SAND--91-0758
DE93 004386

SAND 91-0758

The Use of Sequential Indicator Simulation to Characterize Geostatistical Uncertainty

Katherine M. Hansen
Department 323
Statistics and Human Factors
Sandia National Laboratories
Albuquerque, NM 87185

Abstract

Sequential indicator simulation (SIS) is a geostatistical technique designed to aid in the characterization of uncertainty about the structure or behavior of natural systems. This report discusses a simulation experiment designed to study the quality of uncertainty bounds generated using SIS. The results indicate that, while SIS may produce reasonable uncertainty bounds in many situations, factors like the number and location of available sample data, the quality of variogram models produced by the user, and the characteristics of the geologic region to be modeled, can all have substantial effects on the accuracy and precision of estimated confidence limits. It is recommended that users of SIS conduct validation studies for the technique on their particular regions of interest before accepting the output uncertainty bounds.

ACKNOWLEDGMENTS

The author thanks Carol Gotway, Bob Easterling, and Brian Rutherford, of the Statistics and Human Factors Department; and Chris Rautman of the Geoscience Assessment and Validation Department, for their many comments and valuable suggestions about the study and the manuscript. In addition, Geoffrey Watson of Princeton University, and Mohan Srivastava of FSS International, helped to review the work. Alex Treadway, of the Performance Assessment Development Division, helped with the computer software. Andre Journel and the members of the Stanford Center for Reservoir Forecasting are thanked for their advice and assistance in the early phases of the project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

Many studies in hydrology, petroleum geology, and the environmental sciences involve modeling the behavior of complex, three-dimensional systems. The hydrologic characterization of Yucca Mountain, a potential site for a nuclear waste repository, is such a project. Often the goal is the evaluation of a particular transfer function (such as a volume of recoverable oil or, in the case of Yucca Mountain, minimum ground water travel-time) over a region of interest. The availability of adequate data is usually a problem: most transfer functions require complete information on a number of variables (or attributes) over an entire region. In practice, such exhaustive data are rarely available, and the investigator must use limited data from a few sampled locations to infer reasonable attribute values over the entire region. The data and inferred values are then input to a transfer code, and the output is an approximation to the true value of the transfer function (the value obtained with perfect data over the entire region.)

Attribute values at unsampled locations may be assigned by estimation or by simulation. Estimation involves using the available data to identify a single value (for each attribute and each unsampled location) that is thought to lie close to the true but unknown value at that location. Simulation uses the available data to construct a probability distribution for each attribute at each location. A full realization of all attributes over the region of interest is then created by sampling from the constructed distributions. In both estimation and simulation, it is important to characterize the uncertainty associated with the use of less-than-perfect data to evaluate the transfer function.

Estimation techniques, like kriging (Olea. 1974; Journel and Huijbregts, 1978; Cressie, 1989), often come with built-in error estimates that are simple to compute but depend heavily on the validity of a number of parametric assumptions that can be quite difficult to verify in practice. In addition, it is often unclear how errors in the estimation of attribute values are propagated through different transfer functions.

When simulation is used, the variability associated with the output of transfer codes is characterized in a Monte Carlo fashion by repeatedly simulating new attribute values at the unsampled locations, and evaluating the transfer function over many such realizations. Simulation can be quite computer-intensive, and may require different

types of parametric assumptions than are required for estimation. In addition, it is not clear that the variance of the set of functional evaluations (over many simulations) will necessarily be a good approximation of the uncertainty associated with the use of simulated (rather than exact) data; or that a response distribution based on multiple simulations will be unbiased. These issues are taken up, for one simulation technique, in this report.

This work was completed under WBS number 1.2.3.2.2.2.2.

1.1 Purpose

The unsaturated tuffs at Yucca Mountain in southern Nevada are being investigated as a host for a potential radioactive waste repository. Licensing the repository will require that performance-assessment calculations show that a repository at Yucca Mountain will meet or exceed a series of federal requirements set forth in 40 CFR 191 (EPA, 1986) and 10 CFR 60 (NRC, 1986). These requirements include a probabilistic assessment of the movement of radionuclides from the repository to the accessible environment; as well as restrictions on the transport rate of radionuclides into rock surrounding the repository, and the preemplacement ground water travel time. In order to address these requirements, analysts must be able to model the hydrologic system at Yucca Mountain, and to quantify the uncertainty associated with the chosen models.

Over the last few years, geostatisticians at the Stanford Center for Reservoir Forecasting (SCRF) have developed a flexible simulation technique, known as sequential indicator simulation (SIS), that may be used for characterizing the uncertainty associated with the prediction of the performance of three-dimensional systems from limited, spatially-correlated data (Journel, 1986 and 1988; Journel and Alabert, 1989). This technique may provide a reasonable approach for the analysis of the data available for characterization of the Yucca Mountain site.

While SIS appears promising, a number of statistical properties of the method remain unstudied. For example, the effects of sample size and location, imperfectly known spatial correlation structure, and modeling errors are unknown. The reliability of performance models will depend on the influence that each of these factors has on simulations, and on the output of transfer functions. In order to make sense of sequential indicator simulations of actual field data from Yucca Mountain and other relevant

sites, it is important to characterize these influences.

Through the use of synthetic data in a controlled, statistical investigation, it is possible to isolate the effect of each relevant factor on sequential indicator simulations of models that are completely known to the investigator (but not to the simulation program.) For example, when the true value of a transfer function over a particular data set is known, it is possible to determine whether confidence intervals estimated by SIS accurately reflect the uncertainty associated with output values.

This report describes an extensive simulation study designed to address some of the statistical issues related to the SIS technique. A large number of simulations were carried out on a variety of synthetic data sets, and the performance of the SIS technique in different experimental situations was characterized both qualitatively and quantitatively.

This type of methodological study can be quite valuable in the regulatory environment within which the Yucca Mountain Site Characterization Project exists. SIS is a new technique, not yet well established in the literature or practice of geostatistics or risk assessment. Without knowledge of the statistical properties of the technique, regulators should be skeptical of failure probability estimates based on multiple sequential indicator simulations. (This is true of other untested simulation algorithms as well.) This study attempts to anticipate and respond to some of the concerns that they may express. The study has identified a number of experimental situations that can cause the SIS technique to produce inaccurate or imprecise results, and the report discusses precautions that may help users to avoid (or at least to be aware of) such situations in a practical study.

2.0 SEQUENTIAL INDICATOR SIMULATION

2.1 Motivation; Outline of the Method

Many of the classical geostatistical simulation techniques are based on an assumption of multivariate Gaussianity (or normality), which can be exceedingly difficult to verify in practice. By contrast, the SIS technique is designed to provide a nonparametric statistical framework for studying geological systems. The approach may be more flexible than the classical techniques, in that an assumption of multi-Gaussianity, with the resulting tendency towards maximum-entropy solutions, is not implicit in the SIS model. (For a discussion of the nonconservative implications of maximum-entropy solutions for performance assessment calculations, see Journel and Alabert, 1989.) Nonparametric variance and confidence interval estimates are found by repeated SIS sampling from a distribution of realizations that are equiprobable with respect to the available data and the specified covariance models. In addition, the SIS framework allows for the inclusion of inequality-type soft data (Journel, 1986) in the analysis, a feature not present in most other geostatistical techniques. It should be noted that strict stationarity, which is not assumed for parametric methods, is required for SIS.

The principle underlying SIS is that it is possible to completely specify the joint probability of a collection of N dependent events, $\{ A_j, j = 1, \ldots, N \}$ as the product of N conditional probabilities:

$$P(A_j, j=1, \ldots, N) = P(A_N \mid A_j, j=1, \ldots, N-1) \times \qquad (1)$$
$$\times P(A_{N-1} \mid A_j, j=1, \ldots, N-2) \times \cdots$$
$$\cdots \times P(A_2 \mid A_1) \times P(A_1) .$$

In order to demonstrate how this principle is applied to the simulation problem, we introduce the following notation:

S	space of interest ($\mathbb{R}^2$, the space of two-dimensional real numbers, in this report)
u	point in S
$Z(u)$	attribute measured at u (for convenience, Z is one-dimensional)
(n)	$\{ Z(u_\alpha), \alpha = 1, \ldots, n \}$; the available data, at n locations $u_\alpha \in S$
ψ	transfer function (or "response function"); defined on $Z(S)$

For example, a simplified model might have S representing a vertical cross-section of rock, Z representing porosity, (n) representing available porosity data at n locations in S, and $\psi(Z(S))$ representing a minimum water travel-time through the slice S, as a function of porosity only.

Suppose that we wish to characterize the behavior of ψ over S. The space S is divided up into N nodes. Data (the value of the attribute Z) are available at n of these nodes. SIS is used to simulate values for the remaining $N - n$ nodes, and the result is referred to as a realization. The transfer function ψ is evaluated over the realization, and the outcome is recorded. The process is repeated many times, yielding a suite of measurements of ψ, which can be used to estimate the value of the transfer function over the field S, and to assess the uncertainty associated with this estimate. A single realization is carried out as follows:

1. Start at a randomly-selected node, u_1. Derive the conditional distribution of $Z(u_1)$, given (n):

$$P[Z(u_1) \leq z \mid (n)] = P[Z(u_1) \leq z \mid Z(u_\alpha) = z_\alpha, \alpha \in (n)].$$

2. Draw a realization of $Z(u_1)$. Add this to the data set, which is now denoted $(n+1)$.
3. Move to a second randomly-selected node, u_2. Derive the distribution of $[Z(u_2) \mid (n+1)]$.
4. Draw a realization from this distribution, and add it to the data set.

Repeat steps 3 and 4 for all $N - n$ nodes to be simulated.

2.2 Deriving Conditional Distributions

Of course, the difficult part of the SIS procedure involves the derivation of the conditional distribution of Z at one node, given the available data at other nodes. Journel and Alabert (1989) discuss how this can be done using binary indicator random variables:

The event $\{ Z(u) \leq z \}$ can be characterized by an indicator variable:

$$I(u;z) = \begin{cases} 1, & \text{if } Z(u) \leq z, \\ 0, & \textit{otherwise.} \end{cases} \tag{2}$$

Then any conditional probability for $Z(u)$ can be written as a conditional expectation:

$$P[Z(u) \leq z \mid (n)] = E[I(u;z) \mid (n)]. \tag{3}$$

Now consider K threshold values, $\{ z_k, k = 1, \ldots, K \}$, over the range of the attribute Z. Each conditioning data point $Z(u_\alpha) = z_\alpha$ is coded into an indicator column with K members, each zero or one:

$$\{ Z(u_\alpha) = z_\alpha \} \rightarrow \{ i(u_\alpha; z_k), k = 1, \ldots, K \}$$

Next, the conditional probability distribution for $Z(u)$ is expressed as conditional to the n indicator columns:

$$\begin{aligned} &P[Z(u) \leq z_{k_0} \mid Z(u_\alpha) = z_\alpha, \alpha \in (n)] \\ &\approx E[I(u;z_{k_0}) \mid I(u_\alpha; z_k) = i(u_\alpha; z_k), k = 1, \ldots, K; \alpha \in (n)], \end{aligned} \tag{4}$$

with z_{k_0} being one of the K threshold values, $\{ z_k \}$. Note that there is a loss of information involved with discretizing the conditioning data, so that Equation (4) is only an approximation to the distribution in Equation (3). Journel and Alabert (1989) argue that it is reasonable to assume that $I(u;z_{k_0})$ is more correlated with $I(u_\alpha; z_{k_0})$ than with any of the other indicator data $I(u_\alpha; z_k)$ with $z_k \neq z_{k_0}$, and simplify Equation (4) to:

$$\begin{aligned} &P[Z(u) \leq z_{k_0} \mid Z(u_\alpha) = z_\alpha, \alpha \in (n)] \\ &\approx E[I(u;z_{k_0}) \mid I(u_\alpha; z_{k_0}) = i(u_\alpha; z_{k_0}), \alpha \in (n)]. \end{aligned} \tag{5}$$

This approximation ignores cross-correlations between indicators at different

thresholds.

In order to compute the conditional expectation in Equation (5), Journel and Alabert (1989) propose the first-order approximation:

$$P[Z(u) \leq z_{k_0} \mid (n)] \approx a_0 + \sum_{\alpha \in (n)} a_1(\alpha) \times i(u_\alpha, z_{k_0}). \tag{6}$$

The weights a_0 and $a_1(\alpha)$ are found using simple or ordinary kriging, versions of weighted least squares regression where the weights are determined by the covariance structure of the data. (For details, see Journel and Huijbregts, 1978.) The covariance structure is specified in terms of the *indicator covariance function*, $C_I(h;z)$:

$$C_I(h;z) = F(h;z) - F^2(z), \tag{7}$$

where:

$$F(h;z) = P[Z(u) \leq z, Z(u+h) \leq z],$$

and:

$$F(z) = F(0;z) = P[Z(u) \leq z]; \tag{8}$$

or the *indicator variogram*, $2\gamma(h;z)$:

$$2\gamma(h;z) = 2 \times [C_I(0;z) - C_I(h;z)]. \tag{9}$$

Note that the indicator covariance and variogram functions may vary as a function of z. Thus it is possible to model different degrees of spatial correlation at different levels of the attribute Z; this is one of the features that makes SIS a very flexible technique. Journel (1988) discusses the importance of this feature for flow modeling.

Once the K indicator covariance functions (one for each threshold value z_k) have been specified, kriging (simple or ordinary) is used to estimate the conditional probability that the value at a specific node will exceed each threshold, given the observed indicator columns. This is done using Equation (6), and provides the user with an approximate conditional distribution for the node. A single sample is taken from this distribution, and, as outlined earlier, this simulated value is added to the data set and attention shifts to the next randomly-chosen node.

Journel and Alabert (1989) state that realizations constructed in this manner honor the observed data and the specified covariance structures; we have checked this

assertion and our findings are discussed in section 4.5.

2.3 Application

As discussed earlier in this section, the SIS technique can be used to simulate a large number of equiprobable realizations of $Z(S)$. Journel (1988) states that the differences between the realizations themselves "should provide a measure of spatial uncertainty about the input" [to a transfer function, response function or flow simulator (used synonymously)]. Journel (1988) continues: "That input uncertainty can then be processed through the flow simulator to yield the corresponding measure of uncertainty on the response function. . ." The procedure is illustrated in Figure 2-1. Although SIS has other applications, the ability to gauge the uncertainty associated with estimates of the response function $\psi(Z(S))$ is the focus of this report.

2.4 Mechanical Issues

A number of mechanical problems exist related to the coding of the SIS algorithm. Some of these, related to the use of kriging to estimate probabilities, may be corrected in a straightforward, if not theoretically-pleasing, fashion. In other situations, the SIS literature gives the user very little guidance in making decisions that may have a large influence on the resulting simulations.

The order-relations problem is the first of two correctable situations. For two threshold values, z_{k_1} and z_{k_2}, with $z_{k_1} < z_{k_2}$, the estimated conditional probability of a node exceeding z_{k_2} may be greater than the estimated conditional probability of that same node exceeding z_{k_1}, when the actual probabilities must be in the opposite order. This problem can occur when the indicator variograms for the two thresholds differ. A simple correction, discussed in Hohn (1988), involves reducing the probability of exceeding the higher threshold to the probability of exceeding the lower threshold. This correction assigns zero probability to the interval between the two thresholds. Other techniques may also be used. Appendix A describes the correction applied in the software issued by SCRF.

The second problem associated with kriging for probabilities is that it is not uncommon for kriged estimates to lie outside of the range [0, 1]. This situation, which

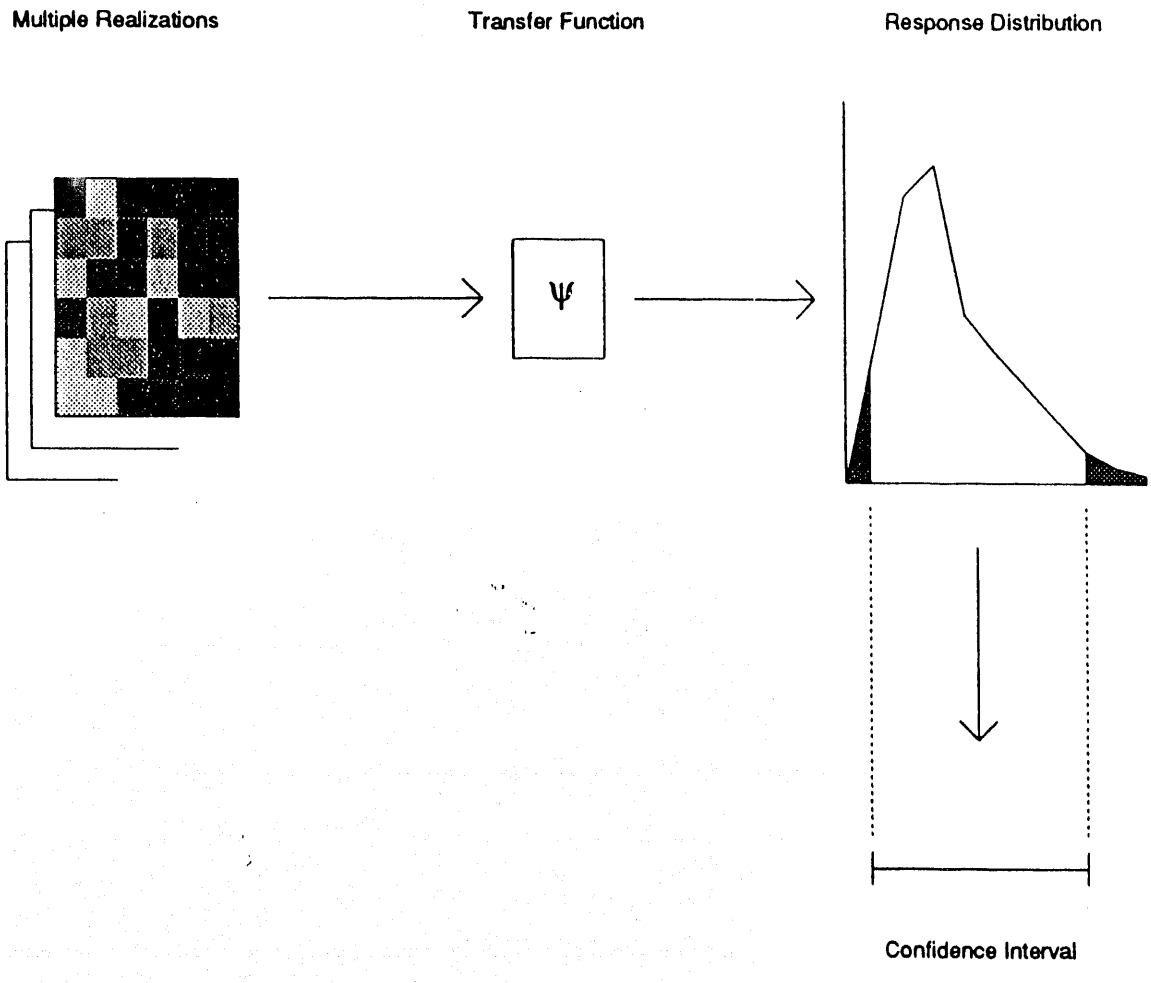


Figure 2-1 (After Journel, 1988.) Schematic diagram of the use of sequential indicator simulation to generate a response distribution for the transfer function ψ.

results from negative kriging weights, is easily corrected by setting negative probabilities equal to zero and probabilities exceeding one equal to one. However, the need for such adjustment calls into question the meaning of kriged probabilities.

The other issues involve the simulation of an attribute value $Z(u)$, once the indicator distribution for the node u has been constructed. Figure 2-2 shows how a bin (the region between two threshold values) is chosen. Once the bin is selected, however, it is unclear how to best simulate a value within that bin. The user can generate a value from a uniform (or any other) distribution across the bin. Unfortunately, any such decision involves completely specifying a within-bin distribution, which runs somewhat counter to the nonparametric philosophy of SIS. Further, the simulation of values in the outer two bins can be problematic. Suppose that uniform within-bin distributions are called for. The user needs to specify minimum and maximum allowable values, which serve as the end limits of the lower and upper bins, respectively. The choice of these values will have a large and direct influence on the tails of the simulated distribution, and yet the user may have little prior knowledge or sample data to act as a guide in selecting appropriate extreme values. Because many relevant transfer functions (e.g., groundwater travel-times) are sensitive to the behavior of extreme values, this situation can become particularly troublesome.

2.5 Modeling Indicator Variograms

Because SIS uses either simple or ordinary kriging in the construction of conditional distributions, it is necessary for the user to specify a covariance model at each indicator threshold used in the kriging. The use of different indicator variograms at different levels of the variable of interest provide the user with a great deal of flexibility in modeling spatial dependency; however, the user may be forced to estimate a large number of indicator variograms from a relatively small data set. As Figure 2-3 shows, fitting reasonable variograms, given only a sparse sample, is often extremely difficult. This problem is by no means unique to SIS, though; it applies to virtually all kriging-related geostatistical simulation and estimation procedures. In some cases, the lack of sample information may be overcome if a geologic analog of the area to be simulated exists, and can be sampled extensively. This borrowing of information from related regions is recommended by the SCRF group. Of course, this method is only

Threshold	Exceedence Probability	
min	1.00	
z_1	.93	
z_2	.84	
z_3	.61	
		← r=.57
z_4	.40	
z_5	.13	
max	0.00	

Figure 2-2 Illustration of how a bin is selected for an unsampled node. In this example, there are K=5 indicator thresholds, in addition to the minimum and maximum values set by the user. The probability of exceeding the minimum is set equal to unity, and the probability of exceeding the maximum is set equal to zero. The probabilities of exceeding the indicator thresholds at the node are estimated by kriging. Next, a random number, r, between zero and one is generated. The selected bin is determined by comparing the generated value with the kriged exceedence probabilities. In the example above, the generated value, r=0.57, lies between the exceedence probabilities of the third and fourth thresholds. The simulated value at this node will thus be between z_3 and z_4.

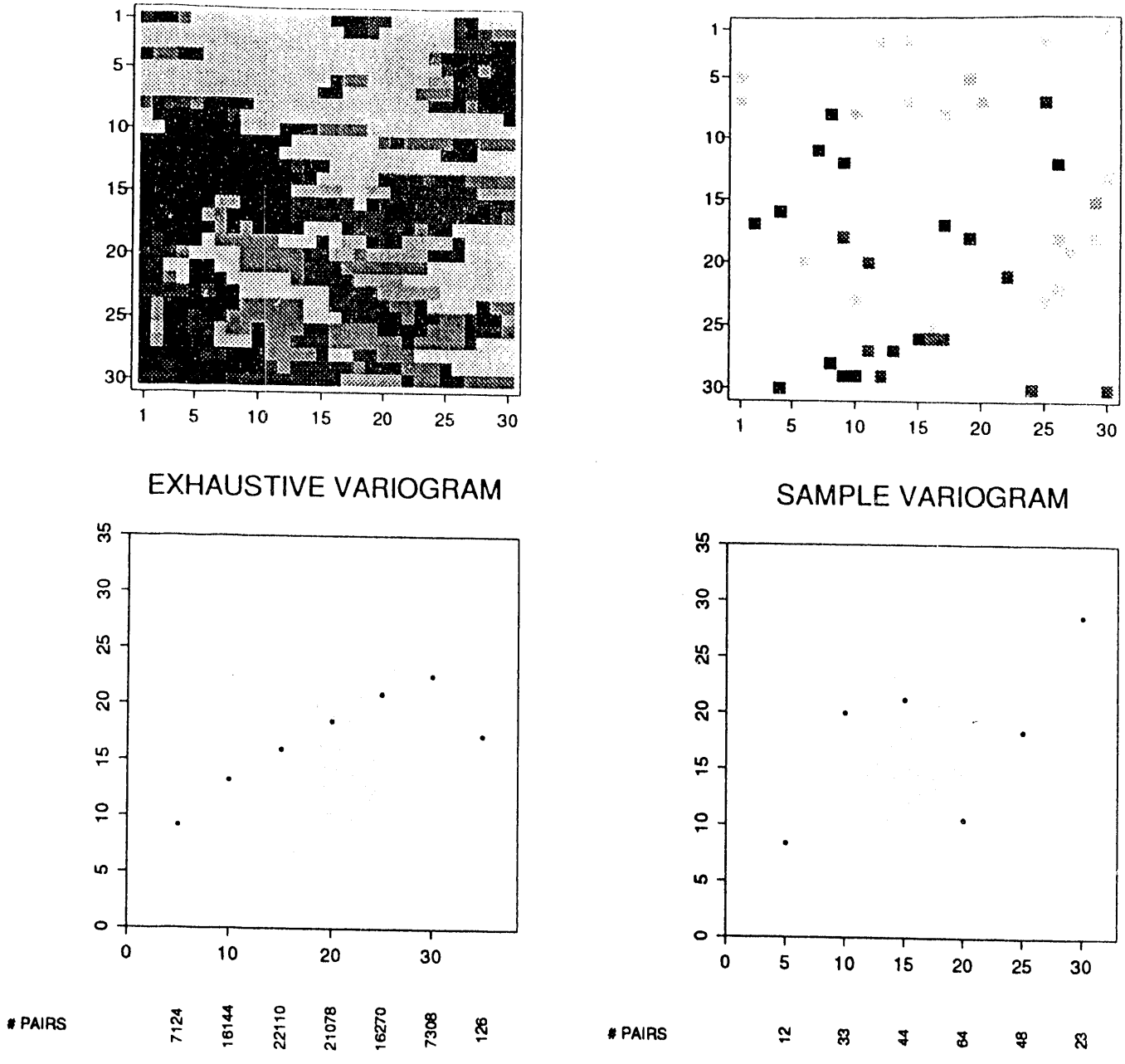


Figure 2-3 Graphical illustration of the difficulty encountered in attempting to model a variogram using sampled data. The left-hand panel shows an exhaustive 30×30 data set, and the variogram of these data. The right-hand panel shows a randomly-selected sample of 45 values from the exhaustive data set, and the sample variogram computed from these 45 points. Even with a relatively high sampling rate of 5%, it would be difficult to correctly model the exhaustive variogram from the sample data.

appropriate when the spatial relationships among locations in the related region are similar to those found in the actual region to be modeled. In cases where a geologic analog site is not available, a combination of sample data, geologic intuition, and other available information, must be used to develop plausible variogram models. The influence of the chosen indicator variogram models on the results of SIS is examined in the simulation study.

3.0 DESIGN OF THE SIMULATION EXPERIMENT

A two-part simulation experiment was carried out to examine the effects of the various factors discussed above on the results of SIS. Interest was focused on the dual issues of the accuracy and precision of transfer function predictions based on multiple simulations. To illustrate the concepts of accuracy and precision, suppose that we have used SIS to construct a 95% confidence interval for the minimum groundwater travel time through a region. (The construction of confidence intervals will be discussed shortly.) A confidence interval is accurate if it does contain the true value, and it is precise if it is provides enough information to constrain the true value. The precision of a confidence interval is directly related to its width: a confidence interval for minimum water travel-time that runs from 3000 to 3050 years is precise, while one that runs from 1 to 10,000 years may not be. The concepts of accuracy and precision are illustrated in Figure 3-1.

The first phase of the experiment was designed to study the effect of a large number of combinations of the chosen experimental factors over multiple simulations. Data obtained from this phase were useful primarily in quantifying the precision of SIS realizations. The second phase of the experiment was a more in-depth examination of a subset of the factor combinations used in the first phase. These latter experiments involved a larger number of simulations, and provided us with information about both accuracy and precision.

3.1 Experimental Factors

The two phases of the experiment were designed to study the effect of changes in a number of factors that may influence the accuracy and precision of SIS simulations. The factors considered were the number and location of the available samples, the variogram information input to the SIS code, and the statistical properties of the exhaustive data sets themselves. The levels at which these factors were set are listed in Table 3-1. Table 3-2 is a design matrix for the two phases of the experiment, showing the factor combinations used in each run. A brief discussion of the factors follows.

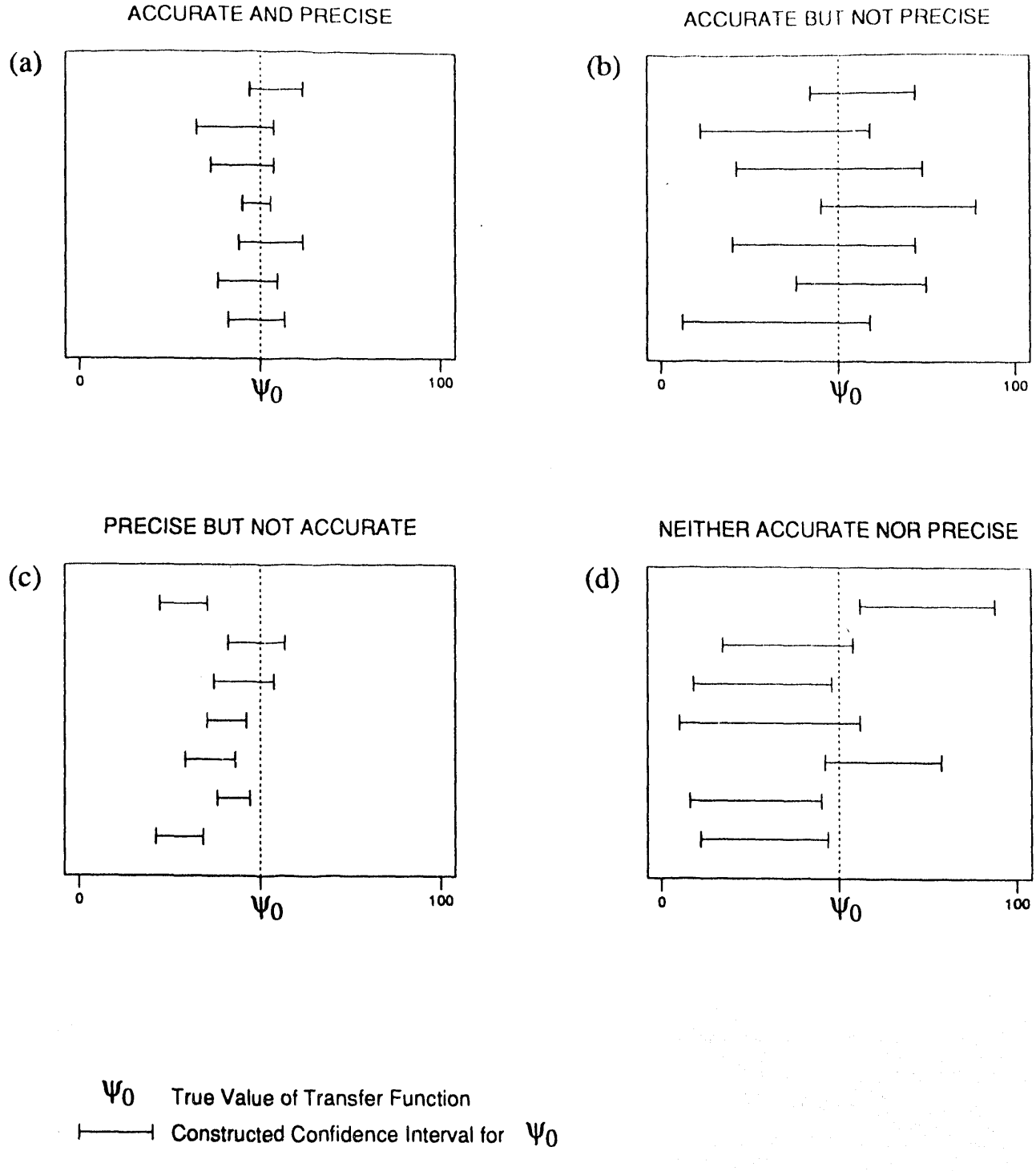


Figure 3-1 Illustration of the concepts of accuracy and precision, as applied to confidence intervals. In this example, suppose that the transfer function may only take on values within the range from 0 to 100, and let ψ_0 represent its true value. The information gained from the imprecise simulation distributions, (b) and (d), is minimal. Ideally, we would like to be in situation (a).

TABLE 3-1

FACTORS USED IN THE SYNTHETIC STUDY OF SEQUENTIAL INDICATOR SIMULATION

Factor Name	Abbreviation	Levels
sample size	*n*	5, 15, 30, 45
sample location	*sl*	(1) random uniform (2) random clustered*
SIS input variogram information source	*iv*	(1) correct theoretical (2) incorrect theoretical** (3) correct exhaustive (4) related exhaustive
method of generating exhaustive data sets	*gm*	(1) SIS (2) Choleski

*(2a) 2:1 preferential sampling in the first and third quadrants, (2b) 1.33:1 preferential sampling in first quadrant, (2c) 2.67:1 preferential sampling in the first quadrant.

**(2a) SIS input variogram information and correct theoretical differs in only in terms of the marginal cumulative distribution function (cdf) specified, (2b) SIS and correct theoretical variogram information differs in min/max thresholds only, (2c) SIS and correct theoretical variogram information differs both in marginal cdf and in max/min thresholds.

TABLE 3-2

DESIGN MATIRX FOR THE SYNTHETIC EXPERIMENT

Run	Phase I	Phase II	*n*	*sl*	*iv*	*gm*
A	y	y	5	1	1	1
B	y	y	15	1	1	1
C	y	y	30	1	1	1
D	y	y	45	1	1	1
E	y		30	2a	1	1
F	y		45	2a	1	1
G	y		45	2b	1	1
H	y	y	45	2c	1	1
I	y		15	1	2a	1
J	y	y	30	1	2a	1
K	y		30	1	2b	1
L	y		30	1	2c	1
M	y	y	30	1	3	1
N	y	y	30	1	4	1
P	y	y	45	1	4	1
Q	y	y	30	1	3	2
R	y		45	1	3	2
S	y		15	1	4	2
T	y		30	1	4	2
U	y		45	1	4	2

A y in column 2 indicates that the variable combination was run in the first phase, and a y in column 3 indicates that the combination was run in the second phase.

3.1.1 Sample Size: *n*

The first factor, n, is the number of sample data points (out of $30 \times 30 = 900$ in an exhaustive data set) that were included as input to the simulation program. Simulations were run with n set at 5, 15, 30, and 45 points. By including this factor, we hope to be able to determine whether (and the extent to which) changing the sample size affects accuracy and precision.

3.1.2 Sample Location: *sl*

The factor sl, for sample location, specifies the way in which the n sample points were chosen. This factor is included to provide a comparison between the statistical properties of simulations conditioned on clustered data, and simulations conditioned on unclustered data.

For sl=1, the sample points were randomly selected from the 30×30 grid, with each point having an equal probability of selection. For $sl = 2a$, $2b$, and $2c$, sampling was done preferentially in certain quadrants of the grid, to examine the effect of clustered data. When $sl = 2a$, points in the first and third quadrants were twice as likely to be chosen as points in the second and fourth quadrants. When $sl = 2b$, the points in the first quadrant were 1.33 times as likely to be chosen as points in the other three quadrants, and when $sl = 2c$ points in the first quadrant were 2.67 times as likely to be chosen as points in the other quadrants. Thus, for $sl = 2a$, we expect about 67% of the samples to lie in the first or third quadrants; for $sl = 2b$, we expect about 57% of the samples to lie in the first quadrant; and for $sl = 2c$, we expect about 73% of the samples to lie in the first quadrant. Figure 3-2 shows typical samples taken using each of the four clustering schemes.

3.1.3 Input Variogram Information: *iv*

In an actual field study, an investigator will not have access to perfect variogram information about the region of interest. He or she must model indicator variograms from geologic intuition, available samples, and data from nearby regions thought to have similar properties. Will this affect the accuracy and precision of uncertainty

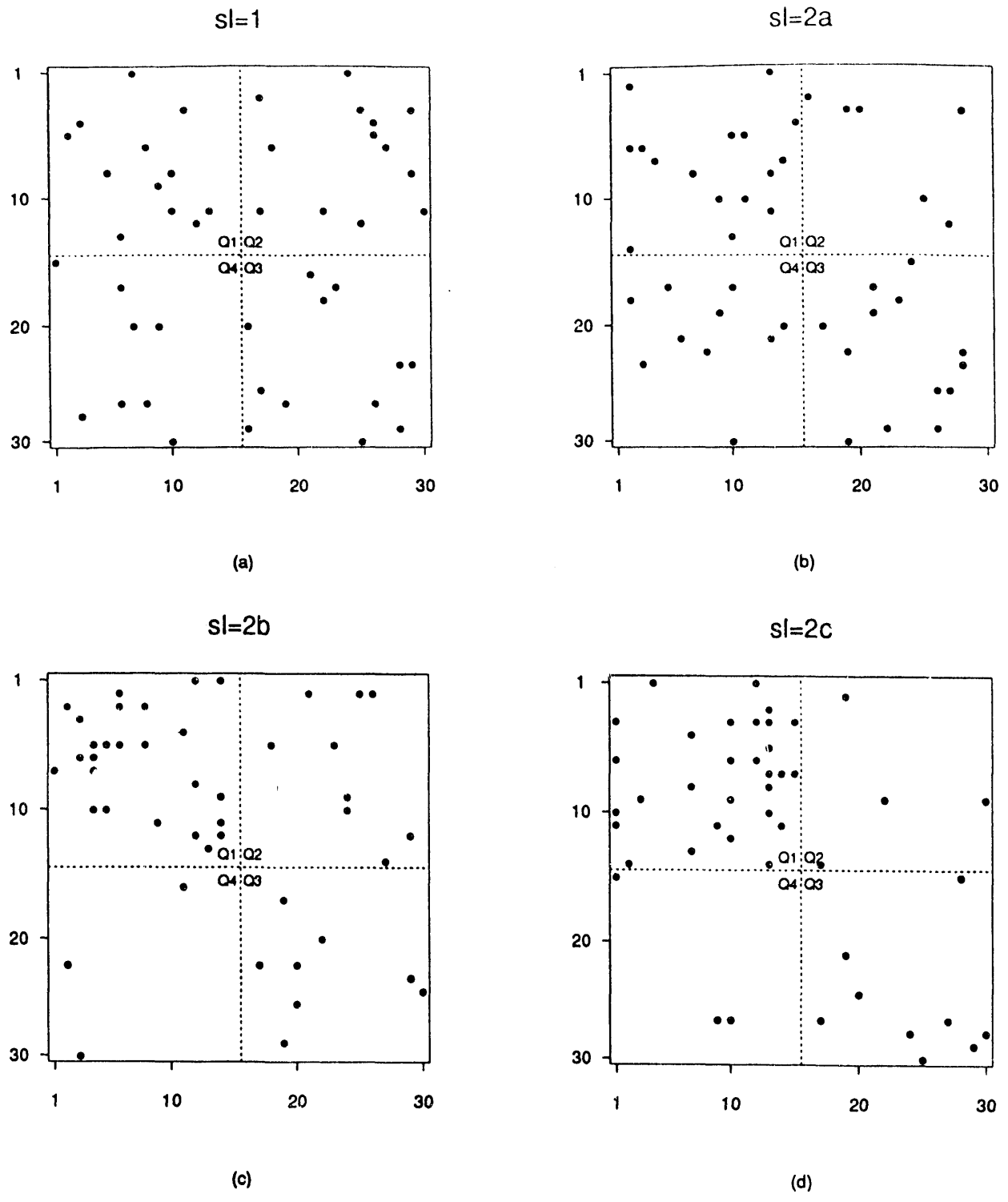


Figure 3-2 Examples of samples drawn using each of the four clustering schemes. (a) sl=1; random sampling from a uniform distribution. (b) sl=2a; mildly clustered random sampling with a 2:1 ratio of points in the first and third quadrants (Q1 and Q3) to points in the second and fourth quadrants. (c) sl=2b; clustered random sampling with a 1.33:1 ratio of points in the first quadrant to points in the other three quadrants. (d) sl=2c; clustered random sampling with a 1.67:1 ratio of points in the first quadrant to points in the other three quadrants. In all of the clustered sampling schemes, the within-quadrant sampling distributions are uniform.

estimates computed by sequential indicator simulations? In order to examine this issue, we included runs of the simulation experiment that used variogram information from a variety of sources.

The first step in specifying an indicator variogram for a particular cutoff value, z_k, is providing $F(z_k)$, the (unconditional) value of the cumulative distribution function of the variable $Z(u)$ at cutoff value z_k (See Equation 8). The next step is choosing nugget and shape parameters for the indicator variogram. A set of indicator variograms, at different levels of a variable of interest, will specify a marginal distribution for that variable, and will also dictate the extent of the correlation to be seen at each level of the variable. Thus, in order to test the influence of input indicator variogram information on simulated results, we study the effects of both the marginal distributions (also referred to as proportions), and the correlation structures specified. These factors are referenced by *iv* and *gm*; *iv* specifies the type and source of indicator variogram information input to the SIS program, and *gm* specifies the simulation method used to generate the exhaustive data sets, and thus whether the different degrees of correlation are present at different indicator levels in the exhaustive data sets. The exact variogram parameters used in each run are discussed in detail in Appendix B; the input given to computer codes is specified in Appendix C.

For runs with $iv = 1$, the exhaustive data sets were generated using SIS itself, and the same theoretical indicator variograms used to generate the exhaustive data set were also used to generate the multiple realizations. These variograms were spherical in shape, with ranges of 7, 5, 3, 2, and 2, at the 0.20, 0.35, 0.50, 0.65, and 0.80 quantiles, respectively. (See Appendix B for details.)

When $iv = 2a$, $2b$, or $2c$, the exhaustive data sets were again generated using SIS, but using input indicator variogram information that differed from the theoretical. For $iv = 2a$, the variogram information used to generate the realizations differed in proportions only from those used to generate the exhaustive data set: that is, the threshold -2 might be the 0.20 quantile of the theoretical distribution for the exhaustive data set, and the 0.25 quantile of the theoretical distribution for the realizations. For $iv = 2b$, the theoretical distributions differed only in the specified minimum and maximum values. For $iv = 2c$, the theoretical distributions of the exhaustive data sets and the realizations differed in the specified minimum and maximum values, as well as in

the proportions.

When $iv = 3$, the input variogram information was computed directly from the exhaustive data set. No functional form was fit to the raw variogram of the exhaustive data set, since evaluation was required only at the distances at which information was available. When $iv = 4$, the input variograms were taken from a data set that is related to the exhaustive data set, as discussed in the previous section. The related data set used for variogram estimation was the same size as the exhaustive data set, and had the same theoretical variograms. Due to the unreliability of variograms modeled from a sparse sample, and the time involved in plotting and examining sample characteristics, no simulations were run using variograms extracted from sample data.

3.1.4 Method of Generating Exhaustive Data Sets: *gm*

Of course, the properties of real geologic regions come about as a result of natural processes, and are not created by sequential indicator simulation. Thus, it is important that we study the behavior of SIS-based confidence intervals on data sets that have not themselves been generated using SIS. The factor *gm* specifies whether or not SIS itself was used to generate exhaustive data sets. When $gm = 1$, the exhaustive data sets were generated using SIS, and were conditional on the same two fixed data points (see Appendix C, Table C.2, for details). When $gm = 2$, the exhaustive data sets were not generated using SIS. They were created by first generating a vector of 900 independent random numbers that are each uniformly distributed with a mean of zero and unit variance. This vector is then multiplied by one-half of the Choleski decomposition of a chosen theoretical covariance matrix (see Anderson, 1984), and the elements of the resulting product are associated with locations on the 30×30 study grid. (The theoretical covariance matrix chosen corresponds to a spherical variogram with a range of ten and a sill of five.) Because they are weighted sums of independent random variables, the transformed values have a univariate histogram that tends to appear normal in shape. Note, however, that this is not a directly Gaussian simulation technique. As discussed in Appendix D, for this technique, with the indicator levels used in this report (the 0.20, 0.35, 0.50, 0.65, and 0.80 quantiles), the same range of correlation is observed at the different indicator levels.

3.1.5 Soft Information

The use of soft information to supplement the available hard data at sample locations was not considered in the simulation study. It is reasonable to suppose that the use of such additional information could be quite valuable in actual field studies.

3.2 Transfer Functions

Table 3-3 lists the seven transfer functions that were evaluated for each complete realization.

3.2.1 Reproducing the Univariate Distribution: ψ_1 - ψ_5

The first five transfer functions represent the percent of the data that exceeds five fixed thresholds, corresponding to the indicator thresholds used by SIS in the control runs of the experiment. These functions were used simply to determine how well simulated data sets reproduce the univariate distribution of an exhaustive data set. While this is an important characteristic of the simulations, it is not a sufficient basis on which to judge the simulation technique. The issue of the spatial relationships among variables is extremely important in real-world hydrologic studies, and must be incorporated into the simulation study. For this reason, the sixth and seventh transfer functions were added. Both ψ_6, a simplified minimum path-finder, and ψ_7, a relative of the spatial covariance, are sensitive to extreme values, and depend on the spatial structure of a realization, in addition to its univariate distribution.

3.2.2 A Simplified Minimum Path Finder: ψ_6

The sixth transfer function, ψ_6, represents a simplified minimum path finder through a two-dimensional grid. The original plan was to consider all paths through the simulated region that start in any grid point on the left-hand side of the simulated area, and at each step move one node to the right and either one node upwards, one node downwards, or horizontally. The path that has the minimum summed node values upon reaching the right-hand edge of the simulated area is considered to be the

TABLE 3-3

THE SEVEN TRANSFER FUNCTIONS USED IN THE SYNTHETIC STUDY

Transfer Functions	
ψ_1	% of data exceeding -2**
ψ_2	% of data exceeding -1**
ψ_3	% of data exceeding 0**
ψ_4	% of data exceeding 1**
ψ_5	% of data exceeding 2**
ψ_6	"minimum path finder"; details in text
ψ_7	related to spatial covariance; details in text

** The values {-2,-1,0,1,2} are the indicator thresholds used for sequential indicator simulation.

minimum path, and the sum of the node values is the value of the transfer function. Let us denote this transfer function by ψ_{6a}. Figure 3-3(a) illustrates a ψ_{6a} minimum path through a small grid. The problem with this transfer function is that it is very difficult to compute for a grid of 30×30 nodes. The number of paths that need to be tested for different grid sizes is listed in Table 3-4. The computer time required to compute ψ_{6a} for a large number of realizations is prohibitive.

For this reason, ψ_{6b} was proposed. This is a simplification of ψ_{6a}, allowing candidate paths that move one node to the right on each step (as before), but can only move horizontally or one node downwards. In other words, paths that move in the upwards direction at any step are no longer acceptable. The ψ_{6b} minimum path through the small grid is drawn in Figure 3-3(b); note that the starting position has moved up one node from the starting position of the less restrictive ψ_{6a} path. Table 3-4 shows that the use of ψ_{6b} leads to a considerable reduction in the number of candidate paths to be tested. However, the computing time for repeated realizations of a 30×30 grid is still unacceptably high.

The transfer function ultimately chosen, ψ_6, is a further simplification of ψ_{6b}. At each step, the path must move over two nodes to the right and either zero or one nodes downward; the ψ_6 path through the grid is shown in Figure 3-3(c). These criteria lead to a large reduction in the number of paths to be tested, as shown in Table 3-4. Run times for ψ_6 are short enough to allow for its use on a large number of simulated data sets.

The function ψ_6 is sensitive to spatial connectivity among extreme values (particularly low values) of the simulated variable. It is designed to test one of the assumed strengths of the SIS technique - namely, that realizations may be generated that have a different correlation extent at different levels of the variable of interest.

It is important to note here that ψ_6 is not being proposed for use in actual hydrologic flow codes. Rather, it is simply a tool to help us to study the extent to which sequential indicator simulations can be used to gauge the uncertainty in non-linear, spatially-dependent functions evaluated over sparsely sampled regions. If the SIS technique has difficulty characterizing the uncertainty in relatively simple functions like ψ_6, it is reasonable to expect that these problems will persist for more complex and realistic codes.

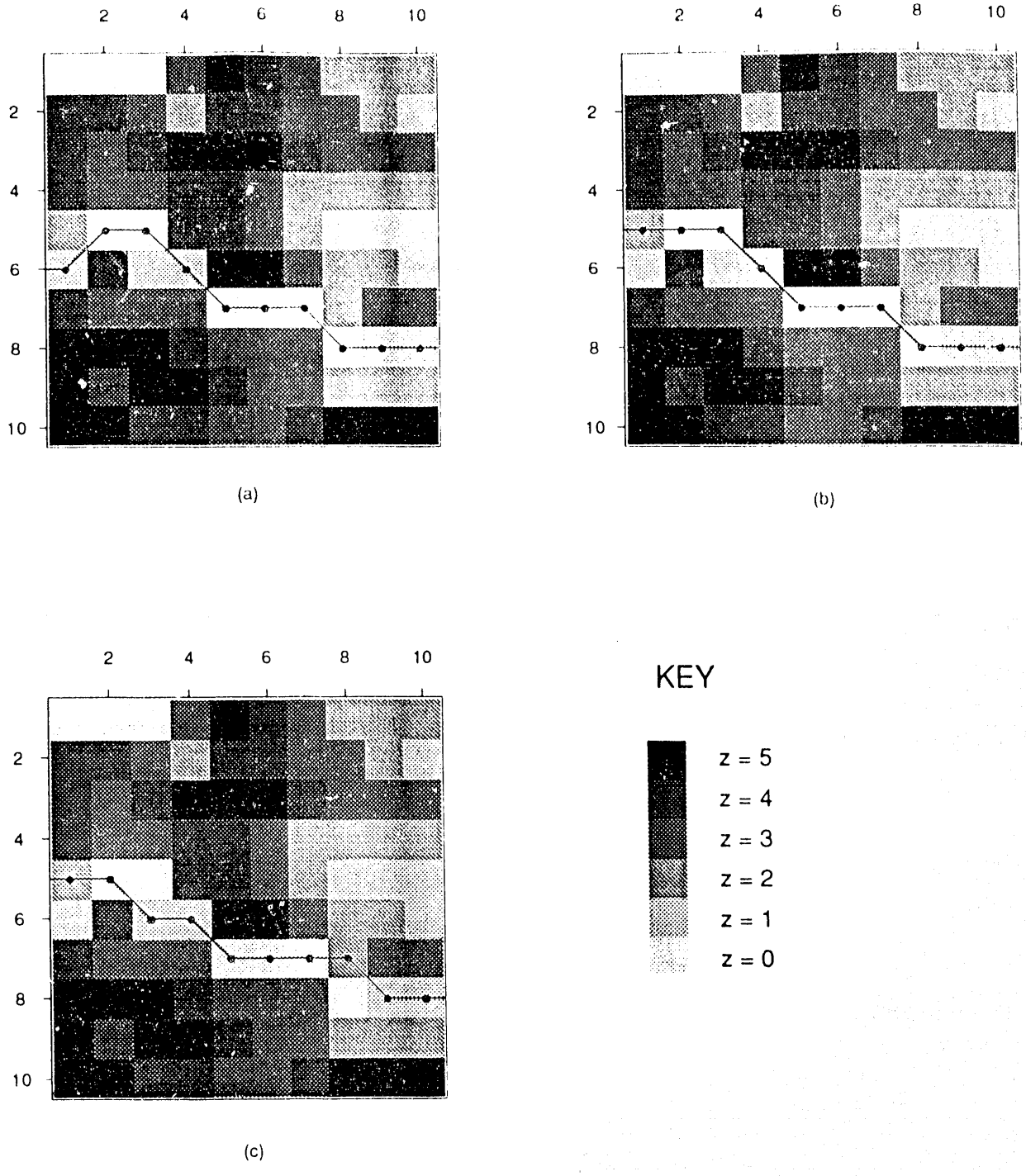


Figure 3-3 "Minimum flow paths" for a 10 × 10 data set using the transfer functions ψ_{6a}, ψ_{6b}, and ψ_6. (a) Flow path for ψ_{6a}, allowing movement both upwards and downwards, has a value of 4. (b) The transfer function ψ_{6b} allows only horizontal or downward movement, and has a value of 5. (c) The most restrictive transfer function, ψ_6 requires horizontal movement at every even-numbered step. The value of the minimum flow path using this definition increases to 8.

TABLE 3-4

THE NUMBER OF CANDIDATE PATHS FOR TRANSFER FUNCTIONS ψ_{6a}, ψ_{6b}, and ψ_6

Grid Size	Number of Candidate Paths**		
	ψ_{6a}	ψ_{6b}	ψ_6
4 × 4	68	20	7
10 × 10	136,946	2,816	128
12 × 12	1,515,296	13,312	160
18 × 18	--	1,245,184	3,584
30 × 30	--	--	376,832

**Computation for multiple simulations becomes unfeasible at around 500,000 tested paths.

3.2.3 A Relative of the Spatial Covariance: ψ_7

The last transfer function, ψ_7, is the average of the lag-1 spatial products in the vertical and horizontal directions. The equation for ψ_7 is as follows:

$$\psi_7 = \frac{1}{2\times 29\times 30}\left[\sum_{r=1}^{30}\sum_{c=1}^{29} Z_{r,c}\, Z_{r,c+1} + \sum_{r=1}^{29}\sum_{c=1}^{30} Z_{r,c}\, Z_{r+1,c}\right], \qquad (10)$$

where $Z_{r,c}$ is the simulated value for the node at row r and column c. As is true with ψ_6, this function is sensitive to spatial patterns among the data. Note however, that ψ_7 is a close relative of the spatial covariance function. Thus, its variability might be better characterized by simulations that use a single z-covariance function than by SIS simulations that use multiple indicator covariance functions. When evaluating results for ψ_7, we should recognize that the SIS algorithm is not optimal for this type of transfer function.

3.3 Experimental Phase I

In the first phase, every run used 10 separate exhaustive data sets, each made up of 900 data points in two dimensions, and each generated with the same theoretical indicator variograms. The nodes in the data sets were equally spaced along a 30 × 30 grid. The seven transfer functions were evaluated over each exhaustive data set; this provides true values of the seven transfer functions. Next, a sample was taken from each data set, and SIS was used to generate 100 separate realizations of the full 30 × 30 grid from this sample. The transfer functions were evaluated over each realization, and from the 100 realizations, 95% confidence intervals for each transfer function were computed. The confidence intervals were constructed using the *percentile method* (Efron and Gong, 1983): the interval between the 0.025 and 0.975 percentiles of the collection of evaluations of the transfer function was taken to be a 95% confidence interval for the value of the transfer function evaluated on the true (exhaustive) attribute values over the region of interest. A single run in this part of the experiment yields ten separate confidence intervals for each transfer function, one for each data set. Figure 3-4 is a schematic outline of the procedures constituting a run. Figure 3-5 shows an example of a single exhaustive data set and several SIS realizations.

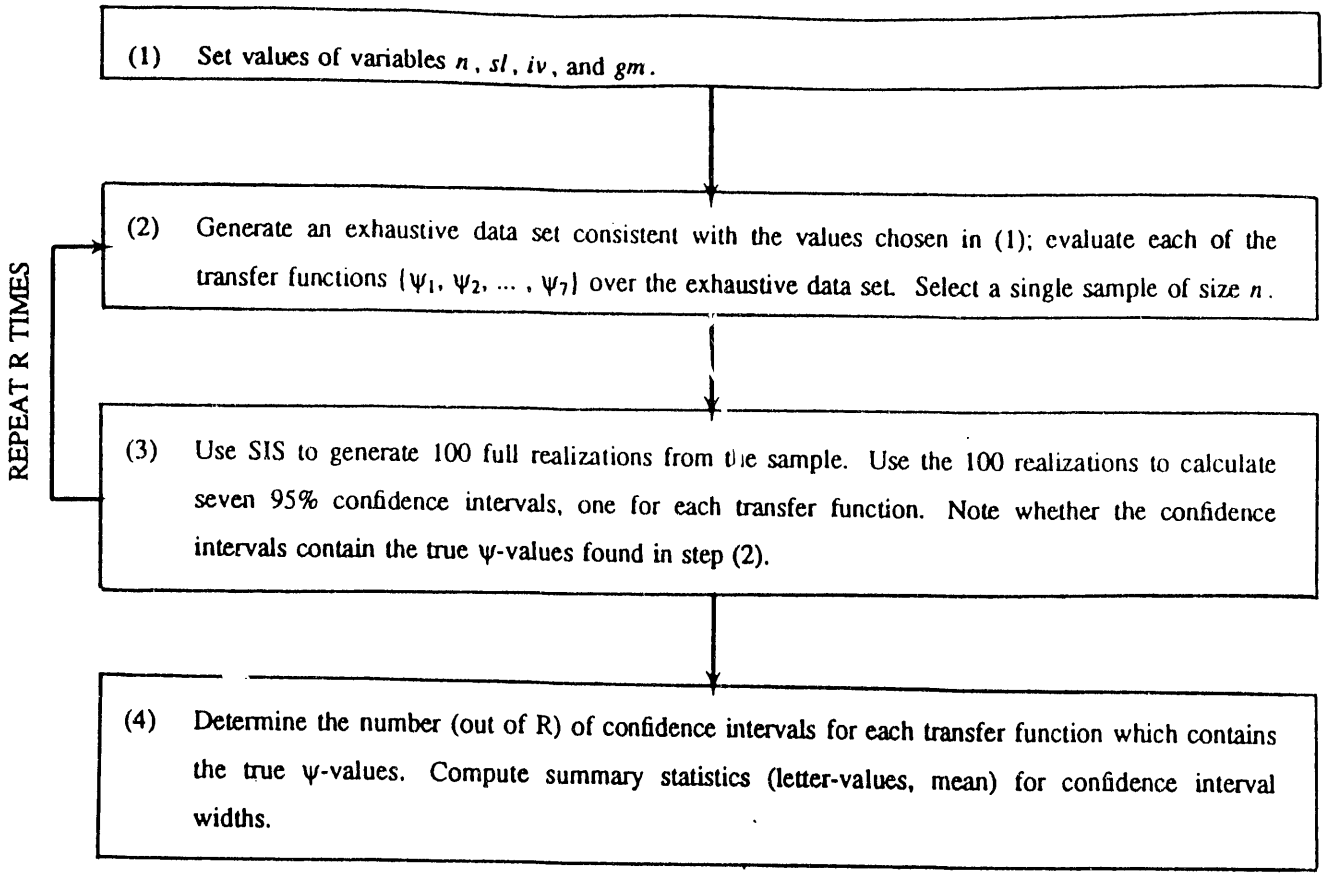


Figure 3-4 Outline of the procedures constituting a single run in the synthetic experiment. For Phase I, R=10, and for Phase II, R=50 repetitions. It follows that each run in Phase I requires 1,010 CIS simulations: 10 exhaustive data sets, plus 100 realizations for each of the exhaustive data sets. Each run in Phase II requires 5,050 simulations.

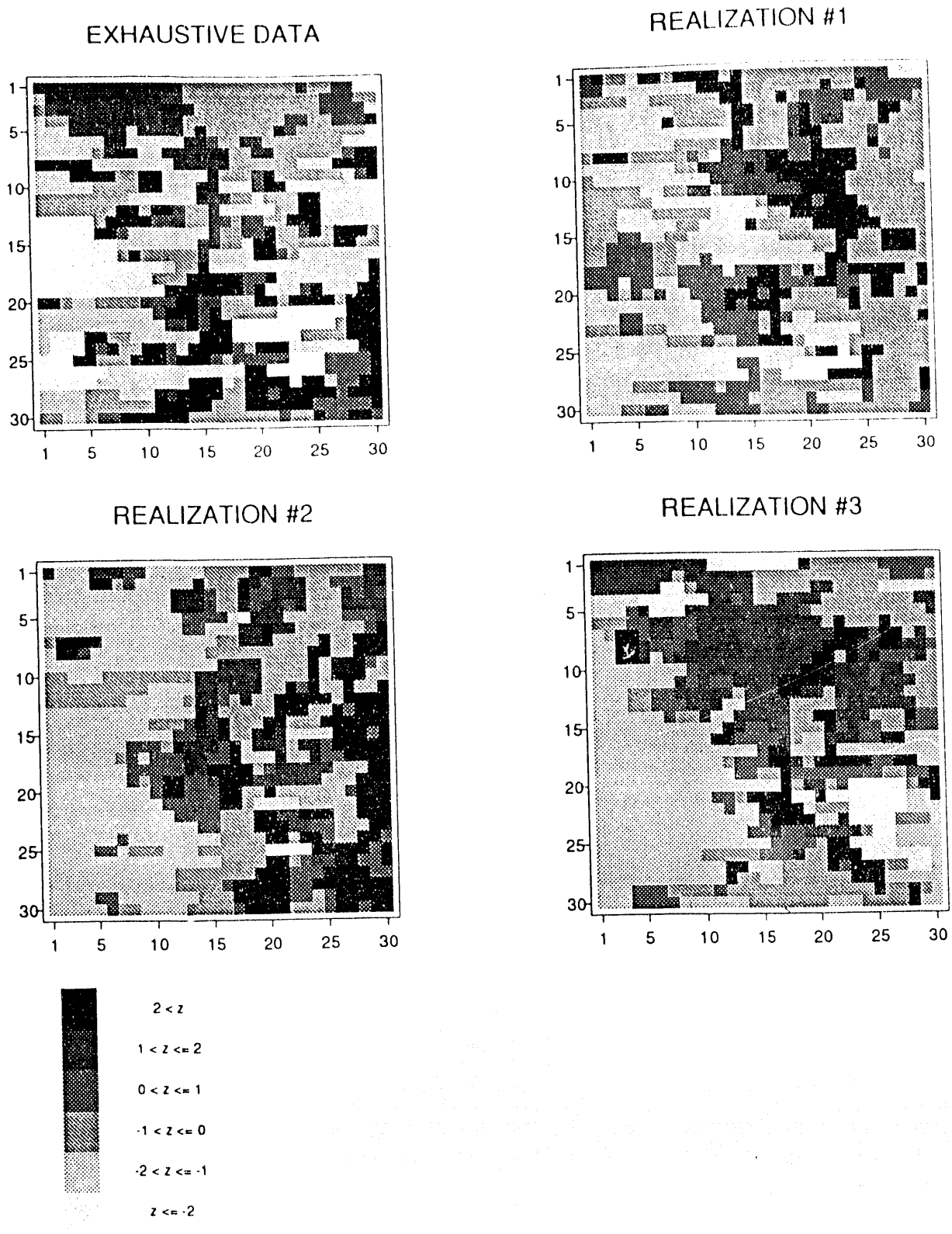


Figure 3-5 An exhaustive data set and three different SIS realizations, on a 30 × 30 grid. The realizations were generated using the theoretical indicator variograms of the exhaustive data set, and data at 30 randomly-chosen sample nodes. Note the variability seen among the different realizations.

If the assumptions used for SIS are valid, and the implementation is sensible, we would expect about 95%, or 9.5, of these confidence intervals to contain the true value of the transfer function over the appropriate data set. As discussed in Appendix E, when the number of inaccurate confidence intervals is equal to three or more (out of ten), the null hypothesis that the intervals are valid 95% confidence intervals may be rejected at a level less than 5%. The actual number of confidence intervals containing the true value, and the widths of the confidence intervals, were recorded for each transfer function on each run. Clearly, narrow confidence intervals (precise) that contain the true ψ-values (accurate) are evidence of a well-behaved simulation algorithm, while very wide confidence intervals, or those not containing the true value of ψ with the expected frequency, indicate problems with the SIS technique. Note that with only ten confidence intervals per run in the first phase, it is difficult to come to any definitive conclusions regarding accuracy: as discussed in Appendix E, the relevant statistical tests lack sufficient power. Thus, the issue of accuracy is largely deferred until the second phase of the experiment.

3.4 Experimental Phase II

Of the 20 factor combinations studied in Phase I, 10 were chosen for further experimentation in Phase II. These combinations are identified with a "y" in the third column of Table 3-2. In this phase, a run used 50 different exhaustive data sets, each on a 30×30 grid. As before, the exhaustive data sets were independently generated, each with the same theoretical indicator variograms. A sample was taken from each, and was used to generate 100 realizations. (That is, there were a total of 5,000 realizations generated: 100 realizations for each of 50 exhaustive data sets.) The large number of runs conducted in Phase II of the experiment allows us to determine, with more power, whether confidence intervals constructed from SIS realizations are accurate. When inaccuracies are found, we attempt to identify probable causes for the failure of SIS distributions.

3.5 Difficulties With Available SIS Computer Code

The simulation study was carried out using version 1.1 of the C language computer program isim3d.c (Gomez-Hernandez, Srivastava, and Serikiotou, 1989) to perform sequential indicator simulation. After the study had been in progress for several months, a bug was discovered in this version of the code. The bug involved the estimation of conditional probability distribution functions (PDFs) using simple or ordinary kriging: with version 1.1, it is possible to generate PDFs that are not monotone increasing. This is a clear violation of one of the most basic tenets of probability theory.

Later versions of the C program corrected the problem, but the time involved in re-running the entire simulation study using new software was prohibitive. Instead, a few important runs were repeated using version 2.21 of isim3d.c, in order to determine whether the accuracy and precision figures obtained using the newer version of the computer program were consistent with those seen for the earlier version. The results were quite similar for the two versions of the software. It does not appear that the bug in version 1.1 of isim3d.c has had a large impact on the simulation study.

Appendix A provides a complete description of the bug in version 1.1, the correction taken in version 2.21, and a comparison made between the results from the two programs.

4.0 RESULTS

Tables 4-1 through 4-7 summarize the results for each of the experimental runs on the seven transfer functions. In the sections that follow, these results are discussed and analyzed in detail.

4.1 Sample Size

4.1.1 Precision in Control Situations

Runs A, B, C, and D of the experiment all represent control situations. In each of these runs, SIS was used to generate the exhaustive data sets, and was used again, with the same theoretical indicator variograms, to generate the multiple realizations. The sampling was done without preferential clustering in any one region of the grid. The four runs differed from one another in sample size only: n was set equal to 5 for run A, 15 for run B, 30 for run C, and 45 for run D. Changing n led to substantial differences in accuracy and precision.

By examining Table 4-8, which groups the results from runs A - D, we can see that increasing the sample size increased the precision of confidence intervals generated using SIS. For the first transfer function, the median confidence interval width decreased by about 8%* as n increased from 5 to 15; by about 38% as n increased from 15 to 30; and by another 19% as n increased from 30 to 45. The sixth transfer function had corresponding changes of an 22% decrease, a 15% decrease, and a small 3% decrease. For ψ_7, the changes in median confidence interval widths were again all decreases, of 3%, 26%, and 24%, respectively. These types of findings, which are evident in the boxplots ** of Figure 4-1, are not surprising: we would expect that, as more information was given to the simulation program, the variability of the output

* All percentages reported are fractions of the larger of the two numbers being compared.

** Boxplots are described in Tukey (1977). For all boxplots presented in this report, the central box runs from the lower quartile of the data to the upper quartile, and the line within the box represents the median observation. The "whiskers" extend to the nearest value within a standard range of the quartiles. Observations lying more than a standard range away from the quartiles are plotted singly. [A standard range is equal to 1.5 × (*upper quartile - lower quartile*)].

TABLE 4-1

RESULTS FOR ψ_1, FOR PHASES I AND II OF THE SYNTHETIC EXPERIMENT

	PHASE I				PHASE II
Run #	lq[a]	med[b]	uq[c]	# outside[d]	# outside
A	0.36	0.37	0.44	1	2
B	0.32	0.34	0.36	0	0
C	0.16	0.21	0.25	1	0
D	0.11	0.17	0.19	1	3
E	0.21	0.25	0.25	0	
F	0.17	0.20	0.22	2	
G	0.19	0.21	0.21	0	
H	0.20	0.23	0.25	2	2
I	0.25	0.30	0.32	1	
J	0.18	0.22	0.27	0	3
K	0.21	0.23	0.25	1	
L	0.20	0.25	0.26	0	
M	0.22	0.24	0.28	2	10
N	0.18	0.22	0.28	3	15
P	0.14	0.17	0.20	4	20
Q	0.20	0.23	0.24	1	10
R	0.18	0.19	0.19	2	
S	0.27	0.28	0.30	0	
T	0.21	0.22	0.22	1	
U	0.13	0.18	0.19	3	

[a] lq = lower quartile of the 10 Phase I confidence interval widths.

[b] med = median of the 10 Phase I confidence interval widths.

[c] uq = upper quartile of the 10 Phase I confidence interval widths.

[d] # outside = number of inaccurate confidence intervals counted (out of 10 for Phase I, and out of 50 for Phase II).

TABLE 4-2

RESULTS FOR ψ_2, FOR PHASES I AND II OF THE SYNTHETIC EXPERIMENT

	PHASE I				PHASE II
Run #	lq[a]	med[b]	uq[c]	# outside[d]	# outside
A	0.42	0.47	0.51	0	2
B	0.32	0.34	0.36	0	0
C	0.27	0.30	0.30	1	2
D	0.20	0.23	0.25	0	6
E	0.28	0.29	0.32	0	
F	0.23	0.25	0.25	2	
G	0.23	0.25	0.27	0	
H	0.27	0.29	0.31	1	2
I	0.35	0.36	0.41	2	
J	0.23	0.26	0.29	2	3
K	0.29	0.32	0.36	0	
L	0.26	0.29	0.31	0	
M	0.28	0.29	0.30	1	6
N	0.23	0.26	0.28	2	8
P	0.18	0.22	0.24	5	14
Q	0.24	0.25	0.29	0	8
R	0.19	0.20	0.22	1	
S	0.30	0.34	0.38	1	
T	0.23	0.24	0.26	2	
U	0.17	0.19	0.22	1	

[a] lq = lower quartile of the 10 Phase I confidence interval widths.

[b] med = median of the 10 Phase I confidence interval widths.

[c] uq = upper quartile of the 10 Phase I confidence interval widths.

[d] # outside = number of inaccurate confidence intervals counted (out of 10 for Phase I, and out of 50 for Phase II).

TABLE 4-3

RESULTS FOR ψ_3, FOR PHASES I AND II OF THE SYNTHETIC EXPERIMENT

	PHASE I				PHASE II
Run #	lq[a]	med[b]	uq[c]	# outside[d]	# outside
A	0.44	0.49	0.51	2	3
B	0.35	0.39	0.42	0	6
C	0.28	0.31	0.32	3	2
D	0.21	0.22	0.25	2	11
E	0.28	0.31	0.31	4	
F	0.22	0.25	0.26	3	
G	0.23	0.26	0.28	1	
H	0.26	0.29	0.30	0	8
I	0.35	0.37	0.43	2	
J	0.25	0.28	0.31	2	5
K	0.29	0.30	0.31	0	
L	0.27	0.30	0.33	3	
M	0.22	0.26	0.31	2	10
N	0.24	0.28	0.29	0	14
P	0.21	0.23	0.24	3	16
Q	0.24	0.25	0.28	2	10
R	0.17	0.20	0.21	1	
S	0.34	0.36	0.38	0	
T	0.23	0.25	0.26	0	
U	0.18	0.20	0.20	1	

[a] lq = lower quartile of the 10 Phase I confidence interval widths.

[b] med = median of the 10 Phase I confidence interval widths.

[c] uq = upper quartile of the 10 Phase I confidence interval widths.

[d] # outside = number of inaccurate confidence intervals counted (out of 10 for Phase I, and out of 50 for Phase II).

TABLE 4-4

RESULTS FOR ψ_4, FOR PHASES I AND II OF THE SYNTHETIC EXPERIMENT

	PHASE I				PHASE II
Run #	lq[a]	med[b]	uq[c]	# outside[d]	# outside
A	0.39	0.41	0.44	0	2
B	0.29	0.34	0.38	0	3
C	0.27	0.29	0.29	1	4
D	0.18	0.19	0.22	3	10
E	0.2[illegible]	0.255	0.27	3	
F	0.18	0.21	0.25	2	
G	0.19	0.235	0.25	0	
H	0.22	0.245	0.26	0	4
I	0.31	0.31	0.37	1	
J	0.22	0.24	0.26	1	4
K	0.23	0.245	0.29	0	
L	0.22	0.25	0.27	1	
M	0.21	0.22	0.25	2	10
N	0.21	0.23	0.24	2	11
P	0.16	0.19	0.23	1	7
Q	0.23	0.25	0.25	0	4
R	0.14	0.17	0.20	2	
S	0.30	0.35	0.37	0	
T	0.21	0.24	0.25	3	
U	0.17	0.17	0.20	0	

[a] lq = lower quartile of the 10 Phase I confidence interval widths.

[b] med = median of the 10 Phase I confidence interval widths.

[c] uq = upper quartile of the 10 Phase I confidence interval widths.

[d] # outside = number of inaccurate confidence intervals counted (out of 10 for Phase I, and out of 50 for Phase II).

TABLE 4-5

RESULTS FOR ψ_5, FOR PHASES I AND II OF THE SYNTHETIC EXPERIMENT

	PHASE I				PHASE II
Run #	lq[a]	med[b]	uq[c]	# outside[d]	# outside
A	0.32	0.35	0.36	0	0
B	0.24	0.29	0.31	0	1
C	0.22	0.23	0.26	0	2
D	0.15	0.17	0.20	1	7
E	0.16	0.18	0.22	0	
F	0.11	0.14	0.19	0	
G	0.15	0.18	0.21	0	
H	0.14	0.19	0.22	0	2
I	0.26	0.28	0.31	1	
J	0.15	0.22	0.26	0	2
K	0.18	0.20	0.22	0	
L	0.16	0.20	0.26	1	
M	0.17	0.18	0.19	3	14
N	0.16	0.18	0.20	4	10
P	0.12	0.15	0.19	2	16
Q	0.19	0.20	0.23	1	8
R	0.12	0.15	0.17	2	
S	0.22	0.29	0.33	1	
T	0.17	0.19	0.22	0	
U	0.12	0.15	0.16	2	

[a] lq = lower quartile of the 10 Phase I confidence interval widths.

[b] med = median of the 10 Phase I confidence interval widths.

[c] uq = upper quartile of the 10 Phase I confidence interval widths.

[d] # outside = number of inaccurate confidence intervals counted (out of 10 for Phase I, and out of 50 for Phase II).

TABLE 4-6

RESULTS FOR ψ_6, FOR PHASES I AND II OF THE SYNTHETIC EXPERIMENT

	PHASE I				PHASE II
Run #	lq[a]	med[b]	uq[c]	# outside[d]	# outside
A	80	87	91	1	2
B	54	68	81	0	3
C	55	58	59	0	2
D	48	56	63	0	6
E	50	59	61	0	
F	50	52	58	1	
G	51	55	73	0	
H	42	49	70	1	2
I	75	78	84	1	
J	59	68	75	0	6
K	32	38	45	10	
L	49	69	82	1	
M	46	52	57	0	6
N	47	59	67	1	8
P	44	52	61	2	6
Q	69	79	85	4	13
R	47	61	66	2	
S	74	88	116	0	
T	65	70	80	4	
U	61	64	73	5	

[a] lq = lower quartile of the 10 Phase I confidence interval widths.

[b] med = median of the 10 Phase I confidence interval widths.

[c] uq = upper quartile of the 10 Phase I confidence interval widths.

[d] # outside = number of inaccurate confidence intervals counted (out of 10 for Phase I, and out of 50 for Phase II).

TABLE 4-7

RESULTS FOR ψ_7, FOR PHASES I AND II OF THE SYNTHETIC EXPERIMENT

	PHASE I				PHASE II
Run #	lq[a]	med[b]	uq[c]	# outside[d]	# outside
A	3.4	3.5	4.1	0	2
B	3.1	3.4	3.6	1	0
C	2.2	2.5	2.7	0	1
D	1.6	1.9	2.1	2	4
E	2.2	2.5	2.9	0	
F	2.0	2.1	2.4	2	
G	2.1	2.2	2.4	0	
H	2.2	2.4	2.8	0	2
I	2.6	3.0	3.6	2	
J	2.4	2.7	3.1	0	2
K	1.1	1.2	1.4	7	
L	3.1	3.4	3.7	4	
M	2.2	2.7	3.2	2	4
N	1.9	2.5	3.2	2	4
P	1.8	2.0	2.2	0	13
Q	3.0	3.6	5.0	8	37
R	2.4	2.9	3.9	3	
S	4.2	5.2	6.0	2	
T	3.0	3.7	4.6	8	
U	2.8	3.1	3.6	7	

[a] lq = lower quartile of the 10 Phase I confidence interval widths.

[b] med = median of the 10 Phase I confidence interval widths.

[c] uq = upper quartile of the 10 Phase I confidence interval widths.

[d] # outside = number of inaccurate confidence intervals counted (out of 10 for Phase I, and out of 50 for Phase II).

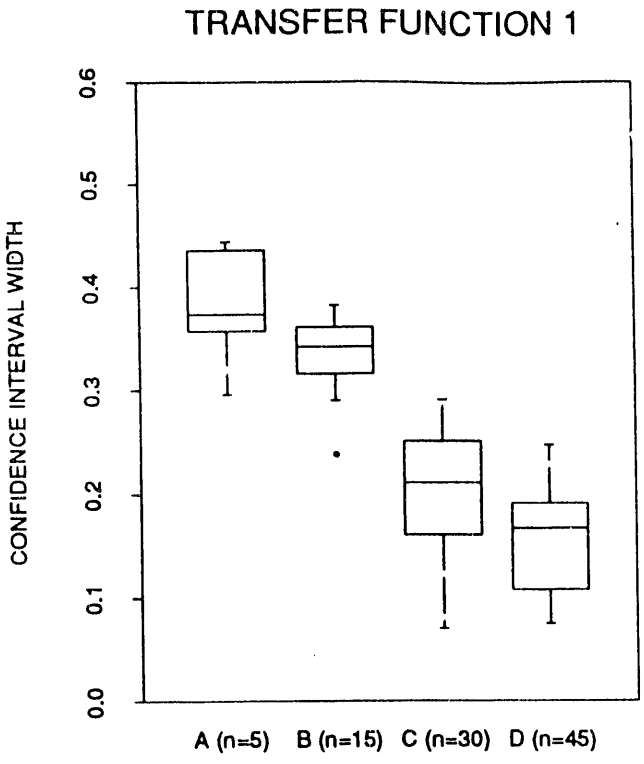
TABLE 4-8

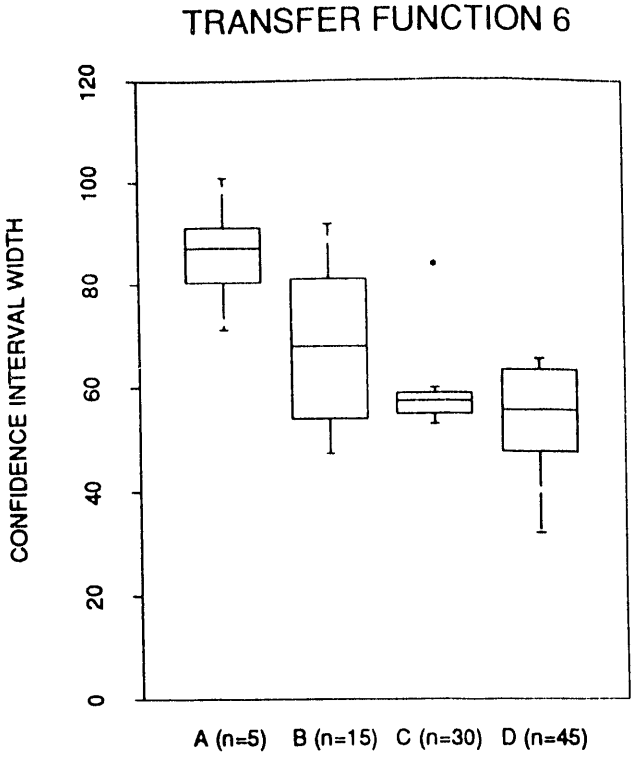
EFFECT OF SAMPLE SIZE ON ACCURACY AND PRECISION FOR THE CONTROL RUNS

Median Confidence Interval Widths - Phase I								
Run	n	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
A	5	.37	.47	.49	.41	.35	87	3.5
B	15	.34	.34	.39	.34	.29	68	3.4
C	30	.21	.30	.31	.29	.23	58	2.5
D	45	.17	.23	.22	.19	.17	56	1.9

Number Inaccurate Confidence Intervals - Phase II								
Run	n	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
A	5	2	2	3	2	0	2	2
B	15	0	0	6*	3	1	3	0
C	30	0	2	2	4	2	2	1
D	45	3	6*	11*	10*	7*	6*	4

* Runs for which the number of inaccurate confidence intervals is statistically significant at a level of less than 5%.





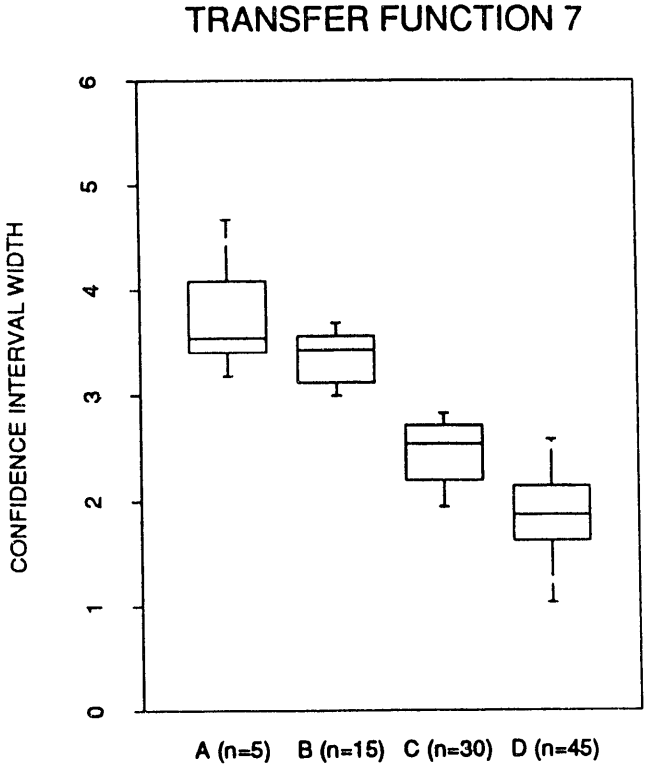


Figure 4-1 Boxplots showing the effect of sample size on precision, for the control runs, A (n=5), B (n=15), C (n=30), and D (n=45). Each boxplot represents the ten confidence intervals constructed during Phase I of the experiment. Results for the first, sixth, and seventh transfer functions are shown; plots for $\psi_2 - \psi_5$ exhibited the same trends. Increasing n led to improved precision, as demonstrated by the narrowing of the confidence intervals.

would decrease.

4.1.2 Accuracy in Control Situations

The findings regarding the effect of sample size on accuracy, for control situations, are somewhat counterintuitive. When n was as large as 45 (5% of the 900 total nodes), the accuracy of SIS predictions decreased significantly. For transfer functions ψ_2, ψ_3, ψ_4, ψ_5, and ψ_6, six or more confidence intervals in Phase II did not contain the true value. (As discussed in Appendix E, when six or more out of 50 intervals fail to contain the true value, the null hypothesis that the intervals are valid 95% confidence intervals for the true value can be rejected at a level less than 5%.) As n increased from 5 to 15 and from 15 to 30, the precision of the simulation distribution increased, as was seen in the decreasing confidence interval widths; these improvements in precision came without a decline in accuracy. However, when the sample size increased from 30 to 45, the resulting decrease in confidence interval widths was often too drastic, and accuracy was lost. Figure 4-2 shows a fairly typical case of an inaccurate confidence interval from run D: the true value of the transfer function lies just outside of the extremes of the simulated distribution.

4.1.3 Nonrepresentative Samples

One possible cause of confidence interval inaccuracy is a skewed, or nonrepresentative sample: the histogram of the sample might be quite different from that of the exhaustive data set from which the sample was drawn. Even when sampling is random, as it is for the control runs, substantial differences in the histograms may exist. These differences, in turn, could influence the histogram of the multiple realizations conditioned on the sample, and thus the values of other transfer functions evaluated over the realizations. For example, if an unusually large percentage of the data were contained in the left-most bin (corresponding to a small value of ψ_1), we would intuitively expect that ψ_6, the minimum-path finder, would also tend to have a small value. This relationship is illustrated in Figure 4-3, a plot of the value of ψ_6 versus that of ψ_1, for 100 realizations of a run D data set. Realizations that are extreme in ψ_6 tend to be extreme in ψ_1, and the two transfer functions are positively correlated.

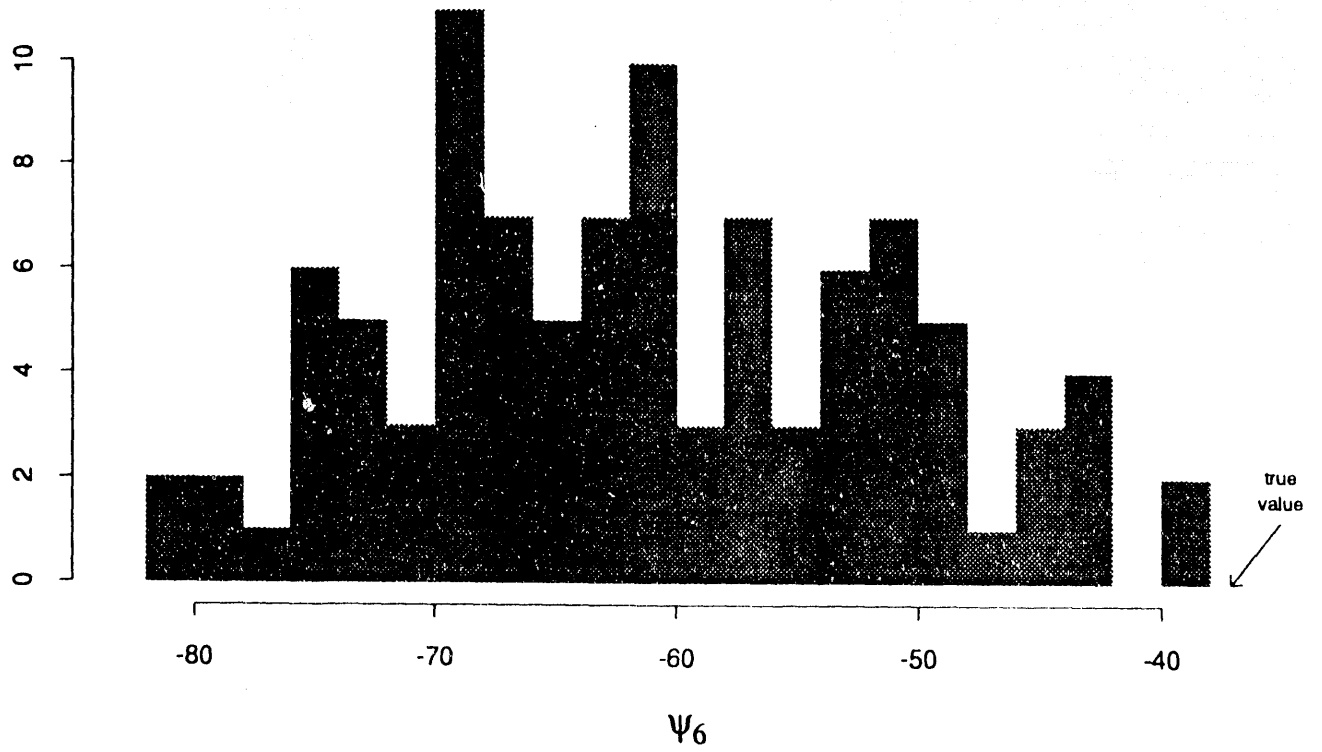


Figure 4-2 Histogram of simulated distribution of ψ_6, for D-3.3, one of the exhaustive data sets in run D. The shaded area represents the values of the transfer function seen in 100 independent realizations generated from the same sample of size n=45. The true value of ψ_6 for this data set was -37.142. The simulated distribution ranged from -81.453 to -39.044, while the 95% confidence interval ranged from -79.284 to -40.112. Note that the true value is very close to one of the extremes of the simulated distribution; this was generally the case with the inaccurate simulations.

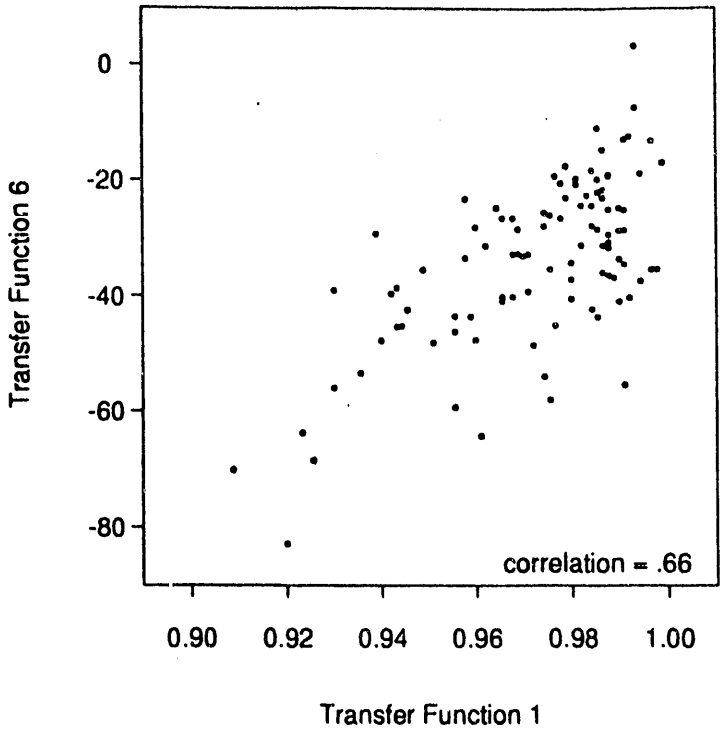


Figure 4-3 Plot of the relationship between ψ_1 and ψ_6, for the 100 realizations of data set D-1.2.

In order to determine whether nonrepresentative samples may have been a factor in the inaccuracy of run D confidence intervals for the higher-level transfer function ψ_6, we recorded the percentage of data exceeding the lowest threshold (-2) for exhaustive data sets, samples, and realizations, for each of the six run D cases giving inaccurate confidence intervals for ψ_6. (Note that this percentage is equal to the value of ψ_1, which can be evaluated on samples, as well as on exhaustive data sets.) The results are shown in Figure 4-4. Note that, in five of the six cases (all except for the case denoted D-4.4) the mean value of ψ_1 over 100 realizations was closer to the sample value than to the exhaustive value; this is not surprising, given that the realizations were conditional on the sample. For four of these five cases (denoted D-1.2, D-1.8, D-2.4, and D-3.3), the realizations of ψ_6 departed from the true value in the direction of the sample value of ψ_1. That is, for D-1.2, D-2.4, and D-3.3, the upper limit of the constructed confidence intervals was less than the true value of ψ_6. For D-1.8, the lower limit of the ψ_6 confidence interval exceeded the true value. It is thus plausible to suppose that the differences in the sample histogram and the exhaustive histogram influenced the minimum flow-paths seen in these conditional realizations.

Case D-4.5 was an interesting exception; the sample value of ψ_1 (.578) was less than the true value (.624); the mean value seen over the hundred conditional realizations (.532) was even smaller. However, the true value of ψ_6 (-125.24) was smaller than the lower limit of the confidence interval (-122.46) and even the minimum value seen over all of the realizations (-123.07). Figure 4-5 shows the exhaustive data set, and three realizations generated from 45 sample data points. On each image, the ψ_6 minimum flow path is shown with a dashed line. Note that the exhaustive data set and one of the realizations have flow paths that pass only through nodes with values in the lowest bin (less than or equal to -2). The fact that the exhaustive data set has a value lower than those of the realizations is simply due to bad luck: the values within the lower bin that were assigned by SIS to the nodes in the exhaustive data set happened to be smaller than those assigned to the nodes in the realizations. The SIS procedure, by default, uses a uniform within-bin distribution, and the exhaustive data set contained a large number of nodes along the flow path having values close to the specified minimum of -5. (The mean node value along the minimum path was equal to -125/30 = -4.17, while the mean of the uniform distribution on the interval between

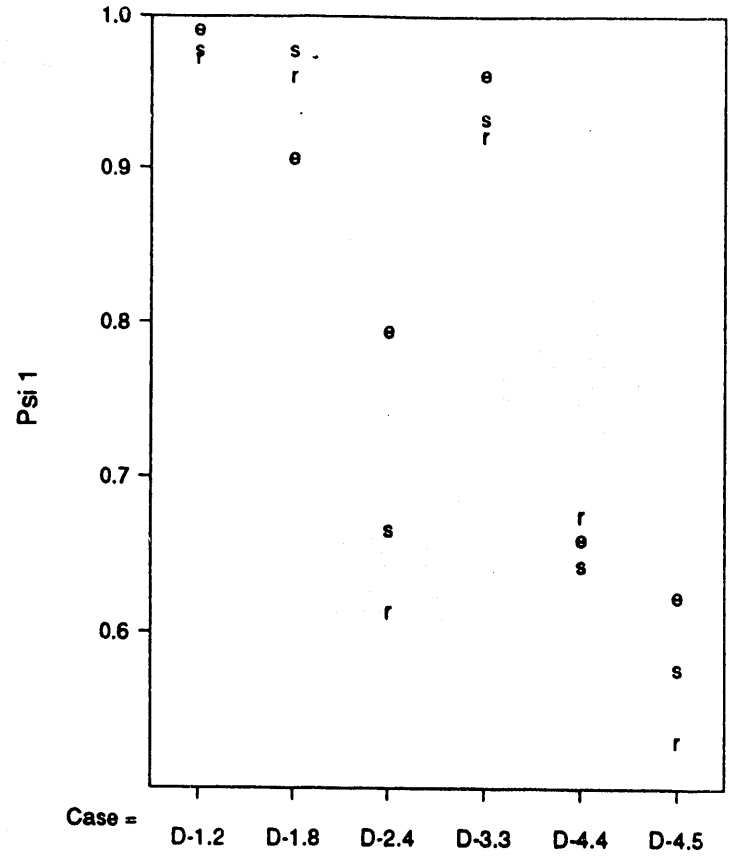


Figure 4-4 Illustration of the relationship between e, the exhaustive (or "true") value of ψ_1, s, the sample value, and r, the mean value seen over 100 SIS realizations, for six data sets giving inaccurate ψ_6 confidence intervals. (The theoretical value of ψ_1 is equal to .80.) Note that in five of the six cases, the mean value of ψ_1 over the realizations departs from the true value in the direction of the sample value. For cases D-1.2, D-1.8, D-2.4, and D-3.3, the mean value of ψ_6 over the realizations departs from the true value of ψ_6 in the direction of the sample value of ψ_1. This suggests that inaccuracies in SIS confidence intervals for ψ_6 may be due, in part, to differences in exhaustive and sample cumulative distribution functions.

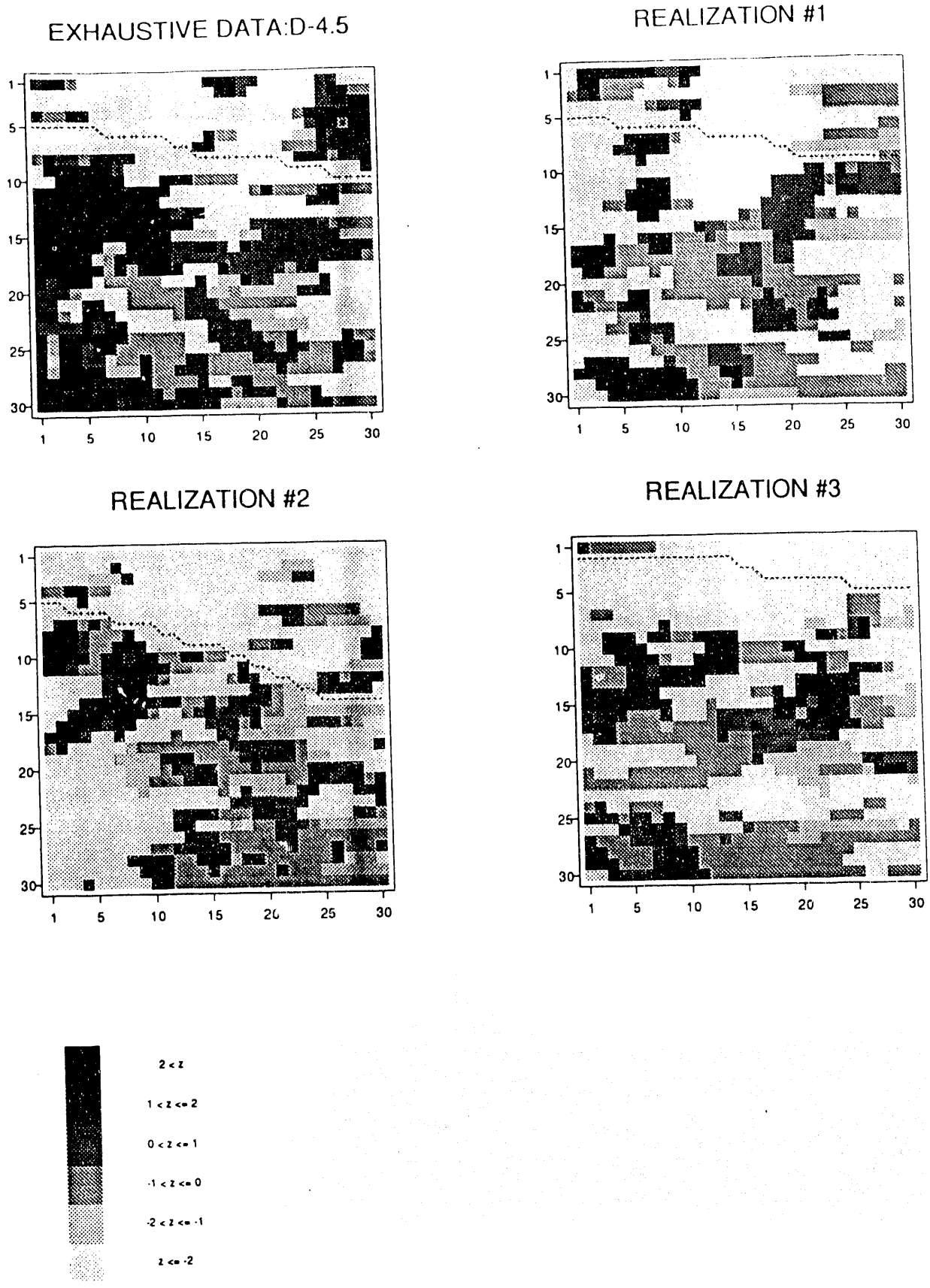


Figure 4-5 Exhaustive data set D-4.5, and three realizations that are conditional on 45 sample values. (The sample data are displayed in the upper right-hand panel of Figure 2-3.) The true value of ψ_6 is -125.24; the realizations #1, #2, and #3 ψ_6 values of -101.43, -108.88, and -116.50, respectively. The SIS confidence interval for ψ_6 was inaccurate in this case.

-5 and -2 is equal to -3.5.) Thus, the exhaustive data set itself seems to lie in a "tail" of the distribution of $Z(S)$, and most simulation techniques would be expected to have difficulty generating such an extreme pattern in only 100 realizations. The inaccuracy of the SIS-generated ψ_6 confidence interval for this particular exhaustive data set is not surprising. It is reasonable to expect that, if a larger number of realizations were carried out from the same sample data, more extreme values of ψ_6 would be found.

4.1.4 The Effect of Simple vs. Ordinary Kriging

In addition to possibly nonrepresentative samples, another cause for inaccurate confidence intervals for run D might be that the confidence intervals are simply too narrow. In conventional independent and identically distributed sampling, confidence interval widths decrease proportionally with $1/\sqrt{n}$. The SIS confidence interval widths seen for runs A through D decrease at a slower rate, because they are influenced by several factors: the sample size, the assumed variograms, and the parameters of the SIS code itself. Some of the inaccuracies observed in run D may be caused by confidence intervals that are overly narrow due to these code parameters.

In version 1.1 of the SIS code used for this study, isim3d.c, the determination of the appropriate bin for a new node is done by simple kriging or ordinary kriging, depending on the number of data points (sample points or nodes already simulated) that are in the neighborhood of the new node. Specifically, simple kriging is used if fewer than five data points are in the neighborhood, whose width is chosen by the user. If five or more data points are in the neighborhood, then ordinary kriging is used. Because ordinary kriging will fit a local mean in the region being simulated, the newly-simulated node will have a stronger tendency to take on a value similar to those of the neighboring points than would be the case if simple kriging were used. As a result, a smaller degree of variability is seen once ordinary kriging is invoked.

When the initial sample size is as large as $n = 45$, the nodes to be simulated are more likely to start out with at least five sample points in their neighborhood than would be the case with smaller initial samples. Thus, ordinary kriging, with its tendency to reduce variability, is invoked early in the simulation process. The resulting set of realizations may not exhibit enough variability to cover the true values of

the transfer functions. Later versions of the SIS code allow the user to reset the value at which the switch from simple kriging to ordinary kriging takes place. However, it may be difficult for the user to choose an appropriate value for a particular study. This issue is discussed in more depth, and is illustrated graphically, in Appendix A.

4.1.5 Non-Control Runs

All of the above discussion relates to the control runs. Several more comparisons may be made to determine if the effects of n on accuracy and precision are consistent over different experimental situations.

Runs I (n=15) and J (n=30) can be compared to examine the effect of changing sample size on simulations for which the indicator variograms used to generate the exhaustive data sets and the realizations differed in terms of the marginal distributions that they specified. Results for each transfer function are given in Tables 4-1 through 4-7 and Table 4-9. The decreasing confidence interval widths seen with increasing the sample size from 15 to 30 are consistent with the decreases seen in the control situation (runs B and C).

Runs S (n=15), T (n=30), and U (n=45) each used exhaustive data sets that were generated using the Choleski (non-SIS) method. Results are summarized in Table 4-10. None of the runs were included in Phase II, but some accuracy problems were severe enough to be detected during Phase I: run T had a statistically significant number of inaccurate confidence intervals for ψ_4, ψ_6, and ψ_7; and run U had a significant number for ψ_1, ψ_6, and ψ_7. These results follow the pattern of the number of inaccuracies increasing with the sample size suggested by run D. Other factors contributing to run T and U inaccuracies will be discussed in the section on the effect of different methods of generating exhaustive data sets. Figure 4-6 shows boxplots of the confidence interval widths for runs S, T, and U. The general trend of precision increasing with sample size is evident.

Runs E and F can be compared to study the effect of changing n on the precision of simulation distributions generated from clustered samples. The runs were not included in Phase II, and it is not possible to come to any conclusions regarding accuracy. Both of the runs used the correct theoretical indicator variogram information to generate the multiple realizations, and both used sampling plans which clustered the

TABLE 4-9

EFFECT OF SAMPLE SIZE ON PRECISION FOR RUNS I AND J

Median Confidence Interval Widths - Phase I								
Run	n	Ψ_1	Ψ_2	Ψ_3	Ψ_4	Ψ_5	Ψ_6	Ψ_7
I	15	.30	.36	.37	.31	.28	78	3.0
J	30	.22	.26	.28	.24	.22	68	2.7

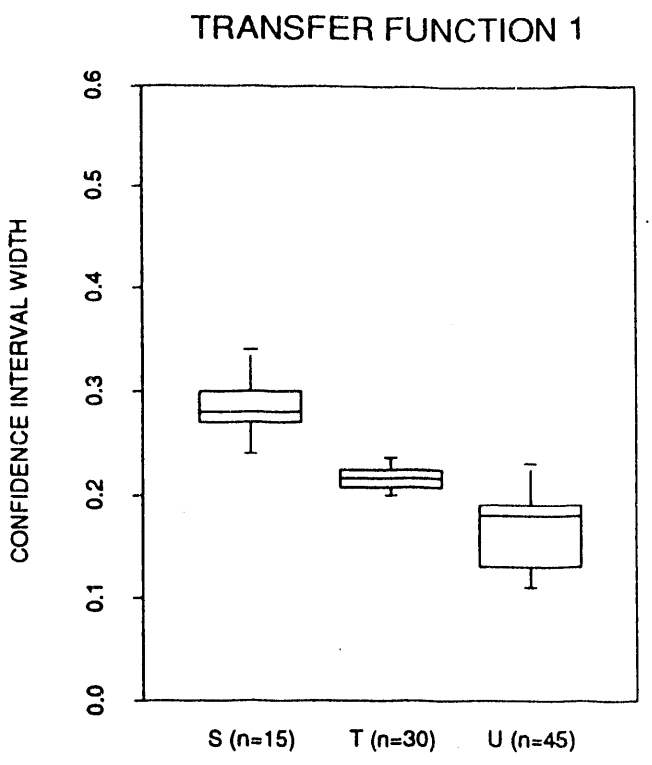
TABLE 4-10

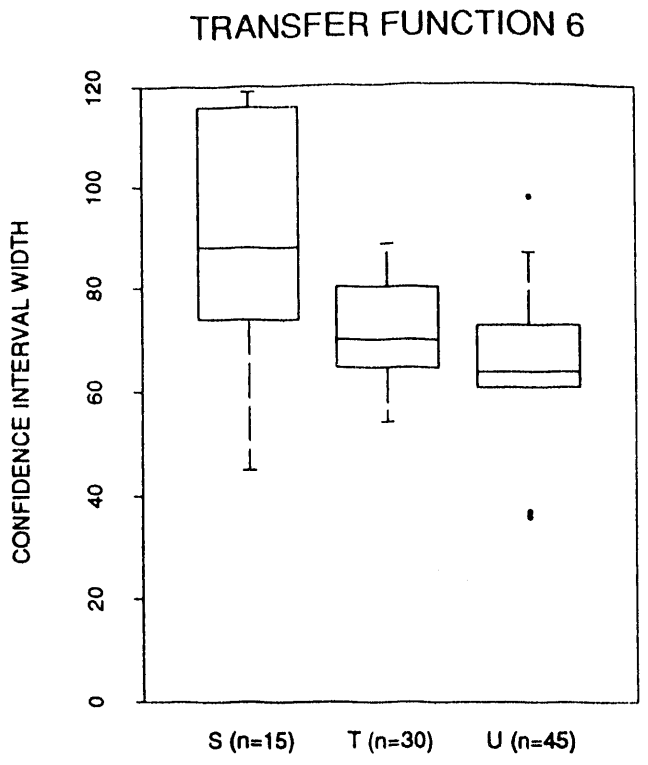
EFFECT OF SAMPLE SIZE ON ACCURACY AND PRECISION FOR THREE RUNS THAT USED EXHAUSTIVE DATA SETS GENERATED BY THE CHOLESKI METHOD

Median Confidence Interval Widths - Phase I								
Run	n	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
S	15	.28	.34	.36	.35	.29	88	5.2
T	30	.22	.24	.25	.24	.19	70	3.7
U	45	.18	.19	.20	.17	.15	64	3.1

Number Inaccurate Confidence Intervals - Phase I								
Run	n	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
S	15	0	1	0	0	1	0	2
T	30	1	2	0	3*	0	4*	8*
U	45	3*	1	1	0	2	5*	7*

* Runs for which the number of inaccurate confidence intervals is statistically significant at a level of less than 5%.





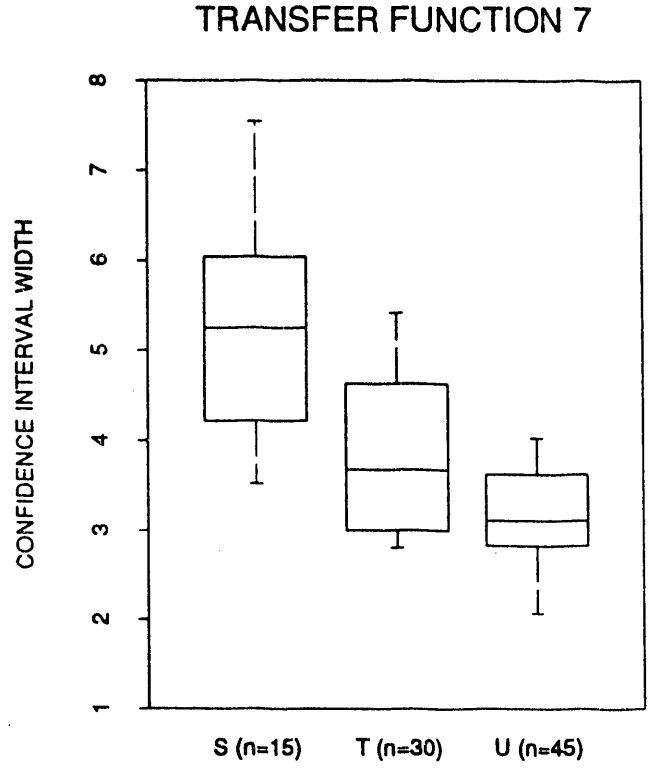


Figure 4-6 Boxplots showing the increase in precision seen with large sample sizes for runs S (n=15), T (n=30), and U (n=45). Each of these runs used exhaustive data sets that were generated by the Choleski method.

samples preferentially in the first and third quadrants. The sample size was 30 for run E, and 45 for run F. The changes in confidence interval widths seen with changing n for these runs are consistent with the results of control simulations.

Runs N (n=30) and P (n=45) were included in both phases of the experiment. Increasing the sample size from 30 to 45 provided improved precision, consistent with trends seen previously. Accuracy problems were severe at both n=30 and n=45; this situation will be discussed in the section on input variogram information.

Runs Q and R both used exhaustive data sets that were generated using the Choleski method, and, for each of these runs, the indicator variogram information input to the SIS program was computed from the exhaustive data sets themselves. Run R (n=45) gave confidence intervals that were more precise than those given by run Q (n=30). The median confidence interval widths for ψ_1, ψ_6, and ψ_7 in run Q were respectively, 17%, 23%, and 19% larger than those for run R. Run Q was included in Phase II of the experiment and the accuracy problems that were encountered will be discussed in the section on different generation methods.

4.2 Sample Location

We now discuss the effect of clustered data samples on the accuracy and precision of simulated distributions. Clustered samples were taken in four different runs, all generated using SIS (gm=1) and all of which used the correct theoretical indicator variograms to generate realizations (iv=1). The clustering was carried out at four different levels: 2:1 clustering in the first and third quadrants of the grid space (runs E and F); 1.33:1 clustering in the first quadrant (run G); and 1.67:1 clustering in the first quadrant (run H). These runs may be compared to the appropriate control runs to study the effect of clustering.

Run E used n=30 samples, and may be compared to run C, which had the same sample size but used a nonclustered sampling scheme. The data of Tables 4-1 through 4-7 and Table 4-11, and the boxplots of Figure 4-7 show that the mild clustering used for run E did not have a large impact on the precision of simulated transfer function distributions, particularly for the two spatially-sensitive transfer functions.

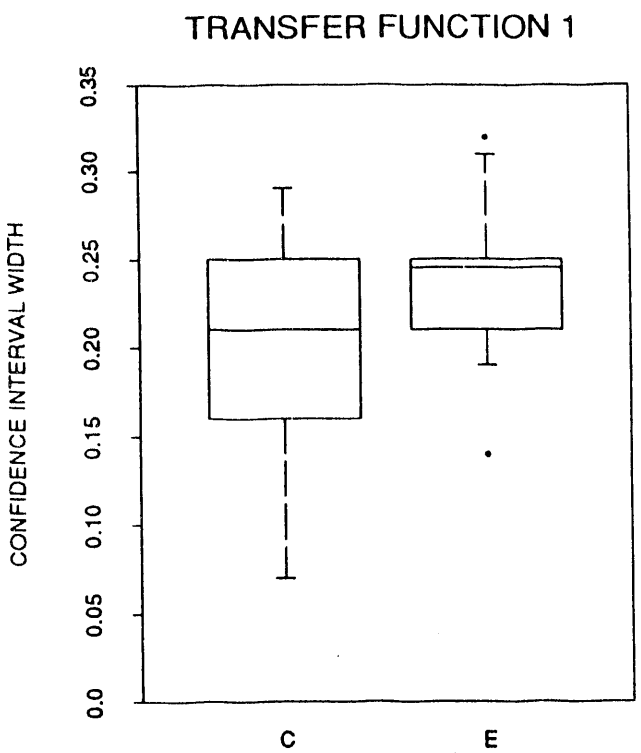
TABLE 4-11

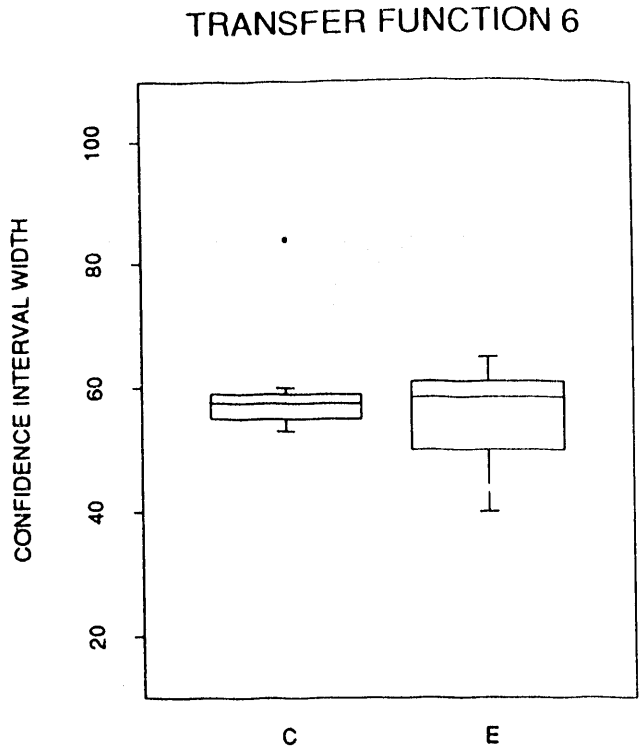
INFLUENCE OF CLUSTERING ON PRECISION FOR SIX RUNS USING EXHAUSTIVE DATA SETS GENERATED BY SIS

Median Confidence Interval Widths - Phase I									
Run	n	Clustering	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
C	30	none	.21	.30	.31	.29	.23	58	2.5
E	30	2:1 (2 quads)	.25	.29	.31	.26	.18	59	2.5
D	45	none	.17	.23	.22	.19	.17	56	1.9
F	45	2:1 (2 quads)	.20	.25	.25	.21	.14	52	2.1
G	45	1.33:1 (1 quad)	.21	.25	.26	.24	.18	55	2.2
H	45	2.67:1 (1 quad)	.23	.29	.29	.25	.19	49	2.4

Number Inaccurate Confidence Intervals - Phase II									
Run	n	Clustering	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
D	45	none	3	6*	11*	10*	7*	6*	4
H	45	2.67:1 (1 quad)	2	2	8*	4	2	2	2

* Runs for which the number of inaccurate confidence intervals is statisticaly significant at a level less than 5%.





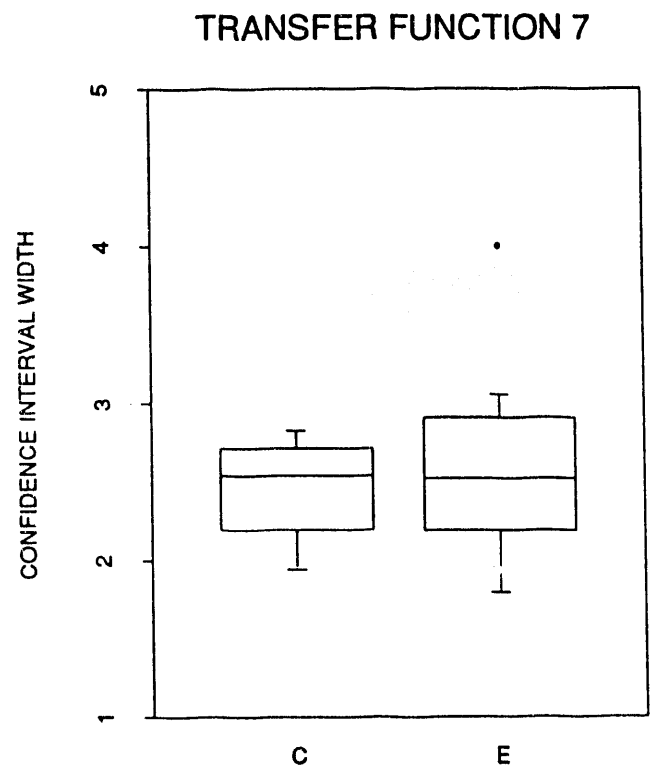


Figure 4-7 Boxplots of Phase I confidence interval widths for runs C and E. The mild clustering in the samples of run E does not seem to have an impact on the precision of SIS confidence intervals.

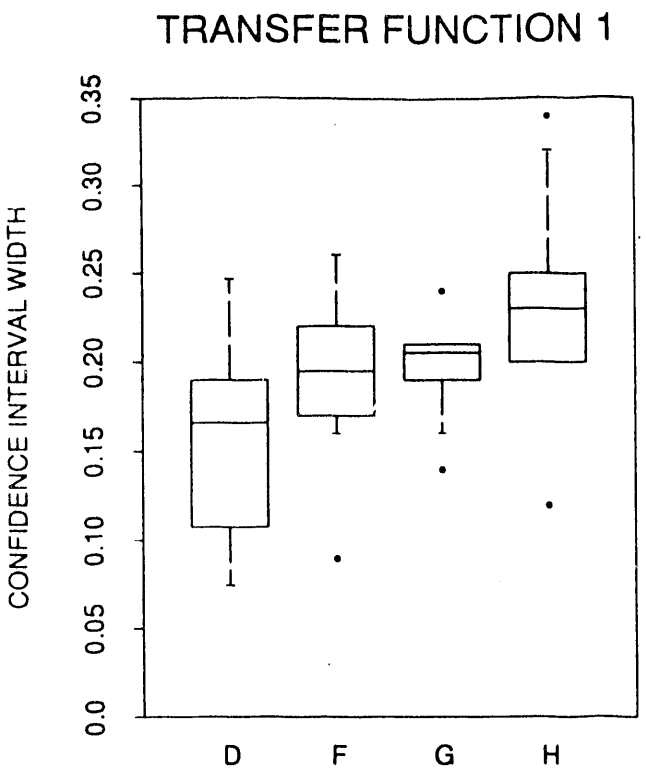
Runs F, G, and H each used n=45 samples, and the degree of clustering among samples increased progressively through these runs. The boxplots of Figure 4-8 show that confidence interval widths increase, and precision decreases, from the nonclustered run D through the clustered runs. It is interesting to note that this effect was evident even for run F, which had the same mild clustering pattern as that of run E, discussed above. It appears that clustering of samples had a greater impact for large sample sizes than for small sample sizes. This is known in statistics as an interraction effect: the factor *sl* has a different effect on precision for different levels of the factor n. This can probably be explained by observing that for fairly large sample sizes, clustered sampling provides a lot of redundant data, which does little to constrain the values of the transfer functions. For small samples, the redundancy is less pronounced, and a loss of precision due to clustering was not evident. Note that the precision of run F was better than that of run C (Tables 4-1 through 4-7, Table 4-11), showing that more information was provided by 45 mildly-clustered samples than by 30 uniformly-distributed samples. In fact, even for the 45 severely-clustered samples of run H, more precision was obtained than for the 30 samples of run C.

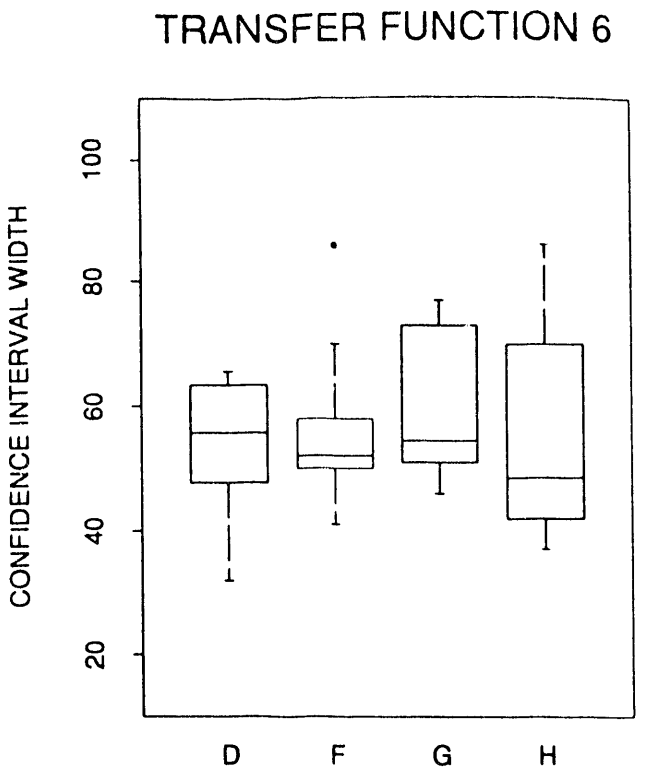
Recall that accuracy problems were encountered for five of the seven transfer functions in run D. Run H was included in Phase II, in order to determine whether this pattern would persist for clustered sampling. Tables 4-1 through 4-7 (last column) show that accuracy remained a problem only for a single transfer function, ψ_3. For the other transfer functions, including the spatially-sensitive ψ_6 and ψ_7, it appears that the loss of accuracy was avoided. This result is consistent with some of the observations made earlier: by sacrificing some precision, clustered sampling plans were able to avoid inaccuracy in simulated distributions.

4.3 Input Variogram Information

The results below demonstrate the various effects that may follow from different types and sources of input variogram information.

Runs C, J, K, L, M, and N each had n=30 samples taken in a nonclustered fashion. For these runs, the exhaustive data sets were generated in an identical manner, as described in Appendix B. The runs differed only in terms of the indicator variogram information input to the SIS code for generating realizations. Results for





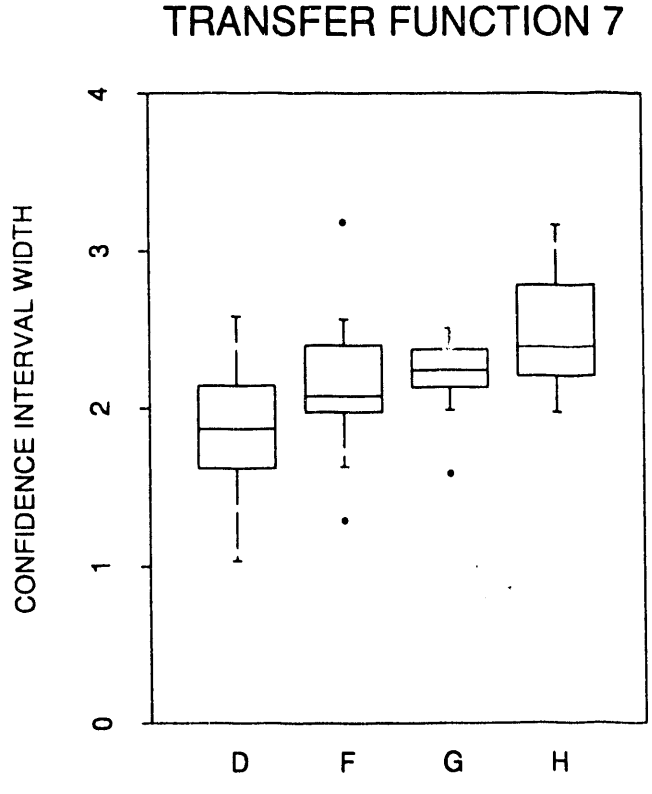
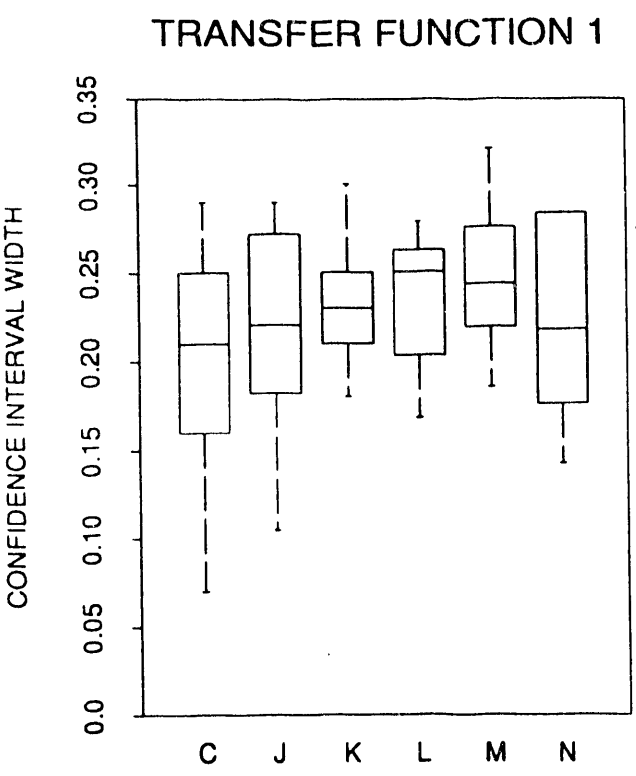


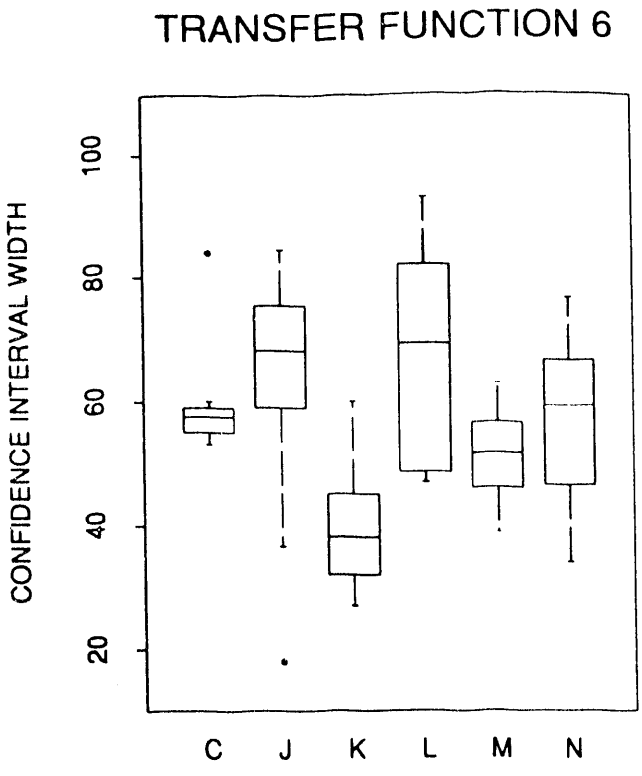
Figure 4-8 Boxplots of Phase I confidence interval widths for several runs with n=45. The degree of clustering in the samples increases steadily from run D (no clustering) to run H (strong clustering.) Precision decreases (confidence interval widths increase) noticeably with stronger clustering for ψ_1 and ψ_7. This trend is not seen for ψ_6.

the latter five runs will be compared to those for run C, the control run. Figure 4-9 is a series of boxplots illustrating precision differences for these runs on transfer functions ψ_1, ψ_6, and ψ_7.

4.3.1 Indicator Class Proportions

In run J, the indicator variogram information used to generate realizations differed from that used to generate exhaustive data sets in proportions only. As stated in Appendix B, when generating the exhaustive data sets, the cumulative distribution function (CDF) at five indicator thresholds, (-2, -1, 0, 1, and 2), was set at (0.20, 0.35, 0.50, 0.65, and 0.80), respectively. For generating the realizations, the CDF at the five indicator thresholds was set to (0.25, 0.40, 0.50, 0.60, and 0.75). Thus, the tails of the distribution were heavier for the realizations than they were for the exhaustive data sets. This change in the marginal distribution of the data values did not lead to substantial changes in either the accuracy or the precision of SIS confidence intervals for $\psi_1 - \psi_5$, the transfer functions related to the histogram of a data set. However, the sixth transfer function gave inaccurate confidence intervals six out of fifty times in Phase II (see Table 4-12). One explanation for this inaccuracy is that the higher expected number of values lying in the lower bin leads to smaller values of ψ_6 in the realizations than in the exhaustive data set. This factor alone did not cause all of the inaccuracies, since the lower limit of the SIS confidence intervals exceeded the true value in two of the six cases. However, for all six of these data sets, the deviations in ψ_6 were consistent with those in the sample value of ψ_1. When the sample had a smaller proportion of values in the lowest bin than the exhaustive data set, the true value of ψ_6 was less than the lower limit of the SIS confidence intervals; and when the sample had a larger proportion of values in the lowest bin, the true value of ψ_6 exceeded the upper limit of the confidence intervals. This suggests that nonrepresentative sampling may again have played a role in the inaccuracies. Note that the precision of run J confidence intervals for ψ_6 and ψ_7 also decreased relative to the precision seen in run C. This was probably caused by the larger tails in the CDF specified by the input variogram information.





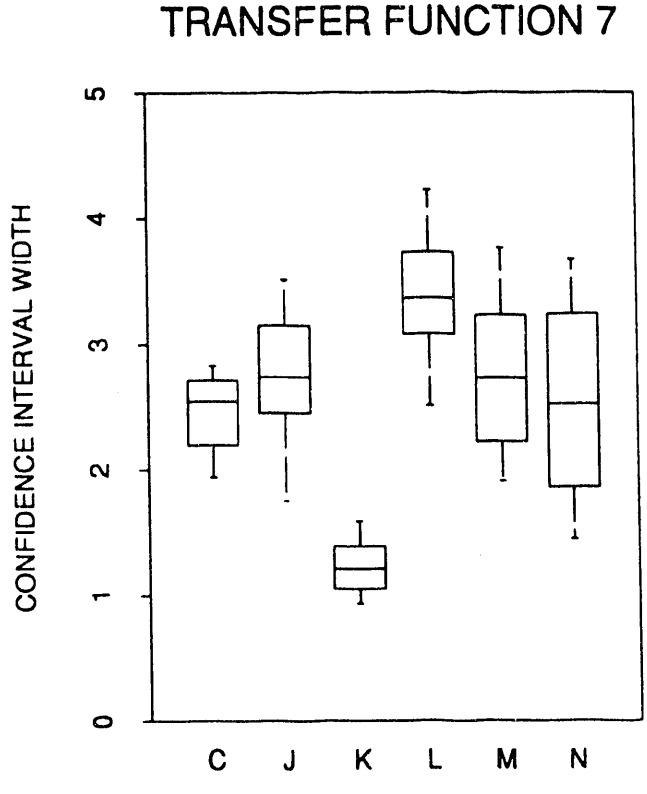


Figure 4-9 Boxplots illustrating precision changes as a function of the input indicator variograms, for six runs all having a sample size of 30.

TABLE 4-12

INFLUENCE OF THE INPUT VARIOGRAM INFORMATION, AND THE METHOD OF GENERATING EXHAUSTIVE DATA SETS, ON ACCURACY AND PRECISION

Median Confidence Interval Widths - Phase I									
Run	n	iv, gm	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
C	30	iv=1, gm=1	.21	.30	.31	.29	.23	58	2.5
J	30	iv=2a, gm=1	.22	.26	.28	.24	.22	68	2.7
M	30	iv=3, gm=1	.24	.29	.26	.22	.18	52	2.7
N	30	iv=4, gm=1	.22	.26	.28	.23	.18	59	2.5
Q	30	iv=3, gm=2	.23	.25	.25	.25	.20	79	3.6
D	45	iv=1, gm=1	.17	.23	.22	.19	.17	56	1.9
P	45	iv=4, gm=1	.17	.22	.23	.19	.15	52	2.0

Number Inaccurate Confidence Intervals - Phase II									
Run	n	iv, gm	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
C	30	iv=1, gm=1	0	2	2	4	2	2	1
J	30	iv=2a, gm=1	3	3	5	4	2	6*	2
M	30	iv=3, gm=1	10*	6*	10*	10*	14*	6*	4
N	30	iv=4, gm=1	15*	8*	14*	11*	10*	8*	4
Q	30	iv=3, gm=2	10*	8*	10*	4	8*	13*	37*
D	45	iv=1, gm=1	3	6*	11*	10*	7*	6*	4
P	45	iv=4, gm=1	20*	14*	16*	7*	16*	6*	13*

* Runs for which the number of inaccurate confidence intervals is statistically significant at a level less than 5%.

4.3.2 Extreme Values

Run K used indicator variogram information for the exhaustive data sets and the realizations that differed only in the specified minimum and maximum values. The exhaustive data sets had a minimum value of -5 and a maximum of 5; in the realizations, the minimum and maximum were -3 and 3, respectively. Thus, the end bins used for realizations were restricted to the inner one-third of the bins used for the exhaustive data sets. It is logical to expect that the confidence intervals seen for $\psi_1 - \psi_5$ would not be affected by this change, but that the confidence intervals for ψ_6 and ψ_7 would increase in precision (due to the smaller range of possible values) and decrease in accuracy. As is seen in Figure 4-9 and Tables 4-1 through 4-7, the data were consistent with this expectation. All ten run K confidence intervals for ψ_6 and seven of the ten for ψ_7 were inaccurate, and the confidence interval widths decreased by 34% for ψ_6 and by 52% for ψ_7. These results demonstrate that misspecification of the minimum and maximum values for SIS can have a drastic effect when transfer functions are sensitive to tail values. It should be noted, however, that the parameters chosen for run K were quite extreme, reducing the tail area by 67%. It may be unlikely that such a gross misestimation of the minimum and maximum values of a distribution would occur in an actual field study, where external information about the range of a variable may be available.

4.3.3 Indicator Class Proportions and Extreme Values

In run L, the indicator variograms used to generate the realizations differed from those used to generate the exhaustive data sets both in the specified cumulative distribution functions, and in the minimum and maximum values. Realizations were generated using the CDFs used in realizations for run J, with the minimum and maximum values changed from -5 and 5, to -6 and 6. Thus, the tails for the realizations were both longer and heavier than those for the exhaustive data sets. As Tables 4-1 through 4-7 show, the accuracy and precision of confidence intervals for $\psi_1 - \psi_5$ were not substantially different in runs L and C. However, the median confidence interval width seen for ψ_6 and ψ_7 increased by 16% and 26%, respectively, from the control run to run K. In addition, four of the ten confidence intervals for ψ_7 were inaccurate; this is

a statistically significant result at a level less than 1%. The loss of accuracy and the changes in precision were evidently caused by the simulation of larger (in absolute value) numbers in both of the end bins. This led to increased values for ψ_7, which is particularly sensitive to the magnitude of simulated values.

4.3.4 Indicator Thresholds from Exhaustive Data

In run M, the percentiles of the exhaustive data set were used as thresholds for sequential indicator simulation, as outlined in Appendix B. Results from Phase I showed relatively small effects on precision for all seven transfer functions: over ψ_1 through ψ_5, there was an average 10% decrease in median confidence interval width from run C to run M; ψ_6 had a 10% decrease; and ψ_7 had 7% increase. Table 4-12 shows that, in Phase II, all transfer functions, with the exception of ψ_7, gave a statistically significant number of inaccurate confidence intervals. This was unexpected, since SIS actually had more information about the exhaustive data sets in run M than in run C.

The six run M data sets giving inaccurate confidence intervals for ψ_6 were examined individually. Five of the six gave inaccurate confidence intervals for ψ_1, and for each of these five, the inaccuracies in ψ_6 were consistent with those in ψ_1 (i.e., the exhaustive values of ψ_1 and ψ_6 both exceeded the upper limit of their respective confidence intervals). This suggests that random bad luck, in the form of nonrepresentative sampling, may have played a role in the inaccuracies, but further investigation showed that this factor was not solely responsible. All six of the exhaustive data sets, with the same samples and the same random numbers (seeds), were used to generate 100 more SIS realizations. The new realizations used the same sequence of random numbers as the original realizations, but had the correct theoretical indicator thresholds input to SIS. That is, the parameters of the control run, C, were used. None of the six new confidence intervals were inaccurate. In addition, the two run C data sets giving inaccurate ψ_6 confidence intervals were also re-run, again using the original sample and random numbers, but with the parameters of run M (i.e., the indicator thresholds were based on percentiles of the exhaustive data sets). Both of the newly-generated confidence intervals were inaccurate. This strongly suggests that the accuracy problems seen in run M can be attributed to the use of exhaustive (rather than theoretical)

indicator variogram information, and are not simply due to nonrepresentative samples.

A possible explanation for these results is that, when the same variogram information is used to generate both an exhaustive data set and a collection of realizations, the realizations may be seen as being drawn from the same random functions distribution from which the exhaustive data set was drawn. Thus, there is a good chance that the exhaustive data set will be covered by the collection of realizations. When the realizations are generated using variogram information modeled from the exhaustive data set itself, additional information about the exhaustive data set is available, but this may be counter-balanced by the fact that the realizations are no longer being drawn from the same distribution as the exhaustive data set. This observation is consistent with a number of results seen throughout the study, and will be considered in more detail in the discussion section.

4.3.5 Indicator Thresholds From Related Data

Run N was similar to run M in that the indicator thresholds used for generating realizations were different from those used in generating the exhaustive data sets. Here, the percentiles of a related data set, generated using the same input parameters as those used to generate the exhaustive data set, were used for input to the realizations. As with run M, precision differences with the control run were small. Accuracy problems persisted: transfer functions ψ_1 through ψ_6 all gave a statistically significant number of inaccurate confidence intervals, ranging from 8 out of 50 inaccurate for ψ_2 and ψ_6, to 15 out of 50 inaccurate for ψ_1. Given the results of run M, these inaccuracies should not be surprising: if the use of indicator variogram information based on the correct exhaustive data set leads to accuracy problems, we would expect that the use of indicator variogram information based on a different exhaustive data set would also have difficulties.

Like run N, run P used indicator variogram information modeled after a related exhaustive data set; the two runs differ in that a larger sample, of size 45, was used for run P. Thus, run D is the appropriate control run for comparison with run P. Precision differences between runs D and P were minimal. Run D had accuracy problems for five of the seven transfer functions, and these were generally worsened in run P, where all seven transfer functions gave a significant number of inaccurate confidence

intervals. The most extreme result was seen for ψ_1, where twenty out of fifty Phase II confidence intervals failed to contain the true value of the function. The combination of a relatively large sample size, and the use of indicator variograms based on a related data set, caused the SIS technique to provide output that is useless in practical terms. The implications of these results will be discussed further shortly.

From the results of this section, it is clear that the source of the indicator variograms used to generate SIS realizations can have profound effects on the accuracy and precision of the generated confidence intervals. This is cause for concern since, in practice, a user of SIS will not have the luxury, as here, of knowing the true variograms; and even with truth, the results for $n = 45$ suggest a problem.

4.4 Method of Generating Exhaustive Data Sets

Most of the runs in the synthetic experiment used exhaustive data sets that were themselves generated by the SIS computer code. Because actual geologic regions of interest are not created in this manner, it is important to test the simulation technique on synthetic data sets that were generated using different algorithms. In order to meet this need, runs Q, R, S, T, and U used exhaustive data sets that were generated using a single z-variogram (the Choleski method) without the presence of different indicator thresholds (see section 3.1.4 for details). The realizations for run Q were generated using indicator variogram thresholds based on the exhaustive data sets, and the realizations for the other three runs used thresholds based on the statistics of a related data set.

Inaccuracy was a problem in all of these runs, except for run S, which had a small sample size and relatively imprecise confidence intervals. Run Q was included in Phase II, and had a significant number of inaccuracies for all transfer functions except for ψ_4 (see Table 4-12). Runs T and U had a significant inaccuracy rate in Phase I for three transfer functions each, including, for both runs, the spatially-sensitive functions ψ_6 and ψ_7. These problems are quite disturbing, because good information was input to the SIS program. For example, the five indicator thresholds used for simulation of realizations were the appropriate (for the variogram sill) percentiles of the exhaustive or related data sets, and the range specified for the indicator variograms was the same as the range specified for the z-variogram used to

generate the exhaustive data sets. (As discussed in Appendix D, the indicator variogram models specified provide an excellent fit to the indicator variograms obtained in practice from data sets generated using the Choleski method.) In a practical study with actual field data, we would not expect to have access to this type of knowledge. Even with a lot of information, the sequential indicator simulation technique has produced a large number of inaccurate confidence intervals, and it is important to try to understand these failures.

The extreme difficulty that SIS encountered in accurately constraining ψ_7 for these runs (37 out of 50 confidence intervals were inaccurate for run Q) was not unexpected. This transfer function is closely related to the z-covariance, and is not sensitive to the type of differing extents of correlation for which SIS is designed. A simulation technique based solely on the z-covariance would be more appropriate than SIS for constraining such functions. SIS was not intended for use with functions similar to ψ_7, and the results demonstrate that it is not useful for this type of function.

The inaccuracies seen for the sixth transfer function are less numerous than those seen for ψ_7, yet they are more disturbing. In run Q, 13 out of 50 (26%) of the Phase II confidence intervals for ψ_6 were inaccurate. The problem reflected by these results cannot be explained simply by skewed samples, since the inaccuracy rate for run C, with the same sample size, was only 4%. It seems likely that a more fundamental difficulty has been encountered: the distribution from which SIS has sampled is simply not representative of the distribution from which the exhaustive data sets were drawn. Inaccurate confidence intervals for ψ_1 through ψ_5 demonstrate that the two distributions produce output realizations with different histograms, and inaccurate confidence intervals for ψ_6 indicate that the differences can affect higher-order properties.

Figures 4-10 and 4-11 show the histograms of an exhaustive data set generated by the Choleski method, and of several SIS realizations generated using a sample from this exhaustive data set. By examining the differences in these histograms, we can better understand the problems that lead to inaccurate confidence intervals. In Figure 4-10, the bins used in constructing the histograms have endpoints corresponding to the extrema and quantiles of the exhaustive data set. Specifically, the extrema and the 0.20, 0.35, 0.50, 0.65, and 0.80 quantiles of the exhaustive data set are used as endpoints. Thus, in the upper left-hand panel, the four middle bins each contain 15% of

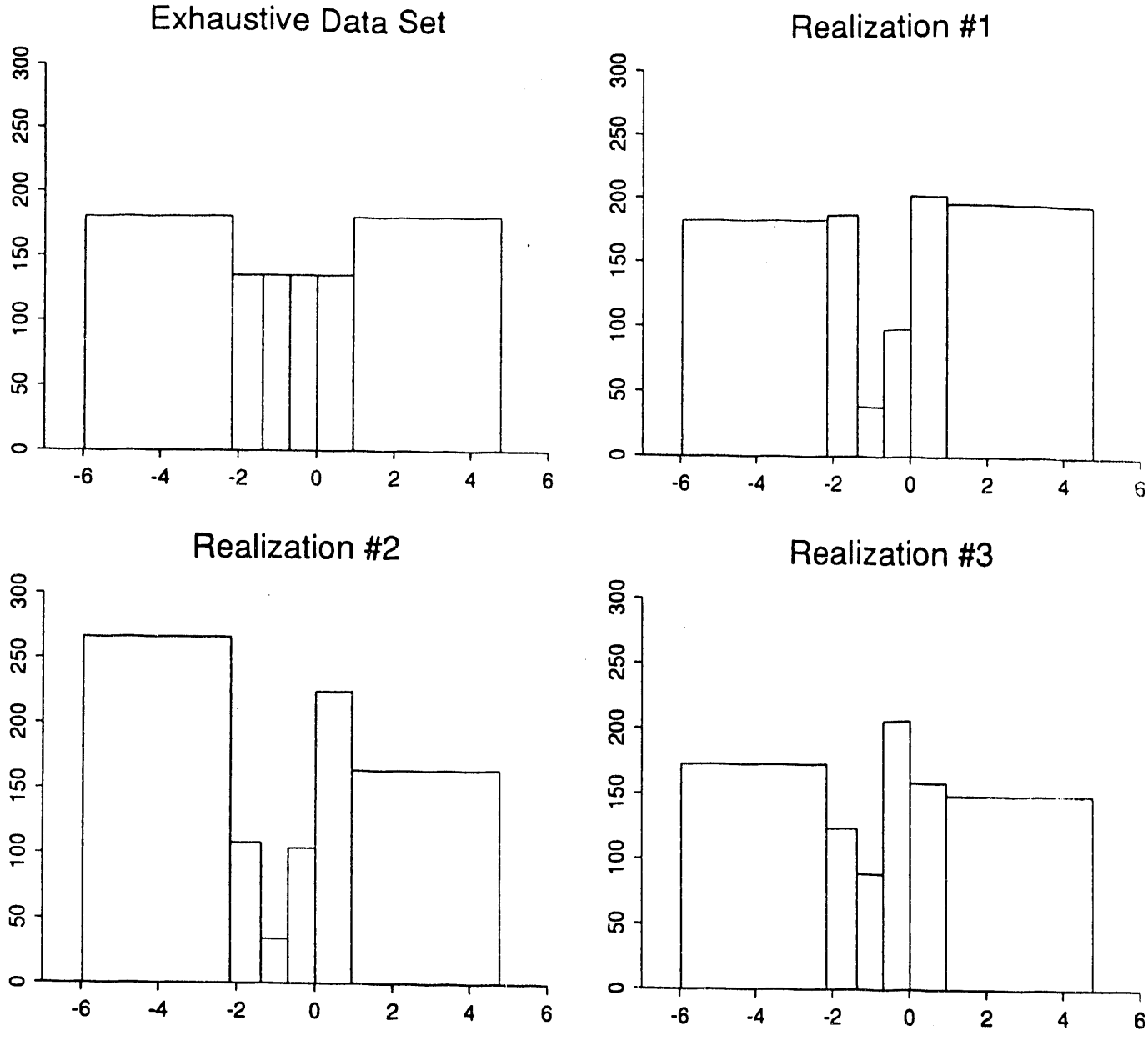


Figure 4-10 Histograms of an exhaustive data set, and three different realizations. The exhaustive data set was generated using the Choleski method, and the realizations were generated by SIS, using a sample of size 30 from the exhaustive data set, in the manner of run Q. The endpoints of the histogram bins are equal to the quantiles of the exhaustive data set, as input to SIS. At this scale, the histograms of the realizations closely resemble that of the exhaustive data set.

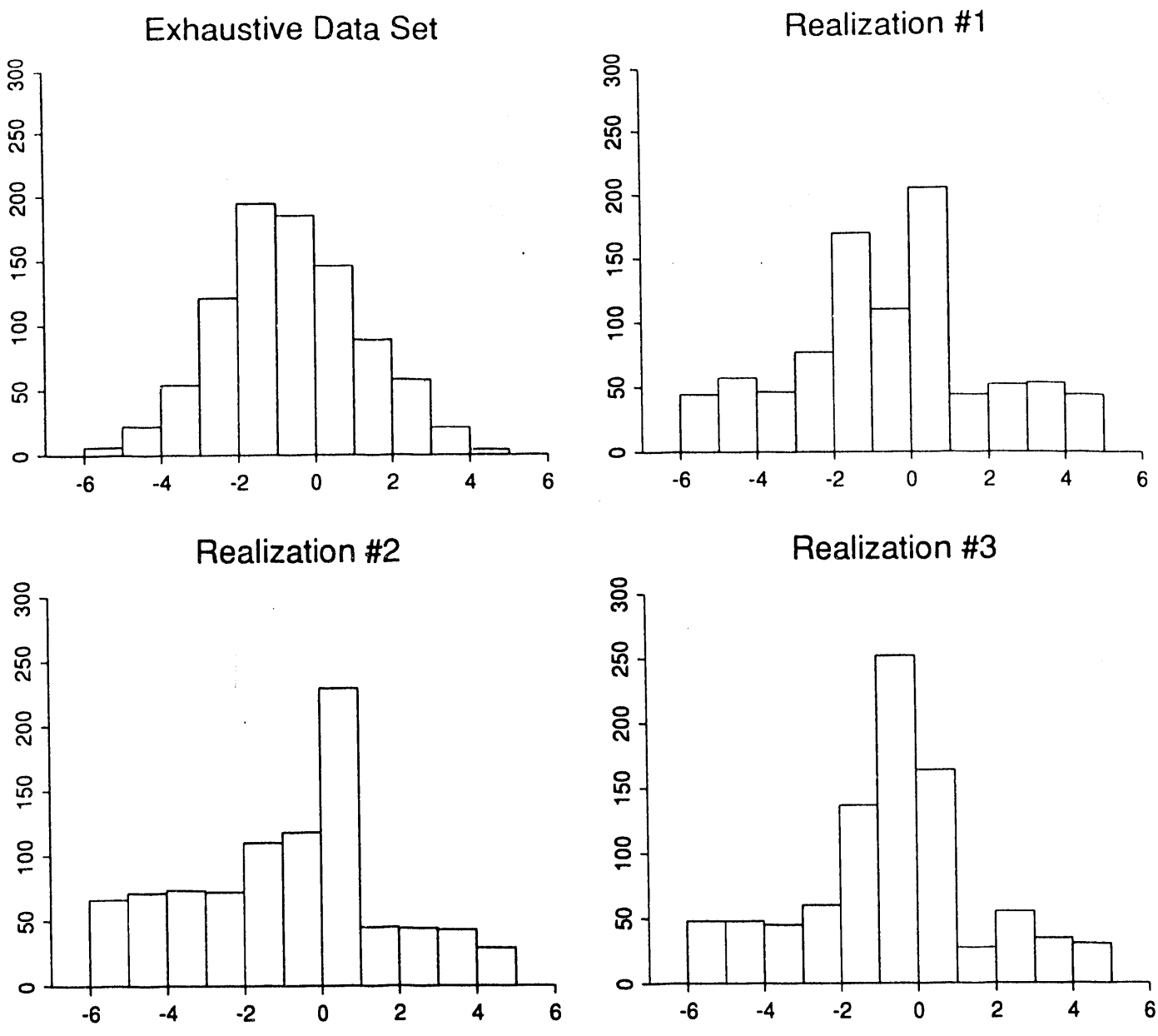


Figure 4-11 A second set of histograms for the same four data sets used in Figure 4-10. Here, the histograms use a finer binning scale, which allows us to see the differences in the way the two simulation techniques (Choleski and SIS) model the distribution of values within a particular indicator bin. Due to the heavier tails seen for SIS-generated distributions, values of ψ_6 and ψ_7 tend to be more extreme for SIS realizations than for the exhaustive data sets, and inaccurate confidence intervals result.

the values in the exhaustive data set, and the two end bins each contain 20%. These quantiles are the same ones provided as indicator thresholds to SIS in generating the realizations. We can see from the other three panels that, at this scale, the characteristics of the exhaustive data set have been re-captured reasonably well in the SIS realizations.

Figure 4-11 shows the histograms of the same four data sets, here plotted at a finer scale than in Figure 4-10. It is evident that the exhaustive data set has a distribution that is thinner in the tails than those of the three realizations. This difference is due to the differing nature of the Choleski and SIS simulation procedures. The SIS method, with the parameters chosen here, models a uniform distribution within indicator thresholds bins. Conversely, the Choleksi method gives final values that are a weighted sum of independent random variables, and thus have a histogram that is roughly Gaussian in shape.

It is likely that this difference in histogram shapes is responsible for many of the inaccuracies seen for runs Q, R, S, T, and U. For example, ten inaccurate confidence intervals constructed for ψ_6 in run Q were examined. In every case, the true value of ψ_6 (a negative number) exceeded the upper limit of the confidence interval. The heavy tails of the realizations resulted in values of ψ_6 that were all smaller than the true value. This situation again points out the importance of accurately modeling the tails of a distribution to be simulated, a task that can be extremely difficult when data are sparse.

4.5 Indicator Variograms of SIS Realizations

One of the appealing features of sequential indicator simulation is that it is based on an indicator kriging estimator that "honors all the indicator covariance model values" (Journel, 1988). In attempting to understand the behavior of the SIS method, it is important to clarify the meaning of this statement. Specifically, what "covariance model values" are referred to, and how closely are they reproduced?

There are several different sets of indicator variograms, corresponding to covariance models, associated with the use of sequential indicator simulation. First, there is the set of what we will call *input variograms,* which are specified by the user in advance of simulation. There are also *sample variograms,* which are usually modeled

from available sample data, and perhaps from additional geological data and/or insight. The *exhaustive variograms,* are calculated from the true, but generally unknown, exhaustive data set from which the sample is taken; and, finally, the *theoretical variograms* describe the spatial continuity seen over the full random functions distribution of which the exhaustive data set is a single element. Note that these different types of indicator variograms are interrelated. For example, the exhaustive variograms represent the mean(s) of the distribution of sample variograms from a particular exhaustive data set; and for large samples, the sample variograms will resemble the exhaustive variograms. Similarly, the theoretical variograms represent the mean(s) of the exhaustive variograms over the random functions distribution. In practice, the input variograms would usually be based on the sample variograms; in this study, input variograms based on the exhaustive and theoretical variograms are used. So - which covariance structures are reproduced in the realizations?

The answer to this question turns out to depend strongly on the size of the sample. Figure 4-12(a) shows the indicator variograms (at a threshold of -2, the theoretical 20^{th} percentile) of five different 30 × 30 SIS realizations, all generated with the same input = theoretical indicator variograms, and constrained by only two sample points. The results are extremely variable, and none of the realizations has a variogram that resembles the input variogram, but the *mean* of the five is quite close to the input function. This is encouraging: we would not want to restrict the simulator to producing only output data sets that exactly match the specified variogram models, yet we would like the realizations, on average, to conform to these models.

Next, an exhaustive data set was generated, with the same input indicator variograms used to generate the realizations in the previous step. A random sample of 45 points was taken from this data set. Five realizations were generated using the same input = theoretical variograms, and constrained at the 45 sample points. Figure 4-12(b) shows the indicator variograms of the exhaustive data set, the five realizations, and the mean of the realizations. The random seeds used for the realizations were the same as those used for part (a) of the figure, so that the difference in the two sets of variograms is due solely to the sample data. In this step, the variability seen among the different realizations has decreased, and the variograms no longer seem to be centered around the input variogram. Rather, the mean variogram over the realizations lies close to the exhaustive variogram, which was not known to the simulation

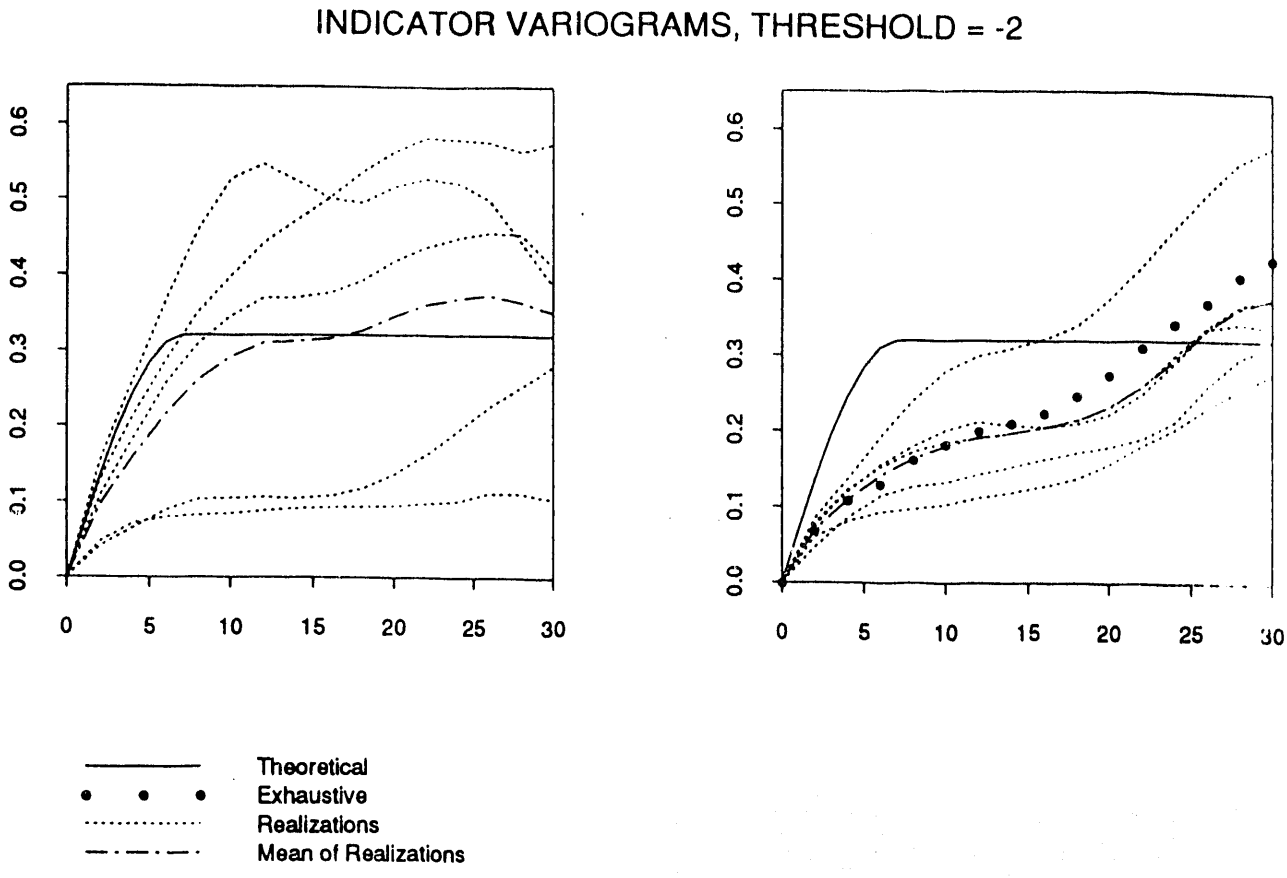


Figure 4-12 Indicator variograms for different SIS realizations. In the left-hand panel, five realizations were independently generated using the theoretical indicator variogram shown by the solid line as input. Each realization was constrained at only two sample points. None of the individual realizations has an indicator variogram resembling the theoretical function, yet the mean over the five (dashed line) shows good agreement. In the right-hand panel, an initial exhaustive data set was generated in the manner that all realizations in the previous step were generated. From this data set, 45 points were sampled. The five dashed lines are indicator variograms from realizations obtained by re-running SIS, using the same theoretical indicator variogram as input, now with the 45 sample points given. (The two sets of realizations used the same five random seeds.) Note that, in the right-hand panel, the mean of the five realization variograms lies close to the variogram (solid dots) of the exhaustive data set from which the samples were taken.

program. Presumably, this was caused by the relatively large sample taken from the exhaustive data set. This is another encouraging result, for it implies that as more information about the exhaustive data is made available, the input variogram becomes less important, and the realizations are less variable and take on the character of the exhaustive data set.

The potential for problems arises when a sample is large enough to influence the variograms of the realizations, yet has properties that differ substantially from those of the underlying exhaustive data. For example, a user could be aware that a particular sample is skewed (e.g., contains a disproportionate number of low values), and might attempt to correct for this by using input variograms that specify an appropriate data histogram. But the characteristics of the sample data could still force the realizations away from the desired properties. Figure 4-13 illustrates the issue. This type of difficulty can never be avoided: when data and a model are both used, and are in any way inconsistent, one must be favored over the other. With relatively large samples, SIS tends to favor data. In some instances in this study (for example in run D), this characteristic has resulted in inaccurate confidence intervals. Note, however, that the models used in this run were *good,* i.e., they were the theoretical variograms used to generate the exhaustive data set. In actual practice, we would never have access to such good models, so that a preference for data over models seems desirable, and should not be viewed as a weakness of SIS. This preference should be understood as an important property of the simulator, which can have either positive or negative consequences, depending on the representativeness of the sample and on the quality of the models.

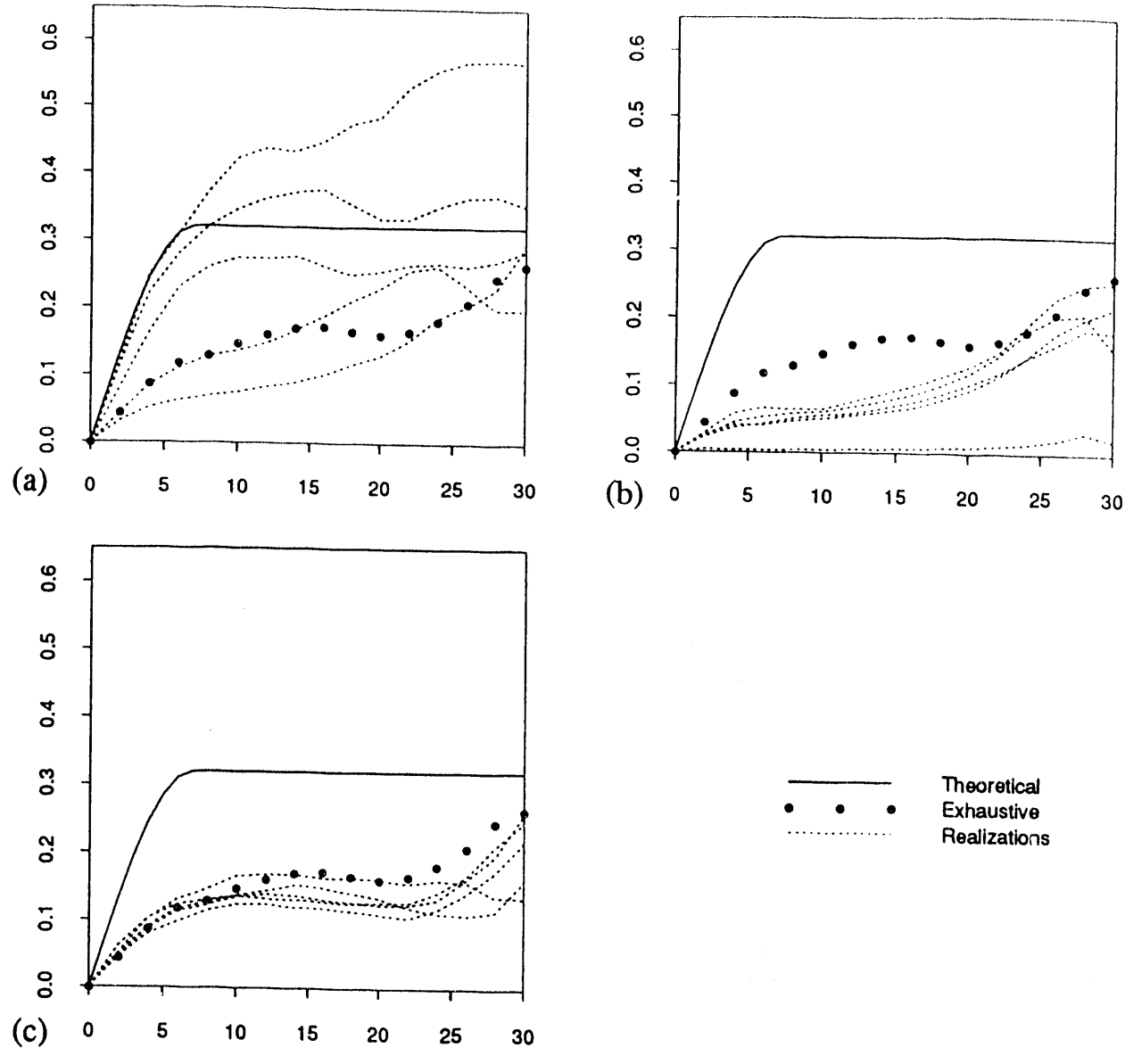


Figure 4-13 Indicator variograms of different SIS realizations, showing the effect that non-representative sampling can have on output variograms. Exhaustive data set D-1.8 was generated, using the theoretical indicator variogram shown by the solid line; the solid dots represent the variogram of this data set. Next, samples drawn from the exhaustive data set and the theoretical variogram were used to generate five different realizations. (a) First, a sample size of five was used. The five resulting realizations are shown by dotted lines. Note that most of the realization variograms lie between the theoretical and exhaustive functions. (b) An additional 40 sample points were provided as input to the realizations, bringing the total sample size to 45. The indicator variograms of the realizations are now quite similar to one another, but they are unlike either the theoretical or exhaustive variograms. A 95% confidence interval for ψ_6, constructed from 100 realizations using this sample, is inaccurate. (c) When another 105 sample points are added, 16.7% of the data points are known to the simulator, and the variograms of the realizations are all close to the exhaustive variogram. (The same five random seeds were used to generate the realizations for each sample size.)

5.0 DISCUSSION AND CONCLUSIONS

The results of this study have illustrated a number of problems that can occur when sequential indicator simulation is used to characterize uncertainty. A few general patterns that persist over the various experimental runs may help to identify situations that lead to difficulties with the simulated response distributions.

Several of the runs giving inaccurate Phase II confidence intervals for various transfer functions were related in that a considerable amount of information about the true, exhaustive data set was provided as input to the SIS program generating the realizations. Specifically: in run D, fairly large samples (5% of the full data sets) were taken; and in run M, the percentiles of the exhaustive data sets, in the form of indicator thresholds, were provided to the simulator. Why would additional information about the exhaustive data result in inaccurate confidence intervals? One possible explanation can best be understood in terms of sampling distributions.

Recall that the goal of SIS is to draw a number of realizations from an underlying distribution that characterizes the state of uncertainty about a particular region of interest. In the control runs (A, B, C, and D) of the experiment, exhaustive data sets were generated by assuming that the value of the random variable Z was known at two locations, and by providing SIS with indicator variograms for five different thresholds. Now suppose that we were to generate realizations by repeating this process one hundred more times, (without sampling from the exhaustive data set that we have just generated). In this case, our realizations would truly be drawn from the distribution that generated the exhaustive data sets, and we would expect that, in almost all cases, the value of a specific transfer function evaluated over the exhaustive data set would lie between the extremes of that transfer function evaluated over the realizations. (On average, the value for the exhaustive data set would lie outside of the range of the values over the realizations only 2 out of 101 times.)

The situation just described is similar to a control run with a low value of n: when the sample drawn from the exhaustive data set is small, and SIS is used to generate both the exhaustive data sets and the realizations, the realizations are essentially drawn from the distribution from which the exhaustive data set was drawn. The resulting confidence intervals for various transfer functions are not very precise, but

they are accurate. This changes when the sample size increases: the distribution used to generate the realizations is conditional on a large number of available data, and is no longer equivalent to the distribution used to generate the exhaustive data. The simulated distributions for various transfer functions become more precise, due to the constraints imposed by the large sample, but accuracy may decrease, as was observed in run D.

Now consider what happens when the indicator variograms used to generate the realizations contain specific information about the exhaustive data set, as was the case in run M. Here again, the realizations are drawn from a distribution different from the one from which the exhaustive data set was drawn. One would hope that, since information about the exhaustive data set was incorporated into the distribution from which SIS realizations are drawn, this distribution would be approximately centered on the exhaustive data set. The number of inaccurate confidence intervals seen in run M demonstrates that this is often not the case.

The problems become more pronounced when the exhaustive data set is not generated using SIS. The results of runs Q, R, S, T, and U were disappointing, and suggest that accuracy problems may be likely to occur when the region of interest has statistical properties that depart from the SIS model of specific dependency relationships at a number of known indicator thresholds. Although the SIS models appear to be more flexible than the familiar, Gaussian-based models, they have had difficulty adequately representing data sets with a fairly simple spatial structure, even when appropriate range and sill information is provided. This is particularly discouraging, since the available data from any target region is likely to be limited, and it will be difficult to check the validity of the indicator models. Accurate modeling of the tails of the exhaustive distribution is necessary, yet not always possible in practice.

It should be clear that qualitative judgments that SIS realizations look to be consistent with one's concepts about an underlying geologic process are not, by themselves, sufficient to justify the use of SIS-generated response distributions to quantify uncertainty. It is necessary that SIS be carefully (and quantitatively) tested in situations close to those that will be encountered in any practical case study. Validation studies must be carried out, for the relevant transfer functions, and for the data and models to be used in practice. In particular, users should attempt to determine the

influence that sample size, input variograms, sample locations, and other factors deemed relevant, have on their output.

Throughout this report, we have seen that SIS can be useful, but may give inaccurate results in a variety of situations. Users of SIS should be aware of its limitations, and should ensure that they are not applying the method, or attaching too much importance to the output, in cases where inaccuracies are likely to occur. One can have confidence in SIS results only by having information that is generally not available - like the theoretical indicator variograms associated with a particular region. In addition, tuning parameters of the SIS algorithm, that are unrelated to either data or geology, can have a major effect on the output. We strongly recommend that when SIS is used in practice to characterize uncertainty, users conduct sensitivity and validation studies for their own transfer functions and data sets.

6.0 REFERENCES

Anderson, T. W., (1984). *An Introduction to Multivariate Statistical Analysis,* Second Edition, John Wiley & Sons, New York. (NNA.910812.0002).

Cressie, N, (1989). "Geostatistics," *The American Statistician,* v. 43, no. 4, p. 197 - 202. (NNA.911007.0005).

Efron, B., and Gong, G., (1983). "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-validation," *The American Statistician,* v. 37, no. 1, p. 36 - 48. (NNA.910620.0120).

EPA (U. S. Environmental Protection Agency), 1986. "Environmental Radaition Protection Standards for Management and Disposal of Spent Nuclear Fuel, High-Level and Transuranic Radioactive Wastes," *Code of Federal Regulations, Protection of Environment.* Title 40, Part 191, Washington, DC. (HQS.880517.2925).

Feller, W., (1950). *An Introduction to Probability Theory and its Applications,* Volume I, Third Edition, John Wiley & Sons, New York. (NNA.910523.0030).

Gomez-Hernandez, J. J., Srivastava, R. M., and Serakiotou, N., (1989). *A Manual for isim3d - Version 1.1,* Stanford Center for Reservoir Forecasting, Department of Applied Earth Sciences, Stanford University, Stanford, CA. (NNA.910620.0123).

Journel, A. G., (1986). "Constrained Interpolation and Qualitative Information - the Soft Kriging Approach," *Mathematical Geology,* v. 18, no. 3, p. 269 - 286. (NNA.910523.0028).

Journel, A. G., (1988). "Fundamentals of Geostatistics in Five Lessons," in *Short Course in Geology,* Volume 8, American Geophysical Union, Washington, DC. (NNA.910405.0057).

Journel, A. G., and Alabert, F, (1989). "Non-Gaussian Data Expansion in the Earth Sciences," *Terra Nova,* v. 1, p. 123 - 134. (NNA.891208.0051).

Journel, A. G., and Huijbregts, Ch. J., (1978). *Mining Geostatistics,* Academic Press, London. (NNA.900919.0192).

Hohn, M. E., (1988). *Geostatistics and Petroleum Geology,* Van Nostrand Reinhold, New York. (NNA.910617.0004).

NRC (U. S. Nuclear Regulatory Commission), 1986. "Disposal of High-Level Radioactive Wastes in Geologic Repositories," *Code of Federal Regulations, Energy,* Title 10, Part 60, Washington, DC. (NNA.870325.0172).

Olea, R. A., (1974). "Optimal Contour Mapping Using Universal Kriging," *Journal of Geophysical Research,* v. 79, p. 695 - 702. (NNA.910523.0029).

Silvey, S. D., (1975). *Statistical Inference,* Chapman and Hall, London. (NNA.910531.0116).

Tukey, J. W., (1977). *Exploratory Data Analysis,* Addison-Wesley, Reading, MA. (NNA.910617.0008).

APPENDIX A - ISSUES RELATED TO UPDATED SIS SOFTWARE

While the simulation study was in progress, an updated version of the SIS software was released by SCRF. In this appendix, we discuss two issues that were raised by the updated program.

A.1 Correcting Order Relations Problems

The simulation study described in the main body of this report was carried out using version 1.1 of the computer program isim3d.c to perform sequential indicator simulation. After the study had been in progress for several months, a bug was discovered in this version of the code. The bug involved the correction procedures for the so-called "order relations problem," discussed in section 2.4. This problem can occur when indicator kriging is used to estimate the probability that the z-value at a particular node will be less than various thresholds: it is theoretically possible that the cumulative distribution function (cdf) specified by a sequence of kriged probabilities might not be a monotone non-decreasing function. For example, indicator kriging might estimate the probability that a node would be less than 0 as 0.50, and the probability that the same node would be less than 2 as .25. Clearly, such a situation violates the basic principles of probability. In an attempt to correct this problem, version 1.1 of SCRF's SIS computer code incorporates the following fix. If there are two or more cutoffs (indicator thresholds), then each pair of consecutive cutoffs is checked, in order from lowest to highest, to see if the estimated cumulative probability of the higher cutoff is less than that for the lower cutoff. When this condition is detected, the two estimated cumulative probabilities are averaged, and this average is assigned to both cutoffs in the pair.

This fix fails to ensure a correct solution, as can be seen in the following example. Suppose that a simulation uses 3 indicator thresholds, z_{k_1}, z_{k_2}, and z_{k_3}, with $z_{k_1} < z_{k_2} < z_{k_3}$. For a particular node, α, suppose that the initial estimated cumulative probabilities for the three thresholds are as follows:

$$P(z_\alpha \le z_1) = 0.7, \quad P(z_\alpha \le z_2) = 0.5, \text{ and } P(z_\alpha \le z_3) = 0.3.$$

When the first pair has been checked and corrected, both the first and second values

will be set to 0.6:

$$P(z_\alpha \le z_1) = 0.6, \quad P(z_\alpha \le z_2) = 0.6, \text{ and } P(z_\alpha \le z_3) = 0.3.$$

When the second pair is checked, both the second and third values will be set equal to .45:

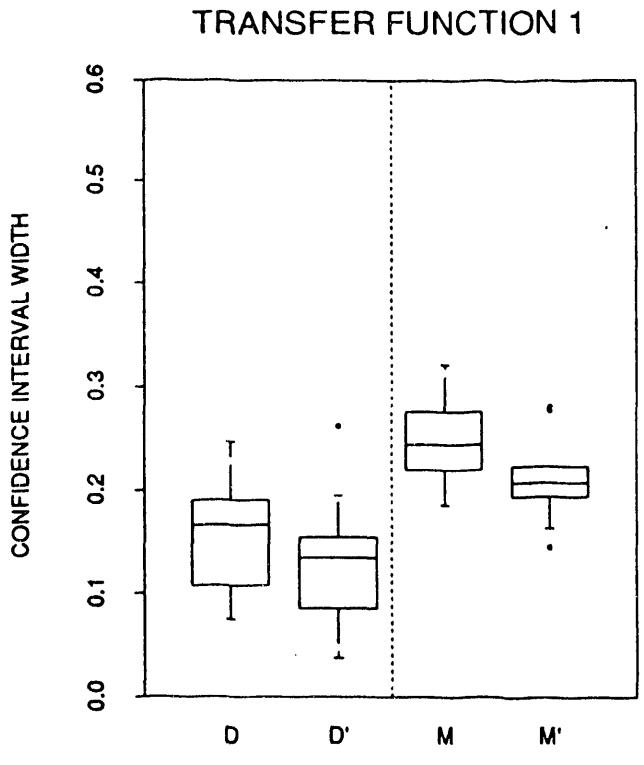
$$P(z_\alpha \le z_1) = 0.6, \quad P(z_\alpha \le z_2) = 0.45, \text{ and } P(z_\alpha \le z_1) = 0.45.$$

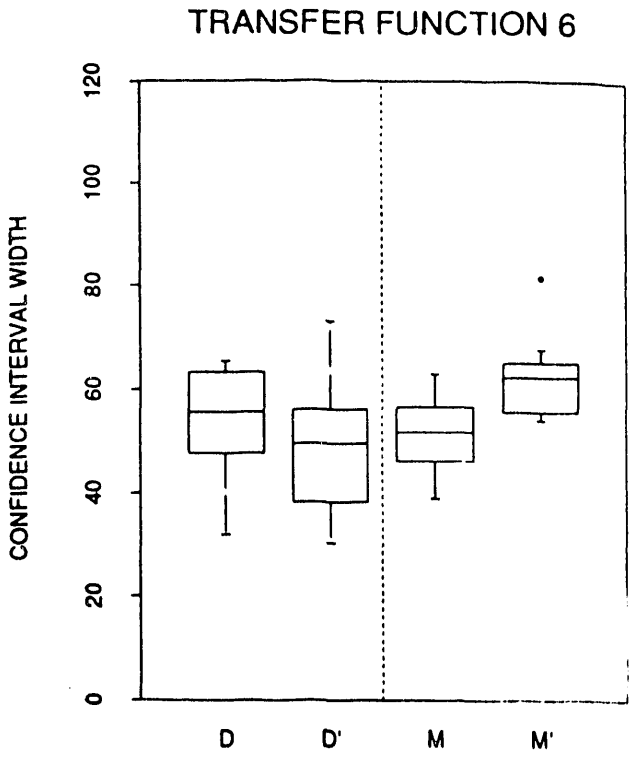
The correction subroutine is now complete, but the order relations problem remains: the resulting cdf is *not* a non-decreasing function. This is the bug that was present in the original implementation (version 1.1) of the software, and was not discovered until the present study was well under way. In December 1990, version 2.21, a debugged update of the program isim3d.c was released by SCRF. In this version, a new subroutine for correcting order relations problems is introduced. The program documentation describes the new correction algorithm as follows:

> After the conditional cdf has been estimated an order relation correction is carried out. This correction consists of performing two passes, one from the lowest threshold to the largest and the other in the opposite direction. In each pass a new estimate of the cdf without order relations is obtained. One of these estimates tends to overestimate the final cdf values and the other tends to underestimate the final cdf values. The arithmetic average of the two estimates is chosen as a better approximation.

The time required to rerun the entire study using version 2.21 was prohibitive. It was decided that a few key runs could be duplicated on the newer, debugged code. If the results did not differ substantially from the original results, it would be reasonable to conclude that the problems occurring in the original implementation were not solely due to the bug in the method of correcting the order relations problem.

Runs D and M were chosen for duplication. The runs carried out on debugged code will be referred to as D' and M', respectively. In the original implementation, runs D and M each gave a significant number of inaccurate Phase II confidence intervals for the majority of the transfer functions. Results for the duplicated runs are summarized in Figure A-1 and Table A-1.





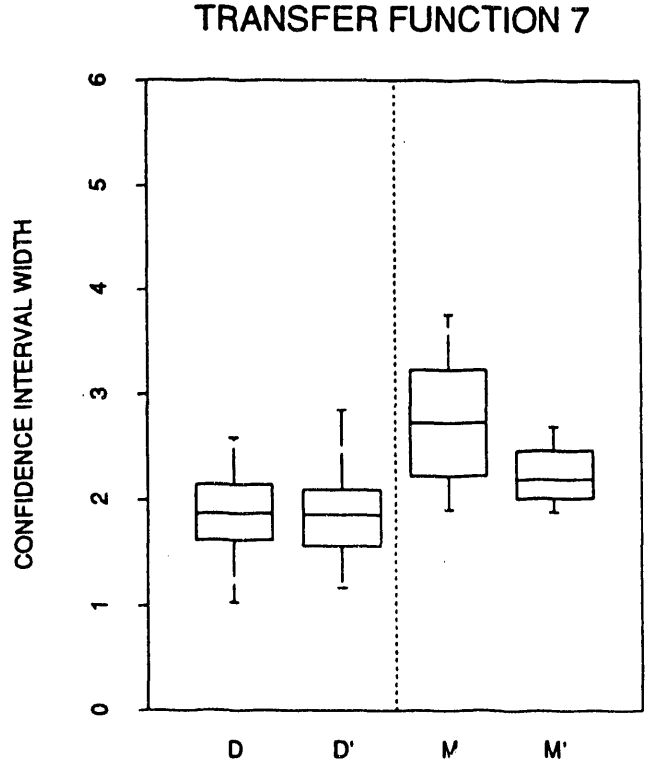


Figure A-1 Boxplots showing precision for two runs duplicated using version 2.21 of the simulation code. There does not appear to be any systematic increase or decrease in precision seen from run D to run D', or from run M to run M'.

TABLE A-1

ACCURACY AND PRECISION IN ORIGINAL RUNS D AND M, COMPARED TO RUNS D' AND M'

Median Confidence Interval Widths - Phase I								
Run	n	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
D	45	0.17	0.23	0.22	0.19	0.17	56	1.9
D'	45	0.13	0.21	0.25	0.23	0.17	50	1.9
M	30	0.24	0.29	0.26	0.22	0.18	52	2.7
M'	30	0.21	0.25	0.27	0.26	0.22	62	2.2

Number Inaccurate Confidence Intervals - Phase II								
Run	n	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7
D	45	3	6*	11*	10*	7*	6*	4
D'	45	6*	5	6*	3	3	5	7*
M	30	10*	6*	10*	10*	14*	6*	4
M'	30	10*	8*	1	4	8*	7*	8*

*Runs for which the number of inaccurate confidence intervals is statistically significant at a level less that 5%.

Runs D and D' gave comparable degrees of precision, as measured by the median confidence interval widths seen over ten Phase I replications. Accuracy improved in run D', where three out of seven transfer functions gave a significant number of inaccurate confidence intervals, (as compared to five out of seven for run D). Note however, that this improvement came largely in the simplest transfer functions, ψ_1 through ψ_5. The performance for the two complex and spatially-sensitive transfer functions (ψ_6 and ψ_7) did not improve. For ψ_6 in run D', five out of fifty confidence intervals were inaccurate; although not a statistically significant count, this was just one fewer than was observed for run D. And for ψ_7, a statistically significant seven out of fifty confidence intervals were inaccurate in run D'.

Runs M and M' also had comparable degrees of precision, and, as with runs D and D', accuracy in ψ_1 through ψ_5 seemed to be improved by the use of the debugged version of the simulation code. However, accuracy problems persisted, and actually worsened, for the spatially-sensitive transfer functions.

From these results, it appears that use of the debugged code may lead to an enhanced ability to accurately reproduce the marginal distribution of the z variable over the region of interest. This is seen in the improved accuracy of the confidence intervals for ψ_1 through ψ_5. However, this improvement does not seem to carry over into the higher-order properties of the realizations, as evidenced in the poor performances seen for the spatially-sensitive transfer functions. It is clear that the accuracy problems observed for many runs in the main body of this report cannot be attributed solely to incorrect solution of the order relations problem.

A.2 The Importance of the SK/OK Flag

When sequential indicator simulation is carried out, kriging is used to estimate the conditional cdf of the z variable at each unsampled node. Either simple or ordinary kriging may be used in this step. In general, if a considerable amount of data are available in the neighborhood of the node, ordinary kriging, with its capacity to fit a local mean, is used. If the information available in the neighborhood of the node is sparse, simple kriging is used. When this is the case, the mean of the conditional cdf is specified solely by the input indicator variograms, and is not fit locally.

The issue of when there are ample data to warrant the use of ordinary kriging turns out to be important to the performance of sequential indicator simulation. In version 1.1 of the program isim3d.c, simple kriging is used whenever five or fewer data points (samples or nodes already simulated) are in the neighborhood of the node to be simulated. (The neighborhood size is specified by the user.) In version 2.21 of the program, a simulation parameter called "OK flag" is introduced. The user sets OK flag equal to an integral value, and whenever more than this number of data points are in a neighborhood, ordinary kriging is used. (Setting OK flag = 5 is equivalent to the algorithm used in version 1.1.)

A brief experiment was carried out to demonstrate the effect of changing the OK flag on output simulations. The experiment followed an approach taken by Lucien Verrezen of Genmin during his visit to SCRF in the spring of 1990. An exhaustive data set was generated using version 2.21 of isim3d.c, with the same indicator variograms used in the control runs (A through D) of the simulation study, and with the OK flag set equal to five. A sample of 15 data points was taken from this exhaustive data set. The sample was then input as conditioning data to version 2.21 of the simulation code four different times. The same random seed and input indicator variograms were used each time, but different values of the OK flag were chosen. The first realization used OK flag = 0, corresponding to the case when ordinary kriging is invoked at all times, and simple kriging is never used. The second realization used OK flag = 5, the value used in version 1.1 of the computer code, (and in the simulation study). The third realization used OK flag = 10, and the last realization used OK flag = 900. Because there are only 900 nodes in the full data set, the latter case implies that simple kriging is used at all times, and ordinary kriging is never invoked.

Figure A-2 shows the exhaustive data set, the sample, and the four realizations. It is clear from this display that the value of the OK flag can have a large impact on the realizations. Note that when the OK flag is set equal to zero and ordinary kriging is used for all nodes, the realization is characterized by large patches of nodes falling within the same bin. This tendency, which is caused by the fitting of local means in all regions of the plane, becomes less evident in subsequent realizations that rely more on simple kriging. The opposite extreme is the realization with OK flag = 900, which demonstrates very little consistency among neighboring nodes.

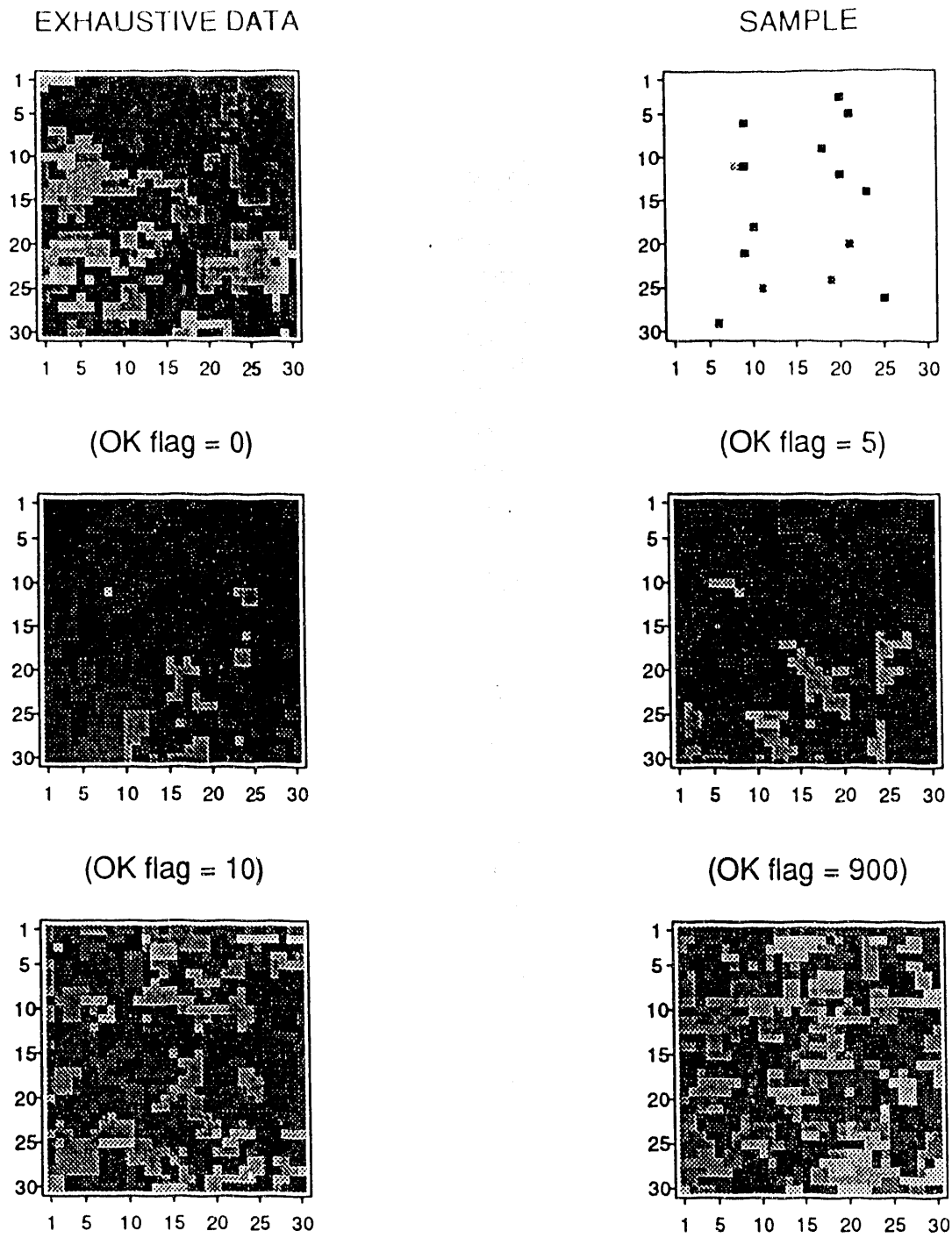


Figure A-2 Plot illustrating the influence of the OK flag on SIS realizations. The exhaustive data set was generated with OK flag = 5. A sample of size 15 was then taken, and used to generate four realizations. The same random seed was used for each realization, but the value of the OK flag varied. The first realization, which used only ordinary kriging, contains large patches of neighboring nodes with similar values. By contrast, the last realization, which used only simple kriging, exhibits very little spatial connectivity.

It is evident that the effect of the OK flag on the continuity of SIS realizations can be as large as the effect of the input indicator variograms. Unfortunately, no theory has been developed to help users to choose an appropriate value. The SCRF documentation suggests setting the OK flag equal to twice the maximum number of nodes per octant to be retained for kriging. This rule would give a value of 8 in the above example. It is not clear whether or why this formula is appropriate; it is clear that the user's choice of the OK flag can have a large impact on the output of SIS, and must be considered carefully.

APPENDIX B - SIMULATION METHODS AND PARAMETERS

This appendix describes, in detail, the simulation methods and parameters used to generate exhaustive data sets and realizations for the simulation study. All of the indicator variograms used in the study were spherical in shape and had zero nugget. In the notation used below, "cdf" refers to cumulative distribution function, "anis" refers to the ratio of the range in the x direction to the range in the y direction, and "a" represents the range in the x direction.

Runs A, B, C, D, E, F, G, H

The exhaustive data sets were generated using sequential indicator simulation. Five indicator thresholds were used: -2, -1, 0, 1, and 2. The parameters of the indicator variograms used for each threshold were as follows:

minimum value=-5, maximum value=5

threshold=-2, cdf=.20, anis=3, a=7, sill=.16

threshold=-1, cdf=.35, anis=3, a=5, sill=.2275

threshold=0, cdf=.50, anis=1, a=3, sill=.25

threshold=1, cdf=.65, anis=1, a=2, sill=.2275

threshold=2, cdf=.80 anis=1, a=2 sill=.16

The realizations were also generated using SIS with the indicator variograms specified above.

Runs I, J

The exhaustive data sets were generated exactly as in run A. The realizations were generated using SIS with the following indicator variogram parameters:

minimum value=-5, maximum value=5

threshold=-2, cdf=.25, anis=3, a=7, sill=.1875

threshold=-1, cdf=.40, anis=3, a=5, sill=.24

threshold=0, cdf=.50, anis=1, a=3, sill=.25

threshold=1, cdf=.60, anis=1, a=2, sill=.24

threshold=2, cdf=.75 anis=1, a=2 sill=.1875

Run K

The exhaustive data sets were generated exactly as in run A. The realizations were generated using SIS with the following indicator variogram parameters:

minimum value=-3, maximum value=3

(All thresholds and variograms as in run A.)

Run L

The exhaustive data sets were generated exactly as in run A. The realizations were generated using SIS with the following indicator variogram parameters:

minimum value=-6, maximum value=6

(All thresholds and variograms as in run I.)

Run M

The exhaustive data sets were generated as in run A. Realizations were generated using SIS, with the following parameters:

minimum value=p_0, maximum value=p_{100}

threshold=p_{20}, cdf=.20, anis=3, a=7, sill=.16

threshold=p_{35}, cdf=.35, anis=3, a=5, sill=.2275

threshold=p_{50}, cdf=.50, anis=1, a=3, sill=.25

threshold=p_{65}, cdf=.65, anis=1, a=2, sill=.2275

threshold=p_{80}, cdf=.80 anis=1, a=2 sill=.16,

where p_α refers to the α percentile of the exhaustive data set.

Runs N, P

The exhaustive data sets were generated as in run A. Realizations were generated using SIS, with the following parameters:

minimum value=q_0, maximum value=q_{100}

threshold=q_{20}, cdf=.20, anis=3, a=7, sill=.16

threshold=q_{35}, cdf=.35, anis=3, a=5, sill=.2275

threshold=q_{50}, cdf=.50, anis=1, a=3, sill=.25

threshold=q_{65}, cdf=.65, anis=1, a=2, sill=.2275

threshold=q_{80}, cdf=.80 anis=1, a=2 sill=.16,

where q_β refers to the β percentile of the related data set.

Runs Q, R

The exhaustive data sets were generated as follows: (1) Independently sample 900 values from a uniform distribution with mean 0, variance 1. Assign each value to a point on the 30 × 30 grid. (2) Use the Choleski decomposition of a theoretical covariance matrix to transform the data so that they exhibit the desired covariance structure (Anderson, 1984). The covariance matrix chosen represents a spherical variogram with a sill of 5 and a range of 10. Realizations were generated using SIS, with the following parameters:

minimum value=p_0, maximum value=p_{100}

threshold=p_{20}, cdf=.20, anis=1, a=10, sill=.16

threshold=p_{35}, cdf=.35, anis=1, a=10, sill=.2275

threshold=p_{50}, cdf=.50, anis=1, a=10, sill=.25

threshold=p_{65}, cdf=.65, anis=1, a=10, sill=.2275

threshold=p_{80}, cdf=.80 anis=1, a=10 sill=.16

where p_α refers to the α percentile of the exhaustive data set.

Runs S, T, U

The exhaustive and related data sets were generated in the same manner as the exhaustive data sets in run Q. Realizations were generated using SIS, with the following parameters:

minimum value=q_0, maximum value=q_{100}

threshold=q_{20}, cdf=.20, anis=1, a=10, sill=.16

threshold=q_{35}, cdf=.35, anis=1, a=10, sill=.2275

threshold=q_{50}, cdf=.50, anis=1, a=10, sill=.25

threshold=q_{65}, cdf=.65, anis=1, a=10, sill=.2275

threshold=q_{80}, cdf=.80 anis=1, a=10 sill=.16,

where q_β refers to the β percentile of the related data set.

APPENDIX C - INPUTS TO isim3d.c

This section describes, in detail, the simulation parameters used to generate the various exhaustive and realization data sets used in the study. The C language program isim3d.c (Gomez-Hernandez et al., 1989) provided to Sandia National Laboratories by the Stanford Center for Reservoir Forecasting, was used to generate all sequential indicator simulations. This program requires information from a number of input files. For all of the sequential indicator simulations carried out in this study, the input files *isim3d.spec,* and *isim3d.layer* appeared as shown in Table C-3 and Table C-4, respectively. The other two input files, *isim3d.var* and *isim3d.dat,* were varied over the different runs.

Runs A, B, C, D

Exhaustive Data Sets:

Generated using isim3d.c, with the following input files:

isim3d.var, isim3d.dat: as shown in Tables C-1, and C-2.

Realizations:

Generated using isim3d.c, with the following input files:

isim3d.var: as shown in Table C-1.

isim3d.dat: random samples from the exhaustive data set.

Runs E, F

Exhaustive Data Sets:

Generated using isim3d.c, with the following input files:

isim3d.var, isim3d.dat: as shown in Tables C-1 and C-2.

Realizations:

Generated using isim3d.c, with the following input files:

isim3d.var: as shown in Table C-1.

isim3d.dat: clustered random samples from exhaustive data sets. Clustered such that, on average, 67% of samples will lie within the first and third quadrants of grid square (upper right and lower left).

TABLE C-1

INPUT FILE isim3d.var*

isim3d.var		1
1.1	nind	5
1.2	z_min, z_max	-5 5
1.3	threshold	-2
1.4	p_cdf	.2
1.5	nugget	0
1.6	cmax	1
1.7	num_struct	1
1.8	type	2
1.9	sill	.16
1.10	a	7
1.11	anis.x	1
1.12	ains.y	3
1.13	anis.z	1
1.14 1.15 1.16	cos	1 0 0 0 1 0 0 0 1
2.3	threshold	-1
2.4	p_cdf	.35
2.5	nugget	0
2.6	cmax	1
2.7	num_struct	1
2.8	type	2
2.9	sill	.2275
2.10	a	5
2.11	anis.x	1
2.12	ains.y	3
2.13	anis.z	1
2.14 2.15 2.16	cos	1 0 0 0 1 0 0 0 1

isim3d.var, cont'd		2
3.3	threshold	0
3.4	p_cdf	.5
3.5	nugget	0
3.6	cmax	1
3.7	num_struct	1
3.8	type	2
3.9	sill	.25
3.10	a	3
3.11	anis.x	1
3.12	ains.y	1
3.13	anis.z	1
3.14 3.15 3.16	cos	1 0 0 0 1 0 0 0 1
4.3	threshold	1
4.4	p_cdf	.65
4.5	nugget	0
4.6	cmax	1
4.7	num_struct	1
4.8	type	2
4.9	sill	.2275
4.10	a	2
4.11	anis.x	1
4.12	ains.y	1
4.13	anis.z	1
4.14 4.15 4.16	cos	1 0 0 0 1 0 0 0 1

isim3d.var, cont'd		3
5.3	threshold	2
5.4	p_cdf	.80
5.5	nugget	0
5.6	cmax	1
5.7	num_struct	1
5.8	type	2
5.9	sill	.16
5.10	a	2
5.11	anis.x	1
5.12	ains.y	1
5.13	anis.z	1
5.14 5.15 5.16	cos	1 0 0 0 1 0 0 0 1

* The table is in the format of the worksheets recommended by Gomez-Hernandez et al. (1989). The actual input file contains only the numerical entries in column 3.

TABLE C-2

INPUT FILE isim3d.dat*

isim3d.dat				
ndata	2			
(x,y,z,value) in order: "y faster-than x faster-than z"				
1	3	3	0	1.201
2	12	12	0	0.672

* The table is in the format of the worksheets recommended by Gomez-Hernandez et al. (1989). The actual input file contains only the numerical entries in column 2.

TABLE C-3

INPUT FILE isim3d.spec*

isim3d.spec		
1	g_delta.x, g_delta.y, g_delta.z	1 1 2
2	g_origin.x, g_origin.y, g_origin.z	0 0 0
3	g_nodes.x, g_nodes.y, g_nodes.z	30 30 1
4	cdf	1
5	out_dummy, in_dummy	-1 1
6	seed	(varies)
7	from.x, from.y, from.z	1 1 1
8	to.x, to.y, to.z	30 30 1
9	radius.x, radius.y, radius.z	10 10 1
10 11 12	direction cosines	1 0 0 0 1 0 0 0 1
13	rotation flag	0
14	max_per_octant	4
15	dbg	0
16	out_flg	2

* The table is in the format of the worksheets recommended by Gomez-Hernandez et al. (1989). The actual input file contains only the numerical entries in column 3.

TABLE C-4

INPUT FILE isim3d.layer*

isim3d.layer		
top, thickness, in order: "y faster-than x axis"		
1	0	1
2	0	1
3	0	1
.	.	.
.	.	.
.	.	.
900	0	1

* The table is in the format of the worksheets recommended by Gomez-Hernandez et al. (1989). The actual input file contains only the numerical entries in column 2.

Runs G, H

Exhaustive Data Sets:

Generated using isim3d.c, with the following input files:

isim3d.var, isim3d.dat: as shown in Tables C-1 and C-2.

Realizations:

Generated using isim3d.c, with the following input files:

isim3d.var: as shown in Table C-1.

isim3d.dat: clustered random samples from exhaustive data sets. Clustered such that, on average, 57% (run G) or 73% (run H) of samples will lie within the first quadrant of grid square (upper right).

Runs I, J

Exhaustive Data Sets:

Generated using isim3d.c, with the following input files:

isim3d.var, isim3d.dat: as shown in Tables C-1 and C-2.

Realizations:

Generated using isim3d.c, with the following input files:

isim3d.var: as shown in Table C-1, with the following changes:

(line 1.4) .25

(line 1.9) .1875

(line 2.4) .40

(line 2.9) .24

(line 4.4) .60

(line 4.9) .24

(line 5.4) .75

(line 5.9) .1875

isim3d.dat: random samples from the exhaustive data set.

Run K

Exhaustive Data Sets:

Generated using isim3d.c, with the following input files:

isim3d.var, isim3d.dat: as shown in Tables C-1 and C-2.

Realizations:

Generated using isim3d.c, with the following input files:

isim3d.var: as shown in Table C-1, with the following change:

(line 1.2) -3 3

isim3d.dat: random samples from exhaustive data sets.

Run L

Exhaustive Data Sets:

Generated using isim3d.c, with the following input files:

isim3d.var, isim3d.dat: as shown in Tables C-1 and C-2.

Realizations:

Generated using isim3d.c, with the following input files:

isim3d.var: as shown in Table C-1, with the following changes:

(line 1.2) -6 6

(line 1.4) .25

(line 1.9) .1875

(line 2.4) .40

(line 2.9) .24

(line 4.4) .60

(line 4.9) .24

(line 5.4) .75

(line 5.9) .1875

isim3d.dat: random samples from exhaustive data sets.

Run M

Exhaustive Data Sets:

Generated using isim3d.c, with the following input files:

isim3d.var, isim3d.dat: as shown in Tables C-1 and C-2.

Realizations:

The quantiles of the exhaustive data set were used to compute indicator variograms for the realizations, which were generated using isim3d.c, with the following input files:

isim3d.var: as shown in Table C-1, with the following changes:

(line 1.2) min and max values of exhaustive data set

(line 1.3) 20^{th} percentile of exhaustive data set

(line 2.3) 35^{th} percentile of exhaustive data set

(line 3.3) 50^{th} percentile of exhaustive data set

(line 4.3) 65^{th} percentile of exhaustive data set

(line 4.3) 80^{th} percentile of exhaustive data set

isim3d.dat: random samples from the exhaustive data set.

Runs N, P

Exhaustive Data Sets:

Generated using isim3d.c, with the following input files:

isim3d.var, isim3d.dat: as shown in Tables C-1 and C-2.

Realizations:

For each exhaustive data set generated, a second, related data set was generated by isim3d.c, using the same input files. The quantiles of the related data set were used to model indicator variograms for the realizations, which were generated using isim3d.c, with the following input files:

isim3d.var: as shown in Table C-1, with the following changes:

(line 1.2) min and max values of related data set

(line 1.3) 20^{th} percentile of related data set

(line 2.3) 35^{th} percentile of related data set

(line 3.3) 50^{th} percentile of related data set

(line 4.3) 65^{th} percentile of related data set

(line 4.3) 80^{th} percentile of related data set

isim3d.dat: random samples from the exhaustive data set.

Runs Q, R

Exhaustive Data Sets:

Generated as follows: (1) Independently sample 900 values from a uniform distribution with mean 0, variance 1. Assign each value to a point on the 30 × 30 grid. (2) Use the Choleski decomposition of a theoretical covariance matrix to transform the data so that they exhibit the desired covariance structure (Anderson, 1984). The covariance matrix chosen represents a spherical variogram with a sill of five and a range of ten; this is the same shape as that of the indicator variograms specified in Table C-1.

Realizations:

Generated using isim3d.c, with the following input files:

isim3d.var: as shown in Table C-1, with the following changes:

(line 1.2) min and max values of exhaustive data set

(line 1.3) 20^{th} percentile of exhaustive data set

(line 1.10) 10

(line 1.12) 1

(line 2.3) 35^{th} percentile of exhaustive data set

(line 2.10) 10

(line 2.12) 1

(line 3.3) 50^{th} percentile of exhaustive data set

(line 3.10) 10

(line 4.3) 65^{th} percentile of exhaustive data set

(line 4.10) 10

(line 5.3) 80^{th} percentile of exhaustive data set

(line 5.10) 10

isim3d.dat: random samples from the exhaustive data set.

Runs S, T, U

Exhaustive Data Sets:

Generated as in run Q.

Realizations:

For each exhaustive data set, a second complete data set was generated using the same theoretical variogram as that of the exhaustive. This second data set is referred to as a "related" data set. The quantiles of the "related" data set were used to compute indicator variograms of the realizations. The realizations were generated using isim3d, with the following input files:

isim3d.var: as shown in Table C-1, with the following changes:

(line 1.2) min and max values of related data set

(line 1.3) 20^{th} percentile of related data set

(line 1.10) 10

(line 1.12) 1

(line 2.3) 35^{th} percentile of related data set

(line 2.10) 10

(line 2.12) 1

(line 3.3) 50^{th} percentile of related data set

(line 3.10) 10

(line 4.3) 65^{th} percentile of related data set

(line 4.10) 10

(line 5.3) 80^{th} percentile of related data set

(line 5.10) 10

isim3d.dat: random samples from the exhaustive data set.

APPENDIX D - INDICATOR VARIOGRAMS FOR EXHAUSTIVE DATA SETS GENERATED BY THE CHOLESKI METHOD

Runs Q, R, S, T, and U of the experiment used exhaustive data sets that were generated using the Choleski method. (For details of the method, see section 3.1.4.) Each of these exhaustive data sets had a theoretical z-variogram that was spherical in shape, with a range of ten and a sill of five. In generating the 100 realizations for these runs, five indicator thresholds were given to SIS, specifying the 0.20, 0.35, 0.50, 0.65 and 0.80 quantiles of the desired cumulative distribution function (cdf). At each threshold, SIS was told that the theoretical indicator variogram should be spherical, with a range of ten (the sill is determined solely by the cdf). In order to demonstrate that this was appropriate, i.e., that, for the exhaustive data sets, the shape of the indicator variograms at each of the five thresholds was well-modeled by the shape of the theoretical z-variogram, we present Figures D-1 through D-5.

The dashed lines of the left-hand panel of Figure D-1 show the observed indicator variograms (at the 0.20 quantile) for ten independent exhaustive data sets generated using the Choleski method. The solid line represents the corresponding indicator variogram model provided to SIS. Although none of the individual indicator variograms appear to be particularly close to the solid line, they do tend to cluster around it, and all have ranges approximately equal to ten. The right-hand panel of the figure shows (solid dots) the mean of the ten observed variograms; here, the correspondence with the indicator variogram model is excellent.

Figures D-2 through D-5 depict the individual and mean indicator variograms, for the same ten exhaustive data sets, for the other four thresholds. In each case, the indicator variogram models provided to SIS closely resemble the means of the individual indicator variograms. Based on this small simulation experiment, we conclude that the indicator variogram models provided to the SIS simulator for runs Q, R, S, T, and U were appropriate.

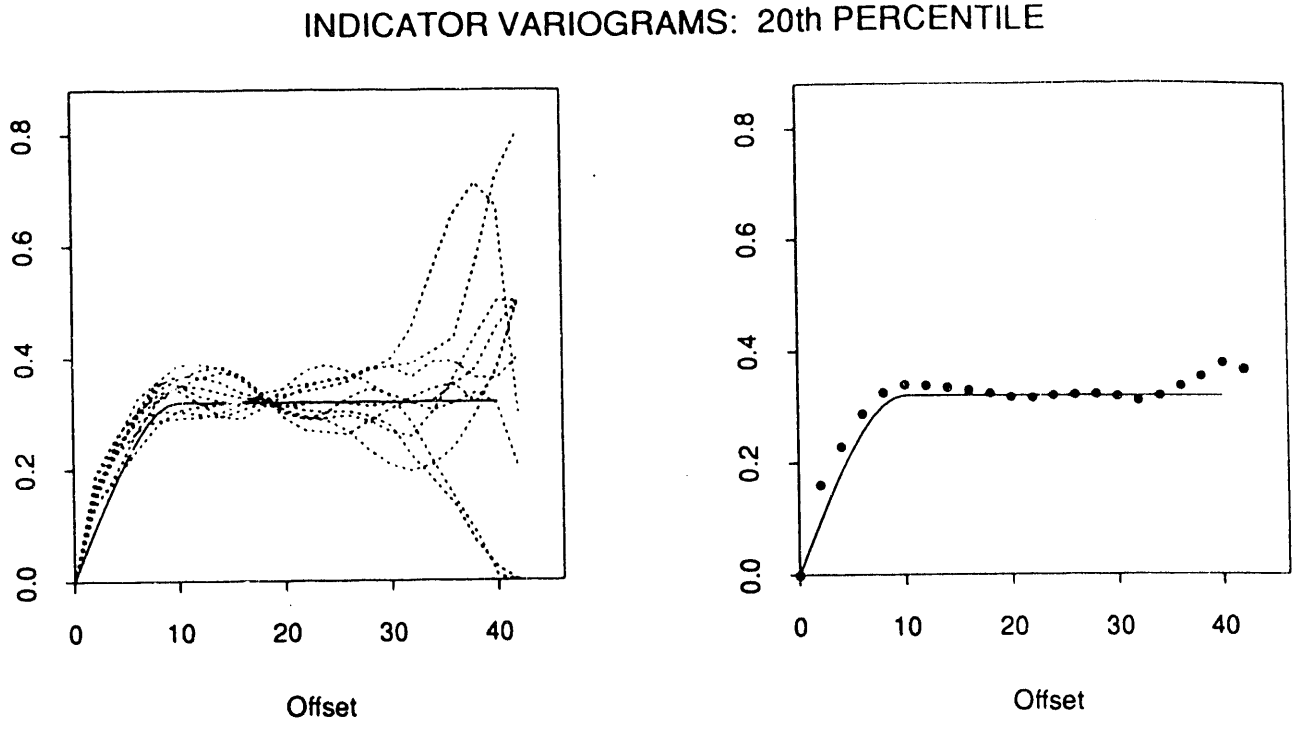


Figure D-1 The left-hand panel of the figure shows the exhaustive indicator variograms, at the 20th percentile, for ten 30 × 30 data sets, independently generated using the Choleski method. The solid line represents the theoretical indicator variogram model that was provided to SIS for generating realizations from such exhaustive data sets. The right-hand panel shows the mean of the ten individual indicator variograms, along with the theoretical model input to SIS. Agreement between the mean and theoretical indicator variograms is excellent.

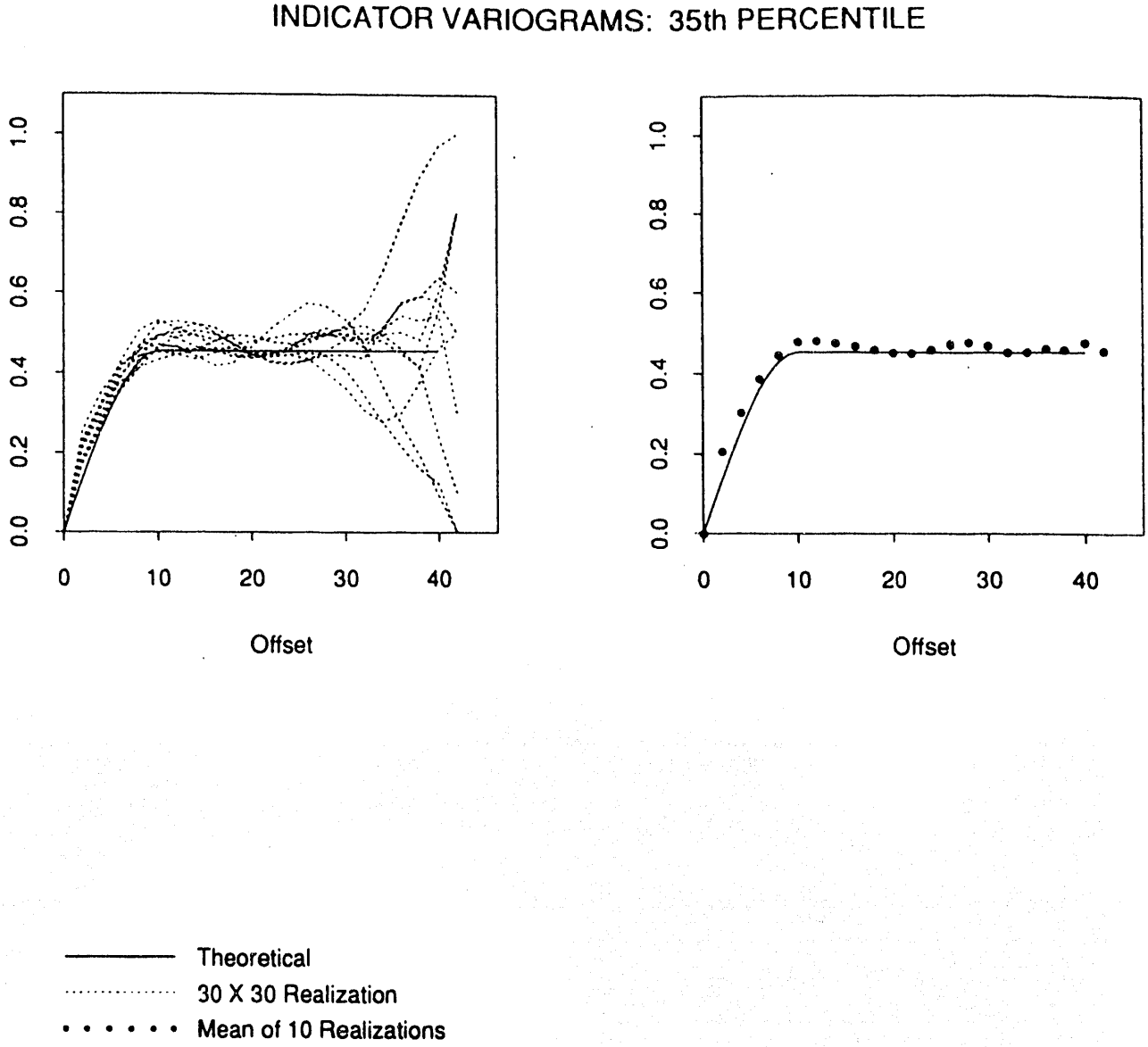


Figure D-2 Indicator variograms, at the 35th percentile, for ten data sets generated using the Choleski method.

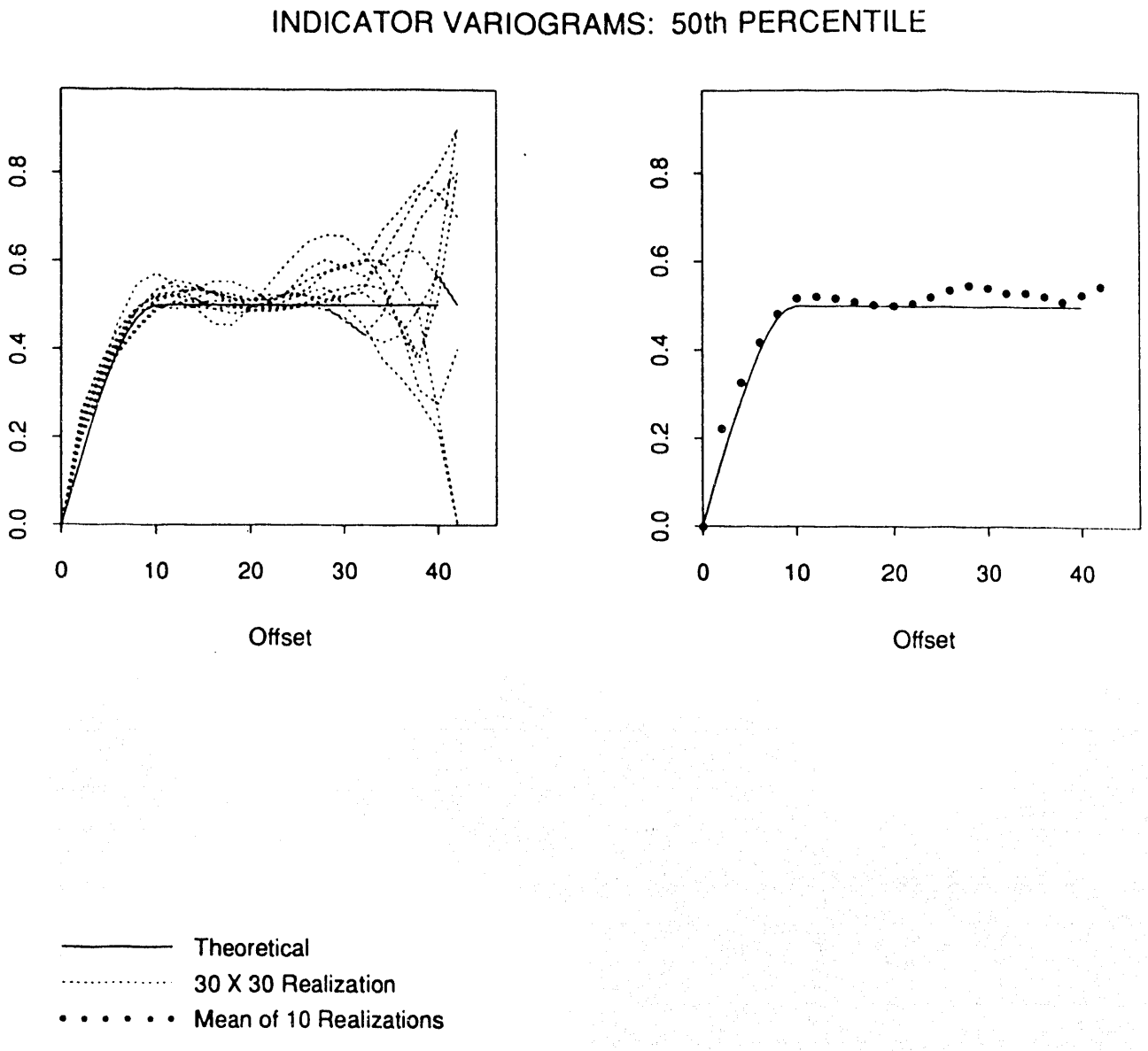


Figure D-3 Indicator variograms, at the 50th percentile, for ten data sets generated using the Choleski method.

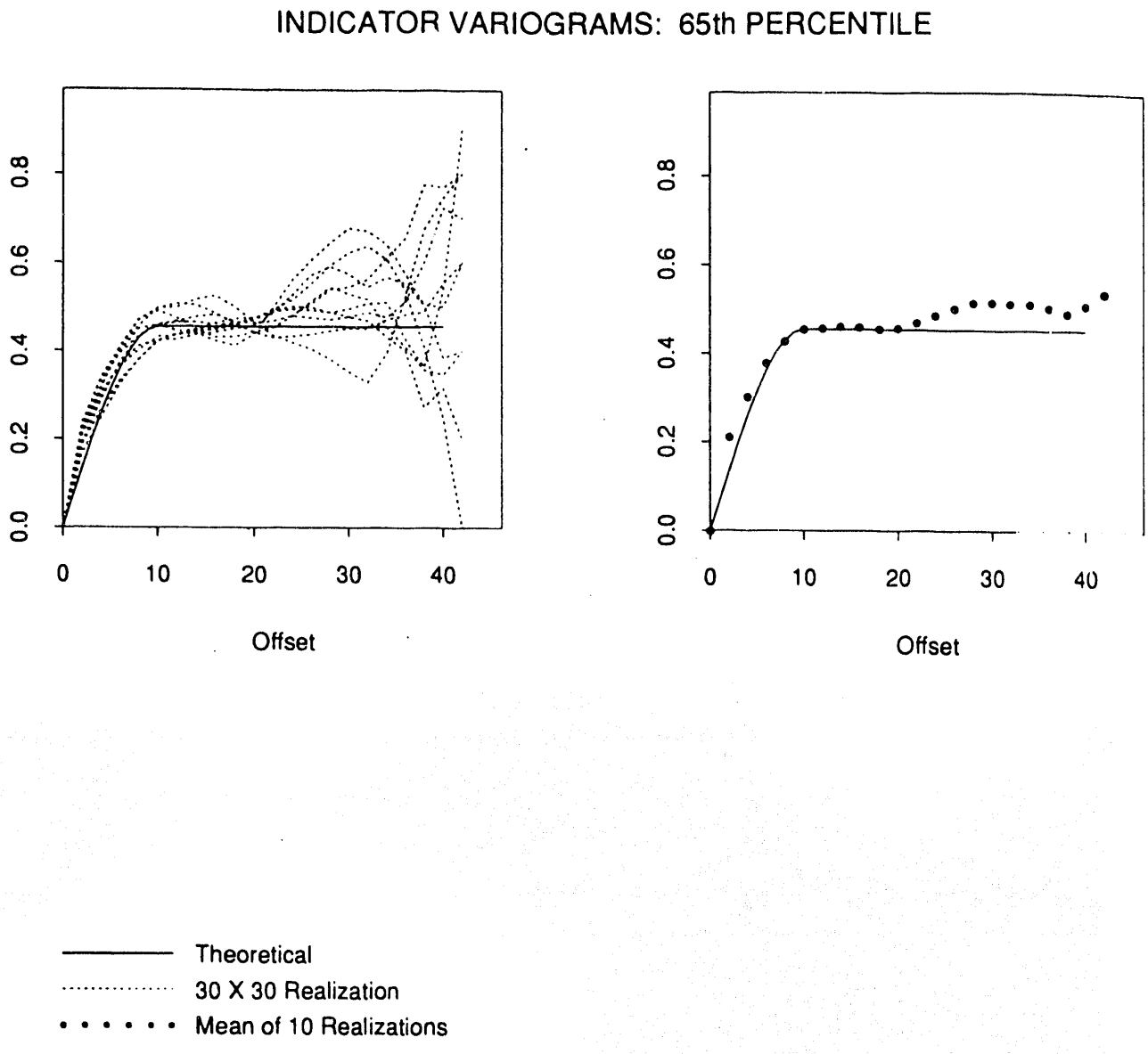


Figure D-4 Indicator variograms, at the 65th percentile, for ten data sets generated using the Choleski method.

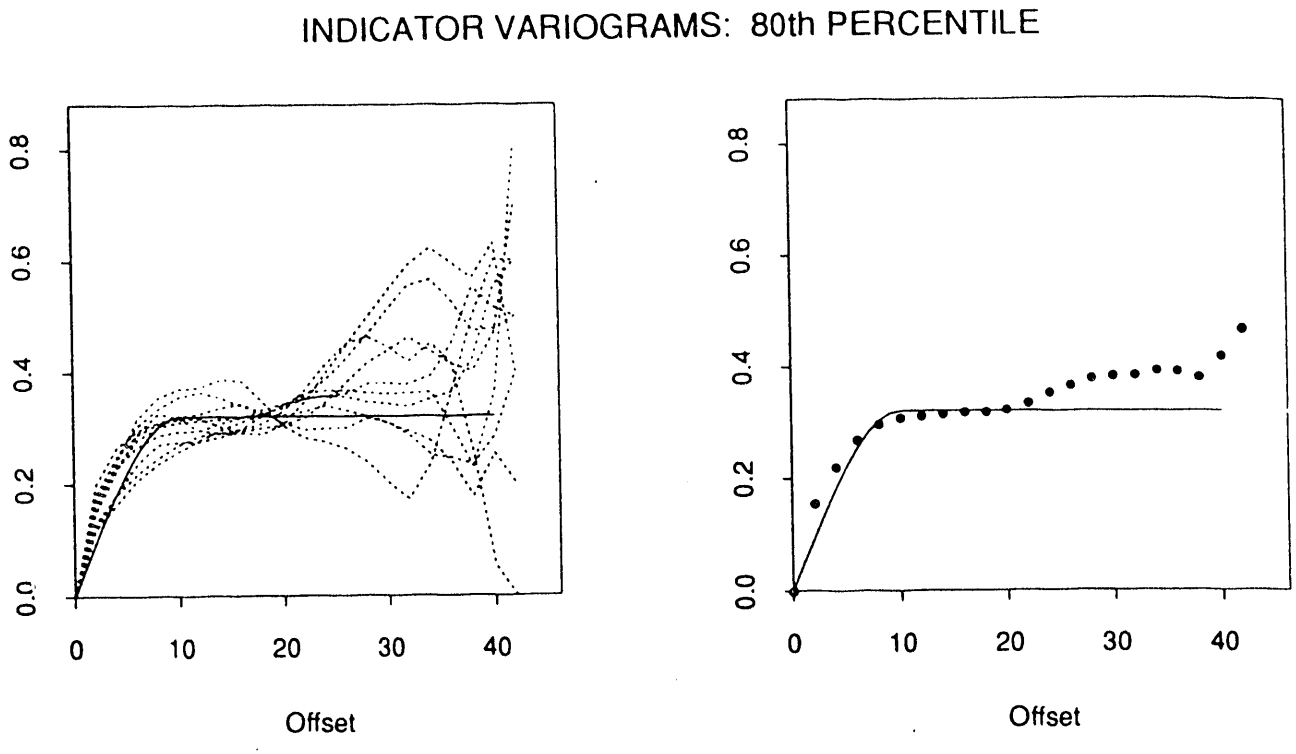


Figure D-5 Indicator variograms, at the 80th percentile, for ten data sets generated using the Choleski method.

APPENDIX E - HYPOTHESIS TESTS FOR CONFIDENCE INTERVAL ACCURACY

Each run in the first phase of the experiment used ten different exhaustive data sets, and generated 100 realizations of each. As the boxplots in Figure 9 (and others) show, ten exhaustive data sets are often sufficient to demonstrate changes in precision, from one run to the next, as characterized by confidence interval widths. However, the calculations below show that ten exhaustive data sets are not enough to determine whether or not the bootstrap confidence intervals obtained by multiple simulations are accurate.

Suppose that the distribution of the multiple simulations perfectly characterizes the uncertainty resulting from the use of simulated (rather than exact) data at a large percentage of nodes in the region of interest. In this case, we would expect that, for any single transfer function, approximately 5% of bootstrap 95% confidence intervals would fail to contain the true value of the transfer function. It follows that if ten confidence intervals are constructed, on average, 0.50 intervals will not contain the true value. Thus, in our simulation experiment, if all ten intervals contain the true value, we are likely to conclude that the simulation results are accurate. However, if at least one confidence interval does not contain the true value, we need to decide whether the results indicate that the simulations are inaccurate. To make this decision, we will use a statistical hypothesis test.

The appropriate null and alternative hypotheses for our test of accuracy are as follows:

$$H_0: \; p \leq 0.05$$

$$H_1: \; p > 0.05,$$

where p represents the percentage of confidence intervals not containing the true value of the transfer function. We will reject H_0 if the number of confidence intervals not containing the true value exceeds some threshold value x^*. The appropriate value for x^* is chosen by considering the level and power desired for the test. In statistical terms, the level of a hypothesis test, denoted α, is the probability that H_0 will be rejected when H_0 is, in fact, true. The power of a test, denoted $1-\beta$, is the probability that H_0 will be rejected when H_1 is true. Ideally, we would like a test with level $\alpha = 0$ and power $1-\beta = 1$. This is usually a practical impossibility, so that we typically choose

a test that has a reasonably low level (0.05 is a common value) and as high a power as possible. For a thorough discussion of the theory of statistical hypothesis testing, see Silvey (1975).

For our tests of confidence interval accuracy, the binomial distribution (Feller, 1950) is used for level and power calculations. If H_0 is true, then the probability that a single confidence interval will not contain the true value is given by $p = 0.05$. The probability that, out of n confidence intervals, exactly x will not contain the true value is given by

$$\binom{n}{x} .05^x \, .95^{n-x} , \tag{E.1}$$

and the probability that x *or fewer* intervals will not contain the true value is given by

$$P(n, x, .05) = \sum_{k=0}^{x} \binom{n}{k} .05^k \, .95^{n-k} . \tag{E.2}$$

In order to test at the 5% level, we will choose x^* such that $P(n, x^*, 0.05)$ exceeds 0.95. In this way, there is at most a 5% chance that H_0 will be rejected when H_0 is actually true. In order to ensure that the test has maximum power for the specified level, x^* is taken to be the smallest value of x such that $P(n, x, 0.05)$ exceeds 0.95.

For $n = 10$, $P(10, 1, 0.05) = 0.914$ and $P(10, 2, 0.05) = 0.989$, so that $x^* = 2$ is the appropriate critical value for testing at a level of at most 5%. (Note that the actual level of the test is not exactly equal to 0.05; it is $1 - 0.989 = 0.011$ which is somewhat conservative.) If ten confidence intervals are constructed using multiple simulations, and more than two fail to contain the true value of the transfer function, then we will reject the null hypothesis, and conclude that the confidence intervals are not accurate.

How powerful is the test? To answer this question, we can evaluate the probability that more than two (out of ten) confidence intervals will not contain the true value, as a function of p. This is the same as the probability that the null hypothesis will be rejected, and is given by:

$$1 - P(10, 2, p) = 1 - \sum_{k=0}^{2} \binom{10}{k} p^k \, (1-p)^{10-k} . \tag{E.3}$$

Table E-1 shows the power of the test for selected values of p, and Figure E-1 (solid line) graphically illustrates the relationship between p and the probability that H_0 will

TABLE E-1

POWER OF THE STATISTICAL TESTS FOR DETECTING INACCURATE CONFIDENCE INTERVALS

p	$1 - P(10, 2, p)$	$1 - P(50, 5, p)$
0	0.000	0.000
0.05	0.012	0.038
0.10	0.070	0.384
0.20	0.322	0.952
0.25	0.474	0.993
0.30	0.617	0.999
0.40	0.833	1.0
0.50	0.945	1.0
0.60	0.988	1.0
0.70	0.998	1.0
0.75+	1.0	1.0

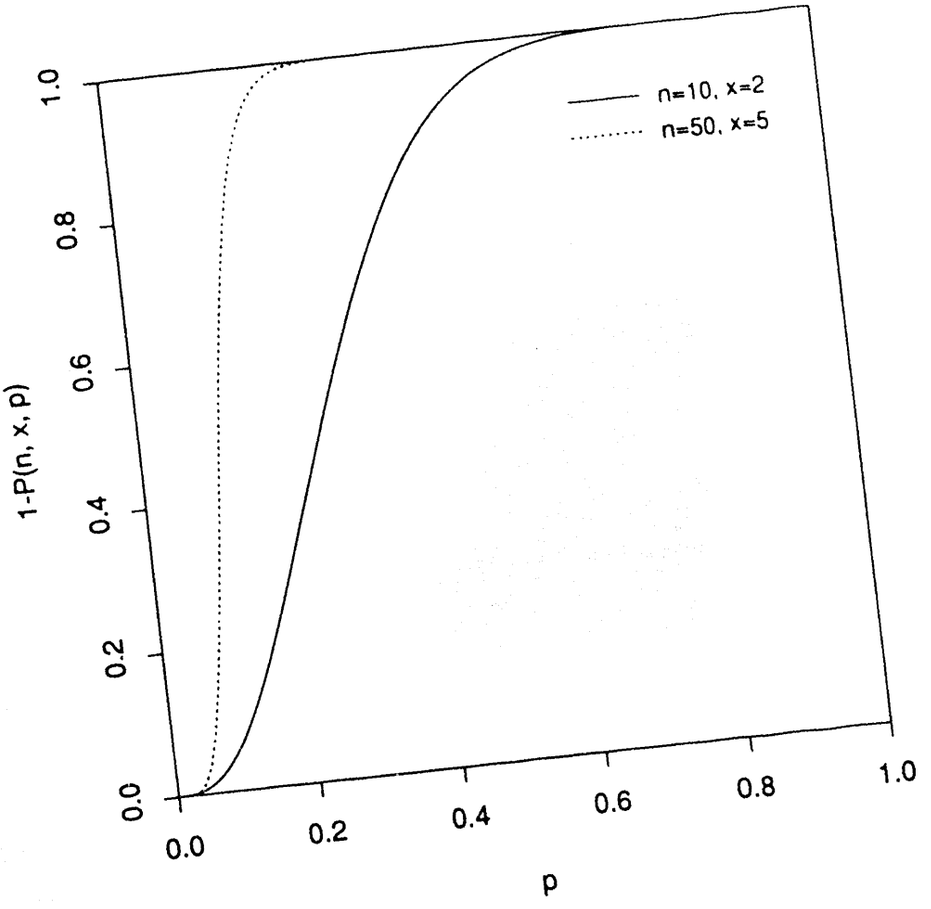


Figure E-1. Power of the statistical tests, as a function of p, the true rate of inaccuracy in constructed 95% confidence intervals.

be rejected. It is clear that departures from H_0 have to be quite substantial before the test has a good chance of detecting inaccuracy in the simulated confidence intervals: for example, even when the actual percent of confidence intervals failing to contain the true value is as high as 25%, there is still less than a 50% chance that the test will detect the departure from H_0. It is for this reason that Phase II of the experiment is necessary. Phase I, with only ten confidence intervals per run, does not provide enough information about the accuracy of inferences made from multiple conditional indicator simulations.

Each run in Phase II uses 50 different exhaustive data sets; thus 50 different bootstrap 95% confidence intervals are constructed for each run. In order to develop a hypothesis test for the accuracy of the confidence intervals, we begin by determining the critical number (out of 50) confidence intervals failing to contain the true value. Using expression (A.2), we find that $P(50, 4, 0.05) = 0.896$, and $P(50, 5, 0.05) = 0.962$. Thus, the appropriate critical value for a test of level at most 5% is $x^* = 5$, the smallest integer such that $P(50, 5, 0.05)$ exceeds 0.95. Here again, the true level of the test is not exactly 0.05; it is equal to $1 - 0.962 = 0.038$. The second data column of Table E-1 shows some values of the power of the test, $1 - \beta = 1 - P(50, 5, p)$, for various values of p. Note that the Phase II test is considerably more powerful than Phase I in detecting inaccuracy in the bootstrap confidence intervals. This can also be seen by comparing the solid and dashed lines on Figure E-1.

APPENDIX F

Information from the Reference Information Base Used in this Report

This report contains no information from the Reference Information Base.

Candidate Information for the Reference Information Base

This report contains no candidate information for the Reference Information Base.

Candidate Information for the Site & Engineering Properties Data Base

This report contains no candidate information for the Site and Engineering Properties Data Base.

DISTRIBUTION LIST

1 J. W. Bartlett, (RW-1)
Director
Office of Civilian Radioactive
Waste Management
U.S. Department of Energy
1000 Independence Avenue, S.W.
Washington, DC 20585

1 F. G. Peters, (RW-2)
Deputy Director
Office of Civilian Radioactive
Waste Management
U.S. Department of Energy
1000 Independence Avenue, S.W.
Washington, DC 20585

1 T. H. Isaacs (RW-4)
Office of Strategic Planning
and International Programs
OCRWM
U.S. Department of Energy
1000 Independence Avenue, S.W.
Washington, DC 20585

1 J. D. Saltzman (RW-5)
Office of External Relations
Office of Civilian Radioactive
Waste Management
U.S. Department of Energy
1000 Independence Avenue, S.W.
Washington, DC 20585

1 Samuel Rousso (RW-10)
Office of Program and Resources
Management
OCRWM
U.S. Department of Energy
1000 Independence Avenue, S.W.
Washington, DC 20585

1 J. C. Bresee (RW-10)
Office of Civilian Radioactive
Waste Management
U.S. Department of Energy
1000 Independence Avenue, S.W.
Washington, DC 20585

1 C. P. Gertz (RW-20)
Office of Geologic Disposal
Office of Civilian Radioactive
Waste Management
U.S. Department of Energy
1000 Independence Avenue, S.W.
Washington, DC 20585

1 S. J. Brocoum (RW-22)
Analysis and Verification Division
Office of Civilian Radioactive
Waste Management
U.S. Department of Energy
1000 Independence Avenue, S.W.
Washington, DC 20585

1 J. Roberts, Acting Assoc. Dir.
(RW-30)
Office of Systems and Compliance
Office of Civilian Radioactive
Waste Management
U.S. Department of Energy
1000 Independence Avenue, S.W.
Washington, DC 20585

1 J. Roberts (RW-33)
Director, Regulatory Compliance
Division
Office of Civilian Radioactive
Waste Management
U.S. Department of Energy
1000 Independence Avenue, S.W.
Washington, DC 20585

1 G. J. Parker (RW-332)
Office of Civilian Radioactive
Waste Management
U.S. Department of Energy
1000 Independence Avenue, S.W.
Washington, DC 20585

1 R. A. Milner (RW-40)
Office of Storage and Transportation
Office of Civilian Radioactive
Waste Management
U.S. Department of Energy
1000 Independence Avenue, S.W.
Washington, DC 20585

1 S. Rousso, Acting Assoc. Director
(RW-50)
Office of Contract Business
Management
Office of Civilian Radioactive
Waste Management
U.S. Department of Energy
1000 Independence Avenue, S.W.
Washington, DC 20585

19-OCT-92

1 T. Wood (RW-52)
Director, M&O Management Division
Office of Civilian Radioactive
Waste Management
U.S. Department of Energy
1000 Independence Avenue, S.W.
Washington, DC 20585

1 Dr. Garry D. Brewer
Nuclear Waste Technical Review Board
The Dana Building, Room 3516
University of Michigan
Ann Arbor, MI 48109-1115

1 Dr. Clarence R. Allen
Nuclear Waste Technical Review Board
1000 E. California Blvd.
Pasadena, CA 91106

1 Dr. John E. Cantlon, Chairman
Nuclear Waste Technical Review Board
1795 Bramble Dr.
East Lansing, MI 48823

1 Dr. Patrick A. Domenico
Nuclear Waste Technical Review Board
Geology Department
Texas A & M University
College Station, TX 77843

1 Dr. Donald Langmuir
Nuclear Waste Technical Review Board
109 So. Lookout Mountain Cr.
Golden, CO 80401

1 Dr. John J. McKetta, Jr.
Nuclear Waste Technical Review Board
Decision Focus, Inc.
4984 El Camino Real
Los Altos, CA 94062

1 Dr. D. Warner North
Nuclear Waste Technical Review Board
Decision Focus, Inc.
4984 El Camino Real
Los Altos, CA 94062

1 Dr. Dennis L. Price
Nuclear Waste Technical Review Board
1011 Evergreen Way
Blacksburg, VA 24060

1 Dr. Ellis D. Verink
Nuclear Waste Technical Review Board
4401 N.W. 18th Place
Gainesville, FL 32605

5 C. P. Gertz, Project Manager
Yucca Mountain Site Characterization
Project Office
U.S. Department of Energy
P.O. Box 98608--MS 523
Las Vegas, NV 89193-8608

1 C. L. West, Director
Office of External Affairs
DOE Field Office, Nevada
U.S. Department of Energy
P.O. Box 98518
Las Vegas, NV 89193-8518

12 Technical Information Officer
DOE Nevada Field Office
U.S. Department of Energy
P.O. Box 98518
Las Vegas, NV 89193-8518

1 P. K. Fitzsimmons, Technical
Advisor
Office of Assistant Manager for
Environmental Safety and Health
DOE Nevada Field Office
U.S. Department of Energy
P.O. Box 98518
Las Vegas, NV 89193-8518

1 D. R. Elle, Director
Environmental Protection Division
DOE Nevada Field Office
U.S. Department of Energy
P.O. Box 98518
Las Vegas, NV 89193-8518

1 Repository Licensing & Quality
Assurance
Project Directorate
Division of Waste Management
U.S. Nuclear Regulatory Commission
Washington, DC 20555

1 Senior Project Manager for Yucca
Mountain
Repository Project Branch
Division of Waste Management
U.S. Nuclear Regulatory Commission
Washington, DC 20555

1 NRC Document Control Desk
Division of Waste Management
U.S. Nuclear Regulatory Commission
Washington, DC 20555

1 P. T. Prestholt
NRC Site Representative
301 E. Stewart Ave., Room 203
Las Vegas, NV 89101

1 E. P. Binnall
Field Systems Group Leader
Building 50B/4235
Lawrence Berkeley Laboratory
Berkeley, CA 94720

1 Center for Nuclear Waste
Regulatory Analyses
6220 Culebra Road
Drawer 28510
San Antonio, TX 78284

3 W. L. Clarke
Technical Project Officer for YMP
Attn: YMP/LRC
Lawrence Livermore National
Laboratory
P.O. Box 5514
Livermore, CA 94551

1 J. A. Blink
Deputy Project Leader
Lawrence Livermore National
Laboratory
101 Convention Center Drive
Suite 280, MS 527
Las Vegas, NV 89109

4 J. A. Canepa
Technical Project Officer for YMP
N-5, Mail Stop J521
Los Alamos National Laboratory
P.O. Box 1663
Los Alamos, NM 87545

1 H. N. Kalia
Exploratory Shaft Test Manager
Los Alamos National Laboratory
Mail Stop 527
101 Convention Center Dr., Suite 820
Las Vegas, NV 89109

1 J. F. Divine
Assistant Director for
Engineering Geology
U.S. Geological Survey
106 National Center
12201 Sunrise Valley Dr.
Reston, VA 22092

1 L. R. Hayes
Technical Project Officer
Yucca Mountain Project Branch--MS 425
U.S. Geological Survey
P.O. Box 25046
Denver, CO 80225

1 V. R. Schneider
Asst. Chief Hydrologist--MS 414
Office of Program Coordination
& Technical Support
U.S. Geological Survey
12201 Sunrise Valley Drive
Reston, VA 22092

1 J. S. Stuckless
Geological Division Coordinator
MS 913
Yucca Mountain Project
U.S. Geological Survey
P.O. Box 25046
Denver, CO 80225

1 D. H. Appel, Chief
Hydrologic Investigations Program
MS 421
U.S. Geological Survey
P.O. Box 25046
Denver, CO 80225

1 E. J. Helley
Branch of Western Regional Geology
MS 427
U.S. Geological Survey
345 Middlefield Road
Menlo Park, CA 94025

1 R. W. Craig, Chief
Nevada Operations Office
U.S. Geological Survey
101 Convention Center Drive
Suite 860, MS 509
Las Vegas, NV 89109

1 D. Zesiger
U.S. Geological Survey
101 Convention Center Dr.
Suite 860, MS 509
Las Vegas, NV 89109

1 G. L. Ducret, Associate Chief
Yucca Mountain Project Division
U.S. Geological Survey
P. O. Box 25046
421 Federal Center
Denver, CO 80225

1 A. L. Flint
U.S. Geological Survey
MS 721
P.O. Box 327
Mercury, NV 89023

1 D. A. Beck
Water Resources Division
U.S. Geological Survey
6770 So. Paradise Road
Las Vegas, NV 89119

1 P. A. Glancy
U.S. Geological Survey
Federal Building, Room 224
Carson City, NV 89701

1 Sherman S. C. Wu
Branch of Astrogeology
U.S. Geological Survey
2255 N. Gemini Dr.
Flagstaff, AZ 86001

1 J. H. Sass
Branch of Tectonophysics
U.S. Geological Survey
2255 N. Gemini Dr.
Flagstaff, AZ 86001

1 DeWayne A. Campbell
Technical Project Officer for YMP
U.S. Bureau of Reclamation
Code D-3790
P.O. Box 25007
Denver, CO 80225

1 J. M. LaMonaca
Records Specialist
U.S. Geological Survey
421 Federal Center
P. O. Box 25046
Denver, CO 80225

1 W. R. Keefer
U.S. Geological Survey
913 Federal Center
P.O. Box 25046
Denver, CO 80225

1 M. D. Voegele
Technical Project Officer for YMP
Science Applications International Corp.
101 Convention Center Dr.
Suite 407
Las Vegas, NV 89109

2 L. D. Foust
Nevada Site Manager
TRW Environmental Safety Systems
101 Convention Center Drive
Suite 540, MS 423
Las Vegas, NV 89109

1 C. E. Ezra
YMP Support Office Manager
EG&G Energy Measurements, Inc.
MS V-02
P.O. Box 1912
Las Vegas, NV 89125

1 E. L. Snow, Program Manager
Roy F. Weston, Inc.
955 L'Enfant Plaza, Southwest
Washington, DC 20024

1 Technical Information Center
Roy F. Weston, Inc.
955 L'Enfant Plaza, Southwest
Washington, DC 20024

1 D. Hedges, Vice President,
Quality Assurance
Roy F. Weston, Inc.
4425 Spring Mountain Road, Suite 300
Las Vegas, NV 89102

1 D. L. Fraser, General Manager
Reynolds Electrical & Engineering Co.
Mail Stop 555
P.O. Box 98521
Las Vegas, NV 89193-8521

1 R. F. Pritchett
Technical Project Officer for YMP
Reynolds Electrical & Engineering Co.
MS 408
P.O. Box 98521
Las Vegas, NV 89193-8521

1 B. W. Colston
President/General Manager
Las Vegas Branch
Raytheon Services Nevada
MS 416
P.O. Box 95487
Las Vegas, NV 89193-5487

1 R. L. Bullock
Technical Project Officer for YMP
Raytheon Services Nevada
Suite P250, MS 403
101 Convention Center Dr.
Las Vegas, NV 89109

1 Paul Eslinger, Manager
PASS Program
Pacific Northwest Laboratories
P.O. Box 999
Richland, WA 99352

1 A. T. Tamura
Science and Technology Division
Office of Scientific and Technical Information
U.S. Department of Energy
P.O. Box 62
Oak Ridge, TN 37831

1 Carlos G. Bell, Jr.
Professor of Civil Engineering
Civil and Mechanical Engineering Department
University of Nevada, Las Vegas
4505 South Maryland Parkway
Las Vegas, NV 89154

1 P. J. Weeden, Acting Director
Nuclear Radiation Assessment Division
U.S. Environmental Protection Agency
Environmental Monitoring Systems Laboratory
P.O. Box 93478
Las Vegas, NV 89193-3478

1 ONWI Library
Battelle Columbus Laboratory
Office of Nuclear Waste Isolation
505 King Avenue
Columbus, OH 43201

1 T. Hay, Executive Assistant
Office of the Governor
State of Nevada
Capitol Complex
Carson City, NV 89710

3 R. R. Loux, Jr.
Executive Director
Nuclear Waste Project Office
State of Nevada
Evergreen Center, Suite 252
1802 North Carson Street
Carson City, NV 89710

1 C. H. Johnson
Technical Program Manager
Nuclear Waste Project Office
State of Nevada
Evergreen Center, Suite 252
1802 North Carson Street
Carson City, NV 89710

1 John Fordham
Water Resources Center
Desert Research Institute
P.O. Box 60220
Reno, NV 89506

1 David Rhode
Desert Research Institute
P. O. Box 60220
Reno, NV 89506

1 Eric Anderson
Mountain West Research-Southwest Inc.
2901 N. Central Ave. #1000
Phoenix, AZ 85012-2730

1 Department of Comprehensive Planning
Clark County
225 Bridger Avenue, 7th Floor
Las Vegas, NV 89155

1 Planning Department
Nye County
P.O. Box 153
Tonopah, NV 89049

1 Lincoln County Commission
Lincoln County
P.O. Box 90
Pioche, NV 89043

5 Judy Foremaster
City of Caliente
P.O. Box 158
Caliente, NV 89008

1 Economic Development Department
City of Las Vegas
400 East Stewart Avenue
Las Vegas, NV 89101

1 Community Planning & Development
City of North Las Vegas
P.O. Box 4086
North Las Vegas, NV 89030

1 Director of Community Planning
City of Boulder City
P.O. Box 61350
Boulder City, NV 89006

1 Commission of the European Communities
200 Rue de la Loi
B-1049 Brussels
BELGIUM

2 M. J. Dorsey, Librarian
YMP Research and Study Center
Reynolds Electrical & Engineering Co., Inc.
MS 407
P.O. Box 98521
Las Vegas, NV 89193-8521

1 Amy Anderson
Argonne National Laboratory
Building 362
9700 So. Cass Ave.
Argonne, IL 60439

1 Steve Bradhurst
P.O. Box 1510
Reno, NV 89505

1 Vernon Poe
P.O. Box 1026
Hawthorne, NV 89415

1 Jason Pitts
Lincoln County Courthouse
Pioche, NV 89043

1 Michael L. Baughman
35 Clark Road
Fiskdale, MA 01518

1 Glenn Van Roekel
Director of Community Development
City of Caliente
P.O. Box 158
Caliente, NV 89008

1 Ray Williams, Jr.
P.O. Box 10
Austin, NV 89310

1 Leonard J. Fiorenzi
P.O. Box 257
Eureka, NV 89316

1 Juanita Hayes
P.O. Box 490
Goldfield, NV 89013

1 Bjorn Selinder
190 W. First St.
Fallon, NV 89406

1 Charles Thistlethwaite, AICP
Associate Planner
Planning Department
Drawer L
Independence, CA 93526

1 Les Bradshaw
Nye County District Attorney
P.O. Box 593
Tonopah, NV 89049

1	6300	D. E. Miller
1	6302	T. E. Blejwas
1	6304	J. T. Holmes
1	6312	F. W. Bingham
1	6313	L. S. Costin
2	6318	Gail Gerstner-Miller for 100/1232222/SAND 91-0758/NQ
1	6319	R. R. Richards
1	0323	B. Rutherford
1	0323	R. Easterling
10	0323	K. Hansen
1	6115	P. J. Hommert, Acting
20	6341	WMT Library
1	6410	D. A. Dahlgren
5	7141	Technical Library
1	7151	Technical Publications
10	7613-2	Document Processing for DOE/OSTI
1	8523-2	Central Technical Files

QAGR 1232222A, Rev. O

END

DATE FILMED

4/7/93