INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

DISSERTATION

MODELING FOREST STAND STRUCTURE USING GEOSTATISTICS, GEOGRAPHIC INFORMATION SYSTEMS, AND REMOTE SENSING

Submitted by

GERHARD HUNNER

Department of Forest Sciences

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2000

COLORADO STATE UNIVERSITY

April 11, 2000

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY **GERHARD HUNNER** ENTITLED **"MODELING FOREST STAND STRUCTURE USING GEOSTATISTICS, GEOGRAPHIC INFORMATION SYSTEMS, AND REMOTE SENSING"** BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Co-Advisor

Advisor

Department Head

ABSTRACT OF DISSERTATION

MODELING FOREST STAND STRUCTURE USING GEOSTATISTICS, GEOGRAPHIC INFORMATION SYSTMES, AND REMOTE SENSING

Forest management requires the estimation and mapping of forest resources. For reasons of time and cost an exhaustive measurement of every individual tree in a forest is not feasible and the variables of interest are measured at single point locations. However, information, such as basal area and timber volume, is usually required for the entire forest. This leads to methods of interpolating data and estimating the mean value within the area. Forest stand structure is traditionally mapped as polygons. This approach assumes that forest parameters are homogeneous within each polygon and change abruptly at boundaries. Many natural phenomena, however, change gradually over space. Spatial interpolation techniques like geostatistical methods, can be applied to represent forest stand structure as a continuous surface. While traditional statistics assume independent data, geostatistics take a different approach by quantifying and modeling this spatial dependence. Underlying this approach is the expectation that, on average, samples close together have more similar values than those that are farther apart (spatial autocorrelation).

This study compared five geostatistical methods of interpolation (ordinary kriging, universal kriging with first-degree trend surface, universal kriging with second-degree trend surface, cokriging, and disjunctive kriging) with three traditional estimation methods (polygonal mapping, inverse distance weighting, and inverse distance weighting squared). These eight techniques were used to spatially interpolate the number of stems, total basal area, and number of seedlings on 82 sample plots in a 121-hectare first-order forest watershed in the USDA Forest Service, Fraser Experimental Forest, Fraser, Colorado. Secondary variables used for cokriging included elevation, a combined value for slope and aspect, and the normalized difference vegetation index (NDVI) from Landsat-TM satellite imagery. The comparison criterion was the mean square error (MSE) calculated by cross validation.

For variable number of stems the MSEs ranged from 44.568 to 49.444 with cokriging being the best estimation method and disjunctive kriging giving the poorest results. However, the differences between the various methods were relatively small. The MSEs for variable total basal area ranged from 3.464 to 4.598. The best results were obtained using polygonal mapping, while the poorest results were given by inverse distance weighting squared. Again, the differences between the various methods were relatively small. Variable number of seedlings had the best estimation results applying inverse distance weighting squared (MSE of 69.881). The worst results were obtained using disjunctive kriging (MSE of 118.995). For this variable, the differences in MSEs for the various interpolation methods were much larger than with the other two variables.

There was no single "best interpolation method" as the performance of the estimation techniques was different between variables. Overall, however, cokriging

performed best, followed by polygonal mapping. By utilizing the spatial cross-correlation between primary and secondary variables the quality of the cokriging estimates was improved as compared to the results of the other kriging methods. Polygonal mapping gave good estimation results for two of the variables under study. Universal kriging with a first- or second-degree trend surface yielded, in general, better results than ordinary kriging. Removing a trend improved the interpolation results for two of the variables in comparison with ordinary kriging. Inverse distance weighting was generally less accurate than the linear kriging methods. Inverse distance weighting techniques outperformed the kriging methods only in the case where the requirements of kriging (approximately normal distribution of the data) were not fulfilled. The disadvantage of inverse distance weighting is that it cannot take the clustering of sample points into account, while kriging methods give less weight to clustered sample points. The nonlinear kriging method (disjunctive kriging) performed least well. This method could only be used if the transformed data were close to being normally distributed. But even in cases where the transformation process was successful, disjunctive kriging results were not better than the other kriging methods. Additional information (in the form of spatially cross-correlated auxiliary variables) seemed to be a more important consideration than whether the estimation method is linear or nonlinear.

Gerhard Hunner
Department of Forest Sciences
Colorado State University
Fort Collins, CO 80523
Summer 2000

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation and gratitude to my advisor, Dr. Robin Reich. He introduced me to the field of geostatistics, and gave me his guidance and advice throughout the years. Special thanks go to my co-advisor, Dr. Todd Mowrer at the Rocky Mountain Research Station, USDA Forest Service. His continuous support and encouragement, hours of discussion about the secrets of geostatistics, and – most important – his friendship helped me in getting through this endeavor. Also thanks to him for providing the data for this project and the financial support. Thanks go to my other two committee members, Dr. Denis Dean and Dr. Melinda Laituri, for their advice and support. It has been a great pleasure to have worked with them during my graduate education.

I owe my deepest thanks to Rosalía Rojas-Sánchez. Her love and moral support gave me strength during the ups and downs of living in a foreign country. I wish to thank all my friends in Germany, notably Martin and Elizabeth Rempfer, Ingrid Moiser, Ludwig Nutz, Sabine Hanisch, Regine Oberstadler, and Axel Moosdorf. They showed me that distance and time are no barriers for true friendship. I am also grateful to my friends in Colorado, notably Jimmy Gaudry, Robin Weber, Rich Muzzy, Claudia Leon, and Julio Zimbron for sharing this journey with me.

DEDICATION

To my parents,

For their everlasting support of my education and goals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Forest management requires the estimation and mapping of forest resources. A forest inventory is the procedure for obtaining information on the quantity and quality of the forest resources (Husch et al., 1982). For reasons of time and cost an exhaustive measurement of every individual tree in a forest is not feasible. Different sampling techniques have been applied to estimate forest population variables. Some of the most widely used sampling methods are simple random sampling, stratified random sampling, and systematic sampling. All of these sampling techniques assume independence among the sampling units. However, this assumption is rarely fulfilled in nature. Husch et al. (ibid., p. 182) write: "*In biological populations, such as a forest, the components are rarely, if ever, arranged completely independent of each other, but instead, show a systematic or periodic variation from place to place.*" Despite the violation of the independence assumption, these traditional sampling techniques are commonly used in forest inventory. While traditional statistics assume independent data, geostatistics take a different approach by quantifying and modeling spatial dependence.

Information gathered from forest sampling is traditionally mapped as polygons. This approach assumes that forest parameters are homogeneous within each polygon and change abruptly at boundaries. Many natural phenomena, however, change gradually over space. Spatial interpolation techniques like geostatistical methods, can be applied to represent forest stand structure as a continuous surface. Most interpolation techniques involve weighted linear combinations of available nearby sample plot values in order to provide estimates at unsampled locations. G. Matheron (1962, 1963) was the first to use the term geostatistics extensively and his definition will be retained:

> "*Geostatistics is the application of the formalism of random functions to the reconnaissance and estimation of natural phenomena.*"

Matheron (ibid.) laid the foundation of geostatistics with his Regionalized Variable Theory. A regionalized variable is a variable that is distributed in space, and its value is a function of its position in space. A regionalized variable often changes so irregularly from place to place that its variation cannot be modeled by a simple deterministic function. A probabilistic approach is chosen where the data values are viewed as the outcomes of a random mechanism (Goovaerts, 1997). However, behind a locally erratic aspect there is a characteristic behavior or structure in the spatial variability of a regionalized variable in the sense that the values at positions near one another tend to be more similar than others farther apart (spatial autocorrelation).

The geostatistical estimation process can be divided into two parts: first, the investigation and modeling of the spatial correlation of the data (structural analysis), and second, the estimation process (kriging). The variogram is the basic tool in geostatistics used to quantify spatial correlation. After having fitted a mathematical function to a

sample variogram, an estimation procedure called kriging can be used to estimate values at unsampled locations. Many different kriging techniques have been developed over the years. Ordinary kriging is the original kriging technique (Matheron, 1962, 1963) resulting in an unbiased estimate with minimized estimation error variances at unsampled locations. Ordinary kriging requires a stationary mean in data values over the study area. Many natural phenomena, however, exhibit a change in average values over the area (called a trend). Universal kriging is a kriging method that takes such local trends into account when computing the estimates. Cokriging utilizes multivariate information to determine estimates at unsampled locations by exploiting the cross-correlation between a variable of interest and a set of spatially cross-correlated secondary variables. These three kriging methods belong to the group of linear geostatistics as they deal only with linear combinations of the variable under study. Disjunctive kriging is a non-linear estimation method. Non-linear methods are more flexible; they allow a much broader range of possible combinations of sample values to estimate values at unsampled locations. Disjunctive kriging allows the calculation of conditional probabilities, that a variable is greater than, or less than, some prescribed cut-off or tolerance value.

Objectives of the Study

Many interpolation techniques have been widely and successfully used in other fields of natural resource assessment, primarily in mining and petrology, but rarely in forestry. To date there has not been any study in forestry that has compared the various techniques. Given their increasing importance, conducting such a comparative study is certainly warranted. The variables of interest chosen for this investigation are the number of trees, the basal area, and the number of seedlings. Auxiliary variables included elevation, a combined value of slope and aspect, and normalized difference vegetation index (NDVI) from Landsat-TM satellite imagery. Data were available for the Lexen Creek watershed in the Fraser Experimental Forest, Fraser, Colorado. The four methods of kriging applied in this project were ordinary kriging, universal kriging, cokriging, and disjunctive kriging. For purpose of comparison the traditional forest mapping technique (here called polygonal mapping, where the variable of interest is averaged over all sample points within a polygon), as well as inverse distance weighting (an interpolation technique where the weight for each sample point is inversely proportional to its distance from the point being estimated) are included. The objectives of this dissertation are summarized as follows:

1) To analyze and model the spatial correlation of the three variables of interest (number of trees, total basal area, number of seedlings), as well as the spatial cross-correlation between those variables and a set of secondary variables (elevation, a combined value of slope and aspect, satellite data).

2) To estimate the three variables of interest using ordinary kriging assuming a stationary mean in the data.

3) To estimate the three variables of interest using universal kriging assuming a non-stationary mean (trend) in the data. In the presence of a strong trend universal kriging should result in better estimates than ordinary kriging.

4) To estimate the three variables of interest using cokriging by exploiting the spatial cross-correlation between these variables and the set of secondary variables. The geostatistical technique of cokriging is used to combine auxiliary information with primary field data that should lead to better estimates.

5) To estimate the three variables of interest using disjunctive kriging. Linear kriging methods may only yield optimal results if the variable under study has a normal distribution. However, non-linear kriging techniques like disjunctive kriging can be applied to any non-normal distribution. Disjunctive kriging should lead to better estimates for variables with a non-normal distribution. In addition, disjunctive kriging allows the calculation of conditional probabilities.

6) To estimate the three variables of interest using inverse distance weighting. This interpolation method is easily implemented and quickly applied to any data set; no stationary mean is required. However, inverse distance weighting does not result in unbiased estimates and minimized error variances as do the geostatistical kriging methods.

7) To estimate the three variables of interest using the traditional polygonal mapping technique where the forest is divided into homogeneous polygons. If the variables of interest change continuously over space this method should be the least accurate of all interpolation methods.

8) To compare the different interpolation techniques using cross-validation. The resulting residuals are used with different measures of validation to quantify and compare the errors associated with each interpolation technique.

CHAPTER 2

SPATIAL STATISTICAL TECHNIQUES

2.1 Principles of Spatial Statistical Techniques

Spatial interpolation techniques use sample information to predict values of interest in areas not sampled. All interpolation methods involve weighted linear combinations of nearby sampled values in order to provide estimates at unsampled locations:

$$Z^*(x_0) = \sum_{i=1}^{n} \lambda_i \cdot Z(x_i) \qquad [2.1]$$

where:

$Z^*(x_0)$ = the estimate of variable Z at the unsampled location x_0,

$Z(x_i)$ = the value of variable Z at the sample location x_i,

λ_i = a weight assigned to the value $Z(x_i)$, and

n = the number of nearby sample values.

Different approaches to assigning the weights to the data values give rise to many different methodologies.

2.2 Polygonal Mapping

A simplistic approach to incorporate the information from nearby samples is by weighting them all equally, using their mean as the estimate. With polygonal mapping a spatially distributed variable of interest is averaged over all sample points within a sampling unit (polygon). External landscape features are used to delineate polygons. These are the themes drawn on most thematic maps of forests, soils, geology, vegetation, land use, etc. This technique assumes that important variation occurs at boundaries; the variation within each polygon is smaller than the variation between the polygons. This kind of mapping leads to a "stepped" model of the landscape as the polygons form a discontinuous surface of plateaus (Burrough, 1986).

Analysis of Variance (ANOVA)

Analysis of variance (ANOVA) techniques can be used to decide whether the differences among the sample means of each polygon are large enough to imply that the corresponding population means are different. ANOVA is a test of the statistical significance of the differences among the mean scores of two or more groups on one (or more) variables (Vogt, 1999). The differences in group means are judged statistically significant (or not) by comparing them to the variation within groups. The procedure in ANOVA involves computing a ratio (F ratio) of the variance between groups to the variance within the groups.

The ANOVA model is described by the following model:

$$Z(x_{ij}) = \mu_j + \varepsilon_{ij} \qquad [2.2]$$

where:

$Z(x_{ij})$ = the variable of interest at location x_i in group j,

μ_j = the mean of group j, and

ε_{ij} = the residual or noise.

The assumptions for the ANOVA model are:

- The samples are independent, random samples.
- The samples are normally distributed.
- The variance is the same for each group.

Ott (1992) has a further discussion on the requirements of these assumptions.

A hypothesis test is applied. The null hypothesis of ANOVA is that all group means μ_j are identical. The alternative hypothesis is that at least one group mean is different from the others. The null hypothesis is tested by forming an F-ratio based on the variance between the groups s_B^2 and the variance within the groups s_W^2 (pooled over all groups):

$$F = \frac{s_B^2}{s_W^2} \qquad [2.3]$$

When the null hypothesis is true, both s_B^2 and s_W^2 estimate the sample variance σ^2, and F would be expected to lie near a value of 1. When the hypothesis of equality is false, s_B^2 will tend to be larger than s_W^2 due to the differences among the group means.

The results of the analysis are summarized in an ANOVA table:

Table 2.1 The ANOVA table.

Source	Sum of Squares	Degrees of Freedom	Mean Square	F Test	p-value
Between groups Within groups Totals	SSB SSW TSS	t - 1 n_T - t n_T - 1	s_B^2 = SBB/(t-1) s_W^2 = SBW/(n_T-t)	s_B^2/s_W^2	p

where:

- TSS is the total sum of squares, and is computed as

$$TSS = \sum_{i,j} y_{ij}^2 - \frac{y_{..}^2}{n_T}$$

where y.. is the sum of all samples.

- SSB is the sum of squares between samples, and is calculated as:

$$SSB = \sum \frac{y_{j.}^2}{n_j} - \frac{y_{..}^2}{n_T}$$

where y_j. is the sum of the samples obtained from group j, and n_j is the number of samples in group j.

- SSW is the within-sample sum of squares, and is computed as:

SSW = TSS – SSB.

- p-value (level of significance) is the probability of observing a F-value greater than the tabulated F-value, assuming the null hypothesis is true.
- t is the number of groups.
- n_T is the total number of sample points.

The null hypothesis of equality of the t group means is rejected if the test statistic F exceeds the tabulated value of F for a = α, df_1 = t-1, and df_2 = n_T - t, where α is the type I error (normally set to 0.1), df_1 is the first degree of freedom, and df_2 is the second degree of freedom.

Multiple Comparison Methods

After conducting an analysis of variances one wants to know which of the group means is significantly different from the others. Multiple comparison methods are used to find significant differences between all possible comparisons of the different groups. Among several multiple comparison methods available, Tukey's Honest Significant Difference (HSD) Test is one of the most widely used techniques. The original Tukey's HSD test requires equal sample numbers per group. A modified version, sometimes referred to as the Tukey-Kramer procedure (Ott, 1992), takes unequal sample numbers per group into account:

$$HSD = q_a(t, df)\sqrt{\frac{s_w^2}{n}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \qquad [2.4]$$

where $q_\alpha(t, df)$ is the upper-tail value of the Studentized range. Its value can be found in tables using number of groups t and the degrees of freedom df for s_w^2. n_i is the number of samples in group i, n_j is the number of samples in group j. Two group means are declared different if the difference between both means is larger than (or equal to) HSD.

Applications of Polygonal Mapping

Most thematic maps show polygons of homogeneous units of forests, soils, land use, etc.. External features like rivers, breaks of slope, and change in vegetation are used to delineate such units. Lowell (1996) assessed the validity of polygons versus a continuous surface representation of forest volume in a boreal forest in Quebec. He found that in virtually no situation in his study – between forest types, different stands of the same forest type, and within a single stand – was it appropriate to model forest volume as

a continuous surface. His evidence suggested that the forest is a set of polygons with inexact boundaries.

2.3 Inverse Distance Weighting

Inverse distance weighting (IDW) is an interpolation technique that gives more weight to nearby samples and less weight to those that are further away. The weight for each sample is inversely proportional to its distance from the point being estimated (Isaaks & Srivastava, 1989):

$$Z^*_{IDW}(x_0) = \frac{\sum_{i=1}^{n} \frac{1}{d_i^p} \cdot Z(x_i)}{\sum_{i=1}^{n} \frac{1}{d_i^p}} \qquad [2.5]$$

where:

$Z^*(x_0)$ = the estimated value at unsampled location x_0,

$Z(x_i)$ = the sample point value at location x_i,

d = the distance from the sample location to the point being estimated,

p = the distance exponent, and

n = the number of sample points.

As one decreases p, the weights given to the samples become more similar. As one increases p, the individual weights become more dissimilar, with most weight given to the nearest sample. Traditionally, the most common choice for the inverse distance exponent is 2 (Isaaks & Srivastava, 1989).

The variance is calculated as (Reich, 1999a):

$$\sigma^2_{IDW}(x_0) = \left[\sum_{i=1}^{n}\left(\frac{1}{d_i^p}\cdot Z(x_i) - \frac{1}{d_i^p}\cdot Z^*(x_0)\right)^2 / (n-1)\right] / n\cdot \bar{x}^2 \qquad [2.6]$$

where $\bar{x}$ is the sample mean.

IDW estimators are very sensitive to the number of neighbors used in the estimation, and to the distance exponent used in the weighting process (Weber & Englund, 1994). Burrough and McDonnell (1998) mention that the quality of IDW is also affected by the clustering in the data and by the presence of outliers. Lam (1983, p. 132) writes that "*Simplicity of the principle, speed in calculation, ease of programming, and good results for many type of data*" make IDW an attractive interpolation method.

Applications of Inverse Distance Weighting

Inverse distance weighting has been applied in many different fields: soil science (Laslett et al., 1987; Laslett & McBratney, 1990; Brus et al., 1996; Gallichand & Marcotte, 1993), hydrology (Rouhani, 1986; Grimm & Lynch, 1991), atmospheric science (Phillips et al., 1997), geochemistry (Kane et al., 1982), and in the creation of digital elevation models (Weber & Englund, 1992; Weber & Englund, 1994; Wood & Fisher, 1993). No application of IDW in forestry was found. Several studies have been conducted in the past comparing different interpolation techniques. While in some studies (Laslett & McBratney, 1990; Grimm & Lynch, 1991; Weber & Englund, 1994; Phillips et al., 1997; Zimmerman et al., 1999) kriging (see below) performed best among the interpolation techniques, other studies (Weber & Englund, 1992; Gallichand & Marcotte, 1993; Brus et al., 1996) found IDW as good as or better than geostatistical methods.

2.4 Geostatistics

Many properties of the earth's surface, such as elevation, precipitation, soil etc., vary continuously in space. Geostatistical methods rest on the recognition that the spatial variation from place to place of many properties, known as "regionalized variables," cannot be modeled by simple mathematical functions. However, there is some spatial structure, i.e. spatial dependence that can be described by a stochastic approach. This spatial structure, or spatial autocorrelation, is the dependence between data values: values located close together show greater similarity than values farther apart. The geostatistical approach proceeds by first exploring and modeling this autocorrelation of the variables. The resulting information is then used to estimate the weights for geostatistical interpolation techniques. For a detailed description on the theory of geostatistics the interested reader is referred to David (1977), Journel & Huijbregts (1978), Rendu (1981), David (1988), Isaaks & Srivastava (1989), Journel (1989), Cressie (1991), Wackernagel (1995), and Armstrong (1998). For an introduction of geostatistics to fields other than mining one should consult Hohn (1998) for the field of petroleum geology, Kitanidis (1997) for hydrogeology, and Goovaerts (1997) for natural resources evaluation.

2.4.1 Regionalized Variable Theory

A *regionalized variable* is a variable that is distributed in space (Journel & Huijbregts, 1978). Many natural phenomena in the earth sciences, e.g. soil properties, elevation, precipitation, etc., can be regarded as regionalized variables. From a mathematical point of view, the value of a regionalized variable at any place on the earth's surface is a function of its spatial position (Webster & Oliver, 1990). A

regionalized variable often changes so irregularly from place to place that its variation cannot be modeled by a simple deterministic function. In many situations the parameters of a data-generating mechanism cannot be determined. A probabilistic approach is chosen where the data values are viewed as the outcomes of a random mechanism (Goovaerts, 1997).

A *random variable* is a variable whose values are generated according to some probabilistic mechanism. Such a random variable can take on values that follow a certain probability distribution, such as normal or uniform distribution (Hohn, 1998). The random variable is denoted by the capital letter Z, while its outcome values are denoted with the corresponding lower case letter z. To show the location dependence of a random variable, the notation Z(x) is used, with x being a vector of location coordinates. A sample value z, measured at location x, represents one particular realization of a single random variable Z(x). At each point x, the nature of the mechanism that produces a value z may be different (Wackernagel, 1995), i.e. each value of z(x) observed at each point x represents a different random variable Z(x) with a different probability distribution. However, behind a locally erratic aspect there is a characteristic behavior or structure in the spatial variability of a regionalized variable in the sense that the values at positions near one another tend to be more similar than others farther apart. Therefore a regionalized variable possesses two characteristics:

- a random, erratic component which is related to the local behavior of the natural phenomenon, and
- a general, structured component which is related to the overall distribution of the natural phenomenon.

The probabilistic interpretation provided by random functions takes these two aspects of randomness and structure into account.

The Random Function Concept

A *random function* is the set of all random variables observed in the area (Deutsch & Journel, 1998). The definition of a random function comprises the random and structured aspects of a regionalized variable (Journel &Huijbregts, 1978):

- locally, at a point x, Z(x) is a random variable;
- regionally, Z(x) is also a random function in the sense that for each pair of points at locations x and x + h (where h is the distance vector between points), the corresponding random variables Z(x) and Z(x+h) are not independent, but are related by some degree of spatial correlation.

A group of spatially distributed samples (random variables) can be considered as one particular realization of a random function. However, this interpretation of a regionalized variable as a particular realization of a certain random function Z(x) has an operative sense only when it is possible to infer the probability law which comprises this random function (Journel & Huijbregts, 1978). This probability law of a random function is defined as the set of all distribution functions observed for all points in a study area. Many realizations of the random function are required in order to infer its probability law. In practice, we are limited to a single realization z(x) of the random function Z(x) at position x, and therefore it is impossible to infer the probability law of the random function. However, exhaustive knowledge of the probability law is not required for most estimation problems, and only the first two moments of the random function are used.

Statistical Moments

Moments are summary parameters that characterize statistical populations.

First-order moment:

The distribution function of a random variable Z(x) observed at position x has an expectation which depends only on x:

- $E\{Z(x)\} = m(x)$ [2.7]

Second-order moments:

Three second-order moments are used in geostatistics:

- The variance (Var):

  $\text{Var}\{Z(x)\} = E\{[Z(x) - m(x)]^2$ [2.8]

- The covariance (C):

  $C(x, x+h) = E\{[Z(x) - m(x)] \cdot [Z(x+h) - m(x+h)]\}$ [2.9]

- The variogram (γ), which is defined as half of the variance of the increments $[Z(x) - Z(x+h)]$:

  $\gamma(x, x+h) = \frac{1}{2} \text{Var}[Z(x) - Z(x+h)]$ [2.10]

Hypothesis of Stationarity

If the phenomenon under study is homogeneous over a certain area, the observed values z(x) and z(x+h) at two different locations x and x+h can be considered as two different realizations of the same random variable. This homogeneity or repetition provides the equivalent of many realizations of the same random function Z(x), and permits a certain amount of statistical inference. The assumption of homogeneity avoids

the problem that usually only one realization of the random variable is available for a given value of x. Several types of homogeneity have been described under the general concept of stationarity. Stationarity is defined through the first- and second-order moments of the observed random function, and various degrees of stationarity correspond to the particular moments that remain invariant across a study area (Hohn, 1998). The type of stationarity that is assumed tells what kind of statistical inference is permitted within the probabilistic model.

Strict stationarity

In the strict sense, a random function is called stationary when its probability law is invariant under translation, i.e. the characteristics of a random function stay the same when shifting a given set of points from one part of the study area to another (Wackernagel, 1995). Two k-component random variables have the same k-variable distribution law for any separation vector h: $\{Z(x_1), Z(x_2),\ldots, Z(x_k)\}$ and $\{Z(x_1+h), Z(x_2+h),\ldots, Z(x_k+h)\}$ are identical in law, i.e. the distribution function remains unchanged with respect to h. This very strong stationarity hypothesis is rarely assumed in geostatistics. Geostatistical applications consider only the previously defined first two moments. Therefore, it is sufficient to assume that these moments exist, and then limit the stationarity assumptions associated with them.

Second-order stationarity

A random function is called second-order stationary if

- the mathematical expectation $E\{Z(x)\}$ exists and does not depend on location x (stationarity of the mean):

$$E\{Z(x)\} = m, \quad \forall x \qquad [2.11]$$

This condition implies that there is no trend (changing mean) in the data.

- For each pair of random variables $\{Z(x), Z(x+h)\}$, the covariance exists and depends only on the separation distance h, and not on the location x and x+h (stationarity of the covariance):

$$C(h) = E\{Z(x) \cdot Z(x+h)\} - m^2, \quad \forall x \qquad [2.12]$$

If the covariance is stationary, then the variance and the variogram are also stationary:

$$Var\{Z(x)\} = E\{[Z(x) - m]^2\} = C(0), \quad \forall x \qquad [2.13]$$

$$\gamma(h) = E\{[Z(x) - Z(x+h)]^2\} = C(0) - C(h), \quad \forall x \qquad [2.14]$$

The third tool for characterizing the autocorrelation between two variables is the correlogram:

$$\rho(h) = C(h)/C(0) = 1 - \gamma(h)/C(0) \qquad [2.15]$$

Under the hypothesis of second-order stationarity, the covariance, the variogram, and the correlogram are equivalent tools for characterizing the correlation between two random variables Z(x) and Z(x+h).

Intrinsic hypothesis

Second-order stationarity assumes the existence of a covariance and, thus, of a finite *a priori* variance. Many physical phenomena and random functions have an infinite capacity for variation (i.e. Brownian motion), and therefore no *a priori* variance or covariance exists (Pannatier, 1996). However, a variogram can be defined by the variance of the increments. The intrinsic hypothesis is a "reduced" second-order stationarity hypothesis, as it limits the second-order stationarity to the *increments* of the random

function Z(x), and it assumes only the existence and stationarity of the variogram. A random function is called intrinsic if:

- the mean of the *increments* [Z(x) - Z(x+h)] exists and does not depend on location x or x+h (stationarity of the mean):

  $$E\{Z(x) - Z(x+h)\} = 0, \quad \forall x \qquad [2.16]$$

- the *increment* [Z(x) - Z(x+h)] has a finite variance which does not depend on the location x or x+h (stationarity of the variogram):

  $$\gamma(h) = \tfrac{1}{2} \mathrm{Var}\{Z(x) - Z(x+h)\} = \tfrac{1}{2} E\{[Z(x) - Z(x+h)]^2\}, \quad \forall x \qquad [2.17]$$

Under the intrinsic hypothesis only the variogram exists and can be used to characterize the correlation between two random variables Z(x) and Z(x+h). The intrinsic hypothesis is sufficient for most geostatistical problems (Journel & Huijbregts, 1978). Regionalized variables that are second-order stationary always satisfy the intrinsic hypothesis, but the converse is not necessarily true.

2.4.2 Structural Analysis

The geostatistical estimation process consists of two parts:

- the *structural analysis*: the investigation and modeling of the patterns of spatial correlation that characterize a regionalized variable, and
- the *kriging* methods: the estimation process using features from the structural analysis to define the weighting factors.

2.4.2.1 The Variogram

Traditionally the variogram has been used for modeling spatial variability rather than the covariance or correlogram. The reason for geostatisticans preferring the variogram is that its definition requires only intrinsic stationarity. Hohn (1998, p. 15) describes the variogram as "*the basic geostatistical tool for visualizing, modeling, and exploiting the spatial autocorrelation of a regionalized variable.*"

The Experimental Variogram

The *semivariance* $\gamma(h)$ is half of the mean squared difference for two data points separated by distance (called the lag) h. The variogram can be estimated without bias from the sample data:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \qquad [2.18]$$

where:

$Z(x_i)$ = the sample value of variable Z at position x_i,

$Z(x_i+h)$ = the sample value of variable Z at position x_i+h, and

N(h) = the number of pairs of observations separated by distance h.

A plot of $\gamma(h)$ against h is known as the experimental *semivariogram*, or simply the *variogram* (Figure 2.1). The variogram measures the average dissimilarity between data points that are separated by a distance h (Goovaerts, 1997). The dissimilarity generally increases with increasing lag distance.

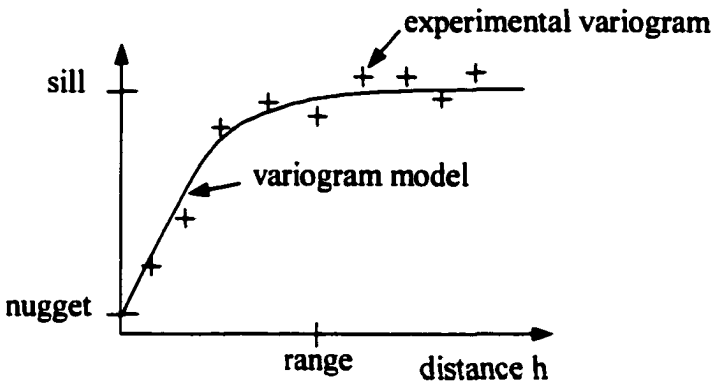

Figure 2.1 The experimental variogram and the variogram model with its typical features: sill, range, and nugget.

A typical variogram shows three features: the sill, the range, and the nugget effect. As the separation distance between data points increases, the variogram values also increase. Finally, the variogram reaches a maximum value called the *sill*. The distance at which the variogram reaches the sill is called the *range*. At this distance there is no longer any spatial correlation between data points. Though the value of the variogram for distance zero is strictly 0, several factors, such as sampling error and short scale variability, may cause sample values separated by extremely small distances to be quite dissimilar. This causes a discontinuity at the origin of the variogram (a non-zero intercept) called the *nugget*.

When computing a variogram two distance parameters need to be chosen: first the difference between two lags (called lag spacing, or lag increment), and secondly a distance tolerance (called lag tolerance). Most of the time sample points are not regularly distributed in space. Hence, the problem one encounters is that for any lag h there is enough randomness in the sample location that very few pairs of samples are separated exactly by h. Therefore in practice one has to specify tolerances on the distance of lag h

and group the sample pairs into distance classes (Issaks & Srivastava, 1989). The lag spacing that gives the "smoothest" variogram structure should be chosen. The most common choice for the lag tolerance is half the lag spacing.

The accuracy of the variogram calculation is proportional to the number of data points available. Hohn (1998) and Olea (1994) mention three rules of thumb to ensure proper variogram estimation:

1) For each computed value of the variogram the number of pairs should be greater than 30.
2) The section of interest of the variogram, usually the increase at relatively small distances, should be represented by three to four values.
3) The maximum lag should be limited to one-half the extreme distance in the sampling area.

The experimental variogram just presented is a measure of spatial correlation independent of orientation or direction (*isotropy*), and is called an omnidirectional variogram. However, sometimes the spatial correlation changes with direction (*anisotropy*) and directional variograms can be calculated. In the case of directional variograms different directions and a directional tolerance (analogous to the lag tolerance) have to be chosen. Issaks and Srivastava (1989) give more details.

Variograms are sensitive to extreme values. Several alternative methods have been proposed that deal with this problem. Some of these robust methods include relative variograms (local relative variograms, general relative variograms, pairwise relative variograms), madogram, and rodogram. However, these methods are not considered a substitute for the evaluation of a traditional variogram. Issaks & Srivastava (1987)

explain these alternative techniques. Another method for making a variogram look "smoother" is either to remove erratic single data points from the sample set, or to delete erratic data pairs from specific lags.

A constant mean is required for estimating the variogram. A variation in the mean is called a drift or trend. It can be proved that for stationary and intrinsic variables

$$\lim_{h \to \infty} \frac{\gamma(h)}{|h|^2} = 0 \qquad [2.19]$$

This result shows that if the variogram increases more rapidly than h^2, the variable is nonstationary, which indicates the presence of a trend (Armstrong, 1998). Proper variogram estimation practice requires the removal of this trend as it overestimates the underlying variogram. One method for eliminating the trend is by subtracting an analytical function, such as a polynomial, from the sample data (Olea, 1994).

The Variogram Models

A mathematical model must be fitted to the experimental variogram. The need for a variogram model comes from the fact that the interpolation algorithm may require a variogram value for some distances at which we do not have an experimental variogram value. Another reason for a variogram model is that the experimental variogram values are subject to error, especially if the sample is small, which may make the sample variogram appear erratic (Oliver & Webster, 1990).

Variogram Model Requirements

As shown before, interpolation techniques involve weighted linear combinations of available nearby sample values in order to provide estimates of variable Z at unsampled locations x_0:

$$Z^*(x_0) = \sum_{i=1}^{n} \lambda_i \cdot Z(x_i) \qquad [2.20]$$

The variance of any linear combination of data values must be positive, or equal to zero:

$$\text{var}\,[Z^*(x_0)] \geq 0 \qquad [2.21]$$

To ensure that the variance is non-negative, the variogram model γ must be *conditionally negative definite* (Goovaerts, 1997). Note, other authors (e.g. Journel & Huijbregts, 1978; Armstrong, 1998) write that the variogram model $-\gamma$ (negative γ!) has to be conditionally positive definite! Only certain functions that fulfill this condition can be used as models for variograms. As it is not easy to recognize functions that have this property or to test for it, it is best to choose variogram models from the range of suitable functions rather than try to create them oneself (Armstrong, 1998).

Permissible Variogram Models

The following four basic variogram models are known to be conditionally negative definite and are the variogram models most commonly used in geostatistics. Figure 2.2 shows the different variogram models. The following notation will be used in the description of the various variogram models: h is the lag distance, a is the range, C_0 is the nugget effect, $C_0 + C_1$ is the sill, and $\gamma(h)$ is the variogram value at lag distance h.

Spherical model:

$$\gamma(h) = C_0 + C_1 \cdot \begin{cases} 1.5 \cdot \frac{h}{a} - 0.5 \cdot \left(\frac{h}{a}\right)^3 & \text{if } h \le a \\ C_0 + C_1 & \text{otherwise} \end{cases} \quad [2.22]$$

The spherical model is the most commonly used variogram model.

Exponential model:

$$\gamma(h) = C_0 + C_1 \cdot \left[1 - \exp\left(\frac{-3h}{a}\right)\right] \quad [2.23]$$

The exponential model reaches its sill asymptotically. The practical range is defined as the distance where the variogram value is 95% of the sill.

Gaussian model:

$$\gamma(h) = C_0 + C_1 \cdot \left[1 - \exp\left(\frac{-3h^2}{a^2}\right)\right] \quad [2.24]$$

The Gaussian model is used to model extremely continuous phenomena. This variogram model reaches its sill asymptotically and the practical range is defined as the distance where the variogram value is 95% of the sill.

Power model:

$$\gamma(h) = C_0 + C_1 \cdot h^{\alpha} \quad [2.25]$$

with $0 < \alpha < 2$. The linear model with $\alpha = 1$ is a special case of the power model. The power model does not reach a sill, but increases proportionally to h.

Other available models include cubic model, hole effect model, cardinal sine model, prismato-magnetic model, prismato-gravimetric model (Armstrong, 1998), and De Wijsian model (Hohn, 1998).

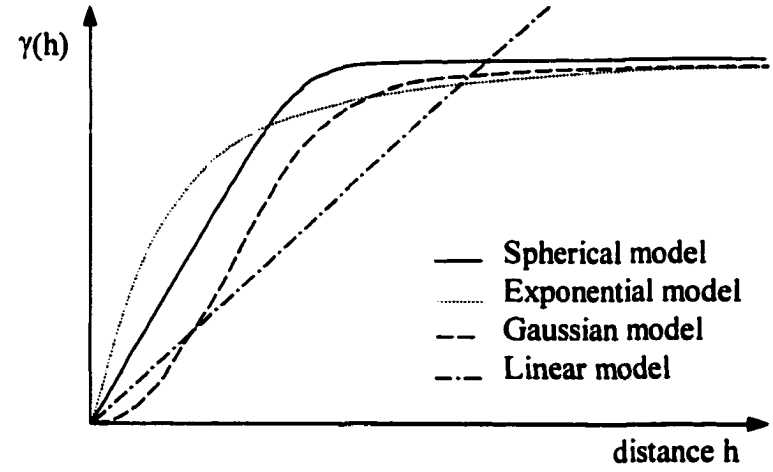

Figure 2.2 The four basic variogram models.

Variogram models can be divided into two groups, namely models with a sill and models without a sill. Models with a sill are called *bounded models* (or transition models). The spherical, exponential and Gaussian models are examples of bounded variogram models. Models without a sill are called *unbounded models* (or non-transition models). The linear model is an example of an unbounded variogram model. Only second-order stationary regionalized variables have bounded variograms, whereas unbounded variograms come from intrinsic variables or non-stationary ones (Armstrong, 1998).

The fitting of the variogram models is done with the help of computer software. Some software packages use weighted least-squares for automatic fitting of the models to the experimental variograms, other programs allow the user to find the model parameters via trial and error.

In some cases the modeling of an experimental variogram calls for a combination of several basic variogram models. The *linear model of regionalization* (Goovaerts, 1997) ensures that any linear combination of permissible models is also admissible, so that the resulting nested variogram is a valid, conditionally negative definite variogram model.

In the case of anisotropy several directional variograms are calculated. Two different types of anisotropy can be distinguished: geometric and zonal anisotropy. *Geometric anisotropy* means that the sills of the variograms are the same in all directions, but the ranges are different. *Zonal anisotropy* means that the ranges of the variograms are the same in all directions, but the sills are different. The various directional models have to be combined into a model that is consistent in all directions. The key idea is to define a transformation that reduces all directional variograms to a common model with a standardized range of 1. Issaks and Srivastava (1987) give a detailed description of this method.

2.4.2.2 The Cross-variogram

A cross-variogram is used in geostatistics to measure how two different variables vary jointly in space. The cross-variogram pairs values of different variables at different locations:

$$\gamma_{ZW}(h) = \frac{1}{2N(h)} \sum_{i,j=1}^{N(h)} [Z(x_i) - Z(x_i + h)] \cdot [W(x_j) - W(x_j + h)] \qquad [2.26]$$

where:

$\gamma_{zw}(h)$ = the cross-variogram value for primary variable Z and secondary variable W separated by distance h,

$Z(x_i)$ = the value of the primary variable Z at position x_i,

$Z(x_i+h)$ = the value of the primary variable Z at position x_i+h,

$W(x_j)$ = the value of the secondary variable W at position x_j,

$W(x_j+h)$ = the value of the secondary variable W at position x_j+h, and

$N(h)$ = the number of pairs of observations separated by distance h.

When dealing with several variables, each variable (the variable of interest, and the auxiliary variables) is characterized by its own variogram, and each pair of variables is characterized by its cross-variogram.

The variance of any linear combination of primary and secondary data values must be positive or equal to zero. To ensure that the variance is non-negative, the variogram and cross-variogram models must be *conditionally negative semi-definite* (Goovaerts, 1997). All of the previously mentioned variogram models fulfill this requirement.

Each of the variograms and cross-variograms can be a combination of several basic variogram models. The difficulty is that these models for the variograms and cross-variograms cannot be built independently of one another. The *linear model of coregionalization* provides a method for modeling the variograms and cross-variograms of several variables so that the estimation variance of any possible linear combination of these variables is always non-negative (Goovaerts, 1997; Issaks & Srivastava, 1989). The linear model of coregionalization is defined as the linear combination of M basic variogram models:

$$\gamma_{ZW}(h) = \sum_{m=1}^{M} b_{ZW}^{m} \cdot \gamma_m(h) \qquad [2.27]$$

where:

$\gamma_{ZW}(h)$ = the nested variogram or cross-variogram of variable Z and/or W,

$\gamma_m(h)$ = the m^{th} basic variogram model, and

b^m_{ZW} = the matrix of coefficients corresponding to the sill of the basic model $\gamma_m(h)$.

The conditions for the matrix $\gamma_{ZW}(h)$ to be a permissible model of coregionalization are (Goovaerts, 1997):

- the basic variogram models $\gamma_m(h)$ are permissible models (as defined above), and
- the matrix of coefficients b_m^{WZ} is *positive semi-definite.*

A symmetric matrix is positive semi-definite if its determinant and all its principal minor determinants are non-negative. In addition, Goovaerts (ibid.) provides four rules for choosing basic models for the linear model of coregionalization:

1) Every basic model that is included in a cross-variogram model must appear in both variogram models.
2) A basic model that is included in both variogram models does not have to be included in the cross-variogram model.
3) A model that is absent on a variogram model must be absent on all cross-variogram involving this variable.
4) None of the variogram and cross-variogram models has to include all basic models.

2.4.3 Kriging Estimators

The simplest application of the regionalized variable theory is an estimation technique known as kriging. Kriging is a "best linear unbiased estimator" (BLUE). This estimation technique is "linear" because its estimates are weighted linear combinations of the available sample data. It is "unbiased" since it tries to have the mean error equal to zero. It is "best" because it aims at minimizing the variance of the errors (Isaaks & Srivastava, 1989). A variogram model is used to determine the kriging weights in such a way that the estimator satisfies the two conditions of unbiasedness and minimized estimation variance. Despite the fact that the spatial correlation is usually modeled using variograms, the mathematical formulas for the kriging methods are normally expressed in terms of covariances. Both correlation measures, however, are related (equation [2.14]). Various kriging methods have been developed over the years. This research focuses on ordinary kriging, universal kriging, cokriging, and disjunctive kriging.

2.4.3.1 Ordinary Kriging

Ordinary kriging (OK) is applied in situations where the mean of the data values is constant (stationary), but unknown. In contrast, simple kriging is a technique where the mean is constant and known.

Unbiasedness Condition

The OK estimator is unbiased. This means that, on average, the estimation error will be zero:

$$E\left[Z^*(x_0) - Z(x_0)\right] = 0 \qquad [2.28]$$

where $Z^*(x_0)$ is the estimated value, and $Z(x_0)$ is the true value at location x_0. To ensure the unbiasedness of the estimates, the sum of the weights λ_i for the n nearest neighbors must equal one:

$$\sum_{i=1}^{n} \lambda_i = 1 \qquad [2.29]$$

Issaks and Srivastava (1987) show the mathematical proof for this unbiasedness condition.

Minimum Variance Condition

In addition to the unbiasedness condition, the OK estimation variance is to be minimized. The estimation variance is the expected value of the square of the difference between the estimated value $Z^*(x_0)$ and the true value $Z(x_0)$ at location x_0:

$$Var\left[Z^*(x_0) - Z(x_0)\right] = \sigma^2 = E\left\{\left[Z^*(x_0) - Z(x_0)\right]^2\right\} = \text{minimum} \qquad [2.30]$$

The OK estimation variance is calculated as (Isaaks & Srivastava, 1989):

$$\sigma_{OK}^2 = C(x_0, x_0) + \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \cdot \lambda_j \cdot C(x_i, x_j) - 2\sum_{i=1}^{n} \lambda_i \cdot C(x_i, x_0) \qquad [2.31]$$

where:

$C(x_0, x_0)$ = the covariance of the point to estimate at location x_0 with itself,

$C(x_i, x_0)$ = the covariance of the sample point at location x_i and the point to estimate at location x_0,

$C(x_i, x_j)$ = the covariance of the sample point at location x_i and the sample point at location x_j,

λ_i = the weight for sample point i at location x_i,

λ_j = the weight for sample point j at location x_j, and

n = the number of sample points.

Issaks and Srivastava (1987) give a detailed mathematical derivation of the estimation variance equation. In order to minimize the estimation variance, subject to the constraint of the unbiasedness condition, a Lagrange parameter μ is introduced into the variance equation. Lagrange parameters are used to convert a constrained minimization problem into an unconstrained one. The equation is then differentiated with respect to λ and μ, and the partial first derivatives are set to zero. This leads to a set of n+1 linear equations with n+1 unknowns. This system of equations is called the ordinary kriging system:

$$\sum_{i=1}^{n} \lambda_i \cdot C(x_i, x_j) + \mu = C(x_i, x_0) \qquad \text{for } i = 1, \ldots, n \qquad [2.32]$$

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

Isaaks and Srivastava (1987) show the matrix formulation of the ordinary kriging system. It should be pointed out that the kriging system has a unique solution only if the variogram models are conditionally negative definite (Goovaerts, 1997).

Ordinary Kriging Estimates

After the weights λ_i and the Lagrange parameter μ have been determined, the OK estimate can be calculated using the following formula:

$$Z^*_{OK}(x_0) = \sum_{i=1}^{n} \lambda_i \cdot Z(x_i) \qquad [2.33]$$

The OK variance is calculated as:

$$\sigma^2_{OK}(x_0) = C(x_0, x_0) - \sum_{i=1}^{n} \lambda_i \cdot C(x_i, x_0) + \mu \qquad [2.34]$$

Ordinary Kriging Flowchart

Figure 2.3 shows the flowchart for the ordinary kriging process. The sample data are used to calculate an experimental variogram, and a variogram model is fitted to the data. Both, the sample data and the variogram model are applied in the ordinary kriging procedure to calculate the kriging weights and Lagrange parameters. The final results are the ordinary kriging estimates and the ordinary kriging variances.

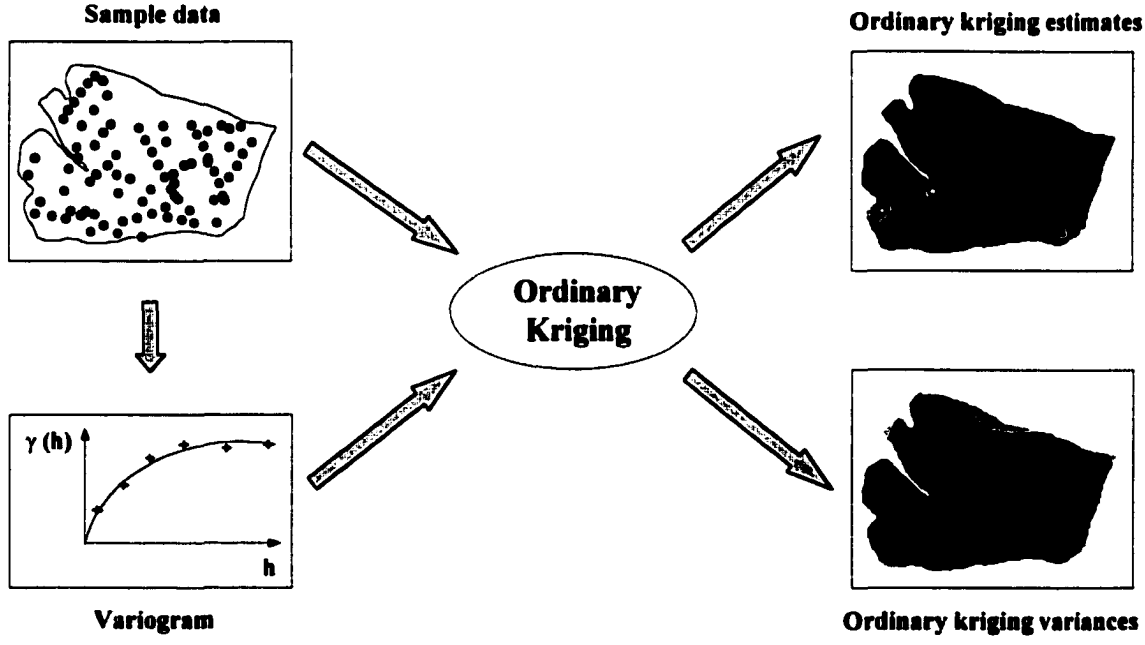

Figure 2.3 Flowchart for ordinary kriging.

Ordinary Kriging Features

The following features of ordinary kriging also apply to other kriging techniques.

1) OK yields unbiased estimates. The kriging weights sum to one to ensure that there is no bias.

2) OK assures that the estimation error variance is minimized.

3) The OK weights take into account the distance from the sample points to the point being estimated (the further away, the less weight) and the clustering of sample points (smaller weights for clustered samples).

4) OK is an exact interpolator. The estimate at a sample point is identical with the data value at that point, i.e. there is no uncertainty at the sample points.

5) OK can yield estimates that are smaller or larger than the smallest or largest sample value, as the kriging weights can be smaller than 0 or larger than 1 (the sum of the weights still has to be 1).

6) Anisotropic patterns of spatial correlation can be taken into account by using directional variograms.

7) The accuracy of OK estimates depends largely on how well the variogram model describes the experimental variogram. Since kriging uses mainly the nearest sample points, the shape of the variogram near the origin is particularly important, e.g. an increasing nugget effect makes the kriging estimates more like a simple averaging of the sample data.

8) OK has a smoothing effect. Smoothing is the result of combining several sample values to calculate an estimate and thereby reduce the variability of the estimated values. This property applies to all interpolation methods, not only kriging.

9) OK can result in negative kriging weights that in turn can lead to negative kriging estimates (Issaks and Srivastava (1987). In many earth science applications, however the variable of interest is always positive. For such variables, negative kriged estimates should be set to zero.

Ordinary Kriging Applications

The estimation method known as kriging was developed by the South African mining engineer Danie G. Krige and the French geomathematician Georges Matheron. D.G. Krige (1951) took the first steps in the development of geostatistics in the Witwatersrand gold mines in South Africa. However, as Cressie (1990, p. 250) writes "*D.G. Krige's contributions in mining engineering were considerable but he did not discover kriging.*" Credit for the development of kriging is given to G. Matheron. He further improved Krige's findings and developed the theory of regionalized variables which describes the important aspects of geostatistics. His two-volume work "Traité de Géostatistique Appliqué" (1962, 1963) laid the foundations of geostatistics. In honor of D.G. Krige's contributions to related areas of mining, Matheron chose the term "kriging" for his new geostatistical estimation method. Many other researchers in many disciplines have contributed to the problem of optimal spatial prediction, including forestry: Matern (1960) relates that a Swedish forester used what is essentially the variogram to express the variation in forest surveys as early as 1926. Matern (1960) applied spatial correlation

measures to the distribution of Swedish forests in 1947. Another contribution in the field of meteorology should be mentioned: Gandin (1963) independently developed "kriging" in the former USSR (Cressie, 1990). For a more comprehensive description on the history of kriging see Cressie (ibid.).

Originally applied in the mining industry, kriging soon found its way into many other disciplines, notably soil science (Burgess & Webster, 1980; Heuvelink & Bierkens, 1992; Goovaerts & Journel, 1995), hydrology (Kitanidis & Vomvoris, 1983; Aboufirassi & Marino, 1983), ecology (Orloci, 1978; Robertson, 1987; Rossi et al., 1992; Pohlmann, 1993), atmospheric science (Lajaunie, 1984; Hudson, 1993), remote sensing (Rossi et al., 1994; Stonge & Cavayas, 1995), and forestry.

The first people applying OK for forestry purposes were Guibal (1973), followed by Marbeau (1976). Both used OK in forest inventory applications. Bouchon (1979) estimated the surface area and conducted a structural analysis of forest stands using OK.

Biondi et al. (1989) applied OK to estimate stem diameters, basal area, and 10-year periodic basal area increments in an old-growth stand of southwestern ponderosa pine (*Pinus ponderosa* Dougl. *Ex.* Laws. *var. scopulorum*) in northern Arizona. The temporal variation of the three variables was evaluated at 10-year intervals from 1920 to 1990. However, no validation was conducted to evaluate the kriging results.

Höck et al. (1993) estimated the site indices in Kaingaroa Forest in New Zealand. Site index was defined as the top height (measured in meters) that a stand of trees attains at age 20 years. A cross-validation procedure using all 722 sample points showed the good agreement between the actual values and the estimated values.

Holmgren and Thuresson (1995) applied OK to predict total wood volume, hardwood volume, and two monetary measures ("inoptimality-losses" for thinning and clearcutting treatments) in a forest estate in northern Sweden with Norway spruce (*Picea abies*) as the dominant tree species. The kriging results were not evaluated. The kriging estimates were combined with estimates from image analysis of digital aerial photos. These combined estimates were used to allocate tactical treatment units in forest management planning.

Gilbert and Lowell (1997) used OK (and an interpolation method called area-stealing) to estimate forest volume in a balsam fir - white birch (*Abies balsamea* L. - *Betula papyrifera* Marsh.) forest in Quebec, Canada. Correlograms were calculated from three relatively dense sampling schemata. The data set was divided into two parts: 80% of the data were used to conduct the interpolation, 20% of the data were used to validate the interpolation results. Despite a spatial correlation of tree volume (as depicted in the correlograms) the resulting kriging surfaces were no more precise than 80% with 95% confidence even for a forest sample having an intensity of 16%. This result was only a slight improvement over using the mean of all sample plots as an estimate for each sample location.

2.4.3.2 Universal Kriging

A serious restriction of ordinary kriging is the requirement of a stationary mean over the study area. Many natural phenomena, however, are known to be non-stationary. Ordinary kriging should not be used in the presence of a strong trend, a change in average value, as it yields erroneous and biased results. Universal kriging is a form of

interpolation that takes account of local trends in the data when minimizing the error associated with the estimation. Universal kriging is also called *kriging with a trend* (Goovaerts, 1997; Journel & Rossi, 1989). A non-stationary variable Z at location x can be expressed as the sum of a varying trend m(x) and the residuals ε(x):

$$Z(x) = m(x) + \varepsilon(x) \qquad [2.35]$$

where ε(x) is assumed to have a mean of zero, and be second-order or intrinsically stationary. Universal kriging consists of three stages: first, the trend must be estimated and subtracted from the original data values. Second, the residual-based variogram has to be derived, and the residuals are estimated. And third, the estimated residuals are combined with the trend surface to obtain the final estimates of the actual surface.

Modeling the Trend

The trend is assumed to be a linear combination of the form:

$$m(x) = \sum_{k=0}^{K} a_k \cdot f_k(x) \qquad \text{for k} = 0, \ldots, \text{K} \qquad [2.36]$$

where f is any known function, and a is the unknown coefficient. The trend is modeled as a function of the spatial coordinates whose unknown parameters are derived from the data. The estimated trend at any location is obtained by combining the estimated coefficients with the values of the functions at that location. First- and second-degree polynomials are typically used for the function f_k. A linear trend consists of three (k = 2) functions and coefficients; a quadratic trend of six (k = 5) functions and coefficients:

$$\text{linear trend: } m(x) = a_0 + a_1x + a_2y, \text{ and} \qquad [2.37]$$

$$\text{quadratic trend: } m(x) = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy, \qquad [2.38]$$

where x are the x-coordinates, and y are the y-coordinates.

Unbiasedness Condition

The estimated trend is subtracted from the original data values. The resulting residuals are then used in an ordinary kriging procedure. In order to obtain unbiased estimates from UK the following constraint must be satisfied:

$$\sum_{i=1}^{n} \lambda_i^R = 1 \qquad [2.39]$$

where λ_i^R are the weights of the n nearby residual values.

Minimum Variance Condition

The estimation variance of the residuals is calculated as:

$$\sigma_R^2 = C^R(x_0, x_0) + \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i^R \cdot \lambda_j^R \cdot C^R(x_i, x_j) - 2\sum_{i=1}^{n} \lambda_i^R \cdot C^R(x_i, x_0) \qquad [2.40]$$

where:

$C^R(x_0, x_0)$ = the covariance of the residual to estimate at location x_0 with itself,

$C^R(x_i, x_0)$ = the covariance of the residual at location x_i and the residual to estimate at location x_0,

$C^R(x_i, x_j)$ = the covariance of the residual at location x_i and the residual at location x_j,

λ_i^R = the weight for the residual i at location x_i,

λ_j^R = the weight for the residual j at location x_j, and

n = the number of residuals.

In order to minimize the estimation variance, subject to the constraint of the unbiasedness conditions, a Lagrange multiplier is introduced into the equation. The equation is differentiated with respect to the weights λ and the Lagrange parameter μ, and the partial first derivatives are set to zero. This leads to a set of n+1 equations with n+1 unknowns:

$$\sum_{i=1}^{n} \lambda_i^R \cdot C^R(x_i, x_j) + \mu = C^R(x_i, x_0) \qquad \text{for } i = 1, \ldots, n \qquad [2.41]$$

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

Universal Kriging Estimates

After this system of equations has been solved and the weights and the Lagrange parameter determined, the estimates for the residuals can be calculated using the following formula:

$$Z^{R*}{}_{UK}(x_0) = \sum_{i=1}^{n} \lambda_i^R \cdot Z^R(x_i) \qquad [2.42]$$

The estimated residuals are then added to the trend component. The UK variance is calculated as:

$$\sigma_{UK}^2(x_0) = C^R(x_0, x_0) - \sum_{i=1}^{n} \lambda_i^R \cdot C^R(x_i, x_0) + \mu \qquad [2.43]$$

Variogram Estimation

In general the inference of the residual variogram γ^R is not straightforward (Goovaerts, 1997; Wackernagel, 1995). The problem is that the available data are the original Z values and not residual values. One solution to the problem (applied in this project) makes use of the residuals to estimate a variogram. Cressie (1991) and Armstrong (1984) point out that the residual-based variogram is biased as the estimated residual-based variogram differs from the underlying variogram of the true residuals. The bias is small at small lag distances near the origin, and larger at distant lags. However, kriging mostly uses the nearest data values where the bias is small. The universally

kriged estimate may be little influenced by the bias, whereas the universally kriged variance may be smaller than it should be (Cressie, 1991). Another approach uses the fact that the available Z variogram is related to the residual variogram (Cressie, 1991). Therefore, the residual variogram can be inferred from the original Z variogram. The method infers the variogram from data pairs that are unaffected or slightly affected by the trend. Data pairs for small separation distances are generally less affected by a trend than data pairs a large distance apart. For larger distances data pairs in subareas or along directions with little or no trend could be chosen. The residual variogram can then be directly inferred from the sample data. A third method differences the data values and thereby filters out the trend. This method is known as "intrinsic random functions of order k" (Matheron, 1973; David, 1988; Cressie, 1991) and it uses generalized covariances. However this method causes problems when data are not gridded, and often results in very large (artifact) relative nugget effects (Goovaerts, 1997).

Universal Kriging Flowchart

Figure 2.4 shows the flowchart for the universal kriging process. The spatial coordinates of the sample data are applied in a trend surface analysis. The resulting trend surface is subtracted from the sample data leaving the residuals for each sample location. These residuals are then used to calculate an experimental variogram with a variogram model fitted to the data. The residual values and the residual variogram model are applied in an ordinary kriging process. The results are estimates of the residuals and estimates of the kriging variances. In a final step, the estimated residuals are added to the trend surface to obtain the final estimates.

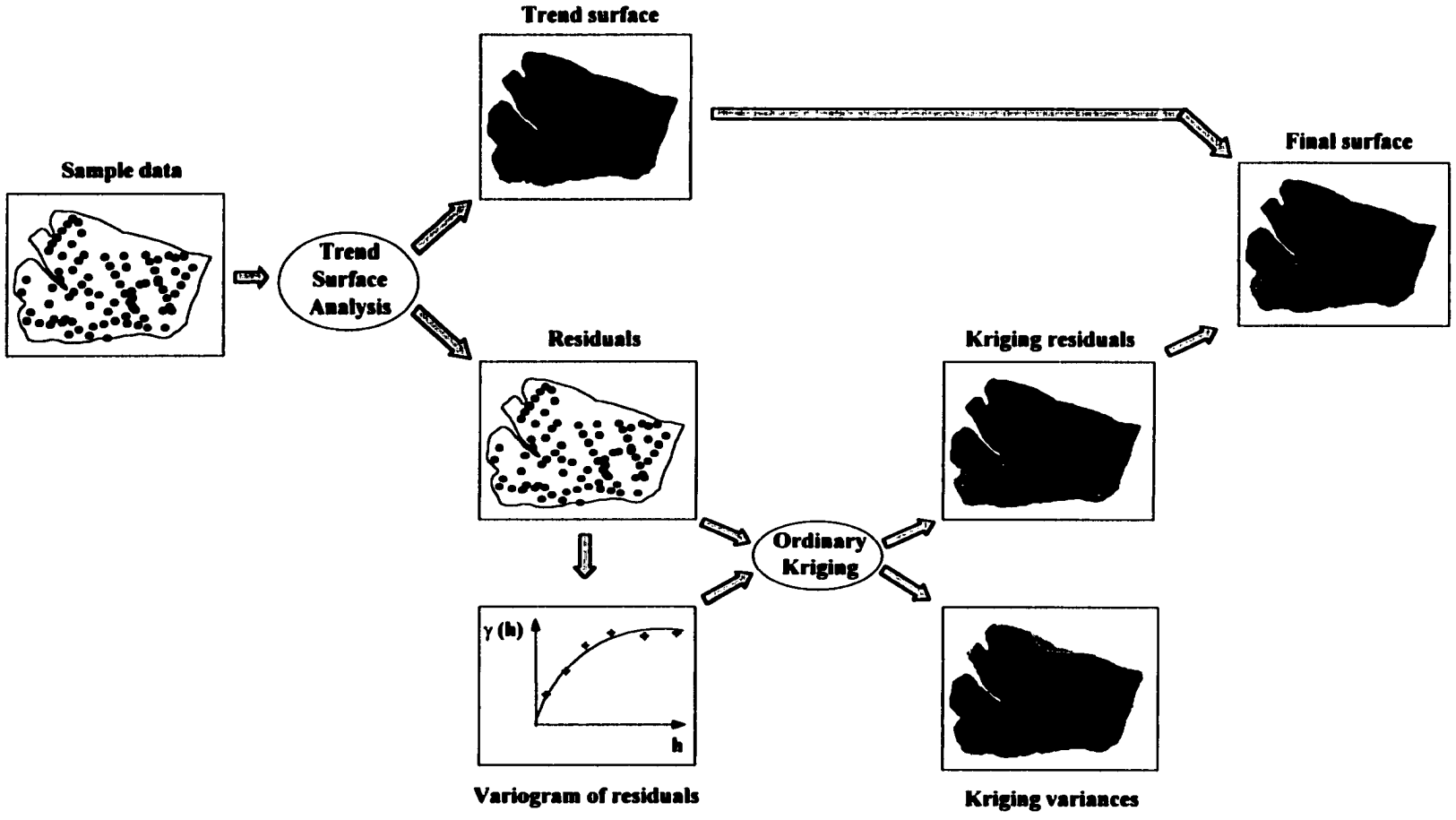

Figure 2.4 Flowchart for universal kriging.

Universal Kriging Features

1) UK yields unbiased estimates with minimized error variances in cases where a trend is present.

2) There are some difficulties in deriving the underlying variogram. Armstrong (1984) gives a good overview of the problems involved.

Applications of Universal Kriging

The principles of UK were an extension by Matheron (1969) of his earlier regionalized variable theory. The first paper on UK published in English was by Huijbregts and Matheron in 1971. Since then UK has been applied in many different fields: mining (Dagbert & David, 1976; Song et al. 1992), soil sciences (Webster & Burgis, 1980; Stein & Corsten, 1991), contour mapping (Olea, 1974), earthquake research (Carr & Roberts, 1989), ecology (Ver Hoef, 1993), and forestry.

Samra et al. (1989) estimated tree heights after 1, 2, and 3 years of growth in a plantation of Dharek (*Melia azedarach* Linn.) in India using UK. A trend in the data was removed by a method called median polishing. Kriging was then conducted on the median-polished values. The trend kept aside was put back by adding the medians to the kriged values. Cross-validation results showed the good quality of the UK procedure.

Kallas (1997) created a hazard rating map of Armellaria root disease in the Black Hill of South Dakota. She fitted a second-degree trend surface to her data including x- and y-coordinates, elevation, slope, aspect, and precipitation as independent variables.

Kriging was applied to the residuals of the trend surface. Cross-validated sample values were used to calculate an R^2 value. An R^2 value of 0.952 could be reached.

Metzger (1997) used UK (and cokriging) to model basal area, percent open canopy, and height of the understory vegetation on the Austin Cary Memorial Forest in northern Florida. The trend in the data was removed by fitting a first- and second-degree trend surface to the data using x- and y-coordinates and satellite data as the independent variables. The residuals of the cross-validated sample values were used to calculate an R^2 value of 0.58 for basal area, 0.69 for percent open canopy, and 0.42 for the height of the understory vegetation.

2.4.3.3 Cokriging

Cokriging (CK) utilizes multivariate information to determine estimates at unsampled locations. It is a method of unbiased estimation that minimizes the variance of the estimation errors by exploiting the spatial cross-correlation between the variable of interest and a set of secondary variables. The estimates are derived using a weighted linear combination of primary and secondary data values. Variograms for the primary and each secondary variable, as well as cross-variograms for each combination of these variables are used to determine the estimation weights. As with ordinary kriging, CK assumes stationarity of both primary and secondary variables. Wackernagel (1995) distinguishes three different situations where cokriging can be applied, depending on the location of the primary and secondary variables:

- Complete heterotrophy: primary and secondary variables have been measured on different groups of sample points and have no sample locations in common. This case causes some problems for inferring the cross-variograms.
- Partial heterotrophy: some of the variables have some sample locations in common.
- Isotrophy: both, primary and secondary variables are available at all sample locations. Although it is possible to have more than one secondary variable, only one secondary variable will be considered here for the explanation of the mathematical background of cokriging.

Unbiasedness Condition

The weights of the primary and secondary variables must be such that the CK estimate is unbiased. Cokriging applies the following two unbiasedness conditions (Isaaks & Srivastava, 1989):

$$\sum_{i=1}^{n} \lambda_i^Z = 1 \quad \textit{and} \quad \sum_{j=1}^{m} \lambda_j^W = 0 \qquad [2.44]$$

where λ_i^Z are the weights for the n data values of the primary variable Z, and λ_j^W are the weights for the m data values of the secondary variable W. The first unbiasedness condition requires that there is at least one sample value of the primary variable Z.

Minimum Variance Condition

The CK weights must be such that the error variance of the estimates is minimized. The CK estimation variance is calculated as (Isaaks & Srivastava, 1989):

$$\sigma_{CK}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i^Z \cdot \lambda_j^Z \cdot C^Z\left(x_i^Z, x_j^Z\right) + \sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i^W \cdot \lambda_j^W \cdot C^W\left(x_i^W, x_j^W\right) - 2\sum_{i=1}^{n}\lambda_i^Z \cdot C^Z\left(x_i^Z, x_0^Z\right) - 2\sum_{j=1}^{m}\lambda_j^W \cdot C^W\left(x_j^W, x_0^W\right) + 2\sum_{i=1}^{n}\sum_{j=1}^{m}\lambda_i^Z \cdot \lambda_j^W \cdot C^{ZW}\left(x_i^Z, x_j^W\right) + C^Z\left(x_0^Z, x_0^Z\right) \quad [2.45]$$

where:

$C^Z(x_i^Z, x_j^Z)$ = the covariance of primary variable Z at location x_i and primary variable Z at location x_j,

$C^W(x_i^W, x_j^W)$ = the covariance of secondary variable W at location x_i and secondary variable W at location x_j,

$C^Z(x_i^Z, x_0^Z)$ = the covariance of primary variable Z at location x_i and primary variable Z at unsampled location x_0,

$C^W(x_j^W, x_0^W)$ = the covariance of secondary variable W at location x_j and secondary variable W at unsampled location x_0,

$C^{ZW}(x_i^Z, x_j^W)$ = the cross-covariance of primary variable Z at location x_i and secondary variable W at location x_j,

$C^Z(x_0^Z, x_0^Z)$ = the covariance of primary variable Z at unsampled location x_0 with itself,

λ_i^Z and λ_j^Z = the weights for primary variable Z at location x_i or x_j,

λ_i^W and λ_j^W = the weights for secondary variable W at location x_i or x_j,

n = the number of primary sample values Z, and

m = the number of secondary sample values W.

In order to minimize the CK estimation variance two Lagrange parameters μ_1 and μ_2 (one for the primary variable, and one for the secondary variable) are introduced into the equation. The equation is differentiated with respect to the weights λ and the two Lagrange parameters μ_1 and μ_2. Then the partial first derivatives are set to zero. This leads to a set of (n+m+2) equations with (n+m+2) unknowns. This set of equations is called the cokriging system:

$$\sum_{i=1}^{n} \lambda_i^Z \cdot C^Z\left(x_i^Z, x_j^Z\right) + \sum_{i=1}^{m} \lambda_j^W \cdot C^{WZ}\left(x_i^W, x_j^Z\right) + \mu_1 = C^Z\left(x_0^Z, x_j^Z\right) \quad \text{for j = 1, ..., n}$$

$$\sum_{i=1}^{n} \lambda_i^W \cdot C^W\left(x_i^W, x_j^W\right) + \sum_{i=1}^{n} \lambda_i^Z \cdot C^{ZW}\left(x_i^Z, x_j^W\right) + \mu_2 = C^{ZW}\left(x_0^Z, x_j^W\right) \quad \text{for j = 1, ..., m}$$

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

$$\sum_{j=1}^{m} \lambda_j^W = 0 \qquad [2.46]$$

For a matrix formulation of the CK system the reader is referred to Myers (1982) and his follow-up paper (Myers, 1983). It should be pointed out that the cokriging system has a unique solution only if the underlying variogram and cross-variogram models are conditionally negative semi-definite (see Chapter 2.4.2.2, "The Cross-Variogram," for details).

Cokriging Estimates

Once the weights and the Lagrange parameters have been determined the CK estimate and the CK variance can be calculated. The CK estimate Z at location x_0 is a weighted linear combination of primary and secondary data values:

$$Z_{CK}^{*}(x_0) = \sum_{i=1}^{n} \lambda_i^Z \cdot Z(x_i) + \sum_{j=1}^{m} \lambda_j^W \cdot W(x_j) \qquad [2.47]$$

The CK variance can be calculated using the following formula:

$$\sigma_{CK}^2(x_0) = C^Z\left(x_0^Z, x_0^Z\right) + \mu_1 - \sum_{i=1}^{n} \lambda_i^Z \cdot C^Z\left(x_i^Z, x_0^Z\right) - \sum_{j=1}^{m} \lambda_j^W \cdot C^{WZ}\left(x_j^W, x_0^Z\right) \qquad [2.48]$$

Cokriging Flowchart

Figure 2.5 shows the flowchart of the cokriging process. Primary and secondary sample values are used to calculate their respective variograms. In addition, both, primary and auxiliary data are utilized to derive a cross-variogram. The sample data of the primary and secondary variable, the variograms for both variables, as well as the cross-variogram are applied in the cokriging procedure. The final results are the cokriging estimates and the cokriging variances.

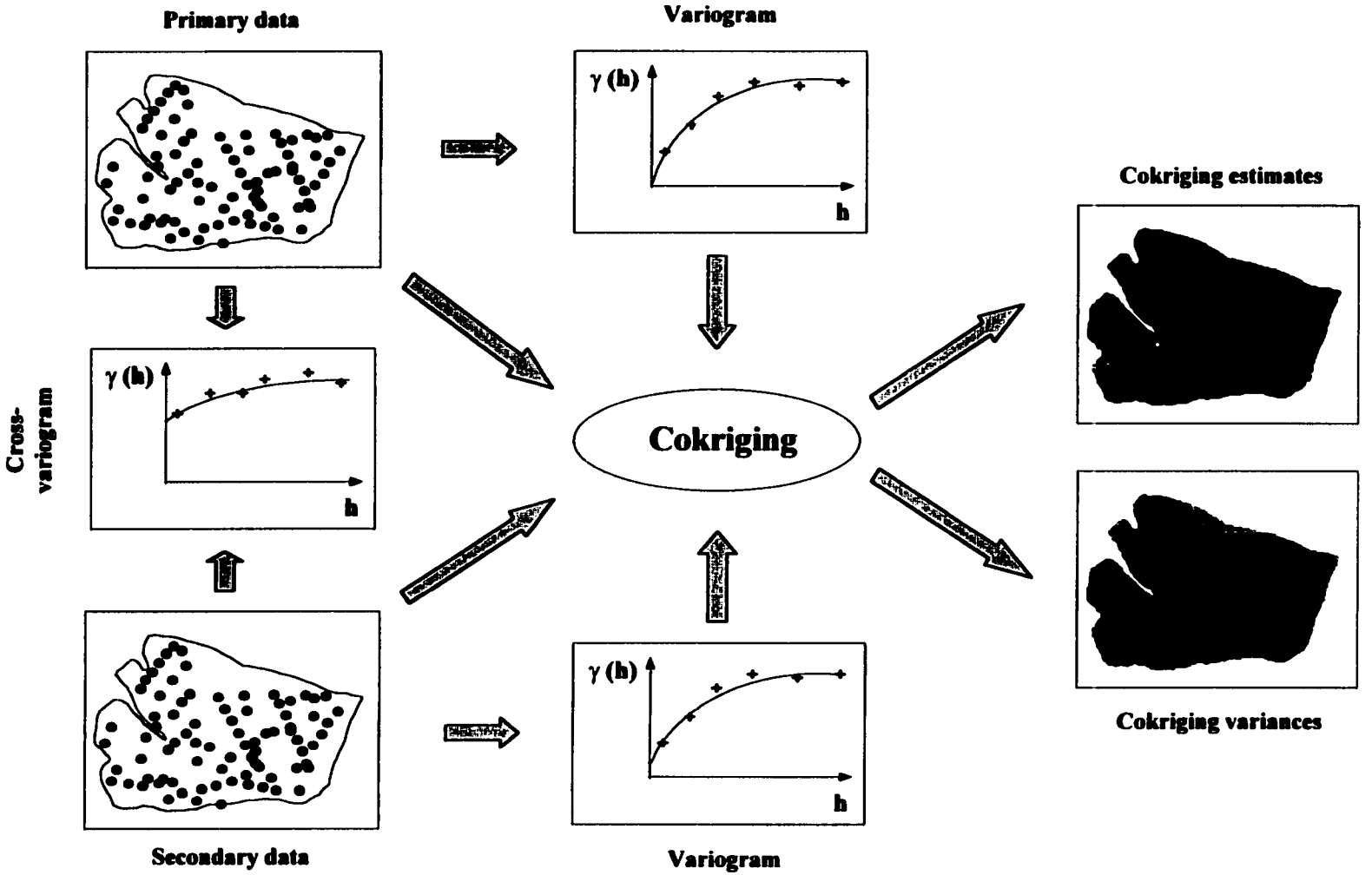

Figure 2.5 Flowchart for cokriging.

Cokriging Features

1) CK uses a primary variable of interest and spatially cross-correlated secondary variables to yield unbiased estimates with minimized error variances.

2) By making use of the spatial cross-correlation between primary and secondary variables the quality of the kriging estimates should be improved and the estimation variances should be smaller than for the ordinary or universal kriging.

3) CK is a most useful estimation technique when the variable of interest is difficult or costly to sample, while a secondary, correlated variable can be easily and inexpensively sampled. The usefulness of secondary variables is often enhanced by the fact that the primary variable of interest is undersampled.

4) The quality of the CK estimates depends largely on the accuracy of the variogram and cross-variogram models. Errors in these models lead to errors in the CK estimates and variances. Furthermore, variograms and cross-variograms cannot be modeled independently from one another. The linear model of coregionalization provides a method for modeling the variograms and cross-variograms of several variables so that the estimation variance of any possible linear combination of these variables is always non-negative (Goovaerts, 1997). With increasing numbers of auxiliary variables the variogram modeling process becomes more difficult and cumbersome.

5) Isaaks & Srivastava (1989) point out that there are certain situations where CK will not improve an ordinary kriging estimate. If all the variogram models are "quite similar" in shape and the primary variable is not noticeably undersampled, CK will not improve the estimation very much.

Cokriging Applications

The first literature describing CK seems to be Journel and Huijbregts in 1978. Since then CK has been applied in many different fields: soil science (Vauclin et al., 1983; McBratney & Webster, 1983; Papritz & Fluhler, 1994), hydrology (Boufirassi & Marino, 1984; Solow & Gorelick, 1986; Leenaers et al., 1989), atmospheric science (Krajewski, 1987), earthquake research (Carr & McCallister, 1985), remote sensing (Atkinson et al., 1992), mining (Pan et al., 1993), and forestry.

Kallas (1997) used CK (and universal kriging) to create a hazard rating map of *Armillaria Root Disease* on the Black Hills National Forest in South Dakota using site productivity as secondary variable. The cross-validation technique was used to derive residuals for each sample point. An R^2 value then computed. Ordinary cokriging yielded an R^2 value of 0.97.

Metzger (1997) used CK (and universal kriging) to model basal area, percent open canopy, and height of the understory vegetation on the Austin Cary Memorial Forest in northern Florida. She used field sample data and combined them with the seven spectral bands of a Landsat-TM satellite image. The CK models had an R^2 value of 0.77 for basal area, an R^2 value of 0.97 for percent open canopy, and an R^2 value of 0.84 for the height of the understory vegeation.

Phillips et al. (1998) used CK for estimating ozone exposure in southeastern forests as the primary variable and monthly data on anthropogenic emissions of nitrous oxides (NO_x), average daily maximum temperature, wind directional frequencies, and distance downwind from anthropogenic NO_x sources as auxiliary variables. In a

comparison with inverse distance weighting and ordinary kriging, CK exhibited the best estimation.

2.4.3.4 Disjunctive Kriging

All the geostatistical methods mentioned so far belong to the field of linear geostatistics, i.e. these methods deal only with linear combinations of the variable under study. This makes it possible to linearly estimate the value of the variable at a given point. Linear kriging methods yield estimates that have the smallest estimation variance among all unbiased linear estimators. However, these linear estimators may only be optimal when the variable under study has a normal distribution (Rendu, 1980). In contrast, non-linear geostatistical methods can be applied to variables that exhibit any non-normal distribution. Disjunctive kriging (DK) is a non-linear unbiased estimation method. It is better, in general, than linear estimators in providing minimum error variance estimates of a property through non-linear combinations of the data and exactness of estimation (Rivoirard, 1994). DK allows the calculation of conditional probabilities, that a variable is greater (or smaller) than some prescribed cut-off or tolerance value. Instead of the weighted linear combination of samples for kriging, DK could be viewed as a form of the more general estimation:

$$Z^*(x_0) = \sum_{i=1}^{n} f_i \cdot Z(x_i) \qquad [2.49]$$

where each f_i is a function of one sample value $Z(x_i)$. In using linear kriging, each f_i is a linear function and only the coefficients need to be determined. For each function f_i in the DK case, as opposed to finding coefficients, the appropriate functions must be found.

Proper choice of the functions f_i assures the unbiasedness of the estimates and the minimization of the estimation error variances.

Two different spaces are involved with DK: one space with the original sample values Z(x), and another space with the corresponding transformed values Y(x). Whereas the original data Z(x) can have any distribution, the transformed values Y(x) have a normal standard distribution with a mean of 0 and a variance of 1. The DK estimator is made up of a series of non-linear functions where each function depends on only one normalized sample value $Y(x_i)$:

$$Z^*_{DK}(x_0) = \sum_{i=1}^{n} f_i \cdot [Y(x_i)] = \sum_{i=1}^{n} \sum_{k=0}^{\infty} f_{ik} \cdot H_k[Y(x_i)] \qquad [2.50]$$

where $f_i[Y(x_i)]$ are the functions to be determined, and n is the number of sample points. These functions are expressed on the right hand side of equation [2.50] as a series of Hermite polynomials $H_k[Y(x_i)]$, where f_{ik} is a constant which depends on i and k. In other words, in DK the variable Z at location x_0 to be estimated is decomposed into a sum of disjoint, uncorrelated components (Hermite polynomials) of sample values.

Normalization of Data

The original data values Z(x) of any arbitrary distribution must be transformed into a new variable Y(x), that has a standard normal distribution with a mean of zero and unit variance. It is further assumed that the transformed variable has a bivariate normal distribution for each pair of variables. For each original data value Z(x) one must calculate a normalized value Y(x). Given n values of Z(x), sorted in increasing order, the probability p that Z(x) is less than or equal to any value $Z(x_0)$ is calculated as:

$$p = (i_0 - 0.5)/n \qquad [2.51]$$

where i_0 is the number of sample values less than or equal to $Z(x_0)$. The value $Y(x_i)$ corresponding to each p_i can be found in a table of the cumulative distribution of the standard normal variate. A graphical representation of the procedure is shown in Figure 2.6.

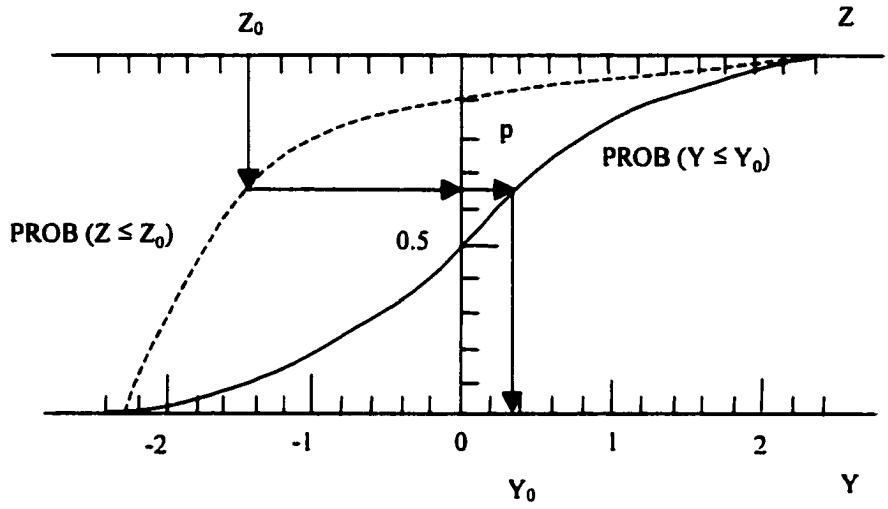

Figure 2.6 Graphical transformation of original data to normal distribution (after Hohn, 1998).

The original values Z(x) and the normalized values Y(x) are related through a function $Z(x) = \phi\,[Y(x)]$, that is assumed to exist, to be unique and to be invertible (Yates et al., 1986). This transformation function can be expressed as a linear combination of Hermite polynomials:

$$Z(x) = \phi[Y(x)] = \sum_{k=0}^{\infty} C_k \cdot H_k[Y(x)] \qquad [2.52]$$

where $H_k[Y(x)]$ is a Hermite polynomial of order k (see below), and the C_k's are the Hermite coefficients (see below). Kim et al. (1977) show the mathematical proof that

such a transformation function always exists between the arbitrary Z(x) values and the normally distributed Y(x) values.

Hermite Polynomials

The normalized values are used in a limited expansion of Hermite polynomials that approximate the transformation function $\phi[Y(x)]$. Many functions can be represented by an infinite series of Hermite polynomials. Hermite polynomials are orthogonal with respect to the standard normal distribution. The Hermite polynomial $H_k[Y(x)]$ of order k is defined by Rodriques' Formula:

$$H_k(y) = (-1)^k \cdot e^{[y^2/2]} \cdot \left(d^k / dy^k\right) \cdot e^{[-y^2/2]} \qquad [2.53]$$

which is easily calculated using the recurrence relationship:

$$H_{k+1}(y) = y \cdot H_k(y) - k \cdot H_{k-1}(y) \qquad [2.54]$$

where $H_0(y) = 1$, and $H_1(y) = y$.

A justification for using Hermite polynomials to express $\phi[Y(x)]$ is beyond the scope of this paper. The interested reader is referred to Kim et al. (1977), and Journel and Huijbregts (1978).

Hermite Coefficients

The Hermite coefficients are derived from the sample value distribution. Using the orthogonality properties of Hermite polynomials the coefficients C_k in equation [2.54] can be calculated through Hermite integration:

$$C_k = \frac{1}{k!\sqrt{(2\pi)}} \sum_{j=1}^{J} \omega_j \cdot \phi(y_j) \cdot H_k(y_j) \cdot \exp[-y_j^2/2] \qquad [2.55]$$

where y_j is an abscissa and each w_j is a corresponding weight. To calculate this numerical integration it is necessary to determine the values of $\phi(y_j)$ for only a few specific values of y; normally J=10 is used (Rendu, 1980). These values y_j and the corresponding w_j values are tabulated in Abramowitz and Stegun (1970). Given the y_j values, the corresponding values $z_j = \phi(y_j)$ have to be derived by using a procedure inverse to the one given in the "Normalization of Data" subsection: from y_j and the cumulative frequency distribution of the standard normal a value p_j is calculated. The value p_j is used to interpolate between the values of z_j that bracket p_j to give $\phi(y_j)$ (Hohn, 1998). The interpolation is done by fitting an n^{th} order polynomial to the [Z(x), Y(x)] data pairs, or some other method (Yates et al., 1986).

Rendu (1980) writes that the choice of K depends on the complexity of the sample value distribution and the correlation between the sample values and the values to be estimated. In practice, only a finite number of coefficients k is used. To find the optimal number K the mean and variance of the transformed data are used. If $\phi[Y(x)]$ is a standard normal function, the mean and variance of the transformed data can be found from the coefficients C_k. The mean value is calculated as:

$$\mu = E[\phi(Y)] = C_0 \qquad [2.56]$$

And the variance of the data is:

$$\sigma^2 = Var[\phi(Y)] = \sum_{k=1}^{\infty} k! \cdot C_k^2 \qquad [2.57]$$

If the mean and the variance of the Hermite polynomials are close to the sample mean and sample variance, the approximation with K Hermite polynomials is good.

Variogram Modeling

For the DK estimation process it is necessary to determine the spatial autocorrelation of the normalized values Y(x). Because the normalized values have a variance of 1, the sill should be 1, and therefore the correlogram can be used to model the spatial variability (see chapter 2.4.1 ("The Regionalized Variable Theory") for more details on correlograms). However, under the assumption of second-order stationarity the correlogram $\rho(h)$ is related to the variogam $\gamma(h)$ by:

$$\rho(h) = 1 - \gamma(h) \qquad [2.58]$$

Hence, as with ordinary kriging an experimental variogram is calculated and a variogram model fitted to the data.

Disjunctive Kriging Estimation

An unbiased estimator with minimum estimation variance is given by the following equation:

$$Z^*_{DK}(x_0) = \sum_{k=0}^{K} C_k \cdot H^*_k[Y(x_0)] \qquad [2.59]$$

where the series has to be truncated to K terms (as explained above), and the $H_k^*[Y(x_0)]$ represent the estimated value of the k^{th} Hermite polynomial at the estimation location x_0. The sum of these estimates multiplied by the coefficients C_k (calculated from the sample value distribution as shown above) makes up the disjunctive kriging estimate at the unsampled location x_0.

The estimated value of the k^{th} Hermite polynomial at the estimation location x_0 is calculated from the surrounding normalized values y_i (i = 1, 2, ..., n) as:

$$H_k^*[Y(x_0)] = \sum_{i=1}^{n} \lambda_{ik} \cdot H_k[Y(x_i)] \qquad [2.60]$$

where λ_{ik} are the disjunctive kriging weights, and $H_k^*[Y(x_i)]$ is the k^{th} Hermite polynomial of the normalized sample value at location x_i. The optimal (in terms of minimum error variance) disjunctive kriging weights λ_{ik} for each normalized sample point and each Hermite polynomial can be found by solving the following system of linear kriging equations:

$$\sum_{j=1}^{n} \lambda_{ik} \cdot \{\rho(x_i, x_j)\}^k = \{\rho(x_0, x_j)\}^k \quad \text{for i} = 1,...,n \quad \text{and} \quad k = 0,...,K \qquad [2.61]$$

where:

$\{\rho(x_i, x_j)\}^k$ = the correlation of the normalized value at location x_i and normalized value at location x_j for the k^{th} Hermite polynomial, and

$\{\rho(x_0, x_j)\}^k$ = the correlation between the normalized value to estimate at location x_0 and normalized value at location x_j for the k^{th} Hermite polynomial.

Hohn (1998) mentions that when large values of K are considered, the value $\{\rho(x_0, x_j)\}^k$ approaches zero, and the weights λ_{ik} also tend to zero. Therefore large values of K are unnecessary. For k=0 equation [2.61] represents the unbiasedness condition:

$$\sum_{i=1}^{n} \lambda_{i0} = 1 \qquad [2.62]$$

The disjunctive kriging variance is given by:

$$\sigma_{DK}^2 = \sum_{k=1}^{K} k! \cdot C_k^2 \left[1 - \sum_{i=1}^{n} \lambda_{ik} \cdot \{\rho(x_0, x_i)\}^k\right] \qquad [2.63]$$

Conditional Probabilities

Many decision-makers (e.g. land managers, legislators) must often make decisions based on critical thresholds. An advantage of disjunctive kriging is the ability to calculate the conditional probability that the value at an estimation site is greater than (or less than) an arbitrary critical value, thereby giving the decision maker a means to judge the risk of taking the estimates at its face value (Oliver et al., 1996). Estimating the conditional probability is possible since the DK estimator is nonlinear.

The probability that the estimate Z at location x_0 is above the cutoff value z_c, given the observed values $Z(x_i)$ with i=1, 2, ..., n, is represented by:

$$P(x_0) = \text{Prob}\{Z(x_0) \geq z_c \mid Z(x_1), Z(x_2), \ldots, Z(x_n)\} \qquad [2.64]$$

The probability $P(x_0)$ can be calculated using Hermite polynomials and an indicator function. The indicator function is based on the transformed cutoff value y_c: $\Theta_{Yc}[Y(x_i)]$. It takes a value of 1 if $Y(x_i) \geq y_c$ and 0 otherwise. The indicator function can be expanded by a series of Hermite polynomials. The following equation gives the probability $P(x_0)$ that the value of Z at x_0 equals or exceeds the threshold z_c in terms of the indicator function (Yates et al., 1986):

$$P(x_0) = \Theta_{Yc}[Y(x_0)] = 1 - G(y_c) + g(y_c)\sum_{k=1}^{K} H_{k-1}(y_c)\cdot H_k^*[Y(x_0)]/k! \qquad [2.65]$$

where $G(y_c)$ is the distribution function for the normal distribution, and $g(y_c)$ is the normal probability density function. Yates et al. (1986) describe the mathematical derivation of the conditional probability equation.

Disjunctive Kriging Flowchart

Figure 2.7 gives an overview for the disjunctive kriging process. Sample data with any arbitrary distribution are transformed to a normal distribution. The normalized values are utilized to calculate the Hermite coefficients, the Hermite polynomials, and a variogram. The normalized variogram model, as well as the Hermite coefficients and polynomials are applied in the disjunctive kriging procedure. The final results are the disjunctive kriging estimates, the disjunctive kriging variances, and the conditional probabilities.

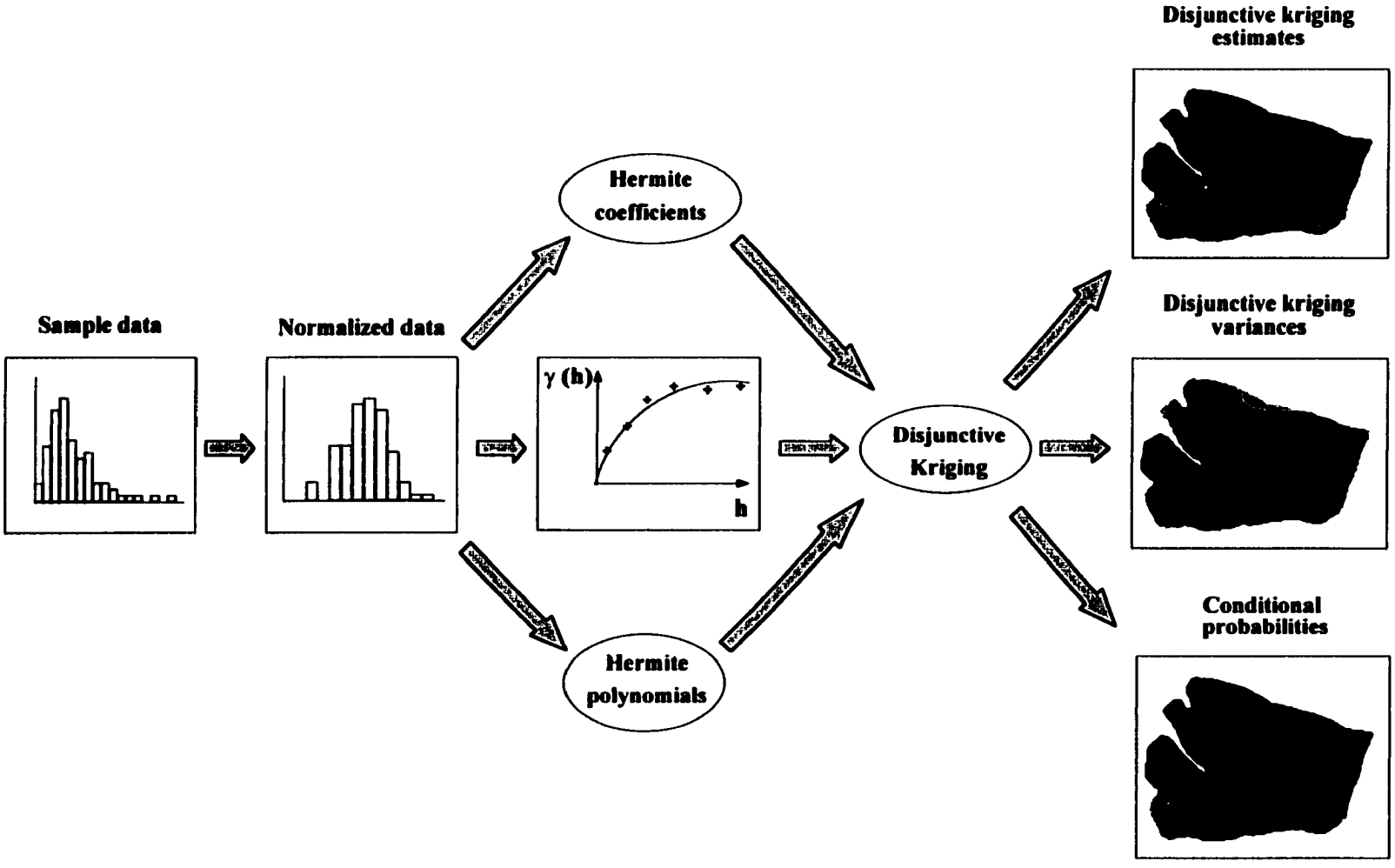

Figure 2.7 Flowchart for disjunctive kriging.

Disjunctive Kriging Features

1) DK is a non-linear unbiased estimator that yields minimized error variances.

2) In DK the variable to be estimated is decomposed into a sum of disjoint (uncorrelated) components of sample values.

3) DK is independent of the sample data distribution. Any arbitrary distribution is transformed to a normal distribution with a mean of zero and unit variance. The transformed data is assumed to have a bivariate normal distribution.

4) The key requirement for the application of DK is second-order stationarity of the sample data (Kim et al., 1977).

5) DK allows the calculation of conditional probabilities, that a variable is greater (or smaller) than some prescribed cut-off or tolerance value.

6) DK is, in general, better than linear estimators in providing minimum error variance estimates of a property through non-linear combinations of the data and exactness of estimation. However, if a random variable is uni- or bivariate normally distributed, the linear ordinary kriging estimator is identical to the DK estimator (Journel & Huijbregts, 1978; Rendu, 1980).

7) DK involves complex calculations and high computational costs. The DK equations must be solved k times (the order of the Hermite Polynomials) for each estimate.

8) The original version of DK was developed for the assumption of a bivariate normal distribution for the transformed variable. Later research (Matheron & Armstrong, 1986a, 1986b) derived transformations for gamma, Poisson, beta, binomial, negative binomial, and hypergeometric distributions.

Disjunctive Kriging Applications

Matheron presented the first paper on DK at the NATO Advanced Study Institute on Advanced Geostatistics in the Mining Industry meeting in Frascati, Italy in October 1975. The proceedings of that meeting contains the first (highly mathematical) description of the DK procedure (Matheron, 1976). The most detailed (and again highly mathematical) description of DK is given by Kim et al. (1977). A more popular account was written by Rendu in 1980 and by Yates et al. in 1986. The most recent detailed description of DK was published by Rivoirard in 1994. He has generalized Matheron's original approach by treating it as one of several involving kriging of orthogonal transforms, the particular choice of which depends on the circumstances.

DK has been used in several different fields: mining (Jackson & Marechal, 1979; Murtly, 1988), earthquake ground motion (Carr, 1983), soil science (Webster & Oliver, 1989; Wood et al., 1990), and fisheries (Petitgas, 1993). DK has never been used in any forestry application. Wood et al. (1990) applied DK to estimate and map the salinity of soil in the Bet Sheau Valley of Israel from measurements of electrical conductivity. Petitgas (1993) modeled areas of high pelagic fish density in acoustic fisheries surveys. Yates and Yates (1988) applied DK as a decision-making tool that helps determine when reclamation action is necessary. They used DK to aid in management decision making of the placement of septic tanks to protect a municipality's drinking water supply.

2.4.4 Cross Validation

In a comparison between different interpolation techniques one would like to check the results of the different approaches and choose the estimation method that works best. Cross-validation uses only the information available in the sample data set. Isaaks & Srivastava (1989, p. 352) give a nice description of cross validation: "*The sample value at a particular location is temporarily discarded from the sample data set; the value at the same location is then estimated using the remaining samples.* [...]. *Once the estimate is calculated we can compare it to the true sample value that was initially removed from the sample data set. This procedure is repeated for all available samples. The resulting true and estimated values can then be compared* [...]." The differences between the estimated values Z^* and the true values Z are called residuals r:

$$r = Z^*(x_0) - Z(x_0) \qquad [2.66]$$

where $Z^*(x_0)$ is the estimated value at location x_0, and $Z(x_0)$ is the true value at that location. Different measures of validation are used to assess the performance of the different interpolation methods.

Measures of Validation

1) Mean of the residuals

The mean of the residuals should be close to zero in order to produce unbiased estimates. If the mean is not far from zero one can say that there is no apparent bias, while a large negative (or positive) average error can represent systematic underestimation (over-estimation, respectively).

2) Variance of the residuals

Another desirable feature in the residuals distribution is a small spread. The variance measures this spread of the errors.

3) q-q plot of the residuals

A third feature is that the residuals should be normally distributed. A q-q plot displays the quantiles of a data set versus the quantiles of a reference theoretical distribution. A q-q plot using a normal distribution as reference distribution is also called a normal plot and is used to check on the normality assumption. If the plot is a straight line the data are normally distributed.

4) Histogram of the residuals

A histogram shows the frequency for each class interval of residuals. A histogram allows for visually checking the unbiasedness, small spread and normality of the residuals.

5) Mean Square Error

In many situations the error distribution might depict a mixture of desirable (e.g. zero mean, small spread) and undesirable (e.g. bias, large spread) features. The mean square error (MSE) is a statistical measure that incorporates both, the bias and the spread of the errors:

$$MSE = variance + bias^2$$

The MSE is calculated as:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} r^2 \qquad [2.67]$$

where n is the number of sample points. The smaller the MSE, the better the estimator.

6) Scatterplot of predicted values versus residuals

Besides analyzing the entire error distribution for unbiasedness, one also wants to check on the properties of any range of errors. This so-called conditional unbiasedness can be checked with a scattergram plotting the estimated values versus the residuals. The errors should evenly spread around the zero line.

7) Scatterplot of predicted values versus true values

Another measure of performance checks on the bivariate distribution of estimated and true values. The perfect estimation would result in a scatterplot with points lying along the 45-degree line. In practice, however, the points appear as a cloud. The closer the cloud of points to the 45-degree line, the better the estimation. A correlation coefficient calculated from the true and estimated values can be applied to summarize how closely the points in the scatterplot come to falling on a straight line.

CHAPTER 3

Data Set

3.1 Location

The Fraser Experimental Forest is a 93 square kilometer outdoor research laboratory maintained by the Rocky Mountain Research Station, USDA Forest Service. The forest is located approximately 160 kilometers northwest of Denver, Colorado (Figure 3.1). The data for this study were collected on 82 1/125 hectare circular plots distributed across the Lexen Creek watershed, a 121 hectare first-order subalpine watershed in the Fraser Experimental Forest.

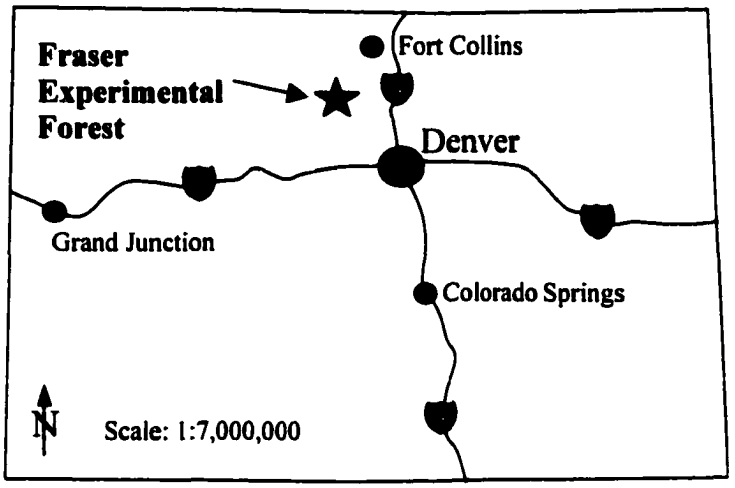

Figure 3.1 Location map of Fraser Experimental Forest.

3.2 Vegetation, Topography, Climate

Native vegetation is typical of the subalpine forest zone of the central Rocky Mountain: Engelmann spruce (Picea Engelmannii, Parry) and subalpine fir (Abies lasiocarpa, Nutt.) are the predominant trees at higher elevations, on north slopes, and along streams. Lodgepole pine (Pinus contorta, Dougl.) is the predominant tree species at lower elevations and on drier upper slopes. Scattered patches of aspen occur in areas opened up by logging or fire. Occasionally, a large, old (450 to 500 years) Douglas fir (Pseudotsuga mentziesii) can be found (Alexander et al., 1985). Measurements from all plots reflect uneven-aged overstory conditions (Mowrer, 1997). By all indicators mortality has only been caused by natural processes (Mowrer, ibid.).

Topography of the Experimental Forest is typical of Southern Rocky Mountain Province. The west side of the forest is characterized by rugged mountains and narrow, steep-sided valleys filled with alluvium and glacial outwash. South and east sides of the

forest are remnants of an old peneplain, dissected by mountain glaciers and characterized by long, gentle, relatively uniform slopes. The north side is a nearly level, broad valley dissected by St. Louis Creek and surrounded by rolling hills (Alexander et al., 1985).

Climate is cool and humid with long, cold winters and short, cool summers. Average yearly temperature at Forest headquarters (2,740 meter elevation) is 0.5° C. Mean monthly temperature for January is -10° C, for July 13° C. Precipitation over the entire Experimental Forest averages about 50 to 75 centimeters, with nearly two-thirds falling as snow from October through May (Alexander et al., 1985).

CHAPTER 4

METHODS

4.1 Field Data

Three sample plot variables were used in this research: number of trees, total basal area, and number of seedlings. The 82 circular sample plots had a radius of 5 m. Plot center locations were determined using a six-channel global positioning system (GPS) receiver, and were subsequently differentially corrected to obtain a nominal 5 m r.m.s. (root mean square error) locational accuracy. Figure 4.1 shows the location of the 82 sample plots within the Lexen Creek watershed.

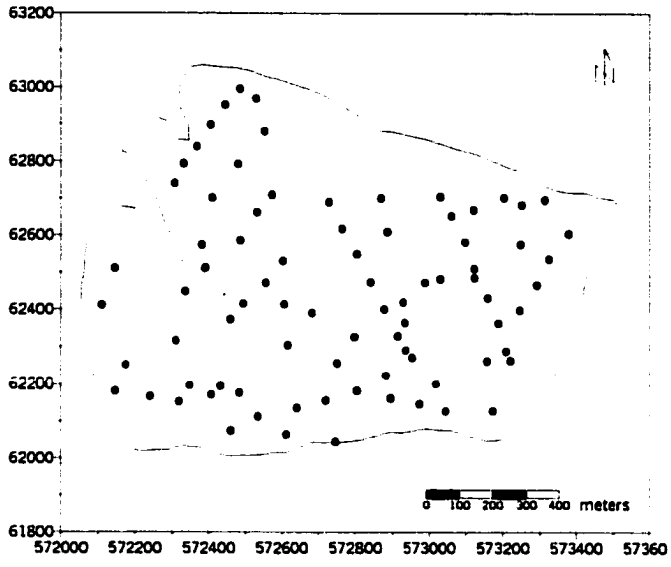

Figure 4.1 Locations of the 82 sample plots in the Lexen Creek watershed.

The following coordinate system was specified: state plane coordinate system, zone North Colorado (no. 3451), North America datum 1927, Clarke 1866 ellipsoid, and meters as the measurement units. All stems greater than 2.54 cm within the plot were counted and measured for dbh (diameter at breast height, 1.37 m above the ground), from which the total basal area was calculated for the plot. Seedling numbers were derived by establishing a line transect across the plot on the slope contour. All seedlings within 45 cm of either side of the line were recorded.

4.2 Remote Sensing Data

The Landsat-TM satellite scene no. 93265028-01 (path 34, row 32) was acquired on 23rd September 1992. The scene was 100 percent cloudfree. Digital image processing was performed using IDRISI 1.0 (Clark Labs, 1995) and ERDAS IMAGINE 8.2 (ERDAS Inc., 1997). The Lexen Creek study area covered a rectangular 267x167 pixel

subscene of the full scene. The image was geometrically corrected using digital orthophotos (0.5 m pixel resolution) as the reference image. The scene was resampled from its original 30x30 m pixel resolution to a 10x10 m resolution. In order to create a moving average of the digital values, a 3x3 moving window was passed over the image. Each new pixel value represents an average value of its eight neighboring pixels. The normalized difference vegetation index (NDVI) was calculated using the following formula: (band4 – band3) / (band4 + band3). For each sample plot location the NDVI value was extracted.

4.3 Digital Elevation Model

Arc/Info (ESRI, 1997) contour line coverages with 2 m elevation intervals were available for the study area. Arc/Info's TIN module was used to create a digital elevation model (DEM) from these contour lines. From the DEM elevation, slope and aspect maps were derived. For each sample plot location elevation, slope and aspect values were extracted. A combined slope/aspect value was computed for each sample point:

$$\text{slope/aspect} = [\sin(\text{aspect}) + \cos(\text{aspect}) - 1] \cdot \tan(\text{slope}) \qquad [4.1]$$

where aspect is measured in radians, and slope is measured in degrees (Bonham et al., 1995; Stage, 1976). Unfavorable, drier aspects (south and west) result in a negative value, whereas the favorable, wetter aspects (north and east) result in positive values. The steeper the slope, the more negative (more positive, respectively) the value.

4.4 Spatial Statistical Techniques

For each of three variables of interest (number of stems, total basal area, and number of seedlings) the following interpolation methods were applied: polygonal mapping, inverse distance weighting, ordinary kriging, universal kriging, cokriging, and disjunctive kriging. Based on the 82 sample plots a total of 20,800 estimates were generated on a 10x10 m grid system. It should be mentioned that the circular sample plots have an area of 78.54 m^2. The value for each of the 100 m^2 grid cells represents an estimate on a per sample plot basis. For values based on the area, each grid cell estimate has to be multiplied by a correction factor of 0.73. To evaluate the interpolation results cross-validation was applied. The residuals from this cross-validation procedure were used to compute mean square error (MSE), scatterplot of estimated values versus true values, scatterplot of estimated values versus residuals, histogram, mean, variance, and q-q plot. Cross-validation was also applied to find the optimal number of nearest neighbors to include in the estimation processes.

The ISATIS (Geovariances, 1999) and GEOPACK (Yates & Yates, 1990) geostatistical software package, as well as the SPLUS (MathSoft, 1998) and SAS (SAS Institute Inc., 1997) statistical software packages were used to conduct the different interpolation techniques and their cross-validations. The SPLUS functions used in this project are custom programs (Reich, 1999b), and are not part of the commercial product. The seven measurements of validation were calculated in SPLUS. The VARIOWIN (Pannatier, 1996) variogram modeling software was used to display the experimental variograms and cross-variograms with their variogram and cross-variogram models. The

SURFER (Golden Software Inc., 1994) graphics program was applied to display the interpolated maps and the standard deviation of estimation.

Polygonal Mapping

Traditional photointerpretation techniques were applied to delineate eight forest stand polygons on digital orthophotos. An average value for each of the three variables of interest within each stand polygon was computed. The SAS statistical software package was used to apply ANOVA for testing the hypothesis of all eight polygonal means being identical. Tukey's HSD multiple comparison test was applied to determine which polygonal mean differed significantly from the others. The residuals were calculated as the deviation between each polygonal mean and each sample value within that polygon.

Inverse Distance Weighting

Two methods of inverse distance weighting were applied: (1) inverse distance weighting calculated interpolated values by weighting sample points inversely proportional to their distance from the unsampled locations. (2) Inverse distance weighting squared weighted the sample points proportional to their squared distance. To conduct cross-validation the SPLUS function CRSSIDW was used. The SPLUS program IDW was then applied to calculate the prediction surface and the standard deviations of prediction.

Ordinary Kriging

Program ISATIS was used to conduct ordinary kriging. Experimental variograms with lag spacings between 20 and 180 m were tested. The lag tolerances were always half of the lag spacings. Due to the small number of sample points available only omnidirectional variograms were computed. At least 30 data pairs were used in the computation of the variogram values at each lag. The maximum lag was limited to half the distance of the study area, about 800 m. Four variogram models were tested for each experimental variogram: spherical model, exponential model, Gaussian model, and power model. A cross-validation procedure in ISATIS was used to find the optimal number of nearest neighbors to include in the kriging process.

Universal Kriging

A first- and second-degree trend surface (using the SPLUS program TRENDLS) was subtracted from the data values in order to remove a possible trend in the data. The first-degree trend surface was of the form:

$$Z(x_0) = a_0 + a_1 x + a_2 y \qquad [4.2]$$

where $Z(x_0)$ was the value of interest at location x_0, x was the x-coordinate, y was the y-coordinate, and a_0, a_1, and a_2 were the unknown coefficients. The second-degree trend surface was of the form:

$$Z(x_0) = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 y^2 + a_5 xy \qquad [4.3]$$

The experimental variograms of the residuals of the trend surface were calculated and a variogram model was fitted to the data using ISATIS. Cross-validation was applied to

find the optimal number of nearest neighbors. Ordinary kriging was conducted on the residuals. In a final step the trend surface and the kriged residual surface were combined.

Cokriging

Primary and secondary variables were available at all sample locations. The cokriging technique used in the case where the available variables are measured at the same locations is sometimes called colocated cokriging (Goovaerts, 1997). Each primary variable was combined with each possible combination of secondary variables (elevation, a combined value of slope and aspect, and NDVI). Variograms for the primary variable and each secondary variable, as well as cross-variograms for each combination of these variables had to be computed. ISATIS' variogram modeling routine checked automatically whether the variogram and cross-variogram models were in compliance with the rules of the linear model of coregionalization.

Disjunctive Kriging

GEOPACK was used to conduct disjunctive kriging. The number of Hermite polynomials, as well as the number of least-square polynomials for the transformation process had to be determined. Hermite polynomials and least-square polynomial with values ranging from 1 to 10 were tested. The optimal number of Hermite polynomials was found when the mean and variance of the original data, and the mean and variance of the transformed data were approximately the same. Experimental variograms were calculated and variogram models were fitted to the transformed data values. No cross-validation program was available. Therefore, 82 data sets were created, dropping one data

value for each data set. The remaining 81 data values were used to estimate the removed data value. In addition to the disjunctive kriging estimates and the standard deviations of estimation, an example for a conditional probability was computed.

CHAPTER 5

Results and Discussion

Three variables of interest were examined: number of stems (STM), total basal area (TBA), and number of seedlings (SEEDL). Several summary statistics are presented in each section for each of the variables:

1) Measures of location: mean, median, minimum, and maximum.

2) Measures of spread: standard deviation and variance.

3) Measures of shape: skewness and kurtosis.

For each variable the spatial interpolation methods compared were: polygonal mapping, inverse distance weighting, inverse distance weighting squared, ordinary kriging, universal kriging with a first- and second-degree trend surface, cokriging, and disjunctive kriging. The residuals of a cross-validation procedure were used to evaluate the interpolation results. Seven measures of validation were applied: mean square error (MSE), scatterplot of estimated data values versus true data values, scatterplot of estimated data values versus residuals, as well as histogram, mean, variance, and q-q plots of the residuals.

5.1 Number of Stems

5.1.1 Summary Statistics

Several summary statistics for *number of stems per plot* (STM) were calculated. Table 5.1 summarizes the minimum and maximum value, mean, median, standard deviation, variance, skewness, and kurtosis for STM. The coefficient of skewness is a measure of symmetry. If the value is close to zero, the distribution is approximately symmetric. If the value is positive, the distribution is positively skewed; if it is negative, then the distribution is negatively skewed. In the case of variable STM the distribution is skewed to the right as shown in Figure 5.1. Kurtosis is a measure of peakedness. It indicates the extent to which a distribution departs from the normal curve by being either pointier or flatter. A normal distribution has a value of 3. Numbers smaller than 3 mean flatter than normal, values greater than 3 mean more peaked than normal.

Table 5.1 Summary statistics for number of stems per plot (STM).

Statistics	Values
Minimum	1.00
Median	9.00
Mean	11.29
Maximum	42.00
Standard Deviation	7.69
Variance	58.45
Skewness	1.53
Kurtosis	3.33

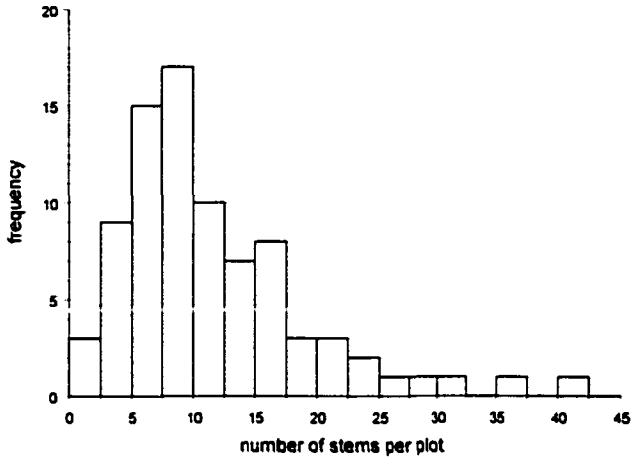

Figure 5.1 Histogram for number of stems per plot (STM).

5.1.2 Polygonal Mapping

Eight forest stands were delineated on digital orthophotos using traditional photointerpretation techniques. Figure 5.2 displays the delineated forest stands, the stand numbers, and the sample point locations. For each stand polygon the average number of stems per plot was calculated. Table 5.2 shows the number of sample plots, the average number of stems, and the standard deviation for each polygon.

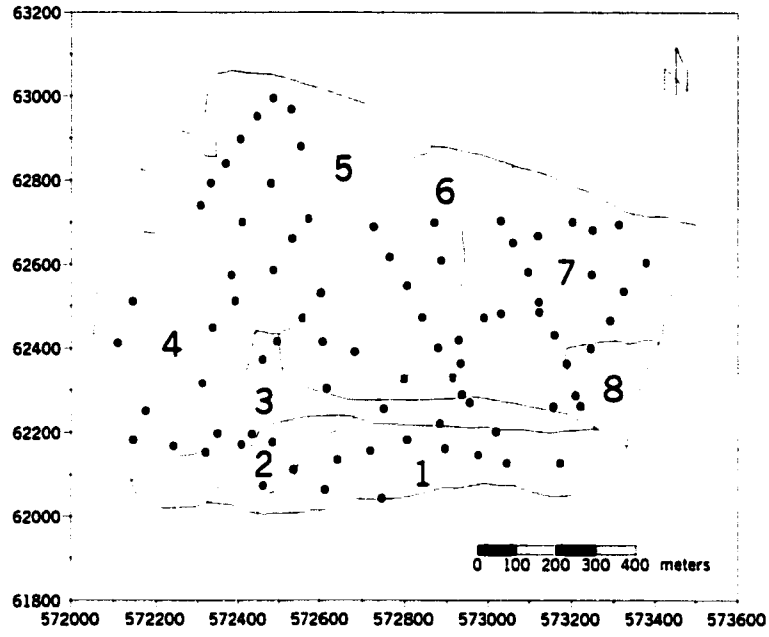

Figure 5.2 Delineated forest stands, stand numbers, and sample plot locations.

Table 5.2 The number of sample plots, the mean, and the standard deviation of each forest stand polygon for number of stems per plot (STM).

Stand number	Number of sample plots	Mean	Standard deviation
1	10	13.30	7.90
2	3	11.33	6.03
3	9	13.67	11.30
4	7	19.71	9.55
5	15	8.13	5.22
6	11	5.54	2.70
7	23	11.87	6.72
8	4	10.50	3.11

An ANOVA table (Table 5.3) was calculated to test the null hypothesis of all eight polygonal means being identical. The alternative hypothesis was that at least one of the means is different from the others. The p-value of 0.0053 was lower than the

type I error set to 0.05. Therefore the null hypothesis of all eight polygonal means being identical was rejected; at least one polygonal mean was different from the others.

Table 5.3 ANOVA table for number of stems per plot (STM).

Source	Sum of Squares	Degrees of Freedom	Mean Square	F Test	p-value
Between groups	1110.711	7	158.673	3.19	0.0053
Within groups	3682.264	74	49.760		
Totals	4792.975	81			

Tukey's HSD test, a multiple comparison method, was used to find out which polygonal means were significantly different from one another. Table 5.4 shows the outcome from Tukey's HSD test. The polygonal means were ranked in descending order. Polygonal means with the same letter in front were not significantly different. Therefore only stand 4 and stand 6 were significantly different from each other, while there was no statistically significant difference between stands 1, 2, 3, 5, 7, and 8. It should be pointed out that both, the ANOVA calculation and Tukey's HSD test require a normal distribution of the data. Variable STM, however, was skewed to the right. Therefore, the results of both tests should be taken with caution.

Table 5.4 Tukey's HSD test for number of stems per plot (STM).

Tukey Grouping		Mean	Stand
	A	19.71	4
	A		
B	A	13.67	3
B	A		
B	A	13.30	1
B	A		
B	A	11.87	7
B	A		
B	A	11.33	2
B	A		
B	A	10.50	8
B	A		
B	A	8.13	5
B			
B		5.54	6

The residuals were calculated as the difference between each polygonal mean and each sample point value within that polygon. The residuals were then used to compute the seven measures of validation. The mean square error (MSE) was 44.906. The cloud of points in the scatterplot of the estimated data values versus true data values (Figure 5.3a) had a correlation coefficient of 0.48. The scatterplot of estimated data values versus residuals (Figure 5.3b) depicted that the errors were fairly evenly distributed (besides one possible outlier with an unusual high value of 28.33). A check on the outlier revealed that it originated from a sample plot with a large number of 42 tree stems. Because this number of stems was judged as a correct value, it was decided to leave this data value in this, and all subsequent analyses. The histogram (Figure 5.3c) and the q-q plot (Figure 5.3d) showed that the residuals were approximately normally distributed (again with the exception of one possible outlier). The frequency distribution of the residuals

had a mean of 0, and a variance of 45.46. The minimum value was -10.71, and the maximum value was 28.33.

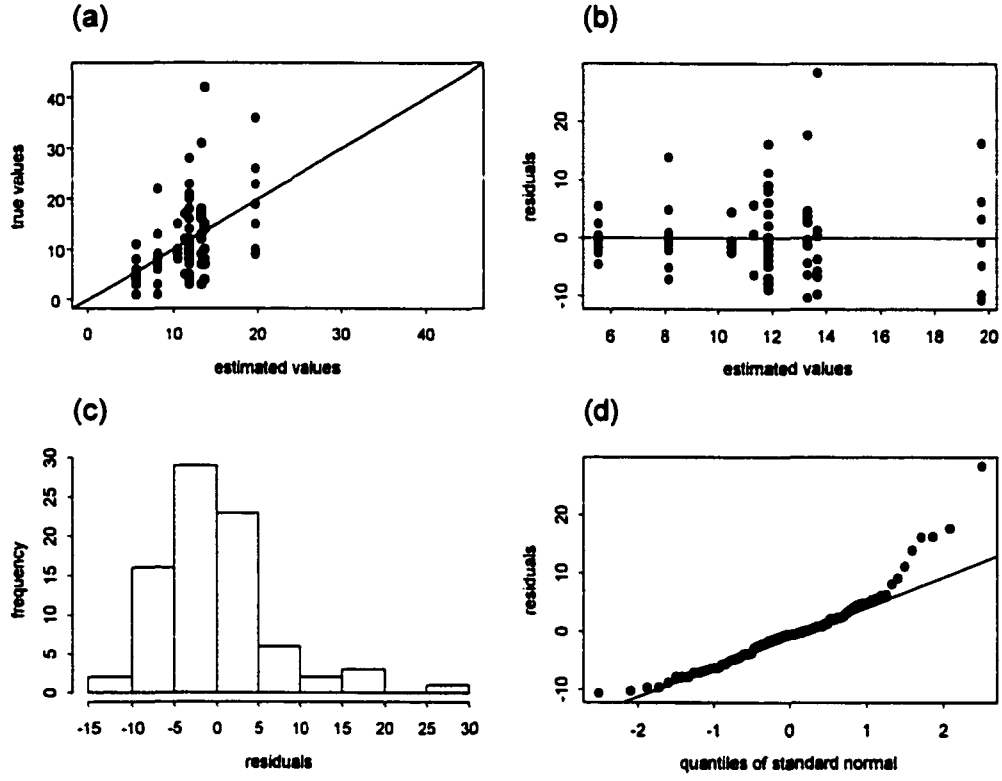

Figure 5.3 Polygonal mapping with number of stems per plot (STM):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A polygonal map and a surface map showing the mean number of stems (STM) per polygon, as well as a polygonal map and a surface map showing the standard deviation per polygon are displayed in Figure 5.4 and Figure 5.5. Polygon number 6 had the smallest mean value with 5.54, and polygon 4 the highest value with 19.71. Polygon number 6 also had the smallest standard deviation of 2.70, whereas polygon 3 had the

highest with 11.30. It should be mentioned that the standard deviations should not be compared with each other as they are based on different means.

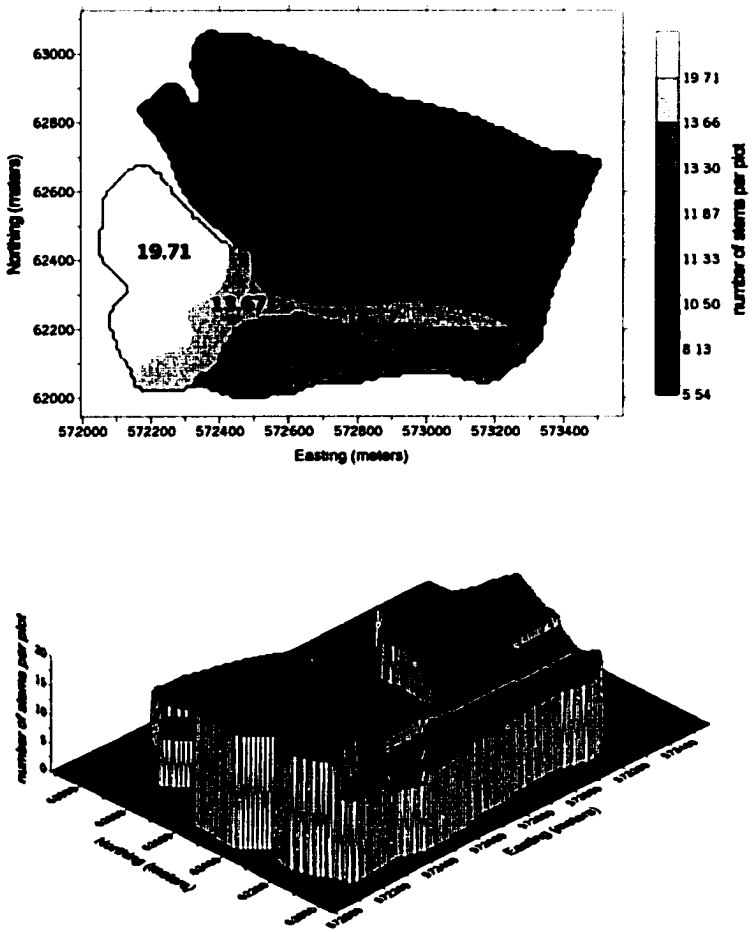

Figure 5.4 Polygonal mapping with number of stems per plot (STM): polygonal map (top graph) and surface map (bottom graph) showing the mean per polygon.

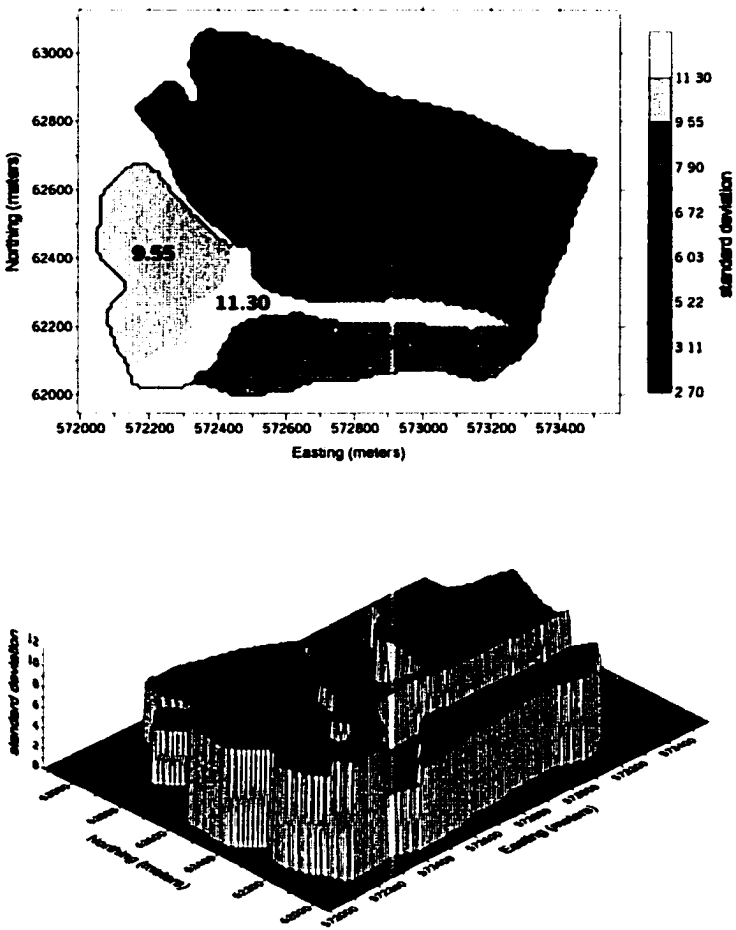

Figure 5.5 Polygonal mapping with number of stems per plot (STM): polygonal map (top graph) and surface map (bottom graph) showing the standard deviation per polygon.

5.1.3 Inverse Distance Weighting

Inverse distance weighting and inverse distance weighting squared were used to obtain estimates for number of stems per plot (STM) at unsampled locations.

Inverse Distance Weighting

Using the residuals from the cross-validation procedure a minimum MSE of 49.140 could be achieved by including the 7 nearest neighbors. The other four measures of validation are displayed in Figure 5.6. The estimated data values versus true values had a correlation coefficient of 0.41. The distribution of the residuals had a mean of -0.25, a variance of 49.68, with a minimum value of -11.76, and a maximum value of 30.52. Without the single possible outlier of value 30.52, the residuals were approximately normally distributed, had a constant variance, and no apparent trend in the data.

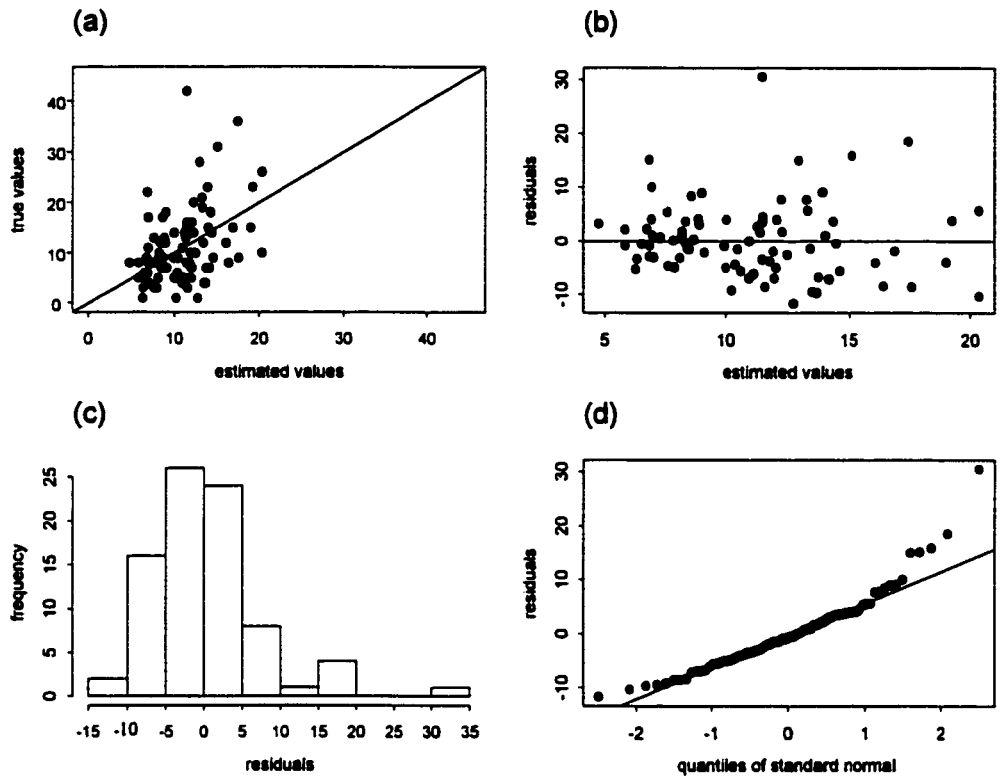

Figure 5.6 Inverse distance weighting with number of stems per plot (STM):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates are displayed in Figure 5.7. The minimum value of the estimates was 1.66, and the maximum value was 37.85. A contour map and surface map of the standard deviations are shown in Figure 5.8. The minimum value of the standard deviations was 0.13, the maximum value was 9.24.

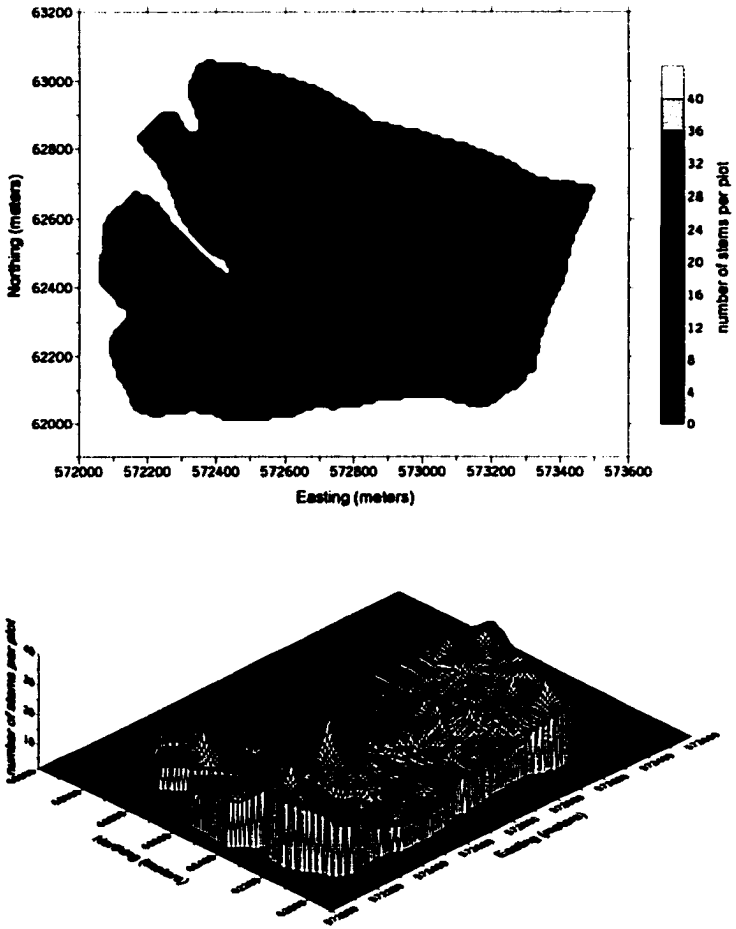

Figure 5.7 Inverse distance weighting with number of stems per plot (STM): contour map (top graph) and surface map (bottom graph) of estimates.

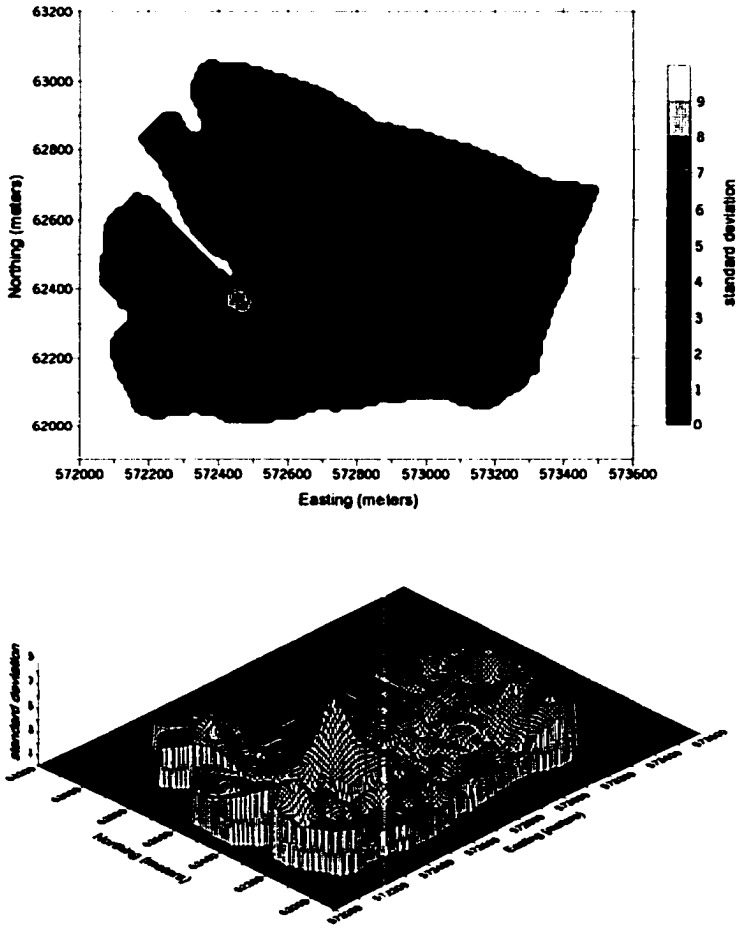

Figure 5.8 Inverse distance weighting with number of stems per plot (STM): contour map (top graph) and surface map (bottom graph) of standard deviations.

Inverse Distance Weighting Squared

A minimum MSE of 48.915 was obtained using 11 nearest neighbors. The other four measures of validation are displayed in Figure 5.9. The correlation coefficient between estimated and true data values was 0.41. Without the single possible outlier of value 29.62, the residuals were approximately normally distributed, with a zero mean, a constant variance, and with no apparent trend in the data. The distribution of the residuals had a mean of -0.23, a variance of 49.46, with a minimum value of -13.86, and a maximum value of 29.62.

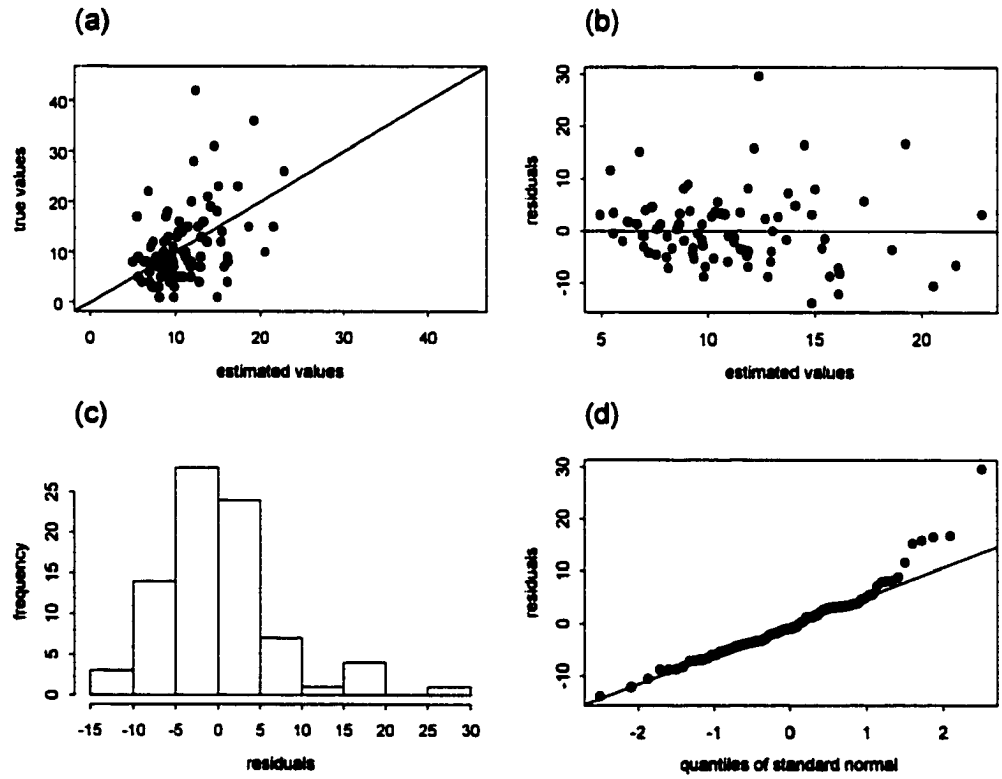

Figure 5.9 Inverse distance weighting squared with number of stems per plot (STM):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates are displayed in Figure 5.10. The minimum value of the estimates was 1.03, and the maximum value was 41.81. A contour map and surface map of the standard deviations are shown in Figure 5.11. The minimum value of the standard deviations was 0.001, and the maximum value was 8.98.

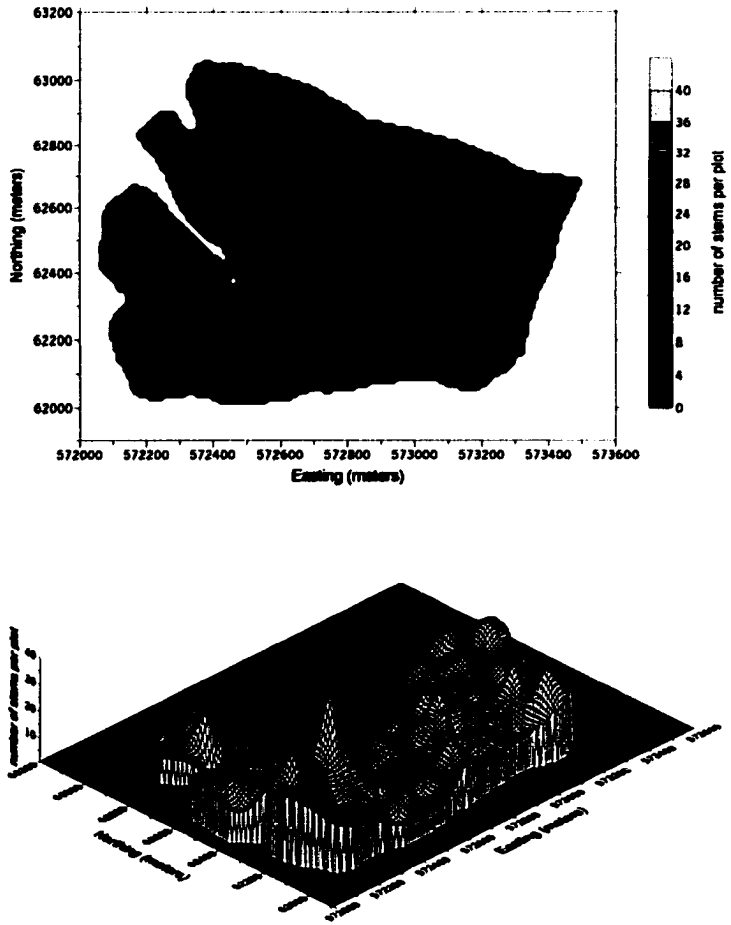

Figure 5.10 Inverse distance weighting squared with number of stems per plot (STM): contour map (top graph) and surface map (bottom graph) of estimates.

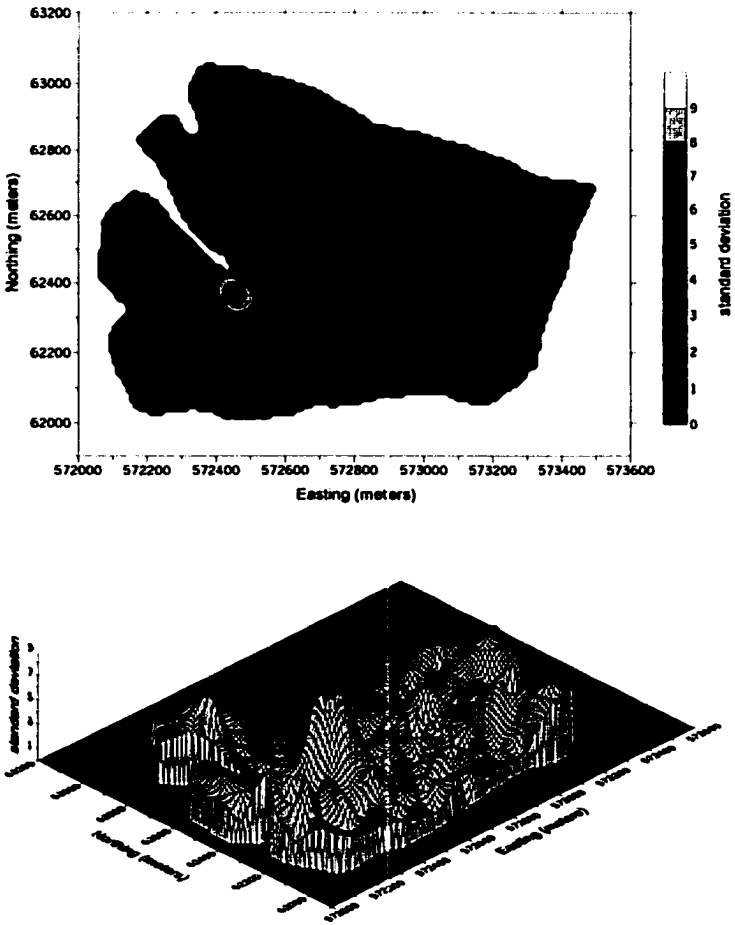

Figure 5.11 Inverse distance weighting squared with number of stems per plot (STM): contour map (top graph) and surface map (bottom graph) of standard deviations.

5.1.4 Ordinary Kriging

An experimental variogram was calculated to examine the spatial correlation of number of stems per plot (STM). A lag distance of 160 m and a lag tolerance of 80 m resulted in the smoothest looking experimental variogram (Figure 5.12). An exponential model was then fitted to the experimental variogram (Figure 5.12). The model was specified with a nugget of 14, a sill of 58, and a range of 270.

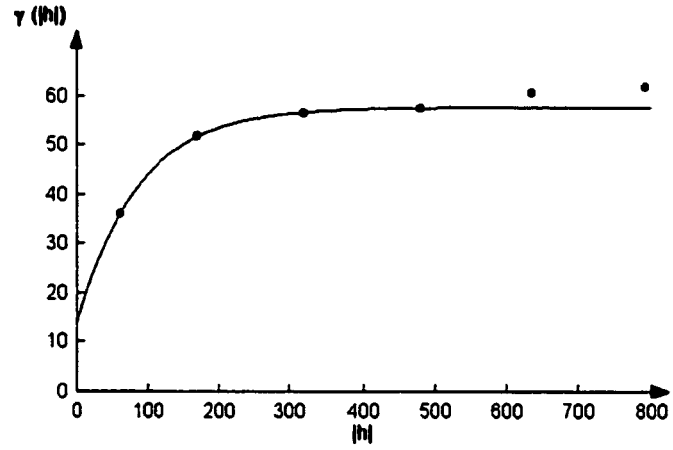

Figure 5.12 Experimental variogram and exponential variogram model for number of stems per plot (STM).

Cross-validation was applied to calculate the residuals. Using 7 nearest neighbors resulted in the smallest MSE of 47.443. Without the possible outlier of value 29.42, Figure 5.13 showed that the residuals had an approximately normal distribution, a constant variance, and no apparent trend in the data. The correlation coefficient between estimated and true data values was 0.44. The distribution of the residuals had a mean of -0.19, a variance of 47.99, with a minimum value of -12.93, and a maximum value of 29.42.

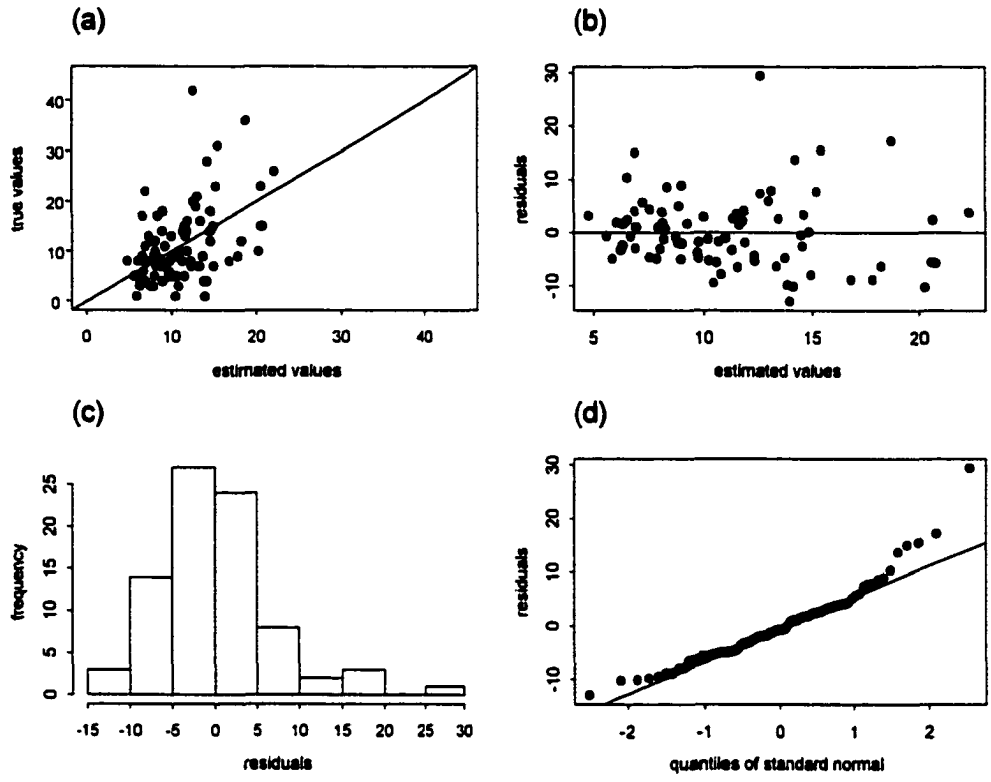

Figure 5.13 Ordinary kriging with number of stems per plot (STM):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates are displayed in Figure 5.14. The minimum value of the estimates was 2.17, and the maximum value was 34.79. A contour map and surface map of the standard deviations are shown in Figure 5.15. The minimum value of the standard deviations was 4.22, the maximum value was 8.28. Note that the scale bar for the maps of the standard deviations of the kriging methods is different form the ones used for the inverse distance weighting techniques.

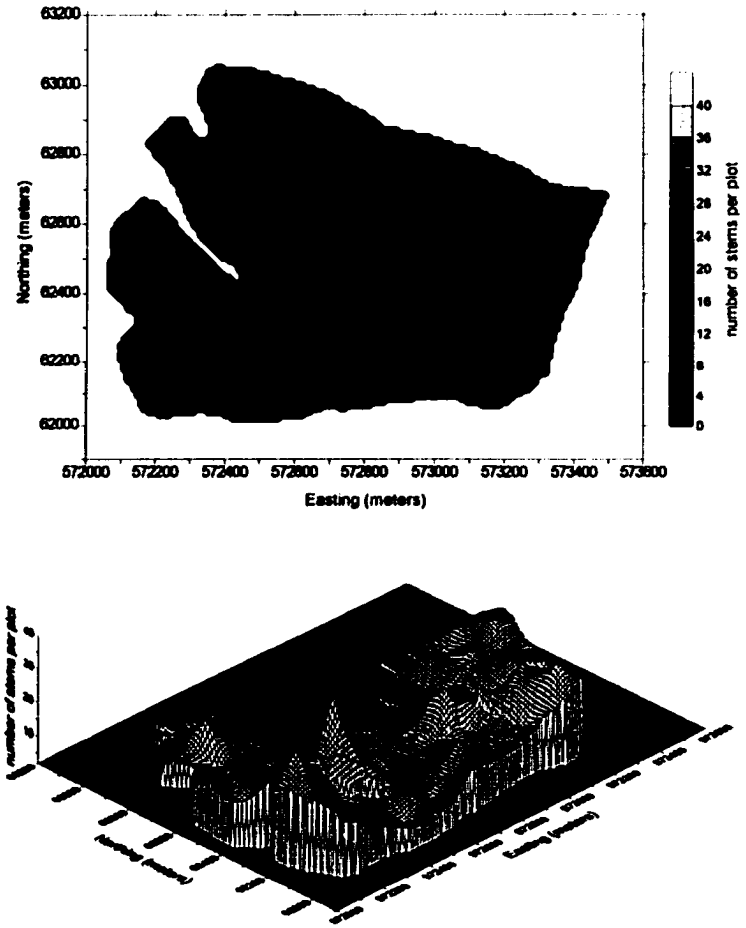

Figure 5.14 Ordinary kriging with number of stems per plot (STM): contour map (top graph) and surface map (bottom graph) of estimates.

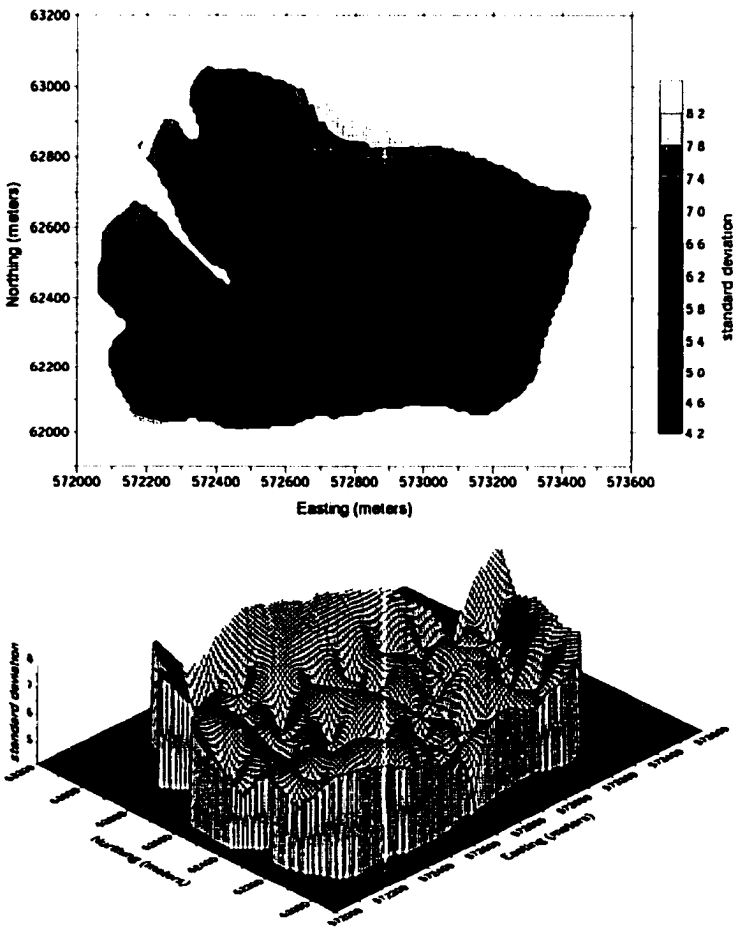

Figure 5.15 Ordinary kriging with number of stems per plot (STM): contour map (top graph) and surface map (bottom graph) of standard deviations.

5.1.5 Universal Kriging

Universal kriging consisted of three stages:

(1) The trend had to be estimated. A first- and second-degree trend surface was calculated as a function of x- and y-coordinates of the sample plot locations. These trend surfaces were subtracted from the original data values.

(2) The residual-based variogram was derived, and the residuals were estimated using ordinary kriging.

(3) The estimated residuals were combined with the trend surface to obtain the final estimates of the actual surface.

First-Degree Trend Surface

A first-degree trend surface using x- and y-coordinates as independent variables was fitted to the STM data set. An R^2 value of 0.095 was achieved, i.e. the variables x-coordinate and y-coordinate accounted for 9.5% of the variability in STM. The trend surface was subtracted from the original data values, and the residuals were then used to calculate the experimental variogram (Figure 5.16). A lag distance of 160 m and a lag tolerance of 80 m were chosen. The spherical variogram model had the following specifications: nugget of 24.5, sill of 53.5, and of range: 225.

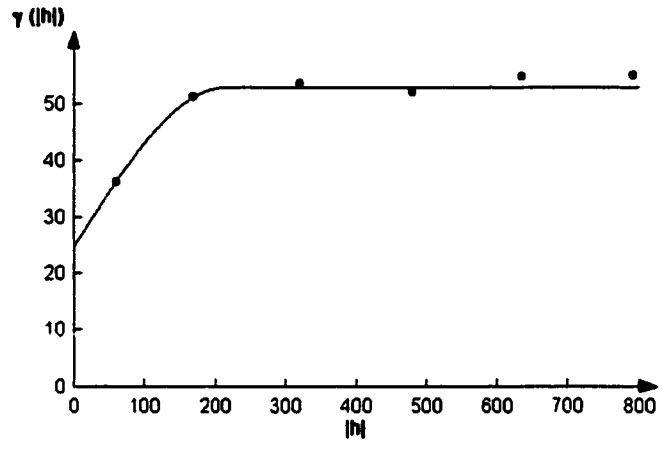

Figure 5.16 Experimental variogram and spherical variogram model for the residuals of the first-degree trend surface for number of stems per plot (STM).

The smallest MSE was obtained using the 7 nearest neighbors: 46.804. The correlation coefficient between estimated and true data values was 0.46. Besides the possible outlier with a value of 29.07, the analysis of the residuals did not exhibit any peculiarities that would suggest violations of the underlying assumptions (Figure 5.17). The distribution of the residuals had a mean of -0.20, a variance of 47.34, with a minimum value of -12.95, and a maximum value of 29.07.

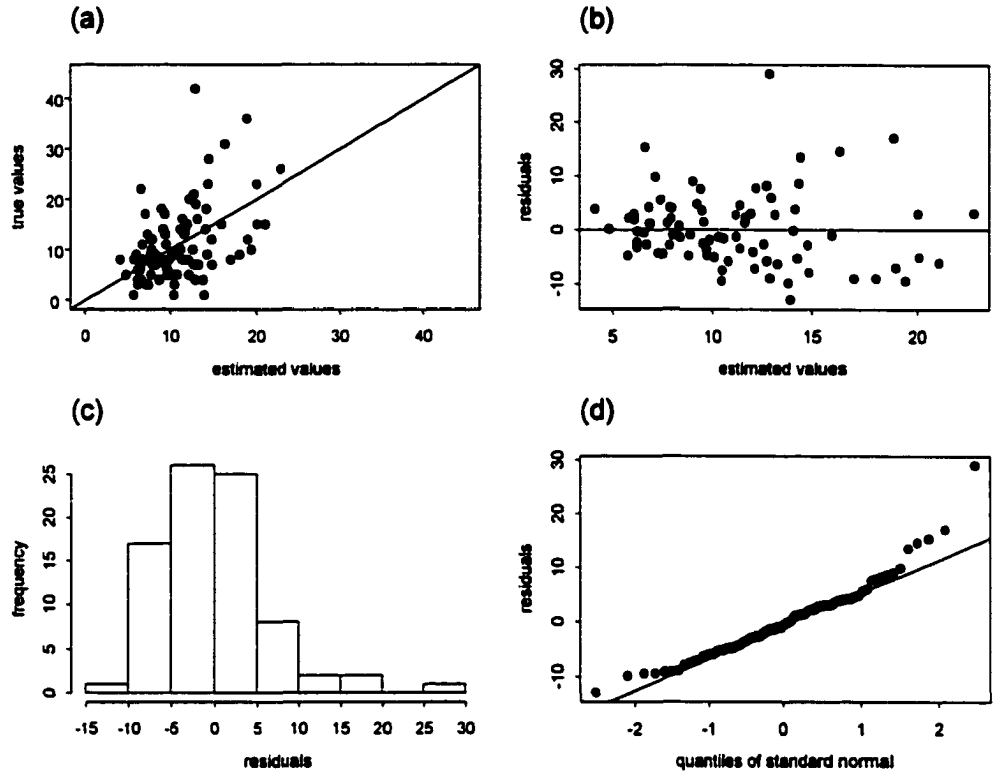

Figure 5.17 Universal kriging with first-degree trend surface for number of stems per plot (STM):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

Ordinary kriging was conducted on the residuals of the first-degree trend surface using the spherical variogram model. The final surface was created by combining the first-degree trend surface with the kriged surface of the residuals (Figure 5.18). The minimum value of the estimates was 2.75, and the maximum value was 28.48. A contour map and surface map of the standard deviations are shown in Figure 5.19. The minimum value of the standard deviations was 5.82, and the maximum value was 7.80.

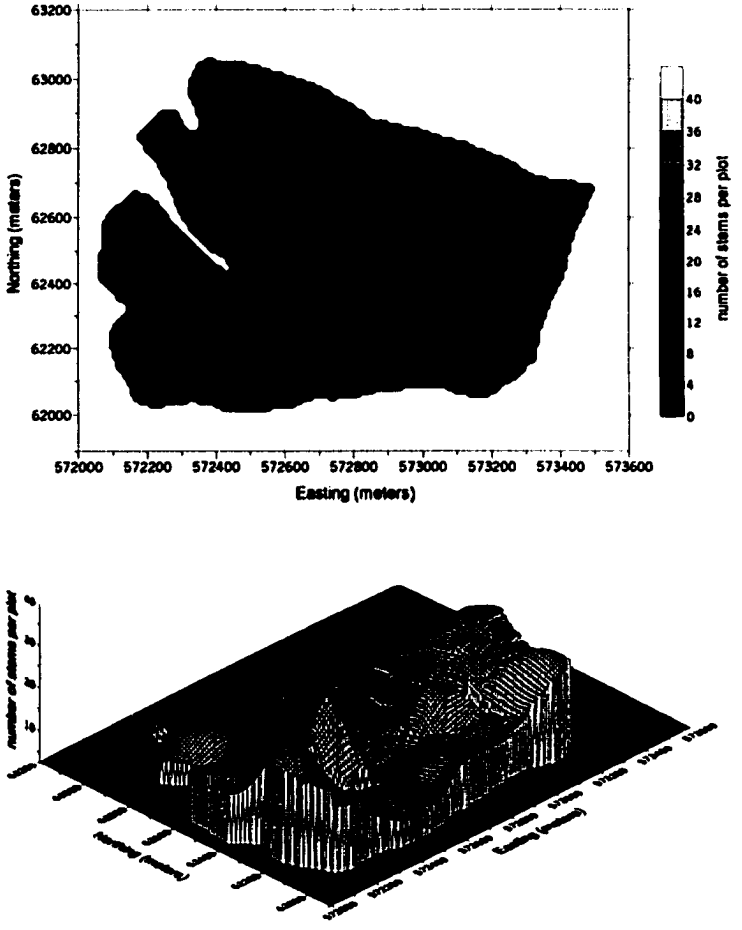

Figure 5.18 Universal kriging with first-degree trend surface for number of stems per plot (STM): contour map (top graph) and surface map (bottom graph) of estimates.

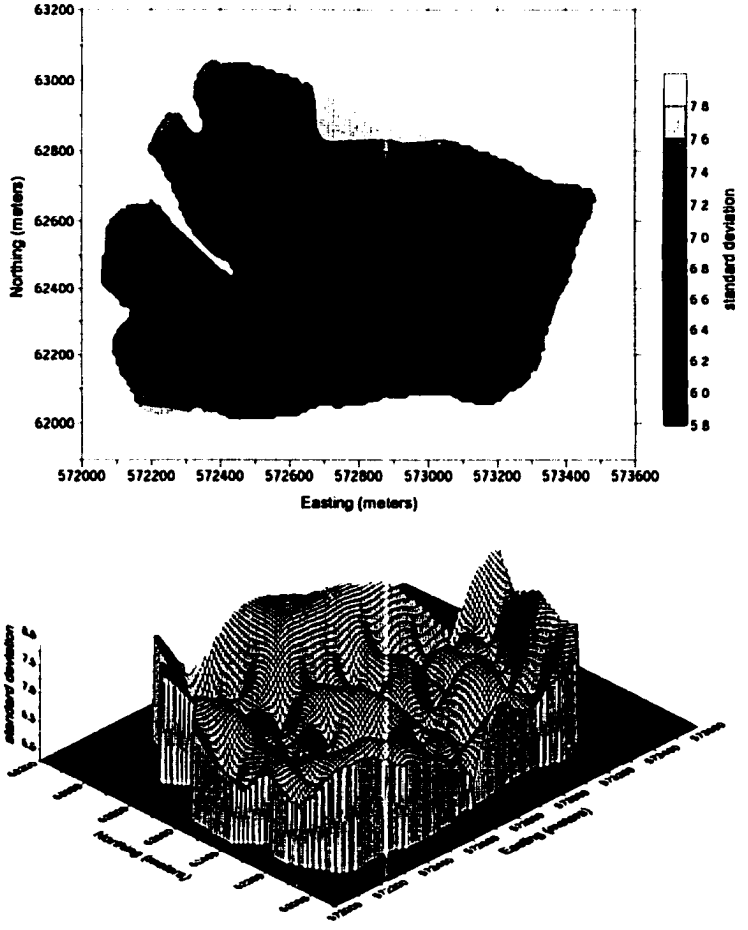

Figure 5.19 Universal kriging with first-degree trend surface for number of stems per plot (STM): contour map (top graph) and surface map (bottom graph) of standard deviations.

Second-Degree Trend Surface

A second-degree trend surface using x- and y-coordinates as independent variables was fitted to the data set. An R^2 value of 0.259 was achieved. The experimental variogram for the residuals of the second-degree trend surface had a lag distance of 160 m and a lag tolerance of 80 m (Figure 5.20). The spherical variogram model had a nugget of 30, a sill of 44, and a range 160. The large nugget is an indication of the data values approaching a random surface. The experimental variogram for the second-degree trend surface has a more erratic appearance than the first-degree trend surface. This can be explained by the fact that the second-degree trend surface accounted for almost 26% of the variability in the data, in comparison with only 9.5% of the first-degree trend surface. Therefore, there was less variability and less spatial correlation left in the residuals of the second-degree trend surface, and hence the experimental variogram looked less smooth.

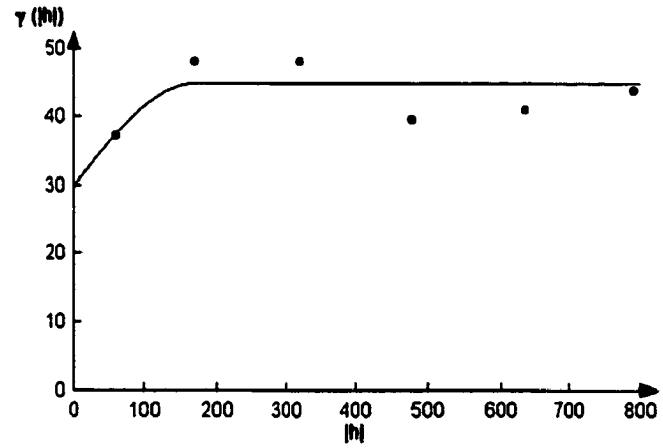

Figure 5.20 Experimental variogram and spherical variogram model for the residuals of the second-degree trend surface for number of stems per plot (STM).

Including the 20 nearest neighbors in the kriging process resulted in the smallest MSE of 45.326. The correlation coefficient between estimated and true data values was 0.48. Besides the possible outlier with a value of 31.27, the analysis of the residuals did not exhibit any peculiarities that would suggest violations of the underlying assumptions (Figure 5.21). The distribution of the residuals had a mean of -0.08, a variance of 45.88, with a minimum value of -13.60, and a maximum value of 31.27.

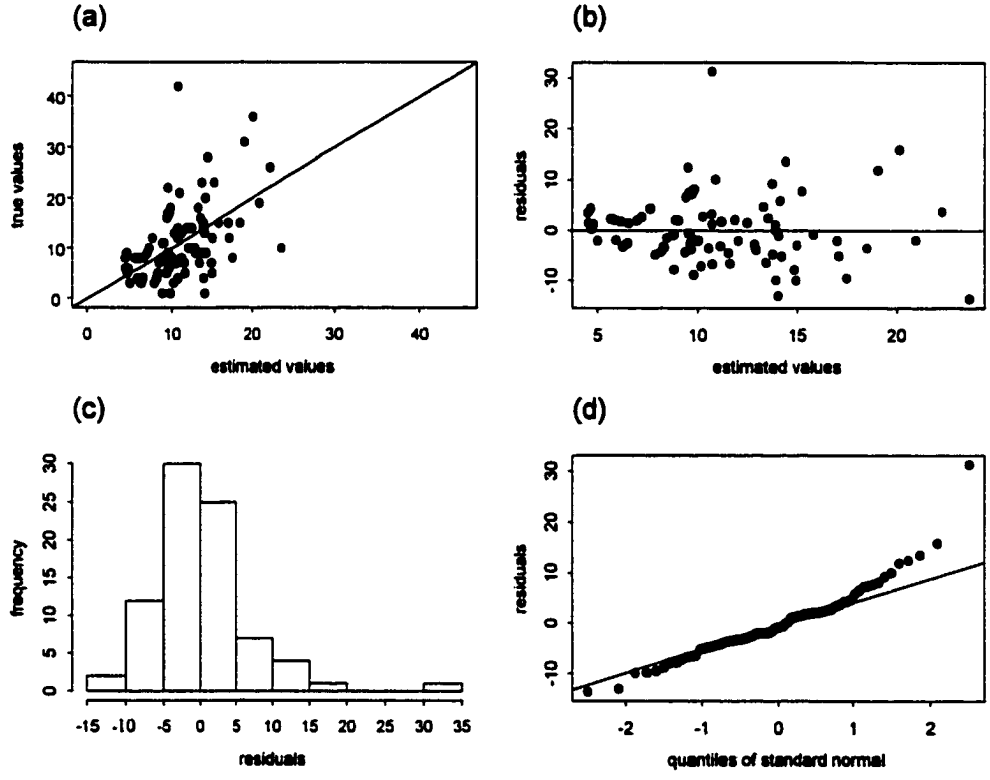

Figure 5.21 Universal kriging with second-degree trend surface for number of stems per plot (STM):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

Ordinary kriging was conducted on the residuals of the second-degree trend surface using the spherical variogram model. The final surface was created by combining

the second-degree trend surface with the kriged surface of the residuals (Figure 5.22). The minimum value of the estimates was 0.23, and the maximum value was 30.38. A contour map and surface map of the standard deviations are shown in Figure 5.23. The minimum value of the standard deviations was 4.81, and the maximum value was 6.90.

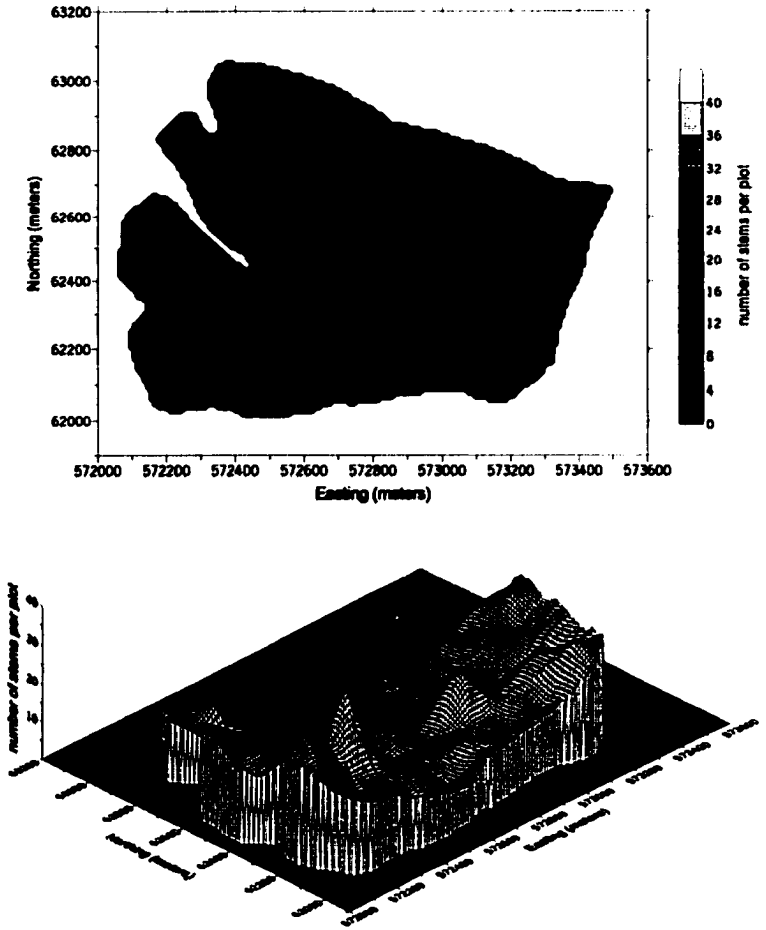

Figure 5.22 Universal kriging with second-degree trend surface for number of stems per plot (STM): contour map (top graph) and surface map (bottom graph) of estimates.

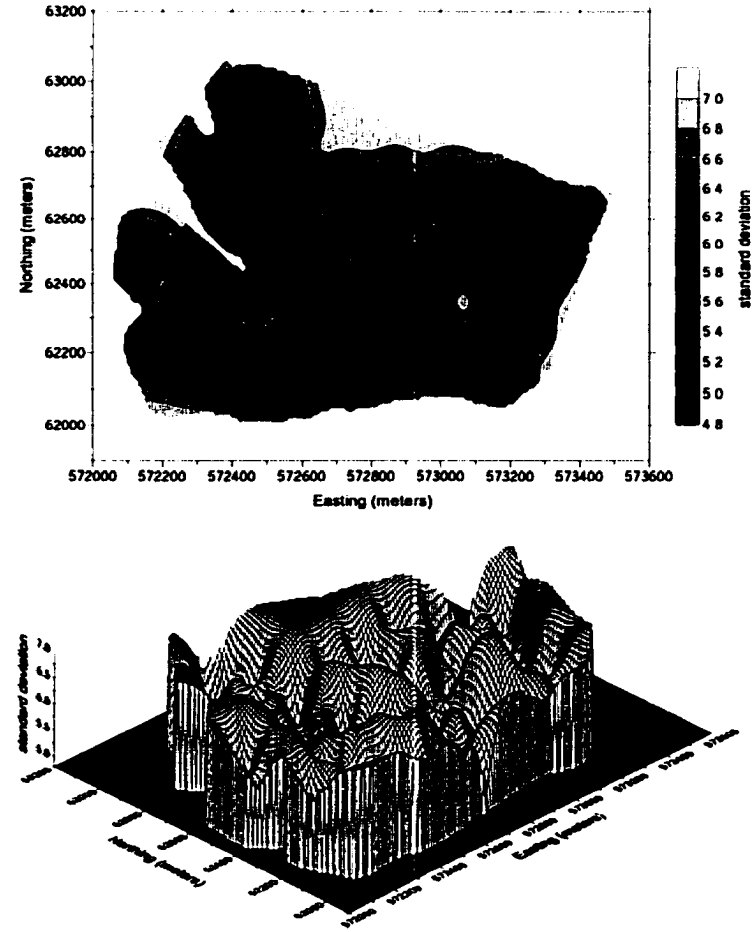

Figure 5.23 Universal kriging with second-degree trend surface for number of stems per plot (STM): contour map (top graph) and surface map (bottom graph) of standard deviations.

5.1.6 Cokriging

The primary variable number of stems per plot (STM) was combined with different sets of the following secondary variables: elevation (ELEV), a combined value of slope and aspect (SLOASP), and the Normalized Difference Vegetation Index (NDVI). Variogram models were fitted to the variograms and cross-variograms of the primary and secondary variables. From all possible combinations of secondary variables, using the combined value of slope and aspect (SLOASP) as auxiliary variable yielded the smallest MSE of 44.568 (Table 5.5).

Table 5.5 Mean square errors (MSE) for primary variable number of stems per plot (STM) using different combinations of secondary variables: elevation (ELEV), slope/aspect (SLOASP), and normalized difference vegetation index (NDVI).

Secondary Variables	MSE
ELEV	46.997
SLOASP	44.568
NDVI	47.891
ELEV and SLOASP	46.811
ELEV and NDVI	48.505
SLOASP and NDVI	44.895
ELEV, SLOASP, and NDVI	44.899

Nested variogram models were fitted to the experimental variograms of STM and SLOASP, and the cross-variogram STM-SLOASP (Figure 5.24). The nested variogram models consisted of a combination of a spherical and a Gaussian variogram model. Table 5.6 shows the specifications for both variogram models and the cross-variogram model.

Table 5.6 Variogram and cross-variogram model specifications for primary variable number of stems per plot (STM) and secondary variable slope/aspect (SLOASP).

Variogram / Cross-variogram	Spherical			Gaussian	
	nugget	sill	range	sill	range
STM	25.456	23.682	200	13.189	650
SLOASP	0.007	0.034	200	0.117	650
STM-SLOASP	0.421	-0.897	200	1.242	650

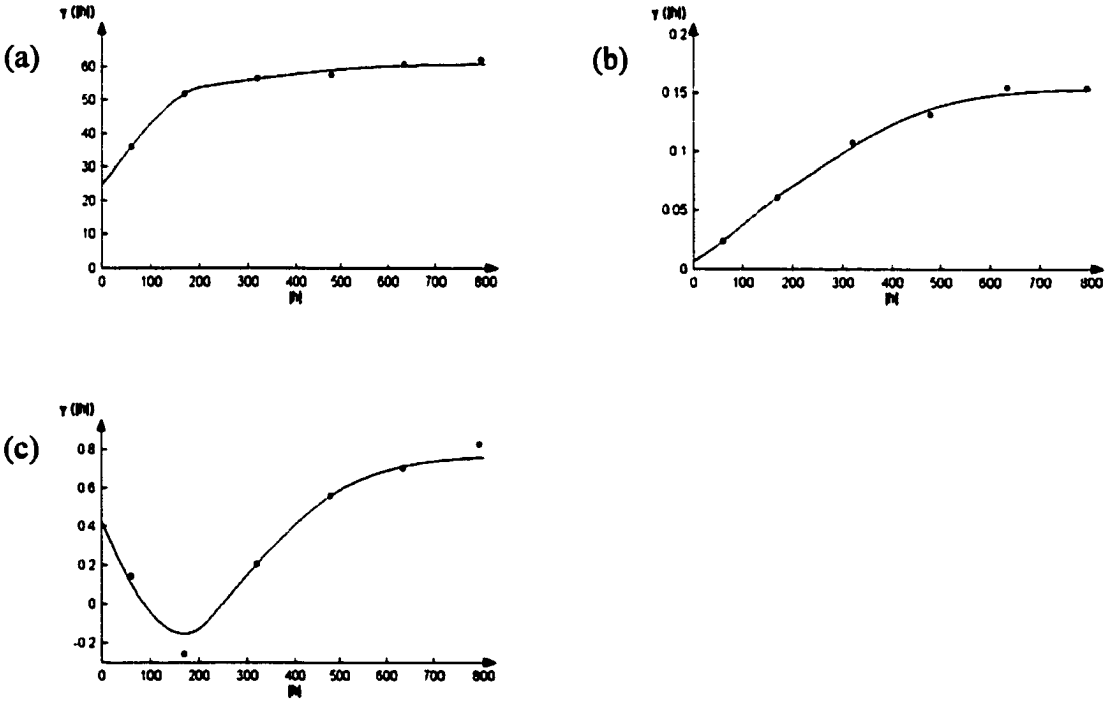

Figure 5.24 (a) Experimental variograms for primary variable number of stems per plot (STM) and (b) secondary variable slope/aspect (SLOASP), as well as (c) experimental cross-variogram with a spherical and a Gaussian (cross-) variogram model.

The cross-variogram values decreased for distances between 0 and 170 m, indicating a negative spatial cross-correlation. For distances greater than 170 m the cross-variogram values increased, indicating a positive spatial cross-correlation.

The smallest MSE could be achieved by using the 20 nearest neighbors: 44.568. The correlation coefficient between estimated and true data values was 0.49. Without the single possible outlier of value 24.97 the residuals were approximately normally distributed, had a constant variance, and had no apparent trend in the data (Figure 5.25). The distribution of the residuals had a mean of -0.16, a variance of 45.09, with a minimum value of -12.22, and a maximum value of 24.97.

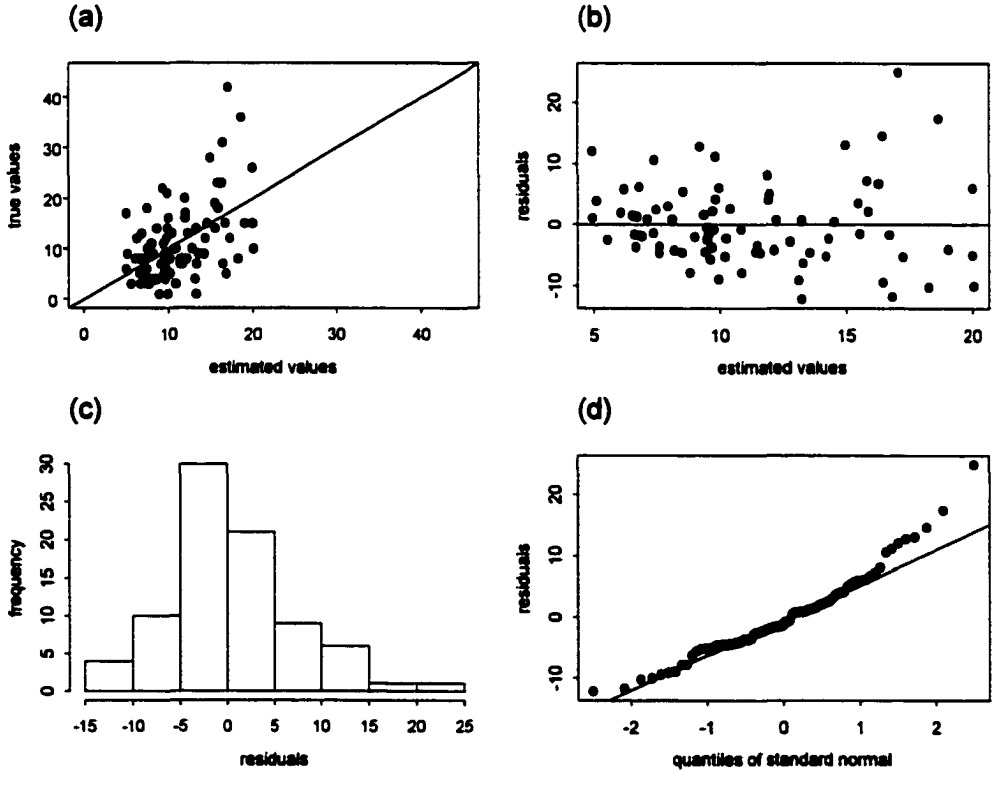

Figure 5.25 Cokriging with primary variable number of stems per plot (STM) and secondary variable slope/aspect (SLOASP):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,

(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates are displayed in Figure 5.26. The minimum value of the estimates was 1.30, and the maximum value was 27.20. A contour map and surface map of the standard deviations are shown in Figure 5.27. The minimum value of the standard deviations was 5.39, and the maximum value was 7.61.

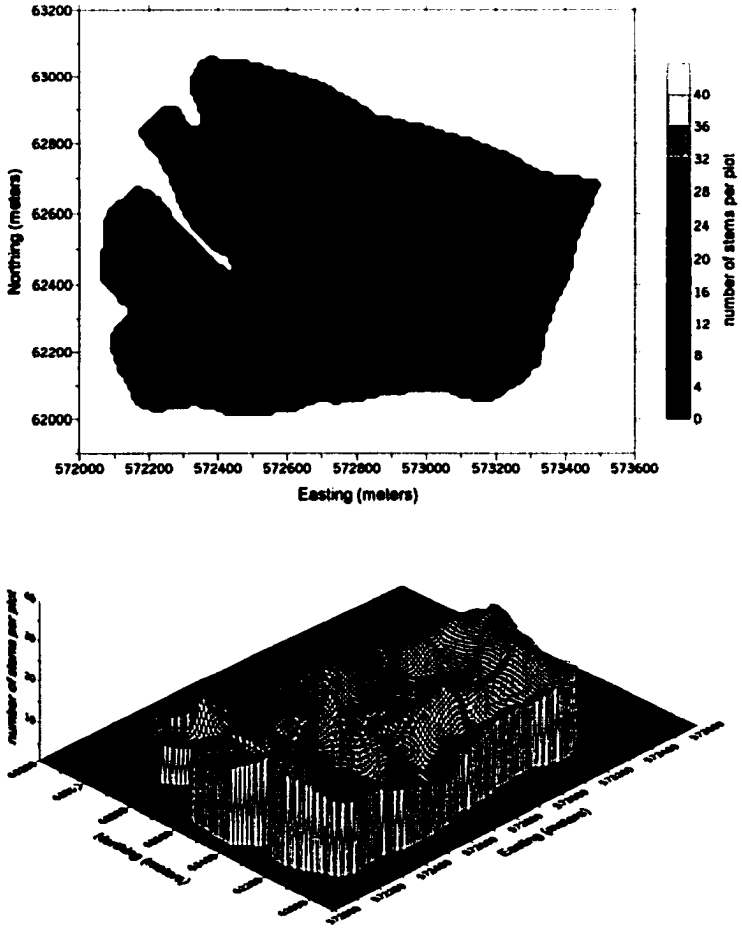

Figure 5.26 Cokriging with primary variable number of stems per plot (STM) and secondary variable slope/aspect (SLOASP): contour map (top graph) and surface map (bottom graph) of estimates.

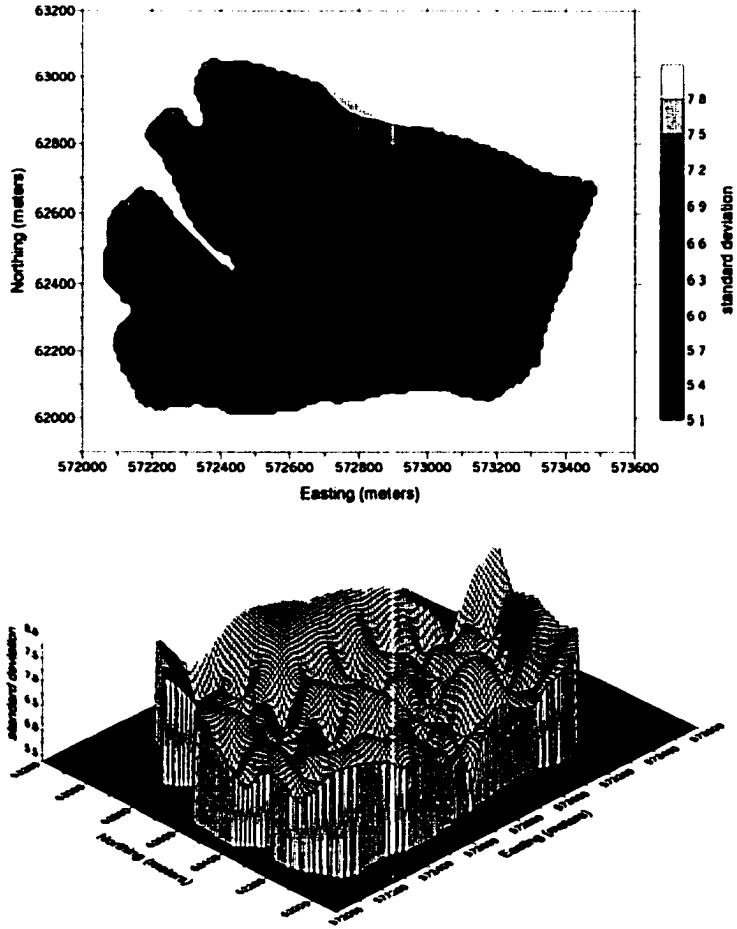

Figure 5.27 Cokriging with primary variable number of stems per plot (STM) and secondary variable slope/aspect (SLOASP): contour map (top graph) and surface map (bottom graph) of standard deviations.

5.1.7 Disjunctive Kriging

The disjunctive kriging technique transformed the original data distribution to a normalized distribution (Figure 5.28).

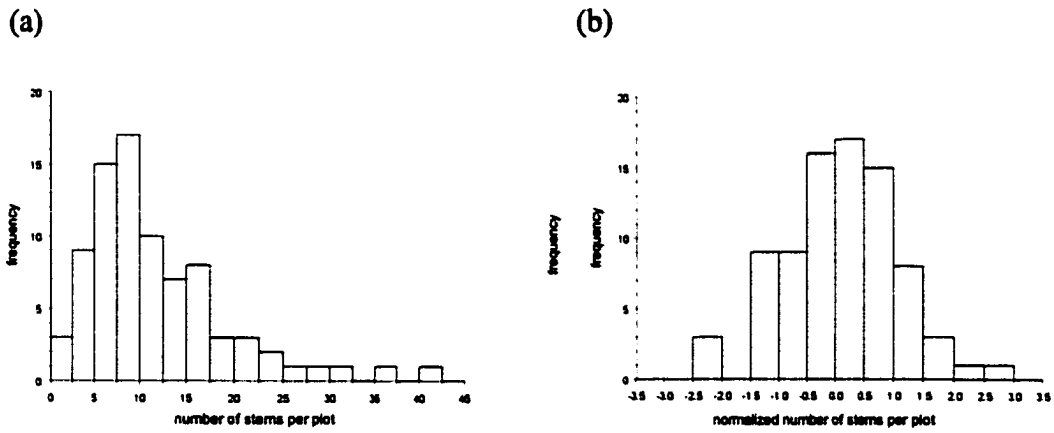

Figure 5.28 Original data distribution (a), and normalized data distribution (b) for number of stems per plot (STM).

Six Hermite polynomials and a seventh-order least-square polynomial were used to transform the original data. The quality of the transformation process was checked with a comparison between the mean and variance of the original data with the transformed data (Table 5.7). Both means and both variances were similar.

Table 5.7 Means and variances of observed and transformed data for number of stems per plot (STM).

	Mean	Variance
Observed data	11.29268	58.45092
Transformed data	10.92424	59.14454

Six Hermite coefficients were calculated and are displayed in Table 5.8.

Table 5.8 The Hermite coefficients for number of stems per plot (STM).

k	C_k
0	10.924240
1	7.197469
2	1.829218
3	0.081426
4	-0.092064
5	-0.058145

The transformed data values were used to calculate an experimental variogram. The lag distance was 160 m and the lag tolerance was 80 m. The spherical variogram model had a nugget of 0.56, a sill of 0.93, and a range of 365 (Figure 5.29).

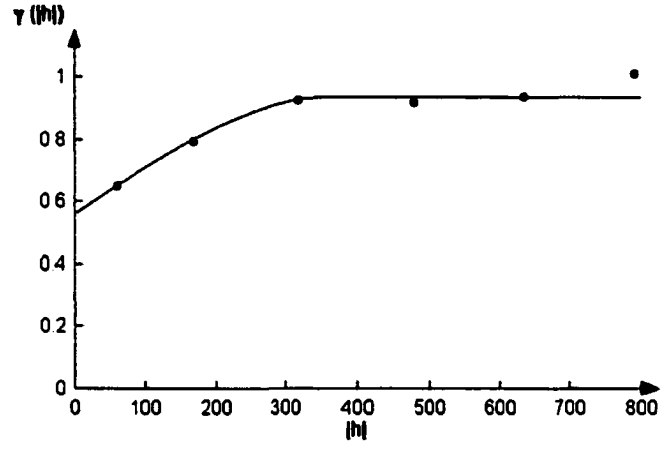

Figure 5.29 Experimental variogram and spherical variogram model for the normalized number of stems per plot (STM).

Including 7 nearest neighbors in the estimation process yielded a MSE of 49.444. Without the possible outlier of value 28.96, the analysis of the residuals did not exhibit any violations of the underlying assumptions (Figure 5.30). The correlation coefficient of

the estimated data values versus the true data values was 0.38. The distribution of the residuals had a mean of - 0.28, and a variance of 49.97. The minimum value was -13.80, and the maximum value was 28.96.

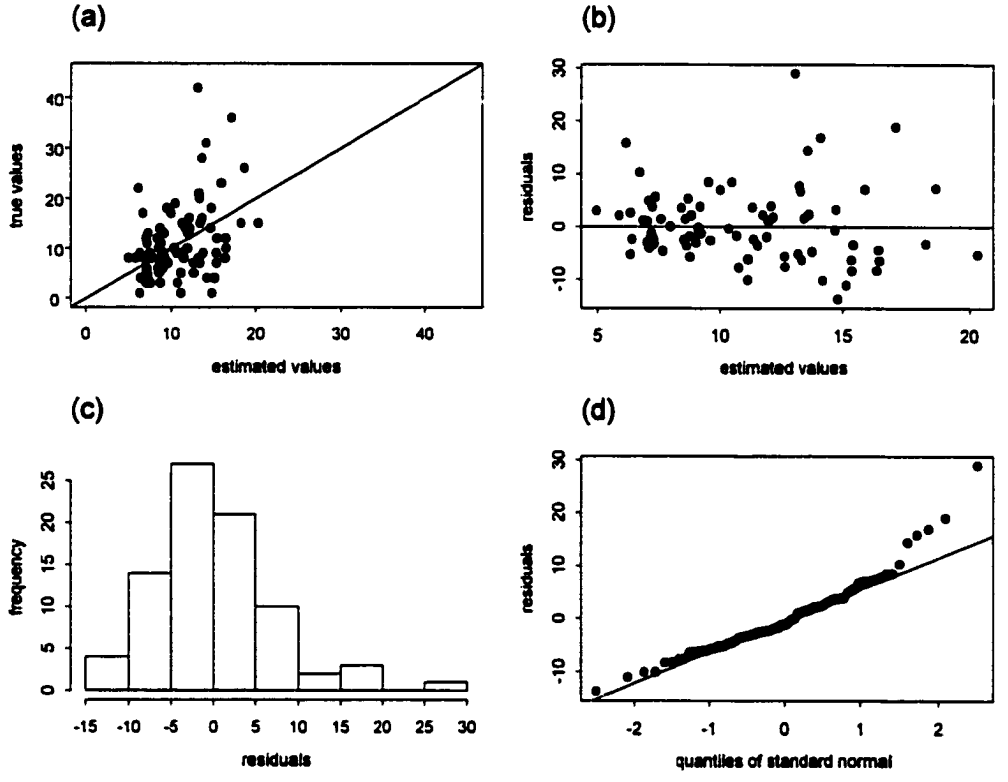

Figure 5.30 Disjunctive kriging with number of stems per plot (STM):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates are displayed in Figure 5.31. The minimum value of the estimates was 5.20, and the maximum value was 20.10. A contour map and surface map of the standard deviations are shown in Figure 5.32. The minimum value of the standard deviations was 6.73, and the maximum value was 7.65. It should be noted that the scale bar is different from the one used to display the standard deviations

maps of the other kriging methods, due to the small range of standard deviations with disjunctive kriging.

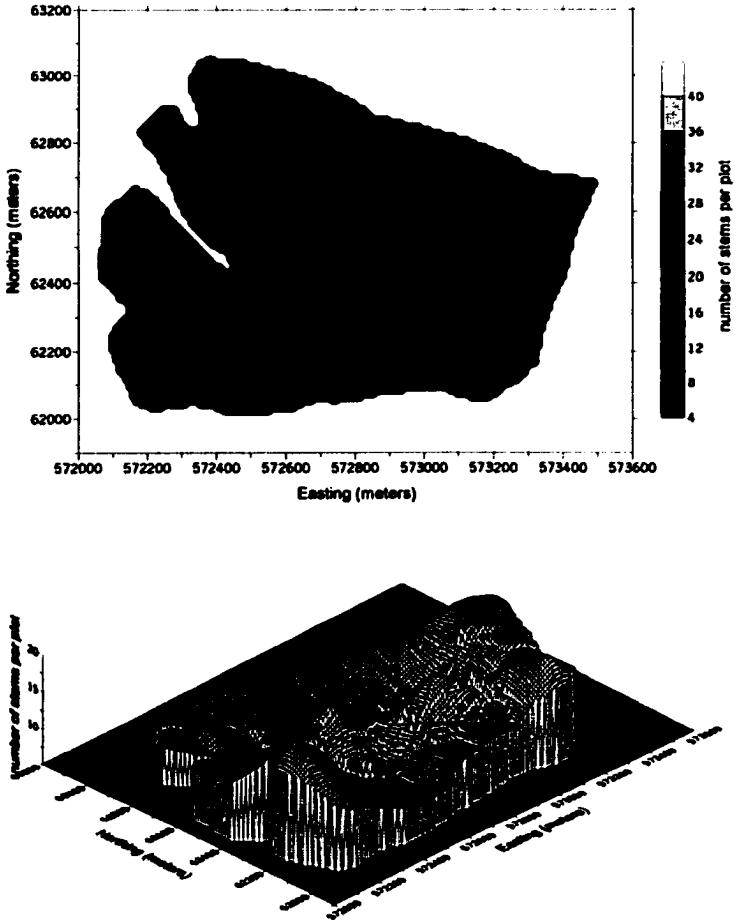

Figure 5.31 Disjunctive kriging with number of stems per plot (STM): contour map (top graph) and surface map (bottom graph) of estimates.

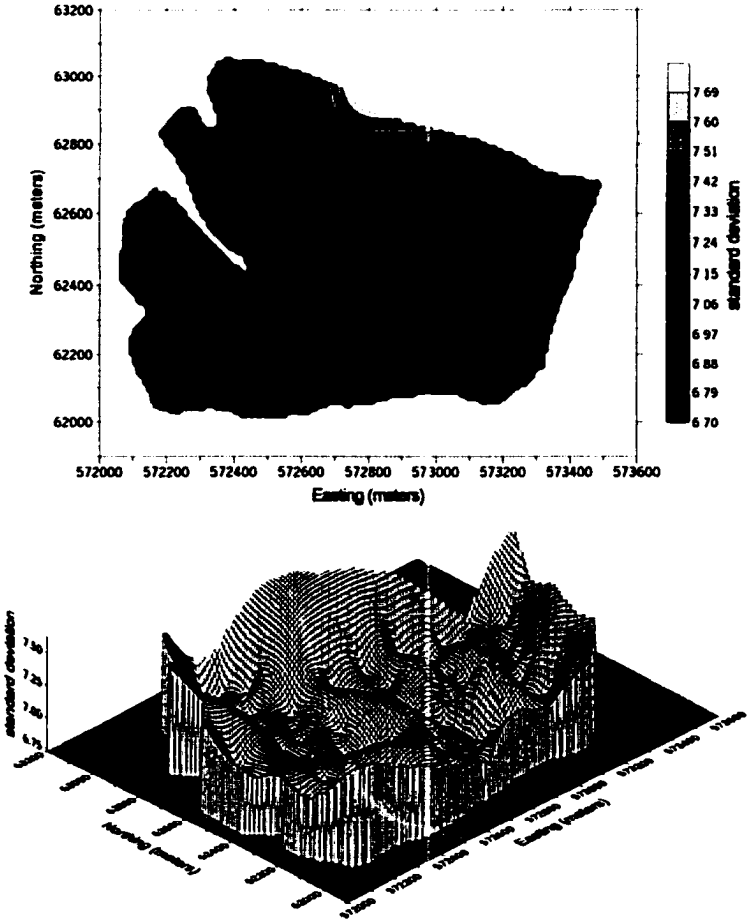

Figure 5.32 Disjunctive kriging with number of stems per plot (STM): contour map (top graph) and surface map (bottom graph) of standard deviations.

As an example for calculating conditional probabilities with disjunctive kriging, a map was created showing the probabilities of each grid cell exceeding 15 stems per plot (Figure 5.33).

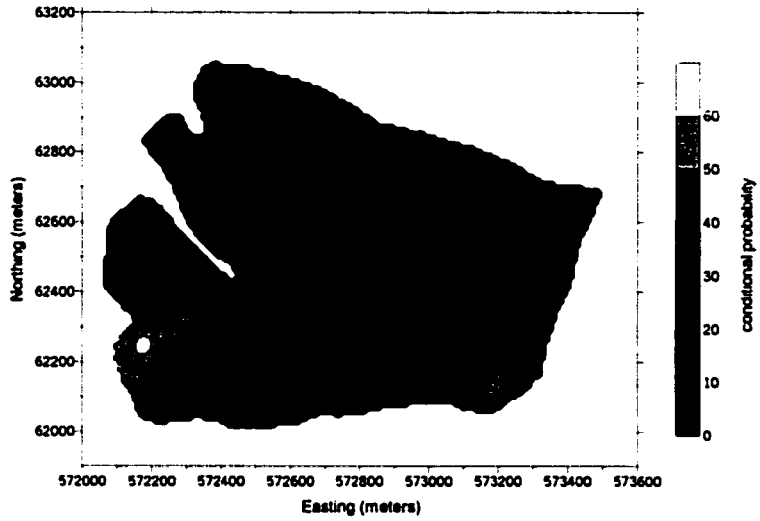

Figure 5.33 Contour map of the conditional probability that the number of stems per plot (STM) is above a value of 15.

5.1.8 Discussion

Eight different interpolation methods were applied with the number of stems per plot (STM) using the mean square error (MSE) for comparison (Table 5.9). Cokriging was the best interpolation technique yielding the lowest MSE of 44.568. The utilization of the spatial cross-correlation between the primary variable STM and secondary variable slope/aspect (SLOASP) helped to improve the estimation accuracy of variable STM. The polygonal mapping technique (MSE of 44.906) was only slightly worse than cokriging. This showed that the spatial distribution of variable STM could be well described with discrete polygons.

Table 5.9 Mean square errors (MSE) of eight interpolation methods for number of stems per plot (STM).

Interpolation Method	MSE
Cokriging	44.568
Polygonal mapping	44.906
Universal kriging (2^{nd} degree trend surface)	45.326
Universal kriging (1^{st} degree trend surface)	46.804
Ordinary kriging	47.443
Inverse distance weighting squared	48.915
Inverse distance weighting	49.140
Disjunctive kriging	49.444

Polygonal mapping was followed in the ranking by universal kriging with a second-degree trend surface (MSE of 45.326), and universal kriging with a first-degree trend surface (MSE of 46.804). The removal of a potential trend in the data by using a first- and second-degree trend surface improved the estimation results in comparison with ordinary kriging (MSE of 47.443). Ordinary kriging outperformed inverse distance weighting squared (MSE of 48.915) and inverse distance weighting (MSE of 49.140). All kriging methods have the advantage that they take the interdependence of the sample points into account by giving less weight to clustered data points. The weights for inverse distance weighting, on the other side, solely depend on the distance from the sample points to the points being estimated. The nonlinear kriging method, disjunctive kriging, scored the lowest with a MSE of 49.444, despite the good results of the transformation process (mean and variance of the original and transformed data were similar, and the transformed data had a normal distribution). Additional information (in the form of spatially cross-correlated auxiliary variables) seems to be a more important consideration

than whether the estimation method is linear or nonlinear. All together, however, the differences in MSEs between the different interpolation methods were relatively small.

When comparing the contour and surface maps of the eight different interpolation methods, one could see that the spatial distribution of STM was approximately the same for all techniques (with the exception of polygonal mapping): zones of low STM and zones of high STM values were located in the same areas. Regions of high values of STM were concentrated in the southwest, southeast, and northeast corner of the study area, while low values occurred in the northern and central part of the area. The levels of STM, however, were different from map to map. Many factors have an influence on the interpolation results and cause different degrees of smoothing. Chapter "Factors Affecting the Results of Interpolation Methods" summarizes all components that have an effect on the interpolation results.

Maps of standard deviation of estimation from inverse distance weighting and inverse distance weighting squared were not comparable with the standard deviation maps resulting from the kriging methods. This is because the calculation of the standard deviation for the inverse distance weighting techniques depends on the estimation mean, whereas the standard deviations for the kriging methods are independent of the estimation mean. The smallest standard deviations occurred for estimation points at, or close to the sample plot locations. The values increased with increasing distance from the sample plot locations, with the largest standard deviations along the boundaries of the study area. The levels of the standard deviation from the various kriging methods were different. A thorough discussion of all components affecting the standard deviations of estimation is given in chapter "Factors Affecting the Results of Interpolation Methods."

5.2 Total Basal Area

5.2.1 Summary Statistics

The following summary statistics for *total basal area per plot* (TBA) were calculated: minimum and maximum value, median, mean, standard deviation, variance, skewness, and kurtosis (Table 5.10). The distribution of variable TBA was slightly skewed to the right as shown in Figure 5.34.

Table 5.10 Summary statistics for total basal area per plot (TBA).

Statistics	Values
Minimum	0.20
Median	3.60
Mean	3.71
Maximum	10.80
Standard Deviation	2.06
Variance	4.20
Skewness	0.79
Kurtosis	3.75

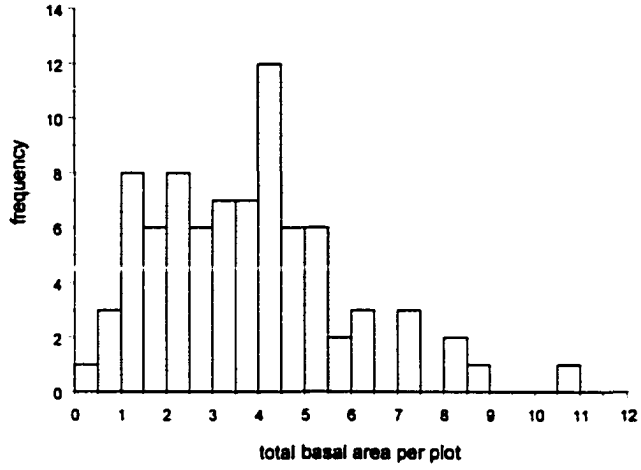

Figure 5.34 Histogram for total basal area per plot (TBA).

5.2.2 Polygonal Mapping

The average total basal area per plot (TBA) was calculated for each forest stand polygon. Table 5.11 shows the number of sample plots, the average total basal area, and the standard deviation for each polygon.

Table 5.11 The number of sample plots, the mean, and the standard deviation of each forest stand polygon for total basal area per plot (TBA).

Stand number	Number of sample plots	Mean	Standard deviation
1	10	3.84	2.19
2	3	3.00	1.28
3	9	3.26	1.75
4	7	4.47	1.38
5	15	5.24	3.02
6	11	3.00	1.19
7	23	3.25	1.59
8	4	2.40	1.20

An ANOVA table (Table 5.12) was calculated to test the null hypothesis of all eight polygonal means being identical. The p-value of 0.0413 was lower than the type I error set to 0.05. Therefore, the null hypothesis of all eight polygonal means being identical was rejected; at least one polygonal mean was different from the others.

Table 5.12 ANOVA table for total basal area per plot (TBA).

Source	Sum of Squares	Degrees of Freedom	Mean Square	F Test	p-value
Between groups	59.942	7	8.563	2.23	0.0413
Within groups	284.573	74	3.845		
Totals	344.515	81			

Table 5.13 shows the outcome from Tukey's HSD test. None of the stands was significantly different from each other. Again, it should be pointed out that both, the ANOVA calculation and Tukey's HSD test are based on a normal distribution of the data. Variable TBA, however, is slightly skewed to the right. Therefore, the results of both tests should be taken with caution.

Table 5.13 Tukey's HSD test for the average total basal area per plot (TBA).

Tukey Grouping	Mean	Stand
A	5.24	5
A		
A	4.47	4
A		
A	3.84	1
A		
A	3.26	3
A		
A	3.25	7
A		
A	3.00	2
A		
A	3.00	6
A		
A	2.40	8

The measures of validation are summarized in Figure 5.35. The mean square error (MSE) was 3.464. The cloud of points in the scatterplot of the estimated data values versus true data values had a correlation coefficient of 0.42. The scatterplot of estimated data values versus residuals depicted that the errors were fairly evenly distributed. The histogram and the q-q plot also showed that the residuals were approximately normally distributed. The frequency distribution of the residuals had a mean of 0.01, and a variance of 3.51. The minimum value was -4.28, and the maximum value was 5.58.

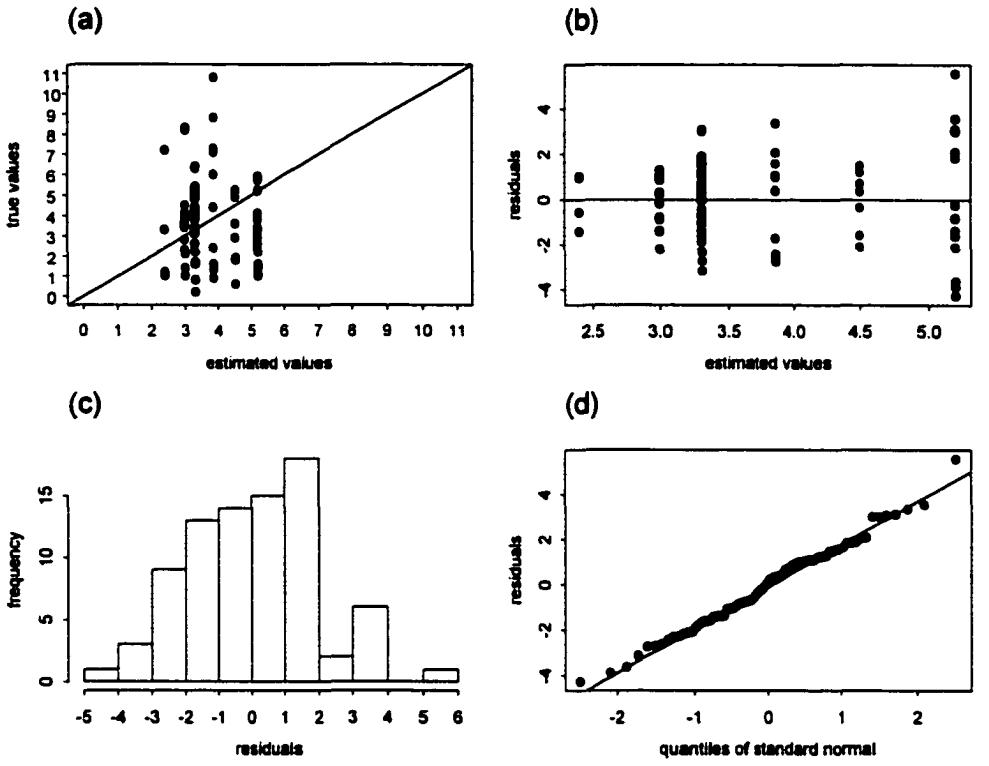

Figure 5.35 Polygonal mapping with total basal area per plot (TBA):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A polygonal map and a surface map showing the TBA per polygon, as well as a polygonal map and a surface map showing the standard deviation per polygon are displayed in Figure 5.36 and Figure 5.37. Polygon number 8 had the smallest value with 2.40, and polygon 5 the highest value with 5.24. Polygon number 6 had the smallest standard deviation of 1.19, whereas polygon 5 had the highest with 3.02. It should be mentioned that the standard deviations should not be compared with each other as they are based on different means.

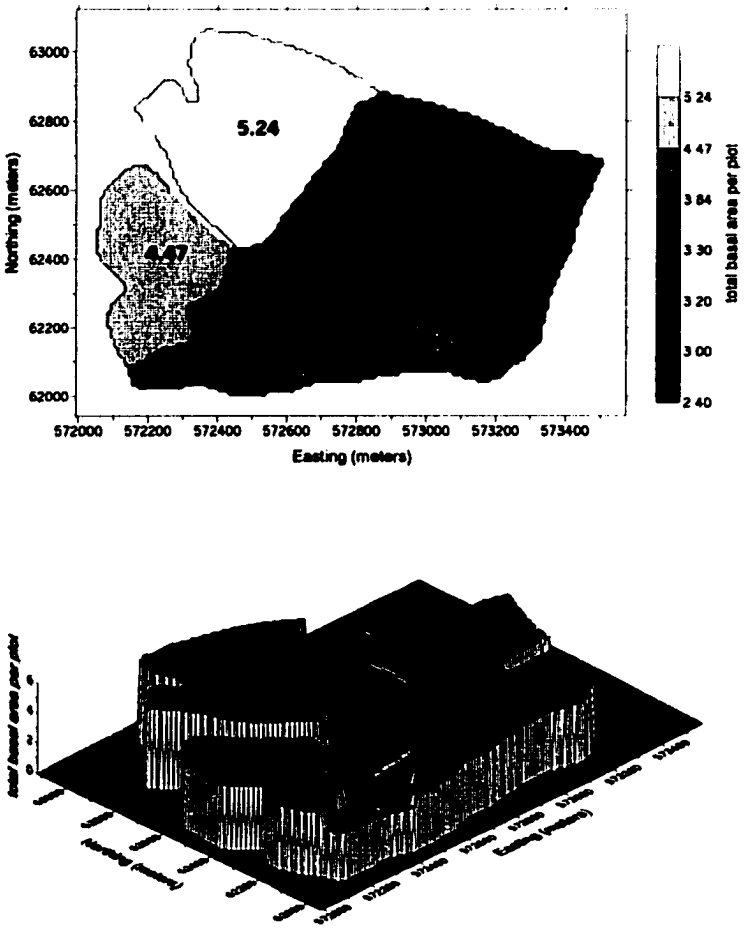

Figure 5.36 Polygonal mapping with total basal area per plot (TBA): polygonal map (top graph) and surface map (bottom graph) showing the mean per polygon.

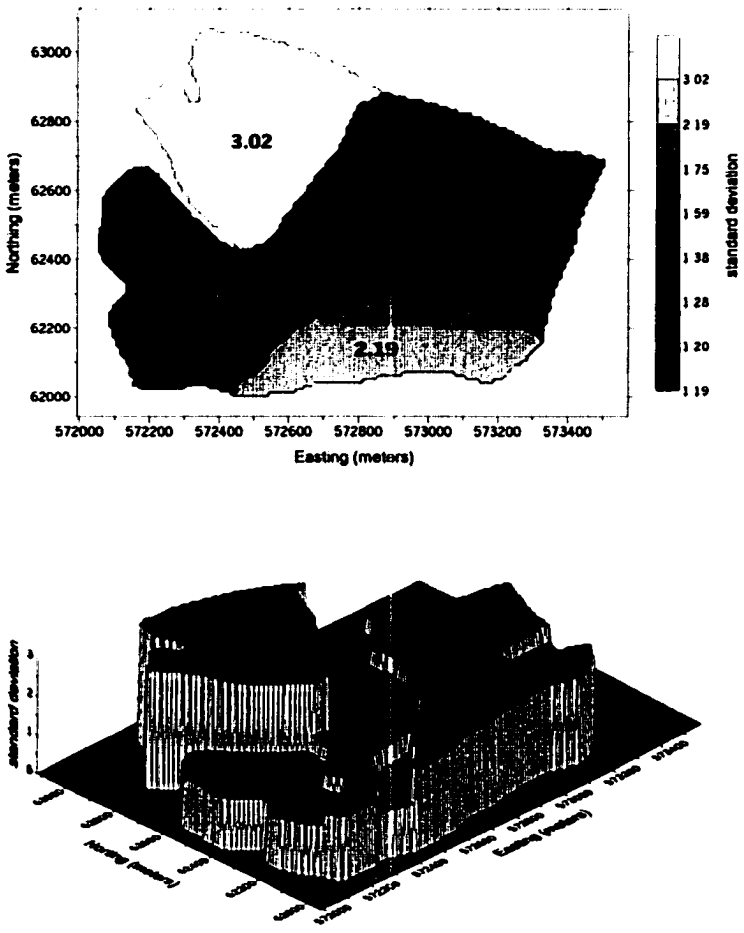

Figure 5.37 Polygonal mapping with total basal area per plot (TBA): polygonal map (top graph) and surface map (bottom graph) showing the standard deviation per polygon.

5.2.3 Inverse Distance Weighting

Inverse distance weighting and inverse distance weighting squared were used to obtain estimates for total basal area per plot (TBA) at unsampled locations.

Inverse Distance Weighting

A minimum MSE of 4.122 could be achieved by including the 20 nearest neighbors. The other four measures of validation are displayed in Figure 5.38. The residuals were approximately normally distributed, had a constant variance, and no apparent trend in the data. The estimated data values versus true values had a correlation coefficient of 0.21. The distribution of the residuals had a mean of -0.07, a variance of 4.17, with a minimum value of -5.02, and a maximum value of 5.99.

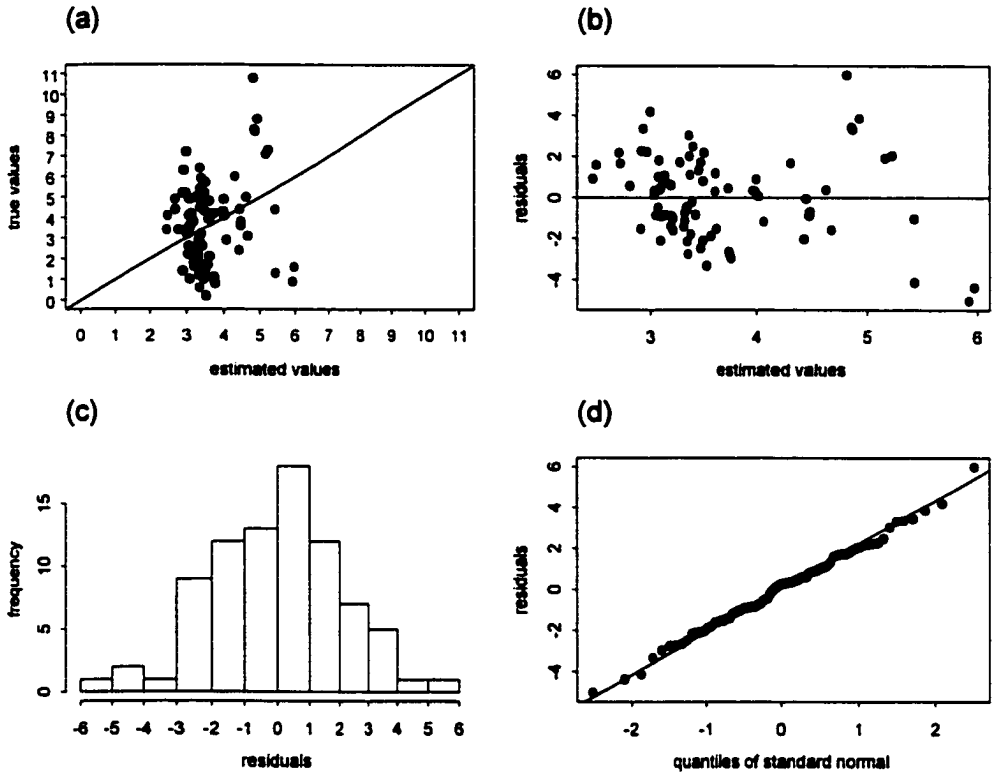

Figure 5.38 Inverse distance weighting with total basal area per plot (TBA):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates are displayed in Figure 5.39. The minimum value of the estimates was 0.98, and the maximum value was 9.45. A contour map and surface map of the standard deviations are shown in Figure 5.40. The minimum value of the standard deviations was 0.05, the maximum value was 1.64.

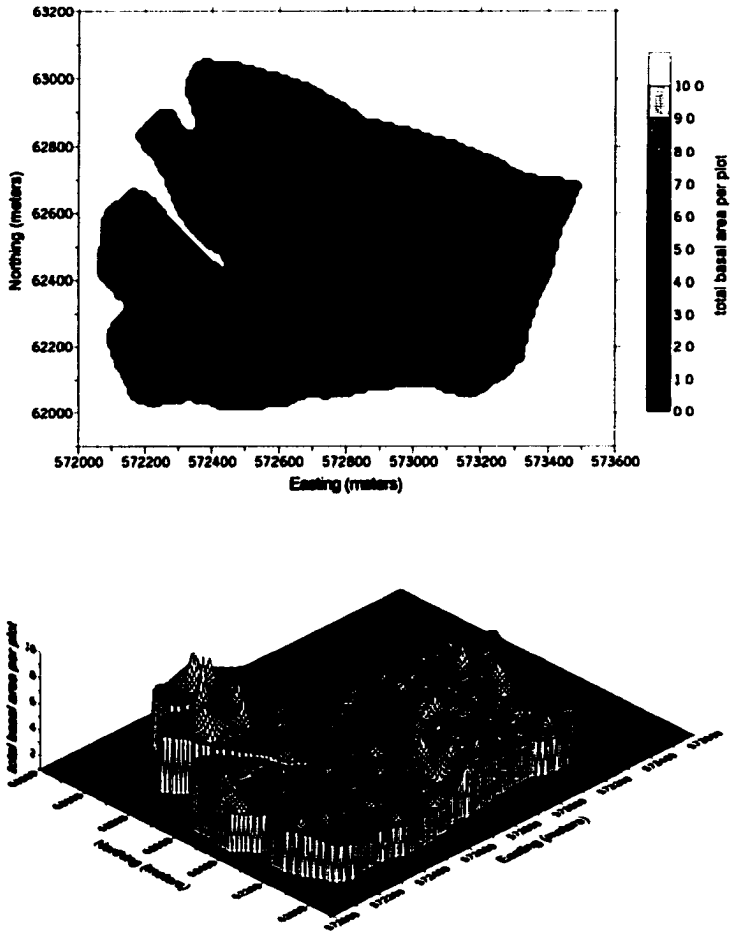

Figure 5.39 Inverse distance weighting with total basal area per plot (TBA): contour map (top graph) and surface map (bottom graph) of estimates.

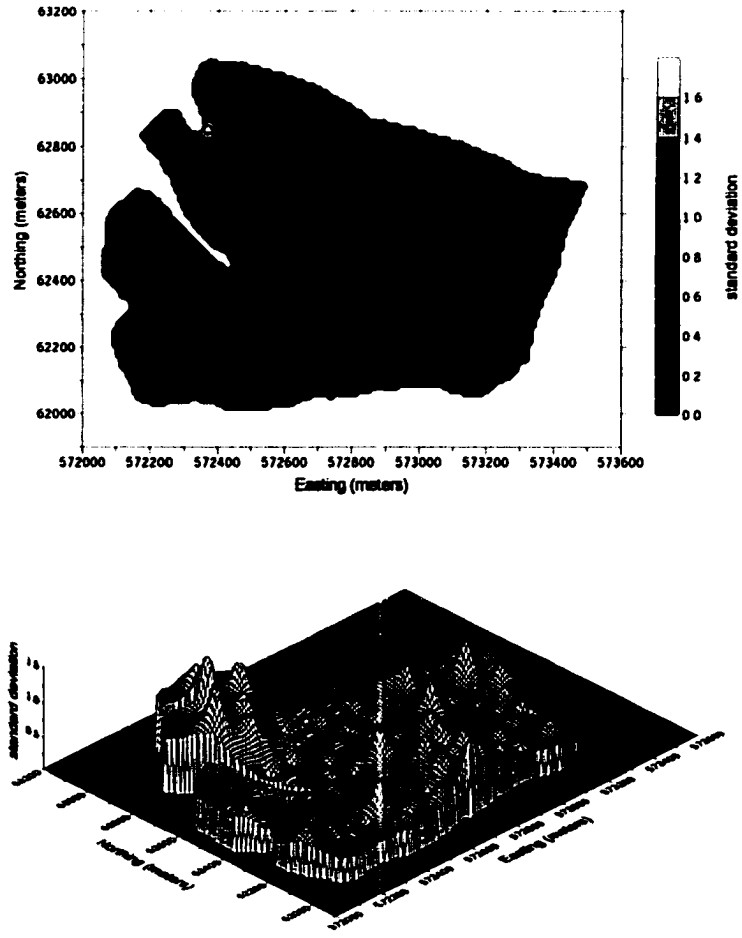

Figure 5.40 Inverse distance weighting with total basal area per plot (TBA): contour map (top graph) and surface map (bottom graph) of standard deviations.

Inverse Distance Weighting Squared

A minimum MSE of 4.598 was obtained using the 20 nearest neighbors. The correlation coefficient between estimated and true data values was 0.15. The residuals were approximately normally distributed, with a zero mean, a constant variance, and with no apparent trend in the data (Figure 5.41). The distribution of the residuals had a mean of -0.05, a variance of 4.65, with a minimum value of -5.89, and a maximum value of 5.87.

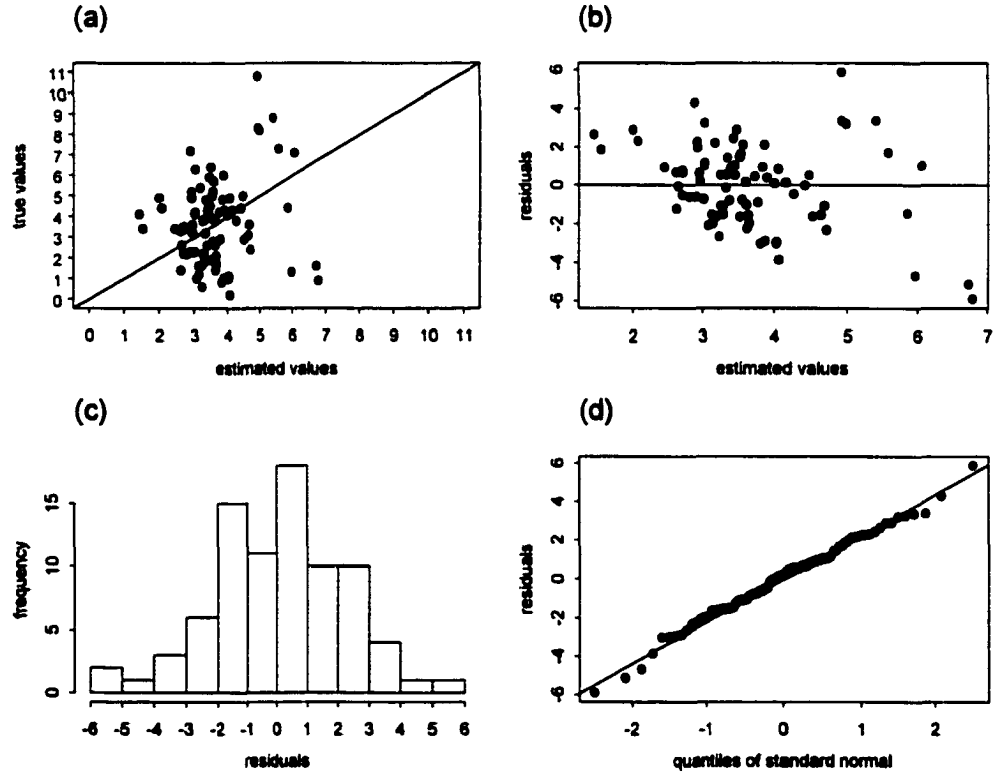

Figure 5.41 Inverse distance weighting squared with total basal area per plot (TBA):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates are displayed in Figure 5.42. The minimum value of the estimates was 0.23, and the maximum value was 10.76. A contour map and surface map of the standard deviations are shown in Figure 5.43. The minimum value of the standard deviations was 0, and the maximum value was 2.47.

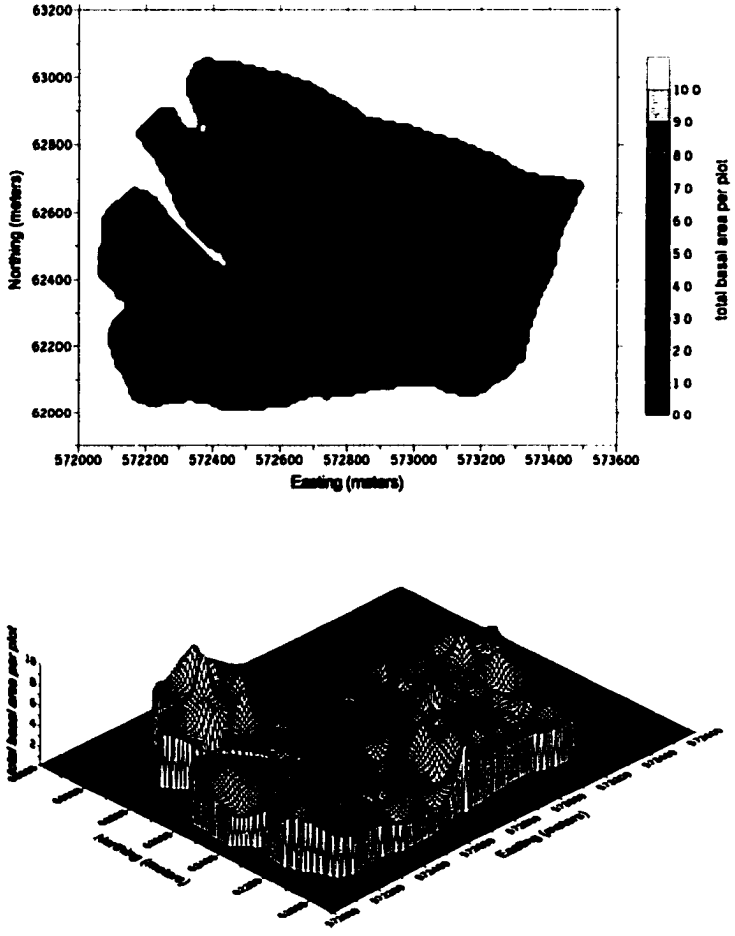

Figure 5.42 Inverse distance weighting squared with total basal area per plot (TBA): contour map (top graph) and surface map (bottom graph) of estimates.

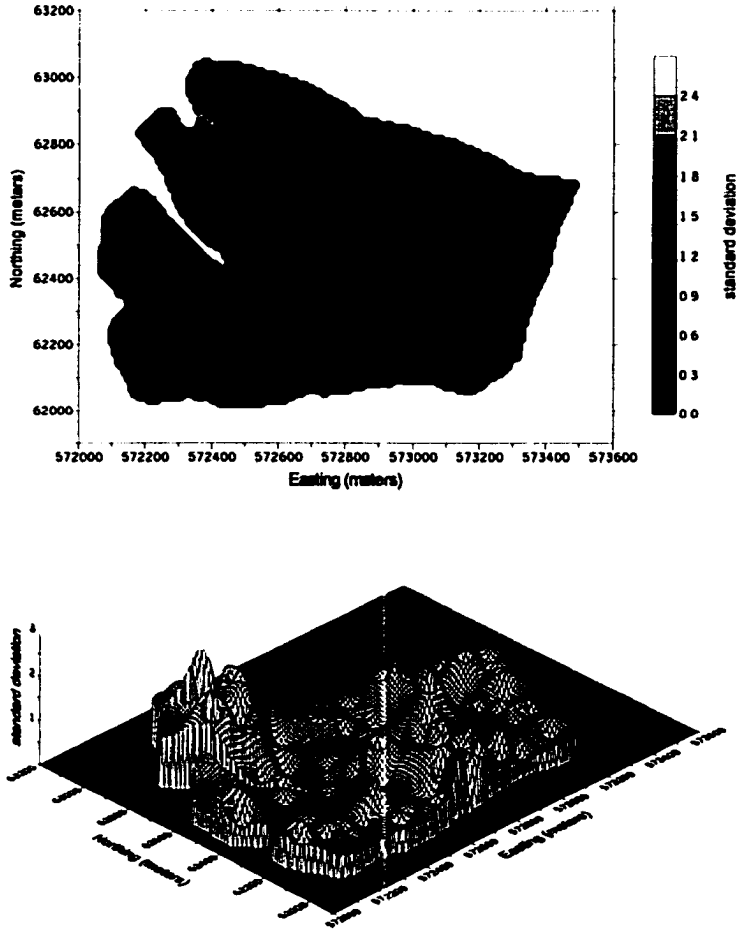

Figure 5.43 Inverse distance weighting squared with total basal area per plot (TBA): contour map (top graph) and surface map (bottom graph) of standard deviations.

5.2.4 Ordinary Kriging

An experimental variogram for total basal area per plot (TBA). The lag distance was 135 m and the lag tolerance was 67.5 m. A power model was then fitted to the experimental variogram (Figure 5.44). The model was specified with a power of 1.84 and a slope of 0.0001.

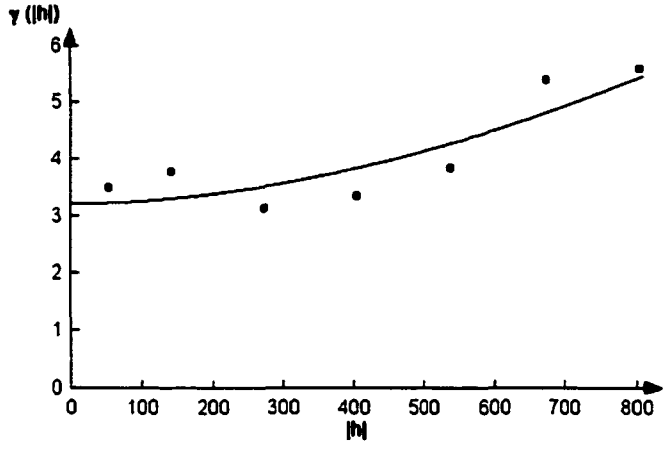

Figure 5.44 Experimental variogram and power variogram model for total basal area per plot (TBA).

Using 20 nearest neighbors resulted in the smallest MSE of 3.951. The correlation coefficient between estimated and true data values was 0.26. The distribution of the residuals had a mean of -0.03, a variance of 4.00, with a minimum value of -4.58, and a maximum value of 5.80. Figure 5.45 showed that the residuals had an approximately normal distribution, a constant variance, and no apparent trend in the data.

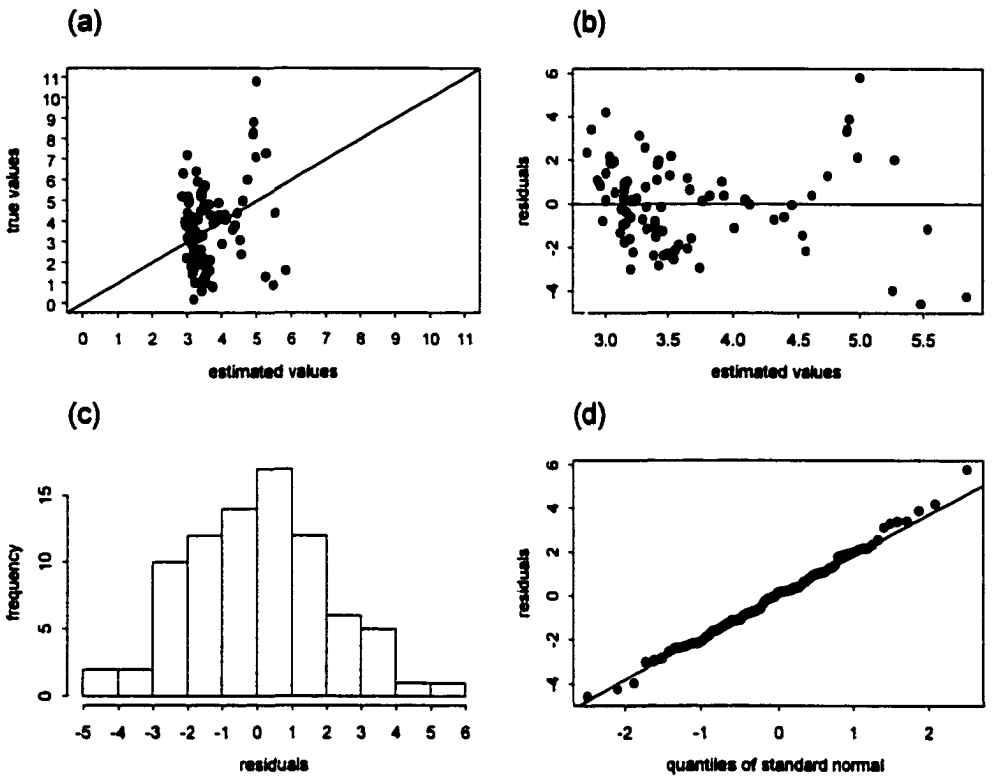

Figure 5.45 Ordinary kriging with total basal area per plot (TBA):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates are displayed in Figure 5.46. The minimum value of the estimates was 2.64, and the maximum value was 5.87. A contour map and surface map of the standard errors are shown in Figure 5.47. The minimum value of the standard errors was 1.66, the maximum value was 1.95.

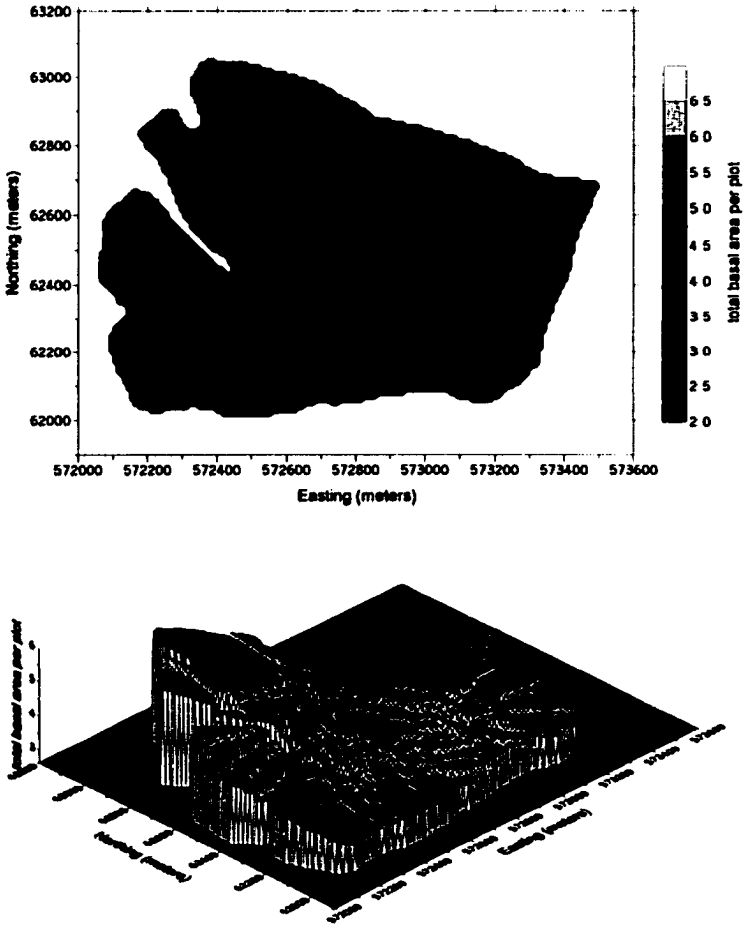

Figure 5.46 Ordinary kriging with total basal area per plot (TBA): contour map (top graph) and surface map (bottom graph) of estimates.

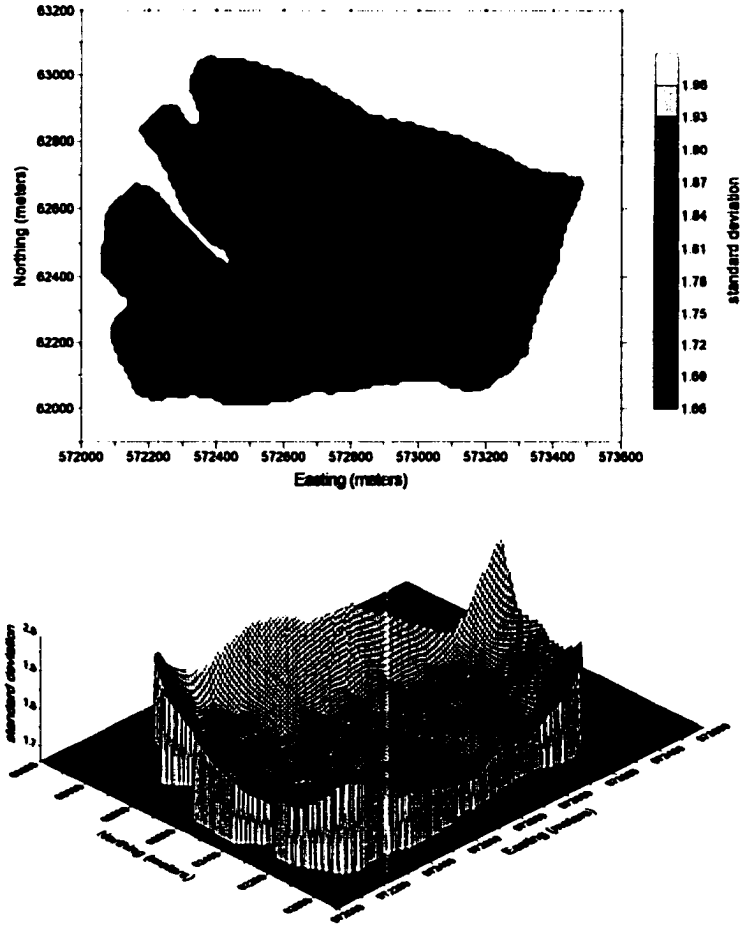

Figure 5.47 Ordinary kriging with total basal area per plot (TBA): contour map (top graph) and surface map (bottom graph) of standard deviations.

5.2.5 Universal Kriging

A first- and second-degree trend surface were calculated as a function of x- and y-coordinates of the sample plot locations for total basal area per plot (TBA).

First-Degree Trend Surface

A first-degree trend surface was fitted to the total basal area per plot (TBA) data set. An R^2 value of 0.12 was achieved. The trend surface was subtracted from the original data values, and the residuals were then used to calculate the experimental variogram (Figure 5.48). A lag distance of 135 m and a lag tolerance of 67.5 m were chosen. The power variogram model had the following specifications: a power of 1.72, and a slope of 0.0001.

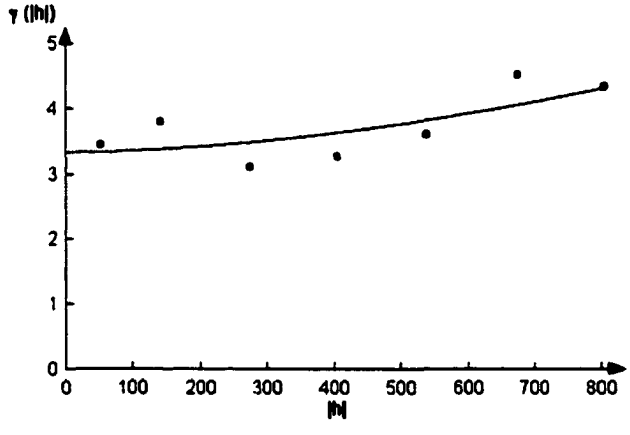

Figure 5.48 Experimental variogram and power variogram model for the residuals of the first-degree trend surface for total basal area per plot (TBA).

The smallest MSE was obtained using the 20 nearest neighbors: 3.916. The correlation coefficient between estimated and true data values was 0.28. The distribution of the residuals had a mean of -0.04, a variance of 3.96, with a minimum value of -4.69,

and a maximum value of 5.63. The analysis of the residuals did not exhibit any peculiarities that would suggest violations of the underlying assumptions (Figure 5.49).

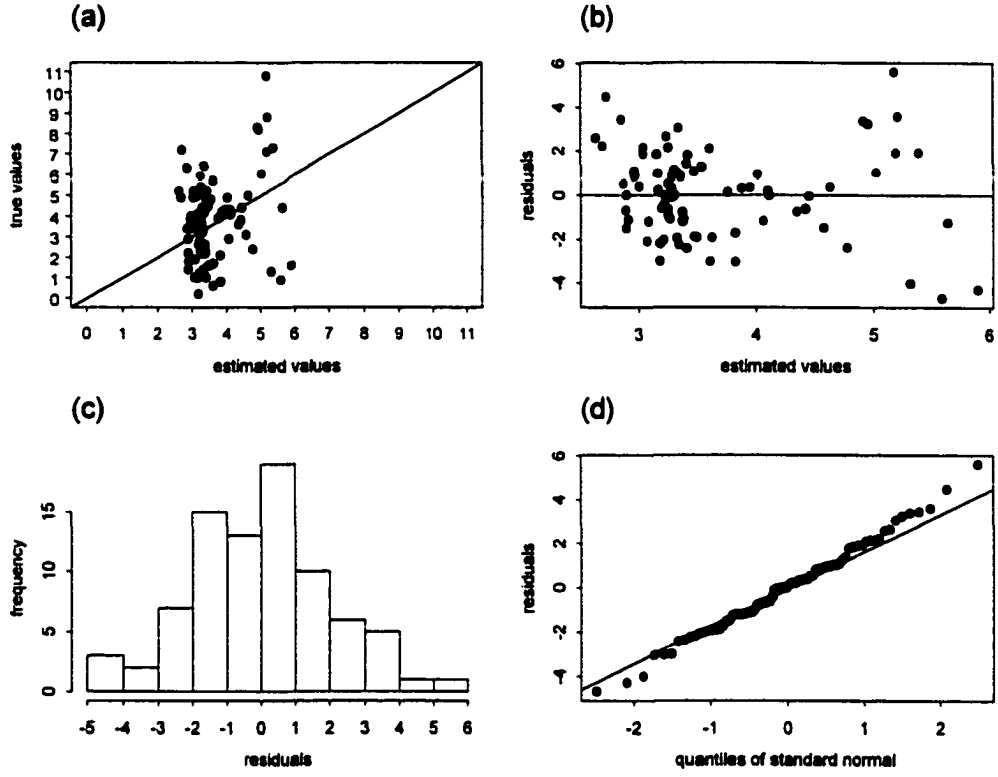

Figure 5.49 Universal kriging with first-degree trend surface for total basal area per plot (TBA):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

Ordinary kriging was conducted on the residuals of the first-degree trend surface using the power variogram model. The final surface was created by combining the first-degree trend surface with the kriged surface of the residuals (Figure 5.50). The minimum value of the estimates was 2.59, and the maximum value was 5.53. A contour map and surface map of the standard deviations are shown in Figure 5.51. The minimum value of the standard deviations was 1.72, and the maximum value was 1.92.

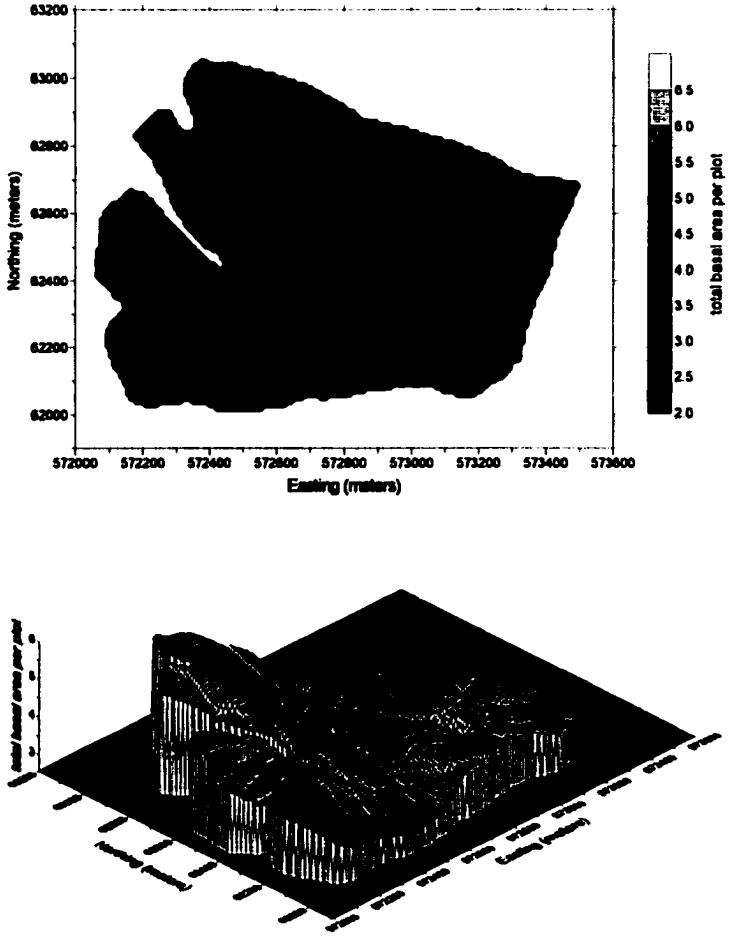

Figure 5.50 Universal kriging with first-degree trend surface for total basal area per plot (TBA): contour map (top graph) and surface map (bottom graph) of estimates.

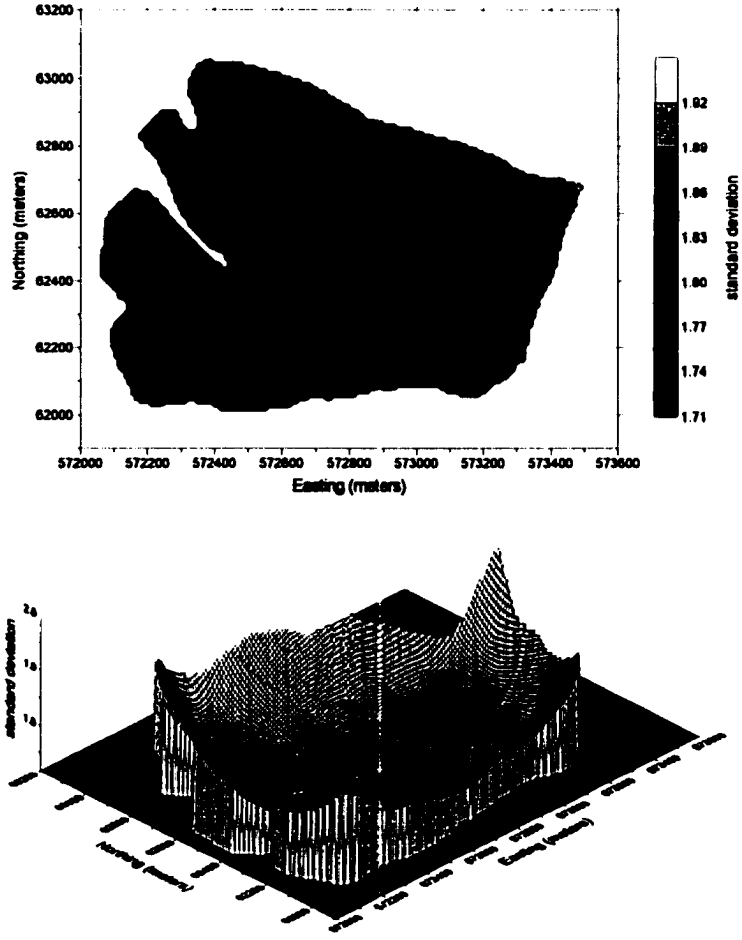

Figure 5.51 Universal kriging with first-degree trend surface for total basal area per plot (TBA): contour map (top graph) and surface map (bottom graph) of standard deviations.

Second-Degree Trend Surface

A second-degree trend surface was fitted to the data set. An R^2 value of 0.17 was achieved. The experimental variogram for the residuals of the second-degree trend surface had a lag distance of 135 m and a lag tolerance of 67.5 m (Figure 5.52). The power variogram model had a power of 1.62 and a slope of 0.00001.

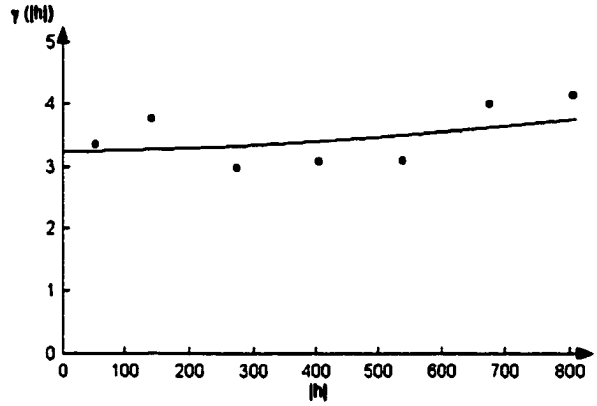

Figure 5.52 Experimental variogram and power variogram model for the residuals of the second-degree trend surface for total basal area per plot (TBA).

The smallest MSE of 3.835 was achieved by including the 20 nearest neighbors in the kriging process. The correlation coefficient between estimated and true data values was 0.32. The analysis of the residuals did not exhibit any peculiarities that would suggest violations of the underlying assumptions (Figure 5.53). The distribution of the residuals had a mean of 0.01, a variance of 3.88, with a minimum value of -4.99, and a maximum value of 5.12.

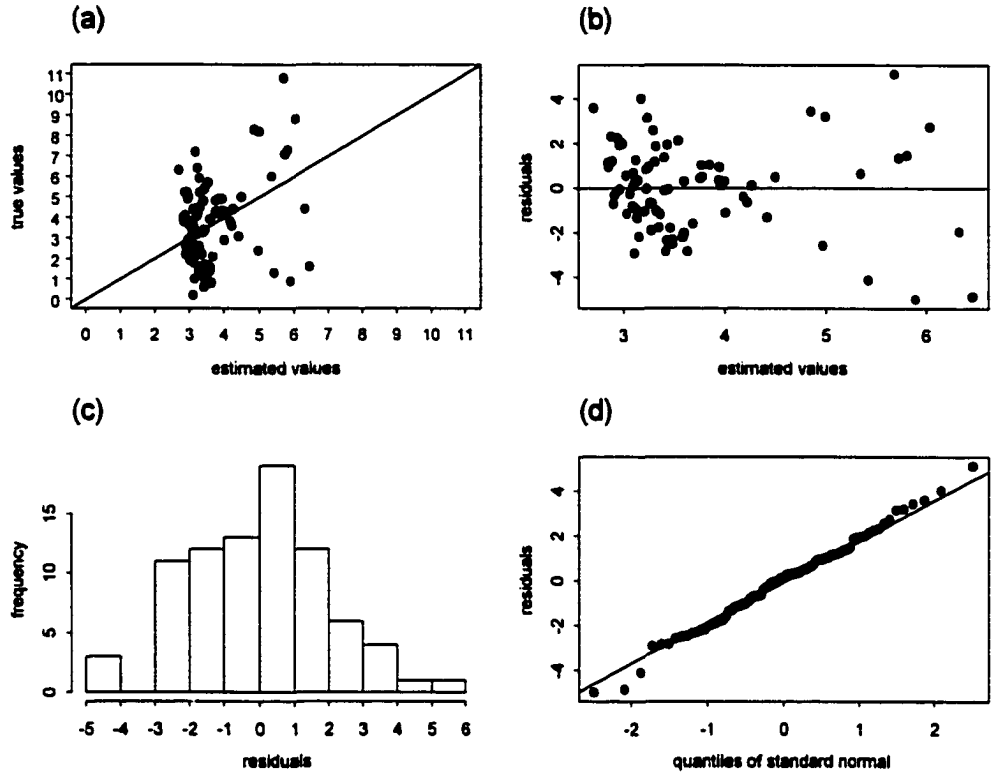

Figure 5.53 Universal kriging with second-degree trend surface for total basal area per plot (TBA):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

Ordinary kriging was conducted on the residuals of the second-degree trend surface using the power variogram model. The final surface was created by combining the second-degree trend surface with the kriged surface of the residuals (Figure 5.54). The minimum value of the estimates was 2.49, and the maximum value was 5.11. A contour map and surface map of the standard deviations are shown in Figure 5.55. The minimum value of the standard deviations was 1.71, and the maximum value was 1.88.

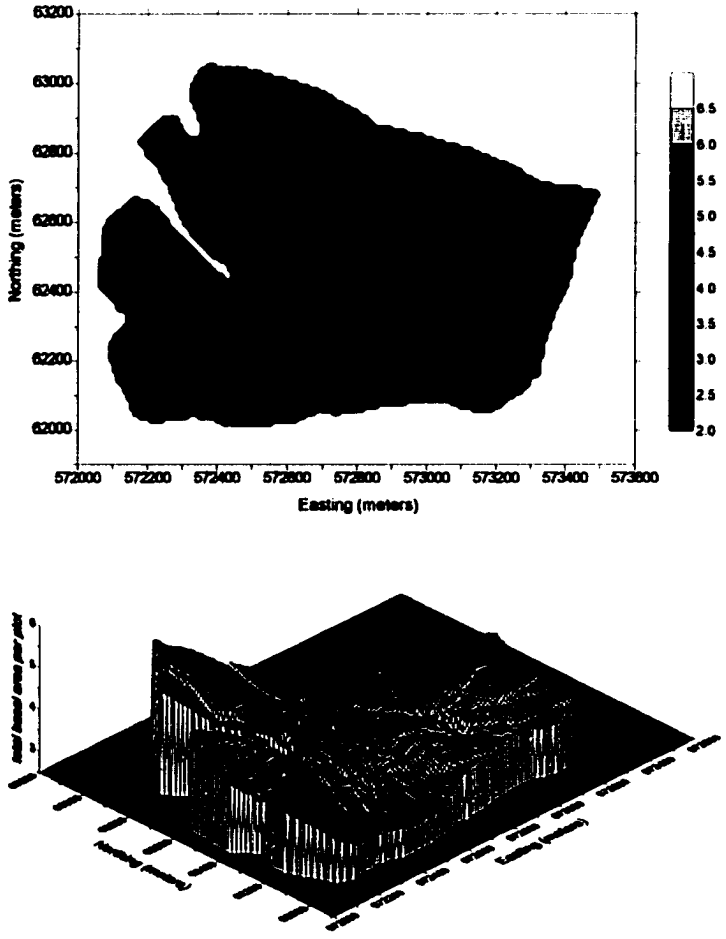

Figure 5.54 Universal kriging with second-degree trend surface for total basal area per plot (TBA): contour map (top graph) and surface map (bottom graph) of estimates.

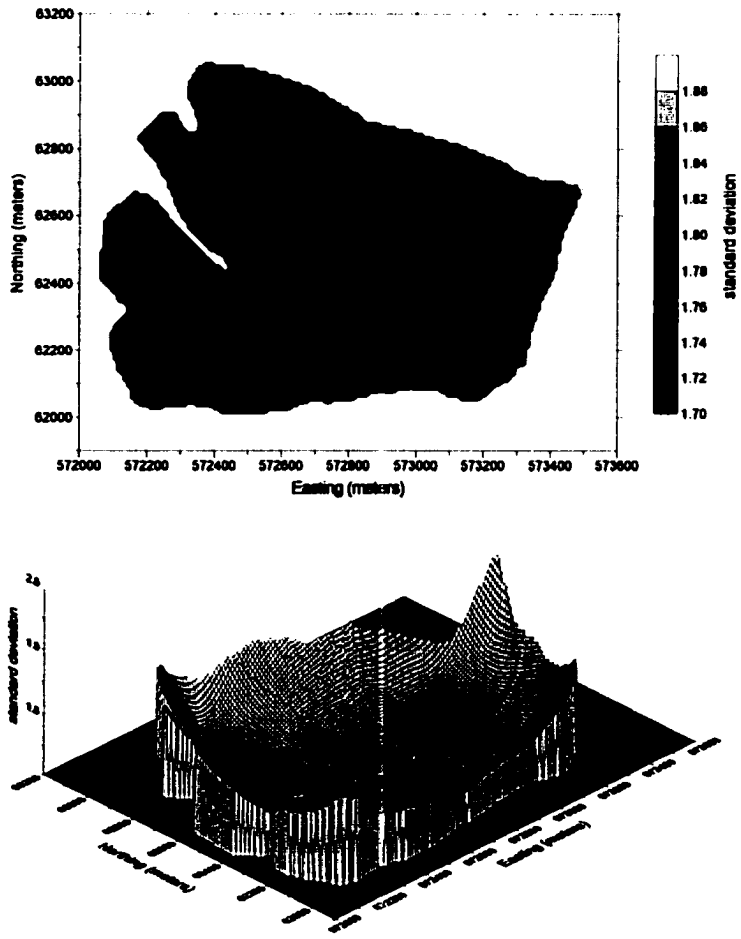

Figure 5.55 Universal kriging with second-degree trend surface for total basal area per plot (TBA): contour map (top graph) and surface map (bottom graph) of standard deviations.

5.2.6 Cokriging

The primary variable total basal area per plot (TBA) was combined with different sets of the following secondary variables: elevation (ELEV), a combined value of slope and aspect (SLOASP), and the Normalized Difference Vegetation Index (NDVI). From all possible combinations of secondary variables, using elevation (ELEV) as auxiliary variable yielded the smallest MSE of 3.911 (Table 5.14).

Table 5.14 Mean square errors (MSE) for primary variable total basal area per plot (TBA) using different combinations of secondary variables: elevation (ELEV), slope/aspect (SLOASP), and normalized difference vegetation index (NDVI).

Secondary Variables	MSE
ELEV	3.911
SLOASP	4.154
NDVI	4.139
ELEV and SLOASP	4.115
ELEV and NDVI	4.118
SLOASP and NDVI	4.153
ELEV, SLOASP, and NDVI	4.127

A power model was fitted to both experimental variograms and the cross-variogram (Figure 5.56). The lag spacing was 135 m, and the lag tolerance was 67.5 m. Table 5.15 shows the specifications of the models. TBA and ELEV were positively spatially cross-correlated as can be seen in the cross-variogram.

Table 5.15 Variogram and cross-variogram model specifications for primary variable total basal area per plot (TBA) and secondary variable elevation (ELEV).

Variogram / Cross-variogram	Power	
	power	slope
TBA	1.84	0.00001
ELEV	1.90	0.03050
TBA-ELEV	1.90	0.00017

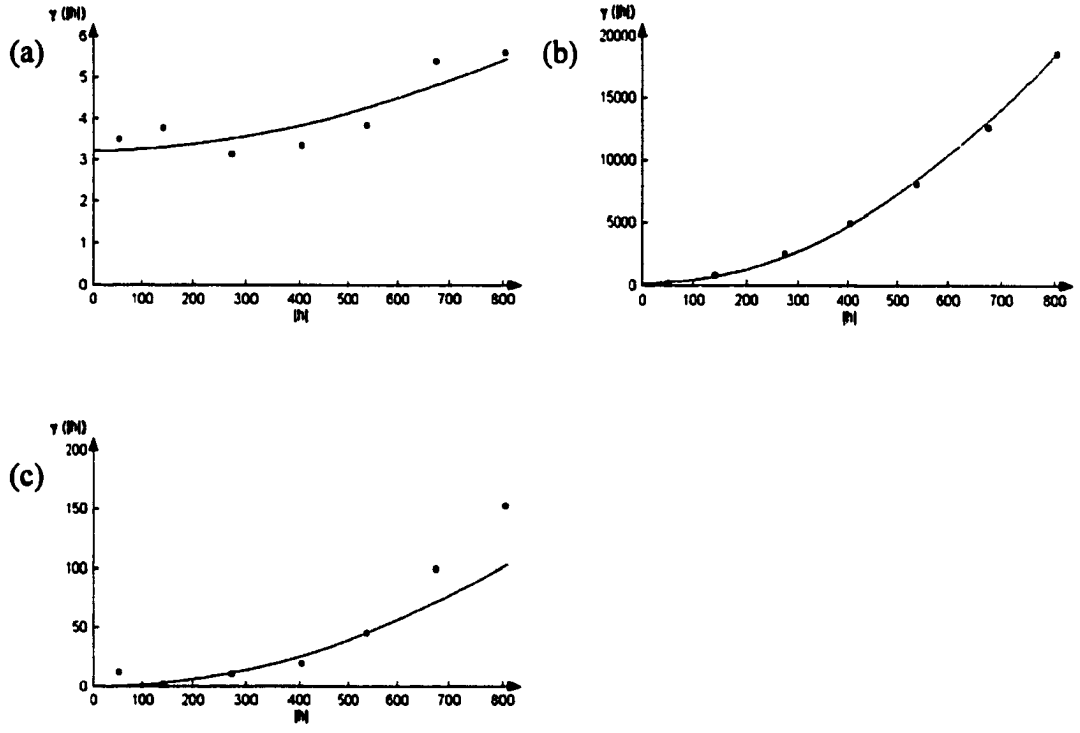

Figure 5.56 (a) Experimental variograms for primary variable total basal area per plot (TBA) and (b) secondary variable elevation (ELEV), as well as (c) experimental cross-variogram with a power (cross-) variogram model.

Including the 20 nearest neighbors resulted in the smallest MSE of 3.911. The residuals were approximately normally distributed, had a constant variance, and had no apparent trend in the data (Figure 5.57). The distribution of the residuals had a mean of

0.02, a variance of 3.96, with a minimum value of -4.69, and a maximum value of 5.66. The correlation coefficient between estimated and true data values was 0.28.

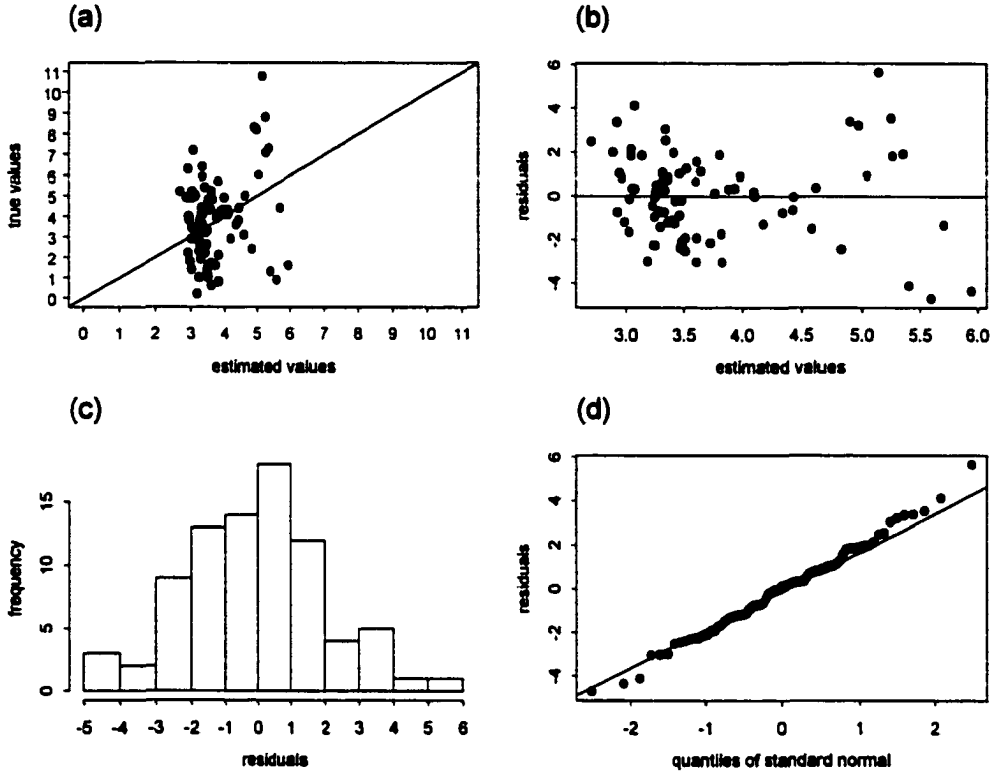

Figure 5.57 Cokriging with primary variable total basal area per plot (TBA) and secondary variable elevation (ELEV):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates are displayed in Figure 5.58. The minimum value of the estimates was 2.65, and the maximum value was 6.12. A contour map and surface map of the standard deviations are shown in Figure 5.59. The minimum value of the standard deviations was 1.84, and the maximum value was 1.94.

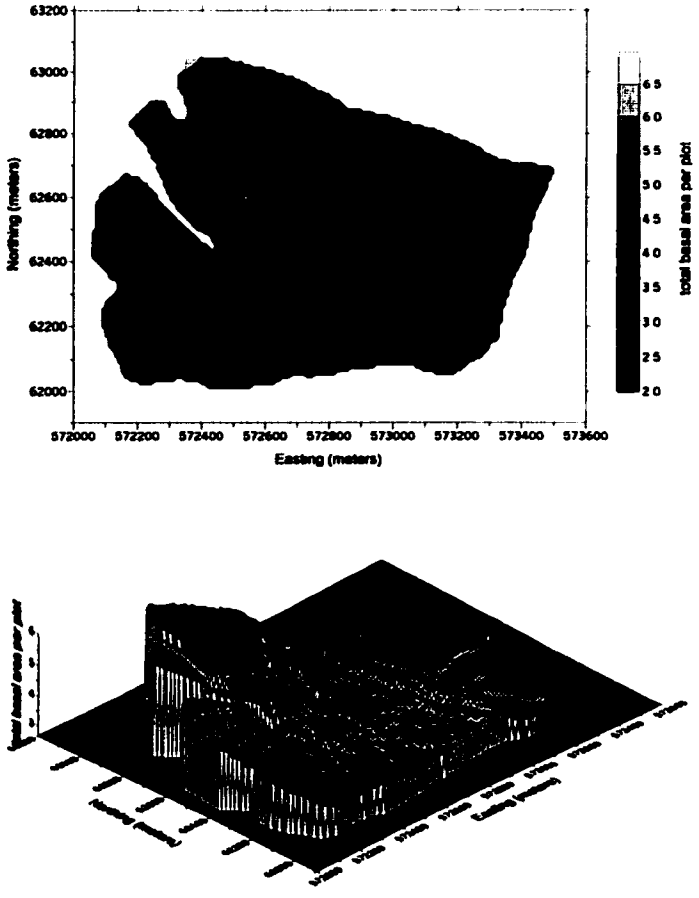

Figure 5.58 Cokriging with primary variable total basal area per plot (TBA) and secondary variable elevation (ELEV): contour map (top graph) and surface map (bottom graph) of estimates.

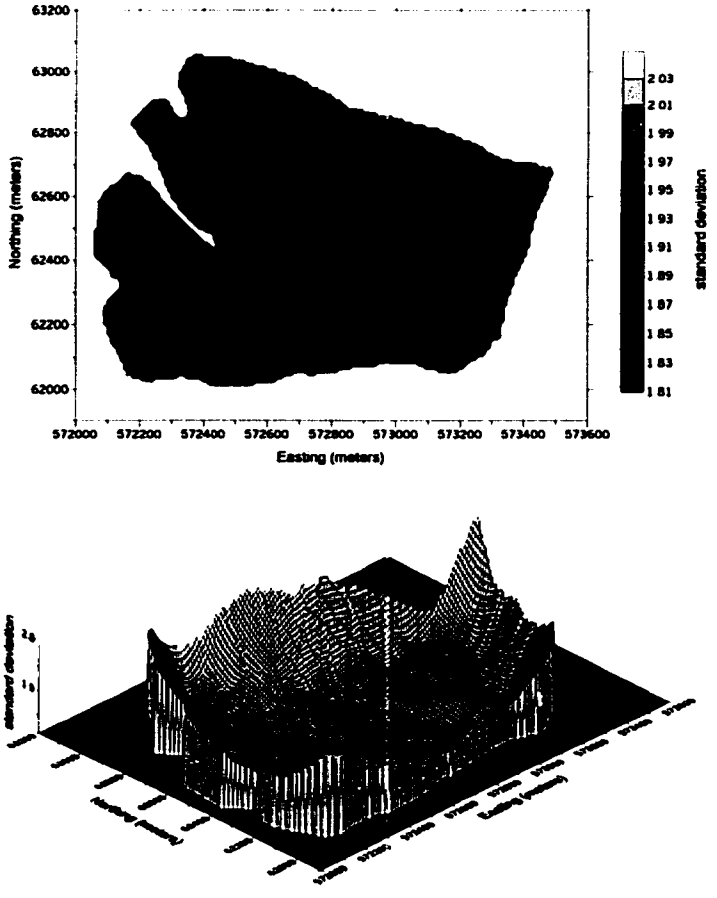

Figure 5.59 Cokriging with primary variable total basal area per plot (TBA) and secondary variable elevation (ELEV): contour map (top graph) and surface map (bottom graph) of standard deviations.

5.2.7 Disjunctive Kriging

The original data distribution and the transformed distribution are shown in Figure 5.60.

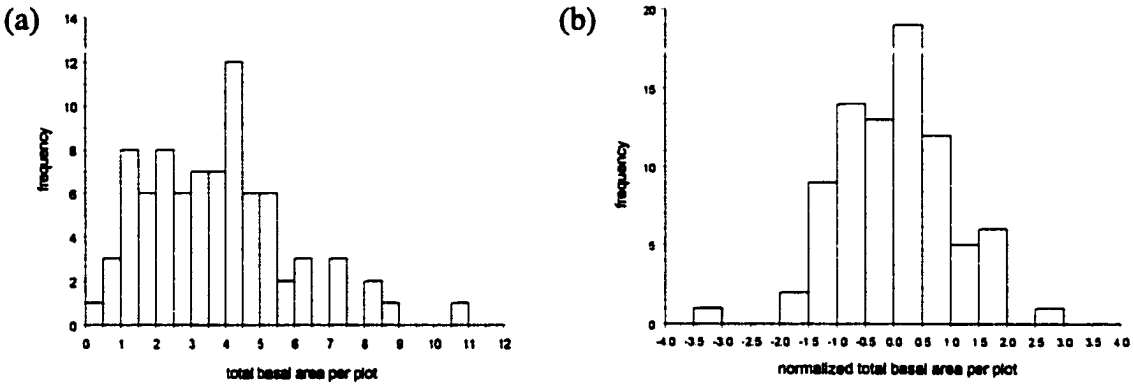

Figure 5.60 Original data distribution (a), and normalized data distribution (b) for total basal area per plot (TBA).

Nine Hermite polynomials and a seventh-order least-square polynomial were used to transform the original data. The mean and variance of the original data and the mean and variance of the transformed data were similar (Table 5.16).

Table 5.16 Means and variances of observed and transformed data for total basal area per plot (TBA).

	Mean	Variance
Observed data	3.70732	4.20141
Transformed data	3.67061	4.20391

Nine Hermite coefficients were calculated and are displayed in Table 5.17.

Table 5.17 The Hermite coefficients for total basal area per plot (TBA).

k	C_k
0	3.670614
1	2.004507
2	0.276611
3	-0.027365
4	-0.007030
5	-0.008796
6	-0.003738
7	0.000487
8	0.000405

The lag distance of the experimental variogram for the transformed data was 135 m and the lag tolerance was 67.5 m (Figure 5.61). The power variogram model had a power of 1 and a slope of 0.00001.

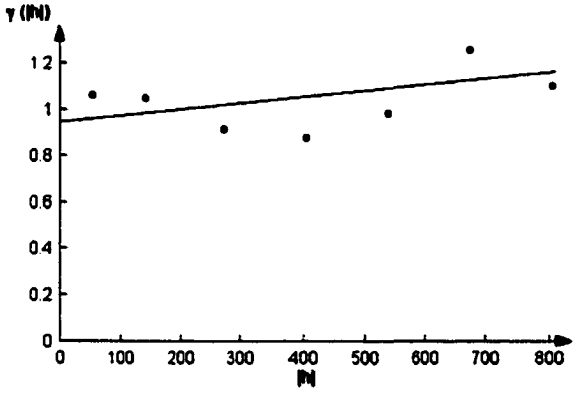

Figure 5.61 Experimental variogram and power variogram model for the normalized total basal area per plot (TBA).

The smallest MSE was 3.993 using 20 nearest neighbors in the estimation process. The analysis of the residuals did not exhibit any violations of the underlying assumptions (Figure 5.62). The correlation coefficient of the estimated data values versus the true data values was 0.23. The distribution of the residuals had a mean of - 0.07, and a variance of 4.04. The minimum value was -3.96, and the maximum value was 6.44.

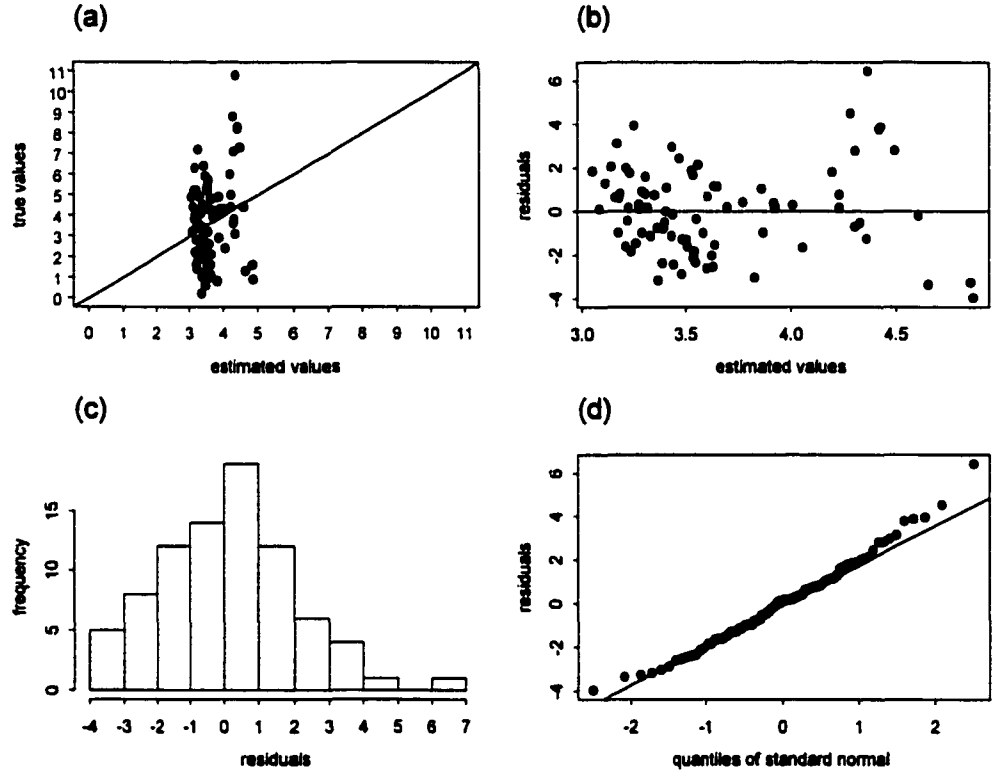

Figure 5.62 Disjunctive kriging with total basal area per plot (TBA):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates are displayed in Figure 5.63. The minimum value of the estimates was 2.83, and the maximum value was 4.73. A contour map and surface map of the standard deviations are shown in Figure 5.64. The minimum

value of the standard deviations was 1.94, and the maximum value was 1.99. It should be noted that the scale bar is different from the one used to display the standard deviations maps of the other kriging methods, due to the small range of standard deviations with disjunctive kriging.

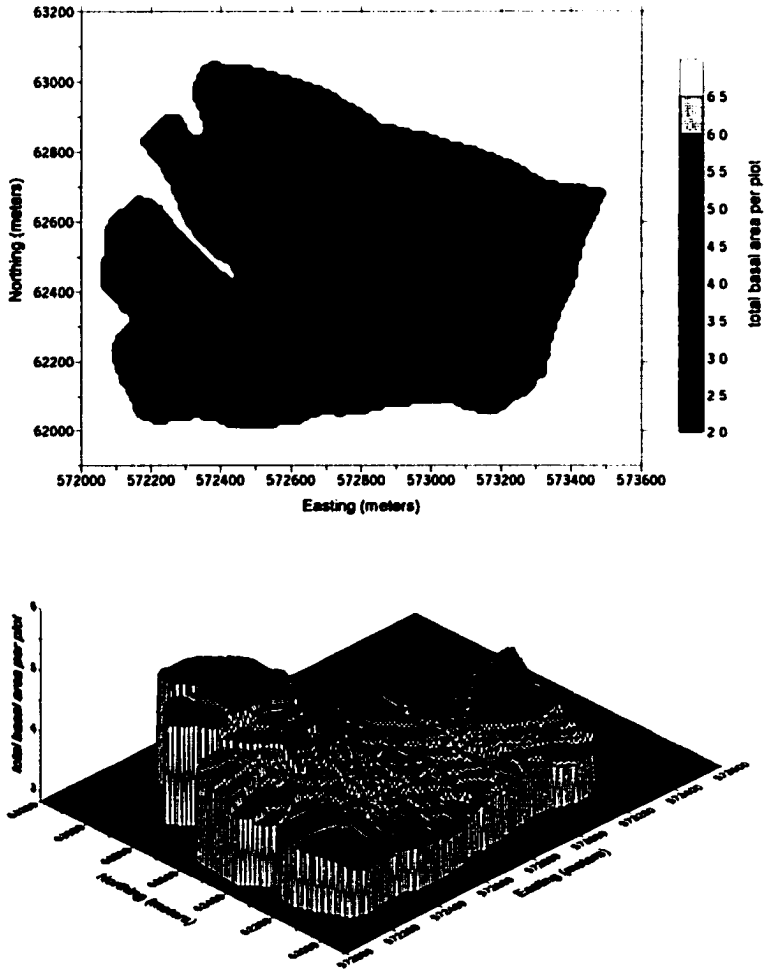

Figure 5.63 Disjunctive kriging with total basal area per plot (TBA): contour map (top graph) and surface map (bottom graph) of estimates.

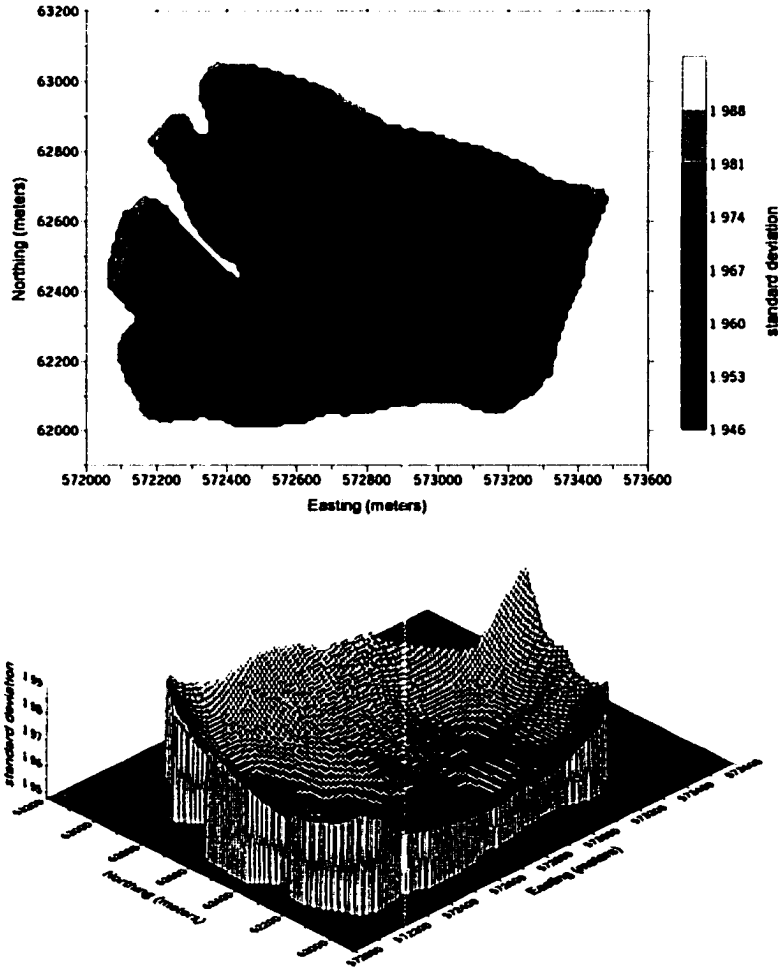

Figure 5.64 Disjunctive kriging with total basal area per plot (TBA): contour map (top graph) and surface map (bottom graph) of standard deviations.

As an example for calculating conditional probabilities with disjunctive kriging, a map was created showing the probabilities for each grid cell exceeding 5 feet^2 total basal area per plot (Figure 5.65).

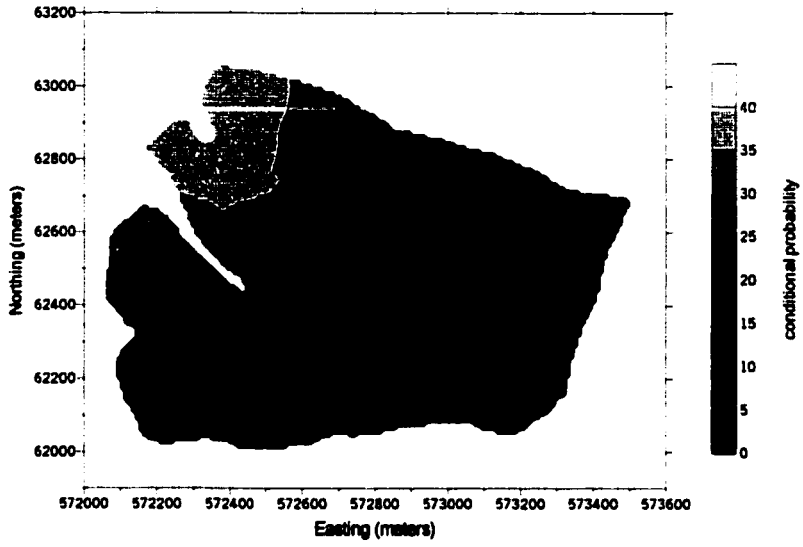

Figure 5.65 Contour map of the conditional probability that the total basal area per plot (TBA) is above a value of 5.

5.2.8 Discussion

Table 5.18 shows the mean square errors (MSE) of all interpolation methods for the total basal area per plot (TBA). Polygonal mapping was the best interpolation technique with a MSE of 3.464. The spatial distribution of variable TBA was better described with discrete polygons than with any continuous surface. Universal kriging with a second-degree trend surface reached the second best place with a MSE of 3.835. A second-degree trend surface was able to remove an underlying trend in the data.

Table 5.18 Mean square errors (MSE) of eight interpolation methods for total basal area per plot (TBA).

Interpolation Method	MSE
Polygonal mapping	3.464
Universal kriging (2nd degree trend surface)	3.835
Cokriging	3.910
Universal kriging (1st degree trend surface)	3.914
Ordinary kriging	3.951
Disjunctive kriging	3.993
Inverse distance weighting	4.122
Inverse distance weighting squared	4.598

Cokriging made good use of the spatial cross-correlation between the primary variable TBA and the auxiliary variable elevation (ELEV), and reached a good estimation result with a MSE of 3.910. Applying a first-degree trend surface with universal kriging was less effective in modeling the underlying trend in the data than the second-degree trend surface: a MSE of 3.914 could be reached. Disjunctive kriging yielded a MSE of 3.993. The transformation from the original data to the normalized data performed well: the mean and variance of the original and transformed data were similar, and the transformed data had a normal distribution. All kriging methods reached very similar MSEs. Both inverse distance methods yielded the worst results: inverse distance weighting had a MSE of 4.122, and inverse distance weighting squared a MSE of 4.598. The reason for the superior performance of the kriging methods over the inverse distance weighting techniques lies in the fact that the kriging methods take the clustering of the data points into account.

The spatial distribution of TBA was approximately the same for all interpolation techniques (with the exception of polygonal mapping): the total basal area per plot

changed gradually from low values in the east, southeast, and south to zones of high values in the northwest part of the study area. All factors that have an influence on the estimation results and the standard deviations of estimation are summarized in Subchapter 5.5, "Factors Affecting the Results of Interpolation Methods."

5.3 Number of Seedlings

5.3.1 Summary statistics

The following summary statistics for *number of seedlings per plot* (SEEDL) were calculated: minimum and maximum value, median, mean, standard deviation, variance, skewness, and kurtosis (Table 5.19). The distribution of SEEDL was highly skewed to the right as shown in Figure 5.66. There were 16 plots with no seedlings.

Table 5.19 Summary statistics for number of seedlings per plot (SEEDL).

Statistics	Values
Minimum	0.00
Median	4.00
Mean	9.13
Maximum	81.00
Standard Deviation	13.94
Variance	191.82
Skewness	2.84
Kurtosis	12.59

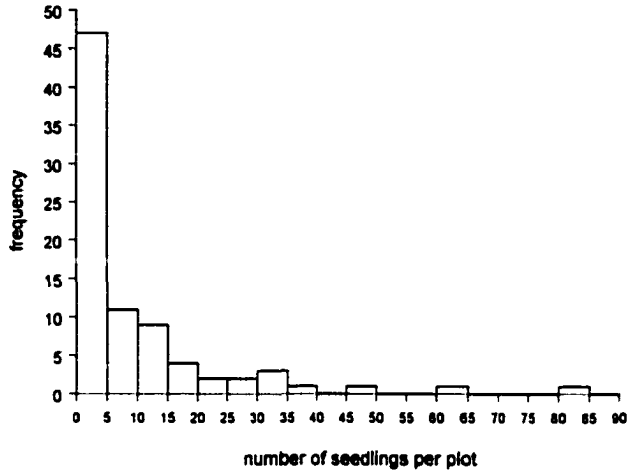

Figure 5.66 Histogram for number of seedlings per plot (SEEDL).

5.3.2 Polygonal Mapping

For each forest stand polygon the number of seedlings per plot (SEEDL) was calculated. Table 5.20 shows the number of sample plots, the average number of seedlings, and the standard deviation for each polygon.

Table 5.20 The number of sample plots, the mean, and the standard deviation of each forest stand polygon for number of seedlings per plot (SEEDL).

Stand number	Number of sample plots	Mean	Standard deviation
1	10	23.10	13.57
2	3	35.33	25.42
3	9	17.44	24.06
4	7	10.14	10.24
5	15	3.20	5.09
6	11	1.09	1.22
7	23	4.52	5.60
8	4	5.00	2.45

An ANOVA table (Table 5.21) was calculated to test the null hypothesis of all eight polygonal means being identical. The p-value was 0.0001. Therefore, the null hypothesis of all eight polygonal means being identical was rejected.

Table 5.21 ANOVA table for number of seedlings per plot (SEEDL).

Source	Sum of Squares	Degrees of Freedom	Mean Square	F Test	p-value
Between groups	6435.830	7	919.404	7.32	0.0001
Within groups	9293.694	74	125.590		
Totals	15729.524	81			

Table 5.22 shows the outcome from Tukey's HSD test. The ANOVA calculation and Tukey's HSD test are based on a normal distribution of the data. Variable SEEDL, however, is highly skewed to the right, and therefore, the results of both tests are misleading.

Table 5.22 Tukey's HSD test for number of seedlings per plot (SEEDL).

Tukey Grouping			Mean	Stand
	A		35.33	2
	A			
B	A		23.10	1
B	A			
B	A	C	17.44	3
B		C		
B		C	10.14	4
B		C		
B		C	5.00	8
B		C		
B		C	4.52	7
		C		
		C	3.20	5
		C		
		C	1.09	6

Figure 5.67 shows the validation results. The mean square error (MSE) was 113.338. The correlation coefficient between the estimated data values and the true data values was 0.64. The scatterplot of estimated data values versus residuals depicted that the errors were fairly evenly distributed (besides one possible outlier with an unusual high value of 63.56). A check on the outlier revealed that it originated from a sample plot with a large number of 81 seedlings per plot. Because this value was judged as correct, it was decided to leave this data value in this, and all subsequent analyses. The histogram and the q-q plot showed that the residuals were approximately normally distributed (again with the exception of one possible outlier). The frequency distribution of the residuals had a mean of -0.001, and a variance of 114.74. The minimum value was -22.33, and the maximum value was 63.56.

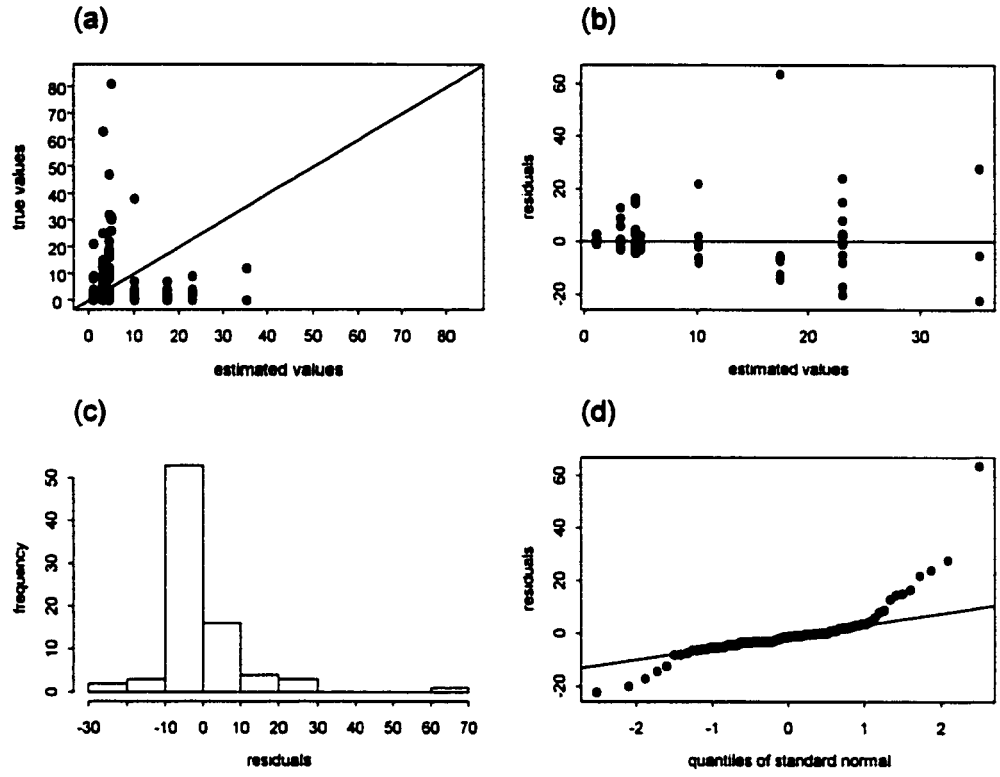

Figure 5.67 Polygonal mapping with number of seedlings per plot (SEEDL):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates are displayed in Figure 5.68. The minimum value of the estimates was 1.09, and the maximum value was 35.33. A contour map and surface map of the standard deviations are shown in Figure 5.69. The minimum value of the standard deviations was 1.22, the maximum value was 25.42.

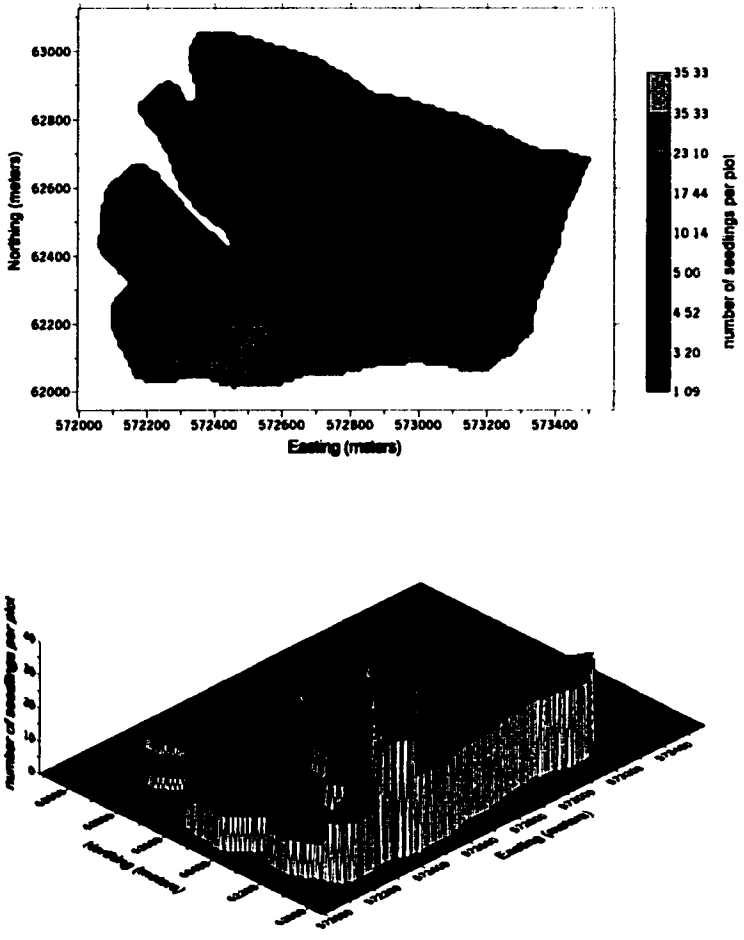

Figure 5.68 Polygonal mapping with number of seedlings per plot (SEEDL): polygonal map (top graph) and surface map (bottom graph) showing the mean per polygon.

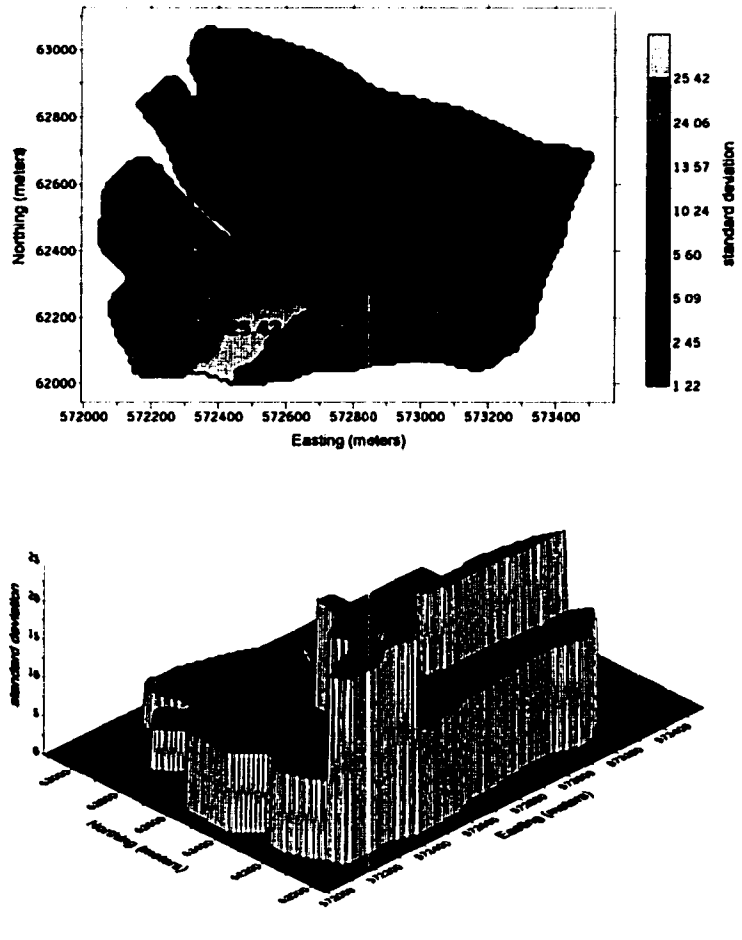

Figure 5.69 Polygonal mapping with number of seedlings per plot (SEEDL): polygonal map (top graph) and surface map (bottom graph) showing the standard deviation per polygon.

5.3.3 Inverse Distance Weighting

Inverse distance weighting and inverse distance weighting squared were used to obtain estimates for number of seedlings per plot (SEEDL) at unsampled locations.

Inverse Distance Weighting

Using the residuals from the cross-validation procedure a minimum MSE of 86.763 could be achieved by including the 9 nearest neighbors. The estimated data values versus true values had a correlation coefficient of 0.74. The distribution of the residuals had a mean of 0.19, a variance of 87.80, with a minimum value of -22.77, and a maximum value of 43.50. Without the two possible outliers of value 30.46 and 43.50, the residuals were approximately normally distributed, had a constant variance, and no apparent trend in the data (Figure 5.70).

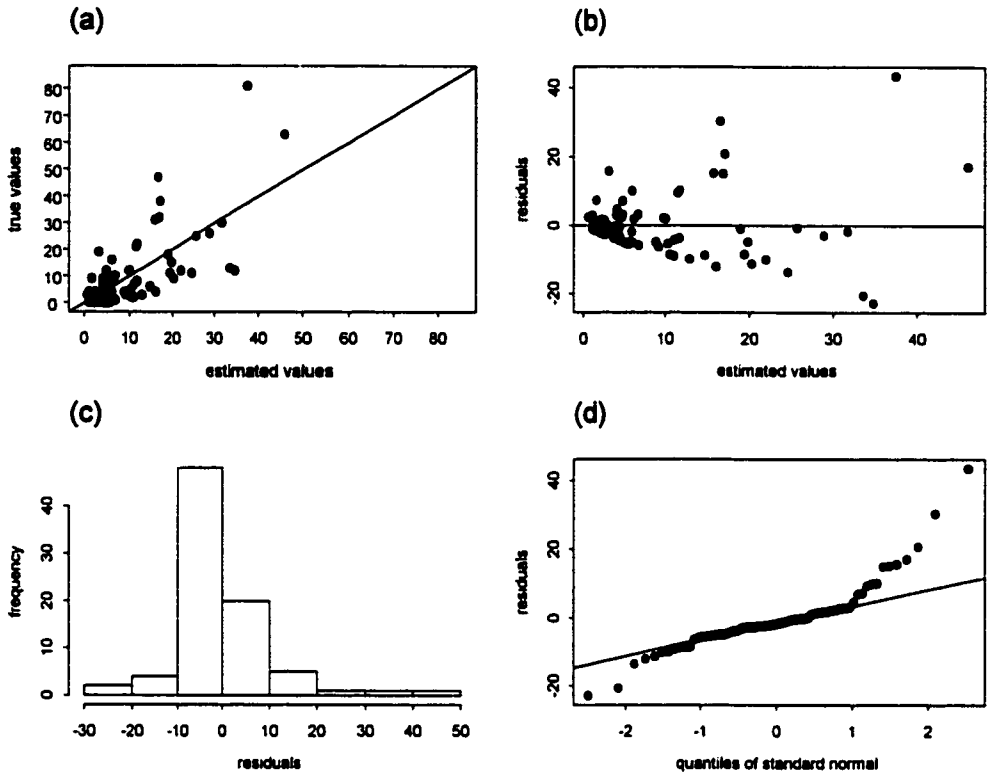

Figure 5.70 Inverse distance weighting with number of seedlings per plot (SEEDL):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates is displayed in Figure 5.71. The minimum value of the estimates was 0.14, and the maximum value was 64.89. A contour map and surface map of the standard deviations is shown in Figure 5.72. The minimum value of the standard deviations was 0.11, the maximum value was 14.76.

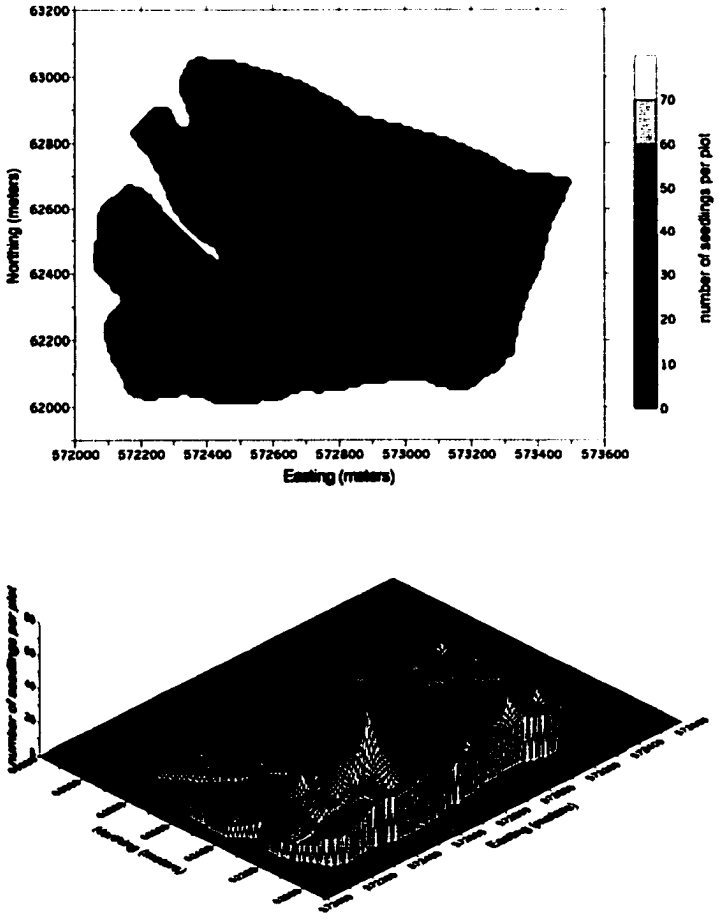

Figure 5.71 Inverse distance weighting with number of seedlings per plot (SEEDL): contour map (top graph) and surface map (bottom graph) of estimates.

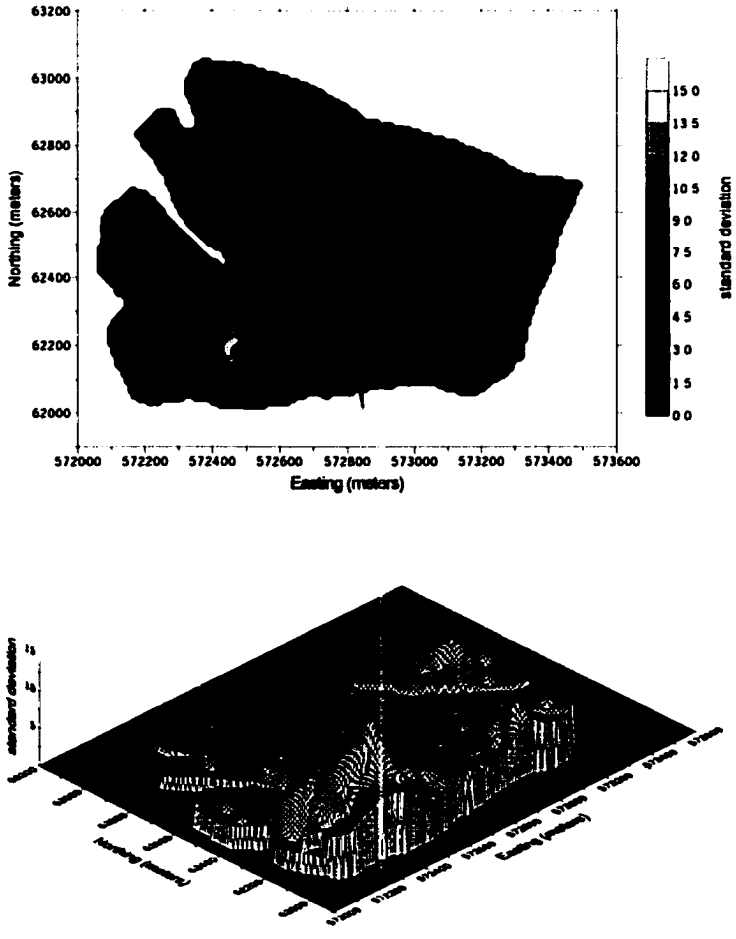

Figure 5.72 Inverse distance weighting with number of seedlings per plot (SEEDL): contour map (top graph) and surface map (bottom graph) of standard deviations.

Inverse Distance Weighting Squared

Using 9 nearest neighbors resulted in a minimum MSE of 69.881. The other four measures of validation are displayed in Figure 5.73. The correlation coefficient between estimated and true data values was 0.80. Without the four possible outliers of values -28.07, -23.31, 26.60, and 29.66, the residuals were approximately normally distributed, with a zero mean, a constant variance, and with no apparent trend in the data. The distribution of the residuals had a mean of 0.46, a variance of 70.53, with a minimum value of -28.07, and a maximum value of 29.66.

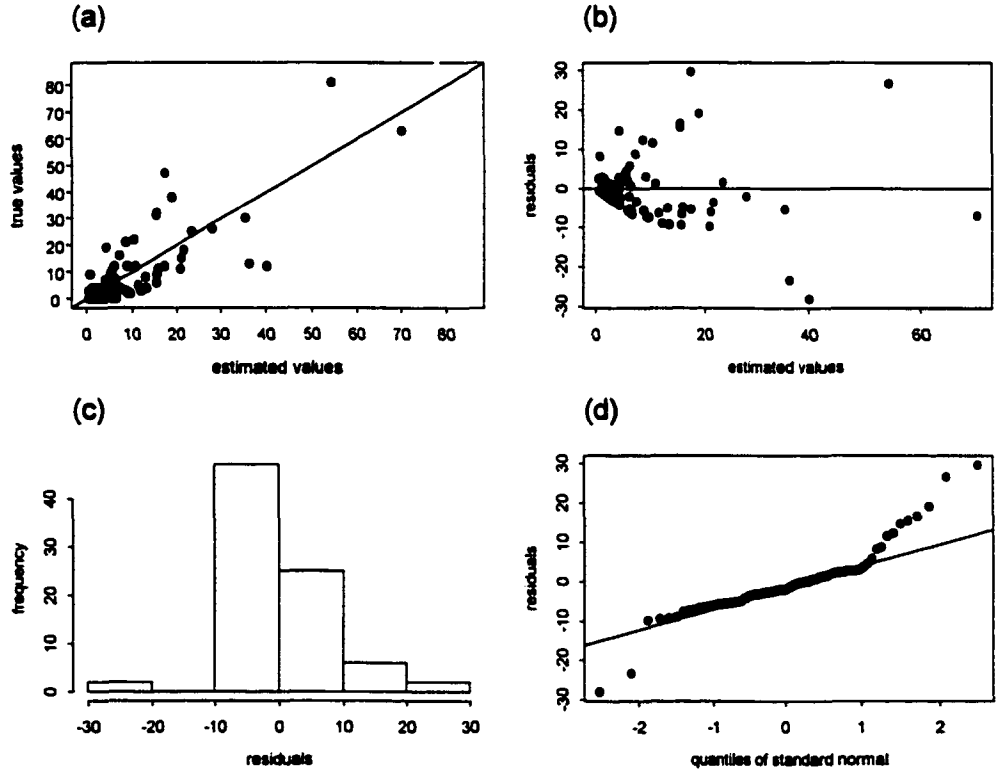

Figure 5.73 Inverse distance weighting squared with number of seedlings per plot (SEEDL):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates are displayed in Figure 5.74. The minimum value of the estimates was 0, and the maximum value was 78.64. A contour map and surface map of the standard deviations are shown in Figure 5.75. The minimum value of the standard deviations was 0, and the maximum value was 18.44.

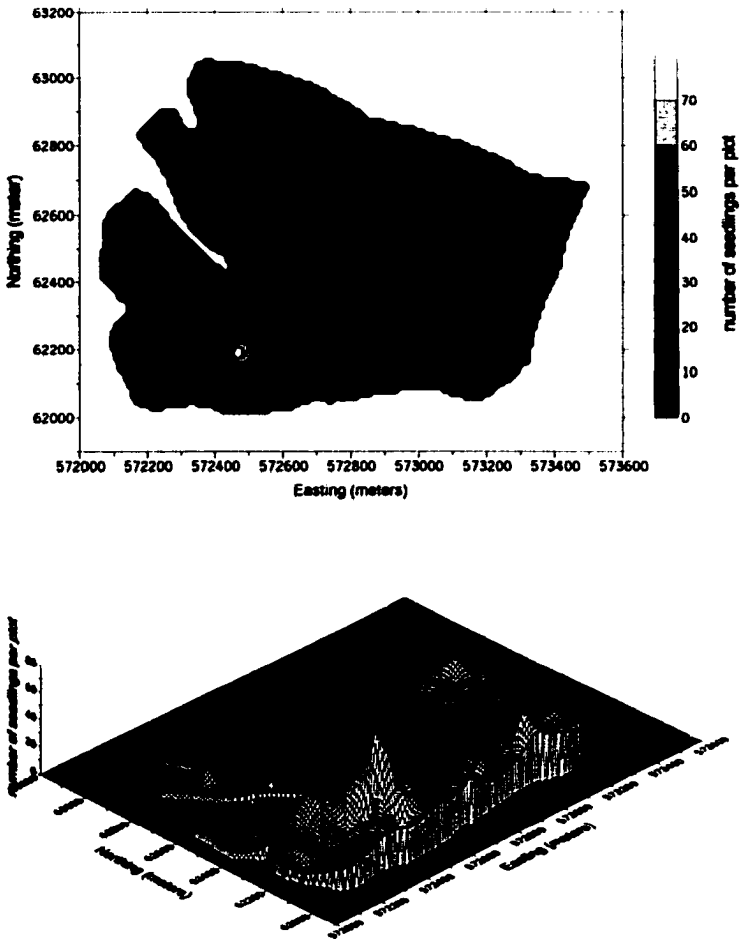

Figure 5.74 Inverse distance weighting squared with number of seedlings per plot (SEEDL): contour map (top graph) and surface map (bottom graph) of estimates.

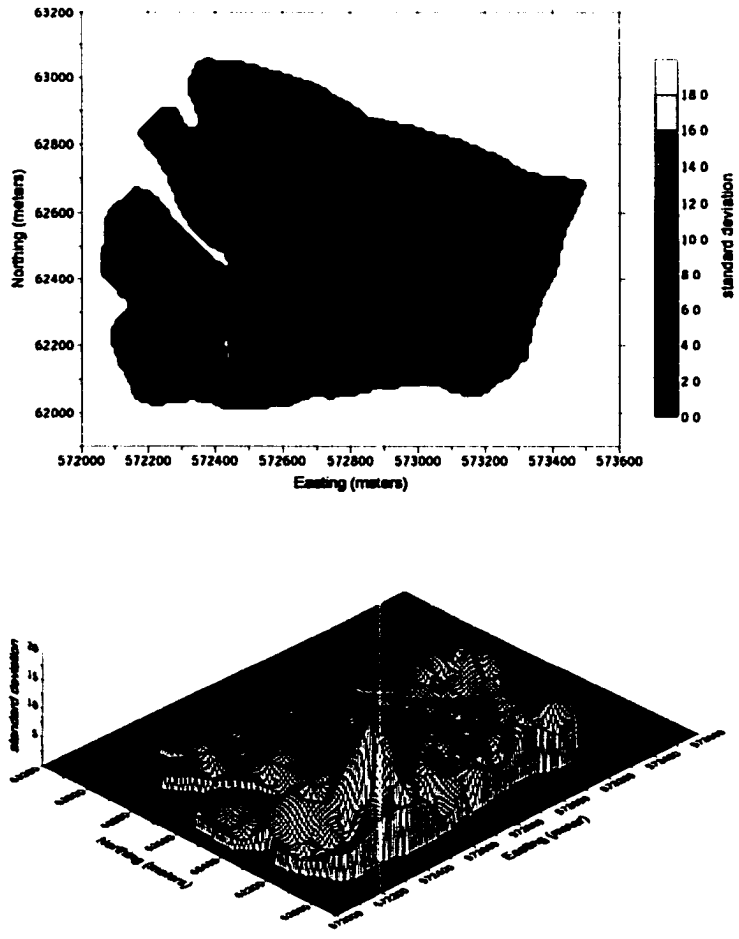

Figure 5.75 Inverse distance weighting squared with number of seedlings per plot (SEEDL): contour map (top graph) and surface map (bottom graph) of standard deviations.

5.3.4 Ordinary Kriging

An experimental variogram was calculated to examine the spatial correlation of number of seedlings per plot (SEEDL). A Gaussian model was fitted to the experimental variogram which had a lag distance of 150 m and a lag tolerance of 75 m (Figure 5.76). The model was specified with a nugget of 96, a sill of 227, and a range of 515.

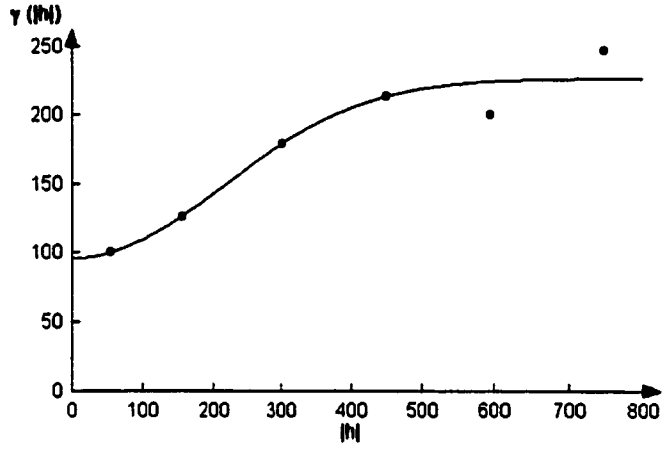

Figure 5.76 Experimental variogram and Gaussian variogram model for number of seedlings per plot (SEEDL).

Using 3 nearest neighbors resulted in the smallest MSE of 95.903. Without the two possible outliers of value -37.20 and 45.62, Figure 5.77 showed that the residuals had an approximately normal distribution, a constant variance, and no apparent trend in the data. The correlation coefficient between estimated and true data values was 0.71. The distribution of the residuals had a mean of -0.47, a variance of 96.86, with a minimum value of -37.20, and a maximum value of 45.62.

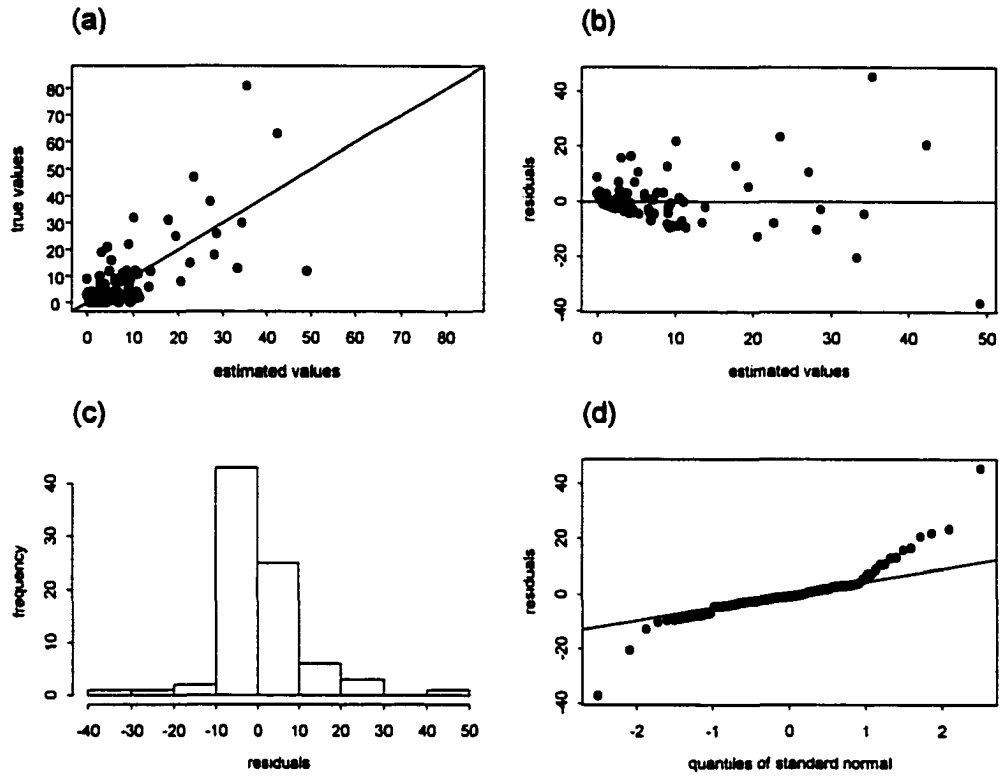

Figure 5.77 Ordinary kriging with number of seedlings per plot (SEEDL):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates are displayed in Figure 5.78. Negative kriging estimates were set to zero. The minimum value of the estimates was 0, and the maximum value was 77.66. A contour map and surface map of the standard errors are shown in Figure 5.79. The minimum value of the standard errors was 1.29, the maximum value was 14.30.

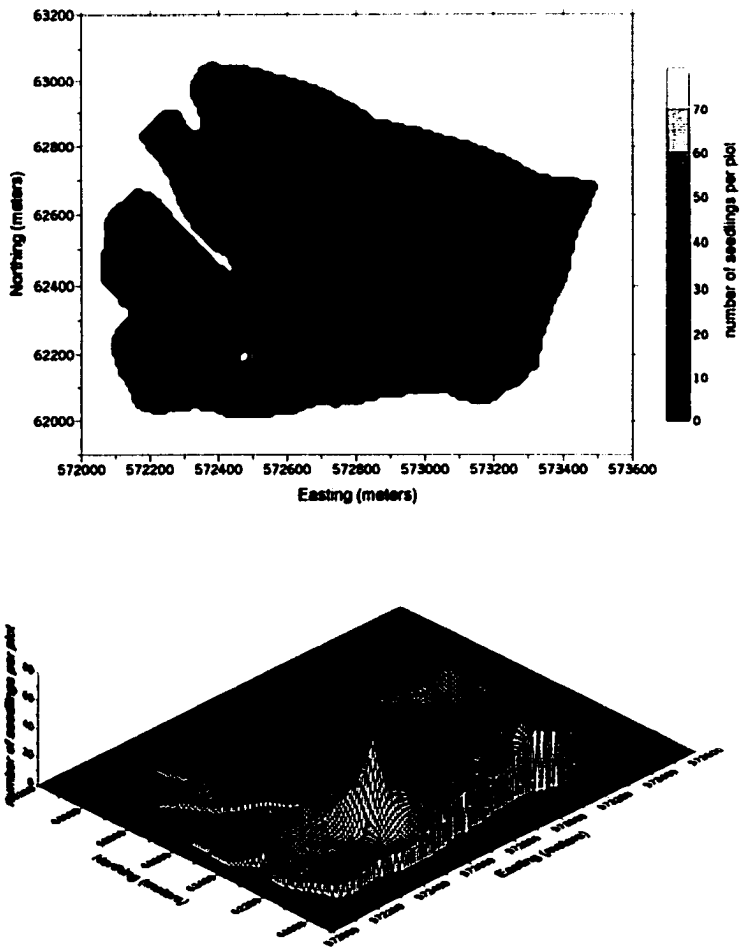

Figure 5.78 Ordinary kriging with number of seedlings per plot (SEEDL): contour map (top graph) and surface map (bottom graph) of estimates.

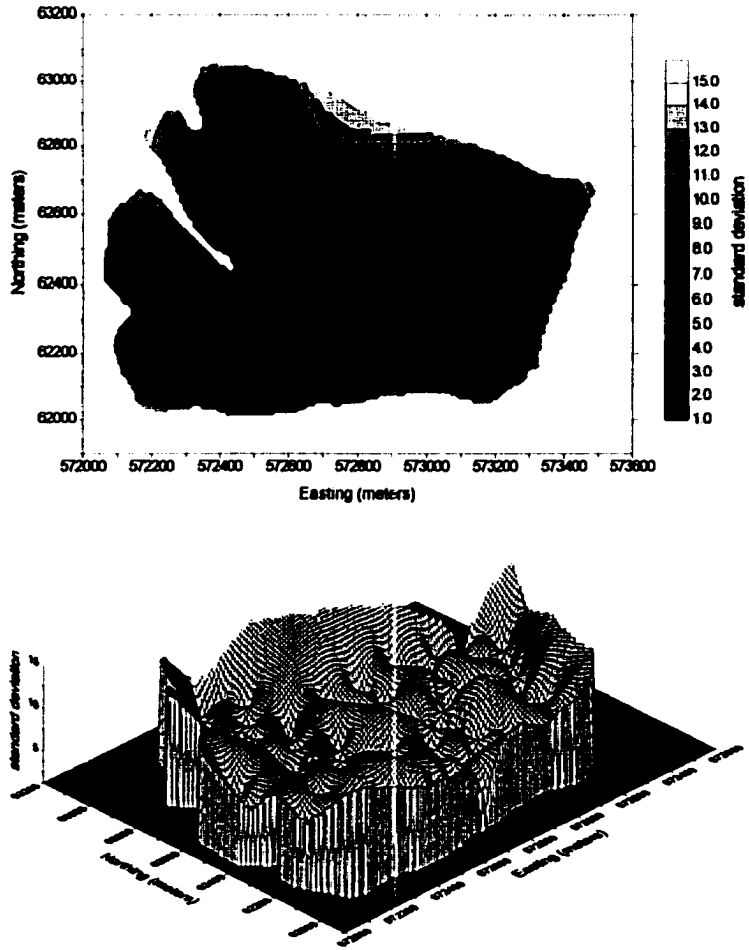

Figure 5.79 Ordinary kriging with number of seedlings per plot (SEEDL): contour map (top graph) and surface map (bottom graph) of standard deviations.

5.3.5 Universal Kriging

A first-and second-degree trend surface were used to estimate an underlying trend in the number of seedlings per plot (SEEDL).

First-Degree Trend Surface

A first-degree trend surface was fitted to the number of seedlings per plot (SEEDL) data set. An R^2 value of 0.28 was achieved, i.e. the variables x-coordinate and y-coordinate accounted for 28% of the variability in SEEDL. A lag distance of 150 m and a lag tolerance of 75 m were chosen for the experimental variogram of the residuals (Figure 5.80). The Gaussian variogram model had the following specifications: nugget of 98, sill of 158.5, and range of 400.

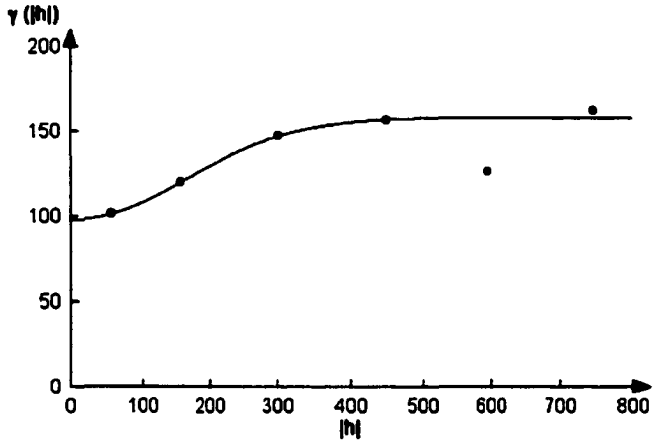

Figure 5.80 Experimental variogram and Gaussian variogram model for the residuals of the first-degree trend surface for number of seedlings per plot (SEEDL).

The smallest MSE of value 99.556 was obtained using the 3 nearest neighbors. The correlation coefficient between estimated and true data values was 0.70. Besides the

two possible outliers with a value of -38.38 and 46.56, the analysis of the residuals did not exhibit any peculiarities that would suggest violations of the underlying assumptions (Figure 5.81). The distribution of the residuals had a mean of -0.39, a variance of 100.63, with a minimum value of -38.38, and a maximum value of 46.56.

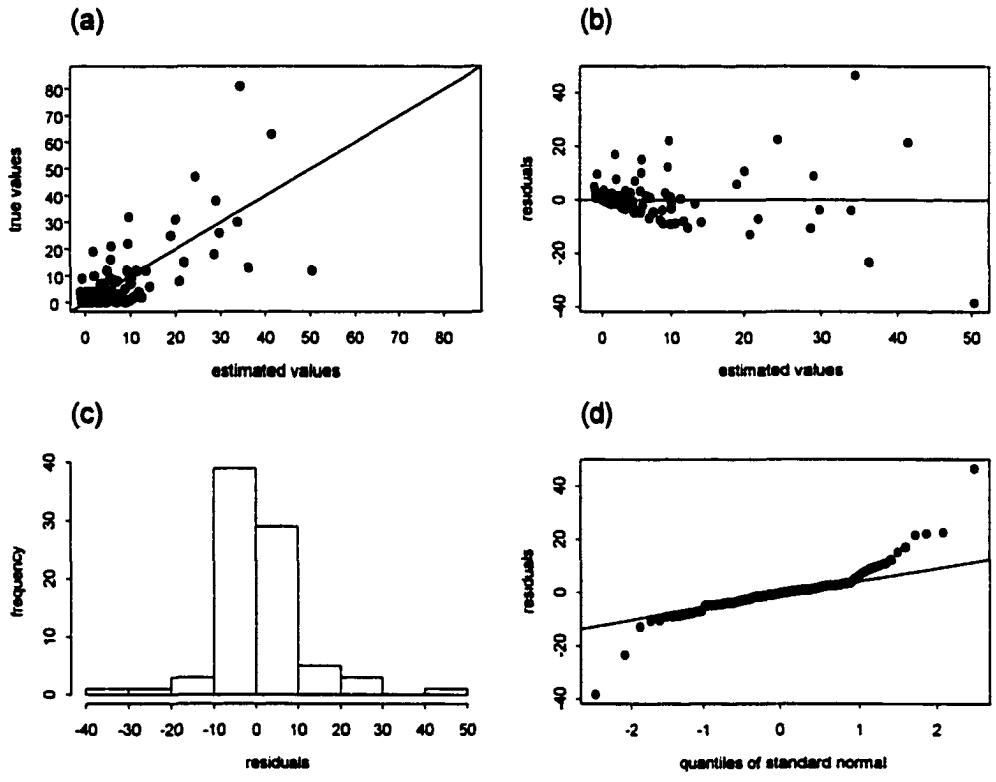

Figure 5.81 Universal kriging with first-degree trend surface for number of seedlings per plot (SEEDL):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

Ordinary kriging was conducted on the residuals of the first-degree trend surface. The final surface was created by combining the first-degree trend surface with the kriged surface of the residuals (Figure 5.82). Negative estimates were set to zero. The minimum value of the estimates was 0, and the maximum value was 76.37. A contour map and

surface map of the standard deviations are shown in Figure 5.83. The minimum value of the standard deviations was 1.42, and the maximum value was 12.71.

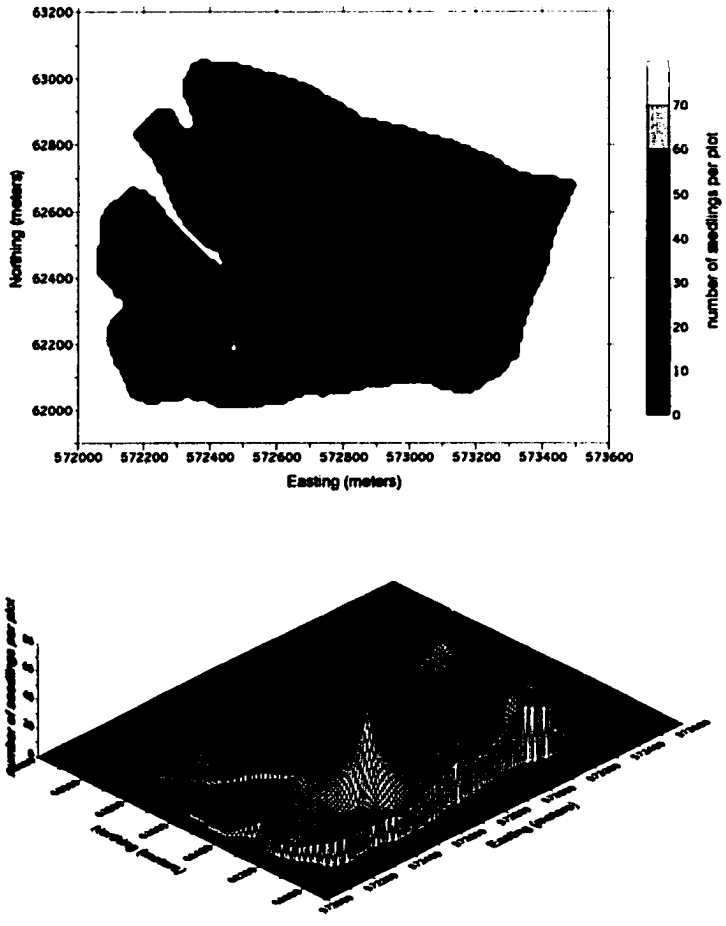

Figure 5.82 Universal kriging with first-degree trend surface for number of seedlings per plot (SEEDL): contour map (top graph) and surface map (bottom graph) of estimates.

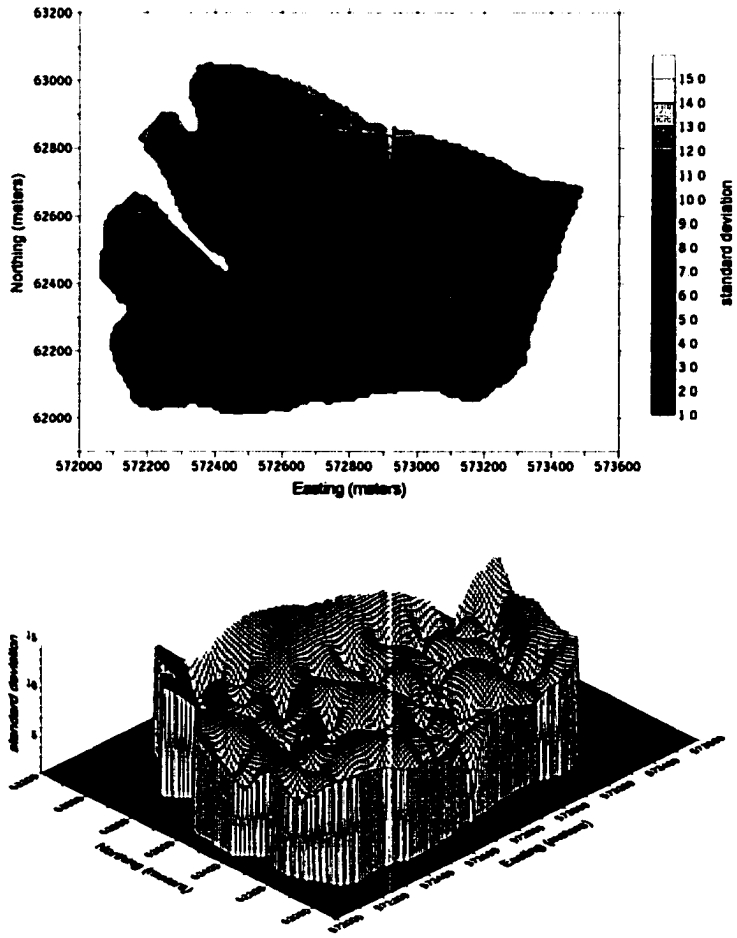

Figure 5.83 Universal kriging with first-degree trend surface for number of seedlings per plot (SEEDL): contour map (top graph) and surface map (bottom graph) of standard deviations.

Second-Degree Trend Surface

A second-degree trend surface was fitted to the data set. An R^2 value of 0.36 was achieved. A lag distance of 150 m and a lag tolerance of 75 m were chosen for the experimental variogram of the residuals (Figure 5.84). The exponential variogram model had a nugget of 91.5, a sill of 141.5 and a range of 700. The large nugget is an indication of the data values approaching a random surface.

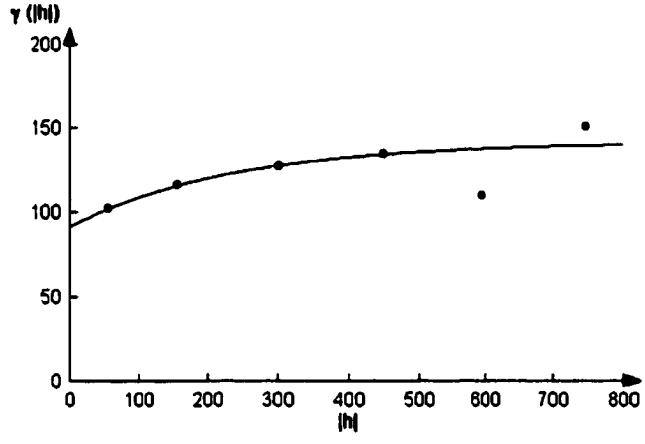

Figure 5.84 Experimental variogram and exponential variogram model for the residuals of the second-degree trend surface for number of seedlings per plot (SEEDL).

Including the 3 nearest neighbors in the kriging process resulted in the smallest MSE of 100.703. The correlation coefficient between estimated and true data values was 0.70. Besides the two possible outliers with a value of -38.96, and 46.47 the analysis of the residuals did not exhibit any peculiarities that would suggest violations of the underlying assumptions (Figure 5.85). The distribution of the residuals had a mean of -0.25, a variance of 101.88, with a minimum value of -38.96, and a maximum value of 46.47.

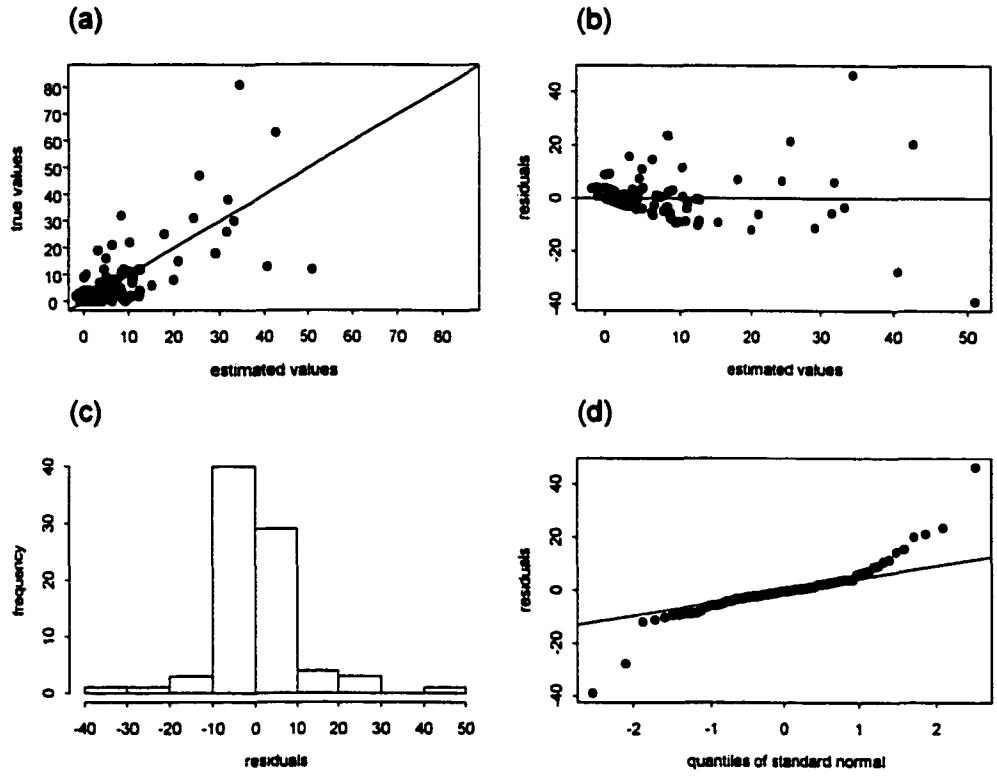

Figure 5.85 Universal kriging with second-degree trend surface for number of seedlings per plot (SEEDL):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

Ordinary kriging was conducted on the residuals of the second-degree trend surface. The final surface was created by combining the second-degree trend surface with the kriged surface of the residuals (Figure 5.86). The minimum value of the estimates was 0 (negative estimates were set to zero), and the maximum value was 75.74. A contour map and surface map of the standard deviations are shown in Figure 5.87. The minimum value of the standard deviations was 1.53, and the maximum value was 11.91.

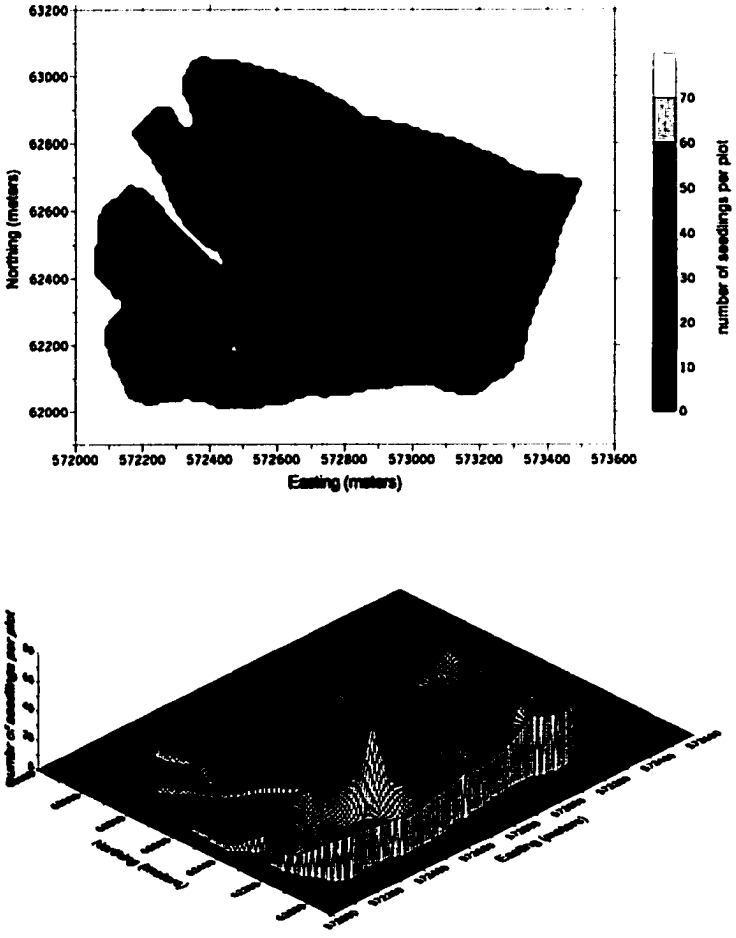

Figure 5.86 Universal kriging with second-degree trend surface for number of seedlings per plot (SEEDL): contour map (top graph) and surface map (bottom graph) of estimates.

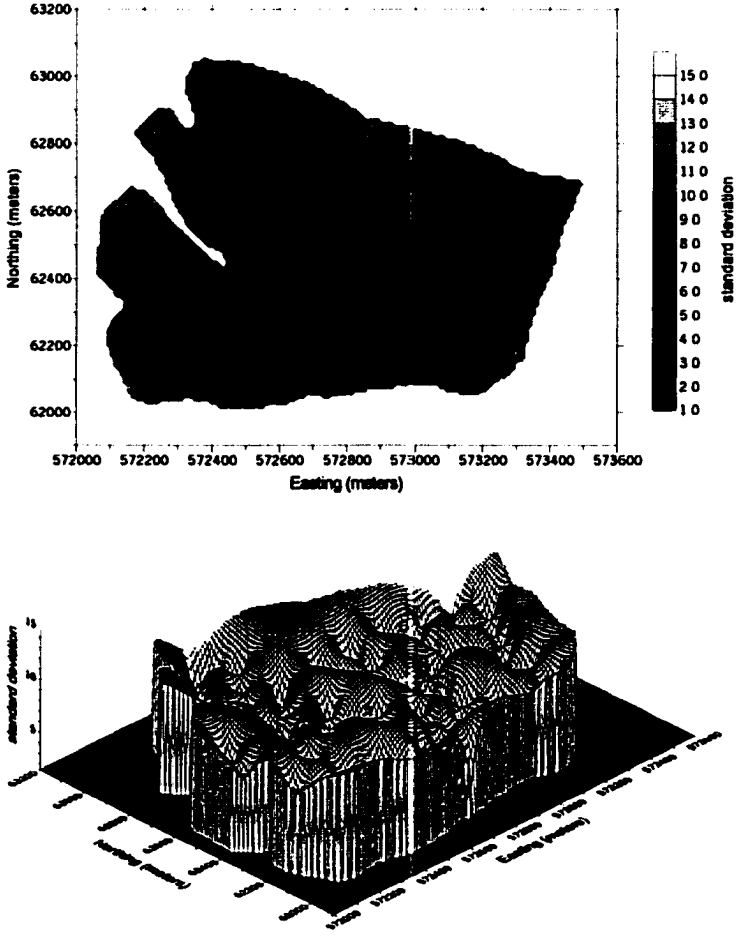

Figure 5.87 Universal kriging with second-degree trend surface for number of seedlings per plot (SEEDL): contour map (top graph) and surface map (bottom graph) of standard deviations.

5.3.6 Cokriging

The primary variable number of seedlings per plot (SEEDL) was combined with auxiliary variables. From all possible combinations of secondary variables, using elevation (ELEV) as auxiliary variable yielded the smallest MSE of 84.646 (Table 5.23).

Table 5.23 Mean square errors (MSE) for primary variable number of seedlings per plot (SEEDL) using different combinations of secondary variables: elevation (ELEV), slope/aspect (SLOASP), and normalized difference vegetation index (NDVI).

Secondary Variables	MSE
ELEV	84.646
SLOASP	98.343
NDVI	91.984
ELEV and SLOASP	87.644
ELEV and NDVI	89.034
SLOASP and NDVI	90.848
ELEV, SLOASP, and NDVI	90.827

A lag distance of 150 m and a lag tolerance of 75 m was applied for the experimental variograms of SEEDL and ELEV, and the cross-variogram SEEDL-ELEV (Figure 5.88). The nested variogram models consisted of a combination of an exponential and a Gaussian variogram model. Table 5.24 shows the specifications for both variogram models and the cross-variogram model. SEEDL and ELEV were negatively spatially cross-correlated as can be seen from the cross-variogram.

Table 5.24 Variogram and cross-variogram model specifications for primary variable number of seedlings per plot (SEEDL) and secondary variable elevation (ELEV).

Variogram / Cross-variogram	Exponential			Gaussian	
	nugget	sill	range	sill	range
SEEDL	67	73	400	100	700
ELEV	10	13	400	7800	700
SEEDL-ELEV	25	30	400	-260	700

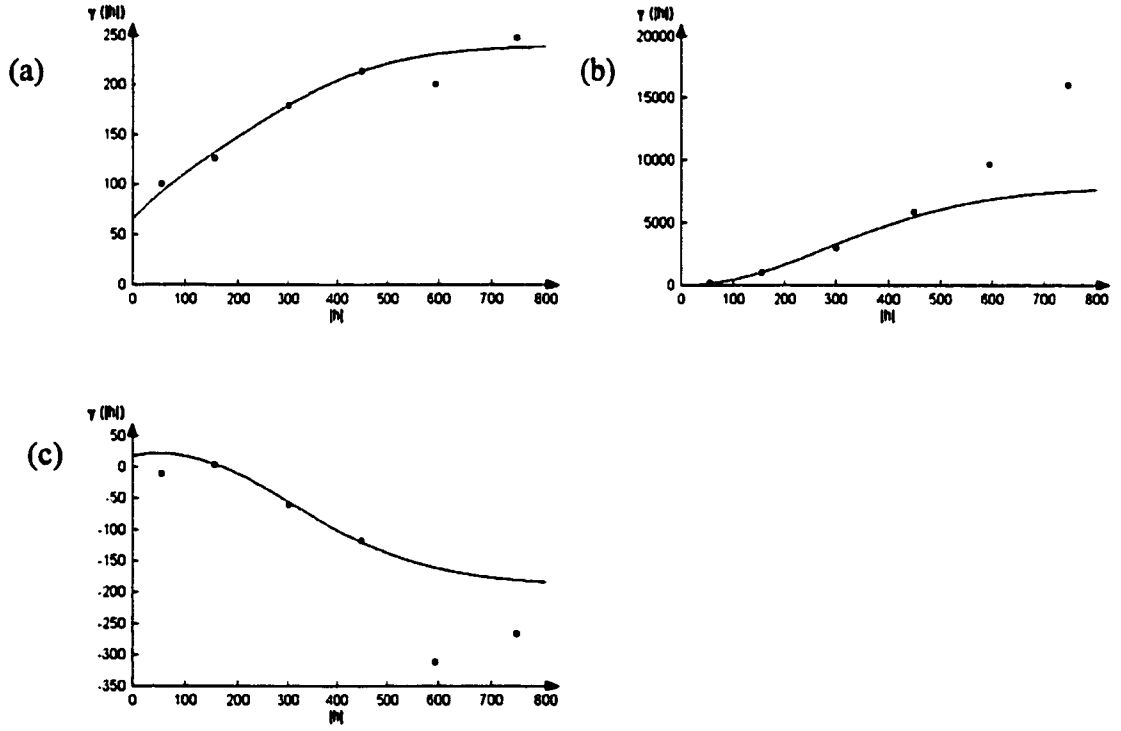

Figure 5.88 (a) Experimental variograms for primary variable number of seedlings per plot (SEEDL) and (b) secondary variable elevation (ELEV), as well as (c) experimental cross-variogram with an exponential and a Gaussian (cross-) variogram model.

The smallest MSE could be achieved by using the 3 nearest neighbors: 84.646. The correlation coefficient between estimated and true data values was 0.75. Without the two possible outliers of value -34.94 and 41.22 the residuals were approximately

normally distributed, had a constant variance, and had no apparent trend in the data (Figure 5.89). The distribution of the residuals had a mean of -0.40, a variance of 85.53, with a minimum value of -34.94, and a maximum value of 41.22.

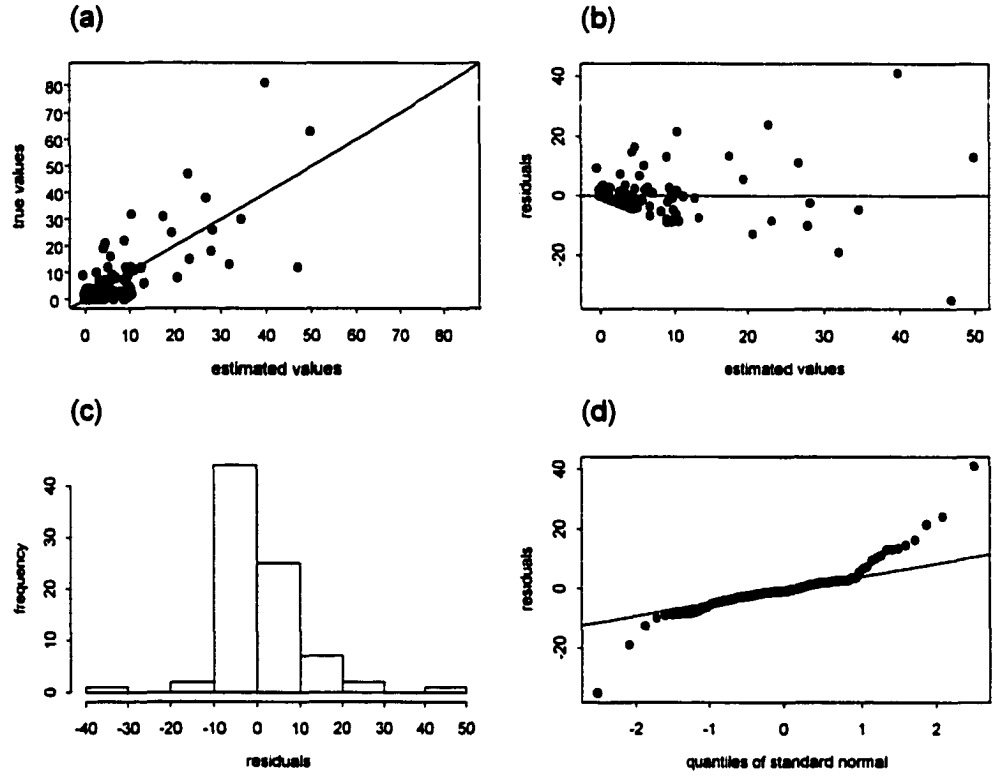

Figure 5.89 Cokriging with primary variable number of seedlings per plot (SEEDL) and secondary variable elevation (ELEV):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates are displayed in Figure 5.90. The minimum value of the estimates was 0 (negative estimates were set to zero), and the maximum value was 61.63. A contour map and surface map of the standard errors are shown in Figure 5.91. The minimum value of the standard errors was 9.91, and the maximum value was 14.93.

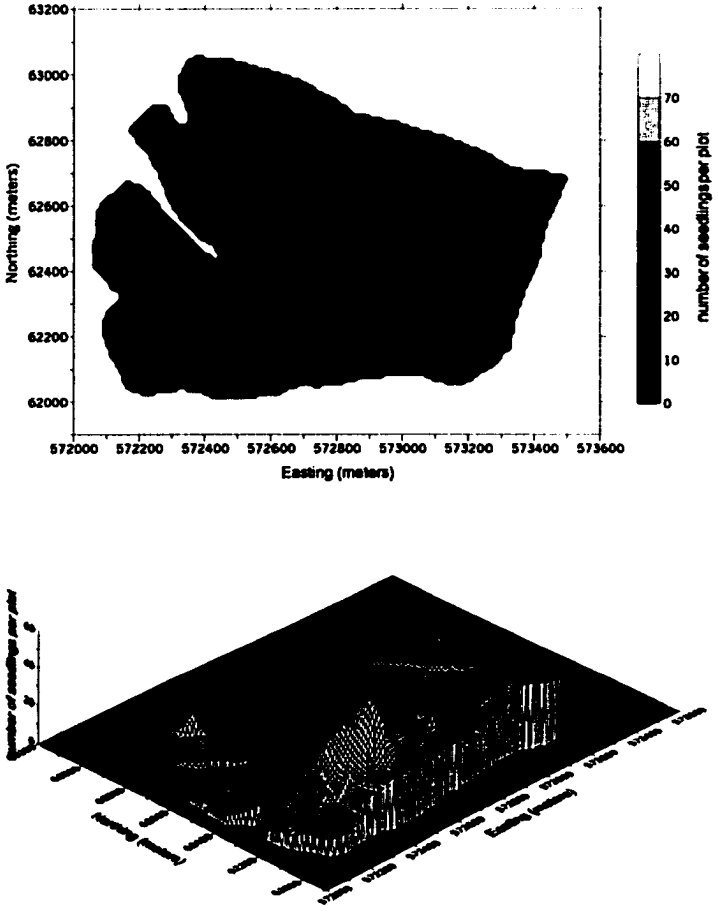

Figure 5.90 Cokriging with primary variable number of seedlings per plot (SEEDL) and secondary variable elevation (ELEV): contour map (top graph) and surface map (bottom graph) of estimates.

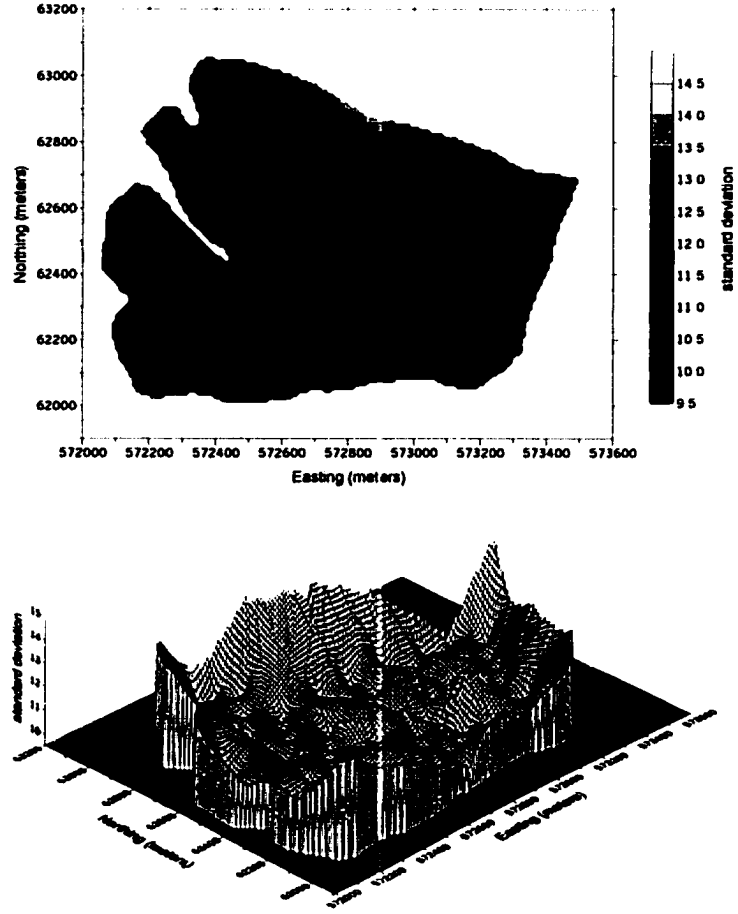

Figure 5.91 Cokriging with primary variable number of seedlings per plot (SEEDL) and secondary variable elevation (ELEV): contour map (top graph) and surface map (bottom graph) of standard deviations.

5.3.7 Disjunctive Kriging

The disjunctive kriging technique transformed the original data distribution to a normalized distribution (Figure 5.92).

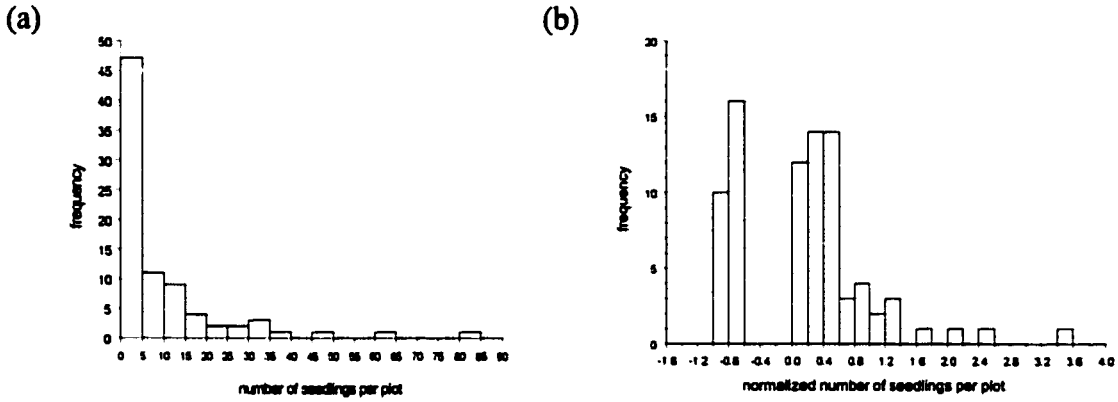

Figure 5.92 Original data distribution (a), and normalized data distribution (b) for number of seedlings per plot (SEEDL).

Ten Hermite polynomials and a third-order least-square polynomial were used to transform the original data. The quality of the transformation process was checked with a comparison between the mean and variance of the original data with the transformed data (Table 5.25). While the means of the original and the transformed data were quite similar, both variances were very different.

Table 5.25 Means and variances of observed and transformed data for number of seedlings per plot (SEEDL).

	Mean	Variance
Observed data	9.13415	191.82350
Transformed data	9.89167	146.88790

Ten Hermite coefficients were calculated and are shown in Table 5.26.

Table 5.26 The Hermite coefficients for number of seedlings per plot (SEEDL).

k	C_k
0	9.891669
1	8.161440
2	5.423038
3	-0.683610
4	-0.762005
5	0.047767
6	0.066391
7	-0.003859
8	-0.005439
9	0.000119

The transformed data values were used to calculate an experimental variogram. The lag distance was 150 m and the lag tolerance was 75 m. The Gaussian variogram model had a nugget of 0.325, a sill of 0.76, and a range of 470 (Figure 5.93).

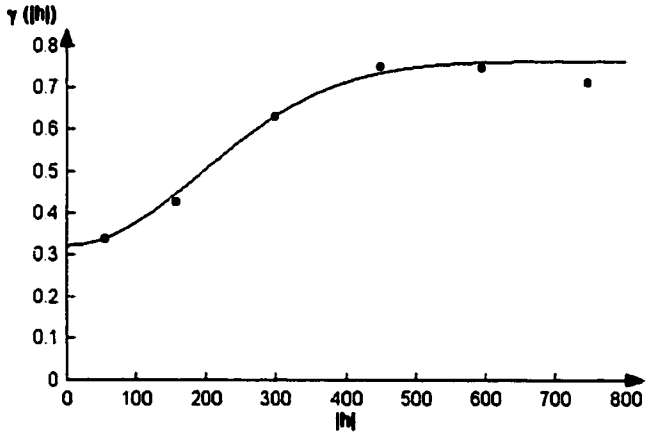

Figure 5.93 Experimental variogram and Gaussian variogram model for the normalized number of seedlings per plot (SEEDL).

Including 10 nearest neighbors in the estimation process yielded a MSE of 118.975. Without the three possible outliers of value 30.98, 40.00, and 59.44, the analysis of the residuals did not exhibit any violations of the underlying assumptions (Figure 5.94). The distribution of the residuals had a mean of 0.68, and a variance of 119.98. The minimum value was -11.45, and the maximum value was 59.44. The correlation coefficient of the estimated data values versus the true data values was 0.68.

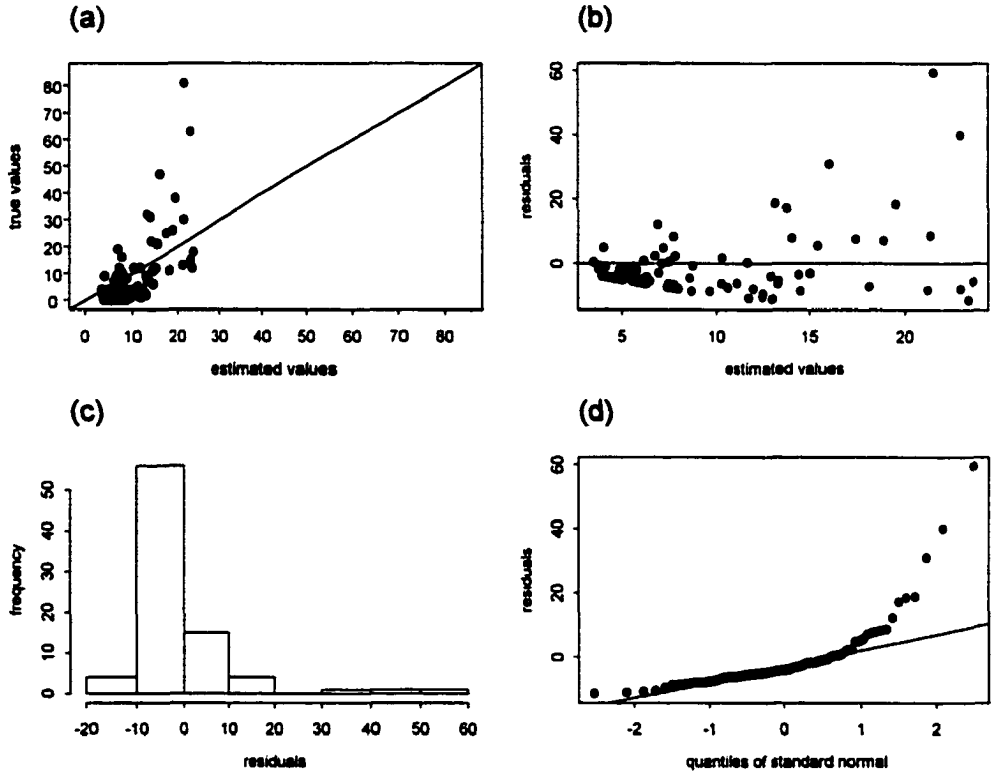

Figure 5.94 Disjunctive kriging with number of seedlings per plot (SEEDL):
(a) scatterplot of estimated data values versus true data values,
(b) scatterplot of estimated data values versus residuals,
(c) histogram of residuals, and
(d) q-q plot of residuals.

A contour map and surface map of the estimates are displayed in Figure 5.95. The minimum value of the estimates was 2.96, and the maximum value was 31.94. A contour

map and surface map of the standard deviations are shown in Figure 5.96. The minimum value of the standard deviations was 10.27, and the maximum value was 11.75.

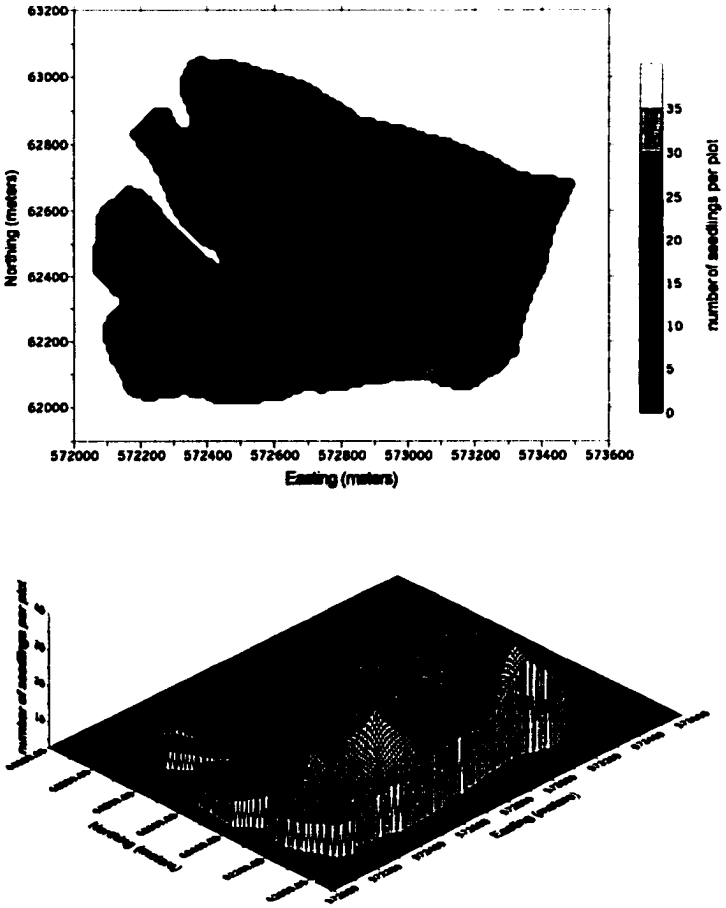

Figure 5.95 Disjunctive kriging with number of seedlings per plot (SEEDL): contour map (top graph) and surface map (bottom graph) of estimates.

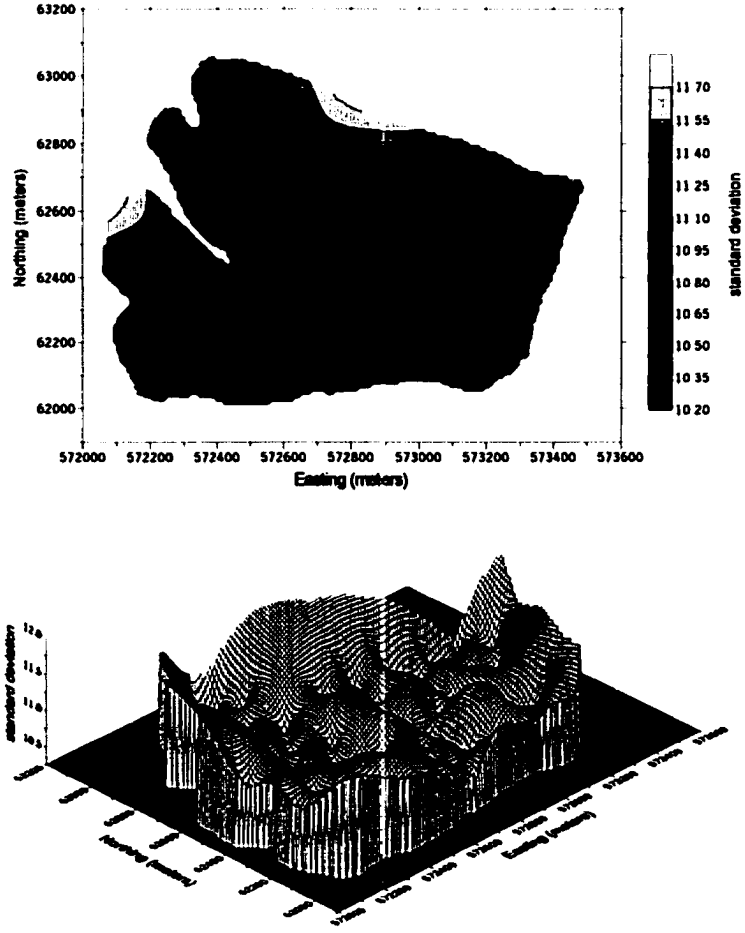

Figure 5.96 Disjunctive kriging with number of seedlings per plot (SEEDL): contour map (top graph) and surface map (bottom graph) of standard deviations.

As an example for calculating conditional probabilities with disjunctive kriging, a map was created showing the probabilities for each grid cell exceeding 10 seedlings per plot (Figure 5.97).

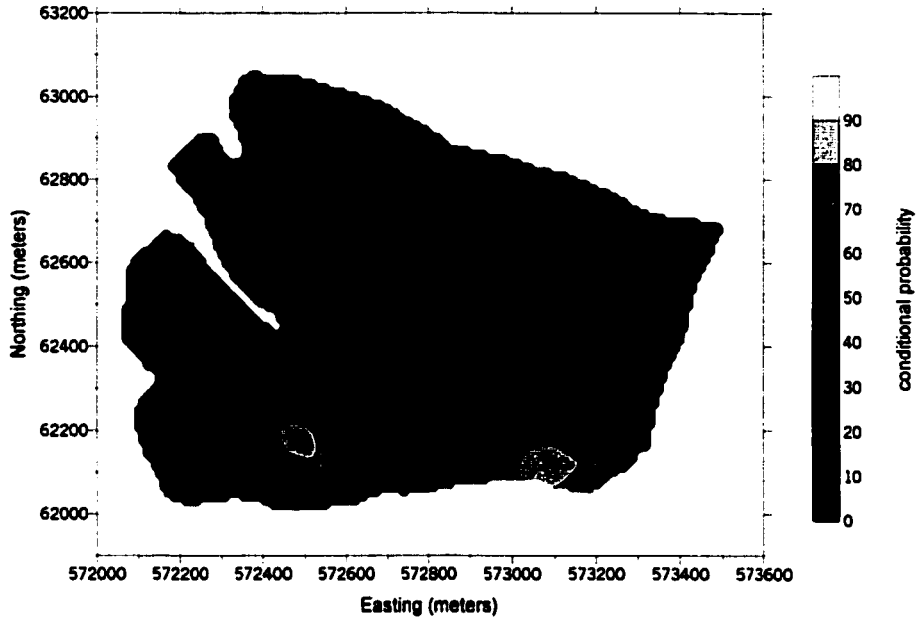

Figure 5.97 Contour map of the conditional probability that the number of seedlings per plot (SEEDL) is above a value of 10.

5.3.8 Discussion

The MSEs of all eight interpolation methods for the number of seedlings per plot (SEEDL) are summarized in Table 5.27. The differences in the MSEs are much larger than for the interpolation results for variables number of stems per plot (STM) and total basal area per plot (TBA).

Table 5.27 Mean square errors (MSE) of eight interpolation methods for the number of seedlings per plot (SEEDL).

Interpolation Method	MSE
Inverse distance weighting squared	69.881
Cokriging	84.646
Inverse distance weighting	86.763
Ordinary kriging	95.903
Universal kriging (1^{st} degree trend surface)	99.556
Universal kriging (2^{nd} degree trend surface)	100.703
Polygonal mapping	113.338
Disjunctive kriging	118.975

The best result could be achieved with inverse distance weighting squared having a MSE of 69.801, while inverse distance weighting was the third best method with a MSE of 86.763. All kriging methods require the sample data to have a normal distribution in order to be an unbiased and optimal interpolation method. Despite the fact that kriging has often been found to be insensitive to lack of normality (Hohn, 1998), the distribution of variable SEEDL, however, is far from a normal distribution (Figure 5.66). This lack of normality might be the reason for the bad results with ordinary kriging: the MSE was 95.903. But why is cokriging the interpolation method with the second best outcome (MSE of 84.646)? This kriging method also suffers from non-normally distributed data. Goovaerts (1997) and Asli and Marcotte (1995) might have the explanation: both report that interpolation results from cokriging can be improved if the primary and secondary variable have a large correlation coefficient. Asli and Marcotte (ibid.) write that a high correlation (correlation coefficients of 0.4 and above) between the principal variable and the secondary variable is important to get the maximum benefit from the information available on the secondary variable. The correlation coefficient

between primary variable SEEDL and auxiliary variable elevation (ELEV) was 0.57. This relatively strong correlation between these two variables might have helped to improve the estimation results of cokriging over ordinary kriging. Both universal kriging methods yielded results worse than ordinary kriging and cokriging: universal kriging with a first-degree trend surface had a MSE of 99.556, whereas universal kriging with a second-degree trend surface had a MSE of 100.703. Zimmerman et al. (1999) point out that for some types of surfaces it is better to completely ignore the modeling of trend (as ordinary kriging does) then to model the trend inappropriately. This might be one of the reasons for the better performance of ordinary kriging. The results for polygonal mapping with a high MSE of 113.338 indicated that variable SEEDL was better described as a continuous surface than with discrete polygons. However, a different set of polygons (polygon delineations based on "seedling characteristics") might have given better (?) estimation results. Disjunctive kriging resulted in the worst outcome: the MSE was 118.995. This result can easily be explained by looking at the poor transformation results: the "normalized" SEEDL data were far from being normally distributed (Figure 5.92). While the mean of the original data was relatively close to the mean of the transformed data, the difference in the variances of the original and the transformed data was very large (Table 5.2).

All interpolation methods yielded, in general, the same estimation maps (with the exception of polygonal mapping). The highest number of seedlings per plot could be found in the southwest part of the study area. Another zone of high values was found in the southeast corner, while a small area of medium values was located in the northeast. Zones of low numbers of seedlings covered the rest of the area. Chapter "Factors

Affecting the Results of Interpolation Methods" summarizes all components that have an effect on the interpolation results and the calculation of the standard deviations.

5.4 Summary

Three variables of interest were examined: number of stems per plot (STM), total basal area per plot (TBA), and number of seedlings per plot (SEEDL). Auxiliary variables included elevation (ELEV), a combined value of slope and aspect (SLOASP), and the normalized difference vegetation index (NDVI). For each variable eight different interpolation methods were compared: polygonal mapping, inverse distance weighting, inverse distance weighting squared, ordinary kriging, universal kriging with a first-degree trend surface, universal kriging with a second-degree trend surface, cokriging, and disjunctive kriging. As a measure of performance the mean square error (MSE) was calculated. Table 5.28 gives a summary of the MSEs for each interpolation method and each of the three variables under study. The secondary variables yielding the best cokriging results were: primary variable STM with secondary variable slope/aspect, and primary variables TBA and SEEDL with secondary variable elevation.

Table 5.28 Summary of mean square errors (MSE) for each interpolation method and each variable of interest: number of stems per plot (STM), total basal area per plot (TBA), and number of seedlings per plot (SEEDL).

Interpolation Method	STM	TBA	SEEDL
Polygonal mapping	44.906	3.464	113.338
Inverse distance weighting	49.140	4.122	86.763
Inverse distance weighting squared	48.915	4.598	69.881
Ordinary kriging	47.443	3.951	95.903
Universal kriging (1^{st} degree trend surface)	46.804	3.916	99.556
Universal kriging (2^{nd} degree trend surface)	45.326	3.835	100.703
Cokriging	44.568	3.911	84.646
Disjunctive kriging	49.444	3.993	118.995

For variable STM the MSEs ranged from 44.568 to 49.444 with cokriging being the best estimation method and disjunctive kriging giving the worst results. However, the differences between the various methods were relatively small. The MSEs for variable TBA ranged from 3.464 to 4.598. The best results were obtained using polygonal mapping, while the worst results were given by inverse distance weighting squared. Again, the differences between the various methods were relatively small. Variable SEEDL had the best estimation results applying inverse distance weighting squared (MSE of 69.881). The worst results were obtained using disjunctive kriging (MSE of 118.995). For this variable, the differences in MSEs of the various interpolation methods were much larger than for the other two variables.

Table 5.29 gives a summary of the rankings for each interpolation method and each of the three variables under study. Column SUM is the sum of all rankings for each interpolation technique across all three variables.

Table 5.29 Summary of rankings for each interpolation method and each variable of interest: number of stems per plot (STM), total basal area per plot (TBA), and number of seedlings per plot (SEEDL).

Interpolation Method	STM	TBA	SEEDL	SUM
Polygonal mapping	2	1	7	10
Inverse distance weighting	7	7	3	17
Inverse distance weighting squared	6	8	1	15
Ordinary kriging	5	5	4	14
Universal kriging (1^{st} degree trend surface)	4	4	5	13
Universal kriging (2^{nd} degree trend surface)	3	2	6	11
Cokriging	1	3	2	6
Disjunctive kriging	8	6	8	22

The performance of the different estimation techniques was different between variables. There was no single best interpolation method. Overall, however, cokriging performed the best, followed, with some distance, by polygonal mapping. Universal kriging with a second-degree trend surface, and universal kriging with a first-degree trend surface yielded, in general, better results than ordinary kriging. Inverse distance weighting squared and inverse distance weighting were generally outperformed by the linear kriging methods. The nonlinear kriging method disjunctive kriging performed least well.

The spatial distribution of two of the variables (STM and TBA) could be well described with discrete polygons. On the other side, variable SEEDL could only be poorly described as a discrete surface. It should be noted, however, that different sets of polygons (based on different forest characteristics) result in different estimation results. **Polygonal mapping** assumes that important variation occurs at boundaries; the variation within each polygon is smaller than the variation between the polygons.

The quality of **inverse distance weighting** is affected by the clustering in the data. Inverse distance weights depend only on the distances between the data locations and the particular location to be estimated. The weights do not take into account the clustering of sample values as do the kriging methods. However, in cases where the prerequisites of kriging are not fulfilled (approximately normal distribution of the data), as it was the case with variable SEEDL, inverse distance weighting can yield better results than the kriging methods.

In theory, **ordinary kriging** is an optimal interpolator in the sense that it minimizes estimation variance when the variogram is known, the expected values of the

mean and variance are constant over the study area, and the variable under study has a normal distribution. In practice, these conditions are never met (Weber & Englund, 1992). However, as Weber and Englund (1992) note, kriging is robust: it generally produces good sub-optimal estimates even when reality departs from the ideal. Variables STM and TBA were slightly skewed to the right and ordinary kriging gave better estimation results than inverse distance weighting. Variable SEEDL, however, was highly skewed to the right and ordinary kriging gave worse results than inverse distance weighting. All kriging methods have the advantage that they, in addition to the distance from the sample points, take the interdependence of the sample points into account by giving less weight to clustered data points.

Universal kriging is a form of interpolation that takes trends into account when minimizing the estimation error. Nonstationarity of the sample data can be recognized in the experimental variogram: if the variogram values increase faster than the lag distance squared (Armstrong, 1999) the variable under study is nonstationary. Despite the fact that there was no apparent trend visible in the variograms, removing a potential trend by applying a first- and second-degree trend surface improved the interpolation results for variables STM and TBA in comparison with ordinary kriging. Zimmerman et al. (1999) point out that for some types of surfaces it is better to completely ignore the modeling of a trend (as ordinary kriging does) then to model the trend inappropriately. This might be the reason for the better performance of ordinary kriging over universal kriging for variable SEEDL.

By utilizing the spatial cross-correlation between primary and secondary variables the quality of the **cokriging** estimates was improved as compared to the results of the

other kriging methods. Interpolation results from cokriging can be improved if the primary and secondary variable have a large correlation coefficient (Goovaerts, 1997; Asli & Marcotte, 1995). The correlation coefficient between primary variable SEEDL and auxiliary variable elevation was 0.57 and cokriging performed much better than ordinary kriging. On the other hand, primary variable STM and secondary variable slope/aspect had a correlation coefficient of 0.33, while primary variable TBA and secondary variable elevation had a correlation coefficient of 0.20. For both variables cokriging was only slightly better than ordinary kriging.

In general, **disjunctive kriging** performed least well. For variable SEEDL the transformed data were far from being normally distributed, a problem encountered when large numbers of identical data values are involved (Armstrong & Matheron, 1986). The requirement of disjunctive kriging having transformed data with a normal distribution was therefore not fulfilled. But also for variables STM and TBA the estimation results were least satisfactory, despite the good results of the transformation process (mean and variance of the original and transformed data were similar, and the transformed data had a normal distribution). Additional information (in the form of spatially cross-correlated auxiliary variables) seems to be a more important consideration than whether the estimation method is linear or nonlinear.

A complete description of all elements of the various interpolation methods that have an influence on the estimation results and their estimation standard deviation is discussed next.

5.5 Factors Affecting the Interpolation Results

Interpolation results are influenced by a large variety of factors. What are these factors that affect the differing outcomes of the various interpolation methods? While some factors affect all interpolation methods in the same way, others are specific to a family of related estimation methods, or specific to a single estimation technique. The following groups of influential factors will be examined:

1) General factors affecting the results of all interpolation methods,

2) Factors affecting the results of the kriging methods, and

3) Specific factors affecting the results of each individual interpolation method.

General factors affecting the results of all interpolation methods

There are components that have an influence on the results of estimation procedures, no matter what interpolation method is chosen. Table 5.30 gives an overview of these factors.

Table 5.30 General factors affecting the results of all interpolation methods.

• Total number of sample points
• Sample point configuration
• Outliers
• Number of sample points included in each estimation process
• Search window

a) Total number of sample points

Spatial interpolation techniques use sample point information to predict values of interest in areas not sampled. Theoretically, the more sample information is available the better the interpolation results. However, the sample points should be representative of the underlying population of the study area.

b) Sample point configuration

Clustered data points might represent the population in some areas very well, but little in parts with few sample points. Some interpolation methods take clustering into account (e.g. kriging), while others do not (e.g. inverse distance weighting). Carefully designed sampling schemes can achieve representativity of the data points. See Burgess et al. (1981), McBratney et al. (1981), and McBratney and Webster (1983) for a discussion on sampling designs for spatial interpolation procedures.

c) Outliers

Another feature affecting the representativity of sample values is the presence of outliers. With some spatial interpolation techniques (e.g. inverse distance weighting) outliers "only" affect the estimation of points in the neighborhood of the outliers. While with other estimation methods an outlier might affect the interpolation results in all of the study area: kriging results depend on the underlying variogram model. If the variogram is based on outliers this might affect the estimation in the whole study area. Robust measures of variogram calculations are available (see below).

d) Number of sample points included in each point estimation process

Theoretically, with increasing number of the estimation process should improve. However, an increasing number of sample points could contain data points that are far away from the point being estimated. Such sample points do not have any spatial correlation with the point being estimated. A large number of samples also increases the smoothing effect. Smoothing is the result of combining several sample values to calculate an estimate and thereby reduce the variability of the estimated values.

e) Search window

The choice of a search neighborhood controls which sample points are included in the estimation process. Usually, in the isotropic case, a circular search window is chosen, while in the case of anisotropic spatial correlation an ellipse is normally defined as the search area (Issaks & Srivastava, 1989; Goovaerts, 1997). When using kriging techniques, the radius of the circle normally equals the range of the variogram. The number of sample data within the search neighborhood can either be a specified single number or a range of numbers (i.e. a minimum and maximum number of sample points). In order to reduce the influence of clustered sample points, the search circle can be divided into equal angle sectors (e.g. quadrants, octants). Again, the number of sample data within each sector can either be a specified single number or a range of numbers.

Factors affecting the results of the kriging methods

The quality of all kriging methods (the estimation, as well as the estimation variances) depends largely on the quality of the underlying empirical variogram and variogram model. Table 5.31 gives an overview of the factors that have an influence on the accuracy of the variogram.

Table 5.31 Factors affecting the results of the kriging methods.

- Number of sample points
- Lag spacing and lag tolerance
- Outliers
- Variogram model
- Variogram sill
- Variogram range
- Nugget effect
- Anisotropy

a) Number of sample points

The empirical variogram is estimated from the sample data. The accuracy of the variogram calculation is proportional to the number of data points available. For each computed variogram value the number of pairs should be greater than 30.

b) Lag spacing and lag tolerance

Using different lag spacings and lag tolerances can change the appearance of the variogram drastically.

c) Outliers

Variograms are sensitive to outliers. Several alternative methods have been proposed that deal with this problem. Some of these robust variogram calculation methods include relative variograms (local relative variograms, general relative variograms, pairwise relative variograms), madogram, and rodogram. Another method for making a variogram look "smoother" is either to remove erratic single data points from the sample set, or to delete erratic data pairs from specific lags.

d) Variogram model

Different variogram models result in different kriging estimates and variances. Since kriging uses mainly the nearest sample points, the shape of the variogram model near the origin is particularly important, e.g., a Gaussian variogram model with a parabolic behavior near the origin gives more weight to the closest sample points (Isaaks & Srivastava, 1989).

e) Variogram sill

If two variogram models differ only in the sill, the model with the larger sill results in larger estimation variances, whereas the resulting kriging estimates are the same for both models (Isaaks & Srivastava, 1989).

f) Variogram range

If two variogram models differ only in the range, the model with the smaller range has the larger kriging variances (Isaaks & Srivastava, 1989). With decreasing

range, the kriging weights become more similar and the estimation process becomes a simple averaging of the sample data (increased smoothing effect).

g) Nugget effect

The nugget effect represents spatial variation at distances smaller than the shortest lag and measurement errors. If two variogram models differ only in the nugget, the model with the larger nugget causes larger estimation variances (Burgess & Webster, 1980b; Isaaks & Srivastava, 1989). As the size of the nugget increases, the kriging weights become more similar and the estimation process becomes a simple averaging of the sample data (increased smoothing effect).

h) Anisotropy

Anisotropic patterns of spatial correlation can be taken into account by using directional variograms. Such variograms give more weight to samples in the direction of the maximum spatial correlation. The estimation variances are less than with omnidirectional variograms (Isaaks & Srivastava, 1989).

Specific factors affecting the results of each individual interpolation method

Table 5.32 gives an overview of the factors that are specific to the different interpolation methods.

Table 5.32 Specific factors affecting the results of each individual interpolation method.

Polygonal Mapping • Delineation process
Inverse Distance Weighting • Clustering • Outliers • Distance exponent
Ordinary Kriging • Variogram • Stationary mean and variance • Normal distribution
Universal Kriging • Variogram
Cokriging • Variogram and cross-variogram • Sampling density • Correlation coefficient
Disjunctive Kriging • Transformation process

Polygonal mapping

Polygonal mapping is usually based on the interpretation of aerial photos (or satellite images). External landscape features are used to map polygons of homogenous areas. This *delineation process* is highly subjective: different photointerpreters might delineate "homogenous" areas differently. With changing polygon delineation, the estimates for each polygon change.

Inverse distance weighting

The quality of inverse distance weighting is affected by the *clustering* in the data and the presence of outliers (Burrough and McDonnell, 1998). Inverse distance weighting weights depend only on the distances between the data locations and the particular location to be estimated. Clustered data are not representative of the study area and therefore should receive less weight than those in sparsely sampled parts of the study area. However, inverse distance weighting is not capable of declustering such data. Unlike with the kriging methods, *outliers* with inverse distance weighting "only" affect the estimation of points in the neighborhood of the outliers.

Inverse distance weighting is also sensitive to the *power of distance* p used in the weighting process (Weber & Englund, 1994). As one decreases the exponent p, the weights given to the samples become more similar. As one increases p, the individual weights become more dissimilar, with most weight given to the nearest sample. In this case, when the location to be estimated is close to a sample location, the weight at that location will dominate completely (Myers, 1994).

Ordinary Kriging

In theory, ordinary kriging is an optimal interpolator in the sense that it minimizes estimation variance when the *variogram* is known, the expected values of the *mean and variance are constant* over the study area, and the variable under study has a *normal distribution.* In practice, these conditions are never met (Weber and Englund, 1992). However, as Weber and Englund (1992) note, kriging is robust: it generally produces good sub-optimal estimates even when reality departs substantially from the ideal. It has often been found to be insensitive to lack of normality (Hohn, 1998).

Universal Kriging

Ordinary kriging should not be applied in the presence of a strong trend (change in average values). In such cases universal kriging which takes a changing mean into account can yield unbiased estimates with minimized error variances.

The problem with universal kriging is that there are some difficulties in deriving the underlying *variogram.* Armstrong (1984) gives a good overview of the problems involved. The method applied in this project (use of residuals to calculate a variogram) has its critics: Cressie (1991) and Armstrong (1984) point out that the residual-based variogram is biased as the estimated residual-based variogram differs from the underlying variogram of the true residuals. The bias is small at small lag distances near the origin, and larger at distant lags. However, kriging mostly uses the nearest data values where the bias is small. The universally kriged estimate may be little influenced by the bias, whereas the universally kriged variance may be smaller than it should be (Cressie, 1991).

Cokriging

In theory, by making use of the spatial cross-correlation between primary and secondary variables the quality of the cokriging estimates should be improved and the estimation variances should be smaller than for the ordinary or universal kriging estimates. However, many factors have an influence on the cokriging results.

a) Variogram and cross-variogram

The quality of the cokriging estimates depends largely on the accuracy of the variogram and cross-variogram models. Errors in these models lead to errors in the cokriging estimates and variances. Cokriging requires the joint modeling of variograms and cross-variograms. The difficulty lies in the fact that variogram and cross-variogram models cannot be built independently of one another. The models have to follow the rules of the linear model of coregionalization to ensure a non-negative estimation variance. This makes the combined modeling of variograms and cross-variograms a difficult and cumbersome task. The complexity increases with increasing number of secondary variables. Cokriging and ordinary kriging yield similar interpolation results in the isotropic case (where primary and secondary variables are sampled at the same locations) if the variogram and cross-variogram models have similar shapes (Issaks & Srivastava, 1989; Goovaerts, 1997). Goovaerts (ibid.) reports that an increasing nugget in the variogram of the primary variable causes the secondary data to have larger weights.

b) Sampling density of primary and secondary variable

If the primary variable is not noticeably undersampled with respect to the secondary variable, cokriging and ordinary kriging will yield similar results (Issaks &

Srivastava, 1989; Goovaerts, 1997). Cokriging is a most useful estimation technique when the variable of interest is difficult or costly to sample, while a secondary, correlated variable can be easily and inexpensively sampled (Issaks & Srivastava, 1989).

c) Correlation coefficient between primary and secondary variable

If the primary and secondary variable are only weakly correlated, cokriging and ordinary kriging have similar interpolation results (Goovaerts, 1997). With increasing correlation coefficient the secondary variable has an increasingly larger weight in the estimation process. Asli and Marcotte (1995) report that a high correlation (correlation coefficients of 0.4 and above) between the principal variable and the auxiliary variable is important to get the maximum benefit from the information available on the secondary variable.

Disjunctive Kriging

In theory, disjunctive kriging is better than linear estimators in providing minimum error variance estimates of a property through non-linear combinations of the data and exactness of estimation. Disjunctive kriging is based upon the assumptions that the data are univariate (and bivariate) normally distributed. A *transformation process* changes any arbitrary distribution to a normal distribution with a mean of zero and unit variance. If, however, the transformed data do not follow a normal distribution, the results of the disjunctive kriging method are less satisfactory. If a random variable is univariate or bivariate normally distributed, the linear ordinary kriging estimator is identical to the disjunctive kriging estimator (Journel & Huijbregts, 1978; Rendu, 1980). The original version of DK was developed for the assumption of a bivariate normal

distribution for the transformed variable. Later research (Matheron & Armstrong, 1986a, 1986b) derived transformations for gamma, Poisson, beta, binomial, negative binomial, and hypergeometric distributions.

CHAPTER 6

CONCLUSIONS

This study compared five geostatistical methods of interpolation (ordinary kriging, universal kriging with a first-degree trend surface, universal kriging with a second-degree trend surface, cokriging, and disjunctive kriging) with three traditional estimation methods (polygonal mapping, inverse distance weighting, and inverse distance weighting squared). The eight methods were used to spatially interpolate the number of stems, the total basal area, and the number of seedlings. The major conclusions of this study can be summarized as follows:

1) No single "best interpolation method" could be found. Each database required a new search for the best estimation technique. Weber and Englund (1992, p. 391) wrote: "The search for the ideal interpolator is far from over." This sentence is as true today as it was in 1992.

2) Ordinary kriging gave good estimation results as long as the data distribution was approximately normal. The method should not be applied in cases of strong

deviations from a normal distribution. The advantage of all kriging methods lies in the fact that these techniques take the clustering of sample points into account.

3) Detrending the data with a first- and second-degree trend surface and applying kriging to the residuals gave good results. Despite the fact that no trend could be detected in the variograms, universal kriging yielded better interpolation results than ordinary kriging for two of the three variables examined.

4) Cokriging was the interpolation method that performed very well across all data sets. Exploiting the spatial cross-correlation between a primary variable of interest and auxiliary variables resulted in better estimation results. A high correlation coefficient between the variables seemed to improve the interpolation results. The disadvantage of cokriging is that the modeling of the variograms and cross-variograms has to be in compliance with the linear model of coregionalization. With increasing number of auxiliary variables the variogram modeling process becomes more difficult and cumbersome.

5) Disjunctive kriging should only be implemented if the transformed data follow a normal distribution, and the mean and variance of the original and transformed data are similar. Because the results of this nonlinear kriging method were not better that the other kriging methods, and because of its increased computational requirements and complex analysis, the use of disjunctive kriging is not recommended for this type of data set. Additional information in the study area, in the form of auxiliary variables, seemed to be a more important consideration than whether an estimator is linear of nonlinear.

6) Inverse distance weighting techniques outperformed kriging methods only in cases where the requirements of kriging (approximately normal distribution of the data) were not fulfilled.

7) Polygonal mapping gave good estimation results for two of the variables under study. The spatial distribution of these two variables could be mapped very well with discrete polygons. Therefore, the representation of a variable as a continuous surface is not always appropriate.

CHAPTER 7

RECOMMENDATIONS

A number of recommendations following this study have been formulated to assist in future research of modeling forest stand structure:

1) One of the advantages of kriging methods is that anisotropic patterns of spatial correlation can be taken into account by using directional variograms. Especially in mountainous areas (like the study site in this research project) one would expect that the spatial continuity changes with direction: spatial correlation should be large in the direction of the contour lines, and small perpendicular to the contour lines. The use of directional variograms, however, requires a large number of sample points in order to calculate reliable experimental variograms.

2) Cross-validation as a measure of evaluation has its critics (e.g. Isaaks & Srivastava; Goovaerts, 1997). A better approach would be to split the data into two independent data sets: the first data set is used in the estimation, while the second data set is used for validation. But, again, this would require a larger data set.

3) Several authors (e.g. Issaks & Srivastava, 1989; Goovaerts, 1997) report that if the primary variable of interest is not noticeably undersampled relative to the auxiliary variable, cokriging would not improve the interpolation process very much. Including a larger number of sample plots with secondary variables should therefore lead to better interpolation results.

4) Other interpolation methods are available that account for secondary information:

a) Standardized cokriging:

While ordinary cokriging has two unbiasedness conditions, standardized cokriging has only one: the weights for the primary variable and the weights for the secondary variable must sum to one. Isaaks and Srivastava (1989) report that standardized cokriging seemed to yield the better estimates. They found that one unbiasedness condition gave better results, both in the bias and the spread of the estimation errors.

b) Universal cokriging:

A trend in the primary data is modeled as a function of the spatial coordinates of the sample points. The residuals of this trend surface are then used in a cokriging process. The cokriged residuals are combined with the trend surface to obtain the final estimates of the actual surface. This method is also called cokriging with trend models (Goovaerts, 1997).

c) Kriging with an external drift:

A trend in the primary data is modeled as a linear function of the secondary variable. This method requires that the values of the secondary variable are known at all primary data locations and all locations to be estimated (Goovaerts, 1997).

5) Other nonlinear kriging methods are available:

a) Disjunctive cokriging:

This method uses spatially cross-correlated auxiliary variables to improve the disjunctive kriging estimates of a primary variable of interest (Yates, 1986).

b) Lognormal kriging:

can be applied to data with a logarithmic distribution (variable number of seedlings per plot (SEEDL) has such a logarithmic distribution).

6) Other methods for calculating conditional probabilities (besides disjunctive kriging) are available: indicator kriging and indicator cokriging. However, it is still necessary to show that a "good" estimate of the conditional probability can be obtained with any of these methods.

7) Interpolation methods are based on mathematical and statistical principles and not on ecological processes. Using ecological knowledge of the organism and environment under study should be used in the interpretation of the interpolation results. In the complex relationship between plants and their environment, factors like climate, topography, soil, moisture content, etc. and the interrelationships between them, have an influence on spatial patterns. Finding the ecological reasons for the spatial distribution of plants, therefore, requires an extensive ecological analysis.

8 Bibliography

Aboufirassi, M., and M.A. Marino. 1983. Kriging of water levels in the Souss aquifer, Morocco. Mathematical Geology 15 (4): 537-552.

Aboufirassi, M., and M.A. Marino. 1984. Cokriging of aquifer transmissivities from field measurements of transmissivity and specific capacity. Mathematical Geology 16 (1): 19-36.

Abramowitz, M., and A. Stegun. 1965. Handbook of Mathematical Functions. Dover, New York.

Alexander, R.R., Troendle, C.A., Kaufmann, M.R., Shepperd, W.D., Crouch, G.L., and R.K. Watkins. 1985. The Fraser Experimental Forest, Colorado: Research Program and Published Research 1937-1985. General Technical Report RM-118, USDA, Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado.

Antle, A.N., and P.L. Marshall. 1996. Filling in missing forestry data: exploring autocorrelational techniques. In: Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. USDA Forest Service, General Technical Report RM-GTR-277, Fort Collins, Colorado. 593-600.

Armstrong, M. 1984. Problems with universal kriging. Mathematical Geology 16: 101-108.

Armstrong, M., and G. Matheron. 1986. Disjunctive kriging revisited: part I. Mathematical Geology 18(8): 711-728.

Armstrong, M. 1998. Basic Linear Geostatistics. Springer, New York. 153 pages.

Asli, M., and D. Marcotte. 1995. Comparison of approaches to spatial estimation in a bivariate context. Mathematical Geology 27(5): 641-658.

Atkinson, P.M., Webster, R., and P.J. Curran. 1992. Cokriging with ground-based radiometry. Remote Sensing of the Environment 41: 45-60.

Biondi, F., Myers, D.E., and C.C. Avery. 1994. Geostatistically modeling stem size and increment in an old-growth forest. Canadian Journal of Forest Research 24: 1354-1368.

Bonham, C.D., Reich, R.M., and K.L. Kimberly. 1995. Spatial cross-correlation of *Bouteloua gracilis* with site factors. Grassland Science 41: 196-201.

Bouchon, J. 1979. Structure des peuplements forestiers. Annales des Sciences Forestières 36(3): 175-209.

Brus, D.J., De Gruijter, J.J., Marsman, B.A., Visschers, R., Bregt, A.K., Breeuwsma, A., and J. Bouma. 1996. The performance of spatial interpolation methods and choropleth maps to estimate properties at points: a soil survey case study. Environmetrics 7 (1): 1-16.

Burgess, T.M., and R. Webster. 1980a. Optimal interpolation and isarithmic mapping of soil properties I: The semivariogram and punctual kriging. Journal of Soil Sciences 31: 315-331.

Burgess, T.M., and R. Webster. 1980b. Optimal interpolation and isarithmic mapping of soil properties II: block kriging. Journal of Soil Science 31: 333-341.

Burgess, T.M., Webster, R., and A.B. McBratney. 1981. Optimal interpolation and isarithmic mapping of soil properties IV: sampling strategy. Journal of Soil Science 32: 643-659.

Burrough, P.A. 1986. Principles of Geographical Information Systems for Land Resources Assessment. Oxford University Press, Oxford. 194 pages.

Burrough, P.A., and R.A. McDonnell. 1998. Principles of Geographical Information Systems. Oxford University Press, New York. 333 pages.

Carr, J.R., Deng, E.D., and C.E. Glass. 1986. An application of disjunctive kriging for earthquake ground motion estimation. Mathematical Geology 18(2): 197-213.

Carr, J.R., and K.P Roberts. 1989. Application of universal kriging for estimation of earthquake ground motion: statistical significance of results. Mathematical Geology 21 (2): 255-266.

Clark Labs. 1995. IDRISI. Clark University Worcester, Massachussetts.

Cliff, A.D., and J.K. Ord. 1981. Spatial Processes: Models and Applications. Pion, London. 262 pages.

Cressie, N. 1990. The origins of kriging. Mathematical Geology 22: 237-252.

Cressie, N. 1991. Statistics for Spatial Data. Wiley, New York. 900 pages.

Dagbert, M, and M. David. 1976. Universal kriging for ore-reserve estimation - conceptual background and application to the Navan deposit. CIM Bulletin February: 80-92.

David, M. 1977. Geostatistical Ore Reserve Estimation. Developments in Geomathematics 2. Elsevier, New York. 364 pages.

David, M. 1988. Handbook of Applied Advanced Geostatistical Ore Reserve Estimation. Developments in Geomathematics 6. Elsevier, New York. 216 pages.

ERDAS, Inc. 1997. ERDAS IMAGINE. Atlanta, Georgia.

ESRI. 1997. ARC/INFO. Redlands, California.

Gallichand, J., and D. Marcottte. 1993. Mapping clay content for subsurface drainage in the Nile delta. Geoderma 58(3/4): 165-179.

Gandin, L.S. 1963. Objective analysis of meteorological fields: Gidrometeorologicheskoe Izdatel'stva (GIMIZ), Leningrad (translated by the Israel Program for Scientific Translations, Jerusalem, 1965). 238 pages.

Geovariances. 1999. ISATIS. Avon Cedex, France.

Gilbert, B., and K.Lowell. 1997. Forest attributes and spatial autocorrelation and interpolation: effects of alternative sampling schemata in the boreal forest. Landscape and Urban Planning 37: 235-244.

Golden Software Inc. 1994. SURFER 5.0. Golden, Colorado.

Goovaerts, P., and A.G. Journel. 1995. Integrating soil map information in modeling the spatial variation in continuous soil properties. European Journal of Soil Science 46: 397-414.

Goovaerts, P. 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, Oxford. 467 pages.

Gown, S.N., Warring, R.H., Dye, D.G., and J Yang. 1994. Ecological remote sensing at OTTER: Satellite Macroscale Observation. Ecological Applications 4: 322-343.

Grimm, J.W., and J.A. Lynch, 1991. Statistical analysis of errors in estimating wet deposition using five surface estimation algorithms. Atmospheric Environment 25 (2): 317-327.

Guibal, D. 1973. L'estimation des Oukoumés du Gabon. Note interne 333, Centre de Morphologie Mathématique, Fontainbleau, France.

Höck, B.K., Payn, T.W., and J.W. Shirley. 1993. Using a geographic information system and geostatistics to estimate site index of Pinus radiata for Kaingaroa Forest, New Zealand. New Zealand Journal of Forestry Science 23 (3): 264-277.

Holmgren, P., and T. Thuresson. 1997. Applying objectively estimated and spatially continuous forest parameters in tactical planning to obtain dynamic treatment units. Forest Sciences 43(3): 317-326.

Hohn, M.E. 1998. Geostatistics and Petroleum Geology. Kluwer Academic Publishers, Dordrecht. 235 pages.

Hudson, G. 1993. Kriging temperature in Scotland using the external drift method. In: Geostatistics Troia'92. Soares, A. (ed.). Kluwer, Dordrecht. 577-588.

Huijbregts, C., and G. Matheron. 1971. Universal kriging - an optimal approach to trend surface analysis. Proceedings of the 9th International Symposium on Techniques for Decision Making in the Mineral Industry. The Canadian Institute of Mining and Metallurgy, Special Volume 12. 159-169.

Isaaks, E.H., and R.M. Srivastava. 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York. 561 pages.

Jackson, M., and A. Marechal, 1979. Recoverable reserves estimated by disjunctive kriging. Proceedings of the 16th APCOM Symposium. 240-249.

Jensen, J.R. 1996. Introductory Digital Image Processing. Prentice Hall, New Jersey. 316 pages.

Jost, A. 1993. Geostatistische Analyse des Stichprobenfehlers systematischer Stichproben. PhD dissertation. University of Freiburg, Freiburg im Breisgau, Germany.

Journel, A.G., and C.J. Huijbregts. 1978. Mining Geostatistics. Academic Press, London. 600 pages.

Journel, A.G., and M.E. Rossi. 1989. When do we need a trend model in kriging?, Mathematical Geology 21(7): 715-740.

Joy, M., Klinkenberg, M., and S. Cumming. 1994. Handling uncertainty in a spatial forest model integrated with GIS. In: Proceedings: 8th Annual Symposium on Geographic Information System in Forestry, Environmental and Natural Resources Management, Vancouver, British Columbia. 359-365.

Kallas, M. 1997. Hazard Rating of Armillaria Root Rot on the Black Hills National Forest. M.Sc. Thesis, Department of Forest Sciences, Colorado State University, Fort Collins, Colorado.

Kane, V., Begovich, C., Butz, T, and D.E. Myers. 1982. Interpretation of regional geochemistry. Cumputers & Geosciences 8: 269-286.

Kim, Y.C., Myers, D.E., and H.P. Knudsen. 1977. Advanced Geostatistics in Ore Reserve Estimation and Mine Planning (Practitioner's Guide). US Department of Energy, Grand Junction, Colorado. 154 pages.

Kitanidis, P.K., and E.G. Vomvoris. 1983. A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one dimensional simulations. Water Resources Research 19: 677.

Krajewski, W.F. 1987. Cokriging radar rainfall and rain gauge data. Journal of Geophysical Research 92: 9571-9580.

Krige, D.G. 1951. A statistical approach to some basic mine valuation problems in the Witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa 52: 119-139.

Krige, D.G. 1966. Two-dimensional moving average trend surfaces for ore valuation. Transactions of the South African Institute of Mining and Metallurgy 66: 13-38.

Lajaunie, C. 1984. A geostatistical approach to air pollution modeling. In: Geostatistics for Natural Resources Characterization. Verly, G., David, M., Journel, A.G., and A. Marechal (ed.). Reidel, Dordrecht. 877-891.

Laslett, G.M., McBratney, A.B., Pahl, P.J., and M.F. Hutchinson. 1987. Comparison of several spatial prediction methods for soil pH. Journal of Soil Sciences 38: 325-341.

Laslett, G.M., and A.B. McBratney. 1990. Further comparison of several spatial prediction methods for soil pH. Soil Science of America Journal 54(6): 1553-1558.

Lillesand, T.M., and Kiefer, R.W. 1994. Remote Sensing and Image Interpretation. John Wiley & Sons, New York. 750 pages.

Longley, P.A., Goodchild, M.F., Maguire, D.J., and D.W. Rhind. 1999. Geographical Information Systems. (ed.). John Wiley & Sons, New York. 1120 pages.

Lowell, K. 1996. Discrete polygons or a continuous surface: which is the appropriate way to model forests cartographically? In: Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. USDA Forest Service, General Technical Report RM-GTR-277, Fort Collins, Colorado. 235-242.

Mandallaz, D. 1993. Geostatistical Methods for Double Sampling Schemes: Application to Combined Forest Inventories. Habilitation thesis. Chair of Forest Inventory and Planning, Department of Forest and Wood Sciences, ETH Zentrum. Zurich, Switzerland. 133 pages.

Marbeau, P. 1976. Géostatistique forestière: état actuel et dévelopements nouveaux pour l'amenagement en forêt tropicale thèse. Centre de Morphologie Mathématique, Fontainebleau, France.

Matheron, G. 1962. Traité de geostatistique appliquée, vol. I. Memoires du Bureau de Recherches Géologique et Miniéres, no. 14. Editions Technip, Paris, France. 333 pages.

Matheron, G. 1963. Traité de geostatistique appliquée, vol. II, Le Krigeage. Memoires du Bureau de Recherches Géologique et Miniéres, no. 24. Editions Technip, Paris, France. 171 pages.

Matheron, G. 1965. Les Variables Regionalisees et leur Estimation. Masson et Cie, Paris.

Matheron, G. 1969. Le Krigeage Universel. Cahiers du Centre de Morphologie Mathématique de Fontainebleau, France. 82 pages.

Matheron, G. 1973. The intrinsic random functions and their applications. Advances in Applied Probability 5: 439-468.

Matheron, G. 1976. A simple substitute for conditional expectation: the disjunctive kriging. In: Advanced Geostatistics in the Mining Industry. Guarascio, M., David, M., and C. Huijbrechts (ed.). Reidel, Dordrecht. 221-236.

Matheron, G, and M. Armstrong. 1986a. Disjunctive kriging revisited: part I. Mathematical Geology 18(8): 711-728.

Matheron, G, and M. Armstrong. 1986b. Disjunctive kriging revisited: part II. Mathematical Geology 18(8): 729-742.

Matern, B. 1960. Spatial Variation. Springer, New York. 151 pages.

Mathsoft. 1998. SPLUS 4.5. Cambridge, Massachussetts.

McBratney, A.B. and R. Webster. 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables. I. Theory and method. Computers & Geosciences 7: 331-334.

McBratney, A.B. and R. Webster. 1983. How many observations are needed for regional estimation of soil properties? Soil Sciences 135 (3): 177-183.

McBratney, A.B. and R. Webster. 1983. Optimal interpolation and isarithmic mapping of soil properties V: Co-regionalization and multiple sampling strategy. Journal of Soils Sciences 34: 137-162.

McBratney, A.B. and R. Webster. 1986. Choosing functions for semivariograms of soil properties and fitting them to sampling estimates. Journal of Soils Sciences 37: 617-639.

Metzger, K.L. 1997. Modeling Forest Stand Structure to a Ten Meter Resolution Using Landsat-TM Data. M.Sc. thesis, Department of Forest Sciences, Colorado State University, Fort Collins, Colorado. 123 pages,

Mowrer, H.T. 1997. Propagating uncertainty through spatial estimation processes for old-growth subalpine forests using sequential Gaussian simulation in GIS. Ecological Modelling 98: 73-86.

Murthy, P.S. 1989. Comparison of ordinary and disjunctive kriging in an Indian iron ore deposit. Mathematical Geology 21 (4): 443-462.

Myers, D.E. 1982. Matrix formulation of cokriging. Mathematical Geology 14: 249-257.

Myers, D.E. 1983. Estimation of linear combinations and cokriging. Mathematical Geology 15 (5): 633-638.

Myers, D.E. 1984. Cokriging - new developments. In: Geostatistics for Natural Resources Characterization. Verly, G., David, M., Journel, A.G., and A. Marechal (ed.). Reidel, Dordrecht. 295-305.

Myers, D.E. 1994. Spatial interpolation: an overview. Geoderma 62: 17-28.

Olea, R.A. 1974. Optimal contour mapping using universal kriging. Journal of Geophysical Research 79 (5): 695-702.

Olea, R.A. 1975. Optimal Mapping Techniques Using Regionalized Variable Theory. Kansas Geological Survey Series on Spatial Analysis, No. 2. Lawrence, Kansas.

Olea, R.A. 1991. Geostatistical Glossary and Multilingual Dictionary. Oxford University Press, New York.

Olea, R.A. 1994. Fundamentals of semivariogram estimation, modeling and usage. In: Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies. Yarus, J.M., and R.L. Chambers (ed.). American Association of Petroleum Geologists, Tusla, OK. 27-35.

Oliver, M.A., and R. Webster. 1990. Kriging: A method of interpolation for geographical information systems. International Journal of Geographic Information Systems 4: 313-332.

Orloci, 1978. Multivariate Analysis in Vegetation Research. Dr. W. Junk Publishers, The Hague.

Ott, R.L. 1992. An Introduction to Statistical Methods and Data Analysis. Duxbury Press, Belmont. 1051 pages.

Pan, G., Gaard, D., Moss, K., and T. Heiner. 1993. A comparison between cokriging and ordinary kriging: case study with a polymetallic deposit. Mathematical Geology 25 (3): 377-398.

Pannatier, Y. 1996. VARIOWIN - Software for Spatial Data Analysis in 2D. Springer, New York. 91 pages.

Papritz, A., and H. Fluhler. 1994. Temporal change of spatially autocorrelated soil properties: optimal estimation by cokriging. Geoderma 25: 1015-1026.

Petitgas, P. 1993. Use of disjunctive kriging to model areas of high pelagic fish density in acoustic fisheries surveys. Aquatic Living Resources 6: 201-209.

Phillips, D.L., Lee, E.H., Herstrom, A.A., Hogsett, W.E., and D.T. Tingey. 1997. Use of auxiliary data for spatial interpolation of ozone exposure in southeastern forests. Environmetrics 8 (1): 43-61.

Pohlmann, H. 1993. Geostatistical modeling of environmental data. CATENA 20: 191-198.

Ramirez-Maldonado, H. 1988. On the Relevance of Geostatistical Theory and Methods to Forest Inventory Problems. PhD dissertation, University of Georgia, Athens, Georgia.

Reich, R.M., and V.A. Bravo. 1998. Combining spatial statistics with GIS and remote sensing in designing multiresource inventory and monitoring systems. SOFOR GIS98, 2nd Southern Forestry GIS Conference, October 28-29, Athens, Georgia. 137-148.

Reich, R.M., and R.A. Davis. 1998. Class handouts for "Quantitative Spatial Analysis". Colorado State University, Fort Collins, Colorado.

Reich, R.M. 1999a. Personal communication. Department of Forest Sciences, Colorado State University, Fort Collins, Colorado.

Reich, R.M. 1999b. Department of Forest Sciences, Colorado State University, Fort Collins, Colorado.

Rendu, J.M. 1980. Disjunctive kriging: a simplified theory. Mathematical Geology 12 (4): 306-321.

Reynolds, M.R., Burkhart, H.R., and R.F. Dabiels. 1981. Procedures for statistical validation of stochastic simulation models. Forest Sciences 27: 349-364.

Rivoirard, J. 1994. Introduction to Disjunctive Kriging and Non-Linear Geostatistics. Clarendon Press, Oxford. 180 pages.

Rossi, R.E., Mulla, D.J., Journel, A.G., and E.H. Franz. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. Ecological Monographs 62 (2): 277-314.

Rossi, R.E., Dungan, J.L., and L.R. Beck. 1994. Kriging the shadows: geostatistical interpolation for remote sensing. Remote Sensing Environment 49: 32-40.

Rouhani, S. 1986. Comparative study of groundwater mapping techniques. Ground Water 24 (2): 207-216.

SAS Institute Inc. 1997. SAS 7.0. Cary, North Carolina.

Samra, J.S., Gill, H.S., and V.K. Bhatia. 1989. Spatial stochastic modeling of growth and forest resource evaluation. Forest Sciences 35 (3): 663-676.

Snee, R.D. 1977. Validation of Regression Models: Methods and Examples. Technometrics 19: 415-428.

Solow, A.R., and S.M. Gorelick. 1986. Estimating monthly streamflow values by cokriging. Mathematical Geology 18 (8): 785-810.

Song, Y., Ding, P., Cheng, Y., Hun, G., and R. Wang. 1992. Application of universal kriging to geochemical prospecting data. Mathematical Geology 24 (6): 609-630.

Stein, A., and L.C.A. Corsten. 1991. Universal kriging and cokriging as a regression procedure. Biometrics 47: 575-587.

Stonge, B.A., and F. Cavayas. 1995. Estimating forest stand structure from high resolution imagery using the directional variogram. International Journal of Remote Sensing 16 (11): 1999-2021.

Vauclin, M., Vieira, S.R., Vachand, G., and D.R. Nielson. 1983. The use of cokriging with limited field soil observations. Soil Science Society of America Journal 47: 175-184.

Ver Hoef, J.M. 1993. Universal kriging for ecological data. In: Environmental Modelling with GIS. Goodchild, M.F., B.O. Parks, and L.T. Steyaert (ed.). Oxford University Press, Oxford. 447-453.

Vogt, W.P. 1999. Dictionary of Statistics and Methodology. SAGE Publications, Thousand Oaks. 318 pages.

Wackernagel, H. 1995. Multivariate Geostatistics: An Introduction with Applications. Springer, Berlin. 256 pages.

Weber, D.D., and E.J. Englund. 1992. Evaluation and comparison of spatial interpolators. Mathematical Geology 24 (4): 381-392.

Weber, D.D., and E.J. Englund. 1994. Evaluation and comparison of spatial interpolators II. Mathematical Geology 26 (5): 589-604.

Webster, R., and T.M. Burgess. 1980. Optimal interpolation and isarithmic mapping of soil properties III: Changing drift and universal kriging. Journal of Soil Science 31: 505-524.

Webster, R., and M.A. Oliver. 1980. Optimal interpolation and isarithmic mapping of soil properties. IV: Disjunctive kriging and mapping the conditional probability. Journal of Soil Science 40: 497-512.

Wood, G., Oliver, M.A., and R. Webster. 1990. Estimating soil salinity by disjunctive kriging. Soil Use and Management 6: 97-104.

Wood, J.D., and P.F. Fisher. 1993. Assessing interpolation accuracy in elevation models. IEEE Computer Graphic and Applications 13 (2): 48-56.

Yates, S.R., Warrick, A.W., and D.E. Myers. 1986. Disjunctive kriging: I. Overview of estimation and conditional probability. Water Resources Research 22: 615-621.

Yates, S.R., and A.W. Warrick. 1987. Estimating soil water content using cokriging. Soil Science Society of America Journal 51: 23-30.

Yates, S.R., and M.V. Yates. 1988. Disjunctive kriging as an approach to management decision-making. Soil Science Society of America Journal 52: 1554-1558.

Yates, S.R., and M.V. Yates. 1990. Geostatistics for Waste Management: a User's Manual for the GEOPACK (version 1.0) Geostatistical Software System. Robert S. Kerr Environmental Research Lab, Office of Research and Development, U.S. Environment Protection Agency, Ada, Oklahoma.

Zimmerman, D., Pavlik, C., Ruggles, A., and M.P. Armstrong. 1999. An experimental comparison of ordinary and universal kriging and inverse distance weighting. Mathematical Geology 31 (4): 375-390.