Estimation of various flow and mass transport parameters can be seen as a problem of spatial statistics. The definition of the properties of porous and fractured media in space and time using the concept of random functions provides means for (i) studying the inherent heterogeneity, (ii) evaluating the spatiotemporal variability of the properties, and (iii) assessing the uncertainty associated with their estimated values. In this chapter, the fundamentals of stochastic subsurface hydrology are presented. Applications include mapping hydrogeological properties, flow and transport in porous and fractured media, and inverse problems.

INTRODUCTION

Stochastic analysis has two main historic roots: One is in the analysis of the so-called random errors in measurements of physical entities, which has gradually embraced the study of irregularities and uncertainties observed in what were believed to be deterministic phenomena and has led to the discipline of mathematical statistics; the other root is in the study of games of chance, which has led to the theory of probability. Originally, both these branches of science were confined to the realm of mathematics and did not interfere much with the deterministic concept of the physical world governed by the laws of Newton’s mechanics and uniquely describable by the differential calculus. At that instance, the only problems seemed to be (i) to develop the theory in sufficient detail, (ii) to ascertain the initial conditions, and (iii) to avoid measurement errors. All discrepancies between theory and observation and uncertainties in the prediction of physical phenomena were attributed to human limitations. All this changed with the establishment of the quantum theory, which introduces the revolutionary concept that uncertainty, apart from being the result of human limitations, is also an intrinsic feature of matter itself. This concept has transformed statistics and probability from a tool dealing with noise in the laws of nature into a tool for the formulation of these laws themselves.

Undeniably, natural porous and fractured formations are heterogeneous, and display spatial variability of their petrophysical properties. This variability is of irregular and complex nature. It generally defies a precise quantitative description because of insufficient information at all relevant scales. In practice, however, only sparse scale-dependent measurements are available mainly due to limited costs. Owing to lack of detailed and exhaustive information, the higher the variability the higher is the uncertainty. For this very reason, stochastic approaches have been developed in subsurface hydrology to “fill in” or bridge the gap between the overall knowledge about the “physical law” and the strong spatial variation of the petrophysical properties, which are known through sparse tainted measurements. The theory of random, or stochastic, processes provides a natural framework for evaluating aquifer uncertainties. In the stochastic formalism, uncertainty is represented by probability or by related quantities like statistical moments. Boundary conditions, initial conditions, and parameters can be treated as random functions or fields whose values are determined by probability distributions conditional to sparse measurements or information.

The definition of the properties of porous media in space and time using the concept of random functions has two major advantages: (i) It conceptually defines the properties in space at a given point, without having to define a volume over which these properties must be integrated. (ii) It provides means for studying the inherent heterogeneity and variability of these properties in space, and for evaluating the uncertainty of any method.
of estimation of their values. The pioneering works of Schvidler (1962) and Matheron (1965) were the backbone of the work of Gedeon Dagan and Lynn Gelhar who have shaped and popularized what stochastic subsurface hydrology is today. A few monographs have since been published to cover the richness, strengths, and weakness of the stochastic methods. These monographs include a wealth of references, and we urge readers to consult them and the references therein. In this article, we will cover the fundamentals or in other words what became the need-to-know in stochastic subsurface flow and transport. For more advanced topics, we urge interested readers to consult most recent publications.

DETERMINISTIC, RANDOM, AND STOCHASTIC CONCEPTS

Of these three, the term deterministic is probably least ambiguous and involves a one-to-one relationship among events. The term random is often used as a synonym for stochastic, however, statistically speaking, there is a difference. The term random will be used for a completely irregular order or arrangement of elements from a given set. Thus, for instance, a random sequence or series will mean one whose terms are mutually independent and therefore uncorrelated. In the literature, this notion is often termed pure random. The term stochastic will be used for a relationship, a variable, or a process, incorporating both an element of randomness and an element of determinism. Thus a stochastic entity Y, that is, the logarithm of the hydraulic conductivity K, will be viewed as a sum \( Y = (Y) + (Y') \) where \((Y)\) and \((Y')\) are the deterministic and random components, respectively. The deterministic component is often interpreted as an overall trend or signal and the random component as error or noise. In agreement with the original meaning of the Greek word \( \sigma \tau \omega \phi \zeta \) (\([\text{stokhos} \text{ target or aim}\)) a stochastic relationship can be viewed as one aiming at the relationship indicated by the deterministic component. It follows that the two concepts, deterministic and random, are only special cases of stochastic and represent its two end limits approached as one or the other of the two components tends to zero. According to this interpretation, strictly deterministic relationships cannot be proved to exist in the real world since any verification involves measurement on the one hand and exact reproducibility on the other. Whether, due to the limited accuracy of measurement or due to the intrinsic impossibility of exact or exhaustive depiction of any specific petrophysical parameter, there is always a noise and purely deterministic relationship.

It may be noted that our intuitive perception of deterministic laws is in fact stochastic; we expect that actual measurements will agree only approximately with a given deterministic law. As a typical example from hydrogeology, one may take the concept of the pumping test. Though classified as a deterministic model, pumping tests in heterogeneous aquifer are expected to yield integrated and approximated value of the transmissivity (or hydraulic conductivity) and the storativity coefficient of the aquifer. Those values minimize the square of the errors between the theoretical model (the physical law) and the measurements (drawdown vs. time) and represent the deterministic component of the intuitively implied stochastic relationship between pumping rate and drawdown.

Similarly, any other deterministic law or model can be interpreted as a stochastic model in which the random component is not explicitly taken into account. It may be too small to be detected by the measuring apparatus or, though detectable, small enough in comparison with the deterministic component to be neglected; or, for the sake of simplicity, it may be filtered out. An important aspect of the above definition of stochasticity is that it does not imply any specific origin of the random element. It does not differentiate between a case where the latter accounts merely for the analyst’s limited knowledge of some deterministic components and one where it might be an inherent feature of the analyzed phenomenon.

Limited knowledge of information in both observed data (i.e. hydraulic head, concentration), and measured petrophysical properties (i.e. hydraulic conductivity, porosity) and their spatial distribution causes uncertain predictions. Spatial variability and uncertainty have led engineers and geologists to use probabilistic theories that translate the uncertainty to a random space function (RSF) or a random field, consisting of an ensemble of “infinite” number of equally probable realizations of parameter values, all having the same spatial statistics, particularly correlation structure. Imbedded in this approach is a geostatistical model.

Geostatistics is commonly used to analyze and interpolate between measurements using methods such as kriging (pronounced as in bridging), where the uncertainties in kriged (pronounced as in bridged) values are also quantified. Usually, these data are collected on different scales that may differ from the required scale of predictions. The task of quantitatively relating measurements and properties on different scales is difficult and intriguing. We briefly present in the following section the essence of geostatistics.

GEOSTATISTICS

Geostatistics has its basis in Matheron’s (1965) Theory of Regionalized Variables. A random variable is one that has a variety of values in accordance with a particular probability distribution (Journel and Huijbregts, 1978). If the random variable is distributed in space and/or time, we say that it is a regionalized variable. These variables, because of their spatial-temporal character, have a random as well as a
structural component. At first sight, a regionalized variable seems to be a contradiction. In one sense, it is a random variable that locally does not have any relation to the nearby variables. On the other hand, there is a structural aspect in the regionalized variable that depends on the distance of separation of the variables. Both characteristics can be described, however, using a random function for which each regionalized variable is a particular realization. By incorporating the random as well as the structural aspects of a variable in a simple function, the spatial variability can be addressed on the basis of the spatial structure shown by these variables. In this sense, a regionalized variable is a variable that qualifies a phenomenon that is distributed through space and/or in time and that presents a certain correlation structure. Following are useful applications of geostatistics:

- **Characterization of large data sets**: For spatially distributed data, the resolution of measurement points is mostly rather low. By contrast, time series often consist of thousands of measurements. Therefore, condensing the information may be necessary. In this context, the probability density function (pdf) of the measured quantity (which is characterized by its mean and variance if the pdf is Gaussian) and some information about the average persistence (like the auto-covariance function, variograms) may be adequate.

- **Interpolation**: Given point-measurements in space, one is asked to create maps of the measured quantity. This requires interpolation. In the geostatistical framework, spatial interpolation is identical to conditioning, that is, in the vicinity of a measurement point, the expected value of the unknown is, due to the spatial correlation, weighted toward the measured value, and the variance is decreased. Taking the mean of the conditional pdf as interpolated value is referred to as kriging. In contrast to other interpolation techniques, kriging allows to evaluate and assess the uncertainty of the interpolated value indicating also where an additional measurement would be most valuable.

- **Inverse modeling**: The geostatistical approach of inverting is essentially identical to that of interpolation. The difference is that cross-covariances describing the spatial correlation between different variables (such as hydraulic head and log-conductivity) are used rather than auto-covariance functions. This is referred to as cokriging. As will be shown later, cross-covariance functions can be calculated from the auto-covariance functions and the governing partial differential equation. Like kriging, cokriging yields not only the best estimate but also the spatial distribution of the uncertainty related to that estimate.

- **Monte Carlo simulations**: (MCSs): MCS is somewhat the experimental numerical apparatus of statisticians. By creating equally probable conditional realizations, preferably conditional to measured data, of the hydraulic conductivity field for example, one can simulate the flow and transport for each realization. All the results are then averaged to predict the probabilistic behavior of the flow and transport in the aquifer. In this context, using multiple realizations is justified by the uncertainty in the interpolation between measurements.

### Basic Statistical Treatment

As a first step, the quality of the data is proved, and the basic statistical calculations from which the statistics or measurements are obtained are performed. These are the numerical values that enable us to characterize and compare the statistical distributions. These statistics can be made on the variable itself, or some transformation of the variable (e.g., logarithmic, log). They can be of several types, as follows:

1. **Measurements of centralization**: These measurements indicate a value around which the distribution values are distributed. They are generically known as means, and have the following forms. The arithmetic mean or mean, \( m \). The median, \( M \), is the value that divides the population into two equal parts. It is the quartile of 50% (Q2). The lower quartile, Q1, is the value with 25% of the population below it and the remaining 75% above it. The upper quartile, Q3, is the value with 75% of the population below it and the remaining 25% above it. The mode, Mo, is the value of the greatest absolute frequency of a distribution. Other means such as harmonic and geometric means are of great interest. The geometric mean is defined as the mean of the log-transformed measurements, and the harmonic mean as the mean of the inverse of the measurements.

2. **Measurements of dispersion**: These measurements indicate the variability of dispersion of the values of a distribution. The variance, \( \sigma^2 \), is the average of the squares of the deviations with respect to the mean. The standard deviation, \( \sigma \), is the square root of the variance and represents the margin of the variation or the error of estimation in which the data analyzed are included. It has the same units as the variable under consideration. The range, \( \forall \), of a distribution is the difference between the extremes, maximum and minimum values. The coefficient of variation, CV, is the quotient of the standard deviation and the mean. It enables us to compare distributions that have different units. There is a direct relation between the value of this parameter and the dispersion of the distribution. The interquartile range, IR, is the difference between the upper quartile and the lower quartile and it indicates the range between which 50% of the central values of the population are distributed.

3. **Distribution moments**: These can be central or with respect to the origin. The central moment of order \( r \),
Table 1 Basic Univariate statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations #</td>
<td>$N$</td>
</tr>
<tr>
<td>Histogram and cumulative histogram</td>
<td>$F(x_i), CF(x_i)$</td>
</tr>
<tr>
<td>Minimum</td>
<td>$\min_i \forall x_i, \min x_i, i = 1, \ldots, N$</td>
</tr>
<tr>
<td>25th% Percentile</td>
<td>$Q_i$</td>
</tr>
<tr>
<td>Mode</td>
<td>$Mo$</td>
</tr>
<tr>
<td>Median or 50th % Percentile</td>
<td>$M$</td>
</tr>
<tr>
<td>Mean</td>
<td>$m = \frac{1}{N} \sum_{i=1}^{N} x_i$</td>
</tr>
<tr>
<td>75th % Percentile</td>
<td>$Q_3$</td>
</tr>
<tr>
<td>Maximum</td>
<td>$\max_i \forall x_i, \max x_i, i = 1, \ldots, N$</td>
</tr>
<tr>
<td>Variance</td>
<td>$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - m)^2$</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - m)^2}$</td>
</tr>
<tr>
<td>Coefficient of variation (%)</td>
<td>$CV = \frac{\sigma}{m}$</td>
</tr>
<tr>
<td>Range</td>
<td>$R = \max - \min$</td>
</tr>
<tr>
<td>Interquartilic range</td>
<td>$IR = Q_3 - Q_4$</td>
</tr>
<tr>
<td>3rd Order central moment $M_3$</td>
<td>$M_3 = \frac{1}{N} \sum_{i=1}^{N} (x_i - m)^3$</td>
</tr>
<tr>
<td>Coefficient of skewness</td>
<td>$CS = \frac{M_3}{\sigma^3}$</td>
</tr>
<tr>
<td>4th Order central moment $M_4$</td>
<td>$M_4 = \frac{1}{N} \sum_{i=1}^{N} (x_i - m)^4$</td>
</tr>
<tr>
<td>Coefficient of Kurtosis</td>
<td>$CK = \frac{M_4}{\sigma^4}$</td>
</tr>
</tbody>
</table>

$M_r$, is the average of the deviation with respect to the mean taken to order $r$.

4. **Asymmetry or skewness measurements**: These provide an idea of the asymmetry of the distribution. A distribution is symmetric when the frequencies corresponding to equidistant values with respect to a central value are equal. In the ideal symmetry condition, the values of the mean, median, and mode coincide. A distribution has a bias to the right or is positive if the frequencies descend more slowly on the right of the histogram. In this case, the mean is greater than the mode. A distribution has its bias to the left or is negative if the frequencies descend more slowly on the left of the histogram. In this case, the mean is less than the mode. The coefficient of skewness, $CS$, calculated in terms of the moments, is the ratio between the moment of order 3 and the standard deviation cubed.

5. **Measurements of kurtosis**: These indicate the degree of kurtosis with respect to a normal or Gaussian distribution. The coefficient of kurtosis, $CK$, calculated in terms of the moments, is the ratio between the moment of order 4 and the standard deviation taken to the fourth power. In a Gaussian or normal distribution, this coefficient $CK$ is 3 and its curve is normal or mesokurtic. If $CK$ is greater than 3, the distribution has a leptokurtic curve, which is sharper than the Gaussian one. If $CK$ is less than 3 the distribution has a platikurtic curve, which is flatter than the Gaussian one. The analytic expressions of the above concepts are given in Table 1.

**Geostatistical Description of Spatially Variable Parameters**

Depicting heterogeneity of aquifers is a daunting task and could be achieved through two approaches. On the one hand, solely on the basis of the geological deposition processes, one can create images (realizations) of the subsurface heterogeneity (Gelhar, 1992; Rubin, 2003). Unfortunately, the geology-based models are limited because it is difficult to integrate the geological observations. However, statistical and geostatistical methods honor the measured data regardless of their origin: hard data (direct measurements of the variable to be mapped, primary variable) or soft data (indirect measurements of the variable in consideration, secondary variable). Honoring data is usually
referred to as conditioning. Conditioning on soft data usually necessitates a physical law and/or a linear or nonlinear relationship or correlation that bind the secondary variable to the primary one (i.e., Ezzeddine et al., 1999).

A statistical description of hydraulic properties is equivalent to interpreting the distribution as an outcome of a random process. As a consequence, an infinite number of different distributions sharing the same statistical description are possible. Introducing conditioning decreases the variability but does not change the fact that an infinite number of realizations are equally probable. Our limited knowledge of the “true” setting through only contaminated sparse measurements of hydraulic properties made the stochastic techniques a must for interpolation and uncertainty assessment.

For example, one will take a certain number of core samples and analyze them to get point information about the hydraulic conductivity. On the basis of these data one may estimate the hydraulic conductivity distribution. At every location within the aquifer, the conductivity value is estimated by interpolation and so its level of confidence or uncertainty. It is rather unlikely that the interpolated value is the true one. Therefore, using the interpolated conductivity field as unique (single) realization in a flow and transport simulator may lead to biased outcome.

**Statistics of Spatial Variables**

In order to describe a spatial field, such as the hydraulic conductivity within an aquifer, by geostatistics, we must conduct as many point-measurements of the quantity of interest as possible. Without loss of generality, we focus on the log-hydraulic conductivity $Y = \ln(K)$. The simplest single-variate statistical description of the data set is the mean (1st moment) and the variance (2nd moment):

$$E[Y] \equiv \bar{Y} = \frac{1}{n} \sum_i Y_i = \int yf_Y(y) \, dy$$

(1)

$$\text{Var}[Y] = E[(Y - E[Y])^2] = \frac{1}{n - 1} \sum_i (Y_i - \bar{Y})^2$$

(2)

where $Y_i(x)$ is the value of $Y$ at a location $x_i$, and $E[]$ and $\text{Var}[]$ are the expected, and the variance operator, respectively. If $Y$ is normal (Gaussian), then the first and second moment are sufficient to describe $Y$, otherwise higher statistical moments are needed. They further enhance the statistical characterization of $Y$, the hydraulic conductivity is considered as a spatial random function (SRF). Its value may differ from one location, $x_1 = (x_1, x_2, x_3)$, to another, $x_2 = (x_1, x_2, x_3)_2$. Consequently the similarity between the values of $Y$ at different locations is expressed by the covariance function, which is defined by

$$\text{Cov}[Y(x_1)Y(x_2)] = E[(Y(x_1) - E[Y(x_1)])(Y(x_2) - E[Y(x_2)])]$$

$$- E[Y(x_2))]$$

(3)

$\text{Cov}[]$ is the covariance operator. It is worth mentioning that the covariance does not always exist, given the type of variability of the variable. When dealing with a single data set, $\text{Cov}$ is called auto-covariance. Otherwise, when dealing with multiple data sets, that is, the hydraulic conductivity and head, the similarity is termed as cross-covariance. Furthermore, when a SRF is said to be stationary, the $\text{Cov}$ operator depends solely on the separation distance $h = x_1 - x_2$. $\text{Cov}[Y(x_1)Y(x_2)] = \text{Cov}[Y(h)]$, regardless their locations, otherwise the SRF is location-dependent and called nonstationary. Stationarity has different levels, that is, weak and strong and we will discuss them later. It is worth noting that when $h$ vanishes, the covariance reduces to $\sigma^2_Y$. Similarly, the correlation function or correlogram $\rho$ of a SRF is also defined. The correlation function $\rho(h) = \text{Cov}[Y(h)]/\sigma_Y^2$ takes values in $[-1, +1]$. When $\rho = 0$ there is no correlation, the data sets are independent. Otherwise when $\rho = \pm 1$, there is a total (perfect) positive/negative (direct/indirect) correlation. The definition of the covariance assumes that the mean is known, which in reality may not be the case, and that a covariance exists. To overcome this limitation, Matheron suggested the use of another similarity measure, the variogram $\gamma$ defined by

$$\gamma_Y(h) = \frac{1}{2}E[(Y(x_i) - Y(x_j))^2] = \frac{1}{2n(h)} \sum_{i,j \neq i, j} (Y_i - Y_j)^2$$

(4)

The word semivariogram stems from 1/2 term in the above notation. The names variogram and semivariogram are used interchangeably. The variogram has two attractive properties: (i) $\gamma$ is linked to the $\text{Cov}$ through the variance $\sigma_Y^2$, (ii) it is defined even when the variance itself is not finite, that is, does not exist. A stationary SRF with a, not necessarily known, constant mean and a variogram is referred to as intrinsic hypothesis. Note that these statistical properties can also be defined in space and time, called spatiotemporal statistical tools and we urge the reader to consult Christakos (1992) for more details. These tools have become the main engine behind geology-related estimation problems since the 1960s. Matheron coined the word geostatistics since it relates the field of geology to the field of statistics as is in geophysics for example. In the next section we will tackle the variogram analysis, that is, the main step behind any geostatistical application.

**Typical Covariances and Variograms**

For the sake of clarity, we introduce only the typical covariances and variograms under stationarity hypothesis. The common models are (i) exponential, Gaussian, spherical,
and power-law models. The exponential model is given by

\[ C_Y(h) = \sigma_Y^2 e^{-h}, \quad h = \sqrt{\sum_i \left( \frac{h_i}{\lambda_i} \right)^2}, \]

\[ \gamma(h) = \sigma_Y^2 (1 - e^{-h}) \]

where \( \lambda_i \) are correlation lengths (range, scaling parameters which can be different in different directions, in case of anisotropy of the variability, Figure 1). Because \( h \) is positive, the variogram slowly reaches a finite limit for large values of \( h \). This limit is called sill and, in this particular case, it is equal to the variance. Because the exponential model has very attractive mathematical properties, various analytical solutions for stationary \( Y \) random fields with an exponential covariance have been developed. One should notice that the derivative at the origin does not vanish. This causes small-scale variability. Such a behavior does not occur in the Gaussian model given by

\[ C_Y(h) = \sigma_Y^2 e^{-h^2}, \quad h = \sqrt{\sum_i \left( \frac{h_i}{\lambda_i} \right)^2}, \]

\[ \gamma(h) = \sigma_Y^2 (1 - e^{-h^2}) \]

Contrary to the exponential model, the Gaussian model converges faster to the sill (Figure 2). The Gaussian model is much smoother and gradual interpolator than the exponential model. This is different in the spherical model:

\[ C_Y(h) = \begin{cases} \sigma_Y^2 \left[ 1 - \frac{3}{2} h + \frac{1}{2} h^3 \right], & h \leq 1 \\ 0, & \text{otherwise} \end{cases} \]

\[ h = \sqrt{\sum_i \left( \frac{h_i}{\lambda_i} \right)^2} \]

\[ \gamma(h) = \begin{cases} \sigma_Y^2 \left[ \frac{3}{2} h - \frac{1}{2} h^3 \right], & h \leq 1 \\ \sigma_Y^2, & \text{otherwise} \end{cases} \]

The behavior of the spherical model near the origin is nearly linear. Unlike the exponential model, there is no correlation when \( h \) is greater than 1. The power-law, also called self-similar or fractal, model in the isotropic case is given by:

\[ \gamma_Y(h) = C h^H, \quad 0 < H < 1 \]

where \( C \) is a constant; \( H \) is called the Hurst number/exponent. For an exponent \( H = 1/2 \), the power variogram reduces to the linear variogram. \( Y \) does not have a finite variance because the variability is unbounded, and a covariance function cannot be defined, only a variogram.

Theoretically, variograms vanish at the origin, however, very often, variograms exhibit a finite jump at the origin. This apparent jump at the origin is called nugget effect (originated from the mining industry). The nugget effect is taken into account by adding a nugget contribution, \( \gamma^0 \), to \( \gamma \) the variogram fitted to the data as if \( \gamma^0 \) were the origin

\[ \gamma_Y(h) = \gamma^0 (1 - \delta(h)) + \gamma'(h) \]

where \( \delta(h) \) is the Kronecker or delta function. A nugget variogram expresses purely random phenomena without any spatial structure. The contribution of the nugget effect may be interpreted as a measurement error or as variability at scales smaller than the measurement scales. Any combination of the mentioned models is possible, and it is called a nested model or nested structure.

Variograms are direction-dependent; therefore it is useful to evaluate the variogram at different directions (i.e. North, East, West, South and any combinations). Generally, variograms do not show anisotropy, different behavior for different direction. If they do, then: (i) it may be a sign that
is defined by graphical depth. To assure that the measurement are correlated with respect the principal directions of the heterogeneities or a transformation into an axisymmetric, otherwise it is anisotropic.

The directional integral scale \( I_i \) of a covariance function is defined by

\[
I_i = \frac{1}{\sigma_y^2} \int_0^\infty \frac{\text{Cov}[h_i]}{\text{Cov}[0]} \, dh_i
\]

where \( I = \lambda \) for the exponential model, \( \lambda \sqrt{\pi}/2 \) for the Gaussian and \( 3\lambda/8 \) for the spherical model. When \( I_1 = I_2 = I_3 \) the field is isotropic, when \( I_1 = I_2 \neq I_3 \) the field is axisymmetric, otherwise it is anisotropic.

For sloped strata, a transformation of coordinates into the principal directions of the heterogeneities or a transformation into the stratigraphic coordinates may be necessary to assure that the measurement are correlated with respect to the geological depositions/layering and not to the geographical depth.

**Calculation of Experimental Variograms**

To estimate a variogram, we use the measurement points \( Y_i \) and assume ergodicity (i.e. space averages can be used to estimate the averages in the entire set of realizations). First, we define a certain number of distance classes between measurement points. Then, taking into account all possible pairs or points for each class of distances, we calculate: (i) the number of pairs present in that class, (ii) the average distance in the class, and (iii) the average square increments \( 1/2(Y_i - Y_j)^2 \). This is referred to as experimental variogram. Generally, the number of pairs is not evenly distributed between the classes of distances due to the scattering of the measurements. There are more pairs at shorter than longer distances. The variogram becomes more uncertain as the separation distance \( h \) increases. Once the experimental variogram is calculated, one may fit one of the theoretical models to the experimental variogram. Curve fitting may be biased by the choice of classes and the averaging step. More details on the determination of a fitted variogram are given in Journel and Huijbregts (1978), Isaaks and Srivastava (1989), Kitanidis (1997) to name a few.

**Kriging**

Consider a set of \( n \) measurements of \( Y \) with \( Y_i \) measured at the location \( x_i \). The variogram \( \gamma_{YY} (h) \) is known. We want to estimate the value \( Y_0 \) at a location \( x_0 \). A linear estimator is a linear combination of the measured data \( Y_0^* = \sum_i \lambda_{0,i} Y_i \) in which \( \lambda_i \) is the weighting factor for measurement \( Y_i \) and the * indicates that \( Y_0 \) is estimated. If the true unknown value at \( x_0 \) is \( Y_0 \), the estimation is said to be an optimal estimate in the minimum variance sense if \( \mathbb{E}[(Y_0^* - Y_0)^2] \) is minimal. We also require an unbiased estimator; in other words, the expected value of the estimation error is zero: \( \mathbb{E}[(Y_0^* - Y_0)] = 0 \). The unbiasedness condition yields a constraint on the choice of the weights, that is \( 1 = \sum_i \lambda_{0,i} \), which assumes a constant mean throughout out the domain \( \mathbb{E}[Y(x)] = m \). The mean \( m \) is not necessarily known. The minimum variance conditions leads to

\[
\begin{align*}
\text{Min } \mathbb{E}[(Y_0^* - Y_0)^2] &= \sigma_Y^2(x_0) = -\sum_i \sum_j \lambda_i \lambda_j \\
\gamma_{YY}(x_i - x_j) + 2 \sum_i \lambda_i \gamma_{YY}(x_i - x_0) \\
\text{subject to } 1 &= \sum_j \lambda_{0,j}
\end{align*}
\]

To minimization problem, an additional unknown referred to as Lagrangian multiplier, \( \mu \), (Matheron, 1965) is introduced to enforce the constraint. Taking the derivatives with respect to \( \lambda_i \) and \( \mu \), and finding the minimum by equating the derivatives to zero, leads to the following system of \( n + 1 \) linear equations:

\[
- \sum_j \lambda_j \gamma_{YY}(x_i - x_j) + \mu = \gamma_{YY}(x_i - x_0) \\
\forall i = 1, n \text{ and } 1 = \sum_i \lambda_i
\]

This is known as kriging system named by Matheron after the South African mining engineer Krige. Because the variogram is a symmetric function, the matrix is symmetric also, however, it is not positive definite. One attractive property of Kriging is that the inverse of the matrix has to be done only once and stored since it does not depend on the location \( x_0 \); however, the right-hand vector does depend on \( x_0 \) and has to be updated for each location. The kriging estimator is also referred to as the Best Linear Unbiased Estimator (BLUE). Expression for the estimation variance is as follows (Matheron, 1965, see also de Marsily, 1986):

\[
\sigma_Y^2(x_0) = \sum_j \lambda_j \gamma_{YY}(x_j - x_0) - \mu
\]

The kriging estimator with a constant unknown mean as described here is called ordinary kriging. When the mean is known it is called simple kriging. One may argue that the contributions of distance measurements are weak and therefore only the immediate measurements within a certain neighborhood of \( x_0 \), should be considered. In this case, the matrix to be inverted is neighborhood-dependent and has to
be assembled for each point-estimation. Readers are urged to look up Deutsch and Journel (1997), Kitanidis (1997), and Chiles and Delfiner (1999).

Cokriging

Hydraulic conductivity measurements are more intensive to obtain than hydraulic head and, therefore, expensive. Generally, $K$ is estimated using pumping tests; and thus it is not a local measurement but rather an integrated measurement. Head measurements, however, are affordable and sometimes abundant in number. It is imperative then to benefit from the cross-correlation between both SRFs to improve point-estimation of either variable through a more generalized kriging process called cokriging. We illustrate it on an exhaustive set of $N_Y$ and $N_H$ measurements of hydraulic conductivity and head, respectively. Assuming that the both $E[Y]$ and $E[H]$ are known the simple Cokriging system for $h$ reads (Rubin, 2003):

$$h(x_0) = E[h(x_0)] + \sum_{i=1}^{N_Y} \lambda_i (Y_i - E[Y]) + \sum_{i=N_Y+1}^{N_Y+N_H} \lambda_i (h_i - E[h]) \quad (14)$$

$$C_{YH}(x_l, x_i) = \sum_{i=1}^{N_Y} \lambda_i C_{YH}(x_l, x_i) + \sum_{i=N_Y+1}^{N_Y+N_H} \lambda_i C_{YH}(x_l, x_i); \quad \forall l = 1, N_Y \quad (15)$$

$$C_{HH}(x_0, x_l) = \sum_{i=1}^{N_Y} \lambda_i C_{YH}(x_l, x_i) + \sum_{i=N_Y+1}^{N_Y+N_H} \lambda_i C_{HH}(x_l, x_i); \quad \forall l = N_Y + 1, N_Y + N_H \quad (16)$$

and the simple Cokriging system for $Y$ reads

$$Y(x_0) = E[Y] + \sum_{i=1}^{N_Y} \lambda_i (Y_i - E[Y]) + \sum_{i=N_Y+1}^{N_Y+N_H} \lambda_i (h_i - E[h]) \quad (17)$$

$$C_{YH}(x_0, x_l) = \sum_{i=1}^{N_Y} \lambda_i C_{YH}(x_l, x_i) + \sum_{i=N_Y+1}^{N_Y+N_H} \lambda_i C_{YH}(x_l, x_i); \quad \forall l = 1, N_Y \quad (18)$$

$$C_{HH}(x_0, x_l) = \sum_{i=1}^{N_Y} \lambda_i C_{YH}(x_l, x_i) + \sum_{i=N_Y+1}^{N_Y+N_H} \lambda_i C_{HH}(x_l, x_i); \quad \forall l = N_Y + 1, N_Y + N_H \quad (19)$$

where $E[h(x_0)]$ is the unconditional head mean; $E[Y]$ is the conditional mean of $Y$; $C_{YH}$ is the $Y-H$ cross-correlation; $C_{YY}$ and $C_{HH}$ are the auto-correlations of $Y$ and $H$, respectively. When $C_{HH}$ is not defined (unbounded variability) $\gamma_{HH}$ is used instead and an additional constraint on the unboundedness is added.

Closed form of $C_{YH}$ and $\gamma_{HH}$ for stationary $Y$ field assuming an exponential covariance function $C_{YY}$, a uniform steady flow in an infinite domain have been obtained by Dagan (1989) using 2nd order perturbation of the flow field for 2D and 3D cases. The reader is referred to Dagan (1989) and Rubin (2003) for the analytical expressions.

The cokriging process can be generalized to $N_Z$ different $Z$ type of measurements. For illustrative applications of cokriging please (see Kitanidis, 1997) and the section on “Inverse Problems as Statistics”. It is important to note that the back transform form $Y$ to $T$ is just $T = \exp(Y)$ and that no additional terms are justified, as some authors choose to correct for the biased mean and add a term related to the variance of $Y$ (de Marsily, 1986).

Unconditional and Conditional Simulations

Matheron proposed and implemented the unconditional simulation called Turning Bands method. From the 1980s onward, many applications appeared, almost all of which were dedicated to geological mining. In the 1990s, many theoretical aspects have been developed, including new more efficient algorithms such as applications relevant to multiple fields, analysis of basins, treatment of images, simulation of porous and fractured media, simulations of geological lithofacies, and so on. This development has been aided by the widespread use of computers, making it possible to deal with the numerous calculations required. The values of the models obtained using geostatistical simulation agree with the experimental information and reproduce the observed variability. The fact that the variograms of the simulated values and the real values coincide implies that both sets of values have the same spatial and/or temporal variability. We will introduce two different ways on generation random fields. The Turning Bands method, which is an unconditional method, is presented. The Sequential Gauss Simulations (SGS) is then introduced as an example of conditional method. A more sophisticated Bayesian random field will be presented in the Section “Numerical methods – Monte Carlo solution”.

Unconditional Simulations: Turning Bands Method

The Turning Bands method involves the simulation of isotropic random fields in two- or higher-dimensional space by using a sequence of one-dimensional processes along lines crossing the space. The algorithm can be described as follows (Figure 3):

1. Choose an arbitrary origin within or near the domain of the field to be generated.
2. Select a line \( i \) crossing the domain having a direction given by the unit vector \( \mathbf{u}_i \), which may be chosen either randomly or from some fixed set.

3. Generate a realization of a one-dimensional process, \( Z_i(l_i) \), along the line \( i \) having zero mean and covariance function \( B_1(d_i) \) where \( l_i \) and \( d_i \) are measured along line \( i \).

4. Orthogonally project each field point \( x_k \) onto the line \( i \) to define the coordinate \( l_{ki} = x_k \cdot \mathbf{u}_i \) of the one-dimensional process value \( Z_i(l_{ki}) \).

5. Add the component \( Z_i(l_{ki}) \) to the field value \( Z(x_k) \), for each \( x_k \).

6. Return to step (2) and generate a new one-dimensional process along a subsequent line until \( L \) lines have been produced.

7. Normalize the field \( Z(x_k) \) by dividing through by the factor \( \sqrt{L} \).

Essentially, the generating equation for the zero-mean discrete process \( Z(x) \) is given by

\[
Z(x) = \frac{1}{\sqrt{L}} \sum_{i=1}^{L} Z_i(x_i \cdot u_i) \quad (20)
\]

which can be an exceptionally fast algorithm, particularly as the number of dimensions of the process increases. It depends on the knowledge of the one-dimensional covariance function, \( B_1(d) \). Once this is known, the line processes can be produced using some efficient 1-D algorithm such as autoregressive, moving average, or FFT techniques. The covariance function \( B_1 \) is chosen such that the multidimensional covariance structure \( B_n \) in \( R^n \) is reflected in each realization or over the ensemble. For two-dimensional isotropic processes Cressie (1993) gives the following relationship between \( B_2 \) and \( B_1 \) for \( r = |d| \),

\[
B_2(p_1, p_2) = \sigma^2 \exp \left\{ -\frac{2}{\gamma} \sqrt{p_1^2 + p_2^2} \right\} \quad (21)
\]

Cressie (1993) supplied explicit solutions for either the equivalent one-dimensional covariance function or the equivalent one-dimensional spectral density function for a variety of common multidimensional covariance structures. In particular, for the exponential type covariance function,

\[
B_2(p_1, p_2) = \sigma^2 \exp \left\{ -\frac{2}{\gamma} \sqrt{p_1^2 + p_2^2} \right\} \quad (22)
\]

The line processes could be constructed using a 1-D FFT algorithm. Line lengths were chosen to be twice that of the field diagonal to avoid the symmetric covariance problem inherent with the FFT method. To reduce errors arising due to overly coarse discretization of the lines, the ratio between the incremental distance along the lines, \( \Delta u \), and the minimum incremental distance in the field along any coordinate, \( \Delta x \), is constrained by \( \Delta u / \Delta x = 1/2 \).

It is worth mentioning here that the unconditional random field can be made conditional with a simple double kriging. Otherwise conditional field can obtained using conditional sequential methods (Cressie, 1993). Figure 4 depicts an example of a 2D random field.

**Conditional Simulations: Sequential Gaussian Simulations**

The procedure required producing a random field where, on one hand, at some specific locations some of the measurements \( Z \) are known and must be honored and, on the other hand, at the other locations, where measurements are lacking, must be estimated based on the inferred
correlation structure of the neighboring measurements. This can be accomplished through the kriging estimation. The interpolation by kriging leads to rather smooth distributions of the spatial variable, that is, the interpolated values themselves, although based on a certain variogram, do not satisfy the underlying variogram. This is due to the fact that at each point of interpolation the most likely value is chosen, neglecting the estimation variance. For many applications such as drawing a map of the spatial variable, the smooth estimate by kriging is adequate. However, using the distribution in numerical simulations of flow and transport processes one may undermine the heterogeneities effects. In contrast to the kriging interpolation, there are an unlimited number of possible realizations satisfying the underlying spatial variability variogram. These multiple realizations can then be used for MCSs. Several methods have been developed for the generation of geostatistical realizations (GSLIB library) (Deutsch and Journel, 1997). Here we will restrict our description to the sequential Gaussian simulation (SGSIM, or SISIM its analogue for indicator data). The basic principle of SGSIM is to perform multiple kriging, but rather than choosing the conditional mean at the point of interpolation, a random value is taken from the Gaussian distribution described by the estimation mean and variance. The algorithm consists of the following steps:

1. Choose a random sequence in which all points of interpolation will be considered.
2. For point \( i \), calculate the kriging estimate and its variance using all measured values and the \( i \) already generated values for conditioning.
3. Choose a random value from the distribution described by the Gaussian distribution with the conditional mean and estimation variance calculated from kriging.
4. Go to step 2 unless all points have been considered.

The algorithm may be made more efficient if conditioning is restricted to points within a certain neighborhood (otherwise the computational effort for the kriging step increases linearly). An example of Sequential Indicator Simulation (SISIM) is given in Figure 5. As expected both realizations are similar around the boreholes.

**STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS AND THEIR SOLUTIONS**

Stochastic modeling deals with solving stochastic partial differential equations (SPDEs). To illustrate the concept of SPDE, we limit ourselves to the flow equation. The flow equation in porous media is based on the continuity of the mass balance concept. It is a partial differential equation (PDE) and, under transient conditions, is given by:

\[
S_s \frac{\partial h}{\partial t} - \nabla.(K \nabla h) = 0
\]  

(24)

If the hydraulic conductivity of the porous media \( K \) and the storativity \( S_s \) are assumed random functions, then the flow equation is called a SPDE. This means that the solution of the equation itself, the head \( h \), is also a random function. Solving the flow equation implies determining the pdf of \( h \), in particular, its first moments from the prescribed pdf of \( K \) and \( S_s \). We introduce in the next section three ways for solving SPDEs, that is, the spectral, perturbation, and the Monte Carlo methods.

**Spectral Methods**

This method is applicable to the 2nd order stationary stochastic processes for both inputs and outputs. If \( Y(x) \) is a 2nd order stationary process, the spectrum (or spectral density) of \( Y \) is the Fourier transform and the inverse Fourier transform of its covariance function are given by:

\[
\varphi(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\omega s} \text{Cov}[Y(x+s), Y(x)] \, ds;
\]

\[
\text{Cov}[Y(x+s), Y(x)] = C(s) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{iks} \varphi(\omega) \, d\omega
\]  

(25)

The representation theorem (Wiener-Kintchine) states that if the 2nd order stochastic process \( Y(x) \) is of zero mean \( \text{E}[Y] = 0 \) and of covariance \( C(s) \), then there is an associated complex process \( Z \) (and \( Z^* \) its conjugate) that satisfies the
Fourier–Stieltjes integral:

\[ Y(x) = \int_{-\infty}^{+\infty} e^{i\omega x} dZ(\omega) \]

\[ E[dZ(\omega_1) dZ^*(\omega_2)] = 0 \quad \text{if} \quad \omega_1 \neq \omega_2 \quad \text{however} \]

\[ E[dZ(\omega_1) dZ^*(\omega_2)] = \varphi(\omega_1) \quad \text{if} \quad \omega_1 = \omega_2 \quad (26) \]

Gelhar (1992) gives a one-dimensional steady state flow in an infinite domain, as example, and it is reproduced here for completeness. The steady state flow equation reads:

\[ \frac{d}{dx} \left[ K(x) \frac{dh}{dx} \right] = 0 \quad (27) \]

We assume that \( K(x) \) is a 2nd order stationary stochastic process, \( h \) is the head. Assuming a constant flow rate, \( q \), throughout the domain, the integration of the flow equation once yields to:

\[ \frac{dh}{dx} = -\frac{q}{K(x)} = -r(x) \quad q \quad (28) \]

because \( K \) is a nonzero finite physical property, one can define the hydraulic resistivity \( r(x) = K(x)^{-1} \). The goal is then to determine the first two statistical moments of the SRF \( h(x) \) and \( r(x) \). Each SRF can be decomposed in two terms: an expected value and a fluctuation around the expected value, \( E[h] \) and \( h' \), and \( E[r] \) and \( r' \) such that \( E[h] = h - h' \); \( E[h'] = 0 \) and \( E[r] = r - r' \); \( E[r'] = 0 \). Substituting these quantities in the above equation leads to:

\[ \frac{dE[h]}{dx} = -E[r] \quad q; \quad \text{thus} \]

\[ E[h] = -qE[r] \quad x \quad \text{constant} \quad (29) \]

\[ \frac{dh'}{dx} = -q \quad r' \quad (30) \]

Assuming that \( h \) and \( r \) are 2nd order stationary processes and using the Wiener–Kintchine theorem, we introduce the following two complex stochastic processes:

\[ h'(x) = \int_{-\infty}^{+\infty} e^{i\omega x} dZ_h(\omega); \]

\[ r'(x) = \int_{-\infty}^{+\infty} e^{i\omega x} dZ_r(\omega) \quad (31) \]

Substituting them in the perturbed equation, spectral density of \( h \) is given by

\[ \varphi_h(\omega) = E[dZ_h(\omega) dZ_h^*(\omega)] = \left( \frac{q}{\omega} \right)^2 E[dZ_r(\omega) dZ_r^*(\omega)] \]

\[ = \left( \frac{q}{\omega} \right)^2 \varphi_r(\omega) \quad (32) \]

which gives the spectrum of \( h \) given a spectrum of \( r \), and hence both moments are determined.

Perturbation Methods

Perturbation methods aim at rewriting a difficult mathematical problem into an infinite series of easy ones (terms). It is most useful when the first few terms of the series are pertinent to the solution of the original problem, the remaining of the series is assumed to be small. To illustrate this technique, let us apply it to the previous one-dimensional problem. First we assume that \( K \) is 2nd order stationary SRF with \( E[K] \) its expected value and \( k' \) its fluctuation around the mean. Similarly, we assume that \( h \) is 2nd order stationary SRF, and \( E[h] \) and \( h' \) are its mean and fluctuation, respectively. We seek a 1st order \( (\epsilon^1) \) solution to the flow equation, thus we assume that \( K \) and \( h \) are expanded in terms of small perturbations: \( K = E[K] + \epsilon k' \) and \( h = E[h] + \epsilon h' \). Substituting those into the flow equation and neglecting term of higher order \( (\epsilon^2) \) yield to (de Marsily, 1986):

\[ E[K] \frac{d^2 E[h]}{dx^2} + \epsilon \left( E[K] \frac{d^2 h'}{dx^2} + \frac{dk'}{dx} \frac{dE[h]}{dx} + k' \frac{d^2 E[h]}{dx^2} \right) \]

\[ = 0 \quad (33) \]

If this holds for any small \( \epsilon \), each term of these two must vanish. Therefore,

\[ E[K] \frac{d^2 E[h]}{dx^2} = 0; \quad \text{thus} \quad E[h] = \frac{q}{E[K]} x + \text{constant} \quad (34) \]

Substituting \( E[h] \) into the second term leads to

\[ \frac{dh'}{dx} = \frac{q}{E[K]^2} k' + \text{constant}; \quad \text{and} \quad \text{Cov} \left[ \frac{dh'}{dx} \right] = -\frac{E^2}{dx^2} \text{Cov}[h] = \left( \frac{q}{E[K]^2} \right)^2 \text{Cov}[K] \quad (35) \]

After integrating twice one can determine the Cov function for a stationary SRF:

\[ \text{Cov}[h(x + s), h(x)] = \left( \frac{q}{E[K]^2} \right)^2 \int_{-\infty}^{x+s} \int_{-\infty}^{y} \times \text{Cov}[K(x + w), K(x)] dw \ dy \quad (36) \]

Thus, the first two moments of \( h \) are defined given the moment of \( K \). It has been shown that the error involved in the method of perturbation compared to the spectral method is less than 10% if \( \sigma^2 < 1 \).
All analytical and numerical solutions of the SPDEs involve the use of either technique (Gelhar (1992), Dagan (1989), Rubin (2003)).

**Numerical Methods – Monte Carlo Solution**

Given the petrophysical properties of an aquifer, one can, using geostatistical methods, create multiple conditional simulations/realizations of the spatial variability of those petrophysical parameters. These simulations are equally probable and the difference between them is a measurement of the estimation uncertainty. Contrary to the perturbation and spectral methods, the MCS necessitates a prior knowledge of the probability distribution and the covariance function of the parameter, that is, \( K \). For each realization/simulation “\( i \)” of the field \( K_i \), the flow equation is solved numerically and for instance, the solution \( h_i \) is obtained. It is then possible to statistically analyze the ensemble of computed solutions: expected value, variance, pdf, and so on, for each location \( x \). Stationarity assumption is alleviated since statistics can be evaluated at every point of the field. Unfortunately, MCSs are time consuming for two main reasons. On the one hand, to capture heterogeneity, MCSs require solving numerically the flow and/or transport equation(s) in many detailed realizations of \( K(x) \). This can be computationally expensive, especially under transient conditions. Each Monte Carlo grid must capture the detailed heterogeneity of parameter fields. On the other hand, to get meaningful statistics, a large number of realizations/simulations is necessary. It is worth mentioning that the experimental variance is proportional to the inverse of the square root of the number of realizations, and therefore the rate of convergence is slow.

**An Application of SPDE to Effective Hydraulic Conductivity**

Effective properties are of great interest for hydrogeological applications and especially for numerical simulations. For example, regional-scale models of groundwater flow and transport often discretize the domain into grid blocks larger than typical integral scales of field data (see Section “Typical covariances and variograms”). This process is purely numerical and it is intended to alleviate the complexity of problem and thus the number of unknowns. For heterogeneous formations, the difference between the sale (size) of heterogeneity and the discretization block size is often handled using effective (upscaled) hydrogeological parameters. For example, in the case of flow in large 3D heterogeneous media domain, the flux is given by Darcy’s law: \(-K \nabla H = q\) and the effective hydraulic conductivity is defined by: \(\langle q \rangle = -K_{\text{eff}}(\nabla H)\), where the brackets denote the expected values, boldface denotes vectors. Second-order analytical solutions were developed by Dagan (see Dagan (1989)) using perturbation methods for axisymmetric anisotropy and by Gelhar and Axness (1983) (see Gelhar (1992)) using spectral methods for fully anisotropic media. The effective tensor of hydraulic conductivity is given by

\[
K_{\text{eff}, ij} = K_G \left( 1 + \frac{\sigma^2}{2} \right) \delta_{ij} - g_{ij},
\]

and

\[
g_{ij} = \frac{1}{\sigma^2} \int \frac{\omega_i \omega_j}{\omega^2} \phi_Y(\omega) \, d^3 \omega
\]  

(37)

where \( \delta_{ij} \) is the Kronecker tensor, \( K_G \) is the geometric mean of the hydraulic conductivity, \( \omega \) is the vector wave (Fourier space), and \( \phi_Y \) is the spectral form of the covariance of the log-hydraulic conductivity, \( Y \). More results are given in Dagan (1989) and Rubin (2003).

**SOLUTE TRANSPORT IN STATIONARY HETEROGENEOUS MEDIA**

A good understanding of the physical processes governing the transport of contaminants in the subsurface is crucial to decision makers and risk assessors. In the simplest case of nonreactive contaminants, the transport processes encompasses advection by regional flow and mixing or dispersion that is evident as the spreading of contaminant front in the subsurface. Early studies of the dispersion of the contaminants showed dispersion rates in the field that were much larger than the laboratory measured ones. The disparity between the laboratory and the field scale dispersivity is attributed to the velocity variation induced by heterogeneity in the hydraulic properties of the subsurface. This phenomenon is known as macrodispersion. The basis for calculating flow velocities for fluids in a heterogeneous porous medium is Darcy’s law, which relates the specific discharge \( q \) with hydraulic conductivity \( K \) and the hydraulic gradient (\( \nabla H \)). Owing to the strong spatial variation of \( K \) and \( \nabla H \), which cannot be known in detail, stochastic approaches have been developed. Stochastic transport theories attempt to develop relationships between macrodispersivities and statistical parameters characterizing the hydraulic conductivity spatial variability. To develop such relationships, we need to quantify the statistical structure of the velocity variations resulting from spatial variation of the hydraulic conductivity. The resulting statistical structure of the velocity variations is then used for quantifying macrodispersivities.

**Macrodispersion Coefficient**

Characterization of the macrodispersion coefficient in terms of the specific discharge spectrum can be approached in two different frameworks. On the one hand, the Eulerian
approach: the solute transport equation serves as the basis for a perturbation approximation to the equation for the ensemble-average concentration field, Gelhar (1992). On the other hand, the Lagrangian approach that relies on the relationship between the dispersion of a passive tracer and the mean square displacement of a particle moving in a random velocity field, originally founded by Taylor (1921) and applied to subsurface hydrology by Matheron and de Marsily (1980) and Dagan (1989). Despite the differences, the results obtained by Gelhar (1992) are identical to those obtained by Dagan (1989). For brevity, we only summarize the main results; more details can be found in the monographs of the mentioned authors. Here we limit ourselves to the Lagrangian approach. Following Dagan (1989), let us assume that at \( t = 0 \) a body of tracer concentration \( C_0 \) is introduced in a fluid and spread subsequently due to both molecular diffusion and convection by the fluid. Taylor suggested that the solute body is regarded as a collection of invisible particles; the concentration is regarded as the relative number per unit volume. For a single particle the motion is described by the displacement:

\[
X_i(t) = \int_0^t v_i(t') \, dt'
\]

where \( v \) is the Lagrangian velocity which is related to the Eulerian one through: \( v(t) = u(X(t)) \). The velocity is a random variable and is the sum of the two uncorrelated velocities: (i) a velocity due to the molecular diffusion and (ii) a fluid velocity (advective). The velocity vector is assumed to be a 2nd order stationary random process in time with the following statistics: \( E[v_i(t)] \) and \( E[(v_i(t) - E[v_i(t)]) (v_j(t) - E[v_j(t)])] = \text{Cov}[v_i, v_j](t - \tau) \). Thus the mean velocity is \( E[dx_i/dt] = E[v_i(t)] \), and the fluctuation of the particle around the mean is \( dX_i/dt = v_i(t) \). Therefore, the covariance of two Cartesian components of \( X' \) is given by

\[
E[X'X'] = \int_0^t \int_0^t E[v_i(t')v_j(t'')] \, dt' \, dt'';
\]

\( \forall i, j = 1, 2, 3 \) \( \quad \) (39)

The relationship between the dispersion coefficient and the Lagrangian velocity covariance is given by (Dagan, 1989):

\[
D_{ij}(t) = \frac{1}{2} \frac{dE[X'X']}{dt} = \int_0^t \left[ \text{Cov}[v_i, v_j](t - \tau) \right] \, d\tau; \quad \forall i, j = 1, 2, 3 \quad (40)
\]

We present here a few solutions of the displacement tensor, \( X_{ij} \).

**Case 1** – 2D zero pore-scale dispersion: planar flow in the horizontal plane, thin aquifer with uniform mean head gradient, exponential covariance for \( Y \), the dimensionless time is given by \( t' = tU_1/I_Y \). The solution is given by

\[
X_{11}(t') = \frac{2}{\sigma_Y^2 I_Y^2} \left[ 2t' - 3 \ln(t') + \frac{3}{2} - 3E \right]
+ \frac{3}{\sigma_Y^2 I_Y^2} \left[ E(-t') + e^{-t'} \right] \left[ (1 + t') - \frac{1}{t^2} \right]
\]

\[
X_{22}(t') = \ln(t') - \frac{3}{2} + E(-t') + \frac{3}{\sigma_Y^2 I_Y^2} \left[ 1 - (1 + t')e^{-t'} \right]
\]

where \( E \) is the Euler constant.

**Case 2** – 3D isotropic case, zero pore-scale dispersion. The longitudinal and transversal displacement tensor are given by

\[
X_{11}(t') = \frac{2t' - 2}{\sigma_Y^2 I_Y^2} \left[ \frac{4}{3} - \frac{4}{t^2} + \frac{8}{t^3} - \frac{8}{t^4} \left(1 + \frac{1}{t} \right) e^{-t'} \right]
\]

\[
X_{22}(t') = \frac{X_{33}(t')}{\sigma_Y^2 I_Y^2}
= 2 \left[ -\frac{1}{3} - \frac{1}{t^2} + \frac{4}{t^3} - \left( \frac{4}{t^3} + \frac{4}{t^2} + \frac{1}{t} \right) e^{-t'} \right]
\]

**Case 3** – 3D anisotropic case, zero pore-scale dispersion. A more general flow case with an exponential axisymmetric covariance for \( Y \), where \( e \) is the anisotropy ratio defined as the ratio between the vertical and the horizontal length scales, \( e = I_v/I_h \). The displacements are given by (Figure 6):

\[
X_{11}(t') = \frac{2}{\sigma_Y^2 I_Y^2} \left[ 2t' + 2(e^{-t'} - 1) + 8e \int_0^\infty \frac{J_0(kt') - 1}{(1 + k^2 - (ek)^2)^2} \right]
\]

\[
\left[ 1 - \frac{ek}{\sqrt{1 + k^2}} - \frac{ek(1 + k^2 - (ek)^2)}{(1 + k^2)^3/2} \right] \, dk
\]

\[
- 2e \int_0^\infty \left\{ J_0(kt') - \frac{J_1(kt')}{kt'} - \frac{1}{2} \right\} \, dk
\]

\[
\left[ \frac{(ek)^3((ek)^2 - 5 - 5k^2)}{(1 + k^2 - (ek)^2)^3(1 + k^2)^3/2} + \frac{1 + k^2 - 5(ek)^2}{(1 + k^2 - (ek)^2)^3} \right] \, dk
\]

\[
X_{22}(t') = \frac{X_{33}(t')}{\sigma_Y^2 I_Y^2}
= -2e \int_0^\infty \left[ \frac{J_1(kt')}{kt'} - \frac{k}{2} \right] \, dk
\]

\[
\left[ \frac{e(ek)^2((ek)^2 - 5 - 5k^2)}{(1 + k^2 - (ek)^2)^3(1 + k^2)^3/2} + \frac{1 + k^2 - 5(ek)^2}{k(1 + k^2 - (ek)^2)^2} \right] \, dk
\]

(41)
contaminant displacements in saturated and unsaturated zones are covered in Zhang (2001).

**Statistical Spatial Moments**

A contaminant plume is generally described using the statistical spatial moments. The \( p \)th \( \leq 1 \) moment of the concentration distribution in space, \( M_{p_1 \ldots p_n}(t) \), is defined here following Aris (1956) by

\[
M_{p_1 \ldots p_n}(t) = \eta \int_{\Omega} C(x, t) \prod_{i} x_{i}^{p_i} \, d\Omega
\]  

(48)

In the next section, we present how to estimate the displacement tensor from concentration measurements using the concept of statistical spatial moments.
where $C(x, t)$ is the concentration at point $x = (x_1, \ldots, x_n)$, and at time $t$. $n$ denotes the space dimension and can be equal to 2 or 3, depending on the type of analysis being performed; $\eta$ is the effective porosity; $\Omega$ denotes the aquifer domain occupied by the plume, and the order of the moment is given by $p = \Sigma p_i$. Here, we shall deal only with the zeroth, first, and second spatial moments of the plume. The 0th moment measures the mass of the plume, while the 1st moment provides a measure of the displacement of the centroid of the plume $\bar{X}(t) = (\bar{x}_1, \ldots, \bar{x}_n)$. In three-dimensional space, the centroid of the plume is given by:

$$\bar{x}_1 = \frac{M_{100}}{M_0}, \quad \bar{x}_2 = \frac{M_{010}}{M_0}, \quad \bar{x}_3 = \frac{M_{001}}{M_0} \quad (49)$$

For the case where $p \geq 2$ the central moments of the plume will be used, which are given by

$$M_{p1\ldots n}(t) = \eta \int_{\Omega} C(x, t) \prod_{i} (x_i^{p_i} - \bar{x}_{i,c})^{p_i} \, d\Omega \quad (50)$$

The 2nd moments provide a measure of the spread of the plume around its centroid. The terms of the displacement covariance tensor $\tilde{X}_{ij}(t)$ are computed using the following expressions:

$$\tilde{X}_{11} = \frac{M_{200}}{M_0}, \quad \tilde{X}_{22} = \frac{M_{020}}{M_0}, \quad \tilde{X}_{33} = \frac{M_{002}}{M_0},$$
$$\tilde{X}_{12} = \frac{M_{110}}{M_0}, \quad \tilde{X}_{13} = \frac{M_{101}}{M_0}, \quad \tilde{X}_{23} = \frac{M_{011}}{M_0} \quad (51)$$

In applications, $\Omega$ is not known exactly, and in many cases it is only partly covered by samplers. Hence, prior to computing the moments, there is a need to estimate $\Omega$. This problem is referred to as the “null points” problem, and it has been approached in previous works using methods such as linear extrapolation or surface fitting using spline functions or kriging. A concise presentation of these methods is presented in Chapter Chapter 147, Characterization of Porous and Fractured Media, Volume 4.

We illustrate the application of the spatial statistical moments on the large-scale field experiment conducted at Cape Cod (Leblanc et al., 1991) The Cape Cod large-scale natural gradient tracer test began in July 1985 with the injection of 7.6 m$^3$ of tracer solution in the aquifer through a volume of dimensions of $1.2 \times 3 \times 4$ m$^3$. The tracer solution contained a nonreactive tracer, bromide, and reactive tracers, lithium and molybdate. For illustration purposes, we shall concentrate on the displacement of the bromide plume. The plume was sampled over a period of 511 days. In each of the 16 sampling sessions, concentration was measured using a large and dense three-dimensional array of samplers (see Figure 7).

Zeroth-, first-, and second-order moments were computed and given in Figures 8, 9(a, b), and 10(a–c) (Rubin and Ezzedine, 1997).
Ergodicity and Plume Size Effect on Displacement Tensor

A solute plume is ergodic if the mean and covariance of the displacement along the various streamlines are equal to the mean and covariance of the displacement of a single particle. In this case, the ensemble-average concentration plume dispersivity is expected to accurately predict the plume growth using single-particle displacement statistics only when the plume size is much larger than the integral scales of the hydraulic conductivity (Dagan, 1989). It is therefore important to investigate the applicability and limitation of the ergodic hypothesis. On the basis of the concept of relative dispersion, Dagan (1989) reached the following relationship:

\[ E[S_{ij}(t)] = S_{ij}(t = 0) + X_{ij}(t) - R_{ij}(t) \]  \hspace{1cm} (52)

where \( S_{ij}(t) \) is the second spatial moment of the plume of a plume of arbitrary dimension as a random function of trajectories of the particles:

\[ S_{ij}(t) = \frac{1}{|\Omega|} \int [X_i(t|\Omega) - R_i(t)][X_j(t|\Omega) - R_j(t)] \, d\Omega; \]

\[ R_i(t) = \frac{1}{|\Omega|} \int X_i(t|\Omega) \, d\Omega \]  \hspace{1cm} (53)

where \( R(t) \) is plume centroid. Because \( X_i \) is a random function, \( R \) is inherently so. \( X_i(t|\Omega) \) is the trajectory of a particle emanating from a source of size \( \Omega \) with and a volume \( |\Omega| \). The variance of the centroid of the plume is given by

\[ R_{ij}(t) = E[(R_i(t) - E[R_i(t)])(R_j(t) - E[R_j(t)])] \]

\[ = \frac{1}{|\Omega|^2} \int \int E[X'_i(t|\Omega)X'_j(t|\Omega')] \, d\Omega \, d\Omega' \]  \hspace{1cm} (54)

If the plume spreads are much larger than the integral scale of \( Y \), the statistical variability of \( Y \) becomes narrow and \( R \) approaches the ensemble mean displacement. An elaborate discussion of ergodicity and operational ergodicity is given in Rubin (2003, Chapter 10). Ergodicity analysis was performed on the Cape Cod data and results are given in Figures 11(a–c).

Spatial moment concepts could be extended to temporal statistical moments. They are introduced in the following section.

Statistical Temporal Moments

The Lagrangian framework is most suitable for modeling and analyzing travel times. One compelling reason is the lack of discrepancy between the measurement scale in the field and the scale of numerical element used for modeling. Samplers in the field measure the concentration at a streamline, and the streamline is also the fundamental unit in the Lagrangian approach. However, the elegance of the Lagrangian-based solutions can sometimes be lost when confronted by the reality of field data analysis. To illustrate this matter, we proceed by presenting temporal moments statistics. Similar to spatial moments, the temporal moments are commonly defined by the following integrals:

\[ \mu_i(x) = \int_{-\infty}^{+\infty} t^i C(x,t) \, dt \]  \hspace{1cm} (55)

where \( \mu_i(x) \) are the \( i \)th temporal noncentral moments. To demonstrate these concepts, let us consider the moments of
Figure 10  (a) The longitudinal central second moment $X_{11}$ as computed in the present study and by Garabedian et al. (1991), (b) The lateral second central moment $X_{22}$ as computed in the present study and by Garabedian et al. (1991), (c) The vertical second moment $X_{33}$ as computed in the present study and by Garabedian et al. (1991) (Reproduced from Rubin and Ezzedine (1997) by permission of American Geophysical Union)

Figure 11  The 3D conditional $S_{33}$ computed using (52), the conditional $X_{33}$, experimental results of vertical spread, and 3D pore scale dispersion ($P_e = 70$, Fiori, 1996). b) the 2D and 3D $S_{11}$ computed using (52), unconditional $X_{11}$, 2D unconditional higher-order approximation $X_{11}[1,2]$ (Hsu et al., 1996) and 3D $X_{11}$ with pore scale dispersion ($P_e = 70$, Fiori, 1996). c) the 2D and 3D $S_{22}$ computed using (52), unconditional $X_{22}$, 2D unconditional higher-order approximation $X_{22}[1,2]$ (Hsu et al., 1996) and 3D $X_{22}$ with pore scale dispersion ($P_e = 70$, Fiori, 1996) (Reproduced from Rubin and Ezzedine (1997) by permission of American Geophysical Union)
the travel time at \( s(L) \) where \( L \) is a distance downstream from the injection source located at \( x_0 \), and \( s \) is a streamline coordinate system. Moreover, we assume that the tracer is injected at \( x_0 \) over a short period of time \( T_0 \), such that the following boundary condition for the advection dispersion equation holds: \( C(x_0, t) = C_0 \) for \( t \in [-T_0/2, T_0/2] \) and \( C(x_0, t) = 0 \) otherwise. Neglecting pore-scale dispersion, the concentration distribution along a streamline is given by the square wave

\[
C(t, \tau) = C_0 \left[ H \left( t - \tau + \frac{T_0}{2} \right) - H \left( t - \tau - \frac{T_0}{2} \right) \right]
\]

(56)

where \( H \) is the Heaviside step function, and \( \tau \), is the time that it takes for a particle emanating from \( x_0 \) to reach \( s(L) \) and is given by

\[
\tau(L) = \int_{s(x_0)}^{s(L)} \frac{ds}{V(s, x_0)}
\]

(57)

where \( V \) denotes velocity. Applying the travel time to the square wave solution leads to the first three travel time moments:

\[
\mu_0 = C_0 T_0, \quad \mu_1 = C_0 T_0 \tau,
\]

and \( \mu_2 = C_0 T_0 \left( \tau^2 + \frac{T_0^2}{12} \right) \)

and thus \( \tau = \frac{\mu_1}{\mu_0} \) and \( \tau^2 = \frac{\mu_2}{\mu_0} - \frac{T_0^2}{12} \)

(58)

The spatial average of \( \tau \) over a control plane of area \( A \) perpendicular to the mean flow direction at \( L \) distance from the source is given by

\[
\bar{\tau}(L) = \frac{1}{A} \int_A \frac{\mu_1(x_0 + L, y, z)}{\mu_0(x_0 + L, y, z)} dy \ dz;
\]

and \( \sigma_\tau^2(L) \equiv (\tau^2) - (\bar{\tau})^2 = \frac{1}{A} \int_A \frac{\mu_2(x_0 + L, y, z)}{\mu_0(x_0 + L, y, z)} dy \ dz
\]

\[= \frac{T_0^2}{12} - (\bar{\tau})^2 \]

(59)

In the Lagrangian approach, mass transfer across streamlines cannot be modeled, yet this aspect of the transport problem must be dealt with in order to make the travel time analysis wieldy. Hereafter, an approach is presented that will allow one to retain the simplicity of the Lagrangian approach yet account realistically for the effects of pore-scale dispersion.

Inspection of the time records of concentration at various samplers reveals that the above theoretical development is somewhat limited (Figure 12). Rather than a single square pulse, or even a diffused, Gaussian-like pulse, we observe trains of signals of different modes, stretching over periods of time much longer than the injection period.

Since the Cape Cod plume was injected over a period of 18 h, and the time-concentration records show pulses that take much longer to travel across individual samplers, it becomes clear that computing the travel time moments based on the entire time-concentration records is the culprit for the mismatch between the theoretical and experimental travel time variances (Figure 13).

To use the time records for inference, the effects of pore-scale dispersion need to be removed nonarbitrarily, and the geometry of the advective pulse reconstructed. To do so, it is suggested that the travel time \( t \) to each sampler be taken as the time at which the maximum concentration \( C_{\text{max}} \) is detected. We shall refer below to that time as \( t_{c_{\text{max}}} \). The parameter \( t_{c_{\text{max}}} \) is taken as the most likely estimator of \( \tau \), instead of using the Aris estimator. The justification for choosing the peak arrival time as the travel time least affected by pore-scale dispersion is based on the observation that this travel time is associated with the plume centroid, where the concentration is generally the largest and the concentration gradients equal to zero. Additionally, if one considers an ensemble of streamlines with a solute pulse moving in any of them, possibly with different velocities, it can be expected that the concentration that develops in any streamline due to lateral mass transfer will be smaller than that of the pulse that was originally injected, at least over a long time, until the concentration variations smooth out. The first two moments of the travel time are now given by

\[
\bar{\tau}(x) = \frac{1}{A} \int_A t_{c_{\text{max}}}(x, y, z) \ dy \ dz;
\]

\[
\sigma_\tau^2(x) = \frac{1}{A} \int_A [t_{c_{\text{max}}}(x, y, z) - \bar{\tau}(x)]^2 \ dy \ dz
\]

(60)

The assumption underlying both moments is that \( C_{\text{max}} \) is the concentration of the pulse’s centroid and that by tracking the displacement of the pulse’s centroid, it is possible to capture the effects of advection while removing the effects of pore-scale dispersion. The viability of the suggested approach predicates upon the existence of a distinguishable \( C_{\text{max}} \). The time \( t_{c_{\text{max}}} \) will become nondistinguishable as the concentration field smoothes out with time due to the effects of pore-scale dispersion.

Figure 14(a) depicts the mean travel time as a function of distance from the source. The results obtained by averaging \( t_{c_{\text{max}}} \) over the planes normal to the mean flow direction are very similar to those obtained using Aris moments of order 0, and 1, for both analyses yield results which are close to those obtained using the average velocity. These results suggest that removing the effects of pore-scale dispersion is inconsequential for first moments analysis.
and that the large distribution of the pulses over the flow domain cancels out the effects of the low-level concentrations due to averaging. The $\langle \tau \rangle$ model following Rubin and Dagan (1992) is based on a transformation of the single-particle displacement variance, and it shows excellent agreement with the observed data. In analyzing the travel time variances, the situation becomes quite different. Figure 14(b) shows the travel time variance computed using the ensemble of the experimental $t_{c_{\text{max}}}$ over consecutive planes of multilevel samplers. The first and second moments were computed over consecutive planes and compared with the Aris-based experimental results. This comparison shows a dramatic reduction in the magnitude of the experimental variances. A comparison of the experimental results with theoretical results shows a good match.

**INVERSE PROBLEMS AS STATISTICS**

Much of the problem and initial cost of subsurface remediation concerns field site characterization. A full three-dimensional “picture” of the heterogeneous subsurface is needed in order to identify the key controls on the flow and contaminant transport processes. Natural heterogeneity and the large spatial variability of the permeability predominantly control the flow field and hence the transport. Moreover, natural heterogeneity exhibits variability over a wide range of scales and hence is difficult to characterize due to the scarcity of data and the costliness of

---

**Figure 12** Examples of the concentration versus time records at several samplers. Typically, a high-level detect is followed or preceded by a long train of low-level detects (Reproduced from Rubin and Ezzedine (1997) by permission of American Geophysical Union)

**Figure 13** Theoretical and experimental travel time variance (Reproduced from Rubin and Ezzedine (1997) by permission of American Geophysical Union)
conventional field sampling techniques such as drilling. With poor site characterization, remediation schemes are unnecessarily expensive, because costly overdesign may be required to compensate for uncertainty. Therefore, inverse problems were formulated in order to estimate the hydraulic properties from a sparse data of measurements. This section describes how to determine model parameter values. Models are assumed to be valid; the only unknowns are parameter values that define the models. For completeness, we introduce some concepts and terminology commonly used in inverse/forward problem community. An elaborate discussion of the deterministic approach to inverse problems is given in Chapter 155, Numerical Models of Groundwater Flow and Transport, Volume 4.

Well-posed versus Ill-posed Problem

Prediction based on a given set of parameter values is called “forward modeling.” Determination of parameter value from observed data is called “inverse modeling.” Inversion can be achieved in two ways. On one hand, a modeler iteratively modifies parameter values (such as hydraulic conductivity), run a forward model until attaining best “fit” or “match” between the measurements and the modeled variable, head for instance. This kind of process falls into the trial and error methods. Such forward modeling is sometimes tedious and time consuming. On the other hand, an inverse algorithm is adopted to automatically or semiautomatically obtain the parameter values from the observed data and an initial set of trial parameters values. The procedure also provides an estimate of parameter uncertainty. In both cases, inversion requires minimizing the discrepancy between predictions and observations (Sun, 1994).

A well-posed inverse problem requires “existence” of the problem, the “uniqueness” and the “stability” of the solution or algorithm. Obviously, in view of observed data and our understanding of a real-world physical system, a problem is presumed to exist, for example, detection of contaminant plume in groundwater suggests contamination must have happened in the past. The question then is how to relate the observation to the migration history of the contaminant. A cause generally has an effect. Can an effect result from different causes? Is it unique in theory or model? Even if it is, have we counted and resolved all parameters that define a model? Uniqueness has two levels: the model itself and the model-defining parameters. The latter is related to the stability of a solution algorithm. How sensitive are parameters to uncertainty of observed data? Are the errors amplified during inversion? Is the inversion algorithm efficient in terms of ease of usage and cost of running the inversion program (complexity)? Mathematically speaking, inverse problems are ill-posed problems.

Deterministic versus Stochastic Inversion

To address the inverse problem, two main frameworks have been developed and they are either deterministic or stochastic. In the deterministic framework, the structure of the spatial variability of the parameters is fixed. For example the aquifer is divided into a number of zones, and each zone is supposed to have a constant hydraulic conductivity; then, the algorithm seeks the best hydraulic conductivity values, for which the solution of the flow equation reproduces the observed data. In a stochastic framework, however, the spatial variability of the parameters is statistically mapped. For example, the overall mean and variance, and the variogram of the final realization are specified. This characterization is not enough to fully determine the parameter values at every cell. Thus, a spatial realization is sought, satisfying the statistical constraints and honoring the observed values. As it has been shown in Section “Geostatistics”, an infinite
number of realizations can meet the statistical constraints and reproduce the observed data. To make these inferences quantitative in either deterministic or stochastic framework, one must answer three fundamental questions:

- How accurately is the data known, that is, what does it mean to “fit” the data?
- How accurately can we model the response of the aquifer system? In other words, have we included all the physics in the model that contribute significantly to the data?
- Finally, what is known about the system besides the data? This is called *a priori* information and is essential since for any sufficiently fine parameterization of an aquifer system there will be unreasonable models that fit the data too. Prior information is the means by which we reject or down-weight unreasonable estimation models.

There are a variety of “recipes” for constructing estimators; perhaps the most common in statistics are, minimum variance estimation, maximum likelihood estimation (MLE) and Bayesian estimation. The latter have asymptotically optimal properties under certain restrictive assumptions. Hereafter we present three inverse problems; they are introduced in order of complexity.

**Illustrations and Examples**

**A Simple Statistical Inference Problem: Least-square Method**

To illustrate the deterministic inversion process, let us consider a classic example of parameters estimation for a pumping test (readers are referred to Chapter 151, Hydraulics of Wells and Well Testing, Volume 4 for more elaborate details). The solution of a pumping test in a confined aquifer is given by Jacob’s semilog approximation for large time:

$$
\delta h \approx \frac{Q}{4\pi T} \left( -0.5772 - \ln \left[ \frac{r^2 S}{4T} \right] + \ln \left[ t \right] \right)
$$

(61)

where $T$ is the transmissivity, $S$ is the storativity, $Q$ is the pumping rate, $t$ is time, $h$ is the hydraulic head and $\delta h$ is the drawdown. The inverse problem is to determine $T$ and $S$ given a series of error-free observations of $\delta h$ at different times. The relation (model) between the drawdown and time is not linear but logarithmic. However, if $\ln \left[ t \right]$ is treated as an independent variable the relation between $\delta h$ and $\ln \left[ t \right]$ is a linear one and the inversion in the least-square sense is straightforward. The sum of the square of discrepancy $\varepsilon_i$ between each predicted $\delta h_i$ and measured $\delta h_i^{\text{obs}}$ is minimized in least-square sense. If the above equation is rewritten as follows:

$$
y = ax + b; \quad x \equiv \ln \left[ t \right]; \quad a = \frac{Q}{4\pi T}; \quad b = \frac{Q}{4\pi T} \left( -0.5772 - \ln \left[ \frac{r^2 S}{4T} \right] \right); \quad y \equiv \delta h
$$

(62)

The goal is then to minimize the sum or the square of errors (minimum variance):

$$
\text{Min } \delta^2 = \text{Min } \left[ \sum_{i=1}^{n} \delta^2_i \right] = \text{Min } \left[ \sum_{i=1}^{n} \left( y_i - y_i^{\text{obs}} \right)^2 \right]
$$

(63)

The solution of the problem is readily available in textbooks and it is given by:

$$
a = \frac{n \sum_{i=1}^{n} x_i y_i^{\text{obs}} - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i^{\text{obs}}}{n \sum_{i=1}^{n} x_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2}; \\
b = \frac{n \sum_{i=1}^{n} x_i^2 y_i^{\text{obs}} - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i^{\text{obs}}}{n \sum_{i=1}^{n} x_i^2 - \left( \sum_{i=1}^{n} x_i \right)^2}
$$

(64)

and $a$ and $b$ can be expressed in univariate statistics as shown on Table 1.

The expected transmissivity value is then determined by $E[T \ln(t)] = a \ln(t) + b$.

**Cokriging Methods**

It has already been pointed out that the final hydraulic property map (realization) cannot disregard the measurement data; they are the only factual knowledge available about the aquifer. However, aquifers are systems, the state of which is described by the spatial distribution of piezometric heads, and by the concentration of the solutes dissolved in water. In general, there is more information about the state of the system than about the parameters that controls it. Therefore, it appears necessary to generate spatial distributions of the parameters that are not only conditional to parameter values, but also consistent with the (partial) knowledge about the state of the system. This matter has already been introduced in the Section “Geostatistics”. We present here two examples for completeness.

**Cokriging Head and Hydraulic Conductivity**

Rubin (2003) used the analytical approach to solve the perturbed flow equation. Rubin calculated $h' = h - E[h]$ and $Y' = Y - E[Y]$ at the points of $h$, head, and $Y$, log of
transmissivity; and determined analytically the covariance function of \( h' \) and the cross-covariance \( (h', Y') \) as a function of the covariance of \( Y \). The covariance of \( Y \) is function of a set of parameters \( \theta \) (i.e. integral scale of \( Y \) and its variance). This is actually sufficient to estimate the transmissivity field by cokriging. The cokriging estimator then gives the optimal estimation of \( Y \) at any point as follows:

\[
Y(x) = \sum_{i=1}^{n_Y} \lambda_i Y_i + \sum_{j=1}^{n_H} \nu_j (h_j - E[h_j]) \tag{65}
\]

where the \( \lambda_i \) and the \( \nu_j \) are optimal weights, that depend on the position \( x \). The cokriging equations that provide the value of the optimal weights simply require that the covariance functions of \( Y \), \( h \) and of \( h - Y \) be known. Rubin then calculate by cokriging all the values of \( Y \) at the measurement points where \( Y \) is known and where it is therefore possible to compare the known value with the one estimated by cokriging – without using the known value of this point in the cokriging equations. As the cokriging estimator is a function of the \( \theta \) parameters, these parameters can thus be optimized to minimize the errors between the estimated and measured \( Y \) values. The Maximum Likelihood method was used for their optimization (see Chapter 156, Inverse Methods for Parameter Estimations, Volume 4). Once the \( \theta \) parameters are known, the cokriging equations give an estimation of \( Y \) at all points and a map of \( Y \) is obtained (Figure 15).

**Cokriging Head and Hydraulic Conductivity for Concentration Estimations**

The goal is to condition the first two moments of the concentration on measurements of conductivity \( K \) and hydraulic head \( H \). In order to model their spatial variability, \( Y \) and \( H \) will be modeled here as space random functions, so that \( Y(x) = m_Y + Y'(x) \), where \( m_Y = (Y(x)) \), and \( H(x) = (H(x)) + H'(x) \). Hence \( Y' \) and \( H' \) represent the fluctuations of \( Y \) and \( H \) about their respective means. The hydraulic head is assumed to be at steady state. Small variability of the log-conductivity is assumed, and the linearized steady state flow equation is used to relate between \( Y' \) and \( H' \) and other hydrogeological variables. Under these conditions, \( H' \) becomes a linear function of \( Y' \), and their cross-covariance can be derived analytically. Similarly, the concentration \( C \) is also viewed as a space random function and we define \( C(x) = (C(x)) + C'(x) \), where \( (C(x)) \) is the unconditional mean of \( C(x) \).

The goal can now be stated as follows: define \( (C^c(x, \tau)) \) and \( \sigma^2_{C^c} \), the conditional mean of the concentration and its estimation variance, where conditioning is done over \( N \) measurements, \( M \) of which are log-conductivities, and the rest are head measurements. These estimates should also reflect the dependence of \( C \) on time, \( \tau \). A superscript “c” denotes estimators, which are conditional to measurements. This task is accomplished using the following relationships for the conditional mean:

\[
(C^c(x, \tau)) = (C(x, \tau)) + \sum_{i=1}^{M} \lambda_i(x, \tau) Y'(x_i) + \sum_{j=M+1}^{N} \mu_j(x, \tau) H'(x_j) \tag{66}
\]

**Figure 15** Estimated transmissivity and its variance (Reproduced from Rubin and Dagan, 1987 by permission of American Geophysical Union)
and the conditional estimation variance:

\[
\sigma^2_{c}(x, \tau) = \sigma^2_{c}(x, \tau) - \sum_{i=1}^{M} \lambda_{i}(x, \tau) C_{CY}(x, \tau; x_{i}) - \sum_{j=M+1}^{N} \mu_{j}(x, \tau) C_{CH}(x, \tau; x_{j})
\]

(67)

where \(\sigma^2_{c}(x, \tau)\) is the unconditional variance of the concentration. The above equations constitute the well-known geostatistical formalism of cokriging. Hence, the estimates of \(C\) and \(H\) are the solutions of the following linear system (Journel and Huijbregts, 1978):

\[
C_{CY}(x, \tau, x_{k}) = \sum_{i=1}^{M} \lambda_{i}(x, \tau) C_{Y}(x_{i}, x_{k}) + \sum_{j=M+1}^{N} \mu_{j}(x, \tau) C_{YH}(x_{j}, x_{k}),
\]

\(k = 1, \ldots, M\)  \hspace{1cm} (68)

\[
C_{CH}(x, \tau, x_{k}) = \sum_{i=1}^{M} \lambda_{i}(x, \tau) C_{YH}(x_{i}, x_{k}) + \sum_{j=M+1}^{N} \mu_{j}(x, \tau) C_{H}(x_{j}, x_{k}),
\]

\(k = M + 1, \ldots, N\)  \hspace{1cm} (69)

In the above equations the following definitions are used to denote spatial moments: \(C_{Y}(x, x') = \langle Y(x')Y'(x') \rangle\), \(C_{YH}(x, x') = \langle H'(x)H'(x') \rangle\), \(C_{YH}(x, x') = \langle Y(x)H'(x') \rangle\), \(C_{CH}(x, \tau, x') = \langle (x, \tau)H'(x') \rangle\) and \(C_{CY}(x, \tau; x') = \langle (x, \tau)Y'(x') \rangle\). Given a set of \(N\) measurements, one can map the conditional expected concentration and its variance. \(C_{CY}\) and \(C_{CH}\) are depicted on Figure 16(a,b), respectively (Ezzedine and Rubin, 1996). Once these cross-covariances are analytically determined, one can proceed by mapping \(C\) conditional to \(Y\) and \(H\) measurements.

**Cokriging Using Displacement and Travel Time for Concentration Estimations**  
As part of the previous developments, the cross-covariances between \(Y\) and \(X(\tau)\) were obtained. That suggests the possibility of using the time-displacement data to update the estimates of \(Y\). Measured displacement can be coupled or perhaps used as an alternative to concentration measurements.

In fact, the use of displacements and travel times may be a better idea than using concentration data. For once, the concentration of a pulse is affected by pore-scale dispersion, while the travel times and displacements of the centroid of the pulse are affected to a much lesser extent. The concentration cross-covariances \(C_{CY}\) and \(C_{CH}\) were developed under the limiting assumption of negligible pore-scale dispersion, and by using travel times and displacements, we can avoid this limitation altogether.

In order to demonstrate the benefits of using travel time and displacements versus concentration, consider the following admittedly simplistic case. Solute of concentration \(C_{0}\) is injected over a very short period. The displacement of the pulse is monitored. Let us assume that at time \(\tau\) the pulse was located at \(\alpha = (\alpha_{1}, \alpha_{2}, \alpha_{3})\). In the absence of pore-scale dispersion, the concentration is equal to the initial one, that is, \(C_{0}\). It is clear that

![Figure 16](image-url)

**Figure 16**  
(a) Cross-covariances \(\tilde{C}_{CY}(x_{C}, \tau; x_{Y})\) for fixed \(x_{Y} = (3, 0, 0)\), \(x_{C} = (x_{C}, 0, 0)\) and \(\tau\) are variables, and \(r = x_{C} - 3\)  
(b) Cross-covariances \(\tilde{C}_{CH}(x_{C}, \tau; x_{H})\) for fixed \(x_{H} = (3, 0, 0)\), \(x_{C} = (x_{C}, 0, 0)\) and \(\tau\) are variables, and \(r = x_{C} - 3\) (Reproduced from (Ezzedine and Rubin, 1996) by permission of American Geophysical Union)
the magnitude of the displacement \( X(\tau) = \alpha \) as well as the measurement \( C(\alpha, \tau) = C_0 \) are equivalent in term of the information content. Next we wish to compute the conditional mean and conditional variance of the pulse’s longitudinal displacement. These moments can be related to the moments of the concentration (Ezzedine and Rubin, 1996). Conditioning on the concentration through kriging yields to \( (X_1') = (X(\tau)) + \lambda (C_0 - \langle C(\alpha, \tau) \rangle) \) and the conditional variance \( X_1' \) is \( X_1 + \lambda \langle C'(\alpha, \tau) X_1' \rangle \) with \( \lambda = \frac{\langle C'(\alpha, \tau) X_1' \rangle}{\sigma^2(\alpha, \tau)} \). The cross-covariance \( \langle C'(\alpha, \tau) X_1' \rangle \) is given by

\[
\langle C'(\alpha, \tau) X_1' \rangle = C_0(\alpha_1 - \langle X_1(\tau) \rangle)
\]

we used \( \text{Prob}[X(\tau) = \alpha] = \langle C \rangle/C_0 \). Recalling that \( \sigma^2(\alpha, \tau) = \langle C \rangle(C_0 - \langle C \rangle) \) (Dagan, 1989), the conditional displacement is reduced to \( (X_1') = \alpha_1 \) and \( X_1' = X_1(\tau) - (\alpha_1 - \langle X_1(\tau) \rangle)^2(C_0 - \langle C \rangle)/\langle C \rangle \). In the particular case of \( \alpha_1 = \langle X_1(\tau) \rangle \) we get \( (X_1') = \langle X_1(\tau) \rangle \) and \( X_1' = X_1(\tau) \) which are the unconditional moments, that is, no reduction at all in the displacement variance is achieved: the result for the first moment is clearly right while the result for the variance is unsatisfactory because it does not show any benefit from the measurement. But this result is not wrong: it is a direct outcome from the \( \langle C'(\alpha, \tau) X_1' \rangle \) cross-covariance being zero for \( \alpha_1 = \langle X_1(\tau) \rangle \): although \( C \) and \( X \) are not independent, they are not linearly correlated, and \( C = 0 \) may occur for \( X_1' \) either positive or negative (Figure 16a). Repeating this exercise but conditioning on the measured displacement instead will lead to \( X_1' = 0 \) due to the exactitude property of the kriging interpolator. The reason the displacement is more effective than the concentration is that unlike concentration, no ambiguity arises from its interpretation: a positive \( X_1' \), for example, can only arise from a positive \( Y' \). Hence by a mere change of the interpretation we are able to make a better use of the data.

**Bayesian Inversion**

For a statistician, an inverse problem is an inference or estimation problem. The data are finite in number and contain errors, as they do in classical estimation or inference problems, and the unknown typically is infinite dimensional, as it is in nonparametric regression. The additional complication in inverse problem is that the data could be directly and indirectly related to the unknown. Bayesian techniques have become more attractive in the hydrogeological communities through the elegant work of Tarantola (1987). One of the fundamental tenets of Bayesian inference is that uncertainty always can be represented as a probability distribution; in particular, the Bayesian approach treats the model as the outcome of a random experiment. The essential defining property of a Bayesian is to talk about the probability \( P(H|E) \) of a hypothesis \( H \), given evidence \( E \). Whether one adheres to a Bayesian view, estimators that arise from the Bayesian approach have an attractive property is that the posterior pdf is at least as informative as the prior one. In this case, the likelihood function is called *diffusive* or *totally noninformative*, and the prior estimates are exactly equal to the posterior estimates. It is emphasized that the method does not always guarantee better estimates for a couple of reasons. First, the Bayesian approach provides a pdf, not a single valued estimate. Second, the improvement achieved in the posterior pdf is dictated by the quality of external factors such as the accuracy of the likelihood function. Bayesian inversion is illustrated in the next Section “Joint geophysical-hydrogeological stochastic methods for subsurface characterization” particularly for combining geophysical survey and hydrogeological data for subsurface characterization.

**JOINT GEOPHYSICAL-HYDROGEOLOGICAL STOCHASTIC METHODS FOR SUBSURFACE CHARACTERIZATION**

**Motivation**

Combining ground-surface or cross-well geophysical surveys with well logs for enhancing the quality of subsurface characterization has been the goal of recent studies. The primary motivation has been the recognition that geophysical surveys offer unique opportunities for enhancing cross-well interpolation and are particularly promising in situations of data scarcity. Incorporation of two- and three-dimensional densely sampled geophysical data with conventional hydrological data increases the amount of data available for the characterization and thus has the potential to significantly improve the hydraulic parameter estimates over those obtained from borehole data alone.

The key problems are two: the nonuniqueness of the relationships between the hydrogeologic and geophysical rock/soil properties, and the absence of universal rock physics relationships that link the geophysical observable (i.e. electrical conductivity) to the hydrogeologic parameters (e.g. permeability, porosity). While these relations are sometimes difficult to obtain, it is critically important to recognize that even weak correlations can lead to a measurable improvement in estimation of the hydrogeologic variables. Integration of multisources of data is case and site-specific; however, the general framework is similar. The purpose of this section is to address the problem of joint hydrogeologic-geophysical site characterization following the study conducted by the author on Lawrence Livermore National Laboratory (LLNL) site data.
A few observations based on these studies are as follows: (i) No universal methods or petrophysical models are available for converting geophysical attributes to hydrogeological ones. (ii) The most challenging problem is tying well-logging measurements to the geophysical surveys. This issue involves problems of scale disparity and inconsistencies in the methods of data acquisition and interpretation. The last problem was studied extensively by Ezzedine et al. (1999) and can be demonstrated by the fact that resistivity at the LLNL site was measured along boreholes using several different tools, each characterized by a different support volume, sometimes leading to dramatically different results. Geophysical characterization is covered in Chapter 148, Aquifer Characterization by Geophysical Methods, Volume 4.

The main challenge that Ezzedine et al. (1999) had to face was in creating the bridge to link between ambiguously related geophysical surveys and well data. The second challenge was imposed by the disparity between the scale of the geophysical survey and the scale of the well logs. Ezzedine et al. (1999) approach is hierarchal and is intended to integrate and transform the well log data to a form where it can be updated by the geophysical survey, and this tends to be a convoluted process. In an ideal situation the geophysically measured attributes correlate well with the hydrogeological ones, for example, permeability, and the conversion of the geophysical survey to a hydrogeological distribution map is straightforward. In more realistic situations, such as the one described here, the conversion of the geophysical attributes to the hydrogeological ones is convoluted and nonunique. The difficulties faced in the implementation of the geophysical survey are several: the survey resistivity is expected to be of a relatively low resolution. At the same time, high-resolution permeability images are needed for flow and transport simulations. Thus the Bayesian framework has been chosen, which allows bridging between measurements of different resolutions.

### Bayesian Data Assimilation

Bayesian data assimilation is developed following a data-driven approach for lithology mapping based on the well log data. The proposed approach is general in its basic principles but at the same time is site specific since the petrophysical models employed are not universal. The general approach is stochastic. The choice is justified given the large uncertainty associated with cross-well interpolation, with the petrophysical models and with the interpretation of the geophysical surveys. The rationale for the approach is based on the following observations:

1. Resistivity and shaliness can be used for lithology identification through the cross-plot (Figure 17). Once a type of lithology is determined, further mapping of hydrogeological properties can be pursued.

2. Facies identification based on the shaliness-resistivity cross-plot is nonunique owing to some overlap between the sand and silt clusters.

3. Borehole resistivity measurements display short correlation range, and it is impractical to develop spatial images of the resistivity using cross-well geostatistical interpolation.

4. Shaliness displays a well-defined spatial correlation structure. It can be used for projecting resistivity measurements indirectly through a combination of geostatistical interpolation-simulation techniques, in conjunction with the nonlinear correlation structure it displays with the resistivity, as expressed through the cross-plot (Figure 17).

On the basis of these observations, the proposed approach consists of sequentially generating a series of collocated
attributes. At the basis of the hierarchy, images of the lithology are generated, conditional to well logs and possibly also to the survey resistivity. Each lithology image then serves as the basis for generating a series of shaliness images, again conditional to well data. The shaliness images are then used to correlate the survey resistivity with the hydrogeological attributes obtained experimentally. The series of generated images all have in common the well data and the same underlying spatial structure, and hence they are all physically plausible. The variations between the images constitute a measure of the spatial variability and estimation uncertainty. The focus is on estimating resistivity, but it can be converted to porosity and conductivity through well-known petrophysical models.

Bayesian Data Integration
The data integration encompasses the following steps:

**Step 1: Generation of the lithology images using sequential indicator simulation (SISIM).**

The lithology is defined through an indicator variable \( I \) according to: \( i = 1 \) if \( x \) is located in a silt body, 0 otherwise. Note that boldface letters denote vectors, that is, \( x \) is the location coordinates vector. Lowercase \( i \) is a realization of the spatial random function (SRF) \( I \). \( I \) is characterized through its expected value conditional to the borehole data, \( p^c = E[I] = E[I|\text{measurements}] \), with a superscript “c” denoting conditional. Since \( I \) is binary, \( p^c \) is statistically exhaustive. Its spatial variability is defined through the semivariogram and is shown in Figure 18(a,b). These statistics are the cornerstone of the SIS algorithm (Deutsch and Journel, 1997) adopted here.

**Step 2: Generation of shaliness images.**

This step is similar in principle to the previous one. The differences are in the fact that (i) the shaliness \( S \) is not a binary variable and (ii) the pattern of spatial variability of the shaliness may be different between the sand and silt lithologies, that is, \( \gamma_{Si} \), the semivariogram of the shaliness \( S \), depends on the lithology \( i = 0 \) or 1. SGS algorithm (Deutsch and Journel, 1997) is adopted here to generate shaliness images. Shaliness \( S \) is defined by its mean \( m_{Si} \), its semivariogram \( \gamma_{Si} \) and its covariance, \( \text{Cov}_{Si} \), for a given facies \( i \) (Figure 19a,b).

**Step 3: Computing the resistivity prior pdf.**

Once \( x \) is identified as being either sand or silt and is assigned a shaliness value, a prior pdf for the resistivity \( f_{R[i]}(r|I = i, S = s) \) can be defined through Figure 17(a). \( R \) and \( S \) denote the SRF of the resistivity and the shaliness, respectively, and \( r \) and \( s \) denote their realizations.

Figure 17(b) illustrates the joint pdf of \( R \) and \( S \) given \( I = 0 \) (i.e. sand lithology) and the marginals \( f_{R[i]}(r|I = 0) \) and \( f_{S}(s|I = 0) \). Conditioning further on \( S = s_0 \) leads to \( f_{R[i]}(r|S = s_0, I = 0) \), which is the Bayesian prior. Scarcity of data leads to condition on ranges of \( S \) values rather than on single values. These pdfs are the Bayesian prior pdfs of the resistivity, and hence the stochastic estimation for the resistivity \( R \) at \( x \) in case no additional data become available through surveying.

**Step 4: Updating \( f_{R[i]}(r|I = i, S = s) \) based on cross-well electromagnetic resistivity survey \( \rho(x) \).**

Defining \( f_{R[i]}(r|I = i, S = s) = f'_{R[i]}(r) \) for brevity, and given a collocated survey resistivity \( \rho(x) \), the posterior pdf \( f'_{R[i]}(r|\rho) \) can be defined through Bayes’ rule:

\[
C_R = \left( \int_{-\infty}^{+\infty} L(\rho|r)f'_{R[i]}(r)dr \right)^{-1}
\]

and

\[
f'_{R[i]}(r|\rho) = C_R L(\rho|r)f'_{R[i]}(r)
\]

![Figure 18](http://www.mrw.interscience.wiley.com/ehs) Experimental and theoretical indicator variograms: (a) vertical and (b) horizontal variograms (Reproduced from Ezzedine et al., 1999 by permission of American Geophysical Union). A color version of this image is available at http://www.mrw.interscience.wiley.com/ehs
where \( L(\rho | r) \) is the likelihood function, and \( C_R \) is a normalized factor. In general, \( \rho \) is defined over a support volume larger than the support volume of \( r \). In the case of a high-resolution geophysical survey \( \rho(x) \rightarrow r(x) \) and Bayesian updating is unnecessary. This, however, is not generally the case and the alternative is to update \( f'_{R(x)}(r) \) given \( \rho \). Typically we are interested in \( R \) representative of a block of scale \( \sim 1 \) m while \( \rho \) is defined by blocks of scale \( \sim 3 \) m or greater. The inference of the likelihood function, \( L(\rho | r) \), is critical for the successes of the updating process. Once \( f''_{R(x)}(r) \) is defined, a realization of \( R \) at \( x \) can be drawn. The whole process is repeated for all \( x \) until a complete image of the resistivity field is completed.

Similarly, the lithology images can be improved through the resistivity survey despite the nonlinear and nonunique relationship displayed in the cross-plot. The approach calls also for Bayesian updating of \( \rho_c \) as well, through the relationship:

\[
p'_{\rho_c} = C_I L(\rho | I) p^c,
\]

where \( L(\rho | I) \) is the likelihood function, of a similar nature to \( L(\rho | r) \), only relating \( \rho \) to \( I \) rather than \( R \). \( C_I \) is a normalized factor similar to \( C_R \).

**Step 5. Measure the effectiveness of the Bayesian updating.**

To evaluate the effectiveness of the updating procedure, the following statistic were analyzed:

\[
\begin{align*}
\mathcal{M}_k = & \frac{|r_k - m''|}{|r_k - m'|}, & \text{if } \mathcal{M}_k < 1 & \text{Successful updating,} \\
\mathcal{M}_k = & 1 & \text{Unsuccessful}
\end{align*}
\]

(72)

where \( k \) is a running index over all the points outside the wells, \( r \) is the actual resistivity, \( m'' \) is the mean of the posterior pdf \( f''_{R(x)}(x) \), and \( m' \) the mean of the prior pdf \( f'_R(x)(x) \). The ratio \( \mathcal{M} \) compares the performance of the posterior and the prior pdfs. \( \mathcal{M} \) smaller than 1 indicates a successful updating procedure. \( \mathcal{M} = 1 \) is a diffuse likelihood and hence a noninformative survey. Figure 20 depicts the variation of \( \mathcal{M} \), as a function of the resolution of the survey. For completeness, statistics were also computed for resistivity surveys of \( (2 \times 2) \) and \( (12 \times 12) \) block resolution. As expected, Figure 20 shows that \( \mathcal{M} \) decreases with decrease in resolution.

**THE USE OF STOCHASTIC CONCEPTS IN MODELING FRACTURED MEDIA**

**Introduction**

Most fractured rock systems consist of rock blocks bounded by discrete discontinuities comprised of fractures, joints,
and shear zones, usually occurring in sets with similar geometries. Fractures may be open, mineral-filled, deformed, or any combination thereof. Open fractures may provide conduits for the movement of groundwater and contaminants through an otherwise relatively impermeable rock mass. Major factors affecting groundwater flow through fractured rock include fracture density, orientation, effective aperture width, and the nature of the rock matrix. Fracture density and orientation are important determinants of the degree of interconnection of fracture sets, which is a critical feature contributing to the hydraulic conductivity of a fractured rock system. Only interconnected fractures provide pathways for groundwater flow and contaminant transport. Fractures oriented parallel to the hydraulic gradient are more likely to provide effective pathways than fractures oriented perpendicular to the hydraulic gradient. The cross-sectional area of a fracture will have an important effect on flow through the fracture. Fracture-flux is proportional to the cube of the fracture aperture, this relationship is valid only for fractures with apertures greater than tens of microns (Bear et al., 1993). Fracture apertures, and therefore flow through fractures, are highly stress-dependent, and generally decrease with depth. In recent years, a number of models have been developed to represent fracture flow. We introduce in the next section the most common geometric concepts of the models.

**Conceptual Model Geometry**

Several different conceptual models have been used to describe flow and transport in fractured media. The most common conceptual picture of flow and contaminant transport in a fractured porous medium is that the advective flow of water and transport of pollutants is largely, or entirely, through the fractures. Water and contaminants may diffuse into and out of the porous rock matrix. This diffusion can act to spread out the contaminant plume in space and time, and to retard it.

In situations where transient water flow is involved, water may also be stored in, and released from, the rock matrix and the dead zones. To the extent that there is sufficient primary porosity in the matrix to allow advective flow and transport, as might be the case for a sandstone, this basic conceptual picture will be in error, as will any model that is based on it. If the rock matrix has very low porosity, such as would be the case for granite, then the role of the rock matrix can often be neglected.

Several different approaches, or concepts, have been used to describe the fractured mass (Figure 21). Models can be roughly classified as equivalent porous media models, multiple interacting continua (MINC) such dual porosity models, and stochastic fracture models. One could also develop models that overlap these categories.

**The Equivalent Porous Medium**

This approach treats the fractured rock system as if it were an unconsolidated porous medium. This approach is most likely to be successful when the spacing of the fractures is small compared to the scale of the system being studied, and the fractures are interconnected. The validity of using the Equivalent Porous Medium (EPM) approach to model pollutant transport in a fractured system is less well established. EPM could be extended to include models that assume an equivalent unique fracture intersecting the well, such fracture could be horizontal, vertical, and so on. This approach, generally used in the petroleum industry, only gives the local properties of a major fracture and does not necessarily assume that an EPM exists.

For the sake of simplicity, let us assume for a moment that fractures in a three-dimensional space are planes of finite extent, with the shape of irregular polygons or ellipses. If the average size of each fracture is small, and if its density is very low, a situation may arise in which none of the fractures intersects another. In this case, the global permeability of the system would be zero, since water cannot flow from one fracture to another. Whatever the size of the Representative Elementary Volume (REV), there is no EPM. A concept other than just size is thus required for the existence of an EPM; this other topological
characteristic is called the “connectivity” of the network and has been thoroughly studied in other areas of physics under the name of “percolation theory”. An important concept in percolation theory is that of the “percolation threshold”, defining a density of fractures above which the connectivity of the fractures is sufficient for flow to take place through a network, even an infinite one. Below this threshold, a few fractures may be connected, forming a “finite cluster” where flow can take place, but if the size of the domain of interest is increased, the system is globally unconnected, and only local pervious clusters exist.

At present, the best criterion for determining if a fracture network is above or below the percolation threshold is probably the one proposed by Charlaix (1984); the threshold is given by \( n r^3 = 0.15 \) to 0.30; \( n \) is the density of fractures (number of fractures per unit volume of rock); \( r^3 \) is the cube of the constant radius of the fractures, assuming that the shape of a fracture is a circular disc. This expression assumes that the orientation of the fractures is random. If fractures are not discs, \( r^3 \) should be taken as the average area of the fractures multiplied by their average half-perimeter divided by two.

**Multiple Interacting Continua (MINC)**

A modification of the EPM is to model the system as if it were composed of two overlapping continua with different porosities and permeabilities (Barenblatt et al., 1960). Low porosity and high permeability are associated with the fractures and high porosity and low permeability are associated with the rock matrix. The model allows for the transfer of contaminants between the fractures and the rock matrix. This MINC approach requires that the fractures be closely spaced relative to the size of the system and that the fractures be frequently interconnected.

A simplified version of MINC is to represent the fractured system by a set of porous matrix blocks of well-defined geometry. The most common examples are parallel prismatic blocks (e.g. cubes, Figure 22) or spheres arranged in a regular array. The spaces between the blocks are the fracture channels. The blocks are assumed to be porous so that solutes can diffuse into and out of the matrix. This approach combines dual porosity with the discrete fracture approach. While no real aquifer has such a well-defined geometry, the model can provide insight into the important factors in solute transport in fractured porous media.

**Stochastic Discrete Fracture Networks (SDFN)**

In the discrete fracture approach, the fracture geometry is explicitly included. Fractures are most often represented as channels with parallel sides, and the individual fractures are combined into fracture networks. The simplest network has a set of parallel fractures in what is basically a one-dimensional problem. A more complex network has two sets of parallel fractures oriented at some angle to each other in a two-dimensional array. A more complex, and one step closer to reality, is to allow the fractures to have varying lengths, locations, and orientations relative to one another (Figure 21).

One obvious problem in the practical application of discrete fracture models is that it is almost impossible to define the fracture system at a site in fine enough detail to apply the model. The best possibility for this approach seems to be through some sort of statistical modeling of the fracture system to duplicate the measured hydrology at the site. Most of the work on complex discrete fracture networks has been done in connection with the disposal of nuclear waste in crystalline rocks and has not included diffusion into the rock matrix. Real rock fractures may have rough surfaces (walls) that are not parallel to each other. The fracture may be partially filled with precipitated minerals or the walls could touch each other under mechanical stresses. In this case, it is better to describe the fracture system with flow through a series of tortuous intersecting channels.

When the SDFN may be simplified to single main channels connecting the center of the intersecting fractures and the center of their respective intersection areas, the SDFN is reduced to 3D network of 1D network of pipes or channels (Figure 21). Integrated hydraulic properties, for example, hydraulic conductivity, are then defined and the problem is solved through classical analogies to electric resistance network. This alternative is very useful when dealing with a large number of fractures and small computational capabilities.

None of the above conceptual pictures is “best” in an absolute sense. Rather, each may be appropriate for a particular situation. Models that are conceptually simpler have the advantage of being easier to implement as a rule, but they may also oversimplify the situation and miss important phenomena that are taking place. More complex models have the potential to provide a more detailed description of what is happening at the site being

![Figure 22](image-url)
modeled, but they are also likely to be more difficult to implement and may require data that cannot be collected with currently available techniques. Concepts introduced in the Section “Geostatistics” can be directly applied to the stochastic continuum approach where fractured mass is reduced to a Rubik’s-cube-like equivalent heterogeneous “porous medium” where each block is defined by local EPM properties. Hereafter we focus only on SDFN models and their applications.

**Stochastic Discrete Fracture Network Models**

In this approach based on the model developed by Cacas (1989) and described in Ezzedine (1994), the fractures are treated as discs. The density and extensions are difficult to determine independently. There are relations between these magnitudes that can be used to test the connectivity of the medium. The hydraulic model is a bond model. The flow is not two-dimensional in the fracture plane but occurs in channels linking the centers of the connected discs. The walls are impervious. An additional parameter called aperture or “hydraulic thickness” is attributed to each disc in the model.

**Characteristics of the Fracture Families**

In natural fractured reservoirs, fractures can usually be grouped into families of directions using Schmidt or Rose diagram (Figure 23). For a single realization, a large number of fractures are drawn sequentially for one family of directions at a time. The families are then combined to

![Figure 23](https://example.com/figure23.png)  
**Figure 23** Stochastic discrete fracture network. Observed fracture in a borehole, observed fracture families (Rose diagram), simulated fracture families, and a 3D SDFN model (Reproduced with permission of Ezzedine, 1994)
create the final network. For a given family, the draw is made as follows:

**Number of Fractures** A Poisson process is used to obtain the number of discs belonging to family \( i \) in the simulated volume \( V (m^3) \) as a function of the volumetric density \( \lambda_i \) (number of discs/m\(^3\)). The following expression gives the probability of having an \( N_i \) number of centers in the volume \( V \).

\[
p(N_i = k) = e^{-\lambda_i V} \left( \frac{\lambda_i V}{k} \right)^k \frac{1}{k!}
\]  

(73)

**Location of the Disc Centers** The centers of the discs are independently distributed. The drawing of the Cartesian coordinates of the centers is done in a domain called the “generation domain”. The \((x_i, y_i, z_i)\) are coordinates of the centers are drawn according to a uniform law in the intervals \([x_{\text{min}}, x_{\text{max}}], [y_{\text{min}}, y_{\text{max}}]\) and \([z_{\text{min}}, z_{\text{max}}]\).

**Direction** The direction of a fracture is defined by its normal unit vector \( \mathbf{N} \). The families of directions in a network consist of fractures with similar directions. Their normal vectors form an approximately cone-shaped segment, the axis of which is the “pole” of the family. Fisher’s pdf is adopted and expressed as follows:

\[
f_\alpha (\alpha) = C(\kappa) \exp[\kappa \cos \alpha \sin \alpha] \quad \text{with} \quad C(\kappa) = \frac{\kappa}{4\pi \sinh(\kappa)}
\]  

(74)

where \( f \) represents the pdf of the angle \( \alpha \) formed with the pole of the family (the pole is here the mean normal vector of the family); \( \alpha \) is the angle between the fracture pole and the pole of the family; \( \kappa \) is the parameter of the law. For \( \kappa \to +\infty \) the distribution is very concentrated around the mean direction. For \( \kappa \to 0 \), this distribution is closer to the uniform distribution.

**Radius** The radius is the parameter describing the extension of the fracture. Several statistical laws are available and given in Table 2 for completeness. Recently, de Dreuzy et al. (2001) suggested random fracture networks following a power-law length distribution.

**Connectivity**

The analysis of the connectivity and the search for continuous paths within such fracture networks use algorithms shared with the theory of graphs. It is, in fact, easy to ascertain whether two fractures are interconnected or not, but in a network of fractures, it becomes very costly to examine all the fracture couples. Therefore, the flow region is cut up into a certain number of blocks. A given disc intersects one or several of these elementary volumes. Two discs will not cut across each other unless both of them together cross at least one of the blocks. It is therefore necessary to test a possible intersection of two discs only if they are geographically fairly close to each other. This method makes it possible to limit the number of tests. Then, in 3D, the number of fracture couples is reduced from \( 1/2N(N-1) \) to \( 1/2N^2h^3k^2/R^3 \) if the flow region is divided into cubes of volume \( h^3 \) and a fracture crosses \( k \)-cubes. Thus, the number of tests is reduced by \( \sim h^3k^2/R^3 \) provided that \( h \) is chosen as small as possible and \( k \) is of the order of 2 or 3 for one fracture.

**Channeling, Integrated Hydraulic Conductivity**

The “channeling hypothesis” is taken into account by defining permeability, said to be “integrated”, into which the effective flow area is introduced. Once the interconnected geometry of the formation has been constructed, one defines the links connecting the centers of the two fractures that pass through the middle of the intersection segment. Thus, the geometry is reduced to a group of one-dimensional elements placed end to end in a three-dimensional domain.
(Figure 24). Between two intersecting fractures, water circulates in a tube linking their centers and passing through the intersection segment. Assuming a fixed, straight, hollow cylinder of length $l$, large compared to the diameter $2R$, the volumetric flow in the tube is proportional to the fall of the pressure and to the fourth magnitude of the radius:

$$Q = \frac{\pi}{(8\mu)} R^4 \nabla h$$

(75)

$\mu$ is the kinematic viscosity and $\nabla h$ the hydraulic gradient between the ends of the tube. The “integrated hydraulic conductivity” is then defined as the ratio between the flow rate and the head gradient: $K = \pi/(8\mu) R^4$. The integrated hydraulic conductivity is the product of the “real” hydraulic conductivity (expressed in m s$^{-1}$) multiplied by the cross-section of the flow, which is difficult to measure by hydrogeological tests. In this model, each disc is characterized by an integrated hydraulic conductivity, which is determined by the aperture (radius) according to a lognormal pdf. Effective integrated hydraulic conductivity between two intersecting fractures is estimated using a weighted harmonic average. The weights are the length of the two segments forming the pipe connection between the fractures (Figure 25). This magnitude represents the mean resistance, which prevents the water from circulating between two intersecting fractures. It integrates and lumps the whole flow geometry at the scale of the fracture: wall roughness, tortuosity, flow line lengths and channeling, and so on.

**Flow in Fractured Network**

Similar to Kirchoff’s law in electricity (Figure 25), when the flow is laminar and steady state in each fracture flowing following Poiseuille’s law, at each fracture intersection the continuity of mass balance equation holds.

Consider a branch of a flow network (Figure 26) limited by two nodes $i$, and $j$, of length $L_{ij}$, where the flow is governed by Darcy’s law. This flow is characterized by a spatial variation of the conductivity $k_{ij}$. $x$ is the distance between the observed section and the node $i$, measured along the branch. Because of the Darcy equation, the flow toward node $i$ in branch $ij$ is written:

$$q_{ij} = -k_{ij}(x) \frac{\mathrm{d}h_{ij}}{\mathrm{d}x} \bigg|_{x=0}$$

and

$$q_{ji} = -k_{ji}(x) \frac{\mathrm{d}h_{ji}}{\mathrm{d}x} \bigg|_{x=L_{ij}}$$

(76)

Assuming that the nodes have no capacity, the mass balance equation at each node $i$ is $\sum_{j \neq i} q_{ij} = Q_i$, which is a system of linear equations in $h_{ij}$ that can be solved readily using classic numerical techniques.

Modeling the flow through each fracture by a single conduit is restrictive and in some cases does not reflect the more complex reality, especially when the fracture network encompasses only a few fractures. To alleviate these restrictions, Ezzedine (1994) introduced “daughter fractures” that are imbedded into the “mother fractures” in order to create a 2D network of channels of interconnected 1D pipes. The introduction of a random perturbation to the normal vector of “daughter fractures” leads to different intersecting flow channels within the “mother fractures” (Figure 27). This technique allows enhancing the flow within each mother fracture from a 1D flow into 2D flow, which is more representative of the reality.
Unsteady Flow and Transport in Fractured Network

In many applications, however, the assumption of steady flow is restrictive. Therefore, Ezzedine (1994) extended the model to simulate unsteady state flow. The flow at the fracture scale is derived analytically using Laplace transform. Flow at the formation level is then achieved by combining the solution of the generated individual fractures. A real flow injection test, performed at the Hot Dry Rock (HDR) site Soultz-sous-Forêts in Alsace/France, is simulated. Results show that the mechanical properties of the media influenced by the flow injection must be incorporated into the proposed model. In particular, changes in the hydraulic conductivity and the storage coefficient function of the effective stress appear to be quite significant. Consequently, the model is adjusted accordingly to include the mechanical effects that resulted in significant improvements. Heat extraction from Hot Dry Rock (HDR) is not limited to the hydrological phenomena discussed thus far but necessitates that thermal and geochemical processes must be addressed. Similar to fluid flow, a one-dimensional geochemical model was developed that takes into account the equilibrium and/or nonequilibrium transport of the chemical species and the water rock interaction. Thermal processes are modeled using a double porosity approach. The hydrothermal and chemical transport at the formation level is obtained by simultaneously combining the solution of the stochastically generated individual mother and daughter fractures.

Acknowledgments

The author would like to thank Ghislain de Marsily for his review. His comments and suggestions have improved the content and the clarity of this chapter.

REFERENCES


