CHAPTER II

THEORIES AND CONCEPTS

This chapter presents the theory and technique used in application of geostatistical simulation for permeability determination from well logging data. It involves two steps, which are geostatistical simulation and permeability equation from porosity.

2.1 Geostatistical method

Geostatistics is a method that describes and models the spatial variability of reservoir properties and the spatial correlation between related properties. The method can be used to the construct spatial distribution of reservoir properties from some sampled locations. The key feature of such techniques is its ability to account for spatial character of the data and incorporate many types of geological and engineering data. Spatial continuity is known to be important in describing reservoir parameters since reservoir properties are correlated in space. Lack of spatial information seriously decreases the accuracy of estimated properties. In general, it is likely to increase the accuracy of reservoir properties when using geostatistics.

In addition, the important characteristic of geostatistics is probabilistic aspect which can describe more heterogeneity in reservoir as well as generating the distribution of reservoir parameters. Another aspect of geostatistics is the estimation of petrophysical properties at unsampled location. It means that geostatistics can recognize and preserve spatial relationships of sample data for estimation of data values or describe uncertainty at unsampled locations.

The porosity is the reservoir property that related to reservoir connectivity. This variable is spatial interdependent and its characteristic can be incorporated into the simulation model.

Geostatistics plays a role in describing the spatial variable that is an essential feature of many natural phenomena and provides adaptations of classical regression techniques to take advantage of this variable. Geostatistics has many techniques that use to analyze or estimate spatial variable, i.e. variogram, cross validation, Kriging estimation, Gaussian simulation, and etc.

Several geostatistical tools are adopted in this study, which are variogram analysis, cross validation and Sequential Gaussian Simulation. Therefore, the steps of analysis can be classified as statistical analysis, structural analysis, cross validation, and Sequential Gaussian Simulation, respectively. The geostatistical model flow chart is presented in Figure 2.1.

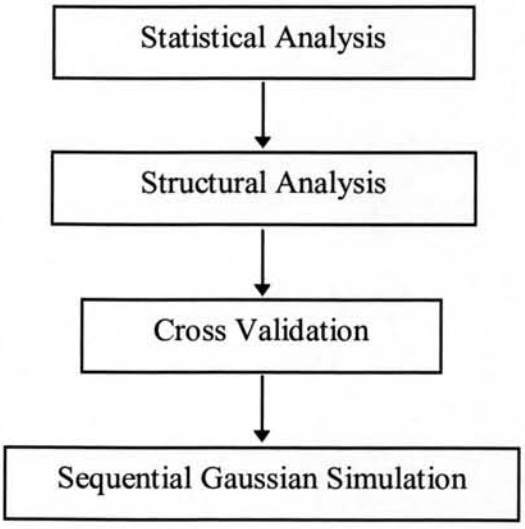

Figure 2.1: Geostatistical model flow chart

2.1.1 Statistical analysis

In statistical analysis, there are tools to analyze and summarize in order to measure certain characteristics of variable such as a measure of location (mean, median, minimum, maximum and quartiles), a measure of spread (variance and standard deviation) and a measure of shape (coefficient of skewness, kurtosis and coefficient of variation). The statistical tools used in this study consist of location plot, histogram plot, and probability plot.

Location plot is the map that exhibits the positions and values of sample in X-Y-Z directions.

Histogram is a frequency distribution that used to present the observed values from a sample set for a particular variable. The histogram rapidly and pictorially provides a good representation of the shape of data's distribution. The histogram displays how often observed values fall within certain intervals or classes. The variables encountered in data sets may assume a number of different distributional shapes and forms.

Probability plot is some petrophysical of the estimation tool to present the distribution of data. Probability plots consist of probability density function and cumulative probability distribution function that can display type of distribution.

2.1.2 Structural analysis

Structural analysis is the most commonly used geostatistical technique for describing the spatial relationship. It involves variogram construction and modeling. Mathematically, the variogram calculation is defined as:

$$\gamma(\vec{h}) = \frac{1}{2N(h)} \sum_{i=1}^{N} [Z(x_i) - Z(x_i + \vec{h})]^2 \qquad (2.1)$$

where

$\gamma(\vec{h})$ = variogram value at distance $\vec{h}$

$N(h)$ = number of data pair

$[Z(x_i) - Z(x_i + \vec{h})]$ = the difference in value between two sample points separated by distance $\vec{h}$

Equation 2.1 is used to calculate the variogram value at any distance ($\vec{h}$) and direction. The plot of variogram values against distance ($\vec{h}$) along a specific direction presents the spatial variability structure of that variable. Figure 2.2 presents a variogram plot and basic components of variogram model.

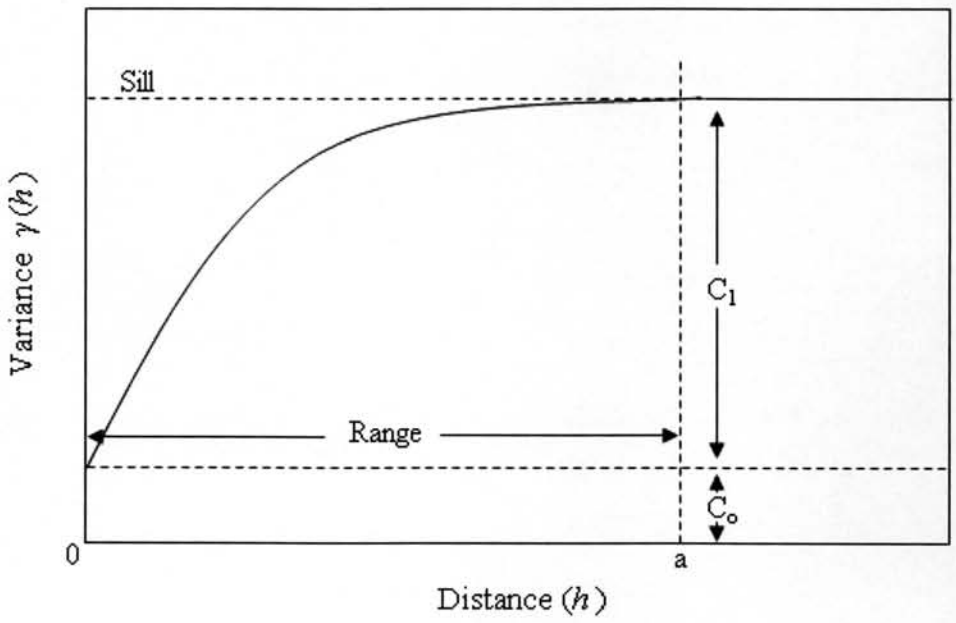

Figure 2.2: Basic components of variogram.

As seen in Figure 2.2, three main components of variogram model are sill, range and nugget. Sill is the maximum variance of variable and theoretically equal to the data variance. Range is the maximum distance in which data still have correlation. The variation at small scale and should be zero at zero distance is called nugget. But nugget in practice comes from two sources, measurement error and small scale variation.

To determine the spatial correlation structure of a variable, the experimental point variogram values at given distances and direction are calculated using Equation 2.1. After that, the variogram model is fitted into the experimental point variogram values. In practice, the goodness of fit is observed by a visual inspection. There are many variogram models to select for variogram modeling such as power, spherical, exponential, Gaussian model, etc. For this study, only the equation and characteristics of interested variogram models (spherical, exponential and Gaussian model) are described.

(i) Spherical Model

The spherical model is, in all probability, the most frequently used function for modeling variograms. The shape of spherical model is represented in Figure 2.3. The beginning portion of the graph, indicating small separation distances, display a linear component. This linear behavior is lost at larger distances, and the curve flattens, eventually reaching a plateau. A line tangent at the origin will intersect the sill at two-thirds the range. The equation for the spherical model is described as follows:

$$\gamma(h) = \begin{cases} C_o + C_1 \left[\frac{3}{2}\frac{h}{a} - \frac{1}{2}\frac{h^3}{a^3} \right] & h \le a \\ C_o + C_1 & h > a \end{cases} \qquad (2.2)$$

where a is the range, C_o is the nugget effect, and $C_1 + C_o$ is the sill.

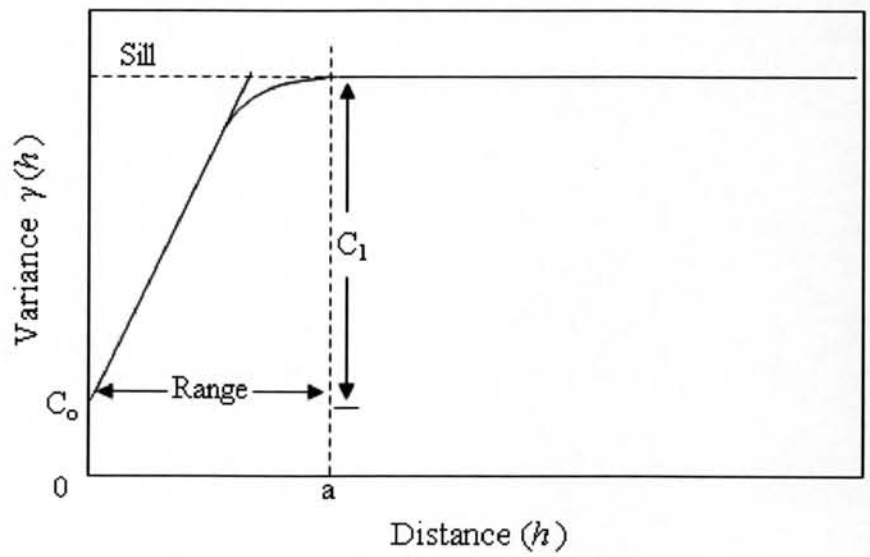

Figure 2.3: Spherical variogram model.

(ii) Exponential Model

The shape of exponential model is very close to spherical model but represents a more continuous process. Its shape is represented in Figure 2.4. It rises more steeply from the origin than spherical model but then flattens at a more leisurely rate. The equation and definitions of the exponential model are as follows:

$$\gamma(h) = \begin{cases} C_o + C_1\left[1 - \exp(-\frac{h}{a})\right] & h \le a \\ C_o + C_1 & h > a \end{cases} \quad (2.3)$$

where a is the range, C_o is the nugget effect, and $C_1 + C_o$ is the sill.

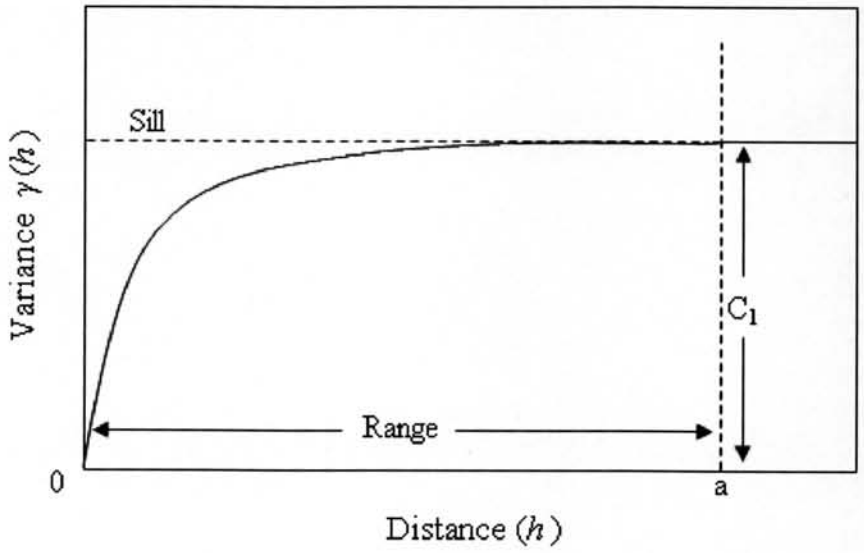

Figure 2.4: Exponential variogram model.

(iii) Gaussian Model

Gaussian model is typically used to model processes that show extremely low variation at short distances, to the extent that the model is parabolic at the origin. The shape of this model is represented in Figure 2.5. The equation and definition of the Gaussian model are as follows:

$$\gamma(h) = \begin{cases} C_o + C_1\left[1 - \exp(-\frac{h^2}{a^2})\right] & h \le a \\ C_o + C_1 & h > a \end{cases} \quad (2.4)$$

where a is the range, C_o is the nugget effect, and $C_1 + C_o$ is the sill.

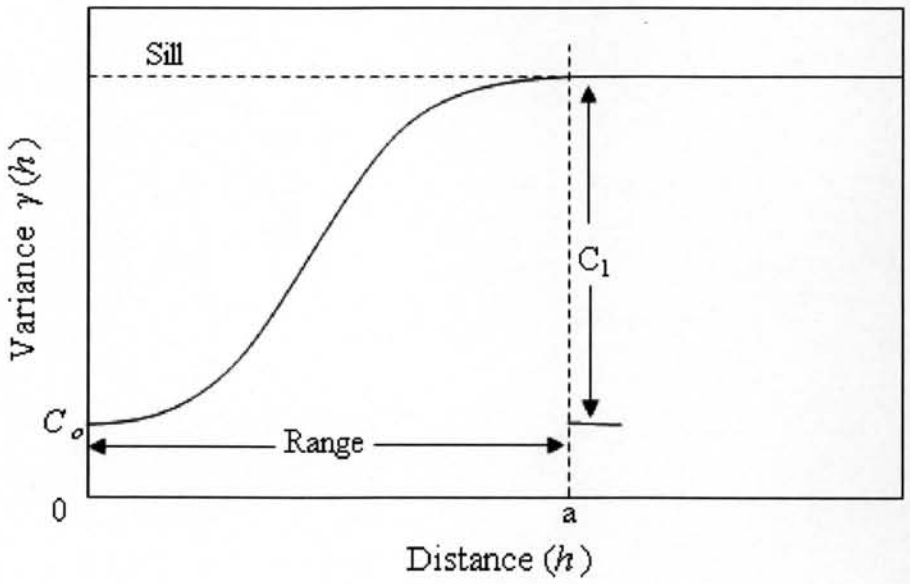

Figure 2.5: Gaussian variogram model.

The variogram calculation depends on sample distribution. Two parameters which are angle and lag distance are introduced to calculate variogram values in less restrictive way. Angle is the direction of variogram calculation. Lag distance is the distance (h) for variogram calculation. This process can be applied well to regularly spaced data but not the irregularly spaced data because there are few pairs of data in irregularly spaced data that have the same distance. Thus, tolerlance angle, bandwidth, and lag tolerance are introduced to find the more number of calculated pairs. Tolerance angle is an angle that deviates from the observed direction. However, at a large distance, the direction of interest loses its consistency. Therefore, bandwidth is introduced because bandwidth is the offset of the direction of interest. It keeps the pairs of sample that are located between the direction vector and bandwidth lines. Finding the variogram value at any distances, lag distance is specified first and the real distance is then calculated from the average distance in each lag. Figure 2.6 shows the variogram parameters such as angle, angle tolerance, lag distance, and bandwidth.

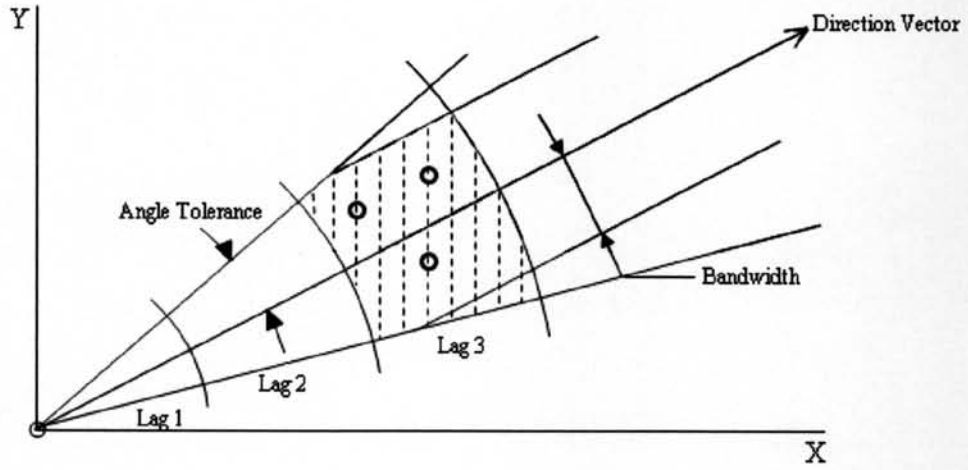

Figure 2.6: Variogram parameters (angle tolerance, lag distance, and bandwidth).

Generally, variogram is a significant tool that is necessary for estimation and simulation process. The next section describes the cross validation process that verifies the reliability of the variogram model.

2.1.3 Cross Validation

Cross validation is a technique to prove the reliability of the variograms from structural analysis by comparing the estimated and true values using only the information available in sample data set. In cross validation, the estimation method is tested at the locations of existing samples. The sample value at a particular location is temporarily discarded from the sample data set. The value at the same location is then estimated using the remaining samples. The procedure, shown in Figure 2.7, can be seen as an experiment in which we mimic the estimation process by pretending that we had never sampled a certain location. Once the estimate is calculated and compared to the true sample value that was initially removed from the sample data set. This procedure is repeated for all available samples. The resulting true and estimated values can then be compared using the statistical and visual tools.

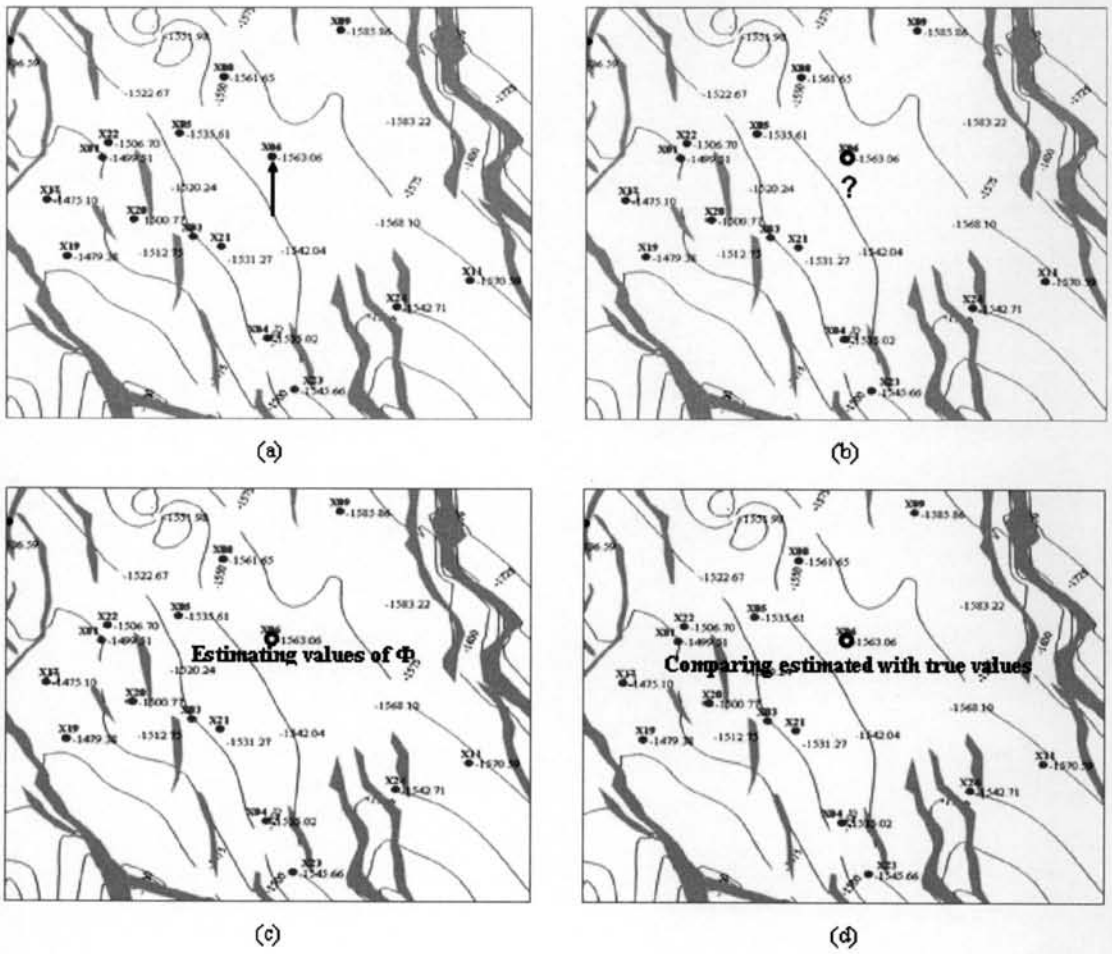

Figure 2.7: An example of cross validation procedure. (a) Sample at selected location being removed, (b) Sample after removed, (c) Sample being estimated, and (d) Comparing estimated with true value.

For this study, cross validation is used to evaluate the differences between estimated calculated by Ordinary Kriging. Ordinary Kriging, is the one type of Kriging estimation and remains an essential tool for geostatistical estimation technique. The distinguishing feature of Ordinary Kriging is minimizing the error variance.

Ordinary Kriging seeks an estimate of a property Z^* at an unmeasured location (point 0) that is based on the $Z_{(xi)}$, $i = 1,\ldots,n$ known values.

$$Z^* = \sum_{i=1}^{n} \lambda_\alpha Z_{(xi)} \qquad (2.5)$$

Equation (2.5) is a common interpolator. The values λ_α, α=1,…,n, are the kriging weights. Clearly, if the λ_α was determined, the Z^* will be calculated.

From the distinguishing feature of Ordinary Kriging to minimize the error variance, the bias can be removed by forcing the λ_α to sum to one as Equation (2.6) below.

$$\sum_{\alpha=1}^{n} \lambda_\alpha = 1 \qquad (2.6)$$

Equation (2.7) is the equation for Ordinary Kriging. μ is the Lagrange multiplier.

$$\sum_{\alpha=1}^{n} \lambda_\alpha \bar{\gamma}(v_\alpha, v_\beta) + \mu = \bar{\gamma}(v_\alpha, \nabla) \quad ; \beta = 1,..,n \qquad (2.7)$$

where

n = number of data

∇ = point to be estimated

v_α = sample at location x_α

which can be written in the following matrix form:

$$\begin{bmatrix} \bar{\gamma}(v_1,v_1)... & \bar{\gamma}(v_1,v_\beta)... & \bar{\gamma}(v_1,v_n) & 1 \\ \bar{\gamma}(v_\beta,v_1)... & \bar{\gamma}(v_\beta,v_\beta)... & \bar{\gamma}(v_\beta,v_n) & 1 \\ \bar{\gamma}(v_n,v_1)... & \bar{\gamma}(v_n,v_\beta)... & \bar{\gamma}(v_n,v_n) & 1 \\ 1 \quad ... & 1 \quad ... & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ ... \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \bar{\gamma}(v_1,\nabla) \\ ... \\ \bar{\gamma}(v_n,\nabla) \\ 1 \end{bmatrix}$$

The variogram values, $\gamma(\vec{h})$, can be calculated from Equations 2.2, 2.3 and 2.4 depend on variogram model. The Kriging weights can be obtained after solving the matrix above. And then substituting the Kriging weights in Equation 2.5, the estimated value is known.

As mention earlier, cross validation is used to compare the estimated value with the true value, so the error distribution is very necessary for this study. The next section describes the error and its distribution.

The true values can be defined as Z_1, Z_2... Z_n, and the estimated values can be defined as Z_1^*, Z_2^*... Z_n^*. The estimation error can be defined as the true value minus the estimate:

$$e = Z - Z^* \tag{2.8}$$

Now that actual errors are being calculated, the concepts of precision, accuracy and bias become increasingly important. Precision is a measure of the variability between individual estimates or measurements and indicates how closely individual calculations agree with one another. Accuracy, summarizes the closeness of measurements to the true value, indicates how close to the target center measurements or estimates come. Bias is defined as any systematic or consistent over- or underestimation of true values of the quantity being estimated.

The average of the estimate should not deviate from the actual value because deviation indicates a lack of accuracy. In fact, it will be much more satisfying if there is no systematic error or bias in the estimates.

For error distributions, it is useful to apply the histogram plot. Figure 2.8 shows three identical histogram plots, each representing the distribution of the errors made during an estimation process. The distribution in Figure 2.8a has an average value of less than zero, indicating that a bias toward underestimation exists. Conversely, the distribution of error in Figure 2.8b has a positive average, indicating overestimation. Figure 2.8c presents a more balanced and acceptable situation, where the average value is equal to zero. It should be noted that each of these diagrams shows the same precision for the error distribution. Only Figure 2.8c can be considered unbiased and also presented more accurate than the others.

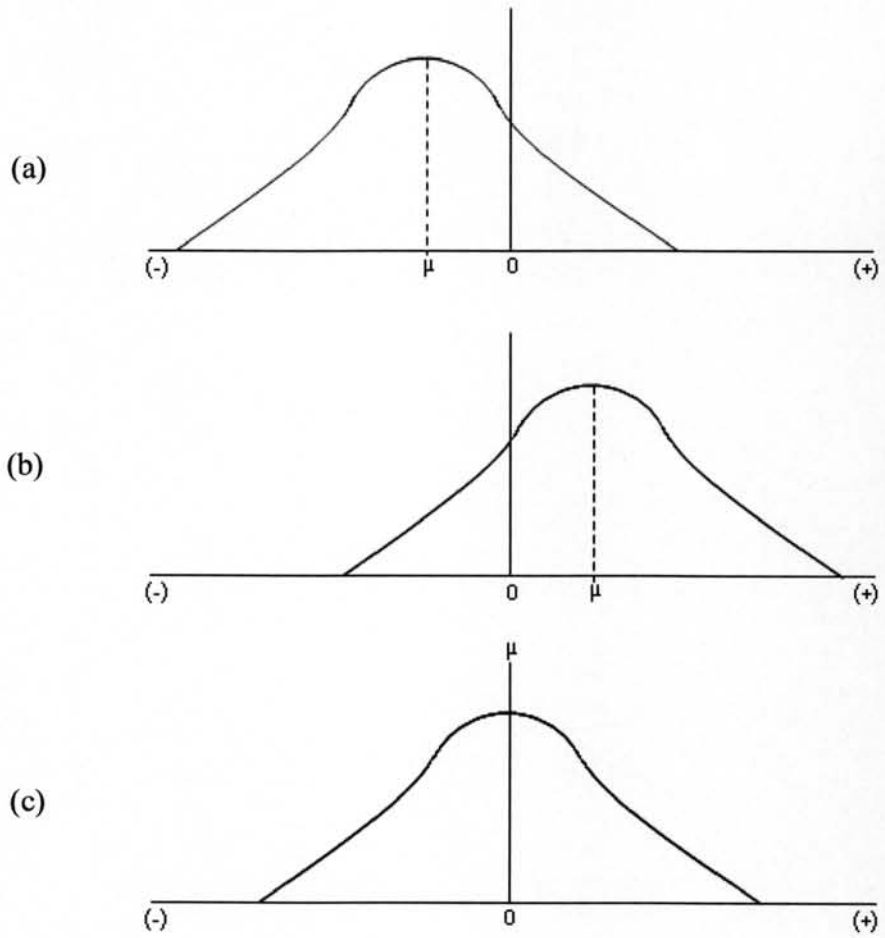

Figure 2.8: Examples of bias in error distributions: (a) negative bias, (b) positive bias, and (c) no bias.

Figure 2.9 is the histogram visualization of precision. Each error distribution has a mean of zero, indicating unbiasness. However the width or spread of the distribution that indicates the variance is smaller in Figure 2.9a than in Figure 2.9b. The smaller variance in error distribution indicates the higher precision of the estimation model, and thus the high reliability of the input variogram model.

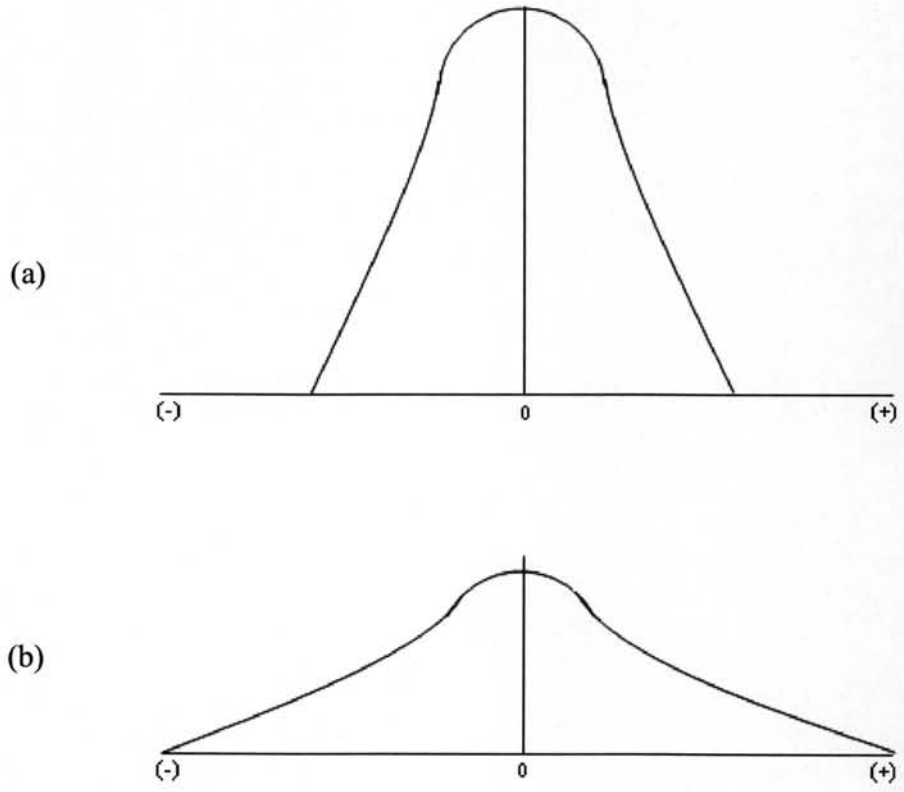

Figure 2.9: Example of precision in error distribution: (a) The narrower distribution indicates greater precision; (b) the wider distribution indicates less precision

2.1.4 Sequential Gaussian Simulation (SGS)

Sequential Gaussian Simulation is a procedure for estimating the reservoir characteristics between data points. Based on the idea of iterating from a first guess and refining through reduction of errors, the procedure generally transforms the model to normality, simulating the normally distributed transform, and then back-transforming to the original variable of interest.

The objective in simulation is to generate multiple realizations of the map each of which would honor the actual data, and approximately the same variogram and distribution.

Because the assumption of Sequential Gaussian Simulation states that the study variable has to follow standard normal distribution with zero mean and unit variance, normal score transform needs to be explained.

The process of transforming original data to standard normal data is carried out using normal score transform function. Let Z and Y be the two data sets and their *cdf* (Cumulative Distribution Function) are $F_Z(z)$ and $F_Y(y)$, respectively. The transform $Y = \phi(Z)$ identifies the cumulative probabilities, which correspond to the Z and Y p-quantiles:

$$F_Y(y_p) = F_Z(z_p) = p, \forall p \in [0, 1] \tag{2.9}$$

Thus,

$$y = F_Y^{-1}(F_Z(z)) \tag{2.10}$$

with $F_Y^{-1}(\cdot)$ being the inverse *cdf*, or quantile function, of Y data set:

$$y_p = F_Y^{-1}(p), \forall p \in [0,1] \tag{2.11}$$

In case that Y is standard normal with *cdf* $F_Y(y) = G(y)$, the transform $G^{-1}(F_Z(\cdot))$ is the normal score transform. Figure 2.10 presents the graphical transformation from real data set to normal score data.

Sequential Gaussian Simulation (SGS) is a procedure that uses the Kriging mean variance to generate realization of a multivariate Gaussian field. By assuming multivariate normal distribution and sequential simulation approach, SGS algorithm is simulated variable value one cell node after another in a sequential manner, subsequently using these values as the conditional data. The procedure is the following:

1. Transforms the data set into a standard normal distribution.
2. Constructs variogram analysis and fits a proper model for the data set.
3. Selects grid node at random.
4. Performs Kriging to estimate the mean and variance values at that node location.
5. Draws a simulated data from the distribution, and adds the simulated data to the data set.

6. Selects another grid node at random and repeats the procedure for Kriging until all grid nodes are simulated.
7. Back transform to the original distribution, the realization map is created.
8. Repeats for all the other realizations using a different random number sequence to generate multiple realization maps.

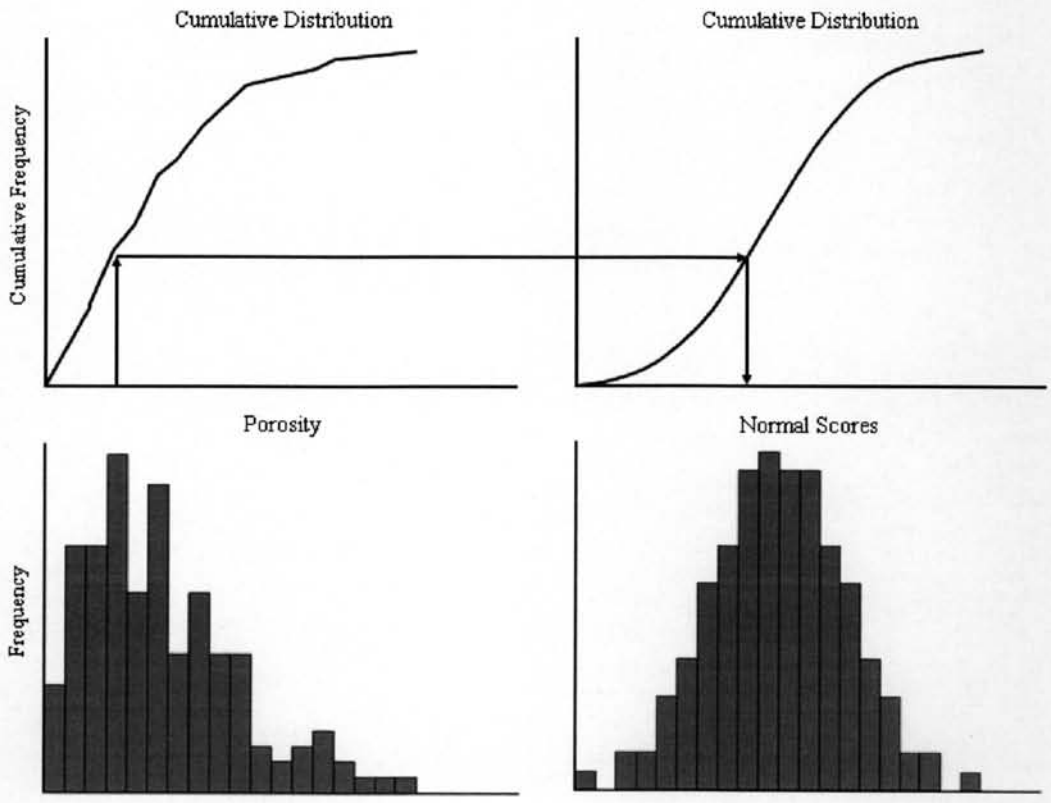

Figure 2.10: Graphical display of normal score transform

Sequential Gaussian Simulation is a high performance to simulate values at unsampling data locations. Figure 2.11 shows the flow chart of Sequential Gaussian Simulation procedure. Practically, this technique consists of several calculations so using computer program is necessary to speed up the computation process.

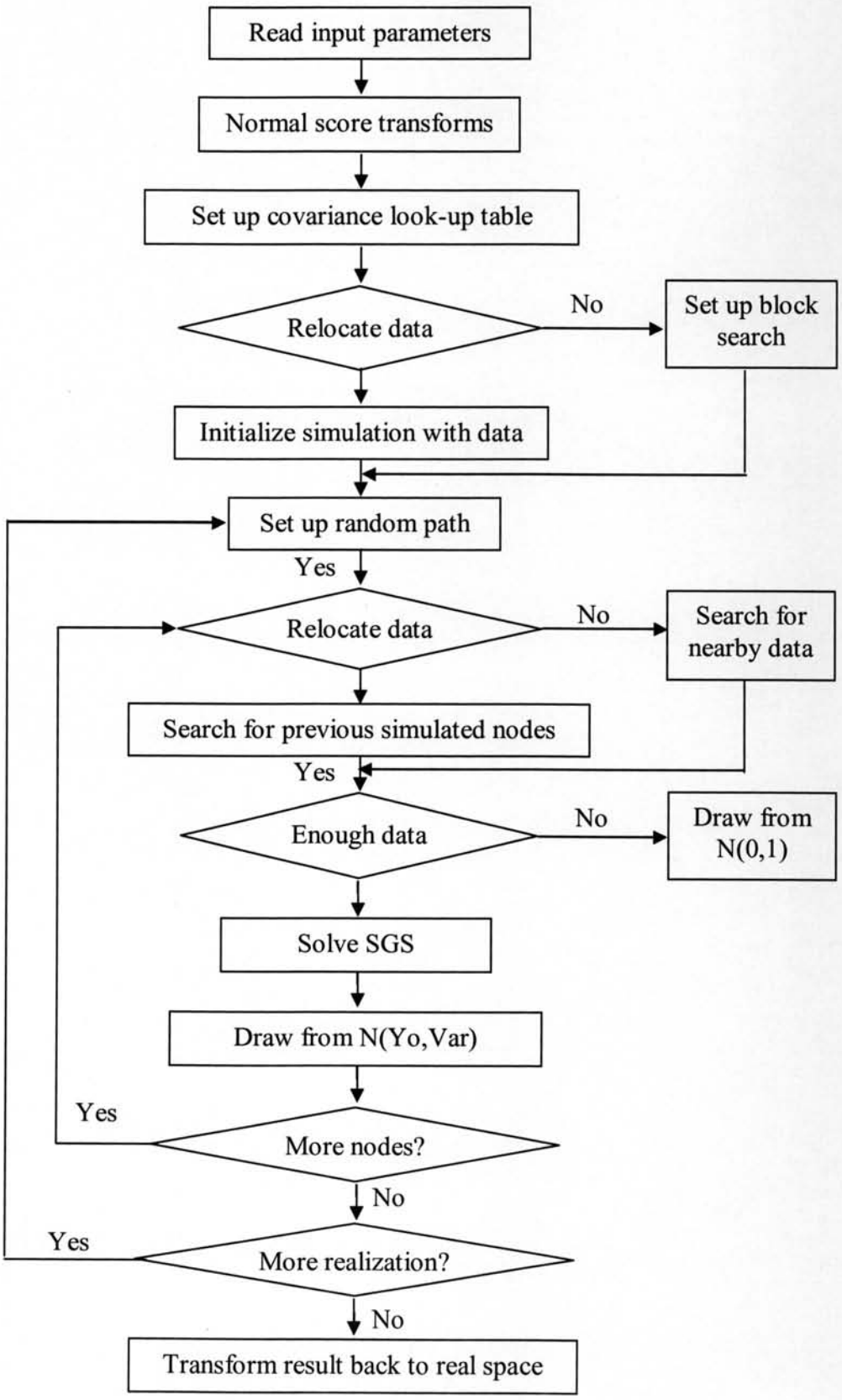

Figure 2.11: Sequential Gaussian Simulation (SGS) flow chart.

2.2 Permeability equation from porosity

Estimation of permeability is based on assumption that porous medium can be represented by a bundle of capillary tubes. Combination of Darcy's law and Poiseuille's law for straight cylindrical tubes yields following relationship:

$$k = \frac{r^2}{8}\Phi \tag{2.12}$$

Equation 2.12 shows that the relationship between porosity and permeability depends on geometrical characteristics of the pore space.

2.2.1 Kozeny-Carman Equation

For a realistic porous medium, Kozeny and Carman have modified Equation 2.12 by adding a tortuousity factor (τ) and using mean hydraulic radius, expressed through surface area per unit grain volume Sgv. The resulting generalized of Kozeny-Carman equation has the following form:

$$k = \frac{\Phi^3}{(1-\Phi)^2}\frac{1}{F_s \tau^2 Sgv^2} \tag{2.13}$$

where F_s is the shape factor, k is permeability in μm^2 and Φ is porosity in fraction.

The group $F_s \tau^2$ is known as the Kozeny constant and was a main limitation in previous attempts to use Equation 2.13 for permeability calculation. Because actual values of the Kozeny constant are usually not known for particular rocks, and the term of surface area per unit grain volume was not accounted in calculations. The hydraulic flow unit approach is used to solve this problem.

2.2.2 Hydraulic Flow Unit Concept

In 2000, the concept of hydraulic flow unit has been used in the petroleum industry to improve prediction of permeability in uncored wells by classification of rock types and prediction of flow properties, based on sensible geological parameters and the physics of flow at pore scale.

As mentioned earlier in Equation 2.13, the HFU approach addresses variability of the Kozeny constant and surface area per unit grain volume term by classificying parameter flow zone indicator, which includes all major geological and geometrical characteristics of a porous medium.

$$F_{zi} = \frac{1}{\sqrt{F_s}\tau Sgv} \tag{2.14}$$

The central idea of the HFU approach is to group data according to the flow zone indicator values and other petrophysical information using various statistical and graphical methods to obtain classification of the HFU. HFU distribution in uncored wells is then estimated using relations between HFU and flow zone indicator and various responses.

Equation 2.13 can be written in field units (i.e., permeability is in md) and combined with Equation 2.14 as the following form:

$$0.0314\sqrt{\frac{k}{\Phi}} = \frac{\Phi}{(1-\Phi)} F_{zi} \tag{2.15}$$

where the constant 0.0314 is the conversion factor from μm^2 to md.

To simplify graphical analysis of the data, two additional parameters, reservoir quality index (RQI) and ratio of pore volume to grain volume (Φ_z) are defined.

$$RQI = 0.0314\sqrt{\frac{k}{\Phi}} \tag{2.16}$$

$$\Phi_z = \frac{\Phi}{(1-\Phi)} \tag{2.17}$$

Using the parameters earlier, Equation 2.15 can be rearranged as follows:

$$\log RQI = \log \Phi_z + \log F_{ZI} \tag{2.18}$$

According to Equation 2.18, a log-log plot of RQI vs. Φ_z yields a straight line with unit slope and the intercept equal to flow zone indicator. Hence, all data points corresponding to one HFU should lie on the same line.

After calculating pore-throat related parameters of reservoir quality index and flow zone indicator from core information, hydraulic flow unit can be identified based on flow zone indicator values. Although there should exist one single flow zone indicator value for each hydraulic flow unit, a distribution for each flow zone indicator around its true mean results because of random measurement errors. When multiple hydraulic flow unit groups exist, the overall flow zone indicator distribution function is a superposition of the individual distribution functions around their means flow zone indicator. Identification of each mean flow zone indicator, or the corresponding hydraulic flow unit, would require decomposition of the overall flow zone indicator distribution into its constituting elements. This is a desuperposition problem and cluster analysis techniques allow for such a decomposition process.

Two approaches of cluster analysis, histogram analysis and probability plot, are discussed in the next section.

2.2.3 Graphical Clustering Method

Graphical clustering methods of histogram analysis and probability plot provide a general visual image of flow zone indicator distribution to determine the number of hydraulic flow units, their mean values, and their distribution types.

(i) Histogram Approach

Because flow zone indicator distribution is a superposition of multiple log-normal distributions, a histogram of flow zone indicator should show "n" number of normal distributions for "n" number of hydraulic flow units. The convolved frequency distribution for a mixture of multiple Gaussian probability density functions is described by;

$$f = \sum_{i=1}^{N_U} \frac{\omega_i}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(z - \bar{z}_i)^2}{2\sigma_i^2}\right] \qquad (2.19)$$

and,

$$\sum_{i=1}^{N_U} \omega_i = 1 \qquad (2.20)$$

where ω_i is the weight factor. In general, ω_i are themselves from a probability density function. For equally weighted distributions,

$$\omega_i = \frac{1}{N_U} \quad \text{for all } i = 1, \ldots, N_U \tag{2.21}$$

When clusters are distinctly separate, the histogram clearly delineates each hydraulic flow unit and provides their corresponding flow zone indicator values.

(ii) Probability Plot Approach

The probability plot or cumulative distribution function, is the integral of histogram, is a smoother plot than the histogram. The scatter in data is reduced on this and the identification of clusters becomes easier. The cumulative distribution function is given by Equation 2.22 below.

$$F = \frac{1}{2}\left[1 + \sum_{i=1}^{N_U} \omega_i \operatorname{erf} \frac{(z - \bar{z}_i)}{2\sigma_i}\right] \tag{2.22}$$

A normal probability plot has a specially arranged coordinate system where a normal distribution forms a distinct straight line. Hence, the number of straight lines and the limiting boundary values of flow zone indicator or each hydraulic flow unit can be obtained from the probability plot of log (F_{zi}). Because means of flow zone indicator values are not calculated from probability plots, the representative of flow zone indicator value for each hydraulic flow unit is obtained by averaging all the flow zone indicator values within the corresponding hydraulic flow unit limits. This exactly corresponds to a linear least-square regression of data where the slope of the regression line is equal to unity. The probability method is more useful than the histogram method because it is easier to identify straight lines visually.

Two main theories and concepts of geostatistical model and permeability equation from porosity that are described in this chapter will use in chapter III and chapter IV. The procedure of this study is shown in Figure 2.12.

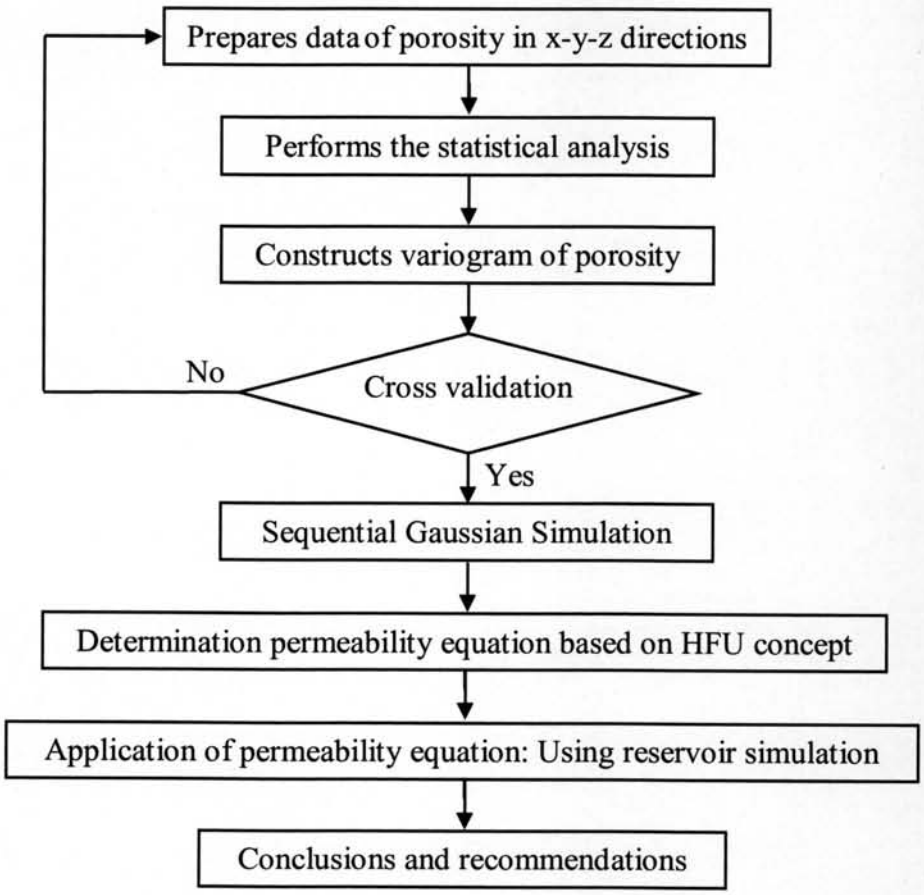

Figure 2.12: Procedure of the study.