AN ABSTRACT OF THE THESIS OF

Mohamad A. Al-Taher for the degree of Doctor of Philosophy

in Soil Science presented on December 17, 1991

Title: Geostatistical Application to Salinity Mapping and Simulated Reclamation

Abstract Approved: Redacted for Privacy

B. P. Warkentin

Geostatistical methods were used to find efficient and accurate means for salinity assessment using regionalized random variables and limited sampling. The random variables selected, sodium absorption ratio (SAR), electrical conductivity (EC), and clay content were measured on samples taken over an area of fifteen square miles. Ordinary kriging and co-kriging were used as linear estimators. They were compared on the basis of average kriging variance and sum of squares for error between observed and estimated values.

The results indicate a significant improvement in the average kriging variance and sum of squares by using co-kriging estimators. EC was used to estimate SAR because of the high correlation between them. This was not true for clay content. A saving of two-thirds of the cost and time was achieved by using electrical conductivity as an auxiliary variable to estimate sodium absorption ratio. The nonlinear estimator, disjunctive kriging, was an improvement over co-kriging in terms of the variances. More information at the estimation site is a more important

consideration than when the estimator is linear. Disjunctive kriging was used to produce an estimate of the conditional probability that the value at an unsampled location is greater than an arbitrary cutoff level. This feature of disjunctive kriging aids salinity assessment and reclamation management.

A solute transport model was used to show how saptially variable initial conditions influenced the amount of water required to reclaim a saline soil at each sampling point in a simulated leaching of the area.

Geostatistical Application to

Salinity Mapping and Simulated Reclamation

by
Mohamad A. Al-Taher

A THESIS
submitted to
Oregon State University

in partial fulfillment of
the requirement for the
degree of
Doctor of Philosophy

Completed December 17, 1991
Commencement June 1992

APPROVED:

Redacted for Privacy

Professor of Soil Science in charge of major

Redacted for Privacy

Head of Department of Crop and Soil Science

Redacted for Privacy

Dean of Graduate School

Date thesis is presented December 17, 1991

Typed by Mohamad A. Al-Taher

ACKNOWLEDGEMENT

I am deeply indebted to my major professors, Dr. Benno Warkentin and Dr. Peter Shouse, for their guidance, support, and intellectual stimulation. I consider myself very fortunate to be their student. I extend my appreciation to other members of my graduate committee, Dr. James Vomocil, Dr. Marvin Pyles, Dr. David R. Thomas, and Dr. Arsalan Mazaheri with whom it was always a pleasure to cooperate. I want to thank Mrs. Janet Preble for her considerable help. My special thanks to Dr. Scott Yates, for all his support, assistance, ideas, and his programs which made these results possible. I am very thankful to Dr. Martinus Th. van Genuchten, and Mr. Walter Russell as well as other U.S. Salinity lab personnel for their help in making my stay in Riverside, California useful and productive. My appreciation is extended to Oregon State University and the University of California, Riverside for their financial assistance.

Finally, and mostly, I want to express my love and gratitude to my family. I owe a very special thanks to you Najat. Without you,, sweetheart, it would have been much harder.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Geostatistical Application to Salinity Mapping and Simulated Reclamation

CHAPTER 1

Methods of Geostatistical Analysis

for Salinity Assessment

Introduction

Maintaining an adequate supply of soil water for plant use is a problem when salt concentrations in the soil exceed threshold levels. Knowledge of the salinity distribution within a region is important for the development of a rational water management system for efficient water use and for delineating possible sites of soil and groundwater pollution. Our tasks in soil salinity assessment are to subdivide large heterogeneous regions into smaller regions which have relatively homogeneous salt concentrations and to associate these salinity levels with edaphic, management, or climatic factors operating within the region. The need for comprehensive methods for the timely assessment of soil salinity has become increasingly important. One such method has recently been developed by Rhoades et al., (1988). This promising new tack for assessing the spatial and temporal distribution of soil salinity uses portable instrumental techniques for measuring bulk soil electrical conductivity (EC_a). Data collected using this new technology lends itself quite well to geostatistical analysis and computer assisted mapping. In this chapter we will develop a geostatistical methodology for computerized salinity assessment. In subsequent chapters, the methodology will be applied to mapping saline and sodic soils and then to simulated remediation of salt affected areas.

The fundamentals of geostatistics were developed empirically by Krige and others during the 1950's to assist in the location, assessment, and management of gold ore reserves in South Africa (Clark, 1979; David, 1977; Cooper and Istok, 1988). A coherent body of theory was later developed by Matheron (1963) for solving ore

estimation problems associated with spatial variation. This theory embraces a set of specific statistical techniques for quantifying the correlation of spatially distributed random variables, and for performing interpolation of these variables (Journal and Huijbregts, 1978; Oliver, 1987; Webster, 1985; Warrick et al., 1986; Stein, 1991).

Geostatistical techniques provide a means for describing the spatial variability of soil properties by quantifying the spatial and inter-variable correlation among samples, provide optimum interpolation schemes for preparation of soil salinity maps, and provide a method of estimating the location of new sampling sites to improve quantitative estimates of spatially variable soil properties (Webster, 1985). Literally hundreds of articles have been published in soil science journals on the use of geostatistics to characterize the spatial variability of soil properties in the natural landscape. Some soil science researchers have used geostatistical methods to estimate and map soil properties such as water content (Morkoc et al., 1985; Yates and Warrick 1987), soil temperature (Vauclin et al., 1982; ten Berge et al., 1983; Morkoc et al., 1985; Davidoff et al., 1986; Yates and Warrick 1987), soil-water pressure head (Saddiq et al., 1985; Hamlett et al., 1986; Yeh et al., 1986), various soil chemical properties (Campbell, 1978; Gajem et al., 1981; McBratney et al., 1982; Yost et al., 1982a and 1982b; van der Zaag et al., 1984; Trangmar et al., 1985; Webster and McBratney, 1987; Samura et al., 1988), soil texture and structural stability (Campbell, 1978; Vauclin et al., 1983; Yates and Warrick, 1987; Ovalles and Collins, 1988; Shouse et al., 1990), mechanical impedance (Selim et al., 1987), infiltration (Sisson and Wierenga, 1981; Vieira et al., 1981; Bautista and Wallander,

1985; Berndtsson and Larson, 1987; Cressie and Horton, 1987), water retention properties (Gajem et al., 1981; Russo and Bresler, 1981 and 1982; Vauclin et al., 1983), salinity (Hajrasulliha el al., 1980; Miyamoto and Cruz, 1987), and soil surface topography (Mulla, 1988). Some soils researchers have focused on determining spatial scales and patterns of soil variation to improve sampling efficiency (Burgess and McBratney, 1981; Campell, 1978; McBratney et al., 1981; McBratney and Webster, 1981).

Few soil scientists have applied geostatistics to large-scale mapping projects, and even fewer have applied geostatistics to salinity assessment and remedial management studies. A well-defined, comprehensive methodology for this latter application has not yet been developed. Although the theoretical basis and practical implications of the methods described in this chapter do exist in the literature, relevant papers are dispersed among journals of widely differing disciplines. Thus, the objectives of this first chapter are 1) to briefly review and bring together all geostatistical theory relevant to salinity mapping; 2) to outline methods to be used in applying this theory to salinity assessment; 3) to assess the uncertainty of various geostatistical estimation procedures; and 4) to evaluate the role of auxiliary variables in the estimation process. Actual applications are discussed in subsequent chapters.

Basic geostatistical principles

Any set of n values, $Z(x_i),...,Z(x_n)$, of a measured soil property, Z, within the landscape is called a regionalized or spatial random variable (Cooper and Istok, 1988). The collection of random variables for all points in the landscape is the random function Z(x). In soil science Z(x) is known only at the actual measurement sites, x_i; Z(x) is considered to be a random variable everywhere else within the landscape. This simply means that we have only one realization of the random function Z(x) within the landscape. The problem in soil science is that the number of observations, $Z(x_i)$, is usually quite limited. Therefore, to accurately map a soil property requires estimation of Z(x) at unvisited (unsampled) coordinates, x_o, in the landscape. The best estimator of Z(x) at a specific location x_o is given by the expected value of $Z(x_o)$ as determined by the values already measured, $Z(x_i),...,Z(x_n)$. This is called the conditional expectation of $Z(x_o)$ and requires the joint probability distribution of the n + 1 random variables $Z(x_o)$, $Z(x_i),...,Z(x_n)$ (Journel and Huijbregts, 1978; Rendu, 1978; Stien, 1991). Knowledge of the joint probability distribution is impossible with only one realization of Z(x). Another estimator with fewer requirements is necessary.

As a direct consequence of having only one realization of Z(x), linear estimators are usually considered adequate in geostatistical analysis (Delfiner, 1976; Olea, 1975). Linear estimators require that the mean and variance of the probability distribution of Z(x) be known (Cooper and Istok, 1988). To estimate these statistical moments of Z(x) stationarity of Z(x) is assumed. In the case of stationarity,

measurements at different locations (x_i) can be considered separate realizations of Z(x) and the measured values can be used to calculate the moments of the probability distribution of Z(x). Linear geostatistics uses linear estimators and the assumptions of stationarity.

The first two moments of Z(x) required for geostatistical analysis are defined by Equations 1-4. The first order moment is simply the mean, m, of Z(x):

$$E[Z(x)] = m \qquad [1]$$

There are three second order moments; the variance, the covariance and the variogram (Equations 2, 3, and 4, respectively):

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

$$C(h) = E\{[Z(x+h)][Z(x)]\} - m^2 \qquad [3]$$

$$2\gamma(h) = E\{[Z(x+h) - Z(x)]^2\} = C(0) - C(h) \qquad [4]$$

where E{ } is the expected value. The form of the variogram most often used in geostatistical practice is called the semivariogram, defined in Eq. 5.

$$\gamma(\mathbf{h}) = \frac{1}{2} E\left\{\left[Z(x+\mathbf{h}) - Z(x)\right]^2\right\} = C(0) - C(\mathbf{h}) \qquad [5]$$

Under stationarity assumptions, the covariance and semivariance are not functions of position within the landscape, but functions only of the vector distance ,**h**, separating pairs of measurement points. Many spatially random variables show a tendency for variance estimates, C(0), to increase without limit as the size of the sampled area is increased. Under these circumstances an *a priori* variance and covariance for the random function Z(x) are undefined (nonstationarity). Matheron (1971) promoted the less restrictive intrinsic hypothesis, namely stationarity of the mean and variance of the differences between samples which are distance **h** apart. The mean and semivariance are defined under the intrinsic hypothesis. Because of this fact, the semivariance is the principal tool used in geostatistics to describe the spatial correlation structure of regionalized variables. The semivariogram (semivariance plotted as a function of lag distance) reveals facts about the continuity of the data in space, the directions of the continuity, the range of the variance structure, or interdependence, and whether or not the data are isotropic.

Methodology

Our proposed methodology of soil salinity mapping consists of six steps. "Preliminary data analysis" is the use of classical statistics procedures to characterize the distribution of measured soil salinity and related variables and to make

appropriate transformations to normality. "Variography", the second step, is calculating the experimental semivariogram. The third step, "modeling", is selecting and fitting appropriate theoretical models to the experimental semivariograms. "Kriging", the fourth step, is using the semivariogram models for interpolation. Co-kriging uses the semivariogram and cross semivariogram models (describing the spatial correlation structure of two or more variables) for interpolation. Co-kriging makes use of more than one regionalized variable. Step six, "disjunctive kriging", is using non-linear geostatistical techniques in making management decisions. The above six steps are comprehensively described below.

1. Characterizing sample data

Preliminary data analysis is an essential part of characterizing a spatial data set. Robust and resistant data analysis techniques are used for this purpose (Tukey, 1977; Hoaglin et al., 1983; Hamlett et al., 1986; Cressie and Horton, 1987). These procedures often give valuable information about the distribution of the data, the types of transformations needed, and identify possible outliers (errors) or atypical values which may be important in a future study or for deeper understanding. Application of robust and resistant statistical techniques is important for testing the stationarity and normality conditions which must be satisfied in order to correctly use linear geostatistics (Hamelett et al., 1986; Cressie and Horton, 1987; Cressie, 1986).

Deviations from normality can often be overcome by making appropriate transformations or by identifying and eliminating erroneous measurements from the

sample data (Armstrong, 1984; Tukey, 1977). Goodness of fit tests such as the chi-squared statistic, the Kolmogorov-Smirnov statistic or the Shapiro-Wilk statistic can be used to determine if a regionalized variable is normally distributed (Henley, 1981; Delhomme, 1976; Russo and Jury, 1987).

Soil salinity and other soil properties have been found to be log-normally distributed (Hajrasuliha et al., 1980; Warrick and Nielsen, 1980; Jury, 1985) and must be transformed to fit a normal probability distribution. The natural log transformation function is given by

$$Y(x) = \ln[Z(x) + A] \qquad [6]$$

where Y(x) is the natural log-transformed data, Z(x) is the original data, and A is a constant which may be added to Z(x) to improve the fit (Rendu, 1978; Tukey, 1977). The relationships of parameters in the normal and log-normal probability istributions are given by

$$m = \exp\left[\alpha + \frac{\beta}{2}\right] \qquad [7]$$

$$\sigma^2 = m^2\left[\exp(\beta^2) - 1\right] \qquad [8]$$

where m and σ^2 are the mean and variance of the random function Z(x), and α and

β^2 are the mean and variance of the random function Y(x) (Rendu, 1978). These relationships are important when transforming Y(x) to Z(x).

2. Variography: Calculating semivariograms

Equation 5 defines the true semivariance of the regionalized variable, Z(x), but with one realization of the random function the true semivariance can only be estimated. This estimate is the experimental semivariance, $\gamma^*(\mathbf{h})$, defined as

$$\gamma^*(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [Z(x_{i+\mathbf{h}}) - Z(x_i)]^2 \qquad [9]$$

where $N(\mathbf{h})$ is the number of sample pairs separated by the vector $\mathbf{h}$.

Two types of estimation errors occur when calculating the true semivariance from a single realization (Journel and Huijbregts, 1978). The first is called the variance of estimation and is a consequence of using a limited number of sample measurements $Z(x_i)$, to estimate the semivariance. This error is inversely proportional to the number of sample pairs $N(\mathbf{h})$ at a given lag distance (Cooper and Istok, 1988). The second type of error is a manifestation of local fluctuations in the mean value of Z(x). These fluctuations cause estimates of the semivariance to vary from point to point in the landscape. This source of error is called the fluctuation variance and is related to the distribution of the separation distances between sample pairs (Cooper and Istok, 1988).

A theoretical analysis of the above two error sources in the experimental

semivariance is used to determine the minimum number of sample pairs, and a maximum value of the separation vector, **h**, for which the true semivariance can be estimated from the experimental semivariance (Journel and Huijbregts, 1978; Webster, 1985; Cooper and Istok, 1988). Two practical rules for calculating the experimental semivariance follow directly from these analyses:

Rule 1: $N(\mathbf{h}) > 50,$ [10]

Rule 2: $|\mathbf{h}| < L/2,$ [11]

where $N(\mathbf{h})$ is the number of pairs at vector distance **h**; $|\mathbf{h}|$ is the magnitude of the separation vector; and L is the longest possible distance in the sampled landscape. These two rules are used to reduce estimation errors.

Rule 2 implies that kriging (the interpolation procedure based on the semivariogram to be described later) should not be used to estimate soil properties that are farther than L/2 away from the nearest sampling site. The sample pairs restriction, (rule 1) affects the calculation of the experimental semivariance. It is often necessary to group sample pairs into distance classes for each specified vector, $\mathbf{h}_o$, in order to obtain adequate numbers for accurately estimating $\gamma(\mathbf{h})$ (David, 1977; Webster, 1985). In two-dimensional problems, like salinity mapping, **h** may be specified by an angle (θ) and a distance ($|\mathbf{h}_o|$) while tolerance intervals for angles and distances can be defined using the method outlined by Webster (1985). Sample pairs can be grouped together if the magnitude and angle describing the vector, $\mathbf{h}_o$, separating the two points fall within the respective tolerance intervals for the specified vector. These sample pairs are then used to calculate the experimental

semivariance ($\gamma^*(\mathbf{h}_o)$.

In soil salinity mapping, the sample numbers within the landscape are usually small, causing difficulty in selecting tolerance intervals which provide sufficient numbers of sample pairs for calculating the entire $\gamma^*(\mathbf{h}_o)$ curve. Under these conditions, the semivariance may be calculated by selecting sample pairs which fall into the tolerance intervals for the specified angles. The selected sample pairs are then sorted into distance classes that contain the same number of sample pairs (specified to be > 50 according to rule 1). For each distance class, $\gamma^*(\mathbf{h})$ is calculated for the average distance within each class. This procedure ensures that the experimental semivariance is based on an "adequate" number of sample pairs.

The method of moments estimator of the semivariance, Eq. 9, is sensitive to extreme values of $Z(x_i) - Z(x_{i+h})$, especially because the squared differences follow the highly skewed chi-squared distribution with one degree of freedom (Webster, 1985). Sometimes the class intervals or tolerance levels must be rather wide to get sufficient numbers of sample pairs for estimating $\gamma^*(\mathbf{h})$. Within a distance class or angle tolerance interval there may be extreme values of $[Z(x_i) - Z(x_{i+h})]^2$. Cressie and Hawkins (1980) proposed a more robust estimator for the semivariance to deal with these cases:

$$\gamma^* = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left\{ \left[\left(Z(x_i) - Z(x_{i+h}) \right)^2 \right]^{\frac{1}{4}} \right\}^4 / (0.457 + 0.494/N) \qquad [12]$$

where [0.457 + 0.494/N] is a bias-correcting term which ensures the expected value

of $2\gamma^*(\mathbf{h}) \simeq 2\gamma(\mathbf{h})$. This robust estimator helps to minimize the effects of artificially grouping data sample pairs into wide class bands or wide tolerance intervals, a procedure which is sometimes necessary to calculate the experimental semivariance.

An analysis of two-dimensional variations, as in salinity mapping, must always allow for possible anisotropy. It is not unusual to find anisotropic experimental semivariograms in which the semivariogram along one direction is different than the semivariogram along another (Figure 1). In these cases, the angles and distances and their respective tolerance intervals can be selected to accentuate suspected anisotropies and to provide enough sample pairs for accurately calculating $\gamma^*(\mathbf{h}_o)$. The experimental semivariogram is an excellent tool for determining the extent of field anisotropy, by calculating the semivariogram for several directions, (e.g. east-west, north-south, northeast-southwest, northwest-southeast) and comparing these "directional" semivariograms. Figure 1 illustrates an example of the detection of anisotropic conditions with directional semivariograms and is an example of geometric anisotropy. The gradient of the semivariance in the north-south direction is much steeper than the semivariance in the east-west direction. Experimental semivariogram models must take into account anisotropic field conditions.

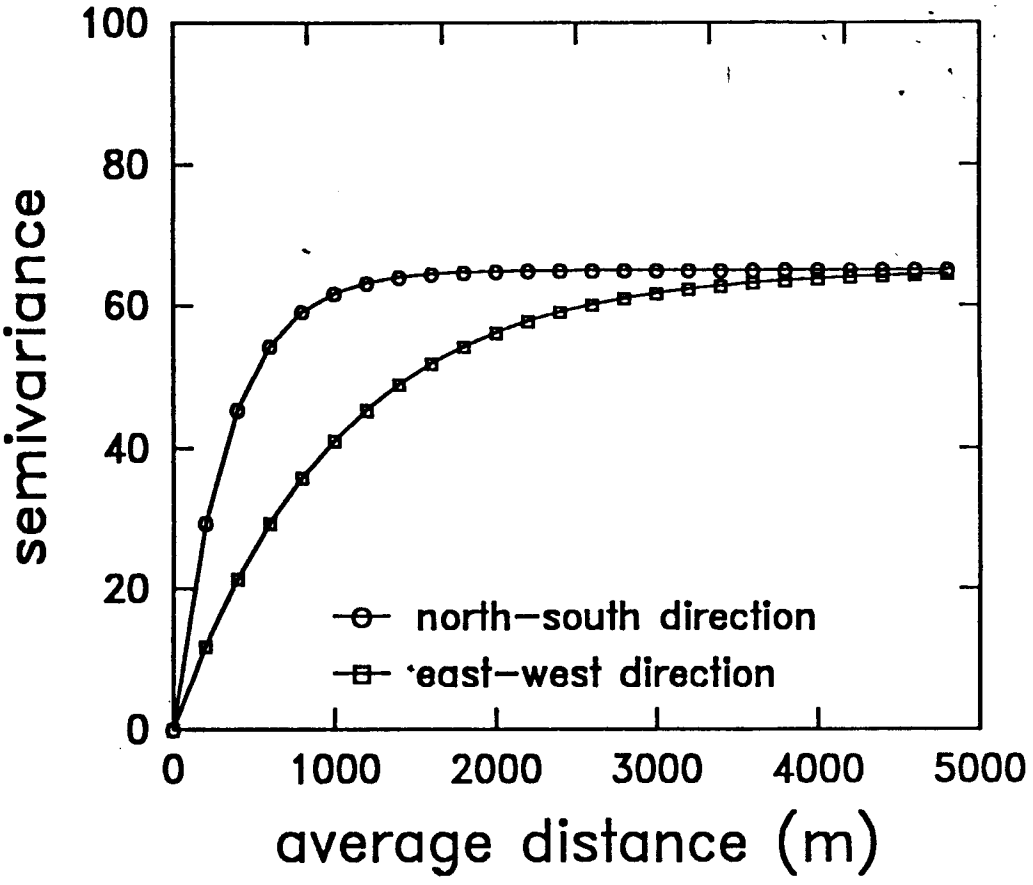


Fig. 1. Directional Semivariograms.

3. Fitting models to experimental semivariograms

Soil salinity usually varies continuously in space (Jury, 1985; Webster, 1985) and semivariograms are usually continuous functions. Structural variance analysis consists of fitting mathematical models to experimental semivariograms. It is generally possible to fit simple mathematical functions to experimental semivariograms (Webster, 1985). We will use the semivariogram function for interpolation (kriging and co-kriging), so its mathematical form is of practical importance.

Two main types of semivariogram models exist. The first type is characterized by a semivariance which increases with increasing lag distance, $|\mathbf{h}|$, to some maximum and then levels off and remains constant with further increases in lag distance. This type of semivariogram is described as transitive. Figure 1 illustrates a generalized example of transitive semivariogram models. As the distance between sampling sites increases, the soil salt concentrations become increasingly dissimilar; there is also a finite distance within which all variation is encountered. The lag distance where the semivariogram reaches a maximum is called the range (a) and marks the limit of spatial interdependence (structured variation). The range of the semivariogram model is an extremely important statistical parameter. It has been interpreted as a measure of the average distance across distinct soil types by Webster and colleagues (Webster, 1973; Webster and Cuanalo, 1975; McBratney and Webster, 1981). The semivariogram maximum is the *a priori* variance of soil salinity within the landscape, known as the sill (C_0+C). A regionalized variable with a transitive

semivariogram model is considered by some to imply that the variable is second-order stationary. However, recent studies (Russo and Jury, 1987; Stark and Fang, 1982) indicate that nonlinear drift can still produce transitive semivariograms.

For the second major type of semivariogram model, the semivariance increases without limit. In this case there is no finite *a priori* variance. This type of semivariogram indicates a soil property with an infinite capacity for variation within the sampled-region (or scale of the study). If sampled on a much larger scale, perhaps a maximum variance could be found.

Spatial correlation can be present at any scale within the soil landscape but this correlation is only defined for distances greater than the smallest separation between sample pairs. A ramification of this restriction is that for some semivariograms the apparent limiting value of $\gamma(\mathbf{h})$ is not zero as $|\mathbf{h}|$ approaches zero. The situation when the y-intercept of the semivariogram is non-zero is known as the nugget effect; the intercept is known as the nugget variance (C_0). The nugget effect is attributed to a combination of measurement errors and/or variations (of the random function) at separation distances smaller than the closest sampling interval. As we notice, the sample semivariogram does not provide any information about the spatial structure of the variance for distances shorter than the minimum spacing between sample data; unless the sampling includes duplicate samples at the same location. So the behavior of the semivariogram near the origin can not usually be determined from the sample semivariogram. We will see in later chapters how the nugget variance affects the resulting estimate of the interpolated values of Z(x).

When there appears to be no spatial correlation over the range of **h** for which the semivariogram is defined, and the semivariogram values are equal to the population variance, then the semivariogram is modeled as a pure nugget variance. The pure nugget effect makes the estimation procedure more like a simple averaging of the available data.

When choosing a model to represent an experimental semivariogram, one must allow for at least three parameters; a y-intercept (nugget variance), a monotonically increasing section of potentially varying shape, and a sill (*a priori* variance) (McBratney and Webster, 1986; Russo and Jury, 1987). In two or more dimensions one must also provide for anisotropy if it exists. Any mathematical function for experimental semivariograms must be conditionally positive-definite, (i.e.,increasing or constant with increasing **h**, and non negative) (Olea, 1975; McBratney and Webster, 1986; Webster, 1985). This condition is a strict one and a model for the semivariogram may not be chosen just because it looks as though it fits the experimental data well (Armstrong and Jabin, 1981). In the mining literature, functions which meet this criterion are called authorized functions (David, 1977; Journel and Hujbregts, 1978; Clark, 1979; McBratney and Webster, 1986). The best advice is to choose an appropriate function from authorized models listed by Clark (1979), Journel and Huijbregts (1978), or McBratney and Webster (1986). Below we describe several authorized models frequently found in analysis of spatial variation of soil properties.

By far the simplest authorized model that can be fitted to one-dimensional

data is the linear model of the general form:

$$\gamma(\boldsymbol{h}) = C_0 + W(\boldsymbol{h}) \qquad [13]$$

where C_0 is the nugget variance and W is the slope. In theory, W can approach zero in which case the semivariogram model is called a pure nugget effect and all of the variance occurs within the smallest interval between sampling sites. There is no redundancy of information between samples, and in terms of statistical distance, no sample is closer to the point being estimated than any other sample. If the pure nugget model is to be used in ordinary kriging, all weights are equal. The pure nugget model describes a complete lack of spatial correlation.

The spherical model has been found to fit semivariograms of mineral deposits (Journal and Huijbregts, 1978; Rendu, 1978; Clark, 1979), and soil properties as well (Burgess and Webster, 1980; McBratney and Webster, 1981; Russo and Bresler, 1981). Mathematically, this function is of the form:

$$\gamma(h) = C_0 + C\left[\frac{3\,h}{2\,a} - \frac{1\,h^3}{2\,a^3}\right] \quad \text{for } 0 < h \leq a$$

$$\gamma(h) = C_0 + C \quad \text{for } h > a \qquad [14]$$

where $C_0 + C$ is the sill, C_0 is the nugget variance and a is the range (see Figure 2).

Another transitive model, popular with soil scientists, is the exponential model

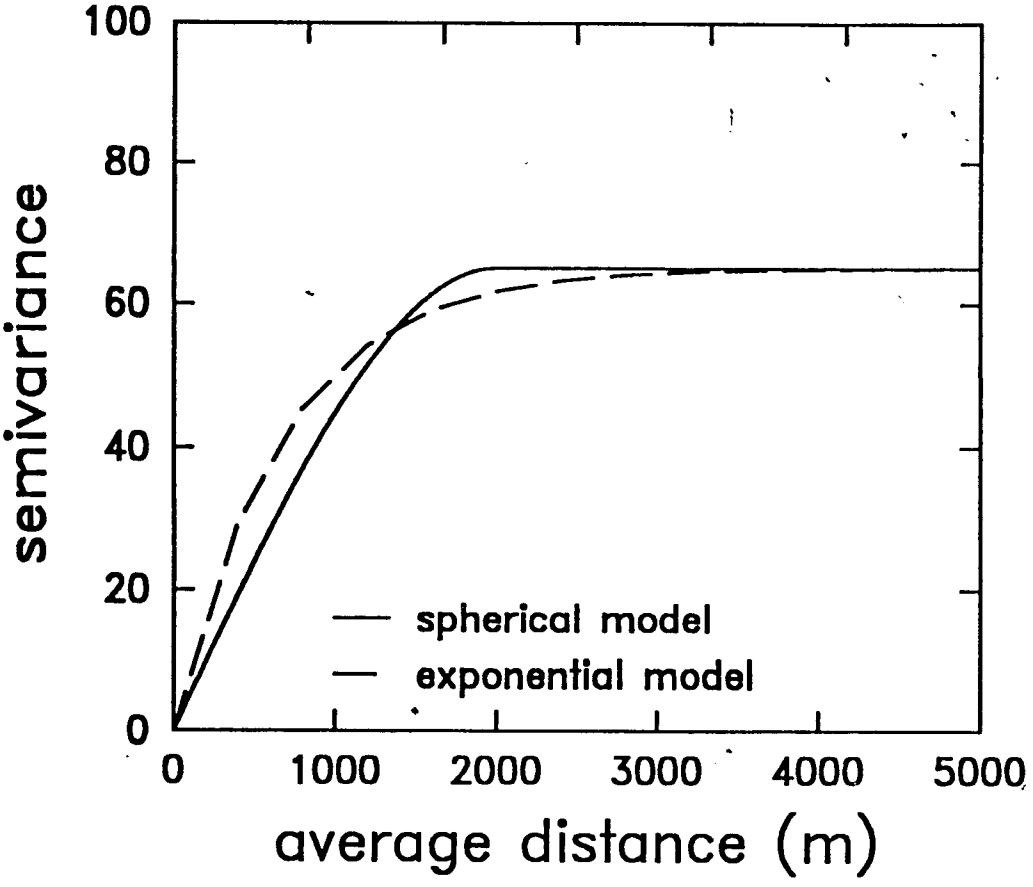


Fig. 2. Spherical and exponential models using similar parameters.

$$\gamma(h) = C_0 + C\left[1 - \exp\left[\frac{-h}{r}\right]\right] \quad \text{for } 0 < h \qquad [15]$$

where C_0, $C_0 + C$ and **h** are as before, and where the parameter r defines the spatial scale of the variability in a way analogous to the range of a spherical model. The sill is approached asymptotically so there is no distinct range, although in practice the effective range (a' = 3r) is defined as the lag at which the semivariogram reaches $0.95(C_0 + C)$.

The exponential model has appealed to statisticians because of its applicability to randomness in space (McBratney and Webster, 1986). First-order auto-regressive and Markov processes are described by exponential semivariograms (Sisson and Wierenga, 1981; Webster, 1985). Transitional phenomena related to the exponential distribution of distances between soil boundaries also give rise to exponential semivariograms (Burgess and Webster, 1981; Oliver, 1987).

Other "authorized" models also exist. We will list some of them here for completeness: 1) the circular model, which is the two-dimensional analog of the spherical model; 2) the one-dimensional Gaussian model, which has a sigmoidal shape; 3) the logarithmic model, which has been popular because of its ease to convert data to linear expressions; and 4) the hyperbolic model, which was used by Vieira et al. (1981) to describe the measured variation in infiltration across an irrigated field in California. Additional discussions of these and other models are given by McBratney and Webster (1986), Webster (1985), and Warrick et al. (1986).

The procedure for fitting an appropriate mathematical model to an experimental semivariogram usually consists of several steps: 1) choosing several candidate models, e.g. the spherical, exponential, circular, etc.; 2) obtaining preliminary estimates of the model parameters (sill, nugget, range) which can be done either by the eyeball method (Clark, 1979; David, 1977) or by using more objective weighted least squares (Cressie, 1985) or maximum likelihood methods (Kitanidis, 1983; Russo and Jury, 1987); and 3) choosing the "best" model from among the available candidate models.

Choosing from among the reasonable candidate models can be difficult because there are rarely any theoretical grounds for choosing a particular model. There are several statistical criteria that can be employed to pick the "best" model for a specific experimental semivariogram. One such method is the Akaike information criterion (AIC). This method helps to determine the model which has the "best" fit with the fewest parameters. The AIC is estimated by

$$AIC^* = n \ln(R) + 2p \qquad [16]$$

where AIC* is the unbiased estimate of the AIC, n is the number of observations (points in the experimental semivariogram), p is the number of independent parameters in the model, and R is the residual sum of squares of deviations from the fitted model. The model having the lowest AIC is judged "best" (Yost et al., 1982a; Russo and Jury, 1987; McBratney and Webster, 1986).

Another method for determining the best model from a set of candidate

models is by means of cross-validation (Delhomme, 1976). This procedure has been misnamed "jackknifing" in the literature (Davis, 1987) and is in reality a "leave-one-out" method. In this method, sample values are deleted one at a time and then kriged (described later) to estimate the missing data point from the remaining sample values. Statistical analysis of the kriging errors (differences between estimated and measured values) and the standardized mean-squared errors (the average of the kriging errors divided by their respective kriging variances) determine if a bias is present in the estimates and if the estimation errors are consistent with the kriging variances. The average kriging error (AKE) must be close to zero for a model to produce unbiased estimates.

$$AKE = \frac{1}{n} \sum_{i=1}^{n} \left[Z^*(x_i) - Z(x_i) \right] = 0 \qquad [17]$$

where $Z^*(x_i)$ is the estimated value at x_i, $Z(x_i)$ is the measured value at x_i, and n is the number of sample values (Cooper and Istok, 1988). For kriging errors to be consistent with the kriging variances, the standardized errors should be normally distributed with a mean (reduced mean, $R\mu$) of zero and a variance (reduced variance, $R\sigma^2$) close to one (Davis, 1987; Vauclin et al., 1983; Yates and Warrick, 1987). However, there is no independent method for testing how close to zero (the mean) and to one (the variance) these two variables should be (Yates and Warrick, 1987). At best, the reduced mean and reduced variance are only indicators for determining the best model from a group of models (Davis, 1987). The use of

cross-validation for model discrimination has been suggested for applications involving kriging, while the AIC method has been preferred for cases where it is necessary to describe the main characteristics of the spatial variability (McBratney and Webster, 1986).

4. Kriging for estimating soil salinity

The prime reason for developing a geostatistical analysis methodology is to use it for estimating soil salinity at unsampled sites so that contour maps of the spatial distribution of salt content within the landscape can be made. The method of estimation is known as kriging (Krige, 1966; Matheron, 1971). It is essentially a means of weighted local averaging based on the semivariogram model, which provides the required spatial information (Oliver, 1987; Olea, 1975; Webster and Burgess, 1980). Kriging has the advantage over regression methods in that the spatial correlation is taken into account during the estimation process. Also, like regression methods, the estimate is unbiased and has a minimum variance between the estimate and the actual value (Journel and Huijbregts, 1978; Clark, 1979).

There are several forms for writing the linear kriging equations which have been given names: simple kriging, ordinary kriging, and universal or intrinsic kriging. Each will be discussed.

Simple Kriging

Simple kriging assumes that the mean value of the random variable is known

a priori. This assumption limits the use of this technique for general estimation problems, and simple kriging is not generally used in spatial estimation. Mean salinity within the landscape is usually not known, therefore simple kriging will not be discussed in detail. Interested readers should consult Journel and Huijbregts (1978) for a detailed explanation of the kriging equations.

Ordinary Kriging

In most instances the mean value of salinity in the landscape is unknown. The kriging equations can be written in such a manner that the mean value becomes part of the solution. In this situation we are restricted to a linear combination of the available data to estimate $Z(x_0)$. Let us consider soil salinity to be a random function, $Z(x)$, defined on a point support (i.e. sample size $<<<$ area sampled), and sampled in two-dimensional space. The ordinary kriging estimator, $Z^*(x_0)$, of an unsampled site is a linear sum of weighted observations within a given neighborhood. The kriging estimator for ordinary kriging is written as (Burgess and Webster 1980):

$$Z^*(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \qquad [18]$$

where $Z^*(x_0)$ is the estimate of Z at x_0, λ_i is the weight assigned to the i^{th} observation and n is the number of observations within the local neighborhood.

An unbiased estimate of $Z(x_0)$ (i.e. $E[Z^*(x_0)-Z(x_0)]=0$) requires that a constraint be placed on the weights, λ_i's such that

$$\sum_{i=1}^{n} \lambda_i = 1 \qquad [19]$$

The error of estimation, or estimation variance, for this linear estimator (Eq. 18) is given by

$$E\left\{\left[Z(x_0) - Z^*(x_0)\right]^2\right\} = -\gamma(x_0, x_0) + 2\sum_{i=1}^{n} \lambda_1 \gamma(x_0, x_i) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(x_i, x_j) \qquad [20]$$

The best linear unbiased estimate of the conditional expectation of $Z(x_0)$ is obtained by using Lagrangian techniques to minimize the estimation variance. The ordinary kriging equations result from the minimization of the estimation variance and the unbiased condition in Eq. 19. The "kriging system" is a set of n+1 linear equations with n+1 unknowns obtained by setting each partial derivative equal to zero

$$\partial \left[E\left\{\left[Z(x_0) - Z^*(x_0)\right]\right\} - 2\mu \sum_{i=1}^{n} \lambda_i\right] / \partial \lambda_i \qquad [21]$$

where μ is the Lagrange multiplier. The system of equations can be written in terms of the semivariogram function with the following system of equations to be solved:

$$\sum_{j=1}^{n} \lambda_j \gamma(x_i, x_j) + \mu = \gamma(x_0, x_i) \quad \text{for } i = 1 \text{ to } n \qquad [22a]$$

$$\sum_{j=1}^{n} \lambda_j = 1 \qquad [22b]$$

This kriging system defines ordinary point kriging because it gives estimates at points within the landscape. The system of equations is easily solved using Gauss elimination and the solution yields the n weights and the Lagrange multiplier (Webster, 1985; Cooper and Istok, 1988). The minimum estimation variance, or kriging variance, σ_E^2, is obtained from

$$\sigma_E^2 = E\left\{\left[Z(x_0) - Z^*(x_0)\right]^2\right\} = \sum_{i=1}^{n} \lambda_i \gamma(x_i, x_0) + \mu - \gamma(x_0, x_0) \qquad [23]$$

Kriging provides both unbiased estimates (with minimum variance), and a measure of the estimation variance. As such, kriging is superior to other interpolation methods (Oliver, 1987; Webster, 1985).

Soil salinity and related soil properties sometimes show large nugget variances which lead to large estimation variances and erratic contours when mapped using point kriging (Marx and Thompson, 1987). Although a contour map drawn from point estimates is the most accurate map that can be made using point data, local discontinuities can obscure longer range trends in soil salinity. Also, the positions

of the discontinuities depend on the locations of the particular data points, and a shift in the orientation of an observation grid can easily result in a substantially different map. This is an artifact caused primarily by sampling (Burgess and Webster, 1980).

The above shortcomings of point kriging can be avoided by using a procedure known as block kriging. Block kriging can be regarded as a more general kriging method of which point kriging is a special case (Webster, 1985). With point kriging we have assumed that both the estimates and the samples were based on point support. Instead of interpolating between points, block kriging considers a region with its center at x_0. The semivariances between the observed data points and the interpolated point are replaced by the average semivariances between the observed points and all points in the region while the average value of Z over the entire block B, is established by

$$\bar{Z}^*(x_B) = \sum_{i=1}^{n} \lambda_i (x_i) \qquad [24]$$

The block estimation variance is written as

$$\sigma^2_{EB} = \sum_{i=1}^{n} \lambda_i \bar{\gamma}(x_i, x_B) + \mu_B - \bar{\gamma}(x_B, x_B) \qquad [25]$$

where $\gamma(x_i,x_B)$ is the average semivariance between the sample point x_i and all the points within B, and $\gamma(x_B,x_B)$ is the within block variance. Intuitively one would

expect the estimation variance to be reduced using block kriging since a variance reduction is associated with averaging the smaller scale fluctuations of the random variable over space. Depending on the assumptions about the salinity distribution within the landscape, the ordinary kriging equations can be written in terms of the covariance, or the semivariogram. If we assume the weaker intrinsic hypothesis, then the kriging equations must be written in terms of the semivariogram because under the intrinsic hypothesis the variance may not be defined. For this reason we have chosen to use the kriging equations in terms of the semivariogram.

Universal Kriging

In some instances soil salinity and related soil variables may exhibit some non-stationary behavior called drift. In this instance, the expected value of soil salinity is not constant even within small neighborhoods but is a function of position within the landscape. Thus Eq. 1. becomes

$$E[Z(x)]=m(x) \qquad [26]$$

The quantity m(x) represents the drift. Universal kriging is a method which provides an unbiased linear estimator when drift is present. One of the difficulties with using universal kriging is the circular nature of the problem. To determine the drift, the semivariogram must be known, but to determine the semivariogram the drift must be known since the semivariogram is calculated using the residuals (Z(x)-m(x)). In practice the drift is assumed to be of the form

$$m(x) = \sum_{i=0}^{p} a_i f_i(x) \qquad [27]$$

where a_i's are the coefficients and f_i's are known functions which describe the drift. First and second degree ploynomials are usually used for the functions f_i. The semivariogram is assumed to be linear over some restricted neighborhood. By using known functions and a linear semivariogram model the coefficients can be evaluated. In principle, a particular combination of drift function and neighborhood size is selected and a_i's are estimated. If the experimental semivariogram (of the residuals) matches the assumed semivariogram, then the combination is judged to be appropriate. If the match is poor, the neighborhood size or drift function is systematically changed until a match is confirmed. Knowledge of the coefficients is essential to estimate the semivariogram, but not necessary for kriging. The universal kriging estimator is of the same form as Eq. 18, but Eq. 22a becomes

$$\sum_{i=1}^{n} \lambda_i \gamma(x_i, x_j) + \sum_{k=0}^{p} \mu_k f_k(x_i) = \gamma(x_0, x_i), \qquad [28]$$

and Eq. 27 becomes

$$\sum \lambda_i f_j(x_i) = f_j(x_0) \quad \text{for } j = 0, \ldots, p \qquad [29]$$

The kriging variance is given by

$$\sigma^2_{UK} = \sum \lambda_i \gamma(x_i, x_0) + \sum_{k=0}^{p} \mu_k f_k(x_0) \qquad [30]$$

where the μ_k's are the Lagrange multipliers corresponding to the p+1 constraints required to ensure the estimator is unbiased. The universal kriging variance is usually larger than with ordinary kriging because of the uncertainty associated with modeling the drift functions (Webster and Burgess, 1980; Warrick et al., 1986). For a detailed mathematical explanation of universal kriging the reader is referred to Olea (1975) and Olea (1977). Webster and Burgess (1980) explain the limitations of universal kriging in soil science.

5. Co-kriging

Co-kriging, which has been introduced into the soil sciences more recently (Yates and Warrick, 1987), uses two or more regionalized variables simultaneously, and in such a manner that the spatial correlation information from each variable aids in the interpolation process. It is not necessary that both variables have the same sampling locations, which brings to mind at least two motivations for using co-kriging, instead of the simpler ordinary kriging: i) the under sampled problem--where one variable is more costly or more difficult to obtain than the other; and ii) increased apparent sampling density--where one variable is sampled at one set of sites and the second variable is sampled at a different set of sites with only minimal coincident sampling necessary to calculate the cross-semivariogram. Theoretically, if these

regionalized variables are correlated with one another, there should be an overall improvement in the quality of the estimate based on the comparison between the kriging and co-kriging variances. Also, there is often a great potential for reducing the sampling density required for appraising and mapping soil salinity and related soil properties by using an under-sampling strategy.

Under-sampling is often of primary interest in salinity mapping because of the desire to reduce labor costs and the time involved. Chemical analyses of saturated soil extract is very costly and time consuming. If the necessity for taking and analyzing soil samples could be reduced without reducing the quality of the salinity assessment, then salinity surveys over large areas could be made more practical.

When two spatially random variables $Z_u(x)$ and $Z_v(x)$ are sampled in space, and a significant correlation between the two variables exists, the added information due to this correlation can be used to improve the estimate at an unsampled location (interpolation). This procedure is known as co-kriging (Journel and Huijbregts, 1978; Vauclin et al., 1983; Yates and Warrick, 1987; Ahmed and De Maesily, 1987). The added information enters the interpolation through the cross-semivariance function which gives the correlation between two variables as a function of separation distance (**h**). This allows us to use the information about how one variable is distributed in space to help estimate the spatial distribution of the other variable. The cross-semivariance function is analogous to the semivariance function and is a measure of the correlation in space between two regionalized variables. The moment estimator of the cross-semivariance is given by

$$\gamma^{*}_{uv}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z_u(x_i) - Z_u(x_i+h)][Z_v(x_i) - Z_v(x_i+h)] \qquad [31]$$

To properly execute co-kriging, it is necessary to have two semivariograms (one for each variable) and one cross-semivariogram for every pair of correlated regionalized variables in the co-kriging system.

The theory and practice of co-kriging is a logical extension of kriging to situations where two of more variables are spatially interdependent. A co-kriged estimate is a weighted average of the available data with weights chosen so that the estimate is unbiased and has minimum variance, directly analogous to ordinary kriging. The co-kriging estimator is defined as

$$Z^{*}(x_0) = \sum_{i=1}^{k} \lambda_{ui} Z_u(x_i) + \sum_{j=1}^{k'} \lambda_{vj} Z_v(x_j) \qquad [32]$$

where k and k′ are the numbers of samples of $Z_u(x)$ and $Z_v(x)$ used to estimate the unsampled point, x_0, λ_u and λ_v are the associated weighing factors for $Z_u(x)$ and $Z_v(x)$. In a similar manner to ordinary kriging of one variable, the co-kriging estimator is required to be unbiased and of minimum error variance. In order to achieve this, the sum of the λ_{ui}'s must equal one and the sum of the λ_{vi}'s must equal zero.

This procedure is commonly applied to the under-sampled problem, where the variable of interest is costly or difficult to measure, while another related variable is relatively easy or inexpensive to measure. Co-kriging allows an improvement in the

estimation of the variable of primary interest without additional sampling. This may translate into a more efficient sampling scheme.

Although in most cases estimates of only one of the random functions is desired, such as in the under-sampled problem, the complete co-kriging solution will enable estimation of all correlated functions simultaneously (Myers, 1982 and 1984; Carr et al., 1985). In theory it is possible to use an unlimited number of variables, in practice however, three or four is the most that can be handled efficiently.

6. Disjunctive Kriging

Linear kriging requires only the knowledge of the semivariogram or the covariance, so it demands the least amount of information on the stochastic nature of the random function, Z(x). The price to pay is that linear kriging estimates may not be, in general, as precise as other estimates which do demand more information. For example, ordinary kriging yields estimates that have the smallest estimation variance among all unbiased linear estimators. This linear estimator may not be optimal unless the regionalized variable has a multivariate normal probability distribution. Non-linear geostatistics offers tools that can be applied to variables which exhibit non-normal probability distributions.

Sometimes the observed data clearly exhibit non-Gaussian characteristics, and whose log-transforms are also non-Gaussian. These cases have motivated the development of a completely new set of assumptions to deal with non-Gaussian random functions. One of the earliest products was the non-linear estimation

procedure of disjunctive kriging that was originally formulated by Matheron (1976).

Matheron (1976) proposed disjunctive kriging as a simple non-linear alternative which only requires that the bivariate distributions for the variable be known. When a random variable is uni- and bi-variate normally or lognormally distributed, the linear kriging estimator for a known mean is identical to the disjunctive kriging estimator (Rendu, 1980; Journel and Huijbregts, 1978). By adding an additional assumption that the distribution is multi-variate normal, then the linear kriging estimator and the disjunctive kriging estimator are the same as the conditional expectation, defined as the best possible (minimum variance) estimator that can be deduced from the available data.

The disjunctive kriging method requires many more steps and computational capacity to obtain the estimates, compared to simple or ordinary kriging. One assumption implicit in the disjunctive kriging method is that the original variable, $Z(x)$ can be transformed into a bivariate normal variable, $Y(x)$, with a zero mean and unit variance. So, $Z=\phi(Y)$, where ϕ is the Gaussian transform function that is assumed to exist, to be unique and to be invertible. These assumptions make the use of disjunctive kriging more difficult by demanding additional verification of the validity of the assumptions required.

The disjunctive kriging estimator is made up of a series of non-linear functions where each function depends on only one transformed value $Y(x_i)$

$$Z^*_{DK}(x_0) = \sum_{i=1}^{n} f_i[Y(x_i)] \qquad [33]$$

where n is the number of samples; and $f_i[Y(x_i)]$ are the functions to be determined. When the functions, f_i, are linear (i.e., $f_i = w_i Y(x_i)$), the disjunctive kriging technique is the same as ordinary kriging.

Disjunctive kriging can also be used when there are two or more random functions of interest and is called disjunctive co-kriging. The disjunctive co-kriging technique is described by Yates (1986). For more information of disjunctive kriging, readers are referred to Rendu (1980), Yates et al., (1986a and 1986b) and Yates and Yates (1989).

The original developments of disjunctive kriging were done using the assumption of a bi-variate normal distribution for the transformed variable. Later, appropriate transformations have been found for Beta, Gamma, Poisson, binomial, hypergeometric, uniform Rayleigh, chi square, and arcsine distributions (Matheron and Armstrong, 1986; Christakos, 1988).

One important advantage of disjunctive kriging over linear kriging methods is that an estimate of the conditional probability (Matheron, 1976; Journel and Huijbregts, 1978; and Kim et al., 1977) is possible. This is defined as the probability that $Z(x_0)$ be above a specified cutoff or threshold value, z_c, that is $P[Z(x_0) \geq z_c \mid Z(x_i)]$. Conditional probability can play an important role when the user is interested in determining the chances that a variable is above a threshold level. This is a common

question encountered in salinity assessment and land reclamation, where growers are concerned with areas of high salinity risk.

The overall objective when using kriging methods is to obtain a moving average of a property that is distributed in space. For ordinary kriging this amounts to determining the constant weighting coefficients. For disjunctive kriging, on the other hand, unknown functions must be determined that may or may not be linear. When these functions are linear and the random function is multivariate normal, the disjunctive kriging method is the same as the ordinary kriging method and , therefore, ordinary kriging can be considered a special case of the more general disjunctive kriging method.

Consider a second order stationary random function Z(x) that has been sampled over two dimensions at n locations: x_1, x_2, ..., x_n,. It is assumed that Z(x) is spatially correlated and this correlation can be described by a semivariogram under a second order stationarity hypothesis.

The disjunctive kriging method utilizes the autocorrelation function in determining the weighting coefficients for a series of Hermite polynomials. Therefore, the second order stationarity conditions are required so that the variance exists, in which case the autocorrelation function can be written in terms of the semivariogram

$$\rho(\mathbf{h}) = 1 - \gamma(\mathbf{h}) / \gamma(\infty) \qquad [34]$$

where $\rho(\mathbf{h})$ is the autocorrelation function, $\gamma(\mathbf{h})$ is the semivariogram, $\gamma(\mathbf{h})$ is the sill

value of the semivariogram, **h** is the vector distance, and $\gamma(\infty)$ is the variance.

To obtain the disjunctive estimator, the original data must be transformed into a new variable, Y(x), with a standard normal distribution where pairs of sample values are bivariate normal. The function, $\phi[Y(x)]$, which describes this transformation is

$$\phi[Y(x)] = Z(x) = \sum_{k=0}^{\infty} C_k H_k[Y(x)] \qquad [35]$$

where the values for Y(x) are obtained by taking the inverse, $Y(x) = \rho^{-1}[Z(x)]$ and $H_k[Y(x)]$ is a Hermite polynomial of order k. The C_k's are the Hermitian coefficients, which are determined using the properties of orthogonality, and are generally determined using numerical integration, as follows

$$C_k = \frac{1}{k!\sqrt{(2\pi)}} \sum_{i=1}^{J} \omega_i \phi(v_i) H_k(v_i) \exp\left[-v_i^2/2\right] \qquad [36]$$

where v_i and ω_i are the abscissa and weight factors for Hermite integration (Hochstrasser, 1965; Yates and Yates, 1989). If $\phi(s)$ is a standard normal random function, the mean and variance of the data can be found from the coefficients, C_k. The mean value is equal to C_0 and variance of the data is

$$\sum_{k=1}^{\infty} k! C_k^2 \qquad [37]$$

The disjunctive kriging estimator is found from a sum of unknown functions of the transformed sample values, $Y(x_i)$. It is required that each unknown function, $f_i[Y(x_i)]$, depend on only one transformed value, $Y(x_i)$. This gives the disjunctive estimator as

$$Z^*_{DK}(x_0) = \sum_{i=1}^{n} f_i[Y(x_i)] = \sum_{i=1}^{n} \sum_{k=1}^{\infty} f_{ik} H_k[Y(x_i)] \qquad [38]$$

where f_i is the unknown function with respect to the transformed variable, $Y(x_i)$, which is to be determined. In the last part of Eq. 38, the unknown function has been written in terms of a Hermite polynomial with coefficients, f_{ik}.

In an analogous manner to ordinary kriging, an unbiased estimator with minimum estimation variance is sought, which results in the following system of equations

$$Z^*_{DK}(x_0) = \sum_{k=0}^{K} c_k H^*_k[Y(x_0)] \qquad [39]$$

and

$$H^*_k[Y(x_0)] = \sum_{i=1}^{n} b_{ik} H_k[Y(x_i)] \qquad [40]$$

where the series in Eq. 40 has been truncated to k terms and b_{ik} are the disjunctive kriging weights. The $H_k^*[Y(x_0)]$ represents the estimated value of the k^{th} Hermite

polynomial at the estimation site. The sum of these estimates multiplied by the coefficient, C_k (which transforms Y(x) into Z(x)) makes up the disjunctive kriging estimate at x_0. To obtain an estimated value for the Hermite polynomial, the disjunctive kriging weights, b_{ik}, must be found by solving the linear kriging equation for each k

$$\sum_{i=1}^{n} b_{ik}(\rho_{ij})^k = (\rho_{oj})^k; \qquad j=1,2,3,\ldots,n. \qquad [41]$$

When k=0, Eq. 41 represents the unbiased condition, that is, that the sum of the weights equals unity. The disjunctive kriging variance is

$$\sigma^2_{DK} = \sum_{k=1}^{K} k!\, c_k^2 \left[1 - \sum_{i=1}^{n} b_{ik}(\rho_{oi})^k\right]. \qquad [42]$$

One advantage the disjunctive kriging method has over ordinary kriging is that an estimate of the conditional probability that the value at an estimation site is greater than an arbitrary critical value, y_c, can be calculated. This conditional probability is a useful means for determining the risk of various management alternatives. The conditional probability is obtained by defining an indicator variable that is equal to one if $Y(x_i) \geq y_c$ and zero otherwise (Yates et al., 1986b). This allows the conditional probability to be written in terms of the conditional expectation and gives the estimator of the conditional probability as

$$P^{*}_{DK}(x_0) = 1 - G(y_c) + g(y_c) \sum_{k=1}^{K} H_{k-1}(y_c) H^{*}_{k}[Y(x_0)]/k! \qquad [43]$$

where $G(y_c)$ and $g(y_c)$ are the cumulative probability and probability density functions, respectively, for a standard normal variable, and $H_k^*[Y(x_0)]$ is found using Eq. 40. The estimated conditional probability density function, $pdf_{DK}^*(x_0)$, is found by taking the derivative of Eq. 43 with respect to y_c which is

$$pdf^{*}_{DK}(x_0) = g(u) \left\{ 1 + \sum_{k=1}^{K} H_k(u) H^{*}_{k}[Y(x_0)]/k! \right\} \qquad [44]$$

Summary

Geostatistical spatial analysis is an advanced analysis technique and a potentially powerful tool for mapping soil resources and environmental contaminants. Detailed descriptions of the theory and practice of geostatistics are available for the statistician, but usually not to the nonspecialist.

We have presented a five-step methodology to help soil scientists in applying geostatistical spatial analysis to resource mapping. The methodology can be used to map a variety of chemical and physical soil properties and their estimation errors within the landscape. The same is also true for optimizing locations for future sampling, and for estimating the extent and distribution of the measured soil properties for an inventory or design of remedial treatment strategies. For kriging to be worthwhile, data must be collected at short spatial intervals so that they are spatially dependent and provide enough pairs for accurately estimating the semivariograms.

CHAPTER 2

Application of Mapping to Soil Salinity

Introduction

Maintaining an adequate supply of soil water for plant use is a problem when the osmotic pressure measured as electrical conductivity (EC) exceeds the threshold level. Maximizing water use efficiency requires a rational soil and water management system. Knowledge of the salinity distribution as well as the soil hydraulic properties within the region is an important input to a water management system for efficient water uses and prevention of groundwater contamination. The decisions to control this problem depend on the spatial structure between the soil parameters and the assessment of current soil salinity information. The need to assess soil salinity is becoming increasingly important for two reasons: i) the need to increase irrigation efficiency for conserving water and reducing agricultural pollution of water supplies; ii) the imposition of restrictions on the discharge of saline drainage waters from irrigation projects. The variability of soil salinity can affect the reclamation, water management, and productivity, as well as optimization of the cost-benefit relationships.

Soil salinity is a dynamic and spatially variable soil property (Jury, 1985; Webster, 1985; Hajrasuliha et. al, 1980). This is because soil is not a homogenous mass, but rather a heterogenous body of material, and also because of the nature of soil forming processes with distinct boundaries between the soil regions or horizons being rare. Electrical conductivity; EC of soil is a physical parameter as well as the main factor in limiting soil water use and total yield or farm production. The sodium adsorption ratio (SAR) is considered to measure the limiting factor in soil

reclamation. The structure of soil and the solute movement through the profile are influenced by this ratio.

The inherent spatial variability of soil complicates salinity characterization within the landscape (Beckett and Webster, 1971; Warrick and Nielsen, 1980). Numerous samples are generally needed to accurately assess just one field. The intensity with which a soil must be sampled to estimate some characteristic within a given accuracy will depend on the magnitude of the variation within the soil population. The more heterogenous the population, the more intense must be the sampling rate to get a given precision. Considerable variation may be expected for characteristics such as pH, available Phosphorus, exchangeable Potassium, exchangeable Sodium, EC, and permeability. The heterogeneity of the salinity parameters can be studied using statistical and mathematical functions and models (Warrick et al, 1986). The statistics of the correlation, variations, and the analysis of variance give a quantitative description for this variability. Samples close to each other often have similar values, or are autocorrelated (Journel and Huijbregts, 1978; Clark, 1979; Burgess and Webster, 1980; Rendu, 1978). The correlation between variables might be significant or non significant. The mathematical models will take these statistics into account in leaching processes or in transport phenomena.

Soil salinity information can become quickly out-dated as management and edaphic factors change. When the need for repeated measurements and extensive

spatial sampling are met, the expenditure of time, effort and money needed to characterize salinity conditions with conventional soil sampling and laboratory analyses can quickly become prohibitive. Geostatistical methods offer a means for reducing the expected extensive sampling requirement by identifying areas with similar salinities and by establishing predictive relationships between neighboring sampling sites (Rhoades et al., 1989). Introducing geostatistical analysis to soil salinity offers an efficient and accurate means for estimating the number of variables required to obtain salinity information, reduces sampling effort and cost, and determines the variance map that indicates the error source and the existence of extreme levels of salt in the soil.

Correlation coefficients between the variables EC, SAR and clay content have been used in a recent soil resource inventory survey conducted by the Agriculture Research Service and the Soil Conservation Service of the U. S. Department of Agriculture. The objectives were: 1) estimate SAR from EC; 2) determine the effect of subsampling on the variogram and estimator procedure; 3) map the spatial distribution of SAR, EC, and clay content; 4) compare kriging and co-kriging in estimating variance; 5) determine the effect of sample number on the estimation of the variable itself; 6) decide whether additional improvement is likely if other correlations exist; 7) understand the cross-correlation between the random variables and the auxiliary variables.

Experimental Methods

Site description

The South-Fork King River Watershed is located in northern Kings County, California east of the Kings River. A 3885 ha study area in the LeMoore, Hanford, Guersey, and Stratford quadrangles was used. Irrigated agriculture predominates in the area and a cross-section of agricultural crops and management practices are represented. Agricultural operations range from well-managed large corporate farming units to small family farms and abandoned parcels of land.

The area also contains a wide distribution of soil types, ranging from fine to coarse textured soil. The water table in the area is at a depth ranging from less than one to greater than three meters below the soil surface. The quality of the groundwater varies from one to 20 ds/m. Generally there was good quality water available for irrigation, although some management units had to rely on water of lesser quality during part of the growing season. The variety of management and edaphic factors contributed greatly to the variability of the salt content within the landscape.

Soil sampling and instrumental methods

The project could afford to take at most 1000 soil samples in the survey area of 8 by 4.8 km. This translated into about 64 soil sampling sites for each 1,609 m by 1,609 m block (section). To ensure a uniform distribution of sampling sites, a stratified systematic unaligned sampling method (Webster, 1977) was used, which

combines the advantages of a regular grid with randomization. The coordinates of sampling sites were determined before the survey was started. Each section was subdivided into 64 blocks (200 by 200 m), each block being further subdivided into 10 rows and 10 columns. Rows and columns were chosen at random, with the restriction that adjacent rows and columns could not be the same. This procedure ensured that adjacent samples were not aligned, which reduced the risk of obtaining biased estimates of the mean values when periodicities are present at the sampling frequency. Sites located on roads, ponds or buildings were not sampled, nor replaced by others. All soil sampling sites were identified and their latitude and longitude were calculated. Although all data were attempted at each sampling site, some sites were deficient in one or several measurements due to instrument malfunction.

For the purposes of salinity mapping, sample sites must be associated with geographic coordinates; latitude and longitude were used. If rapid methods of salinity measurement are to be used to advantage, a rapid means of determining sample site location must be used. For this purpose the LORAN-C (long range navigation) system used in marine and aviation navigation was employed. The LORAN-C system is operated by the United States Coast Guard. The coverage is good regardless of the terrain and is not limited to line-of-sight transmission.

Soil samples were taken at every sample site at a depth between 10 and 25 cm below the soil surface, using a 6.5 cm diameter soil auger. Each sample was approximately 2500 cm^3 in volume and had a mass of approximately 3000 g. Soil samples were sealed in water vapor tight plastic containers, and weighed using a

portable balance. Samples were then transported to the laboratory where they were air dried, ground, and passed through a 2 mm sieve. The Rhoades and Oster (1986) method was used to extract the soil solution at saturated water content, and to measure the electrical conductivity of the extracted solution as an estimate of the soil salinity.

The traditional method of estimating soil salinity is to measure the electrical conductivity of the saturation extract (EC_e) of a soil sample taken in the field (U. S. Salinity Laboratory, Staff, 1954). The chemical analysis for calcium, sodium, potassium, boron, and magnesium have been included. Sodium absorption ratio was calculated as:

$$SAR = \frac{Na^{+}}{\sqrt{\frac{Ca^{++} + Mg^{++}}{2}}} \qquad (45)$$

Clay content was estimated by hand texturing.

Figures 3-7 indicate the sampling scheme for the three variables, SAR, EC, and clay content, chosen for this experiment. The spatial distribution of SAR, EC, and clay content was determined in the following steps: 1) The variograms and sample semi-variograms were calculated and fitted to theoretical models after validation tests. Cross-semivariograms, and the kriging and co-kriging for mapping purposes were calculated. 2) SAR was selected for the estimation procedure and the effect of sample number on the spatial parameters, because of the expense in determining it in the laboratory compared to clay content or electrical conductivity.

SAR values were put into subfiles containing 901, 700, 500, 300, and 100 samples, with a random generator program. Kriging and co-kriging was applied to each subfile with the 901 samples of EC and clay content. The purpose was to determine the smallest number of SAR samples required to estimate its distribution in the field with the aid of other auxiliary variables. Statistical analysis and normality tests were done to insure the conditions required for using geostatistical procedures. Spatial structure requires the normality distribution assumption; the results in Tables 1 and 2 suggest that this assumption is satisfied by the lognormal transformation LEC for EC and lognormal LSAR for SAR.

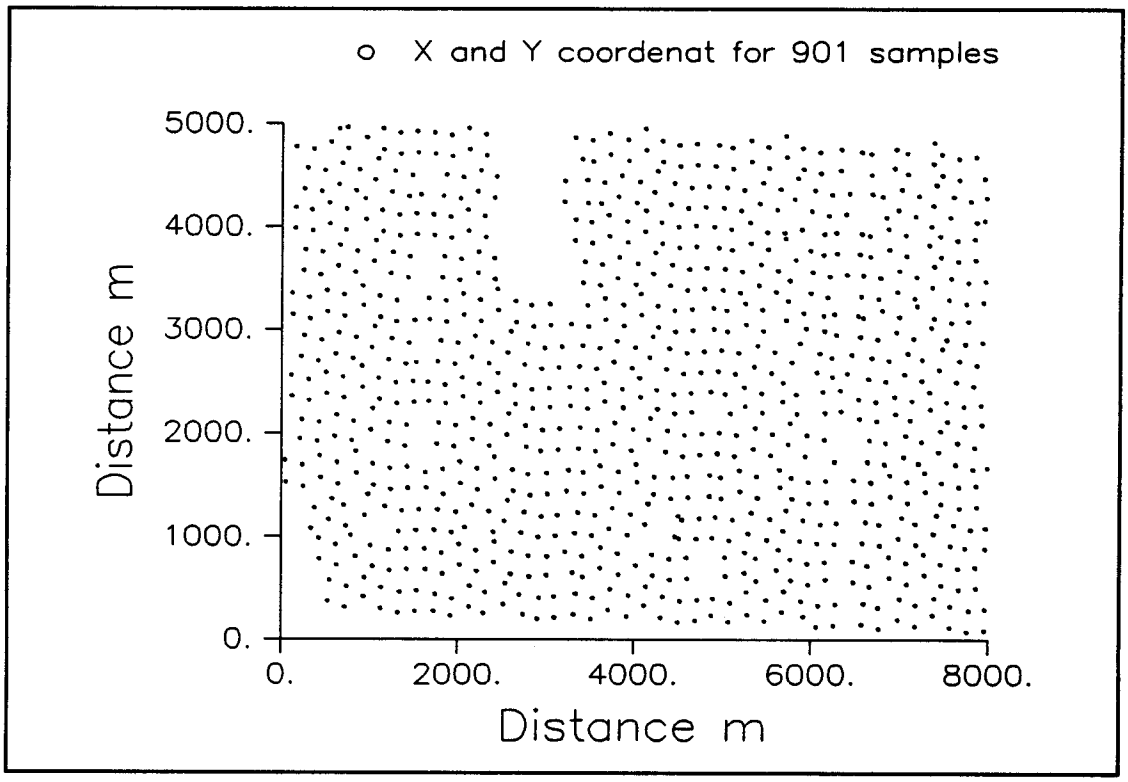


Figure 3. Field sampling scheme for 901 EC, SAR, and caly content.

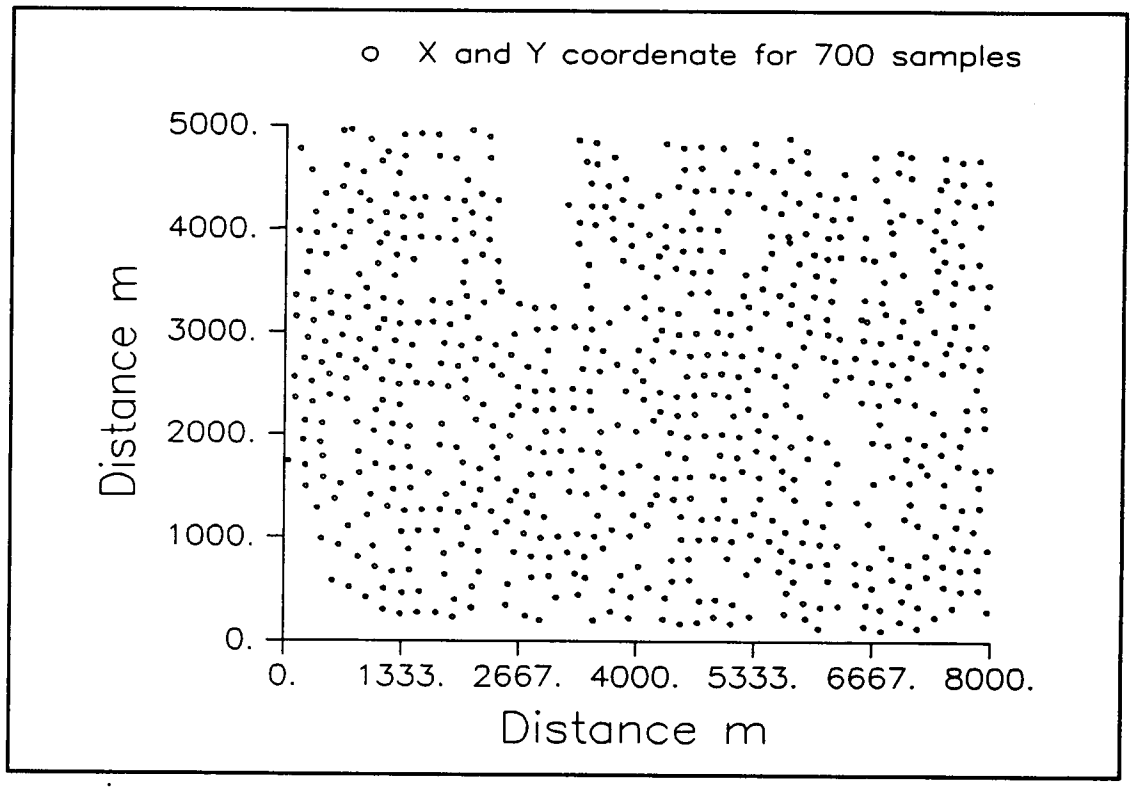


Figure 4. Field sampling scheme for 700 EC, SAR, and clay content.

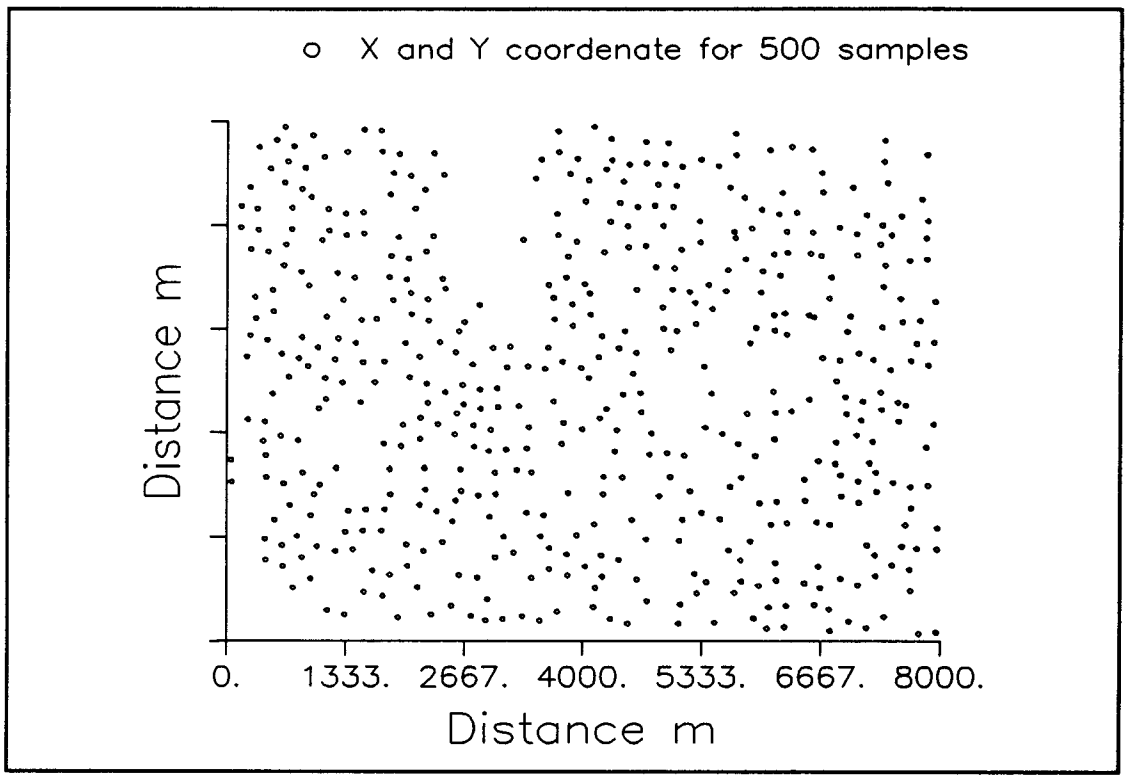


Figure 5. Field sampling scheme for 500 EC, SAR, and clay content.

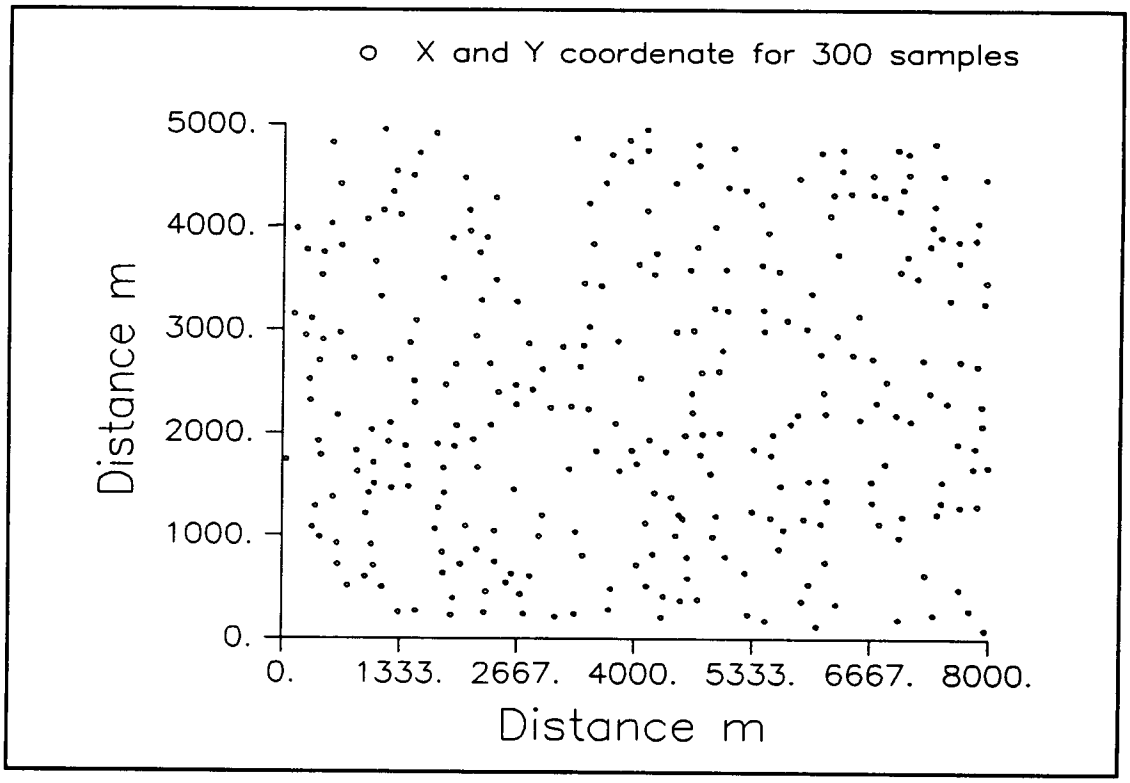


Figure 6. Field sampling scheme for 300 EC, SAR, and clay content.

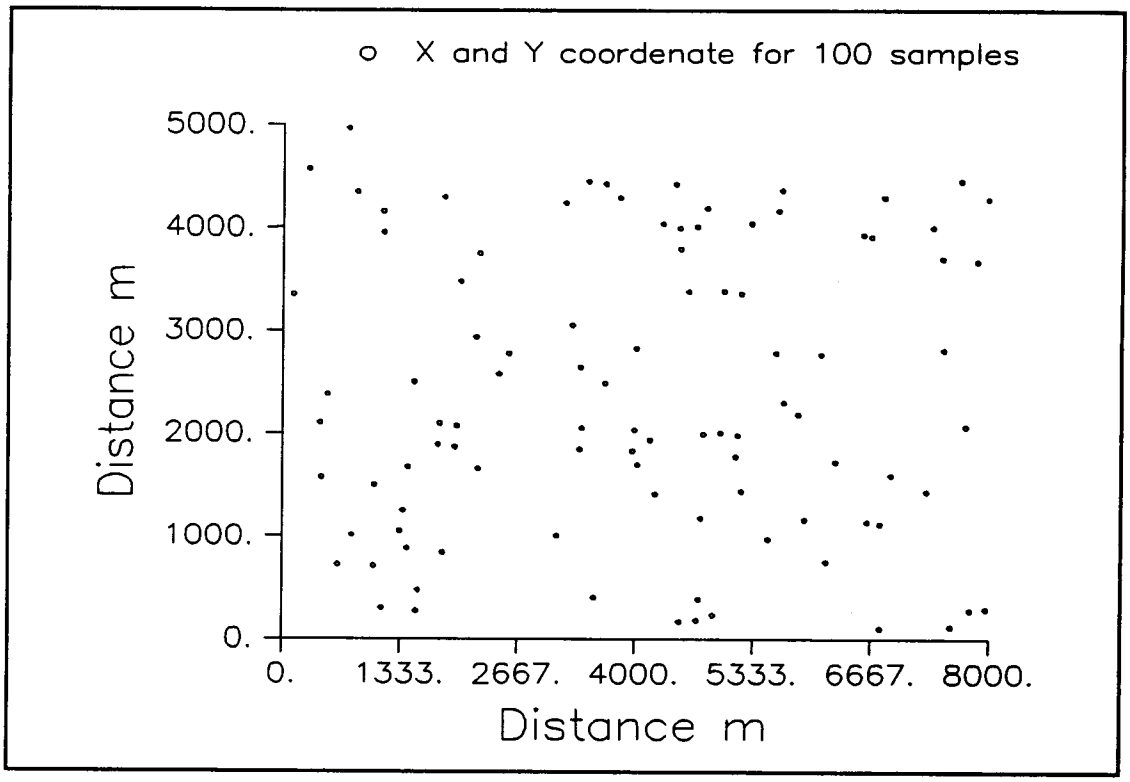


Figure 7. Field sampling scheme for 100 EC, SAR, and clay content.

Statistical description

The statistical parameters for the variables EC, SAR and clay content were determined. Measurement of dispersion, standard deviation, range, and coefficient of variation were calculated. The normality distribution tests KS and W^2 were also performed. Correlation between the variables was also determined and finally, the skeweness and kurtosis tests were done.

The descriptive statistics were calculated using a SAS pc computer program version and Geopack program.

Measures of location

Both mean and median are measures of the location of the center of the distribution. The mean is sensitive to erratic high values. The median however will not be affected as much because it depends only on how many values are above or below, and not on the magnitude. Table 1 shows the values of the two statistics. The uniformity of the transformed data for EC and SAR show that the log transform is less noisy than the actual observations.

Table 1
Statistical Description

Variables	Mean	Median
100	17.320	7.625
300	18.496	6.980
500	22.260	7.420
700	22.663	7.440
901	22.447	7.710
LSAR No.		
100	2.082	2.023
300	2.037	1.943
500	2.073	2.004
700	2.089	2.007
901	2.108	2.042
EC	6.675	3.560
LEC	1.214	1.269
CLAY	22.921	22.000

Measures of dispersion

The sample standard deviation is the most widely used and the best measure of variability. The variance is the average squared difference of the observed values from the means. Since it involves squared differences, the variance is sensitive to erratic high values. The interquartile range is another useful measure of the spread of the observed data. It is the difference between the upper and lower quartiles,

given by:

$$IQR = Q_3 - Q_1$$

Unlike the variance and the standard deviation, the interquartile range does not use the mean as the center of the distribution, and is therefore often preferred if a few erratically highly values strongly influence the mean.

Table 2 shows the results for these three statistics for EC, SAR and clay content. The transformed data, EC and SAR, are more uniform with lower variance.

Table 2

Statistical Description

SAR No.	Standard Deviation	Variance	Range
100	23.346	539.597	1.33760
300	35.648	1266.515	290.060
500	57.329	3280.004	773.900
700	50.853	2582.427	705.750
901	52.635	2767.416	773.900
LSAR No.			
100	1.322	1.730	6.207
300	1.299	1.680	6.721
500	1.389	1.926	7.871
700	1.417	2.007	7.868
901	1.393	1.938	7.961
EC	9.157	83.759	83.500
LEC	1.188	1.411	5.631
CLAY	9.613	92.312	52.000

Measures of shape

1. Coefficient of skewness

The coefficient of skewness is the statistic that measures the symmetry:

$$coefficient\ of\ skewness = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - m)^3}{\sigma^3} \qquad (46)$$

The coefficient of skewness suffers even more than the mean and the variance from sensitivity to erratic high values. Often only the sign and not the magnitude of the coefficient is used to describe the symmetry. A positively skewed histogram has a long tail of high values to the right, making the median less than the mean. If the skewness is close to zero, the histogram is approximately symmetric and the median is close to the mean.

2. Kurtosis

The kurtosis, like the skewness, measures the nature and amount of departure from normality. It describes the shape and peakedness of the frequency distribution.

$$\text{Skew} = m_3/(m_2)^{3/2}$$

$$\text{Kurt} = m_4/(m_2)^2-3$$

where:

$$m_2 = \sum_1^n (x_i - \bar{x})^2/n$$

$$m_3 = \sum_1^n (x_i - \bar{x})^3/n \qquad (47)$$

$$m_4 = \sum_1^n (x_i - \bar{x})^4/n$$

For a normal frequency distribution, both skew and kurt are zero (Snedecor and

Cochran, 1980). They showed that if skewness or kurtosis is higher than a standard tabulated value, the null hypothesis of no significant skew or kurtosis will be rejected at a specified significance level. Thus significant skew and kurtosis indicate a non-normal frequency distribution.

3. Coefficient of variation; CV

This statistic is often used as an alternative to skewness to describe the shape of the distribution. Data with large values of CV tend to vary widely, i.e. values of CV greater than one indicate the presence of erratic values in the data set. It is defined as:

$$CV = (SD/\bar{X}) \tag{48}$$

In percentage form, this becomes:

$$100CV = 100 (SD/\bar{X}) \; percent \tag{49}$$

where SD is the sample standard deviation and $\bar{X}$ is the sample mean. It has been used as a quantitative index of the measured variability (Rogowski, 1972). However the CV does not provide a good insight into the nature of the statistical distribution of measurements like the range.

Warrick and Nielsen (1980) classified the CV into :

1. Low variation; CV $\leq$ 10% .

2. Medium variation 10 % < CV > 60 % .

3. Large variation CV > 60 % .

Table 3 provides a quick summary of variation among the variables measured in this study. The CV for LEC, SAR, and clay content fall in the low variation class. High kurtosis indicates the cumulative distribution curve is shifted to the right, i.e. right tailed.

Table 3

Statistical description

SAR No.	CV	Skewness	Range	Kurtosis
100	0.135	2.462	1.33760	9.926
300	1.927	4.791	290.060	30.264
500	2.575	9.021	773.900	105.004
700	2.244	7.231	705.750	76.583
901	2.345	8.368	773.900	98.357
LSAR No.				
100	0.635	-0.121	6.207	2.483
300	0.637	0.205	6.721	2.740
500	0.670	0.241	7.871	2.658
700	0.678	0.195	7.868	2.605
901	0.661	0.266	7.961	2.672
EC	1.372	3.307	83.500	18.429
LEC	0.979	0.141	5.631	2.147
CLAY	0.419	0.266	52.000	2.672

4. Normal and lognormal probability plots

The estimation tools, kriging and co-kriging, work better if the distribution of data values is close to a Gaussian or normal distribution. The Gaussian distribution is one of many distribution for which a concise mathematical description exists. Also it has properties that favor its use in theoretical approaches to estimation. The normal probability plot, a type of cumulative frequency plot, helps decide this question. Where data sets are not close to this distribution, the lognormal distribution is a good alternative. A variable is distributed lognormally if the distribution of the logarithm of the variable is normal.

The assumption about the distribution has a great impact on the effect of extreme values. There is the danger of casually disregarding extreme deviations in the probability plot.

The two most frequently used methods for establishing normality of the measured values are: 1) visual inspection for skewness (or lack of fit) in the sample frequency distribution using a histogram (Nielsen et al., 1973). 2) examination of the fractile diagram obtained by plotting the measured values on probability paper (Rogowski, 1972; Nielsen et al.,1973; Biggar and Nielsen, 1976). In spite of the intuitive appeal and ease of application, these graphical methods have specific drawn-backs. First, because these methods often rely on visual inspection to establish the appropriateness of the hypothesized distribution, they are subjected to human error. Considerable care is needed in selection of class intervals and scale when using method one and scale when using method two. Secondly, because these graphical

methods are not based on quantitative measurements, objective statistical evaluation of the goodness of fit of the theoretical distribution to the measured data is not possible.

Geopack provides an easy program to use the first test for normality distribution of the data to be used, and SAS program is used for the second method.

Kolmogorov-Smirnov statistics

The Kolmogorov-Smirnov (KS_a) test for goodness of fit is a non parametric test applied to a continuous random variable to judge whether the distribution is normal or lognormal.

$$KS_a = MAX|F_i - \hat{F}_i| \qquad (50)$$

KS_a is the maximum difference between F_i, the true distribution for the variable and $\hat{F}_i$, the hypothesized distribution. If KS_a is greater than a critical value (KS_c), the null hypothesis that F_i is the true distribution should be rejected. KS_c values are given by Rohlf and Sokal (1981) for the 0.1, 0.05, and 0.01 significance level. For intrinsic hypothesis and $n > 30$ the calculated critical value is $KS_c = C/\sqrt{n}$, where C is a constant that changes with the selected level of significance. C values for levels of 0.1, 0.05, and 0.01 as well as the KS_a are shown in Table 4.

The Geopack program has been used to determine KS_a. Table 4 shows that the distribution is more likely to be log normal for SAR and EC for some subfiles, while other subfiles were neither normal nor lognormal, depending on which critical

values are used. Clay content shows a normal distribution.

Cramer Von Mises statistics

This is defined as:

$$W^2 = \int_{-\infty}^{+\infty} [F_n(x) - F(x)]^2 f(x)\, d(x) \qquad (51)$$

where f(x) is the density function corresponding to F(x). W^2 is a weighed integral of the squared deviation between $F_n(x)$ and F(x).

It should be noted from the KS_a and W^2 equations that the values of the test statistics are likely to be small when the hypothesized distribution F(x), is correctly specified; larger values of KS_a and W^2 suggest the possibility that the hypothesized distribution is incorrect. A SAS program was used to determine W^2. The results in Table 4 suggest the lognormal distribution for most LSAR subfiles, and LEC and clay content are normally distributed without transformation. The closer the test value is to unity, the higher is the P-value for that data to have a normal distribution.

With either test parameter, our judgement is that SAR and EC are lognormally distributed and the clay content is normally distributed.

Table 4

Kolmogorov-Smirnov test

No.	KS_a		$KS_{c‡}$			W^2	
	SAR	LSAR	0.1	0.05	0.01	SAR	LSAR
100	0.223	0.057	0.080	0.089	0.103	0.697	0.979
300	0.303	0.050	0.047	0.051	0.060	0.481	0.976
500	0.349	0.037	0.036	0.040	0.046	0.351	0.976
700	0.329	0.030	0.030	0.034	0.039	0.426	0.977
901	0.337	0.036	0.027	0.030	0.036	0.389	0.978
CLAY.901	0.058		0.027	0.030	0.034	0.973	
EC.901	0.243		0.027	0.030	0.034	0.653	
LEC.901	0.075		0.027	0.030	0.034	0.957	

‡ is the probability of type I error (α)

Linear correlation between variables

Correlation is the measure of how closely two variables are associated. The correlation coefficient, r, of X_1 and X_2 is given by:

$$r = \left(\sum X_1 X_2\right) / \left(\sum X_1^2 X_2^2\right)^{1/2} \qquad (52)$$

$$\text{where } X_1 = X_1 - \bar{X}_1^2$$

$$X_2 = X_2 - \bar{x}_2^2$$

Calculated r values can be compared with critical values (Snedecor and Cochran, 1980; Rohlf and Sokal, 1981) to test the null hypothesis of no correlation (E(r) = 0) at a chosen level of significance.

Three patterns can be observed on a scatterplot, the variables are positively

correlated, negatively correlated, or uncorrelated. Two variables are positively correlated if the larger values of one variable tend to be associated with larger values of the other variable, and similarly with the smaller values of each variable. For example the porosity and permeability are positively correlated. For negatively correlated variables, the larger value of one is associated with smaller values of the other. Finally, the increase in one variable may have no effect on the other, these are uncorrelated variables. Presence of extreme pairs has an effect on the correlation coefficient.

The correlation coefficient is a measure of how close the observed values fall on a straight line. If the correlation coefficient $r = 1$ the scatterplot will be a straight line with positive slope; if $r = -1$ the scatterplot will have a negative slope. If the $|r| < 1$ the scatterplot appears as a cloud of points that becomes fatter and more diffuse as $|r|$ deceases from 1 to 0.

The r value provides a measure of the linear relationship between two variables. If the two variables are not linearly related, the correlation coefficient may be a very poor summary statistic. So the rank correlation coefficient can be useful to identify the linearity between the two variables.

Tables 5 and 6, calculated with a SAS program, show the correlation between the variables. LEC has highly significant values of correlation with LSAR. LSAR has less significant values, especially with clay content, which is a negative correlation. Strong significant positive correlation is shown between the LEC and LSAR for different subset files as well as for the LEC and clay content. Correlation

between clay content and LSAR was very small and negative in most subset files, but was significant.

The correlation coefficients in Tables 5 and 6 illustrate a major role in calculating the cross-semivariogram as indicated in Figures 8-15. Highly correlated variables allow distinguishing theoretical models as in the LEC/LSAR cross-semivariogram, but poor correlation raises some difficulties in fitting the theoretical model. All this has an effect on kriging variance or co-kriging estimation (Figure 17).

Table 5

Correlation coefficient for 901 data

Variables	CLAY	EC	SAR	LEC	LSAR
CLAY	1.000	0.059	-0.107	0.118	-0.036
EC		1.000	0.568	0.799	0.673
SAR			1.000	0.465	0.614
LEC				1.000	0.828
LSAR					1.000

Table 6

Correlation coefficient for subfiles

Variables	LSAR				
	901	700	500	300	100
CLAY	-0.036	-0.051	-0.061	-0.017	0.132
EC	0.673	0.667	0.718	0.685	0.686
SAR	0.614	0.641	0.588	0.704	0.801
LEC	0.828	0.831	0.842	0.821	0.839
LSAR	1.000	1.000	1.000	1.000	1.000

Spatial Statistics

This technique is used when data are spatially dependent. Data are spatially dependent when a neighboring observation gives information on the value or magnitude of a parameter. In classical statistics such as regression analysis, independence of the sample data is assumed.

Geostatistics methods measure the relationships between distance and variance. This is a preferred method of analyzing spatially dependent data: 1) it assures that the observed values are returned by the interpolation method; 2) it is an unbiased estimation procedure; 3) it provides an estimate of the variance of the interpolated value, which is the minimum among all other estimators; 4) if the kriging error is normally distributed, it can be used to put confidence limits on average concentrations of individual blocks of land.

Fit to a Gaussian distribution is important for two reasons: 1) The pragmatic

one which is that all prediction, estimation, and distribution theory is relatively straightforward. 2) The asymptotic one where the net result of many small-order effects is approximately Gaussian (central limit theorem). When data are multivariate normally distributed the ordinary kriging is known to be statistically optimal. Similarly, kriging is generally considered less sensitive for extreme values. Lognormal kriging is more resistant against outliers, but it is extremely sensitive to slight fluctuations and to error in estimating the sill of the variogram at the lags, which determines the bias correction factors (Armstrong and Boufassa, 1988).

Stationarity

When a parameter is homogeneous over a study area, observed values $Z(x_i)$ and $Z(x_i+h)$ can be considered realizations of the same random variable with a particular distribution function. The assumption of homogeneity avoids the problem that usually only one realization of a random function is available for a given value of x_i. An understanding of stationary leads to a greater appreciation of the semi-variogram and spatial autocorrelation.

Stationarity is defined through the first and second order moments of the observed random function, and the degree of stationarity corresponds to the particular moments that remain invariant across a study area.

I- First order moment

$$E\{Z(x)\} = m(x)$$

II- Second order moment

1) Variance of the random variable Z(x):

$$VAR\{Z(x)\} = E[Z(X)-m(x)]^2$$

2) Covariance:

$$C(x_1,x_2) = E\{Z(x_1)- m(x_1)][Z(x_2) - m(x_2)]\}$$

3) Semivariogram:

$$\gamma(x_1,x_2) = VAR\{Z(x_1) - Z(x_2)\}/2$$

III- Second order stationarity requires the following:

1) $E\{Z(x)\} = m(x)$. Does not depend upon x.

2) The covariance depends only on separation distance h;

$$C(h) = \Sigma\{Z(x)*Z(x+h)\} - m^2 \text{ for all } x$$

The vector h may be one to three dimensional.

If the covariance C(h) is stationary, the variance and the semivariogram are also stationary:

$$C(0) = E\{[Z(x)-m]^2\} = VAR\ \{Z(x)\}$$

$$\gamma(h) = E\{[Z(x+h)-Z(x)]^2\}/2$$

$$= C(0) - C(h)$$

The correlogram is:

$$\rho(h) = C(h)/C(0)$$

$$= 1 - \gamma(h)/C(0)$$

Under the second order of stationarity, the semivariogram and the covariance are alternative measures of spatial autocorrelation.

The intrinsic hypothesis requires that the expected values of the first moment and the semivariogram are invariant with respect to x. By not requiring stationarity of the covariance and therefore the existence of a finite variance C(0) of the random function, the intrinsic hypothesis is sufficient for most geostatistical problems.

Transformation

Assumptions about variance stability, normality and linearity need to be satisfied in order to make inferences. If these assumptions are violated, the underlying inferential basis is eroded. The deviation from the assumption of normality in regionalized variable theory has been a criticism of kriging (Henley, 1981). Transformation to normality prior to geostatistical analysis results in a nonlinear function of the original data, so that kriging estimates may not be made with minimum estimation variance and without bias (Trangmar et al., 1985). Lognormal distributions have been widely used in soil science (van der Zaag et al., 1984; Yost et al., 1982a, b; Trangmar et al., 1985). It simply involves computing semivariograms and kriging natural log transformed values of the original data using the same procedures as for simple linear kriging. The estimation of log-kriged values is smaller than values of the original data (Journel and Huijbregts, 1978). This approach has been found to be practical (Rendu, 1979; Yost et al., 1982b).

Transformed SAR and EC values such as range, standard deviation, skewness, and kurosis are lower (Tables 1-5).

Variogram

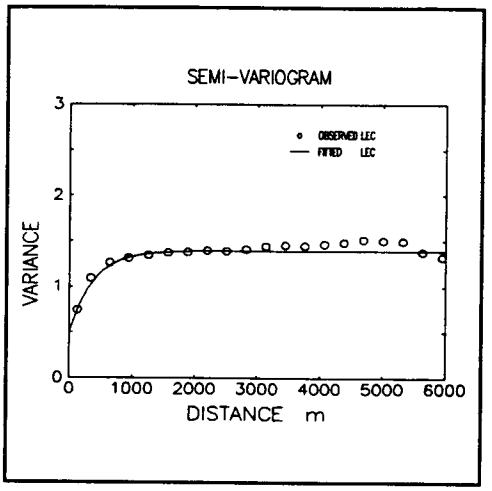
The quantity $\gamma(h)$, a function of the increment h_1 - h_2, is called a semivariogram by Matheron (1963) or structure function by Kolmogorov (1941) in physics and by Jowett (1952) in time series. In Matheron's geostatistics and in Gandin's analysis the variogram is used to define coefficients and weighting factors, λ_i, in an optimal linear predictor. Hence, estimating and modeling the semivariogram is a very important part of kriging that can be difficult if the data are sparse or outliers are present. Kriging depends on an accurate variogram for estimating weights for interpolation. Using incorrect variograms can lead to unfavorable effects on the precision of the kriging estimate, even to a negative kriging variance. Krige and Magri (1982) reported a study on the effect of the outliers. Cressie and Hawkins (1980) and Hawkins and Cressie (1984) studied several estimators of $\gamma(h)$ and found they were less affected by outliers than C(h). If the variable is anisotropic, the semivariogram will be anisotropic.

Figures 8-15 represent semivariograms and cross semivariograms for LEC, LSAR, and clay content variables. The first set includes 10 lag points out of 20. The second shows the 20 lag points used to create spatial structure. Each lag has 200 m distance. Table 7 includes all the semivariogram and cross semivariogram coefficients as well as the theoretical fitted model type.

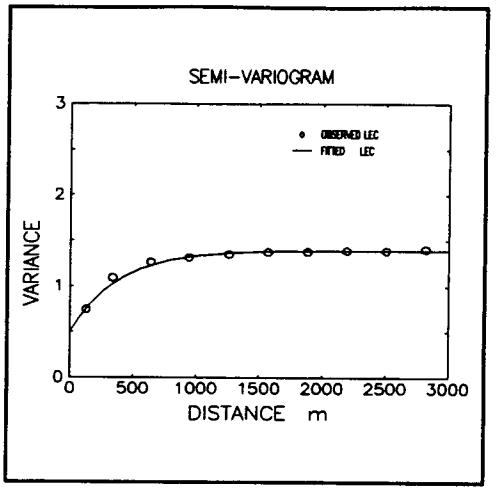
Table 7

Variogram coefficients

Variables	No.	Model	Nugget	Sill - Nugget	Range
LSAR \ LSAR	100	EXP.	1.100	1.160	3760.0
LSAR \ LSAR	300	EXP.	0.380	1.210	1500.0
LSAR \ LSAR	500	EXP.	1.180	0.689	1060.0
LSAR \ LSAR	700	EXP.	0.560	1.410	700.00
LSAR \ LSAR	901	EXP.	1.000	0.717	423.00
LSAR \ LEC	100	SHP.	0.500	0.850	1300.0
LSAR \ LEC	300	SHP.	0.200	0.950	1750.0
LSAR \ LEC	500	EXP.	0.470	0.937	430.00
LSAR \ LEC	700	EXP.	0.590	0.811	800.00
LSAR \ LEC	901	EXP.	0.167	1.127	226.55
LSAR \ CLAY	100	EXP.	0.134	1.550	300.00
LSAR \ CLAY	300	SHP.	0.001	0.710	650.00
LSAR \ CLAY	500	EXP.	1.140	-0.43	1300.0
LSAR \ CLAY	700	SHP.	1.100	-0.780	1350.0
LSAR \ CLAY	901	EXP.	1.500	-1.150	600.00
CLAY \ LEC	901	SHP.	1.396	1.000	1500.0
CLAY \ CLAY	901	SHP.	36.90	54.20	1500.0
LEC \ LEC	901	EXP.	0.990	0.490	680.00

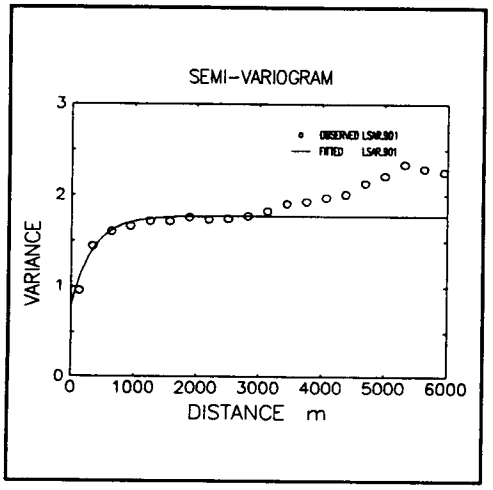


A

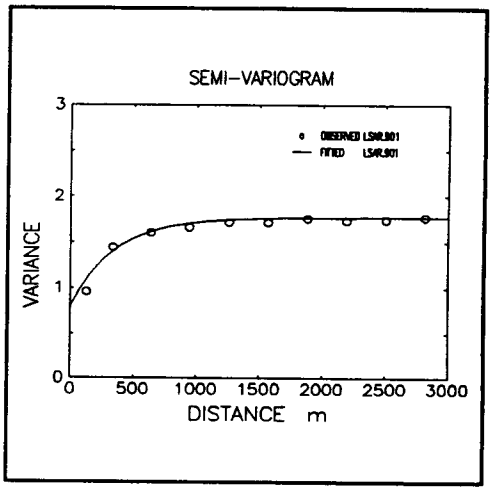


B

Figure 8.A and 8.B Exponential model fitted to LEC.901

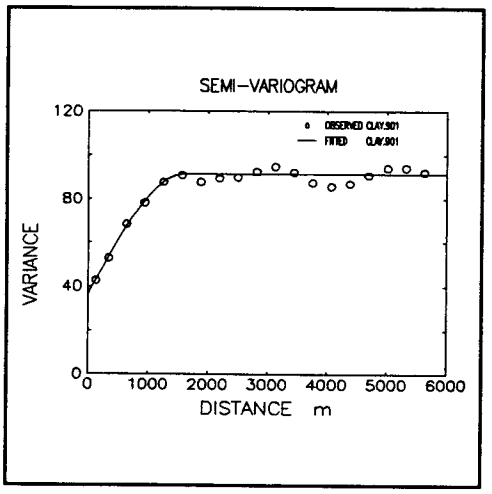


A

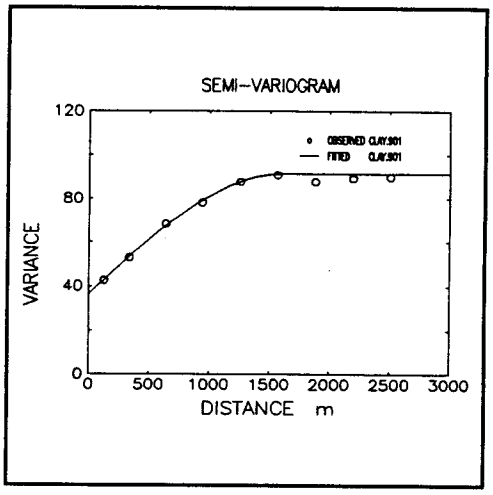


B

Figure 9.A and 9.B Exponential model fitted to LSAR.901

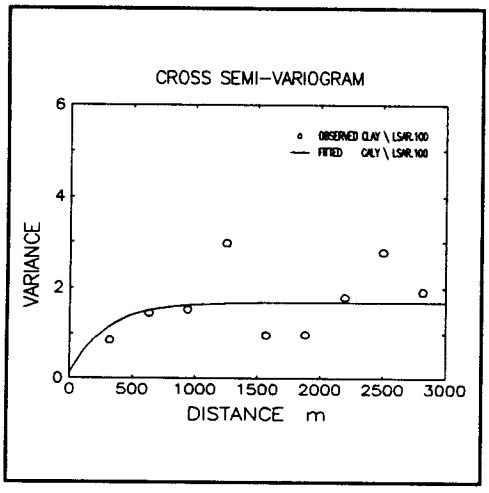


A

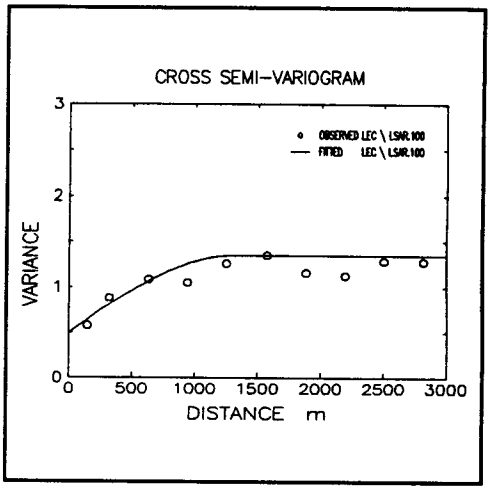


B

Figure 10.A and 10.B Spherical model fitted to clay.901



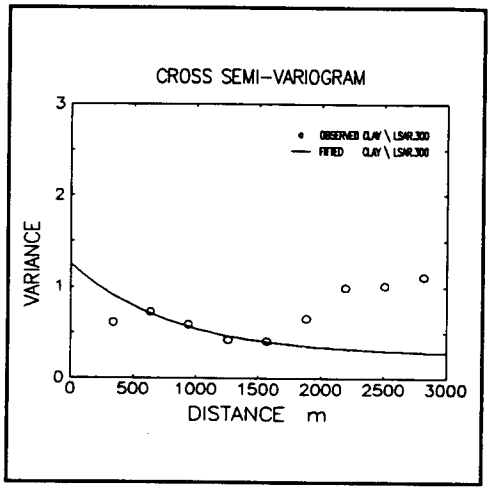
A

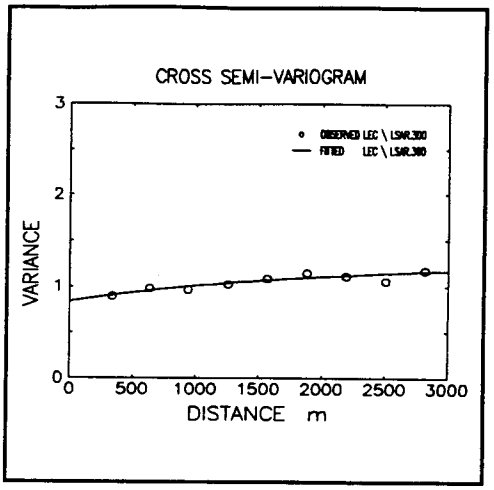


B

Figure 11.A Exponential model fitted to clay content and LSAR.100

Figure 11.B Spherical model fitted to LEC and LSAR.100

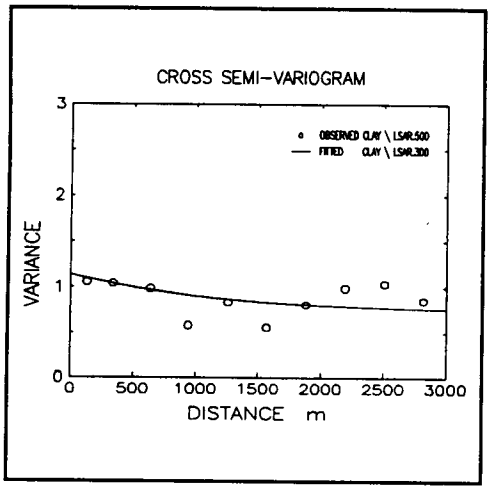


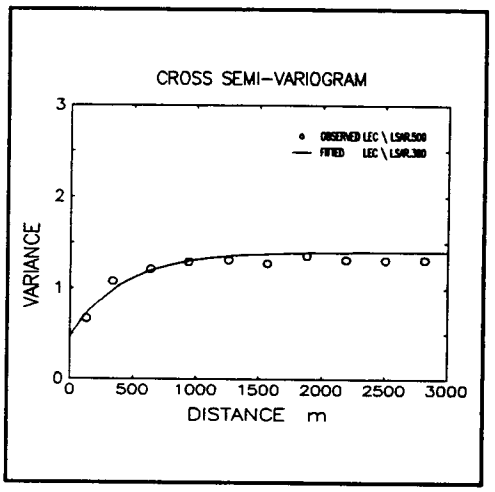


A B

Figure 12.A Spherical model fitted to clay content and LSAR.300

Figure 12.B Spherical model fitted to LEC and LSAR.300

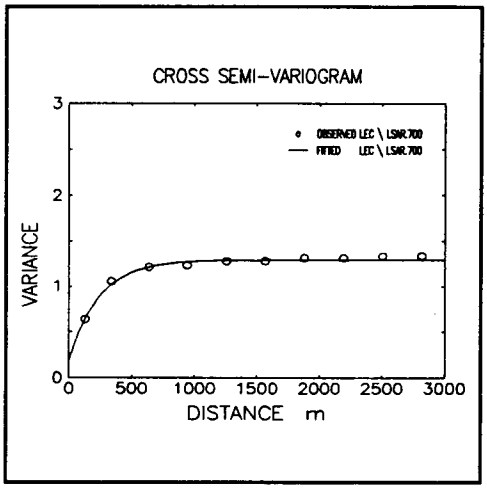


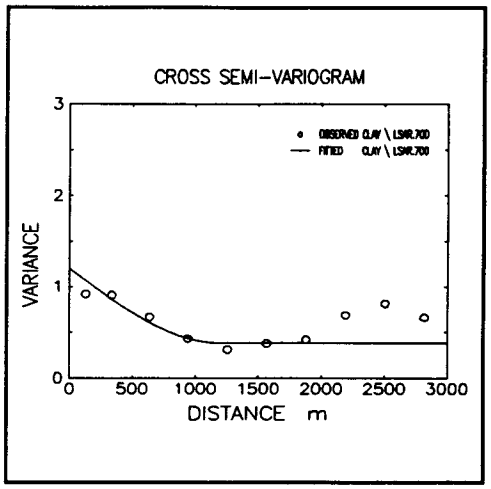


A B

Figure 13.A Exponential model fitted to clay and LSAR.500

Figure 13.B Exponential model fitted to LEC and LSAR.500

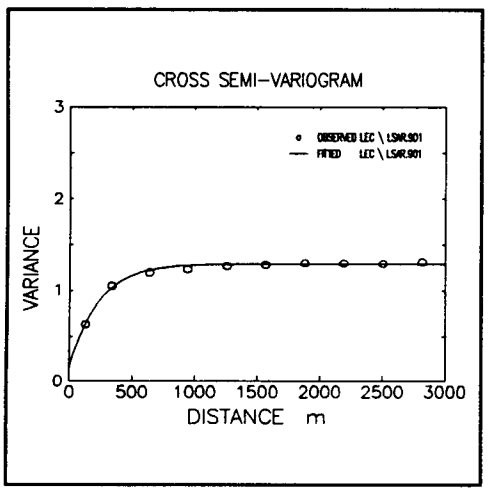


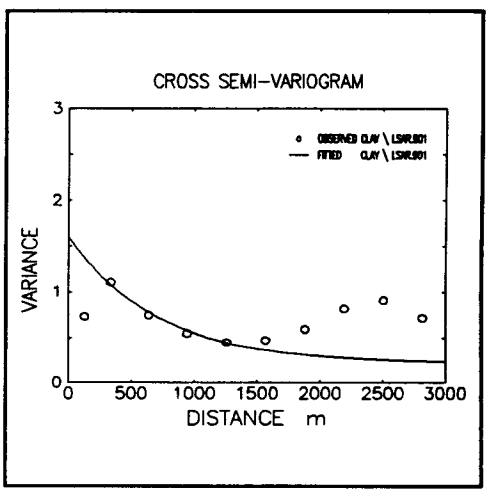


A B

Figure 14.A Exponential model fitted to LEC and LSAR.700

Figure 14.B Spherical model fitted to clay content and LSAR.700





A B

Figure 15.A Exponential fitted model to LEC and LSAR.901

Figure 15.B Exponentional fitted model to clay content and LSAR.901

Nugget effect

Sampling at a given site more than once would yield different measurements because of low precision in the measurement or fine scale variation in the phenomenon under study. This would depend on: 1) geology of the site, 2) time and duration of measurement, 3) the order and the border limit of the sampling area, 4) the scale of sampling relative to the geographic scale of variation in the phenomenon.

Journel and Huijbregts (1978) provided a general definition of the nugget effect as the sum of all sources of variation with ranges much smaller than the distances between samples. The nugget effect is the discontinuity at the origin in the semivariogram model. Since we are interested in the local behavior of the Z(X) field, it follows that we are most interested in γ (h) over short distances. If the nugget is constant over the entire distance, then we have a "pure nugget" or Z(X) is a white noise process. In this case there is no spatial structure and we interpolate all the places identically where there is no observation. One can conclude that the samples are spatially independent.

The nugget model is:

$$\gamma(0)=0 \quad h \leq a$$

$$\gamma(h)=C_o \quad h \rangle a$$

Ways to reduce the nugget effect include better sampling control, reexamining the data for sources of error, and applying adjustments and correction factors.

Range

As the separation distance between pairs increases, the corresponding variogram value will also generally increase. However, when the increase in the separation distance no longer causes a corresponding increase in the average squared difference between pairs of values, the variogram reaches a plateau. The distance at which the variogram reaches this plateau is called the range.

The range of correlation is a distance within which all measurements are correlated. Because of this correlation, a sample is representative of a neighborhood measured by the range of correlation. This phenomena can be useful to quantify the range of the influence of a contaminated region.

Sill

The plateau that the variogram reaches at the range is called the sill. The shape of the semivariogram and the size of the nugget are affected by the support of the measurements. The support is the shape, volume, and orientation of individual samples or measurements. If the support changes, the semivariogram will change.

Estimating the variogram

There are several ways to estimate the variogram:

1) Nonparametric estimation methods or the arithmetic mean of the squared differences (Matheron, 1963).

$$\hat{\gamma}(h) = \frac{1}{2n(h)}\sum_{i=1}^{n}(h)[u(x_i)-u(x_i+h)]^2 \qquad (53)$$

where n(h) is the number of data pairs separated by the lag h. Omer (1984) indicated the above equation is the optimal estimator when $[u(x_i),u(x_i+h)]$ is binormally distributed and the observations $[u(x_i),u(x_j)]$, $(i=1,n;j=1,n;i\neq j)$ are all uncorrelated. This ideal case is rarely encountered in practice, and generally deviations from this ideal case will occur.

Three main deviation types were identified (Omer, 1984). i) distributional deviations where $[u(x),u(x+h)]$ can not be properly represented by a binormal distribution. ii) Sampling deviations where $[u(x_i);i=1,n]$ are not sampled in a randomized way, and the sampling is biased. iii) Outlier deviation where the estimator might be sensitive to the presence of errors. In general, $[u(x_i),u(x_j)]$ are not totally uncorrelated independent variables.

Several other estimators for the variogram were suggested (Dowd, 1982, 1984).

Cressie and Hawkins (1980) suggested a more familiar estimator:

$$\hat{\gamma}(h) = \frac{1}{2[0.47+0.494/n(h)]}\left[\frac{1}{n(h)}\sum_{i=1}^{n}(h)(|u(x_i)-u(x_i+h|)^{1}/2\right]^4 \qquad (54)$$

which is based on normal distribution assumptions. Omer (1984) showed that the Matheron estimator is superior to Equation 54 in the ideal case i.e. no distributional

or sampling deviations occur. Another robust estimator suggested by Omer (1984) is:

$$\hat{\gamma}(h) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ij}[u(x_i)-u(x_j)]^2 \qquad (55)$$

where α_{ij} ($i \neq j$) are weights assigned according to the relative location of two observations in the distribution of values forming the domain of h. This estimator was found to be superior to the Matheron estimator only when distributional and sampling deviations occur simultaneously Omer (1984).

2) Parametric estimation methods

Kriging and co-kriging require the estimated variogram to be fitted to a theoretical model of a positive definite type. Given the structure of a model, estimation of the parameters of the model from the available data is a well-defined algorithmic problem. The most advanced parametric estimation methods are the maximum likelihood (ML), restricted maximum likelihood (RML), minimum variance, unbiased quadratic, and weighed least squares procedures. Each of these methods has been presented and evaluated by Kitanidis (1983) Kitanidis and Lane (1985) and Hoeksema and Kitanidis (1985). When the data are Gaussian distributed or transformed, the ML are known to be asymptotically unbiased with minimum variance, consistent, and normally distributed with a covariance matrix given by the inverse of the Fisher information matrix. In the case of non-normal data the procedure can be viewed as a fitting process based on minimization of a

weighed sum of squares of prediction errors (Kitanidis, 1985).

The variogram $\gamma(h)$ can be defined as

$$\gamma(h) = (1/2)\mathrm{Var}[Z(x+h)-Z(x)]^2 \tag{56}$$

under the zero drift assumption $E[Z(x+h)\} = E[(Zx)]$ so the above equation becomes.

$$\gamma(h) = E[Z(x+h)-Z(x)]^2 \tag{57}$$

An estimate of γ is given by

$$\gamma^*(h) = [\frac{1}{2n(h)}]\sum_{n=1}^{n(h)} [(x_i+h)-Z(x_i)]^2 \tag{58}$$

where n(h) is the number of pairs separated by a distance h.

The valid choice of a certain variogram is restricted such that the negative of $\gamma(h)$ is a positive-definite function (Journel and Huijbregts, 1978; Armstrong and Jabin, 1981).

There are several types of semivariograms models with sill:

exponential

$$\gamma(h) = 1-\exp(-h/a) \tag{59}$$

spherical

$$\begin{aligned} \gamma(h) &= C_o+C[(3/2)(h/a)-(1/2)(h/a)^3] \quad 0<h<a \\ &= C_o+C \qquad\qquad\qquad\qquad\qquad h \geq a \end{aligned} \tag{60}$$

Guassian

$$\gamma(h) = 1-\exp(-h^2/a^2) \tag{61}$$

The ones without a sill are a power

$$\gamma(h) = h^{\alpha} \qquad 0<\alpha<2 \tag{62}$$

and logarithmic model

$$\gamma(h) = \log(h) \tag{63}$$

The linear model with a finite sill is

$$\begin{aligned} \gamma(h) &= C_o+C(h/a) \qquad h<a \\ &= C_o+C \qquad h\geq a \end{aligned} \tag{64}$$

The variogram that is a function only of the separation distance (h) is called isotropic, and the variogram which is a function of direction is called anisotropic.

Selection of the final model from the candidates should be carried out by a systematic model discrimination procedure, such as that which minimizes the Akaike information criterion (AIC). Minimization of sum of squares of error; SSQ; cross validation test and the best-fit model to the experimental variogram were the main procedures to select the theoretical model.

The cross variogram and the variogram must be such that the co-kriging matrix formed with them is positive-definite, (Dunn, 1983; Myers, 1984a). This means the cross semivariogram function should satisfy the Cauchy-Schwartz inequality; $|\gamma_{ij}(h)| \leq |\gamma_{ii}(h)\gamma_{jj}(h)|$.

Measured variograms

The variogram represents the average rate of change of a property with distance and its shape describes the pattern of the spatial variation in terms of

magnitude, scale and general form. Figures 8-15 as well as Table 7 show the coefficients of the variogram and the theoretical model used in this study. The analysis of spatial dependence using semivariograms has contributed to our understanding of variability of salinity management and interpretation. The interpretation focuses on the range of influence, i.e. the size of the area in which variables (SAR) are statistically associated with the variable in the center of the area. This concept provides an estimate of the minimum distance required to designate the area of different levels of salt hazard or the reclamation area as well as the sample spacing or design.

Some semivariogram applications are:

1) Isotropic and anisotropic variation

Soil properties are isotropic if they vary in a similar manner in all directions, in which case the semivariogram depends only on the distance between samples, **h**. Geometrical anisotropy occurs when variance for a given distance **h** in one direction is equivalent to variations for a distance **kh** in another direction. The anisotropy ratio, **k**, indicates the relative size of directional differences in variation.

2) Trends

Many regionalized variables do not vary randomly, but show local trends or components of broader regional trends. Regionalized trends are indicated by semivariograms that increase with distance of sample separation and either do not approach a sill (Gajem et al., 1981) or have a sill which considerably exceeds the general variance S^2 (Bresler et al., 1984).

3) Periodic phenomena

Periodicity of parent material deposition and repetition of land form sequences are often quoted as sources of soil variation. This is expressed in semivariograms as a "hole effect" which is indicative of non monotonic growth of the semivariogram with distance (Journel and Huijbregts, 1978). The hole effect can appear in models with or without sills. This could occur for clay content as a continuous deposition of different mineral types or different mineralization. Its characteristic is a succession of rich and poor zones (David, 1977).

4) Management effect on salinity distribution and reclamation

Management practices may considerably alter the inherent spatial structure of soil properties, as well as the reclamation process. Directional application of water, as in furrow irrigation, may introduce anisotropy of soil moisture content (Gajem et al., 1981) and will affect the EC distribution in the field. Non-uniform fertilizer application to the farmland leads to heterogeneity of the salinity.

Validation Tests

Once the parameters of the model have been estimated, the validity of the model should be tested. Several tests may be conducted:

1) The cross validation test or Jackknife test, (Gambolati and Volpi, 1979). In this test two conditions must be meet i) there is no systematic error i.e. mean maximum error ME of the reduced error vector

$$RE(x_i) = [u^{\star}(x_i) - u(x_i)] / Var[u^{\star}(x_i) - u(x_i)]^{1/2} \text{ must be zero:}$$

$$ME = \frac{1}{n}\sum_{n=1}^{n} RE(x_i) \simeq 0$$
$$ME = \frac{1}{n}\sum_{n=1}^{n} (Z_i^* - Z_i) \qquad (65)$$

ii) The kriging variance $Var[u^{\star}(x_i)-u(x_i)]$ must be consistent with the corresponding error $[u^{\star}(x_i)-u(x_i)]$. The variance is expected to be close to one. Thus the mean square reduced error MRS is required to be unity:

$$MRE = [\frac{1}{n}\sum_{i=1}^{n} RE(x_i)^2]^1/2 \simeq 1 \qquad (66)$$

2) Mean square error; MSE

$$MSE = [\frac{1}{n}\sum_{i=1}^{n} [u^{\star}(x_i)-u(x_i)^2]^1/2 \qquad (67)$$

Figure 16 shows the mean error squre values for kriging and co-kriging estimation for LSAR subfiles.

3. Average kriging variance; AKV

$$AKV = \frac{i}{n}\sum_{i=1}^{n} \sigma_k^2 \qquad (68)$$

Table 9 indicates the results of the average kriging values. Figure 17 suggests that co-kriging has smaller AVK than kriging.

4) Reduced mean error; RME:

$$RME = \frac{1}{n}\sum_{i=1}^{n}(Z_i^* - Z_i) \;/\; \sigma_k \tag{69}$$

5) Reduced variance error; RVE:

$$RVE = VAR[Z^*(x) - Z(x) \;/\; \sigma_k] \tag{70}$$

Table 8 shows the values for reduced mean error and reduced variance obtained from one cross-validation technique.

Another test is a transformation of the measured residual vector $\epsilon(x) = u(x) - m(x)$ into a vector of uncorrelated residuals y(x) and analyzing y(x) (Kitanidis and Vomvoris, 1983). It does not lead to a test of hypotheses in a strong statistical sense, but it does provide a mechanism for ascertaining whether the variogram model reproduces the characteristics of the data. For more detail refer to Davis (1987), Chung (1984), and Gray and Schucany (1972).

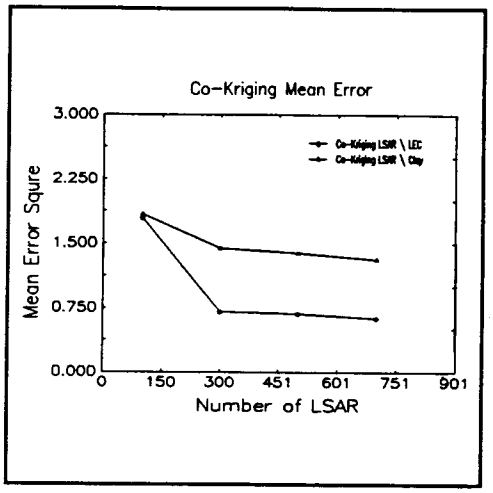


Figure 16 Variance difference between kriging and co-kriging estimation

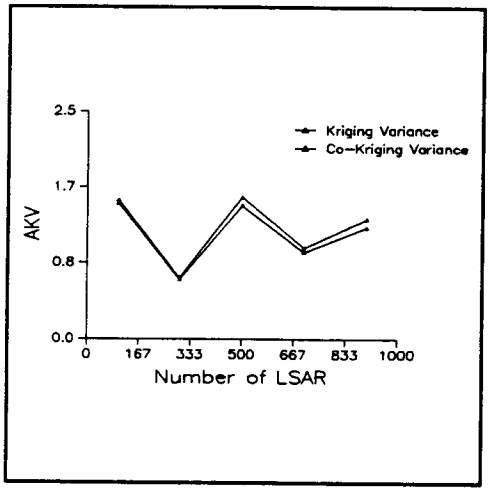


Figure 17 The effect of sampling number on the average kriging variance by using kriging and co-kriging estimation.

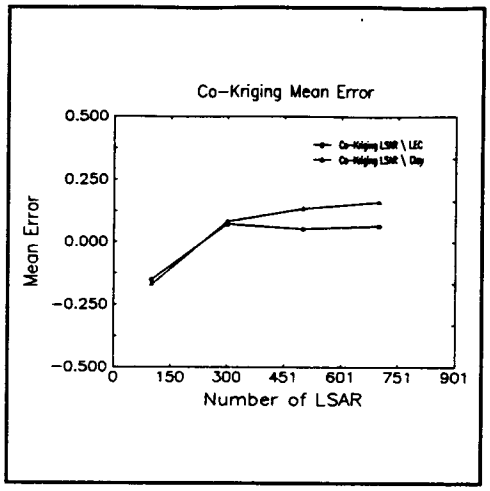


Figure 18 Effect of sampling number on mean error by using kriging and co-kriging estimation.

Table 8

Cross-validation parameters:

Variables	No.	Model	SSQ	Reduced Mean	Reduced Variance
LSAR \ LSAR	100	EXP.	0.014	0.048	0.950
LSAR \ LSAR	300	EXP.	0.009	0.014	1.057
LSAR \ LSAR	500	EXP.	0.117	0.000	0.833
LSAR \ LSAR	700	EXP.	0.160	0.040	0.851
LSAR \ LSAR	901	EXP.	0.0003	0.004	0.918
LSAR \ LEC	100	EXP.	0.048	0.004	0.521
LSAR \ LEC	300	SHP.	0.012	0.002	0.962
LSAR \ LEC	500	EXP.	0.005	0.000	1.089
LSAR \ LEC	700	EXP.	0.590	0.081	0.981
LSAR \ LEC	901	EXP.	0.0004	0.005	0.981
LSAR \ CLAY	100	EXP.	0.433	0.005	0.826
LSAR \ CLAY	300	SHP.	0.688	0.006	0.846
LSAR \ CLAY	500	EXP.	1.361	0.000	0.852
LSAR \ CLAY	700	SHP.	1.682	0.000	0.844
LSAR \ CLAY	901	EXP.	0.508	0.005	0.872
CLAY \ LEC	901	SHP.	0.007	0.004	0.856
CLAY \ CLAY	901	SHP.	0.985	0.005	0.871
LEC \ LEC	901	EXP.	0.001	0.004	0.788

Kriging

Kriging is a technique of making optimal, unbiased estimates of regionalized variables at unsampled locations using the structural properties of the semivariogram and the initial set of data values. The advantage of this method is the measurement of variance error. Gajem et al., (1981) and Sisson and Wierenga, (1981) used it to determine spatial variability of several soil properties. Burgess and Webster, (1980a,b) used it for mapping soil properties. Flatman et al., (1988) and McBratney and Webster, (1983) and Trangmer et al., (1985) also discuss optimal sampling schemes. Campbell, (1978) described the pH distribution. Yost et al., (1982a) examined the drift of several soil parameters in Hawaii. Burgess and McBratney, (1981) and Webster and Burgess, (1980) extended their earlier result to include other types of kriging. Palumbo and Khalee, (1983) estimated the transmissivity of New Mexico soils. Russo and Bresler, (1982), Vieira et al., (1981) and Warrick and Nielsen, (1980) looked at the physical properties of soils in the field. Vauclin et al., (1983) compared the result of undersampled soil properties to each other.

Kriging depends on an accurate variogram for estimating weights for interpolation. Using incorrect variograms can lead to unfavorable effects on the precision of the kriged estimate, even to a negative kriging variance. Figures 19-23, indicated very clearly the distribution of estimated LSAR subfiles. The north east zone had a higher salt concentration.

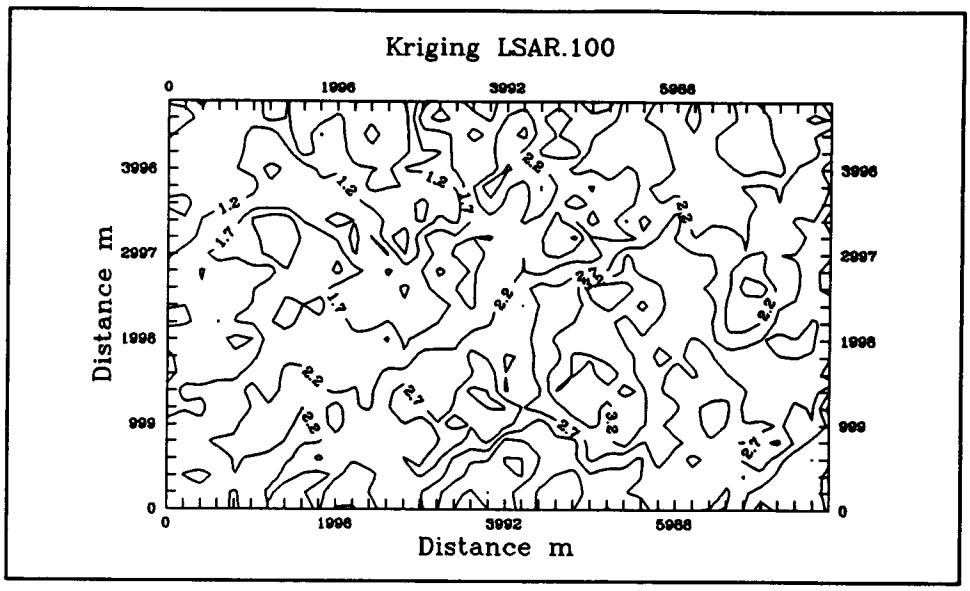


Figure 19 Estimated LSAR.901 from LSAR.100

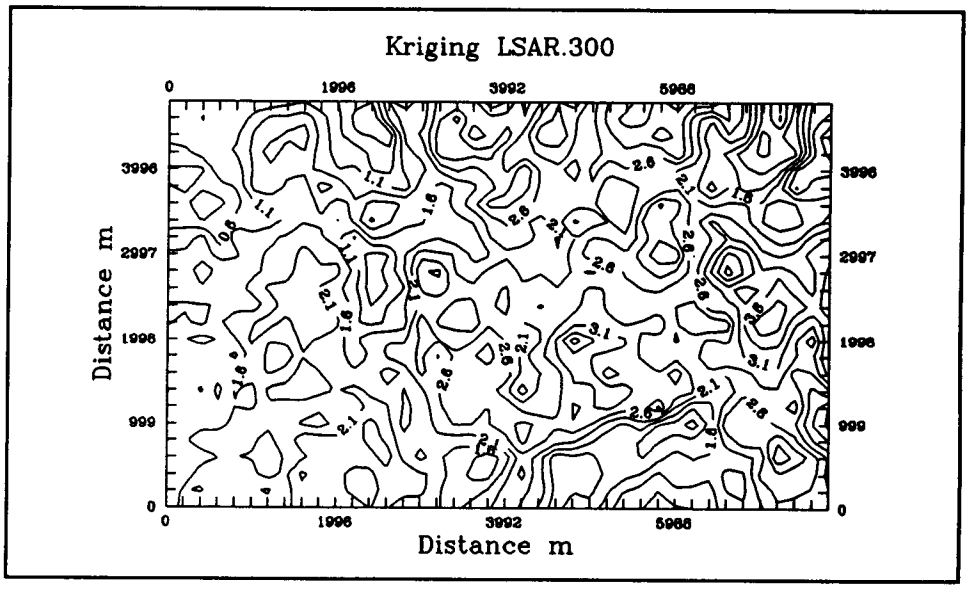


Figure 20 Estimated LSAR.901 from LSAR.300

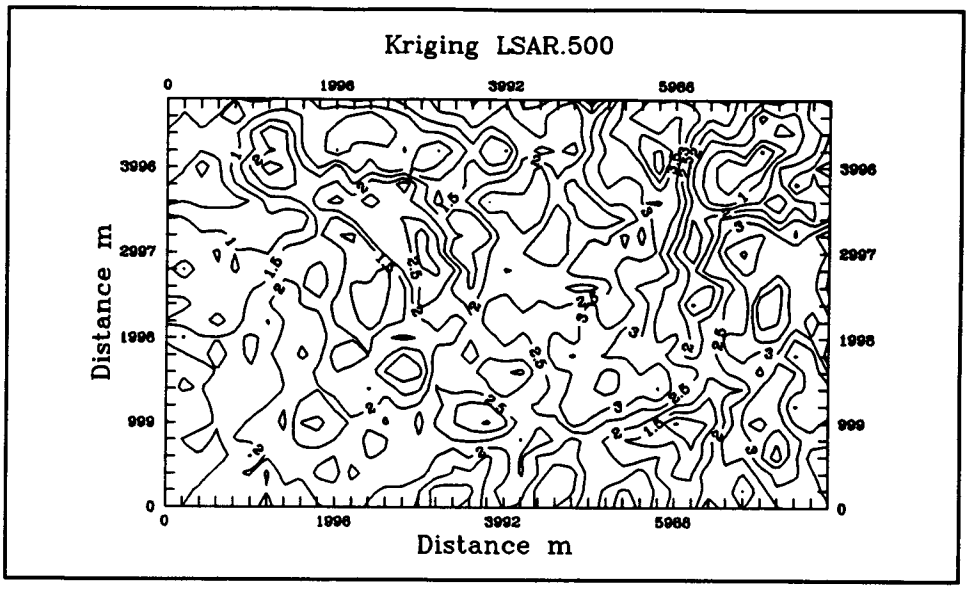


Figure 21 Estimated LSAR.901 from LSAR.500

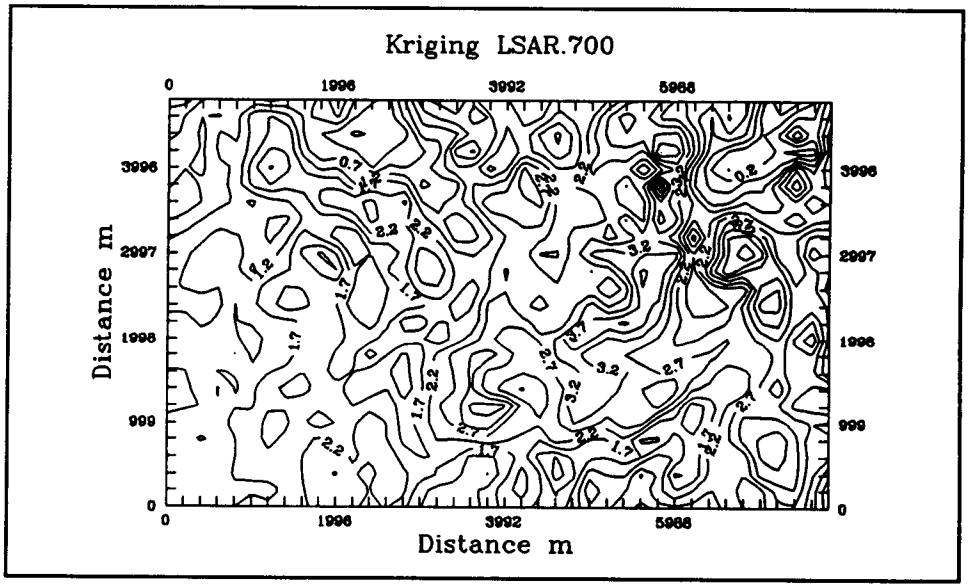


Figure 22 Estimated LSAR.901 from LSAR.700

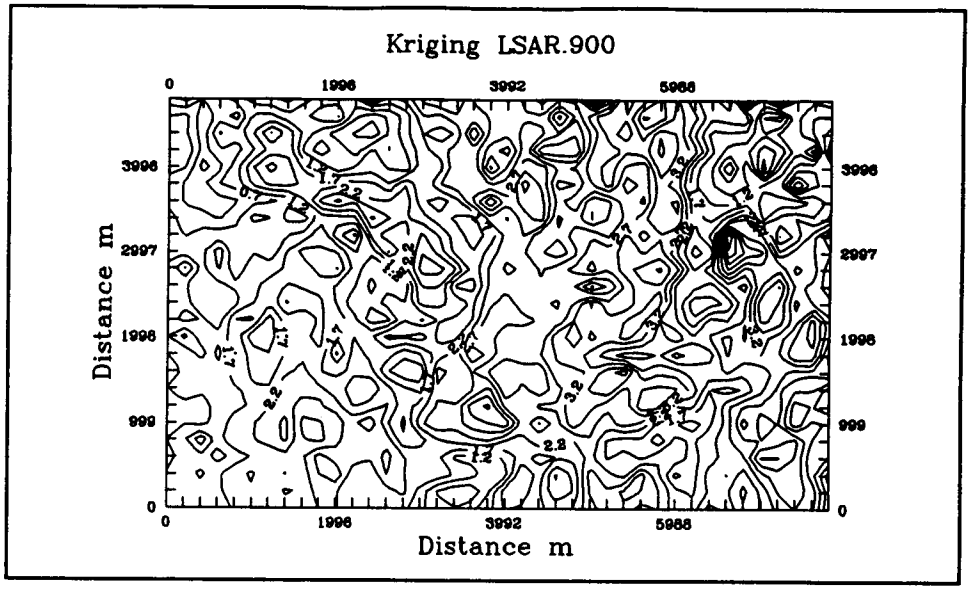


Figure 23 Estimated LSAR.901 from LSAR.901

Co-Kriging

The spatial distribution of any given property may often be closely related to that of other properties affected by the same regionalized phenomenon or spatial process. Such properties are said to be co-regoinalized and are spatially dependent on one another. Co-kriging extends the principle of optimal estimation using regionalized variable theory from a single property to situations where there are two or more co-regoinalized properties.

Matheron, (1971) developed the concept of "cross-semivariogram"; which gives the correlation between the two variables as a function of separation distance between two variables and a set of kriging equations which he called " co-kriging " (Journel and Huijbregts, 1987; Vauclin et al., 1983). For more details refer to Myers

(1982, 1983, 1984). Carr et al., (1985a) give the co-kirging computer program. The Yates, (1986) program "Geopack" performs co-kriging, kriging and disjunctive kriging. Co-kriging or joint estimation can improve the estimation of an under-sampled variable.

Co-kriging is certainly the method having the best theoretical foundation because no assumptions are made about the nature of the correlation between the two variables. The degree of this correlation as well as its spatial structure are taken into account by the cross-variogram (Ahmed and De Marsily, 1987). The technique of co-kriging improves the estimation and reduces the variance of the estimation error, but at the same time the calculation of the cross-variogram and the fitting of a theoretical model sometimes becomes very difficult, particularly when the two variables are not strongly correlated (Ahmed and De Marsily, 1987).

Co-kriging can be used to increase sampling efficiency by: 1) Increasing the sampling of the auxiliary function with respect to the primary variable by transferring the information of the auxiliary function to the primary through the cross-correlation function. 2) Increasing the apparent density of the sampling pattern. This will be useful when each random function is of equal interest and sampling difficulty. An improvement in the overall estimation process for both random functions may be obtained if most sample locations for each random function are independent of each other. This will have the effect of increasing the apparent sampling density in the field with respect to the function, and when a high cross-correlation exists, will be approximately the same as sampling for both function at all the location (Yates and

Warrick, 1987).

Recently co-kriging has been applied to soil hydraulic properties. Shouse et al. (1990) and Vauclin et al. (1983) used ordinary and co-kriging in the analysis of water content and sand content, where sand content was included to improve the estimation process. McBratney and Webster (1983b) used co-kriging to study the spatial variability of silt content where the subsoil silt and sand content were used as auxiliary variables. Carr and Myers (1985) annualized landsat data by using co-kriging. Yates and Warrick (1987) used co-kriging to estimate soil water content, with surface temperature and sand content as auxiliary variables.

On the other hand there are several limitations of co-kriging: 1) It requires a large sample number, because the cross-correlation function will be affected by reduced samples. 2) The cross-correlation function is difficult to obtain and interpolate. It is very hard to fit the theoretical model to it automatically. 3) It requires increased computer time Yates and Warrick, (1987).

Figures 24-27 are co-kriged LSAR for different subfiles by using clay content and LEC as auxiliary variables.

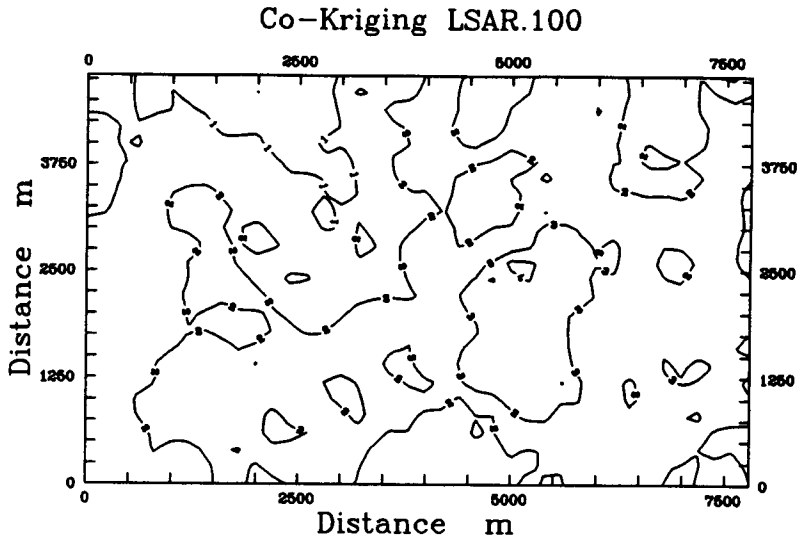


Figure 24 Estimated LSAR.901 from LSAR.100

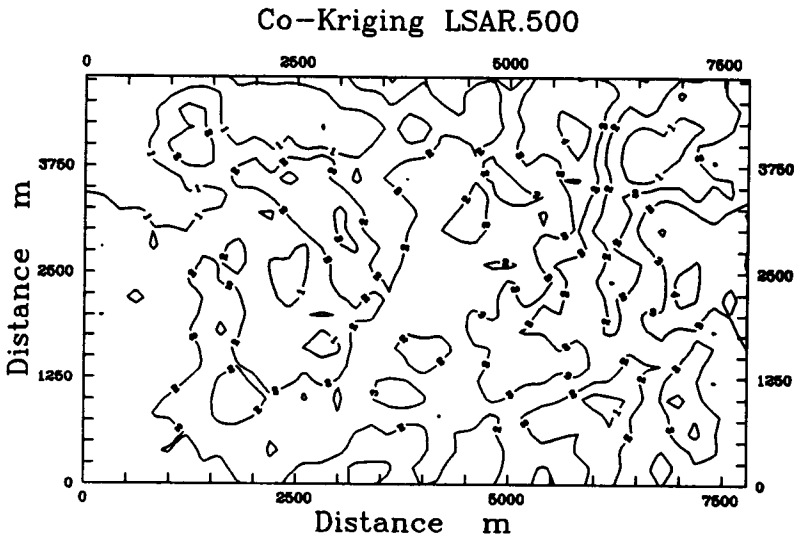


Figure 25 Estimated LSAR.901 from LSAR.500.

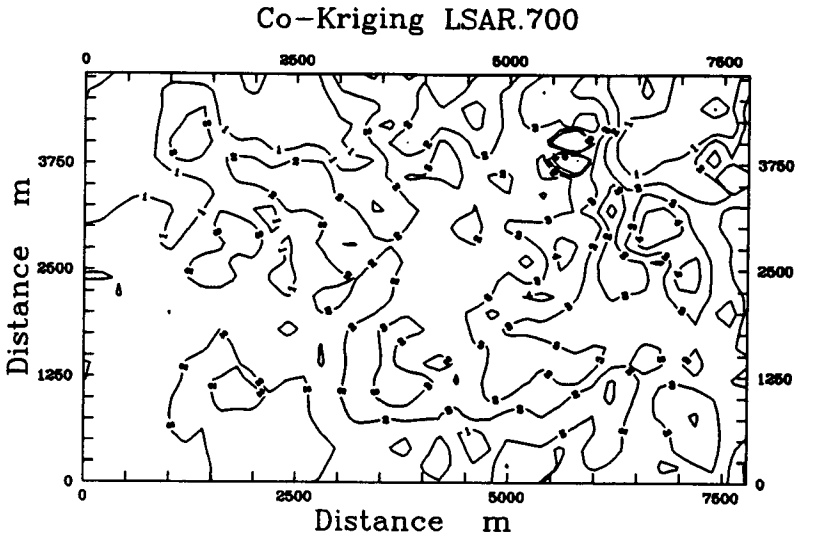


Figure 26 Estimated LSAR.901 from LSAR.700

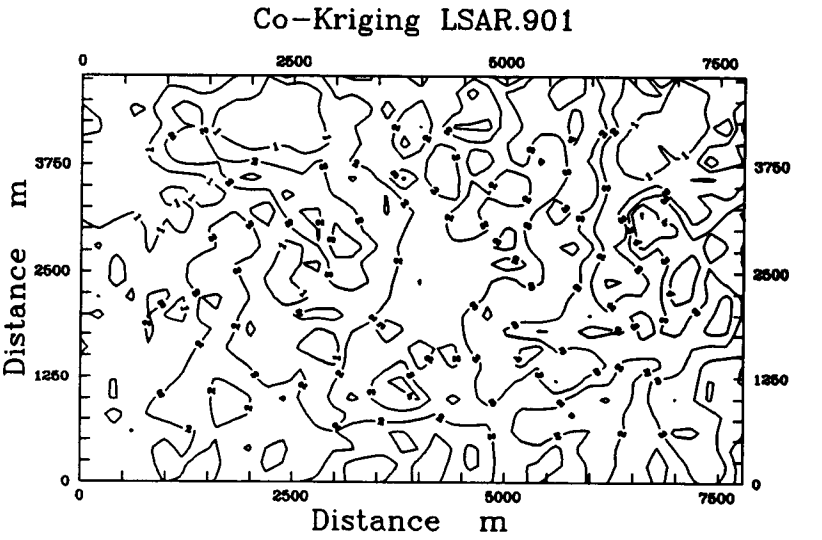


Figure 27 Estimated LSAR.901 from LSAR.901

Mapping

Maps are an important analytical aids in soil science. They are based on a specific number of samples drawn over the land surface of interest. Such analyses are made to estimate sodium absorption ratio at every point within the sampled space. Figures 24-27 show the co-kriging contour maps. Subfiles with LSAR.500 reveal the same distribution of LSAR in both kriging and co-kriging. Concentration of LSAR is higher at the north east zone and around the center of the map. The high value area is located east of the lake, which may show the adverse effect of precipitation of salt at the soil surface. The area around the lake has a better chance for leaching salinity. The lake enhances the drainage system and creates a higher hydraulic gradient toward the water table.

Comparison between kriging and co-kriging

The two methods yield the same maps in general. Co-kriging gives smoother and more uniform regions than kriging. The results from co-kriging may be more conservative than ordinary kriged estimates. Figures 16-18 show the difference between the two estimators variances. Figure 18 shows the minimum mean error calculated for the LSAR.300 subfile by the co-kriging procedure. It illustrates the effect of the correlation between the estimated parameter and the auxiliary one.

Factors affecting the average kriging variance are total number of samples,

number of samples used in the estimation, and maximum radius. As the radius increases AKV increases.

Figure 16 shows the minimum calculated mean square error for the LSAR.300 subfile using co-kriging. This illustrates the effect of correlation between the estimated and the auxiliary parameter. Table 9 gives the variances.

Table 9

Kriging variance for kriging and co-kriging estimation

Number of variables	Average Kriging Variance	
	Kriging	Co-kriging
LSAR.100	1.622	1.384
LSAR.300	0.871	0.645
LSAR.500	1.658	1.255
LSAR.700	1.040	0.948
LSAR.900	1.411	1.101

Summary

The optimization decisions requried in management of agricultural fields can be aided by the geostatistical procedures used. The complexity and heterogeneity of soil and the soil-water-plant relationship requires knowledge of the association as well as the differences among these parameters in the field. The correlation between some parameters and the variance among others requires investigation of all these relationships. This enhances the probability that the best assessment is applied, and at the same time that it will avoid major cost, time, and unpredicted quantitative results. In field management, this relationship can be determined by understanding the spatial structure of the parameters.

As shown in the tables, the integrated information between LEC, clay content and LSAR enhanced the out-come decision for estimating LSAR for the entire field. Co-kriging gave a lower sum of squares mean error than kriging. The number of SAR measurements required to estimate the whole field could be reduced to 300 samples, instead of the 901 samples taken. This reduction in sampling optimizes the cost of reclamation by reducing the cost and time for sampling and analysis, and has a direct input in the final decision for reclamation. The co-kriged estimation maps showed how to apply different techniques for different regions. The contour maps indicated the specific limit of treatment, i.e amount of water, the quality of water and type of amendment required for each regionalized or block area.

CHAPTER 3

Simulated Reclamation of Saline Soils Using Spatially Variable Input Parameters and Initial Conditions

Introduction

The factors most limiting agricultural development in arid and semiarid regions is soil and water salinity. Soils in arid and semiarid regions usually contain salt. Salinity not only limits agricultural production, but is a source of potential groundwater and surface water pollution where irrigation is practiced. The water supply available for irrigation contains dissolved salts, and frequently contains phytotoxic trace elements (Hoffman et al., 1989). Over 90% of the consumptive use of water in the western United States, some 250 x 10^9 L d^{-1} goes to irrigated agriculture (Solley et al., 1988). Of 15 x 10^6 ha currently irrigated in the western U S, between 25 and 35% are seriously affected by salt. The annual cost of damages caused by excess soil salinity is estimated to be $30 million in California's San Joaquin Valley alone; an additional $100 million is lost in productivity in the Colorado River Basin. Salinity is the most pervasive environmental problem stemming from irrigation in the United States; all western river basins except the Columbia have salt levels high enough to reduce agricultural production.

The total dissolved inorganic ions and molecules define the level of salinity in irrigation water and soil solution. The major components of salinity are the cations Ca^{++}, Mg^{++}, Na^{+}, and the anions Cl^{-}, $SO^{=}_{4}$,and HCO_{3}, while K^{+} and NO^{-}_{3} are considered minor components. Soil salinity must be quantified for diagnosis and remediation. There are many methods for quantifying the salinity status of soils and waters (Rhoades and Oster, 1986). The method used most frequently is the measurement of electrical conductivity in soil solution extract obtained at water

contents higher than normally found in the field. The world standard method is described in Handbook 60 (U. S. Salinity Lab. Staff, 1954) as the electrical conductivity of the saturation extract. This is the primary assessment tool available for diagnosis of soil salinity conditions.

Assessment of soil salinity under field conditions is complicated by its spatially variable nature. Soil salinity is dynamic due to the many processes which influence salt accumulation and movement in soil. The major factors influencing soil salinity include: varying management practices, water table depth, soil permeability, evapotranspiration rate, rainfall amount, and solute transport parameters. Recommendations for reclamation must be based on the assessment of soil salinity over an area. An evaluation of the initial level of soil salinity and its distribution in space must be combined with knowledge of the transport properties affecting solute movement in order to choose the best procedure and to predict the cost of reclaiming saline soils.

Reclamation of salt affected soils is frequently required when arid lands are developed for agricultural use, when saline groundwater persists near the soil surface, or when the combined effects of rainfall and irrigation do not meet the leaching requirement for the land. Presently the only method for reclaiming salt affected soils is to leach salts below the crop root zone by adding large amounts of water. Adequate drainage, then, is essential for proper reclamation. Where natural drainage is not sufficient, artificial drainage systems need to be provided for reclamation to be feasible.

Leaching soil for removal of excess salt accumulation requires an understanding of water and solute movement through soils. Solute transport is affected by a number of chemical, physical and microbiological processes. In general, the best reclamation procedure for a specific site will depend on the soil type, the amount and types of salts present, irrigation water quality, hydrology of the region, topography, and irrigation management.

The amount of water that must leach through the soil profile to remove the excess soluble salts depends on the initial salt concentration, the technique of applying irrigation water and the soil type. As a general rule of thumb, 70% or more of soluble salts initially present in a soil profile will be removed by leaching with a depth of water equal to the depth of soil to be reclaimed, when water is continuously pounded on the soil surface. The soil type governs the rate at which water moves through soil and the amount of soil water retained in smaller pores.

Water movement and hence solute transport, is governed by the hydraulic conductivity of the soil and by its water retention. These hydraulic properties of field soils are highly variable in space and time. Hydraulic properties can be measured using laboratory or field techniques. There is an abundance of measurement techniques, but most methods are expensive and/or time consuming. Indirect methods have been developed to estimate soil physical properties from index properties such as grain size distribution.

The purpose of this chapter is to use spatially variable initial salinity and water content field conditions as model input to simulate the spatial distribution of

the required reclamation for a 15 square mile region of salt-affected soils, to determine the spatial variability of applied water necessary for reclamation, the spatial distribution of total salt leached out the root zone, and the total time required for water to pass through the 0.3 m soil profile.

Procedure

The salt-affected region chosen has been described in Chapter 2, and the spatial distributions of salt (EC_e), clay content, and SAR have been previously analyzed using the geostatistical methods described in Chapter 1. The volumetric water content were also measured at 901 sampling points using volumetric soil sampling. The soil texture was determined by hand texturing and independently calibrated against hydrometer measurements of particle size in the laboratory. The textural class data were used to determine the amount of sand and silt, given the measured amount of clay content, from the texture triangle.

The spatial distribution of the initial salinity content and the initial water content profile was measured at the 901 sample sites over the 15 square miles. The soil hydraulic properties at each of the 901 sites were estimated using a model which estimates the soil hydraulic properties from soil texture, salt content and other routinely measured soil properties listed in Table 10. The model uses old and new empirical as well as theoretical relationships between more easily measured soil properties and soil hydraulic properties. The input data required are listed in Table 10. The model produces an output data set which includes the soil water retention

function for each horizon as listed in Table 11. The spatial distribution of soil hydraulic properties as input to the solute transport ie the retention factors, alpha and beta, and the soil hydraulic conductivity function were used in the Hydrus solute transport model which was used to calculate the solute movement through our 901 soil profiles, to simulate salt leaching and reclamation. Hydrus is a one dimensional subsurface water flow and solute transport model which may be used to simulate the movement of water and dissolved solutes in saturated-unsaturated soils. The solution to the flow problem, in general, considers the effects of root water uptake and hysteresis in the soil hydraulic properties. The solute transport equation incorporates the processes of ionic or molecular diffusion, hydrodynamic dispersion, linear or nonlinear equilibrium adsorption, and first order decay. The boundary conditions for the flow and transport equations may be constant or time-varying. The model uses fully implicit Galerkin type linear finite element solutions of the convection dispersion equation and Richards' equation.

The Hydrus model was used to determine the linearly averaged, mean root zone salinity of the soil profile during simulated reclamation at each of the 901 sampling sites. A ponded water surface boundary condition was used because ponding of water is used extensively in this area for leaching of salts. The program outputs were total amount of water applied, time required for leaching to an average salinity of 4 ds m^{-1}, and the total salt leaving the 0.30 m soil depth.

The input parameters used in Hydrus are listed in Table 12.

Table 10

The input file for sample number 100

Name	: GENERAL CLAY DATA SET		
Horizons	: 3		
Root Zone Depth (cm)	: 30.00		
Depth (cm)	: 33.0	66.0	100.0
Rock Fragments (vol. %)	: .00	.00	.00
Clay Activity	: .35	.35	.35
Clay (weight %)	: 12.00	12.00	12.00
Silt (weight %)	: 21.00	21.00	21.00
Sand (weight %)	: 77.00	77.00	77.00
Organic Matter (weight %)	: 1.50	.50	.00
Coarse Sand (weight %)	: .00	.00	.00
Macroscopic Porosity Code	: 2	2	2
Compaction Code	: 2	2	2
Bulk Density Flag	: 1	1	1
Bulk Density (g/cc)	: 1.461	1.461	1.461
Sat. Conductivity (cm/hr)	: .00	.00	.00
Salinity Flag	: 1	1	1
E.C. Soil Extract (mmhos/cm):	1.11	1.11	1.11
E.C. Irrig. Water (mmhos/cm):	1.80	1.80	1.80
W.C. of Soil Paste (g/g)	: 40.02	40.02	40.02
Sodium Adsorption Ratio	: 2.84	2.84	2.84

Table 11

The out put for sample number 100

SOIL PROPERTIES
Name: GENERAL CLAY DATA SET
Number of Horizons: 3
Root Zone Depth : 30.

dp	rf	ca	clay	silt	om	cs	mp	cp	dbmom	flag	Ksat
33.	.0	.35	12.0	21.0	1.5	.0	2	3	1.46	1	25.90
66.	.0	.35	12.0	21.0	.5	.0	2	3	1.46	1	25.04
100.	.0	.35	12.0	21.0	.0	.0	2	3	1.46	1	24.61

DEPTH OF WATER TABLE (cm)	DRAINED WATER VOLUME (cm)	UPFLUX FROM WATER TABLE (cm/hr)	GREEN-AMPT PARAMETERS A (sq.cm/hr)	GREEN-AMPT PARAMETERS B (cm/hr)	WATER CONTENT (THETA) (vol %)	MATRIC WATER SUCTION (HEAD) (cm)
.0	.00	.2000	.0000	.0000	.4279	.0
10.0	.10	.2000	2.4629	25.9038	.4181	-5.0
20.0	.45	.2000	5.1619	25.9038	.4053	-10.0
30.0	1.03	.2000	7.4597	25.9038	.3806	-20.0
40.0	1.81	.2000	9.3126	25.7523	.3596	-30.0
50.0	2.75	.2000	10.8374	25.6096	.3421	-40.0
60.0	3.83	.1911	12.1277	25.5144	.3152	-60.0
70.0	5.03	.1304	13.2230	25.4221	.2953	-80.0
80.0	6.33	.0929	14.1579	25.3211	.2799	-100.0
90.0	7.71	.0698	14.9805	25.2425	.2528	-150.0
100.0	9.17	.0524	15.7118	25.1796	.2347	-200.0
120.0	12.28	.0311	16.9603	25.0853	.2111	-300.0
140.0	15.60	.0200	17.9933	25.0179	.2043	-340.0
160.0	19.09	.0135	18.8675	24.9674	.1959	-400.0
200.0	26.49	.0068	20.2794	24.8966	.1764	-600.0
250.0	36.37	.0032	21.6242	24.8400	.1550	-1000.0
300.0	46.78	.0000	22.6702	24.8023	.1311	-2000.0
400.0	68.78	.0000	24.2220	24.7551	.1067	-5000.0
500.0	91.95	.0000	25.3429	24.7269	.0926	-10000.0
700.0	140.73	.0000	26.9009	24.6945	.0854	-15300.0
1000.0	217.97	.0000	28.3880	24.6703	.0632	-102000.0

Table 12

Salinity data for salinity reclamation and management for sample number 100

```
  2   0   0   0   0   1   0   0   0   0
5000   1   1   1   6   0     0.0
   0.001   0.00001     0.2    900.0     0.1     0.5   15    0.001
  1   1     0.0    30.00     1.0
 0.03880     1.45   41.300     1.30    9.500
    0.3    10.00    1.461     0.0     0.0    .000      1.0
    0.0   -800.0     0.0    1.11
   30.0   -800.0     0.0    1.11
    0.0   2   2   1000.0     0.0     0.0    0.00
  10.00     20.0     30.0    40.0    50.0    60.0
```

****

ONE-DIMENSIONAL FLOW AND TRANSPORT MODEL
HYDRUS v. 3.2

Data input file: FF100.IN

Salinity data for salinity reclamation and management

****

PROBLEM CONTROL PARAMETERS
= =
SIMULATION CONTROL CODE .....(ITKOD) = 2
HYSTERESIS MODELING CODE .....(IHKOD) = 0
TRANSPORT BOUNDARY COND. CODE .....(ICOKOD) = 0
ROOT WATER UPTAKE CODE .....(IRUKOD) = 0
CONDUCTIVITY UPSTREAM WEIGHTING .....(IUPKOD) = 0
FLOW MASS MATRIX OPTION .....(ILKOD) = 1
FLOW INITIAL CONDITION CODE .....(ICKOD) = 0
BOUNDARY CONDITION CODE .....(IBCKOD) = 0
PLOT OUTPUT CODE .....(IOKOD) = 0
RESTART OUTPUT CODE .....(IRSKOD) = 0

Table 12 Continued

TIME STEPPING PARAMETERS
==========
INITIAL TIMESTEP .....(DELIN) = .100E-02
MINIMUM TIMESTEP .....(DELMIN) = .100E-04
MAXIMUM TIMESTEP .....(DELMAX) = .200
TOTAL SIMULATION TIME .....(TMAX) = 900.
REL. PR. HEAD CONVERGENCE TOLERANCE ....(TOL1) = .100
ABS. „ „ „ ....(TOL2) = .500
NUMBER OF NONLINEAR ITERATIONS .....(NITMAX) = 15
CONCENTRATION CONVERGENCE TOLERANCE ...(CTOL) = .002

PROBLEM SPECIFICATION PARAMETERS
==========
MAXIMUM NUMBER OF TIMESTEPS .....(NSTEPS) = 5000
NUMBER OF SOIL MATERIALS .....(NMAT) = 1
NUMBER OF SOIL LAYERS .....(NLAYR) = 1
NUMBER OF BOUNDARY COND. TIME VALUES ....(NBC) = 1
NUMBER OF OUTPUT TIME VALUES .....(NPRINT) = 6
NUMBER OF OBSERVATION POINTS .....(NOBS) = 0
SOIL DEPTH .....(TDEPTH) = 30.0
GROUNDWATER SOLUTE CONCENTRATION .....(CNN) = .000

PROBLEM GEOMETRY
==========
LAYER NUMBER 1
MATERIAL INDEX .....(MATL) = 1
LAYER THICKNESS .....(THICK) = 30.0
BEGINNING DEPTH .....(TOPL) = .000
ENDING DEPTH .....(BOTL) = 30.0
NODAL SPACING .....(DELZ) = 1.00

SOIL HYDRAULIC AND TRANSPORT PROPERTIES
==========
HYDRAULIC PROPERTIES FOR MATERIAL: 1

ALPHA	BETA	WCS	WCR	SATK
.0162	1.3000	44.8000	5.7900	1.200E+02

Table 12 Continued

TRANSPORT PROPERTIES FOR MATERIAL: 1

DIF	DISP	RHO	DONE	DSONE	KD
.3000	10.0000	1.4490	.0000	.0000	.0000

MAXIMUM VALUE OF GRID PECLET NUMBER IS .1 FOR LAYER

BOUNDARY CONDITION DATA

TIME	IRTYP	IDRTYP	BCN1	BCNN	CN1	POTET
.000	2	2	1000.000	.000	.000	.000

OUTPUT TIME VALUES

.100E+02 .200E+02 .300E+02 .400E+02 .500E+02 .600E+02

INITIAL CONDITIONS

Depth	P	WC	C	Depth	P	WC	C
.00	-8.000E+02	23.7320	8.900E-01	1.00	-8.000E+02	23.7320	8.900E-01
2.00	-8.000E+02	23.7320	8.900E-01	3.00	-8.000E+02	23.7320	8.900E-01
4.00	-8.000E+02	23.7320	8.900E-01	5.00	-8.000E+02	23.7320	8.900E-01
6.00	-8.000E+02	23.7320	8.900E-01	7.00	-8.000E+02	23.7320	8.900E-01
8.00	-8.000E+02	23.7320	8.900E-01	9.00	-8.000E+02	23.7320	8.900E-01
10.00	-8.000E+02	23.7320	8.900E-01	11.00	-8.000E+02	23.7320	8.900E-01
12.00	-8.000E+02	23.7320	8.900E-01	13.00	-8.000E+02	23.7320	8.900E-01
14.00	-8.000E+02	23.7320	8.900E-01	15.00	-8.000E+02	23.7320	8.900E-01
16.00	-8.000E+02	23.7320	8.900E-01	17.00	-8.000E+02	23.7320	8.900E-01
18.00	-8.000E+02	23.7320	8.900E-01	19.00	-8.000E+02	23.7320	8.900E-01
20.00	-8.000E+02	23.7320	8.900E-01	21.00	-8.000E+02	23.7320	8.900E-01
22.00	-8.000E+02	23.7320	8.900E-01	23.00	-8.000E+02	23.7320	8.900E-01
24.00	-8.000E+02	23.7320	8.900E-01	25.00	-8.000E+02	23.7320	8.900E-01
26.00	-8.000E+02	23.7320	8.900E-01	27.00	-8.000E+02	23.7320	8.900E-01
28.00	-8.000E+02	23.7320	8.900E-01	29.00	-8.000E+02	23.7320	8.900E-01

Table 12 Continued

```
  30.00  -8.000E+02 23.7320   8.900E-01

 INITIAL MOISTURE IN PROFILE :  711.9607
 INITIAL SALT IN PROFILE     :  633.6450

CONVERGENCE FAILURE AT TIME 1.0000E-03 DELT= 1.000E-03    NODE
=  1 PE= 0.000E+00 T= -1.000E+02 WP= 3.333E-01

NORMAL TERMINATION TIME = 5.3067 AND STEP NUMBER = 5000
 CAVLIM =     .10000
```

Results of the Simulation

Figure 28 shows the spatial distribution of measured initial salinity conditions in the field prior to the onset of simulated reclamation. The figure indicates that there are several areas which have high salt concentrations. The area of high salt concentrations are located primarily on land which has been poorly managed, had shallow water table during part of the season or on abandoned land. The range of salt content is very large, this is part of the reason for having sampled this area.

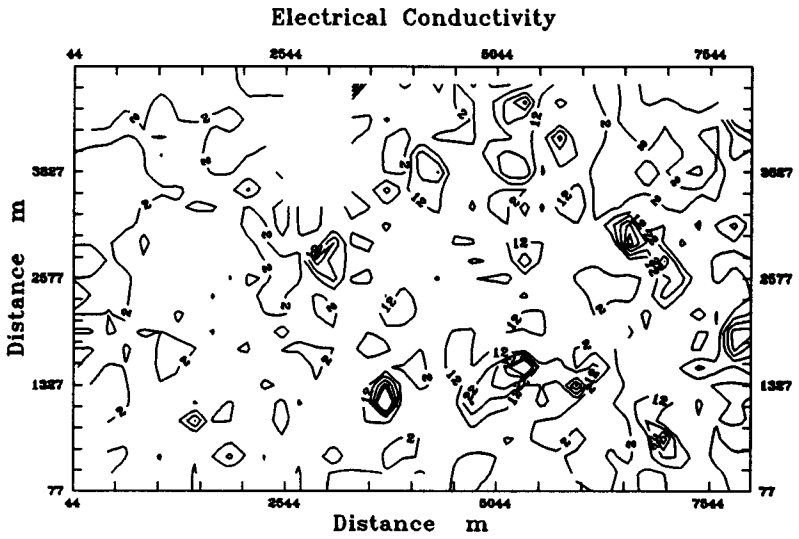


Figure 28 Spatial distribution of EC in the field.

Figure 29 illustrates the spatial distribution of clay content across the study area. The area has a range of soil types, from sands to clays. The clay content data and the soil type information are determined at the time of sampling by state soil scientists working with the Soil Conservation Service. These scientists used hand texturing to determine the "apparent" clay content and the overall soil textural classification of each individual soil sampling site. These estimates were confirmed using a limited number of soil samples (usually several for each textural class) for particle size analysis in the laboratory. The correspondence between measured and estimated was quite good. Silt and sand content were estimated using the clay content and the soil textural classification.

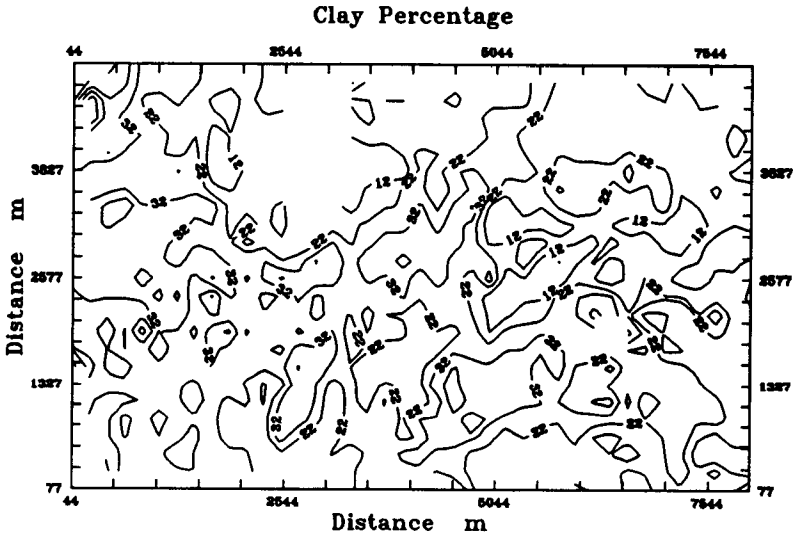


Figure 29 Spatial distribution of clay content in the field.

Figure 30 shows the spatial distribution of the initial soil water content used as the initial condition in the reclamation simulation. The water content covered a relatively wide range (0.1-0.45). The range was indicative of the soil type and recent management practices. In general soils with higher water content were recently irrigated and/or had a higher clay content. The soils with relatively lower water contents were usually sandy soils and/or are soils that have been abandoned or are in fallow.

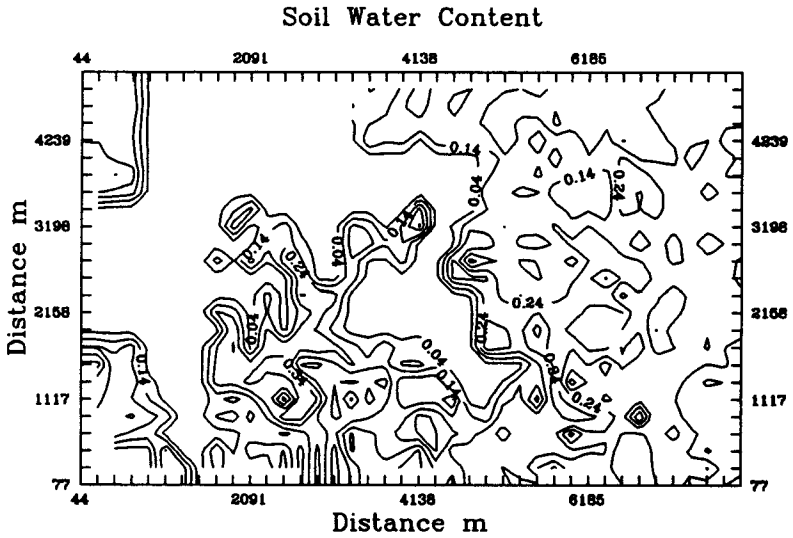


Figure 30 Spatial distribution of moisture content in the field.

Figures 28 and 30 give an indication of the spatial variability of soil properties in the study area. The properties are highly variable and are a result of many processes active in irrigated agriculture. By far the most dominant factor active in the area is the management practices employed by individual farmers and land owners.

The outputs of Hydrus are, total applied water, total amount of drainage water, and total salt transported through the soil profile. These outputs are essential for assessing the reclamation process. The cost and availability of good quality irrigation water are important factors that can help determine the applicability of the reclamation simulation. Figure 31 shows that two-thirds of the area needs to have at least 10 cm of water to be reclaimed to 4 ds m^{-1}. The higher applied water was on the higher zones of salinity.

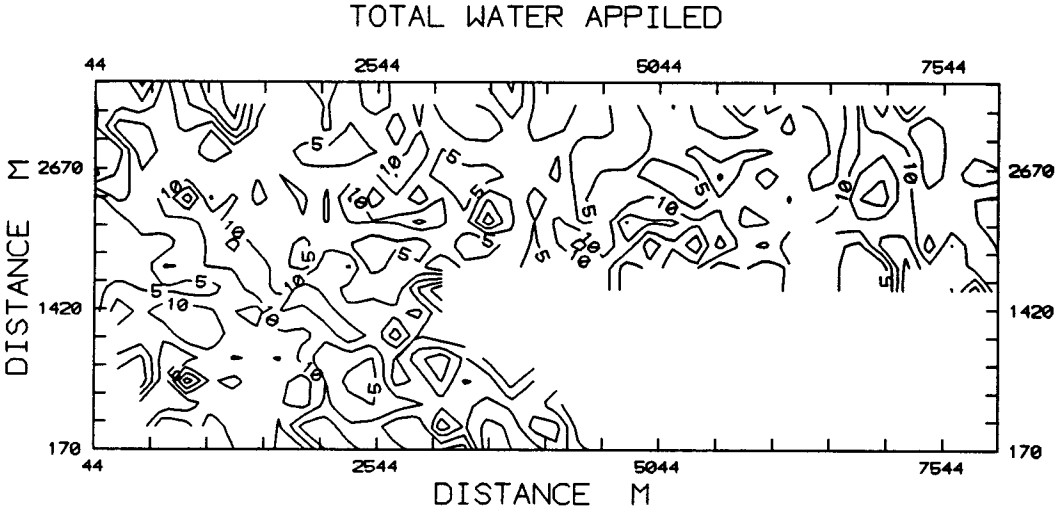


Figure 31 Spatial distribution of applied water.

In any reclamation assessment, drainage water amount is also an important factor to consider because, drainage is essentially waste water and is getting harder to dispose of. Cost of disposal of drainage effluent may be a consideration when analyzing the total economic impact of a proposed reclamation project. In the future, drainage water disposal may become an over-whelming environmental issue to farmers and land owners.

The main purpose of leaching is to relocate the salts from the root zone to some other area. Figure 32 indicates the areas expected to have large amounts of water draining from the profile. In general the larger amounts of drainage water are from areas of high initial salt concentrations and areas of higher clay content. These areas need more applied water for reclamation.

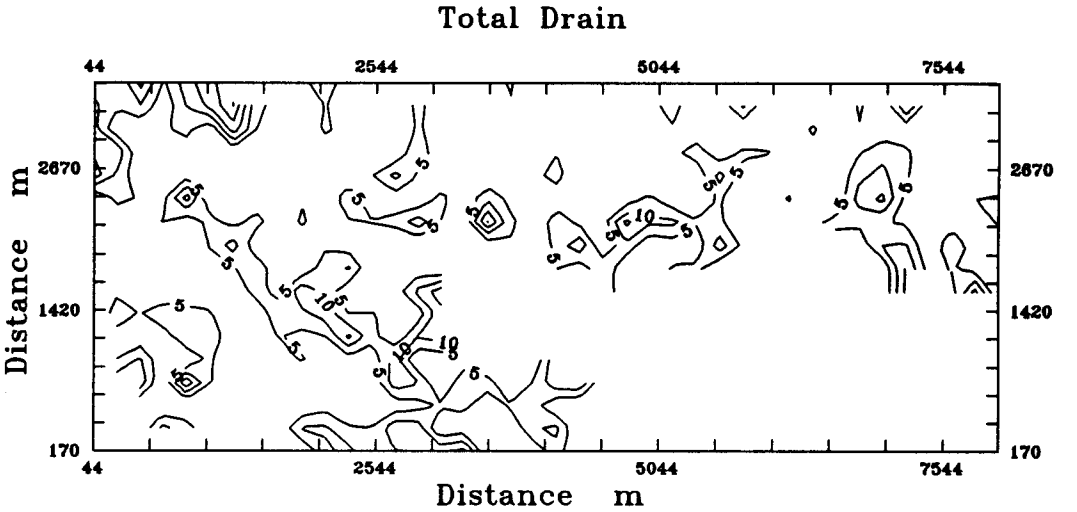


Figure 32 Spatial distribution of total drainage.

The accumulation of salts below the root zone indicates the amount of potential salt loading to groundwater, and is another indicator for possible source of pollution that must be considered when analyzing the total impact of reclamation of a saline soil using leaching principles. Figure 33 shows the spatial distribution of the salt loading potential within the landscape. We can use this to help make the decisions about the feasibility of using leaching to accomplish reclamation.

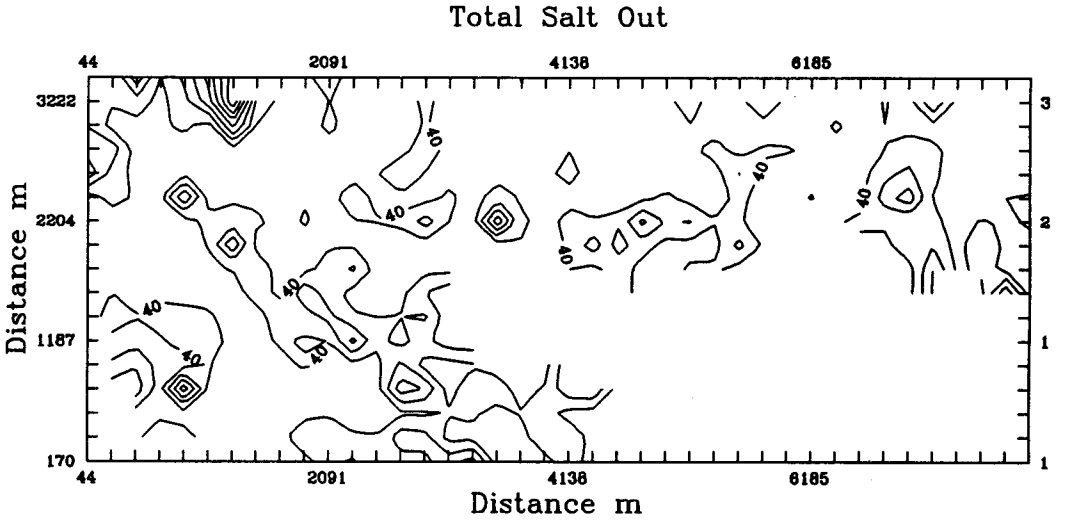


Figure 33 Spatial distribution of salt loaded at subsurface profile.

Conclusions

For a greater productivity and higher quality food, soil reclamation management plays an important role in the future. Soil management and irrigation management practices control soil salinity. Water quality, the irrigation method and crop rotation also play roles in soil salinization. The analysis of salinity distribution in this study indicates the variability of EC over the 15 square mile area. To a lesser extent physical characteristics determine the soil salinity levels. Clay soil had higher salt contents than did the sandy soils.

This simulation was an attempt to use the spatially variable input and initial condition data to analyze the reclamation of a rather large-scale saline soil area. The simulation output, water applied, salt loading to groundwater and drainage water volume can be used in an economic analysis of salt leaching in this area. The simulation did not include specific ions such as sodium or boron. Leaching of these elements from soil is much more difficult than leaching of "salt".

Simulation indicated that areas with large salt concentrations were more likely to produce areas with more drainage water and more salt loading to groundwater. Intuitively this is consistent with known concept of soil science, but the simulation relating effectiveness of leaching can be quantified. Areas of potential pollution can also be identified. With newer and more sophisticated simulation models becoming available, different management practices will be able possible.

Summary and Conclusions

1. Electrical conductivity and sodium adsorption ratio were found to be spatially correlated. The natural log of electrical conductivity, (EC) and sodium adsorption ratio, (SAR) were found to fit exponential theoretical semivariogram models while clay content fits a spherical model.

2. All theoretical models used in semi-variograms and cross-semivariograms have been validated by i) Jack-knife procedure; ii) the difference of error between the actual measured points and estimated points; and iii) The best eye fit curves with the smallest sum of squares error. All results suggested that the models chosen were valid.

3. Smaller subsets of the measured variable affected the magnitude of the variogram model coefficients.

4. Ordinary kriging was the simplest and easiest method to apply. It took less time than co-kriging and disjunctive kriging. The results were adequate for a preliminary study of the spatial distribution.

5. Co-kriging gave the smallest variances and errors and is the best method for estimating SAR from EC. The higher the correlation for these variables, the more realistic and accurate the estimation will be. This criterion helps in using a less expensive measured variable to estimate the more expensive variable.

6. Co-kriging can optimize sample numbers, cost and time. The results suggested that 300 samples could be adequate for this field. This reduction from 900 sample measurements of SAR will save two-thirds of the cost.

7. Disjunctive kriging indicates the possibility of using nonlinear estimators, with better results than linear estimators. It also provides an estimate of conditional probability that a variable at a given site is greater or less than a certain cutoff value. This additional information can provide a basis for risk management.

8. Simulated reclamation with the 901 data points using a Hydrus program and disjunctive kriging gave a very good estimate of applied water and time required to leach the whole field. From disjunctive kriging management, units can be drawn based either on the amount of applied water or the probability of achieving the threshold electrical conductivity.

References Cited

Ahmed, S. and G. De Marsily. 1987. Comparison of geostatistics for estimating transmissivity using data on transmissivity and specific capacity. Water Resour. Res. 24:1717-1737.

Armstrong, M., and R. Jabin. 1981. Variogram models must be positive definite. Math. Geol. 13:455-459.

Armstrong, M., and A. Boufassa. 1988. Comparing the robustness of ordinary kriging and lognormal kriging: outlier resistance. Math. Geo. 20:447-457.

Armstrong, M. 1984. Common problems seen in variograms. J. Int. Assoc. Math. Geol. 16:305-313.

Bautista, E. and W. W. Wallender. 1985. Spatial variability of infiltration in furrows. Trans. ASAE 28:1846-1851 and 1855.

Beckett, P. H. T., and R. Webster. 1971. Soil variability: a review. Soils Fert. 34:1-15.

Berndtsson, R. and M. Larson. 1987. Spatial variability of infiltration in a semi-arid environment. J. Hydrol. 90:117-133.

Biggar, J. W. and D. R. Nielsen. 1976. Spatial variability of the leaching characteristics of a field soil. Water Resour. Res. 12:78-84.

Bresler, E., G. Dagon, R. J. Wagenet, and A. Laufer. 1984. Statistical analysis of salinity and texture effects on spatial variability of soil hydraulic conductivity. Soil Sci. Soc. Am. J. 84:16-25.

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

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

Burgess, T. M. and A. B. McBratney. 1981. Optimal interpolation and isarithmic mapping of soil properties VI. Sampling strategy. J. Soil Sci. 31:643-659.

Campbell, J. B. 1978. Spatial variation of sand content and pH within single contiguous delineation of two soil mapping units. Soil Sci. Soc. Am. J. 42:460-464.

Carr, J., and D. Myers. 1985. COSIM: A fortran IV program to co-conditional simulation. Computers and Geoscience. 11:675-679.

Carr, J., Myers, D. E., and C. E. Glass. 1985a. Co-kriging: a computer program. Computers and Geoscience. 11:111-127.

Christakos, G. 1988. On-line estimation of nonlinear physical systems. J. Int. Assoc. Math. Geol. 20:111-133.

Chung, C. 1984. Use of the jack-knife method to estimate autocorrelation functions (or variograms). In G. Verly et al.,(Eds.) Geostatistics for Natural Resources Characterization. by D. Reidell. Dordecht. 55-70 pp.

Clark, I. 1979. Practical Geostatistics. Applied Science Publishers. London, U.K. 129 pp.

Cooper, R. M. and J. D. Istok. 1988. Geostatistics applied to groundwater contamination. I: Methodology. J. Env. Eng. 144:270-286.

Cressie, N. A. C. 1985. Fitting variogram models using weighted least squares. J. Int. Assoc. Math. Geol. 17:563-586.

Cressie, N. A. C. 1986. Kriging nonstationary data. J. Am. Stat. Assoc. 81:625-634.

Cressie, N. A. C. and D. M. Hawkins. 1980. Robust estimation of the variogram. I. J. Int. Assoc. Math. Geol. 12:115-125.

Cressie, N. A. C. and R. Horton. 1987. A robust-resistant spatial analysis of soil water infiltration. Water Resour. Res. 23:911-917.

David, M. 1977. Geostatistical Ore Reserve Estimation. Elsevier Scientific Publishing Co., Amsterdam, The Netherlands.

Davidoff, B., J. W. Lewis, and H. M. Selim. 1986. A method to verify the presence of a trend in studying spatial variability of soil temperature. Soil Sci. Soc. Am. J. 50:1122-1127.

Davis, B. M. 1987. Uses and abuses of cross-validation in geostatistics. J. Int. Assoc. Math. Geol. 19:320-328.

Delfiner, P. 1976. Linear estimation of nonstationary spatial phenomena. p. 49-68. *In* M. Guarascio et al., (ed.) Advanced geostatistics in the mining industry. Reidel Publ. Co., Dordrecht, The Netherlands.

Delhomme, J. P. 1976. Kriging in the hydrosciences. Adv. Water Resour. 1:251-266.

Dowd, P. 1982. Lognormal kriging: the general case. Math. Geol. 14:474-500.

Dowd, P. 1984. The variogram and kriging: Robust and resistant estimators. In Verly et al. (Eds.), Geostatistics for Natural Resources Characterization. by D. Reidel. Amsterdam. 91-106 pp.

Dunn, M. R. 1983. A simple sufficient condition for a variogram model to yield positive variance under restrictions. Math. Geo. 15:553-564.

Flatman, G. T., E. J. Englund, and A. A. Yfantis. 1988. Geostatistical approaches to the design of sampling regimes. Chapter 4. In L. H. Keith (Ed) Principles of environmental sampling. American Chemical Society Professional reference book.

Gajem, M. R., A. W. Warrick and D. E. Myers. 1981. Spatial dependence of physical properties of a Typic Torrifluvent soil. Soil Sci. Soc. Am. J. 45:709-715.

Gambolati, G., and G. Volpi. 1979. Groundwater contour mapping in Venice by stochactic interpolators. I. Theory. Water Resour. Res. 15:218-297.

Gambolati, G., and G. Volpi. 1979. A conceptual deterministic analysis of the kriging technique in hydrology. Water Resour. Res. 15:625-629.

Gray, H. and W. Schucany. 1972. The generalized jack-knife statistics. Marcel Dekker. New York.

Hajrasuliha, S., N. Banibabbassi, J. Mettey, and D. R. Nielsen. 1980. Spatial variability of soil sampling for salinity studies in southwest Iran. Irrig. Sci. 1:197-208.

Hamlett, J. M., R. Horton, and N. A. C. Cressie. 1986. Resistant and exploritory techniques for use in semivariogram analysis. Soil Sci.Soc. Am. J. 50:868-875.

Hawkins, D. M. and N. Cressie. 1984. Robust kriging: a proposal. Math. Geolo. 16:3-18.

Henley, S. 1981. Nonparametric geostatistics. Applied Science Publishers, Halsteaad Press, John Wiley and Sons, New York, NY.

Hoaglin, D. C., F. Mosteller, and J. W. Tukey. 1983. Understanding robust and exploratory data analysis. John Wiley and Sons, New York, NY.

Hochstrasser, U. W. 1965. Orthogonal polynomials. p. 771-802. *In* M. Abramowitz and I.A. Stegun (ed.) Handbook of mathematical functions. Dover Publ. Co. New

York, NY.

Hoeksema, R. J., and P. K. Kitanidis. 1985. Analysis of the spatial structure of properties of selected aquifers. Water Resour. Res. 21:563-570.

Hoffman, G. J., J. D. Rhoades, J. Letey, and Fang Sheng. 1989. Chapter 18. Salinity Management. ASCE Salinity Handbook .

Journel, A. G. and Ch. J. Huijbregts. 1978. Mining geostatistics. Academic Press. New York.

Jowett, G. H. 1952. The accuracy of systematic sampling from conveyor belts. Appl. Stat. 1:50-59.

Jury, W. A. 1985. Spatial variability of soil physical parameters in solute migration: A critical literature review. EPRI topical report EA 4228. Electric Power research Institute, Palo Alto, CA.

Kim, Y. C., D. E. Myers, and H. P. Knudsen. 1977. Advanced geostatistics in ore reserve estimation and mine planning (practitioner's guide). Report to the U. S. Energy Research and Developement Administration, Subcontract No. 76-003-e, Phase II.

Kitanidis, P. K. 1983. Statistical estimation of polynomial generalized covariance functions and hydrologic applications. Water Resour. Res. 19:909-921.

Kitanidis, P. K., and R. W. Lane. 1985. Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method. J. Hydrology 79:53-71.

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 Resour. Res. 19:677-690.

Kolmogorov, A. N. 1941. The local structure of turbulence in an incompressible fluid at very large Reynolds Numbers. Dokl. Akad. Nauk SSR. 30:229-303.

Krige, D. G. 1966. Two dimensional weighted moving average trend surfaces for ore-evaluation. J.South African Inst. Min. Metal. 66:13-38.

Krige, D. K., and E. J. Magri. 1982. Geostatistical case studies of the advantages of lognormal De Wijsian Kriging with known mean for a Base Metal Mine and a Gold Mine. Math. Geol.14:547-556.

Marx, D. and K. Thompson. 1987. Practical aspects of agricultural kriging. Arkansas Agric. Exp. Stn. Bull. 903.

Matheron, G. 1963. Principles of Geostatistics. Econ. Geol. 58:1246-1266.

Matheron, G. 1971. The theory of regionalized variables and it applications. Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, No. 5.

Matheron, G. 1976. A simple substitute for conditional expectation: The disjunctive kriging. p. 221-236. *In* M. Guarascia et al., (ed.) Advanced geostatistics in the mining industry. Reidel Publ. Co., Dordrecht, The Netherlands.

Matheron, G. and M. Armstrong. 1986. Disjunctive kriging revisited, Parts I and II. J. Int. Assoc. Math. Geol. 18:711-742.

McBratney, A. B., R. Webster., and T. M. Burgess. 1981. The design of optimal sampling schemes for local estimating and mapping of regionalized variables. I. Theory and method. Computers Geosci. 7:331-334.

McBratney, A. B. and R. Webster. 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables. II. Program and examples. Computers Geosci. 7:335-365.

McBratney, A. B., R. Webster, R. G. McLaren, and R. B. Spiers. 1982. Regional variation of extractable copper and cobalt in the topsoil of South-east Scotland. Agronomy 2:969-982.

McBratney, A. B., and R. Webster. 1983. How many observations are needed for regional estimation of soil properties. Soil Sci. 135:177-183.

McBratney, A. B. and R. Webster. 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. J. Soil Sci. 37:617-639.

Miyamoto, S. and I. Cruz. 1987. Spatial variability of soil salinity in furrow-irrigated Torrifluvents. Soil Sci. Soc. Am. J. 51:1019-1025.

Morkoc, F., J. W. Biggar, R. J. Miller, and D. R. Nielsen. 1985. Statistical analysis of sorghum yield: A stochastic approach. Soil Sci. Soc. Am. J. 49:1342-1347.

Mulla, D. J. 1988. Using geostatistics and spectral analysis to study spatial pattern in the topography of Southern Washington State, U.S.A. Earth Surface Processes and Landform. 13:389-405.

Myers, D. E. 1982. Matrix formulation of co-kriging. J. Int. Assoc. Math. Geol. 14:248-257.

Myers, D. E. 1983. Estimation of linear combinations and co-kriging. Math. Geol. 15:633-637.

Myers, D. E. 1984. Co-kriging - new developements. p. 295-305. *In* G. Verly et al., (ed.) Geostatistics for natural resources characterization. Part I. Reidel Publ. Co., Dordrecht, The Netherlands.

Myers, D. E. 1984a. A simple sufficient condition for a variogram model to yield positive variance under restrictions. Math. Geo. 16:431-432.

Nielsen, D. R., J. W. Biggar, and K. T. Eeh. 1973. Spatial variability of field measured soil water properties. Hilgardia, 42:215-259.

Olea, R. A. 1975. Optimum mapping techniques using regionalized variable theory. Series on Spatial Analysis No. 2. Kansas Geol. Surv. Lawrence, KS.

Olea, R. A. 1977. Measuring spatial dependence with semi-variograms. Series on Spatial Analysis No. 3 Kansas Geol. Surv. Lawrence, KS.

Oliver, M. A. 1987. Geostatistics and its application to soil science. Soil Use Manage. 3:8-19.

Omer, H. 1984. The variogram and its estimation. In: Verly et al. (Ed.) Geostatistics for Natural Resouces Characterization. by D. Reidel, Amsterdam. 107-125 pp.

Ovalles, F. A. and M. E. Collins. 1988. Evaluation of soil variability in northwest Florida using geostatistics. Soil Sci. Soc. Am. J. 52:1702-1707.

Palumbo, M. R., and R. Khaleel. 1983. Kriged estimates of transmissivity in the Mesilla Bolson, New Mexico. Water Resources Bulletin. No. 131.

Rendu, J. M. 1978. An introduction to geostatistical methods of mineral evaluation, S. Africa Inst. of Mining and Metallurgy. Johannesburg, South Africa. 84 pp.

Rendu, J. M. 1979. Normal and lognormal estimation. Math. Geol. 11:407-422.

Rendu, J. M. 1980. Disjunctive kriging: A simplified theory. J. Int. Assoc. Math. Geol. 12: 306-321.

Rendu, J. M. 1981. An introduction to geostatistical methods of mineral evaluation. South Africa Instit. Min. and Metal., Johannesburg, South Africa.

Rhoades, J. D., D. L. Corwin, and P. J. Shouse. 1988. Use of instrumental and computer assisted techniques to assess soil salinity. p. 50-103. *In* Proc. Symposium on Solonetz Soils, Osijek, Yugoslavia. 15-20 June 1988.

Rhoades, J. D. and J. D. Oster. 1986. Solute content. In A. Klute (ed.) Methods of soil analysis. Part1. 2nd ed. Agronomy 9:985-1006.

Richards, L. A., and Wadleigh, C. H. 1952. Soil water and plant growth. In Soil Physical Conditions and plant Growth, pp. 73-251. Academic Press, New York.

Rogowski, A. S. 1972. Watershed physics. Soil variability criteria. Water Resour. Res. 8:1015-1023.

Rohlf, F. J., and R. K. Sokal. 1981. Statistical tables. 2nd ed. W. H. Freeman and Co., New York. 219 pp.

Russo, D. and E. Bresler. 1981. Soil hydraulic properties as stochastic processes: I. An analysis of field spatial variability. Soil Sci. Soc. Am. J. 45:682-687.

Russo, D. and E. Bresler. 1982. Soil hydraulic properties as stochastic processes. II. Errors of estimates in a heterogenous field. Soil Sci. Soc. Am. J. 46:20-26.

Russo, D. 1984. A geostatistical approach to solute transport in heterogeneous fields and its applications to salinity management. Water Resour. Res. 20:1260-1270.

Russo, D. and W. A. Jury. 1987. A theoretical study of the estimation of the correlation scale in spatially variable fields. I. Stationary fields. Water Resour. Res. 23:1257-1268.

Saddiq, M. H., P. J. Wierenga, J. M. H. Hendrickx, and M. Y. Hussain. 1985. Spatial variability of soil water tension in an irrigated soil. Soil Sci. 140:126-132.

Samura, J. S., K. N. S. Sharma, and N. K. Tyangi. 1988. Analysis of variability in a sodic soil: Structural analysis. Soil Sci. 145:180-187.

SAS Institute. 1985. SAS procedures guide for personal computers. SAS intitute, Inc., Cary, N.Y.

Selim, H. M., B. Davidoff, H. Flühler, and R. Schulin. 1987. Variability of in situ measured mechanical impedance for a fragipan soil. Soil Sci. 144:442-452.

Shouse, P. J., T. J. Gerik, W. B. Russell, and K. D. Cassel. 1990. Spatial distribution of soil particle size and aggregate stability index in a clay soil. Soil Sci. 149:351-360.

Sisson, J. B., and P. J. Wierenga. 1981. Spatial variability of steady-state infiltration rates as a stochastic process. Soil Sci. Soc. Am. J. 45:699-704.

Snedecor. G. W., and W. G. Cochran. 1980. Statistical methods. 7th ed. Iowa State University Press. Ames IA.

Stark, T. H. and J. F. Fang. 1982. The effect of drift on the experimental semivariogram. J. Int. Assoc. Math. Geol. 14:309-319.

Stein, A. 1991. Spatial interpolation. Ph.D. diss. Agricultural University, Wageningen, The Netherlands.

ten Berge, H. F. M., L. Stroosnijder, P. A. Burrough, A. K. Bregt, and M. J. de Heus. 1983. Spatial variability of physical properties influencing the temperature of the soil surface. Agric. Water. Manage. 6:213-226.

Trangmar, B. B., R. S. Yost, and G. Uehara. 1985. Applications of geostatistics to spatial studies of soil properties. Adv. Agron. 38:45-94.

Tukey, J. W. 1977. Exploratory data analysis. Addison-Wesley, Reading, MA.

U. S. Salinity Laboratory Staff. 1954. Diagnosis and improvement of saline and alkali soils. Handbook 60. US Goverment Print Office, Wash. D.C.

van der Zaag, P., R. S. Yost, B. B. Trangmar, K. Hayashi, and R. L. Fox. 1984. An assessment of chemical properties for soils of Rwanda with the use of geostatistical techniques. Geoderma 34:293-314.

Vauclin, M., S. R. Vieira, G. Vachaud, and D. R. Nielsen. 1983. The use of co-kriging with limited field soil observations. Soil Sci. Soc. Am. J. 47:175-184.

Vauclin, M., S. R. Vieira, R. Bernard, and J. L. Hatfield. 1982. Spatial variability of surface temperature along two transects of bare soil. Water Resour. Res. 18:1677-1686.

Vieira, S. R., D. R. Nielsen, and J. W. Biggar. 1981. Spatial variability of field measured infiltration rate. Soil Sci. Soc. Am. J. 45:1040-1048.

Warrick, A. W., G. J Mullen, and D. R. Nielsen. 1977. Scaling field measured soil hydraulic properties using a similar media concept. Water Resour. Res. 13:355-362.

Warrick, A. W., and D. R. Nielsen. 1980. Spatial variability of soil physical properties in the field. pp. 319-344. In: D. I. Hillel (Ed.), Applications of Soil Physics, Academic, New York.

Warrick, A. W., D. E. Myers, and D. R. Nielsen. 1986. Geostatistical methods applied to soil science. pp. 53-82. In: A. Klute (Ed.) Methods of Soil Analysis. Part 1. Monograph 9, American Society of Agronomy, Madison, WI.

Webster, R. 1973. Automatic soil boundary location from transect data. J. Int. Assoc. Math. Geol. 5:27-37.

Webster, R., and Cuanalo de la C., H. E. 1975. Soil correlograms of North Oxford shire and their interpretation. J. Soil Sci. 26:176-194.

Webster, R. 1977. Spectral analysis of gilgai soil. Aust. J. Soil Res. 15:191-204.

Webster, R., and T. M. Burgess. 1980. Optimal interpolation and isarithmic mapping of soil properties. III. Changing drift and universal kriging. J. Soil Sci. 31:505-524.

Webester, R. 1985. Quantitative spatial analysis of soil in the field. IN: B. A. Stewart (ed.), Advances in soil science 3. Springer Verlag, New York, pp. 1-70.

Webster, R. and A. B. McBratney. 1987. Mapping soil fertility at Broom's Barn by simple kriging. J. Sci. Food Agric. 38:97-115.

Yates, S. R. 1986. Disjunctive kriging. III. Co-kriging. Water Resour. Res. 22:1371-1376.

Yates, S. R. and M. V. Yates. 1989. Disjunctive kriging as an approach to decision making. Soil Sci. Soc. Am. J. 52:1554-1558.

Yates, S. R. and A. W. Warrick. 1987. Estimating soil water content using co-kriging. Soil Sci. Soc. Am. J. 51:23-30.

Yates, S. R., A. W. Warrick, and D. E. Myers. 1986a. Disjunctive kriging. I. Overview of estimation and conditional probability. Water Resour. Res. 22:615-622.

Yates, S. R., A. W. Warrick, and D. E. Myers. 1986b. Disjunctive kriging. II. Examples. Water Resour. Res. 22:615-622.

Yeh, T. C. J., L. W. Gelhar, and P. J. Wierenga. 1986. Observations of spatial variability of soi-water pressure in a field soil. Soil Sci. 142:7-12.

Yost, R. S., G. Uehara, and R. L. Fox. 1982a. Geostatistical analysis of soil chemical properties of large land areas. I. Semi-variograms. Soil Sci. Soc. Am. J. 46:1028-1032.

Yost, R. S., G. Uehara, and R. L. Fox. 1982b. Geostatistical analysis of soil chemical properties of large land areas. II. Kriging. Soil Sci. Soc. Am. J. 46:1033-1037.