Evaluation of three spatial interpolation methods to estimate forest volume in the municipal forest of the Greek island Skyros

Dionissios P. KALIVAS*, Vassiliki J. KOLLIAS and Ektoras H. APOSTOLIDIS

Laboratory of Soils and Agricultural Chemistry, Agricultural University of Athens, Botanikos, 75, Iera Odos, 118 55, Greece

(Received 11 March 2011; final version received 12 July 2011)

Forest volume is of great interest to forest managers. Plot observations of forest volume are available from planning reports. For managers, however, what is relevant are forest volume surfaces. In this study, three interpolation methods were employed to construct continuous surfaces for the evaluation of forest volume at positions for which no measurements are available. For the Municipal Forest of Skyros Island, we compared spatial predictions derived from Inverse Distance Weighting (IDW), Block Kriging (BK), and Block Co-Kriging (BCK) interpolation methods, as applied to data from 120 sample plots. The existence of spatial autocorrelation in the data was identified using correlograms of Moran’s I index. Spatial outliers were identified and excluded from the analysis using the local Moran’s I index. Only slope, of the examined environmental factors, showed an acceptable correlation with forest volume and was used as an auxiliary variable for BCK. The performance of the three methods was evaluated, using an independent validation set and comparing these indices: mean error (ME) and root mean square error. Additionally, for BK and BCK, the indices, standardized ME, average standard error, and the standardized root-mean-square error were used. For the three interpolation methods, the BK method gave the more accurate results; BCK is the next more accurate method and IDW the third.

Keywords: forest modeling; spatial prediction; autocorrelation; geostatistics

1. Introduction

The structure of a forest is a complicated phenomenon. It is the outcome of a number of factors, such as the geomorphology of the region, soil type and depth, rainfall levels, various natural disturbances (e.g. snow, fire, and wind), the competition between various species, and the activity of pathogenous organisms. In addition, forestry-related activities, both legal and illegal, such as clearing, reforestation, and animal breeding can have a significant impact on the structure of a forest.

Biophysical factors, such as altitude, aspect, slope gradient, and soil morphology, have also an indirect influence on the development of the forest species of a certain area. This influence is affected through climatic and soil factors (photoclimate, humidity, and temperature for air and soil, soil genesis, etc.) which, in turn, transform the ecological environment.

Forest volume is the attribute of greatest interest to forest managers. It is defined as the wood volume of a stand of trees, a group of stands or an entire forest. It is the total volume of all existing trees growing in a specific area at the recording time. Forest volume is measured in cubic meters of wood volume (with or without bark) per area unit (m³/ha).

The traditional ways in use today for the calculation of forest volume are based on the division of a forest into stands and sections, according to various morphological and physiographic characteristics (roads, streams, forest species, etc.). The evaluation of their volume and growth characteristics is achieved using sample plots and by applying the average measurements of these areas to the total area for each stand or section. This method is time-consuming and calculation-intensive. Furthermore, it deals in small, discreet segment rather than considering the forest as one interrelated whole as it would be too expensive to inventory all parts of a forest, and cannot create continuous surfaces of different forest characteristics. By using interpolation methods to sample plots that have been inventoried, a continuum of estimated variables can be developed.

One interpolation method that has gained wide acceptance is Kriging (1) developed in the domain of geostatistics, and is in relatively widespread usage. Applications of geostatistical methods in forestry include estimation and mapping of forest properties such as biomass, diameter distribution, forest volume, age and forest growth, etc. (2–9).

However, very often, there exist few sampled sites in relation to the area of study. This condition can result in a considerable uncertainty in the Kriging estimates. In this sense, a combination of Kriging with an auxiliary variable (co-Kriging) can be used to enhance the

*Corresponding author. Email: kalivas@aua.gr

© 2013 Wuhan University
accuracy estimation. Kriging techniques require good judgement, experience, expertise, and effort from the users.

Another interpolation method is the Inverse Distance Weighting (IDW) which is commonly employed in forest inventories \((10-11)\). IDW is a nongeostatistical method, easy and quick to use, comparing to kriging, since it does not require the initial step of variogram estimation. The main drawback is that it does not provide a variance estimate at each interpolated value as kriging does.

The fundamental concept in geography is that nearby entities often share more similarities than entities which are far apart \((12)\). Spatial autocorrelation is a general property of variables measured across geographic space and it is observed when the value of a variable at any point in space is dependent on values at neighboring points. Spatial autocorrelation statistics measure and analyze the degree of dependency among observations. The problem of spatial autocorrelation which is present to all ecological communities or even in any geographical phenomenon can be of great interest in forest inventories as it can extend the local results. Therefore, in order to conduct spatial interpolation on forest attributes, those must be positively spatially autocorrelated. If the sampling schema does not detect spatial autocorrelation positively, it is not appropriate to conduct spatial interpolation among sample data \((2)\).

The above interpolation methods create continuous surfaces from sample points. However, our data consist of sample plots which cannot be treated as points. In that case, block interpolation methods must be used \((13)\).

The main objective of this study is the implementation and the comparison of three spatial interpolation methods and the construction of continuous surfaces for the evaluation of forest volume in blocks for which no measurements were available. Additionally, we examined the existence of the spatial autocorrelation and spatial outliers in the data-set. Also, several environmental factors affecting the development of the forest were studied.

2. Materials and methods

2.1. Study area

Skyros is an island in the central Aegean Sea. It is the largest of the Northern Sporades complex with a total area of 21,000 ha. The island is made up of two major mountainous areas, divided from each other by a long, narrow strip of land (Figure 1(a)). The capital of the island is the village of Chora.

The forest of Skyros is situated in the north-northwestern part of the island (Figure 1(a)). Its area is approximately 1/3 of the island. The forest is a valuable natural resource, especially compared to other islands of the Aegean Sea, where similar forests are very rare. Its importance is primarily biological-ecological, rather than financial.

The forest is made up primarily by pine trees \((Pinus Halepensis)\), with the presence of evergreen sclerophilious shrubland (maquis), either as understory or as the dominant vegetation.

The study area is covered, to the largest extent, by forest, 62% are thick forest (with a density of over 30%) and 15% is thin forest (partially covered forest area with forest density below 30%). The partially covered forest areas in Skyros are pine tree forests \((P. Halepensis)\). They are to be found in low-quality positions, throughout the forest, usually areas with a high slope and shallow, rocky soils.

The slopes in the forest are mostly mild to medium. Higher slopes can be found in the northern, western, and southern part of the forest. The altitude of the study area varies from sea level to 413 m. The orientation is northern-northwesterly, but all aspects can be found in the interior of the forest.

There are a number of small streams in the study area, springing from the mountainous central part of the forest and flowing into the sea. The majority of the forest streams flow in the winter months.

The most common minerals in the forest are limestone, lime slate, and lime dolomite marbles.

2.2. Data collection

The main data source of this study was the Planning Report for the forest of Skyros for the years 2007–2016 \((14)\).

For the collection of data, the study area was divided into stands according to various morphological and physiographic characteristics such as roads, streams, forest species, etc. For each stand, a number of sample plots were designed. Data from the sample plots were collected between November 2007 and February 2008 in forest-covered areas. Each sample plot was of rectangular shape \((40 \times 50 \text{ m})\), with the longer side, in most cases, parallel to the contours. A total of 120 sample plots were collected, at 0.2 ha each. So, the total sampling area was 24 ha, or about 0.6% of the forest-covered area of the Skyros forest. Every effort was made to ensure that the sample plots covered the whole area in a random way. This was not always possible, however, due to the absence of a forest road network in many places and the abrupt terrain profile.

The trees of each sample plot were divided into three classes, according to their trunk diameters at breast height (DBH). The three classes are shown in Table 1.

In every sample plot, the thickness of all trees with diameter equal or greater than 8 cm was measured. Based on these measurements, the average test trunks for every one of the three diameter classes were determined within the study area. For the average trunks, we measured the height and the thickness of the bark at a height of approximately 1.30 m.

The calculation of volume data for the pine trees of each sample plot were carried out using the form factor \(f = V/W\) \((15)\). The form factor \(f\) expresses the relationship between the real volume \(V\) of a tree to the volume of a comparable cylinder \(W\).
As mentioned, in every sample plot, there was a complete measurement of thickness, which was used to determine the average test trunks per diameter class. The height of these trunks was also measured. For the calculation of volume, we used the form factors that the Greek Ministry of Agricultural Development and Food proposes for the planning reports for the *P. halepensis*: 0.41 for I Diameter Class, 0.40 for II Diameter Class, and 0.385 for III Diameter Class. As has also been mentioned, the bark thickness of each average sample trunk was also measured, in order to determine the volume, excluding bark.

From each average sample trunk of each diameter class, we calculated the wood volume for each class separately. To this, we added the volume of the wood of the branches, at 5% for I Diameter Class and 10% for the other two. The sum of the volume of the wood of the trunk and the wood of the branches for each class gave us the total forest volume for each sample plot.

Table 1. The three DBH classes.

| Diameter class | DBH (cm) |
|----------------|----------|
| I              | 8–20     |
| II             | 22–34    |
| III            | >36      |

Figure 1. Land uses and sampling sites of the study area.
2.3. Geographical Information Systems (GIS) database

A digital map of the 120 sample plots was created and rectified under the Hellenic Geodetic Reference System (HGRS 87). Additionally, a point file was created with the coordinates of the center of each plot using GPS compatible with HGRS 87 coordinate system (Figure 1(b)). A unique code for each sample plot was inserted in the database of the point file, based on the stand and section it belongs to, as well as its forest volume. For the creation of land use map of the study area (Figure 1(b)), we used the digital orthophotos of the Ministry of Agricultural Development and Food (at a scale 1:5000). Six different categories of forest land uses were identified (Table 2). Additionally, the forest road network was digitized from the orthophotos.

From the digitized contours (20 m) of the topographical diagrams of the Hellenic Military Geographical Service at a scale 1:5000, a Digital Elevation Model (DEM) of the area was created. From the DEM, the slope and the aspect maps of the area were extracted. We also digitized the coastline and the hydrographical network of the forest area from the topographical diagrams.

A digital map was developed using the geological map of Skyros Island of the Institute of Geological and Mineral Research, of a scale of 1:20,000. All the maps were georeferenced under the HGRS 87 coordinate system.

2.4. Factors influencing forest volume

Factors influencing forest development and, therefore, forest volume are complex and often difficult to model and quantify, as mentioned previously. Especially in Mediterranean ecosystems (such as Greece’s), the situation is further complicated by frequent, destructive summer forest fires, intense and haphazard human intervention in the forest as well as the complexity and the variety of the ecosystems themselves.

In the present study, we have attempted to examine a number of parameters linked to forest development. We examined elevation, slope and aspect, and geological formations and distances from streams and road network. Distance from road network was taken into consideration as a separate factor, in an attempt to model human intervention, based on the principle that the closer a given area is to the road network, the easier and more likely are human access and intervention to this area.

Table 2. The land uses of the study area.

| Land use classes                  | Area (ha) |
|----------------------------------|-----------|
| Forest-covered areas             | 4220      |
| Partially forest-covered areas   | 1049      |
| Fields                           | 619       |
| Bushes                           | 874       |
| Grazing pastures                 | 60        |
| Residences, quarries, rubbish heap | 23   |
| Total area                       | 6845      |

Two factors that were not examined, although they tend to be highly influential in ecosystem development, were rainfall and soil depth. In the first case, we considered that the forest area (smooth and small in size) made any significant differences in rainfall highly unlikely. In the second case, there were no relevant data, as no complete soil map of the island of Skyros exists.

2.5. Autocorrelation

The spatial autocorrelation problem violates standard statistical techniques that assume independence among observations. In forest environment, it can be expected that spatial autocorrelation will be present at various scales. It can be described using correlograms and variograms (16). Correlograms are preferable over variograms for two main reasons: (a) the significance of the correlation coefficient can be tested and (b) correlograms are standardized, so different cases can be compared (16, 17). This study uses correlograms of the Moran’s I coefficient for the examination of autocorrelation. Computation of Moran’s I is achieved by the division of the spatial covariation by the total variation. It is calculated for n values of a variable y and the locations i, h from the following formula:

\[ I = \frac{1}{W} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (y_i - \bar{y})(y_j - \bar{y}) \]

where \( y_i \)'s and \( y_j \)'s are the observed values of the variable y at sites i and j, respectively, \( y \) is the mean of the variable y, and \( w_{ij} \) are the weights given from a weight matrix according to the relationship between the value of a variable and its neighborhood. \( W \) is the sum of the weights for a given distance (18). Moran’s I varies from –1 to 1 although values lower than –1 and higher than +1 may be obtained, with a value approaching zero in the absence of autocorrelation. Positive autocorrelation means that spatially nearby values of a variable tend to be similar while negative autocorrelation indicates that neighboring values are more dissimilar than expected by chance. Moreover, a highly significant Moran’s I hides several distinct local patterns of spatial clustering and complete spatial randomness (19). Therefore, the local Moran’s I can be used in order to study the spatial patterns of spatial association. That is, the local clusters and the spatial outliers. The spatial outliers can cause erratic behavior of variograms, so must be identified and excluded. Spatial outliers are magnitudes of a given parameter distinct from their neighbors. In this study, the local Moran’s I is used as a diagnostic for outliers in global spatial patterns.

Local Moran’s I detects local spatial autocorrelation (19) and decomposes Moran’s I into contributions for each location, \( l_i \). The sum of all \( l_i \) for all observations is proportional to Moran’s I. The local Moran’s index is given from the following formula:
\[ I_i = \frac{y_i - \bar{y}}{\sigma^2} \sum_{h=1,h \neq i}^n [w_{ih}(y_h - \bar{y})] \]  

(2)

where \( \sigma^2 \) is the variance of variable \( y \) and \( w_{ih} \) is a distance weight between locations \( h \) and \( i \). In this study, the inverse distance of the two locations was used. A high negative value of local Moran’s \( I \) implies that the sample plot under study is a spatial outlier, while a high positive value shows that the sample plot under study has similar values as its neighbors.

If the distribution of the data is not heavily skewed, the local Moran’s \( I \) can be standardized to a z-score so that its significance level can be tested on an assumption of normal distribution (20). The threshold of a z-score is usually defined as 1.65 or 1.96 which is equivalent to the significance level of 0.1 and 0.05, respectively, in a two-tailed test. In this study, the significance level of 0.05 was used. For 95% confidence level, if the z-score value of local Moran’s \( I \) is smaller than \(-1.96\), the sample plot is considered as spatial outlier. If the z-score value of the local Moran’s \( I \) is greater than \(1.96\), the sample plot under study is clustered with the neighbor sample plots.

2.6. Prediction methods

Since in-depth discussions about interpolation techniques are given by Journel and Huijbregts, (21) Isaaks and Srivastava (22), and Burrough and McDonnell (23), only an outline of the interpolation methods used will be given here.

The values of the forest volume of the sample plots were used for the prediction of values at unknown points using the interpolation methods: (a) IDW, (b) Block Kriging, and (c) Block co-Kriging.

2.6.1. Model IDW

IDW is based on the assumption that the nearby values contribute more to the interpolated values than distant observations. In other words, for this method, the influence of a known data point is inversely related to the distance from the unknown location that is being estimated.

The IDW interpolation estimate is a linear combination of the observed values, inversely weighted by the distances of the observation locations from the interpolation point.

The general IDW spatial model is as follows:

The spatial prediction of the values of a variable \( Z \) at an unsampled point \( x_0 \) is given by:

\[ \hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \]  

(3)

\[ \sum_{i=1}^n \lambda_i = 1 \]  

(4)

\[ \lambda_i = \frac{d_i^n}{\sum_i d_i^n} \]  

(5)

where \( x \) denotes the set of spatial coordinates \( \{x_1, x_2\} \), \( \lambda_i \) are the weights associated with the sampling points \( X_i \) and the \( i \)th observation point.

2.6.2. Block Kriging (model BK)

The most commonly used type of Kriging is point Kriging. The spatial prediction of the values of a variable \( Z \) at an unsampled point \( x_0 \) is given by Equation (3) except that rather than using weights based on an arbitrary function of distance, the weights used in Kriging are based on the model variogram. To ensure that the prediction is unbiased, the weights placed on each neighboring point must satisfy the Equation (4).

BK is similar to the point Kriging. BK is used in cases where the sampling support is an area. The point Kriging equations are modified to estimate an average value of the variable \( Z \) over a block of area \( B \). The average value of \( Z(B) \) is estimated by the Equation (6) (23).

\[ \hat{Z}(B) = \sum_{i=1}^n \lambda_i Z(x_i) \text{ with } \sum_{i=1}^n \lambda_i = 1 \]  

(6)

2.6.3. Block Co-Kriging (model BCK)

Co-Kriging is a geostatistical technique developed to improve the estimation of a variable using the information of other spatially correlated variables which are generally more densely sampled. The variables are called co-regionalized and are spatially dependent (24).

We consider that \( U \) is an expensive to measure and, therefore, undersampled variable and \( V \) is a cheap to measure variable with more observations. If \( U \) and \( V \) are spatially and mutually correlated, then it may be possible to use the spatial variation of \( V \) (or \( V_k \)) to help the mapping of \( U \). A co-kriged prediction is a weighted average in which the value of \( U \) at location \( x_0 \) is estimated as a linear weighted sum of co-variables \( V_k \) (23).

\[ \hat{U}(x_0) = \sum_{k=1}^V \lambda_{i,k} Z(x_k) \text{ for all } V_k \]  

(7)

To avoid bias, the weights \( \lambda_{i,k} \) must sum as follows

\[ \sum_{i=1}^n \lambda_{i,k} = 1 \text{ for } U = V \text{ and } \sum_{i=1}^n \lambda_{i,k} = 0 \text{ for } U \neq V \]  

(8)

Co-Kriging is the method having the best theoretical foundation, meaning that no assumptions are made on the nature of the correlation between the two variables. It exploits more fully the auxiliary information by directly incorporating the values of the auxiliary variable and measuring the degree of spatial association with the primary variable through the cross-variogram. The calculation of the cross-variogram and the fitting of a
theoretical model become very difficult, particularly when the two variables are not strongly correlated.

BCK, as in block Kriging, estimates the average value of the variable $Z$ over a block $B$. If $n_1$ is the number of sample points of variable $U$ in the block $B$ and $n_2$ is the number of sample points of variable $V$ in the same block, the

$$
\hat{Z}_U(B) = \sum_{k=1}^{2} \sum_{i=1}^{n_1} \lambda_{ik}(B)Z(x_k) \quad (9)
$$

To avoid bias the weights $\lambda_{ik}$ must sum as follows

$$
\sum_{i=1}^{n_1} \lambda_{ik}(B) = 1 \quad \text{and} \quad \sum_{i=1}^{n_2} \lambda_{ik}(B) = 0 \quad (10)
$$

The constraint $\sum_{i=1}^{n_2} \lambda_{ik}(B) = 1$ implies that not all the weights of the variable $U$ are zero and at least one sampling point must be used for the estimation of each block (12).

2.7. Variograms

The first step in geostatistical interpolation methods is to construct a variogram from the point set to be interpolated. Actually, we construct two variograms; (a) the experimental variogram and (b) the theoretical variogram. The experimental variogram is a graph relating the variance of the difference in value of a variable at pairs of sample points to the separation distance between those pairs. The theoretical variogram is a mathematical function that models the trend in the experimental variogram.

The type of the theoretical model which fitted best to the experimental variogram was selected for further geostatistical analysis.

The cross-variogram is used to describe the correlation of different variables. It allows computing a related empirical function in order to model which can then be used in the co-Kriging equations in lieu of the cross-variogram (25).

The simple variogram is a function having always positive or null values since it is a form of variance. The cross-variogram can take positive, negative or null values since it is a form of covariance. A cross-variogram will be positive when the values of the two variables have the tendency to vary jointly. It will be negative when the values of the two variables have tendency to vary in opposite directions. It will be null when the two variables tend to vary independently.

2.8. Criteria of evaluation of the spatial interpolation results

For the evaluation of the results of the spatial interpolation methods and the determination of the accuracy of model predictions in unknown areas, we used the independent validation set to compare predicted values with real values and to assess the accuracy of predictions.

The statistical parameters that were calculated for comparison between all interpolation methods were Mean Error (ME) and the Root-Mean-Square Error (RMSE).

Additionally, for the block Kriging and block co-Kriging methods, we calculated the Standardized Mean Error (SME), the Average Standard Error (ASE), and the Standardized Root-Mean-Square Error (SRMSE).

The ME measures the bias of the prediction and should be close to zero for unbiased methods. It indicates whether the model is, on average, producing estimates that are overestimating or underestimating the observed values. In a well-adapted model, ME and SME should be close to zero for unbiased methods. The RMSE measure the average precision of the prediction and should be as small as possible. The model that performs the best will be the one with the smallest RMSE. This would suggest that the predictions are impartial and close to the respective real values. The values of ASE are used in order to evaluate the prediction divergence from real values. Therefore, ASE should be the same as RMSE, in order to evaluate the divergence of predictions correctly. If the value of the ASE is greater than that of the RMSE this suggests that the variability of the predictions is overestimated. Conversely, if the RMSE is greater than the ASE the variability of the predictions are underestimated. The values of SRMSE should be close to 1. If the SRMSE are greater than 1, then the variability of the predictions is underestimated; if the SRMSE are less than 1, the variability of the predictions is overestimated.

2.9. Software used

The descriptive statistical parameters were estimated with SPSS ver. 17. The spatial analysis, autocorrelation, and geostatistical analysis were performed with GIS software ArcGIS ver. 9.3.1 and its extension Geostatistical Analyst.

3. Results and discussion

3.1. Exploratory non spatial data analysis

The initial data-set was divided in two different in a random way using the GIS software capabilities. The first one includes 100 samples and was used for the interpolation, and the rest 20 were used for the validation of the results.

The descriptive statistics of the examined variables for the 100 sampling plots and the 20 validation plots are shown in Table 3. From the mean and standard deviation values, we conclude that the validation set has the same distribution characteristics with the 100 plots.

The statistical distribution of the forest volume data was tested using two testing methods histograms and Q-Q Plot diagrams. The histogram of forest volume plotted with normal distribution curve (Figure 2). The Q-Q Plot diagram (Figure 3) also tests the distribution of
a variable. If the variable matches the normal distribution the sample plots are clustered on a straight line. Both figures (Figures 2 and 3) show that the distribution of the values of the forest volume that is very close to being normal.

We then examined if there is any linear correlation between forest volume and the rest of the variables. That is if the change (increase or decrease) in altitude, slope, aspect, and distance, from streams and road network, has a significant impact on the values of forest volume.

From the analysis carried out using Pearson’s linear correlation coefficient (Table 4), it is apparent that there exists a statistically significant correlation between forest volumes only with respect to slope. The correlation coefficient between slope and forest volume is \(-0.31\). This means that an increase in slope leads to a reduction in forest volume, which is what we expected to find.

For the other factors under consideration, we found no statistically significant correlation with forest volume. Especially for the factors:

- Distance from streams: There exists high density of the water network within the forest. The average distance from a stream, for the forest as a whole, is 185 m and the regional beneficial microclimatic conditions could not influence forest development, to any significant degree, compared to more remote areas.
- Distance from road network: The majority of the sample plots were near the road network (average distance 99 m), so no significant differences could be found in this case either.
- Altitude and aspect: The area under consideration is, in most cases, smooth and the altitudes are low (Table 3). Therefore, no significant impact on forest development could be detected here either.

In order to, then, examine whether there exists a statistically significant difference for forest volume with respect to the geological formations we performed analysis of variance. The analysis showed that for forest volume, there is no statistically significant difference between geological formations, as the level of significance is quite high (0.697). The variety of geological formations within the forest in conjunction with the ecology of *P. Halepensis*, which is adaptable to a variety of soils, demonstrates that geology could not play a significant role in forest development either.

### 3.2. Exploratory spatial data analysis

The spatial distribution of the variables was examined by posting. We created a dot map of forest volume.

| Table 3. Descriptive statistics of forest volume and the factors that influenced it. |
|-----------------------------------|-----------------|-----------------|---------|--------|--------|--------|
| N = 100                           | Minimum         | Maximum         | Mean    | SD     | CV (%) |
| Forest volume (m³/plot)           | 5.47            | 20.34           | 12.10   | 3.45   | 28.50  |
| Elevation (m)                     | 39.00           | 349.00          | 157.97  | 78.23  | 49.52  |
| Slope (%)                         | 1.00            | 59.00           | 23.66   | 14.40  | 60.86  |
| Aspect (°)                        | 0.00            | 351.00          | 195.26  | 111.86 | 57.29  |
| Distance from river (m)           | 0.00            | 694.00          | 184.07  | 153.93 | 83.63  |
| Distance from road (m)            | 20.00           | 531.00          | 100.24  | 94.53  | 94.30  |

| N = 20                            | Minimum         | Maximum         | Mean    | SD     | CV (%) |
| Forest volume (m³/plot)           | 6.04            | 20.79           | 12.86   | 3.30   | 25.69  |
In this map, after classify all values into classes, every sample plot was noted according to the class it belongs to, in order to determine, by visual inspection, the status of the spatial distribution of the variable (26). Data posting (Figure 4) showed that the values of forest volume are spatially correlated. There seems to be a concentration of higher values in the central-northeastern section of the forest, while the lower values are predominant in the southern section as well as in the perimeter of its limits.

![Map showing forest volume data](image)

**Figure 4.** Posting of the forest volume data.

Table 4. The correlation coefficient of the forest volume with the factors that influence it.

|                  | Forest volume | Elevation | Slope   | Aspect | Distance from river | Distance from road |
|------------------|---------------|-----------|---------|--------|--------------------|-------------------|
| Forest volume    | 1.000         | -0.035    | -0.308**| -0.028 | 0.003              | 0.010             |
| Elevation        |               | 1.000     | 0.155   | 0.221* | 0.486**            | -0.262**          |
| Slope            |               |           | 1.000   | -0.017 | -0.096             | 0.055             |
| Aspect           |               |           |         | 1.000  | 0.196*             | -0.028            |
| Distance from river |           |           |         |        | 1.000              | -0.0009           |
| Distance from road |             |           |         |        |                    | 1.000             |

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).
However, Bailey and Gatrell (27) have noted that the visual inspection of dot maps can only provide a preliminary impression of the status of the spatial distribution. It would be rash to reach conclusions based on visual inspection alone.

3.3. Autocorrelation
We used Moran’s I global and local indexes to examine the autocorrelation of the forest volume data.

The normal distribution of the forest volume data allows us to use these indexes. Global Moran’s I index was calculated for different distances between sample plots. A correlogram was created using these distances, (Figure 5). Correlograms plot the spatial autocorrelation against a number of distances from each sample plot. The values of Moran’s I are greater in smaller distances than for larger distances.

From the analysis, it emerges that Global Moran’s I Index takes a maximum value of 0.69, when the lag size is 250 meters, with a z-score 2.20 at 95% confidence level. These results confirmed that the values for forest volume are spatially correlated and thus we can conduct interpolation. Although the positive autocorrelation is a necessary condition for precise results from interpolation it is not sufficient condition. The interpolation must assume additionally that the phenomenon being considered is continuous at the scale sampled (2). In our case, there a lack of uniformity in forest development. This is related to a series of factors, such as natural disasters, illicit wood-cutting, intense grazing, lack of management, and deserted farm areas which have become forested, thus creating irregularities in the forest volume. However, the forest values are considered continuous since we use the mean forest volume for each plot.

Before proceeding to the interpolation, we check for spatial outliers which usually cause erratic behavior in variograms. Using the local Moran’s I, three spatial outliers with z-scores −2.16, −2.08, and −2.22 were detected and excluded from the analysis.

3.4. Interpolation
A first examination of the experimental variograms for forest volume showed that there was no significant difference in the variation according to direction. Therefore, omni-directional variograms were calculated. The selection of the theoretical variogram model was based on visual inspection and the comparison of the RMSE with ASE of different models. The model with the minimum RMSE and the closest values of the two statistical indexes (RMSE and ASE) was selected as the optimal. For both geostatistical methods used, the spherical theoretical model was selected. This model shows a progressive decrease of spatial dependence of the variable when the distance increases, which fades away altogether when the model reaches a sill.

The parameters of the fitted variogram and of the cross-variogram with auxiliary variable the slope are presented in Table 5. The values of slope from 300 observation points obtained from the DEM of the area were used as auxiliary variable in CK.

For IDW, the RMSE was used as the performance assessment criterion. Several different powers and neighborhoods were applied in order to identify the power that produces the minimum RMSE. The lowest RMSE was found with an optimizing power of 2.24 and neighborhood of minimum 10 sample plots. For Kriging and co-Kriging, the lowest RMSE was found with a neighborhood of minimum 8 sample plots.

Figure 5. Correlogram of forest volume.
Table 5. Variogram and cross-variogram characteristics.

| Property | Lag size | No. of lags | Model  | Nugget | Range  | Sill  |
|----------|----------|-------------|--------|--------|--------|-------|
| FV       | 250      | 10          | Spherical | 3.48   | 996.65 | 9.6   |
| FV–SL    | 250      | 10          | Spherical | 2.72   | 2050.52| -10.75|

FV: forest volume, SL: slope.

Table 6. Validation results for forest volume predictions.

| Method | ME     | RMSE  | ASE   | SME    | SRMSE |
|--------|--------|-------|-------|--------|-------|
| IDW    | -1.5044| 3.6625| –     | –      | –     |
| BK     | -1.0859| 3.3880| 1.2907| -0.6482| 2.0320|
| BCK    | -1.1715| 3.5853| 1.2166| -0.7797| 2.4722|

Figure 6. Interpolated maps of forest volume.
The Kriging and co-Kriging geostatistical layers were converted to block interpolated grids using cell size equivalent to plot size (0.2 ha). The average of 81 (9 × 9) points within the block was used as block value. The selection of the number of points per block was based on the mean value of the number of trees per plot.

The validation of the results of the interpolation methods was conducted using the independent validation set (Table 6).

The two geostatistical interpolation methods BK and BCK performed better than IDW. Predictions are increasingly precise with lower RMSE values (28). If we use the RMSE as a performance assessment criterion of the three methods, BK is the interpolation method with the smaller RMSE. Block co-Kriging is the next more accurate method and IDW the third. The ME of IDW is quite high which reveals a prediction bias. The value of ASEs for BK and BCK are less than RMSEs, this suggests that the variability of the predictions is underestimated. The same suggests and the SRMSE since it is greater than 1 in both geostatistical methods.

In Figure 6, the maps of the prediction values from the IDW, BK, and BCK are presented, while Figure 7 shows the block standard error maps of the predicted values from BK and BCK.

The descriptive statistics of the predicted values from the three interpolation methods are presented in Table 7. IDW method is forced to be an exact interpolator because it produces infinities at the data points ($\sum d_i = 0$). Therefore, the original values are saved in the data points (23).

Kriging is usually an exact interpolator (23). However, when variogram and covariance models have a nugget effect which incorporate measurement error and micro variance (variance at lag 0), Kriging can produce filtered predictions at the data locations, so the produced maps are smooth and free of discontinuities at the sample locations. The algorithms incorporated in Geostatistical Analyst extension of ArcGIS software provide filtered value predictions at sample locations (ESRI, ArcGIS 9.3.1).

The range of the predictions from BK and BCK is smaller than that of the original values. The smallest variation of the predicted surfaces, according to CV (Table 7), is observed from the BCK method due to slope data set used. The range of the standard error in the two geostatistical methods is different. In BCK, the standard error range is less than that of BK.

The spatial distribution of the standard errors presented in Figure 7 is linked to the density of the observation points (Figure 1). In areas with high density of sampling plots, the standard error is small. While in areas with sparse sampling or without sampling points, the standard error is high. In the maps shown in Figure 7, the standard error is high in the edges of the forest area where the sample plots are sparse while is smaller in the...
center of the forest. The calculation of the standard error of the predictions is an advantage of the geostatistical methods. On the other hand, IDW is a simple to implement interpolation method and avoids the difficulties and subjectivity in the definition of variograms and interpolation surfaces from the BK and BCK.

The main shortcoming in this study which was crucial for the interpolation predictions was that the shape and the density of the sample plots could not be collected for the purposes of the study. We used data already collected for the needs of the Forest Management Study for the years 2007–2016 (14). The sample plots used for the implementation of the models were not, therefore, uniformly distributed in space and their density exhibited significant irregularities. In many cases, we had no data for large areas of forest and multiple data points for other, much smaller, areas. Another significant problem is the morphology and the lack of continuity in the forest itself. The pine forest is interrupted in places by agricultural areas, partially forested areas and bushes of evergreen broadleaves. In a number of cases, these areas are of significant size. The problem that arises is that for these areas it is not possible to have data measurements, thus, affecting the shape and density of the whole data-set.

4. Conclusion
The collection of forest data is not an easy task due to various morphological and physiographic characteristics of the forest. Therefore, for the efficient mapping of forest parameters, which are often undersampled, auxiliary data are used. In this study three interpolation methods, IDW, BK, and BCK were implemented and compared for the mapping of forest volume data measured across the municipal forest of Skyros island in the central Aegean Sea.

The spatial autocorrelation and the spatial outliers were examined using global and local Moran’s I indexes, respectively. An improvement in the variograms was obtained when three spatial outliers were detected and excluded from the data-set.

An independent validation set was used for the comparison of the three interpolation methods. BK showed a better overall performance. Its accuracy has been tested using five statistical indexes. The performance of the BCK was not as good as is expected, probably due to low relationship of forest volume with slope which used as the auxiliary variable. Finally, the geostatistical techniques outperformed the IDW method. Small differences were observed between the three interpolation methods due to relatively small number of sample plots and their irregular distribution, since we used data already collected for the needs of the Forest Management Study for the years 2007–2016. Comparing the simplicity of IDW with the skills needed for BK and BCK, IDW can be also considered as an acceptable interpolation method for the problem examined.

Acknowledgement
The authors would like to thank the YLI Ltd for providing the data of this study.

Notes on contributors
Dionissios P. Kalivas is an associated professor of Land Evaluation, GIS and Spatial Analysis at the Agricultural University of Athens. He has authored or co-authored book chapters and more than 80 publications in refereed journals and conference proceedings. His research interests include spatial statistics, geostatistics and spatial technologies in natural resources management.

Vassiliki J. Kollias recently retired as a professor of Geographical Information Systems in Agriculture at the Agricultural University of Athens, in Greece, where she taught for more than 40 years. Her research interests include spatial statistics, spatial databases, expert systems and land-use change models. She has written academic textbooks on the field and has published a large number of research papers in most major international journals and conferences.

Ektoras H. Apostolidis is a forester with an MSc on GIS for Land Resources Assessment from Agricultural University of Athens. He works as an independent consultant and researcher specializing in forest management and more generally in management of natural resources.

References
(1) Matheron, G, The Theory of the Regionalized Variables and its Applications; Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau No. 5; Published by the École Nationale Supérieure des Mines de Paris, 1971; p 211.
(2) Gilbert, B.; Lowell, K. Forest Attributes and Spatial Autocorrelation and Interpolation: Effects of Alternative Sampling Schemata in the Boreal Forest. Landscape Urban Planning. 1997, 37, 235–244.
(3) Holmgren, P.; Thuresson, T. Applying Objectively Estimated and spatially Continuous Forest Parameters in Tactical Planning to Obtain Dynamic Treatment Units. For. Sci. 1997, 43 (3), 317–326.
(4) Gunnarsson, F.; Holm, S.; Holmgren, P.; Thuresson, T. On the Potential of Kriging for Forest Management Planning. Scand. J. For. Res. 1998, 13, 237–245.
(5) Mandallaz, D. Estimation of the Spatial Covariance in Universal Kriging: Application to Forest Inventory. Environ. Ecol. Stat. 2000, 7, 263–284.
(6) Hudak, A.T.; Lefsky, M.A.; Cohen, W.B.; Berterretche, M. Integration of Lidar and Landsat ETM+ for Estimating and Mapping Forest Canopy Height. *Remote Sens. Environ.* 2002, 82, 397–416.

(7) Nanos, N.; Montero, G. Spatial Prediction of Diameter Distribution Model. *For. Ecol. Manage.* 2002, 161, 147–158.

(8) Maselli, F.; Marta, C. Evaluation of Statistical Methods to Estimate Forest Volume in a Mediterranean Region. *IEEE Trans. Geosci. Remote Sens.* 2006, 44 (8), 2239–2250.

(9) Sales, H.S.; Souza, C.M.; Kyriakidis, P.C.; Roberts, D.A.; Vidal, E. Improving Spatial Distribution Estimation of Forest Biomass with Geostatistics: A Case Study for Rondonia, Brazil. *Ecol. Model.* 2007, 205, 221–230.

(10) Clark, M.L.; Clark, D.B.; Roberts, D.A. Small-footprint Lidar Estimation of Sub-canopy Elevation and Tree Height in a Tropical Rain Forest Landscape. *Remote Sens. Environ.* 2004, 91, 68–89.

(11) Palmer, D.J.; Hock, B.K.; Kinberley, M.O.; Watt, M.S.; Lowe, D.J.; Payn, T.W. Comparison of Spatial Prediction Techniques for Developing Pinus Radiate Productivity Surfaces across New Zealand. *For. Ecol. Manage.* 2009, 258, 2046–2055.

(12) Tobler, W. A Computer Movie Simulating Urban Growth in the Detroit Region. *Econ. Geogr.* 1970, 46 (2), 234–240.

(13) Gertner, G.; Wang, G.; Anderson, A.; Howard, H. Combining Stratification and Up-scaling Method-Block Cokriging with Remote Sensing Imagery for Sampling and Mapping an Erosion Cover Factor. *Ecol. inf.* 2007, 2, 373–386.

(14) Apostolidis, H.; Adamopoulos Th.; Apostolidis, E. Planning Report for the Forest of Skyros for the Years 2007–2016; YLI Ltd: Athens, 2008; p 159.

(15) Evert, F. Use of Form Factor in Tree-volume Estimation. *J. For.* 1969, 67 (2), 126–128.

(16) Overmars, K.P.; Koning, G.H.J.; Veldkamp, A. Spatial Autocorrelation in Multi-scale Land Use Models. *Ecol. Model.* 2003, 164, 257–270.

(17) Meisel, J.E.; Turner, M.G. Application of Semivariograms to Real and Artificial Landscapes. *Landscape Ecol.* 1998, 13, 347–362.

(18) Legendre, P.; Legendre, L. *Numerical Ecology. Developments in Environmental Modelling* 20; Elsevier: Amsterdam, 1998; p 853.

(19) Anselin, L. Local Indicators of Spatial Association-LISA. *Geog. Anal.* 1995, 27 (2), 93–115.

(20) Levine, N. *CrimeStat III: A Spatial Statistics Program for the Analysis of Crime Incident Locations*; Ned Levine & Associates: Houston, TX and the National Institute of Justice, Washington, DC, 2004; pp 7.33–7.39.

(21) Journel, A.; Huijbregts, C. *Mining Geostatistics*; Academic Press: London, 1978; p 600.

(22) Isaaks, E.H.; Srivastava, R.M. *An Introduction to Applied Geostatistics*; Oxford University Press: New York, NY, 1989; p 561.

(23) Burrough, P.A.; McDonnell, R.A. *Principles of Geographical Information Systems*; Oxford University Press: Oxford, UK, 1998; pp 132–161.

(24) McBratney, A.B.; Webster, R. Optimal Interpolation and Isarithmic Mapping of Soil Properties: V. Co-regionalization and Multiple Sampling Strategy. *J. Soil Sci.* 1983, 34, 137–162.

(25) Meyers, D.E. Pseudo-cross Variograms, Positive Definiteness, and cokriging. *Math. Geol.* 1991, 23 (6), 805–817.

(26) Goovaerts, P. *Geostatistics for Natural Resources Evaluation*; Oxford University Press: New York, NY, 1997; p 483.

(27) Bailey, C.T.; Gatrell, A.C. *Interactive Spatial Data Analysis*; Longman Scientific & Technical, Harlow Essex: England, 1995; p 413.

(28) Odeh, I.O.A.; McBratney, A.B.; Chittleborough, D.J. Spatial Prediction of Soil Properties from Landform Attributes Derived from a Digital Elevation Model. *Geoderma* 1994, 63, 197–215.