2. HETEROGENEITY, PROBABILITY, AND RANDOM FIELDS

2.1 Introduction: Heterogeneity and Stochastic Analysis

Spatial heterogeneity refers to the variation of a physical property in two- or three-dimensional space. This physical variation is encountered in many earth science applications; it is of particular interest when studying flow and transport processes in the unsaturated zone. When examining soil media, spatial heterogeneity is observed on many different scales such as the microscale of a single pore, the intermediate scale of laboratory experiments, the scale of field experiments, and the megascale, which encompasses entire regions. This work is not concerned with the spatial heterogeneity on the microscale or pore scale because the governing physical laws for porous media flow (chapter 4) are only valid on a scale larger than the microscale. Bear (1972) defined "Representative Elementary Volume" as the smallest volume over which there is a constant "effective" proportionality factor between the flux and the total pressure gradient or total head gradient. This proportionality factor is called the hydraulic conductivity of the REV. By definition of the REV, the hydraulic conductivity does not rapidly change as the volume to which it applies is increased to sizes larger than the REV. This is based on the conceptual notion that either no heterogeneity is encountered at a scale larger than the REV or that heterogeneity occurs on distinctly scales, the smallest of which is the REV (Marsily, 1986). The latter model assumes that within each scale relatively homogeneous regions exist. Within these homogeneous units heterogeneities can only be defined on a significantly smaller scale. Geologists refer to these different scales as facies (Anderson, 1991) while hydrologists commonly speak in terms of hydrologic units (Neuman, 1991). Analysis of a large number of hydrologic and geologic data from different sites associated with different scales has shown that the existence of discrete hierarchical scales for any particular geologic or hydrologic system vanishes in the global view as the multitude of different geologic or hydrologic units allows for a continuous spectrum of scales (Neuman, 1990). This study is

limited to the scale of a typical research field site; this is representative for many field contamination studies i.e., roughly the meter and decameter scale (10^0-10^2 m).

For the scale of the REV, mathematical models based on the physics of flow and transport in homogeneous porous media have been well-established in the literature and their accuracy has been verified in many laboratory experiments (c.f. Hillel, 1980). The physical meaning of the underlying model parameters is already well-understood (c.f. Jury, 1991). It is the fundamental mathematical treatment of flow and transport in **heterogeneous** porous media, which is of concern in this study.

Spatially heterogeneous properties can belong in either one of the following two classes depending on the problem formulation:

(a) porous medium properties that are **measurable** and that are seen as the **cause** of flow and transport behavior in soils such as pore geometry, the saturated permeability of the soil, the soil textural properties, and the soil water distribution;

(b) porous medium properties that are **predictable** based on physical laws or **functions** of class (a) properties e.g., the distribution of soil moisture flow and the solute concentration at some future time.

In very general terms, this dissertation is about the spatial heterogeneity of class (b) properties given some knowledge about the heterogeneity of class (a) properties. More specifically, the spatial heterogeneity of (and hence the uncertainty about) soil water tension, soil moisture flux, and solute transport in soils is computed based on some information about the spatial heterogeneity of the hydraulic properties of unsaturated porous media. Mathematically, spatial heterogeneity can be dealt with in one of three ways:

1. The local (REV-scale) porous media properties (soil property class (a)) are described at every point throughout the field area of interest. Then the classic flow and transport equation are used in a numerical model to obtain the output variables (soil property class (b)). This approach requires an enormous amount of measurement data and is for all practical purposes impossible to implement.

2. Spatial heterogeneity is neglected and instead some form of "effective" parameter is determined to define the flow and transport problem on the scale of interest. Classical flow and transport solutions (analytical or numerical) are applied to solve the problem for a quasi-homogeneous domain with "effective" parameters. This is probably the most widely used approach in both soil and groundwater hydrology due to its relative simplicity and low computational cost. The disadvantage of the method lies in the uncertainty of the prediction, since the "real" field parameters may differ significantly from those assumed in the model.

3. An entirely new mathematical approach is developed that considers the limitations of our knowledge about the field site and that quantifies the uncertainty in the prediction of soil property class (b) given that only a small and economically reasonable amount of measurements are available from the field site (class (a) data or class (b) data). To that end, **spatial stochastic analysis** has been developed over the past three decades for a wide variety of similar problems not only in the treatment of unsaturated zone flow and solute transport but in the treatment of many earth science problems.

The stochastic approach is adopted in this study since the primary interest lies not only in making a best prediction but also in quantifying the uncertainty of the prediction. In this chapter probability theory is introduced together with some of its most important lemmas to show, how - in principle - spatial heterogeneity is conceptualized in form of a mathematical model. Based on probability theory, stochastic analysis (which is a particular form of mathematical analysis) allows the derivation o f the probabilistic parameters describing the spatial heterogeneity of class (b) properties given the probabilistic parameters describing the heterogeneity of class (a) properties. A particular challenge arises, when deterministic measurement data of either class (a) or class (b) properties are given in addition to the parameters describing the degree of heterogeneity. The additional deterministic information must be reflected in the derivation of the probabilistic parameters describing the spatial heterogeneity of class (b) properties. Conditional probability theory and conditional stochastic

analysis will be introduced for this type of application.

Stochastic analysis is closely associated with the theory of random processes, which is a branch of mathematics called probability theory. Probability theory itself is a branch of mathematics called measure theory. "*Probability theory and measure theory both concentrate on functions that assign real numbers to certain sets in an abstract space according to certain rules.*" (Gray and Davisson, 1986 p.27). The treatment of spatial heterogeneity in terms of random processes is a highly abstract procedure, the appropriateness of which has been questioned. However, this treatment is justified by Athanasios Papoulis (1984, p. xi):

> *"Scientific theories deal with concepts, not with reality. All theoretical results are derived from certain axioms by deductive logic. In physical sciences the theories are so formulated as to correspond in some useful sense to the real world, whatever that may mean. However, this correspondence is approximate, and the physical justification of all theoretical conclusions must be based on some form of inductive reasoning."*

For a complete derivation of the concepts of random variables, random processes, and stochastic differential equations there is a vast amount of literature that has been published in this area for many different applications (see e.g. Gray and Davisson, 1986; Papoulis, 1984; Priestley, 1981). The intent of this chapter is to give the reader a full appreciation of the theoretical basis of random processes. This should allow a better understanding of the scope of stochastic modeling.

2.2 Principles of Probability Theory

Probability theory is a construct that allows rigorous quantification of rather imprecise statements such as "tomorrow there may be rain", or "the contaminant plume may already have reached the groundwater table". Such statements are mostly based on past experiences in similar circumstances. The statements are about the chances of one particular result or outcome out of a total of two possible outcomes (a "sample space" of two outcomes), for example, groundwater

is either contaminated or not.

There could be more than two possible results e.g., *a*: "groundwater is contaminated to the extent that it affects a well-field", or *b*: "it is contaminated without affecting any wells", or *c*: "it is not contaminated at all". In this example there are three possible outcomes, which form a set called the "sample space" $\Omega=\{a,b,c\}$ of all possible outcomes. The two outcomes with contaminated groundwater can be grouped together in a set or "event" $F=\{a,b\}$, which is equal to the event "groundwater is contaminated" in the very first example. The complementary set to the event $F=\{a,b\}$ is the event $F^c=\{c\}$: "groundwater is not contaminated". The latter event is a singleton set or a set with only one element (outcome). Another possible event is "no well is contaminated" i.e., the set $F=\{b,c\}$. This event also possesses a complementary set within the sample space Ω: $F^c=\{a\}$ or "groundwater is contaminated to the extent that it affects some wells". Yet another possible event is $F=\{a,c\}$ "some wells are contaminated or the groundwater is not contaminated at all", which has the complementary event $F^c=\{b\}$ "groundwater is contaminated without affecting any wells". Finally there is the trivial event $F=\{a,b,c\}$ "the contamination has either not reached the groundwater, has reached the groundwater but no wells, or has already reached the wells", which is equal to the sample space Ω. This event possesses the complementary event $F^c=\{\phi\}$, the so-called empty set with no elements at all. All possible events for the sample space $\Omega=\{a,b,c\}$ are now defined. Each of these events is a set. A set of all eight possible events (or sets) in the sample space Ω, can also be defined. In set theory "sets of sets" are called "classes" of sets, and the class of *all* eight possible sets mentioned above is called the "power set" of Ω. Since each set is also an event, the power set is called a "class **F** of events" or simply an "event space". To be precise, each set is called an event only, if a probability measure P is assigned to the set. Only then the above probabilistic "experiment" is completely defined. As in this example, the theory of probability rests upon the principles of set theory.

In this simple example all basic elements necessary to define a probability problem, or - in mathematical terms - a "probability space" are encountered. Formally, a probability space

consists of three basic, well-defined objects: The sample space Ω, which is a set of all possible elementary outcomes; the event space **F**, which is the group of all possible events F such that each event F_i is some combination of the elementary outcomes and such that the event space **F** is closed under certain set-operations e.g., if $F_i \in \mathbf{F}$ then $F_i^c \in \mathbf{F}$. Finally, the space $(\Omega,\mathbf{F})$ must be "measurable" such that there exists a probability measure P, which assigns a probability to each event F_i in the event space. The three fundamental axioms of the probability measure P are:

$$P(F) \geq 0 \qquad \text{for all } F \in \mathbf{F}$$

$$P(\Omega) = 1 \tag{2-1}$$

$$P\left(\bigcup_{i=1}^{\infty} F_i\right) = \sum_{i=1}^{\infty} P(F_i) \qquad \text{if all } F_i \text{ are mutually exclusive}$$

where the last equation is for both a finite event space (finite countable number of events) and an infinite event space (infinite number of events).

The above contamination example was a finite event space. An infinite number of events can occur in a discrete sample space e.g., the sample space $\Omega(\mathbb{N})$ of all integer numbers, or in a continuous sample space e.g., the sample space Ω of all spatial points in a particular soil cross-section. An event space **F** with an infinite number of events is called a "Borel Field" **B**. One of the most common Borel fields is the class **B(R)** of all open intervals on the real line.

A very useful probability measure for discrete events can be derived from the

$$p(\omega) \geq 0 \qquad \text{for all } \omega \in \Omega$$

$$\sum_{\omega \in \Omega} p(\omega) = 1 \tag{2-2}$$

"probability mass function" (pmf) $p(\omega)$, which assigns a real number $p(\omega)$ to each elementary outcome ω in the sample space Ω, and which has the following properties:

The probability measure P(F) is defined as:

$$P(F) = \sum_{\omega \in F} p(\omega) \ , \ all F \in \mathbf{F} \qquad \textbf{(2-3)}$$

Note that the pmf is not a probability measure since it is defined for the elementary outcomes themselves and not for a collection of sets. An example of a commonly used pmf is the *uniform pmf*: $\Omega = \mathbb{N}_n = \{0, 1, 2, ..., n-1\}$ and $p(k) = 1/n$, $k \in \mathbb{N}$.

Similarly the "probability density function" (pdf) $f(\omega)$ for continuous sample spaces is defined by:

$$\begin{aligned} f(\omega) &\geq 0 \quad , \text{ all } \omega \in \Omega \\ \int_{\Omega} f(\omega)\, d\omega &= 1 \\ P(F) &= \int_{F} f(\omega)\, d\omega \quad , F \in \mathbf{B} \end{aligned} \qquad \textbf{(2-4)}$$

where $P(F)$ is the probability measure for the continuous sample space **B**. The probability density function of a random variable is often expressed in functional form. The following pdfs are of importance in this study:

$$\textit{uniform pdf:} \quad f(\omega) = \frac{1}{b-a} \qquad \omega \in [a,b], \quad [a,b] \in \mathbb{R}$$

$$\textit{normal pdf:} \quad f(\omega) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\omega-\mu)^2}{\sigma^2}\right) \qquad \omega \in \mathbb{R}$$

$$\textit{lognormal pdf:} \quad f(\omega) = \frac{1}{\sqrt{2\pi\sigma^2\omega^2}} \exp\left(\frac{-(\log\omega-\mu)^2}{2\sigma^2}\right) \qquad \omega \in \mathbb{R}^+ \tag{2-5}$$

$$\textit{Laplace pdf:} \quad f(\omega) = \frac{1}{2\beta} \exp(-|\omega-\alpha|) \qquad \omega \in \mathbb{R}$$

where μ and σ^2 are parameters of the normal and lognormal pdf. The normal pdf is also called the "Gaussian" pdf. α and β are parameters of the Laplace or double exponential pdf.

In the previous paragraphs the properties of the probability space $(\Omega,\mathbf{F},P)$ are defined and the pmf and pdf are introduced as tools to compute the probability measure P. The entire framework of stochastic analysis, which is a part of probability theory, rests like probability theory itself upon these definitions of the probability space, the basic operations of set theory, and the principles of mapping or functions. Mapping one probability space into another via some functional relationship is the key to the work presented in this dissertation, since the essence of stochastic analysis is the "connection of a system to a probability space with a description of the output" (Gray and Davisson, 1986, p.29) and the main objective of probability theory is "to find the probability of some new event formed by set-theoretic operations on given events, given a probabilistic description of a collection of events" (ibid.).

The first quoted statement was illustrated in the contamination example above. The second statement concerns itself both with the additivity property of the probability measure and with the mapping of a probability space into another probability space. The additivity property of P can be exemplified again with the contamination problem: Assuming the following probabilities are known: $P(\{c\})$ "the groundwater is not contaminated" is 20%, and the probability $P(\{b\})$ that "the groundwater is contaminated, but the contamination does not affect any wells" is 10%. What is the probability $P(\{a\})$ that "the groundwater is contaminated, and the contamination has also reached some wells"? From (2-3) it can easily be seen that

$P(\{a\})=P(\Omega)-P(\{b\})-P(\{c\})=70\%$. Similarly the probability $P(\{a,b\})$ of the event {a,b} "groundwater is contaminated" is then $P(\{a,b\})=P(\{a\})+P(\{b\})=80\%$.

Mapping (also called filtering, sampling, estimating, averaging, or measuring) is the process of mapping each element ω_A in a sample space Ω_A into another sample space Ω_B. An example is Ω_A = {a,b,c} of the contamination example. A second sample space Ω_B = {d,e,f} is defined such that *d*: "remedial action taken by EPA", *e*: "remedial action taken by fire department", *f*: "no remedial action taken". Also the following mapping (or function) is defined: $g:\omega \rightarrow \Omega_B$, such that g(a)=d, g(b)=d, g(c)=f i.e., if the groundwater is contaminated at all, EPA will take remedial action; if it is not (yet) contaminated, no remedial action will be initialized. The sample space Ω_A = {a,b,c} is called the domain of the function g, the sample space Ω_B = {d,e,f} is called the range of g, and the set of all $g(\Omega_A)$ in Ω_B, Ω_C = {d,f} is called the range space of g. g is a completely deterministic process, because it defines an exact mapping of the sample space Ω_A into the sample space Ω_B. Since each element of Ω_A is associated with a probability, g provides a tool to determine the probabilities in the range Ω_C of the domain Ω_A: The chances of {d} "EPA taking remedial action" are equal to the chances of {a,b} "groundwater is contaminated", which is 80%. Then the chances of "no action taken" are 20%.

Notice that the above example is not a one-to-one mapping and that there is no inverse mapping $g^{-1}(d)$. Defining g(a)=d, g(b)=e, and g(c)=f, Ω_A is mapped "onto" Ω_B and the range Ω_B is equal the range space Ω_C = {d,e,f}. g is said to have an inverse function g^{-1}, because the mapping is one-to-one.

A mapping can occur between a discrete domain and a discrete range, between a discrete domain and a continuous range (albeit the range space is still discrete), between a continuous domain and a continuous range, and also between a continuous domain and a discrete range. The contamination example was of the first category. Considering, for example, the amount of contaminant mass in the aquifer as a function of the outcome Ω_A would be a mapping between a discrete domain and the (continuous) real line $\mathbb{R}$. The stochastic analysis of unsaturated flow

and transport processes is in principal the mapping of the sample space of soil hydraulic properties into the sample space of flow and transport properties.

2.3 Independence and Conditional Probabilities

Reconsider the initial contamination example with the sample space $\Omega_A=\{a,b,c\}$. The elementary events of this sample space are "mutually exclusive", since given one of the events {a}, {b}, or {c}, none of the others can occur at the same contamination site (in the same experiment): Either the groundwater and the wells are contaminated, or only the groundwater is contaminated or the groundwater is not contaminated at all. It is impossible that the groundwater is contaminated AND that it is not contaminated. The probability of one event F_a occurring, if any other mutually exclusive event F_b has occurred is therefore zero:

$$P(F_a \cap F_b) = P(F_a)P(F_b) = 0 \qquad \textbf{(2-6)}$$

In contrast, one speaks of "independent" events, if the probability of one occurring is independent of whether the other event occurred. Independent events must be from at least two different experiments, each of which has a well-defined sample space. Say, for example, there exist two identical contamination sites at different locations, with a given probability space $(\Omega,\mathbf{F},P)$ for each of the two sites. The two events occurring at the two sites are called "independent", since the probability of the event occurring at one site is independent of the probability of the event occurring at the other site. Then the probability of the event F_a at site A AND the event F_b at site B to occur concurrently is:

$$P(F_a \cap F_b) = P(F_a)P(F_b) \geq 0 \qquad \textbf{(2-7)}$$

Similarly a collection of events $\{F_i, i=1,k\}$ is called "mutually independent", if

$$P\left(\bigcap_{i=1}^{k} F_i\right) = \prod_{i=1}^{k} P(F_i) \qquad \textbf{(2-8)}$$

An example of the latter is the probability of wells contaminated at all of several contamination sites throughout an area, each of which has a probability space (Ω,**F**,P). If, however, two contamination sites are located very close to each other, the resulting events may not be independent. Then their combined probability space has to be taken into account. The combined probability space is also called the 'multivariate' probability space.

The conditional probability is the probability that an event F_a occurs given that another event F_b has already been determined to have taken place. It must be emphasized that conditional probabilities can be defined for events at the same site or of the same experiment, but also for events from two different sites or experiments if their combined or multivariate probability space is considered. For example, the probability P({a}|{a,b}) of {a} "a well is affected by groundwater contamination", given that somehow it is known that the event {a,b} "groundwater is contaminated" has occurred is defined as:

$$P(\{a\}|\{a,b\}) = \frac{P(\{a\}\cap\{a,b\})}{P(\{a,b\})} = \frac{P(\{a\})}{P(\{a,b\})} \qquad \textbf{(2-9)}$$

More generally, the conditional probability of an event F_a given the occurrence of event F_b is:

$$P(F_a|F_b) = \frac{P(F_a\cap F_b)}{P(F_b)} \qquad \textbf{(2-10)}$$

It can be shown that conditional probabilities satisfy all the basic axioms of a probability space (2-1). An important property of conditional probabilities, which is derived from the above definition is "Bayes' theorem":

$$P(F_{A_i} \mid F_b) = \frac{P(F_b \mid F_{A_i})}{P(F_b \mid F_{A_1})P(F_{A_2}) + \ldots + P(F_b \mid F_{A_n})P(F_{A_n})} \qquad \textbf{(2-11)}$$

where the events F_{Ai} are mutually exclusive and the union of all events U $F_{Ai} = \Omega$. The event F_b is an arbitrary event in Ω. The denominator on the right hand side is called the "total probability" of the event F_b. It is the sum of all conditional probabilities of the event F_b given the collection of events $\{F_{Ai}, i=1,n\}$.

In the contamination example, the class of events {{a},{b},{c}} is one of several possible classes that are mutually exclusive and exhaustive of the sample space Ω (a "partition" of Ω). Assume soil samples were taken nearby the well. It is further assumed that the following conditional probabilities are known: If the well is affected by groundwater contamination, chances are 40% that the soil sample is also contaminated. If the groundwater is contaminated, but no wells are affected, chances are 30% that the soil sample is contaminated. If the groundwater is not contaminated at all, chances are 5% that the soil sample is contaminated. Bayes' theorem is used to determine the conditional probability of the event {a} "groundwater contaminated to the extent where it affects wells" given that the event {d} "soil sample contaminated" has occurred: The total probability of {d} (denominator of (2-11)) is:

$$P(\{d\}) = 0.4\ 0.7 + 0.3\ 0.1 + 0.05\ 0.2 = 0.32$$

Then the conditional probability

$$P(\{a\}|\{d\}) = 0.4\ 0.7\ /\ 0.32 = 0.875$$

In subsequent chapters, the concept of conditional probabilities is applied extensively to random variables and functions of random variables. Bayes' theorem plays a fundamental role in the development of this study.

2.4 Random Variables and Random Vectors

2.4.1 Random Variables

With the basic definitions of the probability space, of independence, of the conditional probability space, and of functions, we are well-equipped to proceed with the definition of a random variable X. The term random variable (RV) is actually improper, since by its formal definition a random variable is neither random nor a variable. Mathematically speaking a random variable $X(\omega)$ is a function that maps one to one any elementary outcome of an experiment (or probability space) $(\Omega,\mathbf{F},P)$ into a subset of the real line:

$$\begin{aligned} &X:\ \omega\in\Omega \rightarrow\mathbb{R} \\ &\text{such that: } X^{-1}(B) = \{\omega:\ X(\omega)\in B\} \in \mathbf{F}, \text{ if } B\in\mathbf{B}(\mathbb{R}) \end{aligned} \qquad \textbf{(2-12)}$$

where $\mathbf{B}(\mathbb{R})$ is a Borel field on a subset of the real line (where the "subset" is an interval and may be the entire real line itself). In other words, every outcome ω in the abstract sample space Ω is assigned a real number B through the random variable $X(\omega)$. Ω is the domain to the random variable, and the subset $\mathbf{B}(\mathbb{R})$ of the real line $\mathbb{R}$ is the range of the random variable X.

As an example, let us consider a small core sample of soil . Ω_{Ks} is the sample space of all saturated permeabilities. Then $X:\omega \rightarrow \mathbf{B}(\mathbb{R})$, $\omega \in \Omega_{Ks}$, is the saturated permeability of this soil sample measured in units of [length/time]. X is a real number corresponding to the physical property in the soil core called saturated permeability. Other random variables measured on the soil core are e.g. the water content, the matric potential, and the unsaturated hydraulic conductivity. Generally random variables can be considered as "measurements of an experiment" of which the outcome is unknown *a priori*.

The probability distribution of the random variable X (which may be a pmf, a pdf, or a mixture of both) can be derived from the probability distribution of the underlying experiment Ω since the probability $P_X(b)$ that X takes on a value in b is the probability that the inverse of X, X^{-1} takes on ω:

$$P_X(b) = P(X^{-1}(b)) = P(\{\omega: X(\omega) \in B\}) \qquad b \in \mathbf{B}(\mathbb{R}) \tag{2-13}$$

For most applications it is convenient to use the probability measure P_X rather than the original probability measure P_Ω i.e., one generally operates with the probability space $(\Omega_X, \mathbf{B}, P_X)$ where Ω_X is the range space of the random variable $X(\omega)$. It is important, however, to keep in mind that the probability space of X is only inherited from the original sample space. There may be other random variables that are derived from the same original sample space Ω. When analyzing the relationship between different random variables, their origins must be considered since common origins generally suggest certain dependencies between RVs of the same sample space. As an example consider the above mentioned soil core itself as being from the sample space Ω_{core}, at which different random variables are measured: the saturated hydraulic conductivity, the unsaturated hydraulic conductivity, the soil water potential, the water content, etc. Each of these random variables is a different type measurement of the exact same physical soil core. Mathematically speaking all random variables are in the same domain Ω_{core}. Although the derived probability measures or probability distributions of these random variables may vary, they are not necessarily independent of each other.

The relationship between the pmf $p_X(x)$ of a discrete RV X and its probability distribution P_X is defined equivalently to (2-2):

$$\begin{aligned} p_X(x) &= P_X(x) \qquad , x \in \mathbb{R} \\ P_X(B) &= \sum_B p_X(x) \qquad , B \in \mathbf{B} \end{aligned} \tag{2-14}$$

and in the case of a continuous X with a pdf $f_X(x)$:

$$P_X(x) = \int_B f_X(x)\,dx \qquad , B \in \mathbf{B}$$

$$f_X(x_i) = \frac{dF_x(x)}{dx}\Big|_{x=x_i} \qquad \textbf{(2-15)}$$

where

$$F_X(x) = P_X((-\infty,x])$$

$F_X(x_i)$ is called the cumulative distribution function (cdf) of X. It represents the cumulative probability of $X \le x_i$. Note the following properties of the cdf:

$$\begin{aligned} F_X(\infty) &= 1 \\ F_X(-\infty) &= 0 \\ F_X(x_1) &\le F_X(x_2) \text{ for } x_1 \le x2 \\ P_X(x>x_i) &= 1 - F_X(x_i) \end{aligned} \qquad \textbf{(2-16)}$$

The definition of the cdf allows for the construction of a relationship between the pdf and the probability measure of X such that one can be defined in terms of the other. The definitions of some of the most important pdfs including all those that will be used through the course of this study are already given in (2-5). Their respective cdfs are found by integration of the pdf over the half open interval (-∞,x].

The conditional cdf of a random variable X, given the event F_b is defined equivalently to (2-10):

$$F_X(x_i \mid F_b) = \frac{P(x \le x_i \cap F_b)}{P(F_b)} \qquad \textbf{(2-17)}$$

The conditional pdf of X is $f_X(x|F_b) = dF_X(x|F_b)/dx$. With these definitions Bayes' theorem for continuous random variables becomes:

$$f_X(x_i \mid F_b) = \frac{P(F_b \mid x=x_i) f_X(x_i)}{\int_{-\infty}^{\infty} P(F_b \mid x=a) f_X(a)\, da} \qquad \textbf{(2-18)}$$

With Bayes' theorem the conditional pdf of a random variable X given an event F_b is determined from the unconditional pdf of the random variable X and from the conditional pdf of the event F_b given the outcome of the random variable X. Bayes' theorem establishes the foundation for conditional simulation (also see chapter 3, chapter 10).

Given the (unconditional or conditional) pdf $f_X(x)$ of a random variable X the (unconditional or conditional) mean or expectation of X are defined as:

$$\mu_X = E(X) = \langle X \rangle = \int_{-\infty}^{\infty} x\, f_X(x)\, dx \qquad \textbf{(2-19)}$$

where the notation E() and < > are interchangeable and stand for 'expectation of'. The (unconditional or conditional) variance of X is defined as:

$$\sigma_X^2 = \langle (X-\mu_X)^2 \rangle = \int_{-\infty}^{\infty} (x-\mu_X)^2\, f_X(x) dx \qquad \textbf{(2-20)}$$

From the definition of the variance of X it follows that

$$\sigma^2_X = \langle X^2 \rangle - (\langle X \rangle)^2 \qquad \textbf{(2-21)}$$

The variance is the second-order central moment. Higher order central moments of X are defined as:

$$\langle (X-\mu_X)^n \rangle = \int_{-\infty}^{\infty} (x-\mu_X)^n f_x(x) dx \qquad \textbf{(2-22)}$$

The skewness of X is obtained for n=3, and the kurtosis of X for n=4. In most of the

applications in this study, it is assumed that a random variable has a Gaussian (or normal) distribution function (2.5). Using the above definitions of the various moments, it is seen that the mean of X corresponds to the parameter μ in the Gaussian pdf, the variance of X corresponds to the parameter σ^2 in the Gaussian pdf, the skewness is 0, and the kurtosis is $3\sigma^4$. If one can reasonably assume that a random variable is Gaussian distributed, the actual form of the pdf is completely determined by the first two moments, the mean and the variance of X.

Before proceeding to describe random vectors and random processes, another important concept related to random variables must be introduced: functions of random variables. Functions of random variables - like the random variable itself - allow the derivation of probabilities of new random variables through functional relationships. Suppose that g(x) is a function of the real variable x. Then the random variable Y defined by

$$Y = g(X) \qquad \textbf{(2-23)}a$$

is also a random variable, since Y is also a function on the original sample space Ω through

$$Y(\omega) = g(X(\omega)) \qquad \textbf{(2-23)}b$$

if the domain of the RV Y is X. Depending on the nature of g(X) various methods exist to derive the probability of Y from the probability of X. In this study, partial differential equations describe the relationship between most random variables of interest. In subsequent chapters methods are introduced to derive the pdf of Y from a given pdf of X if Y and X are related through a partial differential equation.

2.4.2 Random Vectors

So far, only one random variable and its probability distribution has been considered. Now we turn to the probability measure (probability distribution) of two or more random variables $X_1, X_2, ..., X_n$. Note that throughout this study vectors are denoted by boldface letters. A vector $\mathbf{X} = (X_1, X_2, ..., X_n)^T$ (^{T} indicates the transpose) is called a random vector if it is a

finite collection of n random variables X defined on a common probability space $(\Omega,\mathbf{F},P)$ The range space of a random vector is $\mathbf{B}(\mathbb{R})^n$ or a discrete subset of $\mathbf{B}(\mathbb{R})^n$, since the vector has n dimensions. The probability measure $P_{\mathbf{X}}$ of a random vector is different from the probability measure P_X of a random variable, since a vector is the joint outcome of several different measurements. Hence, a "joint probability distribution" of the random vector **X** must be defined. The formal definition of the cumulative probability distribution function $F_{\mathbf{X}}(\mathbf{x})$ of a random vector **X** with a continuous range space is:

$$F_{\mathbf{X}}(\mathbf{x}) = F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)$$

$$= P_{\mathbf{X}}(\mathbf{x}: \mathbf{x}\in(-\infty,x_i]; \; i=1,..n) = \int_{-\infty}^{x_1}\ldots\int_{-\infty}^{x_n} f_{\mathbf{X}}(\mathbf{x})d\mathbf{x} \qquad \textbf{(2-24)}$$

The corresponding joint pdf $f_{\mathbf{X}}(\mathbf{x})$ is obtained by taking the total derivative of $F_{\mathbf{X}}(\mathbf{x})$:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\partial F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 \partial x_2 \ldots \partial x_n} \qquad \textbf{(2-25)}$$

The joint cumulative distribution describes the probability that the random vector **X** takes on a particular value $\mathbf{x} = (x_1, x_2,\ldots, x_n)^T$ or less. But the joint probability distribution can also be used to derive the probability distribution P_{X_i} of a random variable X_i within a random vector. This is called the "marginal probability distribution" $f_{X_i}(x_i)$ i.e., the probability distribution of the random variable X_i without regard for the outcome of any of the other random variables in the random vector **X**:

$$F_{X_i}(a_i) = P_{X_i}(x_i \le a_i) = F_{\mathbf{X}}(\infty,\infty,\ldots,\infty,a_i,\infty,\ldots,\infty) \qquad \textbf{(2-26)}$$

The marginal probability density function $f_{X_i}(x_i)$ is found by integrating the joint pdf over all random variables other then X_i:

$$f_{X_i}(a_i) = \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x},x_i=a_i)\,dx_1\ldots dx_{i-1}dx_{i+1}\ldots dx_n \qquad \textbf{(2-27)}$$

Equivalent definitions for the marginal and joint distributions can be derived for discrete range spaces of **X** (i.e. if the components X of **X** can take on a finite discrete set of real numbers only).

Like for random variables, mathematical models are used to describe the joint pdf of **X**. In section 2.2 several pdfs are introduced for single random variables. An important joint pdf is the "joint Gaussian" pdf or "multidimensional Gaussian" pdf, where "multidimensional" refers to the dimensions of the random vector. The formal definition of the multidimensional Gaussian pdf involves two parameters similar to the two parameters μ and σ^2 in the one-dimensional or univariate Gaussian pdf: If **m** is a n-dimensional column vector and **C** a n by n matrix that is symmetric and positive definite, then a joint pdf is said to be Gaussian if it has the following form for any **m** and **C**:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\exp[-\tfrac{1}{2}(\mathbf{x}-\mathbf{m})^T \mathbf{C}^{-1}(\mathbf{x}-\mathbf{m})]}{\sqrt{(2\pi)^n \det \mathbf{C}}} \qquad \textbf{(2-28)}$$

where det**C** is the determinant of C. The vector **m** corresponds to the mean of the random vector **X**. It can be shown that each entry $C_{ij} = C_{ji}$ (symmetry!) can be found by determining the "covariance" of X_i and X_j. The covariance is a second order moment defined as:

$$\mathrm{Cov}(X_i, X_j) = \langle (X_i - m_i)(X_j - m_j) \rangle \qquad \textbf{(2-29)}$$

Then $C_{ij} = \mathrm{Cov}(X_i, X_j)$. The covariance is a measure of the physical correlation between the random variable X_i and X_j e.g., between the saturated hydraulic conductivity and the matric potential in a soil core. Notice that $\mathrm{Cov}(X_i, X_j) = \mathrm{Var}(X)$ for i=j.

If the random variables of a random vector are independent of each other (2-8) then the joint pdf for continuous random variables becomes simply the product of the marginal probability density functions:

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f(x_i) \qquad \textbf{(2-30)}$$

Similarly, the joint pmf for a discrete random variable is the product of the marginal pmf. The probability distributions related to these independent pdfs and pmfs are called "product distributions". If the pdf or pmf for each of the independent random variables in the random vector are the same, the vector is called an "independent and identical distributed" (i.i.d.) random vector.

Like for random variables, one may define a function of a random vector:

$$\mathbf{Y = g(X) = g(X_1, X_2, ..., X_n)} \qquad \textbf{(2-31)}$$

where Y is a new random variable. This definition is then used to derive the probability distribution of Y in terms of the joint probability distribution of **X**.

Random vectors are used in this study in two different ways that are mathematically equivalent, but differ in their physical interpretation. Random vectors of random variables may represent different physical properties such as $\mathbf{X}=(K,h,\Theta)^T$, where K denotes the saturated hydraulic conductivity, h the matric potential, and Θ the moisture content of the soil core mentioned in previous examples. If, for example, this random vector has a joint Gaussian pdf, the joint probability measure is determined by specifying the mean $\mathbf{m} = (m_K, m_h, m_\Theta)^T$ of each of the random variables and the (cross-) covariance C between each of the physical properties in the same soil core.

Alternatively, a random vector may represent the same physical property at different spatial locations in the soil: $\mathbf{X}=(\Theta(\mathbf{x}_1),\Theta(\mathbf{x}_2),...,\Theta(\mathbf{x}_n))^T$. Since each spatial location - statistically speaking - represents the outcome of a joint experiment with its own (marginal) probability space Ω_x, each physical property at each spatial location **x** forms a random variable by itself. If, for example, the water content throughout a field site possesses a joint Gaussian pdf, then the joint probability distribution is uniquely determined by the mean vector $\mathbf{m} = (<\Theta(\mathbf{x}_1)>, <\Theta(\mathbf{x}_2)>,...,<\Theta(\mathbf{x}_n)>)^T$ of Θ at each location separately (i.e. it doesn't have to be the same everywhere!), and the covariance matrix $\mathbf{C}_\Theta$ that defines the covariance between the Θs of each location pair.

While it is quite obvious that the saturated hydraulic conductivity and the soil matric potential at a location **x** in a soil domain are two different random variables with different marginal probability spaces, it must be emphasized here that without further assumptions the saturated hydraulic conductivity at one location is NOT the same random variable as the saturated hydraulic conductivity at another location and does not a priori possess the same marginal probability space! Similarly, the random variable "soil moisture" Θ at a time t_1 is not considered to be the same as the random variable Θ at a time t_2. Also note that there generally is an infinite number of either discrete or continuous physical locations **x** or times t, while the different physical properties are always a finite number of discrete variables. To distinguish between random vectors of different physical variables and random vectors of random variables in space and/or time, the terms 'random process' or 'random field' are used for the latter interpretation.

2.5 Random Processes and Random Fields

2.5.1 Definition

Random processes are an infinite collection of random variables where the random variables are indexed on a discrete or continuous "index set" **I**. In our applications this index set always corresponds to time t or spatial location **x**. The spatial location **x** is always denoted as a (lower case, bold print) vector of spatial coordinates and must not be confused with the probability $P_X(x)$ that the random variable X takes on a value x or the joint probability $P_X(\mathbf{x})$ that the random vector **X** takes on a vector value **x**!

The term "random process" or "stochastic process" is mostly used if the index set is the time variable, while the term "random field" is commonly applied for index sets of spatial locations. Formally, a random process is an indexed family of random variables $\{X(\omega,t); \omega\in\Omega, t\in\mathbf{I}\} = X_t = X(t)$ defined on a common probability space $(\Omega,\mathbf{F},P)$. Equivalently a random field is denoted as an indexed family of random variables $\{X(\omega,\mathbf{x}); \omega\in\Omega, \mathbf{x}\in\mathbf{I}\} = X_{\mathbf{x}} = X(\mathbf{x})$ on a

common probability space $(\Omega,\mathbf{F},P)$, where $\mathbf{I}=\mathbb{R}^n$ or a discrete subset of $\mathbb{R}^n$, $n\leq 3$, indicates the spatial dimensionality. Since the only difference in the definitions of random fields (random processes) and random vectors is the number of components (infinite vs. finite family of random variables) equivalent probability measures are defined for random fields and random processes: process cumulative distribution functions, process density functions, marginal cumulative distribution functions and marginal probability density functions.

Realizations (samples) of random fields are a basic element of the numerical stochastic analysis as will be shown in subsequent chapters. Often, the realizations themselves are referred to as random fields. To avoid confusion and to distinguish the random fields from random realizations of random fields subsequent chapters will use the term "random field variable" (RFV) to denote random fields that are families of random variables as defined above.

In numerical applications, random fields are always discretized in a finite domain. How do these finite discrete subsets relate to the infinite continuous random field? The "Kolmogorov extension theorem" shows that given a consistent family of finite-dimensional (joint) distributions

$$\{P(X(\mathbf{x}_1),X(\mathbf{x}_2),...,X(\mathbf{x}_n));\ \mathbf{x}_i \in \mathbf{I},i=1,...n\} \qquad \textbf{(2-32)}$$

there exists a random process or random field $\{\mathbf{X}(\mathbf{x}),\ \mathbf{x}\in\mathbf{I}\}$ described by these distributions. The term "consistent" distribution refers simply to fact that the joint distribution and the marginal distributions must be consistent in that one can be derived from the other through (2-26). This also includes the condition that "boxes" in n-space have positive probability. Probability distributions are consistent, if they are described, for example, by the multidimensional Gaussian pdf. From the above theorem it then follows that a random field $\{\mathbf{X}(\mathbf{x}),\ x\in\mathbf{I}\}$ is a "Gaussian random field" if ALL finite collections of samples of the random field $(X(\mathbf{x}_1), X(\mathbf{x}_2),...,X(\mathbf{x}_n))^T$ are Gaussian random vectors i.e., satisfy (2-28) and the conditions stated for $\mathbf{m}$ and $\mathbf{C}$.

To further distinguish between the covariances of the same physical property at different

times or locations (covariance of a random field or random process) and the covariance between two different physical properties at the same or at different locations, the latter is from now on referred to as a "cross-covariance".

Random fields and processes - like random vectors - may consist of independent random variables. If each independent random variable possesses the same pdf or pmf, the random field (process) is called an i.i.d. random field (process). Note that for independent random variables $\langle X_i\ X_j \rangle = \langle X_i \rangle \langle X_j \rangle$. The the covariance of two uncorrelated random variables is 0.

2.5.2 Stationarity and Ergodicity of Random Processes

All basic probabilistic concepts encountered in the study of heterogeneous porous media via stochastic analysis are now defined. Before continuing with the introduction of two rather intriguing properties of random fields, two important questions are raised: What is the justification for treating porous medium properties as random variables? And how does the heterogeneous environment of a porous medium i.e., a soil cross-section, a field-lysimeter, the unsaturated zone underneath a particular field site, or the aquifer underneath a certain region, relate to the properties of random variables and random fields?

While the first question is often posed, it fails to address the central problem of environmental modeling, which is not the heterogeneity of natural systems, but the measurement and/or estimation of the heterogeneous properties. Indeed the porous medium as it exists can be interpreted as being completely deterministic i.e., there is nothing random about the properties of the porous medium at any of its locations. The genesis of the pore morphology follows physical laws. All derived porous properties such as the permeability and the water content are also determined by physical principles. Hence, the heterogeneity of the soil does not in itself is completely deterministic.

The randomness lies in the lack of knowledge, and inability to acquire it fully, about

what these porous medium properties exactly are. Soil physical or chemical properties are commonly determined by either an actual **measurement** of soil properties or by the intuitive, graphical, or mathematical **estimation** of soil properties from related data (inverse distance interpolation, kriging, etc.). Both measurement and estimation are associated with errors. The (physically deterministic) errors occurring during the measurement and/or estimation process have the properties of random variables and thus allow a rigorous analysis with statistical tools. This is the key to stochastic analysis and the bridge between reality and conceptual model. Stochastic analysis in subsurface hydrology is about modeling the limitations of our knowledge! How limited our knowledge is will in turn depend on the porous medium heterogeneity. The focus of this study are the **estimation** errors (and NOT the **measurement** errors) occurring in predictions of soil water tension, soil water flux, and solute transport. Without loss of generality measurement errors are neglected.

The beauty of the stochastic analysis is that it provides both a best estimate of the properties of interest (hydraulic conductivity, soil moisture, solute transport, etc.) and a quantitative measure describing the uncertainty of the best estimate. The probability distributions encountered in stochastic modeling are essentially a reflection of the fuzziness or uncertainty of our knowledge about the soil properties. Hence, the justification for treating porous media as random fields lies NOT in the physical nature of the porous medium (which is deterministic) but in the limitation of our knowledge ABOUT the porous medium. This is not to say however that heterogeneity is unrelated to the statistical analysis. Indeed, the estimation error is a direct function of the soil heterogeneity: If the porous medium is relatively homogeneous, the properties of the soil at unmeasured locations are estimated with great certainty given a few sample data. On the other hand, if the porous medium is very heterogeneous and soil properties are correlated over only short distances, an estimation of the exact soil properties at unmeasured locations is associated with large errors. Hence, the heterogeneity of the soil is a measure of the estimation error or prediction uncertainty.

The second question addresses the practical problem of translating field measurements

(a "sample") into statistical parameters defining random variables i.e., into a probability space that is representative of the spatial variability and hence the estimation error with regard to the physical property of interest. This leads to the general problem of deriving "ensemble" statistical parameters of random fields (which consist of an infinite number of random variables, each of which has an infinite number of possible outcomes) from a small sample that gives ONE measurement of each of an INFINITE number of random variables. At the most, using the definitions of the mean, the variance, the (cross-)covariance, and the higher order moments (2-22) "sample" statistical parameters and a "sample probability distribution" or histogram of the measured random field parameters can be computed. The sample statistics give a quantitative estimate of the degree of heterogeneity in the porous medium, which also is an estimate of the expected estimation error. Then two problems need to be addressed:

1. The sample taken from measuring MANY random variables ONCE must be related to the MANY possible outcomes of any particular ONE random variable $X(\mathbf{x})$ at location $\mathbf{x}$.
2. The sample statistics must be related to the ensemble statistics of the random field.

These two points are crucial to the stochastic analysis and in particular the first one must not be underestimated. Recall that a random field consists of an infinite number of random variables, each of which has its own marginal pdf. The random variables in a random field need not have identical probability distributions. As will be seen in chapter 10, estimates of soil properties that are conditioned on field data are indeed always random fields with random variables whose pdf is a function of the location in space, since the uncertainty about field properties may vary from location to location (depending on whether the estimation is close to a measurement point or not)!

First, the question is addressed of how the measurement sample of different random variables (same physical property at different locations in the same single realization of a random field, namely the actual field site) can be taken to be equivalent to many measurements

of the same random variable (same physical property at one location in many different hypothetical realizations of the site including the actual one). The definition of a random field as a collection of random variables in space says *a priori* nothing about the **spatial** relationship of the marginal probability distributions of the random variables that make up the random field, except for the condition that they must form a proper joint probability distribution. But an entire probability distribution for each and every random variable in the random field must be found. This poses a severe dilemma for the statistical treatment of many earth science problems: Only a single realization of the random field is available since all regional and subregional geologic, pedologic, and other environmental phenomena are unique and do not repeat themselves elsewhere. This is a very different problem from flipping a coin, an experiment that can easily be repeated (and measured) as many times as necessary to determine its sample probability distribution. To circumvent the dilemma it is assumed that the marginal probability distribution function of each random variable is identical at every location in the random field. In other words, one must assume that the likelihood that a physical property takes on a particular value B, is exactly the same everywhere in the field. This implies that the mean, the variance, and the other moments of the probability distribution are identical for every location in the random field. This property is called "stationarity" or "strict stationarity". A formal definition is given:

$$P(X(\mathbf{x})\in B;\ \mathbf{x}\in I) = P(X(\mathbf{x}+\Delta\mathbf{x})\in B;\ \mathbf{x}\in I) \qquad \textbf{(2-33)}$$

where B is an event of the Borel field $\mathbf{B}(\mathbb{R})$. In all the applications of this study, a weaker form of stationarity is assumed: "second order stationarity" or "weak stationarity" or "wide-sense stationarity", which requires that the mean and covariance (but not any higher order moment) are identical everywhere in the random field:

$$\langle X(\mathbf{x})\rangle = \langle X\rangle \quad , \text{ all } \mathbf{x}\in I \qquad \textbf{(2-34)}$$

$$Cov_X(X(\mathbf{x}),X(\mathbf{x}+\Delta\xi)) = Cov_X(\Delta\xi) \quad , \text{ all } \mathbf{x},\Delta\xi\text{: } \mathbf{x},\ \mathbf{x}+\Delta\xi \in I \qquad \textbf{(2-35)}$$

Two important examples of strictly stationary processes are the i.i.d. random field, which by definition has identical distributions for each of the random variables in the random field. A Gaussian random field is called weakly stationary if the mean $\mu_X(\mathbf{x})=\mu$ for all $\mathbf{x}$, and the covariance $Cov(\mathbf{x}_1,\mathbf{x}_2)=Cov(\xi)$, $\xi = \mathbf{x}_2 - \mathbf{x}_1$ for all $\mathbf{x}_1$, $\mathbf{x}_2$ on the index set $\mathbf{I}$. Since the Gaussian random field is completely defined by its first two moments, all higher order moments of the Gaussian random field must be stationary if the first two moments are stationary. Hence weakly stationary Gaussian fields are also strictly stationary.

The existence of stationarity in porous medium properties cannot be proven rigorously at any single field site. Data are often sparsely distributed. In the best of cases a linear or higher order trend can reasonably be removed from the data. For all practical purposes, it is therefore convenient to *hypothesize* that the field site is a realization of a weakly stationary random field (after removing an obvious trend). This is a reasonable assumption in many field applications. Once this working hypothesis is postulated, the sample of measurements at different locations is treated as if it were a sample of several realizations of the same random variable (i.e. at the same location).

Next, the sample moments must be related to the ensemble moments of the random variable. This problem is treated by stochastic theorems related to convergence and to the law of large numbers (cf. Gray and Davisson, 1986). The definition of "convergence in the mean square" is: A sequence of random variables X_i, i=1,2,... (e.g. a random field) converges in the mean square to a random variable X if

$$\lim_{n\to\infty} \langle (X_i - X)^2 \rangle = 0 \qquad \textbf{(2-36)}$$

Convergence in the mean square sense is mathematically also written as

$$\underset{n\to\infty}{\text{l.i.m.}}\ X_i = X \qquad \textbf{(2-37)}$$

where l.i.m. stands for "limit in the mean". To solve the problem of relating sample statistics to ensemble parameters it is necessary that the sample statistics taken from a single realization

indeed converge to the ensemble statistics of the random variables as the number of samples is increased:

$$\underset{n\to\infty}{\text{l.i.m.}}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu_X \qquad \textbf{(2-38)}$$

A random field or random process that satisfies this theorem is called "mean ergodic". A sufficient condition for weakly stationary random fields to be mean ergodic with a limiting sample average μ_X is that $C_X(0) < \infty$ and that $\lim_{n->\infty} C_X(n\Delta\mathbf{x}) = 0$. In other words a weakly stationary random field is mean ergodic, if the variance is finite and if random variables are uncorrelated at large separation distances (Papoulis, 1984).

Like stationarity, mean ergodicity cannot be measured in a single field site i.e., a single realization of the hypothetical random field. Rather mean ergodicity is taken as a working hypothesis i.e., it is assumed *a priori* that the measured sample statistics converge in the mean square to the true ensemble parameters as the number of samples increases.

Note that the above definition of a mean-ergodic random field is only a special case of a more general ergodic theorem that states that the sample expectations converge to a constant (not necessarily the mean of any random variable in the random field) as the number of samples increases. Ergodic processes need not be stationary (e.g. a random field with underlying periodic trend) and similarly stationary random fields need not be ergodic (the sample expectations may not converge as the sample size increases such as in the flipping a coin experiment). For all applications in this study, however, both weak stationarity and mean-ergodicity are postulated as working hypotheses. Thus limited knowledge of a deterministic reality can be related to an abstract probability space. Once the step has been made from the sample to the probability space (via the working hypotheses stationarity and ergodicity) the tools available from the definitions of probability theory as stated in the previous sections are used to make probabilistic predictions about the current status of the porous medium at locations other than those from where measurements are available and to make probabilistic predictions

about the future status at both unmeasured and measured locations. It should now be clear that a "probabilistic prediction" will not determine an actual value of a physical property. Rather it will give the moments (or probability distribution) of the random variable defined on the error of estimating a physical property.

Finally it is emphasized that the assumption of mean ergodicity (2-38) does not imply identity of the sample mean m_X based on n samples of a random variable X and the ensemble mean μ_X. For the same reason, the sample variance var_X or sample covariance cov_X are not identical to their respective ergodic limit σ_X^2 and C_X. The difference between the sample statistical parameters of X and its ensemble moments is generally referred to as parameter estimation error and will subsequently be neglected. Such parameter estimation errors, however, are recognized to be an important source of uncertainty in field applications of the stochastic approach.

2.5.3 Conditional Random Fields and Kriging

In section 2.3 the conditional probability space of random variables was introduced. The Bayesian theorems stated there are readily extended to random fields. Conditional probabilities in random fields are defined for separately for each random variable, given the exact outcome of other random variables. In this study, the conditional first and second moment (conditional mean and covariance) of a random variable are of particular interest. These two moments are sufficient to describe the conditional pdf if the underlying unconditional joint probability distribution is Gaussian.

Recall the (multivariate) **joint** cumulative probability distribution function $F_X(x_1,x_2,x_3,...,x_n)$ is defined as:

$$P_X(x_1,x_2,x_3,...,x_n) = \int\int\int...\int f_X(x_1,x_2,x_3,...,x_n)\, dx_1 dx_2 dx_3 ... dx_n \qquad \textbf{(2-39)}$$

with the unconditional expectation (first moment)

$$\langle X_i \rangle = \int_{-\infty}^{\infty} x_i f_X(x_i)\, dx_i \qquad \textbf{(2-40)}$$

and the unconditional covariance (second moment):

$$C_{ij} = \langle X_i' X_j' \rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_i' x_j' \, f_X(x_i,x_j) dx_i dx_j \qquad \textbf{(2-41)}$$

where $x_i' = x_i - \langle X_i \rangle$ is the perturbation around the mean. The variance σ_X^2 is defined by (2-41) for i = j. The joint pdf is the derivative of the joint cumulative distribution function. Assume that m datapoints of the n datapoints of interest were already measured. Then the **marginal** probability density function $f_X(x_{m+1},x_{m+2},...,x_n)$ of the unknown data (RVs) $X_{m+1},...,X_n$ in the unknown ensemble of data $X_1,...,X_n$ is defined by (see 2-27):

$$f_X(x_{m+1},x_{m+2},...,x_n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_X(x_1,x_2,...,x_n) dx_1 dx_2 ... dx_m \qquad \textbf{(2-42)}$$

With the help of these two distribution functions the **conditional** probability density function $f_X(x_{m+1},x_{m+2},x_{m+3},...,x_n|x_1,x_2,x_3,...,x_m)$ of the unknown data is defined given the actual values for the data at points $\mathbf{x}_1,...,\mathbf{x}_m$:

$$f_X(x_{m+1},x_{m+2},...,x_n|\ x_1,x_2,...,x_m) = \frac{f_X(x_1,x_2,...x_m,x_{m+1},...,x_n)}{f_X(x_1,x_2,...,x_m)} \qquad \textbf{(2-43)}$$

The conditional expectation $\langle X_i \rangle^c$ of X_i is defined as:

$$\langle X_i \rangle^c = \langle X_i|x_1,x_2,...,x_m \rangle = \int_{-\infty}^{\infty} x_i f_X(x_i|x_1,x_2,...,x_m)\ dx_i \qquad \textbf{(2-44)}$$

and the conditional covariance (second moment) by

$$E_{ij} = \langle X_i X_j | x_1, x_2, .., x_m \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_i x_j f_X(x_i, x_j | x_1, x_2, .., x_m)\, dx_i dx_j \quad i,j = 1,...,n \tag{2-45}$$

To make complicated matters simple, it is assumed that the unconditional joint pdf is multivariate normal and hence fully characterized by its first and second moments, the mean and the covariance. For practical reasons it is also assumed that the unconditional random process is stationary i.e., the first and second moment are not functions of the spatial location **x**. It is important to understand, however, that the conditional random process is NOT stationary, even if the unconditional probability field is stationary. In other words, even if the unconditional mean $\langle X \rangle$ and variance σ^2_X are independent of location **x**, the conditional mean $\langle X_i \rangle^c$, the conditional variance E_{ii}, and the conditional covariance E_{ij} are functions of location $\mathbf{x}_i$.

Matheron (1971) contributed extensively to the theory of random variables in space, and developed a "best, linear, unbiased estimator" to estimate random variables in space from a few known data, which has become widely known as "kriging" (c.f. Journel and Huijbregts, 1978). Essentially his analysis includes a derivation of the conditional moments of the random variables based on the concept of "regionalized variables" (Matheron's term for random variables in space). If the random field is Gaussian, then the algorithm for determining the conditional expectations in a random field is identical to kriging. Kriging is one of the main tools in geostatistics (Dagan, 1982). Kriging techniques have been developed for second order stationary fields with known constant mean (simple kriging), for intrinsic stationary fields i.e., random fields with constant but unknown mean and weakly stationary increments $(X_i - X_j)$ (ordinary kriging), for intrinsic stationary random fields with an underlying trend of known order (universal kriging), and for some non-Gaussian random fields (log-kriging, probability kriging, disjunctive kriging, and indicator kriging) (Matheron, 1971; Journel, 1983; Armstrong and Matheron, 1986; Journel, 1988; Kim, 1988). For the purpose of this study, it will be convenient to restrict ourselves to the case of a weakly stationary random process (random

field) with a constant unconditional mean $\langle X \rangle$ and finite unconditional variance σ^2_X (simple kriging). The conditional expectation $\langle X_i \rangle^c$ (2-44) can then be computed by a linear estimate $\langle X_i \rangle^k$ (i.e. $\langle X_i \rangle^k$ is a linear function of the given data) such that the mean square of the estimation error $\langle e^2 \rangle$, $e = (X - X^k)$ is minimized (c.f. Papoulis, 1984, p.167ff.). The estimation of the conditional expectation is given by (Dagan, 1982):

$$\langle X_i \rangle^k = \langle X_i | x_1, x_2, .., x_m \rangle = \langle X \rangle + \Sigma \lambda_{ij} (x_j - \langle X \rangle)$$

alternatively: **(2-46)**

$$\langle \mathbf{x_2} \rangle^k = \langle \mathbf{X_2} | \mathbf{x_1} \rangle = \langle \mathbf{X_1} \rangle + \Lambda_{21} (\mathbf{x_1} - \langle \mathbf{X_1} \rangle)$$

where $\mathbf{x_1}$ is the vector of known data of the RVs $\mathbf{X_1} = (X_1,...,X_m)^T$ (at locations $\mathbf{x_1},...,\mathbf{x_m}$) and $\mathbf{X_2}$ is the vector of unknown RVs $(X_{m+1},...,X_n)^T$ (at locations $\mathbf{x_{m+1}},...,\mathbf{x_n}$). In the geostatistics literature, $\langle \mathbf{X_2} \rangle^k$ is referred to as the (simple) kriging estimator (Journel, 1988). The weights λ_{ij} in the weight matrix Λ_{21} are obtained by minimizing $\langle \mathbf{e}^2 \rangle$, which leads to a linear system of equations called the (simple) kriging system of equations:

$$\sum_{j=1}^{m} C_{ij} \lambda_{jk} = C_{ik} \qquad i = 1,2,...,m \qquad k = m+1,...,n$$

alternatively: **(2-47)**

$$\mathbf{C_{11}} \Lambda_{12} = \mathbf{C_{12}}$$

where $\mathbf{C_{11}}$ is the covariance matrix between the known datapoints of $\mathbf{X_1}$ and $\mathbf{C_{12}}$ is the covariance matrix between known datapoints of $\mathbf{X_1}$ and points of unknown data $\mathbf{X_2}$. The kriging system has a solution only if $\mathbf{C_{11}}$ is a positive definite matrix (Journel and Huijbregts, 1978). To assure positive definiteness, the sample covariance data obtained from analysis of $\mathbf{x_1}$ are fitted to an optimal (i.e. best-fitting), valid (i.e. assuring positive definiteness) functional form of the covariance, such as the exponential, spherical, or gaussian models (Isaaks and Srivastava, 1989). The minimized "estimation error covariance" or "mean square error" corresponds exactly to the conditional covariance or simple kriging covariance and is given by:

$$E_{ij} = \langle X_i' X_j' | X_1, X_2, \ldots, X_m \rangle = C_{ij} - \sum_{k=1}^{m} C_{ik} \lambda_{kj} \qquad i,j = m+1,\ldots,n$$

alternatively:

$$\mathbf{E_{22}} = \mathbf{C_{22}} - \Lambda_{21} \mathbf{C_{12}} \qquad \textbf{(2-48)}$$

Note, that the individual entries in the conditional covariance or error covariance matrix $\mathbf{E}_{22}$ are equal to or smaller than the entries in the unconditional covariance matrix $\mathbf{C}_{22}$!

2.6 Spectral Representation of Random Variables

In the analysis of random processes (time series), "spectral analysis" has been an important tool for many different tasks and is a well-established field of probability theory (c.f. Priestley, 1981). Recently, spectral analysis has also become important for the study of spatially variable processes (random fields). Introduced into the field of subsurface hydrology by Gelhar et al., (1974) to study groundwater systems, it has since been applied to a great variety of subsurface hydrologic problems (e.g. Bakr et al., 1978; Gutjahr et al., 1978; Gelhar and Axness, 1983; Yeh et al., 1985a,b; Li et al, 1992).

In principle, spectral analysis is founded on the concept that a single realization of a random process (RVs defined on a 1-dimensional time index) or of a random field (RVs defined on a n-dimensional location index, $n \leq 3$) is nothing but a superposition of many (even infinitely) different (n-dimensional) sine-cosine waves, each of which has different amplitude and frequency. Then any particular realization of a random field can be expressed either in terms of a spatial function or in terms of the frequencies and amplitudes of the sine-cosine waves and their amplitudes (called 'Fourier series' of a discrete process and 'Fourier transform' of a continuous process). The latter are collectively called the "spectral representation" of the random field. The spectral representation of a single random field realization can intuitively be understood as a field of amplitudes, where the coordinates are the frequencies of the sine-

cosine waves. In other words, instead of an actual value for each location in space, the spectral representation gives an amplitude for each possible frequency (wave-length). Note that in n-dimensional space, $n \leq 3$, sine-cosine waves are defined by n-dimensional frequencies (with one component for each spatial direction) and therefore the spectral representation of a n-dimensional random field is also n-dimensional.

The spectral representation is defined deterministically i.e., it is not defined in a probability space and has by itself little to do with a stochastic solution: Each realization of a random field has its own spectral representation, since the amplitudes of the underlying sine-cosine waves are different for each realization. But obviously and following the rules already established in the previous sections the amplitudes of the sine-cosine waves can be defined as random variables with the frequency domain as the index field. In other words, a function of a spatial random field (which is defined on a probability space) is established rather than a function of a realization of a random field (which is a deterministic function). One then deals with the probability space of the spectral representation, which in turn also is a random field, but defined in the frequency domain. Statistically speaking, the probability space of the spatial random field is mapped onto the probability space of the spectral random field.

The advantages of representing a random field in terms of its underlying spectral properties i.e., in terms of the probabilities of amplitudes and frequencies of the "waves" composing a random field, are many. But within the framework of this study two properties are particularly important:

1. The spectral representation of a spatially correlated random field i.e., of random variables with a joint probability distribution is - under certain conditions - a random field with random variables (amplitudes) that are uncorrelated i.e., they are completely defined by their univariate marginal distribution, the analysis of which is much easier than that of random variables with a multivariate joint distribution function.
2. Under certain conditions, the spectral transformation of a partial differential equation is a polynomial whose solution is found much easier than the solution to the partial

differential equation in the spatial domain.

In this study, the tools of spectral analysis are used for three different but related tasks:

1. in a probabilistic sense to analytically derive the joint probability distributions of functions of random fields,
2. in a deterministic sense to numerically generate realizations of random fields of spatially variable parameters,
3. in a deterministic sense to obtain explicit solutions to partial differential equations defined by a particular (deterministic) realization of random fields.

In this section, the basic theorems of spectral analysis are introduced. In the following chapters they are applied to generate random fields (chapter 3), and to derive the joint probability distribution functions of parameters of interest in unsaturated flow and transport by applying spectral analysis to the governing unsaturated flow equation (chapter 4). Finally, in chapter 7 a method is introduced that combines spectral and numerical analysis to efficiently obtain solutions of the unsaturated flow equation given a particular random field realization of the constitutive parameters.

For reasons discussed earlier, this study is solely concerned with stationary random fields. The spectral analysis of stationary random fields has been well-established in the literature and many fine texts can be found on the general subject (e.g. Priestley, 1981). Here only the basic theorems are introduced, which are necessary to understand the techniques applied to the stochastic analysis of flow and transport processes. For complete proofs and a broad introduction to the topic the reader is referred to the established literature (c.f. Priestley, 1981).

The spectral representation of a single realization $X(\mathbf{x})$ of a random field with mean 0 is formally defined in terms of the Fourier-Stieltjes integral (Wiener, 1930):

$$X(\mathbf{x}) = \int_{-\infty}^{\infty} e^{i\mathbf{k}\mathbf{x}} \, dZ(\mathbf{k}) \qquad \textbf{(2-49)}$$

where the integral is n-dimensional, $n \leq 3$, and $Z(\mathbf{k})$ is a (complex valued) function, called the Fourier-Stieltjes transform of $X(\mathbf{x})$. The Fourier-Stieltjes integral must be chosen over the more common Fourier-Riemann integral

$$f(\mathbf{x}) = \int_{-\infty}^{\infty} e^{i\mathbf{k}\mathbf{x}}\, g(\mathbf{k})\, d\mathbf{k} \qquad \textbf{(2-50)}$$

where $g(\mathbf{k})$ is the Fourier transform of $f(\mathbf{x})$ since the Fourier-Stieltjes transform $Z(\mathbf{k})$ of the random field $X(\mathbf{x})$ is generally not differentiable such that $dZ(\mathbf{k}) = z(\mathbf{k})\, d\mathbf{k}$. $Z(\mathbf{k})$ can be understood as an integrated measure of the amplitudes of the frequencies between $(-\infty, \mathbf{k}]$ contributing to the realization $X(\mathbf{x})$.

As already mentioned above, Z can also be interpreted as a random field consisting of random variables $Z(\mathbf{k})$ defined in the frequency domain, where the random field Z is a stochastic function of the random field X i.e., each realization X_i is mapped into a realization of the spectral representation Z_i. In this probabilistic sense (2-49) essentially expresses the fact that "*(virtually) any stationary [random] process [random field] can be represented as (the limit) of the sum of sine and cosine functions with random coefficients dZ(**k**), or more precisely, with random amplitudes |dZ(**k**)| and random phases arg{dZ(**k**)}*" (Priestley, 1981, p.245). The new probability space $(\Omega_Z, \mathbf{F}_Z, P_z)$ of the random variables $Z(\mathbf{k})$ in (2-49) has several very important properties:

$$\begin{aligned} &1.\quad \langle dZ(\mathbf{k}) \rangle = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\mathbf{k}\mathbf{x}} \langle X(\mathbf{x}) \rangle\, d\mathbf{x} \\ &2.\quad \langle |dZ(\mathbf{k})|^2 \rangle = S(\mathbf{k})\, d\mathbf{k} \\ &3.\quad \langle dZ(\mathbf{k}_1)\, dZ^*(\mathbf{k}_2) \rangle = 0 \qquad \text{all } \mathbf{k}_1 \neq \mathbf{k}_2 \end{aligned} \qquad \textbf{(2-51)}$$

The first property states that the mean $\langle dZ(\mathbf{k}) \rangle$ of the random variables $dZ(\mathbf{k})$ is equal to the Fourier transform of the mean of the random variables $X(\mathbf{x})$. In subsequent applications, only

zero-mean random processes are considered, hence the spectral representations are also of zero mean. The second property defines the variance ($S(\mathbf{k})\,d\mathbf{k}$) of the random variable $dZ(\mathbf{k})$. The term $S(\mathbf{k})\,d\mathbf{k}$ is a measure of the average "energy per unit area" or "power" contribution of the amplitude of a frequency $\mathbf{k}$ to the random field $X(\mathbf{x})$. $S(\mathbf{k})$ is called the "spectral density" or "spectrum" of the random field X. $S(\mathbf{k})$ depends purely on the probabilistic properties of the random field $X(\mathbf{x})$ and it can be shown that it is simply the Fourier transform of the covariance $C(\xi)$ of X. The third property states that the increments $dZ(\mathbf{k}_1)$ and $dZ(\mathbf{k}_2)$ at two different frequencies $\mathbf{k}_1$ and $\mathbf{k}_2$ are uncorrelated. Such a random field is also called an "orthogonal" random field.

Through (2-51) the first two moments of the random field $dZ(\mathbf{k})$ are defined solely in terms of the first two moments of the stationary random field $X(\mathbf{x})$. Hence, if the first two moments of the random field $X(\mathbf{x})$ are known, then the first two moments of its spectral representation $dZ(\mathbf{k})$ are known. Note that the spectral representation $dZ(\mathbf{k})$ of a weakly stationary random field $X(\mathbf{x})$ is only stationary to first order: The mean $\langle dZ(\mathbf{k})\rangle$ is constant (first property), but the variance $S(\mathbf{k})$ of the random field $dZ(\mathbf{k})$ is a function of the location $\mathbf{k}$ in the frequency domain (second property).

In summary of this last section, a new probability space, called the spectral representation of a random field, was defined on the known probability space of a random field. The mapping of a stationary, correlated random field X into its spectral representation dZ provides the important advantage of creating an equivalent dZ to the random field X that consists of orthogonal or uncorrelated random variables!