IWTC- 5th International Water Technology Conference, Alexandria, Egypt 2000. Pp. 171-192

Stochastic Models for Characterizing Variability of Hydro-Geological Parameters

Amro M.M. Elfeki
Lecturer at Irrigation and Hydraulics Dept.,
Faculty of Engineering,
Mansoura University, Egypt

Abstract

This review highlights some of the progresses achieved in modelling spatial variability of hydro-geological parameters using the stochastic approach. The reason for this approach is two folds. First, the erratic nature of the hydro-geological parameters observed at field data. Second, the uncertainty due to the lack of information about the subsurface structure that is known only at sparse sampled locations. In this paper the frequently used techniques for unconditional and conditional stochastic models in porous formations have been reviewed. These techniques are the tools to solve stochastic differential equations of flow and transport in porous media either analytically or numerically. In the presented techniques, spatial variability of aquifer parameters is modelled as stationary correlated random fields. These fields are used as input to the classical differential equations to evaluate the effect of spatial variability on flow and transport characteristics. The algorithms used for site characterization and generating heterogeneous descriptions of porous formations are reviewed. These algorithms can be divided broadly into discrete facies models, continuous models and hybrid models. In discrete facies models, the heterogeneity is described as geometrical shapes of geological units, parameter values to which can later be assigned. In continuous models, the heterogeneity is described as random parameters governed by joint probability distributions. In hybrid models both continuous and discrete models are used.

The Monte-Carlo approach is a powerful tool to estimate uncertainty. It handles realistic situations where the flow is controlled by complex boundary conditions in comparison with analytical approaches. It handles the non-linearities of the systems. An extensive description of some numerical methods such as multivariate method, nearest neighbour method, and turning bands method has been presented in this paper. An evaluation of these methods is summarized in a tabular form (see Table (1)). Advantages, disadvantages and limitations of each of these methods are presented. The methods are evaluated on the basis of information available in the literature and by application of these methods to some simple characteristic problems.

1. Introduction

Observations of geological media show significant spatial variability at various scales. Geological formations contain different features of variability such as faults, fractures, lithologies, horizontal layering, inclined bedding, inclusions, lamination and variability of the physical and chemical properties within individual units. Field measurements of hydraulic properties such as hydraulic conductivity, porosity, storativity, dispersivity, etc. exhibit a large degree of spatial variability [Freeze, 1975; Delhomme, 1979; Gelhar, 1986]. The observed variability suggests that it may be useful to describe such parameters in a stochastic context rather than in the traditional deterministic one. The traditional deterministic approach in which the aquifer properties are represented as a unique local parameter throughout the entire flow domain or represented by multi-layered system each of that can be characterized by distinct parameter is not realistic in many geological settings. In reality, subsurface hydrogeological parameters rarely possess uniform properties; on the contrary, their properties usually vary in a discrete or continuous manner on a multiplicity of scales from one location to another. The often encounters random spatial fluctuations cannot be adequately described by smooth deterministic functions [e.g. Bakr, 1976 and Sudicky, 1986]. On the other hand, aquifer parameters are uncertain. This uncertainty is due to the fact that parameters are measured only at some sampled locations such as selected well locations and depth intervals that are often sparse and/or the intrinsic complexity of the geological process that causes the variability. Groundwater flow and transport are therefore more realistically modelled via the stochastic approach. The word stochastic has its origin in the Greek adjective $\sigma\tau o\chi\alpha\sigma\tau\iota\kappa o\varsigma$ which means skilful at aiming or guessing [Haldorsen, Brand and MacDonald, 1987]. The classical theory of flow through porous media is derived, on deterministic bases, from the principles of continuum mechanics. The principle of mass conservation of flow through a saturated control volume of the porous medium leads to an expression for the continuity through the porous medium in terms of the specific discharge. By introducing Darcy's law, one can get the governing equation of flow through porous media in terms of pressures. This equation is in the form of an

ordinary differential equation in one-dimensional steady flow problems, and in the form of partial differential equations in two- and three-dimensional flow problems. Stochastic techniques either analytical or numerical are available to solve these differential equations with stochastic parameters. These techniques are the tools to evaluate the effects of spatial variability of the hydro-geological parameters on flow and transport characteristics in porous formations. Before discussion of some selected techniques a brief review of the concepts and properties of stochastic processes is given.

2. Definition of Stochastic Processes

A stochastic process may be defined, according to Bartlett [1960, referenced by Krumbein, 1967] " some possible actual, e.g. physical process in the real world, that has some random or stochastic element involved in its structures". If a given process operating through time or space, it is thought of as a system comprising a particular set of states. Then in a classical deterministic model the state of the system in time or space can be exactly predicted from knowledge of the functional relation specified by the governing differential equations of the system (deterministic regularity). On the other hand, in the purely stochastic model, the state of a system at any point in time or space is characterized by the underling fixed probabilities of the states in that system (statistical regularity).

A stochastic process in which the variation of a property of a physical phenomenon is represented in one co-ordinate dimension is called uni-dimensional stochastic process. The co-ordinate dimension can be time as in time series, or space as in space series. A multi-dimensional stochastic process is called a random field. A random field is defined as a mathematical way to describe spatial variations of properties of a physical phenomenon. These spatial variations can be studied by means of stochastic processes representing these variations in a continuous sense over the space considered or at discrete points in it. A stochastic process can be defined mathematically as a collection of random variables. This definition could be given in a mathematical form as the set $\{[\mathbf{x}, Z(\mathbf{x},\zeta_i)], \mathbf{x} \varepsilon R^n\}$, $i = 1,2,3...,m$ [Marsily, 1986]. $Z(\mathbf{x},\zeta)$ is stochastic process, (random function), $\mathbf{x}$ is the co-ordinates of a point in n-dimensional space, ζ is a state variable (the model parameter), $Z(\mathbf{x},\zeta_i)$ represents one single realization of the stochastic process, i= 1,2,...,m (i: number of realizations of the stochastic process Z), $Z(\mathbf{x}_0,\zeta)$ = random variable, i.e., the ensemble of the realizations of the stochastic process Z at $\mathbf{x}_0$, and $Z(\mathbf{x}_0,\zeta_i)$= single value of Z at $\mathbf{x}_0$. For simplification the variable ζ is generally omitted and the notation of this stochastic process is $Z(\mathbf{x})$.

3. Stationarity, Non-stationarity, Intrinsic Hypothesis and Ergodicity

(i) Stationarity (Statistical Homogeneity)

Stationarity is a statistical property describing the state of variability of the stochastic process. The stationarity can be tested for all moments of the stochastic process. The stochastic process is said to be second-order stationary if the mean value of a stochastic process is constant at all points in the field that means the mean does not depend on the position. This can be expressed mathematically as,

$$E\{Z(\mathbf{x})\} = \mu_Z \tag{1}$$

and if the covariance of a stochastic process depends only on the difference between the position vectors of two points $(\mathbf{x}_i-\mathbf{x}_j)= \mathbf{s}_{ij}$ the separation vector, and does not depend on the position vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ themselves. This can be expressed mathematically as,

$$Cov(Z(\mathbf{x_i}), Z(\mathbf{x_j})) = E\{[Z(\mathbf{x_i})-E\{Z(\mathbf{x_i})\}][Z(\mathbf{x_j})-E\{Z(\mathbf{x_j})\}]\} = Cov(\mathbf{s}) \tag{2}$$

this implies that the variance is independent of $\mathbf{x}$,

$$Var[Z(x)] = Cov(0) = \sigma^2_Z \tag{3}$$

In the physical sense, one thinks of a medium whose variability is the same throughout the formation of interest, so that the covariance is independent of the position but depends on the separation vector [Bakr et al., 1978].

(ii) Non-stationarity

A stochastic process is called non-stationary, if the moments of the process are variants in space, i.e., from one position to another. On other words, the moments are not only depends on the separation vector, but also, on the position of the point in space.

(iii) Intrinsic Hypothesis

This is another statistical property that is weaker than the second order stationarity. The intrinsic hypothesis assumes that even if the variance of $Z(\mathbf{x})$ is not finite, the variance of the first-order increments of $Z(\mathbf{x})$ is finite and these increments are themselves second-order stationary. This hypothesis postulates that: (1) the mean is the same everywhere in the field; and (2) for all distances, $\mathbf{s}$, the variance of the increments, $\{Z(\mathbf{x}+\mathbf{s})-Z(\mathbf{x})\}$ is a unique function of $\mathbf{s}$ so independent of $\mathbf{x}$. A stochastic process that satisfies the stationarity of order two also satisfies the intrinsic hypothesis, but the converse is not true. The above two properties can be expressed by writing,

$$Var\{Z(\mathbf{x}+\mathbf{s})-Z(\mathbf{x})\} = 2\gamma(\mathbf{s}) \tag{4}$$

where, $\gamma(\mathbf{s})$ is called the semi-variogram.
Eq.(5) may be written as,

$$\gamma(\mathbf{s}) = \frac{1}{2} E\{[Z(\mathbf{x}+\mathbf{s})-Z(\mathbf{x})]^2\} \tag{6}$$

which means, that $2\gamma(\mathbf{s})$ is the mean squared difference for two points separated by a distance $\mathbf{s}$.

From practical point of view, the intrinsic hypothesis is appealing, because it allows the determination of the statistical structure, without demanding the prior estimation of the mean. Furthermore, for a stationary random process, where both the covariance and the sime-variogram are exist, it is easy to show the relationship between them as,

$$\gamma(s) = Cov(0) - Cov(s) \tag{7}$$

(iv) Ergodicity

Ergodicity is a statistical property that implies that the statistics of a single realization in space (spatial statistics) are equivalent to the ensemble of all possible realizations (ensemble statistics). In other words, by observing the variability in space of a property from one realization in enough detail, it is possible to determine the probability distribution function of the random process for all possible realizations. This equivalence is achieved when the size of the space domain is sufficiently large or tends to infinity.

A multi-dimensional stochastic process is said to be isotropic, if the process does not have a preferred direction, i.e., the variability in the process is the same in all directions. On the other hand, the process is said to be anisotropic, if the variability changes from one direction to another. Properties of stationary stochastic processes may be represented in a lag domain either as an auto-correlation function of the lag $\mathbf{s}$, or as cross-correlation function of $\mathbf{s}$. The diagram used to display this function is called correlogram which represent the auto-correlation coefficients versus the lag $\mathbf{s}$ between the auto-correlated values of the process. The spatial auto-correlation is a measure of the spatial correlation structure of a process. It gives the degree to which a process is correlated with itself as a function of separation lag. The auto-correlation of stochastic process $Z(\mathbf{x})$ is expressed mathematically by,

$$\rho_{ZZ}(\mathbf{s}) = \frac{Cov\big(Z(\mathbf{x}+\mathbf{s}),\, Z(\mathbf{x})\big)}{\sigma^2_Z} \tag{8}$$

(i) Integral Scale

The integral scale I_z of auto-correlation function is defined by Dagan [1982] in analogy to Lumley and Panofsky [1964] as,

$$I_z = \int_0^{\infty} \rho_{ZZ}(\mathbf{s})\; d\mathbf{s} \tag{9}$$

which implies that the average distance over which the process is auto-correlated in space. For practical applications, the integration is calculated over a certain limits [0, $\mathbf{S}_o$] where, $\mathbf{S}_o$ is the smallest value of **s** at which the auto-correlation function becomes practically zero.

(ii) Correlation Scale (Range)

The correlation scale is defined as the distance over which the process is auto-correlated in space. It is calculated as the distance at which the auto-correlation function tends to zero. There are various ways [e.g. Smith, et al., 1979 a,b], some authors suggest the threshold value taken as e^{-1} to others [e.g. Gelhar, 1986]. In case of 2D isotropic exponential auto-correlation function is defined by,

$$\rho_{ZZ}(\mathrm{s}) = \mathrm{e}^{-\frac{|\mathrm{s}|}{\lambda}} \tag{10}$$

5. Scales of Natural Variability (Heterogeneity)

Heterogeneity can be found at various scales. Definitions of such scales differ from author to author. According to [Weber, 1986] these scales are described as:

(i) Microscopic Scale: this scale of variability is over the grains and pores. It is in order of mm. At this scale flow inside the pores and between the grains is governed by Navier-Stokes equations.

(ii) Macroscopic Scale: at this scale one is interested in cores and samples which contains many grains and pores. This scale is the average of all various microscopic variability. Darcy's law is emerged at this scale from the average over the enormous complexity of Navier-Stockes equations governing the flow at the pore scale. This variability is known also by Dagan [1986] as Darcy or laboratory scale where one is interested in material properties, like, porosity, permeability, and dispersivity. This scale is in order of ms that is also named by Bear [1979] the representative elementary volume (*REV*).

(iii) Megascopic Scale: this scale describes the internal architecture of reservoir units and lithologies. In this scale one is interested in the dimensions, shapes, orientations and spatial disposition of lithofacies. This kind of variability is sometimes called field or local scale that is in order of 100 m.

(iv) Gigascopic Scale: at this scale one is interested in regional geological features such as faults, fractures, aquifer size, depositional environments or regional tectonic events. This scale is also known as regional or formation scale that is in order of kms. Fig.1 shows the scales of variability according to Weber [1986].

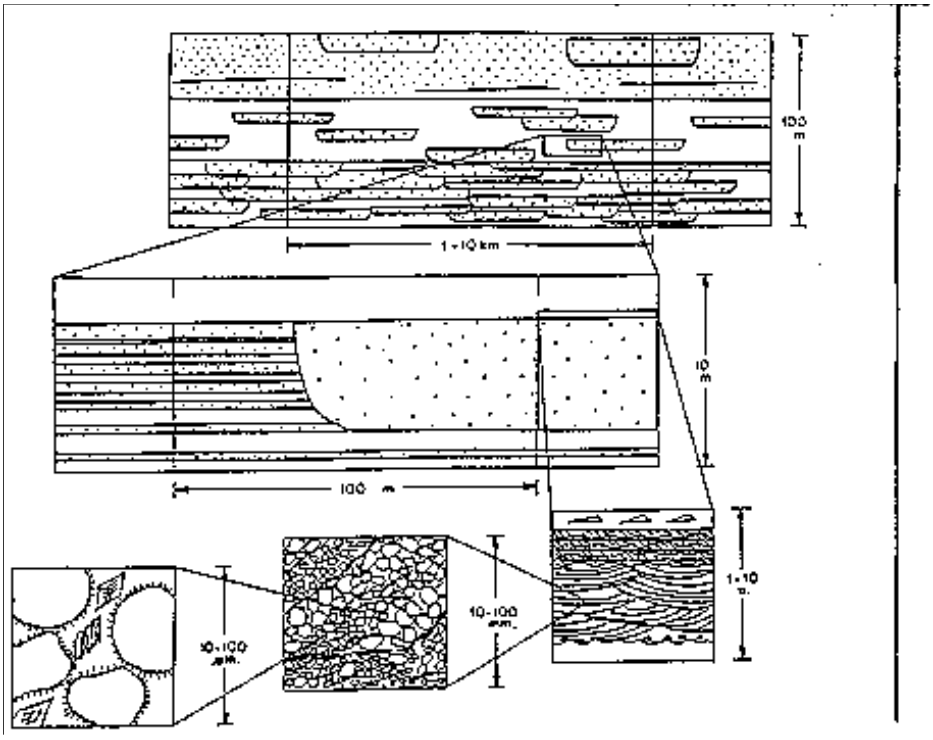

Fig.(1) Scales of Variability According to [Weber, 1986].

6. Site Characterization Using The Stochastic Approach

The complex structure of the geological formations has attracted considerable attention from geologists, geostatisticians, hydrogeologists, oil reservoir engineers, civil engineers, mining engineers, environmental engineers and others who are interested in geoscience. The literature on techniques to describe heterogeneity is abundant and rapidly increasing. The basic idea behind the stochastic approach in site characterization is that a large number of synthetic geological structures are generated based on a stochastic description of the system. In case of single stochastic simulation, the geological judgement should be involved to decide for the most probable image of the geological system based on geological experience, intuition, common sense etc. However, in case of a multi-realization approach, Monte-Carlo approach is applied to estimate the uncertainty in the output (response) variables. A comprehensive review of the stochastic nature of reservoirs can be found in the literature of oil reservoirs, such as Haldorsen, Brand and MacDonald [1987], and Fayers and Hewett [1992]. In the stochastic approach two lines of thought are found in the literature. The first is the facies or discrete models and the second is the continuous models. This kind of classification can also be viewed as description of the different scales mentioned above. For instance, the discrete model is aiming to characterize megascopic scales and the continuous models are used to characterize the macroscopic scales. Description of these models is given next.

6.1 Mosaic Facies (Discrete) Models

These models focus on the geological description of the discrete features of natural formations. In this approach one is aiming to construct formation geological units, its geometric characteristics, lithologies, units dimensions (lengths, thickness, and widths), orientations and frequency of occurrence, etc. There are a variety of techniques for simulating geological patterns. These models are often used by petroleum geologists. One may distinguish two types:

(*i*) *Object-based models, Boolean models or Binary models*:

In these models, one may consider two states or phases e.g. sand and shale formation. Sand bodies are generated in the foreground while the background is shale or vice versa. The shape of the objects can be rectangles, ellipses or any other shape. The objects are distributed randomly in space with random sizes. There are two main parameters needed for such models, the density of the random objects per unit of volume and the statistical distribution of the sizes of the objects. The orientation of the objects with respect to the horizontal can also be considered in these models. Boolean models are very simple and flexible. The limitations come from the randomness in the location of the objects in space and the ignorance of the spatial connectivity between the objects in space. More detailed information can be found in a recent work by Chessa [1995]. The limitations of the object-based models motivated the development of the following models.

(*ii*) *Sequential-based models*:

These types of models can be divided broadly into two categories: (*a*) the first category is based on the theory of two-point auto and cross-covariance description of the spatial process; (*b*) the second category is based on a local conditional probability description of the spatial process. A model based on the first category is the so-called " *Sequential Indicator Simulation*" proposed by Journel [1989]. A reproduction of all facies proportions and auto and cross covariances could be ensured by this method. From a practical point of view, the only difficulty of this method is the inference of reliable covariance functions from the available data. Models based on the second category are:

" *Classical Markov Chain Model*". In this approach the stratigraphic sequence of geological units is described in terms of transition probability matrix. A one-dimensional Markov chain has been used by Krumbein [1967]. The sequence of layers is random but conditional on the preceding layer. A generalization of this approach is possible where the conditional dependence includes layers earlier than the preceding layer. For more details about this approach see Harbaugh and Bonham-Carter [1970]. A complete description of the Markov chain theory is presented in the next chapter (section 3.3). The limitation of this method is that it only characterize the variability in sedimentological sequence in vertical direction. An extension of this method into higher dimensions is the core of the next chapter where a new methodology is proposed.

" *Poisson Random Lines Model*", its theoretical work has been provided by Switzer [1965, referenced by Lin and Harbaugh, 1984] and has been applied by Lippman [1973]. The lines are used to represent boundaries between different soils. In this approach an estimation of a parameter governing the Poisson distribution is required. The parameter is used to determine the number of random lines used in the simulation step. Attempts have been done to estimate this parameter from the transition probabilities, but the question of the most appropriate procedure for estimating this parameter has not been resolved. Some attempts have been done to use this model, but no satisfactory results are found comparable with natural geological formations.

" *Markovian Random Field Model*" used by Cross and Jain [1983] to model 2D surface texture of natural materials. This method was originally developed for the application in the field of image processing. There is a similarity between image description and a reservoir description. It seems that such an approach is well suited to the characterization of underground reservoirs. The method does not use variogram or auto-correlation to describe the relationships between neighbouring locations, but it is based on the theory of conditional probability. The idea in this method is that the probability of a given state being present at a given node on a lattice depends only on the states in the immediate neighbourhood of the simulated grid point. Realizations of Markovian fields can be generated iteratively using a simulated annealing algorithm or Metropolis algorithm. Cross and Jain [1983] applied Metropolis algorithm to simulate Markovian fields. The Metropolis algorithm works briefly as follows. The states in the systems are generated by arbitrary distribution over the lattice. Two grid points are selected at random from the lattice. Then, simple exchange of grid points states based on conditional probabilities is performed by Metropolis algorithm. After each trial step in Metropolis algorithm a new permutation of the states of the grid points is created. The procedure continues iteratively until the marginal and the transitional

probabilities of the states in the system are stabilized. The method seems promising in the field of reservoir characterization but the disadvantages of this method are its computer-intensive iterative procedure required to achieve equilibrium of the system and the difficulty to perform conditioning to honour measurements at their location.

6.2 Continuous Models

The continuous models are another way of describing heterogeneity. These models focus on rock property or parametric variability to describe the local variations of certain parameters (hydraulic conductivity, porosity, dispersivity... etc.). These types of models are frequently used in the field of stochastic subsurface hydrology. There are several methods for the generation of stochastic fields. Some of these methods are: " *Multi-variate Method*" developed by Iman [1980, referenced by Peck, et al., 1988], " *Nearest Neighbour Method*" developed by Whittle [1954], " *Turning Bands Method*" developed by Matheron [1973], " *Spectral Methods*" by Shinozuka and Jan [1972] and Mejia and Rodriguez-Iturbe [1974], " *Fast Fourier Transform Method*" by Borgman, [1984]; and Gutjahr [1989] and " *Source Point Method*" by Ghori, Heller and Singh [1992]. The first three methods are discussed in detail in this paper. The other methods are based more or less on the same background. Descriptions of these methods, their advantages, disadvantages limitations and algorithms for implementation are presented in the following sections. All these methods are based on the theory of regionalized variables developed by Matheron [1971] and they are used to generate realizations of stationary random fields of the model parameters.

6.3 Hybrid Models

The hybrid stochastic model is a combination of the two models described above. The essence of the hybrid model is that geological knowledge is used to describe mega-scopic variability in the form of the architecture of the formation, reservoir discontinuities and zonal boundaries with average hydraulic properties, while parametric data are used to model macro-scopic heterogeneity in a form of continuous spatial variability within each lithology. Each of these units would be treated as a separate, statistically homogeneous unit. A similar assumption was used by Brannan, et al. [1993]. In their model they assume that the geological formation is given and they only generate conductivity fields inside each unit. They also introduce lenses with different shapes and orientations in the generated fields. Elfeki, [1997] has developed a hybrid model combining soft information in Markov chain framework and the hard data in a classical Gaussian correlated random field. The model is promising to model two-scale of variability for flow and transport studies [see e.g. Elfeki et al. 1996 and 1998].

7. Quantifying Uncertainty by The Monte-Carlo Approach and Generation of Random Fields:

Monte-Carlo approach is based mainly on generating random fields of the hydrogeological parameters to represent the heterogeneity of the formation. Then, the usual groundwater flow and/or transport equations can be solved numerically in this geometrical structure. One can assume the probability density function of the model parameters or joint probability density function for a number of parameters in the model. The assumptions of these density functions are based on some field tests and/or laboratory tests. By using a random number generator, one generates a realization for each one of these parameters. The parameter generation can be correlated or uncorrelated depending on the type of the problem. With this parameter realization a classical numerical flow or/and transport model is run and a set of results is obtained. Another random selection of the parameters is made and the model is run again, and so on. It's necessary to have a very large number of runs, and the output model results corresponding to each input is obtained which can be represented mathematically by the stochastic process $\Phi(\mathbf{x}, \zeta_i)$. Statistical analysis of the ensemble of the output (i.e. $\Phi(\mathbf{x}, \zeta_i)$ for i = 1,2,..., m) can be done to get the mean, the variance, the covariance or the probability density function for each node with a location $\mathbf{x}$ in the grid. There are several methods in the literature used for generating random sequences with some auto-correlation structure. The most famous methods will be described in this section. Attention will be given to the Gaussian random fields, where most of the geostatistical parameters exhibit variability that follows a normal or log-normal probability

density function. These methods can be classified as: (1) Direct methods or matrix methods, such as, multi-variate normal distribution and nearest neighbour method. (2) Indirect or transformation methods, such as, turning bands algorithm.

(i) Multi-Variate Method

In the uni-variate case as discussed in the preceding sections, only one random variable was considered. The multi-variate case given in this section is a natural extension of the uni-variate case, i.e., more than one random variable or what is called random vector will be discussed. A *p*-dimensional random vector $\mathbf{Z}=\{Z_1,Z_2,Z_3,...,Z_p\}^T$ is defined to have a multi-variate normal distribution if and only if the *p*-components of **Z** have a uni-variate normal distribution. The distribution of **Z** is denoted $N_p(\boldsymbol{\mu},\mathbf{C})$. The multi-variate normal density function $N_p(\boldsymbol{\mu},\mathbf{C})$, *p*-dimensional normal random variate, is given by Mood and Graybill [1963],

$$f(\mathbf{Z}) = \frac{1}{|\mathbf{C}|^{1/2}(2\pi)^{p/2}} \exp\left[-\frac{1}{2}(\mathbf{Z}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{Z}-\boldsymbol{\mu})\right] \tag{11}$$

where, $N_p(\boldsymbol{\mu},\mathbf{C})$ is a multi-variate normal distribution with mean $\boldsymbol{\mu}$, and covariance matrix **C**, *p* is the number of parameters (nodes of the model), $\mathbf{Z}=\{Z_1, Z_2,..., Z_p\}^T$, *p*-dimensional random vector, ($p\times1$), $\boldsymbol{\mu} = \{\mu_1,\mu_2,\ldots,\mu_p\}^T$, *p*-dimensional mean values vector, ($p\times1$), *T* is superscript transpose operation of a matrix, -1 superscript is inverse operation of a matrix, **C** is a $p\times p$ covariance matrix, $|\mathbf{C}|$ is the determinant of the covariance matrix **C**, and σ^2_{Zi} is the variance of the random parameter Z_i.

All the diagonal elements of the covariance matrix are the variances of the individual random variables and the off-diagonal components are covariances of two random variables Z_i, Z_j where, i=1,...p, and j=1,...p. The covariance matrix can be defined as the matrix which lists the correlations between the parameters. It should be noted that all the covariance matrices are symmetric. This property allows us to apply the theory of symmetric matrices. Sometimes it is convenient to normalize covariance matrices by converting the individual covariance in terms of the correlation coefficient ρ_{ij} as,

$$\rho_{ij} = \frac{Cov(Z_i, Z_j)}{\sigma_{Z_i}\sigma_{Z_j}} \tag{12}$$

where, $Cov(Z_i,Z_j)$ is the covariance between Z_i and Z_j.

The generation of varieties with multi-variate distributions is more than with uni-variate distributions. The obvious difference is that in the generation of multi-variate distributions the dependencies (correlations) among the all components of the random vector must be handled. The general technique for the generation of multi-variate distributions is called "*Conditional Distribution Approach*". A brief description of this approach is given here. For more details about this method reference is made to Johnson [1987]. Let $\mathbf{Z} = \{Z_1,Z_2,Z_3,...Z_p\}^T$ be the *p*-dimensional random vector of interest. The conditional distribution approach involves the following steps: (1) generate $Z_1 = z_1$ from the marginal distribution of Z_1 (uni-variate distribution of Z_1); (2) generate $Z_2 = z_2$ from the conditional distribution of Z_2 given $Z_1 = z_1$; (3) generate $Z_3 = z_3$ from the conditional distribution of Z_3 given $Z_1 = z_1$ and $Z_2 = z_2$... and so forth through the *p* steps.

In certain statistical applications the covariance matrix gives the dependencies between the components of the random vectors. In case of random fields the same philosophy is applied. Each node in the field is considered as a component of a random vector that contains all the nodes in the field. The dependencies between the components are described by the auto-covariance function of the field. The method of multi-variate random number generation with a given covariance matrix of the system has been

developed by [Iman, 1980, referenced by Peck, A. et al., editors, 1988]. It has been applied to groundwater flow by Townley [1984, 1985,1988]. The algorithm for generating random fields with a given covariance structure based on the covariance matrix of the system is as follows [Neuman, 1984]:
1) Build the covariance matrix **C** of the system. The elements of **C** are denoted by,

$$c_{ij} = Cov(Z_i, Z_j) \tag{13}$$

and if $i=j$ the covariances becomes the variances.

If the random field is assumed to be statistically homogeneous (stationary), i.e., the mean is constant, the standard deviation is constant, and the covariance depends only on the distance vector between the points in the domain, then one can write the elements of the covariance matrix as,

$$c_{ij} = \sigma^2_Z \, \rho(s_{ij}) \tag{14}$$

where, σ^2_Z is the variance of the process Z, $\rho(s_{ij})$ is the auto-correlation function, and s_{ij} is the distance vector between point i and point j.
2) One has to decompose the covariance matrix by the Cholesky factorization method Westlake [1968] into

$$\mathbf{C} = \mathbf{L}\,\mathbf{U} \tag{15}$$

where, **L** is a unique lower triangular matrix, **U** is a unique upper triangular matrix, and **U** is $\mathbf{L}^T$ which means that **U** is the transpose of **L**.
Cholesky method is sometimes called " *Square-Root*" method which is used only for the decomposition of symmetric matrices. For more details about Cholesky method see Westlake [1968].
3) Generation of normally distributed p-dimensional sequence of independent random numbers with zero mean and unit standard deviation $N(0,1)$ which can be expressed as, $\boldsymbol{\varepsilon} = \{\varepsilon_1, \varepsilon_2, .., \varepsilon_p\}^T$ where, $\boldsymbol{\varepsilon}$ is vector of normally distributed random numbers, and ε_i is the i-th random number drawn from $N(0,1)$.
4) Multiplication of the independent random vector $\boldsymbol{\varepsilon}$ by the triangular matrix $\boldsymbol{U}$ to get a vector of correlated random numbers. This vector can be expressed by matrix multiplication convention as,

$$\mathbf{X} = \mathbf{U}\,\boldsymbol{\varepsilon} \tag{16}$$

where, **X** is a vector of multi-variate normal random $N_p(\mathbf{0},\mathbf{I})$, **0** is zero mean vector ($p\times1$), and **I** is the identity matrix ($p\times p$). The desired random field is then obtained by

$$\mathbf{Z} = \boldsymbol{\mu} + \mathbf{X} \tag{17}$$

The random process **Z** is the one for which the samples $Z_1, Z_2, ..., Z_p$ are jointly distributed according to a multi-variate probability density function that is given by Eq.(13). In the Monte-Carlo approach, the matrix decomposition by the Cholesky method is performed only once, then a large number of correlated fields can be generated at very low computer cost because this only involves the product of a vector of random numbers by the decomposed matrix.

(ii) Nearest Neighbour Method

Another method that is used for generating a correlated random field is called " *Nearest Neighbour Method*". The general framework required for this method was presented by Whittle [1954]. This method is sometimes called Whittle Model. The idea of this method is based on generating an independent random field with one of the methods described in the previous sections. Then, the value of a parameter at a given node is replaced by taking weighted average of the previous random values at the given node and

a few surrounding ones. This model was first applied to one-dimensional steady state groundwater flow in a bounded domain by Smith and Freeze [1979a] and then they also extended the same approach to the two-dimensional steady state groundwater flow in [1979b]. This method can be expanded to three-dimensions. A general description of this spatial model used by Smith and Freeze can be written in a generalized form proposed by Whittle [1954],

$$Z_i = \sum_{j \neq i} W_{ij} \; Z_j + \varepsilon_i \tag{18}$$

where, Z_i is random variable satisfying the nearest neighbour relation, ε_i is uncorrelated normal random number with $E(\varepsilon_i) = 0$, and $Var(\varepsilon_i) = \sigma_i^2$, i=1,2,.....p, and W_{ij} are the weighting coefficients.

In case of the anisotropic first-order auto-regressive scheme, Eq.(18) can be rewritten according to Smith and Freeze [1979b] as,

$$Z_{ij} = \alpha_x (Z_{i-1j} + Z_{i+1j}) + \alpha_y (Z_{ij-1} + Z_{ij+1}) + \varepsilon_{ij} \tag{19}$$

where, α_x is an auto-regressive parameter expressing the degree of dependence of Z_{ij} on its two neighbouring values in the horizontal direction, Z_{i-1j} and Z_{i+1j}, (α_x<1), and α_y is an auto-regressive parameter expressing the degree of dependence of Z_{ij} on its two neighbouring values in the vertical direction, Z_{ij-1} and Z_{ij+1}, (α_y<1).

Eq.(19) should be applied with modifications at the boundaries of the domain of interest where necessary, because the blocks do not extend across the boundaries i.e. the equation will be truncated near the boundary. Eq.(19) can be recanted in matrix notation as,

$$\mathbf{Z} = \mathbf{W}\,\mathbf{Z} + \boldsymbol{\varepsilon} \tag{20}$$

where, matrix **W** is called the $p \times p$ connectivity matrix, or the $p \times p$ spatial lag operator of scaled weights, w_{kl}. The elements of the connectivity matrix w_{kl} are defined as,

$$w_{kl} = \frac{w^*_{kl}}{N} \tag{21}$$

where, k = 1,..., p, l = 1,..., p, and $k \neq l$, $w^*_{kl} = \alpha_x$ if the blocks k and l are contiguous in the x-direction, $w^*_{kl} = \alpha_y$ if the blocks k and l are contiguous in the y-direction, and $w^*_{kl} = 0$ otherwise, i.e., if $k = l$, or if blocks k and l are not contiguous, and N is the total number of blocks surrounding block k, i.e., N=4 if block k is located inside the domain that is being modeled, N=3 if block k is located on the boundary of the domain, and N=2 if block k is located at a corner of the domain.

The scaling of N is required to ensure stationarity in the generated sequence of Z_{ij} values near the domain boundaries because Eq.(20) is truncated near the boundaries.

The **Z** sequence has a mean μ_Z and standard deviation σ_Z. Similarly, $\boldsymbol{\varepsilon}$ has mean μ_ε and standard deviation σ_ε. In order to simulate the predetermined standard deviation σ_Z, one should start from a random sequence $\boldsymbol{\varepsilon}$ with a standard deviation of one. This vector can be pre-multiplied by an appropriate factor to yield the desired value of σ_Z. This factor is denoted by η. Eq.(20) becomes,

$$\mathbf{Z} = \mathbf{W}\,\mathbf{Z} + \eta\,\boldsymbol{\varepsilon} \tag{22}$$

Eq.(22) can be solved for **Z** sequence as follows

$$\mathbf{Z} = (\mathbf{I} - \mathbf{W})^{-1}\,\eta\,\boldsymbol{\varepsilon} \tag{23}$$

To determine the simulation parameter η, one can follow the procedure of Smith and Freeze [1979] or the procedure of [Baker, 1984] as given in brief here. The auto-correlation matrix of the process **Z** can be written as

$$\mathbf{R} = \frac{E(\mathbf{Z}\,\mathbf{Z}^{\mathrm{T}})}{\sigma_Z^{\,2}} \tag{24}$$

Substitution of Eq.(23) into Eq.(24) and with some algebraic manipulations, one gets

$$\mathbf{R} = \frac{1}{\sigma_Z^2}\,\mathbf{V}\,\eta^2.\sigma_\varepsilon^2 \tag{25}$$

where, $\mathbf{V} = (\mathbf{I} - \mathbf{W})^{-1}.((\mathbf{I} - \mathbf{W})^{-1})^{\mathrm{T}} = ((\mathbf{I} - \mathbf{W}).(\mathbf{I} - \mathbf{W})^{\mathrm{T}})^{-1}$
Let σ_ε be equal to one in Eq.(25), then

$$\mathbf{R} = \frac{1}{\sigma_Z^2}\,\mathbf{V}\,\eta^2 \tag{26}$$

All the diagonal elements of **R** are equal to unity by definition, hence, by taking the trace (the sum of the elements on the main diagonal of the square matrix) and solving the resultant scalar equation for η gives,

$$\eta = \frac{\sigma_Z}{\sqrt{V_m}} \tag{27}$$

where, $V_m = tr\,\mathbf{V}/p$, and the symbol "*tr*" is the trace of the matrix, which is given by, $tr\,\mathbf{V} = \Sigma\, v_{ii}$, $i = 1,...,p$
Substituting Eq.(27) into Eq.(23), the final nearest neighbour generator is obtained,

$$\mathbf{Z} = (\,\mathbf{I} - \mathbf{W}\,)^{-1}\ \frac{\sigma_Z}{\sqrt{V_m}}\ \boldsymbol{\varepsilon} \tag{28}$$

where, the **Z** sequences have a mean of zero. By adding the constant $\boldsymbol{\mu}_{\mathbf{Z}}$ to each element of **Z** the system of equations for the nearest neighbour model can be written as

$$\mathbf{Z} = \boldsymbol{\mu}_{\mathbf{Z}} + (\,\mathbf{I} - \mathbf{W}\,)^{-1}.\frac{\sigma_Z}{\sqrt{V_m}}\,\boldsymbol{\varepsilon} \tag{29}$$

The analysis of the covariance function describing the generated random field with first-order dependence is approximately an exponential decay function Smith and Freeze [1979]. By adjusting the number of neighbours (higher-order nearest neighbour models), the weighting coefficients and the variance of the initial random independent parameters, it is possible to approximately fit any given real covariance function as observed on the data. The advantage of this technique can be seen in Eq.(29). At the beginning of any simulation the matrix (**I - W**) must be inverted only once. For each realization of the process **Z**, the inverted matrix $(\mathbf{I} - \mathbf{W})^{-1}$ is simply multiplied by the generated random vector $\eta\boldsymbol{\varepsilon}$. The drawback of this method is computing the inverse matrix.

(iii) Turning Bands Method (TBM)

The turning bands method is one of the techniques which is designed to generate a realization of stationary, correlated, and multidimensional Gaussian random field from a normal distribution with zero

mean and a specified covariance structure. The *TBM* was first proposed by Matheron [1973] and applied by the Ecole des Mines de Paris [e.g. Journel, 1974; and Delhomme, 1979]. The *TBM* is based on the theory of random fields (multidimensional stochastic process). Its basic concept is to transform a multidimensional simulation into the sum of a series of equivalent uni-dimensional simulations [Mantoglou and Wilson, 1982]. The basic idea of the algorithm in brief is, generating two- and three-dimensional fields by subsequent projection and combining values found from a series of one-dimensional simulations along several lines radiating outward from an arbitrary origin in space. This procedure yields discrete values or realizations of the random field. This method has been widely used in porous flow and transport studies. *TBM* is a repetition of a two step procedure. Firstly, a realization of a random process with a prescribed auto-covariance function and zero mean is generated on one line. The Cholesky decomposition method can be used (but with much smaller correlation matrix dimensions) or by auto-regression methods, like nearest neighbour. Secondly, orthogonal projection of the generated line process to each point in the simulated two- or three-dimensional random field. The two steps are repeated for a given number of lines and then a final value is assigned to each grid point in the field by taking a weighted average over the total number of lines.

There are two main approaches for generating the one-dimensional line process in *TBM*. The first one is space domain approach that was first proposed by Matheron and applied by Journal and Huijbregts [1978] and Delhomme [1979]. This approach can handle only particular forms of auto-correlation functions. The second one is the spectral (frequency) domain approach *STBM*. This approach has been implemented by Mantoglou and Wilson [1982]. It is a more general approach that can handle a wide variety of two-dimensional processes, multi-variate (cross-correlated) processes, as well as spatial averaged (integrated) processes.

(a) Theoretical Background of TBM:

Let $Z_i(u)$, i = 1,..., L a set of N independent realizations of a one-dimensional, second order stationarity stochastic process on a line u with an auto-correlation function $\rho_1(u_o)$, where u_o is the spatial lag on the line. Then the values given by the relation,

$$Z_s(x,y,z) = \frac{1}{\sqrt{L}} \sum_{i=1}^{L} Z_i(u) \tag{30}$$

is a realization of a two- or three-dimensional process. The subscript *s* represents the term " *simulated*" or " *synthetic*". The field generated by Eq.(30) has zero mean as well. The relation between the auto-correlation function on the line process $\rho_1(u_o)$ and the auto-correlation function in the three-dimensional random field $\rho(u_o)$ is given by [for the derivation, refer to Mantoglou and Wilson, 1982; and Mantoglou, 1987],

$$\rho_1(u_o) = \frac{d}{du_o}\left[u_o\, \rho(u_o)\right] \tag{31}$$

and for two-dimensional field the relationship becomes,

$$\int \frac{\rho_1(u_o)\, du_o}{\sqrt{(\mathbf{s}^2 - u_o^2)}} = \frac{\pi}{2}\, \rho(\mathbf{s}) \tag{32}$$

where, **s** is the spatial lag in two-dimensional field.

It can be observed from Eq.(32) that, it is not easy to obtain $\rho_1(u_o)$ directly as a function of $\rho(\mathbf{s})$. Sironvalle [1980] wrote [referenced by Baker, 1984] " This integral equation is too difficult to solve, so we discard this (the *TBM*) method." Therefore, Mantoglou and Wilson [1982] used a spectral method to generate uni-dimensional process along the lines for different types of auto-correlation functions. In the following section this method will be given more attention.

(b) Spectral Turning Bands Method (STBM):

Generation of 2-D random fields by *TBM* needs the solution of the integral equation given by Eq.(32). This integral equation cannot be directly expressed as a function of $\rho(\mathbf{s})$. Particular solutions can be found for certain two-dimensional auto-correlation functions [Mantoglou and Wilson, 1982]. To circumvent this difficulty, an expression for the spectral density function of the one-dimensional processes as a function of the radial spectral density function of the two-dimensional process is used. This expression is given in Fourier space by,

$$S_1(\omega) = \frac{\sigma_Z^2}{2} S(\boldsymbol{\omega}) \tag{33}$$

This means, that the spectral density function of the uni-dimensional process $S_1(\omega)$ along the turning bands lines is given by one half of the radial spectral density function $S(\boldsymbol{\omega})$ of the two-dimensional process multiplied by the variance of the two-dimensional process. Steps used in implementing the *STBM* generator in simulating a two-dimensional random field are given below:

(1) Generation of One-dimensional Uni-variate Process on The Turning Bands Line:

In the literature, there are two main techniques used for generation of the line process. The first, is called the "*Fast Fourier Transform*" (*FFT*) algorithm, which can be used to construct a complex process $X(u) = Z(u) + i\, Y(u)$ given by Tompson, et al. [1989],

$$X(u) = \int_{all\omega} e^{i\omega u}\, dW(\omega) \approx \sum_{all\omega} e^{i\omega_j u}\, dW(\omega_j) \tag{34}$$

where, in the approximate form, X is the sum of a complex series of sinusoidal functions of varying wavelength, each magnified by complex random amplitude with zero mean $dW(\omega_j)$, $\omega_j = j.\Delta\omega$.

The second is the "*Standard Fourier Integration*" method. The real part of the complex process $X(u)$ is given by

$$Re\ X(u) = Z(u) = \int_{all\omega} |\, dW(\omega)\,| \cos(\omega u + \phi_\omega) \tag{35}$$

can be used to develop a straightforward discrete approximation using positive frequencies,

$$Z_i(u) = \sum_{j=1}^{M} |\, dW(\omega_j)\,| \cos(\omega_j u + \phi_j) \tag{36}$$

where, φ_j represents independent random angles which is uniformly distributed between 0 and 2π, M is the number of harmonics used in the calculations, $\omega_j = (j - .5)\ \Delta\omega$, $j = 1,2,...M$, and $\Delta\omega$ is the discretized frequency which is given by ω_{max}/M, and ω_{max} is the maximum frequency used in the calculations.

The magnitude $|\, dW(\omega_j)\,|$ is taken to be deterministic from the spectrum as,

$$|\, dW(\omega_j)\,| = \left[4\, S_1(\omega_j).\Delta\omega\right]^{1/2} \tag{37}$$

where, $S_1(\omega_j)$ is the spectral density function of the real process $Z(u)$ on the line.

$S_1(\omega)$ is assumed to be insignificant outside the region $[-\omega_{max},+\omega_{max}]$. Substitution of Eq.(37) into Eq.(36) gives the generator of the uni-dimensional process on line *i* as,

$$Z_i(u) = 2\sum_{j=1}^{M}\left[S_1(\omega_j).\Delta\omega\right]^{1/2}\cos(\omega_j u + \phi_j) \tag{38}$$

This is the classical method proposed by Rice [1954]. The form of Eq.(38) is slightly modified by Shinozuka and Jan [1972] to give,

$$Z_i(u) = 2\sum_{j=1}^{M}\left[S_1(\omega_j).\Delta\omega\right]^{1/2}\cos(\omega'_j u + \phi_j) \tag{39}$$

where, $\acute{\omega}_j = \omega_j + \delta\omega$

The frequency $\delta\omega$ is a small random frequency added here in order to avoid periodicities. $\delta\omega$ is uniformly distributed between $-\Delta\acute{\omega}/2$ and $\Delta\acute{\omega}/2$, where, $\Delta\acute{\omega}$ is a small frequency, $\Delta\acute{\omega}<<\Delta\omega$. $\Delta\acute{\omega}$ is taken equal to $\Delta\omega/20$ according to Shinozuka and Jan [1972]. This approach involves a discrete summation over frequency increments $\Delta\omega$ up to a maximum cutoff frequency of $\omega_{max} = M\ \Delta\omega$. The method involves a larger computational effort than *FFT*, but it is more flexible in the choice of the parameters, M, $\Delta\omega$, Δu, ω_{max} and u_{max} [Tompson, et al., 1989].

(2) Distribution of The Turning Bands Lines and The Number of Lines:

Theory of TBM is based on an infinite number of lines. The lines are assumed to be randomly oriented, as taken from a uniform distribution on a unit circle in 2-D space or sphere in 3-D space. It has been shown by Mantoglou and Wilson [1982] that by spacing the lines evenly on the unit circle or sphere, with prescribed directions, the simulated correlation function converges much faster to the theoretical function. Mantoglou and Wilson show that a number of 8 -16 lines are generally a satisfactory choice in isotropic auto-correlation function. In anisotropic case, however, a larger number of lines might be required.

(3) Spectral Discretization:

A realization of the line process $Z_i(u_n)$ at point *n* on line *i* is constructed from a discrete integration of a series of random components over all the frequency domain. The frequency discretization $\Delta\omega$ must be kept small enough to ensure a sufficient degree of accuracy, while the number of harmonics M must be kept large enough to account for the contributions of the spectral tail at $\omega_{max} = M\ \Delta\omega$. Mantoglou and Wilson [1982] considered the case where M is varying between 50 and 100 and ω_{max} is 40 times the correlation length of isotropic auto-correlation function in both cases. They found that while the accuracy for $M = 50$ is poor at large distances, the accuracy shown for $M = 100$ improves rapidly.

(4) Physical Discretization:

The physical increment Δu used along the line should be chosen less than the domain discretization Δx, Δy as a pragmatic rule of thumb in order to avoid some numerical problems [Mantoglou and Wilson, 1982].

(5) Length of The Turning Bands Lines:

The minimum length of the line is determined by the orientation of the line and the size of the simulation domain.

(6) Generation of The Simulated Field:

The construction of a simulated field $Z_s(x,y,z)$ will include the selection of a finite number of lines L and their orientation, the discrete simulation of one-dimensional process $Z_i(u)$ on each line, the subsequent projection of these values onto all simulation points x in the domain, and the division of each sum at each point by $L^{1/2}$ to yield $Z_s(x,y,z)$.

8. Conditional Simulations of Random Fields upon Measurements of Field Parameters

All the stochastic methods presented thus far are used for unconditional simulation of stationary random fields. These methods reproduce the first and second moments or the *PDF* of the simulated field. From a practical point of view, it is desirable that the random fields not only reproduce the spatial structure of the field but also honour the measured data and their locations. This requires an implementation of some kind of conditioning, so that the generated realizations are constrained to the available field measurements. One of the advantages of conditioning is to reduce uncertainty in the simulated fields. The uncertainty will be reduced significantly (perhaps to zero in the absence of measurement errors) at the sampled locations and it will be reduced in the vicinity of the sampled locations. The conditional simulation can be achieved by spatial estimation approaches such as Kriging. Kriging is known as the best linear, unbiased estimator (*BLUE*). Therefore, it has advantages over the conventional spatial interpolator techniques (e.g. least square method, weighted residuals method, etc. [see for more details Christakos, G., 1992]) in providing not only the estimates of point values, but also the variance of the corresponding errors of estimation (uncertainty associated with the estimate). One drawback of the Kriging method is that the estimated field is smoother than the real data. Therefore using conditional simulation with Kriging reproduces closely the true variability of the field and honours the measurements. The methods of conditioning can be broadly divided into:

(*i*)" *Direct Approaches or Matrical Methods*": these approaches draw the realizations directly from the sub-ensemble of conditional realizations. The famous algorithms are Gaussian conditional simulations via the *LU* triangular decomposition of the autocovariance. This method has been applied to subsurface hydrology by Neuman [1984]. The limitation of this method is that it is restricted in the size of the grid it can handle. Another method is called sequential simulation algorithm suggested by Journal and Huijbreghts [1981]. This method works by calculating the conditional distribution of grid point values given the data values at the sampled locations and by assuming a multivariate normal distribution of the given data.

(*ii*)" *Indirect Approaches*": these approaches are considered as two steps. Firstly an unconditional realization is generated. Secondly, a modification is used to honour the data at their locations. This method works simply by adding a correlated error obtained from the simulation to a kriged map from the sampled data. In brief, the procedure is the following:

(1) A kriged map is generated from the field data with the sampled locations that will be smoother than reality. (2) An unconditional simulated field is generated by *TBM* from the data that reproduces the spatial structure of the underlying random function. (3) Allocation of the unconditional values (pseudo measurements) at the sites of measurements is done on the simulated map in step 2. (4) Another kriged map is generated from the pseudo measurements. (5) A pseudo error is calculated by subtracting the kriged map in step 4 from the unconditional simulation in step 2. (6) The conditional simulation map is generated by adding the pseudo error in step 5 to the kriged map in step 1. So,

$$Z_{cs} = Z_{kd} + (Z_{us} - Z_{kus}) \tag{40}$$

where, Z_{cs} is the required conditional simulation, Z_{kd} is the kriged map from the real data, Z_{us} is the unconditional simulation, Z_{kus} is the kriged map with the pseudo measurements.

Conditional simulation has been implemented analytically in flow models by [Dagan, 1982a] under the assumption of multi-Gaussian distribution where conditional probabilities can be expressed analytically or by Kriging equations.

9. Evaluation of Some Techniques

In this section, a comparison of the techniques has been worked out and a discussion of the advantages and disadvantages of these methods are tabulated in Table (1). Table (1) displays a comparison between the different techniques applied for generation of random fields that are used input for the numerical simulation approach (Monte-Carlo). Fig.(2) shows three single realizations of log-hydraulic conductivity generated in a domain of dimensions 15 ×15 m by Nearest Neighbour Generator, NNG, Multi-Variate Generator, MVG, and Turning Bands Generator, TBG, worked by Elfeki [1996]. The domain is

discretized by 1 m in both horizontal and vertical directions. The parameters used for the simulations are $\langle K \rangle$=1 m/day and σ_K=2 m/day and the corresponding logarithmic transform are $\langle Y \rangle$=-0.8 and σ_Y=1.3. The first order auto-regressive parameters of the NNG are chosen to be α_x=.98 and α_y =.5 which generate statistically anisotropic field. Average values of the correlation lengths are calculated from the generated fields over 100 realizations resulting in λ_x=1.2 m and λ_y=0.73 m which show a relatively isotropic field although the ratio of the auto-regressive parameters is about 0.5. These values are used to generate random fields by the other methods (MVG, TBG) assuming anisotropic exponential auto-correlations. From Fig.(2), the three realizations more or less identical. The auto-correlation function is calculated over ensemble average of 100 realizations and plotted in Fig.(3) with the theoretical auto-correlation for comparison which seems reasonably good. Also the calculated PDF of a single realization is plotted with the given PDF in the same figure which is fairly good as well. The CPU time of 100 realization with NNG or MVG on PC with 486 processor was few seconds but for the TBG was about 30 minutes.

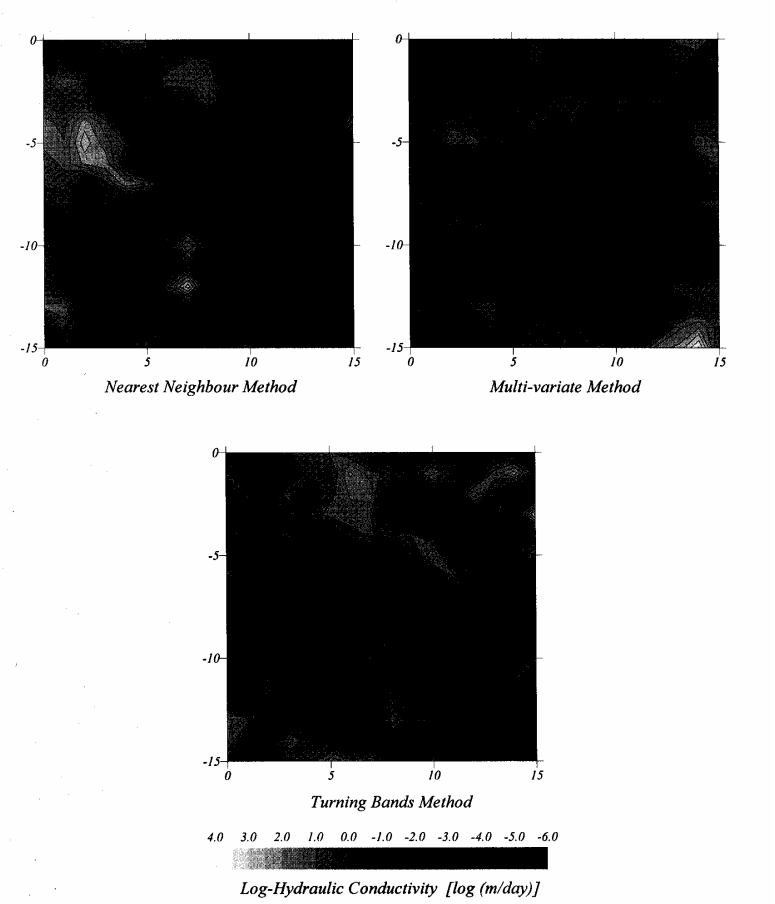

Fig.(2) Three Single Realizations Generated by Different Methods.

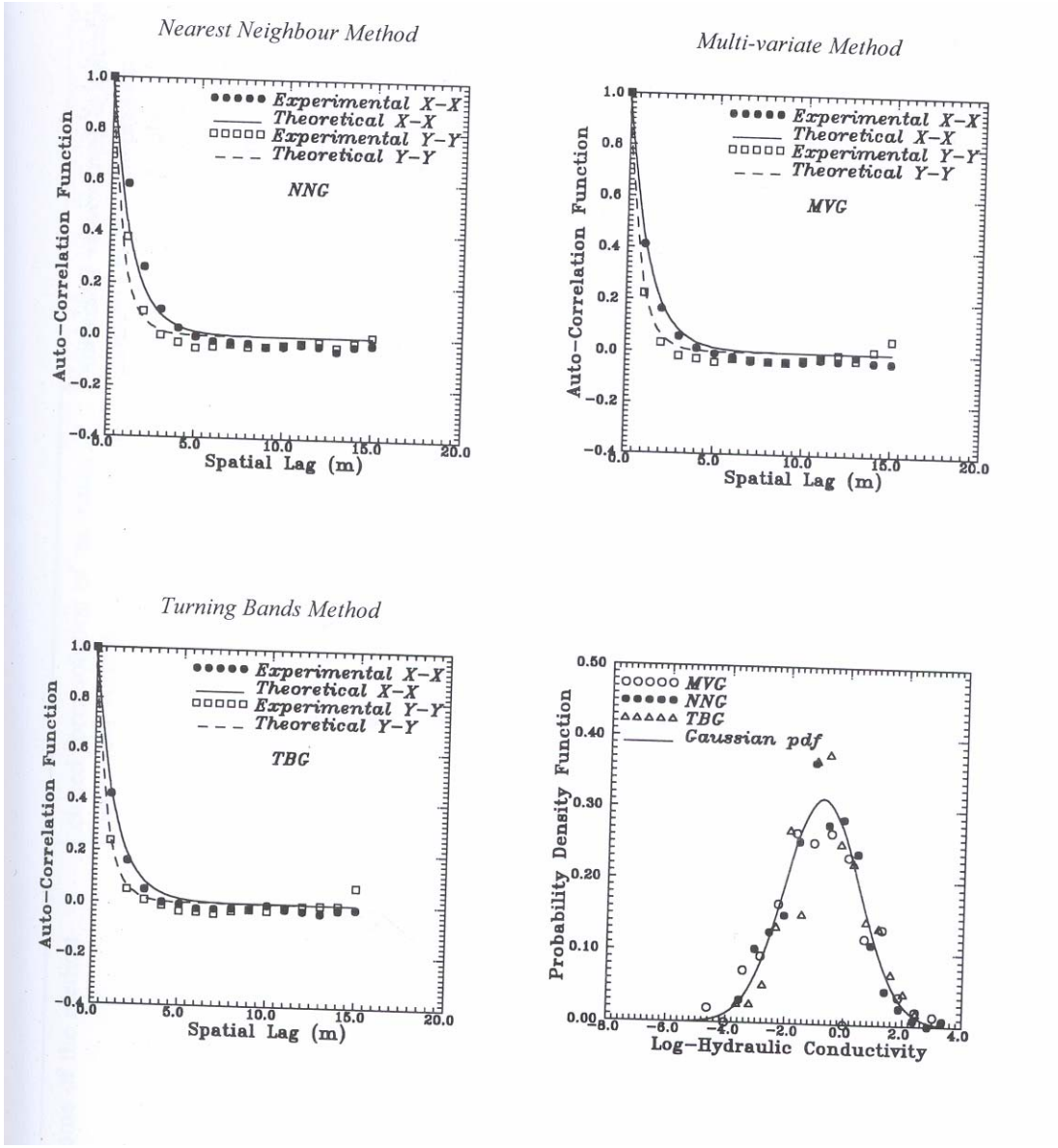

Fig.(3) Auto-correlation and PDF of the Generated Fields in Fig.(2).

Table(1) Comparison between Different Methods for The Generation of Stationary Random Fields.

Method of Simulation / Item of Comparison	Direct Methods (Matrix Methods)		Indirect Methods
	Multi-Variate (MVG)	Nearest Neighbour (NNG)	Turning Bands (TBG)
(1) field data needed for simulation	PDF of the hydrogeological parameters, autocorrelation function, and correlation lengths	PDF of hydrogeological parameters, autoregressive parameters, order of the nearest neighbour model to specify a certain auto-correlation function	PDF of hydrogeological parameters, autocorrelation function, and correlation lengths
(2) type of probability density functions (PDF) that can be handled by the method	any probability density function	any probability density function	normal probability density functions (Gaussian fields)
(3) handling auto-correlation structure of any kind	it is straightforward by filling the covariance matrix of the system using the desired auto-correlation function.	a trial and error procedure by adjusting the auto-regressive parameters and the order of the nearest neighbour till it fits the desired auto-correlation; not easy.	it needs a solution of an integral equation that relates the auto-correlation in the field with the auto-correlation on the line process which in some cases is too difficult to solve. In some cases it may be resorted to spectral TBM to tackle this problem.
(4) stationarity of the simulated field	it can handle stationary and non-stationary fields.	it can handle stationary and non-stationary fields, but the method does not guarantee stationarity even if the field is stationary	the field should be stationary
(5) statistical anisotropy	it handles anisotropy in a simple way by introducing anisotropic auto-correlation functions	it handles a limited range of anisotropy in a simple way by introducing different values of auto-regressive parameters in different directions, but it is difficult to handle highly anisotropic fields. The	it handles anisotropy using transformation method, i.e., an isotropy problem is first transformed into isotropic problem using linear transformation. Second, an isotropic

		method does not guarantee isotropy or anisotropy of the simulated field.	field in the transformed space is generated. Third, transforming back; it needs additional computational efforts and this sort of transform works only for ellipsoidal types of auto-correlation (exponential model).
(6) matrix operations used	L-U decomposition of the covariance matrix by Cholesky method once, and multiplication of the decomposed matrix by random vector for the generation of each realization.	inversion of a banded matrix (the identity matrix minus spatial lag operator matrix) once, and multiplication by a random vector for each realization.	the line process can be generated by matrix methods, but with a much smaller matrix dimensions or by spectral methods, and generation of a line process several times and its projection onto the problem domain.
(7) storage needed	order of ($\sim N^2$), where, N is the number of simulated points.	order of ($\sim N^2$).	order of ($\sim N$).
(8) computer time and requirements for generating one realization	the major part of the computer time is spent on generating the covariance matrix and its decomposition, once the decomposition is available the generation of one realization is very fast, just by multiplication of the decomposed matrix by uncorrelated random vector.	the major part of the computer time is spent on generating the matrices and its inversion, once the inversion is available the generation of one realization is very fast, just by multiplication of the inverted matrix by uncorrelated random vector.	there is no investment in computer time, but the generation of one realization is "expensive".
(9) ergodicity of the simulated field	non-ergodic because it handles a limited number of simulated points.	non-ergodic because it handles a limited number of simulated points.	ergodic, with lines evenly spaced at pre-specified directions even for a finite number of lines.
(10) conditioning	direct	direct	indirect, by Kriging
(11) efficiency of the method	efficient for a few points and a large number of realizations	efficient for few points and a large number of realizations.	efficient for many points and a few number of realizations.

10. Conclusions

The following conclusions can be drawn from this review,

(1) Selection of one of the presented techniques for applications to field problems depends on the available field data. Each technique requires specific information. The application depends also on the computer facilities available in terms of storage and speed.

(2) There is no clear proof that one of the numerical simulation techniques is significantly superior to any other, but it is obvious from Table (1) that the multi-variate method is more general, although it needs much storage compared with the turning bands method.

(3) Solution of stochastic differential equations governing subsurface fluid flow and transport are based on the assumption of Gaussian characteristics and stationarity of the input parameters. Recently, in the hydrogeological field, a considerable attention is devoted to the deviations from these assumptions. It has been shown from geological survey [Krumbein, 1967] that many geological patterns exhibit Markovian properties. Elfeki et al. [1995] has focused on this area of research.

(4) The presented stationary stochastic field models are attractive from a statistical point of view, but most of them are less applicable because they are either far from being realistic from geological point of view or they need intensive hard data (direct measurements of the hydrogeological parameters) to characterize the geological attributes in a proper way. Therefore, this conclusion has motivated new ideas to characterize geological features see e.g. Elfeki, et al. [1995], Elfeki, [1997] and Dekking et al. [1999].

11. References

Agterberg, F. P. (1974). *Geomathematics: Developments in Geomathematics 1*. Elsevier Scientific Publishing Co., Amsterdam, The Netherlands.

Baker, R. (1984). Modelling Soil Variability as a Random Field. *Mathematical Geology*. 16(5): 435-448.

Bakr, A. A. (1976). *Stochastic Analysis of The Effects of Spatial Variations of Hydraulic Conductivity on Groundwater Flow*, Ph.D. dissertation, New Mexico Institute of Mining and Technology, Socorro.

Bakr, A. A., Gelhar, L.W., Gutjahr, A. L.& MacMillan, J.R. (1978). Stochastic Analyses of Spatial Variability in Subsurface Flows. 1. Comparison of One and Three-dimensional Flows. *Water Resour. Res.* 14(2):263-271.

Bartlett, M. S. (1975). *The Statistical Analysis of Spatial Pattern.* Chapman and Hall, London.

Bear, J. (1979). *Hydraulics of Groundwater*, McGraw-Hill. New York.

Bennett, R. J. (1979). *Spatial Time Series*. Pion Limited, 207 Brondesbury Park, London.

Beran, M. J. (1968). *Statistical Continuum Theories.* Interscience, New York.

Bras, R. L. & Rodriguez-Iturbe, I. (1984) *Random Functions and Hydrology.* Addison-Wesley publishing company, Inc.

Borgman, L. & Taheri, M. & Hagan, R. (1984). Three-dimensional Frequency-domain Simulation of Geological Variables. In Verly, G. David, M. Journel, A. G. and Marechal, A. (Eds.), *Geostatistics for Natural Resource Characterization*, Part 2: Reidel, Dordrecht, Netherlands.

Brannan, J. R. & Haselow, J. S. (1993). Compound Random Field Models of Multiple Scale Hydraulic Conductivity. *Water Resour. Res.* 29(2), pp.365-372.

Chessa, A. (1995). *Conditional Simulation of Spatial Stochastic Models for Reservoir Heterogeneity.* Ph.D. Thesis. Delft University of Technology. The Netherlands.

Christakos, G. (1992). *Random Fields Models in Earth Sciences*. Academic, San Diego, Calif.

Clifton, P. M. & Neuman, S. P. (1972). Effect of Kriging and Inverse Modelling on Conditional Simulation of Avra Valley Aquifer in Southern Arizona. *Water Resour. Res*. 18(4): 1215-1234.

Cross, G. R. & Jain, A.K. (1983). Markov Random Field Texture Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* Vol. PAMI-5, no. 1.

Dagan, G. (1982a). Analysis of Flow Through Heterogeneous Random Aquifers, 2. Unsteady Flow in Confined Formations. *Water Resour. Res.* 18(5): 1571-1585.

Dagan, G. (1986). Statistical Theory of Groundwater Flow and Transport: Pore to Laboratory. Laboratory to Formation and Formation to Regional Scale. *Water Resour. Res.* 22(9): 120S-134S.

Davis, J. C. (1973). *Statistics and Data Analysis in Geology*. J. Wiley, New York.

Davis, M. W. (1987). Production of Conductional Simulations Via The LU Decomposition of The Covariance Matrix. *Math. Geol.* Vol. 19, no. 2, pp 91-98.

Dekking, F.M., Kraaikamp, C., Elfeki, A.M. & Bruining, J. (1999). Synthesis and Simulations of Digital Images by Tree-Indexed Markov Chains. *Proceeding of SPIE*, vol. 3815-23, 20-21 July, 1999.

Delhomme, J. P. (1979). Spatial Variability and Uncertainty in Groundwater Flow Parameters: a Geostatistical Approach. *Water Resour. Res*. 15(2): 269-280.

Delhomme, J. P. (1978b). Kriging in Hydroscience. *Advances in Water Resources* 1(5): 251-266.
Dettinger, M. D. & Wilson, J. L. (1981). First Order Analysis of Uncertainty in Numerical Models of Groundwater Flow, 1, Mathematical Development. *Water Resour. Res.*, 17(1): 149-161.
Elfeki, A. M. (1993). *Literature Review of Stochastic Simulation Models in Groundwater and Contaminant Transport*. Tech. Report # 355. Geotechnical Laboratory. Faculty of Civil Engineering. TU Delft. The Netherlands.
Elfeki, A. M. & Uffink, G. J. M. & Barends, F. J. B. (1995). Stochastic Simulation of Heterogeneous Geological Formations Using Soft Information, with An Application to Groundwater. In *Groundwater Quality: Remediation and Protection, QG'95.* Edited by Kovar, K.and Krasny, IAHS Publication No. 225.
Elfeki, A.M. (1996). *Stochastic Characterisation of Geological Heterogeneity and Its Impact on Groundwater Contaminant Transport.* A. A. Balkema Publishers, Rotterdam, The Netherlands.
Elfeki, A. M. M.& Uffink, G. J. M. & Barends, F. B.J (1996). Solute Transport in Single and Multiple Scale Heterogeneous Formations: Numerical Experiments*, geoENV 96*, *First European Conference on Geostatistics for Environmental Applications,* Lisbon, Portugal. Ed. by A. Soares, J. Hernandez & R. Froidevaux. Kluwer Academic Pub.,pp. 51-63.
Elfeki, A. M. M.& Uffink, G. J. M. & Barends, F. B.J (1998). A Coupled Markov Chain Model for Quantification of Uncertainty in Transport in Heterogeneous Formations. *GeoENV 98*, Second *European Conference on Geostatistics for Environmental Applications*. Ed. by A. Soares, J. Hernandez Kluwer Academic Publishers.
Elfeki, A.M.(1997). A Hybrid Stochastic Model for Characterisation of Subsurface Heterogeneity. *Mansoura University Engineering Journal (MEJ*), Vol. 22, No. 3, Dec. 1997.
Everitt, B. S. & Hand, D. J. (1981). *Finite Mixture Distributions.* Chapman and Hall. p. 143.
Farmer, C. L. (1988). The Generation of Stochastic Fields of Reservoir Parameters with Specified Geostatistical Distributions. Proceedings of The IMA Conference "*The Mathematics of Oil Production*". Oxford University Press.
Fayers, F. J. & Hewett T. A. (1992). Review of Current Trends in Petroleum Reservoir Description and Assessing The Impact on Oil Recovery. *Computational Methods in Water Resources IX Conference. Vol.2. Mathematical Modelling in Water Resources.* Computational Mechanics publications. Southhampton. Boston.
Freeze, R. A. (1975). A Stochastic-Conceptual Analysis of One-dimensional Groundwater Flow in Nonuniform Homogeneous Media. *Water Resour. Res.* 11(5): 725-741.
Gelhar, L. W. (1986). Stochastic Subsurface Hydrology: from Theory to Applications. *Water Resour. Res.* 22(9): 135S-145S.
Ghori, S. G., Heller, J. P., and Singh, A. K. (1992). An Efficient Method of Generating Random Fields. *J. Mathematical Geology*.
Gutjahr, A. L. (1989). *Fast Fourier Transforms for Random Field Generation.* Project Report for Los Alamos grant to New Mexico Tech. Contract number 4 R58 - 2690R.
Gutjahr, A. L., Gelhar, L. W., Bakr, A. A. & McMillan, J. R. (1978). Stochastic Analysis of Spatial Variability in Subsurface Flows. Part II: Evaluation and Application. *Water Resour. Res.* 14(5): 953-960.
Gutjahr, A. L., & Gelhar, L. W. (1981). Stochastic Models of Subsurface Flow: Infinite Versus Finite Domain and Stationarity. *Water Resour. Res.* 17(2): 337-350.
Haldorsen, H. H., Brand, P. J., MacDonald, C. J. (1987). Review of The Stochastic Nature of Reservoir. In *Mathematics in Oil Production.* Eds. S. Edwards and P. R. King. Oxford Science Publications.
Haldorsen, H. H. (1983). *Reservoir Characterization Procedures for Numerical Simulation.* Ph.D. dissertation. University of Texas, Austin.
Haldorsen, H. H., Brand, P. J., Macdonald, C. J. (1987). Review of the Stochastic Nature of Reservoir. In *Mathematics in Oil Production.* Eds. S. Edwards and P. R. King. Oxford Science Publications.
Haldorsen, H. H. & Damsleth, E. (1990). Stochastic Modeling. *Society of Petroleum Engineers. SPE 20321*. Distiguished Author Series.
Harbaugh, J. W. and Bonham-Carter (1970). *Computer Simulation in Geology*, J.Wiley, New York.
Hemmerle, W. J. (1967). *Statistical Computations on a Digital Computer*. Blaisdell publishing company.
Hewett, T. A. (1986). Fractal Distributions of Reservoir Heterogeneity and Their Influence on Fluid Transport. *Paper SPE 15386,* presented at the Sixty-First Annual Technical Conference. Society of Petroleum Engineers. New Orleans.
Johnson, M. E. (1987). *Multivariate Statistical Simulation.* John Wiley & Sons, Inc.
Journel, A. G. (1974). Geostatistics For Conditional Simulation of Ore Bodies. *Economic Geology*, 69(5): 673-687.
Journel, A. G. & Huijbregts, C. (1978). *Mining Geostatistics.* New York: Academic Press, 600p.
Kennedy, Jr. W. J., and Gentle, J. E. (1980). *Statistical Computing*. Marcel Dekker. New York.
Krumbein, W. C. (1967). Fortran Computer Program for Markov Chain Experiments in Geology. *Computer Contribution 13,* Kansas Geologic Survey, Lawrence, Kansas.
Krumbein, W. C. (1970). Geological Models in Transition in Geostatistics. In *Geostatistics* Ed. by D. F. Merriam.
Lin, C. & Harbaugh, J. W. (1984). *Graphical Display of Two and Three-Dimensional Markov Computer Models in Geology*. Van Nostrand Reinhold, New York.
Lippman, M. J. (1973). *Two-dimensional Stochastic Model of a Heterogeneous Geologic System*. PhD. Thesis. University of California, at Berkeley, California.
Lumley, J. L., and Panofsky, H. A. (1964). *The Structure of Atmospheric Turbulence*. J. Wiley, New York.
Mantoglou, A. & Wilson, J. L. (1982). The Turning Bands Method for Simulation of Random Fields Using Line Generation by Spectral Method. *Water Resour. Res.* 18(5): 1379-1394.
Mantoglou, A. (1987). Digital Simulations of Multivariate Two- and Three- dimensional Stochastic Processes with a Spectral Turning Bands Method. *Math. Geology*. 19(2): 129-149.

Matheron (1971). *Theory of Regionalized Variables and Its Applications*. 211pp., Ecole des Mines. Fontainebleau. France.
Matalas, N. C. (1967). Mathematical Assessment of Synthetic Hydrology. *Water Resour. Res.* 3(4): 937-945.
Mejia, J. M. & Rodriguez-Iturbe I. (1974). On The Synthesis of Random Field Sampling from The Spectrum: An Application to The Generation of Hydrologic Spatial Processes. *Water Resour. Res.* 10(4): 705-712.
Mood, A. M. and Graybill, F. A. (1963). *Introduction to The Theory of Statistics*.McGraw-Hill Book Co., Inc., New York.
Neuman, S. P. (1984). Role of Geostatistics in Subsurface Hydrology. In (Eds. Verly, G., David, M., Journel, A. G. & Marechal, A.) *Geo-statistics for Nature Resources Characterization*. Proc. NATO-ASI, Part 1, pp. 787-816. Reidel, Dordrecht, The Netherlands.
Peck, A., Gorelick, S., Marsily, G. de, Foster, S. & Kovalevsky, V. (1988). *Consequences of Spatial Variability in Aquifer Properties and Data Limitations for Groundwater Modelling Practice*. IAHS publication 175, IAHS Press, Oxfordshire.
Ripley, B. D. (1981). *Spatial Statistics*, J. Wiley, & Sons, Inc.
Ripley, B. D. (1987). *Stochastic Simulation*, J. Wiley & Sons, Inc.
Rubinstein, R. Y. (1981). *Simulation and the Monte-Carlo Method.* J. Wiley, New York.
Schwartz, F. W. (1977). Macroscopic Dispersion in Porous Media: The Controlling Factors. *Water Resour. Res.*, 13(4), pp. 743-752.
Shinozuka, M. & Jan, C. M. (1972). Digital Simulation of Random Processes and Its Applications. *J. of Sound and Vibration*, 25(1): 111-128.
Shinozuka, M. (1971). Simulation of Multivariate and Multidimensional RandomProcesses. *J. Acoust. Soc. Am.*, vol.(49), no. 1 (part 2), pp. 357-367.
Smith, L. & Freeze, R. A. (1979a). Stochastic Analysis of Steady State Groundwater Flow in a Bounded Domain. 1. One-dimensional Simulations. *Water Resour. Res.* 15(3): 521-528.
Smith, L. & Freeze, R. A. (1979b). Stochastic Analysis of Steady State Groundwater Flow in a Bounded Domain. 2. Two-dimensional Simulations. *Water Resour. Res.* 15(6): 1543-1559
Tompson, A. F. B. & Ababou, R. & Gelhar, L. W. (1989) Implementation of The Three-dimensional Turning Bands Random Field Generator. *Water Resour. Res.* 25(10): 2227-2243.
Tong, Y. L. (1990). *The Multivariate Normal Distribution*. Springer-Verlag. New York Inc.
Townley, L. R. (1984). Second Order Effects of Uncertain Transmissivities on Prediction of Piezometric Heads. In Proc. 5th Intern. Conf. on *Finite Elements in Water Resources*, Burlington, Vermont: Springer-Verlag.
Townley, L. R. & Wilson, J. L. (1985). Computationally Efficient Algorithms for Parameter Estimation and Uncertainty Propagation in Numerical Models of Groundwater Flow. *Water Resour. Res.* 21(12): 1851-1860.
Townley, L. R. (1988). Modelling Flow in Heterogeneous Aquifers: Identification of The Important Scales of Variability. In: *Developments in Water Science 35*. Proc. International Conference, MIT, USA, vol.(1) pp. 197-202. Elsevier.
Vanmarcke, E. (1977). Probability Modelling of Soil Profiles: Proc. *ASCE. Jour. of Geotechnical Eng. Div.* Vol.(103) No. GT11, pp. 1227-1247.
Vanmarcke, E. (1983). *Random Fields: Analysis and Synthesis*. Cambridge, Mass: MIT Press.
Warren, J. E., and Price, H. S. (1961). Flow in Heterogeneous Porous Media. *Soc. Pet. Eng. J.*, vol.(1), pp. 153-169.
Warren, J. E., and Price, H. S. (1964). Macroscopic Dispersion: Transactions AIME, *Soc. Pet. Eng. J.*, vol.(231), pp. 215-230.
Weber, K. J. (1986). How Heterogeneity Affects Oil Recovery. *Reservoir Characterization*. Academic Press, NY. pp. 487-544.
Westlake, J. R. (1968). *A Handbook of Numerical Matrix Inversion and Solution of Linear Equations.* Control Data Corporation, USA.
Whittle, P. (1954). On Stationary Processes in The Plane. *Biometrika,* 41, pp. 434-449.
Yevjevich, V. (1972). *Stochastic Processes in Hydrology*. Water Resources Publications. Fort Collins, Colorado, USA.