12-1995

Evaluating the Feasibility of Sequential Indicator Simulation in Reproducing Spatial Connectivity in a Heterogeneous Transmissivity Field

D. Duane Kenyon

EVALUATING THE FEASIBILITY OF SEQUENTIAL INDICATOR SIMULATION IN REPRODUCING SPATIAL CONNECTIVITY IN A HETEROGENEOUS TRANSMISSIVITY FIELD

THESIS

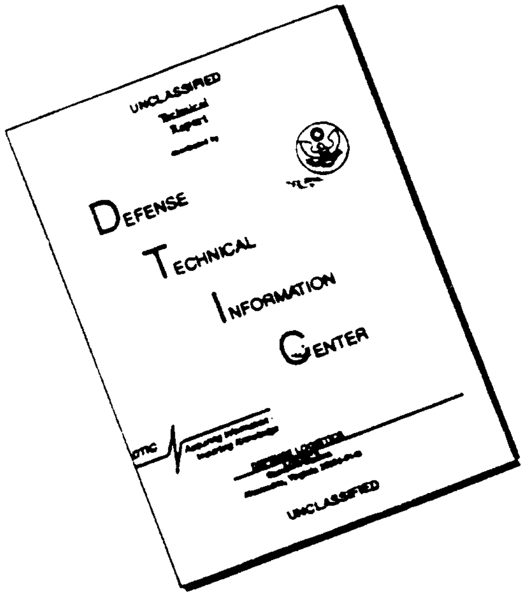
D. Duane Kenyon, 1st Lieutenant, USAF

AFIT/GEE/ENP/95D-03

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

AFIT/GEE/ENP/95D-03

EVALUATING THE FEASIBILITY OF SEQUENTIAL INDICATOR SIMULATION IN REPRODUCING SPATIAL CONNECTIVITY IN A HETEROGENEOUS TRANSMISSIVITY FIELD

THESIS

D. Duane Kenyon, 1st Lieutenant, USAF

AFIT/GEE/ENP/95D-03

AFIT/GEE/ENP/95D-03

EVALUATING THE FEASIBILITY OF SEQUENTIAL INDICATOR SIMULATION IN REPRODUCING SPATIAL CONNECTIVITY IN A HETEROGENEOUS TRANSMISSIVITY FIELD

D. Duane Kenyon

1st Lieutenant, USAF

Approved:

Richard A Hartley 1 Dec 1995

Lt Col Richard Hartley

Edward Heyse 1 DEC 95

Maj Edward Heyse

Dan R. Reynolds 1 Dec 95

Prof. Dan Reynolds

DISCLAIMER:

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. government.

EVALUATING THE FEASIBILITY OF SEQUENTIAL INDICATOR SIMULATION IN REPRODUCING SPATIAL CONNECTIVITY IN A HETEROGENEOUS TRANSMISSIVITY FIELD

THESIS

Presented to the School of Engineering
of the Air Force Institute of Technology
Air University
In Partial Fulfillment of the Requirements for the
Degree of Master of Science in Engineering and Environmental Management

D. Duane Kenyon, B.S.
1st Lieutenant, USAF

December 1995

Approved for public release; distribution unlimited

Acknowledgments

I'd like to thank God for giving me the abilities necessary to complete this project; my wife Kelli for having gobs of patience while providing me sustenance and support; Col Hartley for his flexibility, guidance, and encouragement; and Maj Heyse and Professor Reynolds for their valuable time and much needed suggestions.

Abstract

A Non-Parametric estimation technique was used to simulate realizations of a heterogeneous transmissivity field based upon sampled values from three different sampling scenarios. These realizations were compared to output from a parametric estimation technique with respect to truth as defined by an exhaustive data set of 6,000 transmissivity values. Estimated transmissivity fields were then used as input into a flow model from which fields of heads and specific discharges were obtained and compared. Given the financial limitations imposed upon the number and quality of samples reasonably available, Sequential Indicator Simulation, a non-parametric technique, was shown to be of considerable value when accompanied with sound geological input.

Table of Contents

Table of Figures

Table of Tables

EVALUATING THE FEASIBILITY OF SEQUENTIAL INDICATOR SIMULATION IN REPRODUCING SPATIAL CONNECTIVITY IN A HETEROGENEOUS TRANSMISSIVITY FIELD

I. Introduction

I.1. Motivation

In light of the recent focus upon environmental restoration, the Air Force is forced to make important decisions concerning groundwater remediation which are heavily stressed by the opposing considerations of financial impact and ethical responsibility. It is necessary, therefore, to supply Air Force decision-makers with information detailing the movement of groundwater that is subject to as little uncertainty as possible and that is accompanied with an indicator of that uncertainty.

The hydrogeological parameter most influential in the synthesis of the velocity vectors that describe groundwater flow is hydraulic conductivity. This parameter, which is a measurement of the ability of soil to conduct water, is also termed *transmissivity* when it is expressed as an average over the depth of an aquifer. The relative magnitude and spatial arrangement of transmissivities across a field of interest will dictate the nature of groundwater flow. Unfortunately, there are severe financial and technological limitations upon the quantity and quality of transmissivity data that can normally be obtained. As a result, it is common practice to simply assume a constant transmissivity value over the entire field of interest. At best, a very limited number of discrete transmissivity measurements within a domain of interest are used to estimate the phenomenon over the entire domain by some moving average technique such as kriging. The resulting map of transmissivities would then be used as input into a flow modeling package (transfer function) to produce a field of head values and a field of specific discharges (response variables). Given this output, decision-makers are presumed to gain an adequate understanding of where contaminated groundwater is heading and when it will arrive. It is naive and potentially dangerous to assume, however, that any modeled flow field has accurately and precisely represented reality and is thus reliable information. It would be advantageous, therefore, to investigate an estimation methodology that will reproduce the *most critical* spatial characteristics of a transmissivity field, that will accommodate the less precise but cheaper and more prolific "soft" information, and that will allow for some quantification of uncertainty.

I.2. The Issues

Consider a phenomenon which is described by the spatial and temporal regionalization of a parameter over some domain. Specifically, the flow of water through a porous medium which is described by the spatial regionalization of transmissivity values, $t(\mathbf{u})$, over the domain, D, of which $\mathbf{u}$, a coordinate vector, is an element.

Classically the regionalization is interpreted as a particular outcome of the random function $\{T(\mathbf{u}), \mathbf{u} \in D\}$ (Journel, 1983:446). Within this domain it is only possible to know a limited number of values for the parameter of interest which are $\{t(\mathbf{u}_\alpha) = t_\alpha, \alpha = 1 \text{ to } N\}$ (Journel, 1983:446). It is from the information contained in these limited number of data that an attempt will be made to estimate reality. If the data are well-behaved (Coefficient of Variation, CV, < 1), then a record of successes has been established using Gaussian estimation techniques such as Ordinary Kriging (Journel, 1983:445). If the distribution of the data departs from normality, problems arise using any of the Gaussian techniques. This fact is relevant when considering the random function $T(\mathbf{u})$, since transmissivities have been shown to be distributed log-normally (Freeze, 1975:728). Two options for normalizing log-normal distributions -- (1) omitting the high-valued data or (2) smoothing out the data by passing it through some transform such as the natural logarithm -- have both been deemed unacceptable (Journel, 1983:445-446). It is, in fact, the characteristic of the log-normal distribution that distinguishes it from normality which contains the most relevant information. Because the spatial arrangement of these extreme transmissivity values contained in the distribution's tail are likely to dictate the flow path, omitting outlying data is unacceptable. A natural-log transformation of the data presents problems because it is non-linear while most estimation techniques are linear, thus creating biased results when the estimated data are transformed back (Gilbert, 1987:103). If the effects of a transformation can be understood and corrected for, a natural-log transform may be helpful, but usually only for data sets in which the ratio of the largest datum to the smallest datum is greater than about 20 (Gilbert, 1987:149). As a result, it is important that the estimation technique be distribution-free, thus adhering to the original intent of geostatistics which is to put the data before the model (Journel, 1983:446).

A map of estimated transmissivities is a poor model of reality if it does not reflect those characteristics of the spatial function which *most* affect the response function (Journel and Alabert, 1990:212). If the transmissivity field being estimated contains a significant amount of connectivity among the extreme values, which is not uncommon given fluvial deposits, it will be this feature that will most affect the response function and will be the most critical to reproduce during estimation. If there is a continuous band of very high transmissivities, it will become a preferential flow channel; but if there are low transmissivities blocking that flow at any point, the flow could be completely stopped and rerouted. As a result, if preferential flow channeling exists or is suspected to exist, an estimation technique which is able to reproduce the connectivity of extreme values would be most appropriate.

Furthermore, it is important that the estimation technique be flexible enough to accept "soft" or "fuzzy" data. Because of the expense involved, hard data are usually meager. Consequently, a "good" reproduction of the spatial distribution of the transmissivity values is unlikely. Soft data, usually characterized by constraint intervals or by probabilities determined *a priori*, are much more available. The sources of soft data are numerous, including expert geological interpretation, seismic data, and soil grain analyses. Although less precise, soft data still contains valuable information about the site that often times goes unused since most estimation techniques are not robust enough to utilize it.

Finally, since estimation results can be misleading, they must be accompanied by some measurement of uncertainty. Most estimation techniques cannot satisfy this requirement primarily because they can only supply a single estimated map based upon some optimization criteria. Techniques that can provide some quantification of uncertainty are those that render any number of equiprobable "simulations," the ensemble of which will depict a picture of uncertainty. Just such a simulation technique that is distribution-free, that has demonstrated success in capturing the connectivity of extreme values, and that lends itself to the utilization of soft information is Sequential Indicator Simulation (SIS).

This work will attempt to investigate the feasibility of SIS in characterizing hypothetical field transmissivity values that suggest the existence of well-connected fluvial channel features which would modify regional flow velocity fields due to controlling high transmissivities. First, a univariate analysis will be conducted on the exhaustive transmissivity field followed by a spatial description of the transmissivity covariance. Then, applying three different sampling scenarios, a limited number of hard and soft data will be extracted from truth which will be used for estimation and simulation. An estimation technique that employs a weighted-moving-average (Ordinary Kriging) will be used to identify the limitations of parametric techniques and SIS will be used to obtain equiprobable realizations of the flow field to identify the effectiveness of a non-parametric technique. The spatial characteristics of the resulting realizations from both the parametric and non-parametric techniques will be compared to truth and, finally, a selected number will be used as input to a transfer function to generate heads and specific discharges for further comparison.

II. Background

Several statistical techniques have been developed and employed for the estimation of unknown values over a field of interest given a subset of known information. The goal of some techniques has been to quantify the uncertainty associated with the quantity and quality of the available sample data. To accomplish this, approaches that describe the random function in probabilistic terms are used to generate conditional distributions which honor the available data at their locations and which reproduce their estimated spatial statistics. Using these conditional distributions, Monte Carlo type simulations can be accomplished sequentially across the entire domain of interest to produce multiple realizations of the random function which honor the original data and reproduce its spatial statistics. This probabilistic approach is succinctly described as a stochastic conditional simulation and it is the basis upon which SIS is built.

Another method commonly used for estimating unknown values is that of direct estimation. These techniques, including Ordinary Kriging, provide a unique image which is based upon some optimization criteria. Direct estimation is conditional in that it does honor the data at their locations, but it does not reproduce the estimated spatial statistics. Given that the goal of direct estimation is to optimize some criteria such as the minimization of error variance, its singular output appears rigid and smoothed. In his thesis effort, Alabert presents a figure similar to Figure 1 which illustrates the difference between stochastic conditional simulation and direct estimation (1987:3).

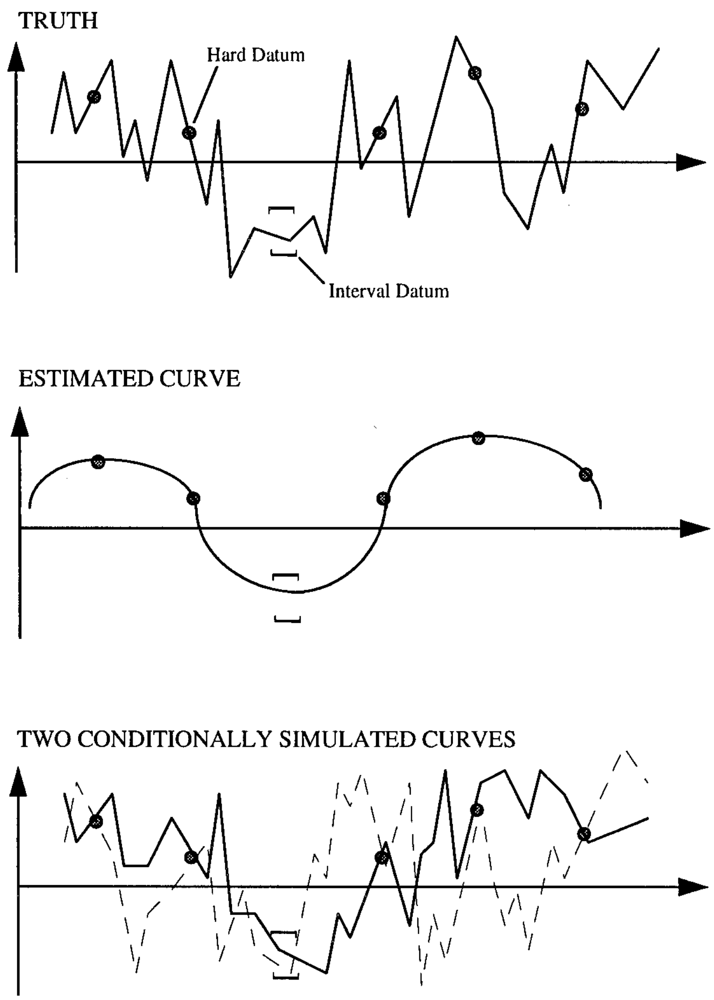


Figure 1 -- Direct Estimation vs. Stochastic Conditional Simulation

II.1. Direct Estimation

The recognition that there is spatial correlation in most Earth Science data sets has allowed for the estimation of unknown values at specific locations. This idea of spatial correlation implies that the data field of interest is continuous across its entire domain and thus allows point estimation using known data points. One of the earliest methods of point estimation was to simply draw bisectors between all known data points defining polygonal areas of influence and then assume that every unknown location within an area was equal to the known value. This method was inadequate, however, because it left major discontinuities between polygons. A technique that overcame this shortcoming was that of assuming a linear relationship between two known values and then drawing lines between those in close proximity to create triangular planes which define the values of unknown points falling within them. While this method of triangulation was an improvement, only three values were used to estimate any one value, thus neglecting the information contained in other surrounding values. As a result, estimation techniques using all nearby known values were developed. The basis for these techniques lies in the simple act of defining some local neighborhood around a unknown point and assigning the neighborhood mean as its value. While this basis, in itself, is not a good estimator, estimates allowing for the weighting of the known values' influence upon the neighborhood mean according to their distance from the point of interest showed improvement. One method of determining weights was that of assigning them in proportion to the inverse of their squared distance from the unknown value. This had the effect of giving the closest samples more influence and the furthest samples less influence and thus provided a better estimate. All of the procedures that had been used up to this point may be appropriate depending upon the goals of the study, however none of them were specifically aimed at reducing the standard deviation of the error residuals; this is, however, the goal of Ordinary Kriging (OK). Before a discussion of geostatistics can take place, it would be helpful to review the major assumptions that must be made prior to its use.

II.1.1. Assumptions of Stationarity and Ergodicity

As was mentioned earlier, the phenomenon of interest, $t(\mathbf{u})$, is a singular realization of the random variable, $T(\mathbf{u})$, the set of which constitutes a random function (Journel and Huijbregts, 1991:29). Ideally, one would like to assume that the phenomenon is homogeneous over the area of interest indicating that the distribution function between any two data points, $t(\mathbf{u}_1)$ and $t(\mathbf{u}_1 + \mathbf{h})$, will remain unchanged with respect to the separation vector, **h**. This assumption is not always possible, however, nor is it always necessary. As a result, lesser degrees of homogeneity, as defined under the concept of stationarity, can be assumed. (The description of stationarity using the notion of homogeneity is used purely for illustrative purposes as stationarity is a property of the model and homogeneity is a subjective impression of the population (Journel, 1986:139).)

The idea of pure homogeneity presented above is defined as strict stationarity; other relevant degrees of stationarity are second order stationarity and quasi-stationarity. If the probability distribution is a function of the vector separating two points then second-order stationarity is implied and if it depends only upon the modulus, it is considered statistically isotropic (Neuman, 1982:84). The assumption of second-order stationarity is usually appropriate for geostatistical applications and will be referred to in this work as simply "stationarity". Quasi-stationarity implies that although there may be a trend present across the site, stationarity may be assumed if it exists within the local area as defined by the magnitude of the vector **h** used in defining the range of the correlation structure. (Journel and Huijbregts, 1991:88).

The assumption of ergodicity means that the probability law governing $T(\mathbf{u})$ can be derived either from repeated sampling of an ensemble of statistically equivalent media at a given point in space (i.e. multiple realizations of the random function), or from samples collected at different points in a single medium (i.e. a single realization of a random function) (Neuman, 1982:83). In reality, this assumption must be adopted since we can only deal with a single realization which is truth (Neuman, 1982:83).

II.1.2. Ordinary Kriging

Ordinary Kriging was introduced in 1962 by Goldberger as an improvement over the classical linear regression model in that it took advantage of the interdependence between the observations of the independent variable to obtain a gain in efficiency or a reduction in the prediction variance (Goldberger, 1962). Because of this characteristic, Ordinary Kriging is commonly described using the acronym BLUE, or Best Linear Unbiased Estimator (Isaaks and Srivastava, 1989). All of the estimation techniques discussed to this point were linear, theoretically unbiased estimators, however none of them specifically minimized the standard deviation of the error residual. This is the goal of Ordinary Kriging, hence the descriptor "best". Ordinary Kriging is accomplished by first describing the spatial continuity of the stationary field with a covariance or variogram structure. Using this structure, weights, determined to ensure unbiasedness and minimized variance, are assigned to known data points surrounding an unknown data point and the unknown locale is estimated using a linear combination of the surrounding values multiplied by their respective weights. Ordinary Kriging has been used extensively for estimation and has been used as a tool and a springboard for many other estimation techniques.

Since Ordinary Kriging is a series of moving averages it tends to produce results with a Gaussian distribution. As a result, the continuity of the extreme values will always be small (Gómez-Hernández, 1991:61). This phenomenon is easily seen on gray-scale kriged maps as the grays will dominate the field while the blacks and whites appear as small discontinuous spots. This smoothing of conductivity fields may be unrealistic and may yield biased responses (Gómez-Hernández and Srivastava, 1990:395). While minimum error variance may be a relevant goal in certain cases, if reproduction of the connectivity of extreme values is the goal, a smooth surface and minimum-error variance may be deemed irrelevant and thus disregarded (Journel and Alabert, 1990:212). Furthermore, kriging does not lend itself to a determination of uncertainty. While it does provide a variance at each point of the kriged field, this is only a measure of local accuracy and not a measure of the joint spatial uncertainty which would be necessary since the calculated value of hydraulic head is dependent not upon one conductivity value but upon the field of values taken together (Gómez-Hernández and Srivastava, 1990:395).

II.2. Conditional Stochastic Simulation

Because any estimation of an unknown parameter must involve some uncertainty, the need was recognized to describe groundwater parameters in statistical terms. This led to the development of theoretical models whose parameters are treated as stochastic variables rather than deterministic ones; therefore the model output is described in probabilistic terms rather than in deterministic terms. (Neuman, 1982). A single discrete realization with kriging variances cannot be used to assess estimation uncertainty; a Monte Carlo approach, however, provides a convenient way to assess uncertainty (Gómez-Hernández and Srivastava, 1990:395). As opposed to a single over-smoothed kriged map, several realizations are put through a transfer function, intended to model some response to the input parameters, and a frequency distribution of the response variable is built which quantifies its uncertainty (Gómez-Hernández, 1990:60).

The earliest attempts at a stochastic model worked under the assumption that the conductivity values were statistically independent (Neuman, 1982). Neuman cites the work of Warren and Price (1961) as they attempted to determine the effects of different heterogeneities as generated by various frequency distributions upon the definition of hydraulic conductivity. A notable conclusion of their work was that it is possible to define an equivalent uniform medium for transient flow in a nonuniform medium. Freeze questioned this theory concluding that there was probably no way at all to replace a heterogeneous medium with an "equivalent" uniform conductivity field (1975:738). Probably a more important contribution of Freeze's work, however, as noted by Neuman (1982:86), is the recognition of the need to account for the correlation structure of the simulated parameter. He notes that "for a stochastic model to yield reliable results, it must take into account the manner in which point values are distributed in space" (Neuman, 1982:86). One approach has been to view the spatial variation of hydraulic conductivities as a random field characterized by spatial covariance functions (Neuman, 1982:86). Smith and Freeze (1979) were able to go beyond the statistical independence of Freeze's 1975 work by introducing autocorrelation between discrete conductivity values using a first-order autoregressive scheme called the "nearest neighbor" model (Neuman, 1982:89). While stochastic simulations had successfully incorporated autocorrelation of parameters and were reproducing some of the global statistical properties of the "true" realization, they were still not completely accurate since they neglected to honor the known values in the simulated fields. Delhomme recognized this shortcoming and introduced "conditional" simulation (1979). Conditional simulation, which was based upon the theories of kriging, emphasized two particular features -- first, simulated T values must have the same autocorrelation as the true T field; and second, locations with measured values must be consistent with those values (Delhomme, 1979).

The generation of equiprobable fields with a common covariance structure is termed "stochastic simulation," and if each of the realizations honors the known data at the sample locations it can be called "conditional stochastic simulation" (Gómez-Hernández and Srivastava, 1990:395). Thus a conditional simulation must meet two requisites: (1) it must produce equiprobable maps with a given covariance structure and (2) it must honor the data at their locations (Gómez-Hernández and Srivastava, 1990:396). This does not indicate however, that any algorithm that meeting the above criteria is appropriate. Journel (1974) developed a Gaussian-related model which was both a conditional and a stochastic simulation, yet Journel and Alabert showed that this algorithm produced an excess of local noise which "blurred" the strong connectivity of extreme values (1989:123-124). Again, it is apparent that the averaging and smoothing characteristics of a Gaussian related technique are relevant and must be considered and understood before it can be used effectively. If there is connectivity of extreme values or if there are extreme values arranged in close proximity to one another, SIS may be an appropriate conditional stochastic simulator. SIS does not share the limitations of a Gaussian-related technique because it is based upon the use of a non-parametric estimation technique called indicator kriging.

II.2.1. Indicator Kriging

Indicator kriging (IK) was introduced in 1983 by Journel as an answer to the problems presented by the assumptions inherent in parametric estimation techniques. He stated that for highly variant data (Coefficient of Variation of 2-5) parametric approaches were useless, and he offered indicator kriging as a solution which would allow a more comprehensive structural analysis and be more robust with respect to outlier values. The use of IK is simple since it utilizes the algorithm of Ordinary Kriging. Instead of averaging measured values however, IK transforms values into a series of zeros and ones based upon their magnitude with respect to a chosen cutoff value, t_c:

$$i(\mathbf{u};t_c) = \begin{cases} 1, & \text{if } t(\mathbf{u}) \le t_c \\ 0, & \text{if not} \end{cases} \tag{1}$$

where $\mathbf{u}$ is the coordinate vector of the estimated point. As a function of t, $i(\mathbf{u};t_c)$ can be seen as a cumulative distribution function (cdf) at $\mathbf{u}$ and with respect to a single cutoff. (Journel, 1983:447). Thus, at each point in the domain, a complete cdf can be generated which describes the probabilities of the unknown value falling within intervals as discretized by the series of selected cutoffs.

Having access to an estimated distribution as opposed to relying upon a prior assumption of normality allows for better estimation of data distributed with long tails. Ordinary kriging is focused on minimizing variance and, as a result, will not reproduce values in the tails of data sets with large variances. As a result, OK becomes an unreliable estimator of a global distribution (Isaaks and Srivastava, 1989:421).

II.2.2. Sequential Indicator Simulation

Gaussian random function models are unable to reproduce the connectivity of extremes since they maximize spatial disorder, or entropy (Journel, 1989:33) Thus the Gaussian model precludes the clustering of high and low values which create the barriers and conduits to flow which are present in many conductivity fields (Desbarats and Srivastava, 1991:688). In SIS, the spatial continuity of the variable of interest is divided into classes and the continuity of each class is explicitly accounted for using IK. As a result, if a single class demonstrates high continuity, the variogram function will describe a long range of correlation (Gómez-Hernández, 1991:61).

SIS was introduced in 1990 by Gómez-Hernández and Srivastava as a simulation technique that honors the initial data and that can feature a "richer" family of spatial structures that are not limited by Gaussianity (395). The goal of their work was to provide an algorithm by which several equiprobable realizations can be generated to evaluate uncertainty without the limitations of parametric techniques (Gómez-Hernández and Srivastava, 1990:395-396). SIS is also different from other conditional stochastic simulation methods in that it is conditional by construction, i.e. each simulated value is added to the field of known values to be included as part of the simulation of the next value (Gómez-Hernández and Srivastava, 1990:396). As a result each simulated value is conditioned upon all known data as well as previously simulated data. The simulation of a point is accomplished by using IK to construct a cumulative distribution function (cdf) for the location and then generating a value by Monte Carlo. The entire SIS algorithm is presented in Chapter III.

Two important advantages of SIS that have been theoretically and experimentally scrutinized are its effectiveness in the reproduction of connectivity of extreme values and its ability to allow for the incorporation of soft data. At Stanford University, Journel and Alabert have written several papers detailing a case study performed using permeability measurements on a slab of Berea sandstone (Journel and Alabert, 1988; 1989; 1990). In these works they suggested ways in which the spatial connectivity might be measured and they concluded that stochastic conditional simulators using the indicator formalism are able to reproduce connectivity as measured by indicator correlograms and that they allow for the incorporation of soft data (Journel and Alabert, 1988; 1989; 1990). In another work, Gómez-Hernández recommends the use of SIS for the generation of hydraulic conductivity fields with significant connectivity of extreme values and he cautions against the use of multi-Gaussian related techniques (1991). While a large amount of research remains regarding the effects of incorporating soft data, Wen and Kung have shown that the addition of large amount of soft data provides better characterization of reference conductivity fields and improved transport simulation results, i.e. improved output from a transfer function (1993).

Furthermore, SIS may prove advantageous over other techniques given its capacity to generate multiple realizations that produce considerably different effects when passed through a transfer function while still honoring the original data. In a study on the Miami-Valley buried aquifer system, Ritzi and others concluded that, because of the range of outcomes that conditional indicator simulation was able to produce, it was preferred over those methods which produced only one possibility based upon some local accuracy sense (1994:672-673). After using a technique similar to SIS at the Hanford Site near Richland, WA, Poeter and Townsend stated that "a true evaluation of the possible subsurface configurations and their impact on the decision at hand is the only honest approach to groundwater analysis," thus concluding that "the era of drawing conclusions on the basis of deterministic flow and transport models has come to an end" (1994;447).

III. Methodology

III.1. Overview

A necessary strategy for evaluating any estimation methodology is to check the composite whole of its results against the single realization which is truth (Desbarats and Srivastava, 1991). Figure 2 illustrates how this approach might be employed to evaluate the estimation of transmissivity values in relation to their impact upon groundwater flow. Provided with a subset of truth from a sampling exercise, an estimation methodology is used to predict remaining values in an effort to approximate the entirety of truth. Each estimated field is then used as input into a transfer function from which the results can be compared to the output of reality as passed through a transfer function and conclusions can be drawn. This side-by-side comparison of what is reality and what is realistic is ideal in that efforts can be exerted toward narrowing the margin of difference between the two and then applying the knowledge to real-life situations.

The obvious problem with this technique is the acquisition of truth; either a site must be sampled so extensively as to reveal truth, or a prototype must be constructed. Detailed field studies, such as the experiments at Borden (Woodbury and Sudicky, 1991) and at the MADE site in Columbus MS (Rehfeldt and others, 1992) have been the source of valuable information, but are very cost intensive. Furthermore, since groundwater flow occurs in the hidden subsurface regions, these field studies do not allow complete

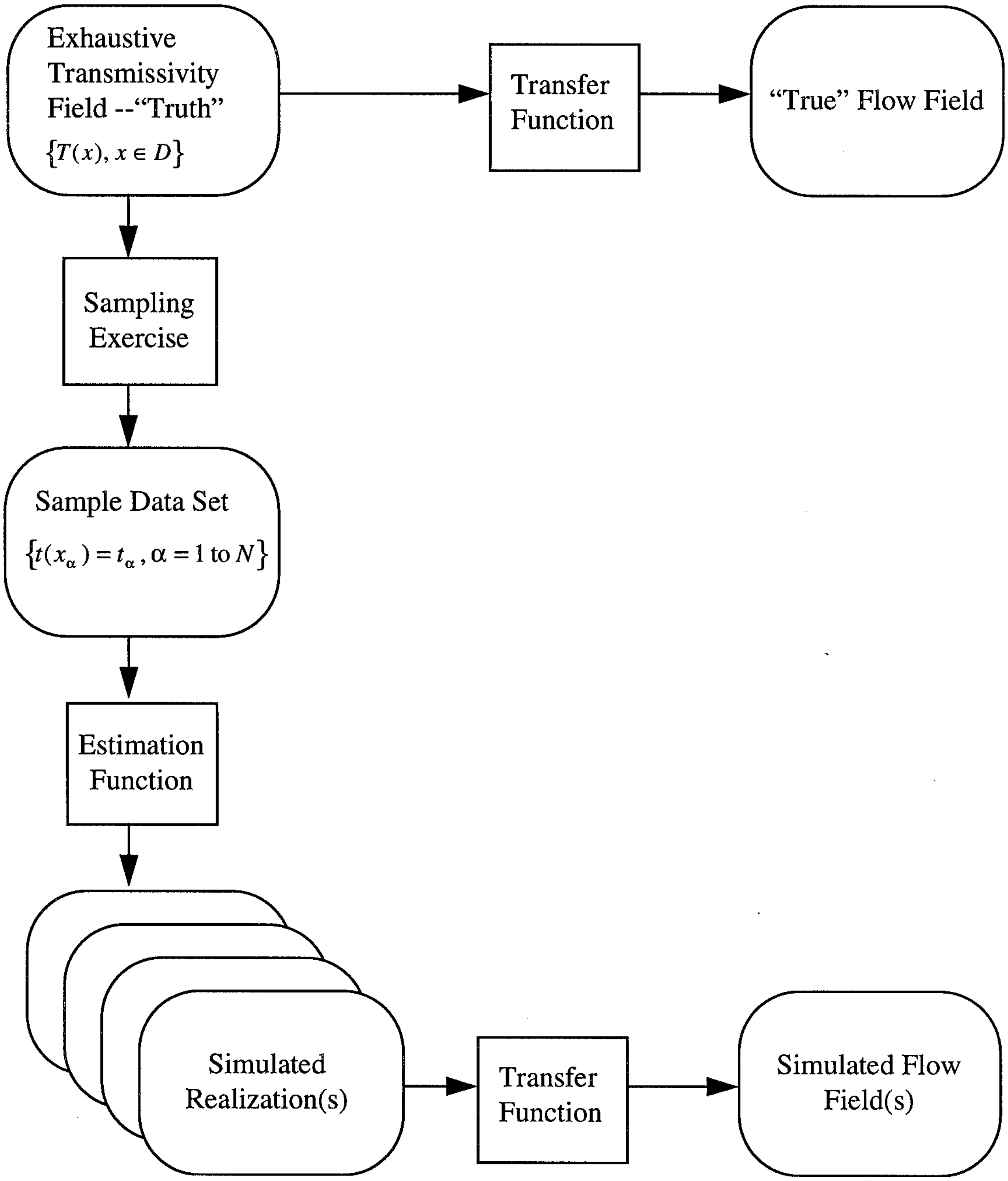


Figure 2 -- Methodology Results vs. Truth

understanding and control over experimental conditions which would allow the explicit interpretation of observations and the assessment of theory (Desbarats and Srivastava, 1991). In other words, they would not allow for a direct assessment of an estimation methodology without some consideration of the transfer function's ability to imitate nature. Fortunately an acceptable alternative is to employ the use of hypothetical transmissivity fields. This option is cost effective and is subject to the complete knowledge and control necessary for the direct assessment of theories of estimation. Using this approach, truth would be exhaustively known, different sampling scenarios could be considered, and the response variables would not be subject to differences in transfer functions.

To accomplish the goals of this work -- that is an evaluation of the ability of SIS to reproduce connectivity, to accommodate soft data, and to quantify uncertainty -- analysis will be applied to a prototype data set, thus permitting a complete statistical audit of the methods considered. It is postulated that the data set constitute a two-dimensional heterogeneous transmissivity field that typifies the connectivity of extreme values as deposited by fluvial outwash. The 6,000 values constituting the field will be regarded as the exhaustive realization of truth and will act as a paragon next to which results will be compared. The first step of the methodology applied in this work will be to conduct a univariate and bivariate analysis of the reference data set whereby the critical statistics and characteristics to be reproduced will be established and understood. Next, three sample data sets will be extracted, statistically examined, and used as input for Ordinary Kriging and SIS. Output from these estimation techniques will be qualitatively and quantitatively compared to the exhaustive data set and will be used as input into MODFLOW for comparison of the resulting fields of heads and specific discharges.

III.2. The Walker Lake Data Set

The data set used in this work is a transformed subset of the Walker Lake data set as used by Alabert (1989) and Isaaks and Srivastava (1989) which was simulated using an analog method to obtain an outcome free of the artifacts of multivariate Gaussian models inherent in most simulation techniques (Desbarats and Srivastava, 1991:687). The Walker Lake data set was originally derived from a digital elevation model of an area in western Nevada (Isaaks and Srivastava, 1989:4). The original data set contained 1.95 million elevation values which were grouped into 5 x 5 blocks providing a regular grid of 260 blocks east-west by 300 blocks north-south. The variable from that data set used for this work is the "*V*" value, as anonymously designated by Isaaks and Srivastava, which was obtained as a function of the block means and variances (1989:545). Since *V* was derived from elevations, it demonstrates the spatial continuity common in most natural data sets (Isaaks and Srivastava, 1989:50-55). It's map exhibits features just like the topographic features shown in Figure 3.

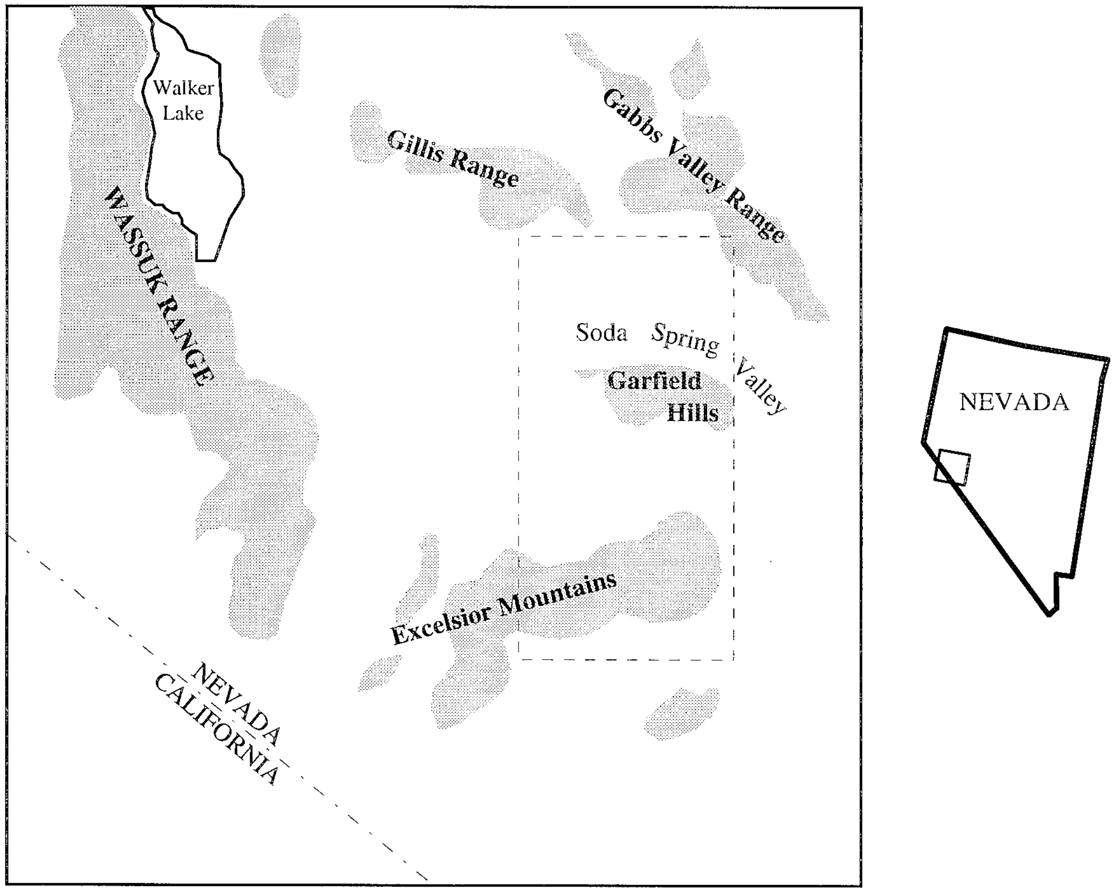


Figure 3 -- Major Features of Walker Lake Data Set (Isaaks and Srivastava, 1989:5)

The variable, *V*, of Isaaks and Srivastava was further transformed into a transmissivity value by Desbarats and Srivastava (1991:688) using the following relation:

$$T = \exp\left(12.4 - 1.47\ln\left(V + 100.0\right)\right) \qquad (2)$$

According to Desbarats and Srivastava this transformation serves three purposes: it transforms low *V* values into high *T* values, thus creating large-scale high-transmissivity features in the hypothetical field; the histogram of the *T* values resembles the standard lognormal model; and the *T* values are scaled to be within the general range of published transmissivity values expressed in square meters per day (1991:688). In a more subjective assessment of the hypothetical field, it is argued that the meandering high-transmissivity region could represent point bar "shoestring" sandstones while the low transmissivity areas could represent floodplain, overband and levee fine-grained deposits (Desbarats and Srivastava, 1991:688).

For the purposes of this work, a 60 x 100 subset was taken from the set of 78,000 *T* values. This subset, identified by the dotted line in Figure 3 as part of the Soda Springs Valley, provides a more manageable data set as well as one with distinct areas of connectivity which resulted from the original low elevation data throughout the valley. Figure 4 is a gray-scale map of this area which will be referred to as the Soda Springs data set. The white regions highlight the areas of high transmissivity.

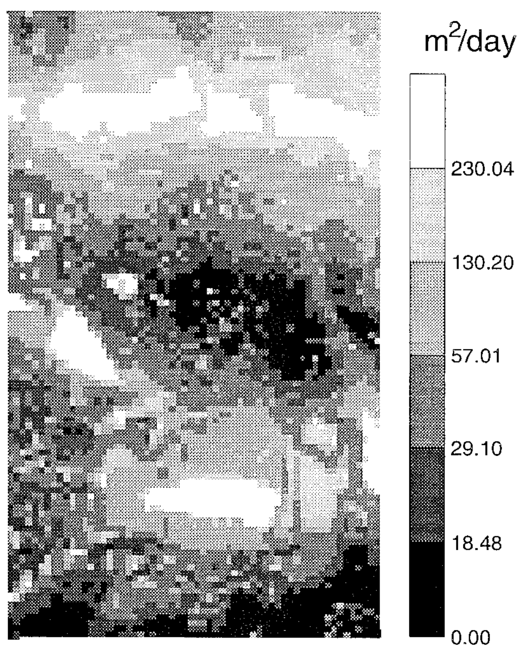


Figure 4 -- Map of 60 x 100 Transmissivity Values Comprising the "Soda Springs" Data Set.

III.3. Statistical Analysis

The statistical analysis of the exhaustive transmissivity field will be accomplished to serve two purposes: first, it will provide an excellent tool to introduce the methods of presenting and describing the data; and second, it will establish the standards for later comparison. The univariate analysis will be conducted to show the characteristics of the distribution of the data, then the important consideration of data location will be incorporated as the spatial features of the data set are described.

III.3.1. Univariate Analysis

Probably one of the most common descriptions of a data set, both because of its simplicity and its importance, is the histogram. The histogram of the exhaustive transmissivity field, shown as Figure 5, reveals some characteristics of the data set which are important to note. First, the data are lognormally distributed as would be expected for a transmissivity field (Freeze, 1975:728). But more significant is the bimodality of the distribution. The spike on the extreme right is a direct result of the large number of the high transmissivity values which, as can be seen in Figure 5, are spatially arranged in a manner that creates a preferential flow path, thus dictating flow direction and time. It is this portion of the distribution that Gaussian estimation techniques should have the most difficulty reproducing.

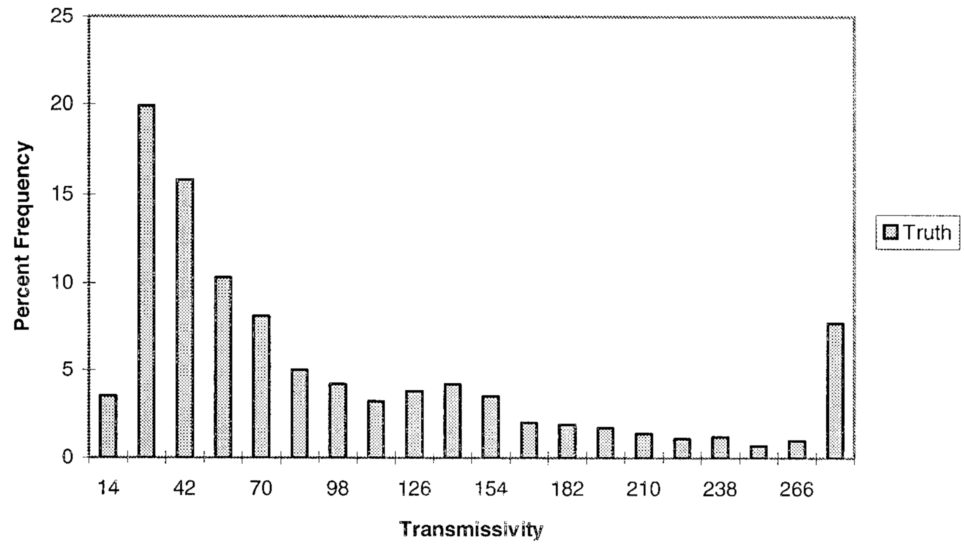


Figure 5 -- Histogram of Exhaustive Transmissivity Values

Statistics which describe the important features of the histogram are also beneficial to consider and understand. Table 1 contains these summary statistics for the exhaustive transmissivity field.

Table 1 -- Summary Statistics for Exhaustive Transmissivity Field

Mean	89.48
Standard Deviation	79.17
Variance	6267.73
Coefficient of Variation	0.88
Minimum	8.43
10th Percentile	18.48
25th Percentile	29.10
Median	57.01
75th Percentile	130.20
90th Percentile	230.04
Maximum	278.77
Coefficient of Skew	1.19

The characteristics of the data set that are most remarkable are the variance and the Coefficient of Variation (CV). The variance is large relative to the mean and, therefore, should be difficult to reproduce using a minimum variance estimation technique. The variance is not so large, however, as to preclude the use of a parametric methodology altogether. As was stated earlier, when the CV is greater than about two, parametric techniques are useless. In fact, when a transformation was applied to the Soda Springs data set increasing the CV to 2.73, estimation attempts using Ordinary Kriging proved futile. By using a data set with a CV of less than two, the opportunity exists to examine the performance and potential advantages of a non-parametric technique in a situation in which a parametric technique is plausible.

Also noteworthy are the characteristics of the cdf as defined by the percentiles. Reproduction of the univariate statistics of the exhaustive data set will be judged, in part, on the adequacy of an estimation methodology to reproduce the cdf. Furthermore, the percentiles noted in the table will be used as cutoffs for the definition of the spatial arrangement of the data set throughout this work; e.g., in Figure 4 the white values are those above 230.04 and the black values are those below 18.48. This will allow for the presentation of results in a fashion which is easily related to "truth." Those values above the 90th percentile (or the white areas on the gray-scale maps) will be the primary focus in this work as they define the preferential flow paths. Because the values at the listed percentiles will carry such importance in the definition of truth they will also be used as the discretizing cutoffs within the SIS algorithm.

III.3.2. Spatial Continuity

The spatial continuity of a data set has traditionally been defined in geostatistics with the semi-variogram function. This function written as:

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} (x_i - y_i)^2 \tag{3}$$

where $N(\mathbf{h})$ is the number of pairs, x_i is the value at the tail of the **h** vector and y_i is the value at the head of the vector, measures the average degree of dissimilarity between some unsampled value, $t(\mathbf{u})$ and a nearby value, $t(\mathbf{u}+\mathbf{h})$ (Deutsch and Journel, 1992:39-40). Semi-variogram values are calculated over the entire site defining the average range and degree of correlation between values at varying directions and magnitudes of **h**. This correlation structure would then be modeled for each direction to provide a mathematical definition of the ranges and degrees of correlation in that direction. Normally the correlation between pairs of values will decrease as distance increases until the site variance, or sill, is reached and independence is assumed. This work will not venture beyond the description of spatial continuity for the exhaustive data set which will be used in both OK and SIS. In practice, transmissivity variograms for a site may be borrowed from a larger data set taken at the same scale from a similar depositional environment or they may be synthesized from geological drawings or interpretations (Journel and Alabert, 1990:213). If these resources are inadequate or unavailable, soil sampling strategies can be implemented which will provide the opportunity to define the spatial continuity of a site, one of which is presented in a work by Flatman and Yfantis (1984).

In order to implement any geostatistical estimation technique using the Soda Springs data set, the assumption of stationarity is requisite and, therefore, must be addressed. The assumption of stationarity is for the model's sake and has little to do with the reality of the situation (Isaaks and Srivastava, 1989:343). In fact, it is difficult to imagine any heterogeneous site that is truly stationary. As a result, the assumption of stationarity must be made cautiously and with specific goals in mind such as the reproduction of major heterogeneities.

The question of stationarity is motivated by the question of the relevance of the nearby samples (Isaaks and Srivastava, 1989:349). Nearby samples are relevant if they contain appropriate information based upon the average spatial continuity of the site. If there are major discontinuities in spatial correlation on the site, as would be the case if extreme values were connected along differing directions, then the estimation may be compromised by the existence of real non-stationarity. It was for this reason that the scope of this study was reduced from an initial analysis of the entire Soda Springs Valley to a portion dominated by only one direction of spatial connectivity.

Another concern when making the assumption of stationarity is the presence of a trend across a site. In this case, it is helpful to understand that the assumption of stationarity does not justify large search windows (Isaaks and Srivastava, 1989:343). Especially on a heterogeneous site, the suitability of conceptualizing the sample data within the search window as an outcome of a stationary random function model often becomes more questionable as the window gets larger (Isaaks and Srivastava, 1989:343). Journel and Rossi addressed this situation in a comparison of estimation techniques that allow for a trend and estimation techniques that do not. They concluded that in interpolation situations when search strategies are retained within local windows, the final estimate will be unaffected by trend specification and the Ordinary Kriging algorithm is recommended (Journel and Rossi, 1989). Local stationarity, or quasi-stationarity, is then seen as a viable assumption in place of global stationarity (Isaaks and Srivastava, 1989:532).

III.4. The Sample Data Sets

Three different sampling scenarios will be used to aid in the comparison of the estimation methodologies. The first sample data set will be comprised of 60 hard data values determined by the random selection of four values in each of 15, 20 x 20 blocks across the site. This ensures a sample data set with univariate statistics very close to those of the exhaustive data set. As a result, the ability of SIS to reproduce the univariate and spatial statistics of truth will be easily and clearly demonstrated. Since 60 data values may be unrealistic, however, the second sampling scenario will be comprised only of 15 hard data points determined by the random selection of one value in each of the 20 x 20 blocks. This limited number of samples will highlight a greater disparity between the results from OK and SIS. Finally, one of the most valuable attributes of SIS will be exploited as 209 soft data values on a regular 5 x 5 grid are added to the 15 hard values. The soft values will be postulated to have come from soil core analysis allowing the determination of whether a particular transmissivity value lies below the 10th percentile of truth, above the 90th percentile of truth, or somewhere in between. This difference amounts to about one order of magnitude or to the difference between gravel and fine sand (Domenico and Schwartz, 1990:65).

III.5. Ordinary Kriging

To demonstrate the limitations of a technique based upon moving averages, Ordinary Kriging will be performed on the sample data sets. Since OK is motivated by the desire to obtain an estimate with a minimum variance, the resulting estimated transmissivity fields should demonstrate a smoothing effect and, as a result, isolated points of extreme transmissivities. Since OK will be used in this capacity and since it is one of the tools utilized in SIS, a review of its construction is appropriate. (The text by Isaaks and Srivastava is an excellent source on this topic and, as such, it contributes extensively to the following condensed discussion (1989:278-290).) The concept of OK is simple -- at every point where there was no sample taken, the value will be estimated using a weighted linear combination of the n available samples. This can be written mathematically using hats for any estimated values:

$$\hat{t} = \sum_{i=1}^{n} w_i \cdot t_i \tag{4}$$

One of the goals of OK is to produce unbiased estimates. This requires the calculation of error residuals. The average error residuals are measured as the average of the difference between the estimated values and the true values at each respective location. Since truth is unknown, however, the concept of a random function must be employed. The estimation of a unknown value can then be modeled by a stationary random function consisting of random variables for each of the sample locations and one for the point to be estimated:

$$\hat{T}(x_0) = \sum_{i=1}^{n} w_i \cdot T(x_i) \tag{5}$$

Now the error residuals can be expressed as a random variable:

$$R(x_0) = \left[\sum_{i=1}^{n} w_i \cdot T(x_i) \right] - T(x_0) \tag{6}$$

To ensure that the bias is zero, or that the expected value of the error at any location is zero, the assumption of stationarity is employed allowing the equivalence of the expected values of transmissivity at different locations:

$$E\{R(x_0)\} = E\left\{ \sum_{i=1}^{n} w_i \cdot T(x_i) \right\} - E\{T(x_0)\} = E\{T\} \cdot \sum_{i=1}^{n} w_i - E\{T\} = 0 \tag{7}$$

To have no bias, then, the sum of the weights must equal one.

The other goal of OK is to produce an estimate with minimum variance of the error residuals. Since truth is unknown, the variance of the actual error cannot be minimized, but the variance of the modeled error, $\tilde{\sigma}_R^2$, can be. Assuming that the mean error is zero and using the rule for the variance of a linear combination of dependent random variables (Devore, 1982:213), the variance of the error can be written in terms of the random function:

$$V\{R(x_0)\} = Cov\{\hat{T}(x_0), \hat{T}(x_0)\} - 2Cov\{\hat{T}(x_0), T(x_0)\} + Cov\{T(x_0), T(x_0)\} \tag{8}$$

which can be rewritten as:

$$\tilde{\sigma}_R^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \tilde{C}_{i,j} - 2\sum_{i=1}^{n} w_i \tilde{C}_{i,0} + \tilde{\sigma}^2 \tag{9}$$

where $\tilde{\sigma}^2$ and $\tilde{C}_{i,j}$ are the variance and covariance of the random function model. What remains then is a series of n equations with n unknowns, and one additional equation resulting from the constraint that all weights sum to one to ensure unbiasedness. The problem of the additional equation is solved by introducing one more unknown in the form of a Lagrange parameter, μ. Adding the Lagrange parameter to equation (9) in such a way that it is like adding a zero results in the following additional term:

$$2\mu\left(\sum_{i=1}^{n} w_i - 1\right) \tag{10}$$

Now, setting the partial first derivatives of the equation for the error variance equal to zero will provide the set of weights required to minimize variance and ensure unbiasedness:

$$\begin{aligned} \frac{\partial(\tilde{\sigma}_R^2)}{\partial w_i} &= 2\sum w_j \tilde{C}_{i,j} - 2\tilde{C}_{i,0} + 2\mu = 0 \Rightarrow \sum w_j \tilde{C}_{i,j} + \mu = \tilde{C}_{i,0} \quad \forall\, i = 1,\ldots,n \\ \frac{\partial(\tilde{\sigma}_R^2)}{\partial \mu} &= 2\sum_{i=1}^{n} w_i - 2 = 0 \Rightarrow \sum_{i=1}^{n} w_i = 1 \end{aligned} \tag{11}$$

The resulting systems of equations are referred to as the ordinary kriging system:

$$\underbrace{\begin{bmatrix} \tilde{C}_{1,1} & \cdots & \tilde{C}_{1,n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{C}_{n,1} & \cdots & \tilde{C}_{n,n} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}}_{\mathbf{C}} \cdot \underbrace{\begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} \tilde{C}_{1,0} \\ \vdots \\ \tilde{C}_{n,0} \\ 1 \end{bmatrix}}_{\mathbf{D}} \tag{12}$$

Solving for the weights then is simply a matter of multiplying both sides equation (12) by $\mathbf{C}^{-1}$. To incorporate the use of the variogram, $\gamma_{i,j}$, the assumptions that the random variables in the random function all have the same mean and variance are employed to establish the following relationship:

$$\gamma_{i,j} = \tilde{\sigma}^2 - \tilde{C}_{i,j} \tag{13}$$

Thus, the ordinary kriging system can be written as:

$$\sum w_j \tilde{\gamma}_{i,j} - \mu = \tilde{\gamma}_{i,0} \quad \forall\ i = 1,\ldots,n \tag{14}$$

plus the unbiasedness condition. Therefore, by using the variograms for a site, every point in a field can be estimated by taking a linear combination of the products of the surrounding points within the search radius and their respective weights as determined using the ordinary kriging system thus ensuring unbiasedness and minimum variance.

III.6. Sequential Indicator Simulation

After the OK results have been obtained, the sample data set will be used as input into the SIS algorithm, from which any number of equiprobable realizations honoring the data at their locations and honoring the spatial covariance structures at a selected number of cutoffs will be obtained. Unlike OK, SIS is not an averaging technique. It uses the averaging of OK only as a means by which to develop the probability of exceeding each of the cutoffs at every point. This is accomplished through the use of indicator coding as was introduced in Chapter II.

III.6.1. Indicator Formalism

The transformation of transmissivity values into zeros and ones creates a Bernoulli random variable, $I(\mathbf{u};\, t_c)$, which has an expected value equal to the probability of exceeding a given cutoff:

$$E\{I(\mathbf{u};t_c)\} = 1 \cdot P\{T(\mathbf{u}) \leq t_c\} + 0 \cdot P\{T(\mathbf{u}) > t_c\} \tag{15}$$

Thus OK can be used to arrive at an estimate of the mean at some location, **u** within any area $A \subset D$, based upon a weighted average of available sample values:

$$E\{I(\mathbf{u};t_c) | t(\mathbf{u}_\alpha)\} = \hat{F}(A;t_c) = \sum_{\alpha=1}^{N} w_\alpha(\mathbf{u},t_c) \cdot i(\mathbf{u}_\alpha,t_c) \tag{16}$$

and so provide the probability that the unknown value will not exceed the given cutoff. To arrive at the cdf for each location then, an indicator variogram function is inferred for each cutoff and an estimate is made for the probability of nonexceedence at each. Consequently, IK provides a cdf conditioned by available data and by the spatial correlation of indicator data that is defined only at chosen cutoff values.

III.6.2. SIS Algorithm

SIS uses IK as a means by which to estimate the distribution of a random variable and to account for the spatial variation of different classes of values as discretized by the cutoffs. At each unsampled node, SIS will obtain a cdf via IK from which it will generate a simulated value using the classical Monte Carlo technique. The simulated node will then be included in the data set of known values and another node will be simulated. The entire algorithm is charted in Figure 6.

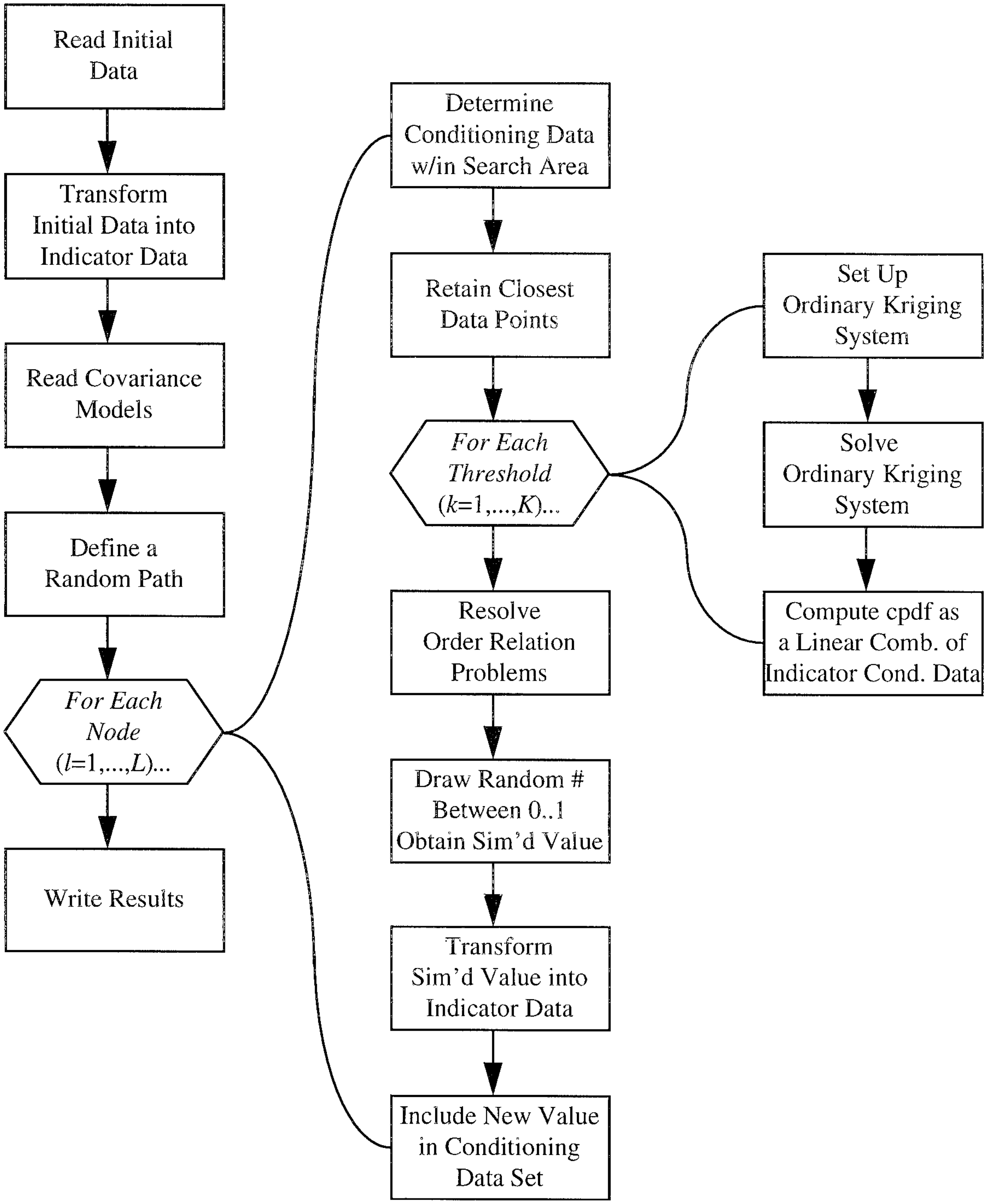


Figure 6 -- The SIS Algorithm

III.6.3. Interpretation and Selection of Realizations

Once a number of realizations have been obtained using the SIS algorithm, the question arises of how to handle them. There are several approaches that can be taken depending upon the goals of the study and upon prior site information. One common approach, and certainly the least advisable has been to use SIS as an improved interpolation algorithm, taking the first realization that is generated and drawing conclusions (Deutsch and Journel, 1992:130). This approach is doomed to failure because it does not appreciate the value of SIS which lies in its ability to reproduce spatial statistics as an *average* over a large number of simulations. If, however, there is some prior knowledge about the site which could define some sort of criteria or ideal from which to evaluate realizations, then a conceivable approach could be to select realizations based upon that criteria, hence conditioning the results by previously unused information (Deutsch and Journel, 1992:130). The selected realizations could then be put into a flow simulator providing results that honor that knowledge. If there are measured head values, then it may be possible to condition the selection of realizations based upon the retention of those that, when put through a flow simulator, most closely model the measured heads.

The most advisable use of SIS lies in its ability to provide a measure of uncertainty. Since each realization is an equiprobable image of reality, those spatial features which are observed on all simulated images may be considered reliable, while those features which are only seen on some images may be judged unreliable (Journel and Alabert, 1989:125). Thus, given a sufficient number of realizations, an accurate picture of spatial features which are probable given the spatial information input into SIS can be obtained. In this work, a methodology employing a combination of the above strategies will be employed. First, a large number of simulations will be run using a particular sample data set from which a quantification of the uncertainty concerning the spatial locations of the extreme transmissivity values will be calculated. Next, those simulations best capturing the characteristics that are deemed most probable will be run through a flow simulator to generate a field of heads. The field most accurately matching the true head field will be retained as a likely configuration of the transmissivity values.

III.7. Flow Simulation

The transfer function used for the generation of heads and specific discharges is the Modular Three-Dimensional Finite-Difference Ground-Water Flow Model, or MODFLOW, as developed by the U. S. Geological Survey. This function employs the standard block-centered flow equation:

$$\frac{\partial}{\partial x}\left(T\frac{\partial \phi}{\partial x}\right)+\frac{\partial}{\partial y}\left(T\frac{\partial \phi}{\partial y}\right)=0 \tag{17}$$

where ϕ ($\mathbf{u}$) is the potentiometric head. Boundary conditions of no-flow to the north and south and constant head to the east and west are established, as well as a constant gradient of 0.0167 across the site from east to west. The equation is then solved iteratively using Slice-Successive Overrelaxation. The output of this procedure includes maps of heads and maps of transverse and longitudinal fluxes. The maps of heads will be utilized by subtracting the gradient to obtain head fluctuation fields and maps of specific discharge will be constructed using the flux data.

III.8. Estimates vs. Truth

Comparisons of the estimated transmissivity fields relative to known truth will be made in several different fashions. First, the univariate statistics will be placed side-by-side including the summary statistics, the histograms, and the cdfs. Next spatial statistics will be compared using the indicator correlogram or the centered standardized covariance which is a measure of linear correlation between two indicators, $I(\mathbf{u},t_c)$ and $I(\mathbf{u}+\mathbf{h},t_c)$ (Journel and Alabert, 1989:126). This function, written

$$\rho_I(\mathbf{h},t_c) = \frac{F(\mathbf{h};t_c) - F^2(t_c)}{F(t_c)[1 - F(t_c)]} \tag{18}$$

where $F(\mathbf{h};\ t_c)$ is the non-centered covariance of the indicator process $I(\mathbf{u};\ t_c)$ and is written:

$$F(\mathbf{h};t_c) = E\{I(\mathbf{u};t_c)\cdot I(\mathbf{u}+\mathbf{h};t_c)\} = P\{T(\mathbf{u}) \le t_c,\ T(\mathbf{u}+\mathbf{h}) \le t_c\} \tag{19}$$

(Journel and Alabert, 1989:126). Finally, the products of the transfer function will be evaluated next to truth to obtain an appreciation for the effects of different estimation techniques upon the output to be used by decision makers.

IV. Results

IV.1. Spatial Description of Exhaustive Transmissivity Field

Analysis of the spatial continuity of the exhaustive transmissivity field was accomplished to provide covariance models for both OK and SIS. A single model using hard transmissivity values was needed for OK while SIS requires models derived from the correlation among indicator data at each selected cutoff. The algorithm requires, in both cases, models for the direction with the largest range of correlation, the major axis, and for the axis orthogonal to it, the minor axis. From this information the algorithm will infer an oval of correlation around a point to be estimated.

To realize the spatial correlation of the Soda Springs data set, the Geostatistical Library, or GSLIB, by Deutsch and Journel was used (1992). First variograms were calculated every five degrees around the tail of **h**, giving a total of 72 directions. These variograms were then plotted as variogram maps, as shown by the example in Figure 7, to determine the major and minor directions of anisotropy. On this map, the white area is equal to the site variance and therefore defines sill. The major axis corresponds with approximately 0 degrees east.

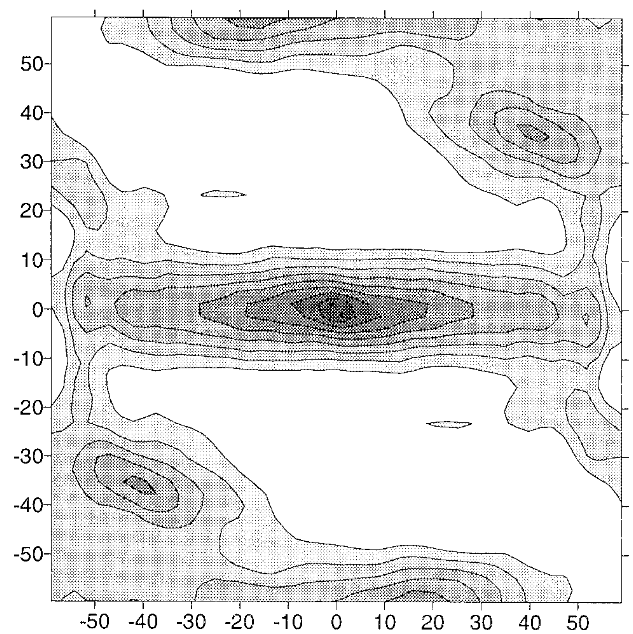


Figure 7 -- Variogram Map for Exhaustive Transmissivity Field

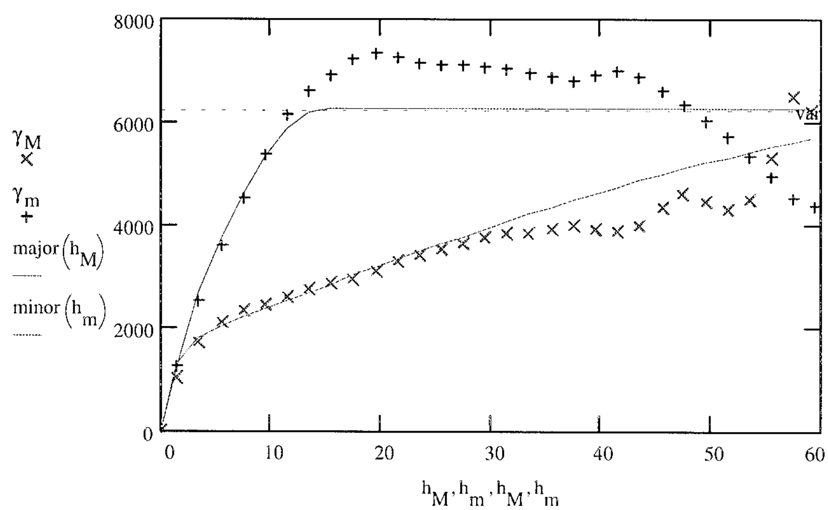


Figure 8 -- Variograms for Principle Axes

Once the principle axes of anisotropy were determined, their respective variograms were plotted and mathematical models were fitted to each for input into the OK and SIS algorithms of GSLIB. Examples of this are shown in Figure 8, where the major axis of 0 degrees east and a minor axis of 0 degrees north from the variogram map above are modeled. The variogram map and models shown are used for Ordinary Kriging; maps and models were completed on the indicator data at each cutoff for use in SIS.

IV.2. Sampling Scenario 1 (60 Hard Data)

The first sampling scenario includes a series of 60 hard data values determined by the random selection of four values in each of 15, 20 x 20 blocks across the exhaustive site. Figure 9 is a plot of the sample values across the domain of interest; the size of the circle is proportional to the relative magnitudes of the sampled values. Given the large number of transmissivity measurements, this scenario may be somewhat unrealistic for Air Force purposes but it will prove informative for experimental purposes.

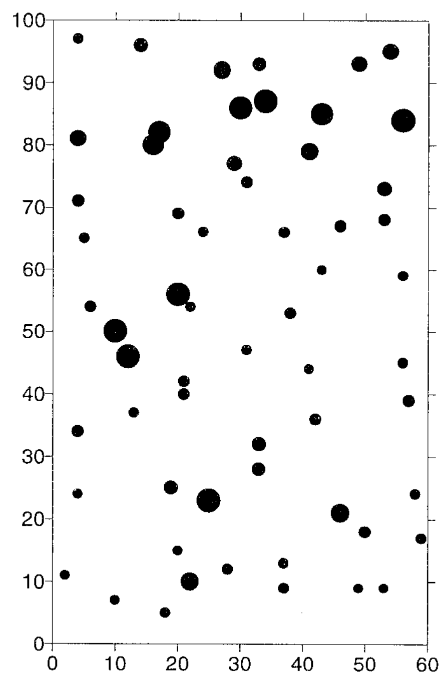


Figure 9 -- Sampling Scenario 1 (60 Hard Data)

IV.2.1. Summary Statistics of Sample Data Set

As can be seen in Table 2 and in Figure 10, the large number of data values provides sufficient information to adequately reproduce the univariate statistics of the exhaustive data set. In the SIS algorithm, the density of the values and the corresponding structure of the histogram, will allow for the use of its cdf in defining the output cdfs.

Table 2 -- Summary Statistics for 60 Hard Data in Sampling Scenario 1

Mean	100.69
Standard Deviation	86.58
Variance	7495.58
Coefficient of Variation	0.86
Minimum	16.87
10th Percentile	22.99
25th Percentile	34.23
Median	63.26
75th Percentile	144.13
90th Percentile	277.73
Maximum	278.77
Coefficient of Skew	1.08

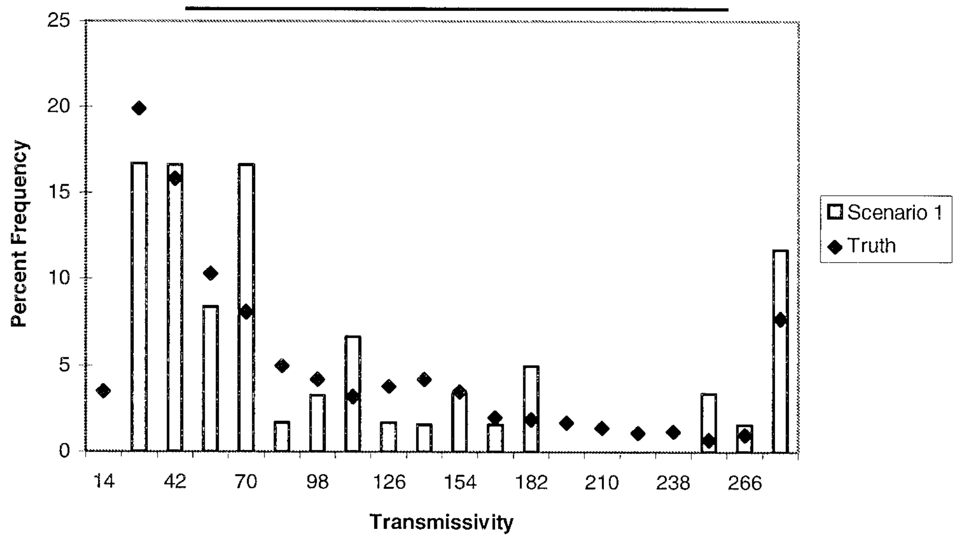


Figure 10 -- Histogram of 60 Hard Data in Sampling Scenario 1

IV.2.2. Summary Statistics of Ordinary Kriging Output

Since the Soda Springs data set was derived from elevation data, it is relatively smooth, i.e. there are no major areas with discontinuities. This characteristic of the data set favors the smoothing that is usually a result of OK. However, since the algorithm is simply a series of reestimated means at each unknown location, it will tend to reproduce more values in the middle of the distribution and less at the extremes. Figure 11, a gray-scale map of the OK output, illustrates this by the evident dominance of the gray values and the limited areas of white and black.

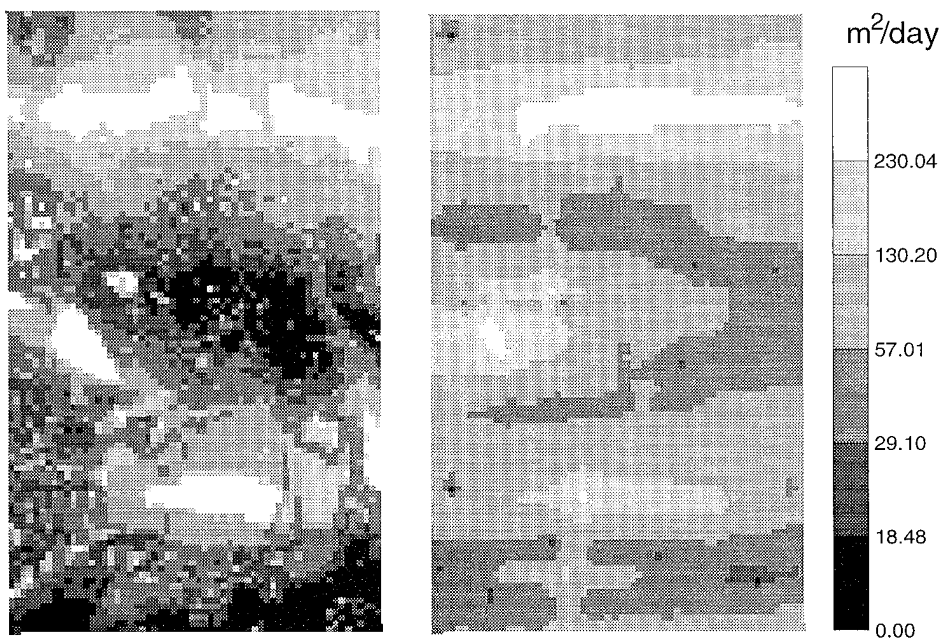


Figure 11 -- Truth vs. OK Output (60 Hard Data)

As will be the standard throughout this work, and unless otherwise noted, the shades of gray will correspond to the cutoffs defined by the 10th, 25th, 50th, 75th, and 90th percentiles of the exhaustive data set. As a result, the above map of estimated transmissivities should be comparable to the true map shown in Chapter III. The OK map appears to have performed well, especially in the reproduction of the higher extremes along the top of the field. The summary statistics and histogram shown in Table 3 and Figure 12 respectively, however, highlight the expected deficiencies of OK. An examination of these statistics reveals a one-half reduction in the variance and a subsequent reduction of the extreme values. The histogram shows that the entire distribution was moved toward center with the median value being shifted to the right. Despite these expected changes however, OK still does relatively well in reproducing the values above the 75th percentile. The reason for this is more apparent in Figure 13, a plot of the exhaustive cdf vs. the OK output cdf.

Table 3 -- Summary Statistics for Ordinary Kriging Output (60 Hard Data)

Mean	99.11
Standard Deviation	56.44
Variance	3185.81
Coefficient of Variation	0.57
Minimum	16.87
10th Percentile	46.89
25th Percentile	56.34
Median	80.48
75th Percentile	125.65
90th Percentile	192.87
Maximum	278.77
Coefficient of Skew	1.17

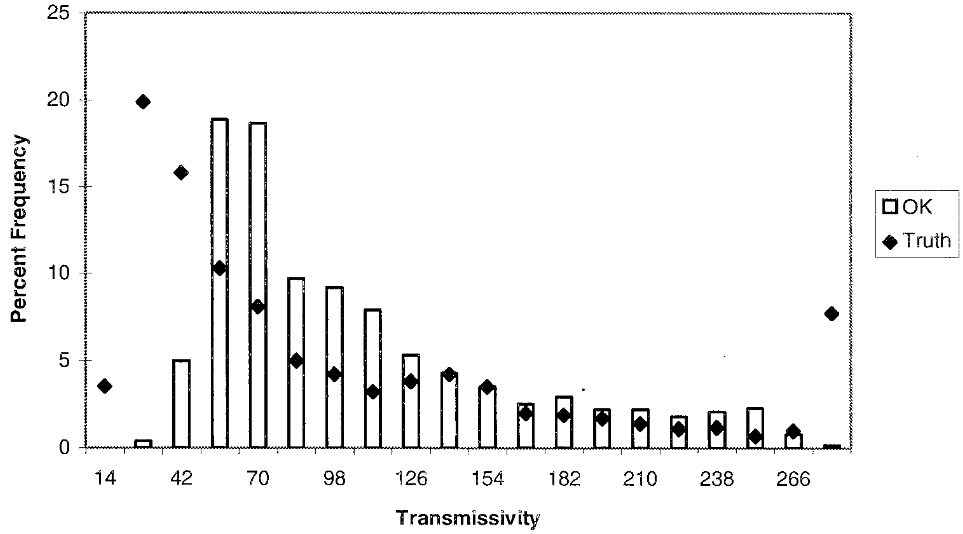


Figure 12 -- Histogram of OK Output (60 Hard Data)

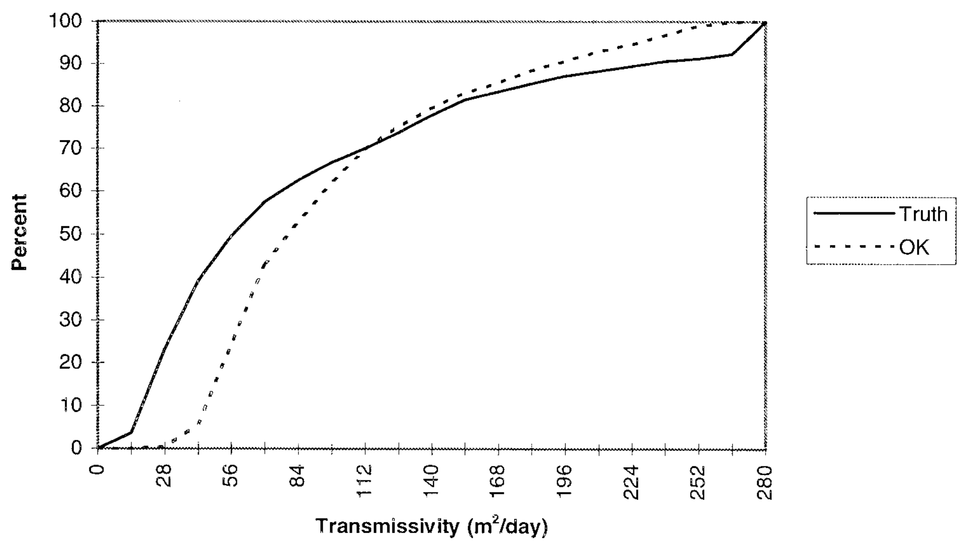


Figure 13 -- Comparison of cdfs for Truth and OK Output (60 Hard Data)

It is apparent in Figure 13 that there is a reduction of the extreme values in the OK output, however there is still a single point, occurring at approximately the 70th percentile, at which the true cdf and the OK cdf are equal. Outward from this point in either direction, OK gradually departs from truth. Consequently, those values close to the 70th percentile, especially those in the increasing direction, are reproduced rather well.

IV.2.3. Average Summary Statistics of 50 SIS Simulations

50 SIS simulations were generated for this sampling scenario. To show the summary statistics of any one of the simulations would be misleading however, since simulation theory only guarantees a match as an average over a large number of realizations (Deutsch and Journel, 1992:130). Consequently, the most efficient demonstration of SIS's effectiveness in this respect is a cdf plot of the exhaustive cdf vs. the average SIS cdf over 50 realizations as is shown in Figure 14. This plot clearly demonstrates the ability of SIS to reproduce the true cdf as an average over a large number of realizations given a sufficient number of hard data points. This is a significant ability of SIS, however, of more significance is the spatial arrangement of the data.

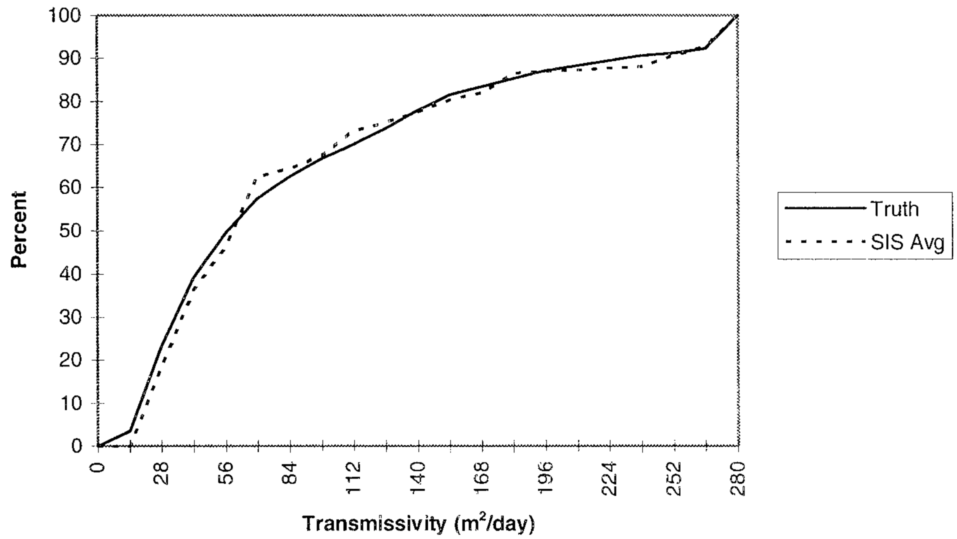


Figure 14 -- Comparison of cdfs for Truth and SIS Output (60 Hard Data)

IV.2.4. Evaluation of Uncertainty

Providing a picture of the spatial arrangement of the 50 SIS realizations first requires the consideration of uncertainty. As was stated earlier, each realization is an equiprobable image of truth, honoring both the location of known data points, as well as the given spatial covariance structure. As a result, the output from the realizations can be evaluated most effectively by noting the spatial arrangements that occur most often. This, however, is not an objective process. Consequently, conditioning the definition of what is certain or uncertain upon other information is helpful.

To illustrate this, a cutoff above or below which the spatial arrangement of values is considered significant must first be chosen. In this work, this is the cutoff corresponding to the 90th percentile of truth, or 230.04 m^2/day. Next, all of the simulations are re-coded so that all values falling above the cutoff are equal to one and all values below the cutoff are equal to zero. Now if each location is averaged across all realizations, some number between zero and one will result, indicating the percentage of the time that a particular location exceeded the cutoff. Consequently, it is necessary to determine a "significant" percentage above which it is probable that a value at a location exceeds the cutoff in reality. Because of the large range of variations possible, it may be incorrect to assume that, because a particular location exceeded the cutoff less than half of the time, it is probably less than the cutoff. In fact, the percentage of values that exceeded the cutoff more often than not is only 2.65. The plot in figure 14 indicates, however, that the average percentage of values above the cutoff is very close to that of truth, or about 10%. Therefore it may be more appropriate to determine the "significant" level of uncertainty based upon the known percentage of values above a given cutoff. This can be determined using sound geological knowledge of the site, using the statistics of the sample data, or using the average of all SIS realizations. In this case, if 10% is chosen as the desired percentage of values above the cutoff, those locations exceeding the cutoff more than three times out of ten, or at the 30% probability level, would be significant. (In other words, 10% of the values exceeded the cutoff at least 30% of the time.) This map, called a probability map, is shown as Figure 15. The white areas would be those with transmissivity values supposed to exceed 230.04 m^2/day.

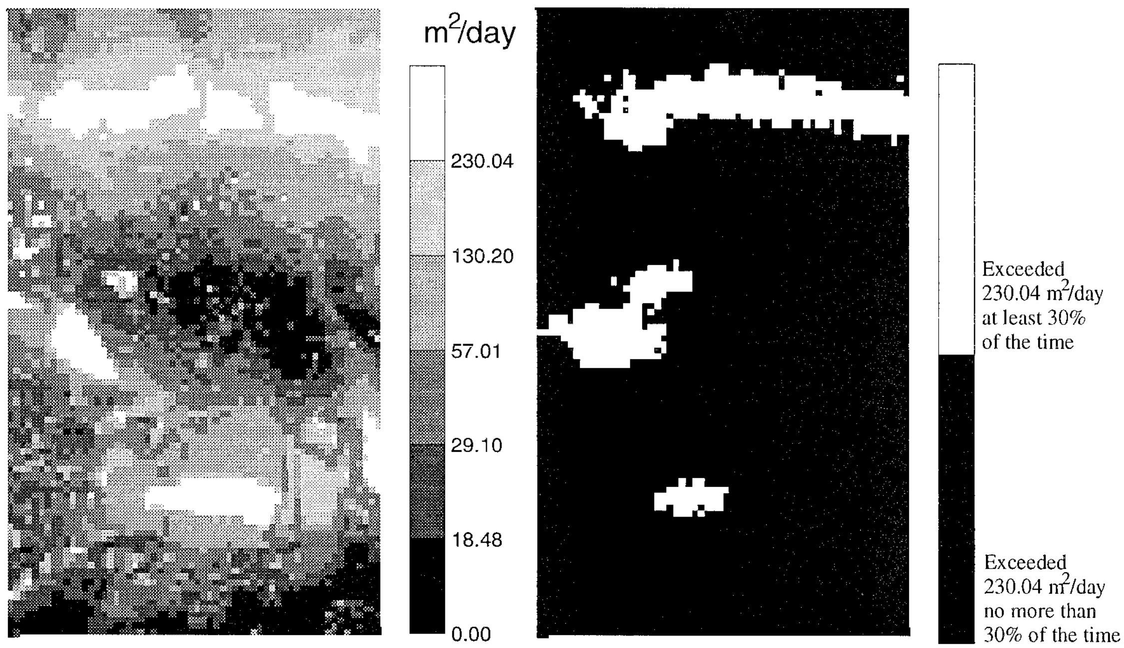


Figure 15 -- Truth vs. SIS Probability Map (60 Hard Data)

It is apparent from this map that the major features of the high-valued data in the exhaustive data set were captured by SIS, but not necessarily to any better extent than was accomplished using OK. A necessary check on the validity of the SIS results, as presented, is to confirm that the covariance structure for the given cutoff is honored. This check is performed on the results in sampling scenario three.

IV.3. Sampling Scenario 2 (15 Hard Data)

Sampling Scenario 2, comprised of only 15 hard data points, is perhaps more realistic given the cost of transmissivity measurements. The locations of the samples were determined by making a random selection of one location in each of the 15, 20 x 20 blocks. The locations and relative magnitudes of the samples are shown in Figure 16.

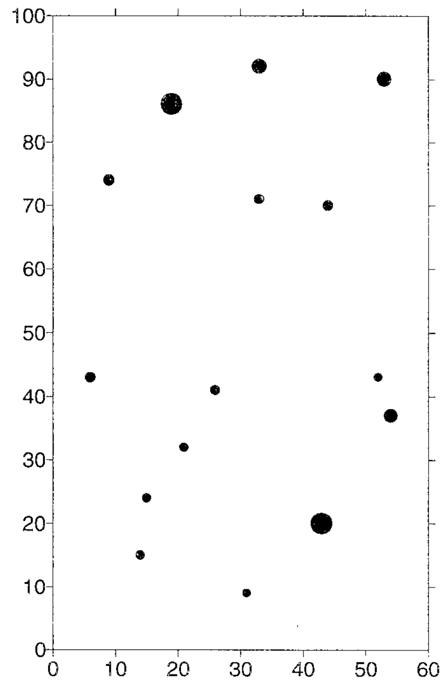


Figure 16 -- Sampling Scenario 2 (15 Hard Data)

IV.3.1. Summary Statistics of Sample Data Set

Because of the sampling strategy, the summary statistics in Table 4 are close to those of truth. The difficulty that is presented with such a small number of data points, however, is more evident in the histogram shown as Figure 17. Since there is a small number of data points, the distribution is not well defined. As a result, SIS will be performed using liner interpolation between the cutoffs and power models at the extremes (Deutsch and Journel, 1992:131-134).

Table 4 -- Summary Statistics for 15 Hard Data in Sampling Scenario 2

Mean	91.41
Standard Deviation	85.71
Variance	7346.20
Coefficient of Variation	0.94
Minimum	20.26
10th Percentile	22.57
25th Percentile	30.45
Median	50.45
75th Percentile	136.56
90th Percentile	278.77
Maximum	278.77
Coefficient of Skew	1.39

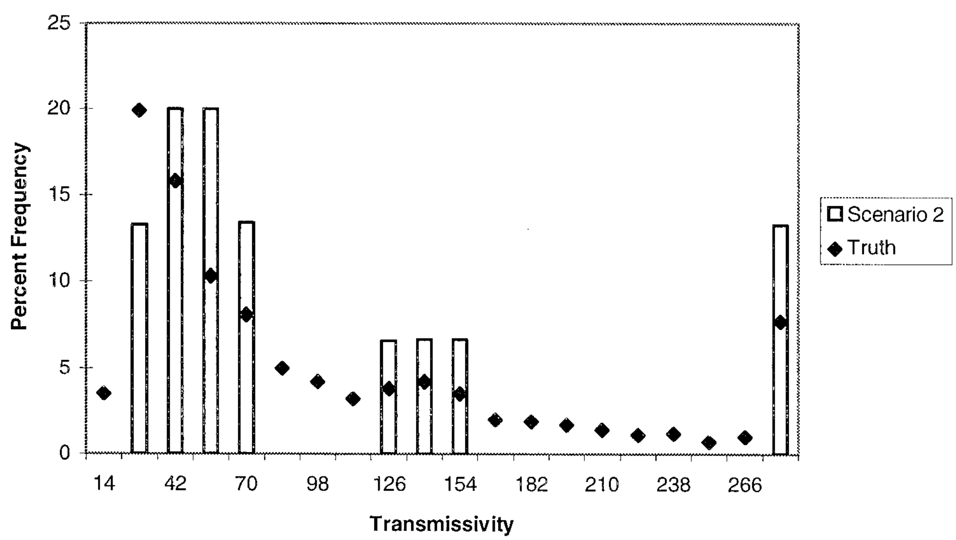


Figure 17 -- Histogram of 15 Hard Data in Sampling Scenario 2

IV.3.2. Summary Statistics of Ordinary Kriging Output

With so few points, the smoothing aspect of OK is much more apparent. The map of the output, Figure 18, illustrates this in that the extremes have been reduced to small isolated spots. This phenomenon is illustrated further by the summary statistics of the output shown in Table 5 and in Figure 19.

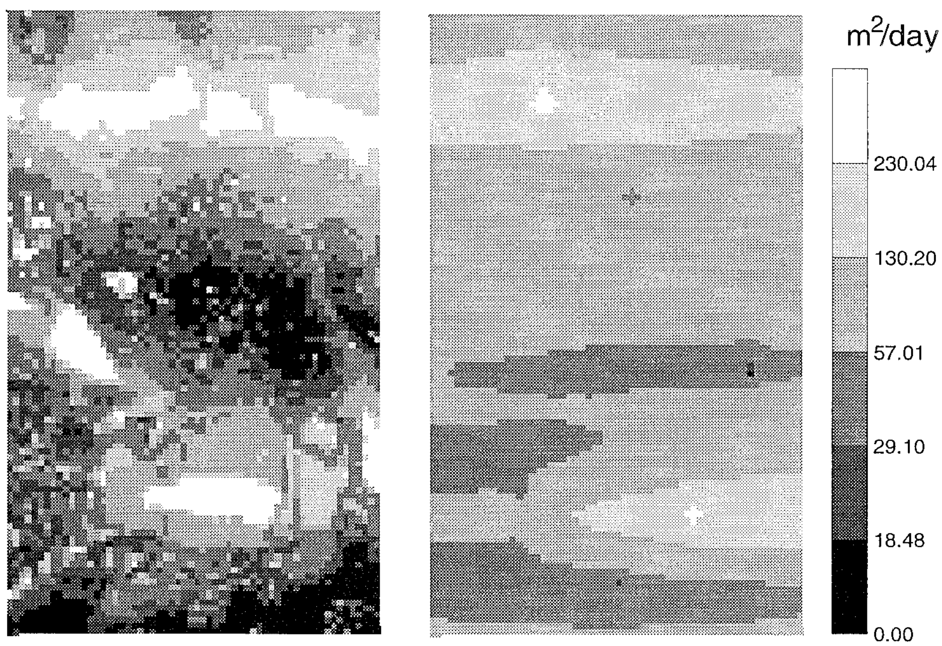


Figure 18 -- Truth vs. OK Output (15 Hard Data)

Table 5 -- Summary Statistics for OK Output (15 Hard Data)

Mean	93.51
Standard Deviation	42.44
Variance	1801.24
Coefficient of Variation	0.45
Minimum	20.26
10th Percentile	49.20
25th Percentile	61.06
Median	83.96
75th Percentile	114.15
90th Percentile	159.86
Maximum	278.77
Coefficient of Skew	1.09

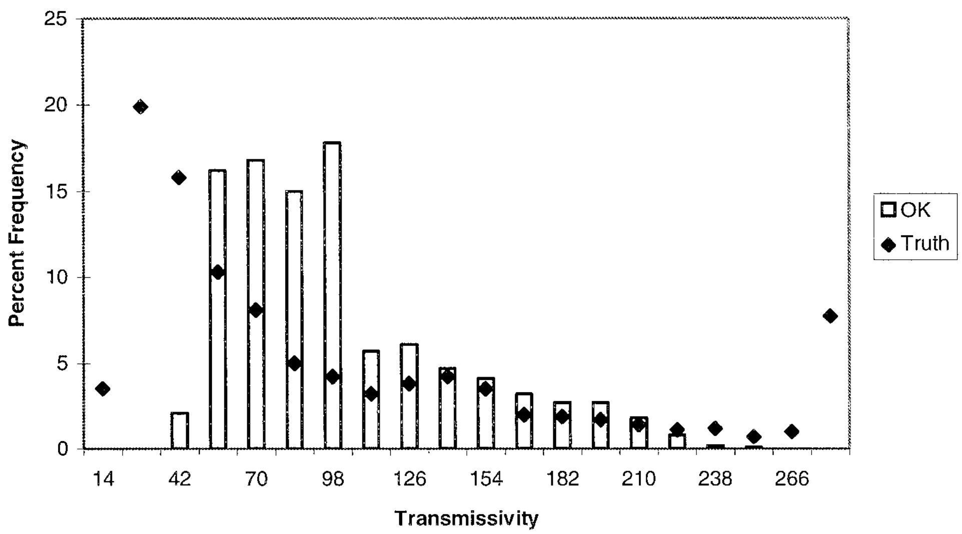


Figure 19 -- Histogram of OK Output (15 Hard Data)

From these statistics, it is apparent that a greater degree of smoothing has occurred because of the small number of data. The variance is less than one-third the exhaustive site variance and the 10th and 90th percentiles have moved much closer to the median. Furthermore, in a comparison of cdfs, Figure 20, the point at which the cdf for the OK output crosses the true cdf has moved further left and on either side of that point, the separation between the two curves increases quickly.

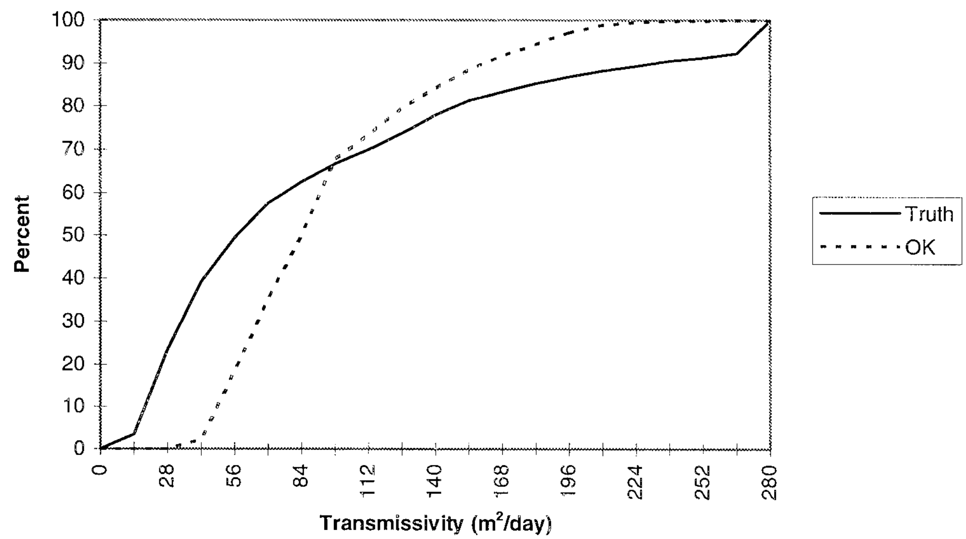


Figure 20 -- Comparison of cdfs for Truth and OK Output (15 Hard Data)

IV.3.3. Average Summary Statistics of 50 SIS Simulations

Again, it will be necessary to show the results from SIS in the comparison of cdfs context. Figure 21 shows the average SIS cdf, determined from the cdfs of 50 realizations, and the cdf of the exhaustive transmissivity field. This plot shows that despite the much lower number of samples, the summary statistics were reproduced quite well. The most notable deficiency is the overestimation of the percentage of higher values. Also, it is interesting to note the effect of the linear interpolation between the cutoffs.

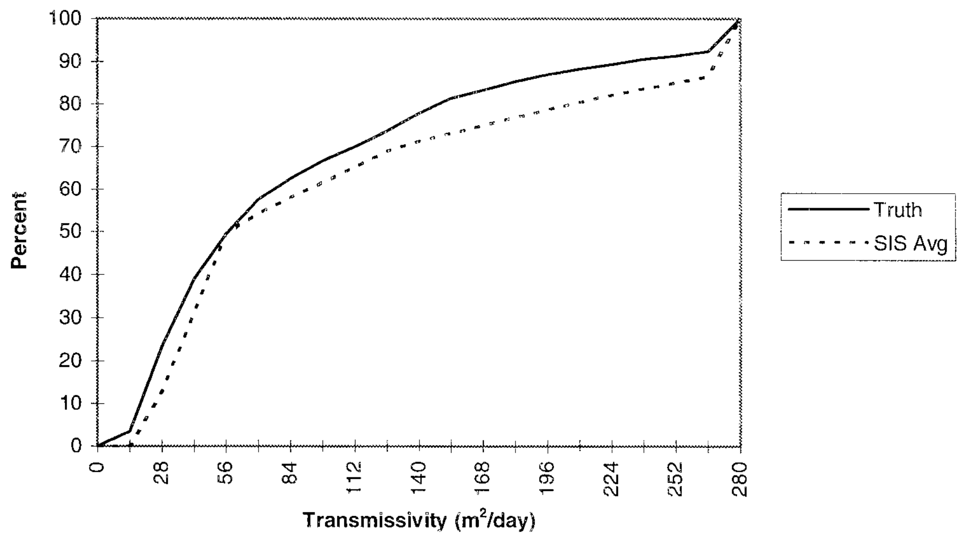


Figure 21 -- Comparison of cdfs for Truth and SIS Output (15 Hard Data)

IV.3.4. Evaluation of Uncertainty

The sample data and the average cdf indicate that there is greater than 10% of the field that exceeds the cutoff value corresponding to the 90th percentile of truth. Since truth is known, however, the probability map shown as Figure 22 is constructed so that the total number of values exceeding the cutoff be close to 10%. As in sampling scenario 1, 10% of the values included only those exceeding the cutoff at least three times out of ten. The map indicates that, in this case, SIS was able to reproduce a strip of connected points across the top similar to that existing in the exhaustive data set. Certainly, as an average over all realizations, SIS has out-performed OK in this respect.

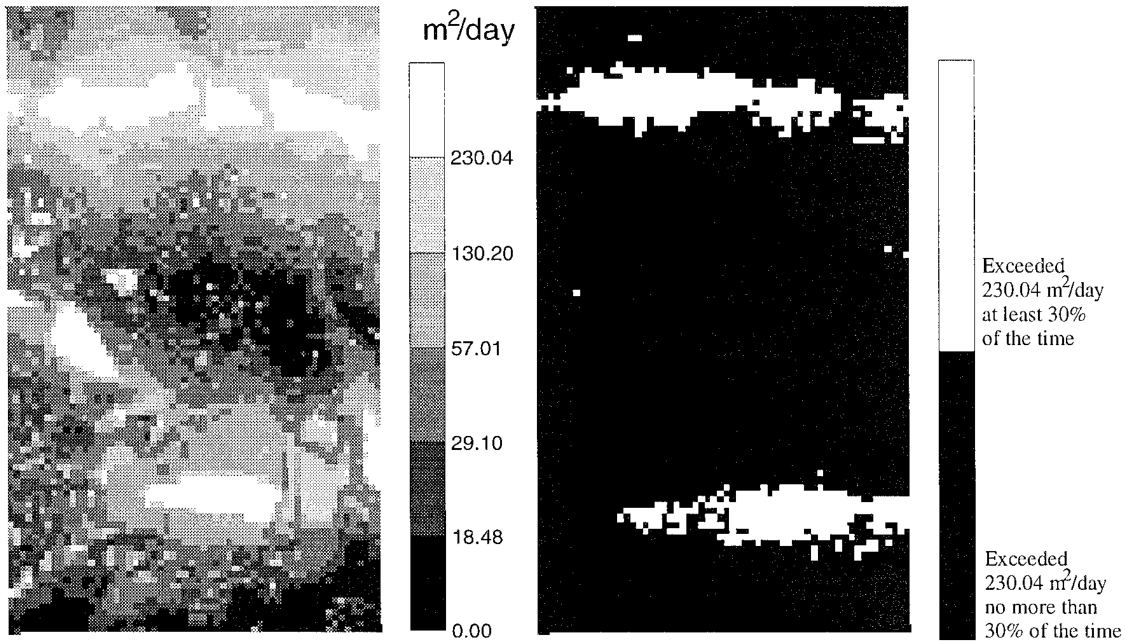


Figure 22 -- Truth vs. SIS Probability Map (15 Hard Data)

IV.4. Sampling Scenario 3 (15 Hard Data & 209 Soft Data)

Sampling Scenario 3 is comprised of the same 15 hard data used in scenario 2, plus the addition of 209 soft data shown in Figure 23. While there is a large number of soft data, they are usually less precise and are therefore much cheaper to obtain. Consequently, depending upon the goals of the study, the opportunity exists to gain a large amount of information at a relatively small cost. The soft data used in this case are supposed to have resulted from simple soil analysis on soil cores taken on a regular 5 x 5 grid across the site. The effect of the soil analysis was to place constraint intervals on the transmissivity at each of the sampled locations. The soil was known to have a transmissivity falling in the top ten percent of the exhaustive transmissivity values (the upright triangles), in the bottom ten percent (the downward facing triangles), or somewhere in between (the circles). This information, while easy to maintain and very informative, is not readily incorporated into the OK algorithm and, as a result, will not affect the OK results from sampling scenario 2. It is however, well suited for incorporation into SIS using the following indicator coding:

$$T(\mathbf{u}_\alpha) \in [a,b] = i(\mathbf{u};t_c) = \begin{cases} 0, \text{ if } t_c < a \\ 1, \text{ if } t_c \geq b \\ \text{undefined otherwise} \end{cases} \tag{20}$$

(Deutsch and Journel, 1992:84; Journel and Alabert, 1989:133).

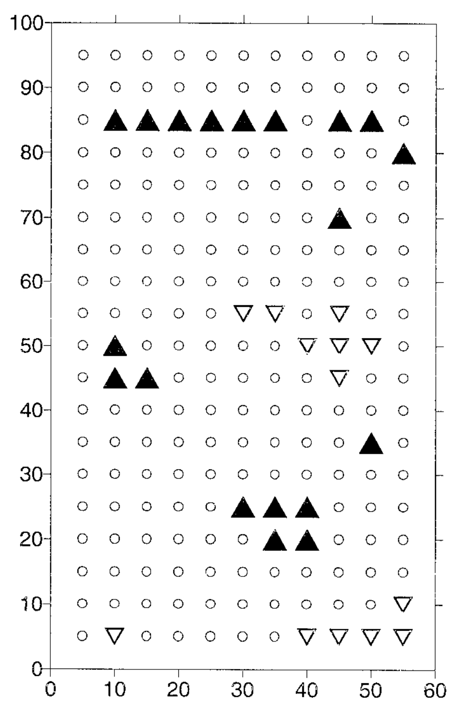


Figure 23 -- 209 Soft Data for Sampling Scenario 3

IV.4.1. Average Summary Statistics of 50 SIS Simulations

The cdf for the SIS output in Figure 24 is interesting in that it is similar to that in scenario 2 except for the tails. The addition of the soft information had the effect of further defining the tails. Thus the cdf for the SIS realizations is virtually equal to the true cdf at both the 10th and 90th percentiles. The uncertainty of the data falling between these two extremes seems to have created a more uniform distribution in the center.

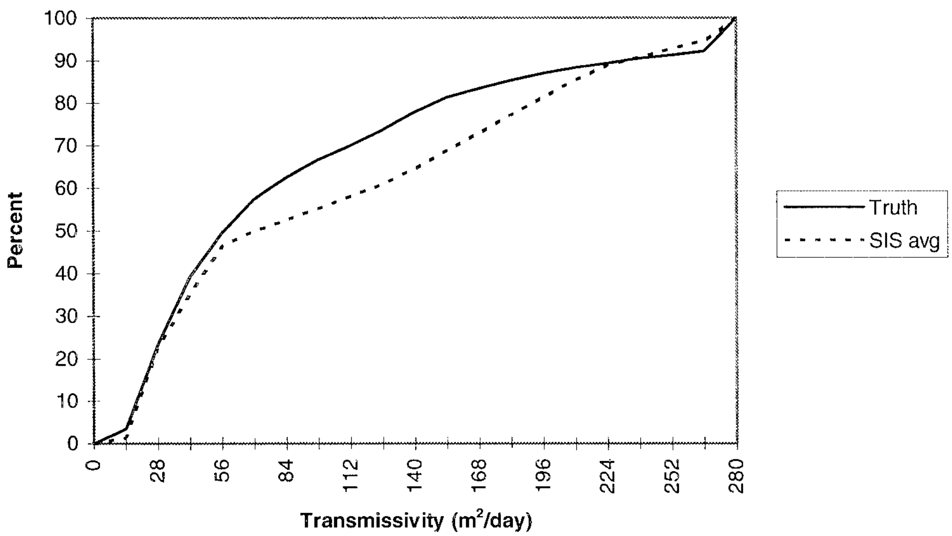


Figure 24 -- Comparison of cdfs for Truth and SIS Output (15 Hard Data and 209 Soft Data)

IV.4.2. Evaluation of Uncertainty

Both the sample data and the cdf indicate that the percentage of values exceeding the 90th percentile of truth is approximately nine. This figure is close to truth, in part, because of the large number of data. For purposes of constructing a probability map with ten percent of the values exceeding the cutoff, it was determined that those exceeding the cutoff at least three times out of ten should be assumed to exceed in reality. The probability map in Figure 25 indicates the effectiveness of SIS in reproducing the major features evident in the Soda Springs data set.

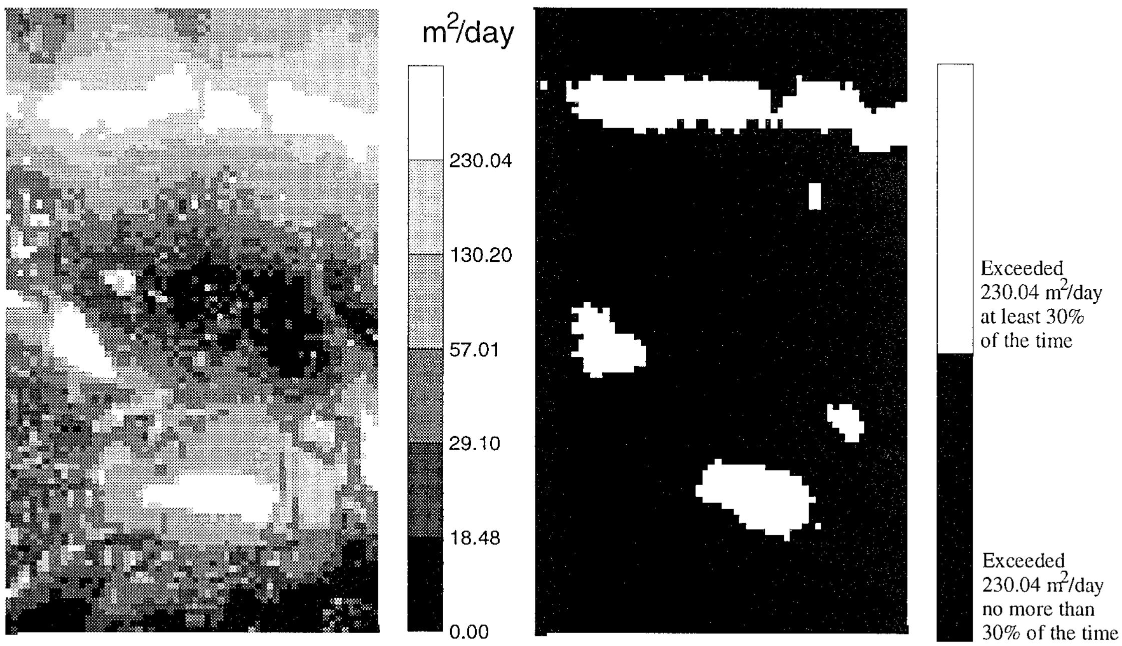


Figure 25 -- Truth vs. SIS Probability Map (15 Hard Data and 209 Soft Data)

IV.4.3. Selection of a Realization

Up to this point, no single realization has been singled out. This has been done to avoid the pitfall of thinking that any one realization can be selected and used. Depending upon the goals of the study, this may or may not be appropriate. Generally, SIS results are best used as an ensemble. It has been suggested that the ensemble results could be run through a transfer function from which statistics could be created that describe the response function (Journel and Alabert, 1990:212). This approach would certainly be valid, however, it would also be computationally intense.

As an alternative to this approach, it may be possible to condition the selection of a single realization based upon the knowledge gained from the probability maps. According to this approach, the realizations best matching the "critical" features as determined before-hand and as modeled on the probability maps could be selected for processing through a flow simulator. The selection could be further narrowed by conditioning the selection of a single realization based upon a comparison of the actual heads measured at the site and the calculated heads from the flow processor. A specific discharge field could then be generated from which the decision maker could gain information concerning the hydrogeologic character of the site. This process was performed upon the realizations obtained using sampling scenario 3. The SIS realization shown in Figure 26 was the realization selected. Summary statistics for this realization are shown in Table 6 and Figure 27.

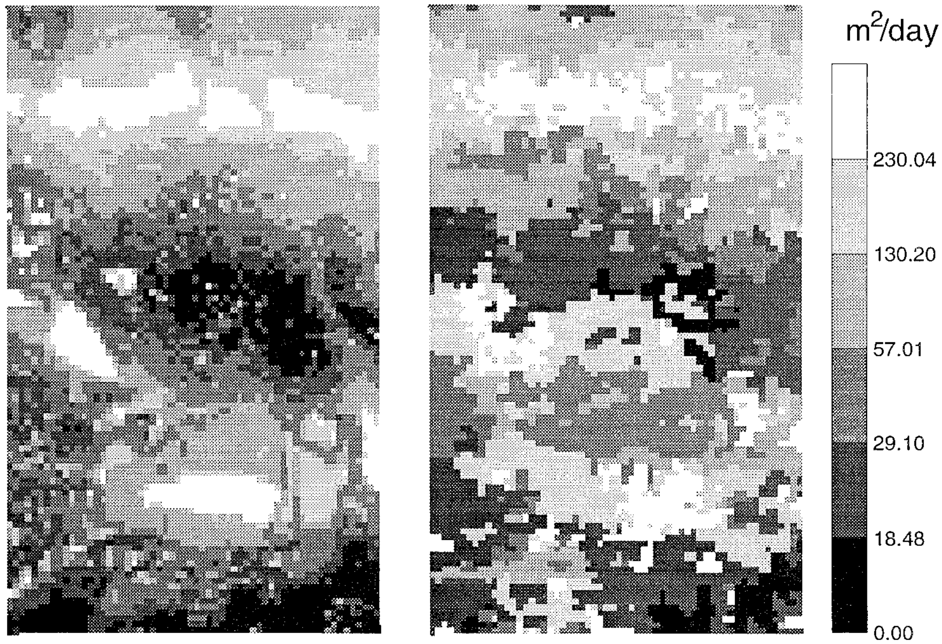


Figure 26 -- Truth vs. a Single SIS Realization (15 Hard Data and 209 Soft Data)

Table 6 -- Summary Statistics for a Single SIS Realization

Mean	109.19
Standard Deviation	84.92
Variance	7211.41
Coefficient of Variation	0.78
Minimum	9.22
10th Percentile	21.343
25th Percentile	27.98
Median	85.93
75th Percentile	178.46
90th Percentile	243.89
Maximum	278.77
Coefficient of Skew	0.58

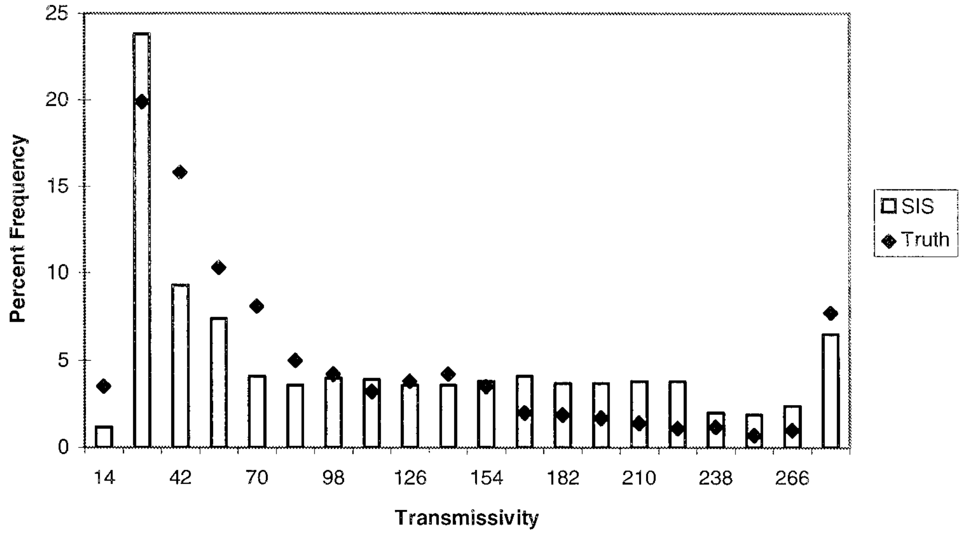


Figure 27 -- Histogram for Selected SIS Realization (15 Hard Data & 209 Soft Data)

IV.5. Connectivity of Extreme Values

The feature of concern in this work is the connectivity of the extreme values. Accordingly an estimated or simulated map is considered good if it reproduces the degree of connectivity existing in the Soda Springs data set. As was mentioned in Chapter III, it has been suggested that an appropriate measurement of the connectivity of values is the indicator correlogram. Using this as a tool to facilitate a comparison of the connectivity of the values exceeding the 90th percentile of the exhaustive distribution, plots were constructed in the same fashion as the variogram plots discussed at the beginning of this Chapter. Figures 28-31 are the indicator correlogram maps for the exhaustive transmissivity field, the OK Output, and the SIS Output. The values of the contours indicate the probability that a point at the head of some vector exceeds the cutoff value of 230.04 m^2/day given that the point at the tail of the vector does.

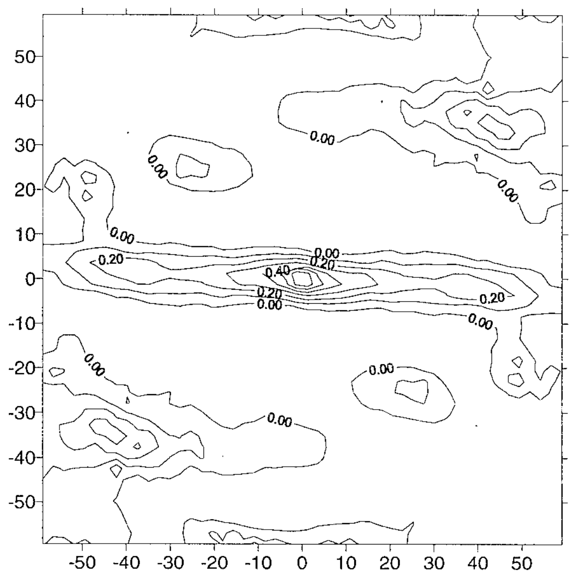


Figure 28 -- Indicator Correlogram Map for Exhaustive Transmissivity Field

Figure 28 indicates that the connectivity of high values in the Soda Springs data set is definitely anisotropic with the major axis traveling east-west. This feature is a direct result of the band of high values running across the top of the field. The indicator correlogram map for the OK results is shown as Figure 29. Obviously, the kriging algorithm was unable to pick up many of the higher values, much less any connectivity.

To create an indicator correlogram map for the ensemble of SIS realizations, the probability map can be used. This map will lend itself well since it can be easily converted to indicator data. Furthermore, this will serve as a check to see if this methodology produces results which honor the exhaustive covariance structure. The indicator correlogram map for the ensemble SIS results is included as Figure 30. It is apparent from Figure 30 that SIS produced results which exhibit a covariance structure quite similar to that of truth.

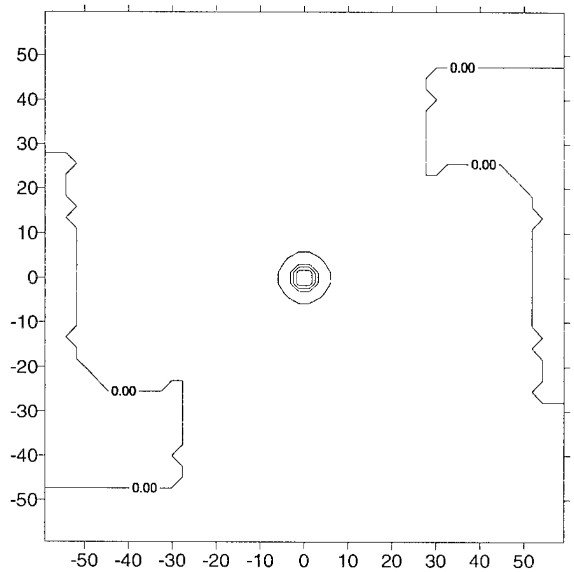


Figure 29 -- Indicator Correlogram for OK Output Given Scenario 3

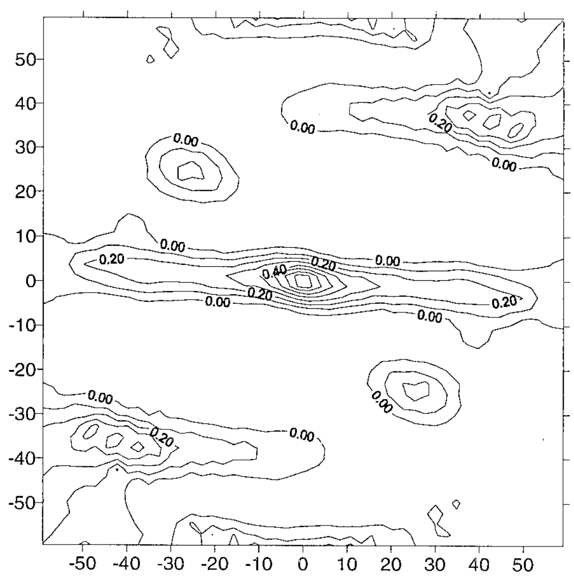


Figure 30 -- Indicator Correlogram for SIS Output Given Scenario 3

Finally, an indicator correlogram map is constructed for the selected SIS realization and is included as Figure 31. Although not quite as good a match as the previous map, the spatial covariance exhibited by this map is still close to that exhibited by truth.

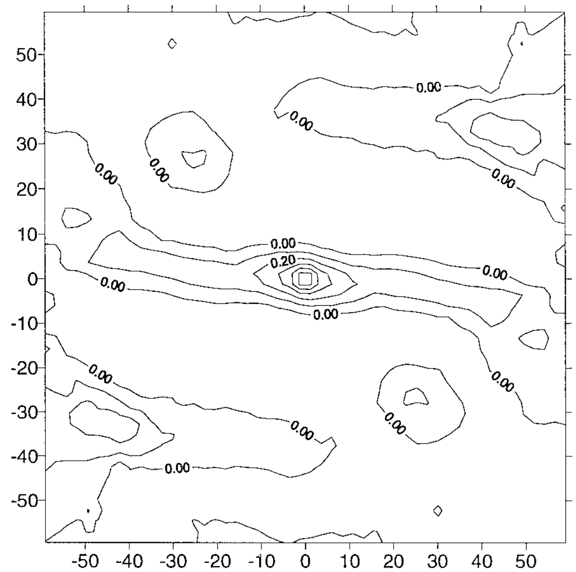
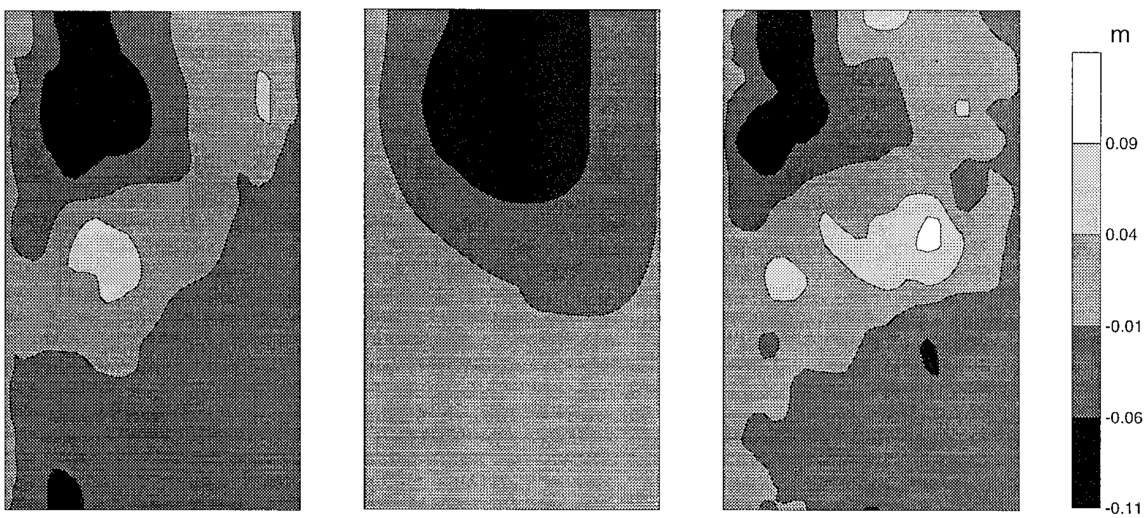


Figure 31 -- Indicator Correlogram for Selected SIS Realization from Scenario 3

IV.6. MODFLOW Output

MODFLOW, representing the transfer function, serves two purposes. First, it is used to generate maps of head values which allows for the comparison of different SIS realizations relative to truth. As a result, the final selection of an SIS realization intended to represent truth is conditioned upon the actual head values measured at the site. Second, MODFLOW is used to generate a particular realization of the response variable, or specific discharge, based upon the selected SIS realization. For comparative purposes, maps of specific discharge were also generated for the exhaustive data set as well as for the OK output.

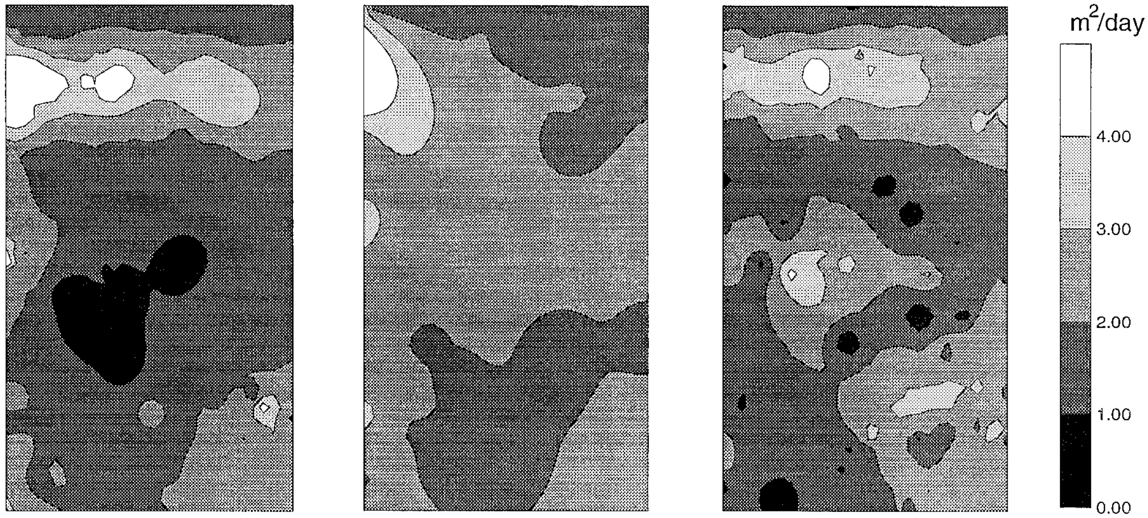
Maps of head fluctuations for truth, for the OK output, and for the selected SIS realization are shown as Figure 32. The head fluctuations were determined by subtracting the applied head gradient from the potentiometric head. Obviously, the map from the SIS realization more closely resembles truth than does that from the OK output.



"Truth" From OK Output From SIS Realization

Figure 32 -- Comparison of Head Fluctuation Maps for Scenario 3

Maps of specific discharge for truth, for the OK output, and for the selected SIS realization are shown as Figure 33. Again, the output from the selected SIS realization more closely resembles truth.



"Truth"   From OK Output   From SIS Realization

Figure 33 -- Comparison of Specific Discharge Fields for Scenario 3

V. Conclusions

A brief analysis was accomplished with the goal of determining the effectiveness of Sequential Indicator Simulation in the reproduction of the connectivity of extreme values in a transmissivity field. The further reaching goal of this study was to investigate a methodology that could fit within typical financial constraints while still providing the decision-maker sufficient information upon which to base a decision as well as a concept of the uncertainty involved.

SIS was used to generate 50 realizations each given three potential scenarios. For comparative purposes, Ordinary Kriging was performed on the sample data sets as well. As an average over all realizations, SIS was able to reproduce the connectivity of extremes fairly well regardless of the sampling scenario. Because of the smooth characteristics of the data set, OK was also able to reproduce the connectivity of the high values in sampling scenario 1, but when a lack of information in sampling scenario 2 forced a greater degree of smoothing, OK failed. In sampling scenario 3, the results of SIS were further improved by the inclusion of soft data.

A methodology for selecting SIS realizations for input into a flow model was also proposed in which the realizations with "critical" features most closely resembling those features deemed probable, were used to generate maps of heads for comparison to the measured head values. This strategy allowed for the conditioning of the SIS output based upon the probability of occurrence and upon the measured head values. The selected realization was then used to generate a map of specific discharges which closely resembles that of truth.

When hard data are limited and soft data are available, and when there exists the potential for the connectivity of extremes, SIS should prove to be a valuable tool in estimating a transmissivity field for subsequent processing through a flow model. As is the case with all geostatistical tools, however, the effective use of SIS must be accompanied by sound geological interpretation and input, and the decision-maker must be made aware of the uncertainties involved.

Bibliography

Alabert, Francois Georges. Stochastic Imaging of Spatial Distributions Using Hard and Soft Information. MS Thesis. Stanford University, August 1987.

Delhomme, J. P. "Spatial Variability and Uncertainty in Groundwater Flow Parameters: A Geostatistical Approach," Water Resources Research, 15: 269-280 (April 1979).

Desbarats, A. J. and R. M. Srivastava. "Geostatistical Characterization of Groundwater Flow Parameters in a Simulated Aquifer," Water Resources Research, 27: 687-698 (May 1991).

Devore, Jay L. Probability and Statistics for Engineering and the Sciences (Third Edition). Pacific Grove, California: Brooks/Cole Publishing Company, 1991.

Deutsch, Clayton V. and André G. Journel. GSLIB -- Geostatistical Software Library and User's Guide. New York: Oxford University Press, 1992.

Domenico, Patrick A. and Franklin W. Schwartz. Physical and Chemical Hydrogeology. New York: John Wiley & Sons, Inc., 1990.

Flatman, George T. and Angelo A. Yfantis. "Geostatistical Strategy for Soil Sampling: The Survey and the Census," Environmental Monitoring and Assessment, 4: 335-349 (1984).

Freeze, R. Allan. "A Stochastic-Conceptual Analysis of One-Dimensional Groundwater Flow in Nonuniform Homogeneous Media," Water Resources Research, 11: 725-741 (October 1975).

Gilbert, Richard O. Statistical Methods for Environmental Pollution Monitoring. New York: Van Nostrand Reinhold, 1987.

Goldberger, Arthur S. "Best Linear Unbiased Prediction in the Generalized Linear Regression Model," American Statistical Association Journal, 57: 369-375 (June 1962).

Gómez-Hernández, J. Jaime. "Generation of Hydraulic Conductivity Fields With Significant Connectivity of Extreme Values," Heterogeneity of Groundwater Flow and Site Evaluation, Proceedings of an NEA Workshop. 59-70. Paris: OECD Publicatoins, 1991.

Gómez-Hernández, J. Jaime and R. Mohan Srivastava. "ISIM3D: An ANSI-C Three-Dimensional Multiple Indicator Conditional Simulation Program," Computers & Geosciences, 16: 395-440 (1990).

Isaaks, Edward H. and R. Mohan Srivastava. An Introduction to Applied Geostatistics. New York: Oxford University Press, 1989.

Journel, A. G. Fundamentals of Geostatistics in Five Lessons. Washington D.C.: American Geophysical Union, 1989.

Journel, A. G. "Geostatistics for Conditional Simulation of Ore-Bodies," Economic Geology, 69: 673-687 (1974).

Journel, A. G. "Geostatistics: Models and Tools for the Earth Sciences," Mathematical Geology, 18: 119-140 (1986).

Journel, A. G. "Nonparametric Estimation of Spatial Distributions," Mathematical Geology, 15: 445-468 (1983).

Journel, A. G. and F. G. Alabert. "Focusing on Spatial Connectivity of Extreme-Valued Attributes: Stochastic Indicator Models of Reservoir Heterogeneities" SPE Paper 18324, Proceedings from the 1988 SPE Annual Technical Conference and Exhibition. 621-632. Houston, Texas: Society of Petroleum Engineers, 1988.

Journel, A. G. and F. G. Alabert. "New Method for Reservoir Mapping," Journal of Petroleum Technology, 42: 212-218 (February 1990).

Journel, A. G. and F. G. Alabert. "Non-Gaussian data expansion in the Earth Sciences," Terra Nova, 1: 123-134 (1989).

Journel, A. G. and Ch. J. Huijbregts. Mining Geostatistics (Fifth Printing). San Diego, California: Academic Press, 1991.

Journel, A. G. and M. E. Rossi. "When Do We Need a Trend Model in Kriging?" Mathematical Geology, 21: 715-739 (1989).

Neuman, Shlomo P. "Statistical Characterization of Aquifer Heterogeneities: An Overview," Recent Trends in Hydrogeology, Special Paper 189. 81-102. Boulder Colorado: Geological Society of America, 1982.

Poeter, Eileen and Peter Townsend. "Assessment of Critical Flow Path for Improved Remediation Management," Ground Water, 32: 439-447 (May-June 1994).

Rehfeldt, Kenneth R., J. Mark Boggs, and Lynn W. Gelhar. "Field Study of Dispersion in a Heterogenous Aquifer: 3. Geostatistical Analysis of Hydraulic Conductivity," Water Resources Research, 28: 3309-3324 (December 1992).

Ritzi, Robert W., Dale F. Jayne, Arthur J. Zahradnik, Adrian A. Field, and Graham E. Fogg. "Geostatistical Modeling of Heterogeneity in Glaciofluvial, Buried-Valley Aquifers," Ground Water, 32: 666-674 (July-August 1994).

Wen, Xian-Huan and Chen-Shan Kung. "Stochastic Simulation of Solute Transport in Heterogeneous Formations: A Comparison of Parametric and Nonparametric Geostatistical Approaches," Ground Water, 31: 953-965 (November-December 1993).

Woodbury, Allan D. and E. A. Sudicky. "The Geostatistical Characterization of the Borden Aquifer," Water Resources Research, 27: 533-546 (April 1991).

Vita

1Lt David Duane Kenyon [redacted]. He graduated from Widefield High School in 1988 and entered undergraduate studies at the United States Air Force Academy. He graduated with a Bachelor of Science in Civil Engineering and received his commission in May 1992. His first assignment was at Falcon AFB as the Environmental Coordinator. In May 1994, he entered the School of Engineering, Air Force Institute of Technology.

Permanent Address: [redacted]

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
December 1995

3. REPORT TYPE AND DATES COVERED
Master's Thesis

4. TITLE AND SUBTITLE
EVALUATING THE FEASIBILITY OF SEQUENTIAL INDICATOR SIMULATION IN REPRODUCING SPATIAL CONNECTIVITY IN A HETEROGENEOUS TRANSMISSIVITY FIELD

5. FUNDING NUMBERS

6. AUTHOR(S)
D. Duane Kenyon, 1Lt, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology,
WPAFB OH 45433-6583

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/GEE/ENP/95D-03

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Dr. Tom Stauffer
AL/EQC-OL
139 Barnes Drive, Suite 2
Tyndall AFB FL 32403-5323

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A Non-Parametric estimation technique was used to simulate realizations of a heterogeneous transmissivity field based upon sampled values from three different sampling scenarios. These realizations were compared to output from a parametric estimation technique with respect to truth as defined by an exhaustive data set of 6,000 transmissivity values. Estimated transmissivity fields were then used as input into a flow model from which fields of heads and specific discharges were obtained and compared. Given the financial limitations imposed upon the number and quality of samples reasonably available, Sequential Indicator Simulation, a non-parametric technique, was shown to be of considerable value when accompanied with sound geological input.

14. SUBJECT TERMS
Geostatistics, Stochastic Processes, Simulation, Ground Water

15. NUMBER OF PAGES
86

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to ***stay within the lines*** to meet ***optical scanning requirements***.

Block 1. Agency Use Only *(Leave blank)*.

Block 2. Report Date. Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C	- Contract	**PR**	- Project
G	- Grant	**TA**	- Task
PE	- Program Element	**WU**	- Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es). Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es). Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. *(If known)*

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."
DOE - See authorities.
NASA - See Handbook NHB 2200.2.
NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.
DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
NASA - Leave blank.
NTIS - Leave blank.

Block 13. Abstract. Include a brief *(Maximum 200 words)* factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code *(NTIS only)*.

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.