SKB

TECHNICAL REPORT 92-18

Stochastic continuum simulation of mass arrival using a synthetic data set. The effect of hard and soft conditioning

Kung Chen Shan[1], Wen Xian Huan[1], Vladimir Cvetkovic[1], Anders Winberg[2]

[1] Royal Institute of Technology, Stockholm
[2] Conterra AB, Gothenburg

June 1992

SVENSK KÄRNBRÄNSLEHANTERING AB
SWEDISH NUCLEAR FUEL AND WASTE MANAGEMENT CO
BOX 5864 S-102 48 STOCKHOLM
TEL 08-665 28 00 TELEX 13108 SKB S
TELEFAX 08-661 57 19

STOCHASTIC CONTINUUM SIMULATION OF MASS ARRIVAL USING A SYNTHETIC DATA SET.
THE EFFECT OF HARD AND SOFT CONDITIONING

Kung Chen Shan[1], Wen Xian Huan[1], Vladimir Cvetkovic[1], Anders Winberg[2]

1 Royal Institute of Technology, Stockholm
2 Conterra AB, Gothenburg

June 1992

Stochastic Continuum Simulation of Mass Arrival using a Synthetic Data Set

-

The effect of hard and soft conditioning

by

Kung Chen Shan[1)], Wen Xian Huan[1)], Vladimir Cvetkovic[1)]
and
Anders Winberg[2)]

[1)] Royal Institute of Technology, Stockholm

[2)] Conterra AB, Göteborg

June 1992

ABSTRACT

The non-parametric and parametric stochastic continuum approaches were applied to a realistic synthetic exhaustive hydraulic conductivity field to study the effects of hard and soft conditioning. From the reference domain, a number of data points were selected, either in a random or designed fashion, to form sample data sets. Based on established experimental variograms and the conditioning data, 100 realizations each of the studied domain were generated. The flow field was calculated for each realization, and particle arrival time and arrival position along the discharge boundary were evaluated. It was shown that conditioning on soft data reduces the uncertainty of solute arrival time, and that conditioning on soft data suggests an improvement in characterizing channelling effects. It was found that the improvement in the prediction of the breakthrough was moderate when conditioning on 25 hard and 100 soft data compared to 25 hard data only.

Keywords:

Conditioning, hard data, indicator, non-parametric, parametric, sequential indicator simulation, soft data, stochastic, synthetic, transport, turning bands, variogram.

SUMMARY

The transport of solute in fractured rock is to a large extent controlled by the spatial variability in fluid advection which are due to heterogeneities in the hydraulic properties of the studied rock formation. This variability is largely unknown, and thus modelling of solute transport is associated with uncertainty. This uncertainty can be accounted for by using stochastic simulation, and can possibly be reduced further by incorporating all available information, quantitative and qualitative.

To study the effects of incorporating qualitative (soft) information, the non-parametric and parametric stochastic continuum approaches, were applied to a realistic synthetic exhaustive hydraulic conductivity field (N=2500) generated in two dimensions. From this reference domain, a number of data points were selected, in a designed or random fashion, to form a sample data set. The points were informed as either hard or soft information. These data correspond to our measured hydraulic conductivity and supporting geological/geophysical information from the field, respectively.

Based on established experimental variograms and the conditioning data, 100 realizations of the studied domain were generated through; a) conditioned parametric method, and b) conditioned non-parametric method (with and without soft information). The flow field was calculated for each realization using a linear hydraulic gradient. A number of particles were released at the upstream end of the domain and particle breakthrough and arrival position distribution along the discharge boundary was evaluated.

The specific conclusions of the study are that; 1) conditioning on soft data reduces the uncertainty of solute arrival time; 2) conditioning on soft data indicates an improvement in characterizing channelling effects, although the latter statement requires further study.

The improvement in the prediction of particle breakthrough when conditioning on 25 hard and 100 soft data was found to be moderate. The number and distribution of actual hydraulic conductivity measurements (hard data) are critical for an accurate description of flow and transport in a heterogeneous formation. Since the statistics of the 25 random sampled hard data corresponded well with that of the reference field, the incorporation of a limited number of soft data (100) did not improve the simulations significantly.

Designed sampling has provided a mild improvement in some of the simulated cases. However, investigations of different sampling strategies have been very limited in this study. A more extensive investigation in this direction could provide more reliable recommendations with regard to efficient sampling of hard and soft data.

The uncertainty associated with the heterogeneity in the hydraulic properties of fractured formations is not the sole one. In order to perform more reliable simulations of flow and transport the uncertainty in boundary conditions and solute source (near-field) characteristics have to be taken into account.

TABLE OF CONTENTS

1. INTRODUCTION

The transport of solute in fractured rock is to a large extent controlled by the spatial variability in fluid advection which is due to heterogeneities in the hydraulic properties of the studied rock formation. When modelling solute transport, the model output is associated with uncertainty. This uncertainty is a result of the fact that we can never establish a full description of the real system based on the often minute number of data points available.

When modelling large scale solute transport in fractured rock a number of different conceptual approaches are available. The most commonly used is that of a deterministic continuum. Here the modelled domain is divided into hydraulic units which are assigned effective material properties. Uncertainty in model output, eg. tracer breakthrough, can only be addressed by sensitivity analysis whereby the effective values of the hydraulic units are changed and/or permutated.

A stochastic continuum approach implies generation of a large number of realizations of heterogeneous material property fields that in a statistical sense are consistent with available data. The simulated realizations could be either unconditioned or conditioned, ie. in the latter case explicitly honouring measured data at their locations. In addition two different formalisms are available, the parametric and the non-parametric approaches. The former assumes and stipulates a distribution function for the material property analysed, eg. log-normal distribution of hydraulic conductivity. The latter method requires no such assumption. Also, in the latter case the geostatistical analysis may be based on either hard (quantitative) or soft (qualitative) information. The uncertainty in a given output may in either case, wether parametric or non-parametric, be acquired by analyzing the statistics of a large number of realizations.

The possibility to incorporate qualitative information implies that geological, geochemical and geophysical information may not only be used in the conceptualization of a hydrogeological system, but can actually be directly used in simulation of flow and transport. It also implies that old hydrogeological data can be reviewed on the basis of their accuracy and be incorporated with different degree of "softness". Another overall implication is a need for integration and optimization of site-characterization programmes, and an even stronger interaction between geoscientists from different disciplines than before.

The study of mass arrival at Finnsjön (Winberg et al 1991) aimed at providing an introduction and test of the non-parametric approach on a real data set, incorporate any available soft information and also to provide a comparison with the parametric technique. The analysis results where found to be encouraging but no definite conclusions could be drawn since no conditioned parametric realizations could be generated at the time of the study. In addition, no independent soft information could be utilized. An analysis of a synthetic exhaustive domain was suggested to further test the non-parametric technique and augment its potential usefulness in SKB:s modelling activities.

2. SCOPE

The main scope of the study is to show the merits of conditioning on soft (qualitative or semi-quantitative) information when performing stochastic continuum simulations of mass arrival.

In doing so, a realistic synthetic hydraulic conductivity field (the exhaustive data set) is generated in two dimensions with well defined statistical and geostatistical characteristics. This domain corresponds to our hydrogeological "reality". From this reference domain, a number of data points are selected, in a designed or random fashion, to form a sample data set. The points are informed either as hard or soft information. These data corresponds to our measured hydraulic and supporting geological/geophysical information from the field, respectively.

The sample data set is used to establish the statistical and geostatistical features of our material property distribution of interest. Experimental variograms are to be generated for the attribute (the hydraulic conductivity) and for defined indicator thresholds. Indicator variograms are to be generated for hard data only and for hard plus soft information.

The experimental variograms and the statistical parameters established on the basis of the sample data set (structured and random sampling scheme) are used to generate realizations of the studied domain through; a) conditioned parametric method (N=100), and b) conditioned non-parametric method (with and without soft information) (N=100), where N equals the number of realizations generated. A linear hydraulic gradient is applied to the domain and the flow field is calculated for each realization. A thousand particles are released at the upstream end of the domain. Particle breakthrough and distribution along the discharge boundary are evaluated.

Correspondingly, the sensitivity to boundary conditions were investigated by rotating the flow conditions 90 degrees.

Convergence in correspondence between results based on generated realizations and results based on the exhaustive data set are compared and discussed on basis of a) the sampling methodology employed, b) the number of informed data (hard, hard and soft), and the simulation method used. Tools used in this comparison are the graphical representation of individual realizations, output from transport calculations.

3. STOCHASTIC CONTINUUM APPROACHES

The two approaches used in this study are similar to the extent that they both adopt the continuum approach, but differ in the aspect of assumptions regarding underlying statistical distribution of hydraulic conductivity/transmissivity. In the Parametric Stochastic Continuum approach, a specific distribution model, in most cases Gaussian, is inherent in the model assumptions. In addition, the spatial correlation is based on the actual (attribute) log(K) or log(T) values. In the case of the Non-Parametric

Stochastic Continuum, no particular distribution model is required. In addition the technique honour the spatial variability and correlation of the studied parameter at different threshold values of hydraulic conductivity.

3.1 Parametric Stochastic Continuum

A scale is selected at which the random variable $Z(u)$ (Z=ln(K)) is to be discretized (eg. finite element block size). This scale is selected larger than the correlation scale of the Z process. Using conditional Monte-Carlo simulation, estimates of Z to be assigned to each of the grid blocks of the model are retained. This process is repeated until an ensemble of realizations have been generated. All realizations are consistent, in a geostatistical sense, with the available sample data set.

Estimates of Z are obtained using the Turning Bands method (Matheron 1973, Mantoglou and Wilson 1982) which requires the probability distribution function and the variogram of the attribute process Z. Conditional simulations are obtained following the procedure presented by Delhomme (1979).

3.2 Non-Parametric Stochastic Continuum

The non-parametric approach to the description of flow and transport differs from the classic stochastic continuum approach in that the spatial correlation is accounted for at different threshold values of hydraulic conductivity/transmissivity. The correlation structures at different thresholds may subsequently be calculated and used in ordinary kriging (OK) to infer not only one discrete estimate in a given point, but a so-called conditional distribution from which an optimal estimate may be retained. This distribution may be used to predict the uncertainty in the estimate before any type of estimation is made.

The conventional kriging algorithm as applied to attribute values (log K) functions as an effective smoothener of spatial connectivity characteristics. If a Gaussian related model is adopted in data expansion, as in the Parametric approach, it has a tendency to blur or underestimate the connectivity of high and low values. In a non-parametric expansion the spatial connectivity modelled through a suite of indicator covariance models (joint probability of two values in space exceeding a given permeability threshold) provides a possibility to emphasize the connectivity of both high and low values (Journel and Alabert 1989).

By applying conditional simulation based on the Non-Parametric Stochastic Continuum, several equiprobable realizations may be generated, all honouring the initial data n. If features (eg. connected bands of high hydraulic conductivity) reoccur consistently in these realizations they may be considered as reliable.

Models of spatial dependence have to fulfil two major requirements with regard to the resulting reconstructed realizations (Journel and Alabert 1989):

- sample values as well as qualitative information must be honoured at their location.

- patterns of connectivity in space, especially those of extreme values must be honoured.

These requirements are met by the non-parametric formalism and simulation procedure presented in the preceding text. Conventional interpolation techniques, eg. kriging, tend to smooth or distort out tendencies towards organized (low entropy) patterns.

3.3.1 Indicator formalism

The indicator transform of the spatially distributed attribute Z(u) is defined as the binary variable:

$$i[Z(u);z] = i(u;z) = \begin{cases} 1, \textit{ if } Z(u) \leq z \\ 0, \textit{ if otherwise} \end{cases}$$

The transform is defined at any location u_j where an attribute (quantitative sample) $Z(u_j)$ is available, providing a complete column of indicator data $i(u_j;z)$ for all possible threshold values z. This is exemplified in Figure 3.1 where; a) attributes Z(u) are shown. In b), c), d) the corresponding binary indicator variables $i(u_j;z)$ are reproduced for three arbitrary threshold values z_k.

The indicator values $i(u;z)$ are interpreted as being outcomes of indicator random variables $I(u;z)$. $I(u;z)$ is characterized by one single parameter, its expected value, which is the marginal distribution of $Z(u)$ for the threshold value z.

$$E\{I(u;z)\} = 1 \times P\{Z(u)\leq z\} + 0 \times P\{Z(u)>z\} = F(z \,|\, (n))$$

Stationarity yields that $F(z)$ is independent of location u and may be obtained from spatial averaging of the indicator values. A further step is that the spatial correlation may be included in that ordinary kriging (OK) is used at each threshold value z_k, using the respective indicator variogram models $\gamma_I(z_k,h)$, to infer an estimate of the conditional probability $P[Z(u)\leq z \,|\, (n)]$.

The structure of spatial correlation may be clearly different for different threshold values. The covariance of the attribute Z can be shown to be the average of all indicator covariances, but underestimates the correlation of lows and overestimates that of highs. Thus, if the detection of extreme values is critical to the problem at hand the attribute covariance is inadequate. Correspondingly, if a Gaussian Z-bivariate model is employed it can be shown that spatial (indicator) correlation at

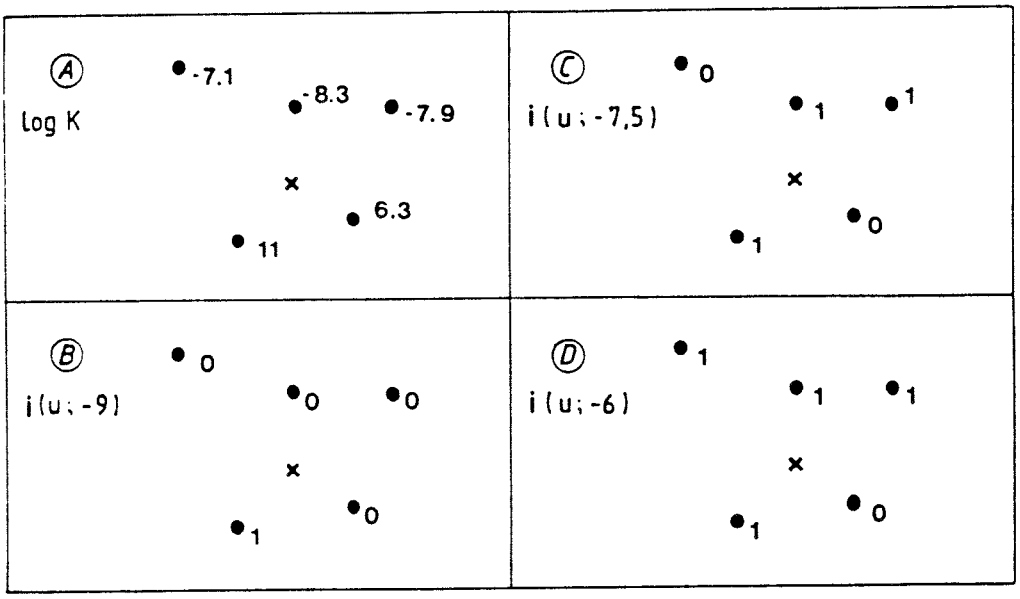


Figure 3.1 Exemplification of indicator coding of (a) hard datum data Z(u) (=log K), (b) indicator at threshold z=-9 , (c) indicator at threshold z=-7.5, and (d) indicator at threshold z=-6.0

extreme values is not accounted for.

The uncertainty in the studied parameter is analysed in terms of probability intervals rather than confidence intervals which is inherent in a model oriented approach. The following exemplifies three such uncertainty relations expressed in terms of probabilities:

$$P[U \in]a,b]|(n)] = F(b|(n)) - F(a|(n))$$
$$P[U \preceq q_p|(n) = p$$
$$P[U \succ t|(n)] = 1 - F(t|(n))$$

3.3.2 Sequential indicator simulation

The sequential indicator simulation scheme (SIS) differs from the Parametric Continuum scheme (turning bands method, TUBA) not only by the fact that multiple covariances (variograms) at different thresholds are used. In addition the Monte Carlo sequence is constructed by informing the modelled domain according to a random walk strategy, whereby each simulated point is informed with a posterior conditional distribution which is consistent with the indicator statistics and geostatistics inferred from the sample data set and the indicator columns at the conditioning points. The simulation scheme is as follows;

1) Introduce a grid with N nodes u_j over which the spatial variability should be modelled.

2) Starting at <u>any</u> node, say u_1, derive the conditional distribution of $Z(u_1)$ by solving the ordinary kriging equation for each threshold value k given all available information (n). The known information (n) is usually known outcomes $Z(u_\alpha) = z_\alpha$ at locations u_α $\{\alpha \in (n)\}$. The conditional or posterior distribution of $Z(u_1)$ is denoted;

$$P\{Z(u_1 \leq z_k | (n)\} = E\{I(;z_k) | (n)\},\ k = 1, \ldots, K$$

3) Draw a realization of $Z(u_1)$ from the distribution. Transfer the realization into the data set which now is of dimension $(n+1)$.

4) Move randomly to a second node u_2. Derive the conditional distribution of $Z(u_2)$ given the available information $(n+1)$, compare 2). Draw a realization of $Z(u_2)$ from the distribution, enter it into the data set which now has dimension $(n+2)$.

5) Loop through all N nodes until they are informed with a simulated value z_j {j=1,...,N}. This set of N values represent one realization of the random field $\{Z(u),\ u \in S\}$. Any number L of such simulations may be obtained by simply repeating the process L times.

3.3.3 Incorporation of soft information

In the above-mentioned simulation scheme, only the hard (quantitative) data Z(u) have been considered. These data, often very sparsely available, may be complemented by various soft (qualitative) information. Coded soft information may readily be entered into the outlined simulation scheme.

Two types of soft information are indicated below;

1) A constraint interval which generates an incomplete indicator column.

$$Z(u) \in [a.b] \Rightarrow i(u;z) = \begin{cases} 0, \text{ if } z \prec a \\ 1, \text{ if } z \succeq b \\ \text{undefined otherwise} \end{cases}$$

2) Local prior distribution coded as "fuzzy" indicator data.

$$Z(u) \in [a,b] \Rightarrow i(u;z) = \begin{cases} 0, \text{ if } z \prec a \\ \in [0,1], \text{ if } z \in [a,b] \\ 1, \text{ if } z \succeq b \end{cases}$$

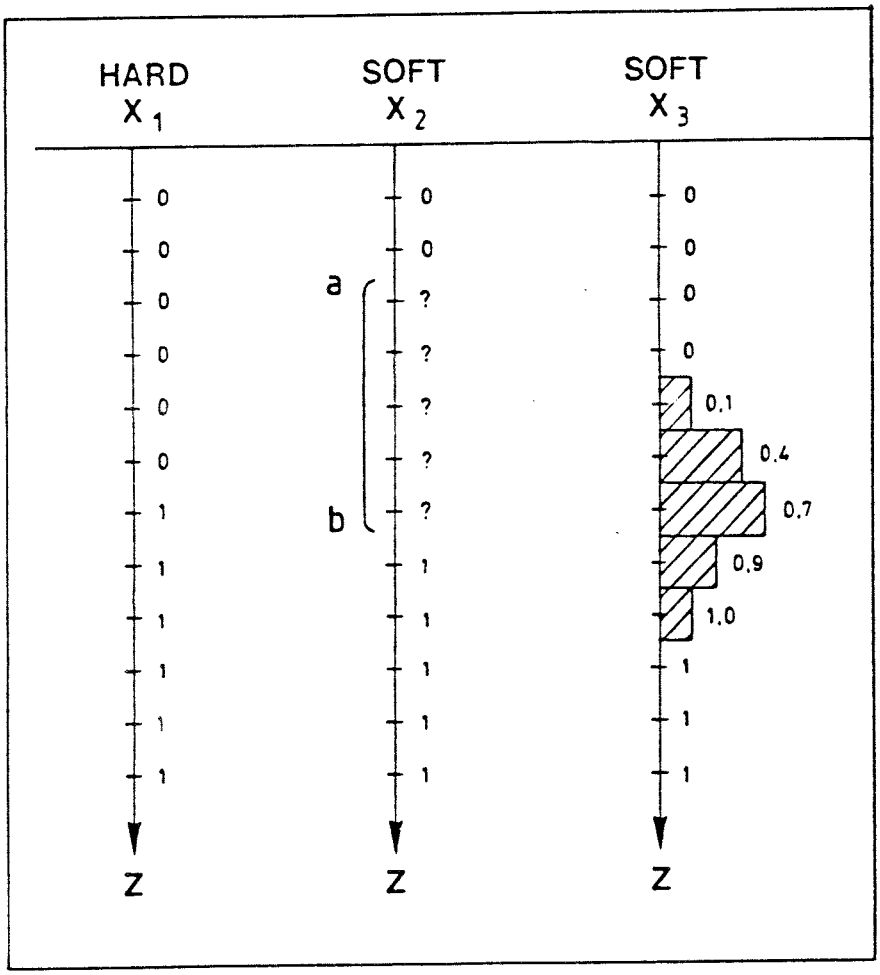


Figure 3.2 Exemplification of indicator column of (a) hard data, (b) soft data (constraint interval, and (c) soft data (prior local distribution).

These soft data with associated coding could eg. be information about rock type (type 1) or a plausible distribution of Z-values deduced for a particular rock type (type 2). In this study only type 1 soft data are used, cf. Chapter 5.

A hard indicator column is presented together with indicator columns of the two types of soft information in Figure 3.2.

4. EXHAUSTIVE DATA SET

Our ambition within this study has been to produce a reference exhaustive data set with a close correspondence to what would be expected in nature for a crystalline bedrock situation. This premise is clearly presumptuous since our problem is that we simply do not know the exhaustive reality of our geological system. However, on the basis of our notion of typical geological and tectonic features and controls of the hydraulic conductivity of crystalline environments plus our understanding of typical range and statistics of material properties, a best "guesstimate" of a real system has been conceived.

The exhaustive reference data set consists of a 250x250 m heterogeneous field with a support (resolution) of 5x5 m. The field was constructed using the non-parametric sequential indicator simulation scheme (SIS), (Gómez-Hernández and Srivastava 1990).

4.1 Geostatistics of the exhaustive reference field

In the construction of the synthetic exhaustive domain using the SIS algorithm, a certain set of statistics and geostatistics have been built in. Within the realm of possible values, the domain has been conditioned to produce a field with a clearly distinguishable anisotropy and banding in the hydraulic conductivity field. Subsequent to scrutiny of several generated candidate fields, the image in Figure 4.1 was selected on purely subjective grounds for further study.

Table 4.1 Univariate statistics of exhaustive reference hydraulic conductivity field.

Geometric mean K (m/s)	Variance (^{10}logK)
$1.63\ 10^{-6}$	0.9784

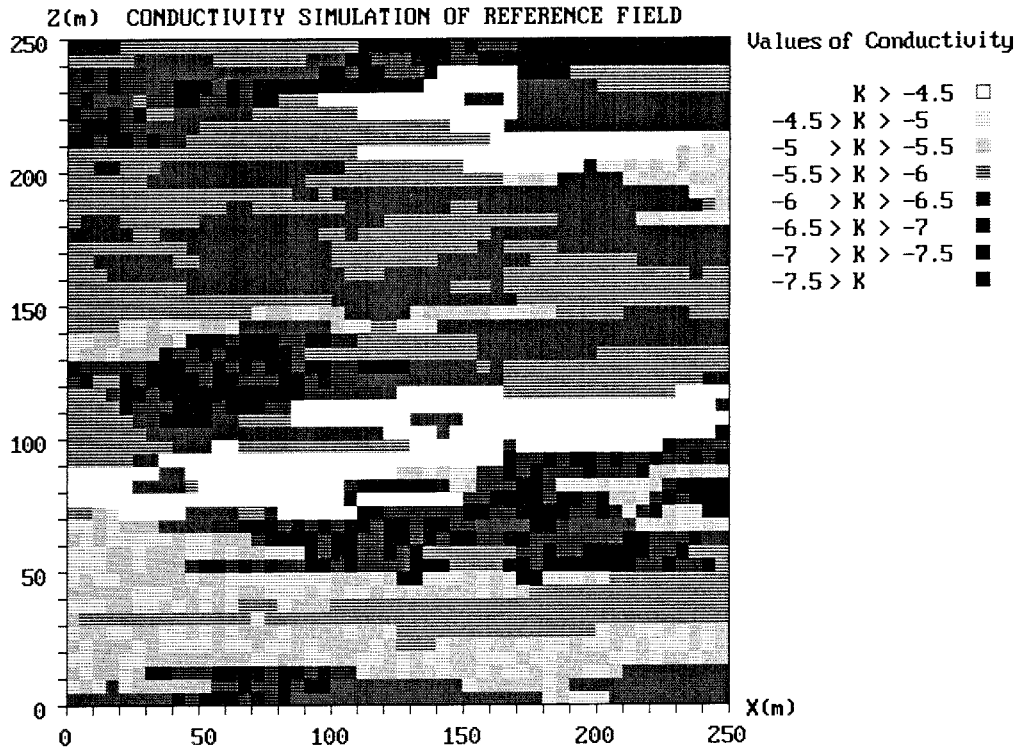


Figure 4.1 Exhaustive reference data set (50x50 cells)

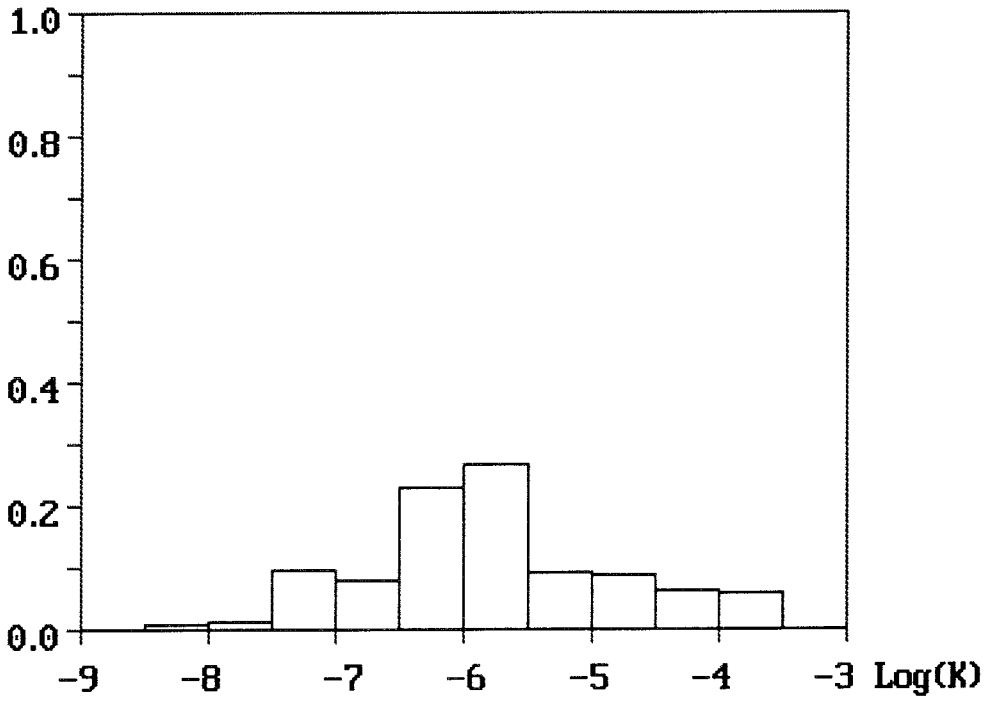


Figure 4.2 Histogram of exhaustive reference data set (N=2500)

Even though the reference exhaustive data set has been constructed with the SIS algorithm, the chosen statistics and the histogram, cf. Table 4.1 and Figure 4.2, imply a near log-normal distribution of hydraulic conductivity. The chosen geometric mean hydraulic conductivity, K_g = 1.6 10^{-6} m/s (log(K) = -5.7885), may be regarded to be fairly high, representative of the upper 100m of a crystalline rock mass. The selected variance (σ^2=0.97845) may be regarded to be low in relation to the geometric mean.

The experimental variogram in the X- and Y- direction based on the attribute (log K) values and fitted exponential models are shown in Figure 4.3. The variogram model has one single structure and no nugget effect. The correlation scales in the X- and Y-direction are 42.5 and 7.5 m, respectively, which implies an anisotropy ratio of 5.7.

Six experimental indicator variograms have been used to describe the geostatistical features of the exhaustive reference data set. Indicator variogram models were conceived for threshold at log(K) = -7, -6.5, -6, -5.5, -5.0, and -4.5. The indicator variograms for the reference field in the X- and Y- direction together with their fitted models are shown in Appendix 1. Although the experimental variograms of thresholds between log(K)= -7 to -5 show some tendency of nugget effects, the fitted indicator variogram models have one single exponential structure as is the case for the attribute variogram. The variograms have a structural anisotropy with a factor 4 to 7 longer correlation scale (range) *a* in the X- than in the Y-direction. The highest anisotropy ratio *R* is found for the highest thresholds and viceversa. Specifics on the experimental variogram models are presented in Table 4.2.

Table 4.2 Parameters of the input indicators variograms in the X- and Y-directions used to create the exhaustive reference data set

	X-direction		Y-direction		
Threshold	Sill	a (m)	Sill	a (m)	R
-7.0	0.140	22.5	0.14	5.0	4.4
-6.5	0.159	50.0	0.159	10.0	5.0
-6.0	0.238	35.0	0.238	5.0	7.0
-5.5	0.228	50.0	0.228	7.5	6.7
-5.0	0.170	30.0	0.170	5.0	6.0
-4.5	0.150	50.0	0.150	7.5	6.7

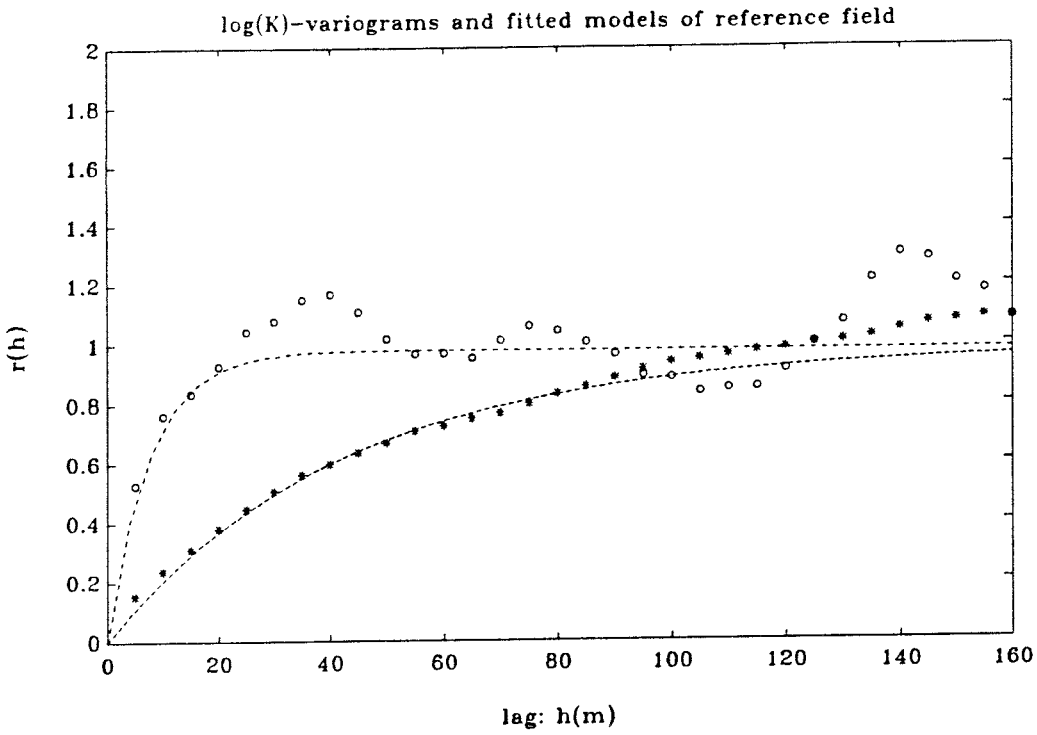


Figure 4.3 Reference data set - Attribute variograms of log(K), (*) = X-direction, (o) = Y-direction.

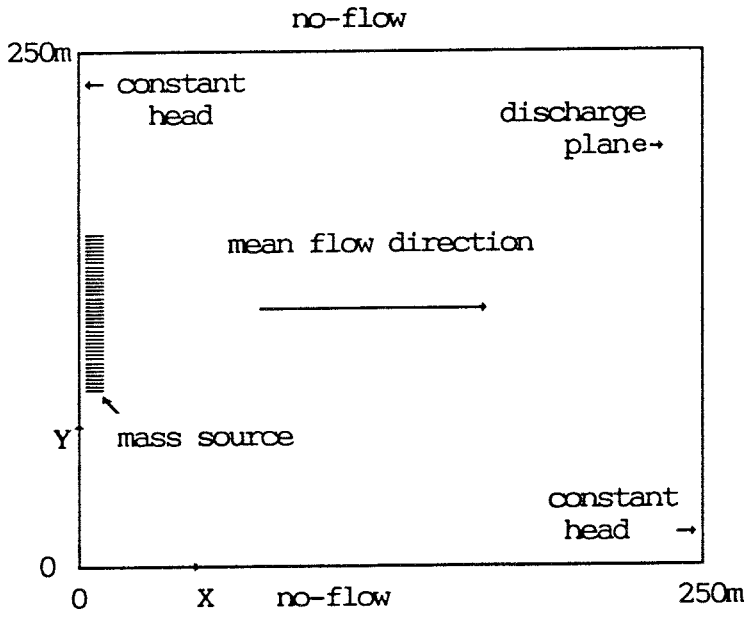


Figure 4.4 Schematic of boundary conditions applied to the studied domain

4.2 Flow and transport simulation of the reference field

The flow and mass transport calculations based on the exhaustive reference field have been simulated under steady-state conditions. The domain is a quadratic area consisting of 50x50 cells with a total domain size of 250x250 m. The upper and lower boundaries are assumed to be no-flow boundaries, whereas the left and right boundaries are constant head boundaries providing a constant head gradient parallel to the X-axis. A schematic diagram of the utilized boundary conditions is shown in Figure 4.4.

The groundwater flow is calculated using a block centred finite difference scheme employing a direct band elimination solver. The harmonic mean is used to average the hydraulic conductivity in neighbouring cells in the calculation. The velocity field is calculated from the head distribution according to Darcy's law. For the sake of simplicity, the porosity of the model domain is assumed constant to ensure that the ratio between gradient and porosity (I/ϕ) is equal to unity.

The transport calculation is then carried out using the output velocity field and a particle tracking routine. The particles are assumed to be conservative, implying that no mass exchange caused by chemical reaction, sorption-desorption, or decay. Since the transport is dominated by advection, local scale dispersion is neglected in the simulations.

At time t=0, 1000 particles are instantaneously released along a segment of the upstream (left) boundary defined by the coordinates (0,100) and (0,150), cf. Figure 4.4. The number of injected particles in different cells is proportional to the flux on the block faces, representing the steady state injection situation. The particle arrival time and cumulative arrival time at the outlet boundary (right) at x=250m is calculated. The distribution of particle arrival (position) along the outlet plane is also recorded to demonstrate the relative probability for a given particle arrival location.

Figure 4.5a shows the cumulative particle breakthrough curve at the outlet plane and Figure 4.5b shows the distribution of particle arrival locations on the outlet plane for the reference field. The results show that the breakthrough curve has a very high peak value with a long tail. Most particles are conveyed in two flow channels parallel to the X-axis which discharge at different locations.

The characteristics of flow and transport of the reference field is thus a distinct two channel flow system with one single high peak, long-tailed breakthrough curve. The observed characteristics are considered to be the real system response of our reference data set to be compared with the Monte-Carlo simulation results which are based on a much more limited amount of information sampled from the exhaustive reference data set.

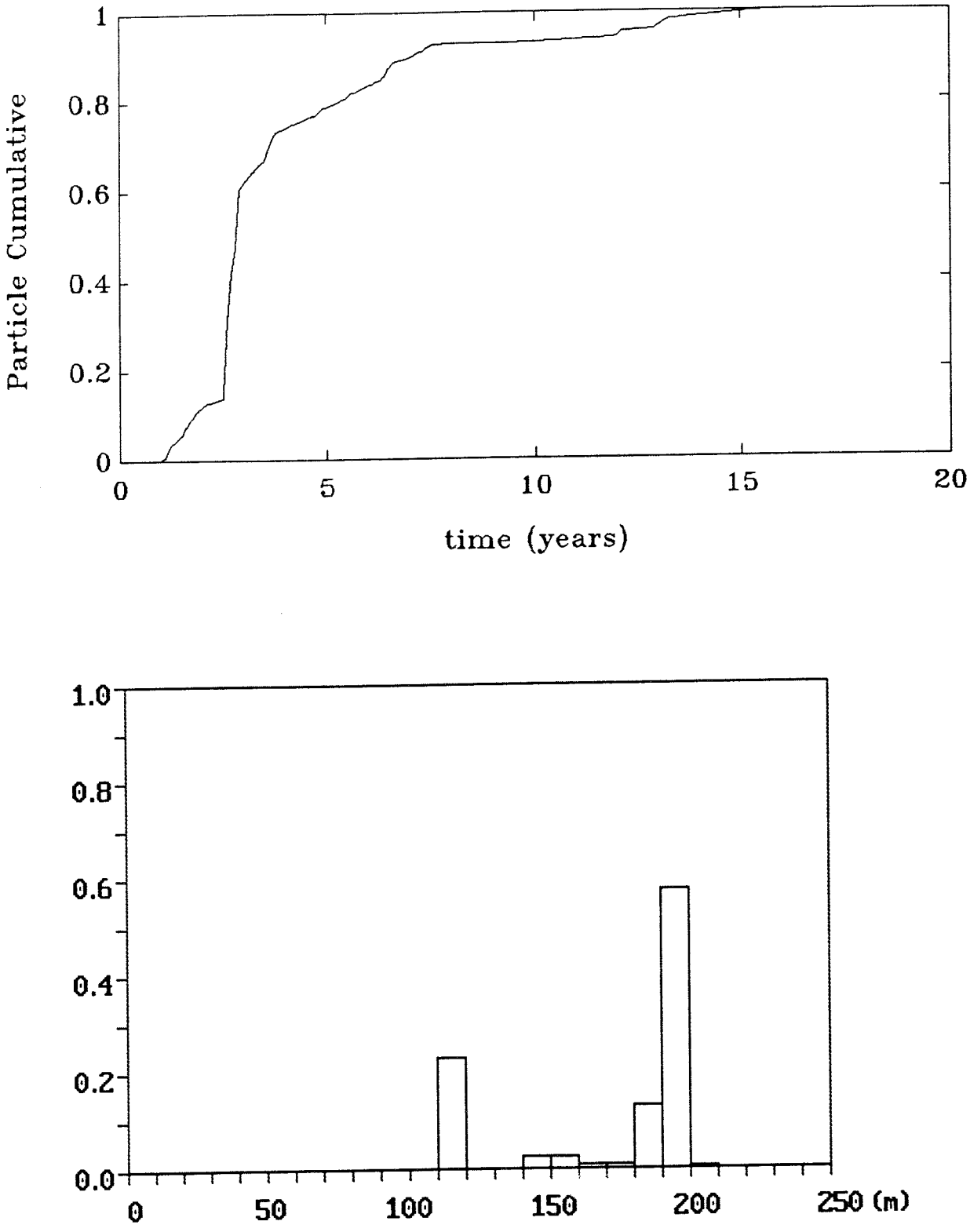


Figure 4.5 Exhaustive reference data set. a) Cumulative breakthrough curve of particle arrival along outlet plane. b) Distribution of particle arrival location along outlet plane.

5. SIMULATION STRATEGY

In order to perform a numerical simulation of groundwater flow and transport in a subsurface system, the distribution of hydraulic conductivity in the analysed system has to be known. In the study of a real field situation, due to time and budget constraints, information of a heterogeneous hydraulic conductivity field is only available in a few locations. The statistical and geostatistical characteristics inferred about the studied domain, and thus the subsequent engineering design and decision making made, depends heavily on this limited information.

In practice, the hydraulic conductivity data is obtained through some measurement layout which is based on the experience of the hydrogeologist and the available geological and geophysical data. In this study we assume two kinds of measurement layouts. The first uses *random selection* of 25 hard (1% of the exhaustive data set) and 100 soft (less precise) which has an uncertainty range (one order of magnitude) centred around the "true" value. The second uses *designed sampling* of 25 hard data selected to capture the perceived spreading of particles from the source area on the left boundary. The 100 soft data are in this case spaced on a regular grid over the model domain. The position of the hard and soft data for the two sampling schemes are shown in Figure 5.1.

In performing the simulations the basic assumption of statistical homogeneity is adhered, ie. the studied stochastic process is assumed to be stationary and ergodic.

5.1 Random sampling

The cumulative distribution of the 25 randomly sampled hard data superimposed on the cumulative distribution of the 2500 data of the exhaustive reference field is shown in Figure 5.2a. The distribution of the random sample data set is found to be very close to that of the reference field. The mean of log(K) is -5.84 which is only about 1% less the mean of reference field. The variance of log(K) is 1.08 which is about 10% higher than the variance of the reference field.

The cumulative distribution of 25 hard plus 100 soft data are also shown in Figure 5.2. From the figure it is evident that the addition of 100 soft data does not improve the quality of the distribution of the random sample data set. Actually, it is slightly worse since the 25 hard data already well represent the reference data field from a statistical point of view.

Due to the limited amount of data in the random sample data set, it was found to be very difficult to construct experimental variograms (attribute and indicator), especially so for the anisotropic situation to be simulated. Hence, in order to enable simulation, the exponential variogram models and correlation scales inferred for the reference field have also been attributed to the random sample data set, though with sample statistics, ie. variance (sill) obtained from the random data set itself.

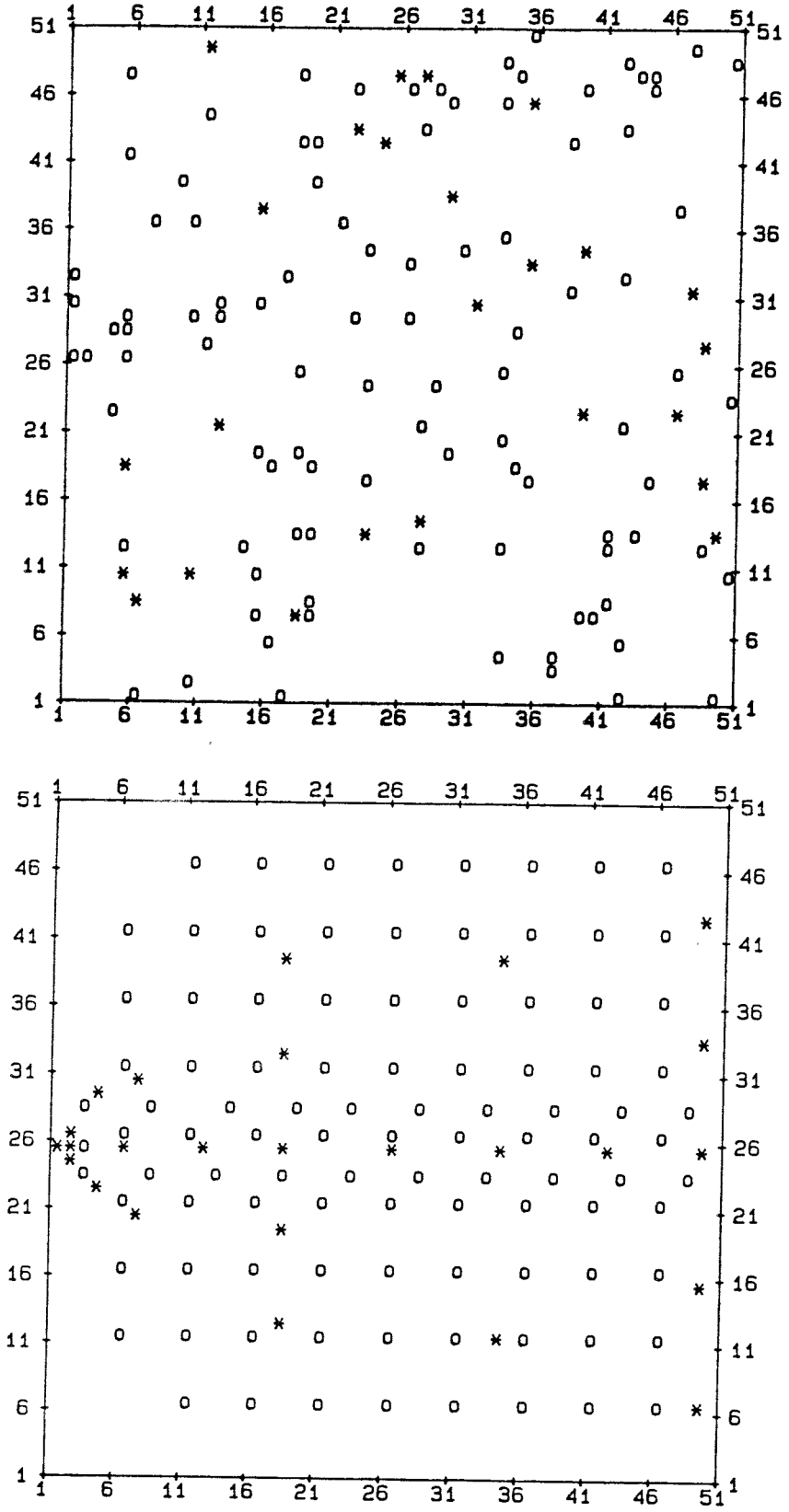


Figure 5.1 Description of the location of the 25 hard data sample and 100 soft data location as distributed on the reference field, a) Random sampling, b) Designed sampling, (*) = hard sample, (o) soft sample.

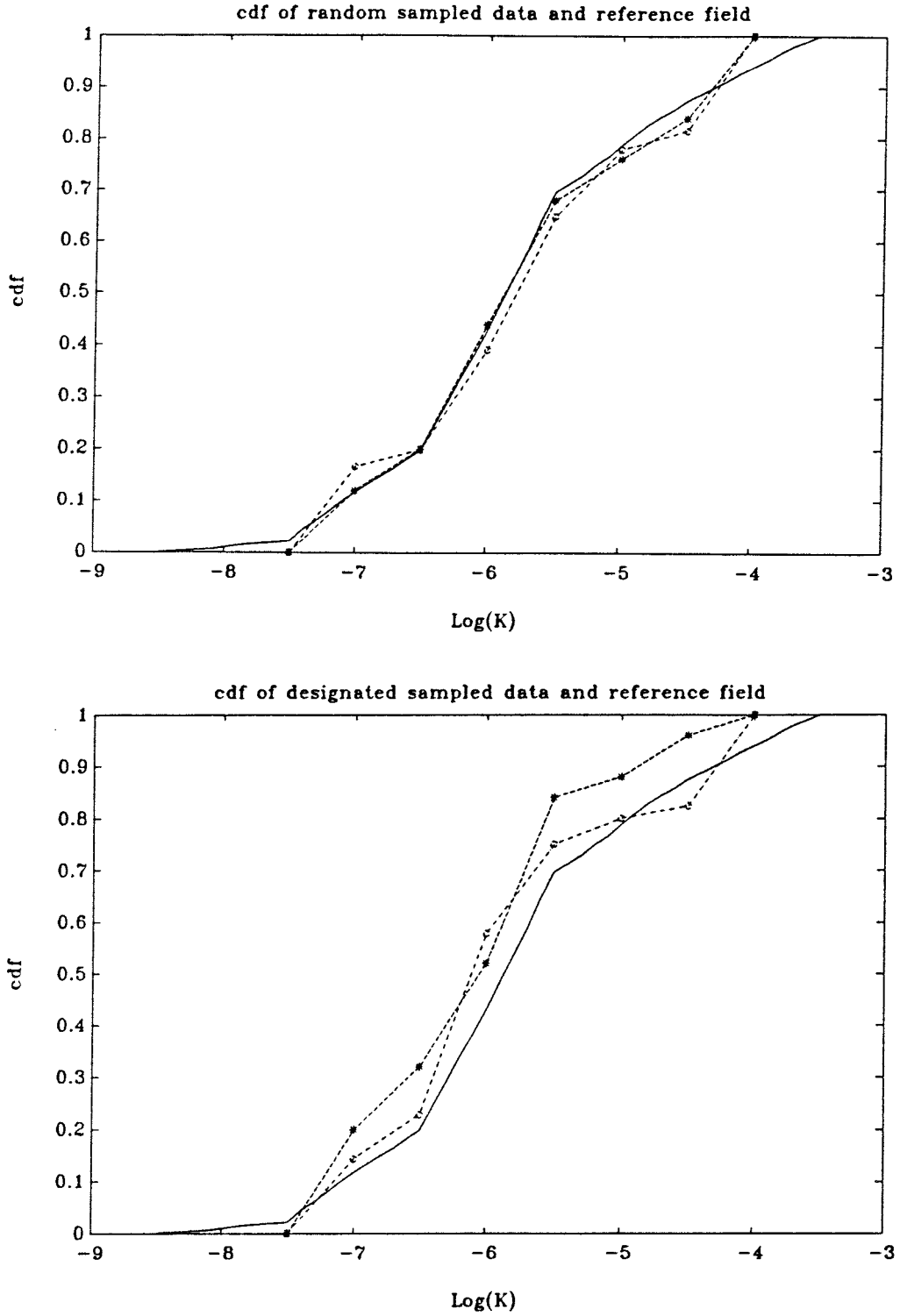


Figure 5.2 Comparison between the cumulative distribution of data in the sample data sets and that of the data in the exhaustive reference data set, a) Random sampling, b) Designed sampling, (*) = 25 hard data, (o) = 25 hard plus 100 soft data.

5.2 Designed sampling

The cumulative distribution of the designed sample data set of 25 hard data superimposed on the cumulative distribution of the reference field are shown in Figure 5.3. The mean of log(K) is -6.14 and the variance is 0.683, 6% and 30% lower than the corresponding statistics of the reference field, respectively. The statistics of the designed hard data sample data set does not adequately represent the reference field because the designed sampling plan puts more weight on the region near the source area.

The cumulative distribution of designed sampled 25 hard and 100 soft data is also shown in Figure 5.2b. The addition of 100 regularly spaced soft data improve the statistics of the design sample data set. Correspondingly to the case of the random sample data set, the corresponding variogram models are identical to those of the reference field, though scaled to the sample statistics of the designed sample data set.

6. STOCHASTIC SIMULATION

In this chapter a description of the simulation procedure is presented for the two approaches. If not stated otherwise, 100 realizations have been generated in each case. The use of the different sample sets is described, as is the incorporation of soft information in the case of the non-parametric continuum approach. The flow and transport calculations follow the same procedure described in Section 4.2. The results are discussed in terms of visual appearance of the generated realizations, comparisons of input and output statistics and geostatistics, mean particle arrival times and a dispersivity index, ensemble particle breakthrough curves and distribution of particles on the outlet boundary. The (tentative) dispersivity index is defined as the coefficient of variation of the mean arrival time. The higher the index, the higher the degree of particle dispersion in the simulated field.

6.1 Parametric continuum simulation

6.1.1 Simulation based on a random sample data set

Using the 25 hard random sample data, Monte-Carlo realisations of the simulated domain have been generated with the parametric approach. The hydraulic conductivity field is on the basis of the sample data set assumed to have a log-normal distribution which is not far off from the reference field, cf. Figures 4.2 and 5.2a.

Three examples of generated equiprobable and statistically consistent hydraulic conductivity realisations are shown in Figure 6.1. From the figure it is obvious that the generated fields do not show the same degree of continuity as the reference field.

The statistics of the generated material property fields show similar statistical characteristics as the reference field. The mean and variance are well reproduced, cf. Table

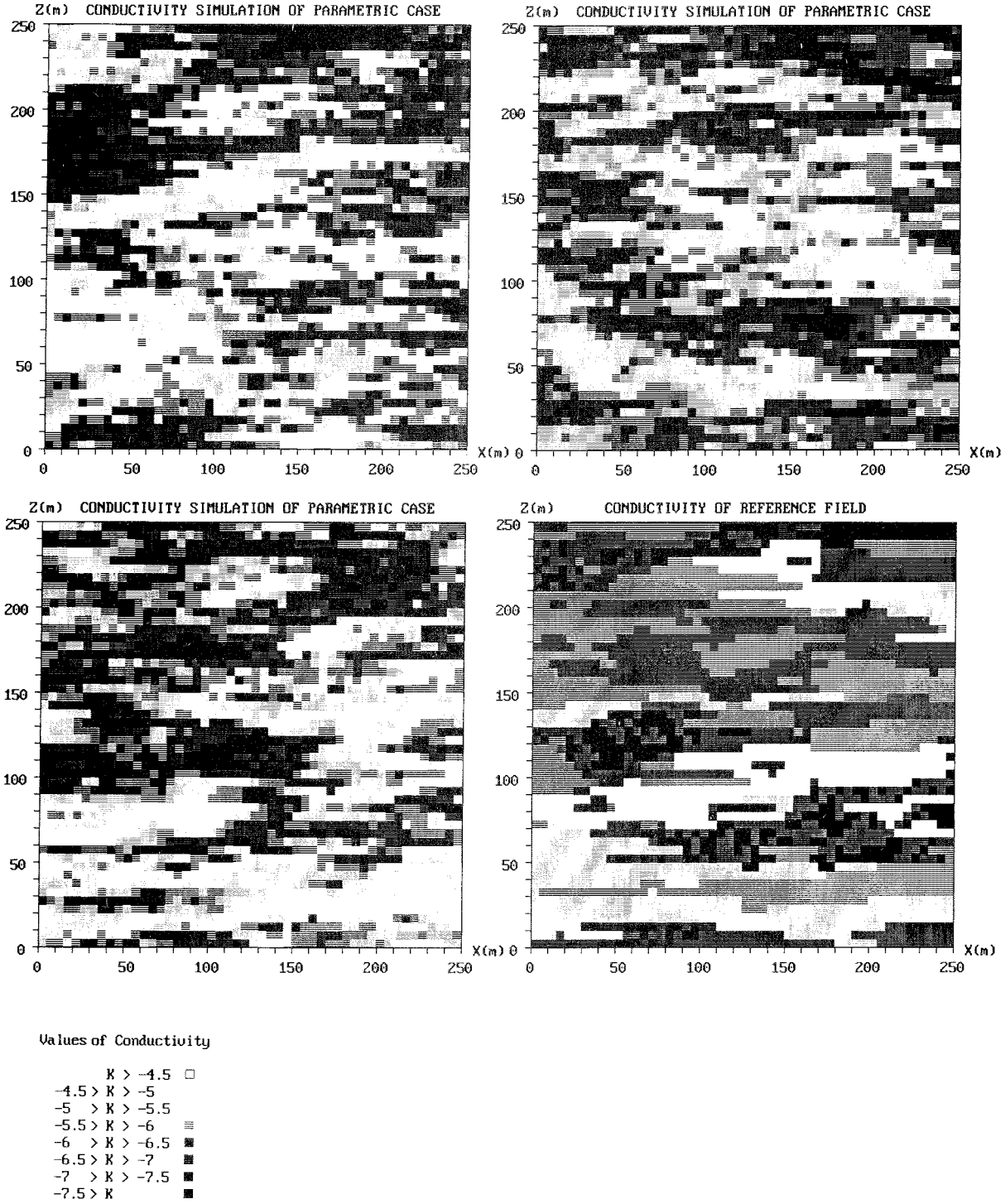


Figure 6.1 Parametric simulation based on 25 hard randomly selected data. Examples of realizations of the hydraulic conductivity field (log(K)), a-c) simulated realizations, d) exhaustive data set.

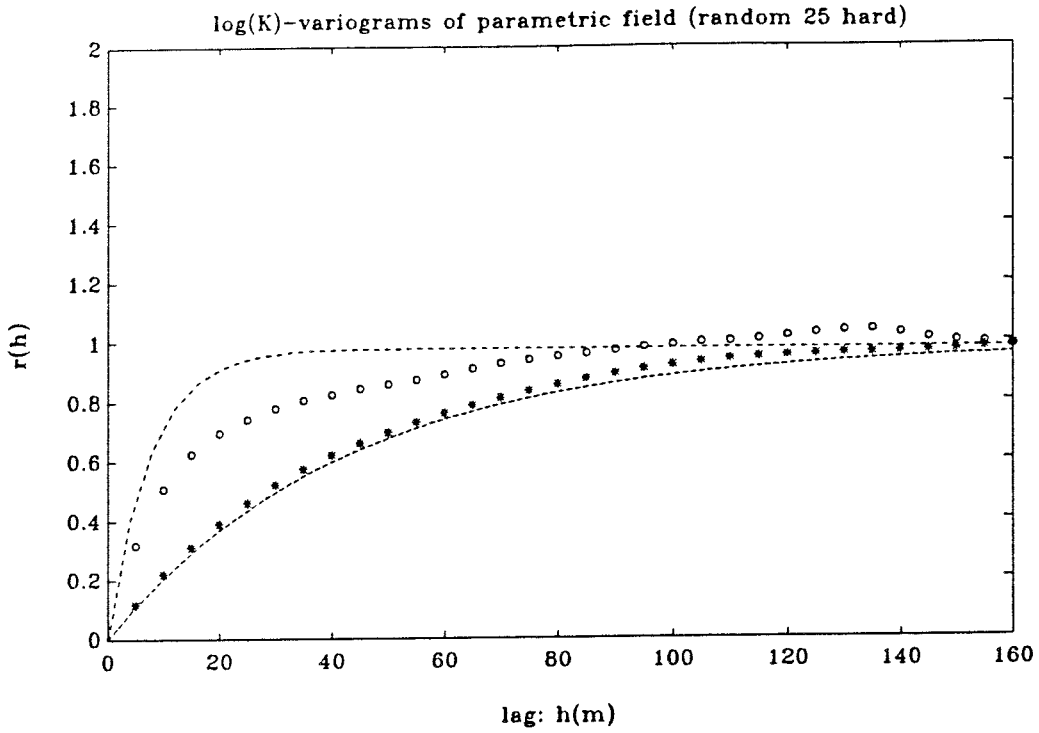


Figure 6.2 Parametric simulation based on 25 hard randomly selected data. Generated ensemble attribute variograms of log(K), (*) = X-direction, (o) = Y-direction, (-----) = reference data set.

6.1. Figure 6.2 show the generated ensemble attribute variogram in the X- and Y-direction together with the corresponding variogram based on the reference field. The correlation length in the X-direction is the same as that of the reference field, whereas in the Y-direction, the generated field exhibits a somewhat longer correlation length than reference field. The noted difference in correlation length in Y-direction may be attributed to a too small number of realisations and/or that the simulation domain is not large enough.

Figure 6.3 shows the resulting ensemble cumulative breakthrough curve of the parametric simulation using the 25 hard random sample data set, plus a one standard deviation envelope of arrival time. The resulting ensemble particle arrival time shows a small deviation from the results based on the reference case, cf. Table 6.2. The dispersivity index based on the ensemble of simulations is 0.72, cf. Table 6.2. Figure 6.4 shows the ensemble distribution of particle arrival locations along the outlet boundary. The results show a single flow channel system which corresponds to the lower flow channel noted for the reference field, cf. Figure 4.5b. The upper flow channel noted for the reference field is not well represented in the simulated fields, plus the fact that the flow path discharge is located much closer to the centre of the outlet plane. In addition, the simulated flow path is more dispersed in the domain compared with the reference field, cf. Figures 4.5b and 6.4.

Table 6.1 Ensemble statistics of generated hydraulic conductivity fields based on 100 realizations. (r) = random sample data set, (d) = designed sample data set.

Simulation case	Mean of log K (m/s)	Variance of log K
Reference field	-5.79	0.98
Para., 25, r	-5.83	1.01
Non-para., 25, r	-5.85	0.89
Non-para., 25+100, r	-5.83	0.98
Para., 25, d	-5.96	0.67
Non-para., 25, d	-6.08	0.73
Non-para., 25+100, d	-5.88	0.92

Figure 6.2 Ensemble statistics of simulated arrival time based on 100 realizations.

Model Case	First Arrival (years)		50% arrival (years)		Mean arrival (years)		Dispersivity index (CoV)	
	Mean	± σ	Mean	± σ	Mean	± σ	Mean	± σ
Ref. field	1.084	-	2.820	-	2.917	-	0.733	-
Random Sampling								
Para., 25	1.195	1.342	2.144	2.670	2.916	4.613	0.716	0.363
Nonp., 25	1.845	2.954	2.728	3.967	3.285	4.261	0.765	0.661
Nonp., 25/100	1.800	1.099	3.518	1.930	4.849	2.169	1.043	0.427
Designed Sampling								
Para., 25	2.810	1.571	4.857	3.003	6.200	3.013	0.687	0.300
Nonp., 25	3.861	2.753	6.738	4.805	8.375	5.318	0.604	0.262
Nonp., 25/100	1.716	1.126	5.467	3.219	6.906	2.792	0.786	0.309

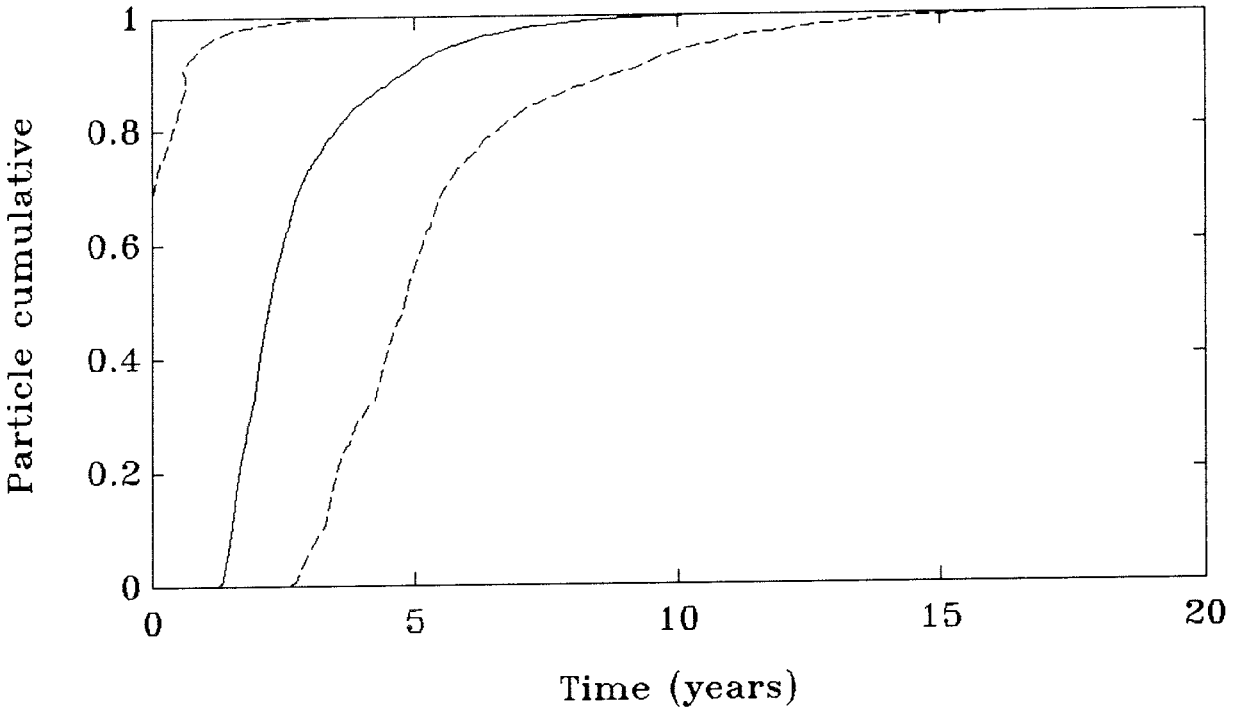


Figure 6.3 Parametric simulation based on 25 hard randomly sampled data. Ensemble cumulative breakthrough curve. Hatched line corresponds to one standard deviation.

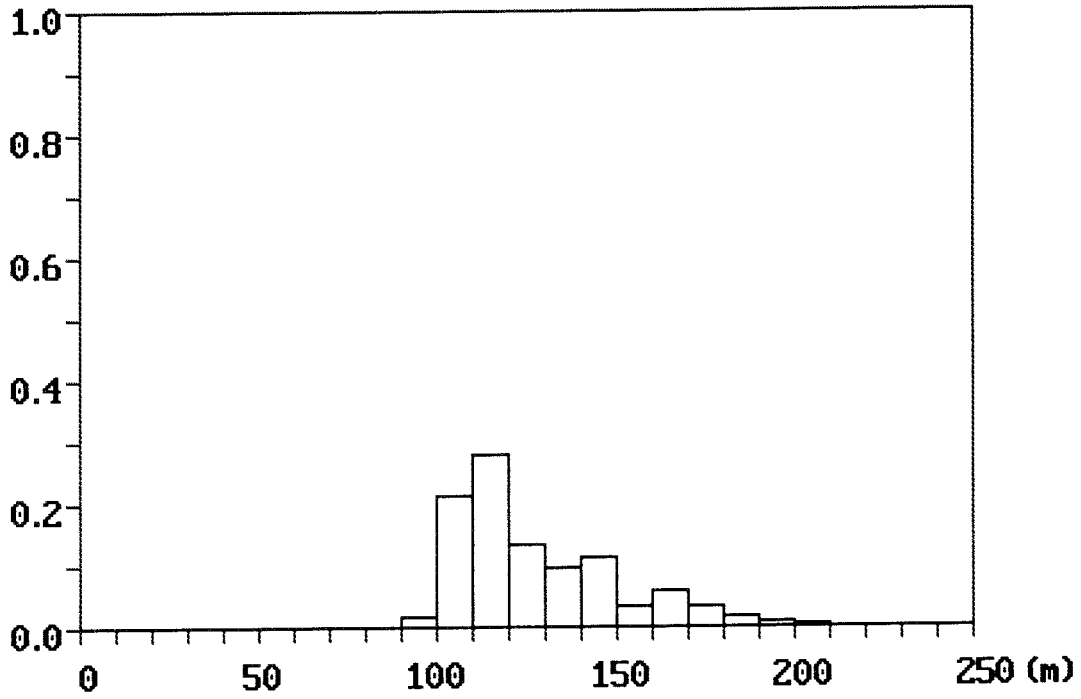


Figure 6.4 Parametric simulation based on 25 hard randomly sampled data. Ensemble particle arrival distribution on the outlet boundary.

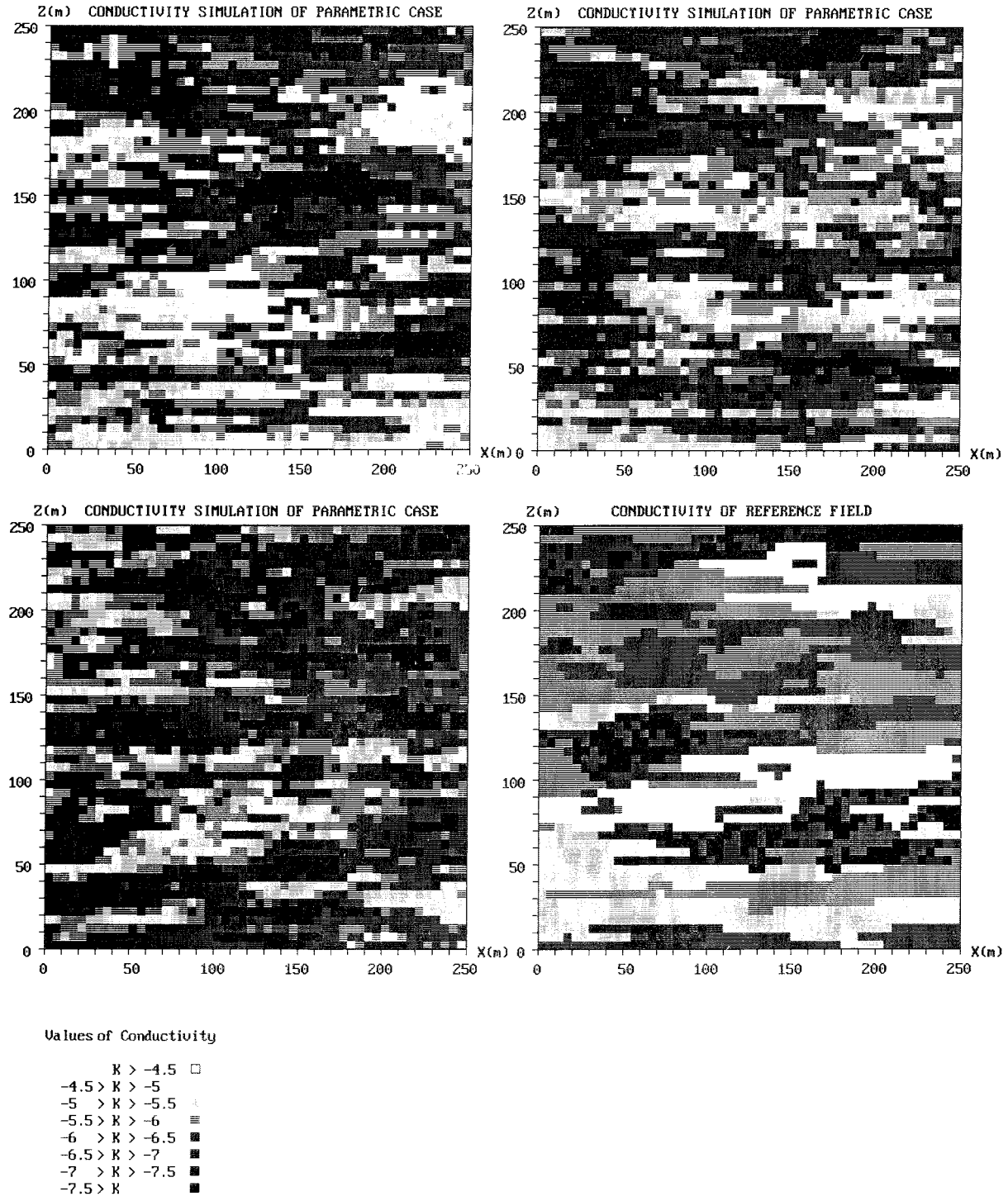


Figure 6.5 Parametric simulation based on 25 hard data obtained from designed sampling. Examples of realization of the hydraulic conductivity field (log(K)), a-c) simulated realizations, d) exhaustive reference data set.

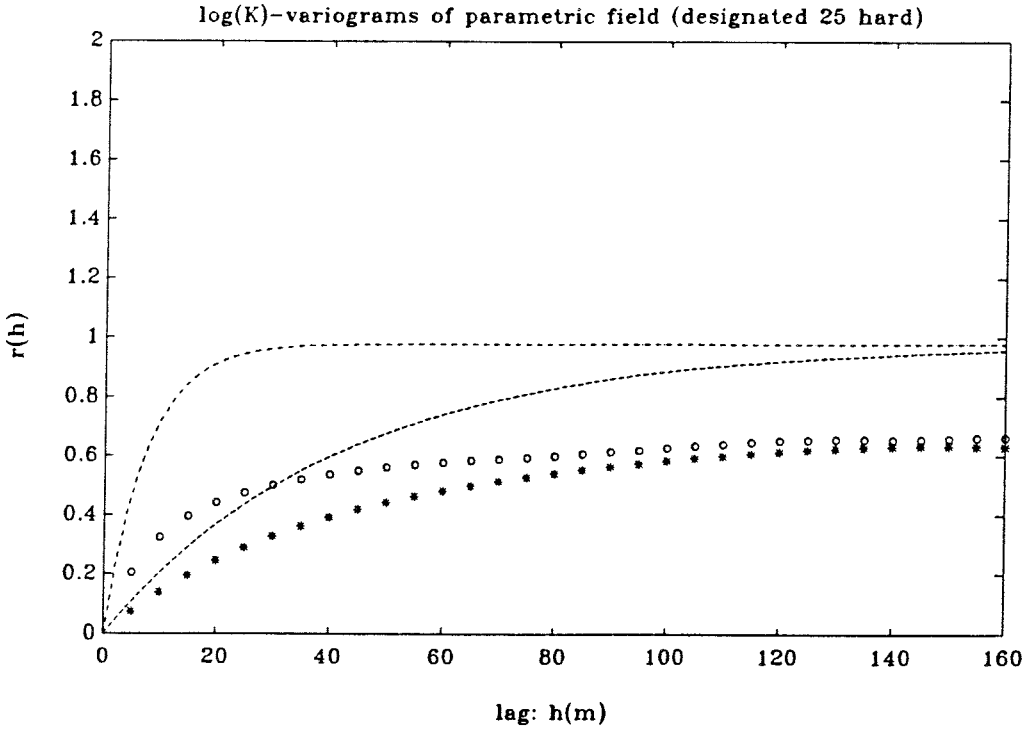


Figure 6.6 Parametric simulation based on 25 hard data obtained from designed sampling. Generated ensemble attribute variogram of log(K). (*) = X-direction, (o) = Y-direction, (-----) = reference data set.

6.1.2 Simulation based on a designed sample data set

Conditioned on the 25 hard data obtained from the designed sampling plan, realizations of the simulated field have been generated with the parametric approach. The hydraulic conductivity field is again assumed to have a log-normal distribution, cf. Figure 5.2b. Three equiprobable realisations are shown in Figure 6.5.

The ensemble mean and variance of generated fields are somewhat lower, but compare well with the statistics of the reference case, cf. Table 6.1. Figure 6.6 shows that the correlation length in the X- and Y-direction compares well with those of the reference field, but the variogram sills are much lower than for the reference field, but compare with the variance of the designed sample data set.

The results of the flow and transport simulation are condensed in the ensemble cumulative breakthrough curve in Figure 6.7, together with one standard deviation envelop of arrival time. The ensemble mean particle arrival time of the parametric simulation based on 25 hard designed sample data is later than for the reference case, cf. Table 6.2. The reason for this later mean arrival time may be due to the fact that the sample mean and variance of Log(K) are lower than for the referenced field. The dispersivity index for the ensemble of simulated fields is lower than for the random sample case described in the previous section, cf. Table 6.2. Figure 6.8 shows the distribution of particle arrival locations along the outlet boundary. The results show two weakly distinguishable flow channel systems, where one corresponds the lower channel in the reference field. The upper flow channel in the reference field is not

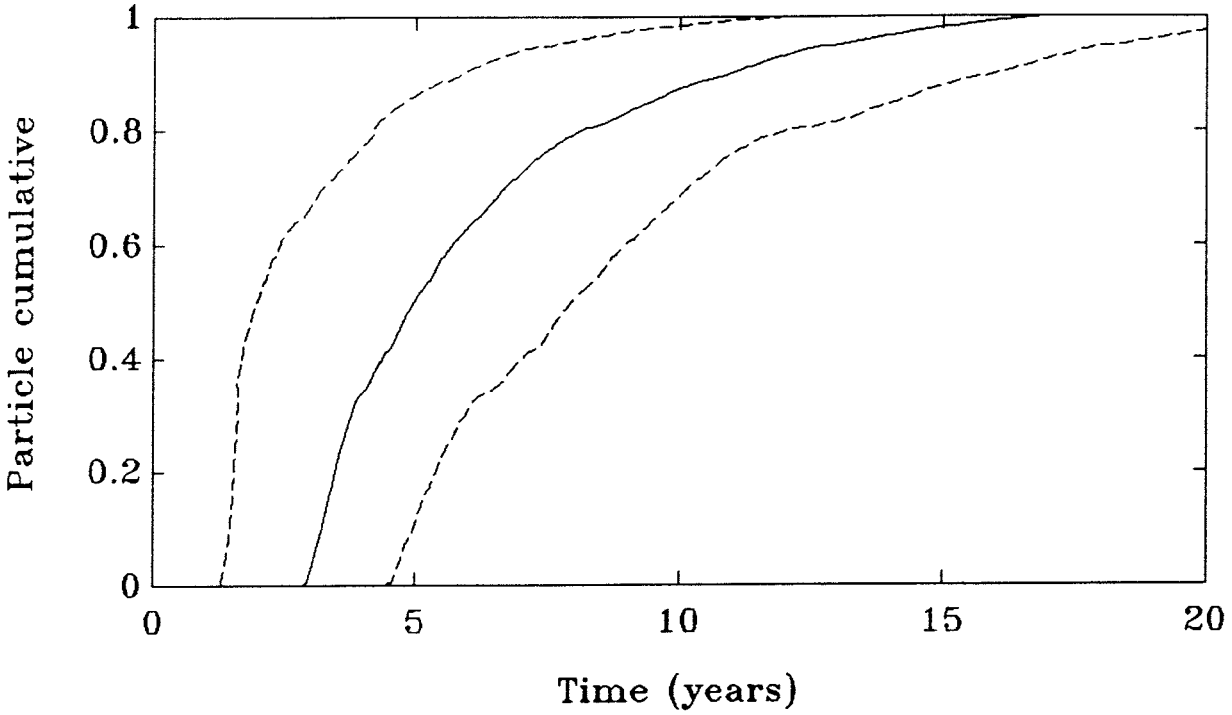


Figure 6.7 Parametric simulation based on 25 hard data obtained from designed sampling. Ensemble cumulative breakthrough curve. Hatched line corresponds to one standard deviation.

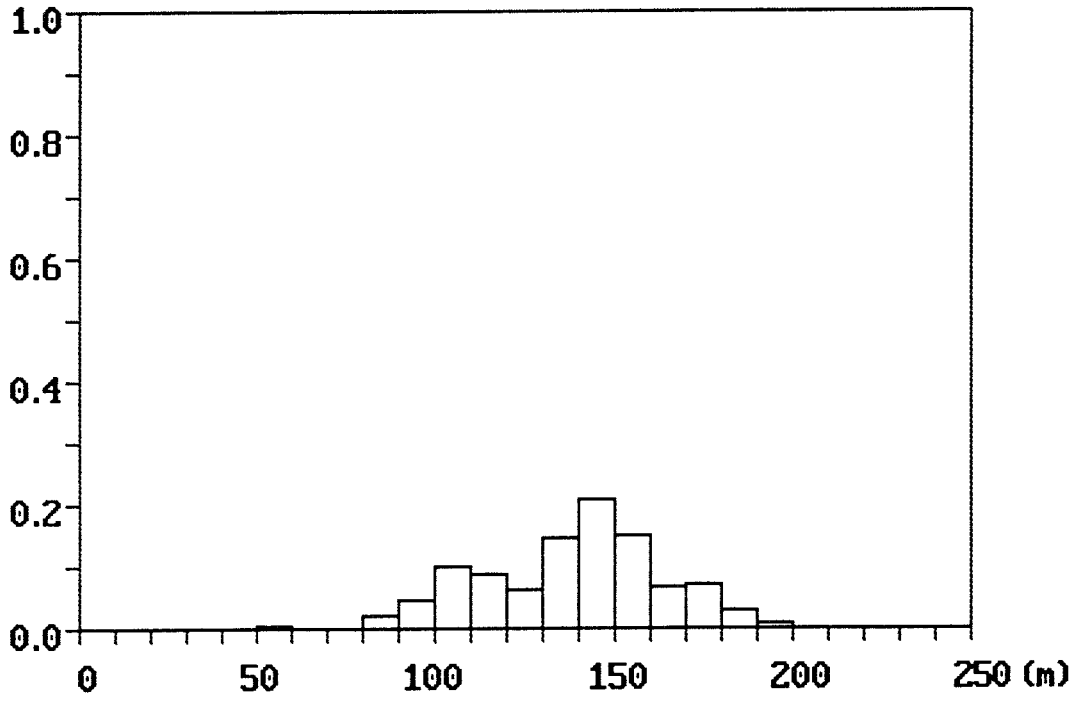


Figure 6.8 Parametric simulation based on 25 hard data obtained from designed sampling. Ensemble particle arrival distribution on the outlet boundary.

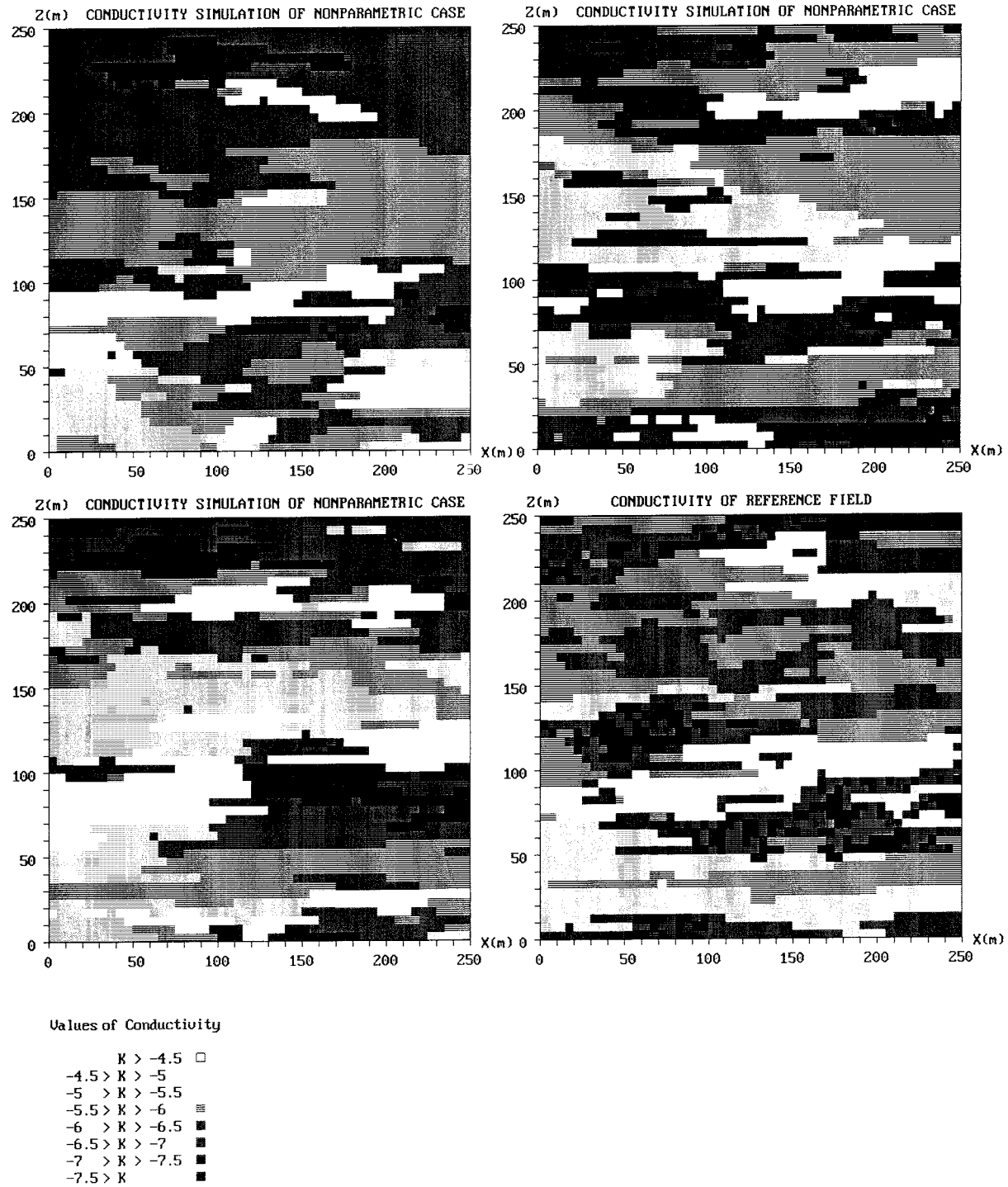


Figure 6.9 Non-parametric simulation based on 25 hard randomly sampled data. Examples of realizations of the hydraulic conductivity field (log(K)), a-c) simulated realizations, d) exhaustive data set.

well represented in the ensemble of simulated fields, where it is located closer to the centre of the outlet plane. The simulated flow paths are more dispersed in the domain compared with the reference field and random sample case described in the previous section, cf. Figure 6.8.

6.2 Non-Parametric continuum simulation

The flow and transport calculations performed on the non-parametric simulations are identical to those performed for the parametric continuum cases. The major difference is that in this section also the effect of incorporating soft information will be studied.

6.2.1 Simulation based on a random sample data set

25 hard data

Based on the 25 randomly sampled hard data, 100 realisations of hydraulic conductivity have been generated with the non-parametric approach. The cumulative distribution curve of the random sample data set is given by Figure 5.2a which has near Gaussian distribution as is the case for the reference field data. Three equiprobable realisations are shown in Figure 6.9. The two bands of high hydraulic conductivity are easily distinguished, and overall the realizations show a higher

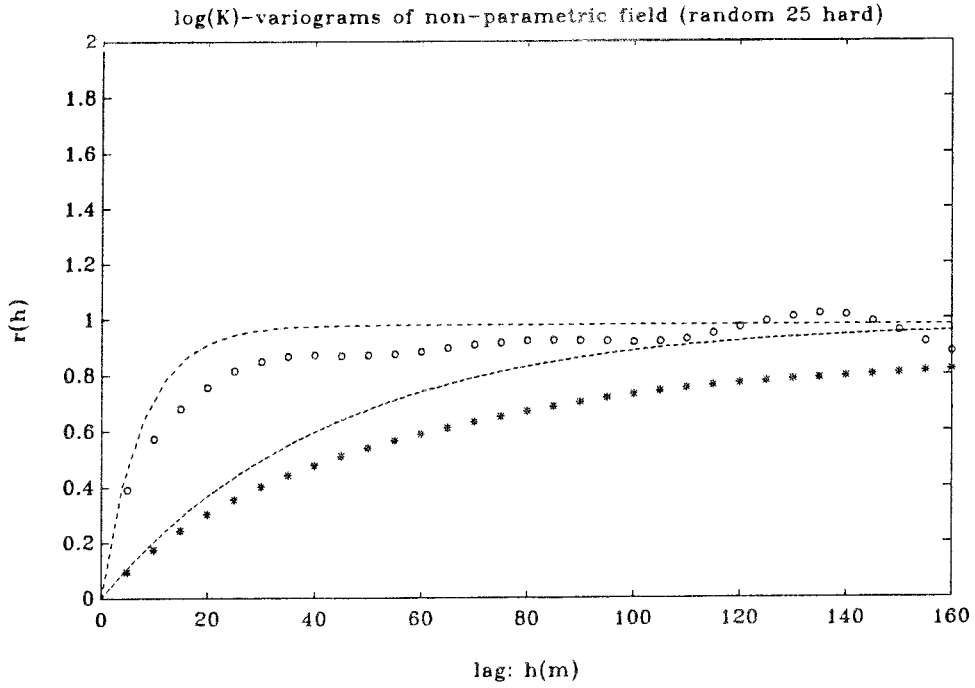


Figure 6.10 Non-parametric simulation based on 25 hard random sample data. Generated ensemble attribute variograms of log(K), (*) = X-direction, (o) = Y-direction, (-----) = reference field.

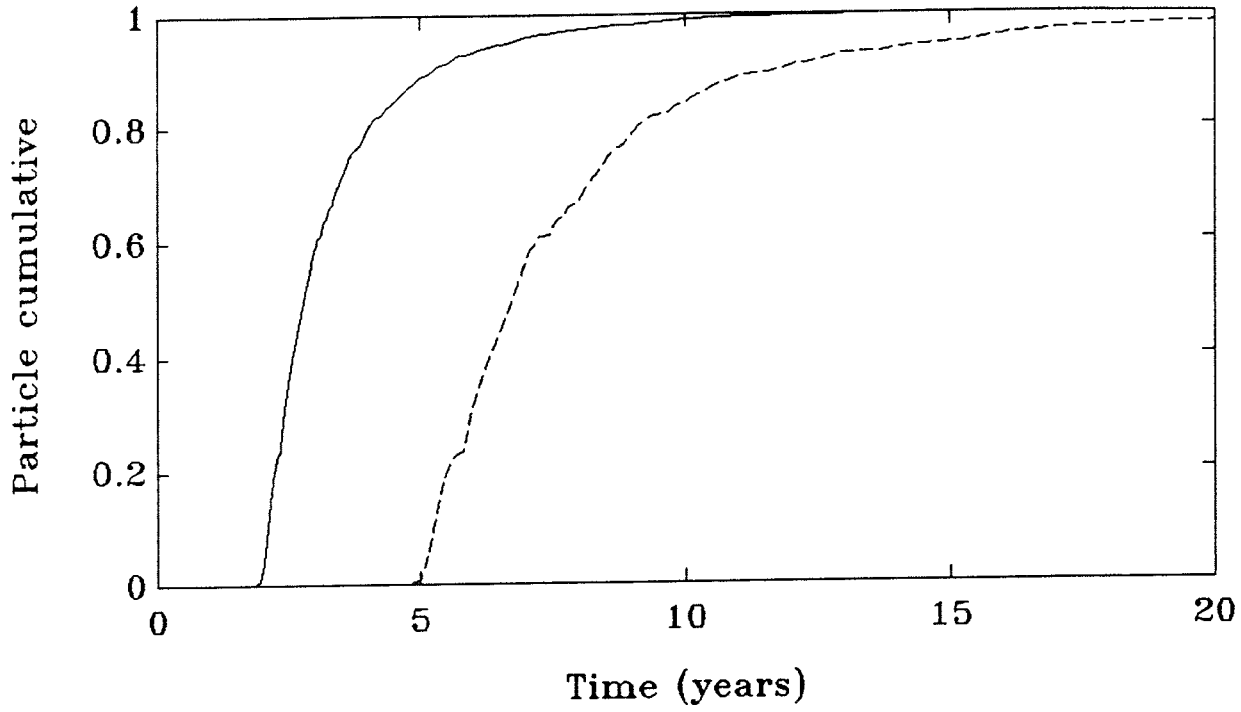


Figure 6.11 Non-parametric simulation based on 25 hard randomly sampled data. Ensemble cumulative breakthrough curve. Hatched line corresponds to one standard deviation.

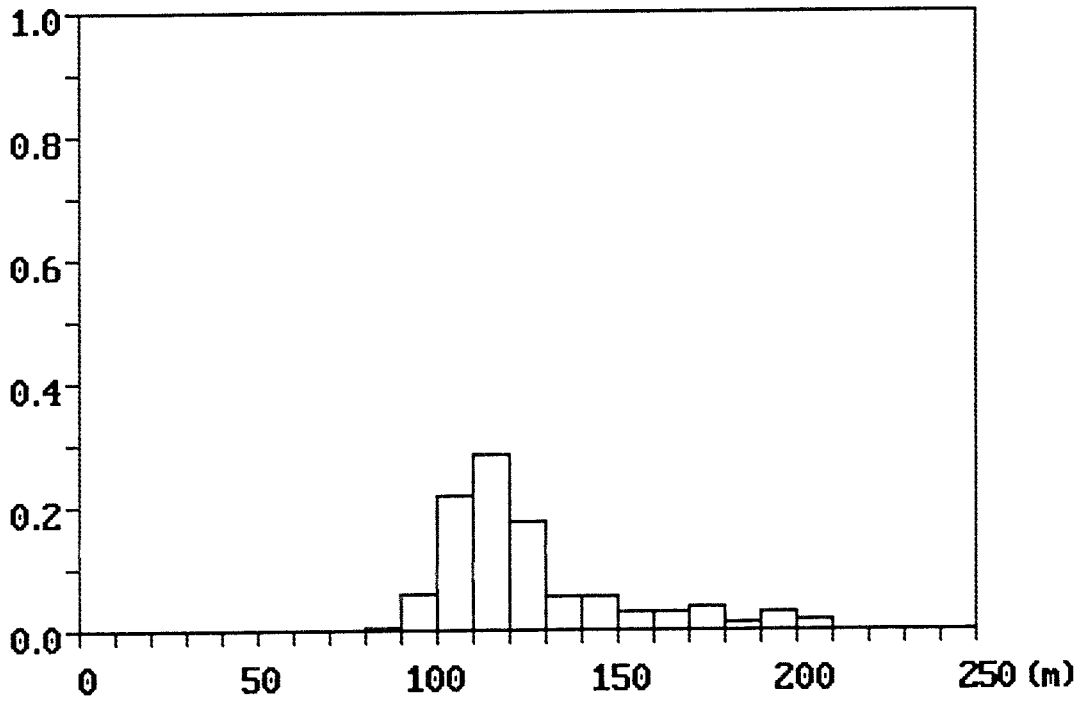


Figure 6.12 Non-parametric simulation based on 25 hard randomly sampled data. Ensemble particle arrival distribution on the outlet boundary.

degree of continuity than the parametric fields shown in the previous section. The statistics of the generated fields are similar to that of the reference field, though the mean and variance are slightly lower, cf. Table 6.1.

The generated ensemble attribute variograms presented in Figure 6.10 show that the correlation lengths in the X- and Y-directions are comparable to those of the reference field though the noted sills are somewhat lower, in accordance with the sample statistics. The ensemble indicator variograms of the generated fields for the different threshold values are shown in Appendix 2. The indicator variances of the realizations in the X- and Y-direction are somewhat lower than the input values, possibly an effect of the fact that the simulation domain is not large enough. The noted differences in the indicator variograms means that the shape of the cumulative distribution of attribute variable is somewhat different than that of the reference field.

The ensemble cumulative breakthrough curve of the non-parametric simulation based on 25 randomly sampled data together with one standard deviation envelop of arrival time is shown in Figure 6.11. A comparison between Figure 6.11 and Figure 4.5a show that the breakthrough curves are similar in shape. The calculated mean arrival time is also more compatible to that of the reference case than the corresponding parametric simulation, cf. Table 6.2. The dispersivity index of the ensemble of simulated fields is somewhat higher than that of the corresponding parametric simulation, cf. Table 6.2. The uncertainty in the calculated mass arrival time, however, is higher for the non-parametric case than for the parametric case. The distribution of particle arrival locations at the outlet boundary is shown in Figure 6.12. The results show one strongly pronounced single flow channel system which corresponds to the lower channel noted for the reference field, cf. Figure 4.5b. The upper channel noted for the reference field is not represented in the ensemble of non-parametric simulations. The simulated flow path centred on the lower flow channel is more dispersed in the simulated domain compared to its appearance in the reference field results.

25 hard plus 100 soft data

Utilizing the random sample data set consisting of 25 hard and 100 soft data, 100 realizations have been generated. Three equiprobable realizations are shown in Figure 6.13. Compared to the 25 hard data case, the visual correspondence with the reference field appears to be improved. The statistics of generated fields corresponds well with that of the reference field, cf. Table 6.1. The generated ensemble attribute variogram shown in Figure 6.14 shows that the correlation lengths in the X- and Y-direction and the corresponding sills of the variograms are comparable to those of the reference field.

The ensemble indicator variograms of the generated fields presented in Appendix 3 shows that the indicator variances in both the X- and Y- direction are lower than the inputed values for higher threshold values, possibly a result of the simulation domain being too small. The difference between the input and output indicator variograms implies that the shape of the cumulative distribution of the attribute variable is different than for the reference field.

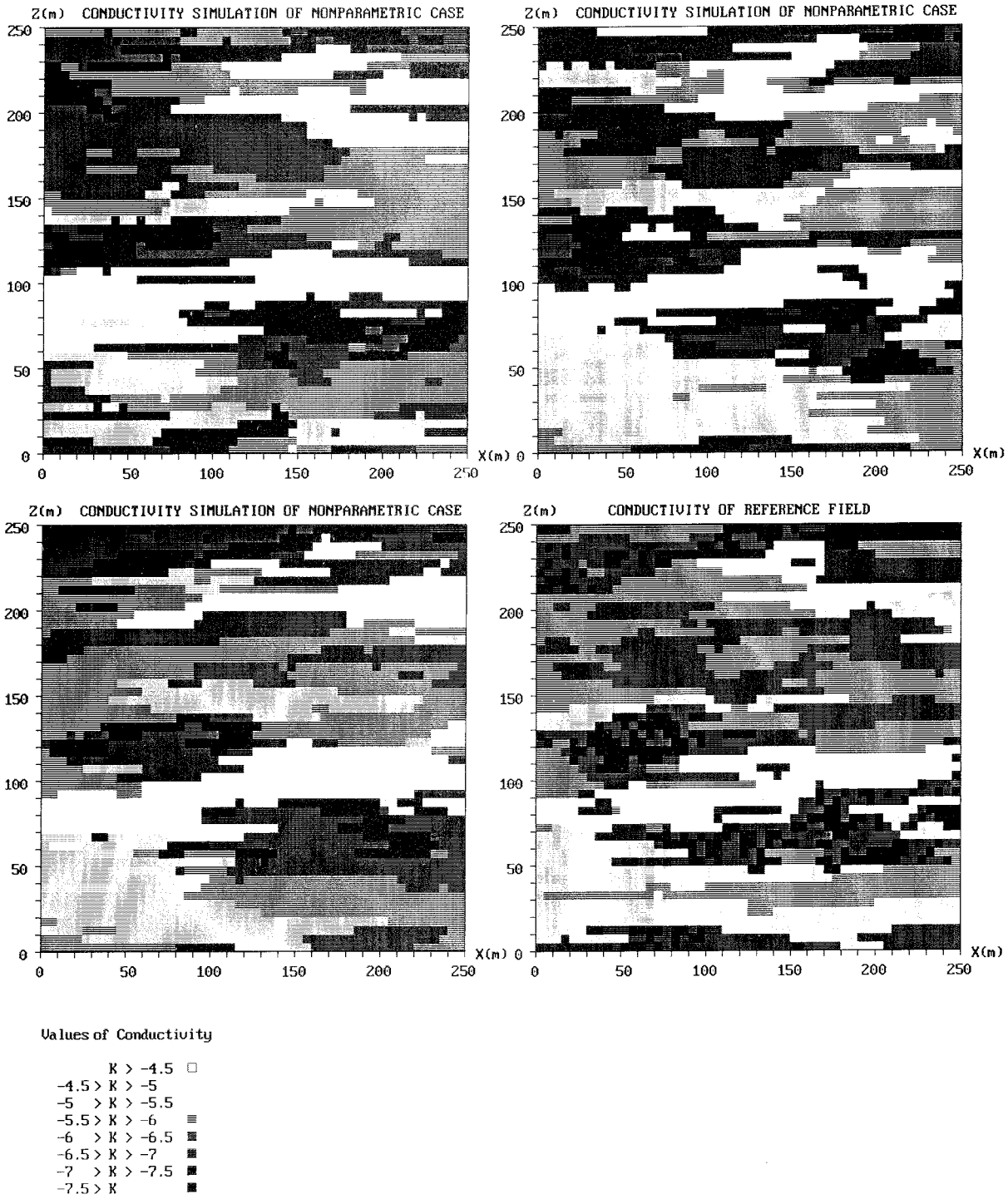


Figure 6.13 Non-parametric simulation based on 25 hard and 100 soft randomly sampled data. Examples of equiprobable realizations of the hydraulic conductivity field (log(K)) consistent with the sample statistics, a-c) simulated realizations, d) exhaustive data set.

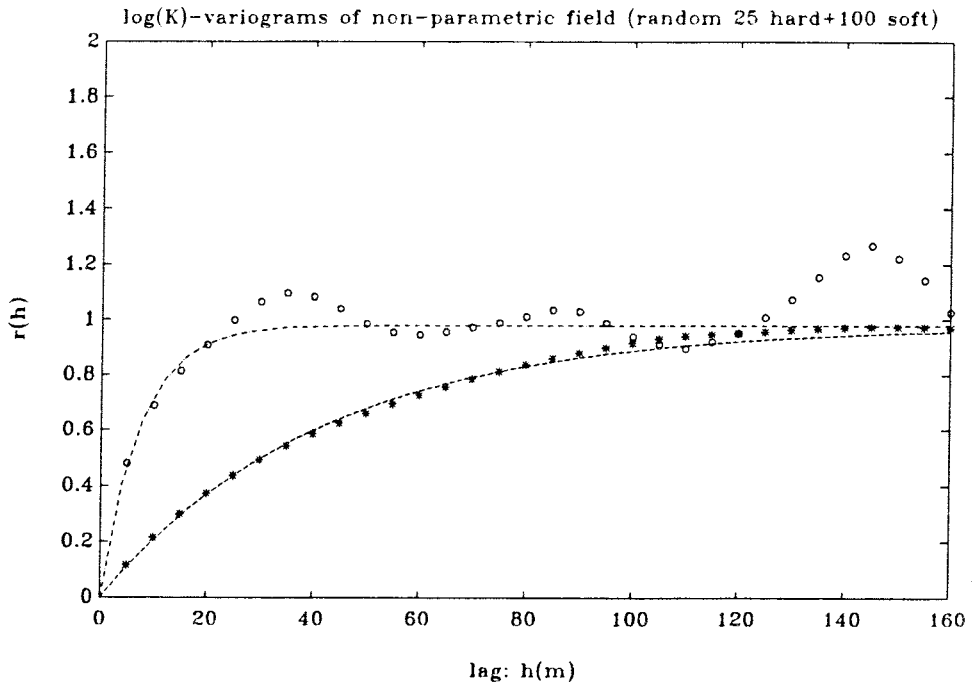


Figure 6.14 Non-parametric simulation based on 25 hard and 100 soft randomly sampled data. Generated ensemble attribute variogram of log(K), (*) = X-direction, (o) = Y-direction, (----) = reference field.

The resulting ensemble cumulative breakthrough curve for the non-parametric simulations conditioned on 25 hard and 100 soft data is shown in Figure 6.15. The particle mean arrival time is in this case somewhat later than for the reference case, cf. Figure 4.5a and Table 6.2.

The dispersivity index of the ensemble of simulated fields is higher than the previous simulation cases, cf. Table 6.2. The uncertainty in the mass arrival time is in this case lower than for the corresponding parametric case and the previous non-parametric case (random 25 hard data). Figure 6.16 shows the distribution of particle arrival locations along the outlet boundary. The results shows two weakly developed flow channel systems where one corresponds to the upper channel noted for the reference field. The lower channel in the reference field is not well represented in the ensemble of simulated fields. The simulated flow paths are more dispersed in the simulated domain compared to those of the reference field, cf. Figure 4.5b.

6.2.2 Simulation based on a designed sample data set

25 hard data

Based on the 25 hard designed sample data, 100 realizations of the simulated field have been generated using the non-parametric approach. The cumulative distribution curve of the sample data set is given by Figure 5.2b and appears near Gaussian, as is the reference field. Three equiprobable realizations which honour the conditioning

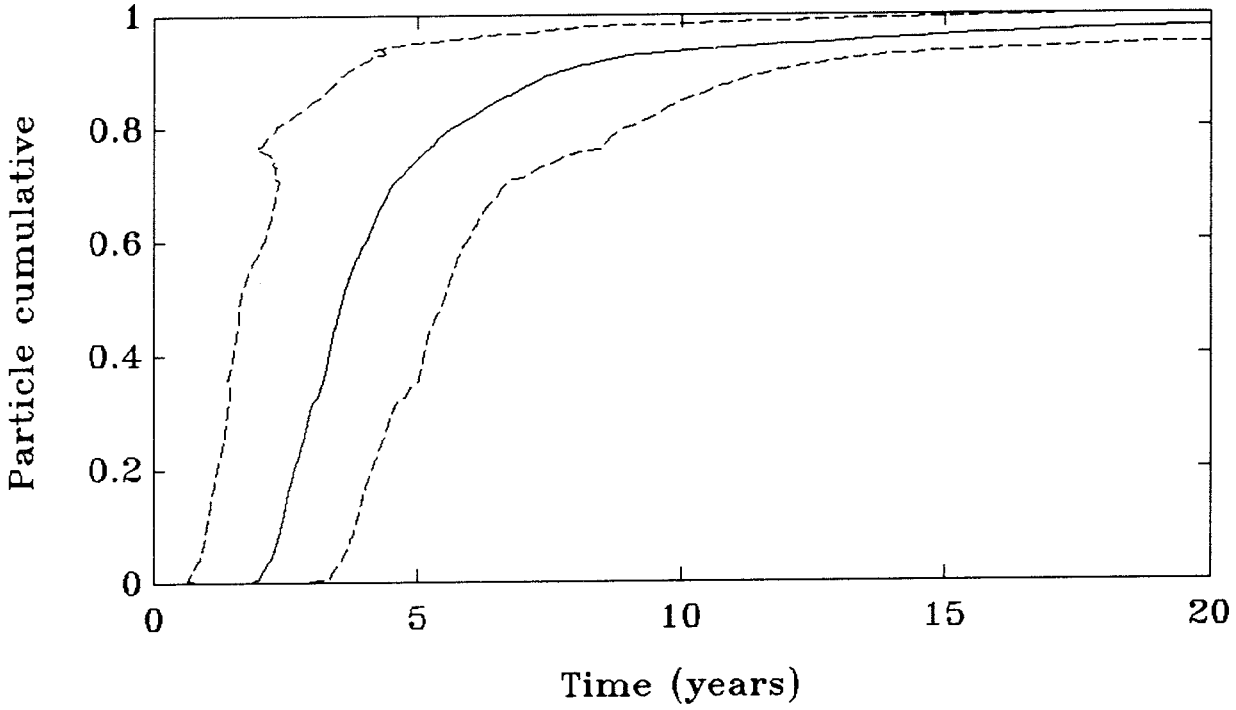


Figure 6.15 Non-parametric simulation based on 25 hard and 100 soft randomly sampled data. Ensemble cumulative breakthrough curve. Hatched line corresponds to one standard deviation.

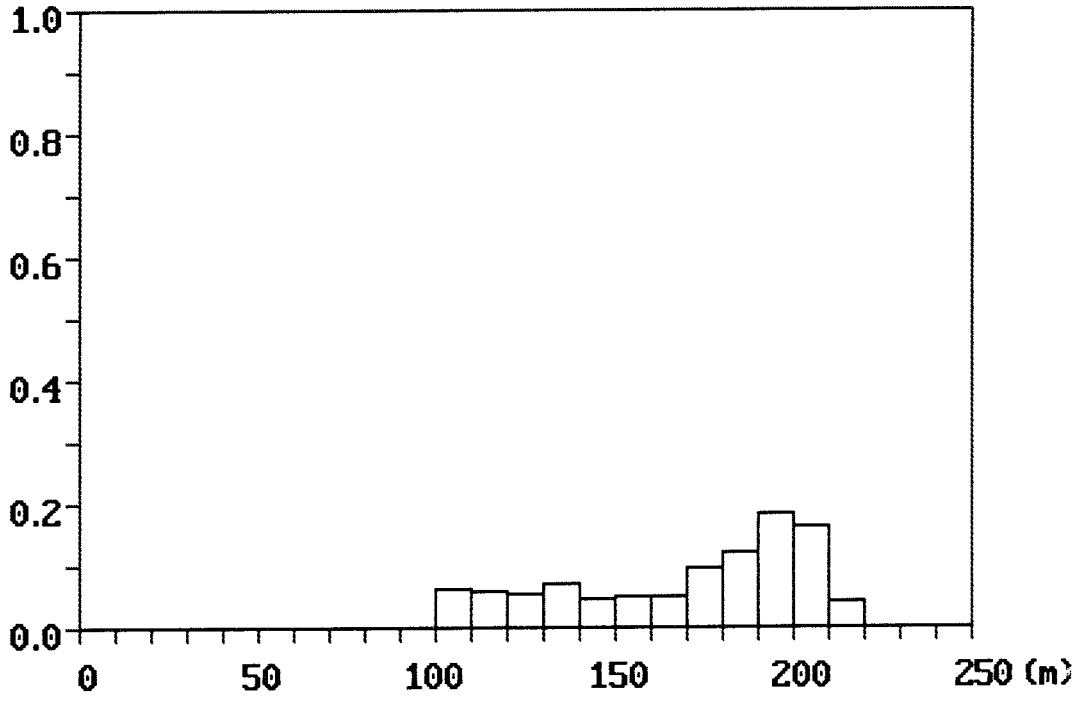


Figure 6.16 Non-parametric simulation based on 25 hard and 100 soft randomly sampled data. Ensemble particle arrival distribution on the outlet boundary.

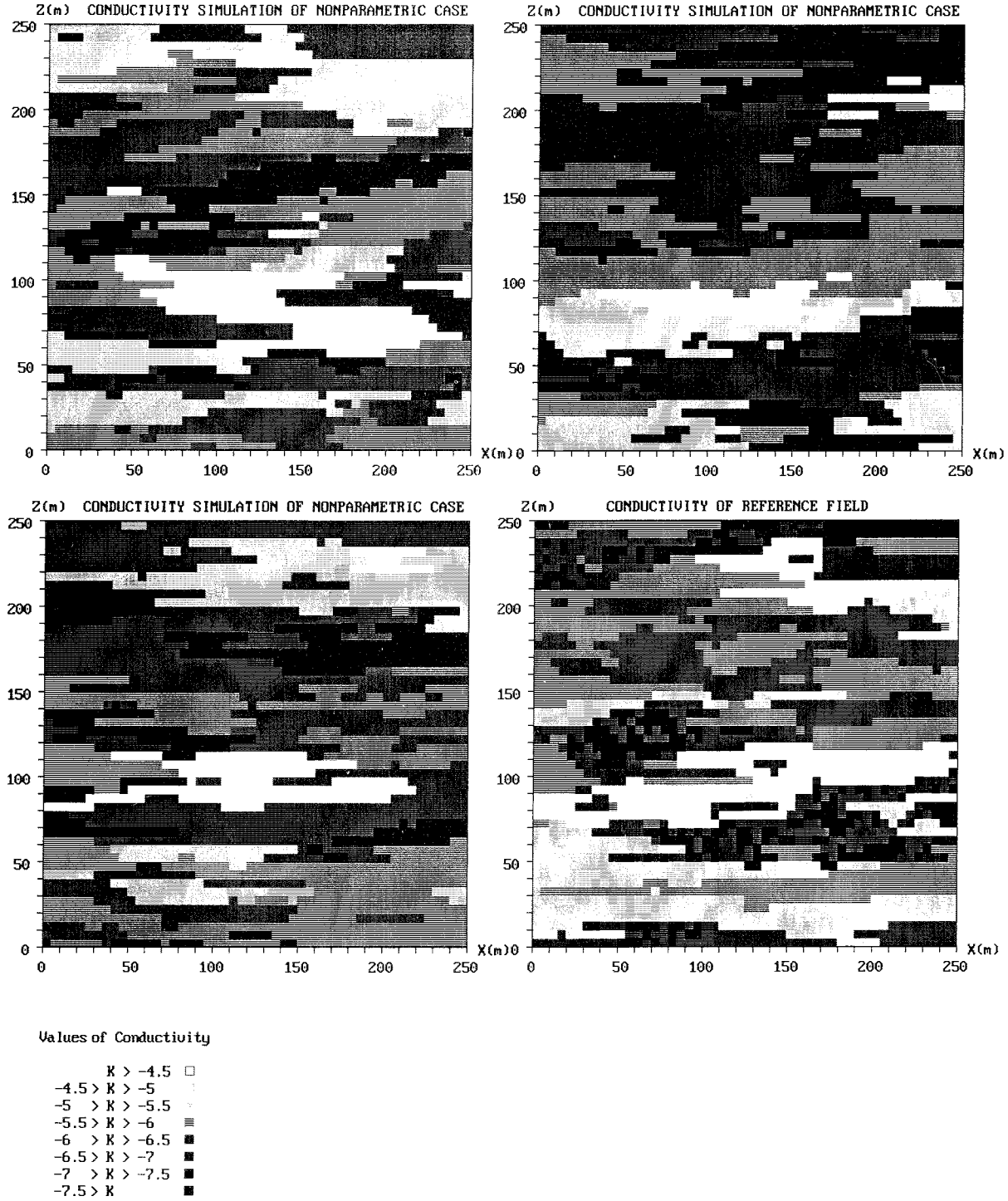


Figure 6.17 Non-parametric simulation based on 25 hard data obtained from designed sampling. Examples of equiprobable realizations of the hydraulic conductivity field (log(K)) with statistics consistent with the sample data set, a-c) simulated realizations, d) exhaustive reference data set.

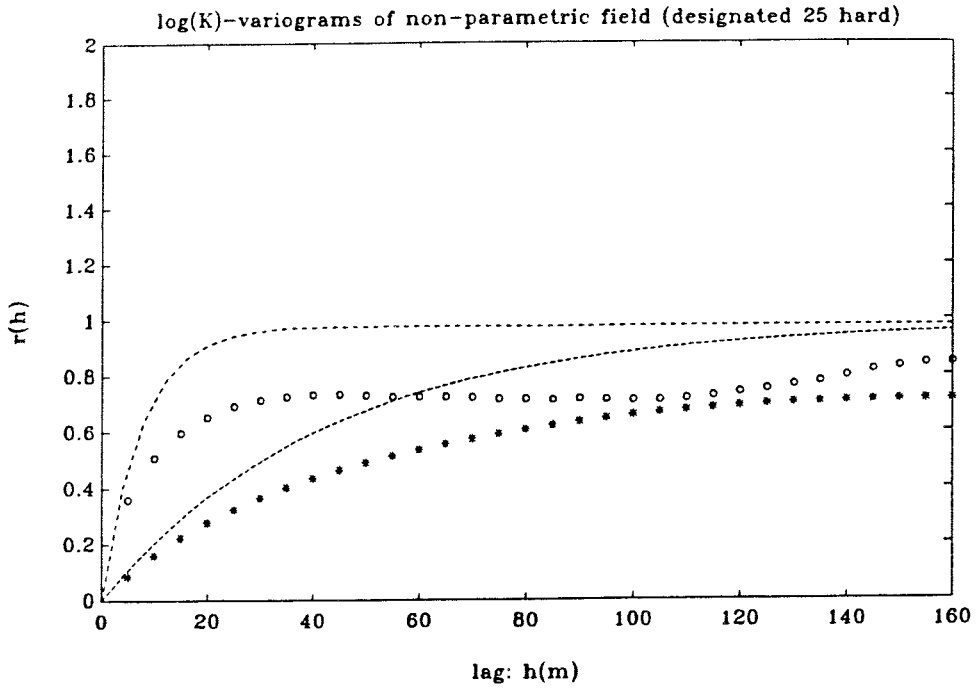


Figure 6.18 Non-parametric simulation based on 25 hard data obtained from designed sampling. Generated ensemble attribute variogram of log(K), (*) = X-direction, (o) = Y-direction, (----) = reference field.

data and the statistics of the sample data set are shown in Figure 6.17. Compared to the corresponding simulations based on the random sample data set, these realizations do not capture the high-conductive bands as good. The mean and variance of the generated fields are lower than for the reference field, cf. Table 6.1.

The generated ensemble attribute variogram shown in Figure 6.18 show that the correlation lengths in the X- and Y-directions are comparable to those of the reference field, though the sills are much lower in accordance with the sample data. The ensemble indicator variograms of the generated field for the various threshold values are shown in Appendix 4. The generated indicator variograms compare well with the input structures for the threshold values -4.5 and -5. For the lower threshold values, -6.5 and -7, the indicator variances are lower than the inputed variograms. For the threshold value -5.5, the generated indicator variance is higher than the inputed one. For the threshold value -6, the correlation lengths of the generated fields are somewhat longer than the input correlation length. The difference between the generated and the inputed structures may be due to the fact that the number of realizations is too small and/or that the simulation domain is not large enough. The noted difference between the in- and output indicator variogram implies that the shape of the cumulative distribution of attribute variable (log(K)) differs from that of the reference field.

Figure 6.19 shows the ensemble cumulative breakthrough curve for the non-parametric simulations together with one standard deviation envelop of arrival time. The mean particle arrival time of the non-parametric simulation based on 25 hard designed sample data is significantly later than the reference case, cf. Table 6.2. The

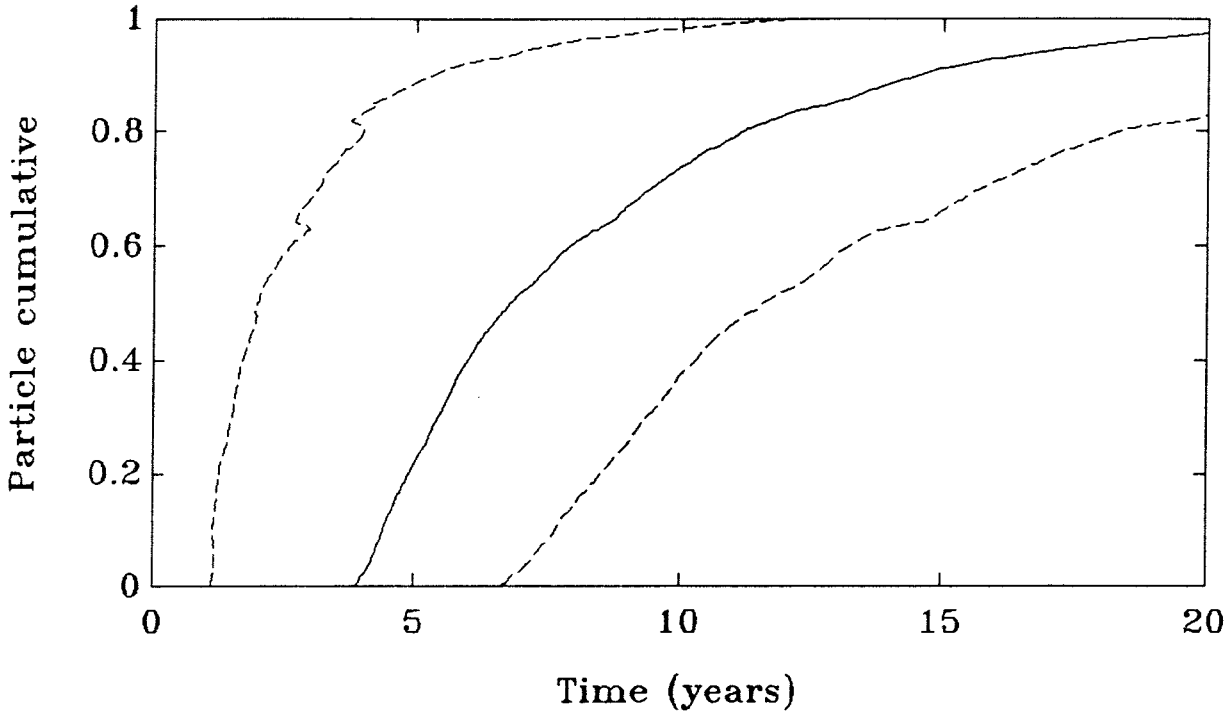


Figure 6.19 Non-parametric simulation based on 25 hard data obtained from designed sampling. Ensemble cumulative breakthrough curve. Hatched line corresponds to one standard deviation.

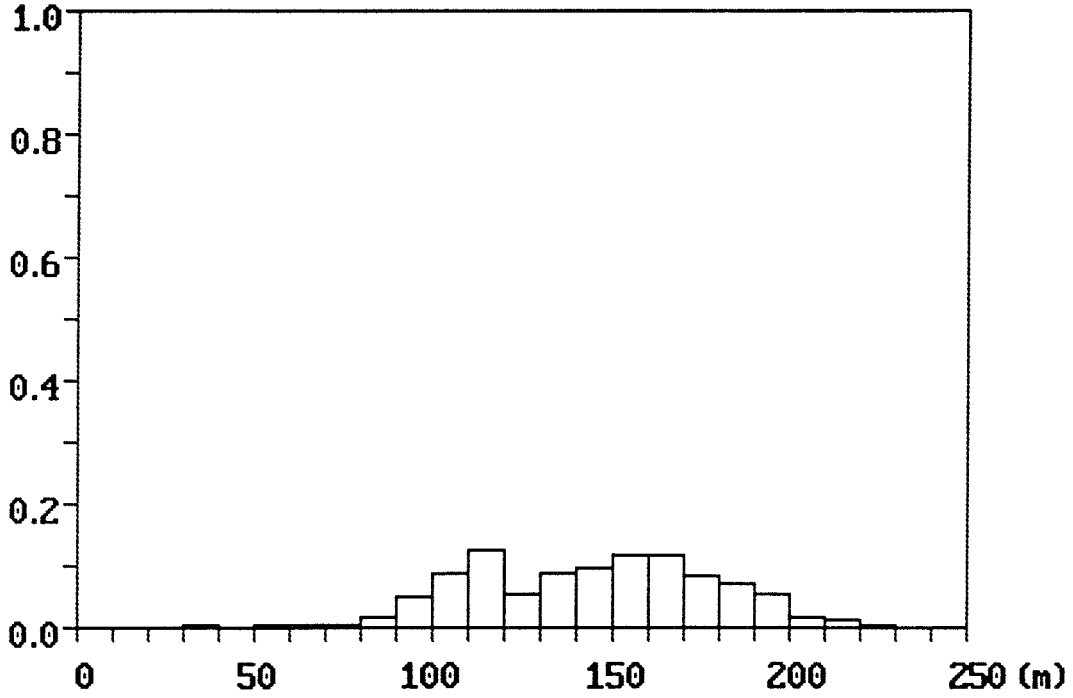


Figure 6.20 Non-parametric simulation based on 25 hard data obtained from designed sampling. Ensemble particle arrival distribution on the outlet boundary.

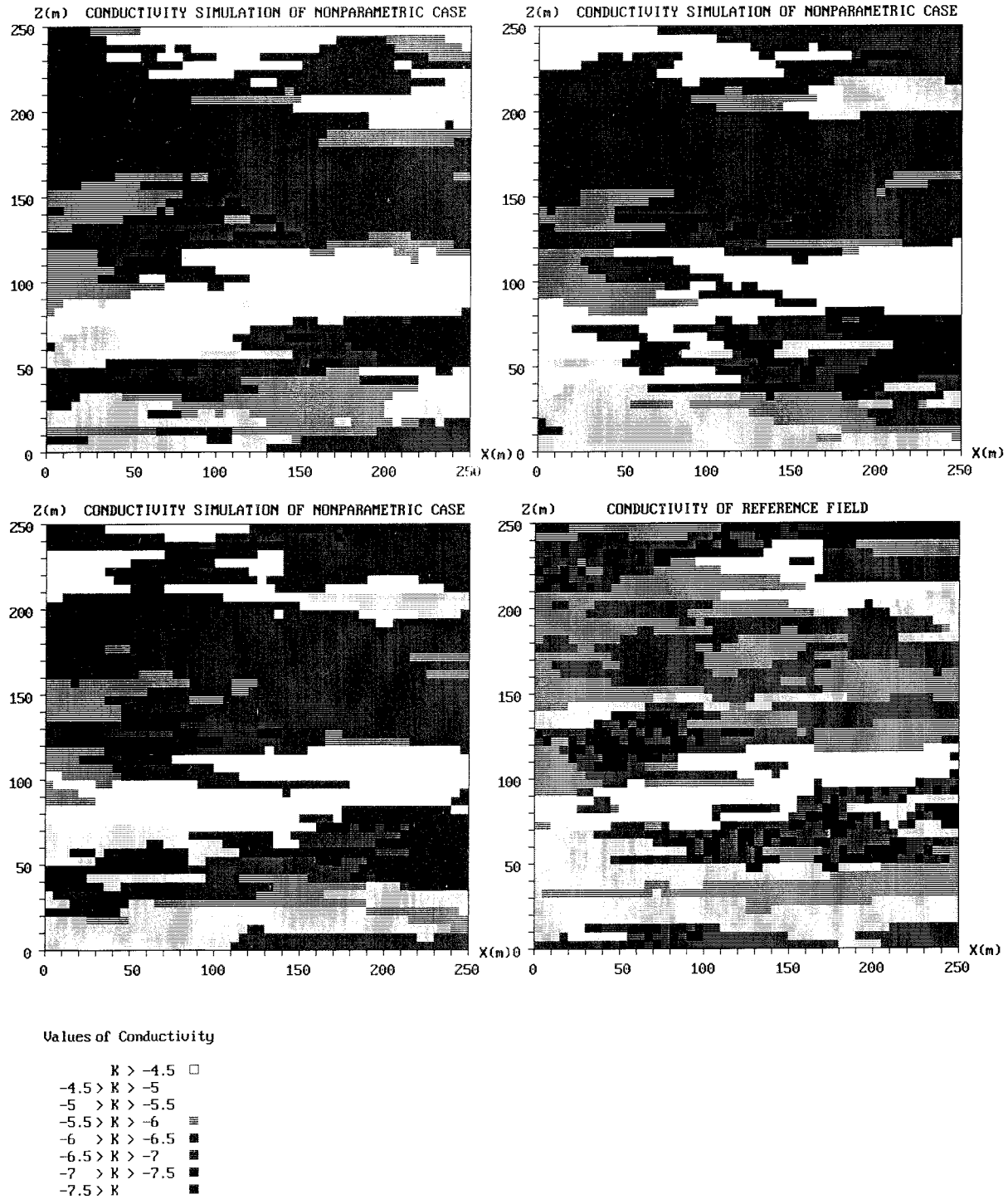


Figure 6.21 Non-parametric simulation based on 25 hard data obtained from designed sampling. Examples of equiprobable realizations of the hydraulic conductivity field (log(K)) with statistics consistent with the sample data set, a-c) simulated realizations, d) exhaustive reference data set.

dispersivity index of the ensemble of simulated fields is lower than for the corresponding ensemble of parametric fields. The uncertainty in mass arrival time of this non-parametric simulation case is higher than for the corresponding parametric case. The distribution of particle arrival locations along the outlet boundary is depicted in Figure 6.20. From the figure two weakly developed flow channel systems can be identified, one corresponding to the lower flow channel noted for the reference field. The upper flow channel identified in the reference case is not well represented in the simulated fields. The simulated lower channel is more dispersed and located closer to the centre of the outlet plane, cf. Figures 6.20 and 4.5b.

25 hard and 100 soft data

For the second designed sampling case, 25 hard data and 100 regularly spaced soft data have been used as a sample data set to generate non-parametric simulations of hydraulic conductivity field. Three examples of equiprobable and statistically consistent hydraulic conductivity realization are shown in Figure 6.21. The generated conductivity fields compare better with the reference field than the previous case based on 25 hard data. The ensemble of generated fields exhibit similar statistical characteristics as the reference field though with somewhat lower mean and variance, cf. Table 6.1. The generated ensemble attribute variogram shown in Figure 6.22 show that the correlation lengths in the X- and Y-directions are somewhat longer than the reference field. The ensemble indicator variograms of the generated fields for the different threshold values are shown in Appendix 5. The generated indicator variances in X- and Y-direction are somewhat lower than the input values for high and low threshold values, whereas it is slightly higher than the input value for the middle range of threshold values.

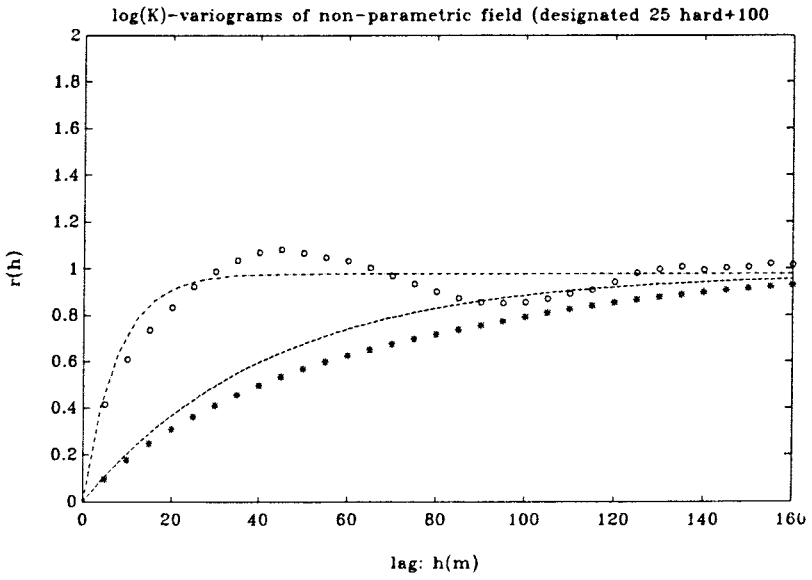


Figure 6.22 Non-parametric simulation based on 25 hard and 100 soft data obtained from designed sampling. Generated ensemble attribute variogram of log(K), (*) = X-direction, (o) = Y-direction, (----) = reference field .

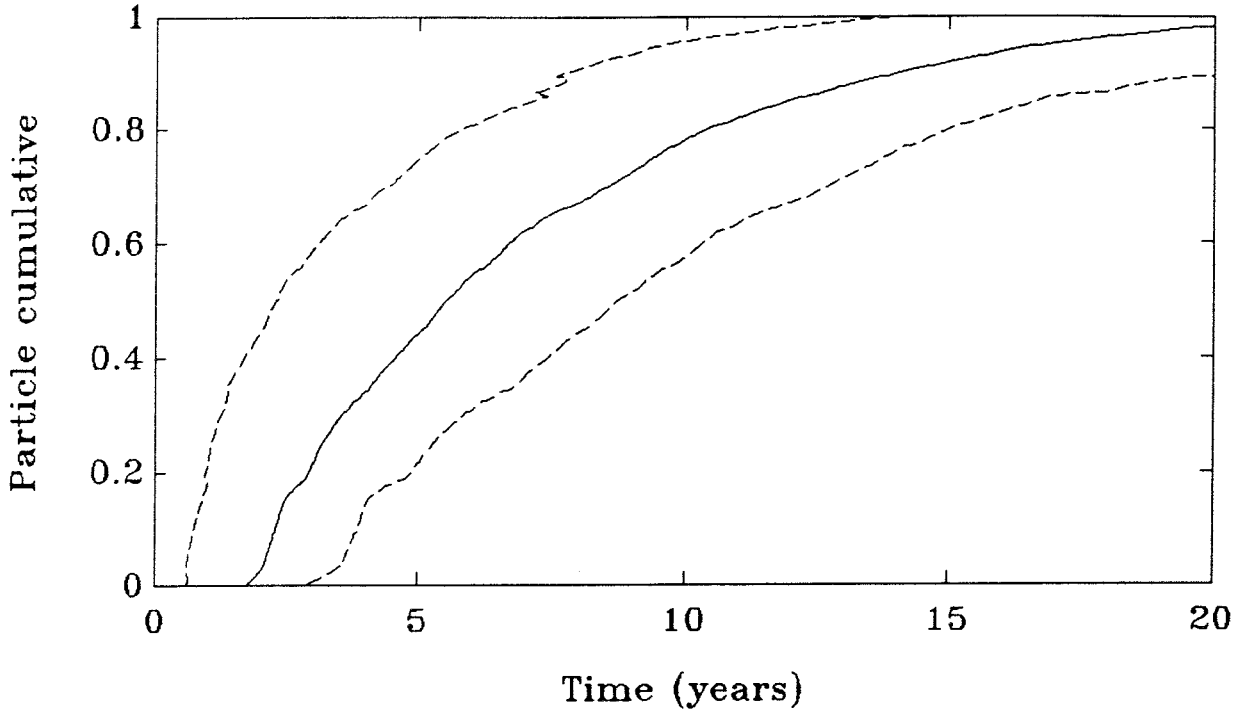


Figure 6.23 Non-parametric simulation based on 25 hard and 100 soft data obtained from designed sampling. Ensemble cumulative breakthrough curve. Hatched line corresponds to one standard deviation.

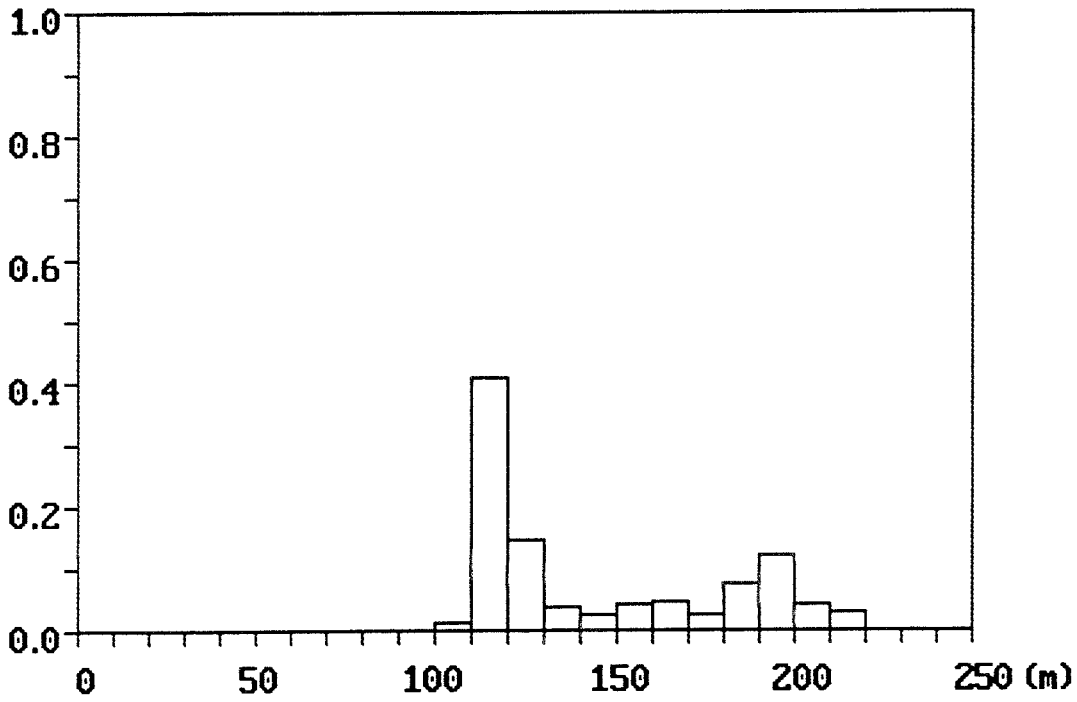


Figure 6.24 Non-parametric simulation based on 25 hard and 100 soft data obtained from designed sampling. Ensemble particle arrival distribution on the outlet boundary.

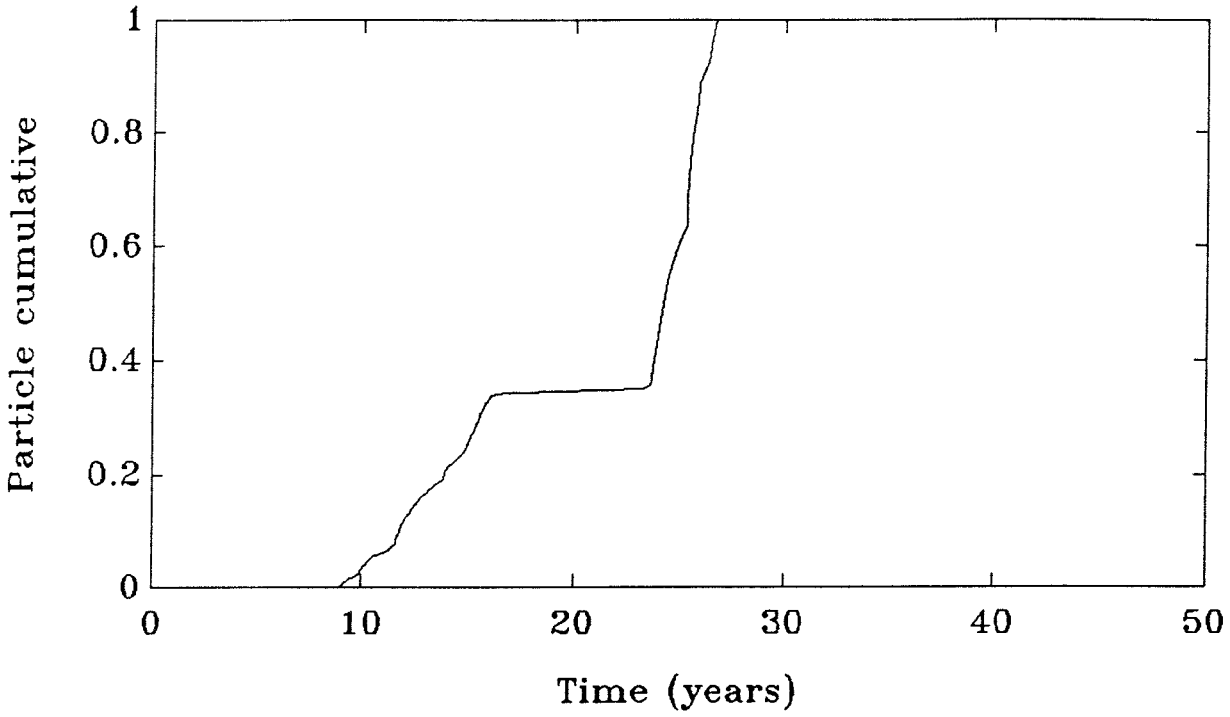


Figure 6.25 Rotated flow field applied to the reference exhaustive field. Cumulative particle breakthrough curve.

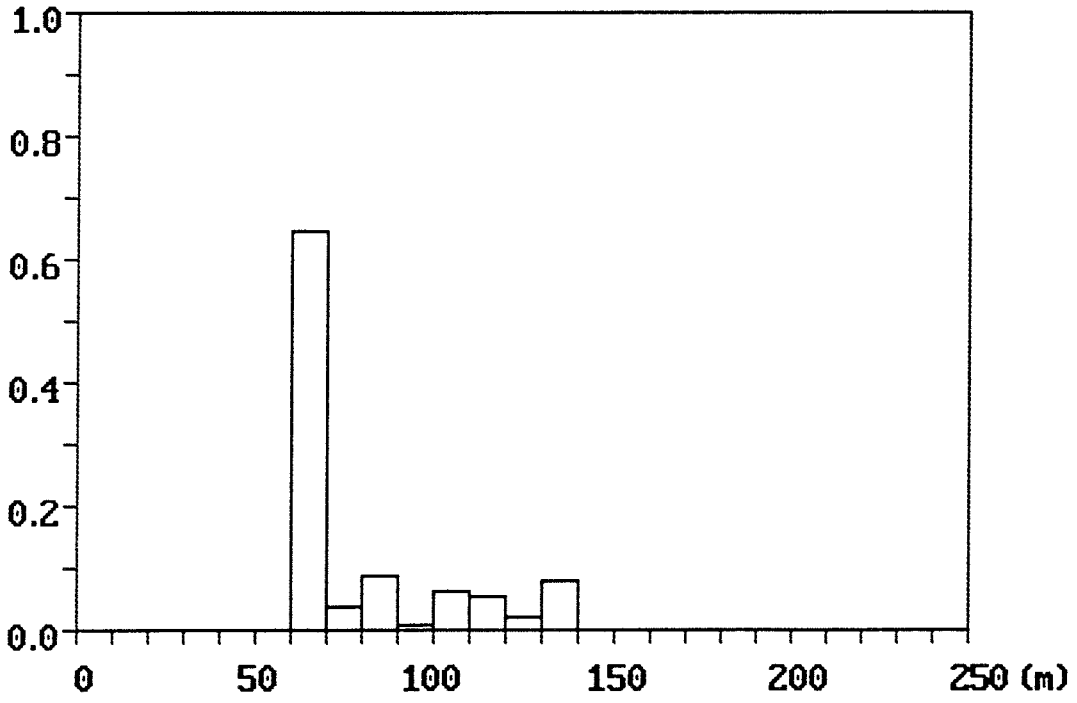


Figure 6.26 Rotated flow field applied to the reference exhaustive field. Distribution of particle arrival location on the outlet (upper) boundary.

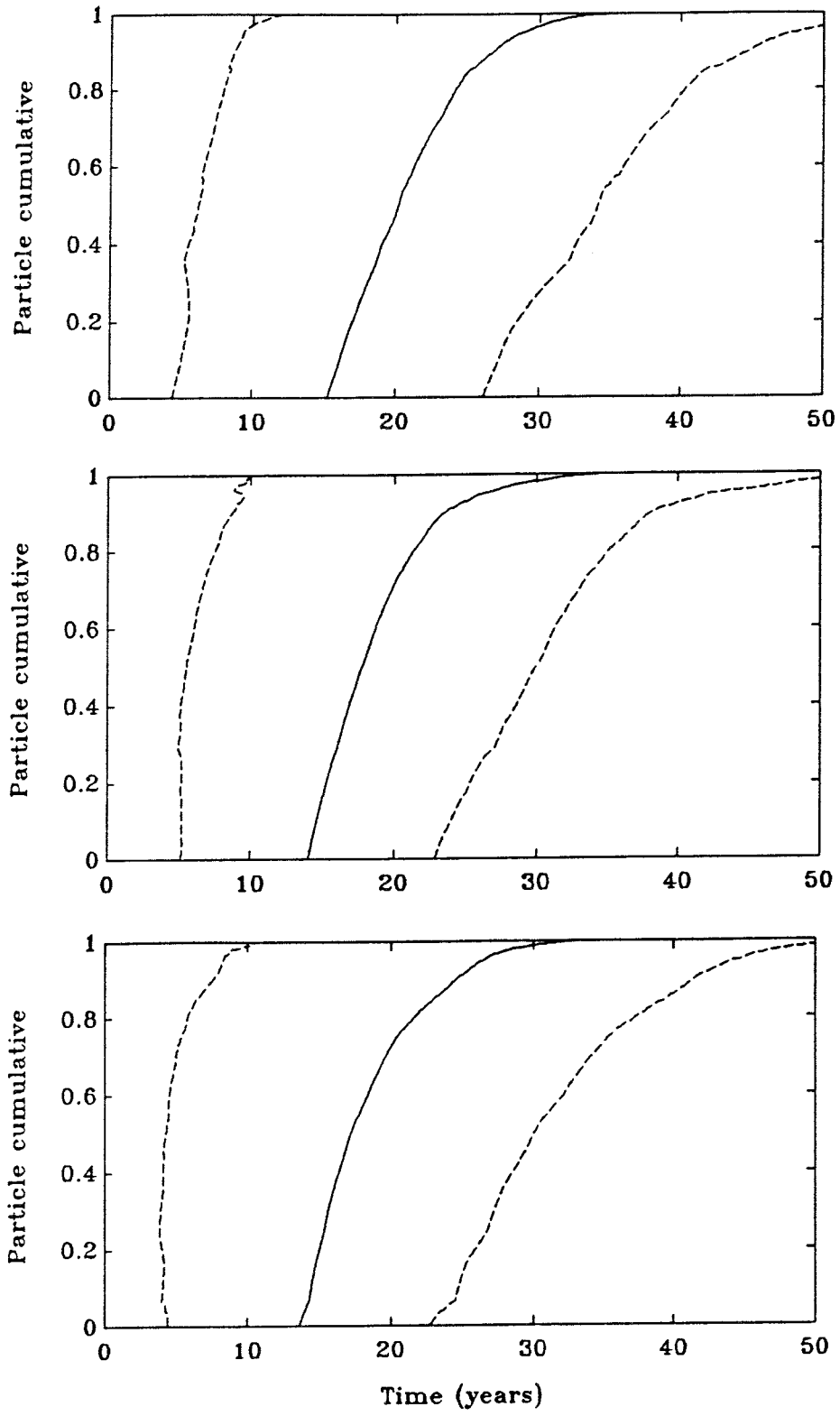


Figure 6.27 Rotated flow field applied to simulated fields conditioned on a random sample data set. Ensemble cumulative breakthrough curves based on a) Parametric (25 hard data), b) Non-parametric (25 hard), c) Non-parametric (25 hard + 100 soft).

The ensemble cumulative breakthrough curve of this non-parametric simulation case together with one standard deviation envelop of arrival time is shown in Figure 6.23. The ensemble mean particle arrival time is slightly lower than that for the reference case. The dispersivity index of the simulated fields is higher than for the previous case. The uncertainty in mass arrival time is similar to that of the corresponding parametric case. The distribution of particle arrival locations along the outlet boundary is shown in Figure 6.24. Evident from the figure is a developed two channel flow system which corresponds well to the reference field results with regard to position. However, the number of particle arriving in the lower channel is much higher than along the upper channel, which is the opposite to what is found for the reference case, cf. Figure 4.5b. In addition the simulated flow paths are more dispersed around the flow channels than is the case for the reference field.

6.3 Simulation using a rotated flow field

The flow and transport characteristics of a given domain are not only dependent on the hydraulic conductivity field, but also on the acting flow conditions in terms of boundary conditions and the source area. Usually, the boundary conditions and flow direction at a local scale do not exactly correspond to the regional flow conditions, the latter which will be influenced by the local hydraulic conductivity distribution pattern, and especially so if we are dealing with a highly heterogeneous field.

In order to better understand the effect of a changing flow direction on the overall flow and transport characteristics of the generated fields, we introduced a 90° rotation of the boundary condition, ie. no flow conditions on vertical faces and constant head conditions on upper and lower boundaries. This boundary condition provides an essentially upward vertical flow direction. Flow and transport calculations have been performed on 10 out of 100 realizations for each of the cases involving the use of a random sample data set presented previously in this chapter. At the centre of the bottom boundary, particles were released in accordance with Section 4.2, and the mass arrival time and particle arrival location at the upper boundary were recorded.

The overall result of the rotated flow field is more meandering flow paths caused by the flow being forced orthogonal to the major anisotropy axis. Figure 6.25 shows the cumulative breakthrough curve of mass arrival at the outlet plane for the reference field. The figure show two distinct arrival peaks with almost no particle arrival in between. The location of particle arrival on the outlet plane shown in Figure 6.26 indicate that c. 65% of the particles discharge through one flow channel and the rest of the particles are conveyed through another more dispersed flow channel. The particles arriving to the first major channel corresponds the later part of breakthrough curve. The reference field can in terms of flow and transport characteristics be described as a two channel system with the prevailing boundary and particle source conditions.

Figure 6.27 shows the resulting ensemble cumulative breakthrough curves for the parametric and non-parametric cases using a random sample conditioning data set. Due to the limited statistics, the results have not been compiled in a table. The

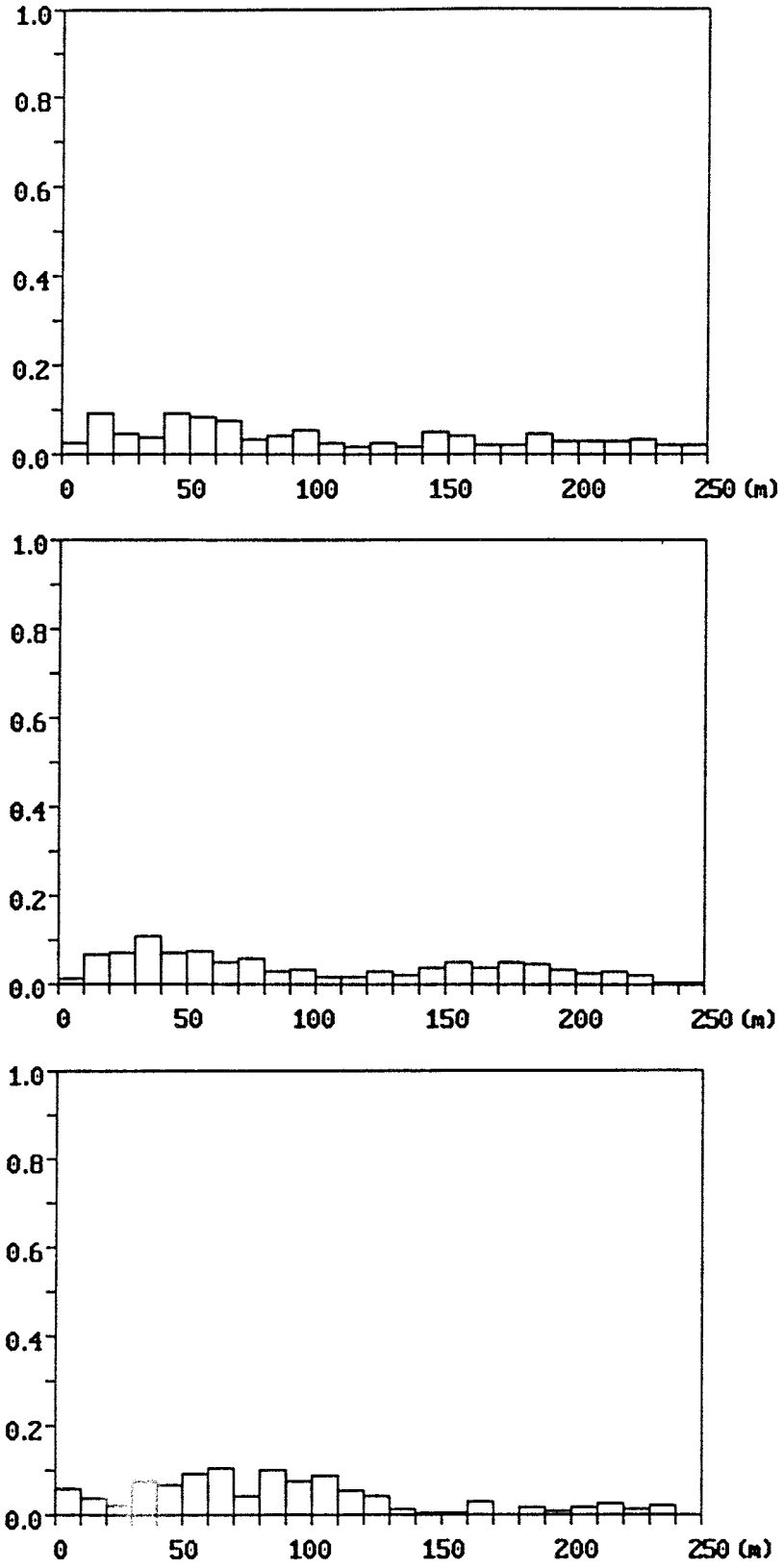


Figure 6.28 Rotated flow field applied to simulated fields conditioned on a random sample data set. Ensemble particle arrival distribution based on a) Parametric (25 hard data), b) Non-parametric (25 hard), c) Non-parametric (25 hard + 100 soft).

ensemble particle arrival location on the outlet (upper) boundary is shown in Figure 6.28. The flow and transport characteristics of the three presented simulation cases are very similar, although very different when compared with the reference case results. The simulated cumulative breakthrough curves of particle arrival time show a higher degree of dispersion than the reference case, but the mean arrival time is approximately identical to that of the reference case.

Also the simulated flow paths are more dispersed than those observed for the reference field, cf. Figures 6.26 and 6.28. The uncertainty in mass arrival time is almost identical for the two simulation approaches. The soft information does not improve the results of the simulation on the rotated flow conditions, possibly because of the complicated flow and transport characteristics and limited number of hard data.

7. DISCUSSION

Comparisons between the different approaches may be performed on the basis of the visual appearance of the individual hydraulic conductivity fields generated. Another way to compare and discuss the results is to use the flow and transport characteristics of the individual or ensemble results. With these characteristics we imply the number of flow channels developed, their absolute and relative geometry and location, and their relative importance in terms of number of particles conveyed, and last, how dispersed the transport is in general, and in particular the dispersion around an individual channel. The above information is contained both in the cumulative breakthrough curve and the plot showing the distribution of particle arrival location. To these characteristics one should of course also add the quantitative information contained in the cumulative breakthrough curve in terms of mean particle arrival time with associated uncertainty and statistics.

7.1 Implications of use of soft information

Our premise in this study has been to test how the addition of 100 soft (imprecise) data may add and sustain the description of the spatial characteristics of our exhaustive reference data set, and also how this addition of soft information may help to narrow down the uncertainty of mass arrival at a given location.

It was found at an early stage that the 25 hard data (1% of the exhaustive reference data set), irrespective of which sampling scheme was employed, were not sufficient to reproduce the spatial characteristics and the anisotropic features of the reference field in terms of variograms, this is particulary true for the construction of indicator variograms. It was thus decided to use the attribute and indicator variograms based on the 2500 data of the exhaustive reference data set in the different stochastic approaches, though with sills scaled down to the variance of the sample data sets.

In the study, 25 hard or 25 hard plus 100 soft data either of designed or random sampling pattern were used to condition the simulation.

From a geostatistical point of view, the generated hydraulic conductivity fields show a better correspondence with the exhaustive reference data set when soft information is incorporated in the conditioning. In the case of the random sample data set the improvement is limited because the 25 hard data already provide a satisfactory representation of the exhaustive representative field. In the case of the designed sample data set there is a noticeable improvement with the addition of soft data compared to using the hard data alone.

The flow and transport characteristics of the exhaustive reference field are best reproduced when using a random sample data set including soft data. Also when using a designed sample data and soft conditioning the correspondence with the exhaustive reference case is acceptable, but the breakthrough is still more disperse than that of the reference case. In all cases the incorporation of the 100 soft data reduces the uncertainty in particle arrival time with about 50%.

The flow and transport characteristics of a heterogeneous field depend on the acting flow conditions (i.a. the boundary conditions), the source area (its location in the domain and surrounding material property distribution), and the distribution of developed preferential flow paths (flow channels). The probably most important factor, in order to capture and quantify these characteristics, is the amount of conditioning data and their particular locations. In our case, the number of hard conditioning (sample) data positioned in the dominant flow paths of the exhaustive reference field is small. As a consequence, the added soft information will not contribute substantially in improving the representation of the reference field, and in reducing the uncertainty of mass arrival.

7.2 Comparison of results based on random and designed sampling

The two sample data sets used in this study, a random sample and a designed sample data set, differ both in the way they originate and also with regard to their respective statistical characteristics. The designed sample data set has been conceived to capture a successively spreading plume of particles downstream from the source area, this without focusing in specifically on the high-conductive portions of the exhaustive reference domain. Compared to the statistics of the random sample data set, the designed sample data set has a too low mean and a too low variance of log(K).

When comparing the parametric simulations based on the designed sample data set, Figure 6.1, with those based on the random sample data set, Figure 6.5, it is evident that the former show a lower amount of high-conductive parts. The same condition also apply to the corresponding non-parametric simulations. The effect of this difference is also evident in the cumulative breakthrough curves and the plots of distribution of particle arrival. Although the flow and transport characteristics of the random sample case differ from the reference case, the cumulative breakthrough curve show similar characteristics to that of the reference case.

When soft information is added to the sample data sets some improvement in the cumulative breakthrough curve based on random sampling is noted, whereas no improvement in correspondence is noted in the results based on the designed sample data set in the case of the non-parametric simulations.

Our observations from the performed simulations show that the designed hard data set of 25 values cannot satisfactorily represent the characteristics of the exhaustive reference field. Due to the inherent low quality in the 25 hard data in the sample, the inclusion of 100 soft data does not help to restore the representativeness of the simulations.

Since the statistics of the random sample data set compare so well with the statistics of the exhaustive reference field, the simulation results show a high degree of similarity even though the flow pattern is not fully represented. The soft information in this case not only help to reduce the uncertainty of particle arrival time, but also to better represent the mean arrival time of the reference field.

7.3 Implications of a rotated flow field

As indicated previously, the flow and transport characteristics of a studied heterogeneous domain do not entirely depend on the material property distribution but also on the boundary conditions acting on the domain. To elude to this fact we have simulated the flow and transport of a limited number of realizations (N=10) for the random sample data set using a 90° rotation of the flow, with all other conditions kept constant.

We found in the simulation of our exhaustive reference field that a dual flow channel system evolved with a distinct dual peak cumulative breakthrough curve. A dual flow channel was also noted for the initial (non-rotated) flow condition, but in the latter case one peak dominated the transport resulting in a single peak breakthrough curve. Overall, the flow and transport orthogonal to the anisotropy is more complex, resulting in meandering flow channels.

It was found that the simulations provided a fair representation of the mean arrival times of the reference field, though with more dispersed flow paths as evident from the produced breakthrough and particle arrival distribution plots, cf. Figures 6.25-6.28. Due to the more complicated flow pattern the addition of 100 soft data did not help to provide a better representation, not even to reduce the uncertainty of mass arrival.

It should however be emphasized that these observations are based on only 10% of the number of simulations used for the initial flow condition. In order to draw firm conclusions, a full suite of at least 100 simulations have to be processed.

7.4 Comparison between parametric and non-parametric techniques

The hydraulic conductivity of our exhaustive reference field has been characterized by its mean, variance and its correlation structures (attribute and indicator variograms).

In our application of the parametric approach, we have employed a Gaussian distribution model and an exponential model of spatial correlation of the attribute. In our application of the non-parametric approach, we have used the sample cumulative distribution, disregarding the actual distribution model, and exponential indicator variograms in our simulations. In both cases the distribution and univariate statistics originate from the respective sample data sets whereas the correlation structure is based on analysis of the exhaustive reference data set (with sills scaled to sample variances).

The attribute variogram of the reference field, used in the simulations, is an exponential model with an anisotropy ratio 5.7, implying a higher continuity in the X-direction.

With regard to the non-parametric approach, the indicator variograms calculated for the lower threshold values have shorter correlation lengths and smaller anisotropy ratios, whereas the higher threshold values have longer correlation lengths and a higher anisotropy ratios than the attribute variogram. These features imply that the exhaustive reference field has a well connected spatial continuity in the X- direction, resulting in features similar to continuous channels of high or low hydraulic conductivity, and also of varying width.

The parametric simulation uses one single correlation length valid for the complete range of hydraulic conductivity. In the simulated realizations this results in the formation of patches of high and low hydraulic conductivity, their size being comparable to the correlation length. These features and the fact that the high-conductive zones are not as continuous as in the reference field is obvious from Figures 6.1 and 6.5. Soft information cannot (presently) be used to improve the correspondence of the parametric simulations.

The realizations of hydraulic conductivity generated using the non-parametric approach and a hard sample data set show the same overall characteristics as the reference field, cf. Figures 6.9 and 6.17. This correspondence is enhanced further by introducing soft data, cf. Figures 6.13 and 6.21.

8. CONCLUSIONS AND RECOMMENDATIONS

Two stochastic continuum approaches for simulating groundwater flow and solute transport in fractured rock have been applied to a synthetic data set to investigate the effect of *conditioning with hard and soft data*. The general conclusion of the study is that conditioning is a viable and potentially useful means for obtaining more reliable solute transport predictions. The specific conclusions of the study are that;

1) conditioning on soft data *reduces the uncertainty* of solute arrival time;

2) conditioning on soft data appears to provide an improvement in *characterizing channelling effects.*

The improvement in the predictions when conditioning on 25 hard and 100 soft data was found to be moderate. The number and distribution of the actual hydraulic conductivity measurements (hard data) are critical for an accurate description of flow and transport in a heterogeneous formation. Since the statistics of the 25 random sampled hard data corresponded well with that of the reference field, the incorporation of a limited number of soft data (100) did not improve the simulations significantly.

With regard to soft data conditioning it should be pointed out that in a real practical application the amount of soft data may be conceivably greater than considered in this study. For instance, seismic data may be obtained over depth and along transects, including crosshole tomographic imaging processing, which could yield a relatively large amount of soft information. Furthermore, the soft data may be characterized in terms of degree of "hardness" (or "softness") depending on its correlation to hard data. For instance, seismic data may be viewed as rather soft since the uncertainty in correlating seismic data to hydraulic conductivity is relatively large, implying a large range around the "true" value for the soft data. It is therefore of interest to analyse further *the effect of conditioning using a larger number of soft data with different uncertainty ranges.*

Designed sampling has provided a mild improvement in some of the simulated cases. However, investigations of different sampling strategies have been very limited in this study. A more extensive investigation in this direction could provide more reliable recommendations with regard to efficient sampling of hard and soft data.

In addition to heterogeneity in the hydraulic properties of fractured formations, an other source of uncertainty are the *boundary conditions* and *solute source characteristics.* For more realistic simulations, these uncertainties have to be taken into account.

Finally, it should be pointed out that the flow and transport calculations performed, and the conclusions drawn are based on, and are specifically valid for *the one single exhaustive reference data set* used in the study. Notwithstanding this fact, we feel that the conclusions drawn with regard to the benefits of soft conditioning have a general validity.

9. REFERENCES

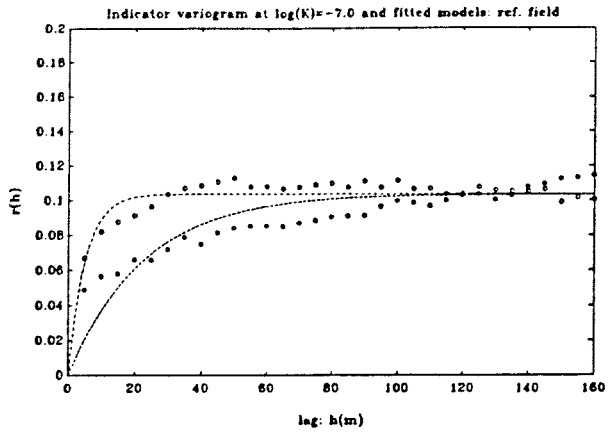
Gómez-Hernández, J.J. and Srivastava, R.M. 1990 : ISIM3D: an ANSI-C Three-dimensional Multiple Indicator Conditional Simulation Program. Computers and Geosciences, vol. 16, no. 4, pp. 395-440.

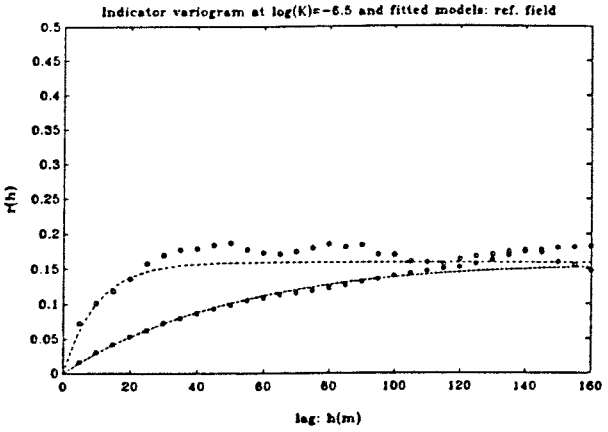
Journel, A.G. and Alabert, F. 1989 : Non-Gaussian Data Expansion in the Earth Sciences. Terra Nova, vol. 1, no. 2, pp. 123-134.

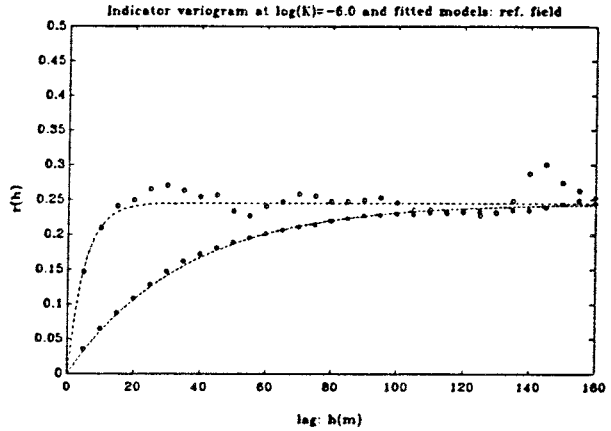
Matheron, G. 1973 : The Intrinsic Random Functions and their Applications. Adv Appl. Prob., vol. 5, pp. 438-468.

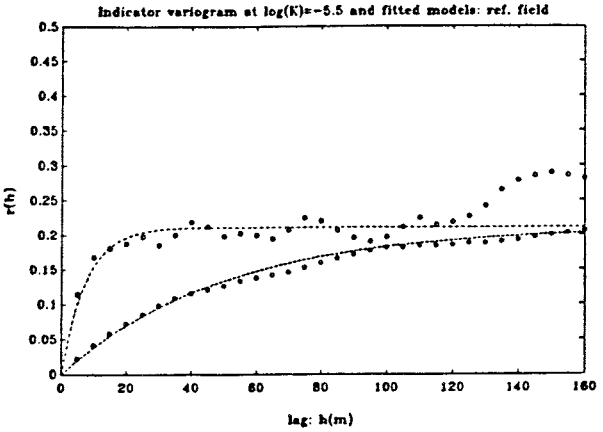
Montaglou, A. 1987 : Digital Simulation of Multivariate Two- and Three-dimensional Stochastic Processes with a Spectral Turning Bands Method. Mathematical Geology, vol. 19, no. 2, pp. 129-150.

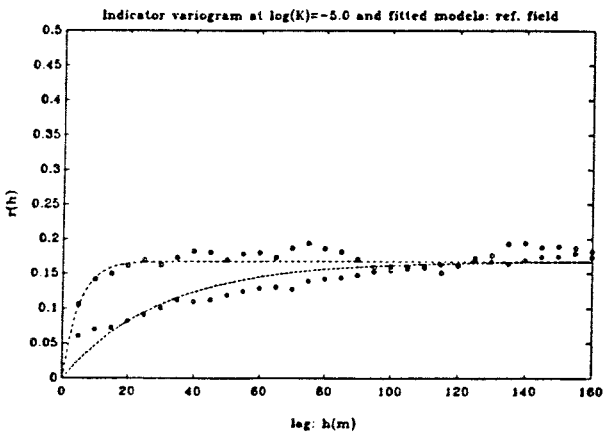
Winberg, A., Kung Chen Shan, Wen Xian Huan and Cvetkovic, V. 1990 : Stochastic Continuum Modelling of Mass Arrival at Finnsjön - Parametric and Non-parametric Approaches. Swedish Nuclear Fuel and Waste Management Company, Progress Report SKB AR 90-40.

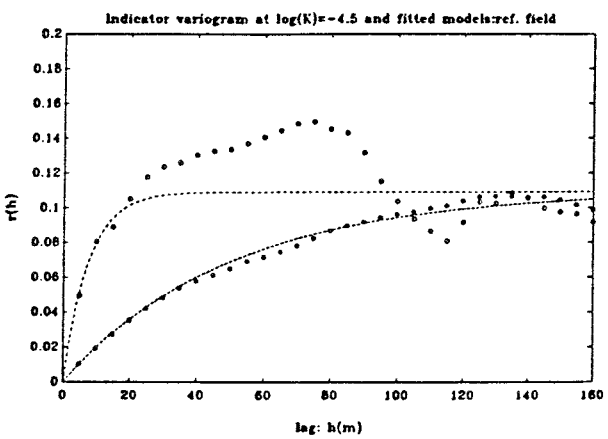




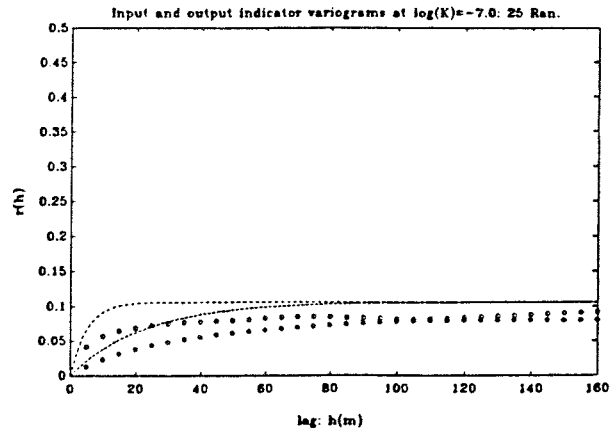


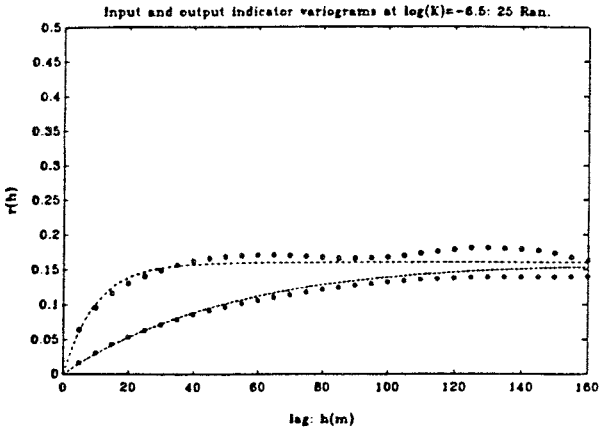


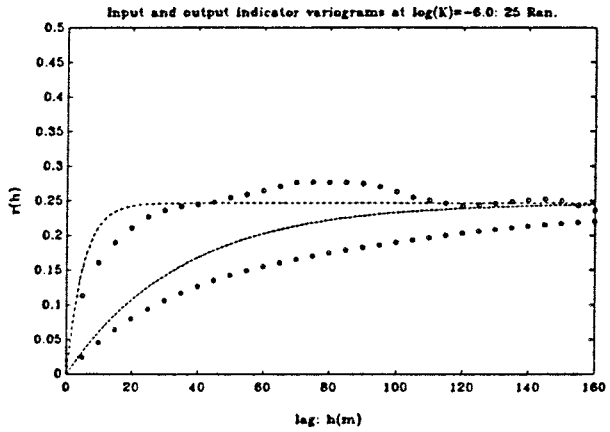


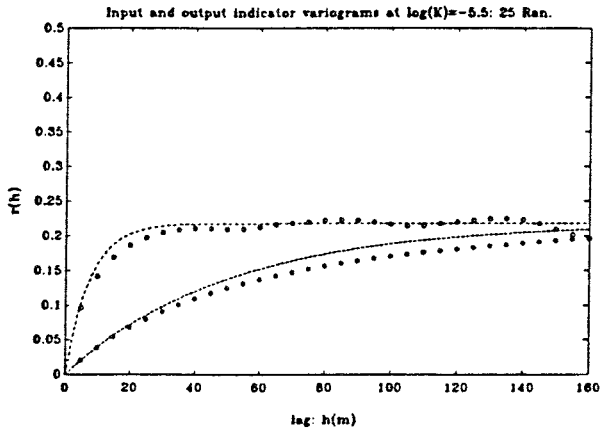


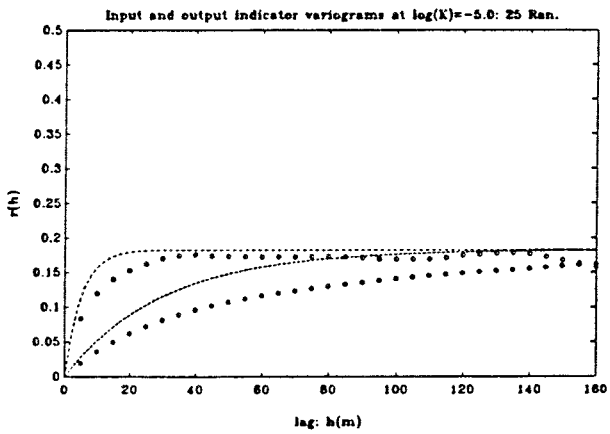
Reference data set - Experimental variograms in the X-direction (*) and in the Y-direction (o) and fitted variogram models (----) for thresholds log(K) =; a) -7.0, b) -6.5, c) -6.0, d) -5.5, e) -5.0, f) -4.5.

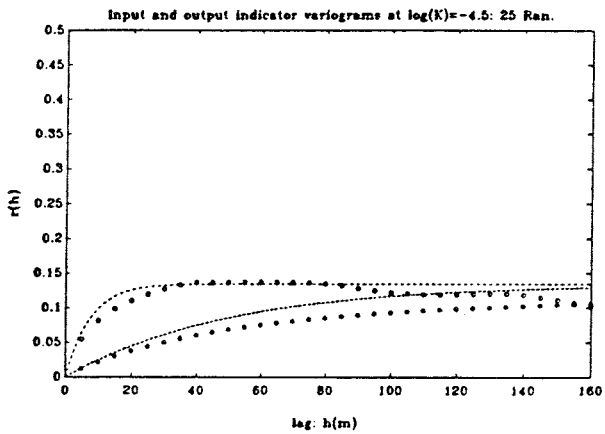




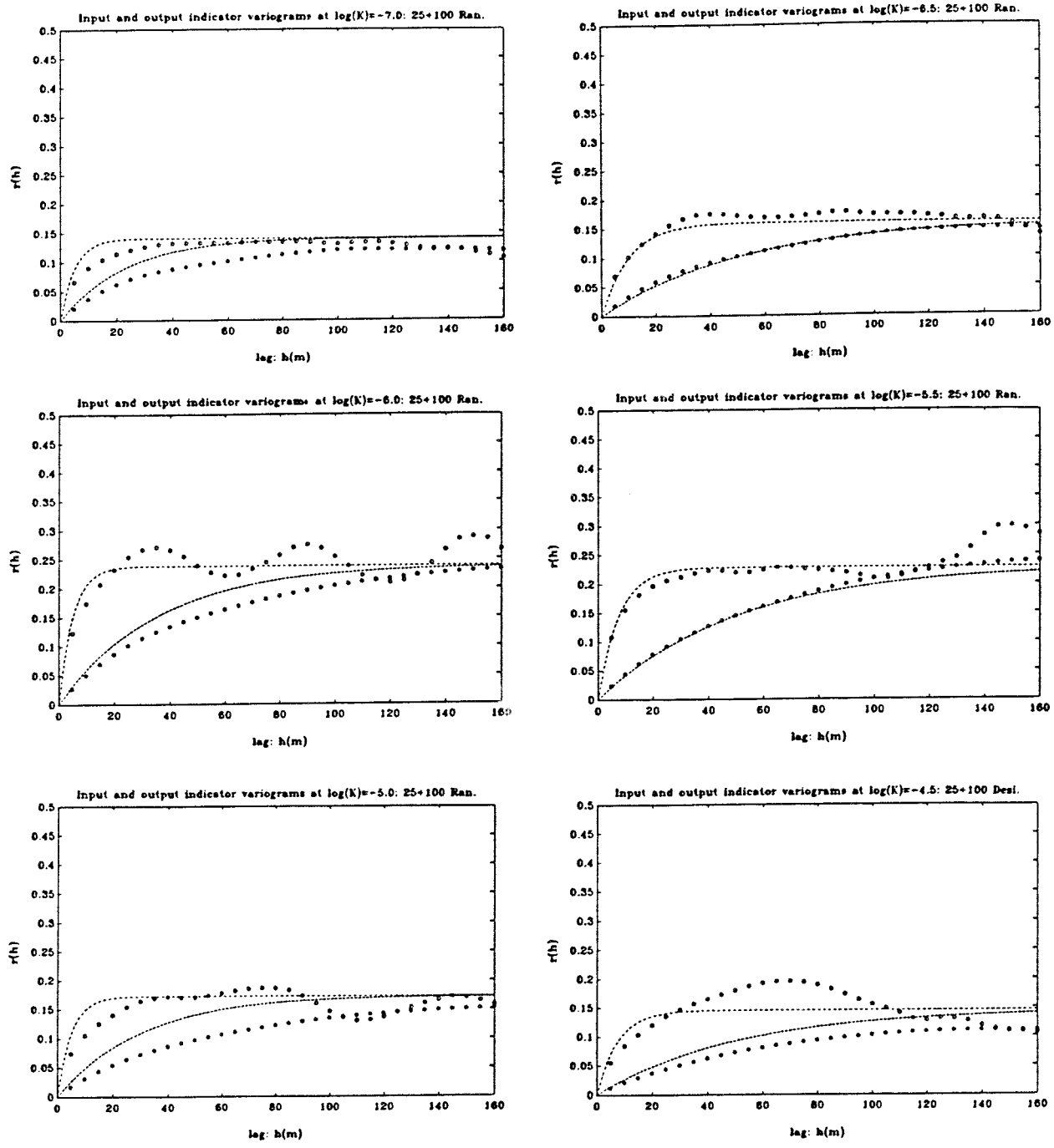




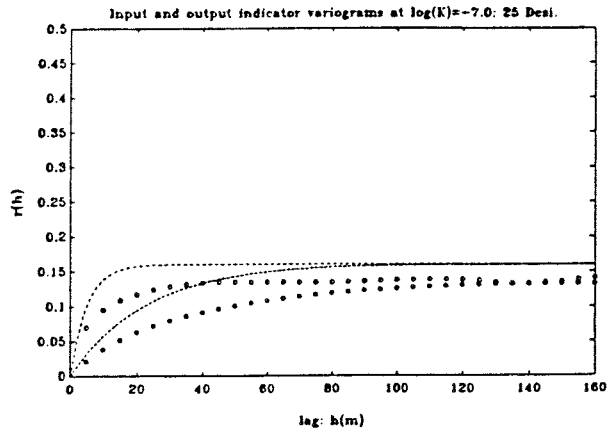


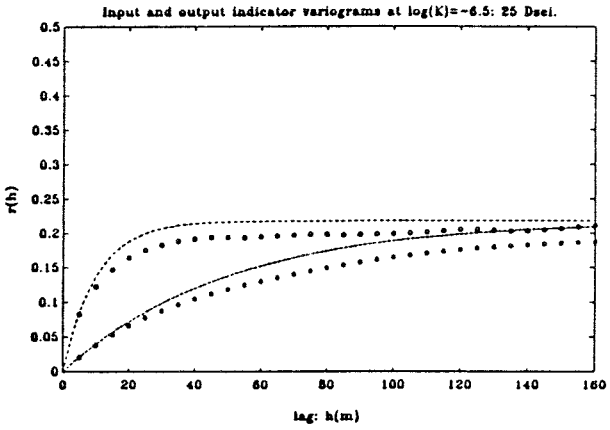


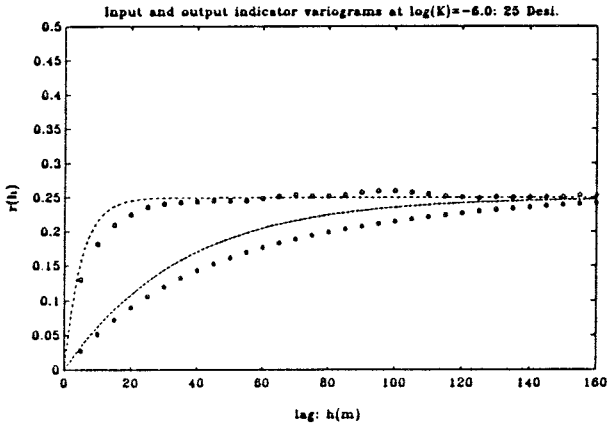
Non-parametric simulation based on a random sample data set of 25 hard data. Comparison between input (reference) (----) and ensemble output indicator variograms in the X-direction (*) and in the Y-direction (o) for thresholds log(K) =; a) -7.0, b) -6.5, c) -6.0, d) -5.5, e) -5.0, f) -4.5.

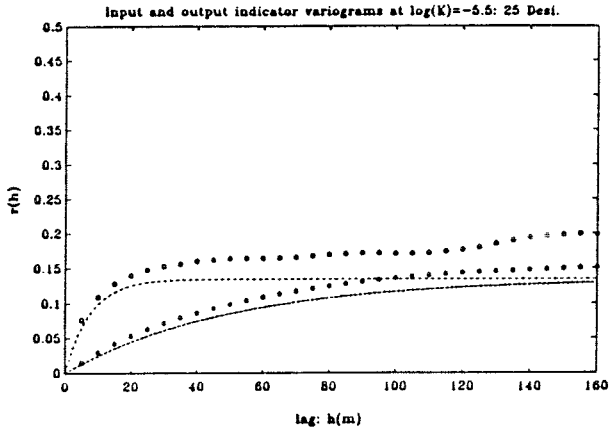


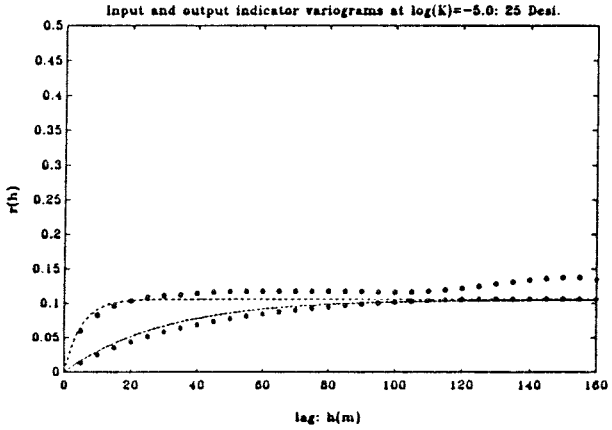
Non-parametric simulation based on a random sample data set of 25 hard + 100 soft data. Comparison between input (reference) (----) and ensemble output indicator variograms in the X-direction (*) and in the Y-direction (o) for thresholds log(K) =; a) -7.0, b) -6.5, c) -6.0, d) -5.5, e) -5.0, f) -4.5.

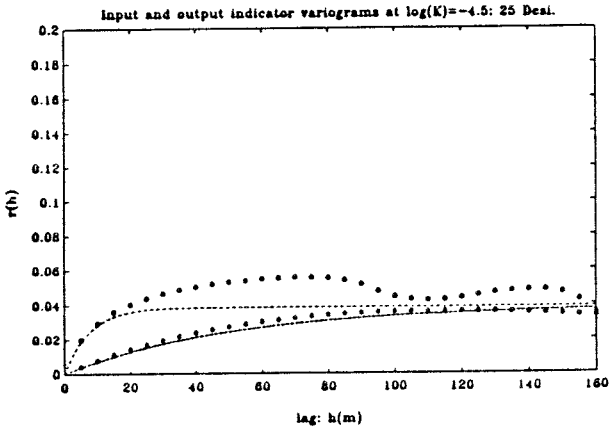




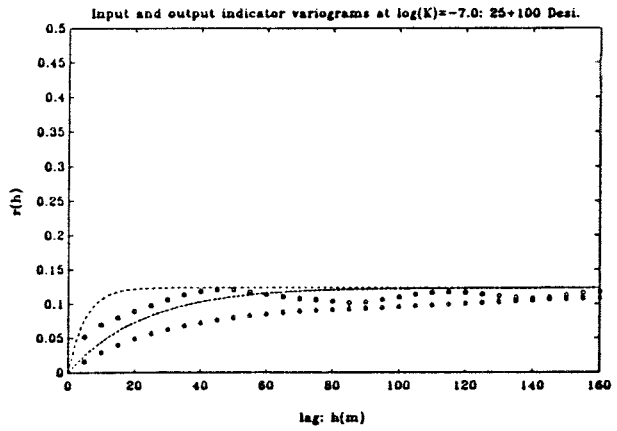


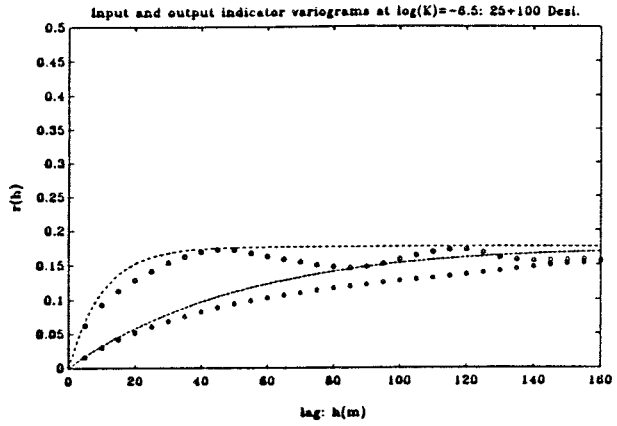


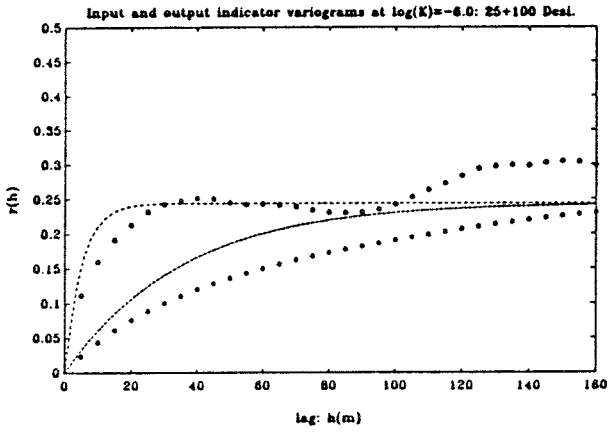


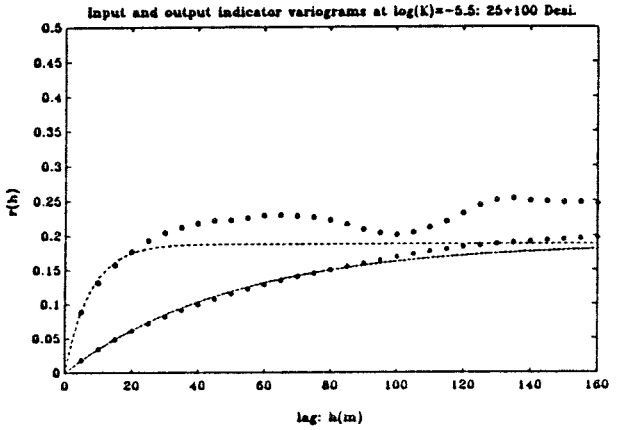


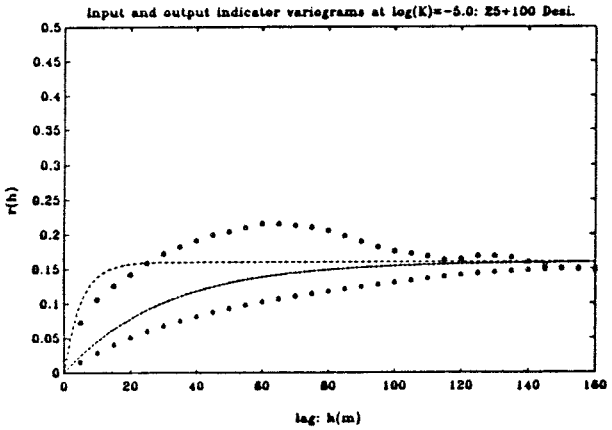
Non-parametric simulation based on a designed sample data set of 25 hard data. Comparison between input (reference) (----) and ensemble output indicator variograms in the X-direction (*) and in the Y-direction (o) for thresholds log(K) =; a) -7.0, b) -6.5, c) -6.0, d) -5.5, e) -5.0, f) -4.5.

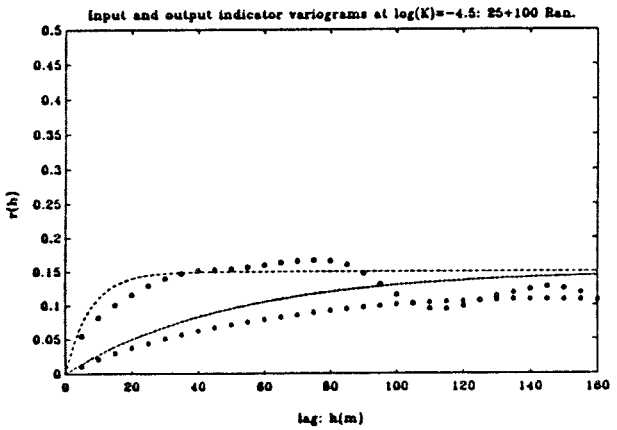












Non-parametric simulation based on a designed sample data set of 25 hard + 100 soft data. Comparison between input (reference) (----) and ensemble output indicator variograms in the X-direction (*) and in the Y-direction (o) for thresholds log(K) =; a) -7.0, b) -6.5, c) -6.0, d) -5.5, e) -5.0, f) -4.5.

List of SKB reports

Annual Reports

1977-78
TR 121
KBS Technical Reports 1 – 120
Summaries
Stockholm, May 1979

1979
TR 79-28
The KBS Annual Report 1979
KBS Technical Reports 79-01 – 79-27
Summaries
Stockholm, March 1980

1980
TR 80-26
The KBS Annual Report 1980
KBS Technical Reports 80-01 – 80-25
Summaries
Stockholm, March 1981

1981
TR 81-17
The KBS Annual Report 1981
KBS Technical Reports 81-01 – 81-16
Summaries
Stockholm, April 1982

1982
TR 82-28
The KBS Annual Report 1982
KBS Technical Reports 82-01 – 82-27
Summaries
Stockholm, July 1983

1983
TR 83-77
The KBS Annual Report 1983
KBS Technical Reports 83-01 – 83-76
Summaries
Stockholm, June 1984

1984
TR 85-01
Annual Research and Development Report 1984
Including Summaries of Technical Reports Issued during 1984. (Technical Reports 84-01 – 84-19)
Stockholm, June 1985

1985
TR 85-20
Annual Research and Development Report 1985
Including Summaries of Technical Reports Issued during 1985. (Technical Reports 85-01 – 85-19)
Stockholm, May 1986

1986
TR 86-31
SKB Annual Report 1986
Including Summaries of Technical Reports Issued during 1986
Stockholm, May 1987

1987
TR 87-33
SKB Annual Report 1987
Including Summaries of Technical Reports Issued during 1987
Stockholm, May 1988

1988
TR 88-32
SKB Annual Report 1988
Including Summaries of Technical Reports Issued during 1988
Stockholm, May 1989

1989
TR 89-40
SKB Annual Report 1989
Including Summaries of Technical Reports Issued during 1989
Stockholm, May 1990

1990
TR 90-46
SKB Annual Report 1990
Including Summaries of Technical Reports Issued during 1990
Stockholm, May 1991

1991
TR 91-64
SKB Annual Report 1991
Including Summaries of Technical Reports Issued during 1991
Stockholm, April 1992

Technical Reports

List of SKB Technical Reports 1992

TR 92-01
GEOTAB. Overview
Ebbe Eriksson[1], Bertil Johansson[2], Margareta Gerlach[3], Stefan Magnusson[2], Ann-Chatrin Nilsson[4], Stefan Sehlstedt[3], Tomas Stark[1]
[1]SGAB, [2]ERGODATA AB, [3]MRM Konsult AB
[4]KTH
January 1992

TR 92-02
Sternö study site. Scope of activities and main results
Kaj Ahlbom[1], Jan-Erik Andersson[2], Rune Nordqvist[2], Christer Ljunggren[3], Sven Tirén[2], Clifford Voss[4]
[1]Conterra AB, [2]Geosigma AB, [3]Renco AB,
[4]U.S. Geological Survey
January 1992

TR 92-03
Numerical groundwater flow calculations at the Finnsjön study site – extended regional area
Björn Lindbom, Anders Boghammar
Kemakta Consultants Co, Stockholm
March 1992

TR 92-04
Low temperature creep of copper intended for nuclear waste containers
P J Henderson, J-O Österberg, B Ivarsson
Swedish Institute for Metals Research, Stockholm
March 1992

TR 92-05
Boyancy flow in fractured rock with a salt gradient in the groundwater – An inititial study
Johan Claesson
Department of Building Physics, Lund University, Sweden
February 1992

TR 92-06
Characterization of nearfield rock – A basis for comparison of repository concepts
Roland Pusch, Harald Hökmark
Clay Technology AB and Lund University of Technology
December 1991

TR 92-07
Discrete fracture modelling of the Finnsjön rock mass: Phase 2
J E Geier, C-L Axelsson, L Hässler,
A Benabderrahmane
Golden Geosystem AB, Uppsala, Sweden
April 1992

TR 92-08
Statistical inference and comparison of stochastic models for the hydraulic conductivity at the Finnsjön site
Sven Norman
Starprog AB
April 1992

TR 92-09
Description of the transport mechanisms and pathways in the far field of a KBS-3 type repository
Mark Elert[1], Ivars Neretnieks[2], Nils Kjellbert[3], Anders Ström[3]
[1]Kemakta Konsult AB
[2]Royal Institute of Technology
[3]Swedish Nuclear Fuel and Waste Management Co
April 1992

TR 92-10
Description of groundwater chemical data in the SKB database GEOTAB prior to 1990
'Sif Laurent[1], Stefan Magnusson[2], Ann-Chatrin Nilsson[3]
[1]IVL, Stockholm
[2]Ergodata AB, Göteborg
[3]Dept. of Inorg. Chemistry, KTH, Stockholm
April 1992

TR 92-11
Numerical groundwter flow calculations at the Finnsjön study site – the influence of the regional gradient
Björn Lindbom, Anders Boghammar
Kemakta Consultants Co., Stockholm, Sweden
April 1992

TR 92-12
HYDRASTAR – a code for stochastic simulation of groundwater flow
Sven Norman
Abraxas Konsult
May 1992

TR 92-13
Radionuclide solubilities to be used in SKB 91
Jordi Bruno[1], Patrik Sellin[2]
[1]MBT, Barcelona Spain
[2]SKB, Stockholm, Sweden
June 1992

TR 92-14
Numerical calculations on heterogeneity of groundwater flow
Sven Follin
Department of Land and Water Resources, Royal Institute of Technology
June 1992

TR 92-15

Kamlunge study site - Scope of activities and main results

Kaj Ahlbom[1], Jan-Erik Andersson[2], Peter Andersson[2], Thomas Ittner[2], Christer Ljunggren[3], Sven Tirén[2]

[1] Conterra AB

[2] Geosigma AB

[3] Renco AB

May 1992

TR 92-16

Equipment for deployment of canisters with spent nuclear fuel and bentonite buffer in horisontal holes

Vesa Henttonen, Miko Suikki

JP-Engineering Oy, Raisio, Finland

June 1992

TR 92-17

The implication of fractal dimension in hydrogeology and rock mechanics. Version 1.1

W Dershowitz[1], K Redus[1], P Wallmann[1], P LaPointe[1], C-L Axelsson[2]

[1] Golder Associates Inc., Seattle, Washington, USA

[2] Golder Associates Geosystem AB, Uppsala, Sweden

February 1992