same methodologies as the analysis for the training set. The predicted values at the sampled locations for testing were recorded in the lab by identifying the points on the map (Figure 5). The Mean-Squared Error (MSE) and the Goodness of prediction (G%) were calculated as measures for accuracy and effectiveness respectively for all the produced topsoil prediction maps (Table 7). The comparison of MSE values between the methods shows for a specific property that the smallest errors are achieved by different methods. In other words, there is no interpolation method that can be considered clearly better than the others. The Goodness of prediction shows positive values only for Fe content and total $CaCO_3$ for all interpolation techniques and slightly positive values only for Organic Matter for IDW. Electric Conductivity and Clay content show negative values in all tested interpolation techniques.

Table 7: MSE and G of the applied interpolation methods per topsoil property.

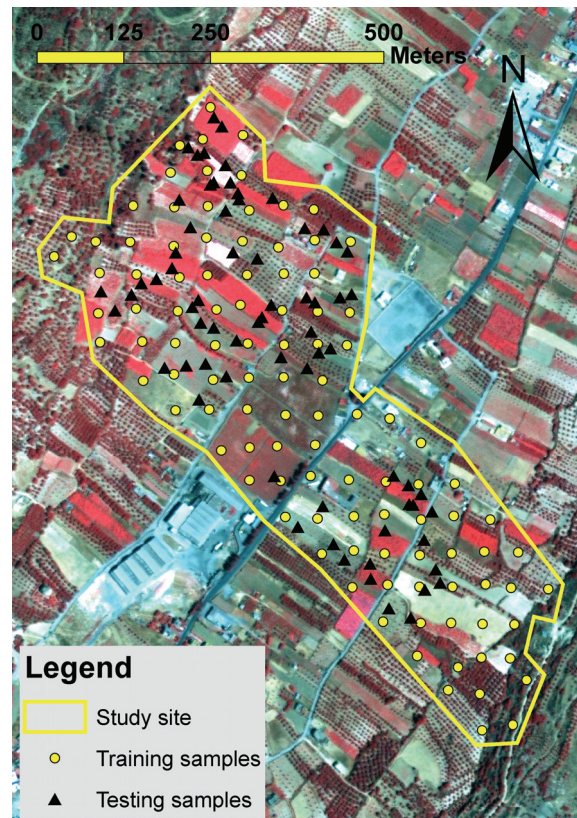| Soil Property		Interpolation methods					
	Mean Value	IDW		RBF		OK	
Organic Matter	2.87%	MSE	G (%)	MSE	G (%)	MSE	G (%)
		0.2	1.54	0.24	-17.8	0.23	-13.4
Clay Content	39.46%	MSE	G (%)	MSE	G (%)	MSE	G (%)
		25.49	-38.1	23.29	-26.2	25.29	-37.1
Electric Conductivity	0.79 mS/cm	MSE	G (%)	MSE	G (%)	MSE	G (%)
		0.19	-73.9	0.21	-91.2	0.19	-73.8
Fe Content	4.67 ppm	MSE	G (%)	MSE	G (%)	MSE	G (%)
		1.88	4.97	1.96	1.00	1.91	3.64
Total CaCO3	35.33 %	MSE	G (%)	MSE	G (%)	MSE	G (%)
		88.6	22.67	86.4	24.58	84.77	26.01

Figure 5: Sampling for testing the accuracy and effectiveness of the prediction maps.

Discussion

With regard to the specific properties, Electric Conductivity (EC) was the only soil attribute for which effectiveness G had a high negative value, signifying that the prediction would have been more reliable if the sample mean had been used instead. Furthermore, in the exploratory analysis this attribute did not reveal any kind of trend and it was impossible to detect any spatial structure. Therefore, it should have been excluded in advance from further analysis. A possible reason for this behaviour is that EC changes rapidly with time as it is affected by soluble salts. The measure of effectiveness G had negative values for Clay content and Organic Matter (in OK and RBF methods) as well. The rest of the soil properties (total $CaCO_3$ and Fe content) had positive G values (though not very high), indicating that the use of interpolation techniques was suitable for mapping these attributes. When comparing errors for each of the maps, the IDW method for Organic Matter and Fe content proved to be more accurate. The RBF method produced more accurate maps for Clay content, whereas the OK technique was best for mapping total $CaCO_3$. However, the differences between the methods for a specific property were not significant.

Concerning the source of errors and uncertainties of the prediction maps in this study, these can be attributed to the following random or systematic reasons:

- The difference in data collection time: As it was described earlier, the data of a part of the training samples was not collected in the same period of time as the main part of the set. This fact possibly affected the results given that soil properties vary over time. Also, the entire testing sample set was collected at a different time than that of the main training dataset.

- The long shape of the study site: The collected soil samples covered an irregular area with 400 metres maximum width and 900 metres length. In the SE side however, the width was only 250 m. If the area had been a square of 600×600 metres, some of the soil attributes might have better revealed their spatial variability and dispersion in both *x* and *y* dimensions.

- The choice of values for Kriging parameters is subjective by default and encloses semantic sources of errors in the final resulting surfaces. Therefore, a different parameter selection might lead to different results.

CONCLUSIONS

In this study, the applied interpolation methods gave similar results in terms of accuracy, without any of them being clearly better than the other methods. This is generally expected when the data-set used is abundant (50×50 m^2), while in a sparse dataset, the assumptions made about the underlying variation and the choice of method and its parameters could be critical (1). The fact that the sampling scheme was regular, i.e. the entire study site was covered with a standard density, excludes the possibility for errors due to the sampling scheme. Moreover, the sampled area included two similar glacises in terms of topography (two sides of the Kampos valley). In terms of effectiveness, only CaCO$_3$ counted to be mapped by any interpolation technique, while for the rest of the properties, mapping could not offer more than the use of mean values.

Fragmentation of the study site resulting in lack of land homogeneity is believed to be the reason for observing high divergences of soil properties, within relatively small distances. This is due to the fact that farmers are used to adopting different ways of management, such as fertilisation, irrigation, or soil tillage. Agricultural fields are surfaces with attributes continuously varying in space though within discrete boundaries. Therefore, their within-field variance is not only controlled by the scale of spatial variations of the surface properties, but also by variance due to crop alteration (10). Future work will use segmentation techniques in order to take account of the different cultivations over space (derived from cadastral data) and local vegetation heterogeneity (vegetation indices derived from updated very high resolution satellite imagery) into the topsoil mapping process originally based on interpolation methods.

ACKNOWLEDGEMENTS

REFERENCES

1   Burrough P A & R A Mcdonnell, 1998. Principles of Geographic Information Systems (Oxford University Press) 356 pp.

2   Mueller T G, N B Pusuluri, K K Mathias, P L Cornelius, R I Barnhisel & S A Shearer, 2004. Map Quality for Ordinary Kriging and Inverse Distance Weighted Interpolation. Soil Science Society of America Journal, 68: 2042-2047

3   Krivoruchko K & C A Gotay, 2003. Using spatial statistics in GIS. In: International Congress on Modelling and Simulation, edited by D A Post, 713-736

4   Kravchenko A & D Bullock 1999. A comparative study of interpolation methods for mapping soil properties. Agronomy Journal, 91: 393-400

5   Kravchenko A N, 2003. Influence of spatial structure on accuracy of interpolation methods. Soil Science Society of America Journal, 67: 1564-1571

6   Schloeder C A, N E Zimmerman & M J Jacobs, 2001. Comparison of methods for interpolating soil properties using limited data. Soil Science Society of America Journal, 65: 470-479

7   Silleos N G, 1990. Mapping and Evaluation of Agricultural lands (Giahoudi-Giapouli) 265 pp.

8   Gitas I Z, C G Karydas & G V Kazakis, 2003. Land cover mapping of Mediterranean land-scapes, using SPOT4-Xi and IKONOS imagery - A preliminary investigation. Options Mediter-raneennes, Series B: 27-41

9   Brady N G & Weil R R, 2008. The Nature and Properties of Soil (Prentice Hall PTR) 965 pp.

10 Friedl M A, 1997. Examining the Effects of Sensor Resolution and Sub-Pixel Heterogeneity on Spectral Vegetation Indices: Implications for Biophysical Modelling. In: Scale in Remote Sensing and GIS, edited by D A Quatrochi & M F Goodchild (Lewis Publishers, Boca Raton, Florida) 113-139