UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Seepage in Heterogeneous Soil Masses

Dr. Felipe Vázquez Guillén

Ciencia Nueva
DOCTORADO UNAM

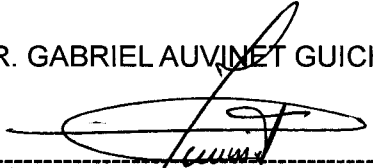**Dr. Gabriel Auvinet Guichard**
Asesor

SUPERVISOR BOARD:

Presidente:	DR. DANIEL RESENDIZ NÚÑEZ
Secretario:	DR. EFRAÍN OVANDO SHELLEY
Vocal:	DR. GABRIEL AUVINET GUICHARD
1 er. Suplente:	DR. ERNESTO HEREDIA ZAVONI
2 d o. Suplente:	DR. ANDRAS BÁRDOSSY

Places where this thesis was developed: INSTITUTE OF GEOTECHNICAL ENGINEERING, UNIVERSITY OF STUTTGART, GERMANY, DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING, UNIVERSITY OF SOUTHERN CALIFORNIA, USA, INSTITUTE OF ENGINEERING, UNAM, MEXICO.

ADVISOR:

DR. GABRIEL AUVINET GUICHARD

SIGNATURE

ABSTRACT

The object of this thesis is to develop numerical tools to characterize heterogeneous soil masses with respect to their hydraulic conductivities from a probabilistic perspective of analysis. The proposed tools can be used to simulate site-specific realizations of random fields conditional to measurements of hydraulic conductivity and histories of hydraulic heads. The document is first devoted to discus the concept of dependence that is defined in the context of random fields then theoretical concepts of the model adopted to represent uncertainty due to spatial variability of the hydraulic conductivity of soil masses is presented. To characterize soil masses taken into account available hydraulic head observations, an inversion technique based on the so-called Ensemble Kalman Filter method is proposed. Both the original and the propose techniques are used to simulate conductivity random fields conditional to histories of hydraulic heads. The developed tools are validated through numerical experiments. In one experiment, the conditional conductivity fields are used to predict seepage velocities in a transversal cross-section of the internal core of an idealized earth dam for the purpose of identifying preferential seepage paths. The characterization of non-homogeneous 1D and 2D hypothetic aquifers is also considered. The presentation of this research finalizes with general conclusions and recommendations that attempt to motivate future investigations.

RESUMEN

El objetivo de esta tesis es desarrollar herramientas numéricas para caracterizar masas térreas heterogéneas en relación con sus conductividades hidráulicas desde una perspectiva de análisis probabilista. Las herramientas propuestas se pueden utilizar para simular campos aleatorios condicionales a mediciones de la conductividad hidráulica y a historias de cargas hidráulicas. El documento se dedica primero a discutir el concepto de dependencia que se define en el contexto de los campos aleatorios y luego presenta los fundamentos teóricos del modelo utilizado para representar la incertidumbre asociada a la variabilidad espacial de la conductividad hidráulica en las masas de suelo. Para caracterizar masas de suelo tomando en cuenta observaciones disponibles de la carga hidráulica, se propone una técnica de modelación inversa basada en el método conocido como filtro de Kalman ensamblado. Tanto la técnica original como la que se propone se utilizan para simular campos aleatorios condicionales a historias de cargas hidráulicas. Las herramientas desarrolladas se validan con experimentos numéricos. En uno de los experimentos, los campos aleatorios condicionales se utilizan para pronosticar velocidades de filtración en la sección transversal del núcleo interno de una presa de tierra idealizada con el propósito de identificar trayectorias preferenciales de filtración. Se considera además la caracterización de acuíferos hipotéticos heterogéneos en una y dos dimensiones. La presentación de esta investigación finaliza con conclusiones generales y recomendaciones que intentan motivar futuras investigaciones.

ACKNOWLEDGEMENTS

CONTENTS

Abstract
Resumen
Acknowledgments
Nomenclature
List of figures
List of tables

ii

NOMENCLATURE

LIST OF VARIABLES

a_x	Correlation scale in the x direction
$\hat{c}(\mathbf{h}, u_1, u_2)$	Bivariate empirical copula density
C	Total number of holes in the objects N of a binary image
$C(u, v)$	Bivariate copula of the random variables U and V
$C(u_1, \ldots, u_n)$	Multivariate copula
C_S	Bivariate spatial copula
$C_{\mathbf{x} \mid n}(\cdot)$	Conditional copula
CV_{Ks}	Coefficient of variation of the saturated hydraulic conductivity
$C(\mathbf{h}; z_p)$	Covariance of the indicator random field $I(\mathbf{x}; z_p)$, called indicator covariance
$C(0; z_p)$	Variance of the indicator random field
$C_Y(\mathbf{x}_1, \mathbf{x}_2)$	Auto-covariance function
$C_Y(\lvert \mathbf{x}_1 - \mathbf{x}_2 \rvert)$	Auto-covariance function of a stationary random field
$C_{H'}(\chi_i, \chi_j)$	Auto-covariance functions between transformed hydraulic heads
$C_{Y'}(\mathbf{x}_i, \mathbf{x}_j)$	Auto-covariance functions between transformed log-conductivities
$C_{Y'H'}(\mathbf{x}_j, \chi_i)$	Cross-covariance functions between transformed log-conductivities and transformed hydraulic heads
$C^G_{\boldsymbol{\Gamma}}(\cdot)$	Multivariate Gaussian copula with correlation matrix $\boldsymbol{\Gamma}$
$C^V_{\boldsymbol{\lambda}, \boldsymbol{\Gamma}}(\cdot)$	Multivariate V-transformed copula with parameters $\boldsymbol{\lambda}$ and correlation matrix $\boldsymbol{\Gamma}$

e_i	i-th finite element
$f(z_1, z_2)$	Bivariate probability density function of the random variables Z_1 and Z_2
$F^{-1}_1, \ldots, F^{-1}_n$	Inverse of the marginal distribution functions
$\hat{F}_n(z)$	n-th empirical distribution function
$F_{\hat{H}'(\chi)}$	Cumulative distribution function of transformed, updated hydraulic heads at location χ
$F_{Yi}(y_i; \mathbf{x}_i)$	i-th marginal distribution function of the random variable Y located at $\mathbf{x}_i$
$F_{\hat{Y}'(\mathbf{x})}$	Cumulative distribution function of transformed, updated log-conductivities at location $\mathbf{x}$
$F_V(v)$	Cumulative distribution function (empirical) of the V-transformed field
g	Gravitational acceleration
$g(\mathbf{x})$	Multi-Gaussian field
G	Theoretical Gaussian cumulative distribution function
h_0	Initial hydraulic head at $t=0$
h_1	Prescribed hydraulic head
$h'_1(\chi_i)$	Transformed value of an observation of the hydraulic head field located at χ_i and time $t=1$
$h'_1(\chi_i)$	Transformed value of the hydraulic head at location χ_i and time $t=1$
$\hat{h}'_1(\chi)$	Realization of the random field of transformed hydraulic heads at time $t=1$ generated by p-field simulation
$\mathbf{h}$	Vector separating any two locations
H	Hydraulic head
H_f	Entropy of the probability density function f
$H(F^{-1}_1(u_1), \ldots, F^{-1}_n(u_n))$	Multivariate distribution function
$H(\chi)$	Random field of hydraulic heads
$H_1(\chi)$	Spatiotemporal random field of hydraulic heads
$\mathbf{H}^f_t$	Joint, reduced vector of forecasted states (prior values at the locations of the observations) of dimension $N_y + N_h$, defined for each realization in the ensemble

$\mathbf{H}_t^u$	Vector of dimension N consisting of back-transformed, updated hydraulic heads realizations
$I(\mathbf{x}, z_p)$	Indicator random field defined for the cutoff value z_p
k	Parameter of the V-Transform
k_s	Value of the saturated hydraulic conductivity
$k_s(\mathbf{x})$	Field of the saturated hydraulic conductivity
$K_s(\mathbf{x})$	Random field of the saturated hydraulic conductivity
$\mathbf{K}_t$	Matrix of size $(n+N)$x(N_y+N_h) called "Kalman gain"
$\mathbf{K}_\alpha$	Sub-matrix containing the weighting functions α of the simple coKriging system of equations. $\mathbf{K}_\beta$, $\mathbf{K}_\lambda$ and $\mathbf{K}_\eta$ are also sub-matrices containing the weighting functions β, λ, η, respectively, of such a system
m	Parameter of the V-Transform
$m(\mathbf{x}, t_1)$	Arithmetic mean at $\mathbf{x}$ and time t_1
n	Number of log-conductivities in the flow domain
n_r	Number of realizations in the ensemble
$\mathbf{n}(\mathbf{x})$	Outward vector normal to the boundary Γ_N
N	Number of isolated objects in a binary image. It is also used to represent the number of hydraulic heads in the flow domain
N_y	Total number of direct measurements of the log-hydraulic conductivity
N_h	Total number of observations of the hydraulic head
$p = F_Z(z_p)$	*p-quantile* of the probability distribution of $Z(\mathbf{x})$
$p_1(\mathbf{x})$	Probability field of log-conductivities at time $t=1$
$P[Y(\mathbf{x}_i) \leq y_i]$	Probability that the value of the random variable Y located at $\mathbf{x}_i$ be less or equal to the value y_i
$P[Y\|Y=y, H=h]$	Conditional probability of the random variable Y
$\mathbf{q}(\mathbf{x}, t)$	Prescribed flux across Neumann boundary segments Γ_N
$s^2_{en}(\mathbf{x}_i)$	Variance at location $\mathbf{x}_i$ determined statistically over the ensemble of realizations
$S \in \mathfrak{R}^n$	Domain of interest in the dimensional space $n=1,2$ o 3
$S(\mathbf{h})$	Set of pairs consisting of distribution function values separated by the vector $\mathbf{h}$
$S_s(\mathbf{x})$	Specific storage

t	Time
$u_1,...,u_n$	Uniformly distributed random variables on the interval $[0,1]$
(u,v)	Values of the distribution functions of the uniform random variables U and V
$(\mathbf{U}_1,\mathbf{U}_2)$	Vectors consisting of realizations of correlated random variables uniformly distributed
$\mathbf{U}_t^u$	Joint vector of dimension $n+N$ consisting of updated realizations
$\mathbf{U}_0$	Joint vector of dimension $n+N$ consisting of forecasted states (prior values)
$v(\mathbf{x})$	Field derived from the V-transform
$V(\mathbf{x})$	Random field of filtration velocities
w	Variance contribution function
$\mathbf{x}$	Spatial coordinate associated to the hydraulic conductivity over the flow domain Ω
$(\mathbf{X}_1,\mathbf{X}_2)$	Vectors consisting of realizations of correlated random variables
y	Variable associated to the log-hydraulic conductivity
$y'=G^{-1}(\cdot)$	Value of the inverse of the theoretical cumulative Gaussian distribution function
$y'(\mathbf{x})$	Log-conductivity field with a univariate Gaussian distribution i.e. $N(0,1)$
$y^*(\mathbf{x}_i)$	Estimated log-conductivity value at location $\mathbf{x}_i$
$y^{ref}(\mathbf{x}_i)$	Log-conductivity value of the reference field at location $\mathbf{x}_i$
$\hat{y}'_1(\mathbf{x})$	Realization of the transformed log-conductivity field at time $t=1$
$y'_0(\mathbf{x}_i)$	Transformed value of the log-conductivity at location $\mathbf{x}_i$ and time $t=0$
$y'_1(\mathbf{x}_i)$	Transformed value of an observation of the log-conductivity field located at $\mathbf{x}_i$ and time $t=1$
$\hat{y}'_1(\mathbf{x})$	Realization of the transformed log-conductivity field at time $t=1$ generated by p-field simulation
$Y(\mathbf{x})$	Random field of the log-conductivity
$Y_0(\mathbf{x})$	Spatiotemporal random field of the log-conductivity at time $t=0$
$\mathbf{Y}_t^u$	Vector of dimension n consisting of back-transformed, updated log-conductivity realizations
z	Elevation, Cutoff value

$(\mathbf{Z}_1, \mathbf{Z}_2)$	Vectors consisting of realizations of correlated random variables log-normally distributed
$z(\mathbf{x})$	Realization of a random field of hydraulic conductivities
$Z(\mathbf{x})$	Random field of the hydraulic conductivity
$\mathbf{Z}_t$	Joint vector of dimension N_h consisting of the observation values

LIST OF SYMBOLS

α	Parameter of the *V*-Transform
$\alpha(\mathbf{x})$	Weighting function of the simple cokriging system of equations
$\beta(\chi)$	Weighting function of the simple cokriging system of equations
χ	Spatial coordinate associated to the hydraulic head field over the flow domain Ω
$\chi(p)$	Euler characteristic
Γ_D	Dirichlet boundary
η_{ij}	Empirical frequencies corresponding to a regular mesh mxm with centers at the coordinates i,j
$\eta(\chi)$	Weighting function of the simple cokriging system of equations
$\lambda(\mathbf{x})$	Weighting function of the simple cokriging system of equations
λ	Vector consisting of copula parameters
μ	Fluid viscosity
μ_Y	Mean value of the random variable Y
$\mu_Y(\mathbf{x})$	Mean value of the random field $Y(\mathbf{x})$
$\mu_{Y|Y,H}$	Conditional mean of the random variable Y
$\mu_{\hat{Y}'}(\mathbf{x})$	Mean value of the distribution function $F_{\hat{Y}'(\mathbf{x})}$ located at $\mathbf{x}$
$\mu_{\hat{H}'}(\chi)$	Mean value of the distribution function $F_{\hat{H}'(\chi)}$ located at χ
μ_{Ks}	Mean value of the saturated hydraulic conductivity
$\hat{\mu}_{Y|Y,H}(\mathbf{x}, t_1)$	Estimator of the conditional mean value of the random field Y at location $\mathbf{x}$ and time t_1
σ_n^2	Variance of a normally distributed random variable

σ^2_Y	Variance of the random variable Y
$\sigma^2_Y(\mathbf{x})$	Variance of the random field $Y(\mathbf{x})$
$\sigma^2_{Y\|Y,H}$	Conditional variance of the random variable Y
$\sigma_{\hat{Y}'}(\mathbf{x})$	Standard deviation of the distribution $F_{\hat{Y}'(\mathbf{x})}$ located at $\mathbf{x}$
$\sigma_{\hat{H}'}(\boldsymbol{\chi})$	Standard deviation of the distribution $F_{\hat{H}'(\boldsymbol{\chi})}$ located at $\boldsymbol{\chi}$
σ^2_Y	Variance of the random variable Y
$\hat{\sigma}^2_{Y\|Y,H}(\mathbf{x},t_1)$	Estimator of the conditional variance of the random field Y at location $\mathbf{x}$ and time t_1
ρ_d	Fluid density
ρ_n	*Pearson*'s correlation coefficient among normally distributed random variables
ρ_{ln}	*Pearson*'s correlation coefficient among log-normally distributed random variables
$\rho(\mathbf{h})$	Correlogram
$\rho_I(\mathbf{h};z_p)$	Indicator correlogram
ρ_s	*Spearman*'s rank correlation
$\rho_s(\mathbf{h})$	*Spearman*'s rank correlation function
τ	*Kendall*'s rank correlation
$\bar{\tau}, \tau^+$	*Kendall*'s conditional correlation coefficient
Φ	Standard Gaussian distribution
Ω	Flow domain with Ω in $\mathbb{R}^p$ for $p=1$ or 2
∇	Gradient operator

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Motivation

The hydraulic conductivity of soil masses is very sensitive to variations of internal structure. In the same type of soil, hydraulic conductivity may vary by several orders of magnitude. Field measurements show clearly that such variations are the rule more than the exception. This fact can be verified both in natural soil formations and in mechanically stabilized soils wherein heterogeneity is likely to occur because of the intrinsic properties of materials at the borrow banks and the procedure adopted for their collocation. In addition, the number of available measurements for determining the hydraulic conductivity of a soil mass in a specific site is often limited. As a result, the hydraulic properties of soil masses are so uncertain that they are best characterized in a probabilistic manner.

Numerical models are often employed in practical situations to analyze seepage conditions. Such models permit to incorporate in the analysis the heterogeneity of the porous media. In real-world problems, the presence of heterogeneity is often critical to the performance or the reliability of soil structures subjected to seepage. For example, K. V. Terzaghi pointed out that minor geological details frequently control the behavior of the soil mass. In this context, it is inconvenient to simplify the analysis of seepage using mean values in idealized homogeneous domains.

Problem definition

A particular case of high relevance in the practice of geotechnical engineering is the identification of preferential seepage paths in the core of earth dams. In heterogeneous media, fluid moves faster along preferred paths of least resistance. The presence of high conductive zones in earthen structures has different sources. It may be the consequence of the inadvertent inclusion of coarser materials or poor mechanical compaction, among other factors.

Although the sole occurrence of preferential seepage pathways in earthen structures does not necessarily lead to failure, it may indicate the risk of internal erosion or the existence of soil cracking. In the event of inadequate performance of filters, preferential seepage paths may facilitate the development of the mechanism of internal erosion in the core. This last phenomenon has been the main cause of failure of earthen structures. The detection of preferential seepage paths in its incipient condition is therefore an issue in dam embankment engineering.

Contribution

The object of this thesis is to develop numerical tools to characterize heterogeneous soil masses with respect to their hydraulic conductivities from a probabilistic perspective of analysis. The propose tools will be used to simulate realizations of random fields conditional to measurements of hydraulic conductivity and/or histories of hydraulic heads. The developed tools during this investigation will be validated through numerical experiments. In one experiment, conductivity fields conditional to histories of hydraulic heads will be generated to predict seepage velocities in a transversal cross-section of the internal core of an idealized earth dam to identify preferential seepage paths.

It is expected that the proposed tools reinforce the experienced engineer's criteria during the process of identification of preferential seepage paths in earth dams and help engineers to make decisions with respect to the safety of such kind of dams.

Organization of the document

The presentation of this research has been organized in five chapters. In chapter 1 the concept of multi-Gaussian dependence in spatial random fields is discussed. Chapter 2 presents the random field model adopted to represent heterogeneous hydraulic conductivities and a simulation algorithm of random fields is proposed. In chapter 3, an extension of the so-called Ensemble Kalman Filter method is proposed and its performance in relation to the original technique is discussed. The characterization of a non-homogeneous hypothetic aquifer with respect to their hydraulic conductivities is conducted in chapter 4. Chapter 5 explores the benefits of histories of hydraulic heads to characterize a non multi-Gaussian conductivity field that mimics continuous zones of hydraulic conductivities in a triangular dam core. The presentation finalizes with general conclusions and recommendations that attempt to motivate future investigations.

CHAPTER 1

FLOW IN CONTINUOUS RANDOM MEDIA: A REVIEW ON THE MULTI-GAUSSIAN DEPENDENCE CONCEPT

Hydraulic conductivity of soils can be determined at least in three major axes of anisotropy by means of any well established standard experiment. However, in practice it is often measured in a limited number of locations within a region of interest treating it as a scalar quantity, namely no as a tensor. As a rule more than an exception, measurements of the hydraulic conductivity of soils exhibit significant spatial variability or heterogeneity even in fairly "homogeneous" formations. It is observed that this spatial variability is very difficult to be described in a deterministic way from a few set of measurements. Statistical analyses of measurements performed in different sites have shown however that such spatial variability presents certain structure or organization.

It is now widely accepted that spatial variability of soil hydraulic conductivity can be interpreted, for practical purposes, as a realization of a random field (Dagan, 1989; Ghelar, 1993; Zhang, 2002; Rubin, 2003). Parameters of the random field can be estimated from local samples taken from the realization itself under the *ergodicity* hypothesis. The spatial structure of the random field can be described by covariance functions. The concept of variogram is also useful under some circumstances. It was introduced by Matheron (1967) who also suggested using the *Kriging* technique for spatial interpolation purposes. All these relevant concepts conform part of the so-called *geostatistics*.

A stochastic approach involves then the adoption of a probabilistic model called random field for the purpose of modeling the spatial variability of the hydraulic conductivity of soils. One or more hydraulic properties can be modeled by random fields. The equations of flow in this approach become *stochastic differential equations*. Dependent variables of flow such as hydraulic head and flux become also random fields. Such fields may also be time dependents in which case they are called *spatial-temporal* random fields. The solution of such equations allows quantifying uncertainty in the flow response. Uncertainty emerges from the fact that spatial distribution of hydraulic conductivities in the flow domain cannot be fully determined. The issue of paramount importance in this approach is not the accurate solution of the stochastic flow equations, but the assignment of hydraulic conductivity values to a flow model and the quantification of uncertainty in the flow response.

Different methods have been developed in order to solve stochastic equations of flow. Among them: 1) Perturbation based methods, 2) Spectral based methods and 3) Monte Carlo based methods. A review of them can be found in Zhang (2002); Lu and Zhang (2004). Variants of perturbation based methods have been developed aimed at improving the computational efficiency of the method (Gainis *et al.,* 2008). The perturbation based methods consist basically of evaluating changes in the response due to variations in the conductivity field near its expected value. In a spectral approach the hydraulic conductivity random field is represented in a functional space by a *Karhunen-Loeve* development and the response field by a *Polynomial chaos* development.

In the simulation or Monte Carlo technique, a particular realization of the hydraulic conductivity random field is simulated first in order to solve a deterministic flow problem numerically by finite differences or finite elements. This process is then repeated a number of times until a sufficiently accurate statistic trend can be defined concerning the variability of the response. This is therefore a quite time-consuming approach in spite of the increasing capacity of modern computers and introduction of variance reduction methods (Curtis, 1949; Yamasaki *et al.,* 1988; Araujo and Awruch, 1994). A simulation approach is often preferred over a perturbation or a spectral approach because the later are just approximates based on the first two moments of the field assuming implicitly that the field is multi-Gaussian and because they provide smooth solutions.

Simulation or Monte Carlo technique for solving seepage problems (Fig. 1) involves three main steps: 1) Generating a large enough number of realizations of the random field adopted for representing heterogeneity; 2) Solving deterministic equations of flow for each one of the realizations in the ensemble and 3) Quantifying uncertainty in the dependent variables of flow through a statistical analysis.

STEP 1	STEP 2	STEP 3
GENERATING REALIZATIONS OF THE RANDOM FIELD	SOLVING EQUATIONS OF THE SEEPAGE MODEL	QUANTIFYING UNCERTAINTY IN THE MODEL RESPONSE
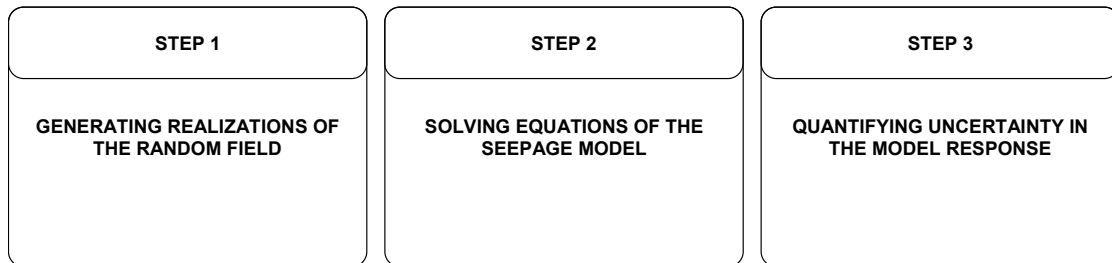
Fig. 1.1 Main steps involved in a stochastic approach based on simulations.

In a stochastic approach, the multi-Gaussian random field is the most common model adopted for representing spatial variability in hydraulic conductivity. A discussion of its suitability to such representation is the main purpose of this chapter. The discussion is based on the study of its spatial dependence characteristics. First, the concept of random field is formally introduced and the particular case of a multi-Gaussian random field is presented. The necessity of considering non multi-Gaussian dependence for representing heterogeneous hydraulic conductivity fields is then established. The chapter finalizes with the main conclusions of the subject.

1.1 The concept of random field

A review of the main definitions involved in the concept of random field is presented in this section.

1.1.1 Definitions

A *random function* is an indexed collection of random variables. When the index is a multi-dimensional spatial one, a random function is also called *random field*. It is in fact a probabilistic model which represents phenomena that vary in space like the distribution of hydraulic conductivity values in a region of interest. In a random field, every point over space is associated to a simple random variable so, globally it is a set of random variables. At each point in space, corresponds not just one value for a property but a whole set of values. From a probabilistic point of view, the value observed at a particular location is regarded as one sample from its probability distribution. A specific configuration of values in space is called a *realization* of the random field. A random field can thus also be considered as an *ensemble* of realizations.

1.1.2 Spatial structure

Spatial structure of a random field can be described by *autocovariance functions*. An autocovariance function expresses linear dependence among random variables separated by a given vector. A standardized autocovariance function called *correlogram* can also be used for such purposes. The correlogram describes the dependence by a linear correlation coefficient, formally the *Pearson correlation coefficient*. In geostatistics, especially in mining applications, the concept of *variogram* (variance of the incremental field) is often used. Autocovariance function and variogram in random fields may vary along different directions. In this case, the random field is said to be *anisotropic*.

In practice, the autocovariance function is estimated from observations as part of a process known as *structural analysis* of the field. In this context the function receives the name of *sample* or *experimental* autocovariance. For random field modeling purposes, a theoretical covariance model has to be adapted to its experimental version. A theoretical covariance model is said to be valid when the variance-covariance matrix of the random field yields positive definite. This must also be true for any linear combination of such valid models. The most common valid covariance models used in practice are (Deutsch and Journel, 1998): exponential[1], spherical[2], Gaussian[3] and nugget effect[4] models. They can be specified through parameters such as *variance* and *correlation scale*, except for the nugget effect model whose correlation range is nil. It is used to represent uncertainty affecting measurements.

The correlation range is the distance at which the correlation value in the correlogram reaches zero. However, asymptotic covariance models such as exponential and Gaussian models reach the value of zero correlation at infinity. In these cases a practical definition of the correlation range, sometimes called *effective range*, has been proposed (Deutsch and Journel, 1992). It is defined as the distance at which 95% of the correlation has been lost.

1.1.3 Stationarity and ergodicity

A random field is said to be *strictly stationary* within the region *A*, if its *n-variate* probability distribution functions are invariant under any translation in such space for any value of *n*. The random field is said to be *stationary of order N* when all *n-variate* probability distributions are invariant under any translation in the space *A* for $n \leq N$. If the expected value is constant throughout the domain and the covariance function depends only on the separation distance between points rather than their actual locations, the random field is said to be *wide sense*

---

[1] $C_{exp}(\mathbf{h}) = \sigma^2 exp(-3\mathbf{h}/a)$
[2] $C_{sph}(\mathbf{h}) = \sigma^2(1.5\mathbf{h}/a - 0.5(\mathbf{h}/a)^3)$ if $\mathbf{h} \leq a$ and $C_{sph}(\mathbf{h}) = \sigma^2$ if $\mathbf{h} \geq a$
[3] $C_{gs}(\mathbf{h}) = \sigma^2 exp(-3\mathbf{h}^2/a^2)$
[4] $C_{nug}(\mathbf{h}) = \sigma^2$ if $\mathbf{h} = 0$ and $C_{nug}(\mathbf{h}) = 0$ if $\mathbf{h} > 0$
where: σ^2; is the variance of the field and *a* is a correlation scale parameter.

stationary. A second order stationary random field is also *wide sense stationary*, but the opposite is not necessarily true because the first two moments may not reveal any information about the bivariate probability distribution, except in the Gaussian case.

In the special case when the statistics of the random field can be obtained from a unique realization, the field is said to be *ergodic*. In other words, *ergodicity* lets the statistical descriptors of the random field to be estimated from *space averages* rather than *ensemble averages*.

1.2 The multi-Gaussian random field

A multi-Gaussian random field is, by far, the most used continuous random function model for representing heterogeneity in hydraulic conductivity. This section introduces its formal definition as well as the reasons for its justification in practical situations.

1.2.1 Definition

A random field is called *multi-Gaussian* if and only if the random variables are jointly Gaussian for any set of *n* points. Therefore, a marginal Gaussian distribution is a necessary but not a sufficient condition for the field to be multi-Gaussian. Further necessary conditions are that the multivariate probability distributions, namely bivariate, trivariate, …, *n-variate* to be also Gaussian.

1.2.2 Practical justification

In practical situations a multi-Gaussian random field model is often assumed to be adequate for modeling uncertainty due to spatial variability. This is generally based on some degree of wide sense stationary behavior of the available observations of the field. But, as a matter of fact, available measurements very often prevent inferring statistical moments higher than those of order two. Therefore, most of the times, a multi-Gaussian random field model cannot really be justified in practice.

In the case of saturated hydraulic conductivities, it has been shown that the histogram can be sufficiently approximated by a lognormal probability distribution. This has been verified from the study of different geologic formations (e.g. Law (1944); Davis (1969); Hoeksema and Kitanidis, (1985); Fogg (1986); Woodbury and Sudicky (1991), López-Acosta and Auvinet, 2011). Therefore, the following transform is commonly applied to saturated hydraulic conductivity k_s:

$$y = \ln\left(k_s\right) \tag{1.1}$$

The empirical univariate distribution of the values of *y* can then be described by a Gaussian (normal) distribution.

However, a marginal Gaussian distribution does not imply a multi-Gaussian distribution. This fact is often overlooked in numerical studies of flow in random media (e.g. Griffiths and Fenton, 1993, 1997; Fenton and Griffiths, 1997; Gui *et al.,* 2000; López-Acosta and Auvinet, 2003, 2004; Ahmed, 2009). In theses cases, the interest is focused only on two subjects: 1) The simulation method of the random field and 2) The method for solving the stochastic equations of flow. The suitability of the random field model used to represent spatial variability of hydraulic conductivity is often unattended. The characteristics of the spatial structure of a multi-Gaussian random field are discussed next.

1.3 Characteristics of the multi-Gaussian dependence

The characteristics of the spatial structure or dependence in multi-Gaussian random fields are studied in this section. The discussion is limited to the bivariate case for practical reasons. The characteristics that are analyzed are: correlation structure, symmetric property, entropy property and connectivity property.

1.3.1 Correlation

The linear correlation coefficient is not invariant to nonlinear transforms of the type of Eq. 1.1. The correlation coefficient $\rho_{ln}(\cdot)$ of a pair of log-normal random variables (Z_1, Z_2) is given by (Mood and Graybill, 1963):

$$\rho_{\ln}(Z_1, Z_2) = \frac{e^{\rho_n \sigma_n^2} - 1}{e^{\sigma_n^2} - 1}; \qquad \in (-1; +1) \qquad (1.2)$$

where σ_n^2 and ρ_n are the variance and linear correlation coefficient of the associated normal random variables, respectively. A plot of Eq. 1.2 is shown in Fig. 1.2. Only the positive branch of the correlation is shown. A dash line with slope 1:1 is also displayed for reference purposes. It is observed that the linear correlation coefficient is not the same between variables of such distributions except at zero and one. Therefore, the correlograms of k_s and y in eq. 1.1 are different. The differences in correlations become significant as variance increases, that is, when the heterogeneity is stronger.
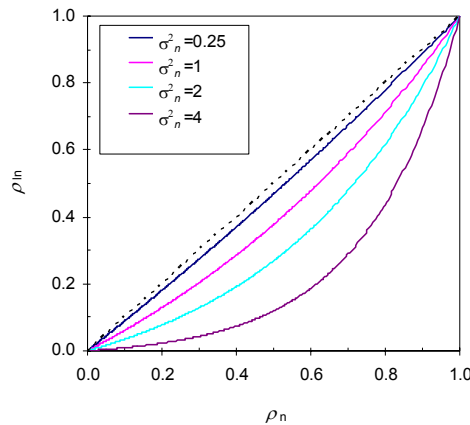
Fig. 1.2 Linear correlation coefficient of log-normal random variables as a function of the linear correlation coefficient of normal random variables with distinct variances. A dash line with slope 1:1 is shown for reference.

1.3.2 Symmetry

In multi-Gaussian random fields, the autocovariance function uniquely defines the spatial structure of all classes[5]. In order to illustrate this fact, consider the following definitions. To any random field $Z(\mathbf{x})$ corresponds a series of indicator random field transforms $I(\mathbf{x},z)$ defined for each cutoff value z as: $I(\mathbf{x},z)=1$ if $Z(\mathbf{x}) \leq z$ or $I(\mathbf{x},z)=0$ otherwise.

---

[5] A class is defined as an interval of values in the marginal distribution of the field.

The decomposition of the random field $Z(\mathbf{x})$ into a series of indicator random fields $I(\mathbf{x};z)$ allows assigning a specific spatial structure to each class of $z(\mathbf{x})$ values. Such indicator random fields are viewed as *co-existents* binary random fields.

The set of *indicator covariances* for a bivariate Gaussian random field has been derived analytically. The covariance $C(\mathbf{h};z)$ of the indicator $I(\mathbf{x};z)$, called indicator covariance, of a standard bivariate Gaussian random field $Z(\mathbf{x})$ is related to the correlogram $\rho(\mathbf{h})$ of $Z(\mathbf{x})$ by (Abramovitz and Stegun, 1964; Chilès and Delfiner, 1999):

$$C(\mathbf{h};z)=\frac{1}{2\pi}\int_{0}^{\rho(\mathbf{h})}\exp\left(-\frac{z^2}{1+u}\right)\frac{du}{\sqrt{1-u^2}} \tag{1.3}$$

Solutions of Eq. 1.3 can be obtained for different correlograms $\rho(\mathbf{h})$, namely: exponential, spherical and Gaussian correlograms and as a function of the *q-quantile* defined by $q=G(z)$, where $G(z)$ is the cumulative Gaussian distribution function and z is the q-quantile of Z. These solutions are plotted in Fig. 1.3 and in Fig. 1.4, respectively. The correlation scale in all correlogramas is unity. Fig. 1.3 expresses results in terms of a standardized indicator covariance called *indicator correlogram*. This indicator correlogram is defined as $\rho_I(\mathbf{h};z)=C(\mathbf{h};z)/C(0;z)$; where: $C(0;z)=q(1-q)$, is the variance of the indicator random field.
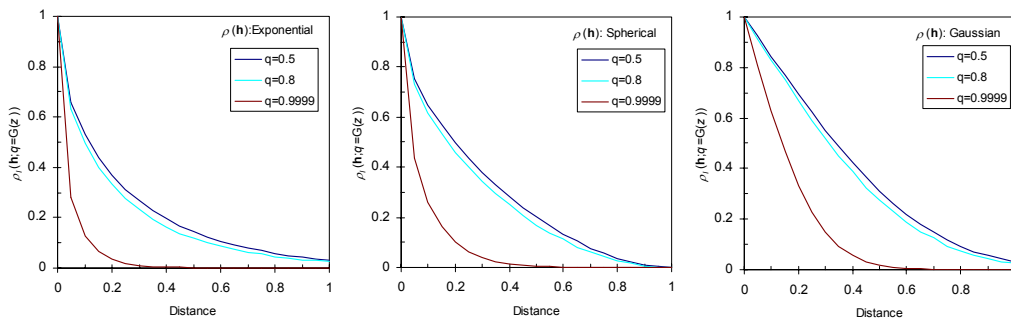
Fig. 1.3 Indicator correlograms of a bivariate Gaussian random field $Z(\mathbf{x})$ as a function of the distance for an exponential correlogram (left), a spherical correlogram (middle) and a Gaussian correlogram (right).
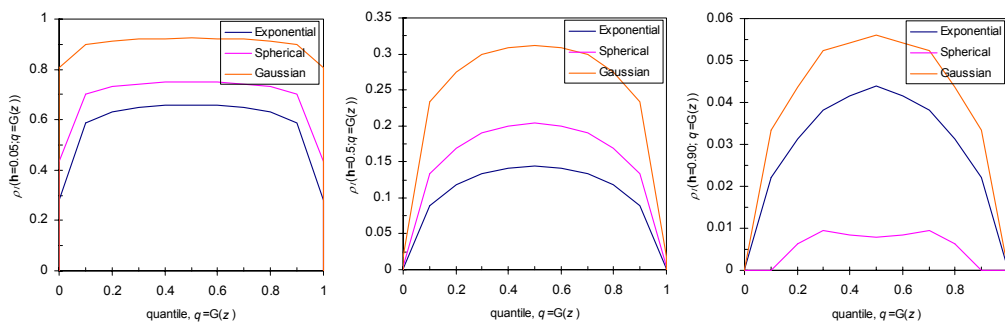
Fig. 1.4 Indicator correlograms of a bivariate Gaussian random field $Z(\mathbf{x})$ as a function of the q-quantile associated to the threshold z at $\mathbf{h}$=0.05 (left), $\mathbf{h}$=0.5 (middle) and $\mathbf{h}$=0.9 (right).

Fig. 1.4 exhibits indicator correlogramas as a function of the quantiles q associated to the threshold z for different distances. It can be observed that correlation is at a maximum for the

6

0.5 quantile and decreases symmetrically as the threshold approaches to its extreme values of either 0 or 1 for all correlogram functions of $Z(\mathbf{x})$ and for all distances. This means that the spatial structure at extreme values tend to a nugget effect (white noise) while mean values present the strongest correlation. Note that the spatial structure of the different classes cannot be modified introducing different correlograms of $Z(\mathbf{x})$, since this structure is defined by the bi-Gaussian probability law itself.

Figure 1.5 presents values of the indicator correlogram $\rho_I(\mathbf{h};z)$ at 0.5, 0.8 and 0.9999 *q-quantiles*, now as a function of the correlogram of $Z(\mathbf{x})$. Note that each value of $\rho(\mathbf{h})$ can be associated with some distance. For example, at $\mathbf{h}=0$ the correlation is the strongest one, so $\rho(\mathbf{h})=1$ whereas it becomes weaker as the distance increases. Recall that 0.2 and 0.0001 are the complementary *q-quantiles* of 0.8 and 0.9999, respectively, so the curves at 0.0001 and 0.9999 coincide. The same behavior occurs at the quantiles 0.2 and 0.8. From all curves it is evident that correlation at both low and high quantiles (0.0001 and 0.9999) is significantly weaker than at the median for all distances. Such behavior is independent of the correlation function. Therefore, extreme values (both low and high) in Gaussian random fields occur randomly within a more continuous mean values structure.
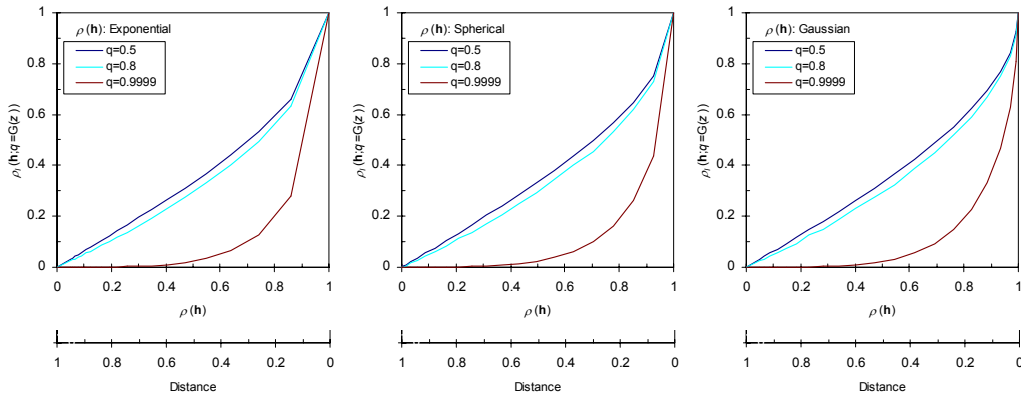
Fig. 1.5 Indicator correlograms of a bivariate Gaussian random field $Z(\mathbf{x})$ as a function of the correlogram of $Z(\mathbf{x})$. For the exponential (left), spherical (middle) and Gaussian (right) correlograms.

1.3.3 Entropy

The term *entropy* was originally introduced in physics to study the implications of the amount of *"disorder"* of a system. Different definitions have been derived since then depending on the discipline. The formally called *Shannon entropy* concept derived by Shannon (1948) as part of the *information theory* has been adopted in the context of random fields (Christakos, 1990; Ababou *et al.,* 1992; Journel and Deutsch, 1993). In information theory, entropy was introduced as a measure to quantify the *"amount of uncertainty"* contained in a *prior* discrete probability distribution, so both terms *"entropy"* and *"uncertainty"* are synonymous (Jaynes, 1957). The extension of the concept to the case of continuous distributions was also performed but the term *differential entropy* instead of *Shannon entropy* is sometimes used.

For a continuous random field $Z(\mathbf{x})$ with bivariate probability density function (PDF) $f(z_1,z_2)$, entropy H_f, is defined as (Shannon, 1948):

$$H_f = -\iint f(z_1, z_2) \log f(z_1, z_2) dz_1 dz_2 \qquad (1.4)$$

For statistical inferences on the basis of partial information it is argued by information theory that one must use that *prior* probability distribution which has maximum entropy subjected to the information that is known. In order to obtain such prior PDF the solution of a maximization problem is considered. According to Journel and Deutsch (1993), maximizing Eq. 1.4 amounts to choose the prior PDF that is maximally uncommitted with respect to unknown information. In other words, it amounts to choose the most conservative prior PDF in view of what is known.

This view has been used to develop a Bayesian/maximum entropy approach to address the spatial estimation problem (Christakos, 1990). A maximum entropy approach has been used in the past in order to derive some analytical results. For example, it has been proved that among all bivariate PDF's sharing the same expectation and covariance function, the prior PDF that maximize entropy in Eq. 1.4 is the bivariate Gaussian distribution (Shannon, 1948; Johnson and Kotz, 1972; Kapur, 1989; Cover and Thomas, 1991). In the multivariate case the corresponding analogous result holds (Christakos, 1990). Practical consequences of the entropy property of a multi-Gaussian random field are closely related to the notion of connectivity. They are discussed in the sequel.

1.3.4 Connectivity

In hydrology the term *connectivity* is used to describe the physical presence of spatially connected zones with similar values of the hydraulic conductivity (Kundby and Carrera, 2005). Particularly relevant is the connectivity of either low or high values because they control the presence of flow paths and flow barriers. Related concepts such as "channeling" and "preferential paths" are used to describe the consequences of connectivity of high values on the flow response. Connectivity and continuity as described by covariance functions are not the same concept. Connectivity attempts to characterize interconnected low/high conductivities "channel-type" paths, whereas continuity describes averaged "discontinuous lens-type" paths.

Different measures have been proposed in order to quantify spatial connectivity (point to point) in random fields, namely: 1) Topological descriptors for random geometric sets such as the Euler characteristic (Mecke and Wagner, 1991; Vögel, 2002); 2) Measures of percolation probability such as the percolation threshold (Hilfer, 1992, 1997); 3) The multi-fractal correlation dimension (Bruderer-Weng *et al.*, 2004) and 4) Some intuitive indicators of flow and transport connectivity such as effective permeability and certain typical flow and transport parameters (Kundby and Carrera, 2005; Trinchero *et al.*, 2008; Vassena *et al.,* 2009).

To illustrate graphically the concept of connectivity, the Euler number $\chi(p)$ is considered in this section. This is a topological measure of binary structures which are obtained after segmentation of a heterogeneous structure using various thresholds p. Its mathematical definition in two dimensions is (Vögel, 2002):

$$\chi(p) = N - C \tag{1.5}$$

It is a scalar whose value is the total number of isolated objects N at threshold p in the binary image minus the total number of holes C in those objects. Hence, $\chi(p)$ has positive values at $p,$ for disconnected structures ($N>C$) and becomes negative ($N<C$) for objects which are more intensely connected.

Fig. 1.6(a) shows a typical realization of a multi-Gaussian random field with isotropic exponential autocovariance function and correlation scale $\approx 1/13$ of the side of the image.

An indicator function is defined such that: $i(\mathbf{x};z)=1$ if $z(\mathbf{x}) \leq z$ (black phase) and 0 otherwise (white phase) where z is an arbitrary threshold defined over the empirical cumulative distribution function $\hat{F}_n(z)$ (ECDF) of the field $z(\mathbf{x})$, such that: $q = \hat{F}_n(z)$ with: $q \in [0,1]$ is the

q-quantile derived from the ECDF. For every threshold *q* a binary image can be obtained by separating the values (phases) below and above the threshold. Then the Euler number is calculated at each threshold.
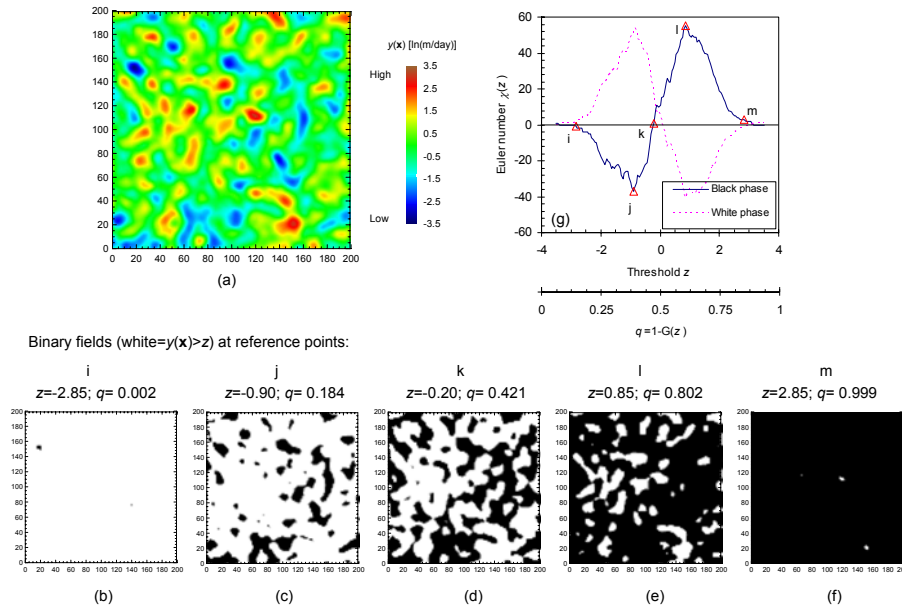
Fig. 1.6 Connectivity of a representative realization of a multi-Gaussian random field with Gaussian auto-covariance function as expressed by the Euler number. a) Realization. b)-f) Binary images at different thresholds. g) Euler functions of the realization.

For example, for a quantile such as $q=0.002$, the lowest values represented in black on Fig 1.6(b) appear as isolated clusters; that is they are not connected and the Euler number takes negative values (point (i) on Fig. 1.6(g)). As the threshold increases, the clusters get closer but they remain unconnected and the Euler number reaches its minimum value (point (j) on Fig. 1.6(g)). Further increments of the threshold create connected isolated zones and the Euler number increases accordingly. When it reaches zero (point (k) on Fig. 1.6(g)) the high values (black phase) becomes well connected (Fig. 1.6(d)), but a continuous high conductive zone spanning two opposite sides of the field can be detected until $q=0.5$. As the threshold keeps increasing, high values become intensely connected thus the Euler number reaches its maximum (point (l) on Fig. 1.6(g)). When the threshold increases even more the highest values (white phase) appear now as isolated clusters, for example, at $q=0.802$ (Fig. 1.6(e)), and the Euler number begins to decrease. For $q=1$ the total number of objects is equal to one and the number of connections is zero and the Euler number tends to unity. An analogous study can be made by analyzing the connectivity behavior of the white phase first (discontinuous line in Fig. 1.6(g)).

The previous connectivity analysis shows that connected paths of extreme values either low or high cannot occur in multi-Gaussian random fields. Values around the median of its univariate distribution present significant spatial connectivity. Similar behaviors were detected previously by other authors at analyzing percolation paths in multi-Gaussian realizations for different values of the relation side of the domain/correlation scale (Allard, 1994). Therefore, it could be anticipated that in isotropic heterogeneous multi-Gaussian media the average flow response will be so close to the flow response of a homogeneous medium with conductivity given by the average of the heterogeneous conductivities.

1.4 Necessity of non multi-Gaussian dependence

Field evidence suggests that natural soil formations are seldom multi-Gaussian but rather present some connected paths of high or low conductivity values that significantly impact the flow response. This section presents such evidence separately, firstly based on actual case studies and then based on numerical modeling. The reader may refer to appendix A for some useful definitions.

1.4.1 Evidence based on actual case studies

Direct measurements and samples collected at different soil formations have shown that the dependence at low quantiles often differs from that at high quantiles of the distribution function. At larger scales, like the scale of hydro-geological facies, comparative analysis between the observed actual behavior and the numerical model suggest the necessity of a better description of the structure of spatial variability of the hydraulic conductivity field.

For example, Journel and Alabert (1989) observed different patterns of spatial correlation at different quantiles of the distribution of air permeability measurements taken on a vertical slab of Berea sandstone. Such behavior corresponds to an asymmetric correlation structure between low and high values. A similar behavior has been observed with hydraulic conductivities. For example, Haslauer *et al.,* (2009) report non-Gaussian dependence structures of core samples of hydraulic conductivities obtained from two different sites. In one of their analysis they used the same data sets used by Sudicky (1986) and Woodbury and Sudicky (1991) which contain hydraulic conductivity measurements of the Borden aquifer.

Furthermore, Fogg (1986) carried out a numerical three-dimensional study of the hydraulic conductivity of a real aquifer consisting of thick complex sequences of sand, silt and clay. His conceptual model was based on interpretative geologic and hydrologic information. He concluded that the flow system in the aquifer was mainly controlled by the continuity and interconnectedness of the sand, rather than by their local conductivity values. Other studies on distinct natural soil formations based on different models for heterogeneity representation agree with Fogg's (1986) conclusions. A general consensus is that the large-scale behavior of natural soil formations is controlled by the way high and low conductive zones are spatially interconnected over the flow domain (Bradbury and Muldoon, 1990; Poeter and Townsend, 1994; LaBolle and Fogg, 2001; Frind *et al.,* 2002; Teles *et al.,* 2004; Zappa *et al.,* 2006; López-Acosta, 2010).

1.4.2 Evidence based on numerical modeling

The necessity of non multi-Gaussian dependence to describe spatial variability of the hydraulic conductivity has also been established through numerical modeling. Several studies have shown that the multi-Gaussian model is not necessarily a conservative assumption. The presence of spatially connected zones of extreme values (either low or high) is the most consequential feature in hydraulic conductivity fields. Most of the studies have analyzed the flow response on different non multi-Gaussian random fields and the responses are then compared to the multi-Gaussian response. The aim is to compare the flow response only in terms of differences on the characteristics of the spatial dependence.

For example, Sanchez-Vila *et al.,* (1996) analyzed steady state, saturated flow in synthetic log-transmissivity random fields with different non multi-Gaussian dependence but sharing normal marginal distributions. The study was carried out in 2D sections under confined boundary conditions by Monte Carlo simulations. They showed that structures favoring increased connectivity of high values consistently yield values of effective conductivities higher than the geometric mean. The differences became higher as the variance increased. For the same flow conditions and solution scheme, Vögel (2002) observed effective hydraulic conductivity values

by a factor 71.5 higher than the geometric mean in non multi-Gaussian fields exhibiting increased connectivity of high values.

In other numerical experiment, Zinn and Harvey (2003) studied groundwater flow and transport in 2D fields. They compared the response among multi-Gaussian and non multi-Gaussian hydraulic conductivity random fields with nearly identical marginal and isotropic covariance functions through Monte Carlo simulations. Their non multi-Gaussian realizations exhibited two particular features: a) Inclusions of low values in a connected high-values background and b) Inclusions of high values in a connected low-values background. As far as to flow concern, the field with connected high-conductivity paths shown an effective conductivity greater than the geometric mean and large variations in specific discharges. In the connected low-conductivity fields, the effective conductivity was less than the geometric mean and the velocity variance was smaller. In both type of conductivity fields, differences in the flow response with respect to the multi-Gaussian case became higher as the variance of the field increased.

Nowak *et al.,* (2008) determined the probability density functions of dependent variables of flow such as hydraulic head and specific discharge. They solved steady state flow in 3D under confined boundary conditions. Heterogeneity in hydraulic conductivity was accounted for through random field models. The study was carried out on the basis of Zinn and Harvey's (2003) comparative scheme by extensive Monte Carlo simulations but now the multi-Gaussian and non multi-Gaussian hydraulic conductivity random fields shared marginal distributions and anisotropic covariance functions. They conclude that, altogether; the connected highly permeable zones seem to adversely affect the fitting of the curves which matched well in the multi-Gaussian case. For the connected low conductivity case they found that the mean total discharge does not significantly differ from the solution in multi-Gaussian fields; moreover the fitted curves followed better the multi-Gaussian case.

Additionally, Journel and Deutsch (1993) analyzed the frequency distributions of effective hydraulic conductivities and times to achieve a 90% water cut from a waterflood numerical experiment. The study was carried out over a 2D section under confined boundary conditions including an inflow and an outflow by simulations. The histogram, covariance function and seven indicator covariance functions were extracted from a synthetic reference field. Then, the information was used to characterize three random fields with different spatial dependences. The response obtained from the synthetic formation was compared against the mean response from the different random fields. All three random fields yielded response distributions whose centers deviate considerably from the reference values. More importantly, the multi-Gaussian model yielded an overoptimistic assessment of uncertainty of such prediction entailing thus a misleading sense of safety.

Comment:

- One can also debate the representativeness of the multi-Gaussian model to represent heterogeneity of hydraulic conductivity in stabilized soils. For example, the presence of wetter bands than the desired average embedded in dryer layers facilitates the creation of continuous zones of high hydraulic conductivity values. Such continuous, highly conductive zones not only may alter the seepage behavior of the structure but also, if combined with certain factors, its safety against an eventual internal erosion event. This phenomenon has been the main cause of failure in this kind of structures over years (ICOLD*,* 1995; 1997). The multi-Gaussian model therefore seems to be no suitable for seepage assessments even in this kind of apparently "homogeneous" structures.

1.6 Conclusions

In the study of flow in random media, the multi-Gaussian random field is a model very commonly used to represent heterogeneity in hydraulic conductivity. Adopting such model is often justified in practice based only on a log-transform normal histogram. A multivariate Gaussian distribution is then "sightlessly" assumed often disregarding the characteristics of multi-Gaussian dependence. Such characteristics were discussed here by analyzing the bivariate case.

Multi-Gaussian random fields exhibit symmetric dependence structure. Both high and low values present exactly the same spatial structure. The highest continuity is presented at mean values (symmetry axis). This property assumes that in nature or man made earth works both high and low conductive zones are created for equal. The Pearson correlation coefficient as expressed by correlogram functions is not preserved by non linear transforms (for example taking the logarithm) in the general case.

The maximum entropy property of the multi-Gaussian random field does not imply maximum uncertainty of the flow response. Such conclusion is based on the fact that with a single autocovariance function there is no way to enhance the organization of neither high nor low values. Nature however presents certain spatial structure or organization for a certain class of values. As a result, a multi-Gaussian random field may not include real spatial features of the hydraulic conductivity fields like better continuity of high values from which a higher uncertainty in the flow response could arise.

Connected paths of extreme values either low or high cannot occur in multi-Gaussian random fields. Values around the median of their univariate distributions present significant spatial connectivity. A general consensus is however that the large-scale flow behavior of natural soil formations is controlled by the way high and low conductive zones are spatially interconnected over the flow domain. In other words, the presence of flow paths and flow barriers controls the flow response. Therefore, multi-Gaussian random fields of the hydraulic conductivity of soils seem to provide a poor representation of reality.

CHAPTER 2

SIMULATION OF NON MULTI-GAUSSIAN RANDOM FIELDS BY COPULAS

The previous chapter established that, to conveniently represent patterns of spatial variability in hydraulic conductivity fields, random fields with non multi-Gaussian dependence characteristics ought to be considered. These kinds of random fields will permit to incorporate a higher degree of realism in descriptions of the heterogeneity of porous media. This chapter proposes an algorithm for the simulation of such fields.

Stochastic simulation of random fields is the process by which alternative, independent configurations or images of the spatial distribution of an attribute with physical meaning such as the hydraulic conductivity of soils are generated. Each configuration or image (often called *realization*) can then be used in numerical models of seepage to quantify uncertainty via the method of Monte Carlo. A random field that is only compatible with their descriptive parameters is called *unconditional random field*, but if the random field is also made specific to a set of location dependent measurements, this is called *conditional random field*. Conditional random fields thus allow simulating plausible spatial configurations that attempt to mimic reality (Deutsch and Journel, 1992).

Different methods can be used to simulate non multi-Gaussian random fields: spectral approaches (e.g Yamazaki and Shinozuka, 1988; Popescu *et al.,* 1998; Grigoriu, 1998), optimization methods (e.g. Srivastaba, 1995), anamorphosis approach (Journel and Huijbregts, 1978; Sanchez-Vila *et al.,* 1996), Bayesian-Maximum Entropy approach (Christakos, 1990) and multiple point geostatistical methods (Guardiano and Srivastaba, 1993; Strebelle, 2002; Journel and Zhang, 2006). The characteristics of the dependence structures achieved by all these approaches cannot be discussed here. Alternatively, a brief discussion of the main characteristics of some non multi-Gaussian random field models and of the methods used to simulate them is presented. Among the most common methods one can find: 1) Series expansions methods (e.g. Sakamoto and Ghanem, 2002a; Sakamoto and Ghanem, 2002b), 2) Multiple indicator decomposition methods (Journel, 1983; Journel and Alabert, 1989) and 3) Isofactorial representation methods (e.g. Chilès and Delfiner, 1999; Emery, 2002).

Series expansion methods for non multi-Gaussian random fields have recently become very popular due to its computational efficiency. In these methods the random field is represented in the form of a *polynomial chaos* expansion[1] which can readily be incorporated into a *spectral stochastic finite element* scheme (Ghanem and Spanos, 1991; Pineda-Contreras and Auvinet, 2013). The random function in the *non-conditional* case is constructed by making use of a *polynomial chaos* expansion to match any kind of marginal distributions while the *Karhunen-Loève* expansion is used to approximate the autocovariance function of the field. In the *conditional* case the autocovariance function is obtained first by a *Kriging* simple estimate, the conditional *eigenvalues* and *eigenfunctions* are then obtained by solving a *Fredholm* integral numerically.

At least, three main drawbacks can be found in this approach. 1) The Karhunen-Loève expansion only approximates the observed covariance. 2) More importantly, there is no control on the characteristics of the no multi-Gaussian dependence. In fact, a non multi-Gaussian random field cannot be entirely described through their first two moments. 3) Since the covariance function is used as the sole descriptor of spatial variability, specific information regarding the continuity of the values at tails of the marginal distributions cannot be incorporated. Therefore, such method seems to be not suitable for hydraulic conductivity fields in which well defined spatial features, namely flow paths and flow barriers control uncertainty in the flow response.

The multiple indicator method meanwhile makes use of the property of the autocovariance function to be decomposed in their indicators[2]. Thus specific information regarding the continuity at a series of quantiles can be incorporated. In practice, information by a set of indicator covariances is only specified. The conditional random field is generated as follows. At any unobserved location over the field, an estimate of the value of the conditional CDF at a particular quantile is obtained by a regression technique applied to an indicator transform (zero or one) of the original random variables. Such technique is called *indicator Kriging* (Deutsch and Journel, 1992). Repeating the process with various specified quantiles, a distinct version of the conditional CDF is obtained at each target point. Hence such approach provides a complete solution to the estimation problem. After the indicator coding (zero or one) the approach becomes *non-parametric*, so it is able to deal with any kind of marginal distributions.

The multiple indicator method is the more common non multi-Gaussian random field model used in geostatistics however it suffers several deficiencies. 1) The indicator covariances cannot be modeled independently of each other (Journel and Posa, 1990). Hence the apparently wide initial flexibility of the model is severely reduced since at certain thresholds some spatial structures are only roughly approximated. 2) The monotonicity of the conditional CDF at each point is not guaranteed; instead it is corrected artificially to meet order relations (Deutsch and Journel, 1992). Consequently, the local conditional CDF at each target point is only approximated.

The indicator formalism is often implemented in a sequential simulation scheme where it receives the name of *sequential indicator simulation* (Deutsch y Journel, 1992). Under such scheme the indicator formalism suffers additional insufficiencies: 1) There is no guarantee in the general case for the reproduction of the autocovariance function of the field, only of their indicator covariances, if and only if, order-relation violations do not occur. 2) The multivariate distributions of the realizations have to be considered as undefined since they depend on factors such as total number of simulated nodes and number and location of the samples (Emery, 2004, 2005), that is, the realizations do not refer to a fully specified random function model therefore their multiple point statistics depend on implementation factors. Hence, incomplete and

---

[1] A polynomial in which the variables are uncorrelated Gaussian random variables.
[2] Covariances and cross-covariances.

approximate representations of bivariate spatial variability as well as undefined multiple point statistics would deliver rough configurations of spatial patterns for seepage modeling purposes.

Instead of only approximating the bivariate dependence by multiple indicators, the isofactorial representation models the entire bivariate probability law via its isofactorial bivariate distribution[3]. Several families of isofactorial distributions exist, each one of them including a series of models able to represent diverse characteristics of the dependence. The *Laguerrian* model for example allows representing asymmetric dependence structures. Such model is an isofactorial representation of the bivariate Gamma distribution (Chilès and Delfiner, 1999) whose factors are given by the normalized *Laguerre* polynomials of order α (Chilès and Delfiner, 1999), where scalar α is the so-called "shape parameter" of the standard univariate Gamma distribution. The factor covariances are randomized in this model and written as function of the correlogram and a positive scalar β. Although a single autocovariance function is used to specify the covariances of all the factors of the bivariate law, the asymmetry of the dependence is controlled by the scalar parameters α and β.

The conditional random function of an isofactorial representation is constructed as follows. An anamorphosis is needed first in order to transform the observed variable into a second variable which follows the univariate distribution of the isofactorial model (e.g. standard Gamma distribution for the Laguerrian model). Then, the conditional probability at a certain threshold is obtained at any point from the anamorphosis function modeled via an expansion in a series of factors (Laguerre polynomials), where each factor can be computed by simple *Kriging* given the values of the same factor at the observed locations. The interpolation technique for the conditional CDF to be achieved is called *disjunctive Kriging* (Kriging of a disjunctive coding of the anamorphosis function). By disjunctive Kriging at different thresholds, the local conditional CDF at the target point can be obtained.

Isofactorial models appear to be an immediate alternative to multiple indicator models since in the former the full bivariate distribution can be taken into account in a consistent way. Furthermore, isofactorial models offer great flexibility to incorporate diverse structural patterns. Isofactorial models share however some deficiencies with multiple indicator models: 1) Disjunctive Kriging does not always give mathematically consistent estimates (Emery, 2002, 2006). For example, the monotonicity of the conditional CDF is not ensured, it still has to be corrected artificially. 2) Only the bivariate distributions of the random field are modeled, the multivariate distributions remain undefined. Therefore multiple point statistics of the realizations may also be implementation dependent at simulating the random field in a sequential scheme *(sequential isofactorial simulation)* (Emery, 2002, 2004).

To overcome the aforementioned second inconvenient, Emery (2005, 2008) formulated a "multivariate *chi-square* distribution", from random fields with bivariate Gamma distribution with half integer shape parameter α, obtained as the sum of squared independent Gaussian random fields. The multivariate dependence becomes now determined furthermore an isofactorial representation of the bivariate distribution is available. The isofactorial representation let the model to be parameterized. The spatial dependence in such model is controlled by the correlogram and the scalar α. The conditional random field is suggested to be constructed by iterative procedures such as simulated annealing (Deutsch and Cockerham, 1994; Deutsch and Journel, 1994) and Gibbs sampler techniques (Geman and Geman, 1984; Casella and George, 1992), avoiding hence the use of disjunctive Kriging due to their often inconsistencies.

It can be shown that the multivariate *chi-square* distribution of Emery (2005, 2008) is in fact a particular case of a more flexible multivariate distribution achieved by copulas. In addition,

---

[3] A factorized joint probability density function expressed in terms of a set of orthonormal functions for $L^2(R,f)$ (Hilbert space of functions that are square-integrable with respect to the measure given by $f(\cdot)$)

models based on copulas are exempt of the aforementioned deficiencies of the Kriging type techniques. Some copula models are presented in this chapter. The exposition of the subject is divided in four main sections. Section one presents a formal introduction to copulas. Section two introduces the concept of spatial copulas as well as the spatial copula modeling process. Section three discusses the subject of non conditional simulation while section four deals with conditional simulation. To illustrate the different subjects discussed, examples over all different sections are presented.

2.1 Definitions

Formally, a copula is a function $C(\cdot)$ on the n-dimensional unit cube i.e., $C:[0,1]^n \rightarrow [0,1]$ that links (couple) a multivariate distribution function $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$ to their marginal distribution functions $F_{X_1}(x_1),\ldots,F_{X_n}(x_n)$. Its mathematical expression is (Sklar, 1959):

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = C\left(F_{X_1}(x_1),\ldots,F_{X_n}(x_n)\right) \tag{2.1}$$

If $F_{X_1},\ldots,F_{X_n}$ are all continuous, then $C(\cdot)$ is unique. Moreover, if $F_{X_1}^{-1},\ldots,F_{X_n}^{-1}$ are the inverse distribution functions of $F_{X_1},\ldots,F_{X_n}$, then:

$$C\left(u_1 = F_{X_1}(x_1),\ldots,u_n = F_{X_n}(x_n)\right) = F_{X_1,\ldots,X_n}\left(F_{X_1}^{-1}(u_1),\ldots,F_{X_n}^{-1}(u_n)\right) \tag{2.2}$$

Since: $U_1 = F_{X_1}(x_1),\ldots,U_n = F_{X_n}(x_n)$ are all uniform RV on $[0,1]$ (Rosenblatt, 1952), a copula is itself a multivariate distribution function with uniform marginals.

Several properties can be derived from copulas (Nelsen, 2006). For example, if $C(\cdot)$ is absolutely continuous, from the definition: $f_X(x) = d/dx(F_X(x))$; it is shows that the copula density $c(\cdot)$ is:

$$c(u_1,\ldots,u_n) = \frac{f_{X_1,\ldots X_n}\left(F_{X_1}^{-1}(u_1),\ldots,F_{X_n}^{-1}(u_n)\right)}{\prod_{i=1}^{n} f_{X_i}\left(F_{X_i}^{-1}(u_i)\right)} \tag{2.3}$$

where $f_{X_1,\ldots,X_n}(\cdot)$ denotes the multivariate density function corresponding to $F_{X_i,\ldots,X_n}(\cdot)$ and $f_{X_i}(\cdot)$ the marginal density function corresponding to $F_{X_i}(\cdot)$ for $i = 1,\ldots,n$.

Moreover, from the definition: $f_{X|Y}(x\,|\,y) = f_{X,Y}(x,y)/f_Y(y)$; it is shown that the density of the conditional copula is:

$$c\left(u_0\,|\,u_1 = F_X(x_1),\ldots,u_n = F_X(x_n)\right) = \frac{f_{X_0,X_1,\ldots,X_n}(x_0,x_1,\ldots,x_n)}{\prod_{i=1}^{n} f_{X_i}(x_i)} \frac{1}{c(u_1,\ldots,u_n)} \tag{2.4}$$

Copulas are of interest as random functions for modeling spatial variability because they express dependence without the influence of their marginal distributions and because of the fact that copulas are invariant to strictly increasing monotonic transforms. For instance, if $Y(\mathbf{x})=ln(K_s(\mathbf{x}))$, then both $K_s(\mathbf{x})$ and $Y(\mathbf{x})$ share the same copula. As a result, correlation and asymmetry measures expressed only in terms of their copulas are also invariant.

The inspection of Eq. 2.2 shows that copulas hold at least the following properties:

- Copulas contain valuable information about the type of dependence that exists between random variables.
- Copulas express dependence between random variables in its purest or essential form, without the influence of the kind of marginal distributions.
- Copulas can be constructed from any kind of continuous marginal distributions.
- Copulas depend only on the ranks of the variables which do not change through strictly increasing monotonic transforms, such as taking the logarithm.
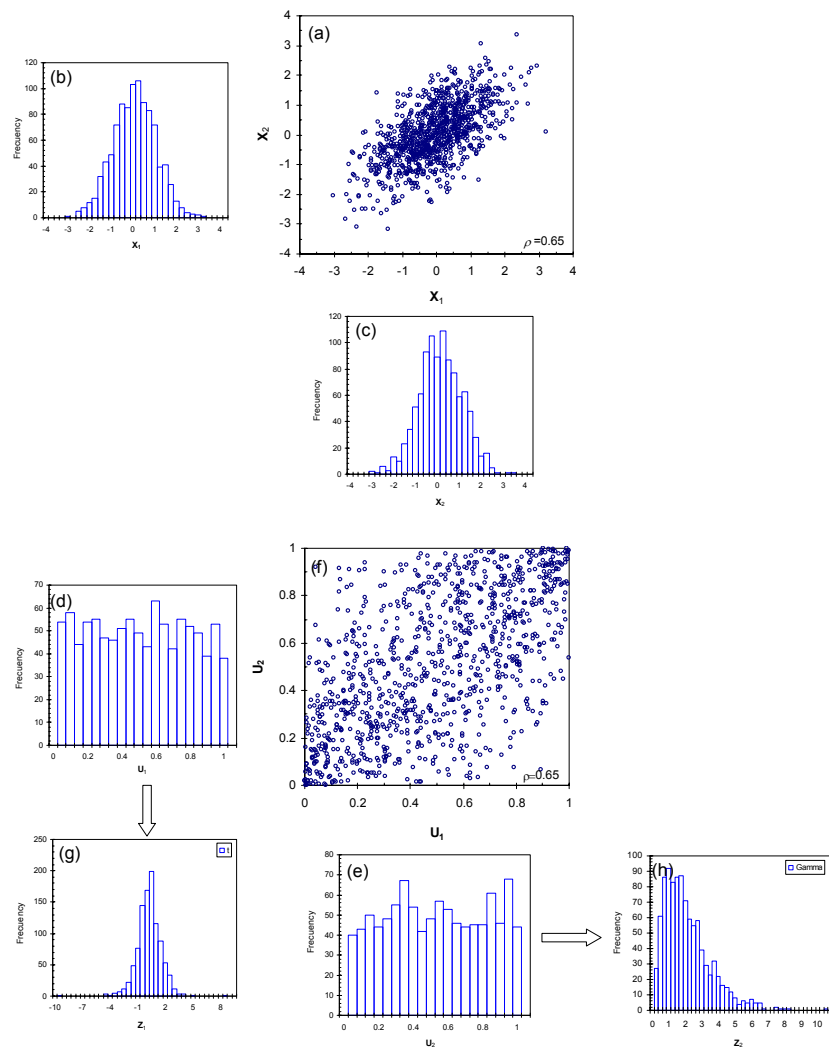
Fig. 2.1 Construction of a bivariate copula using a pair of correlated Gaussian random vectors. a) Scattergram of the correlated Gaussian random vectors. b) and c) Distribution functions of the random vectors. d) and e) Probability values of the distributions functions in b) and c), respectively. f) The bivariate copula of the random vectors $(\mathbf{X}_1,\mathbf{X}_2)$. g) and h) The distribution functions of Student's t and Gamma, that are obtained with the probability values in d) and e) respectively.

To illustrate these properties, consider the following example. Fig. 2.1(a) shows the scattergram of a pair of correlated Gaussian random vectors $(\mathbf{X}_1,\mathbf{X}_2)$ with Pearson correlation coefficient $\rho=0.65$. The distribution functions of each vector is shown to the left (Fig. 2.1b) and below

(Fig. 2.1(c)) of Fig. 2.1(a), respectively. After extracting probability values from each distribution function, a uniform distribution function is obtained for each random vector as shown in Figs. 2.1(d),(e). Every pair of such uniform distribution $(u_1,u_2) \in [0,1]$ from the vectors $(\mathbf{U}_1,\mathbf{U}_2)$, is plotted onto the unit square $[0,1]^2$ as illustrated by Fig. 2.1(f). That plot shows the set of points of a bivariate copula. The uniformly distributed distributions of each random vector can then be used to generate a pair of random vectors $(\mathbf{Z}_1,\mathbf{Z}_2)$ each one holding any kind of marginal distribution. For example, the vectors $(\mathbf{U}_1,\mathbf{U}_2)$ are used in Figs. 2.1(f),(g) to generate the vectors $(\mathbf{Z}_1,\mathbf{Z}_2)$ from the *Student's t* distribution and from the Gamma distribution, respectively.

The bivariate copula expresses the dependence between random variables independently of the kind of marginal distributions. In fact, a pair of Gaussian random vectors was used to construct the bivariate copula but any other kind of continuous distribution for the random vectors $(\mathbf{X}_1,\mathbf{X}_2)$ would have been used similarly. Moreover, under any strictly increasing monotonic transform of the variables in the vectors $(\mathbf{X}_1,\mathbf{X}_2)$ the copula would have remained invariant, since the transform $(\mathbf{X}_1,\mathbf{X}_2) \rightarrow (\mathbf{U}_1,\mathbf{U}_2)$ is in fact monotonous.

Comment:

- The correlation coefficient of the random vectors with *Student's t* and *Gamma* distribution functions in the previous example is no longer ρ=0.65. In fact, Pearson coefficient is dependent of the kind of marginal distributions of the vectors $(\mathbf{Z}_1,\mathbf{Z}_2)$ (Hoeffding, 1940; Lehmann, 1966). Therefore, invariant measures of dependence are more convenient at modeling spatial dependence by copulas, as it will be explained in section 2.3.1.

2.2 Spatial modeling

In this section, the concept of bivariate spatial copulas is presented. A method for constructing bivariate spatial copulas from a sample is also described. Once spatial copulas from a sample are determined, a multivariate copula has to be adopted for spatial variability modeling purposes. This subject is discussed in this section. For illustration purposes, the manner in which the multivariate Gaussian copula is constructed is explained first. Then the multivariate copula adopted in this research for heterogeneity representation of hydraulic conductivity fields is described in detail.

2.2.1 Bivariate spatial copulas

Consider a strictly stationary random field $\{Z(\mathbf{x}) \mid \mathbf{x} \in S\}$, where $S \in \mathfrak{R}^n$, is the domain of interest and *n*=1,2 o 3 the dimensional space. The available experiments on S are interpreted as a realization of the random field assuming ergodicity. Let F_z denote the marginal distribution functions of the random field. The stationarity of the field ensures that F_z is the same for each location $\mathbf{x} \in S$; so $F_z=F_1=F_2=\ldots=F_n$. Similar to a description with autocovariance functions, *bivariate spatial copulas* can be used to describe spatial variability. Namely, the bivariate copula C_s for any two locations separated by a vector $\mathbf{h}$ can be written via Sklar`s theorem as (Bárdossy, 2006):

$$C_S(\mathbf{h};u_1,u_2)=P[Z(\mathbf{x}) \leq z_1, Z(\mathbf{x} + \mathbf{h}) \leq z_2]$$
$$=C(P[Z(\mathbf{x}) \leq z_1], P[Z(\mathbf{x} + \mathbf{h}) \leq z_2])$$
$$=C(F_Z(Z(\mathbf{x})), F_Z(Z(\mathbf{x} + \mathbf{h}))) \qquad (2.5)$$

Hence, the copula becomes a function of the separating vector $\mathbf{h}$ or of the distance $|\mathbf{h}|$. Note that u_1 and u_2 are the quantiles of $Z(\mathbf{x})$ and $Z(\mathbf{x}+\mathbf{h})$, respectively. For a given $\mathbf{h}$ or $|\mathbf{h}|$, the spatial

copula C_S, describe thus the spatial dependence between the quantiles u_1,u_2 of pairs of random variables.

2.2.2 Bivariate empirical copula densities

To determine *bivariate empirical copulas* from a sample $\{z(\mathbf{x}_1),z(\mathbf{x}_2),...,z(\mathbf{x}_n)\}$; the empirical distribution function $\hat{F}_n(z)$, is obtained first. Then, for any given vector $\mathbf{h}$, a set of pair values $S(\mathbf{h})$, of the empirical distribution function of the investigating parameter is obtained by (Bárdossy, 2006):

$$S(\mathbf{h})=\{\hat{F}_n(z(\mathbf{x}_i)), \hat{F}_n(z(\mathbf{x}_j)) \qquad \forall\ i,j \in S \mid |\mathbf{x}_i\text{-}\mathbf{x}_j| \quad \text{or} \quad |\mathbf{x}_j\text{-}\mathbf{x}_i| \approx \mathbf{h}\} \tag{2.6}$$

$S(\mathbf{h})$ is thus a set of points in the unit square $[0,1]^2$. Note that $S(\mathbf{h})$ is by definition symmetrical regarding the major axis $u_1=u_2$ of the unit square; namely: if $(u_1,u_2)\in S(\mathbf{h})$, then $(u_2,u_1)\in S(\mathbf{h})$.

The *bivariate empirical copula density* $\hat{c}$ of $S(\mathbf{h})$ can be estimated by (e.g. Nelsen, 2006; Bárdossy, 2006):

$$\hat{c}\left[\frac{2i\text{ - }1}{2m}, \frac{2j\text{ - }1}{2m}\right] = \frac{m^2}{|S(\mathbf{h})|}\eta_{ij} \tag{2.7}$$

where η_{ij}, denotes the empirical frequency of the values corresponding to a regular mesh (mxm) with coordinates (i,j), where $i,j=1,...,m$. For a given pair (i,j), the empirical frequency η_{ij} is equal to the cardinality (number of elements) of the following set:

$$\eta_{ij}= \left|\left\{(u_1,u_2)\in S(\mathbf{h}); \frac{i-j}{m} < u_1 \leq \frac{i}{m} \text{ and } \frac{j-1}{m} < u_2 \leq \frac{j}{m}\right\}\right| \tag{2.8}$$

with $|\cdot|$ denoting the cardinality of a set. In this way a *bivariate empirical copula density* $\hat{c}(\mathbf{h},u_1,u_2)$ for a given vector $\mathbf{h}$ can be estimated.

To illustrate these concepts, consider the following example. Fig. 2.2 shows three realizations of three different random fields (RFs) which share univariate Gaussian distributions. The histogram and experimental (exhaustive) correlation functions (correlograms) of each realization are shown on top of Fig. 2.2. Images (a) and (b) are realizations of a multi-Gaussian RF with isotropic exponential and Gaussian autocovariance functions, respectively. The range in both functions is a=60 units. Image (c) is one realization of a transform-based RF introduce by Vögel, 2002 (see also Zinn and Harvey, 2003; Knudby and Carrera, 2005; Nowak *et al.,* 2008). Such transform-based RF imposes to the realizations non multi-Gaussian dependence with higher continuity for high/low values. The autocovariance function of the random field is also in this case an isotropic, exponential one with range a=60 units.

The dependence structure of each realization is examined through the assessment of their empirical bivariate copula density plots at different distances. Only horizontal vectors are considered. The plots are shown by columns below each realization in Fig. 2.2.

Empirical copulas were determined at distances $|\mathbf{h}|$=1*,* 5 and 20 units. Observe that copula's density shape of both multi-Gaussian realizations is elliptical and that tends to be lost as distance increases. These observations illustrate two interesting properties of copulas. The elliptical shape is typical of the Gaussian copula density. It is symmetric with respect to the axis u=1-v indicating symmetric dependence. The second observation illustrates that as distance approaches the correlation range the copula shape become less specific indicating nearly

independence. Comparing the copula densities of both multi-Gaussian realizations at the same distance, it is observed that the Gaussian autocovariance function imposes stronger association than the exponential one which means higher continuity of the autocovariance function. This is a well-known result in geostatistics (Deutsch and Journel, 1992).
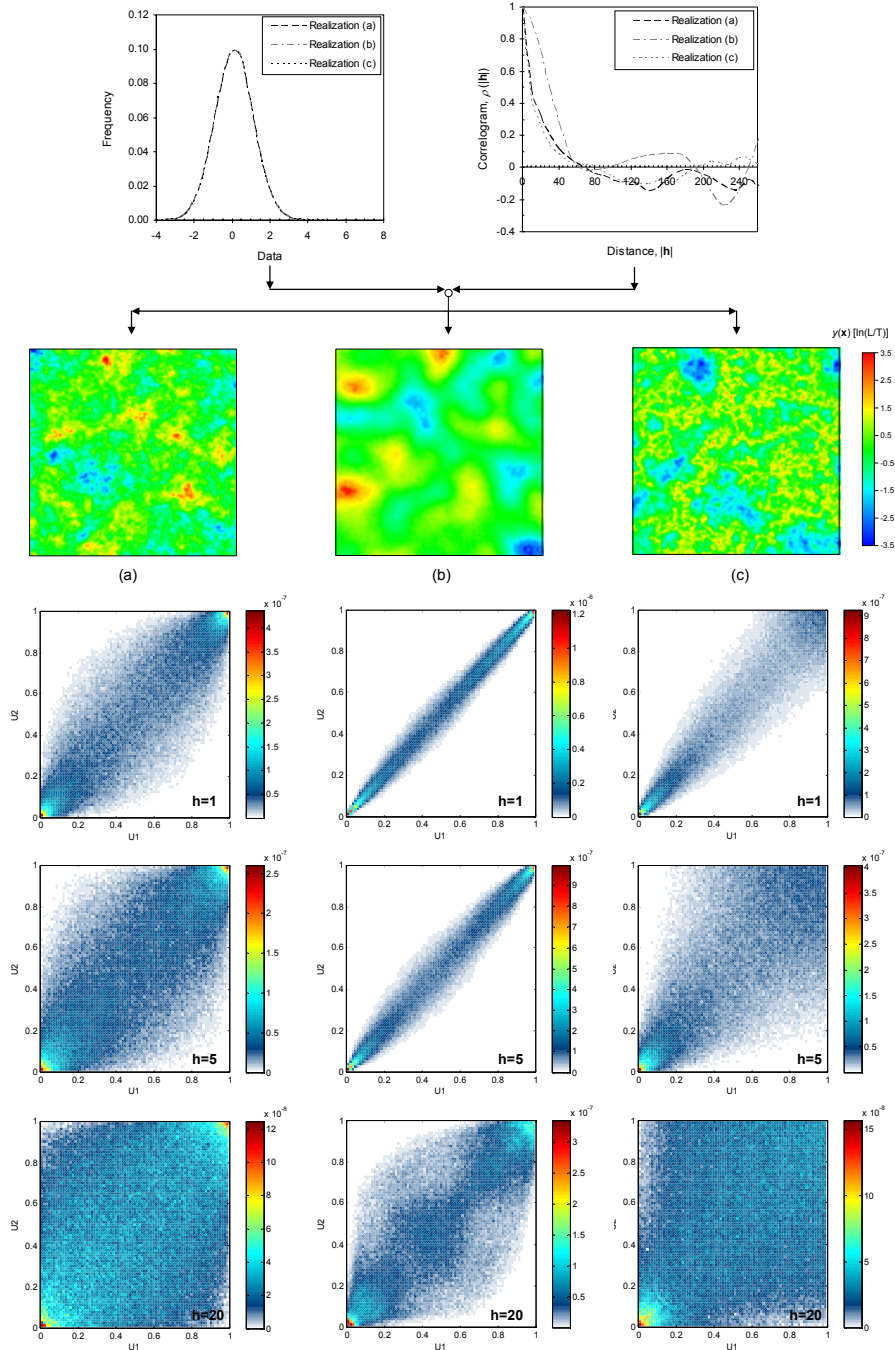
Fig. 2.2 Representative realizations of different random fields and their corresponding empirical copula density plots. a) Realization of a Gaussian RF with exponential auto-covariance function. b) Realization of a Gaussian RF with Gaussian auto-covariance function. c) Realization of a non-Gaussian RF with exponential auto-covariance function. The corresponding histograms and experimental correlograms of each realization are shown on top.

The empirical copula densities of the non multi-Gaussian realization are no longer symmetric (Fig. 2.2(c)). Instead they show a stronger association for low quantiles than for high quantiles.

Such asymmetry imposes higher continuity for the high values as can be observed by inspection in that realization. It is important to say that although the visual inspection of realization (c) suggests certain connectivity for high values, it cannot be said anything about connectivity at examining the copula densities. Intuitively, connectivity may be associated to the wide spread of density values at high quantiles. However, in its actual state of the art, copulas do not provide any information regarding connectivity. Alternatively, connectivity functions can be employed for such purpose (section 2.4.2).

From the previous example the following advantages of bivariate copula density over indicator variograms to examine dependence among random variables can be verified (Bárdossy, 2006; Bárdossy and Li, 2008):

1) Copula densities allow a joint handling of the dependence structure over the entire distribution of quantiles of the variables at once, for a given distance.
2) Differences in types of association between variables are readily identified by copula density shape.
3) Indicator variogram/covariance functions are no longer needed to examine dependence at different quantiles. Copula densities achieve this task to a less inference cost.

Comment:

- For spatial modeling purposes empirical copulas have to be replaced by theoretical ones, in the same way that empirical variograms have to be fitted to theoretical ones to ensure a variance-covariance matrix of the field to be positive definite (Chilès and Delfiner, 1999). Moreover, to describe the multivariate dependence structure in the random field a theoretical multivariate copula is needed. The construction of such copula is the subject of the next section.

2.2.3 The multivariate Gaussian copula

The multivariate Gaussian copula can easily be formulated. In fact, if $H(\cdot)=\Phi_{\Gamma}(\cdot)$ in Eq. 2.1 is the multivariate Gaussian distribution with zero mean and correlation matrix Γ and if $F_1=\ldots=F_n=\Phi$ is the standard Gaussian distribution, then the multivariate Gaussian copula $C^G_{\Gamma}(\cdot)$ is given by:

$$C^G_{\Gamma}(u_1,\ldots,u_n)=\Phi_{\Gamma}(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_n)) \qquad (2.9)$$

Note that in this case the multivariate Gaussian copula is fully parameterized by the correlation matrix Γ. It is worth to mention that for the bivariate case such matrix only contains the correlation coefficient among two random variables.

For the Gaussian copula density an analytical expression is available (e.g. Bárdossy, 2006). Gaussian copula density plots can be easily constructed from this. As an example, Fig. 2.3 displays copula density plots of the bivariate Gaussian copula for different correlation coefficients (ρ=0.95, 0.85 y 0.45). Observe its typical shape and symmetry property.
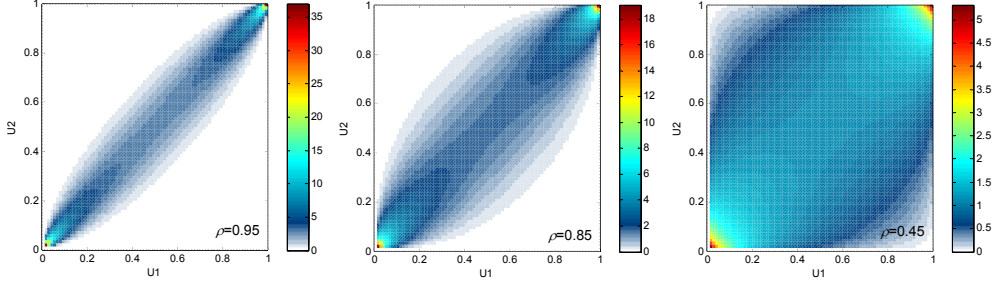
Fig. 2.3 Bivariate Gaussian copula density plots for different correlation coefficients.

2.2.4 The multivariate V-transformed copula

The multivariate *V*-transformed copula is obtained through a non monotonic transformation of a multi-Gaussian random field $G(\mathbf{x})$ with zero mean, unit variance and correlation matrix Γ such that (Bárdossy and Li, 2008; Li and Bardossy, 2009):

$$Y(\mathbf{x})= \begin{cases} k\,(G(\mathbf{x}) - m)^{\alpha} & \text{if } G(\mathbf{x}) \geq m \\[2mm] m - G(\mathbf{x}) & \text{if } G(\mathbf{x}) < m \end{cases} \tag{2.10}$$

where k is a positive constant and m and α are arbitrary real numbers. Note that if $m=0$, $k=1$ and $\alpha=1$, then $Y(\mathbf{x})=|G(\mathbf{x})|$. In that case, the transform in eq. 2.10 corresponds to the transform-based random function used to generate the realization (c) in Fig. 2.3. When $k=1$ and $\alpha=1$ transform in Eq. 2.10 leads to the multivariate non centered *chi-square* distribution. Furthermore, the effect of non linear transform vanishes when $m \rightarrow \pm\infty$ and the resulting copula converges to the Gaussian copula.

The effect of transformation in ec. 2.10 leads to a random field $Y(\mathbf{x})$ such that their marginal distribution functions are identical and they are given by:

$$F_Y= \Phi[\,(y\,/\,k)^{1/\alpha} + m\,] - \Phi[\,m - y\,] \tag{2.11}$$

The multivariate *V*-transformed copula with parameters $\lambda=\{k,m,\alpha\}$ can be written as:

$$C^{V}_{\lambda,\Gamma}(u_1,...,u_n)= H_{\lambda,\Gamma}(F^{-1}_Y(u_1),...,F^{-1}_Y(u_n)) \tag{2.12}$$

where $H_{\lambda,\Gamma}(\cdot)$ is the multivariate distribution function of $Y(\mathbf{x})$ (Appendix C).

For $n=2$, the bivariate *V*-transformed copula is:

$$C^{V}_{\lambda,\Gamma}(u_1,u_2)= \Phi_{\Gamma}[\,(y_i\,/\,k)^{1/\alpha} + m,\ (y_j\,/\,k)^{1/\alpha} + m\,] - \Phi_{\Gamma}[\,m - y_i,\ (y_j\,/\,k)^{1/\alpha} + m\,]$$

$$- \Phi_{\Gamma}[\,(y_i\,/\,k)^{1/\alpha} + m,\ m - y_j\,] + \Phi_{\Gamma}[\,m - y_i,\ m - y_j\,] \tag{2.13}$$

where $\Phi_{\Gamma}(\cdot)$ stands for the bivariate normal distribution with standard normal marginals and correlation matrix Γ.

Different asymmetric dependence structures can be represented with the *V*-transformed copula by simply modifying the set of copula parameters λ and the correlation matrix Γ. Fig. 2.4 shows

22

copula density plots for different sets of parameters computed with eq. 2.4. Observe the wide flexibility of the *V*-transformed copula to represents asymmetric dependences.
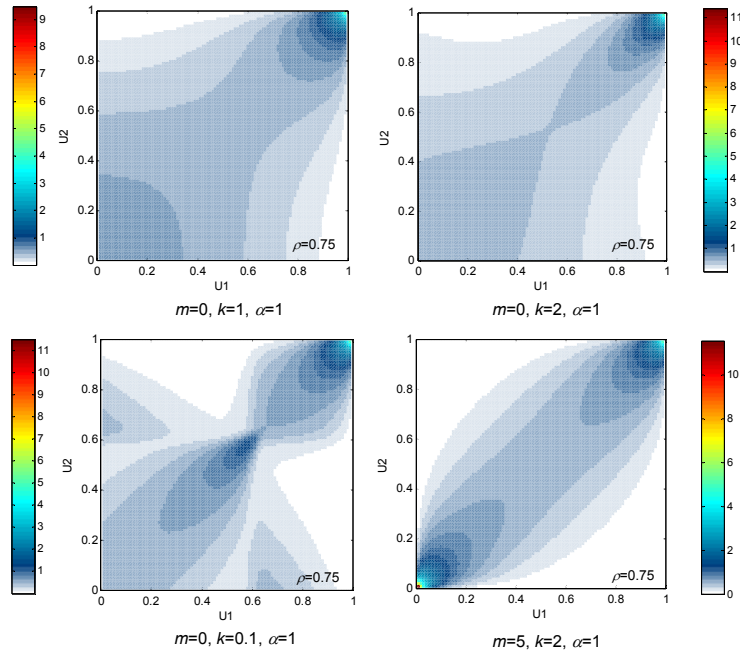
Fig. 2.4 Bivariate theoretical copula density plots for different parameters of the *V*-transform.

The bivariate dependence of random fields using copulas is controlled by their bivariate copulas and the multivariate dependence of such fields will be explicitly accounted for by the multivariate copulas.

Comment:

- Dependence structures in Fig. 2.4 shown stronger association for the upper quantiles. The opposite dependence structure is obtained by taking the complement to one of every value of *u*, that is by computing the copula density $c(1\text{-}u_1, 1\text{-}u_2)$.
- In spite of its wide flexibility, the multivariate *V*-transformed copula might not fit well a given set of observations. In that case other copula models should be considered. Li and Bárdossy (2009) proposed a so-called *maximum copula* which was constructed similarly as the *V*-transformed copula but in this case the transformation involves the maximum of two Gaussian random fields.

2.3 Non-conditional simulation

An unconditional realization using the *V*-transformed copula can easily be generated by simply applying the *V*-transform in Eq. 2.7 to a multi-Gaussian realization. However, by the effect of such transformation, the correlation matrix of $G(\mathbf{x})$ is not preserved in $Y(\mathbf{x})$. Therefore, an appropriate correlation model for $G(\mathbf{x})$ has to be found in order to get a prescribed rank correlation for $Y(\mathbf{x})$ after the transformation. The set of copula parameters *k*, *m* and α will control the asymmetry of the dependence by imposing strong/weak association at tails. The asymmetry of the spatial dependences can be quantified using asymmetry functions. The rank values of the marginal distribution of $Y(\mathbf{x})$ can be used to impose a marginal distribution of any kind to the field. Details on these subjects are explained next.

2.3.1 The correlation function

Alternative measures of dependence to the *Pearson* correlation coefficient are the *Spearman* rank correlation (Hoeffding, 1940; Quesada-Molina, 1992):

$$\rho_s = 12 \int_{[0,1]^2} C(u,v)\, dudv - 3 \tag{2.14}$$

and the *Kendall's* tau rank correlation (Hoeffding, 1948):

$$\tau = 4 \int_{[0,1]^2} C(u,v)\, dC(u,v) - 1 \tag{2.15}$$

both of which rely on the ranks (u,v) of the random variables of the bivariate copula $C(u,v)$ and have the invariance property to monotonic transforms such as $Y(\mathbf{x})=ln(K_s(\mathbf{x}))$. Thus, both $K_s(\mathbf{x})$ and $Y(\mathbf{x})$ will share identical rank correlation functions. The correlograms of $K_s(\mathbf{x})$ and $Y(\mathbf{x})$, on the other hand, are in fact not the same.

From Eqs. 2.14-2.15 it is clear that both *Spearman* and *Kendall's* tau correlation coefficients can be obtained only in terms of the bivariate copula $C(u,v)$. Each coefficient ρ of the bivariate copula $C(u,v)$ can be related to some separation distance in the correlogram of $G(\mathbf{x})$. Hence, by computing ρ_s for different ρ, the curve ρ_s-ρ can be used to determine a prescribe rank correlation function for $Y(\mathbf{x})$, given an appropriate correlation model for $G(\mathbf{x})$. Namely, a prescribed *Spearman* rank correlation function for $Y(\mathbf{x})$ can be achieved following the instructions:

1) Fit a curve to the relationship ρ_s-ρ.
2) Substitute the ρ_s values in the fitted curve in order to obtain *Pearson* values at the distances in ρ_s. This is the *p-h* curve.
3) Fit a curve to the previous *p-h* one using a linear combination of correlation functions derived from valid covariance functions.

It is worth to mention that the prescribed rank correlation function for $Y(\mathbf{x})$ is achieved in the ensemble sense and that such curve is usually only the best fitting in a least square sense.

2.3.2 The asymmetry of the dependence

The dependence at upper and lower quantiles of the marginal distributions of a non multi-Gaussian random field can be different, that is, can be asymmetric. Some measures of asymmetry can be found in the scientific literature (Haslauer *et al.,* 2008; Li and Bárdossy, 2009; Manner, 2010). They are helpful at quantifying asymmetry over distance, similarly to rank correlation functions, by collecting pairs of rank values (u,v) from the marginal distribution separated at the same distance. The concept of exceedance correlation is used in this section for this purpose.

Exceedance correlation is defined as a conditional correlation at a given exceedance level *c*. Manner (2010) suggested an definition of this concept based on the *Kendall's* tau correlation coefficient, and derived the following analytical expressions to evaluate the asymmetry at the exceedance level given by the median *c*=0.5:

$$\bar{\tau} = \cfrac{1}{C(0.5,0.5)\, E[4C(u,v) - 2C(u,0.5) - 2C(0.5,v)|u < 0.5,\, v < 0.5]} + 1 \tag{2.16}$$

$$\tau^+ = \cfrac{1}{C(0.5,0.5)\, E[4C(u,v) - 2C(u,0.5) - 2C(0.5,v) \text{ - } 1\,|\,u > 0.5,\, v > 0.5]} + 1 \qquad (2.17)$$

where $\bar{\tau}$ and τ^+ stand for the correlation of the values below and above the median, respectively. the asymmetry of the dependence is given thus by $|\bar{\tau}\text{-}\tau^+|$.

For example, Fig. 2.5 shows the asymmetry of a theoretical V-transform copula $C(u,v)$ with parameters $m=0$, $k=2.5$ and $\alpha=0.25$ over a sequence of τ values. Such copula was implemented in the program provided by Manner (2010) to evaluate Eqs. 2.16 and 2.17. It can be seen that the spatial dependence of the transformed field $Y(\mathbf{x})$ will be indeed asymmetric, with a stronger correlation for values above the median than for values below the median. Note that the opposite asymmetry can be obtained simply by taking $Y(\mathbf{x})=-Y(\mathbf{x})$, and that marginal distribution and rank correlation function will be preserved because such transformation is monotonous.

2.3.3 The marginal distribution

The rank values of the marginal distribution of $Y(\mathbf{x})$ can be used to impose a marginal distribution of any kind by a memoryless transformation like: $Y(\mathbf{x})=\Phi^{-1}(u=F_Y(y(\mathbf{x})))$ where $\Phi(\cdot)$ is the univariate Gaussian distribution. Therefore, a random field $Y(\mathbf{x})$ may have marginal Gaussian distributions, the prescribed rank correlation function of the conductivity field $K_s(\mathbf{x})$ and non multi-Gaussian dependence as controlled by the parameters m, v and α in the V-transform. Moreover, by taking: $Y'(\mathbf{x})=-Y(\mathbf{x})$ both random fields can share first and second moments but exhibit the opposite asymmetry.
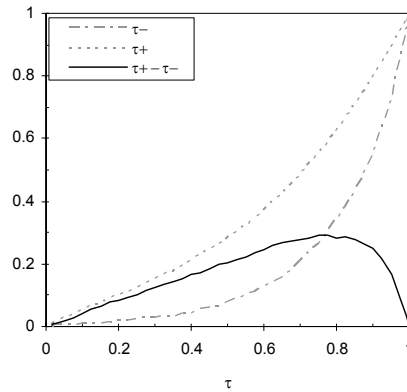
Fig. 2.5 Conditional Kendall's tau correlation functions for values below the median $\bar{t}(0.5)$ and above the median $t^+(0.5)$ for the V-transformed copula with parameters $m=0.0$, $k=2.5$ and $\alpha=0.25$.

2.3.4 Illustrative examples

The distribution of log-hydraulic conductivities $Y(\mathbf{x})=ln(K_s(\mathbf{x}))$ in a triangular domain is interpreted as a realization of a non multi-Gaussian random field with marginal Gaussian distribution functions. The mean of such field is zero and the variance is unity. Thus, the distribution of the saturated hydraulic conductivities has a geometric mean $K_G=1.0$ m/day and a coefficient of variation CV=131%. The hydraulic conductivity field is also assumed to be strictly stationary with rank correlation function given by:

$$\rho_s(\mathbf{h}) = 6 \,/\, \pi \; sin^{-1}(\rho(\mathbf{h}) \,/\, 2) \qquad (2.18)$$

where $\rho(\mathbf{h})$ is a spherical correlogram with correlation range $a = 20$ m (Deutsch and Journel, 1992). $\mathbf{h}$ is the separation vector. The above equation simply expresses the *Spearman* rank correlation function in terms of the *Pearson* correlation function (Krusal, 1954) because, strictly speaking, there are no correlation functions in terms of the *Spearman* rank correlation coefficient. Note that the differences in both sides of the equation are too small for practical purposes.

The asymmetry of the dependence structure of the field is assumed to be described by the *V*-transform copula with parameters $m=0.0$, $k=2.5$ and $\alpha=0.25$. The prescribed rank correlation function in Eq. 2.18 for the field $Y(\mathbf{x})$ is achieved with a nested correlation model corresponding to the Gaussian and Spherical covariances. The variance contribution functions are $w_1=0.90$, $w_2=0.10$ and the isotropic correlation ranges are $a_1=21.5$m and $a_2=20.0$m, respectively. Ten unconditional multi-Gaussian realizations are generated over a square domain with side of 102.5m evenly discretized in 41x41 elements using the SGSIM code (Deutsch and Journel, 1992). The realizations are then transformed by means of the *V*-transform copula.

Fig. 2.6 shows the rank correlation functions in the horizontal direction of all ten realizations. The theoretical correlation function is also shown for comparison purposes. One representative realization is then selected and only the values over the upper triangular section of the image are subsequently considered. A second realization is generated from the former as: $Y(\mathbf{x})=-Y(\mathbf{x})$. Thus, both realizations share first and second statistical descriptors but exhibit an opposite non multi-Gaussian asymmetry.

The realizations of the random field over the aforementioned triangular domain are shown in Fig. 2.7 and their statistical descriptors in Fig. 2.8. Fig. 2.7(a) shows isolated low values within a more continuous spatial structure of high values. Fig. 2.7(b) exhibits the opposite asymmetry. Fig. 2.8(b) indicates a slight anisotropy with higher continuity in the horizontal than in the vertical direction. However, the dependence structure of the realizations cannot be explained by such anisotropy. It should be emphasized that it is the asymmetry in the dependence (Fig. 2.8(c)) which generate such structures, since the more continuous zones in the realizations are clearly associated to the high (Fig. 2.7(a)) or low (Fig. 2.7(b)) values.
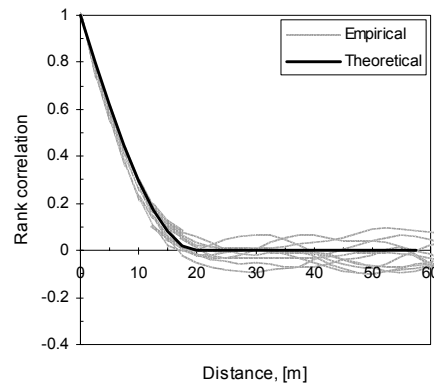
Fig. 2.6 Spearman rank correlation functions of 10 unconditional non multi-Gaussian realizations (dashed lines) and the theoretical rank correlation function of a spherical model (Deutsch and Journel, 1998) (solid line).
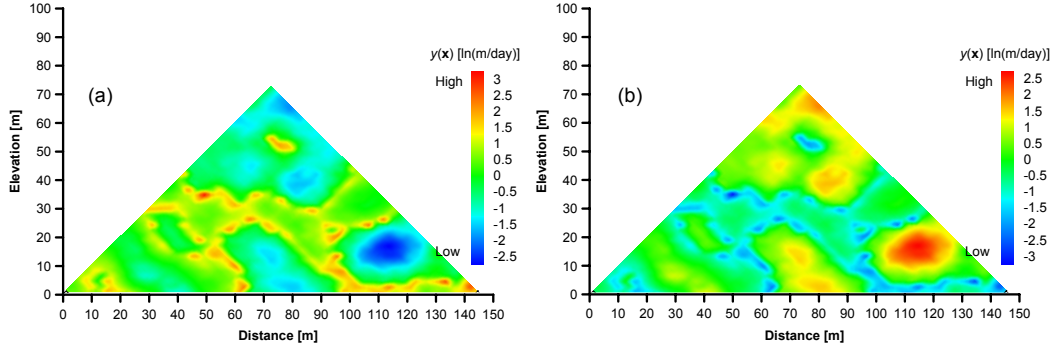
Fig. 2.7 Representative realizations of the log-conductivity random field $Y(\mathbf{x})=ln(K_s(\mathbf{x}))$. a) High values better structured. b) Low values better structured.
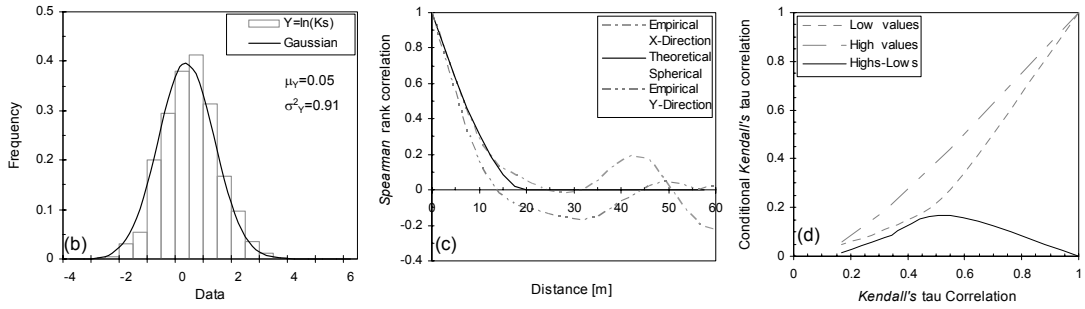
Fig. 2.8 Statistical descriptors of the realizations in Fig 2.7. a) Histogram. b) Experimental rank correlation functions in the horizontal and vertical directions. c) Experimental conditional Kendall's tau correlation functions in the horizontal direction (Manner, 2010).

2.4 Conditional simulation

An algorithm for simulating conditional random fields by copulas is proposed in this section. To validate the simulator, the spatial empirical copulas and connectivity functions of a set of simulated realizations is analyzed.

2.4.1 Statement of the problem

Consider the conditioning of the random field $Z(\mathbf{x})$ at N locations i by the set α with n observations. The corresponding N-variate conditional cumulative distribution function (CCDF) of $Z(\mathbf{x})$ can be written:

$$F_N(Z(\mathbf{x}_1),...,Z(\mathbf{x}_N)\,|\,(n))= P(Z(\mathbf{x}_i) \leq z(\mathbf{x}_i),\ i=1,...N\,|\,Z(\mathbf{x}_\alpha)=z(\mathbf{x}_\alpha),\ \alpha=1,...,n) \quad (2.19)$$

The above equation is the general expression for a CCDF. It shows that a value at each simulated location i for $i=1,...,N$ can be generated by drawing a sample from the CCDF in successive steps, involving a univariate CCDF at each location i with increasing level of conditioning; that is:

$$F_1(\mathbf{x},z_1)=P(Z(\mathbf{x}_1)\leq z(\mathbf{x}_1)\,|\,Z(\mathbf{x}_\alpha)=z(\mathbf{x}_\alpha),\ \alpha=1,...,n)$$

$$F_2(\mathbf{x},z_2)=P(Z(\mathbf{x}_2)\leq z(\mathbf{x}_2)\,|\,Z(\mathbf{x}_\alpha)=z(\mathbf{x}_\alpha),\ \alpha=1,...,n+1)$$

$$F_3(\mathbf{x},z_3)=P(Z(\mathbf{x}_3)\leq z(\mathbf{x}_3)\,|\,Z(\mathbf{x}_\alpha)=z(\mathbf{x}_\alpha),\ \alpha=1,...,n+2)$$

$$\vdots \qquad\qquad \vdots$$

$$F_N(\mathbf{x},z_N)=P(Z(\mathbf{x}_N)\leq z(\mathbf{x}_N)\,|\,Z(\mathbf{x}_\alpha)=z(\mathbf{x}_\alpha),\ \alpha=1,...,n+N\text{-}1) \qquad (2.20)$$

Equations 2.20 describe in fact a very general simulation process known as sequential simulation approach (Devroye, 1983). Such approach presents the following principal characteristics:

1) It is independent of the kind of random function used to establish the sequence in Eq. 2.20.
2) It requires simply the determination of the i for $i=1,...,N$ univariate CCDF's at each simulated location.
3) The univariate CCDF's can be of any type.

These properties let to readily formulate a copula based approach for simulation purposes without additional assumptions regarding the shape of the univariate CCDF's or their monotonicity (as in multiple indicator and isofactorial representation formalisms). Moreover, in a copula model both bivariate and multivariate distributions are defined. This is not the case of the isofactorial representation which only relies on the bivariate distribution. All these arguments seem to provide sufficient bases for the generation of random fields by copulas adopting a sequential simulation scheme.

Comment:

- In the context of random field simulation methods, some authors seek for an exact reproduction of the statistical descriptors of the realizations (histogram and experimental autocovariance function). They argue that if certain statistical descriptors are specified, then they have to be reproduced exactly. It should be recognized however that because statistical descriptors are inferred in practice from a limited number of samples, they cannot be considered exact estimators of the population statistics. The departures from the model statistics, called e*rgodic fluctuations* (Deutsch and Journel, 1992) become smaller as the relative dimension of the random field with respect to the range increases. Such ergodic fluctuations provide in fact more conservative assessments of uncertainty in schemes of the Monte Carlo type.

2.4.2 Simulation process

It terms of copulas, the sequential simulation approach can be formulated as follows (Vázquez and Auvinet, 2014):

$$F_1(\mathbf{x},z_1)=C_{\mathbf{x}|n}(u_1=F(Z(\mathbf{x}_1)\leq z(\mathbf{x}_1))\,|\,u_\alpha=F(z(\mathbf{x}_\alpha)),\alpha=1,...,n)$$

$$F_2(\mathbf{x},z_2)=C_{\mathbf{x}|n}(u_2=F(Z(\mathbf{x}_2)\leq z(\mathbf{x}_2))\,|\,u_\alpha=F(z(\mathbf{x}_\alpha)),\alpha=1,...,n+1)$$

$$F_3(\mathbf{x},z_3)=C_{\mathbf{x}|n}(u_3=F(Z(\mathbf{x}_3)\leq z(\mathbf{x}_3))\,|\,u_\alpha=F(z(\mathbf{x}_\alpha)),\alpha=1,...,n+2)$$

$$\vdots \qquad\qquad \vdots$$

$$F_N(\mathbf{x},z_N)=C_{\mathbf{x}|n}(u_N=F(Z(\mathbf{x}_N)\leq z(\mathbf{x}_N))\,|\,u_\alpha=F(z(\mathbf{x}_\alpha)),\alpha=1,...,n+N\text{-}1) \qquad (2.21)$$

where: $F(\cdot)$ is the marginal distribution and $C_{\mathbf{x}|n}(\cdot)$ is the conditional copula (Appendix C). The simulation process is restricted to local neighborhoods $\mathbf{x}_i$, for $i=1,\ldots,n$ closest to the node to be simulated. The argument is that further away conditioning data is "screened" by the information content of nearest data. The values u_α, include both original data (*prior* distribution) and previously simulated nodes. The simulation process can be performed by visiting unsampled locations at random over a mesh. After visiting all nodes of the mesh the process is completed.

2.4.3 Algorithm

The conditional simulation of random fields by copulas can be performed according to the following self explained scheme. For illustrative purposes, the multivariate *V*-transformed copula model is considered.

1) Obtain the CDF values $u_i \in [0,1]$, for $i=1,\ldots,n$ from the empirical one-dimensional distribution of the random field $Z(\mathbf{x})$, by $u_i = \hat{F}(z_i)$.

2) Obtain the theoretical values z_i^t, using the u_i's values and the one-dimensional distribution of the theoretical copula, for example, the *V*-transformed copula by $z_i^t = F^{-1}_Y(u_i)$ for $i=1,\ldots,n$.

3) Set each node j for $j=1,\ldots,N$ to be simulated over a random path l.

4) Set $j=1$.

5) Select n closest observations to the simulated node j and obtain the conditional copula over the entire interval of values of $u_i \in [0,1]$ by (Appendix C):

$$C_{\mathbf{x}|n}(u|\, u_1 = F_Y(z_1^t),\ldots,u_n = F_Y(z_n^t)) \qquad (2.22)$$

6) Draw a CDF value u_j^* from the conditional copula by Monte Carlo simulation:

$$u_j^* = C^{-1}_{\mathbf{x}|n}(p\,|\, u_1 = F_Y(z_1^t),\ldots,u_n = F_Y(z_n^t)) \qquad (2.23)$$

where p is an uniform random variate $p \in [0,1]$.

7) Assign the sampled value u_j^* to the corresponding node j. Now, it can be considered in the neighborhood of subsequent simulated nodes.

8) Set $j=j+1$.

9) If $j \le N$, get $z_j^t = F^{-1}_Y(u_j^*)$ and go through steps 5 to 7. Otherwise go to step 10.

10) Obtain the simulated value z_j^l at each location j for the realization l using the empirical marginal distribution:

$$z_j^l = \hat{F}^{-1}_Z(u_j^*) \qquad (2.24)$$

Multiple, independent realizations are obtained by visiting the nodes in distinct random sequences l. Note that the first node to be simulated at each new realization is conditioned to n closest available observations. Moreover, it can be observed from step 6 that the simulated value $u_j^* \in [0,1]$. Therefore, the univariate distribution of the simulated random field is the uniform distribution. Then using the *prior* distribution at step 10 the original marginal distribution is retrieved.

The search strategies to establish which nearby data should be considered in the simulation sequence are those implemented in the program SGSIM from the GSLIB library (Deutsch and Journel, 1992). To show some results provided by the simulator, consider the following examples.

2.4.4 Illustrative examples

The point is to generate realizations of a random field $Z(\mathbf{x})$ with marginal Gaussian distribution functions and spatial structure given by a *Spearman* rank correlation function, such that: $\rho_s(\mathbf{h}) = 6/\pi \cdot sin^{-1}(\rho(\mathbf{h})/2)$ with $\rho(\mathbf{h})$ given by the exponential correlogram with isotropic range $a=10$ units. It is of interest to generate a set of realizations for different sets of copula parameters $\boldsymbol{\lambda} = \{m, k, \alpha\}$, but matching prescribed marginal distributions and *Spearman* rank correlation function.

Given the constrained statistics for the realizations in terms of the *Spearman* rank correlation function, the correlation structure $\boldsymbol{\Gamma}$ of the multivariate *V*-transformed copula $C_{\boldsymbol{\lambda}, \boldsymbol{\Gamma}}^{V}$ has to be parameterized as discussed in section 2.3. The prescribed function in each case was achieved with the nested correlation model indicated in Table 2.1.

Table 2.1 Nested correlation model for each analyzed case. The contribution variances and correlation ranges correspond to the Gaussian, spherical and exponential auto-covariance functions, respectively.

Case	Copula parameters			Contribution variances			Correlation ranges[1]		
	m	k	α	w_1	w_2	w_3	a_1	a_2	a_3
1	0	1	1	0.4	0.2	0.4	20	13	20
2	0	0.1	1	0.7	0.05	0.25	15	11	15
3	5	2	1	0.1	0.1	0.8	9	9	11

[1]The correlation ranges are isotropic and given in any consistent units.

15 realizations of each random field indicated in Table 2.1 were generated. The simulation was performed over a square mesh with size equal to 75 units consisting of regularly spaced nodes at every 1 unit. The second realization of each set was selected. For illustration purposes, these are shown in Fig 2.9(b). The histogram and experimental rank correlation functions of each realization are shown on top of such figure (Fig. 2.9(a)). The spatial copula densities of the selected realizations were examined at several distances but only those corresponding to the distance $|\mathbf{h}|=1$ units, that is, where $\rho_s(1) \approx 0.74$, are shown for illustration purposes in Fig. 2.9. To the left of the empirical copula densities, the corresponding theoretical densities are shown for visual reference.

As can be seen in Fig. 2.9(a), the histograms of the realizations are very close to the Gaussian density. This means that the distribution of values of the realizations is quite approximate to the uniform distribution, as expected. In addition, the experimental rank correlation function of each realization follows approximately the corresponding prescribed theoretical function (Fig. 2.9(a)). As observed from a visual inspection of the realizations in the copula space, the continuity at high quantiles seems to be higher than at low quantiles except for the case 3, that is, where $m=5$, $k=2$ and $\alpha=1$. In this case, high/low quantiles appears as isolated clusters. In fact, this case approximately corresponds to a Gaussian random field.

The bivariate empirical copula densities of the realizations clearly show the characteristics of the dependence. All empirical copula densities resemble very well their corresponding theoretical versions. The asymmetry observed at the empirical copulas confirms what the visual inspection of the realizations in the copula space initially suggested, namely higher continuity at high quantiles. The propose scheme seems to provide thus satisfactory results in that the bivariate copula densities of the realizations are rather well approximated.

(a)

m=0, k=1, α=1 m=0, k=0.1, α=1 m=5, k=2, α=1

(b)

m=0, k=1, α=1

m=0, k=0.1, α=1

m=5, k=2, α=1

(c)
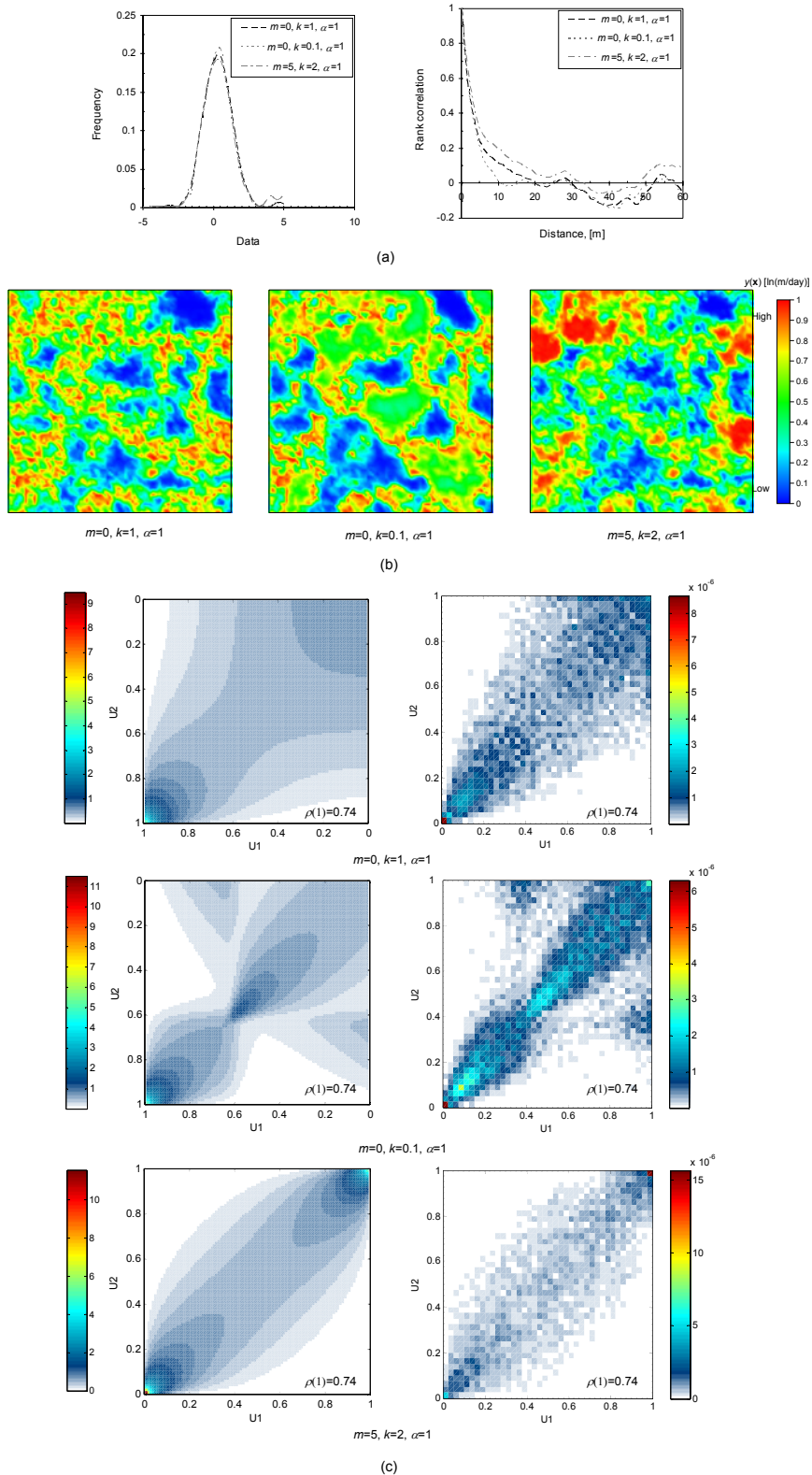
Fig. 2.9 Representative realizations of RFs generated by copulas with the sequential simulation approach. All RFs share identical Gaussian marginal distributions and Spearman rank correlation functions yet differ in the asymmetry that imposes the copula parameters m, k and α. a) The corresponding histograms and empirical rank correlation functions in the horizontal direction. b) Realizations from the RFs whose copula parameters are indicated below them. c) Empirical copula densities (right) and the corresponding theoretical versions (left).
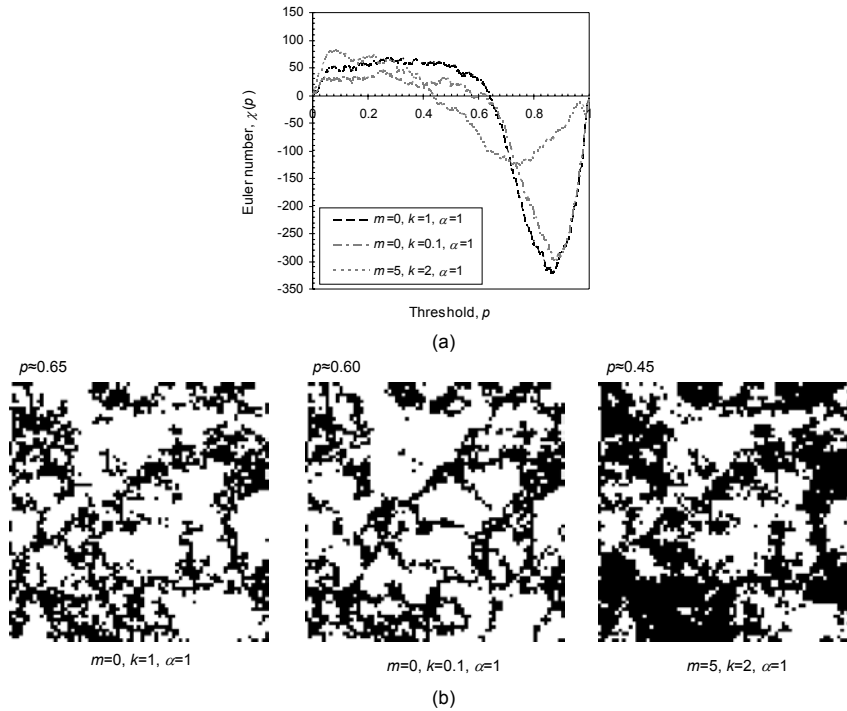
Fig. 2.10 Euler number as a function of the threshold p and the corresponding binary images at $\chi(p)=0$. a) Euler functions for the simulated realizations with the V-transformed copula whose parameters are indicated below each figure. b) Binary images at $\chi(p)=0$.

To explore the connectivity characteristics of the realizations, connectivity functions were constructed by computing the Euler characteristic at different thresholds (Mecke and Wagner, 1991; Vögel, 2002). The thresholds correspond to a set of quantiles. At each threshold a binary image is obtained by separating the values below (white phase) and above (black phase) that threshold. Then the Euler number is obtained at each threshold (Fig. 2.10(a)). The set of values thus obtained is known as the connectivity function. A value of the function equal to zero means the field begins to be well connected with perhaps a continuous loop spanning two opposite sides of the field (see binary plots in Fig. 2.10(b)). The minimum value of the function corresponds to the threshold at which the image becomes intensely connected.

The realizations with asymmetric dependence clearly become more intensely connected than the realization with symmetric dependence (Fig. 2.10(a)). The structures more intensely connected appear at significant higher quantiles. The verge of connectivity ($\chi(p)=0$) in these cases occurs at quantiles higher than the median and the connected phase is wider (Fig. 2.10(b)). Therefore, it can be argued that the proposed algorithm indeed generate realizations whose connectivity functions deviate from the connectivity functions of the mutli-Gaussian realizations (chapter 1).

The higher connectivity of the asymmetric dependences can be related to the higher spread of density values at high quantiles (Fig. 2.9(c)). It would be very interesting to formulate a measure of connectivity in terms of the bivariate copula density.

2.5 Conclusions

Along this chapter some basic concepts about copulas were introduced. The emphasis was paid on concepts like dependence and asymmetry. The spatial modeling process via copulas was addressed in some detail. An algorithm for the simulation of random fields by copulas was proposed. Random fields were simulated with such algorithm and the spatial empirical copula

densities of the realizations were examined. Satisfactory results were found in terms of the reproduction of the corresponding theoretical densities.

Characteristics of the spatial structure such as dependence and asymmetry were described quantitatively by means of single scalars which only depend on the bivariate copulas. By expressing such scalars over different distances, dependence and asymmetry functions were constructed. These functions and plots of bivariate copula densities were used to examine and visualize the spatial structure of the simulated realizations. The connectivity functions of the realizations were also determined in terms of the Euler number.

The spatial variability modeling process by copulas differs from more traditional geostatistical approaches such as the multiple indicator formalism, in that copulas make use of the entire bivariate probability law and not only of an approximation based on indicator covariances.

It should be recognized however that the simulation of random fields by copulas may be very time consuming particularly for large meshes. The computational time needed for simulating non multi-Gaussian random fields is often considerably higher than the time needed for simulating multi-Gaussian random fields. Further research is needed in order to reduce the computational burden.

CHAPTER 3

SOLUTION TO THE INVERSE PROBLEM WITH A MODIFIED ENSEMBLE KALMAN FILTER

The Ensemble Kalman Filter (EnKF) is a technique utilized in science an engineering to update (in a Bayesian sense) estimates of physical variables consistently with the physics of the phenomena under study and also with measurements available at specific locations and times. The method was originally developed for Geophysical applications by Evensen (1994) but in the last decade it has been transferred and adapted to address estimation problems in a variety of disciplines where EnKFs are coupled to physical models to update estimates of estate variables, parameter variables or both (Evensen, 2003; Ghanem and Ferro, 2006; Butala *et al.*, 2009). The estimation of parameter variables is also known in engineering as parameter identification and inverse modeling.

The functional relationship between response variables and parameter variables in the EnKF approach is represented by the equations of the system dynamics, which are often in the form of some kind of stochastic partial differential equations. The EnKF method can be summarized in two main steps which include a forecast step and an update step. The forecast step utilizes the method of Monte Carlo to approximate the probability density function (PDF) of the system response through an ensemble of realizations. The update step approximates the posterior PDF of the parameter to be estimated by conditioning of each prior realization to available time series of observations using a linear estimation technique. As EnKF is in fact a Monte Carlo method, it is suitable even for complicated non-linear systems, but because of the linear update step, it is exact only when the random variables involved are jointly Gaussian.

For applications of EnKFs in the context of random fields, Bertino *et al.* (2003) proposed to apply univariate transformation techniques before the update step to generate a geometrical space in which, at least, the marginal distributions are Gaussian. The Gaussian transformation is in fact a standard procedure in geostatistical applications known as normal-score transform (Deutsch and Journel 1996). Although recent research in the field of groundwater flow in random porous media has found this variant of the EnKF to perform better than the original one and also to be more convenient to handle non Gaussian distributions of parameter variables (Zhu *et al.*, 2011; Shöniger *et al.*, 2012; Xu *et al.*, 2013), the method remains suboptimal and some deficiencies associated to linear estimation techniques such as the use of the covariance

function as the solely descriptor of an evolving spatial dependence are transferred to this alternative (Vázquez and Auvinet, 2015).

In the update step of the EnKF method applied to random fields, instead of modeling the *N*-variate Conditional Cumulative Distribution Function (CCDF), a univariate CCDF is established at each of the *N* locations to be estimated taking into account only the available time series of observations. This strategy lends the scheme to be efficient computationally and also advisable for large dimensional problems. In addition, such scheme enables the incorporation of a simulation technique known as *p*-field simulation (Srivastava, 1992; Froidevaux, 1993) without altering the computational efficiency of the method. In *p*-field simulation, samples from the local Conditional Cumulative Distribution Function (CCDF) are drawn using uniformly distributed, auto-correlated random numbers (*p*-values) so that the resulting simulated values approximate the local CCDF over the ensemble of realizations.

In this chapter is shown how an update step based on *p*-field simulation can be readily implemented within the EnKF context. To develop this approach, the local CCDFs of the random field to be estimated are determined first with the EnKF method and realizations from these distributions are generated then by *p*-field simulation. The proposed scheme denoted here by pf-EnKF, is expected to be more flexible than the EnKF in the sense that the former may refine the EnKF approximation by introducing fluctuations around the mean estimates. The performance of the proposed scheme will be evaluated with respect to different quantitative criteria.

An alternative formulation of the EnKF procedure based on available results of linear estimation theory applied to random functions is used to incorporate the proposed methodology. A one-dimensional, single phase flow problem in continuous random porous media is considered to illustrate the effects of the update step of the pf-EnKF and to compare the results with those from the EnKF, in the estimation of conductivities and heads.

3.1 Groundwater flow equations

In the following analysis, the dynamic model describing fluid flow in a one-dimensional, fully saturated porous media with spatially variable hydraulic conductivity is considered:

$$\frac{\partial}{\partial x}\left[K_s\left(\mathbf{x}\right)\frac{\partial H}{\partial x} \right] = S_s \frac{\partial H}{\partial t} \tag{3.1}$$

subject to initial and boundary conditions:

$$H_{t=0} = h_0, \quad H_{\Gamma_D} = h_1 \tag{3.2}$$

where, *H* is the hydraulic head [L] in the domain Ω_x, $\mathbf{x}$ is the spatial coordinate ($\mathbf{x}=x_3$ [L], where x_3 represents the vertical coordinate which is positive upward), $K_s(\mathbf{x})$ is the saturated hydraulic conductivity [L/T], S_s is the specific storage [L^{-1}], h_0 represents the initial head and h_1 the prescribed head at Dirichlet boundary Γ_D.

Although initial and boundary conditions are considered deterministic constants, the hydraulic conductivity $K_s(\mathbf{x})$ is considered a random field. Therefore, equation 3.1 becomes a stochastic differential equation and the flow response *H* will also be a random field. To obtain the response of the model dynamics satisfying equations 3.1 and 3.2, the method of Monte Carlo is utilized. First, independent realizations of the hydraulic conductivity random field are generated and flow equations are solved numerically with the deterministic values of each realization using a

modified version of a finite element code (Smith and Griffiths, 2004). Statistics of the flow response such as auto and cross-covariance functions between conductivities and heads are determined over the ensemble of realizations through well-known formulas of descriptive statistics, at each time the conductivity fields are updated by the following simulation process.

3.2 Simulation of spatiotemporal random fields

In this section, an alternative formulation of the EnKF method is presented. Additionally, it is shown how an update step based on *p*-field simulation can be readily implemented within the EnKF method.

3.2.1 Definitions

Let $Y_t(\mathbf{x})$ denote the collection of n continuous random variables of the natural logarithm of the saturated hydraulic conductivity K_s i.e. $Y=\ln(K_s)$ indexed at the spatial locations $\mathbf{x}$ in the domain Ω_x with $\Omega_x \subset \mathbb{R}^d$ for $d=1,2$ or 3 and at the discrete index of times $t \in \{0,1,2,\ldots\}$. This n-variate spatio-temporal random field of the log-conductivity is written by:

$$\left\{ Y_t(\mathbf{x}) : \mathbf{x} \in \Omega_x, t = 0,1,2,\ldots \right\} \tag{3.3}$$

Define $H_t(\boldsymbol{\chi})$ the collection of N continuous random variables of the hydraulic head indexed at the spatial locations $\boldsymbol{\chi}$ in the domain Ω_χ with $\Omega_\chi \subset \mathbb{R}^d$ for $d=1,2$ or 3 and at the discrete index of times $t \in \{0,1,2,\ldots\}$. This N-variate spatio-temporal random field of the hydraulic head is written:

$$\left\{ H_t(\boldsymbol{\chi}) : \boldsymbol{\chi} \in \Omega_\chi, t = 0,1,2,\ldots \right\} \tag{3.4}$$

The multivariate Cumulative Distribution Function (CDF) of $H_t(\boldsymbol{\chi})$ at the particular time t depends on the multivariate distributions at all previous times. However, the evolution of such function can be determined assuming the dynamic model in Eqs. 3.1-3.2 behaves like a first order Markov process i.e. $P[H_t(\boldsymbol{\chi})|H_{t-1}(\boldsymbol{\chi}),H_{t-2}(\boldsymbol{\chi}),H_{t-3}(\boldsymbol{\chi}),\ldots]=P[H_t(\boldsymbol{\chi})|H_{t-1}(\boldsymbol{\chi})]$. Hence, only the most recent past determine the multivariate conditional CDF of $H_t(\boldsymbol{\chi})$ given the whole past. This simplified evolution of $H_t(\boldsymbol{\chi})$ is given by:

$$H_t(\boldsymbol{\chi}) = \Im\left(H_{t-1}(\boldsymbol{\chi}), Y_{t-1}(\mathbf{x}) \right) \tag{3.5}$$

where $\Im(\cdot)$ represents the equations of the model dynamics. Observe in Eq. 3.5 that in order to advance $H_t(\boldsymbol{\chi})$ to the first time i.e. $H_1(\boldsymbol{\chi})$, the initial conditions $H_0(\boldsymbol{\chi})$, the random field of parameters $Y_0(\mathbf{x})$ and the boundary conditions have to be specified.

Suppose now that N_h observations of $H_t(\boldsymbol{\chi})$ are taken with no error at the spatial locations $\boldsymbol{\chi}_\alpha$, for $\alpha \in \{1,\ldots,N_h\}$ and at the discrete index of times $t \in \{1,2,3,\ldots\}$. The set of histories of available observations is hence represented by:

$$\left\{ H_t(\boldsymbol{\chi}_\alpha) = h_{\alpha,t}, \ \alpha = 1,2,\ldots,N_h, \ t = 1,2,3,\ldots \right\} \tag{3.6}$$

A random field that is made specific to a set of direct/indirect observations at their exact locations is called conditional. The random fields $Y_t(\mathbf{x})$ and $H_t(\boldsymbol{\chi})$ given the set of observations $H_t(\boldsymbol{\chi}_\alpha)$ are written as:

$$\left\{ Y_t(\mathbf{x}) \big| H_t(\boldsymbol{\chi}_\alpha): \ \mathbf{x} \in \Omega_x, \boldsymbol{\chi}_\alpha \in \Omega_\chi, \ \alpha = 1, 2, \ldots, N_h, \ t = 1, 2, 3, \ldots \right\} \tag{3.7}$$

$$\left\{ H_t(\boldsymbol{\chi}) \big| H_t(\boldsymbol{\chi}_\alpha): \ \boldsymbol{\chi}_\alpha \in \Omega_\chi, \ \alpha = 1, 2, \ldots, N_h, \ t = 1, 2, 3, \ldots \right\} \tag{3.8}$$

The analysis of the conditional random fields in equations 3.7 and 3.8 involves the determination of all marginal Conditional Cumulative Distribution Functions (CCDF) and all n-variate (and N-variate) CCDFs at each location $\mathbf{x}$ or $\boldsymbol{\chi}$ and at the particular time t. This task is simplified by replacing each n-variate (and N-variate) CCDF with a univariate CCDF at each of the n or N spatial locations at the particular time t. The conditional random fields of Eqs. 3.7 and 3.8 can thus be written:

$$\left\{ Y_t(\mathbf{x}_i) \big| H_t(\boldsymbol{\chi}_\alpha): \mathbf{x} \in \Omega_x, \boldsymbol{\chi}_\alpha \in \Omega_\chi, \ i = 1, 2, \ldots, n, \ \alpha = 1, 2, \ldots, N_h, \ t = 1, 2, 3, \ldots \right\} \tag{3.9}$$

$$\left\{ H_t(\boldsymbol{\chi}_i) \big| H_t(\boldsymbol{\chi}_\alpha): \boldsymbol{\chi}_\alpha \in \Omega_\chi, \ i = 1, 2, \ldots, N, \ \alpha = 1, 2, \ldots, N_h, \ t = 1, 2, 3, \ldots \right\} \tag{3.10}$$

If the joint relationship of $Y(\mathbf{x})$ and $H(\boldsymbol{\chi})$ is assumed to be multi-Gaussian (i.e. if a multi-Gaussian model is assumed), then the conditional random field $Y_t(\mathbf{x}_i) \big| H_t(\boldsymbol{\chi}_\alpha)$ can be written as a sum of the two independent random fields (Journel and Huijbregts, 1978):

$$Y_t(\mathbf{x}_i) = Z^*(\mathbf{x}_i) + \left[Y_s(\mathbf{x}_i) - Y_s^*(\mathbf{x}_i) \right] \tag{3.11}$$

and both random fields in Eq. 3.11 can be calculated with the Kriging technique:

$$
\begin{aligned}
Z^*(\mathbf{x}_i) &= \sum_{i=1}^{n} \sum_{\alpha=1}^{N_h} \lambda_i(\boldsymbol{\chi}_\alpha) h_{\alpha,t} \\
Y_s(\mathbf{x}_i) &= Y_{t-1}(\mathbf{x}_i) \\
Y_s^*(\mathbf{x}_i) &= \sum_{i=1}^{n} \sum_{\alpha=1}^{N_h} \lambda_i(\boldsymbol{\chi}_\alpha) H_t(\boldsymbol{\chi}_\alpha)
\end{aligned}
\tag{3.12}
$$

Substituting Eqs. 3.12 into Eq. 3.11 yields:

$$Y_t(\mathbf{x}_i) = Y_{t-1}(\mathbf{x}_i) + \sum_{i=1}^{n} \sum_{\alpha=1}^{N_h} \lambda_i(\boldsymbol{\chi}_\alpha) \left[h_{\alpha,t} - H_t(\boldsymbol{\chi}_\alpha) \right] \tag{3.13}$$

where $\lambda_i(\mathbf{x}_\alpha)$ are weighting functions representing the relative importance of the observations $H_t(\boldsymbol{\chi}_\alpha)$ in estimating the value of $Y_t(\mathbf{x}_i)$. The weighting functions are solutions of the following systems of linear equations:

$$\sum_{i=1}^{n} \sum_{\alpha=1}^{N_h} \lambda_i(\boldsymbol{\chi}_\alpha) C_H(\boldsymbol{\chi}_\alpha, \boldsymbol{\chi}_j; t, t) = C_{YH}(\mathbf{x}_i, \boldsymbol{\chi}_\alpha; t-1, t), \quad j = 1, 2, \ldots, N_h \tag{3.14}$$

with $C_H(\boldsymbol{\chi}_\alpha, \boldsymbol{\chi}_j; t, t)$ representing the spatio-temporal auto-covariance functions between hydraulic heads; $C_{YH}(\mathbf{x}_i, \boldsymbol{\chi}_\alpha; t-1, t)$ represents the spatio-temporal cross-covariance functions between log-conductivities and hydraulic heads. In order the solution of Eq. 3.14 to be unique, the spatio-temporal covariance functions involved have to be strictly positive definite. The equations like

3.13 and 3.14 can be written for the conditional random field $H_t(\chi_i)\big|H_t(\chi_\alpha)$ (Vázquez and Auvinet, 2015).

The determination of the spatio-temporal covariance functions in Eq. 3.14 is however far from being trivial because non linear effects associated to the flow problem impact the statistical relationships between $Y_t(\mathbf{x})$ and $H_t(\chi)$. Although some general theoretical models of spatio-temporal covariance functions could be adopted to quantify the dependences, it remains unclear how suitability such models are to address transient flow problems in porous media with non multi-Gaussian properties. Thereby, the method of Monte Carlo is used to solve Eq. 3.5 and the necessary covariance functions are then determined over the ensemble of realizations. This is the core of the ensemble methods. The covariance functions thus obtained are in fact empirical, but on the average over several realizations they lead to positive definite matrices and may be used directly without modeling.

It is worth mentioning that Eq. 3.13 can be interpreted as a Bayessian mechanism in the sense that the conditional random field $Y_t(\mathbf{x})$ is dependent on the *a priori* knowledge consisting of the random field $Y_{t-1}(\mathbf{x})$, covariance functions and observations.

3.2.2 The EnKF method

The Ensemble Kalman Filter (EnKF) utilizes the method of Monte Carlo to approximate the marginal distributions of $H_t(\chi)$ assuming the flow model is a first order Markov process. Then simulates the distributions of $Y_t(\mathbf{x}_i)$ conditioning each realization of $Y_{t-1}(\mathbf{x}_i)$ to the observations of $H_t(\chi_\alpha)$ assuming a multi-Gaussian model. This update step can be written:

$$\mathbf{U}_t^u = \mathbf{U}_0 + \mathbf{K}_t \left[\mathbf{Z}_t - \mathbf{H}_t^f \right] \tag{3.15}$$

where $\mathbf{U}_t^u = [\hat{y}_t(\mathbf{x}_1), \hat{y}_t(\mathbf{x}_2),\ldots,\hat{y}_t(\mathbf{x}_n)]$ is a n-dimensional vector of updated realizations of the log-conductivity, $\mathbf{U}_0 = [y_{t-1}(\mathbf{x}_1), y_{t-1}(\mathbf{x}_2),\ldots,y_{t-1}(\mathbf{x}_n)]$ is a n-dimensional vector of simulated realizations of the log-conductivity (realizations *a priori*), $\mathbf{Z}_t = [h_{1,t}, h_{2,t},\ldots,h_{Nh,t}]$ is the vector of observations with dimension N_h and $\mathbf{H}_t^f = [h_t(\chi_1), h_t(\chi_2),\ldots,h_t(\chi_{Nh})]$ is a reduced vector of forecasted states (realizations of the hydraulic head at the locations of the observations) of dimension N_h.

The matrix $\mathbf{K}_t$ ($n \times N_h$) is the so-called "Kalman gain", which can be assembled after obtaining the weighting functions associated to each location $\mathbf{x}_i$ at the updating time t, as:

$$\mathbf{K}_t = \begin{bmatrix} \lambda_1(\mathbf{x}_1) & \lambda_1(\mathbf{x}_2) & \ldots & \lambda_1(\mathbf{x}_{N_h}) \\ \vdots & \vdots & & \vdots \\ \lambda_n(\mathbf{x}_1) & \lambda_n(\mathbf{x}_2) & \ldots & \lambda_n(\mathbf{x}_{N_h}) \end{bmatrix} \tag{3.16}$$

The update step of the EnKF is repeated at the next time for which observations are available but the new *a prior* random field of the log-conductivity is the *a posteriori* one. The update step is thereby performed sequentially incorporating past observations well as current observations in the inference process.

3.2.3 The modified EnKF method

This variant of the EnKF method performs the update step within a transformed space in which the marginal distributions of the random fields involved are Gaussian with zero mean and unit variance. After performing the update step the distributions of the variables are back-transformed to their original distributions. The additional steps of the modified EnKF method are explained in the sequel.

3.2.3.1 Gaussian transformations

For the purpose of Gaussian transformations, CDFs of log-conductivities at locations $\mathbf{x}$ i.e. $F_Y(y;\mathbf{x}_i)$ for $i=1,2,\ldots,n$ and CDFs of hydraulic heads at locations $\boldsymbol{\chi}$ i.e. $F_H(h;\boldsymbol{\chi}_j)$ for $j=1,2,\ldots,N$ are established statistically over the ensemble of realizations. Then, the Gaussian transformation G^{-1} (with zero mean and unity variance) is applied to each case i.e. $y'=G^{-1}[F_Y(y;\mathbf{x}_i)]$ and $h'=G^{-1}[F_H(h;\boldsymbol{\chi}_j)]$, where y' and h' are the transformed Gaussian values. In addition, a standardization of the observations using the forecasted values at each location $\boldsymbol{\chi}_\alpha$ for $\alpha=1,\ldots,N_h$ is applied. The function relating y to y' or h to h' in the x-y Cartesian plane, is called the Gaussian anamorphosis function (Fig. 3.1(a)).
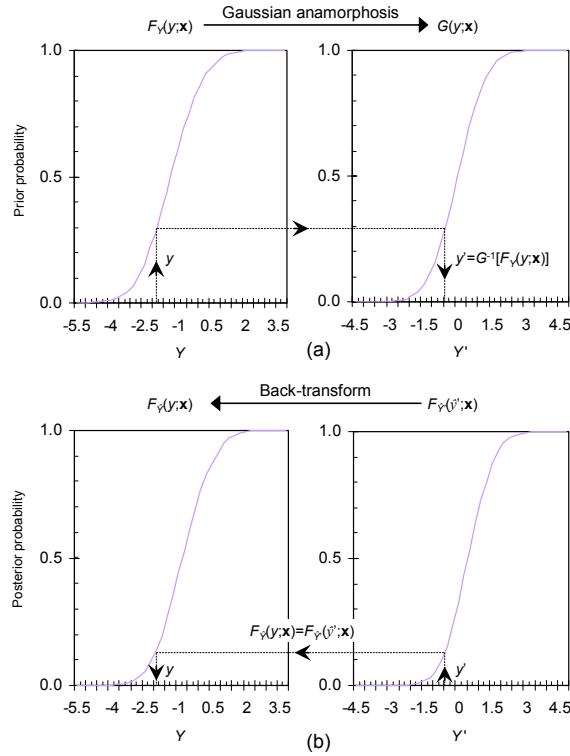
Fig. 3.1 Transformation step of the EnKF and pf-EnKF. a) Gaussian transformation process. b) Back-transformation process.

3.2.3.2 Update step

After performing all transformations described in the previous section, the update step of the modified EnKF can now be written by:

$$\tilde{\mathbf{U}}_t^u = \tilde{\mathbf{U}}_0 + \tilde{\mathbf{K}}_t \left[\tilde{\mathbf{Z}}_t - \tilde{\mathbf{H}}_t^f \right] \tag{3.17}$$

where $\tilde{\mathbf{U}}_t^u = [\hat{y}'_t(\mathbf{x}_1), \hat{y}'_t(\mathbf{x}_2), \ldots, \hat{y}'_t(\mathbf{x}_n)]$ is a n-dimensional vector of updated realizations of the log-conductivity, $\tilde{\mathbf{U}}_0 = [y'_{t-1}(\mathbf{x}_1), y'_{t-1}(\mathbf{x}_2), \ldots, y'_{t-1}(\mathbf{x}_n)]$ is a n-dimensional vector of simulated realizations of the log-conductivity (realizations *a priori*), $\tilde{\mathbf{Z}}_t = [h'_{1,t}, h'_{2,t}, \ldots, h'_{Nh,t}]$ is the vector of observations with dimension N_h and $\tilde{\mathbf{H}}_t^f = [h'_t(\boldsymbol{\chi}_1), h'_t(\boldsymbol{\chi}_2), \ldots, h'_t(\boldsymbol{\chi}_{Nh})]$ is a reduced vector of forecasted states (realizations of the hydraulic head at the locations of the observations) of dimension N_h. The matrix $\mathbf{K}_t$ (nxN_h) is assembled after obtaining the weighting functions associated to each location $\mathbf{x}_i$ at the updating time t, using the transformed values. The update step of the modified EnKF is also performed sequentially.

3.2.3.3 Back-transformations

The conditional values corresponding to the original variables can be retrieved by means of a back-transformation process in which the CCDF of back-transform values of the log-conductivities at each location $\mathbf{x}$ i.e. $F_{\hat{Y}}(y;\mathbf{x})$ is obtained by doing $F_{\hat{Y}}(y;\mathbf{x}) = F_{\hat{Y}}(\hat{y}';\mathbf{x})$. This means that the CCDF value of the original variable y (see section 3.2.3.1) is identified with the CCDF value at its corresponding Gaussian transform value $\hat{y}'$. This process is illustrated in Fig. 3.1(b).

3.2.4 The proposed pf-EnKF method

In addition to the update step of the modified EnKF method presented above, an update step based on p-field simulation can be performed as proposed in this chapter. This extension of the modified EnKF method is called here pf-EnKF method and is presented in the following paragraphs.

A CCDF of log-conductivities at each location $\mathbf{x}$ i.e. $F_{\hat{Y}}(\hat{y}';\mathbf{x}_i)$ for $i=1,2,\ldots,n$ is established statistically over the ensemble of realizations updated with the modified EnKF method (Eq. 3.17). Samples of log-conductivities at each location $\mathbf{x}$ i.e. $\hat{y}'_t(\mathbf{x}_i)$ for $i=1,2,\ldots,n$ can then be drawn from their corresponding conditional distributions by the p-field simulation technique as (Srivastava, 1992; Froidevaux, 1993):

$$\hat{y}'_t(\mathbf{x}_i) = F_{\hat{Y}'}^{-1}(p_{t-1}(\mathbf{x}_i)), \quad i=1,2,\ldots,n \tag{3.18}$$

where $p_{t-1}(\mathbf{x}_i)$ are so-called probability fields (p-fields) of the log-conductivity. The p-fields can be established with the values of the empirical distribution functions of each realization before the update step. Namely, if the update step is at $t=1$, then the p-fields are derived from the realizations at $t=0$. The simulation process described by equation 3.18 is illustrated in Fig. 3.2.

The correlation structure of the updated simulated fields $\hat{y}'_t(\mathbf{x})$ is accounted for by the correlation structure of the fields $p_{t-1}(\mathbf{x})$, as explained by Goovaerst (1997). The back-transformation process of the simulated values $\hat{y}'_t(\mathbf{x}_i)$ is performed as explained in section 3.2.3.3.
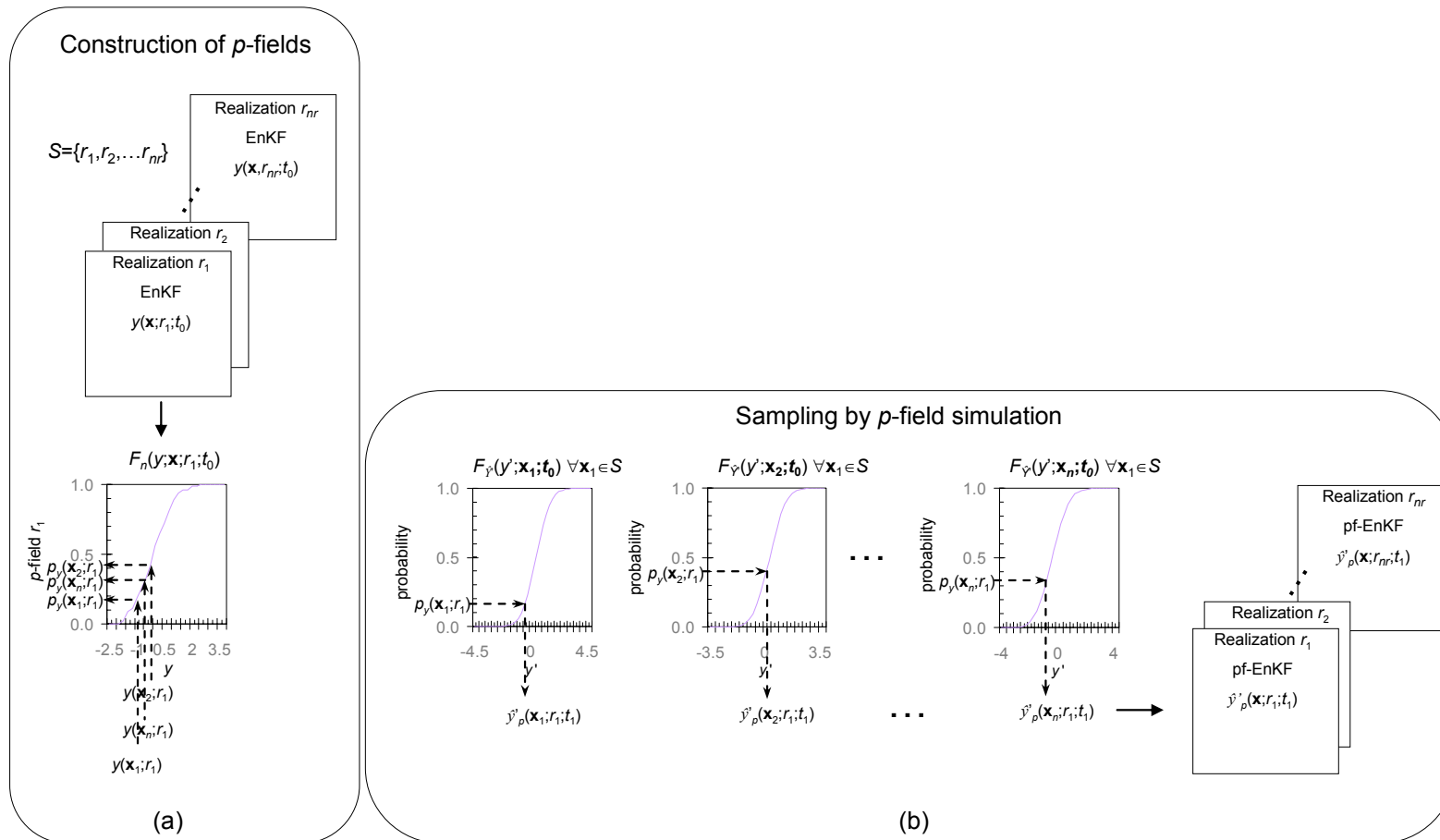
Fig. 3.2 Simulation process described by equation 3.16. a) Generation of *p*-fields. b) Sampling by *p*-field simulation.

3.3 Description of the numerical experiment

To explore the performance of the proposed pf-EnKF method, a hypothetical, one dimensional, vertical, heterogeneous, fully saturated aquifer is considered. The aquifer is 40 m depth and is discretized into 80 finite elements each of which with a length of 0.5 m. Each finite element e_i for $i=1,...,80$ is assigned a log-conductivity value $y^{ref}(\mathbf{x}_i)$ according to the following procedure. First, a multi-Gaussian field $g(\mathbf{x})$ with exponential autocovariance function and correlation scale $a_x=2.5$ m is simulated using a modified version of the SGSIM random field generator (Deutsch and Journel, 1996) (Fig. 3.3(a)). Second, the V-transform (Bárdossy and Li, 2008):

$$v(\mathbf{x}) = \begin{vmatrix} k(g(\mathbf{x})-m)^{\alpha} \text{ if } g(\mathbf{x}) \geq m \\ (m-g(\mathbf{x})) \quad \text{ if } g(\mathbf{x}) < m \end{vmatrix} \tag{3.19}$$

with arbitrarily chosen parameters $m=0$, $k=1$ and $\alpha=2$ is applied to the $g(\mathbf{x})$ field to obtain the transformed $v(\mathbf{x})$ field shown in Fig. 3.3(b). Third, a Gaussian distribution is imposed to the $v(\mathbf{x})$ field as $y'=G^{-1}[F_V(v)]$ where $F_V(v)$ is the empirical CDF of the $v(\mathbf{x})$ field and G the theoretical Gaussian CDF. Finally, such $y'(\mathbf{x})$ field is scaled to a normally distributed $y^{ref}(\mathbf{x})$ field with mean value $\mu_Y=-1.654$ and variance $\sigma^2_Y=1$ as: $y^{ref}(\mathbf{x})=\mu_Y+y'(\mathbf{x})\sigma_Y$. Each one of these values is assumed to be constant within its finite element e_i. This log-conductivity field, which is displayed in Fig. 3.3(c), is considered a "true state of nature" and is called the reference aquifer.
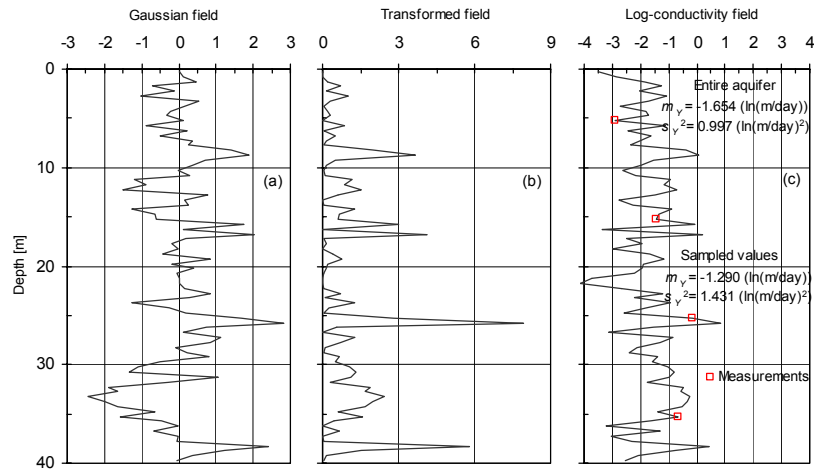
Fig. 3.3 One dimensional fields. a) Initial Gaussian field. b) Field after applying the V-transform with parameters $m=0$, $k=1$ and $\alpha=2$ to the initial Gaussian field. c) Final log-conductivity field after imposing a marginal normal distribution with expected value $\mu_Y=-1.654$ and variance $\sigma_Y^2=1$ to the V-transformed field. Statistics of sampled values (empty squares) are also reported.

Several interesting properties of the V-transformation should be mentioned. Firstly, the symmetric density function of the Gaussian field $g(\mathbf{x})$ is transformed into a non symmetric density function through the parameters m, k and α. Secondly, the empirical autocovariance function of $g(\mathbf{x})$ is not preserved in $v(\mathbf{x})$ because the V-transformation is non monotonous (Fig. 3.4). Thirdly, the spatial correlation of $v(\mathbf{x})$ is stronger for the values above the median than for the values below the median i.e. the spatial correlation of $v(\mathbf{x})$ is asymmetric. This last characteristic of the field holds after imposing to it the Gaussian (normal) distribution function because the Gaussian transformation is monotonous (Deustch and Journel, 1996). Since $y^{ref}(\mathbf{x})$ is normal distributed, $k_s(\mathbf{x})=\exp(y^{ref}(\mathbf{x}))$ is lognormal distributed with expected value $\mu_{Ks}=0.315$

m/day and coefficient of variation CV_{Ks}= 1.31. Conductivity fields with one-dimensional lognormal distributions and asymmetric correlation structures are considered to be more representative of the natural aquifers (Gómez-Hernández and Wen, 1998; Journel and Zhang, 2006), as discussed in chapter 1.
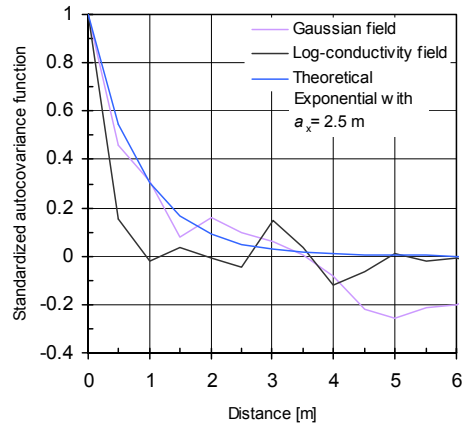
Fig. 3.4 Standardized auto-covariance functions of the initial Gaussian field and reference log-conductivity field. An exponential function is also shown for comparison.

Using the reference aquifer of conductivities, groundwater head responses are generated by solving a transient flow condition with finite elements. At t=0 days, the initial distribution of heads is hydrostatic. At $t{\geq}0$ days, hydraulic head decreases with time at a rate of 0.15 m per day during 150 days at the lower boundary. For the purpose of the present numerical example, the distribution of heads at t=90 days is assumed to be the initial condition (denoted as t=0 days in Fig. 3.5 and henceforth). It is further assumed that groundwater head responses are available at times t=3, t=18 and t=60 days at the two locations indicated in Fig. 3.5. Thus, two histories with three hydraulic head values are generated. These indirect, informative variables of the hydraulic conductivity of the aquifer are considered available transient piezometric observations. At each one of those three times, the updating process of both the EnKF and pf-EnKF schemes is performed. Storage coefficient is assumed to be equal to 0.001 overall the aquifer.
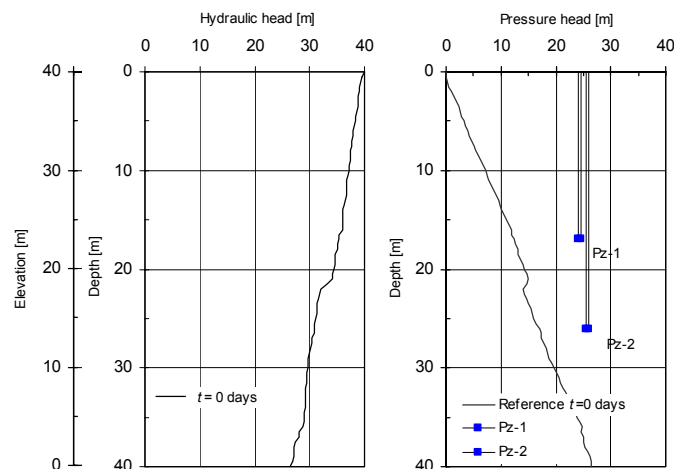
Fig. 3.5 Profile of hydraulic heads in the reference aquifer at t=0 days. The depths of the tips of two piezometers (Pz-1 and Pz-2) are also indicated with filled squares.

44

The reference $y^{ref}(\mathbf{x})$ aquifer is sampled at four locations indicated in Fig. 3.3(c) and the values are considered direct log-conductivity measurements. The mean and variance of the set of sampled values are reported in the same figure. Observe that these statistics overestimate the mean and variance of the reference aquifer. Note in Fig. 3.4 that an exponential autocovariance function with correlation scale $a_x=2.5$ m overestimate the correlation scale of the reference aquifer.

To model a situation in which the statistical parameters of the aquifer are only roughly known *a priori*, the sampled mean and variance as well as the autocovariance function mentioned above are used to simulate two thousand, conditional multi-Gaussian log-conductivity realizations. The number of simulated realizations was established to ensure the stability of the following two error measures, according to some preliminary computations.

To assess the performance of both the EnKF and pf-EnKF filters in the estimation of log-conductivities and hydraulic heads, the root mean square error (RMSE) and the SPREAD measure are considered in this work. The RMSE is evaluated by:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y^*(\mathbf{x}_i) - y^{ref}(\mathbf{x}_i) \right)^2} \qquad (3.20)$$

where n is the number of log-conductivities in the flow domain; $y^*(\mathbf{x}_i)$ is the estimated mean log-conductivity at location $\mathbf{x}_i$ and $y^{ref}(\mathbf{x}_i)$ is the reference log-conductivity also at location $\mathbf{x}_i$.

The SPREAD is computed as:

$$SPREAD = \sqrt{\frac{1}{n} \sum_{i=1}^{n} s_{en}^2(\mathbf{x}_i)} \qquad (3.21)$$

where $s^2_{en}(\mathbf{x}_i)$ is the variance of the estimation of the log-conductivity at location $\mathbf{x}_i$ computed statistically over the ensemble of realizations.

RMSE is a measure of the difference of the means of the estimated and the reference fields and SPREAD is a measure of the dispersion of the estimated field around the reference field. Therefore, they can be viewed as measures of accuracy and precision of the estimations, respectively.

3.4 Results and discussion

The characteristics of the realizations of the conditional mean value of the log-conductivity and hydraulic head fields are discussed in this section. The results from the modified EnKF against those from the proposed pf-EnKF are compared. Before analyzing the effects of conditioning, the effects of the Gaussian transformation are analyzed.

3.4.1 Effects of the Gaussian transformation

Fig. 3.6 illustrates the effect of the Gaussian transformation process of the hydraulic heads at an arbitrarily selected node before the first updating step. In the general case, the shape of the local distributions will depend on the location of the node in the flow domain and on the boundary conditions of the problem at hand. In any case, the local distribution functions can be transformed into Gaussian distributions by building local Gaussian anamorphosis functions

numerically, as explained. For example, observe in Fig. 3.6(a) that although the original values exhibit a skewed distribution, the transformed variable becomes symmetric around the mean showing the well-known bell-shape after the Gaussian anamorphosis (Figs. 3.6(b),(c)).
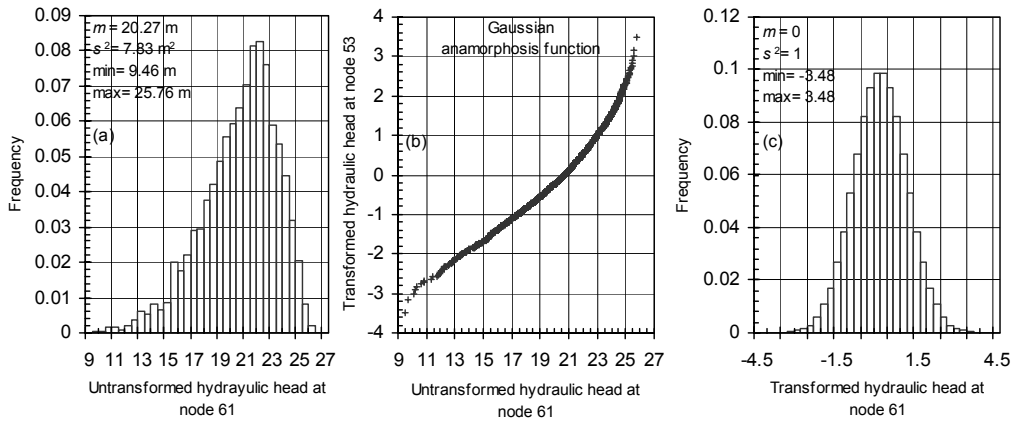
Fig. 3.6 Gaussian transformation of hydraulic heads at node 61. a) Histogram of untransformed hydraulic heads. b) Gaussian anamorphosis function (with zero mean and unity variance). c) Histogram of hydraulic heads after the Gaussian anamorphosis.

Fig. 3.7 represents the relationship between log-conductivities and heads at arbitrarily selected locations, before (Fig. 3.7(a)) and after (Fig. 3.7(b)) applying the respective Gaussian transformations. Given that the Gaussian transformation is monotonous, the bivariate characteristics of the dependence such as the correlation structure at different percentiles are no modified (Deutsch and Journel, 1996; Chilès and Delfiner, 1999). However, the linear correlation coefficient of Pearson, which depends on the kind of marginal distributions of the random variables, might be different before and after transformations. In the particular case of the variables at the locations indicated in Fig. 3.7, it is noted, at the upper right corner of each figure, that such coefficient presents nearly the same value before and after transformations. Therefore, the implicit pseudo-linearization effect associated to the Gaussian anamorphosis reported by Shöniger *et al*. (2012) should be considered application dependent.
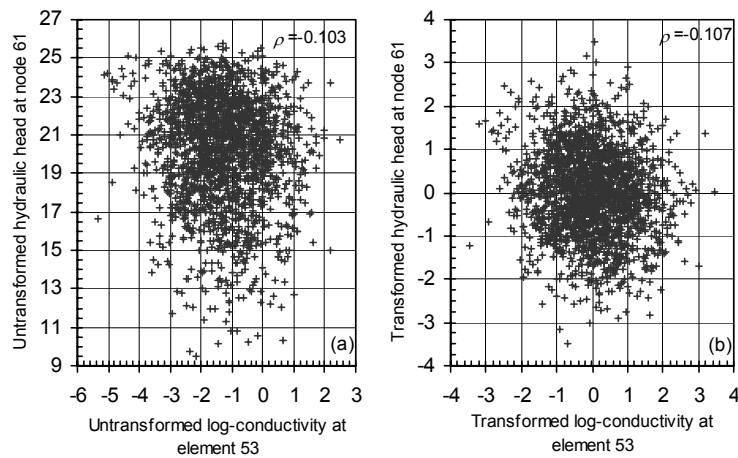
Fig. 3.7 Relationship between log-conductivity and hydraulic head at two arbitrary selected locations (ρ is the Pearson correlation coefficient). a) Before the Gaussian transformation of both variables. b) After the Gaussian transformation of both variables.

3.4.2 Effects of conditioning on log-conductivities alone

The impact of conditioning realizations of log-conductivities to direct measurements only is first analyzed. Figs. 3.8(a),(b) display comparisons of the reference field of log-conductivities against the mean of the conditional realizations of log-conductivities of the EnKF and pf-EnKF filters, respectively. Contrasting both conditional fields, it is observed that the pf-EnKF filter yields, to some extent, more variability between measurements than the EnKF filter. Looking at the RMSE and SPREAD values shown at the bottom right corner of the Figures, it can be established that the EnKF filter is more accurate and that the pf-EnKF is more precise.
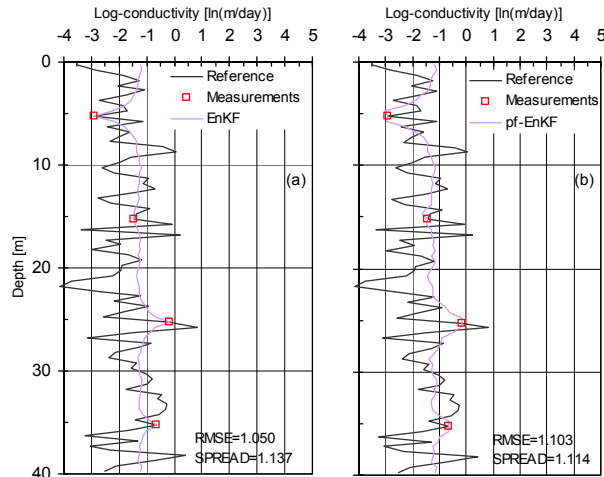
Fig. 3.8 Log-conductivity fields conditional to log-conductivities alone. a) With the EnKF method. a) With the pf-EnKF. The reference field is also shown.
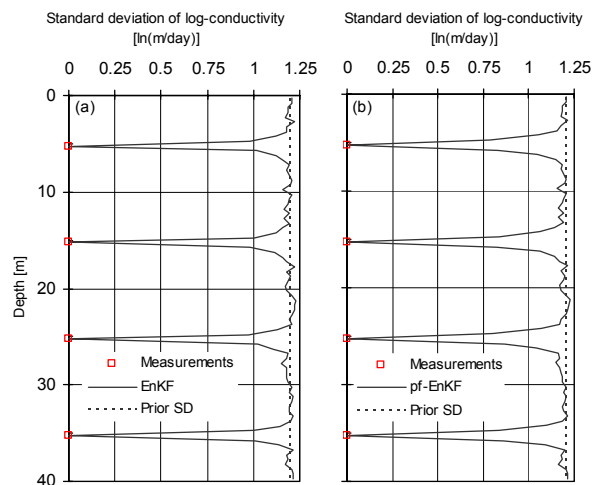
Fig. 3.9 Profiles of conditional standard deviations of log-conductivities with respect to depth at different times (empty squares indicate the locations of known values). a) From the fields of the EnKF. b) From the fields of the pf-EnKF.

Figs. 3.9(a),(b) reproduce profiles of standard deviations (uncertainty) computed with the conditional realizations of log-conductivities of the EnKF and pf-EnKF filters, respectively.

Observe that the effect of conditioning is to reduce, overall, the prior uncertainty and to collapse it to zero at the locations of the measurements. The profiles of both filters are quite similar.

3.4.3 Effects of conditioning on log-conductivities and transient heads

The additional impact of conditioning the realizations of log-conductivities to transient heads responses is now examined. Comparisons of the reference field of log-conductivities with the mean of the conditional realizations of log-conductivities of the EnKF at times t=3 days, t= 18 days and t=60 days are shown in Figs. 3.10(a),(b),(c), respectively. Figs. 3.11(a),(b),(c) shown the same comparisons but the conditional mean field is obtained with the realizations of the pf-EnKF.
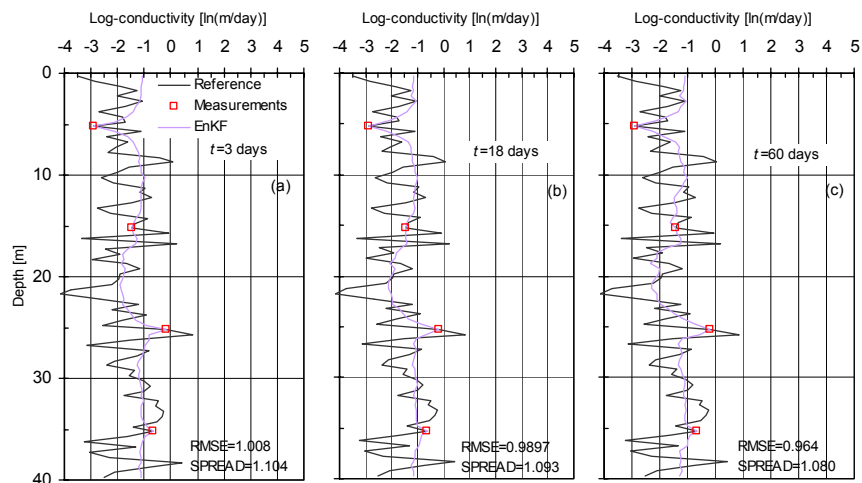
Fig. 3.10 Log-conductivity fields conditional to histories of hydraulic heads with the EnKF method. The reference field is also shown. a) At t=3 days. b) At t=18 days. c) At t=60 days.
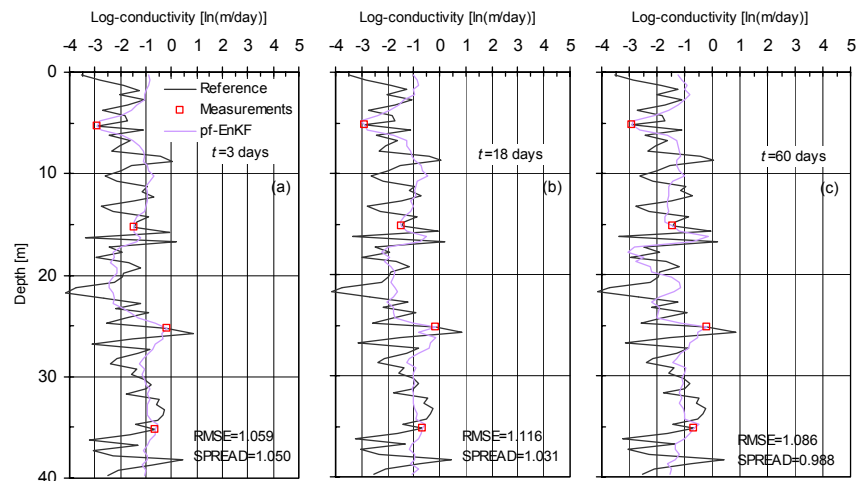
Fig. 3.11 Log-conductivity fields conditional to histories of hydraulic heads with the pf-EnKF method. The reference field is also shown. a) At t=3 days. b) At t=18 days. c) At t=60 days.

Looking at the RMSE and SPREAD measures indicated in each Figure (bottom right corner), it can be noticed that the EnKF yields more accurate results than the pf-EnKF at all times,

although the estimates of the pf-EnKF are more precise. As an example, the RMSE and SPREAD values for the conditional mean log-conductivity field of the EnKF at time *t*=60 are 0.964 and 1.080, respectively, whereas the values of such measures for the conditional mean log-conductivity field of the pf-EnKF are 1.086 and 0.988, respectively. The higher accuracy of the EnKF can be attributed to the fact that the pf-EnKF produces higher fluctuations between measurements that do not follow closer the variability of the reference aquifer. However, it should be recalled that the RMSE and SPREAD values measure the quality of the local estimation only, i.e. they do not indicate anything about the quality of the multivariate estimation. To explore the quality of the multivariate estimation, bivariate empirical copulas will be examined in chapter 4.

The profiles of standard deviations calculated with the realizations of log-conductivities of the EnKF and pf-EnKF are reported in Figures 3.12(a),(b), respectively. The overall uncertainty in both cases decreases as groundwater heads observations in more times are used into the updating process, except at the locations of direct measurements where uncertainty is zero at all times. It is illustrated in those Figures that uncertainty becomes smaller around depths 17 m and 26 m (where the tips of the two piezometers are located) than at other depths and that this reduction is more significant in the profile of the pf-EnKF than in the profile of the EnKF.
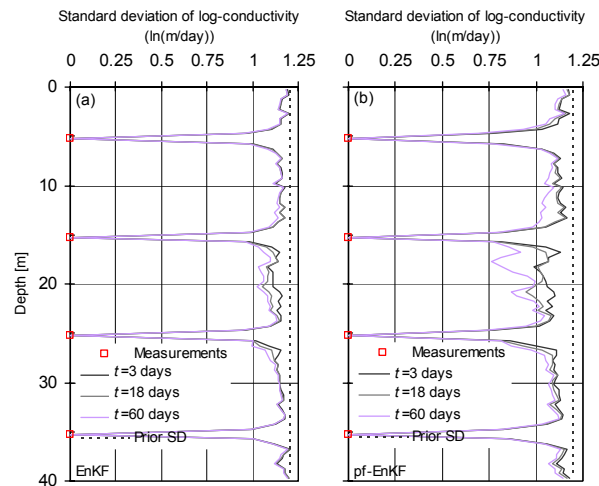
Fig. 3.12 Profiles of conditional standard deviations of log-conductivities with respect to depth at different times (empty squares indicate the locations of known values). a) From the fields of the EnKF. b) From the fields of the pf-EnKF.

Fig. 3.13 displays the frequency distributions of log-conductivities of the reference field, prior ensemble and posterior ensembles at the end of the conditioning process. Recall that the mean value of the conductivity of the reference aquifer was overestimated by the prior random field hence the distribution function of this field is located to the right of the aquifer's distribution function. Looking at the distributions of the posterior fields, it is observed that they exhibit some features of the reference distribution (like some of the "peaks" of both branches) and that are slightly displaced toward the left of the distribution of the prior random field. This indicates the attempt of the conditioning process of both filters to lead their prior distributions toward the reference distribution. Observe that the posterior distributions of both filters are very similar.
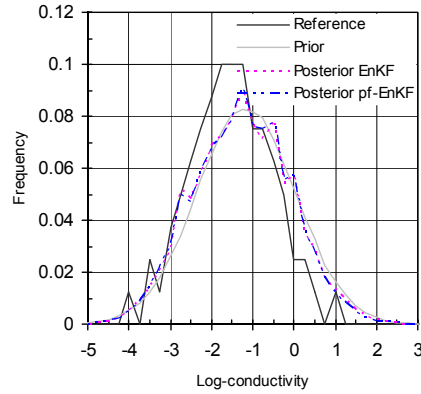
Fig. 3.13 Frequency distributions of log-conductivities of the reference field, set of prior realizations and sets of posterior realizations of the EnKF and pf-EnKF at t=60 days.

The conditional pressure head fields are now reviewed. Fig. 3.14(a),(b) and (c) plot comparisons of reference profiles with the mean profiles calculated with the conditional realizations of hydraulic heads of the EnKF at times t=3, t=18 and t=60 days, respectively. It is seen that such profiles become closer to the reference ones as a higher number of observations of the piezometric heads records are taken into account in the update step. The same behavior is observed in Figs. 3.15(a),(b) and (c), where comparisons of the reference profiles with the conditional mean profiles of the pf-EnKF at the aforementioned times are shown, except that in this case the pf-EnKF introduces higher fluctuations between observations that deviate more from the reference values. Hence, the estimations of the pf-EnKF are less accurate than those of the EnKF. This can be verified by observing the values of the RMSE measure indicated in the figures. The estimated pressure head profiles of the pf-EnKF nevertheless correspond well with the reference profiles. It can also be noted from the SPREAD values at the specified times that the estimations of the pf-EnKF are more precise.
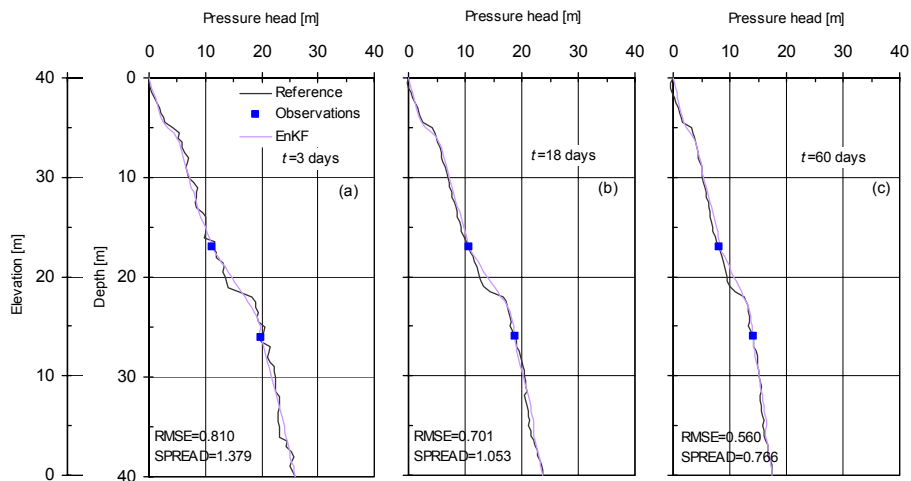
Fig. 3.14 Conditional pressure head fields of the EnKF method. The reference field is also shown. a) At t=3 days. b) At t=18 days. c) At t=60 days.
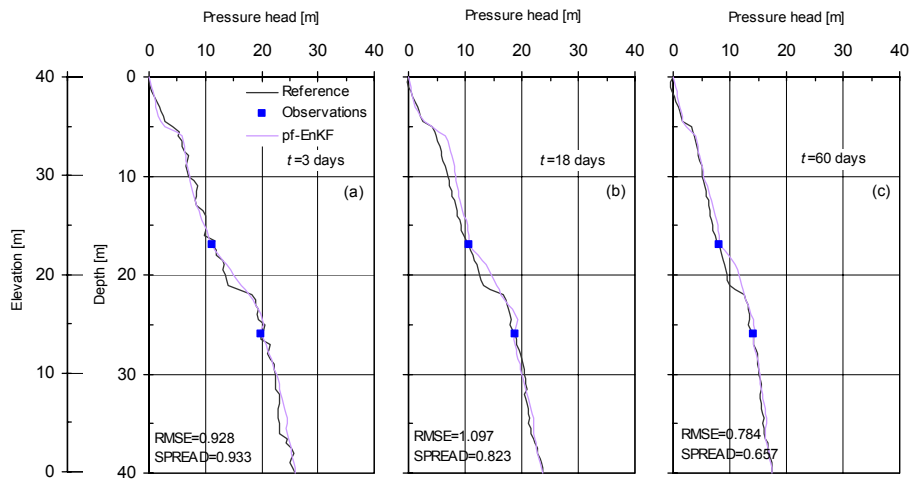
Fig. 3.15 Conditional pressure head fields of the pf-EnKF method. The reference field is also shown. a) At *t*=3 days. b) At *t*=18 days. c) At *t*=60 days.

Figs. 3.16(a),(b) illustrate profiles of conditional standard deviations of pressure heads at times *t*=3, *t*=18 and *t*=60 days quantified with the realizations of the EnKF and of the pf-EnKF, respectively. As expected, uncertainty is nil at depths where pressure heads are recorded and at the upper and lower boundaries where hydraulic heads are prescribed. As depicted in both Figures, standard deviation decreases overall with time, but locally it increases at a few depths. This behavior is evident in the profile calculated with the fields of the pf-EnKF. Such behavior may be explained by the numerical nature of the calculated covariance functions in the EnKF schemes. As explained by Xu *et al*. (2013), numerically calculated covariance functions result in fluctuating covariance estimates about zero at distances for which it should be zero. One way to overcome this problem is through the use of covariance localization techniques.
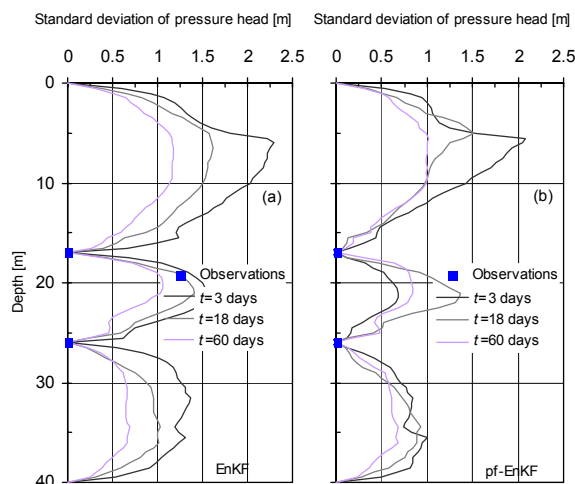

Fig. 3.16 Profiles of conditional standard deviations of pressure heads with respect to depth at different times (solid squares indicate the locations of two observations). a) From the fields of the EnKF. b) From the fields of the pf-EnKF.

3.5 Conclusions

Details about the implementation of an update step within the EnKF procedure based on *p*-field simulation were described in this chapter. This extension of the EnKF method was called pf-EnKF method. To illustrate the effects and to compare the results of the proposed method against those from the modified EnKF, a one-dimensional, single phase flow problem in continuous random porous media was considered.

The modified EnKF and pf-EnKF methods involve transformations of the local distribution functions into standard Gaussian distribution functions before the update steps. Each update step is thus performed within a geometrical space of univariate Gaussian distributions. This provides the filters more numerical stability during the update steps and the possibility of utilizing non Gaussian distributions for the parameters of the dynamic model. The back-transformation of the local conditional distributions to their original distributions after the update steps is also required by both filters.

The most notorious result of the proposed methodology is that induces higher fluctuations between observations in the conditional realizations. It was found however that inducing higher fluctuations around the mean estimates did not provide better results (in terms of local accuracy) at estimating log-conductivities and hydraulic heads. These conclusions can be explained by the fact that the fluctuations did not necessarily follow closer the variability of the corresponding reference fields. Further research is needed to fully assess the performance of the proposed filter. Chapter 4 will explore the performance of the proposed methodology on the reproduction of the spatial copulas of a reference 2D log-conductivity field.

CHAPTER 4

CHARACTERIZATION OF NON-HOMOGENEOUS HYDRAULIC CONDUCTIVITY FIELDS

It was established since the beginning of this study that the hydraulic conductivity of porous formations is very sensitive to variations of internal structure. It was mentioned that even in a same type of soil, hydraulic conductivity may vary by several orders of magnitude. Furthermore, it was pointed out that the number of available measurements for determining the hydraulic conductivity of a particular site is usually limited. It has also been mentioned that available measurements shown that, for practical purposes, heterogeneity of the hydraulic conductivity in the porous formations can be best characterized in a probabilistic manner.

Several authors have accepted that the conductivity of natural porous formations can be interpreted as a realization of a random field (Dagan, 1989; Gelhar, 1993; Zhang, 2002; Rubin, 2003). The descriptive parameters of the random field can be obtained from the realization itself assuming the field is ergodic (Chilès and Delfiner, 1999). Estimates of the hydraulic conductivity of the porous formation at unsampled locations as well as a measure of the uncertainty of the estimation can then be made by estimation or simulation geostatistical techniques (Chilès and Delfiner, 1999).

A random field that is made specific to a set of location dependent direct measurements or indirect observations of the hydraulic conductivity of the porous formation such as sets of hydraulic heads is called conditional random field. In practice, conditional random fields are often desirable because they provide site-specific estimates of the conductivity of the porous formation and because conditioning reduces, on average, the prior uncertainties.

As mentioned before, conditional random fields can be generated by estimation or simulation techniques (Deutsch and Journel, 1996). In the later case, the challenge is to integrate the available information into a prior model of uncertainty to obtain posterior estimates of the conductivity of the porous media. Different schemes can be found in the scientific literature to address this problem under steady state flow conditions (RamaRao *et al.*, 1995; Gómez-Hernández *et al.*, 1997; Janssen *et al.*, 2006; Hendricks and Gómez-Hernández, 2002) and under transient flow conditions (Hendricks-Franssen *et al.*, 1999; Zhu and Yeh, 2005; Capilla and Llopis-Albert, 2009). In the last decade, Ensemble Kalman Filters have captured more attention

to perform the stochastic simulation of random fields conditional to dynamic observations due to its easy of implementation and acceptable computational time at dealing with large dimensional problems. The method is based on a Monte Carlo framework and permits the physical model to be incorporated in the characterization process. Thus predictions with the groundwater flow model can also be conducted.

In this chapter, the modified EnKF and the proposed pf-EnKF presented in chapter 3 are employed to perform the characterization of a non-homogeneous hypothetic aquifer with respect to their hydraulic conductivities. The distribution of conductivities in the aquifer is generated by simulation of a random field. A simulated field, as opposite to a real one, is entirely known and it is thus possible to evaluate the performance of both techniques at recovering some of their relevant statistical descriptors such as histogram, rank correlation functions and bivariate copulas. To generate the simulated field, the *V*-transformed normal copula is utilized. The hypothetic aquifer is used to generate histories of hydraulic heads by simulating a pumping test. A sub-set of such histories, which are associated to certain observations locations, is used to perform the conditioning process of both filters. To evaluate the quality of the characterization, different quantitative criteria are employed.

4.1 Statement of the problem

In the following analysis, it is considered groundwater flow in two-dimensional, fully saturated, continuous porous media with spatially variable hydraulic conductivity described by the equation:

$$\nabla\left[K_s\left(\mathbf{x}\right)\cdot\nabla H\right]+Q\left(\mathbf{x}\right)=S_s\frac{\partial H}{\partial t} \tag{4.1}$$

subject to:

initial condition; $\qquad\qquad H|_{t=0}=h_0$ $\qquad\qquad\qquad\qquad$ (4.2)

and boundary conditions:

Constant pressure head; $\qquad\quad H|_{\Gamma_D}=h_D$ $\qquad\qquad\qquad\qquad$ (4.3)

No normal flow, specific discharge; $\quad\mathbf{q}(\mathbf{x},t)\cdot\mathbf{n}(\mathbf{x})=0 \qquad\quad \mathbf{x}\in\Gamma_N$ $\qquad\qquad$ (4.4)

where: ∇ is the gradient operator; H is the hydraulic head [L], h_0 is the initial head in domain Ω [L], h_D is the prescribed head on Diritchlet boundary segments Γ_D [L]; $Q(\mathbf{x})$ is the pumping rate per unit volume of the aquifer [1/T]; $K_s(\mathbf{x})$ is the hydraulic conductivity [L/T], $\mathbf{q}(\mathbf{x},t)$ is the prescribed flux across Neumann boundary segments Γ_N, $\mathbf{n}(\mathbf{x})$ is an outward vector normal to the boundary Γ_N and S_s is the specific storage [L^{-1}].

Since hydraulic conductivity $K_s(\mathbf{x})$ is a random field, equation 4.1 becomes a stochastic differential equation and the flow response H will also be a random field at a specific time. To obtain the response of the model dynamics satisfying equations 4.1 to 4.4 in the EnKF contexts, the method of Monte Carlo is utilized in this study. First, independent, identically distributed realizations of a prior conductivity random field are generated and flow equations are solved numerically with the deterministic values of each realization using a modified version of a finite element code (Smith and Griffiths, 2004). The expectation and variance of the flow response as well as auto and cross-covariance functions between conductivities and heads are determined

over the ensembles of realizations through well-known formulas of descriptive statistics, at each time the prior random fields are updated with the EnKF and pf-EnKF.

4.2 Description of the numerical experiment

4.2.1 Spatial distribution of conductivities

The aquifer is a square domain with side of 1280 m discretized into 64x64 finite elements each of which with a length of 20 m (Fig. 4.1). A log-conductivity value $y^{ref}(\mathbf{x}_i)$ is assigned to each finite element e_i for $i=1,\ldots,4096$ according to the following procedure. First, an unconditional multi-Gaussian random field with exponential autocovariance function and isotropic correlation scale $a_x=a_y=160$ m is simulated using a modified version of the random field generator SGSIM (Deutsch and Journel, 1996). Second, the V-transform (Bárdossy, 2006):

$$v(\mathbf{x}) = \begin{vmatrix} k(g(\mathbf{x})-m)^{\alpha} \text{ if } g(\mathbf{x}) \geq m \\ (m-g(\mathbf{x})) \quad \text{if } g(\mathbf{x})<m \end{vmatrix} \tag{4.5}$$

with arbitrarily chosen parameters $m=5$, $k=2$ and $\alpha=1$ is applied to a realization $g(\mathbf{x})$ of the multi-Gaussian random field to obtain the transformed $v(\mathbf{x})$ field of equation 4.5. Third, the Gaussian distribution is imposed to the $v(\mathbf{x})$ field as $y'=G^{-1}[F_V(v)]$ where $F_V(v)$ is the empirical CDF of the $v(\mathbf{x})$ field and G[·] the theoretical Gaussian CDF. Finally, such $y'(\mathbf{x})$ field is scaled to a normally distributed y field with mean value $\mu_Y=1.223$ and standard deviation $\sigma_Y=1$ as: $y^{ref}(\mathbf{x})=\mu_Y+y'(\mathbf{x})\sigma_Y$. Each one of these values is assumed to be constant within its finite element e_i. It is worth mentioning that since $y^{ref}(\mathbf{x})$ is normal distributed, $k_s(\mathbf{x})=\exp(y^{ref}(\mathbf{x}))$ is lognormal distributed.
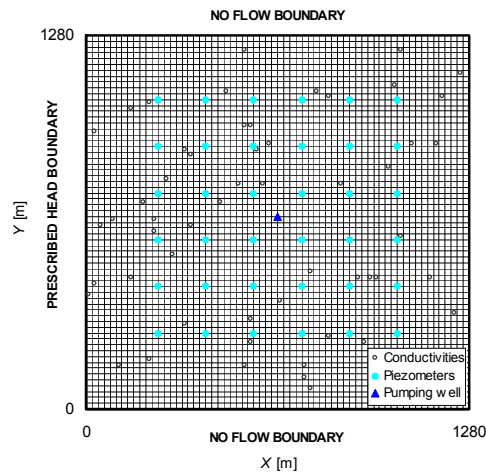
Fig. 4.1 Dimensions of the flow domain. The locations of conductivity measurements (empty circles) and of hydraulic heads observations (filled circles) are indicated. The triangle marks the location of the pumping well. Boundary conditions of the problem at $t \geq 0$ are indicated on the sides of the flow domain.

Because the statistical parameters of a simulated random field are reproduced over an ensemble of realizations, the individual realizations may show slightly different statistics but all will preserve the same characteristics of spatial dependence. These characteristics are given by the copula model with the aforementioned parameters. The bivariate density plot of this theoretical copula model is displayed in Fig. 4.2. Note that this copula density is very close to the Gaussian

copula density. Therefore, the results of the present study are limited to those cases of non-homogeneous aquifers with symmetric spatial structure of conductivities. The resulting log-conductivity field $y^{ref}(\mathbf{x})$ mentioned above is considered to represent a "true state of the nature" and is called in this study the reference field. This represents the distribution of conductivities in a hypothetic aquifer. Their statistical descriptors will be shown later (Table 4.1).
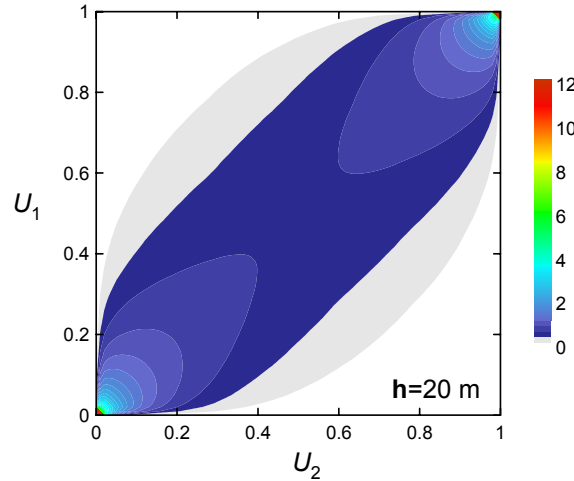
Fig. 4.2 Bivariate V-transformed theoretical copula density model with parameters m=5.0; k=2.0 and α=1.0 used to generate the reference field (hypothetic aquifer).

4.2.2 Histories of hydraulic heads

The reference field of the synthetic aquifer is utilized to generate groundwater head responses by solving a transient flow condition with finite elements. The north and south sides of the aquifer are considered no-flow boundaries whereas east and west sides are Dirichlet boundaries with prescribed heads of 200 m and 198 m, respectively. At t=0 days the initial hydraulic head is of 200 m at each node and for t>0 days a volumetric constant rate of 150 m^3/day is pumped out of the aquifer with a well located at X=640 m and Y=640 m until flow reaches steady state at approximately t=5 days. This period of time is subdivided in 20 equally sized time steps of Δt=0.25 days intervals. For the purpose of the present numerical example, it is assumed that groundwater head responses are available at times t=0.25, t=0.75, t=1.5, t=2.5, t=3.75 and t=5 days at the 36 locations indicated with filled circles in Fig. 4.1. Thus, 36 records with six head responses are generated. These indirect, informative variables of the hydraulic conductivity of the aquifer are considered available transient piezometric observations. At each one of the aforementioned six times, the updating process of both the EnKF and pf-EnKF schemes is performed, at the other times only the forecast step is performed. Storage coefficient is assumed to be equal to 0.0001 overall the aquifer.

4.2.3 Definition of the *prior* ensemble of realizations

The reference field is sampled at the locations marked with empty circles in Fig. 4.1. Such values are considered direct conductivity measurements. The mean and variance of these samples as well as the frequency distributions are reported in Fig. 4.3(a). The empirical rank correlation functions of the samples in the directions X, Y and in all directions are shown in Fig. 4.3(b). It can be observed in this figure that it is possible to fit a common theoretical correlation model to all functions, since the reference field steam from an isotropic random field. The statistical parameters of mean, variance and correlation function displayed in Fig. 4.3(a),(b), are used to parameterize a multi-Gaussian random field and generate two thousand unconditional realizations with a modified SGSIM random field simulator (Deutsch and Journel, 1996). Such

set of realizations represents an initial approximation to the hydraulic conductivity of the reference aquifer. In this way, a practical situation is modeled in which the descriptive parameters of the aquifer are roughly estimated *a priori* in the sense that they differs from the real statistics of the reference field (Table 4.1). The number of simulated realizations of the prior field ensures the stability of the error measures presented next, according to some preliminary simulations.
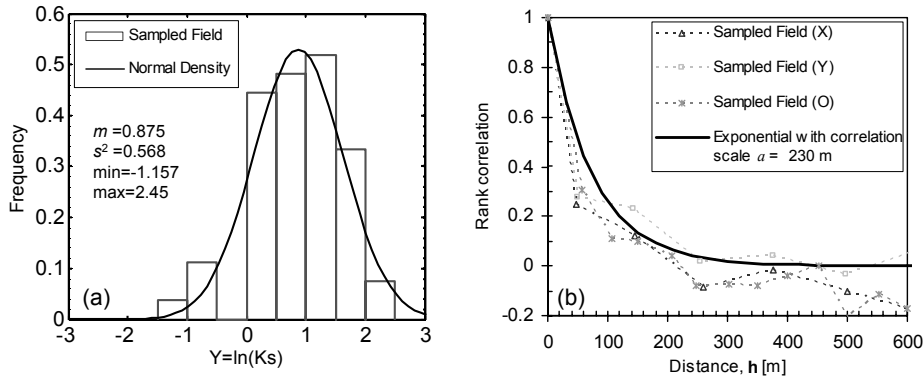
Fig. 4.3 Statistical parameters of the samples taken from the reference field. a) Histogram of the log-conductivity of the sampled field. The solid thick-line represents the fitted normal distribution function with mean and variance indicated in the figure. b) Empirical rank correlation functions of the sampled field in the *X*-direction (X) and *Y*-direction (Y). The omni-directional (O) empirical rank correlation function is also indicated. The solid thick-line represents the fitted theoretical auto-correlation model corresponding to the exponential function with the indicated correlation scale *a*.

4.3 Performance assessment criteria

The average at each location over the set of conditional realizations of any of the filters at a particular time provides an estimate of the conditional expected value of the random field at that spatiotemporal location. The map of such values is referred to as the mean log-conductivity field and it can be viewed as a "minimum variance" estimated map of the corresponding random field.

To evaluate quantitatively the accuracy of the estimation, three error measures are considered: the root mean square error (RMSE), mean absolute error (MAE) and linear error in probability space (LEPS).

The RMSE (*L*1) is evaluated by:

$$L1 = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y^{ref}(\mathbf{x}_i) - y^*(\mathbf{x}_i) \right)^2} \tag{4.6}$$

where $y^{ref}(\mathbf{x}_i)$ is the reference log-conductivity at location $\mathbf{x}_i$; $y^*(\mathbf{x}_i)$ is the estimated mean log-conductivity at location $\mathbf{x}_i$ and n is the number of log-conductivities in the flow domain. RMSE is therefore a measure of the spread of the estimated field around the reference field.

The MAE (*L*2) is computed by:

$$L2 = \frac{1}{n} \sum_{i=1}^{n} \left| y^{ref}\left(\mathbf{x}_i\right) - y^*\left(\mathbf{x}_i\right) \right| \tag{4.7}$$

The MAE is a measure of the difference of the means of the estimated and reference fields.

The LEPS ($L3$) is calculated by (Bárdossy and Li, 2008):

$$L3 = \frac{1}{n} \sum_{i=1}^{n} \left| F\left(y^{ref}\left(\mathbf{x}_i\right)\right) - F\left(y^*\left(\mathbf{x}_i\right)\right) \right| \tag{4.8}$$

where $F(y^{ref}(\mathbf{x}_i))$ is the value of the empirical distribution function of the reference field associated to the log-conductivity at location $\mathbf{x}_i$; $F(y^*(\mathbf{x}_i))$ is the value of the empirical distribution function of the estimated mean field associated to the log-conductivity at the same location $\mathbf{x}_i$. Different from the RMSE and MAE, the LEPS is no affected by the units of the variables. It varies from zero to one representing a perfect and an imperfect estimation, respectively.

To evaluate the variance of the estimation (precision), three additional error measures are determined: The *SPREAD*, average absolute deviation (AAD) and total (spatially averaged) estimation variance (Var_{tot}).

The SPREAD ($L4$) is computed as:

$$L4 = \sqrt{\frac{1}{n} \sum_{i=1}^{n} s_{en}^2(\mathbf{x}_i)} \tag{4.9}$$

where $s^2_{en}(\mathbf{x}_i)$ is the variance of the estimation obtained statistically over the ensemble of realizations at location $\mathbf{x}_i$. SPREAD is thus a measure of the dispersion of the estimated field around the reference field.

The *AAD* ($L5$) is defined as (Zhou *et al.*, 2011):

$$L5 = \frac{1}{n} \frac{1}{n_r} \sum_{i=1}^{n} \sum_{j=1}^{n_r} \left| y^*\left(\mathbf{x}_{ij}\right) - m^*\left(\mathbf{x}_i\right) \right| \tag{4.10}$$

where $m^*(\mathbf{x}_i)$ is the average of the estimation over the ensemble of realizations at location $\mathbf{x}_i$ and $y^*(\mathbf{x}_{ij})$ is the estimated log-conductivity at location $\mathbf{x}_i$ of realization j.

The *Var*$_{tot}$ ($L6$) is defined as (Shöniger *et al.*, 2012):

$$L6 = \frac{1}{n} \frac{1}{n_r} \sum_{i=1}^{n} \sum_{j=1}^{n_r} \left(y^*\left(\mathbf{x}_{ij}\right) - m^*\left(\mathbf{x}_i\right)\right)^2 \tag{4.11}$$

Var$_{tot}$ is often interpreted as a measure of the absorbed information in spatial estimation.

As mentioned earlier, criteria $L1$ to $L3$ and $L4$ to $L6$ are measures of the local accuracy and precision of the estimations, respectively. Generally speaking, the smaller the values of such norms, the more locally accurate and precise the estimations. However, these measures do not provide any information about the goodness of the estimation in terms of the reproduction of the patterns of spatial variability in the reference field. Indeed, several characteristics of the spatial

variability can be analyzed (asymmetry, entropy, connectivity, etc.) but in this study the analysis is limited to evaluate the reproduction of the empirical rank correlation functions and bivariate empirical copulas of the reference field.

To evaluate the reproduction of the rank correlation functions, correlation scales are determined:

$$I = \int_0^\infty \rho_s(\mathbf{h}) d\mathbf{h} \tag{4.12}$$

where $\rho_s(\mathbf{h})$ is the empirical rank correlation function associated to a vector $\mathbf{h}$. Integral scales of the mean rank correlation functions over all functions of the conditional fields are compared to the integral scales of the reference field. The smaller the differences between both scales, the closer the reproduction of the rank correlation function.

The rank correlation function is comparable to the semivariogram or correlogram. It expresses the *Spearman* correlation coefficient among pairs of rank values for various vectors $\mathbf{h}$. The distance where the function reaches zero is in fact the correlation scale (range) defined in geostatistics. Empirical rank correlation functions can be calculated by:

$$\rho_s(\mathbf{h}) = \frac{\frac{1}{n(\mathbf{h})} \sum_{\mathbf{x}_i - \mathbf{x}_j \approx \mathbf{h}} \left(R(\mathbf{x}_i) - m_{R(\mathbf{x}_i)} \right) \left(R(\mathbf{x}_j) - m_{R(\mathbf{x}_j)} \right)}{\sigma_{R(\mathbf{x}_i)} \sigma_{R(\mathbf{x}_j)}} \tag{4.13}$$

where $R(\mathbf{x}_i)$ and $R(\mathbf{x}_j)$ are values of the empirical distribution functions associated to the log-conductivity at locations $\mathbf{x}_i$ and $\mathbf{x}_j$ respectively; $\sigma_R(\mathbf{x}_i)$ and $\sigma_R(\mathbf{x}_j)$ are standard deviations of such values and $n(\mathbf{h})$ is the number of pair of values approximately separated by vector $\mathbf{h}$.

To check the reproduction of the bivariate distributions, Kolmogorov distances are computed:

$$D1 = \sup\left\{ \left| \hat{C}(u_1, u_2) - C(u_1, u_2) \right| ; (u_1, u_2) \in [0,1]^2 \right\} \tag{4.13}$$

where $\hat{C}(u_1, u_2)$ is the mean empirical copula over all copulas of the conditional fields and $C(u_1, u_2)$ is the copula of the reference field. These copulas are determined for different vectors $\mathbf{h}$.

4.4 Results and discussions

This section shows that groundwater head responses are worth to be taken into account in prior uncertainty models to reveal information about the spatial variability of the hydraulic conductivity of its source field. This fact is illustrated with the results of the evolution of the conditional mean estimates and of its uncertainty as well as with the results of the reproduction of the spatial variability of the so-called reference field once the conditioning process is completed (t=5 days).

4.4.1 Estimations of the conditional mean and of its uncertainty

In the experiment established in the previous section, at the end of each update step, an ensemble consisting of two thousand realizations of the log-conductivity conditional to hydraulic heads is obtained with each filter. Each set of realizations is utilized separately to calculate the conditional mean and standard deviation (uncertainty) at each location.

Fig. 4.4 plots the estimated mean log-conductivity fields of the EnKF and pf-EnKF after integrating groundwater head observations until times $t= 0.75$ days and $t= 2.5$ days. The mean log-conductivity field of the prior ensemble of realizations is also illustrated for comparison purposes. Since the prior random field is unconditional, local mean values over its ensemble of realizations approach to the expectation of the prior random field, consequently the mean prior field appears to be homogeneous. However, the information provided by groundwater head responses gradually modifies the prior spatial distribution of log-conductivities. Observe that the values of all measures of error reported to the right of the mean fields of the EnKF indicate that such fields become closer to the reference field. This does not occur with the estimated mean fields of the pf-EnKF. As a matter of fact, the norms indicate greater errors than the prior errors. Nevertheless, it will be shown further that the estimates of the pf-EnKF reproduce better the statistical parameters of the reference field than the prior estimates despite of the behavior of the norms reported in figure 4.4.
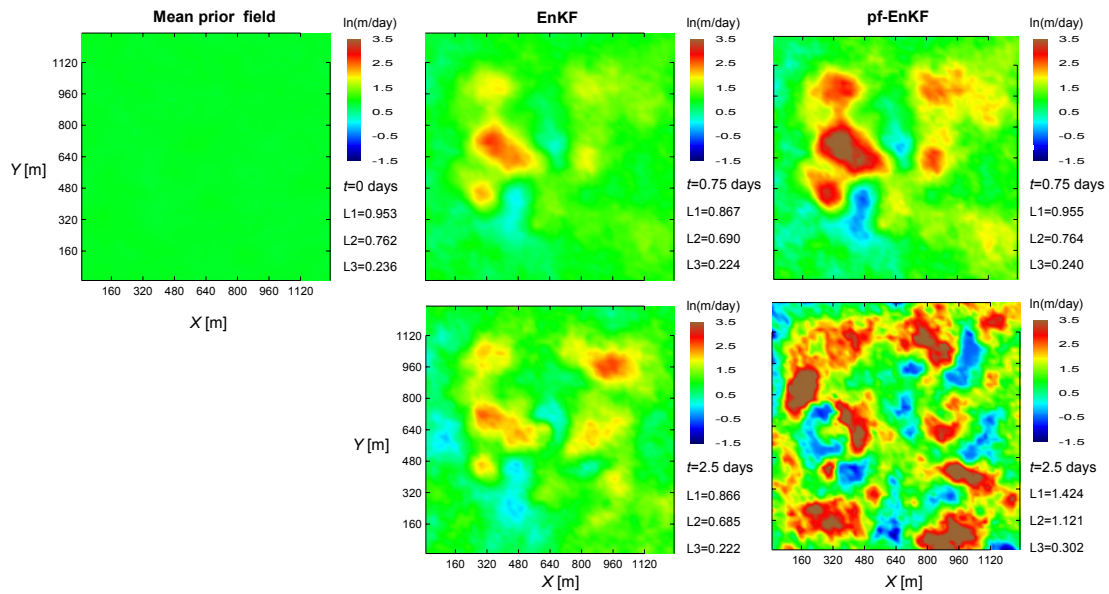
Fig. 4.4 The estimated mean log-conductivity fields of the EnKF (middle) and pf-EnKF (right) at times $t= 0.75$ days and $t= 2.5$ days. The mean prior log-conductivity field is also shown for comparison (left). The numerical values of the measures of local accuracy of the estimation are indicated to the right of the estimated mean fields. $L1$=RMSE (root mean square error), $L2$=MAE (mean absolute error) and $L3$=LEPS (linear error in probability space).
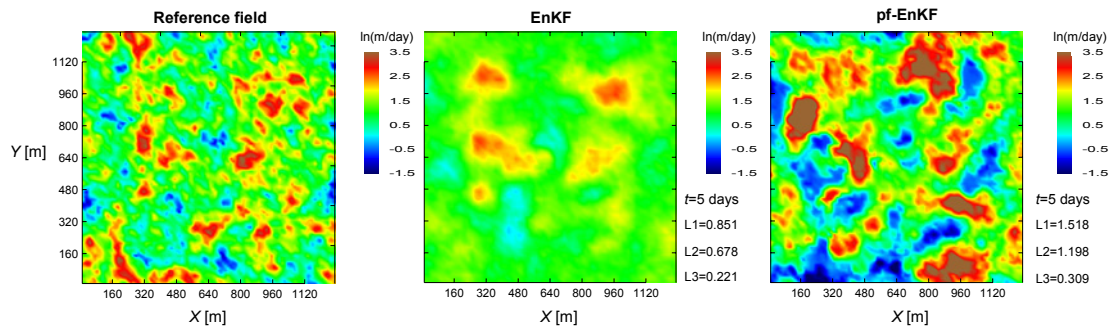
Fig. 4.5 The estimated mean log-conductivity fields of the EnKF (middle) and of the pf-EnKF (right) at time t=5 days. The reference field is shown for comparison purposes (left). The numerical values of the measures of local accuracy of the estimation are indicated to the right of the estimated mean fields. The meaning of $L1$, $L2$ and $L3$ is explained in Fig. 4.4.

After the conditioning process is completed, that is, after integrating the six groundwater head responses of each piezometer, the estimated mean log-conductivity fields reaches the final patterns displayed in Fig. 4.5. It is evident that the spatial distribution of conductivities observed at previous times (Fig. 4.4) evolve and more details of the reference field are reveled. The measures of local accuracy reported to the right of the fields, indicate that the estimation of the EnKF is more accurate than that of the pf-EnKF however the spatial distribution of conductivities of the later appears to be closer to the spatial structure of the reference field. This will be evaluated quantitatively further.

As groundwater head responses are integrated in the conditioning process, an evolution of the local prior variances also occurs. Fig. 4.6 shows the evolution of the standard deviation (uncertainty) of the log-conductivity field. It can be seen in the pictures that, overall, standard deviation decreases with time. Looking at the measures of uncertainty reported to the right of each picture, it is found that this reduction is more significant in the estimates of the pf-EnKF than in the estimates of the EnKF. The pf-EnKF is therefore more efficient than the EnKF but this efficiency does not correspond to higher accuracy, as mentioned earlier. The evolution of the standard deviation in the general case will depend on a number of factors: error in the initial approximation, number and location of piezometers, boundary conditions, etc. and the pf-EnKF seems to provide more precise estimates, although no necessarily such estimates will be more accurate locally.
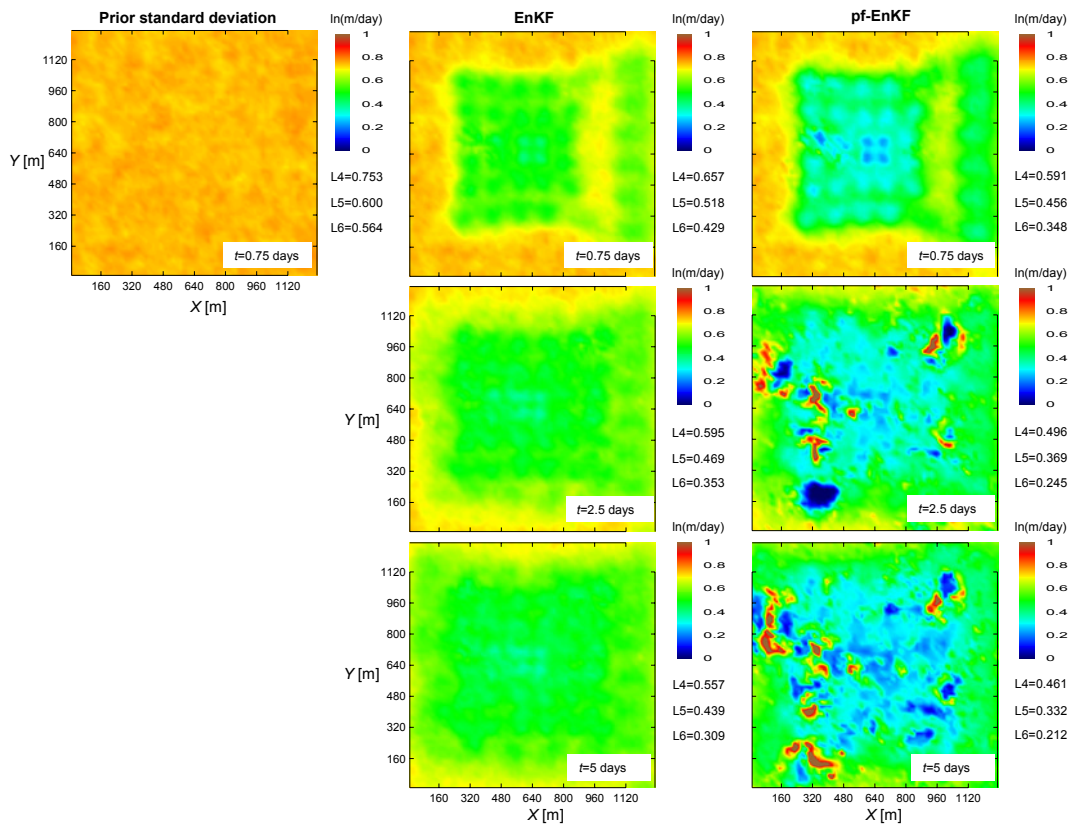
Fig. 4.6 Evolution of the standard deviation of the estimation of the log-conductivity at times $t=0.75$ days, $t=2.5$ days and $t=5$ days which is calculated with the estimates of the EnKF (left) and pf-EnKF (right). The measures of the uncertainty of the estimation are reported to the right of each image. $L1$=SPREAD, $L2$=AAD (average absolute deviation) and $L4$=Var$_{tot}$ (total prediction variance).

Figs. 4.7(a),(b) display the frequency distributions of log-conductivities of the reference field, prior ensemble and final posterior ensembles of the EnKF and pf-EnKF respectively. The global mean and global variance of each set of realizations are used to fit a theoretical normal distribution function to each frequency distribution. Observe that the prior distribution function is located to the left of the distribution of the reference field indicating that the prior distribution underestimates the global mean of the log-conductivity of the reference field. It is also noticed that the posterior distributions functions are displaced toward the right of the prior distribution function. This indicates the attempt of the conditioning process to lead the prior distribution function toward the distribution function of the reference field. It is worth to mention that the posterior distribution function of the pf-EnKF encompasses better the frequencies of extreme values of the reference field, but its variance is larger than the variance of the reference field.
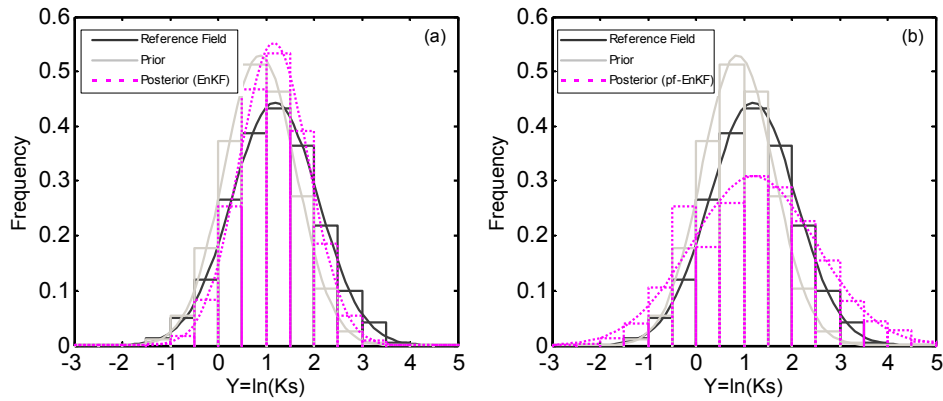
Fig. 4.7 Frequency distributions of log-conductivities. a) From the reference field, the set of prior realizations and of the set of posterior realizations of the EnKF at time t=5 days. b) From the reference field, the set of prior realizations and of the set of posterior realizations of the pf-EnKF at time t= 5 days.

4.4.2 Reproduction of the spatial variability

The empirical rank correlation functions of a set of conditional realizations at time t=5 days are compared against those of the reference field for the X-direction in Fig. 4.8(a) and for the Y-direction in Fig. 4.8(b). The average rank correlation functions of the two thousand conditional realizations of the EnKF and of the pf-EnKF are also illustrated in that figures. The average rank correlation functions of both filters exhibit stronger correlation than the correlation of the reference field in both directions, particularly at short distances. This can be explained by the performance of the linear interpolation technique of both filters. In fact, this technique utilizes a single autocovariance function as the solely descriptor of spatial variability optimizing thus the interpolation around mean values, but disregarding the interpolation at specific percentiles making the interpolation coarser. Nevertheless, the average rank correlation functions of the conditional realizations of both filters approximate better the rank correlation of the reference field than the average rank correlation functions of the prior field. Therefore, groundwater head responses certainly help to identify the spatial variability of its source field.

The results of the reproduction of the empirical copulas of the reference field are now examined. Plots of empirical copula densities (ECD) for a distance of 20 m in the X-direction are shown in Fig. 4.9. The ECD of the reference field is shown for comparison purposes in Fig. 4.9(a). Figs. 4.9(b),(c) are ECD of the mean fields of the EnKF and pf-EnKF, respectively. Observe that the higher densities of these copulas correspond well with the higher densities of the ECD of the reference field yet the ECD of the mean fields are narrower. This indicates that the correlation of the estimated mean fields is stronger than the correlation of the reference field at that distance. The correlation of the mean field of the pf-EnKF is slightly stronger than that of the

mean field of the EnKF because its ECD is narrower, but the shape of its copula resembles better the copula density of the reference field. The better reproduction of this copula density will be confirmed quantitatively further. Figs. 4.9(d),(e) illustrate the mean ECD of an ensemble of two thousand copula densities each of which established with the values of each conditional realization of the EnKF and pf-EnKF, respectively. It is evident that the shape of these ECD is quite better defined than the ECD of the mean fields. Note that these plots make more evident what has been mentioned previously in this paragraph.
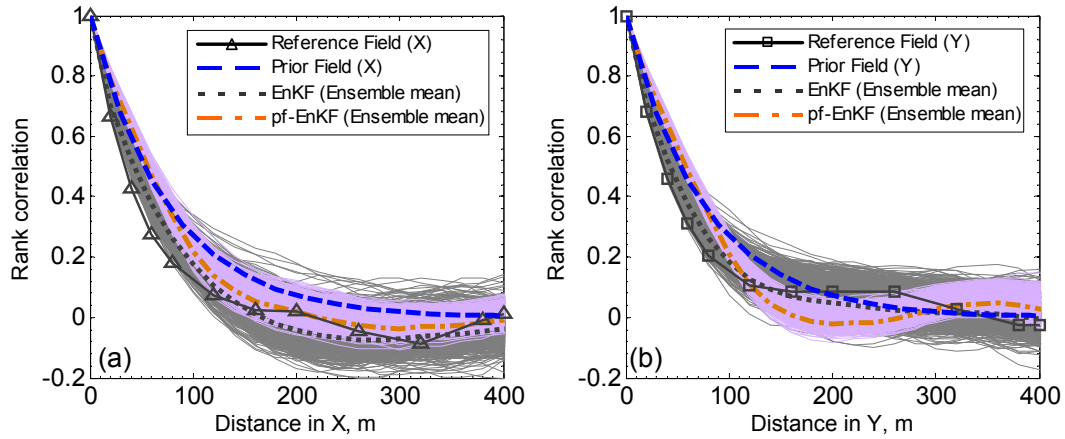
Fig. 4.8 Empirical rank correlation functions of the reference field (line with symbols), of a set of conditional realizations at time t= 5 days (thin lines) and of the average of two thousand rank correlation functions at time t= 5 days (thick lines). a) In the X-direction. b) In the Y-direction.
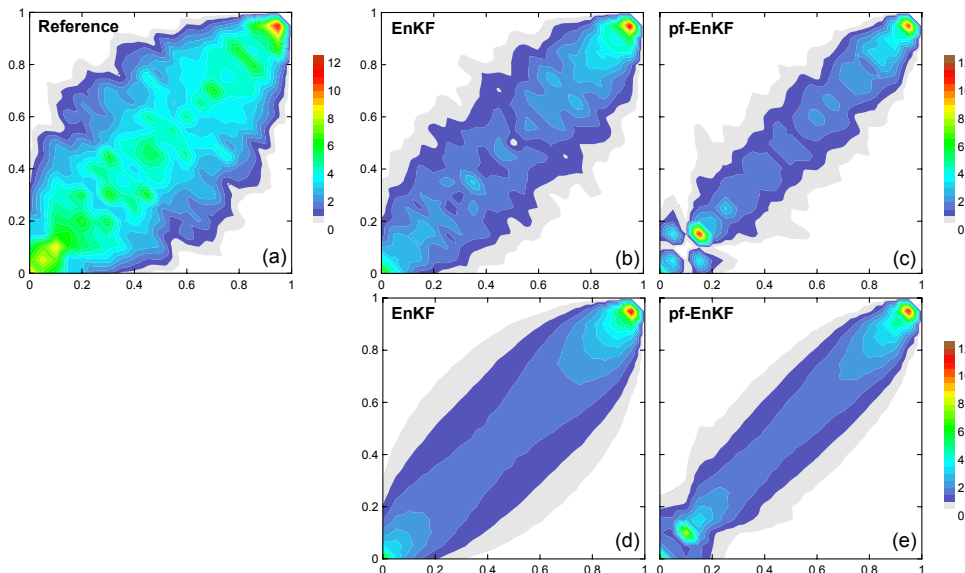
Fig. 4.9 Bivariate empirical copula densities in the X-direction direction for a distance of 20 m. a) From the reference field. b) and c) From the estimated mean field at time t= 5 days. d) and e) Average of two thousand copula densities from the conditional realizations at time t=5 days.

Fig. 4.10 displays comparisons between the copulas of the reference field and the mean copulas of an ensemble of two thousand copulas. The comparisons are representative of the copulas in the X-direction for distances of 20 m, 40 m and 80 m. To evaluate quantitatively the goodness of

63

the reproduction of the copulas of the reference field, the value of the Kolmogorov distance (*D*1) is indicated at the lower left corner of each figure. Based on these numerical values, it can be say that the fields of the pf-EnKF reproduce better the copulas of the reference field than the fields of the EnKF. Therefore, the bivariate spatial structure of the reference field appears to be better captured by the pf-EnKF.
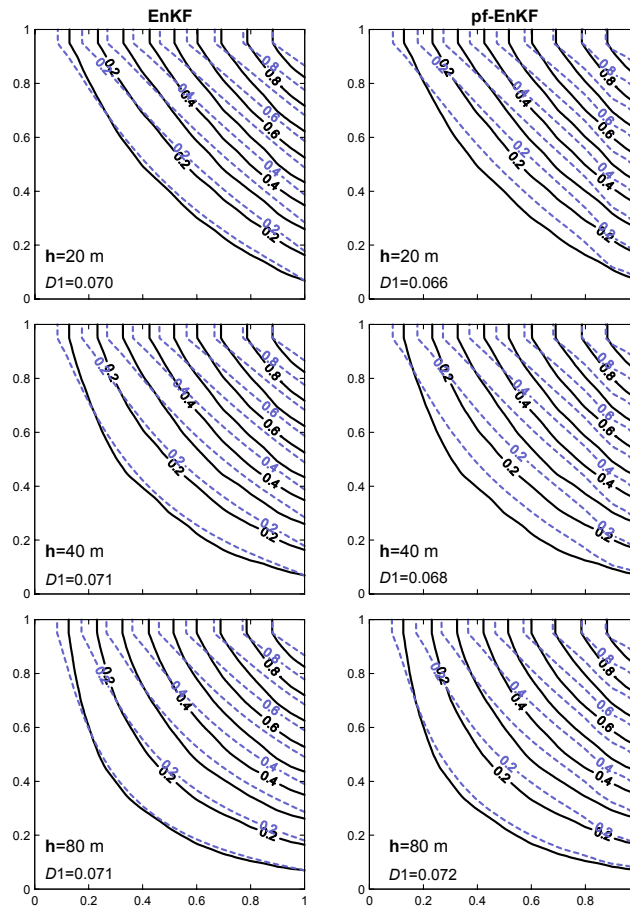
Fig. 4.10 Bivariate empirical copulas for three different distances in the *X*-direction at time *t*=5 days. In the first column are shown the mean copulas from the conditional realizations of the EnKF. In the second column are reported the mean copulas from the conditional realizations of the pf-EnKF. The values at the lower left corner of each figure are the Kolmogorov distances (*D*1).

The values of the criteria *D*1 between the copulas of the reference field and the mean copulas of an ensemble of two thousand copulas with respect to several distances in the *X*-direction is shown in Fig. 4.11(a). The same results but now from the copulas in the *Y*-direction are shown in Fig. 4.11(b). It is clear that the fields of the pf-EnKF reproduce better the copulas of the reference field at short distances in the *X*-direction and at all distances in the *Y*-direction. At long distances in the *X*-direction, the bivariate copulas of the reference field are better approximated by the fields of the EnKF.

4.4.2 Summary of results

At last, the statistical parameters of the estimated fields are compared against those of the reference field in Table 4.1. Generally speaking, the global mean of the reference field is rather well captured by both filters however its variance is underestimated by the EnKF and overestimated by the pf-EnKF. The correlation scales of the reference field are overestimated by

both filters. Observe in Table 4.1 that none of both filters performs consistently better, but they provide better estimates of the spatial structure of the reference field than the prior estimates. Therefore it seems to be convenient to consider both filters in applications.
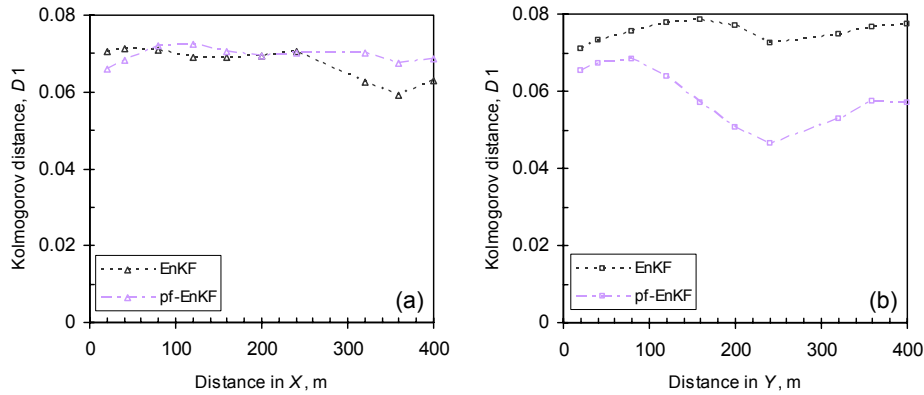
Fig. 4.11 Average of the Kolmogorov distance (D1) computed with two thousand copulas from the conditional realizations of the log-conductivity at time t= 5 days for ten different separations. a) In the X-direction. b) In the Y-direction.

Table 4.1 Comparison of the statistical parameters of the reference field against those of the prior and estimated fields after completed the conditioning process (t= 5days).

	Reference Field	Prior Fields	Estimated Fields (EnKF)	Estimated Fields (pf-EnKF)
Mean, ln(Ks(m/day))	1.19	0.87	1.16	1.20
Standard dev., ln(Ks(m/day))	0.90	0.75	0.72	1.29
Correlation scale in dir. X, m	147.5	230.0	162.1	196.5
Correlation scale in dir. Y, m	190.0	230.0	188.0	188.2

4.5 Conclusions

The benefits of integrate groundwater head responses into prior uncertainty models to characterize non-homogeneous aquifers with respect to their hydraulic conductivities employing the modified EnKF and the proposed pf-EnKF presented in chapter 3 were illustrated in this chapter. A hypothetic aquifer, called the reference field, wherein the distribution of conductivities was generated by simulation of a random field with the V-transformed normal copula was considered for illustrative purposes. The descriptive parameters of a prior multi-Gaussian model were estimated by sampling of the reference field. Groundwater head responses were generated from the solution of a transient groundwater flow problem over the reference field and a set of histories of heads recorded at specified locations was integrated with the conditioning process of both filters.

The quality of the reproduction of the univariate and bivariate statistics of the reference field was evaluated with different quantitative criteria. None of both filters was found to perform consistently better, but the results provided by them showed that groundwater head responses were capable of revealing information about the spatial variability of the distribution of conductivities in the reference field. The conditioning processes of both filters enhanced the prior estimates.

The reference field used in this study was in fact only an idealization of a true aquifer which will certainly be quite more intricate and also three-dimensional, but the numerical techniques that have been presented will provide useful approximations in practical situations.

CHAPTER 5

DETECTION OF PREFERENTIAL SEEPAGE PATHS IN A TRIANGULAR DAM CORE

Preferential seepage is a term utilized to describe concentrated movements of fluid in the flow domain at contrastingly high seepage velocities. This type of seepage occurs in heterogeneous media because fluid moves faster along paths of least resistance. The least resistive pathways are associated to continuous zones of higher hydraulic conductivities than those of the surrounded zones. The presence of continuous high conductive zones in earthen structures has different sources (ICOLD, 1995, 1997, 2004; Fell and Wan, 2005). They can be the result of one or some of the following events: soil cracking, ongoing internal erosion and continuous permeable zones containing coarse and/or poorly compacted materials, to mention a few.

To observe preferential seepage paths in the core of an earth dam, it is sufficient in principle to characterize the distribution of conductivities and to analyze seepage velocities for a specific seepage condition. However, this is not a simple matter because in practice only a limited number of direct conductivity measurements are available. In addition, the distribution of conductivities in embankments does not depend on the intrinsic properties of materials and compaction quality only, but also on events such as those aforementioned.

To characterize hydraulic conductivity fields, its spatial variability is often interpreted as a realization of a random field (Dagan, 1989; Gelhar, 1993; Zhang, 2002; Rubin, 2003). The parameterization of the random field is obtained based on the observations of the realization itself assuming the field is *ergodic* (Deutsch and Journel, 1992; Chilés and Delfiner, 1999). Then estimations and the uncertainty associated to them are determined at unobserved locations by either estimation or simulation geostatistical techniques (e.g. Journel and Huijbregts, 1978; Deutsch and Journel, 1992; Chilés and Delfiner, 1999).

The multi-Gaussian random field model has been widely used for representing spatial variations of the hydraulic conductivity of soils (Fenton and Griffiths, 1996, 1997; Gui *et al.*, 2000; López-Acosta and Auvinet, 2003, 2004; Ahmed, 2009). Such model is often justified in practice only on the basis of a univariate lognormal distribution of conductivities (e.g. Freeze, 1975; Hoeksema and Kitanidis, 1985; Fogg, 1986; Woodbury and Zudicky, 1986; Gelhar, 1993; López-Acosta and Auvinet, 2011). The higher order densities, i.e., the bivariate, trivariate, *n*-

variate densities are "sightlessly" assumed. In other words, the decision of considering multi-Gaussian dependence is not data based but it has the advantage of its mathematical simplicity and easy inference.

The phenomena of preferential seepage paths suggest that spatial variability of conductivities holds characteristics of non multi-Gaussian dependence (Ch. 1). In fact, it is well-known that in multi-Gaussian random fields the most continuous paths have conductivities around the mean (Silliman and Wright, 1988). Thus, connected paths of extreme values do not occur in multi-Gaussian random fields (Journel and Alabert, 1989; Journel and Deutsch, 1993; Gómez-Hernández and Wen, 1998; Journel and Zhang, 2006). However, preferential seepage paths are the result of the way high and low conductivities are spatially interconnected over the flow domain (Fogg, 1986; Bradbury and Muldoon, 1990; Webb and Anderson, 1996; Teles *et al.*, 2004; Zappa *et al.*, 2006; Sánchez-Vila *et al.*, 2006). Therefore, random fields with non multi-Gaussian dependence have to be considered to conveniently represent continuity of extreme values in hydraulic conductivity fields.

Conductivity random fields with non multi-Gaussian dependence characteristics such as higher continuity for high/low values can be generated using different approaches (Journel and Huijbregts, 1978; Journel, 1983; Yamazaki and Shinozuka, 1988; Christakos, 1990; Grigoriu, 1998; Sánchez-Vila *et al.*, 1996; Popescu *et al.*, 1998; Chilès and Delfiner, 1999; Emery, 2008). In practice, the parameterization of these models is not always straightforward due to lack or inexistence of direct conductivity measurements. Therefore, multi-Gaussian models have to be considered a priori most of the times. These models however can be improved by integrating indirect observations of the hydraulic conductivity.

In fact, hydraulic head observations can be related to the actual distribution of conductivities by solving the inverse problem. An inverse solution requires that the hydraulic head field obtained from the seepage model honor observations of the hydraulic head themselves at the observed locations. Different inversion techniques are found in the scientific literature yet the stochastic simulation of conductivity fields conditional to dynamic variables such as piezometric heads has received considerable attention (e.g. Gómez-Hernández *et al.*, 1997; Zhu and Yeh, 2005; Alcolea *et al.*, 2006; Capilla and Llopis-Albert, 2009).

The purpose of this chapter is to characterize a non multi-Gaussian field that mimics continuous zones of high hydraulic conductivities inside a hypothetic earth dam. The characterization of such conductivity field is of great interest in practice because this would permit the detection of preferential seepage paths by visual inspection of the associated velocity field. For such purpose, a solution to the stochastic inverse problem is obtained by means of the modified Ensemble Kalman Filter (EnKF) presented in chapter 3. The results are compared against those obtained from the proposed pf-EnKF. In the examples presented in this chapter, the prior distribution of conductivities in the flow domain is represented by an ensemble of unconditional multi-Gaussian realizations. Hence a situation will be modeled wherein the presence of continuous zones of high conductivities is ignored *a priori* and the only source of information about the actual distribution of conductivities in the core will be provided by histories of piezometric heads.

The methodology is illustrated on a reference synthetic field that describes the distribution of saturated log-conductivities in a triangular dam core (Harr, 1962; Poluvarinova-Kochina, 1962; Bear *et al.*, 1968). Such field is assumed to represent a "true state of the nature" in which continuous zones of high conductivities are apparent. To generate synthetic histories of hydraulic heads, a transient flow condition is solved by modeling a "rapid" drawdown of the reservoir. Then the hydraulic head field is sampled at selected locations with specified intervals of times. The set of conductivity realizations conditional to hydraulic head observations are used to make predictions of velocity fields. Preferential seepage paths are identified by visual inspection of these fields. The error of the predictions is also analyzed.

The chapter is organized as follows. The problem is stated in section 5.1; both reference fields of conductivities and hydraulic heads are described in section 5.2 and the stochastic inversion methods are presented in section 5.3; results are discussed in section 5.4, before concluding the study in section 5.5.

5.1 Statement of the problem

The study is conducted on the internal core of a rockfill dam which is idealized with a triangular cross-section with full head of water as in Harr (1962) (Fig. 5.1). Seepage through triangular sections has been considered by several authors (Poluvarinova-Kochina, 1962; Bear *et al.*, 1968; Collins, 1971). In these cases seepage through the internal core is analyzed only, since hydraulic head losses through rock slopes are negligible for practical purposes.
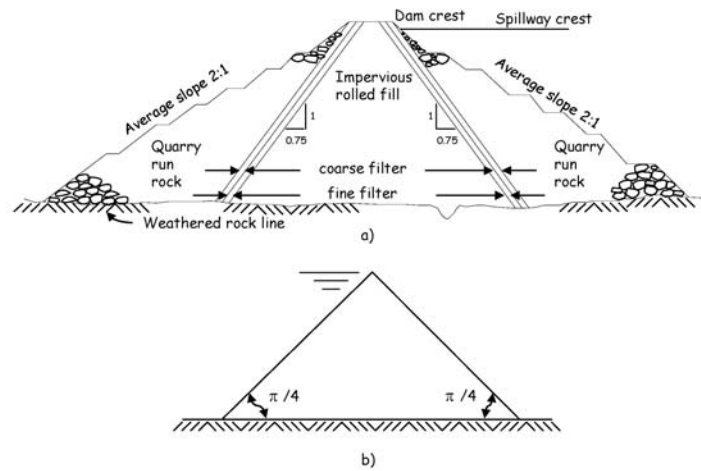
Fig. 5.1 Rockfill dam and the idealization of the cross-section of its internal core (Harr, 1962). a) Rockfill dam. b) Idealization.

In the following analysis, it is assumed that seepage through the dam core is described by the following continuity equation and Darcy's law (e.g. Bear, 1972; Freeze and Cherry, 1979; de Marsily, 1986):

$$\frac{\partial}{\partial x}\left[K_s(\mathbf{x})\frac{\partial H}{\partial x}\right] + \frac{\partial}{\partial y}\left[K_s(\mathbf{x})\frac{\partial H}{\partial y}\right] = S_s\frac{\partial H}{\partial t} \tag{5.1}$$

subject to:

initial condition;
$$H_{t=0} = h_0 \tag{5.2}$$

and boundary conditions:

Constant hydraulic head;
$$H_{\Gamma_D} = h_D \tag{5.3}$$

No normal flow, specific discharge; $\mathbf{q}(\mathbf{x},t)\cdot\mathbf{n}(\mathbf{x}) = 0 \qquad \mathbf{x} \in \Gamma_N \tag{5.4}$

Seepage face;
$$H_{\Gamma_D} = z_D \tag{5.5}$$

where: H is the hydraulic head [L], h_0 is the initial head in domain Ω [L], h_D is the prescribed head on Diritchlet boundary segments Γ_D [L]; $K_s(\mathbf{x})$ is the hydraulic conductivity [L/T], $\mathbf{q}(\mathbf{x},t)$ is the prescribed flux across Neumann boundary segments Γ_N, $\mathbf{n}(\mathbf{x})$ is an outward vector normal to the boundary Γ_N and S_s is the specific storage [L^{-1}]. In this study, the log-hydraulic conductivity is considered as a random field, while specific storage is treated as a deterministic constant. The numerical solution of the seepage problem described by equations 5.1-5.5 is obtained by finite elements (Istok, 1989; Smith and Griffiths, 2004).

5.2 Description of the numerical experiment

5.2.1 Log- conductivity field

The distribution of saturated log-conductivities $Y(\mathbf{x})=ln(K_s(\mathbf{x}))$ in the core is assumed to be described by a non multi-Gaussian random field with marginal Gaussian distribution functions. The mean of such field is -1.193 and its variance is unity. Thus the distribution of saturated conductivities have a mean of K_s=0.5 m/day and a coefficient of variation CV=131%. The random field is also assumed to be strictly stationary characterized by a *Spearman* rank correlation function given by the spherical correlogram with isotropic correlation range a=15 m (Deutsch and Journel, 1992; Genest and Favre, 2007). In addition, the random field is considered to present an asymmetric dependence structure which can be described by means of the *V*-transformed copula (Bárdossy and Li, 2008; Li and Bárdossy, 2009) with parameters m=0.0, k=2.5 and α=0.25.
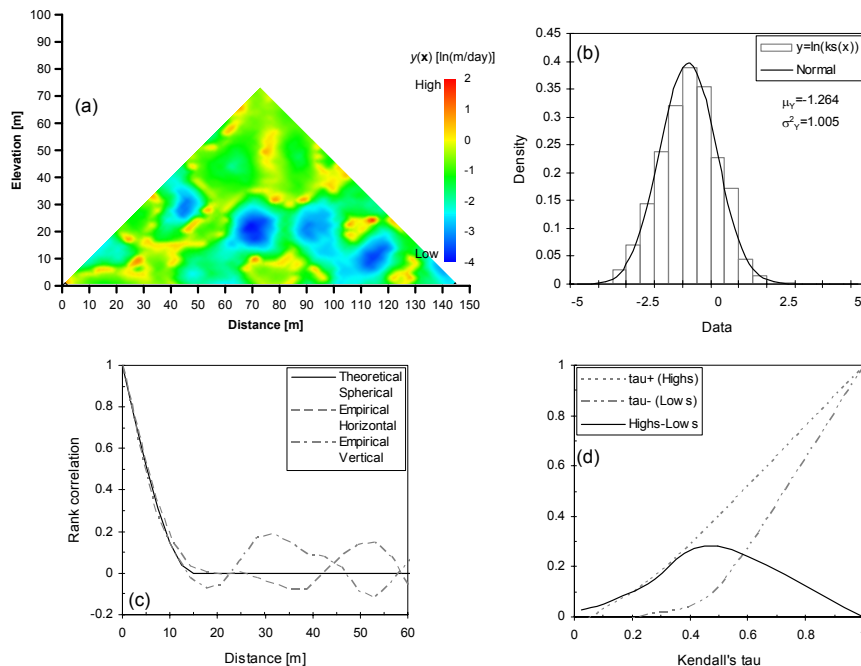
Fig. 5.2 Log-conductivity realization of a random field generated by means of the *V*-transformed copula with parameters m=0.0; k=2.5 and α=0.25. Their statistical descriptors are also shown. a) Realization of log-conductivities $Y(\mathbf{x})=ln(Ks(\mathbf{x}))$. b) Histogram. c) Rank correlation functions in the horizontal and vertical directions. d) Conditional Kendall's tau correlation functions (Manner, 2010) in the horizontal direction.

Based on the aforementioned information, the *V*-transformed copula is used to generate a set of unconditional log-conductivity realizations over a square domain of 102.5 m of side evenly

subdivided in 41x41 square elements. One representative realization of that set is chosen and the values over the dam core are mapped onto a finite element mesh assuming each value remains constant within its element. The characteristic of the finite element mesh are explained further. This realization is assumed to represent a "state of the nature" in which continuous zones of high conductivities are present in the core. Another realization in the ensemble would in fact show a different distribution of log-conductivities but it would preserve the same characteristics of the spatial variability structure.

The representative realization and its statistical descriptors are shown in Fig. 5.2. An exceedance correlation definition based on the *Kendall's* tau correlation coefficient is considered for quantifying the asymmetry of such field (Manner, 2010). The solid line in Fig. 5.2(d) represents the asymmetry of the spatial dependence. The more continuous and pervious zones in the realization have conductivities $K_s(\mathbf{x})$ about 20 times higher than the mean. This characteristic is consistent with a recent study on the spatial distribution of conductivities in the core of an actual earth dam where preferential seepage has been detected (Smith and Konrad, 2011). In addition, the specified correlation range is believed to be reasonable because under current construction practices, no significant continuity of log-conductivities is expected in the cross-section of the core, since material is compacted in lifts along its longitudinal section.
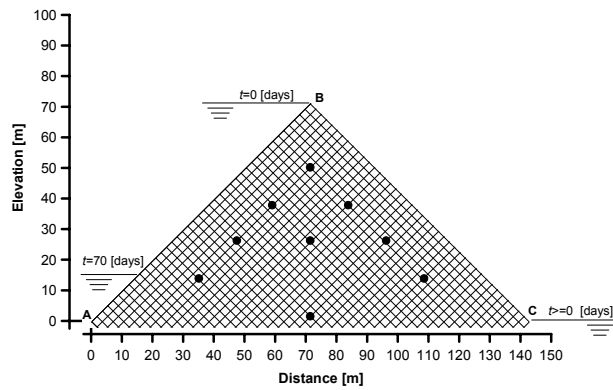
Fig. 5.3 The discretized flow domain and the locations of nine piezometers (filled circles).

5.2.2 Histories of hydraulic heads

To generate histories of hydraulic heads, a transient flow condition in the core is solved by prescribing hydraulic heads at upstream and downstream surfaces and assuming an impervious boundary at the base. The core is subdivided in 861 isoparametric finite elements of 2.5x2.5 m per side (Fig. 5.3). These are populated with the log-conductivity field of Fig. 5.2(a), as explained. The transient flow condition is modeled as follows. At t=0 day, the initial head is at steady state (Fig. 5.4(a)). At t>0 day, a drawdown with constant rate of 0.81 m/day is prescribed to the upstream surface until it reaches the elevation shown in Fig. 5.3. This transitory seepage condition is modeled assuming full saturation of the core during the drawdown of the reservoir (rapid drawdown) (Fig. 5.4(b)).
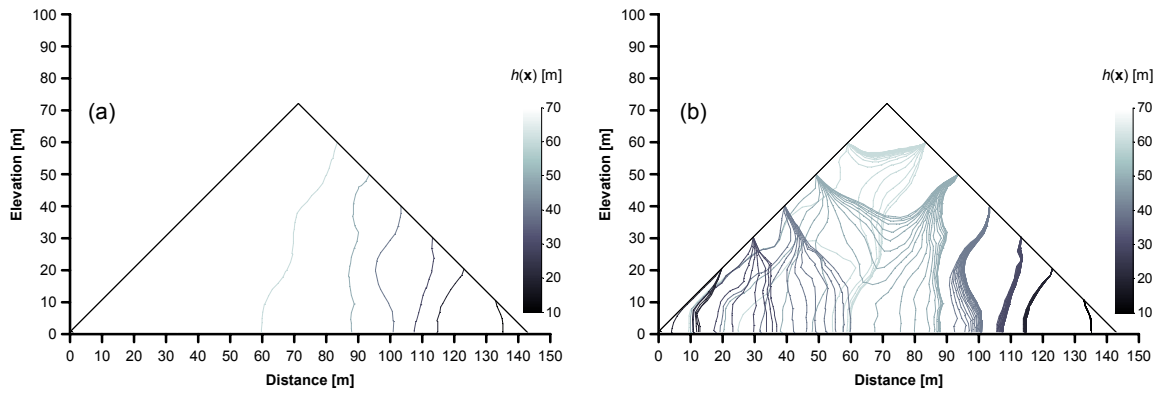
Fig. 5.4 The hydraulic head reference fields [m]. a) At steady state. b) At transient state after a "rapid drawdown" of the reservoir i.e. assuming the material remains saturated.

The prescribed drawdown is reached at t=70 days. This period is chosen as the simulation duration and is subdivided in 40 time steps of 1.75 days interval. Piezometric heads are observed at the nine locations shown in Fig. 5.3, at times t=t_1, t=t_2 and thereafter at an evenly spanned time interval of 3.5 days. Hence each data set considered for stochastic inversion consists of 21 observations. The history of each piezometer is shown in Fig. 5.5. The observations are assumed to be error-free.
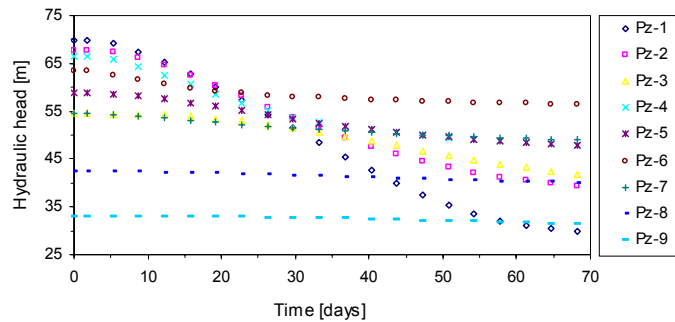
Fig. 5.5 Histories of hydraulic heads.

5.2.3 Initial ensembles of log-conductivities

Two sets of one thousand unconditional realizations of the log-conductivity random field are generated with a modified version of the SGSIM code (Deutsch and Journel, 1992). The statistical descriptors of each ensemble are shown in Table 5.1. The ensemble size is chosen based on the stability of the numerical results to different error measures (chapter 3). Note that in both cases considered in Table 5.1 the actual asymmetry of the reference field is ignored, since the realizations are multi-Gaussian. In other words, the presence of continuous zones of high conductivities in the core is ignored *a priori*. For case 2, is also assumed a less heterogeneous and less continuous spatial structure than that of the reference field. It is worth to mention also that in both cases shown in Table 5.1, only the histories of hydraulic heads by themselves will modify the prior spatial structure of log-conductivities in the core; that is, it is considered that direct conductivity measurements are unavailable or inexistent.

72

Table 5.1 Statistics of the initial ensembles of unconditional log-conductivities $Y(\mathbf{x})=ln(K_s(\mathbf{x}))$ used for inverse modeling.

Case	Ensemble Mean	Ensemble variance	Correlation structure[1]	Asymmetry structure
1	-1.193	1.0	a=15 m	none
2	-1.193	0.8	a=13 m	none

[1]Spherical with isotropic correlation range a

5.2.4 Numerical implementation

The stochastic inverse model is coded in FORTRAN programming language on LINUX platform and run in the HPC cluster "Tonatiuh" at the Institute of Engineering, UNAM.

5.3 Conductivity fields conditional to histories of hydraulic heads

In the subsequent analysis, an ensemble of n_r realizations of the prior random field of the log-conductivity at a particular time $t=0$ i.e. $Y_0(\mathbf{x})$ is considered. The dimension of this field is n. Each realization of such random field is used to solve flow equations of the seepage model and an ensemble of realizations of the random field of hydraulic heads at time $t=1$ i.e. $H_1(\boldsymbol{\chi})$ is thus obtained. The dimension of this field is N.

Once all marginal distributions of both $Y_0(\mathbf{x})$ and $H_1(\boldsymbol{\chi})$ have been transformed numerically into Gaussian distributions according to the procedure discussed in chapter 3, one can define for each realization in the ensemble, a vector of dimension n having prior values of the log-conductivity i.e. $\tilde{\mathbf{U}}_0 =[y'_{t-1}(\mathbf{x}_1),y'_{t-1}(\mathbf{x}_2),\ldots,y'_{t-1}(\mathbf{x}_n)]$, a vector of dimension N_h constituting by observations i.e $\tilde{\mathbf{Z}}_t =[h'_{1,t},h'_{2,t},\ldots,h'_{Nh,t}]$ and a reduced vector of dimension N_h comprising forecasted states (prior values of the hydraulic head at the locations of the observations) i.e. $\tilde{\mathbf{H}}_t^f =[h'_t(\boldsymbol{\chi}_1),h'_t(\boldsymbol{\chi}_2),\ldots,h'_t(\boldsymbol{\chi}_{Nh})]$.

An updated realization of log-conductivities at time $t=1$ i.e. $\tilde{\mathbf{U}}_t^u =[\hat{y}'_t(\mathbf{x}_1), \hat{y}'_t(\mathbf{x}_2),\ldots,\hat{y}'_t(\mathbf{x}_n)]$ can be computed by (Vázquez *et al.*, 2014):

$$\tilde{\mathbf{U}}_t^u = \tilde{\mathbf{U}}_0 + \tilde{\mathbf{K}}_t\left[\tilde{\mathbf{Z}}_t - \tilde{\mathbf{H}}_t^f\right] \tag{5.6}$$

where $\mathbf{K}_t$ is a matrix of size n x N_h consisting of weighting functions $\lambda_l(\boldsymbol{\chi}_\alpha)$ obtained from the solutions of the following system of equations:

$$\sum_{l=1}^{n}\sum_{\alpha=1}^{N_h} \lambda_l\left(\boldsymbol{\chi}_\alpha\right)C_{H'}\left(s,\tau\right) = C_{Y'H'}\left(s,\tau\right) \tag{5.7}$$

where $C_{H'}(s,\tau)$ are auto-covariance functions between piezometric heads at $s=(\boldsymbol{\chi}_\alpha,\boldsymbol{\chi}_j)$ for $j=1,\ldots,N$ and $\tau=t=1$; $C_{Y'H'}(s,\tau)$ are cross-covariance functions between log-conductivities and hydraulic heads at $s=(\mathbf{x},\boldsymbol{\chi}_j)$ and $\tau=(t=0,t=1)$. The functions are determined statistically over the ensembles of realizations.

After obtaining the weighting functions $\lambda_l(\chi_\alpha)$ for each location $\mathbf{x}$, the matrix $\mathbf{K}_t$ can be assembled as:

$$\mathbf{K}_t = \begin{bmatrix} \lambda_1(\chi_1) & \lambda_1(\chi_2) & \cdots & \lambda_1(\chi_{N_h}) \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_n(\chi_1) & \lambda_n(\chi_2) & \cdots & \lambda_n(\chi_{N_h}) \end{bmatrix} \tag{5.8}$$

In addition to the update step of the EnKF method presented above, an update step based on *p*-field simulation can be performed as proposed in chapter 3. This extension of the EnKF method is called from now on pf-EnKF method.

A CCDF of log-conductivities at each location $\mathbf{x}$ i.e. $F_{\hat{Y}'}(\hat{y}';\mathbf{x}_i)$ for $i=1,2,\ldots,n$ is established statistically over the ensemble of realizations updated with the modified EnKF method and samples of log-conductivities at each location $\mathbf{x}$ i.e. $\hat{y}'_t(\mathbf{x}_i)$ for $i=1,2,\ldots,n$ can then be drawn from their corresponding conditional distributions by the *p*-field simulation technique as (Srivastava, 1992; Froidevaux, 1993):

$$\hat{y}'_t(\mathbf{x}_i) = F_{\hat{Y}'}^{-1}\left(p_{t-1}(\mathbf{x}_i)\right), \quad i = 1,2,\ldots,n \tag{5.9}$$

where $p_{t-1}(\mathbf{x}_i)$ are so-called probability fields (*p*-fields) of the log-conductivity. The *p*-fields can be established with the values of the empirical distribution functions of each realization before the update step as explained in Ch. 3.

After back-transforming each local cumulative distribution of the updated log-conductivity random field following the procedure described in chapter 3, a vector $\mathbf{Y}_t^u(\hat{y}_1(\mathbf{x}),r)$ of dimension n consisting of n_r back-transformed, log-conductivity realizations $\hat{y}_1(\mathbf{x})$ is formed. Such vector consist of an ensemble of posterior realizations of the random field of the log-conductivity at time $t=1$ i.e. $Y_1(\mathbf{x})$. The conditional mean $\hat{\mu}_{Y|H}(\mathbf{x};t_1)$ and conditional variance $\hat{\sigma}_{Y|H}^2(\mathbf{x};t_1)$ of such random field can be estimated statistically over the set of realizations in $\mathbf{Y}_t^u$:

$$\hat{\mu}_{Y|H}(\mathbf{x};t_1) = \frac{1}{n_r} \sum_{r=1}^{n_r} \mathbf{Y}_t^u\left(y(\mathbf{x};t_1),r\right)$$

$$\hat{\sigma}_{Y|H}^2(\mathbf{x};t_1) = \frac{1}{n_r} \sum_{r=1}^{n_r} \left(\mathbf{Y}_t^u\left(y(\mathbf{x};t_1),r\right) - m(\mathbf{x};t_1)\right)^2 \tag{5.10}$$

where $m(\mathbf{x};t_1)$ is the arithmetic mean at $\mathbf{x}$ and time t_1; n_r is the number of realizations in the ensemble.

The process described above is repeated at the next time for which observations are available but the new prior ensemble of log-conductivity realizations is the posterior one at time $t=1$ i.e. $Y_1(\mathbf{x})$.

5.4 Results and discussion

The results in this section shown that the mean of the velocity fields obtained from ensembles of log-conductivity realizations conditional to histories of hydraulic heads provide useful information on the existence and location of preferential seepage paths in the core, despite of the

incomplete prior knowledge of the statistical parameters of the reference log-conductivity field. The confidence and the error of the estimations are also analyzed in this section.

5.4.1 Prediction of velocity fields

In the experiments of this chapter, at the end of each update step, an ensemble consisting of 1000 thousand realizations of log-conductivities conditional to piezometric observations is obtained for each analyzed case. The set of realizations after the conditioning process is completed (at $t=70$ days) is used within a Monte Carlo framework to predict seepage velocities and to quantify the uncertainty of the prediction. In the method of Monte Carlo, the forward seepage problem is solved numerically by finite elements for the seepage and boundary conditions of equations 5.1 to 5.5 and using the conductivity values of each conditional realization. The mean and variance of the velocity field are determined statistically over the ensemble of realizations.

In the following discussion, only the mean velocity field at the end of the drawdown (at $t=70$ days) is used to detect, by visual inspection, preferential seepage paths in the core and to compare the locations of such paths against those in the velocity field associated to the reference field of conductivities. Both the EnKF and the pf-EnKF methods are employed for this purpose.

Fig. 5.6(a) shows the velocity field associated to the reference conductivity field at the end of the drawdown (at $t=70$ days). The patterns of preferential seepage in such figure clearly correspond to the continuous zones with higher log-conductivities in the reference field of Fig. 5.2(a). The analysis of a velocity field is therefore useful to detect preferential seepage paths. The reference velocity field is illustrated for the purpose of comparing the mean of the velocity fields obtained from the conditional realizations of conductivities at that particular time.
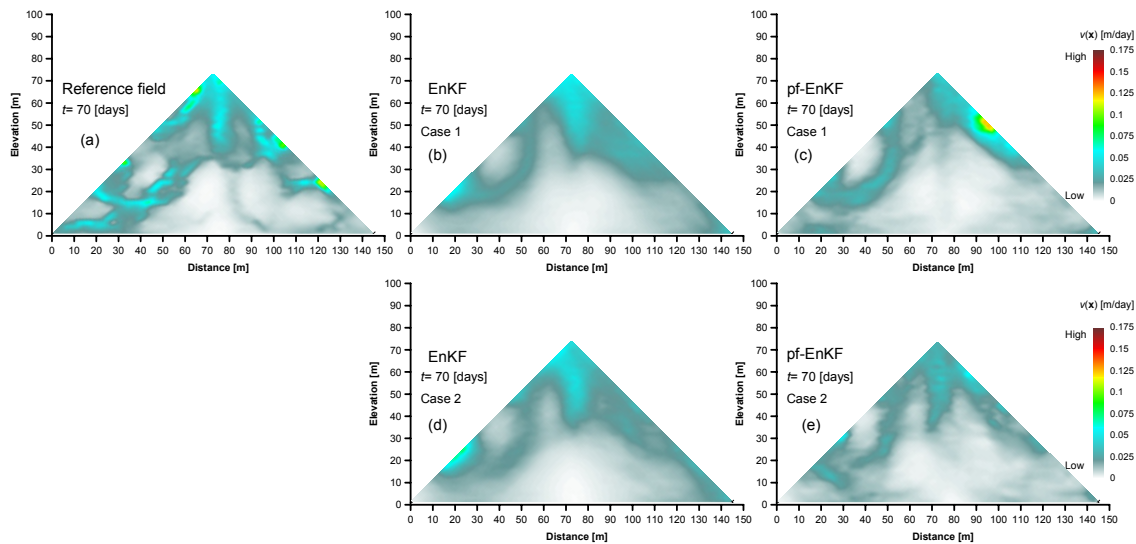
Fig. 5.6 Mean velocity fields [m/day] obtained from the conductivity realizations conditional to nine histories of hydraulic heads. The velocity field of the reference field is also shown for comparison purposes.

Figures 5.6(b),(c) shown the mean of the velocity fields corresponding to the conditional ensembles of the log-conductivity of case 1 that are obtained with the EnKF and pf-EnKF, respectively, after integrating the nine histories of hydraulic heads sequentially until time $t=70$ days. It can be noticed on such figures that the velocity fields shown preferential seepage pathways in the core (clearer zones). This observation holds also for the velocity fields of case 2

(Fig. 5.6(d),(e)). As can be observed in all aforementioned figures, the results of both filters are very similar in both analyzed cases.

5.4.2 Confidence intervals

The confidence of the predicted velocity field after the conditioning process is completed (at *t*=70 days) is also considered in the analysis. Confidence intervals with the mean estimation plus and minus one standard deviation are determined. The results in two control sections are discussed next. One of such sections is along the upstream surface and other is at the middle of the core.

The confidence interval of the velocities along the upstream surface that is obtained for case 1 with the conditional realizations of the EnKF shows that overall the reference velocities are included in the interval, except at some elevations with extreme velocities that corresponds to seepage pathways that leaves the upstream surface of the core (Fig. 5.7(a)). At the middle of the core the reference velocities are well bracket in the confidence interval (Fig. 5.7(b)). The confidence intervals calculated with the conditional realizations of the pf-EnKF are narrowed than those of the EnKF and failed to include reference velocities in several elevations of the control sections, as can be seen in Figs. 5.7(c),(d).
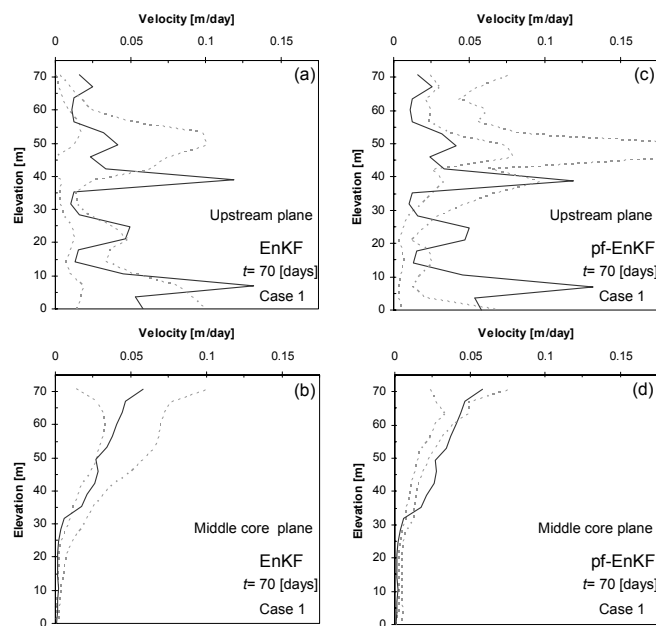
Fig. 5.7 Case 1. Confidence intervals for the forecasted velocities at time *t*=70 days (dashed lines). Reference velocities are indicated by solid lines. Confidence intervals of the EnKF (left). Confidence intervals of the pf-EnKF (right). a) and c) Profiles along the upstream plane. b) and d) Profiles along the middle core plane.

For case 2, confidence intervals obtained with the conditional realizations of the EnKF shown more irregular profiles that deviate more of the reference velocities than in case 1. This behavior may be explained by the poorer initial approximation assumed in case 2. However, the estimated confidence interval is still representative of the reference velocities (Fig. 5.8(a),(b)). The confidence intervals of the pf-EnKF are narrowed and failed to include the reference velocities at several elevations. Therefore, despite the velocity fields of Fig. 5.6 looks very similar, the velocity fields of the EnKF are closer to the reference velocity field than that of the

pf-EnKF. This conclusion will be reviewed next by means of an error analysis of the estimation of the reference velocity field.
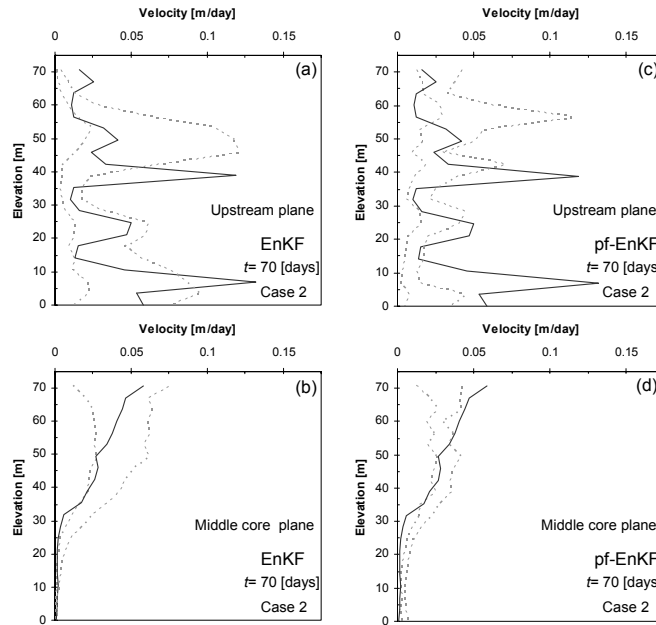
Fig. 5.8 Case 2. Confidence intervals for the forecasted velocities at time $t=70$ days (dashed lines). Reference velocities are indicated by solid lines. Confidence intervals of the EnKF (left). Confidence intervals of the pf-EnKF (right). a) and c) Profiles along the upstream plane. b) and d) Profiles along the middle core plane.

5.4.3 Cross-validation

In this section, the cross-validation between of the estimated and reference velocity fields is conducted. For the sake of completeness, cross-validations of the estimated log-conductivity and hydraulic head fields are also considered. Besides two well-known measures of *goodness-of-fit*, namely the mean absolute error (MAE) and the square root of the mean square error (RMSE), the so-called linear error in probability space (LEPS) is evaluated (Bárdossy and Li, 2008). It is defined by:

$$\text{LEPS} = 1/n \sum_{i=1}^{n} |F(y^*(\mathbf{x}_i)) - F(y^{ref}(\mathbf{x}_i))| \qquad (5.10)$$

where $|\cdot|$ means absolute value; $F(y^{ref}(\mathbf{x}_i))$ is the value of the cumulative distribution function (CDF) corresponding to location $\mathbf{x}_i$ of the reference field; $F(y^*(\mathbf{x}_i))$ is the value of the CDF also at location $\mathbf{x}_i$ of the estimated field. The fields $y^*(\mathbf{x}_i)$ and $y^{ref}(\mathbf{x}_i)$ represent, interchangeably, the estimated and reference fields respectively of hydraulic log-conductivity, hydraulic head and velocity.

The results of cross-validation of each analyzed case are shown in Table 5.2. The cross-validation is performed after completing the conditioning process (at $t=70$ days). The LEPS is easier to compare than the MAE or RMSE because the former is not affected by the units of the variables. It varies from zero to one representing a perfect and an imperfect estimation respectively.

Table 5.2 Results of crossvalidation of the estimated mean log-conductivity $Y(\mathbf{x})$ field and predicted mean hydraulic head $H(\mathbf{x})$ and velocity $V(\mathbf{x})$ fields after complete the conditioning process (t=70 days). The mean hydraulic head and velocity fields were obtained statistically from the results of the forward solution using the conditional fields of hydraulic conductivities. Results of each analyzed case are indicated.

Case	Field	EnKF			pf-EnKF		
		RMSE	MAE	LEPS	RMSE	MAE	LEPS
	$y(\mathbf{x})$	0.887	0.674	0.211	0.933	0.715	0.207
1	$h(\mathbf{x})$	1.519	1.242	0.016	1.377	1.151	0.017
	$v(\mathbf{x})$	0.015	0.008	0.142	0.018	0.010	0.155
	$y(\mathbf{x})$	0.897	0.688	0.215	1.066	0.827	0.247
2	$h(\mathbf{x})$	1.699	1.406	0.018	2.035	1.477	0.025
	$v(\mathbf{x})$	0.015	0.008	0.138	0.016	0.009	0.155

In terms of the LEPS, the best estimation corresponds to the hydraulic head field in both analyzed cases. The estimated log-conductivity field presents the highest errors in both cases. These results illustrate the well-known "*smoothing effect*" of the flow equations; namely similar hydraulic head behaviors can be produced by different hydraulic conductivity fields (Delhomme, 1979; Chilès and Delfiner, 1999). The LEPS suggests that one should expect similar errors in the estimation of both the log-conductivity and velocity fields, although some statistical compensation effect may be present at estimating velocities. The LEPS also indicates that overall the EnKF performs better than the pf-EnKF, except at estimating the log-conductivity of the reference field of case 1. Results from the other measures shown very similar trends.

5.5 Conclusions

In this chapter, a stochastic inverse modeling approach to address the issue of preferential seepage paths detection in earth dams was presented. A simulated field with apparent continuous zones of high hydraulic conductivities was assumed to represent the actual distribution of conductivities in the core of a hypothetic earth dam. Such reference field was used to generate histories of piezometric heads by solving a transient flow condition associated to a "rapid" drawdown of the reservoir. Then solutions to the stochastic inverse problem were obtained via the modified EnKF and the proposed pf-EnKF methods. Both methods provided realizations of the log-conductivity field conditional to piezometric heads at specified times.

The results of both filters showed that even an incomplete prior knowledge of the actual statistical parameters of the reference field as well as in the absence of direct conductivity measures, the identification of the main features of the spatial distribution of conductivities in the core is possible by integrating a limited set of hydraulic head observations in the prior multi-Gaussian model. The mean velocity fields associated to the characterizations were able to provide useful information about the existence and location of preferential seepage paths in the core.

The spatial variability of the hydraulic conductivity in actual earthen structures is in fact three-dimensional and more complicated seepage patterns than a merely "rapid" drawdown of the reservoir take place during their life time period. The presented methodology can readily be extended to address such conditions.

GENERAL CONCLUSIONS
AND PERSPECTIVES

Summary

This thesis focused on the development of numerical tools to characterize heterogeneous soil masses with respect to their hydraulic conductivities from a probabilistic perspective of analysis. For this purpose, two algorithms were developed; one based on the so-called *V*-transformed copula and other based on the Ensemble Kalman Filter (EnKF) method. The algorithms were validated through numerical experiments.

The discussion was first oriented to the study of the concept of dependence defined in the context of random fields. This examination paid on evidence the necessity of considering random fields with non multi-Gaussian dependence to represent uncertainty due to spatial variability of the hydraulic conductivity of soil masses. The *V*-transformed copula was adopted for this purpose.

An inversion technique based on the Ensemble Kalman Filter was proposed. Both the original (EnKF) and the proposed (pf-EnKF) techniques were used to simulate conductivity random fields conditional to histories of hydraulic heads. In one numerical experiment, the conditional conductivity fields were used to predict seepage velocities in a transversal cross-section of the internal core of an idealized earth dam to identify preferential seepage paths.

Findings and discussions

The numerical experiments conducted in this research have shown that:

- The *V*-transformed copula model is indeed a very flexible tool to represent structures of heterogeneity of soil hydraulic conductivity, in the sense that the maps generated with this model exhibit realistic characteristics of conductivity fields such as asymmetry and connectivity. It is recommended therefore for estimation and simulation purposes in geostatistical applications.

- In the experiments aimed to characterize heterogeneous hydraulic conductivities, the original filter (EnKF) outperformed the proposed one (pf-EnKF) in terms of local accuracy, but the last captured better the bivariate empirical copulas of the reference field. It would be desirable to develop methods that are not only more accurate globally,

but also more accurate locally than their predecessors. However, this challenge is left for further research. The proposed method provided more accurate estimates of the histogram and correlation functions of the reference field than the prior estimates. Both the original and proposed filtering techniques were found to yield useful results even in conductivity fields with non multi-Gaussian dependence.

- Histories of hydraulic heads are definitely worth to be taken into account in characterization process of the hydraulic conductivity of soil masses. All experiments showed that estimates of the hydraulic conductivity are improved at integrating hydraulic head responses in the characterization task.

- Since the developed tools are based on the Monte Carlo simulation method, their execution may be time consuming in large dimensional problems. The use of high performance computing should be considered for their implementation.

Recommendations for future research

- The quality of the realizations generated with the models based on Ensemble Kalman Filtering could be improved by incorporating non lineal estimation techniques in the update step such as the indicator Kriging technique.

- The models based on Ensemble Kalman Filtering could be extended to analyze more elaborated seepage patterns that occur in earthen dams like the first filling of the reservoir.

- It is recommended to investigate in more detail the effect of the initial approximation on the performance of both inversion techniques based on the Ensemble Kalman Filter.

- The copula and filtering models were used separately in the experiments but they could be combined in practical applications. For example, conductivity fields conditional to available measurements of conductivities themselves could be simulated first with the copula model then they could be made conditional to histories of hydraulic heads through Ensemble Kalman Filtering.

- Tools based on the Ensemble Kalman Filter developed in this research were able to estimate hydraulic heads at different locations and times taken into account histories of the hydraulic head itself. In these cases non deformable porous media were considered. Histories of settlements could be incorporated into a model that account for deformability of the media in an analogous manner to that presented in this research to predict subsidence in soil masses subjected to withdraw of groundwater and to estimate uncertainty of the prediction.

BIBLIOGRAPHY

Ababou, R., Sagar, B. and Wittmeyer, G. (1992). "Testing procedures for spatially distributed models", *Adv. Water Resour.*, 15, pp.181-198.

Abramovitz, M. and Stegun, I. (1964). "Handbook of mathematical functions", Dover Publ., 1046 p.

Ahmed, A. (2009). "Stochastic analysis of free surface flow through earth dams", *Computers and Geotechnics*, 36(7), pp. 1186-1190.

Allard, D. and Heresim Group (1994). On the connectivity of two random set models: the truncated Gaussian and the Boolean, In: Soares A, editor. *Geostatistics Troia'92*, Vol. 1. Norwell, MA: Kluwer, pp. 467–478.

Alcolea, A., Carrera, J. and Medina, A. (2006). "Inversion of heterogeneous parabolic-type equations using the pilot points method", *Int. J. Numer. Methods Fluids*, 51(9–10), pp. 963–80.

Araujo, J. M. and Awruch, A. M. (1994). "On stochastic finite elements for structural analysis", *Computers and Structures*, 52(3), pp. 461-469.

Auvinet, G. (2002). "Uncertainty in geotechnical engineering", *Proc. 16th Nabor Carrillo Lecture*, SMMS, Queretaro, México, pp. 58.

Bárdossy, A. (2006). "Copula-based geostatistical models for groundwater quality parameters", *Water Resour. Res.*, 42, W11416 (1 of 12).

Bárdossy, A. and Li, J. (2008). "Geostatistical interpolation using copulas", *Water Resour. Res.*, 44, W07412 (1 of 15).

Bárdossy, A. and Pegram, G. (2009). "Copula base multisite model for daily precipitation simulaiton", *Hydrol. Earth Syst. Sci. Discuss.*, 6, pp. 4485–4534.

Bear, J. (1972). "Dynamic of fluids in porous media", New York, Elsevier.

Bear, J., Zaslavsky, D. and Irmay, S. (1968). "Physical principles of water percolation and seepage", UNESCO.

Bertino, L., Evensen, G. and Wackernagel, H. (2003). "Sequential Data Assimilation Techniques in Oceanography ", *International Statistical Review*, 71(2), pp. 223–241.

Bradbury, K. R. and Muldoon, M. A. (1990). "Hydraulic conductivity determinations in unlithified glacial and fluvial materials", In: D.M. Nielsen and I. Johnson (Editors), Ground Water and Vadose Zone Monitoring, ASTM STP 1053. American Society for Testing and Materials, Philadelphia, pp. 138-151.

Bruderer-Weng, C., Cowie, P., Bernabe, Y. and Main, I. (2004). "Relating flow channeling to tracer dispersion in heterogeneous networks", *Adv. Water Resour.*, 27, pp. 843–855.

Butala, M. D., Frazin R. A., Chen,Y. and Kamalabadi, F. (2009)."Tomographic Imaging of dynamic objects with the ensemble Kalman filter", IEEE *Transactions on Image Processing*, 18, pp. 1573–1587.

Capilla, J. and Llopis-Albert, C. (2009). "Gradual conditioning of non-Gaussian transmissivity fields to flow and mass transport data: 1, Theory", *J. Hydrol.,* 371, pp. 66–74.

Casella, G. and George, E. (1992). "Explaining the Gibbs sampling", *American Statistical Association*, 46(3), pp. 167-174.

Chilès, C. and Delfiner P. (1999). "Geostatistics: modeling spatial uncertainty", Wiley, New York, 695 pp.

Christakos, G. (1990). "A Bayesian/Maximum-Entropy view to the spatial estimation problem", *Math. Geology*, 22(7), pp. 763-777.

Collins, M. (1971). "Seepage through triangular dam without tailwater", *J. Soil Mech. and Foundations Division*, 97(9), pp. 1355-1358.

Cover, T. and Thomas, J. (1991). "Elements of Information Theory", Wiley series in Telecommunications.

Dagan, G. (1989). "Flow and transport in porous formations", Springer-Verlag, Berlin, Germany.

Davis, S. N. (1969). "Porosity and Permeability of Natural Materials", In: R.J.M. de Wiest (Ed.), Flow through Porous Media. Academic, New York, pp. 54-89.

De Marsily, G. (1986). "Quantitative hydrogeology", San Diego, CA., Academic Press.

Delhomme, J. P. (1979). "Spatial variability and uncertainty in groundwater flow parameters: a geostatistical approach", *Water Resour. Res.,* 15(2), pp. 269–280.

Deutsch, C. and Cockerham, P. (1994). "Practical considerations in the application of simulated annealing to stochastic simulation", *Math. Geology*, 26(1), pp. 67-82.

Deutsch, C. V. and Journel A. G. (1992). "GSLIB, Geostatistical Library and User's Guide", Oxford University Press, Oxford, N. Y., USA.

Deutsch, C. and Journel, A. (1994). "The application of simulated annealing to stochastic reservoir modeling", *SPE Advanced Technology Series*, 2(2).

Devroye, L. (1986). "Non-Uniform random variate generation", *Springer-Verlag*, 843 pp.

Emery, X. (2002). "Conditional simulation of non Gaussian random functions", *Math. Geology*, 34(1), pp. 79–100.

Emery, X. (2004). "Properties and limitations of sequential indicator simulation", *Stoch. Environ. Res. Risk Assess.*, 18(6), pp. 414–424.

Emery, X. (2005). "Conditional simulation of random fields with bivariate Gamma isofactorial distributions", *Math. Geology*, 37(4), pp. 419-445.

Emery, X. (2006). "A disjunctive Kriging program for assessing point-support conditional distributions", *Comp. Geosci.*, 32, pp. 965-983.

Emery, X. (2008). "Substitution random fields with Gaussian and Gamma distributions: Theory and application to a pollution data set", *Math. Geosci.*, 40, pp. 83-99.

Erdely, A. (2007). "Diagonal properties of the empirical copula and applications. Constructions of copulas with given restrictions", *Ph.D. dissertation*, Posgrado en Ciencias Matemáticas, Facultad de Ciencias, UNAM, México, 105 p.

Evensen, G. (1994). "Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics", Journal of Geophysical Research, 99, pp. 10143–10162.

Evensen, G. (2006). "Data assimilation: the ensemble Kalman Filter", New York, Springer.

Fell, R. and Wan, C. F. (2005). "Methods for estimating the probability of failure of embankment dams by internal erosion and piping in the foundation and from embankment to foundation", UNICIV report No. R-436, January, The University of New South Wales, Sydney, Australia.

Fenton, G. A. and Griffiths, D. V. (1997). "Extreme hydraulic gradient statistics in a stochastic earth dam", *J. Geot. Geoenv. Eng.*, ASCE, 123(11), pp. 995-1000.

Fogg, G. E. (1986). "Groundwater flow and sand body interconnectedness in a thick, multiple-aquifer system", *Water Resour. Res.*, 22(5), pp. 679-694.

Freeze, R. A. (1975). "A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media", *Water Resour. Res.,* 11(5), pp 725–741.

Freeze, A. and Cherry, J. (1979). "Groundwater", New Jersey, Prentice Hall.

Frind, E., Muhammad, D. and Molson, J. (2002). "Delineation of three dimensional well capturing zones for complex multi-aquifer systems", *Ground Water*, 40(6), pp. 585-598.

Froidevaux, R. (1990). "Geostatistical toolbox primer, version 1.30", FSS International, Troinex, Switzerland.

Gainis, B., Klie, H., Wheeler, M. F., Wildey, T., Yotov, I. and Zhang, D. (2008). "Stochastic Collocation and Mixed finite elements for Flow in Porous Media", *Comput. Methods Appl. Mech. Engrg.,* 197, pp. 3547–3559.

Gelhar, L. W. (1993). "Stochastic subsurface hydrology", Pentrice Hall, New Jersey.

Geman, S. and Geman, D. (1984). "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI, 6(6), pp. 721-741.

Genest, C. and Favre, A. (2007). "Everything you always wanted to know about copula modeling but were afraid to ask", *J. Hydrol. Eng.,* July/August, pp. 347-368.

Ghanem, R. and Ferro, G. (2006). "Health monitoring for strongly non-linear systems using the Ensemble Kalman Filter", *Structural control and health monitoring*, 13, pp. 245–259.

Ghanem, R. and Spanos, P. (1991). "Stochastic finite elements: A spectral approach", Springer, Berlin.

Gómez-Hernández, J. Sahuquillo, A. and Capilla, J. (1997). "Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data, 1, Theory", *J. Hydrol.,* 203(1–4), pp. 162–74.

Gómez-Hernández, J. and Wen, X. (1998). "To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology", *Adv. Water Resour.,* 21(1), pp. 47-61.

Griffiths, D. V. and Fenton, G. A. (1993). "Seepage beneath water retaining structures founded on spatially random soil", *Geotechnique*, 43(4), pp.577-587.

Griffiths, D. V. and Fenton, G. A. (1997). "Three-dimensional seepage through a spatially random soil", *J. Geot. Geoenv. Eng.*, ASCE, 123(2), pp. 153-160.

Grigoriu, M. (1998). ''Simulation of stationary non-Gaussian translation processes'', *J. Eng. Mech.*, 124(2), pp. 121–126.

Guardiano, F. and Srivastava, R. (1993). "Multivariate geostatistics: beyond bivariate moments", *In: Soares, A., Ed., Geostatistics Troia*, Vol. 1, Kluwer Academic, Dordrecht, pp. 133–144.

Gui, S., Zhang, R., Turner, J. P. and Xue, X. (2000). "Probabilistic slope stability analysis with stochastic soil hydraulic conductivity", *J. Geot. Geoenv. Eng.*, ASCE, 126(1), pp. 1-9.

Harr, M. E. (1962). "Groundwater and seepage", New York, McGraw Hill.

Haslauer, C., Bárdossy, A. and Sudicky, E. (2008). "Geostatistical analysis of hydraulic conductivity fields using copulas", *VIII international Geostatistics Congress*, GEOSTASTS, Santiago, Chile.

Haslauer, C. P., Guthke, P., Bárdossy, A. and Sudicky, E. A. (2009). "Effects of Non-Gaussian Spatial Dependence of Hydraulic Conductivity on Hydrodynamic Macrodispersion", *American Geophysical Union*, Fall Meeting, Abstract #H43F-1093.

Hendricks-Franssen, H. J., Gómez-Hernández, J. J., Capilla, J. E. and Sahuquillo, A. (1999). "Joint simulation of transmissivity and storativity fields conditional to steady-state and transient hydraulic head data", *Advances in Water Resources*, 23(1), pp. 1–13.

Hendricks Franssen, H. J. and Gomez-Hernandez, J. (2002). "3D inverse modeling of groundwater flow at a fractured site using a stochastic continuum model with multiple statistical populations", *Stochastic Environ. Res. Risk Assess.*, 16, pp. 155–174.

Hilfer, R. (1992). "Local-porosity theory for flow in porous media", *Physical Review B*, 45, pp. 7115–7121.

Hilfer, R., Rage, T. and Virgin, B. (1997). "Local percolation probabilities for a natural sandstone", *Physica A*, 241, pp. 105–110.

Hoeffding, W. (1940). "Masstabinvariante Korrelationstheorie", *Schriften des Matematischen Instituts und des Instituts für Angewandte Mathematik der Universität*, Berlin, 5, pp. 1979-223

Hoeffding, W. (1948). "A class of statistics with asymptotically normal distribution", *Ann. Math. Stat.*, 19(3), 293–325.

Hoeksema, R. and Kitanidis, P. (1985). "Analysis of the spatial structure of properties of selected aquifers", *Water Resour. Res.*, 21(4), pp. 563–572.

ICOLD (1995). "Dams failures statistical analysis", *Inter. Commission on Large Dams*, Bull. 99.

ICOLD (1997). "Internal erosion: Typology, detection and repair", *Bulletin du Comité Français de Grandes Barrages*, 6, Italy.

ICOLD (2004). "Workshop on dam safety problems and solutions-sharing experience", *In: 72nd Annual Meeting, National Committee on Large Dams,* W1-08-A077, Seoul, Korea.

Istok, J. (1989). "Groundwater modeling by the finite element method", Washington, DC., American Geophysical Union.

Jaynes, E. T. (1957). "Information theory and statistical mechanics", *The Physical Review*, 106(4), pp. 620-630.

Janssen, G., Valstar, J. R., and Van der Zee, S. (2006). "Inverse modeling of multimodal conductivity distributions", *Water Resour. Res.*, 42, W03410, doi:10.1029/2005WR004356.

Joe, H. (1997). "Multivariate models and dependence concepts", *Chapman Hall*, London.

Johnson, N. L., and Kotz, S. (1972). "Distributions in Statistics: Continuous Multivariate Distributions", Wiley, New York, 333 p.

Journel, A. (1983). "Non parametric estimation of spatial distributions", *Math. Geology*, 15, pp. 445-468.

Journel, A. and Alabert, F. (1989). "Non-Gaussian data expansion in the earth sciences", *Terra Nova*, 1, pp. 123-134.

Journel, A. and Deutsch, C. (1993). "Entropy and spatial disorder", *Math. Geology.*, 25(3), pp. 329-355.

Journel, A. G. and Huijbregts, C. (1978). "Mining geostatistics", *Academic Press*, San Diego, Calif., 600 pp.

Journel, A. and Posa, D. (1990). "Characteristic behavior and order relations for indicator variograms", *Math. Geology*, 22(8), pp. 1011-1025.

Journel, A. and Zhang, T. (2006). "The necessity of a multiple point statistical model", *Math. Geology*, 38(5), pp. 591-610.

Kapur, J. N. (1989). "Maximum entropy models in science and engineering", Wiley Eastern Limited, India.

Knudby, C. and Carrera, J. (2005). "On the relationship between indicators of geostatistical flow and transport connectivity", *Adv. Water Resour.*, 28, pp. 405–421.

Kruskal, W. H. (1958). "Ordinal measures of association", *J. Am. Stat. Assoc.*, 53(4), pp. 814–861.

LaBolle, E. M. and Fogg, G. E. (2001). "Role of molecular diffusion in contaminant migration and recovery in an alluvial aquifer system", *Transport Porous Media*, 42(1–2), pp. 155–179.

Law, J. (1944). "A statistical approach to the interstitial heterogeneity of sand reservoirs", *Trans. AIME*, 155, pp. 202-222.

Lebrun, R. and Anne Dutfoy, A. (2009). "Do Rosenblatt and Nataf isoprobabilistic transformations really differ?", *Probabilistic Eng. Mech.,* 24, pp. 577-584.

Lemhann, L. E. (1966). "Some concepts of dependence", *Ann. Math. Statist.*, 37, pp. 1137-1153.

Li, J. and Bárdossy, A. (2009). "Stochastic simulation of hydraulic conductivity fields using non Gaussian dependence", *American Geophysical Union*, Fall Meeting, Abstract #H54B-06.

López-Acosta, N. and Auvinet, G. (2003). "Comparison between first and second order approximations in stochastic analysis of groundwater seepage in soils", *Procc. of the 12th Panamerican Conference on Soil Mechanics and Geotechnical Engineering*, Cambridge Mass. USA, Vol. 2, pp. 1291-1296.

López-Acosta, N. and Auvinet, G. (2004). "Flujo de agua en medios heterogéneos", *Memorias de la XII Reunión Nacional de Mecánica de Suelos*, SMMS, Guadalajara México, Vol. 1, pp. 215-222.

López-Acosta, N. (2010). "Uncertainty in the analysis of seepage in soils", *PhD Dissertation*, DEPFI, UNAM (In Spanish).

López-Acosta, N. and Auvinet, G. (2011). "Uncertainty in analyses of one-dimensional steadystate seepage through random porous media", *Probabilistic Eng. Mech.*, 26(3), pp. 501-510.

Lu, Z. and Zhang, D. (2004). "A comparative study on uncertainty quantification for flow in randomly heterogeneous media using Monte Carlo simulations and conventional and KL-based moment-equation approaches", *Siam J. Sci. Comput.*, 26(2), pp. 558–577.

Manner, H. (2010). "Testing for asymmetric dependence", *Studies on Nonlinear Dynamics and Econometrics*, 14(2), pp. 1-30.

Matheron, G. (1965). "Les variables regionalices et leur estimation", Masson et Cie, Paris, France.

Matheron, G. (1967). "Eléments pour une théorie des milieux poreux", *Masson et Cie*, Paris.

Mecke, K. and Wagner, H. (1991). "Euler characteristic and related measures for random geometric sets", *Journal of Statistical Physics*, 64(3-4), pp. 843-850.

Mood, A. and Graybill, F. (1963). "Introduction to the theory of statistics", McGraw-Hill.

Nelsen, R. B. (2006). "An introduction to copulas", Springer-Verlag, New York, 2nd Edition, pp. 269.

Nowak, W., Schwede, R., Cirpka, O. and Neuweiler, I. (2008). "Probability density functions of hydraulic head and velocity in three dimensional heterogeneous porous media", *Water Resour. Res.*, 44, W08452, pp. 1-15.

Pineda-Contreras, A. R. and Auvinet, G. A. (2013). "Stochastic finite element in geotechnical engineering. Spectral approach", *Ingeniería Investigación y Tecnología,* XIV(1), pp. 11-22. (In spanish).

Poeter, E. and Townsend, P. (1994). "Assessment of critical flow path for improved remediation management", *Ground Water*, 32(3), pp. 439–447.

Poluvarinova-Kochina, P. (1962). "Theory of groundwater movement", New Jersey, Princeton University Press.

Popescu R., Deodatis, G. and Prevost, J. H. (1998). "Simulation of non-Gaussian homogeneous stochastic vector fields", *Probabilistic Eng. Mech.*, 13(1), pp. 1–13.

Quesada-Molina, J. J. (1992). "A generalization of an identity of Hoeffding and some applications", *J. Ital. Stat. Soc.*, 3, pp. 405–411.

RamaRao, B., LaVenue, A., De Marsily, G. and Marietta, M. (1995). "Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields, 1, theory and computational experiments", *Water Resour. Res.*, 31(3), pp. 475–93.

Rosenblatt, M. (1952). "Remarks on a multivariate transformation", *The Annals of Mathematical Statistics*, 23(3), pp. 470-472.

Rubin, Y. (2003). "Applied Stochastic Hydrogeology", Oxford Univ. Press, New York, 391 pp.

Salvadori, G., Michele, C., Kottegoda, N. and Rosso, R. (2007). "Extremes in nature. An approach using copulas", Springer, New York.

Sánchez-Vila, X. (1997). "Radially convergent flow in heterogeneous porous media", *Water Resour. Res.*, 33(7), pp. 1633–1641.

Sanchez-Vila, X., Carrera, J. and Girardi, J. (1996). "Scale effects in transmissivity", *Journal of Hydrology*, 183(1), pp. 1–22.

Sánchez-Vila, X. Guadagnini, A. and Carrera, J. (2006). "Representative hydraulic conductivities in saturated groundwater flow", *Rev. Geophys.,* 44(RG3002), pp. 1-46.

Sakamoto, S. and Ghanem, R. (2002a). "Polynomial chaos decomposition for the simulation of non-Gaussian non-stationary stochastic processes", *J. Eng. Mech.*, 128(2), pp. 190–201.

Sakamoto, S. and Ghanem, R. (2002b). "Simulation of multi-dimensional non Gaussian non Stationary random fields", *Probabilistic Eng. Mech.*, 17, pp. 167–176.

Shannon, C. (1948). "A mathematical theory of communication", *Bell System Tech. J.*, 27, pp. 379-623.

Silliman, S. E. and Wright, A. L. (1988). "Stochastic analysis of paths of high hydraulic conductivity in porous media", *Water Resour. Res.*, 24(11), pp. 1901-1910.

Sklar, A. (1959). "Fonctions de répartition à n dimensions et leurs marges", *Publ. Inst. Statist. Univ. Paris*, 8, pp. 229-231.

Schöniger, A., Nowak, W. and Hendricks Franssen, H. (2012). "Parameter estimation by ensemble Kalman filters with transformed data: Approach and application to hydraulic tomography", *Water Resour. Res.,* 48(W04502), pp. 1-18.

Smith, I. and Griffiths, D. (2004). "Programming the finite element method", England, John Wiley and Sons Ltd.

Smith, M. and Konrad, J. M. (2011). "Assessing hydraulic conductivities of a compacted dam core using geostatistical analysis of construction control data", *Canadian Geotechnical Journal*, 48, pp. 1314–1327.

Srivastava, R. (1992). "Reservoir characterization with probability field simulation", In *SPE annual conference and exhibition*, Washington, D.C., 24753, pp. 927-938.

Srivastava, R. (1995). "An overview on stochastic methods for reservoir characterization", In J. Yarus and R. Chambers, eds. *Stochastic modeling and geostatistics: Principles, methods and Case Studies*, pp. 3-16, AAPG Computer Applications in Geology, No. 3.

Strebelle, S. (2002). "Conditional simulation of complex geological structures using multiple-point statistics", *Math. Geology*, 34(1), pp. 1-21.

Sudicky, E. (1986). "A natural gradient experiment on solute transport in a sand aquifer: Spatial variability of hydraulic conductivity and its role in the dispersion process", *Water Resour. Res.*, 22(13), pp. 2069-2082.

Teles, V., Delay, F. and de Marsily, G. (2004). "Comparison of genesis and geostatistical methods for characterizing the heterogeneity of alluvial media: Groundwater flow and transport simulations", *Journal of Hydrology*, 294, pp. 103–121.

Trinchero, P. Sánchez-Vila, X. and Fernández-García, D. (2008). "Point-to-point connectivity, an abstract concept or a key issue for risk assessment studies?", *Adv. Water Resour.*, 31, pp. 1742-1753.

Vassena, C. Cattaneo, L. and Giudici, M. (2009). "Assessment of the role of facies heterogeneity at the fine scale by numerical transport experiments and connectivity indicators", *Hydrogeology Journal*, 18(3), pp. 651-668.

Vázquez, F. and Auvinet, G. (2014). Simulation of random fields with non multi-Gaussian dependence using copulas (In Spanish), *Ingeniería, Investigación y Tecnología*.

Vázquez, F., Auvinet, G. and Vermeer, P. (2014). Stochastic inverse modeling applied to earth dams for detecting preferential seepage paths, NUMGE, Delft, The Netherlands.

Vázquez, F. and Auvinet, G. (2015). Simulation of spatio-temporal random fields using a modified Kalman filter (In Spanish), *Ingeniería, Investigación y Tecnología*.

Vögel, H. (2002). "Topological characterization of porous media", In Morphology and condensed matter. Physics and geometry of spatially complex systems. Lecture notes in physics, pp. 75–92.

Wackernagel, H. (1995). "Multivariate geostatistics", Berlin, Springer-Verlag.

Webb, E. K. and Anderson, M. P. (1996). "Simulation of preferential flow in three-dimensional heterogeneous conductivity fields with realistic internal architecture", *Water Resour. Res.,* 32(3), pp. 533–545.

Woodbury, G. E. and Zudicky, E. A. (1986). "The geostatistical characteristics of the Borden aquifer", *Water Resour. Res.*, 22(5), pp. 679-694.

Xu, T., Gómez-Hernández, J., Zhou, H. and Li, L. (2013). "The power of transient piezometric head data in inverse modeling: An application of the localized normal-score EnKF with covariance inflation in a heterogeneous bimodal hydraulic conductivity field", *Advances in Water Resources*, 54, pp. 100–118.

Yamazaki, F., and Shinozuka, M. (1988). ''Digital generation of non-Gaussian stochastic fields'', *J. Eng. Mech.*, 114(7), pp. 1183–1197.

Yamazaki, F., Shinozuka, M. and Dasgupta, G. (1988). ''Neumann expansion for stochastic finite element analysis'', *J. Eng. Mech.*, 114(8), pp. 1335–1354.

Zappa, G., Bersezio R., Felletti, F. *et al.* (2006). "Modeling heterogeneity of gravel-sand, braided stream, alluvial aquifers at the facies scale", *Journal of Hydrology*, 325(1–4), pp. 134–153.

Zhang, D. (2002). "Stochastic methods for flow in porous media", Academic Press, San Diego, C.A., USA.

Zhou, H., Gómez-Hernández, J., Hendricks Franssen, H. and Li, L. (2011). "An approach to handling nongaussianity of parameters and state variables in ensemble Kalman filtering", *Adv. Water Resour.,* 34(7), pp. 844–864.

Zhu, J. and Yeh, T. (2005). "Characterization of aquifer heterogeneity using transient hydraulic tomography", *Water Resour. Res.*, 41, W07028, doi:10.1029/2004WR003790.

Zinn, B. and Harvey, C. (2003). "When good statistical models of aquifer heterogeneity go bad: A comparison of flow, dispersion, and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields", *Water Resour. Res.*, 39(3), SBH (4-1 to 4-18).

APPENDIX A

BASIC CONCEPTS OF SEEPAGE IN SATURATED RANDOM MEDIA

A.1 Darcy's law

The experimentally derived form of *Darcy's law* was limited to one-dimensional flow. When the flow is three dimensional, one possible generalization of this is (Bear, 1972):

$$\mathbf{q} = K_s \mathbf{J} = -K_s\, grad\ \varphi \qquad\qquad (A.1)$$

where; q is the *specific flux* vector with components $\mathbf{q}_x$, $\mathbf{q}_y$, $\mathbf{q}_z$ in the directions of the Cartesian *x*, *y*, *z* coordinates, respectively, and $\mathbf{J}$ is the *hydraulic gradient* with components in the *x*, *y*, *z* directions, respectively. The flow described by Darcy's law is *irrotational* and *laminar*. Equation A.1 remains valid for *three-dimensional flow* in heterogeneous isotropic media. In such cases: $K_s = K_s(x,y,z)$. The coefficient of proportionality K_s is called the *hydraulic conductivity*. In isotropic media, it is a scalar (dimensions L/T). When expressed solely on properties of the solid matrix the term *permeability* or *intrinsic permeability* (dimension L^2) is used. Under certain conditions *permeability* may also vary with time. Subsidence and consolidation are phenomena associated with changes in *permeability* with time. In the present work, the term permeability or hydraulic conductivity is used evenly to refer to the hydraulic conductivity concept and only 1D and 2D analyses of seepage are considered.

A.2 Hydraulic conductivity tensors

Natural soil formations and embankment structures are, in fact, *anisotropic* with respect to hydraulic conductivity; that is, the magnitude of this property depends on direction. Permeability and hydraulic conductivity in these cases are, in fact, *symmetric second rank tensors*; therefore they need six independent experiments in order to be fully determined. In practice, principal directions of anisotropy are unknown both with respect to the direction of flow and to the direction of the gradient. Therefore, it is necessary to assume three major axes in which case only three independent experiments are needed. In certain porous media the complexity of the problem can be further reduced assuming only two principal directions, say *x*-

y and *z*. In such cases, any direction in the *x-y* plane is a principal direction and only two independent experiments are needed.

In practice, hydraulic conductivity measurements are often scarce and their corresponding directional measurements are virtually never determined. For this reason, *geostatistical* analyses of hydraulic conductivity tensors are practically impossible to carry out, although scalar hydraulic conductivity values can be used to represent anisotropic hydraulic conductivity through *ellipses of anisotropy*. An ellipse of anisotropy can be defined such that it describes higher continuity in the horizontal direction than in the vertical direction. Continuity is defined in geostatistics by the so-called *correlation scale* of the directional *variograms* or *covariance* functions. The numerical experiments presented in this thesis could be extended to incorporate Ellipses of anisotropy in further researches.

A.3 Filtration velocity

Hydraulic conductivity either a scalar or a tensor is a macroscopic characteristic of the porous medium. Thus, $v = q/A$ is a *filtration velocity* and not the mean flow velocity in the pores. At the microscopic scale the flow is governed by the equations of *Navier-Stokes*. These equations are valid at the pore scale and are subjected to intricate boundary conditions. The effective integration of these equations would require a detailed description of the geometry of the pore space. Only macroscopic seepage flow is considered in this work.

A.4 Scale effects

The hydraulic conductivity of porous media depends on the scale of observation. Dagan (1989) defines three different scales: the *macroscopic scale* of soil samples tested in the laboratory, the *megascopic scale* that involves the entire aquifer and several pumping wells and the *scale of blocks* in models of the finite element type (value defined inside an element of a finite element mesh). Different tensors of conductivities are defined at these scales. The reason for this scale effect to appear can be explained by the fact that hydraulic conductivity is not an *additive variable*. For example, two hydraulic conductivity values in series combine by harmonic averaging. Additive variables are water content and porosity since either both of those quantities combine by simple arithmetic means.

At passing from the microscopic scale to the macroscopic scale of a specimen of laboratory, heterogeneity of the pore space is averaged and a single hydraulic conductivity value appears in the Darcy's law. An interesting question that arises is whether a new Darcy's law would emerge at passing from the macroscopic to the megascopic scales such that it makes possible to characterize the global formation with a single scalar "average" value of the hydraulic conductivity. This leads to the concept of *effective conductivity*.

APPENDIX B

EFFECTIVE CONDUCTIVITY AND EQUIVALENT CONDUCTIVITY

B.1 Effective hydraulic conductivity

Effective hydraulic conductivity K_{eff}, is defined as the second order tensor that relates the expected values of flux and hydraulic head through the generalized *Darcy's law*:

$$E\{q\} = -K_{eff} E\{\nabla \varphi\} \tag{B.1}$$

where expectations mean averaging all the possible head and specific discharge fields that could be obtained with the ensemble of hydraulic conductivity fields.

There are only two exact results for K_{eff} available in the literature. The first one is that K_{eff} is bounded by the *harmonic mean* $K_h = \left(E[K^{-1}] \right)^{-1}$ and the *arithmetic mean* $K_a = E[K]$ of the local hydraulic conductivities (Matheron; 1967):

$$\left(E[K^{-1}] \right)^{-1} \le K_{eff} \le E[K] \tag{B.2}$$

The second exact result is also due to Matheron (1967) who demonstrated that in a two-dimensional infinite domain, K_{eff} is equal to the *geometric mean* $K_G = \exp(E[Y])$, provided that Y is a multi-Gaussian random field with isotropic correlation function.

The name effective is used only when the value K_{eff} is constant throughout the domain. In this case, it is considered as a characteristic property of the medium. The presence of boundaries or sink/sources would cause K_{eff} to be variable in space. In such cases, the term *pseudoeffective* conductivity is used, which emphasizes its local nature (e.g. Sánchez-Vila, 1997).

B.2 Equivalent conductivity

Equivalent conductivity is defined as the conductivity of a fictitious homogeneous medium that conveys the same discharge as the actual, heterogeneous one for a given pressure head drop. It is also called the *upscaled* or *block conductivity* when it is used to replace a block of heterogeneous media by an equivalent, homogeneous one in numerical simulations. In this case, ensemble averages are replaced by spatial (in the volumetric sense) averages:

$$\bar{q} = -K_b \overline{\nabla \varphi} \tag{B.3}$$

where:

$$\bar{q} = \frac{1}{V} \int_V q \, dV \qquad\qquad \overline{\nabla \varphi} = \frac{1}{V} \int_V \nabla \varphi \, dV \tag{B.4}$$

When the averaging volume is very large, with a representative size that is several times larger than the volume given by the integral scale of the heterogeneous (random) parameter, the averaging volume comprises all scales of heterogeneity. In such a case, a stationary random function that represents the media heterogeneity is called *ergodic*. *Ergodicity* imply hence that equivalent and effective (or pseudoeffective) parameters are identical.

APPENDIX C

SOME DEFINITIONS AND PROPERTIES OF MULTIVARIATE COPULAS

C.1 Multivariate copulas

An *n-variate copula (n-copula)* is a mapping C from the unit hypercube $[0,1]^n$ onto the unit interval $[0,1]$; that is:

$$C : [0,1]^n \to [0,1] \tag{C.1}$$

with the following properties:

1) The range of $C(u_1,\ldots,u_n)$ is the unit interval $[0,1]$;
2) $C(u_1,\ldots,0,\ldots,u_n) = 0$ if any $u_i = 0$, for $i = 1,2,\ldots,n$;
3) $C(1,\ldots,u_i,\ldots,1) = u_i$, for all $u_i \in [0,1]$;
4) $C(u_1,\ldots,u_n)$ is n-*increasing* in the sense that for every n-dimensional hypercube in the unit cube the corresponding measure assigned by C has to be nonnegative. Bárdossy (2006) writes such condition by:

$$\sum_{i=0}^{2^n-1} (-1)^{n-\sum_{i=1}^n j_i} C(u_1 + j_1\Delta_1,\ldots,u_n + j_n\Delta_n) \geq 0$$

$$\text{If} \qquad 0 \leq u_i \leq u_i + \Delta_i \leq 1 \qquad \text{and} \qquad i = \sum_{k=0}^{n-1} j_k 2^k$$

$$\tag{C.2}$$

From the properties above, it follows that C is non-decreasing in each variable and uniformly continuous. Existence (almost everywhere) of all partial derivatives is also a useful property of copulas (Nelsen 2006).

C.2 Copulas and multivariate distributions

Multivariate distribution functions and one dimensional distribution functions are linked to each other through a copula. Let consider an n-variate distribution $F_Z \in F(F(z_1),\ldots,F(z_n))$ with j^{th} univariate margins $F(z_j)$. By a multidimensional version of Sklar (1959) theorem (e.g. Nelsen, 2006), it is proved that there is an associate copula $C:[0,1]^n \to [0,1]$ that satisfies:

$$
\begin{aligned}
F_Z(z_1,\ldots,z_n) &= C(F(z_1),\ldots,F(z_n)) \\
&= P\{Z_1 \le z_1,\ldots,Z_n \le z_n\} \\
&= C(P\{Z_1 \le z_1\},\ldots,P\{Z_n \le z_n\})
\end{aligned}
$$

$$(C.3)$$

If the distribution is continuous, then the copula is unique. Conversely, if $F_Z(\cdot)$ is a continuous n-variate distribution function with univariate margins $F(z_1),\ldots,F(z_n)$ and quantile functions $F^{-1}(z_1),\ldots,F^{-1}(z_n)$, then:

$$
C(u) = C(u_1,\ldots,u_n) = F_Z(F^{-1}(u_1),\ldots,F^{-1}(u_n))
$$

$$(C.4)$$

C.3 Copula densities

Reminding that the derivative of a probability distribution function $F_Z(z)$, equals the probability density function $f_Z(z)$, by differentiation of eq. C.4 with respect to all variables, the *multivariate copula density* is obtained as follows:

$$
\begin{aligned}
f_Z(z_1,\ldots,z_n) &= \frac{\partial^n C(F(z_1),\ldots,F(z_n))}{\partial F(z_1),\ldots,\partial F(z_n)} \prod_{i=1}^{n} f_i(z_i) \\
&= c(F(z_1),\ldots,F(z_n)) \cdot \prod_{i=1}^{n} f_i(z_i)
\end{aligned}
$$

$$(C.5)$$

where $f_Z(z_1,\ldots,z_n)$ is the joint probability density function. Thus, defining $u_1 = F(z_1),\ldots,u_n = F(z_n)$, the multivariate copula density can be expressed by:

$$
c(u_1,\ldots,u_n) = \frac{\partial^n C(u_1,\ldots,u_n)}{\partial u_1,\ldots,\partial u_n}
$$

$$(C.6)$$

Then, the multivariate copula density is:

$$
c(F(z_1),\ldots,F(z_n)) = \frac{f_Z(z_1,\ldots,z_n)}{\prod_{i=1}^{n} f_i(z_i)}
$$

$$(C.7)$$

C.4 Conditional copula density

The *multivariate conditional copula density* can be expressed by (Bárdossy and Li, 2008; Salvadori *et al.,* 2007):

$$c\Big(u \Big| u_1 = F(z_1),\ldots,u_n = F(z_n)\Big) = \frac{f_{n+1}(z,z_1,\ldots,z_n)}{f_1(z)f_1(z_1)\cdots f_1(z_n)} \frac{1}{c_n(u_1,u_2,\ldots,u_n)}$$

$$\propto \frac{f_{n+1}(z,z_1,\ldots,z_n)}{f_1(z)f_1(z_1)\cdots f_1(z_n)}$$

$$(C.8)$$

which expresses the relationship among the conditional joint density function $f_{n+1}(z,z_1,\ldots,z_n)$, of the n conditioned measurements plus one, divided by the product of the $n+1$ corresponding one-dimensional density functions $f_1(z_i)$; $i = 1,2\ldots,n$. The $n+1$ variable is that of the variable u at the conditioned point. The conditional copula density can then be viewed as the conditional joint distribution of n conditionally uniform random variables. The value u at the conditioned point can be obtained by evaluating the conditional copula density over the entire interval of values of $u \in [0,1]$.

It is noteworthy that eq. C.8 defines a conditional copula density. For interpolation or simulation purposes, any sample at a target point should be drawn from the conditional copula which is equivalent to the conditional distribution function. Such conditional copula can be computed numerically by integration of the conditional copula density. Using a finite difference scheme the final equation for the conditional copula can be written:

$$C\left(u^*|u_1,\ldots,u_n\right) \propto \sum_{u=0}^{u^*} c\Big(u \Big| u_1 = F(z_1),\ldots,u_n = F(z_n)\Big)\Delta u \qquad (C.9)$$

for Δu, small. More explicitly:

$$C\left(u^*|u_1 = F(z_1),\ldots,u_n = F(z_n)\right) \propto \sum_{u=0}^{u^*} \frac{f_{n+1}(z,z_1,\ldots,z_n)}{f_1(z)f_1(z_1)\ldots f_1(z_n)} \cdot \Delta u \qquad (C.10)$$

APPENDIX D

MULTIVARIATE FUNCTIONS FOR THE *V*-TRANSFORMED COPULA

D.1 The multivariate distribution function

Bárdossy and Li (2008), Li and Bárdossy (2009), Bárdossy and Pegram, 2009 write the multivariate distribution function of the random field $V(\mathbf{x})$ in eq. 2.7 of chapter 2 as (see foot note)[1]:

$$H_n(x_1,\ldots,x_n) = P\{X_1 \le x_1,\ldots,X_n \le x_n\}$$

$$= \sum_{i=0}^{2^n-1}(-1)^{\sum_{j=1}^{n}i_j}\,\Phi(\zeta_i + \overline{m}) \tag{D.1}$$

where:

$$\zeta_i^T = \left(b\left((-1)^{i_1}\right)x_1{}^a,\ldots,b\left((-1)^{i_n}\right)x_n{}^a\right) \tag{D.2}$$

$i_j = 0 \;\; or \;\; 1$; and

$$i = \sum_{j=1}^{n}i_j 2^{j-1} \tag{D.3}$$

with:

$$b = \begin{cases} -1 & if \;\; (-1)^{ij} = -1 \\[2ex] \dfrac{1}{v} & if \;\; (-1)^{ij} = 1 \end{cases} \tag{D.4}$$

and

---

[1] The variable *X* is used in this appendix instead of the variable *V*(**x**) in eq. 2.7 of chapter 2 in order to facilitate the reader to follows the original formulation more easily.

$$a = \begin{cases} 1 & if \quad (-1)^{ij} = -1 \\ \dfrac{1}{\alpha} & if \quad (-1)^{ij} = 1 \end{cases} \tag{D.5}$$

All elements in the vector $\overline{m}$ are assumed to be equal[2].

D.1.1 Expansion of the series in eq. D.1 for the bivariate case

In order to evaluate eq. D.1, the coefficients i, in eq. D.3 have to be determined. Then, the vectors ζ_i^T, in eq. D.2 can be constructed with both coefficients b and a in eq. D.4, which in turn depend on whether the transformed values x_i, are positive or negative and depending on the signs in front of each term in ζ_i^T, determined by the superscripts i's in $(-1)^{i_1},\ldots,(-1)^{i_n}$. Now, to obtain the i's in eq. D.1, the i_j's have to be make either 0 or 1 in order to fulfill both sides of the equality. The sequence is explicitly written for the bivariate case as follows:

Para $n = 2$:

$$i = \sum_{j=1}^{2} i_j 2^{j-1}$$

Expanding the expression above, it is obtained:

$$i = i_1 2^0 + i_2 2^1$$

From eq. D.1, i ranges from 0 to 3. So, it can be written:

$$0 = 0_1 2^0 + 0_2 2^1 \rightarrow \quad 0 = 0_1 + 0_2 2 \rightarrow \quad i_j = 0_1 = 0; \ i_j = 0_2 = 0$$
$$1 = 1_1 2^0 + 1_2 2^1 \rightarrow \quad 1 = 1_1 + 1_2 2 \rightarrow \quad i_j = 1_1 = 1; \ i_j = 1_2 = 0$$
$$2 = 2_1 2^0 + 2_2 2^1 \rightarrow \quad 2 = 2_1 + 2_2 2 \rightarrow \quad i_j = 2_1 = 0; \ i_j = 2_2 = 1$$
$$3 = 3_1 2^0 + 3_2 2^1 \rightarrow \quad 3 = 3_1 + 3_2 2 \rightarrow \quad i_j = 3_1 = 1; \ i_j = 3_2 = 1$$

Making the i_j's either 0 or 1, the signs in front of the x's as well as the values of the b's and a's can be determined. For example:

$$(-1)^{0_1} = 1 \quad \rightarrow \quad b = \frac{1}{v}, \ a = \frac{1}{\alpha} \quad ; \quad (-1)^{0_2} = 1 \quad \rightarrow \quad b = \frac{1}{v}, \ a = \frac{1}{\alpha}$$

---

[2] Note that equation $i = \sum_{j=0}^{n-1} i_j 2^j$, in Bárdossy and Li's (2008) paper is replaced here by eq. D.3 for consistency. Moreover, the superscript i, of eq. D.1 in Bárdossy and Li's (2008) paper is corrected here by writing $\sum_{j=1}^{n} i_j$ in order to get the correct sequence of signs in the expansion.

$$(-1)^{1_1} = -1 \quad \rightarrow \quad b = -1,\ a = 1 \qquad ; \qquad (-1)^{1_2} = 1 \quad \rightarrow \quad b = \frac{1}{v},\ a = \frac{1}{v}$$

$$(-1)^{2_1} = 1 \quad \rightarrow \quad b = \frac{1}{v},\ a = \frac{1}{\alpha} \qquad ; \qquad (-1)^{2_2} = -1 \quad \rightarrow \quad b = -1,\ a = 1$$

$$(-1)^{3_1} = -1 \quad \rightarrow \quad b = -1,\ a = 1 \qquad ; \qquad (-1)^{3_2} = -1 \quad \rightarrow \quad b = -1,\ a = 1$$

Then, the vectors ζ_i^T, can be constructed by:

$$\zeta_0^T = \left[\left(\frac{x_1}{v} \right)^{1/\alpha} ; \left(\frac{x_2}{v} \right)^{1/\alpha} \right]$$

$$\zeta_1^T = \left[-x_1 ; \left(\frac{x_2}{v} \right)^{1/\alpha} \right]$$

$$\zeta_2^T = \left[\left(\frac{x_1}{v} \right)^{1/\alpha} ; -x_2 \right]$$

$$\zeta_3^T = \left[-x_1 ; -x_2 \right]$$

Finally, the expansion of eq. D.1 for $n=2$ results:

$$H_2(x_1, x_2) = P\{X_1 \le x_1, X_2 \le x_2\}$$
$$= \Phi(\zeta_0 + \overline{m}) - \Phi(\zeta_1 + \overline{m}) - \Phi(\zeta_2 + \overline{m}) + \Phi(\zeta_3 + \overline{m})$$

(D.6)

A central step is then the determination of the sequence of 0's and 1's, of the i_j's, which in turn determine the signs in front of the values for the x's inside the vectors ζ_i^T. Such sequence is presented in the table below for $n = 3$.

Table E.1 Binary combinations of i_j and resulting signs for $n=3$.

i	i_j	x_1	x_2	x_3
0	000	+	+	+
1	100	-	+	+
2	010	+	-	+
3	110	-	-	+
4	001	+	+	-
5	101	-	+	-
6	011	+	-	-
7	111	-	-	-

D.2 The multivariate density function

The corresponding multivariate density function of X_j is obtained by differentiation of eq. D.1 (Bárdossy and Li, 2008; Li and Bárdossy, 2009; Bárdossy and Pegram, 2009):

$$h_n(x_1,\ldots,x_n) = \frac{1}{(2\pi)^{\frac{n}{2}}\left|\Gamma^{-1}\right|^{\frac{1}{2}}} \sum_{i=0}^{2^n-1} \frac{1}{(\alpha\cdot v)^{n-\sum_{j=0}^{n-1} i_j}} \left[\left(\frac{x_p}{v}\right)^{\frac{1}{\alpha}-1}\right]^{n-\sum_{j=0}^{n-1} i_j} \exp\left(-\frac{1}{2}(\zeta_i+m)^T \Gamma^{-1}(\zeta_i+m)\right)$$

(D.7)

where $\left|\Gamma^{-1}\right|$, denotes the determinant of the inverse matrix of Pearson correlation coefficients of the Y normal random field to be transformed, T, is the transpose of the vectors (ζ_i+m) and x_p is the product of the x's variables according to the sequence exemplified below.

D.2.1 Expansion of eq. D.7 for the bivariate case

For $n=2$, the expansion of eq. D.7 is:

$$h_2(x_1,x_2) = \frac{1}{2\pi\left|\Gamma^{-1}\right|^{\frac{1}{2}}} \cdot \left\{ \frac{1}{(\alpha v)^2}\left(\frac{x_1 x_2}{v^2}\right)^{\frac{1}{\alpha}-1} \exp\left(-\frac{1}{2}(\zeta_0+\overline{m})^T \Gamma^{-1}(\zeta_0+\overline{m})\right) + \ldots \right.$$

$$\ldots + \frac{1}{(\alpha v)}\left(\frac{x_1}{v}\right)^{\frac{1}{\alpha}-1} \exp\left(-\frac{1}{2}(\zeta_1+m)^T \Gamma^{-1}(\zeta_1+m)\right) + \ldots$$

$$\ldots + \frac{1}{(\alpha v)}\left(\frac{x_2}{v}\right)^{\frac{1}{\alpha}-1} \exp\left(-\frac{1}{2}(\zeta_2+\overline{m})^T \Gamma^{-1}(\zeta_2+\overline{m})\right) + \ldots$$

$$\left. \ldots + \exp\left(-\frac{1}{2}(\zeta_3+\overline{m})^T \Gamma^{-1}(\zeta_3+\overline{m})\right) \right\}$$

APPENDIX E

EQUATIONS OF SOME CONNECTIVITY FUNCTIONS

E.1 The connectivity function and the two-point cluster function

In addition to the Euler function used in chapters 1 and 2, the connectivity of simulated realizations can be quantified through connectivity functions. Although there is not a single mathematical definition for this concept (Renard & Allard, 2013), the so-called connectivity function of Allard (1994) is also employed in applications. The connectivity function of a stationary random field is defined as:

$$\tau_1\left(\mathbf{h};z\right) = P\left[\mathbf{x} \leftrightarrow \mathbf{x}+\mathbf{h} \mid Z\left(\mathbf{x}\right), Z\left(\mathbf{x}+\mathbf{h}\right) \in X\right] \qquad \text{(E.1)}$$

The connectivity function express the probability that a value at $\mathbf{x}$ of X is connected "$\leftrightarrow$" with another value of X at $\mathbf{x}+\mathbf{h}$; where X is the subset of Ω with Ω in $\mathfrak{R}^{1,2 \text{ o } 3}$ in which $I(\mathbf{x};z)$=1. $I(\mathbf{x};z)$ is a binary indicator random variable such that: $I(\mathbf{x};z)$=1 if $Z(\mathbf{x})$≤z or $I(\mathbf{x};z)$=0 if $Z(\mathbf{x})$>z.

In equation (E.1), $\left\{\mathbf{x} \leftrightarrow \mathbf{x}+\mathbf{h}\right\}$ necessitates $\left\{Z(\mathbf{x}), Z(\mathbf{x}+\mathbf{h}) \in X\right\}$, therefore the unconditional probability $P[\mathbf{x} \leftrightarrow \mathbf{x}+\mathbf{h}]$ can be decomposed by (Torquato $et\ al.$, 1988):

$$P\left[\mathbf{x} \leftrightarrow \mathbf{x}+\mathbf{h}\right] = P\left[\mathbf{x} \leftrightarrow \mathbf{x}+\mathbf{h} \mid \mathbf{x}, \mathbf{x}+\mathbf{h} \in X\right] P\left[\mathbf{x}, \mathbf{x}+\mathbf{h} \in X\right]$$
$$\tau_2\left(\mathbf{h};z\right) = \tau_1\left(\mathbf{h};z\right) K_I\left(\mathbf{h}\right) \qquad \text{(E.2)}$$

where $K_I(\mathbf{h})$ is the non centered covariance function of X. Equation (E.2) is called the two-point cluster function (Torquato $et\ al.$, 1988).

Recalling that: $K_I(\mathbf{h}) = P[Z(\mathbf{x}) \leq z_1; Z(\mathbf{x}+\mathbf{h}) \leq z_2]$; equation (E.2) can be writing in terms of copulas as:

$$\tau_2\left(\mathbf{h};z\right)=\tau_1\left(\mathbf{h};z\right)C\left(u_1=F_{Z_1}(z_1),u_2=F_{Z_2}(z_2)\right) \tag{E.3}$$

where: $C(u_1,u_2)$ is the bivariate copula.

To evaluate the average distance over which two points are connected, the integral scale of eq. (E.3) can be computed:

$$I_{\tau_2}=\int_0^\infty \tau_2\left(\mathbf{h};z\right)dh \tag{E.4}$$