



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA
INGENIERÍA CIVIL - ANÁLISIS Y DISEÑO DE PRESAS DE TIERRA

**FILTRACIÓN DE AGUA EN MASAS TÉRREAS
HETEROGÉNEAS**

TESIS
QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN INGENIERÍA

PRESENTA:
FELIPE VÁZQUEZ GUILLÉN

TUTOR PRINCIPAL
GABRIEL, AUVINET, GUICHARD, INSTITUTO DE INGENIERÍA, UNAM

COMITÉ TUTOR
DANIEL, RESENDIZ, NÚÑEZ, INSTITUTO DE INGENIERÍA, UNAM
EFRAÍN, OVANDO, SHELLEY, INSTITUTO DE INGENIERÍA, UNAM
ERNESTO, HEREDIA, ZAVONI, PROGRAMA DE MAESTRÍA Y DOCTORADO EN
INGENIERÍA
ANDRAS, BÁRDOSSY, PROGRAMA DE MAESTRÍA Y DOCTORADO EN
INGENIERÍA

MÉXICO, D. F., OCTUBRE 2014



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

JURADO ASIGNADO:

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

Lugares donde se realizó la tesis: INSTITUTO DE INGENIERÍA GEOTÉCNICA, UNIVERSIDAD DE STUTTGART, ALEMANIA, DEPARTAMENTO DE INGENIERÍA CIVIL Y DEL MEDIO AMBIENTE, UNIVERSIDAD DEL SUR DE CALIFORNIA, EUA, INSTITUTO DE INGENIERÍA, UNAM.

TUTOR DE TESIS:

DR. GABRIEL AUVINET GUICHARD

A handwritten signature in black ink, appearing to read 'G. Auvinet', is written over a horizontal dashed line.

FIRMA

RESUMEN

El objetivo de esta tesis es desarrollar herramientas numéricas para caracterizar masas térricas 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.

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 discuss 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 hypothetical aquifers is also considered. The presentation of this research finalizes with general conclusions and recommendations that attempt to motivate future investigations.

CONTENIDO

Resumen	
Abstract	
Agradecimientos	
Nomenclatura	
Lista de figuras	
Lista de tablas	

Introducción

Motivación.....	iii
Definición del problema.....	iii
Contribución.....	iv
Organización del documento.....	iv

1. Flujo de agua en medios aleatorios continuos: Una revisión del concepto de dependencia multi-gaussiana.....	1
1.1 El concepto de campo aleatorio.....	3
1.2 El campo aleatorio multi-gaussiano.....	4
1.3 Características de la dependencia multi-gaussiana.....	5
1.4 Necesidad de una dependencia no multi-gaussiana.....	10
1.5 Conclusiones.....	12
2. Simulación de campos aleatorios no multi-gaussianos por medio de cópulas.....	15
2.1 Definiciones.....	18
2.2 Modelado espacial.....	21
2.3 Simulación no condicional.....	26
2.4 Simulación condicional.....	30
2.5 Conclusiones.....	35
3. Solución al problema inverso con un filtro de Kalman ensamblado modificado....	37
3.1 Ecuaciones del flujo de aguas subterráneas.....	38
3.2 Simulación de campos aleatorios espacio-temporales.....	39
3.3 Descripción del experimento numérico.....	45
3.4 Resultados y discusiones.....	48
3.5 Conclusiones.....	55
4. Caracterización de campos heterogéneos de la conductividad hidráulica.....	57
4.1 Exposición del problema.....	58
4.2 Descripción del experimento numérico.....	59
4.3 Criterios para evaluar el error.....	61
4.4 Resultados y discusiones.....	64
4.5 Conclusiones.....	70

5. Detección de trayectorias preferenciales de filtración en el núcleo de una presa triangular.....	73
5.1 Exposición del problema.....	75
5.2 Descripción del experimento numérico	76
5.3 Campos de la conductividad condicionales a historias de cargas hidráulicas...	79
5.4 Resultados y discusiones.....	81
5.5 Conclusiones.....	85
Conclusiones generales y perspectivas.....	87
Bibliografía.....	89
Apéndices	
A. Conceptos básicos de la filtración en medios aleatorios saturados.....	A-1
B. Conductividad efectiva y conductividad equivalente.....	B-1
C. Algunas definiciones y propiedades de las cópulas multivariadas.....	C-1
D. Funciones multivariadas de la cópula V -transformada.....	D-1
E. Ecuaciones de algunas funciones de conectividad.....	E-1

NOMENCLATURA

LISTA DE VARIABLES

a_x	Escala de correlación en la dirección x
$\hat{c}(\mathbf{h}, u_1, u_2)$	Densidad de la copula empírica bivariada
C	Número total de agujeros en N objetos de una imagen binaria
$C(u, v)$	Cópula bivariada de las variables aleatorias U y V
$C(u_1, \dots, u_n)$	Cópula multivariada
C_S	Cópula espacial bivariada
$C_{\mathbf{x} n}(\cdot)$	Cópula condicional
CV_{K_S}	Coefficiente de variación de la conductividad hidráulica saturada
$C(\mathbf{h}; z_p)$	Covarianza del campo aleatorio indicador $I(\mathbf{x}; z_p)$, llamada covarianza indicadora
$C(0; z_p)$	Varianza del campo aleatorio indicador
$C_Y(\mathbf{x}_1, \mathbf{x}_2)$	Función de auto-covarianza
$C_Y(\mathbf{x}_1 - \mathbf{x}_2)$	Función de auto-covarianza de un campo aleatorio estacionario
$C_H(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)$	Funciones de auto-covarianza entre cargas hidráulicas transformadas
$C_Y(\mathbf{x}_i, \mathbf{x}_j)$	Funciones de auto-covarianza entre valores transformados del logaritmo natural de la conductividad hidráulica
$C_{YH}(\mathbf{x}_j, \boldsymbol{\chi}_i)$	Funciones de covarianza cruzada entre valores transformados del logaritmo natural de la conductividad hidráulica y valores transformados de la carga hidráulica

$C_{\Gamma}^G(\cdot)$	Cóputa gaussiana multivariada con matriz de correlación Γ
$C_{\lambda\Gamma}^V(\cdot)$	Cóputa V -transformada multivariada con parámetros λ y matriz de correlación Γ
e_i	i -ésimo elemento finito
$f(z_1, z_2)$	Función de densidad de probabilidad bivariada entre las variables aleatorias Z_1 y Z_2
$F_{1, \dots, n}^{-1}$	Inverso de las funciones de distribución marginales
$\hat{F}_n(z)$	n -ésima función de distribución empírica
$F_{\hat{H}}(\boldsymbol{\chi})$	Función de distribución acumulativa de la carga hidráulica transformada y actualizada en la ubicación $\boldsymbol{\chi}$
$F_{Y_i}(y_i; \mathbf{x}_i)$	i -ésima función de distribución marginal de la variable aleatoria Y ubicada en \mathbf{x}_i
$F_{\hat{Y}}(\mathbf{x})$	Función de distribución acumulativa del logaritmo natural de la conductividad hidráulica transformada y actualizada en la ubicación \mathbf{x}
$F_V(\mathbf{v})$	Función de distribución acumulativa (empírica) del campo V -transformado
g	Aceleración de la gravedad
$g(\mathbf{x})$	Campo multi-gaussiano
G	Función de distribución acumulativa gaussiana teórica
h_0	Carga hidráulica inicial en el tiempo $t=0$
h_1	Carga hidráulica preescrita
$h'_1(\boldsymbol{\chi}_i)$	Valor transformado de una observación del campo de cargas hidráulicas ubicado en la posición $\boldsymbol{\chi}_i$ y en el tiempo $t=1$
$h'_1(\boldsymbol{\chi}_i)$	Valor transformado de la carga hidráulica ubicada en la posición $\boldsymbol{\chi}_i$ y en el tiempo $t=1$
$\hat{h}'_1(\boldsymbol{\chi})$	Realización del campo aleatorio de cargas hidráulicas transformadas en el tiempo $t=1$ generada con la técnica de simulación “ p -field”
\mathbf{h}	Vector entre dos ubicaciones cualquiera
H	Carga hidráulica
H_f	Entropía de la función de densidad de probabilidad f
$H(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n))$	Función de distribución multivariada
$H(\boldsymbol{\chi})$	Campo aleatorio de cargas hidráulicas
$H_1(\boldsymbol{\chi})$	Campo aleatorio espacio-temporal de cargas hidráulicas en el tiempo $t=1$

\mathbf{H}_t^f	Vector conjunto de estado reducido (solo incluye valores a priori en las posiciones de las observaciones) de dimensión N_y+N_h , definido para cada realización en el ensamble
\mathbf{H}_t^u	Vector de dimensión N que contiene realizaciones de la carga hidráulica actualizada y transformada a las variables originales
$I(\mathbf{x},z_p)$	Campo aleatorio indicador definido para el valor del umbral z_p
k	Parámetro de la V -transformación
k_s	Valor de la conductividad hidráulica saturada
$k_s(\mathbf{x})$	Campo de conductividades hidráulicas saturadas
$K_s(\mathbf{x})$	Campo aleatorio de la conductividad hidráulica saturada
\mathbf{K}_t	Matriz de tamaño $(n+N)\times(N_y+N_h)$ llamada “Kalman gain”
\mathbf{K}_α	Sub-matriz que contiene las funciones ponderadoras α de un sistema de ecuaciones de coKriging simple. \mathbf{K}_β , \mathbf{K}_λ y \mathbf{K}_η son también sub-matrices que contiene las funciones ponderadoras β , λ , η , respectivamente, de dicho sistema
m	Parámetro de la V -transformación
$m(\mathbf{x},t_1)$	Media aritmética en \mathbf{x} y en el tiempo t_1
n	Número total de conductividades hidráulicas en una región de flujo
n_r	Número total de realizaciones en un conjunto o ensamble
$\mathbf{n}(\mathbf{x})$	Vector saliente, normal a la frontera Γ_N
N	Número de objetos aislados en una imagen binaria. También se utiliza para representar el número total de cargas hidráulicas en una región de flujo discretizada
N_y	Número total de mediciones directas de la conductividad hidráulica
N_h	Número total de observaciones de la carga hidráulica
$p=F_Z(z_p)$	Percentil de la distribución de probabilidad de $Z(\mathbf{x})$
$p_1(\mathbf{x})$	Campo de probabilidades $\in[0,1]$ de una realización del logaritmo natural de la conductividad hidráulica en el tiempo $t=1$
$P[Y(\mathbf{x}_i)\leq y_i]$	Probabilidad de que el valor de la variable aleatoria Y localizada en \mathbf{x}_i sea menor o igual que el valor y_i
$P[Y Y=y,H=h]$	Probabilidad condicional de la variable aleatoria Y
$\mathbf{q}(\mathbf{x},t)$	Velocidad preescrita en los segmentos de frontera de Neumann Γ_N

$s_{en}^2(\mathbf{x}_i)$	Varianza en la ubicación \mathbf{x}_i determinada estadísticamente con el conjunto de realizaciones en el ensamble
$S \in \mathfrak{R}^n$	Dominio de interés en el espacio de dimensión $n=1,2$ o 3
$S(\mathbf{h})$	Pares de valores entre las funciones de distribución separadas por el vector \mathbf{h}
$S_s(\mathbf{x})$	Almacenamiento específico
t	Tiempo
u_1, \dots, u_n	Variables aleatorias uniformemente distribuidas en el intervalo $[0,1]$
(u, v)	Valores de las funciones de distribución de las variables aleatorias uniformes U y V
(U_1, U_2)	Vectores que contienen realizaciones de variables aleatorias correlacionadas uniformemente distribuidas
U_t^u	Vector conjunto de dimensión $n+N$ que contiene realizaciones actualizadas
U_0	Vector de estado conjunto de dimensión $n+N$ que contiene valores a priori de las variables por estimar de cada realización en el ensamble
$v(\mathbf{x})$	Campo que se obtiene a partir de la V -transformación
$V(\mathbf{x})$	Campo aleatorio de velocidades de filtración
w	Función de contribución de varianza
\mathbf{x}	Coordenadas espaciales asociadas a la región de flujo Ω en donde se define la conductividad hidráulica
(X_1, X_2)	Vectores que contienen realizaciones de variables aleatorias correlacionadas
y	Valor asociado al logaritmo natural de la conductividad hidráulica
$y'=G^{-1}(\cdot)$	Valor del inverso de la función de distribución acumulativa gaussiana
$y'(\mathbf{x})$	Campo del logaritmo natural de la conductividad hidráulica cuya función de distribución univariada es gaussiana con media cero y varianza unitaria; es decir, $N(0,1)$
$y^*(\mathbf{x}_i)$	Valor estimado del logaritmo natural de la conductividad hidráulica en la posición \mathbf{x}_i
$y^{ref}(\mathbf{x}_i)$	Valor del logaritmo natural de la conductividad hidráulica del campo de referencia en la posición \mathbf{x}_i
$\hat{y}'_1(\mathbf{x})$	Realización del campo transformado del logaritmo natural de la conductividad hidráulica en el tiempo $t=1$
$y'_0(\mathbf{x}_i)$	Valor transformado del logaritmo natural de la conductividad hidráulica en la ubicación \mathbf{x}_i y en el tiempo $t=0$

$y'_1(\mathbf{x}_i)$	Valor transformado del logaritmo natural de la conductividad hidráulica en la ubicación \mathbf{x}_i y en el tiempo $t=1$ donde se dispone de observaciones
$\hat{y}'_1(\mathbf{x})$	Realización del campo transformado del logaritmo natural de la conductividad hidráulica en el tiempo $t=1$ generado con la técnica de simulación “ p -field”
$Y(\mathbf{x})$	Campo aleatorio del logaritmo natural de la conductividad hidráulica
$Y_0(\mathbf{x})$	Campo aleatorio espacio-temporal de la conductividad hidráulica en el tiempo $t=0$
\mathbf{Y}_t^u	Vector de dimensión n que contiene valores actualizados del logaritmo natural de la conductividad hidráulica de cada una de las realizaciones en el ensamble
z	Elevación, Valor de un umbral
$(\mathbf{Z}_1, \mathbf{Z}_2)$	Vectores que contienen realizaciones de variables aleatorias correlacionadas log-normales
$z(\mathbf{x})$	Realización de un campo aleatorio de conductividades hidráulicas
$Z(\mathbf{x})$	Campo aleatorio de la conductividad hidráulica
\mathbf{Z}_t	Vector conjunto de dimensión N_h que contiene los valores de las observaciones

LISTA DE SÍMBOLOS

α	Parámetro de la V -transformación
$\alpha(\mathbf{x})$	Función ponderadora de un sistema de ecuaciones de coKriging simple
$\beta(\boldsymbol{\chi})$	Función ponderadora de un sistema de ecuaciones de coKriging simple
$\boldsymbol{\chi}$	Coordenadas espaciales asociadas a la región de flujo Ω en donde se define la carga hidráulica
$\chi(p)$	Función de Euler
Γ_D	Frontera con cargas hidráulicas preescritas
η_{ij}	Frecuencias empíricas definidas dentro de una malla regular de $m \times m$ con centros en las coordenadas i, j
$\eta(\boldsymbol{\chi})$	Función ponderadora de un sistema de ecuaciones de coKriging simple
$\lambda(\mathbf{x})$	Función ponderadora de un sistema de ecuaciones de coKriging simple
$\boldsymbol{\lambda}$	Vector que contiene los parámetros de la cópula V -transformada

μ	Viscosidad de un fluido
μ_Y	Valor medio de la variable aleatoria Y
$\mu_Y(\mathbf{x})$	Valor medio del campo aleatorio $Y(\mathbf{x})$
$\mu_{Y H}$	Valor medio condicional de la variable aleatoria Y
$\mu_{\hat{Y}}(\mathbf{x})$	Valor medio de la función de distribución $F_{\hat{Y}(\mathbf{x})}$ ubicada en \mathbf{x}
$\mu_{\hat{H}}(\boldsymbol{\chi})$	Valor medio de la función de distribución $F_{\hat{H}(\boldsymbol{\chi})}$ ubicada en $\boldsymbol{\chi}$
μ_{Ks}	Valor medio de la conductividad hidráulica saturada
$\hat{\mu}_{Y H}(\mathbf{x}, t_1)$	Estimador del valor medio condicional del campo aleatorio Y en la ubicación \mathbf{x} y en el tiempo t_1
σ_n^2	Varianza de una variable aleatoria con distribución normal
σ_Y^2	Varianza de la variable aleatoria Y
$\sigma_Y^2(\mathbf{x})$	Varianza del campo aleatorio $Y(\mathbf{x})$
$\sigma_{Y H}^2$	Varianza condicional de la variable aleatoria Y
$\sigma_{\hat{Y}}(\mathbf{x})$	Desviación estándar de la función de distribución $F_{\hat{Y}(\mathbf{x})}$ ubicada en \mathbf{x}
$\sigma_{\hat{H}}(\boldsymbol{\chi})$	Desviación estándar de la función de distribución $F_{\hat{H}(\boldsymbol{\chi})}$ ubicada en $\boldsymbol{\chi}$
σ_Y^2	Varianza de la variable aleatoria Y
$\hat{\sigma}_{Y H}^2(\mathbf{x}, t_1)$	Estimador del valor medio condicional del campo aleatorio Y en la ubicación \mathbf{x} y en el tiempo t_1
ρ_d	Densidad del fluido
ρ_n	Coefficiente de correlación de <i>Pearson</i> entre variables aleatorias normalmente distribuidas
ρ_{ln}	Coefficiente de correlación de <i>Pearson</i> entre variables aleatorias con distribución log-normal
$\rho(\mathbf{h})$	Correlograma
$\rho_I(\mathbf{h}; z_p)$	Correlograma indicador
ρ_s	Coefficiente de correlación de <i>Spearman</i>
$\rho_s(\mathbf{h})$	Función de correlación de <i>Spearman</i>
τ	Coefficiente de correlación de <i>Kendall</i>

$\bar{\tau}, \tau^+$	Coeficientes de correlación condicional de <i>Kendall</i>
Φ	Función de distribución acumulativa Gaussiana
Ω	Región de flujo con Ω en \mathbb{R}^p donde $p=1$ o 2
∇	Operador gradiente

LISTA DE FIGURAS

Fig. 1.1	Principales etapas involucradas en un enfoque probabilista basado en simulaciones.....	2
Fig. 1.2	Coefficiente de correlación lineal entre variables aleatorias log-normales como una función del coeficiente de correlación lineal entre variables aleatorias normales con distintas varianzas. Se muestra una línea interrumpida con pendiente 1:1 para referencia.....	5
Fig. 1.3	Correlogramas indicadores de un campo aleatorio bivariado gaussiano $Z(\mathbf{x})$ en función de la distancia. Para un correlograma exponencial (izquierda), un correlograma esférico (centro) y un correlograma gaussiano (derecha).....	7
Fig. 1.4	Correlogramas indicadores de un campo aleatorio bivariado gaussiano $Z(\mathbf{x})$ en función del cuantil q asociado al umbral z . Para: $\mathbf{h}=0.05$ (izquierda), $\mathbf{h}=0.5$ (centro) y $\mathbf{h}=0.9$ (derecha).....	7
Fig. 1.5	Correlogramas indicadores de un campo aleatorio bivariado gaussiano $Z(\mathbf{x})$ en función del correlograma de $Z(\mathbf{x})$. Para los correlogramas: exponencial (izquierda), esférico (centro) y gaussiano (izquierda).....	7
Fig. 1.6	Conectividad expresada por medio del número de Euler de una realización representativa de un campo aleatorio multi-gaussiano con función de autocovarianza gaussiana. a) Realización. b)-f) Imágenes binarias en diferentes umbrales. g) Función de Euler de la realización.....	9
Fig. 2.1	Construcción de una cópula bivariada usando un par de vectores aleatorios gaussianos correlacionados. a) Diagrama de dispersión de los vectores gaussianos correlacionados. b) y c) Funciones de distribución de los vectores aleatorios. d) y e) Valores de la probabilidad de las funciones de distribución en b) y c), respectivamente. f) La cópula bivariada de los vectores aleatorios $(\mathbf{X}_1, \mathbf{X}_2)$. g) y h) Las funciones de distribución t de <i>Student</i> y <i>Gamma</i> que se obtienen con los valores de probabilidad en d) y e), respectivamente.....	20
Fig. 2.2	Realizaciones representativas de diferentes campos aleatorios y sus correspondientes gráficos de densidades de cópulas empíricas. a) Realización de un campo aleatorio gaussiano con función de auto-covarianza exponencial. b) Realización de un campo aleatorio gaussiano con función de auto-covarianza gaussiana. c) Realización de un campo aleatorio no gaussiano con función de auto-covarianza exponencial. Los histogramas y correlogramas experimentales de cada realización se muestran en la parte alta de la figura.....	23
Fig. 2.3	Gráficos de densidades de cópulas teóricas gaussianas bivariadas para diferentes	

	coeficientes de correlación.....	24
Fig. 2.4	Gráficos de densidades de cópulas teóricas V -transformadas bivariadas para diferentes parámetros m , k y α	25
Fig. 2.5	Funciones de correlación condicional de <i>Kendall</i> para valores debajo de la mediana $t(0.5)$ y arriba de la mediana $t^+(0.5)$ para una cópula V -transformada con parámetros $m=0.0$, $k=2.5$ y $\alpha=0.25$	28
Fig. 2.6	Funciones de correlación de <i>Spearman</i> de diez realizaciones no condicionales, no multi-gaussianas (líneas interrumpidas) y la función de correlación teórica de un modelo esférico (línea sólida) (Deutsch y Journel, 1998).....	29
Fig. 2.7	Realizaciones representativas del campo aleatorio del logaritmo natural de la conductividad $Y(\mathbf{x})=\ln(K_s(\mathbf{x}))$. a) Valores grandes mejor estructurados. b) Valores pequeños mejor estructurados.....	29
Fig. 2.8	Descriptores estadísticos de las realizaciones en la Fig. 2.7. a) Histogramas. b) Funciones de correlación experimentales en las direcciones horizontal y vertical. c) Funciones de correlación condicional de <i>Kendall</i> en la dirección horizontal (Manner, 2010).....	30
Fig. 2.9	Realizaciones representativas de campos aleatorios generadas por medio de cópulas utilizando la técnica de simulación secuencial. Todos los campos aleatorios comparten funciones de distribución gaussianas y funciones de correlación de <i>Spearman</i> idénticas sin embargo difieren en la asimetría que imponen los parámetros m , k y α de la cópula. a) Histogramas y funciones de correlación empíricas en la dirección horizontal. b) Realizaciones de los campos aleatorios con los parámetros de la cópula que se indican debajo de ellas. c) Densidades de las cópulas empíricas (derecha) y las correspondientes versiones teóricas (izquierda).....	34
Fig. 2.10	Número de Euler como una función del percentil p y las correspondientes imágenes binarias en $\chi(p)=0$. a) Funciones de Euler de las realizaciones simuladas con la cópula V -transformada cuyos parámetros se indican debajo de cada figura. b) Imágenes binarias en $\chi(p)=0$	35
Fig. 3.1	Ilustración de la etapa de transformación del EnKF y el pf-EnKF. a) Proceso de transformación gaussiana. b) Proceso de retransformación.....	43
Fig. 3.2	Ilustración del proceso de simulación descrito por la ecuación 3.16. a) Generación de los campos p . b) Muestreo de las funciones de distribución condicionales utilizando los campos p	44
Fig. 3.3	Campos unidimensionales. a) Campo gaussiano inicial. b) Campo después de aplicar la V -transformación con parámetros $m=0$, $k=1$ y $\alpha=2$ al campo gaussiano inicial. c) Campo final del logaritmo natural de la conductividad después de imponerle una distribución normal con valor esperado $\mu_Y=-1.654$ y varianza $\sigma_Y^2=1$ al campo V -transformado. También se reportan las estadísticas de los valores muestreados (cuadrados vacíos).....	45
Fig. 3.4	Funciones de auto-covarianza estandarizada del campo gaussiano inicial y del logaritmo natural de la conductividad del campo de referencia. También se muestra una función exponencial teórica con fines comparativos.....	46
Fig. 3.5	Perfil de cargas hidráulicas en el acuífero de referencia al tiempo $t=0$ días. También se indican con cuadrados sólidos las profundidades de las puntas de dos piezómetros (Pz-1 y Pz-2).....	47
Fig. 3.6	Transformación gaussiana de las cargas hidráulicas en el nodo 61. a) Histograma de las cargas hidráulicas no transformadas. b) Función de la anamorfosis gaussiana (con	

	media cero y varianza unitaria). c) Histograma de cargas hidráulicas después de la anamorfosis gaussiana.....	48
Fig. 3.7	Relación entre el logaritmo natural de la conductividad y la carga hidráulica en dos ubicaciones seleccionadas arbitrariamente (ρ es el coeficiente de correlación de <i>Pearson</i>). a) Antes de la transformación de ambas variables. b) Después de la transformación gaussiana de ambas variables.....	49
Fig. 3.8	Campos del logaritmo natural de la conductividad condicionados a mediciones de ésta misma variable. a) Con el método EnKF. b) Con el método pf-EnKF. También se muestra el campo de referencia.....	50
Fig. 3.9	Perfiles de la desviación estándar condicional del logaritmo natural de la conductividad hidráulica con respecto a la profundidad para diferentes tiempos (los cuadrados vacíos indican las ubicaciones de los valores conocidos). a) De los campos del EnKF. b) De los campos del pf-EnKF.....	50
Fig. 3.10	Campos del logaritmo natural de la conductividad condicionados a historias de cargas hidráulicas con el método EnKF. También se muestra el campo de referencia. a) Al tiempo $t=3$ días. b) Al tiempo $t=18$ días. c) Al tiempo $t=60$ días.....	51
Fig. 3.11	Campos del logaritmo natural de la conductividad condicionados a historias de cargas hidráulicas con el método pf-EnKF. También se muestra el campo de referencia. a) Al tiempo $t=3$ días. b) Al tiempo $t=18$ días. c) Al tiempo $t=60$ días.....	51
Fig. 3.12	Perfiles de la desviación estándar condicional del logaritmo natural de la conductividad hidráulica con respecto a la profundidad para diferentes tiempos (los cuadrados vacíos indican las ubicaciones de los valores conocidos). a) De los campos del EnKF. b) De los campos del pf-EnKF.....	52
Fig. 3.13	Distribución de frecuencias del logaritmo natural de la conductividad del campo de referencia, del conjunto de realizaciones <i>a priori</i> y de los conjuntos de realizaciones <i>a posteriori</i> del EnKF y pf-EnKF al tiempo $t=60$ días.....	53
Fig. 3.14	Campos condicionales de la carga de presión que se obtienen con el método EnKF. También se muestra el campo de referencia. a) Al tiempo $t=3$ días. b) Al tiempo $t=18$ días. c) Al tiempo $t=60$ días.....	53
Fig. 3.15	Campos condicionales de la carga de presión que se obtienen con el método pf-EnKF. También se muestra el campo de referencia. a) Al tiempo $t=3$ días. b) Al tiempo $t=18$ días. c) Al tiempo $t=60$ días.....	54
Fig. 3.16	Perfiles de la desviación estándar condicional de la carga de presión con respecto a la profundidad para diferentes tiempos (los cuadrados sólidos indican las ubicaciones de dos observaciones). a) De los campos del EnKF. b) De los campos del pf-EnKF.....	54
Fig. 4.1	Dimensiones de la región de flujo. Se indican las ubicaciones de las mediciones de la conductividad (círculos vacíos) y de las observaciones de la carga hidráulica (círculos llenos). El triángulo señala la ubicación del pozo de bombeo. Las condiciones de frontera del problema en el tiempo $t \geq 0$ se indican en los lados de la región de flujo.....	59
Fig. 4.2	Modelo teórico de cópula bivariada <i>V</i> -transformada con parámetros $m=5.0$, $k=2.0$ y $\alpha=1.0$ utilizado para generar el campo de referencia (acuífero hipotético).....	60
Fig. 4.3	Parámetros estadísticos de las muestras tomadas del campo de referencia. a) Histograma del logaritmo natural de la conductividad del campo muestreado. La línea sólida gruesa representa una distribución normal ajustada con la media y la varianza que se indican en la figura. b) Funciones de correlación empíricas del campo muestreado en las direcciones X y Y . También se indica la función de correlación	

empírica omni-direccional (O). La línea sólida gruesa representa el modelo teórico ajustado de auto-correlación correspondiente a una función exponencial con la escala de correlación a que se indica..... 61

- Fig. 4.4** La media del campo del logaritmo natural de la conductividad estimado con el método EnKF (al centro) y con el método pf-EnKF (derecha) en los tiempos $t=5$ días y $t=2.5$ días. Se muestra además la media del campo *a priori* del logaritmo natural de la conductividad con fines comparativos (izquierda). Los valores numéricos de las medidas de exactitud local de la estimación se indican a la derecha de los campos estimados. $L1=RMSE$ (raíz cuadrada del error cuadrático medio), $L2=MAE$ (valor absoluto del error medio) y $L3=LEPS$ (error lineal en el espacio de las probabilidades)..... 64
- Fig. 4.5** La media del campo del logaritmo natural de la conductividad estimado con el método EnKF (al centro) y con el método pf-EnKF (derecha) en el tiempo $t=5$ días. El campo de referencia se muestra con fines comparativos (izquierda). Los valores numéricos de las medidas de exactitud local de la estimación se indican a la derecha de los campos estimados. El significado de $L1$, $L2$ y $L3$ se explica en la Fig. 4.4..... 65
- Fig. 4.6** Evolución de la desviación estándar de la estimación del logaritmo natural de la conductividad al tiempo $t=0.75$ días, $t=2.5$ días y $t=5$ días calculada con las estimaciones del EnKF (izquierda) y pf-EnKF (derecha). Las mediciones de la incertidumbre de la estimación se reportan a la derecha de cada imagen. $L1=SPREAD$, $L2=AAD$ (Valor absoluto de la desviación media) y $L4=Var_{tot}$ (Varianza total de la predicción)..... 66
- Fig. 4.7** Distribución de frecuencias del logaritmo natural de la conductividad. a) Del campo de referencia, del conjunto de realizaciones *a priori* y del conjunto de realizaciones *a posteriori* del EnKF al tiempo $t=5$ días. b) Del campo de referencia, del conjunto de realizaciones *a priori* y del conjunto de realizaciones *a posteriori* del pf-EnKF al tiempo $t=5$ días..... 67
- Fig. 4.8** Funciones de correlación empíricas del campo de referencia (línea con símbolos), de un conjunto de realizaciones condicionales al tiempo $t=5$ días (líneas delgadas) y del promedio de dos mil funciones de correlación en el tiempo $t=5$ días (líneas gruesas). a) En la dirección X . b) En la dirección Y 68
- Fig. 4.9** Densidades de la cópula bivariada empírica a una distancia de 20 m en la dirección X . a) Del campo de referencia. b) y c) De la media del campo estimado en el tiempo $t=5$ días. d) y e) Del promedio de dos mil densidades de cópulas que se obtienen de las realizaciones condicionales en el tiempo $t=5$ días..... 68
- Fig. 4.10** Cópulas empíricas bivariadas para tres diferentes distancias en la dirección X al tiempo $t=5$ días. En la primera columna se muestran las cópulas medias de las realizaciones condicionales del EnKF. En la segunda columna se reportan las cópulas medias de las realizaciones condicionales del pf-EnKF. Los valores en la parte inferior izquierda de cada figura son las distancias de Kolmogorov ($D1$)..... 69
- Fig. 4.11** Promedio de la distancia de Kolmogorov ($D1$) calculada con dos mil cópulas obtenidas de las realizaciones condicionales del logaritmo natural de la conductividad hidráulica en el tiempo $t=5$ días para diez diferentes separaciones. a) En la dirección X . b) En la dirección Y 70
- Fig. 5.1** Presa de enrocamiento y la idealización de la sección transversal de su núcleo interno (Harr, 1962). Presa de enrocamiento. b) Idealización..... 75
- Fig. 5.2** Realización del campo aleatorio del logaritmo natural de la conductividad hidráulica generada por medio de la V -transformación con parámetros $m=0.0$, $k=2.5$ y $\alpha=0.25$. Se muestran además sus descriptores estadísticos. a) Realización de log-conductividades $Y(\mathbf{x})=\ln(Ks(\mathbf{x}))$. b) Histograma. c) Funciones de correlación en las

	direcciones horizontal y vertical. d) Funciones de correlación condicional de <i>Kendall</i> (Manner, 2010) en la dirección horizontal.....	77
Fig. 5.3	La región de flujo discretizada y las ubicaciones de nueve piezómetros (círculos rellenos).....	78
Fig. 5.4	Campos de referencia de la carga hidráulica [m]. a) En flujo establecido. b) En flujo transitorio después de un “vaciado rápido” del embalse, es decir, suponiendo que el material permanece saturado.....	78
Fig. 5.5	Historias piezométricas.....	78
Fig. 5.6	Campos de velocidades promedio [m/día] que se obtienen de las realizaciones de la conductividad condicionadas a nueve historias piezométricas. También se muestra el campo de velocidades de referencia con fines comparativos.....	82
Fig. 5.7	Caso 1. Intervalos de confianza de las velocidades pronosticadas en el tiempo $t=70$ días (líneas interrumpidas). Las velocidades de referencia se indican con líneas sólidas. Intervalos de confianza del EnKF (izquierda). Intervalos de confianza del pf-EnKF (derecha). a) y b) Perfiles a lo largo del talud de aguas arriba. b) y d) Perfiles a lo largo del eje central del núcleo.....	83
Fig. 5.8	Caso 2. Intervalos de confianza de las velocidades pronosticadas en el tiempo $t=70$ días (líneas interrumpidas). Las velocidades de referencia se indican con líneas sólidas. Intervalos de confianza del EnKF (izquierda). Intervalos de confianza del pf-EnKF (derecha). a) y b) Perfiles a lo largo del talud de aguas arriba. b) y d) Perfiles a lo largo del eje central del núcleo.....	83

LISTA DE TABLAS

Tabla 2.1	Modelo de correlación anidado para cada caso analizado. Las contribuciones de la varianza y escalas de correlación corresponden a las funciones de auto-covarianza gaussiana, esférica y exponencial, respectivamente.....	33
Tabla 4.1	Comparación entre los parámetros estadísticos de los campo de referencia y de los campos <i>a priori</i> y estimados después de completar el proceso de condicionamiento ($t=5$ días).....	70
Tabla 5.1	Estadísticas de los conjuntos iniciales de realizaciones no condicionales del logaritmo natural de la conductividad $Y(\mathbf{x})=\ln(K_s(\mathbf{x}))$ utilizados en el modelación inversa.....	79
Tabla 5.2	Resultados de la validación cruzada de la media estimada del campo del logaritmo natural de la conductividad $Y(\mathbf{x})$ y de las predicciones de la media de los campos de carga hidráulica $H(\mathbf{x})$ y velocidad $V(\mathbf{x})$ después de completar el proceso de condicionamiento ($t=70$ días). La media de los campos de la carga hidráulica y velocidad se obtuvieron estadísticamente a partir de los resultados de la solución del problema directo usando los campos condicionales de la conductividad hidráulica. Se indican los resultados de cada caso analizado.....	84

INTRODUCCIÓN

Motivación

La conductividad hidráulica de las masas térreas es muy sensible a variaciones de estructura interna. En un mismo tipo de suelo, la conductividad hidráulica puede diferir en varios órdenes de magnitud. Las mediciones de campo muestran claramente que estas variaciones son la regla y no la excepción. Este hecho se cumple tanto en las formaciones naturales como en los materiales mecánicamente estabilizados en donde las heterogeneidades son favorecidas, principalmente, por las propiedades intrínsecas de los materiales en los bancos de préstamo y por las imperfecciones cometidas durante la colocación de los mismos. Además, el número de mediciones disponibles para determinar la conductividad hidráulica de una masa de suelo en un sitio específico suele ser limitado. En consecuencia, las propiedades hidráulicas de las masas de suelo son tan inciertas que se pueden caracterizar mejor desde una perspectiva probabilista.

Los modelos numéricos se utilizan con frecuencia para analizar el flujo de agua en suelos en los casos de interés práctico. Estos modelos permiten incorporar en los análisis la heterogeneidad del medio. En los casos de la práctica, la presencia de heterogeneidades es crítica para el correcto funcionamiento o incluso la seguridad de las masas térreas sujetas a filtración de agua. K. V. Terzaghi por ejemplo advirtió que detalles geológicos aparentemente insignificantes frecuentemente condicionan el comportamiento de la masa térrea. Ante estas condiciones, es inconveniente continuar simplificando los análisis de flujo de agua utilizando valores medios en dominios homogéneos idealizados.

Definición del problema

Un caso particular de alta relevancia práctica en ingeniería geotécnica es la identificación de trayectorias preferenciales de filtración en el núcleo interno de las presas de tierra. En medios heterogéneos, el movimiento del fluido se concentra a lo largo de trayectorias preferenciales con menor resistencia al paso del mismo. La existencia de zonas más permeables en presas de tierra tiene diversos orígenes. Pueden ser el resultado de la colocación de un material más grueso que el especificado o el de una compactación inadecuada, por mencionar sólo algunos.

Si bien la existencia de trayectorias preferenciales de filtración en presas de tierra no necesariamente implica una condición de inestabilidad, ésta puede señalar el riesgo de erosión interna o la presencia de agrietamiento interno. En el caso de un comportamiento inadecuado de los filtros, la filtración preferencial puede facilitar el desarrollo de un proceso de tubificación. La tubificación ha sido la principal causa de falla en presas de tierra. Por lo tanto, la detección

de trayectorias preferenciales de filtración en su estado incipiente es de alta relevancia en la ingeniería de presas de tierra.

Contribución

El objetivo de esta tesis es desarrollar herramientas numéricas para caracterizar masas téreas heterogéneas en relación con sus conductividades hidráulicas desde una perspectiva de análisis probabilista. Las herramientas propuestas se utilizarán para simular campos aleatorios condicionales a mediciones de la conductividad hidráulica y/o a historias de cargas hidráulicas. Las herramientas desarrolladas durante la presente investigación se validarán con experimentos numéricos. En uno de los experimentos, se generarán campos aleatorios de la conductividad hidráulica condicionados por historias de cargas hidráulicas 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 espera que las herramientas propuestas complementen al experimentado criterio ingenieril durante la tarea de identificación de trayectorias preferenciales de filtración en presas de tierra y ayuden al ingeniero en la toma de decisiones en relación con la seguridad de dicho tipo de presas.

Organización del escrito

La presentación de esta investigación se ha organizado en cinco capítulos. En el capítulo 1 se discute el concepto de dependencia multi-gaussiana en el contexto de los campos aleatorios espaciales. El capítulo 2 presenta el modelo de campo aleatorio adoptado para modelar la heterogeneidad en la conductividad hidráulica y propone un algoritmo de simulación de campos aleatorios. En el capítulo 3 se propone una extensión al filtro de Kalman ensamblado y se discute su desempeño en relación con la técnica original. La caracterización de la conductividad hidráulica de un acuífero heterogéneo se efectúa en el capítulo 4. En el capítulo 5 se exploran los beneficios de utilizar historias piezométricas en la caracterización de un campo no multi-gaussiano de conductividades que simula zonas continuas de conductividades hidráulicas en el núcleo interno de una presa de tierra idealizada. El escrito finaliza con las conclusiones generales y algunas recomendaciones que pretenden orientar futuras investigaciones.

CAPÍTULO 1

FLUJO DE AGUA EN MEDIOS ALEATORIOS CONTINUOS: UNA REVISIÓN DEL CONCEPTO DE DEPENDENCIA MULTI-GAUSSIANA

La conductividad hidráulica de los suelos se puede determinar por lo menos en tres direcciones principales de anisotropía por medio de cualquier experimento estandarizado bien establecido. Sin embargo, en la práctica ésta se mide con frecuencia en un número limitado de ubicaciones dentro de la región de interés tratándola como una cantidad escalar, es decir, no como un tensor. Como una regla más que una excepción, las mediciones de la conductividad hidráulica de los suelos exhiben variaciones espaciales o heterogeneidades considerables aun en formaciones razonablemente “homogéneas”. Se observa que es muy difícil describir esta variabilidad espacial en forma determinista a partir de un número reducido de mediciones. Análisis estadísticos de mediciones realizadas en diferentes sitios han mostrado sin embargo que tales variaciones espaciales presentan cierta estructura u organización.

Hoy en día se acepta ampliamente que la variabilidad espacial de la conductividad hidráulica de los suelos se puede interpretar, para fines prácticos, como una realización de un *campo aleatorio* (Dagan, 1989; Ghelar, 1993; Zhang, 2002; Rubin, 2003). Los parámetros del campo aleatorio se pueden estimar a partir de las muestras locales tomadas de la realización misma bajo la hipótesis de *ergodicidad*. La estructura de variabilidad espacial del campo aleatorio puede ser descrita por medio de *funciones de covarianza*. El concepto de *variograma* también puede ser útil en ciertas circunstancias. Éste fue introducido por Matheron (1967) quién sugirió además utilizar la técnica del *Kriging* con el propósito de interpolar espacialmente entre mediciones. Todos estos conceptos relevantes forman parte de lo que se denomina *geoestadística*.

Con el propósito de modelar la variabilidad espacial de la conductividad hidráulica de los suelos, un enfoque estocástico involucra entonces a un modelo probabilista llamado campo aleatorio. Una o más propiedades hidráulicas se pueden modelar mediante campos aleatorios. Las ecuaciones de flujo de agua en suelos dentro de este enfoque se convierten en *ecuaciones diferenciales estocásticas*. Variables dependientes del flujo tales como la carga hidráulica y velocidad de filtración serán también campos aleatorios. Tales campos pueden depender además del tiempo y se denominan campos aleatorios *espacio-temporales*. La solución de las

ecuaciones estocásticas permitirá cuantificar la incertidumbre en la respuesta de las variables dependientes del flujo. La incertidumbre surge por el hecho de que la distribución espacial de las conductividades hidráulicas dentro de la región de flujo no puede ser determinada por completo. El tema central en este enfoque no es la solución exacta de las ecuaciones estocásticas del flujo de agua sino la asignación de los valores de la conductividad hidráulica al modelo además de la cuantificación de la incertidumbre en las variables dependientes del flujo.

Se han desarrollado diferentes métodos para resolver ecuaciones estocásticas del flujo de agua en suelos. Entre ellos: 1) Métodos basados en las técnicas de perturbaciones; 2) Métodos basados en las técnicas espectrales y 3) Métodos basados en las técnicas de Monte Carlo. Un repaso de estas técnicas se puede consultar en Zhang (2002); Lu y Zhang (2004). Se han desarrollado variantes de los métodos basados en las técnicas de perturbaciones con la intención de mejorar la eficiencia computacional del método (Gainis *et al.*, 2008). Los métodos basados en las técnicas de perturbaciones consisten básicamente en evaluar los cambios en la respuesta debidos a variaciones alrededor del valor medio de la conductividad hidráulica. En un enfoque espectral el campo aleatorio de la conductividad hidráulica se representa en un espacio de funciones por medio de un desarrollo en serie de *Karhunen-Loeve* y el campo de la respuesta por medio de un desarrollo de lo que se denomina *caos polinomial*.

En las técnicas de simulación o de Monte Carlo se simula primero una realización particular del campo aleatorio de la conductividad hidráulica y se resuelven numéricamente las ecuaciones deterministas del problema, ya sea por medio de diferencias finitas o elementos finitos. Este proceso se repite un cierto número de veces hasta que se definan tendencias estadísticas suficientemente aproximadas de la variabilidad de las variables dependientes del flujo. Esta es, por tanto, una técnica que consume bastante tiempo de cálculo a pesar de que la capacidad de los equipos modernos de cómputo continúa en aumento y a pesar de que se han propuesto métodos de reducción de varianza (Curtis, 1949; Yamasaki *et al.*, 1988; Araujo y Awruch, 1994). Con frecuencia se prefiere un esquema basado en simulaciones en lugar de otro basado en perturbaciones o en el enfoque espectral, debido a que estas últimas técnicas sólo son aproximaciones basadas en el conocimiento de los primeros momentos del campo aleatorio, lo que supone implícitamente que el campo aleatorio es multi-gaussiano y además porque tales técnicas proveen soluciones suavizadas.

La técnica de simulaciones o de Monte Carlo aplicada a la solución de problemas de flujo de agua en suelos consiste en tres etapas principales (Fig. 1): 1) Generar un número suficientemente grande de realizaciones del campo aleatorio adoptado para representar heterogeneidad; 2) Resolver las ecuaciones deterministas del flujo de agua en suelos con los valores de cada una de las realizaciones en el ensamble y 3) Cuantificar la incertidumbre en las variables dependientes del flujo a través de un análisis estadístico.

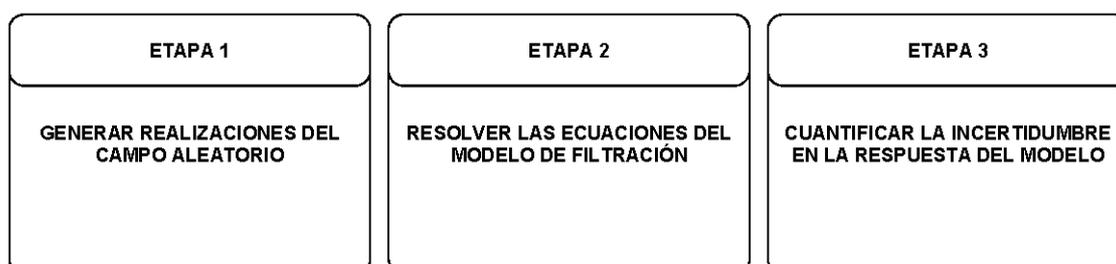


Fig. 1.1 Principales etapas involucradas en un enfoque probabilista basado en simulaciones.

En un enfoque estocástico, el campo aleatorio multi-gaussiano es el modelo más comúnmente adoptado para representar la variabilidad espacial de la conductividad hidráulica. El objetivo

principal de este capítulo es discutir la conveniencia del modelo. La discusión se basa en el estudio de sus características de dependencia. Primero se introduce formalmente el concepto de campo aleatorio, después se presenta el caso particular del campo aleatorio multi-gaussiano. Posteriormente se establece la necesidad de considerar dependencia no multi-gaussiana para representar la heterogeneidad de la conductividad hidráulica de los suelos. El capítulo finaliza con las conclusiones principales del tema.

1.1 El concepto de campo aleatorio

En esta sección se presenta una revisión de las definiciones principales involucradas en el concepto de campo aleatorio.

1.1.1 Definiciones

Una *función aleatoria* es una colección indexada de variables aleatorias. Cuando el índice es espacial multi-dimensional, una función aleatoria se denomina también *campo aleatorio*. Éste es de hecho un modelo probabilista que representa fenómenos que varían en el espacio tal como la distribución de valores de la conductividad hidráulica en una región de interés. En un campo aleatorio, a cada punto sobre el espacio se le asocia una variable aleatoria simple, por ello, globalmente es un conjunto de variables aleatorias. A cada punto en el espacio, no le corresponde sólo un valor de la propiedad sino todo un conjunto de valores. Desde una perspectiva probabilista, el valor observado en una ubicación particular es considerado una muestra de su función de probabilidad. Una configuración espacial específica de valores se denomina *realización* del campo aleatorio. Un campo aleatorio puede ser considerado entonces como un *conjunto de realizaciones (ensamble)*.

1.1.2 Estructura espacial

La estructura espacial de un campo aleatorio se puede describir por medio de *funciones de autocovarianza*. Una función de autocovarianza expresa la dependencia lineal entre variables aleatorias separadas por un vector dado. Una función de autocovarianza estandarizada llamada correlograma también se puede utilizar para los mismos propósitos. El correlograma describe la dependencia por medio de un coeficiente de correlación lineal, formalmente el *coeficiente de correlación lineal de Pearson*. En geoestadística, especialmente en las aplicaciones mineras, se utiliza con frecuencia el concepto de *variograma* (varianza del campo incremental). En campos aleatorios, la función de autocovarianza y el variograma podrían variar según diferentes direcciones. En tal caso se dice que el campo aleatorio es *anisótropo*.

En la práctica, la función de autocovarianza se estima a partir de las observaciones como parte de un proceso conocido como *análisis estructural* del campo. En este contexto la función recibe el nombre de autocovarianza *muestral* o *experimental*. En modelaciones con campos aleatorios se tiene que ajustar un modelo de autocovarianza teórico a la versión experimental. Se dice que una función de covarianza teórica es válida cuando la matriz de varianzas-covarianzas del campo aleatorio resulta ser positiva definida. Esto se debe cumplir además para cualquier combinación lineal de tales modelos válidos. Los modelos de covarianza válidos más utilizados en la práctica son (Deutsch y Journel, 1998): el exponencial¹, el esférico², el gaussiano³ y el de “efecto pepita”⁴ (nugget effect). Estos modelos se pueden especificar a través de parámetros tales como *varianza* y *escala de correlación*, excepto el modelo de efecto pepita cuya escala de

¹ $C_{exp}(\mathbf{h}) = \sigma^2 \exp(-3\mathbf{h}/a)$

² $C_{sph}(\mathbf{h}) = \sigma^2(1.5\mathbf{h}/a - 0.5(\mathbf{h}/a)^3)$ si $\mathbf{h} \leq a$ y $C_{sph}(\mathbf{h}) = \sigma^2$ si $\mathbf{h} \geq a$

³ $C_{gs}(\mathbf{h}) = \sigma^2 \exp(-3\mathbf{h}^2/a^2)$

⁴ $C_{nug}(\mathbf{h}) = \sigma^2$ si $\mathbf{h} = 0$ y $C_{nug}(\mathbf{h}) = 0$ si $\mathbf{h} > 0$

donde: σ^2 es la varianza del campo aleatorio y a es un parámetro que representa el rango de correlación.

correlación es nula. Este último se utiliza para representar la incertidumbre que afecta a las mediciones.

La escala de correlación es la distancia en la que el valor de la correlación en el correlograma alcanza el cero. Sin embargo, los modelos asintóticos exponenciales y gaussianos alcanzan el valor de correlación nula en el infinito. En estos casos se ha propuesto una definición práctica de la escala de correlación llamada algunas veces *escala efectiva* (Deutsch y Journel, 1992). Esta se define como la distancia en la que se ha perdido el 95% de la correlación.

1.1.3 Estacionariedad y ergodicidad

Se dice que un campo aleatorio es *estacionario en el sentido estricto* dentro de la región A , si sus funciones de distribución n -variadas son invariantes ante cualquier traslación en tal espacio para cualquier valor de n . Se dice que el campo aleatorio es *estacionario de orden N* cuando todas las distribuciones de probabilidad n -variadas son invariantes ante cualquier traslación en el espacio A donde $n \leq N$. Si el valor esperado es constante en todo el dominio y la función de autocovarianza sólo depende de la separación entre los puntos en lugar de sus ubicaciones reales, se dice que el campo aleatorio es *estacionario en el sentido amplio*. Un campo estacionario de segundo orden es además estacionario en el sentido amplio, pero lo opuesto no necesariamente se cumple porque los primeros momentos no pueden revelar información acerca de la distribución de probabilidad bivariada, excepto en el caso gaussiano.

En el caso especial de que las estadísticas del campo aleatorio se puedan obtener a partir de una realización única, se dice que el campo es *ergódico*. En otras palabras, la hipótesis de ergodicidad permite que los descriptores del campo aleatorio sean estimados a partir de *promedios espaciales* en lugar de *promedios de realizaciones*.

1.2 El campo aleatorio multi-gaussiano

El campo aleatorio multi-gaussiano es por mucho el modelo de función aleatoria continua más utilizado para representar heterogeneidad en la conductividad hidráulica. En esta sección se introduce su definición formal así como las razones por las que se justifica en situaciones prácticas.

1.2.1 Definición

Un campo aleatorio se denomina *multi-gaussiano* si y solo si las variables aleatorias son conjuntamente gaussianas para cualquier conjunto de n puntos. Por lo tanto, una distribución marginal gaussiana es una condición necesaria pero no suficiente para que un campo aleatorio sea multi-gaussiano. Condiciones necesarias adicionales son que las distribuciones de probabilidad multivariadas, es decir, la bivariada, trivariada, ..., n -variada sean también gaussianas.

1.2.2 Justificación práctica

En situaciones prácticas, un modelo de campo aleatorio multi-gaussiano frecuentemente se considera adecuado para modelar incertidumbre por variabilidad espacial. Comúnmente el modelo se justifica con algún tipo de comportamiento estacionario en el sentido amplio de las observaciones del campo. Pero, si nos referimos a los hechos, con frecuencia las mediciones disponibles impiden inferir momentos estadísticos superiores a los de orden dos. Por lo tanto, en muchas ocasiones un campo aleatorio multi-gaussiano no puede ser justificado en la práctica.

En el caso de las conductividades hidráulicas saturadas, se ha observado que el histograma se puede representar con suficiente aproximación por medio de una distribución de probabilidad lognormal. Esto se ha verificado en el estudio de diferentes formaciones geológicas (p. ej. Law

(1944); Davis (1969); Hoeksema y Kitanidis, (1985); Fogg (1986); Woodbury y Sudicky (1991), López-Acosta y Auvinet, 2011). Por lo tanto, se aplica comúnmente la siguiente transformación a la conductividad hidráulica saturada k_s :

$$y = \ln(k_s) \quad (1.1)$$

La distribución univariada empírica de los valores de y se podrá describir entonces con una distribución gaussiana (normal).

No obstante, una distribución marginal gaussiana no implica una distribución multi-gaussiana. Con frecuencia, este hecho se pasa por alto en estudios numéricos del flujo de agua en suelos con permeabilidad aleatoria (p. ej. Griffiths y Fenton, 1993, 1997; Fenton y Griffiths, 1997; Gui *et al.*, 2000; López-Acosta y Auvinet, 2003, 2004; Ahmed, 2009). En estos casos, el interés se limita a dos temas: 1) El método de simulación del campo aleatorio y 2) El método utilizado para resolver las ecuaciones estocásticas del flujo de agua. La conveniencia del modelo de campo aleatorio utilizado para representar la variabilidad espacial de la conductividad hidráulica con frecuencia es desatendida. Las características de la estructura espacial de un campo aleatorio multi-gaussiano se discuten a continuación.

1.3 Características de la dependencia multi-gaussiana

En esta sección se discuten las características de la estructura espacial o dependencia de los campos aleatorios multi-gaussianos. Por razones prácticas, la discusión se limita al caso bivariado. Las características que se analizan son: la estructura de correlación, la propiedad de simetría, la propiedad de entropía y la propiedad de conectividad.

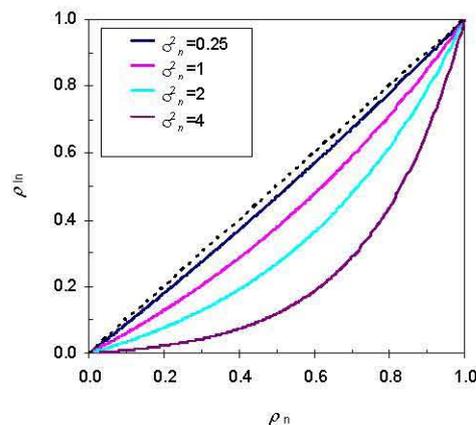


Fig. 1.2 Coeficiente de correlación lineal entre variables aleatorias log-normales como una función del coeficiente de correlación lineal entre variables aleatorias normales con distintas varianzas. Se muestra una línea interrumpida con pendiente 1:1 para referencia.

1.3.1 Correlación

El coeficiente de correlación lineal no es invariante ante transformaciones no lineales del tipo de la ec. 1.1. El coeficiente de correlación $\rho_{ln}(\cdot)$ de un par de variables aleatorias lognormales (Z_1, Z_2) está dado por (Mood and Graybill, 1963):

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

donde σ_n^2 y ρ_n son la varianza y el coeficiente de correlación lineal de las variables aleatorias normales asociadas, respectivamente. Un gráfico de la ec. 1.2 se muestra en la Fig. 1.2. Sólo se muestra la rama positiva de la correlación. Para referencia se indica además una línea interrumpida con pendiente 1:1. Se observa que el coeficiente de correlación lineal no es el mismo entre las variables de tales distribuciones excepto en cero y uno. Por lo tanto, los correlogramas de k_s y de y en la ec. 1.1 son diferentes. Las diferencias entre las correlaciones llegan a ser importantes conforme la varianza aumenta, es decir, cuando la heterogeneidad es mayor.

1.3.2 Simetría

En un campo aleatorio multi-gaussiano la función de autocovarianza define unívocamente la estructura espacial de todas las clases⁵. Para ilustrar este hecho, considérense las siguientes definiciones. A cualquier campo aleatorio $Z(\mathbf{x})$ le corresponde una serie de campos aleatorios indicadores $I(\mathbf{x}, z)$ definidos para cada valor del umbral z como: $I(\mathbf{x}, z) = 1$ si $Z(\mathbf{x}) \leq z$ o $I(\mathbf{x}, z) = 0$ en cualquier otro caso.

La descomposición del campo aleatorio $Z(\mathbf{x})$ en una serie de campos aleatorios indicadores $I(\mathbf{x}; z)$ permite asignar una estructura espacial específica a cada clase de valores $z(\mathbf{x})$. Tales campos aleatorios indicadores se pueden visualizar como campos aleatorios binarios *coexistentes*.

El conjunto de *covarianzas indicadoras* de un campo aleatorio bivariado gaussiano ha sido derivado analíticamente. La covarianza $C(\mathbf{h}; z)$ del indicador $I(\mathbf{x}; z)$, denominada *covarianza indicadora*, de un campo aleatorio bivariado gaussiano $Z(\mathbf{x})$ se relaciona con el correlograma $\rho(\mathbf{h})$ de $Z(\mathbf{x})$ por medio de (Abramovitz y Stegun, 1964; Chilès y Delfiner, 1999):

$$C(\mathbf{h}; z) = \frac{1}{2\pi} \int_0^{\rho(\mathbf{h})} \exp\left(-\frac{z^2}{1+u}\right) \frac{du}{\sqrt{1-u^2}} \quad (1.3)$$

Se pueden obtener soluciones de la ec. 1.3 para diferentes correlogramas $\rho(\mathbf{h})$, por ejemplo, para los correlogramas exponencial, esférico y gaussiano, y también en función de los cuantiles q definidos como: $q = G(z)$, donde $G(z)$ es la función de distribución acumulativa gaussiana y z es el cuantil q de Z . Estas soluciones se grafican en la Fig. 1.3 y en la Fig. 1.4, respectivamente. La escala de correlación en todos los correlogramas es unitaria. En la Fig. 1.3 los resultados se expresan en función de la covarianza indicadora estandarizada llamada *correlograma indicador*. El correlograma indicador se define como: $\rho_I(\mathbf{h}; z) = C(\mathbf{h}; z) / C(0; z)$; donde: $C(0; z) = q(1-q)$, es la varianza del campo aleatorio indicador.

La Fig. 1.4 muestra correlogramas indicadores en función de los cuantiles q asociados al umbral z para diferentes distancias. Se puede observar que la correlación es máxima en el cuantil 0.5 y que decrece simétricamente conforme el umbral se aproxima a los valores extremos de 0 y 1 para todos los correlogramas de $Z(\mathbf{x})$. Esto significa que la estructura espacial en los valores extremos tiende al “efecto pepita” (“ruido blanco”) mientras que los valores medios presentan una estructura fuertemente correlacionada. Note que la estructura espacial de las diferentes clases no se puede modificar introduciendo un correlograma diferente para $Z(\mathbf{x})$, porque esta estructura queda definida por la ley de probabilidad bivariada gaussiana.

⁵ Una clase se define como un intervalo de valores en la distribución marginal del campo.

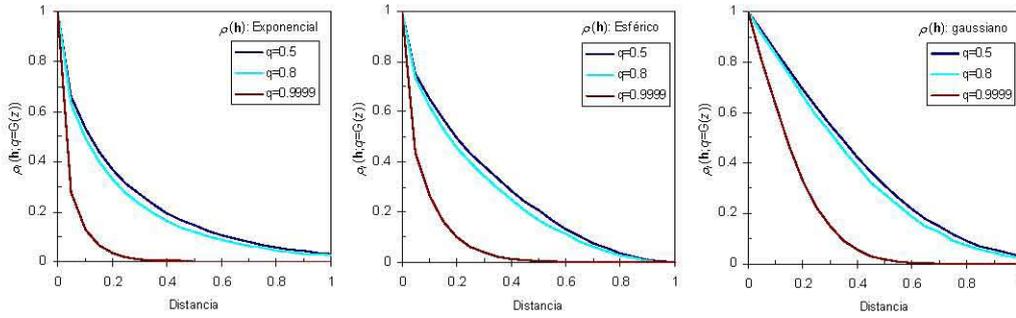


Fig. 1.3 Correlogramas indicadores de un campo aleatorio bivariado gaussiano $Z(\mathbf{x})$ en función de la distancia. Para un correlograma exponencial (izquierda), un correlograma esférico (centro) y un correlograma gaussiano (derecha).

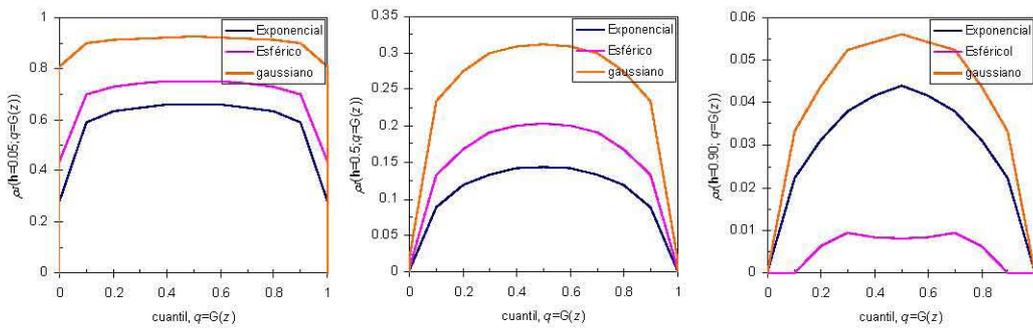


Fig. 1.4 Correlogramas indicadores de un campo aleatorio bivariado gaussiano $Z(\mathbf{x})$ en función del cuantil q asociado al umbral z . Para: $h=0.05$ (izquierda), $h=0.5$ (centro) y $h=0.9$ (derecha).

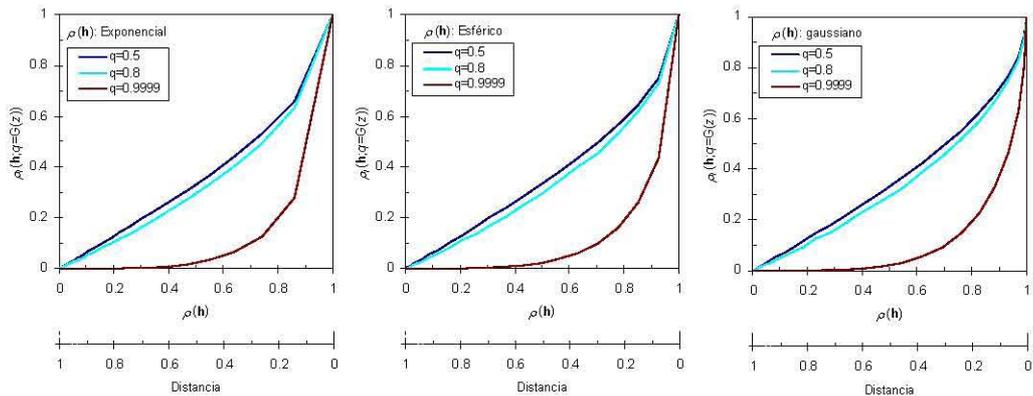


Fig. 1.5 Correlogramas indicadores de un campo aleatorio bivariado gaussiano $Z(\mathbf{x})$ en función del correlograma de $Z(\mathbf{x})$. Para los correlogramas: exponencial (izquierda), esférico (centro) y gaussiano (izquierda).

La figura 1.5 presenta valores del correlograma indicador $\rho(h; z)$ en los cuantiles 0.5, 0.8 y 0.9999, ahora en función del correlograma de $Z(\mathbf{x})$. Note que cada valor de $\rho(h)$ se puede asociar con una distancia. Por ejemplo, en $h=0$ la correlación es la más fuerte, por ello $\rho(h)=1$ mientras que esta disminuye conforme la distancia aumenta. Recordemos que los cuantiles: 0.2

y 0.0001, son los cuantiles complementarios de: 0.8 y 0.9999, respectivamente; por lo tanto, las curvas en 0.0001 y 0.9999 coinciden. Lo mismo ocurre con las curvas en los cuantiles: 0.2 y 0.8. En todas las curvas es evidente que la correlación, tanto en los cuantiles inferiores como en los superiores (0.0001 and 0.9999) es significativamente más débil que en la mediana para todas las distancias. Tal comportamiento es independiente de la función de correlación. Por lo tanto, los valores extremos (grandes y pequeños) en campos aleatorios gaussianos ocurren aleatoriamente dentro de una estructura más continua de valores medios.

1.3.3 Entropía

El término entropía fue introducido originalmente en física para estudiar las implicaciones de la cantidad de “*desorden*” en un sistema. Desde entonces se han derivado diferentes acepciones dependiendo de la disciplina. El concepto denominado *entropía de Shannon* propuesto por Shannon (1948) como parte de la *teoría de la información* ha sido adoptado en el contexto de los campos aleatorios (Christakos, 1990; Ababou *et al.*, 1992; Journel y Deutsch, 1993). En teoría de la información, la entropía fue introducida como una medida para cuantificar la “*cantidad de incertidumbre*” contenida en una distribución discreta de probabilidad, por lo tanto los términos entropía e incertidumbre son sinónimos (Jaynes, 1957). También se extendió el concepto al caso de funciones continuas pero en estos casos algunas veces se utiliza el término de *entropía diferencial* en lugar del término entropía de Shannon.

En un campo aleatorio continuo $Z(\mathbf{x})$ con función de densidad de probabilidad bivariada (PDF) $f(z_1, z_2)$, la entropía H_f se define como (Shannon, 1948):

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

En inferencias estadísticas basadas en información parcial, la teoría de la información argumenta que uno debe utilizar *a priori* la distribución de probabilidad que tenga la entropía máxima con respecto a la información que se conoce. Para obtener tal PDF se considera la solución de un problema de maximización. Según Journel y Deutsch (1993), la maximización de la ec. 1.4 equivale a elegir *a priori* la PDF “menos comprometida” con respecto a la información conocida. En otras palabras, esto equivale a seleccionar *a priori* la PDF más conservadora en vista de lo que se conoce.

Este enfoque ha sido utilizado para desarrollar un esquema bayesiano de máxima entropía para enfrentar el problema de estimación espacial (Christakos, 1990). Un enfoque de máxima entropía ha sido utilizado en el pasado para obtener algunos resultados analíticos. Por ejemplo, se ha demostrado que entre todas las PDFs bivariadas que comparten la misma esperanza y covarianza, la PDF que maximiza la entropía en la ec. 1.4 es la densidad de probabilidad bivariada gaussiana (Shannon, 1948; Johnson y Kotz, 1972; Kapur, 1989; Cover y Thomas, 1991). En el caso multivariado se cumple el resultado análogo correspondiente (Christakos, 1990). Las consecuencias prácticas de la propiedad de entropía de un campo aleatorio multigaussiano están estrechamente relacionadas con la noción de conectividad. Tales consecuencias se discuten en seguida.

1.3.4 Conectividad

En hidrología el término *conectividad* se utiliza para describir la presencia física de zonas espacialmente conectadas con valores similares de la conductividad hidráulica (Kundby y Carrera, 2005). En particular, la conectividad de los valores pequeños y grandes es importante porque controla las trayectorias y barreras del flujo. Para describir las consecuencias de la conectividad de los valores grandes en la respuesta de las variables dependientes del flujo, se utilizan conceptos relacionados tales como “*acanalamiento*” y “*trayectorias preferenciales*”. La conectividad y la continuidad que se describen por medio de funciones de autocovarianza no se

refieren al mismo concepto. La conectividad intenta caracterizar trayectorias de conductividades interconectadas, pequeñas o grandes, “tipo canales”, mientras que la continuidad describe trayectorias promediadas “tipo lentes discontinuos”.

Se han propuesto diferentes medidas para cuantificar la conectividad espacial (punto por punto) en campos aleatorios, a saber: 1) Descriptores topológicos de conjuntos geométricos aleatorios tales como el número de Euler (Mecke y Wagner, 1991; Vögel, 2002); 2) Medidas de la probabilidad de percolación tales como el umbral de percolación (Hilfer, 1992, 1997); 3) La dimensión de correlación multi-fractal (Bruderer-Weng *et al.*, 2004) y 4) Algunos indicadores intuitivos de la conectividad de flujo y transporte tales como la permeabilidad efectiva y ciertos parámetros típicos de flujo y transporte (Kundby y Carrera, 2005; Trinchero *et al.*, 2008; Vassena *et al.*, 2009).

Para ilustrar gráficamente el concepto de conectividad, en esta sección se utiliza el número de Euler $\chi(p)$. Se trata de una medida topológica de estructuras binarias que se obtienen después de segmentar un mapa heterogéneo utilizando diferentes umbrales p . En dos dimensiones, su definición matemática es (Vögel, 2002):

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

Se trata de un escalar con un valor dado por el número total de objetos aislados N en el umbral p de una imagen binaria, menos el número total de agujeros C en dichos objetos. De esta manera, $\chi(p)$ tiene valores positivos en p para estructuras desconectadas ($N > C$) y se vuelve negativo ($N < C$) para objetos que están más intensamente conectados.

La Fig. 1.6(a) muestra una realización típica de un campo aleatorio multi-gaussiano con función de autocovarianza exponencial isotrópica y escala de correlación aproximadamente igual a 1/13 del tamaño del lado de la imagen.

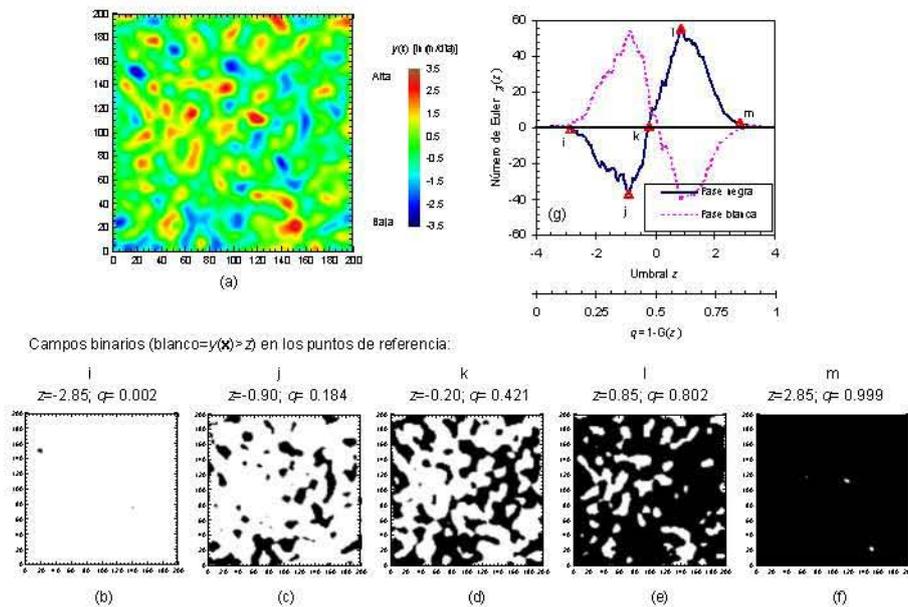


Fig. 1.6 Conectividad expresada por medio del número de Euler de una realización representativa de un campo aleatorio multi-gaussiano con función de autocovarianza gaussiana. a) Realización. b)-f) Imágenes binarias en diferentes umbrales. g) Función de Euler de la realización.

Se define una función indicadora tal que: $i(\mathbf{x};z)=1$ si $z(\mathbf{x})\leq z$ (fase en color negro) y 0 en cualquier otro caso (fase en color blanco) donde z es un umbral arbitrario definido sobre la función de distribución empírica $\hat{F}_n(z)$ del campo tal que: $\hat{F}_n(z)=q$, $q\in[0,1]$ es un cuantil que se obtiene de la distribución empírica. Separando los valores (fases) debajo y arriba del umbral q se puede construir una imagen binaria. Luego, el número de Euler se calcula en cada umbral.

Por ejemplo, para un percentil tal que $q=0.002$, los valores más bajos representados en negro en la Fig. 1.6(b) aparecen como cúmulos aislados, es decir no están conectados y el número de Euler toma valores negativos (punto (i) en la Fig. 1.6(g)). Conforme el umbral aumenta, los cúmulos se acercan pero permanecen desconectados y el número de Euler alcanza su valor mínimo (punto (j) en la Fig. 1.6(g)). Aumentos adicionales del umbral crean zonas aisladas conectadas y el número de Euler aumenta en consecuencia. Cuando éste alcanza el cero (punto (k) en la Fig. 1.6(g)) los valores grandes (fase negra) llegan a estar bien conectados (Fig. 1.6(d)), pero una zona continua de conectividades grandes que cubra dos lados opuestos de la imagen se puede detectar hasta el umbral $q=0.5$. Conforme el umbral continúa aumentando, los valores grandes llegan a estar intensamente conectados y el número de Euler alcanza su valor máximo (punto (l) en la Fig. 1.6(g)). Cuando el umbral aumenta aún más, los valores más altos (fase blanca) parecen ahora cúmulos aislados, por ejemplo en $q=0.802$ (Fig. 1.6(e)) y el número de Euler comienza a disminuir. Para $q=1$ el número total de objetos es igual a uno y el número de conexiones es cero por tanto el número de Euler tiende a la unidad. Se puede realizar un estudio analogo analizando primero la conectividad de la fase blanca (línea discontinua en la Fig. 1.6(g)).

El análisis previo de conectividad muestra que trayectorias conectadas de valores extremos pequeños o grandes no tienen lugar en campos aleatorios gaussianos. Los valores alrededor de la mediana de la distribución marginal presentan una considerable conectividad espacial. Comportamientos similares fueron detectados por otros autores que analizaron trayectorias de percolación en campos multi-gaussianos con diferentes valores de la relación lado del dominio/escala de correlación (Allard, 1994). Por lo tanto, se podría anticipar que en medios heterogéneos multi-gaussianos isotrópicos el promedio de alguna de las variables dependientes del flujo será muy próximo al valor de la misma variable determinado en un medio homogéneo con conductividad dada por el promedio de las conductividades heterogéneas.

1.4 Necesidad de una dependencia no multi-gaussiana

La evidencia de campo sugiere que las formaciones naturales de suelos rara vez son multi-gaussianas y que con frecuencia presentan estructuras conectadas de conductividades pequeñas o grandes que impactan considerablemente la respuesta del flujo. Esta sección presenta tal evidencia por separado, primero la que se basa en casos de estudio reales y después la que se obtiene por modelación numérica. Al lector le podría resultar útil referirse a los apéndices A y B.

1.4.1 Evidencia basada en casos de estudio reales

Mediciones directas y muestras recolectadas en diferentes formaciones de suelos han mostrado que la dependencia en los percentiles inferiores de la función de distribución con frecuencia difiere de la que se observa en los percentiles superiores. A grandes escalas, por ejemplo, a la escala de las facies hidrogeológicas, los análisis comparativos entre el comportamiento real observado y el del modelo numérico sugieren la necesidad de una mejor descripción de la estructura de variabilidad espacial del campo de la conductividad hidráulica.

Por ejemplo, Journel y Alabert (1989) observaron diferentes estructuras de correlación espacial en diferentes percentiles de la distribución de permeabilidades al aire efectuadas sobre un corte vertical de arenisca de Berea. Tal comportamiento corresponde con una estructura de

correlación asimétrica entre valores pequeños y grandes. Un comportamiento similar ha sido observado con las conductividades hidráulicas. Por ejemplo, Haslauer *et al.*, (2009) reporta estructuras de dependencia no gaussianas en las conductividades hidráulicas que obtienen de muestras recolectadas en dos sitios diferentes. En uno de sus análisis ellos utilizaron las bases de datos de Sudicky (1986) y de Woodbury y Sudicky (1991) que contienen mediciones de conductividades hidráulicas del acuífero de Borden.

Además, Fogg (1986) realizó un estudio numérico tridimensional de la conductividad hidráulica de un acuífero real que está conformado por secuencias complejas gruesas de arena, limo y arcilla. Su modelo conceptual se basó en la interpretación de información geológica e hidrológica. Concluyó que el flujo en el acuífero estaba controlado principalmente por la continuidad e interconexión de la arena, más que por los valores locales de la conductividad hidráulica. Otros estudios realizados con diferentes modelos para representar heterogeneidades en distintas formaciones naturales de suelos coinciden con las conclusiones de Fogg (1986). Un consenso general es que el comportamiento a gran escala de las formaciones naturales de suelo está controlado por la manera en la que las zonas más y menos conductivas están interconectadas espacialmente en la región de flujo (Bradbury y Muldoon, 1990; Poeter y Townsend, 1994; LaBolle y Fogg, 2001; Frind *et al.*, 2002; Teles *et al.*, 2004; Zappa *et al.*, 2006; López-Acosta, 2010).

1.4.2 Evidencia basada en modelos numéricos

La necesidad de describir la variabilidad espacial de la conductividad hidráulica por medio de campos aleatorios no multi-gaussianos también ha sido establecida a través de modelos numéricos. Varios estudios han mostrado que suponer aplicable el modelo multi-gaussiano no necesariamente es una hipótesis conservadora. La presencia de zonas con valores extremos espacialmente conectados (ya sean grandes o pequeños) es la característica con mayores implicaciones en los campos de la conductividad hidráulica. Muchos de los estudios han analizado la respuesta de las variables dependientes del flujo en campos aleatorios no multi-gaussianos para compararla con la respuesta multi-gaussiana. El propósito es comparar la respuesta del flujo de agua sólo en términos de las diferencias en las características de la dependencia espacial.

Por ejemplo, Sanchez-Vila *et al.*, (1996) analizaron flujo de agua en suelos saturados en régimen establecido utilizando campos aleatorios simulados del logaritmo natural de la transmisividad con diferentes dependencias no multi-gaussianas pero compartiendo distribuciones marginales normales. El estudio se realizó en secciones bidimensionales bajo condiciones de frontera confinadas por medio de simulaciones con el método de Monte Carlo. Ellos mostraron que aquellas estructuras que favorecen la conectividad de los valores grandes producen sistemáticamente valores de la conductividad efectiva mayores que la media geométrica. Las diferencias fueron cada vez mayores conforme aumenta la varianza. Para las mismas condiciones de flujo y metodología de investigación, Vögel (2002) observó valores de la conductividad hidráulica efectiva 71.5 veces mayores que la media geométrica en campos no multi-gaussianos que exhiben conectividades más significativas de los valores grandes de la conductividad.

En otro experimento numérico, Zinn y Harvey (2003) estudiaron flujo de agua y transporte en campos bidimensionales. Compararon la respuesta de las variables dependientes del flujo de agua en campos aleatorios multi-gaussianos y no multi-gaussianos de la conductividad hidráulica con funciones de autocovarianza isotrópicas casi idénticas, a través de simulaciones con el método de Monte Carlo. Las realizaciones de los campos no multi-gaussianos exhibieron dos características particulares: a) Inclusiones de valores pequeños dentro de una estructura de valores grandes conectados y b) Inclusiones de valores grandes dentro de una estructura de valores pequeños conectados. En lo que se refiere al flujo de agua, los campos con estructuras de valores grandes conectados mostraron una conductividad efectiva mayor que la media

geométrica así como grandes variaciones en la velocidad de descarga. En los campos con valores pequeños conectados, la conductividad efectiva fue inferior a la media geométrica y la varianza de las velocidades fue pequeña. En ambos tipos de campos de conductividades, las diferencias en las respuestas del flujo de agua en relación con el caso gaussiano resultaron ser mayores conforme la varianza del campo aumenta.

Nowak *et al.*, (2008) determinaron funciones de densidad de probabilidad de variables dependientes del flujo de agua tales como carga hidráulica y descarga específica. Resolvieron el flujo de agua establecido con condiciones de frontera confinadas en dominios tridimensionales. Tomaron en cuenta la heterogeneidad en la conductividad hidráulica por medio de un modelo de campo aleatorio. El estudio se realizó con base en el esquema comparativo de Zinn y Harvey (2003) utilizando un gran número de simulaciones en el método de Monte Carlo pero en su estudio los campos aleatorios multi-gaussianos y no multi-gaussianos de la conductividad hidráulica compartieron ahora distribuciones marginales idénticas y funciones de autocovarianza anisótropas. Ellos concluyen que, en términos generales, las zonas conectadas altamente permeables parecen afectar desfavorablemente el ajuste de las curvas que se obtiene en el caso multi-gaussiano. Encontraron que en el caso de conductividades pequeñas conectadas el promedio del gasto de descarga total no es significativamente diferente del que se obtiene en campos multi-gaussianos; además en este caso las curvas ajustadas son más parecidas a las del caso multi-gaussiano.

Adicionalmente, Journel y Deutsch (1993) analizaron distribuciones de frecuencias de conductividades hidráulicas efectivas y de tiempos para conseguir el 90% de saturación de agua en un experimento numérico. El estudio se realizó en secciones bidimensionales con condiciones de frontera confinadas que incluyeron fuentes y sumideros. De un campo de referencia sintético se obtuvieron el histograma, la función de autocovarianza y siete funciones de covarianzas indicadoras. Posteriormente, la información fue utilizada para caracterizar tres campos aleatorios con diferentes dependencias espaciales. La respuesta que se obtuvo del campo sintético fue comparada con la respuesta media de los diferentes campos aleatorios. Los valores medios de las distribuciones de la respuesta de los tres campos aleatorios se desviaron considerablemente del valor de referencia. Más importante aún, el modelo multi-gaussiano proporcionó un intervalo de incertidumbre sumamente optimista de la predicción sugiriendo así un mal entendido sentido de seguridad.

Comentario:

- También se puede debatir la representatividad del modelo multi-gaussiano para representar la heterogeneidad de la conductividad hidráulica en suelos estabilizados. Por ejemplo, la presencia de “bandas” más húmedas que el valor medio deseado dentro de estratos más secos favorece la creación de zonas continuas con conductividades hidráulicas elevadas. Tales zonas continuas, altamente conductivas, no sólo pueden alterar el comportamiento de la estructura ante el flujo de agua, además si se combinan con ciertos factores, también la seguridad de la estructura ante una eventual erosión interna. Por años, este fenómeno ha sido la principal causa de falla en este tipo de estructuras (ICOLD, 1995; 1997). El modelo multi-gaussiano parece por tanto no ser adecuado para evaluar la filtración aún en este tipo de estructuras aparentemente “homogéneas”.

1.6 Conclusiones

En el estudio del flujo de agua en medios aleatorios, el campo aleatorio multi-gaussiano es un modelo muy comunmente utilizado para representar heterogeneidad en la conductividad hidráulica. La adopción del modelo se justifica en la práctica con frecuencia únicamente con

base en un histograma normal de valores logaritmicamente transformados. La distribución gaussiana multivariada después se supone “a siegas”, desatendiendo frecuentemente las características de la dependencia multi-gaussiana. En este capítulo se discutieron tales características analizando el caso bivariado.

Los campos aleatorios multi-gaussianos exhiben una estructura de dependencia simétrica. Tanto los valores grandes como los pequeños presentan exactamente la misma estructura espacial. La mayor continuidad se presenta en los valores medios (eje de simetría). Esta propiedad supone que en la naturaleza o en las estructuras de tierra hechas por el hombre las zonas con conductividades altas y bajas son creadas por igual. Por otra parte, el coeficiente de correlación de *Pearson* que se expresa por medio del correlograma, no es invariante a transformaciones no lineales en general (incluyendo funciones que involucran logaritmos).

La propiedad de máxima entropía del campo aleatorio multi-gaussiano no implica máxima incertidumbre en la respuesta de las variables dependientes del flujo. Esta conclusión se basa en el hecho de que no hay manera de mejorar la organización de los valores pequeños ni de los grandes con una función de autocovarianza individual. La naturaleza, sin embargo, presenta cierta estructura espacial u organización para ciertas clases (en el sentido estadístico) de valores. Por lo tanto, un campo aleatorio multi-gaussiano podría ignorar características espaciales realistas de los campos de conductividades tales como una mejor continuidad de los valores grandes que podrían dar lugar a una incertidumbre mayor en las variables dependientes del flujo.

En campos aleatorios multi-gaussianos no pueden ocurrir trayectorias conectadas de valores extremos pequeños o grandes. Los valores alrededor de la mediana de las distribuciones univariadas presentan la conectividad espacial más significativa. Un consenso general es, sin embargo, que el comportamiento a gran escala del flujo de agua en las formaciones naturales de suelos está controlado por la manera en que las zonas más y menos conductivas están interconectadas espacialmente en la región de flujo. En otras palabras, la presencia de trayectorias preferenciales y barreras de flujo controlan la respuesta de las variables dependientes del flujo. Por lo tanto, los campos aleatorios multi-gaussianos de la conductividad hidráulica de los suelos parecen proporcionar una representación pobre de la realidad.

CAPÍTULO 2

SIMULACIÓN DE CAMPOS ALEATORIOS NO MULTI-GAUSSIANOS POR MEDIO DE CÓPULAS

El capítulo anterior estableció que para representar convenientemente las estructuras de variabilidad espacial de los campos de conductividades hidráulicas se deberían considerar campos aleatorios con características de dependencia no multi-gaussiana. Estos tipos de campos aleatorios permitirán incorporar un mayor grado de realismo en las descripciones de la heterogeneidad de los medios porosos. En este capítulo se propone un algoritmo para simularlos.

La *simulación estocástica de campos aleatorios* es el proceso por medio del que se generan configuraciones o imágenes alternativas e independientes de la distribución espacial de un atributo con significado físico tal como la conductividad hidráulica de los suelos. Cada configuración o imagen (frecuentemente llamada *realización*) se puede utilizar en modelos numéricos de la filtración de agua para cuantificar incertidumbre con el método de Monte Carlo. Un campo aleatorio que sólo es compatible con sus parámetros descriptivos se denomina *campo aleatorio no condicional*, pero si el campo aleatorio se hace además específico a un conjunto de mediciones que dependen de la posición, éste se denomina *campo aleatorio condicional*. Los campos aleatorios condicionales por tanto permiten simular configuraciones espaciales plausibles que pretenden imitar la realidad (Deutsch y Journel, 1992).

Para simular campos aleatorios no multi-gaussianos se pueden utilizar diferentes métodos: enfoques espectrales (p. ej. Yamazaki y Shinozuka, 1988; Popescu *et al.*, 1998; Grigoriu, 1998), técnicas de optimación (p.j. Srivastaba, 1995), esquemas de anamorfosis (Journel y Huijbregts, 1978; Sánchez-Vila *et al.*, 1996), enfoques bayesianos de máxima entropía (Christakos, 1990) y métodos geoestadísticos que hacen uso de estadísticas multivariadas a través de imágenes de entrenamiento (Guardiano y Srivastaba, 1993; Strebelle, 2002; Journel y Zhang, 2006). Las características de las estructuras de dependencia que se logran con todos estos métodos no se pueden discutir aquí. Alternativamente, se presenta una discusión breve de las características principales de algunos modelos de campos aleatorios no multi-gaussianos y de los métodos utilizados para simularlos. Entre los métodos más populares podemos encontrar: 1) Métodos de expansión en series (p. ej. Sakamoto y Ghanem, 2002a; Sakamoto y Ghanem, 2002b); 2)

Métodos de descomposición en indicadores múltiples (Journel, 1983; Journel y Alabert, 1989) y 3) Métodos de representaciones iso-factoriales (p. ej. Chilès y Delfiner, 1999; Emery, 2002).

Los métodos de expansión en series para campos aleatorios no multi-gaussianos se han vuelto muy populares recientemente debido a su eficiencia de cómputo. En estos métodos el campo aleatorio se representa en la forma de una expansión llamada *caos polinomial*¹ que puede incorporarse fácilmente a un *enfoque espectral de elementos finitos* (Ghanem y Spanos, 1991; Pineda-Contreras y Auvinet, 2013). En el caso no condicional la función aleatoria se construye haciendo uso de un caos polinomial para ajustar cualquier tipo de distribuciones marginales mientras que una expansión en serie de *Karhunen-Loève* se utiliza para aproximar la función de autocovarianza del campo. En el caso condicional la función de autocovarianza se obtiene primero por medio de una estimación por *Kriging simple*, después los *Eigen-valores* y *Eigen-funciones* se obtienen resolviendo una integral de *Fredholm* en forma numérica.

Se pueden encontrar por lo menos tres desventajas en este enfoque. 1) La expansión en serie de Karhunen-Loève solo aproxima la función de autocovarianza del campo. 2) Más importante aún, no se tiene control de las características de la dependencia no gaussiana. En efecto, un campo aleatorio no multi-gaussiano no se puede describir por completo a través de sus primeros momentos. 3) Debido a que la función de autocovarianza del campo se utiliza como descriptor único de la variabilidad espacial, no se puede incorporar información específica respecto a la continuidad de los valores en las colas de las distribuciones marginales. Por lo tanto, este método parece no ser conveniente para simular campos aleatorios de la conductividad hidráulica en los que características estructurales bien definidas tales como trayectorias preferenciales y barreras de flujo, controlan la incertidumbre en las variables dependientes del flujo de agua en suelos.

Los métodos de descomposición en indicadores múltiples, por su parte, hacen uso de la propiedad de descomposición de la función de autocovarianza en sus indicadores². De esta manera, se puede incorporar información específica respecto a la continuidad en diferentes percentiles. En la práctica tal información sólo se especifica por medio de un conjunto de covarianzas indicadoras. El campo aleatorio condicional se genera de la siguiente manera: en cualquier punto no observado del campo, se obtiene una estimación del valor de la función de distribución acumulativa condicional en un percentil especificado por medio de una técnica de interpolación lineal que se aplica a variables aleatorias binarias (cero o uno), las cuales se obtienen por transformación de las variables aleatorias originales. La técnica de interpolación se conoce como *Kriging indicador* (Deutsch y Journel, 1992). Repitiendo el proceso para varios percentiles se obtiene una versión discreta de la función de distribución acumulativa condicional en cada ubicación de interés. Por lo tanto, el método proporciona una solución completa al problema de estimación. Después de transformar las variables aleatorias originales en variables aleatorias binarias, el método se vuelve *no paramétrico* por lo que es capaz de tratar con cualquier tipo de distribuciones marginales.

Este tipo de campo aleatorio es el modelo no multi-gaussiano más utilizado en geoestadística sin embargo adolece varias deficiencias. 1) Las covarianzas indicadoras no se pueden modelar en forma independiente una de la otra (Journel y Posa, 1990). Por lo tanto, la amplia flexibilidad inicial aparente del modelo se ve severamente reducida debido a que la estructura espacial en ciertos percentiles sólo se aproxima en forma burda. 2) La monoticidad de la función de distribución acumulativa local no está asegurada por lo que ésta se corrige artificialmente para satisfacer ciertas relaciones de orden (Deutsch y Journel, 1992). Consecuentemente, la función de distribución acumulativa condicional local en cada ubicación de interés sólo es aproximada.

¹ Un polinomio en el que sus variables son variables aleatorias gaussianas no correlacionadas.

² Covarianzas y covarianzas cruzadas.

El formalismo indicador se implementa con frecuencia a través de un esquema de simulación secuencial en donde recibe el nombre de *simulación secuencial indicadora* (Deutsch y Journel, 1992). Bajo este esquema el formalismo sufre deficiencias adicionales: 1) La reproducción de la función de autocovarianza del campo no está garantizada en el caso general, sólo la de las covarianzas indicadoras, si y solo si, no ocurren violaciones a las relaciones de orden. 2) Las distribuciones multivariadas de las realizaciones tienen que considerarse indefinidas ya que dependen de factores tales como el número total de nodos simulados así como del número y ubicación de las muestras (Emery, 2004, 2005), es decir, las realizaciones no provienen de un modelo de campo aleatorio especificado por lo que sus estadísticas multivariadas dependen de factores de implementación. Por lo tanto, representaciones incompletas de la variabilidad espacial bivariada así como estadísticas multivariadas no definidas proporcionarían configuraciones burdas de las estructuras de variabilidad espacial en modelos que tienen el propósito de modelar la filtración de agua en suelos.

En lugar de sólo aproximar la dependencia bivariada del campo por medio de indicadores múltiples, los métodos de representaciones iso-factoriales modelan la ley de probabilidad bivariada completa a través de su distribución iso-factorial bivariada³. Existen varias familias de distribuciones iso-factoriales, cada una de ellas incluye una serie de modelos capaces de representar diferentes características de la dependencia. El modelo *laguerriano* por ejemplo permite representar estructuras con dependencia asimétrica. Tal modelo es una representación iso-factorial de la función de distribución bivariada tipo gamma (Chilès y Delfiner, 1999) cuyos factores están dados por los polinomios normalizados de *Laguerre* de orden α (Chilès y Delfiner, 1999), donde el escalar α es el denominado “factor de forma” de la distribución estandarizada univariada gamma. En este modelo los factores de las covarianzas se vuelven aleatorios y se escriben en función del correlograma y de un escalar positivo β . Aunque se utiliza una función de autocovarianza individual para especificar las covarianzas de todos los factores de la ley bivariada, la asimetría en la estructura de dependencia está controlada por los parámetros escalares α y β .

En una representación iso-factorial el modelo de campo aleatorio condicional se construye de la siguiente manera: primeramente se requiere de la anamorfosis para transformar las variables observadas en variables que describan la distribución univariada del modelo iso-factorial (por ejemplo, la distribución gamma estandarizada en el caso del modelo *laguerriano*). Posteriormente, la probabilidad condicional en un cierto umbral y en cualquier ubicación de interés se obtiene por medio de la función de anamorfosis modelada a través de una expansión en una serie de factores (polinomios de *Laguerre*), en donde cada factor se puede calcular por medio de una estimación con *Kriging* simple dados los valores del mismo factor en las ubicaciones observadas. La técnica de interpolación que permite determinar la función de distribución acumulativa se conoce como *Kriging disyuntivo* (*Kriging* de un código disyuntivo de la función de anamorfosis). A través de un *Kriging* disyuntivo en diferentes umbrales se obtiene la función de distribución condicional local en la ubicación de interés.

Los modelos iso-factoriales parecen ser una alternativa inmediata a los modelos de múltiples indicadores ya que en los primeros se puede tomar en cuenta de manera consistente la distribución bivariada completa. Además, los modelos iso-factoriales ofrecen una amplia flexibilidad para incorporar diversas estructuras de variabilidad espacial. Los modelos iso-factoriales sin embargo comparten algunas deficiencias con los modelos de indicadores múltiples: 1) El *Kriging* disyuntivo no siempre proporciona estimaciones matemáticamente consistentes (Emery, 2002, 2006). Por ejemplo, la monoticidad de la función de distribución acumulativa condicional no está asegurada, ésta se tiene que corregir artificialmente. 2) Solo se modelan las distribuciones bivariadas del campo aleatorio, las distribuciones multivariadas

³ Una función de densidad de probabilidad conjunta factorizada expresada en términos de un conjunto de funciones ortogonales $L^2(R, f)$ (espacio de Hilbert de funciones que son cuadrático-integrables con respecto a la medida dada por $f(\cdot)$).

permanecen indefinidas. Por lo tanto las estadísticas multivariadas de las realizaciones también dependerán de factores de implementación cuando el campo aleatorio se simula con un enfoque secuencial (*simulación secuencial iso-factorial*) (Emery, 2002, 2004).

Para resolver el segundo inconveniente antes mencionado, Emery (2005, 2008) formuló una “distribución multivariada *chi*-cuadrada”, a partir de campos aleatorios con distribuciones bivariadas tipo gamma con factor de forma $\alpha=0.5$ que se obtienen como la suma de campos aleatorios gaussianos independientes elevados al cuadrado. La dependencia multivariada resulta ahora determinada, además de que se dispone de una representación iso-factorial para la distribución bivariada. La representación iso-factorial permite que el modelo sea parametrizado. La dependencia espacial en este modelo está controlada por el correlograma y el escalar α . Se sugiere que el campo aleatorio condicional sea simulado por medio de técnicas iterativas tales como simulación “annealing” (Deutsch y Cockerham, 1994; Deutsch y Journel, 1994) y por técnicas de muestreo de Gibbs (Geman y Geman, 1984; Casella y George, 1992) evitando así el uso de Kriging disyuntivo debido a sus frecuentes inconsistencias.

Se puede mostrar que la distribución multivariada *chi*-cuadrada de Emery (2005, 2008) es de hecho un caso particular de una distribución multivariada más flexible que se obtiene por medio de cópulas. Además, los modelos basados en cópulas están exentos de las deficiencias antes mencionadas respecto a las técnicas de interpolación del tipo Kriging. En este capítulo se presentan algunos modelos de cópulas. La exposición del tema se divide en cuatro secciones principales. La sección 1 presenta una introducción formal de las cópulas. La sección 2 introduce el concepto de cópula espacial así como el proceso de modelado espacial mediante cópulas. La sección 3 discute el tema de la simulación no condicional de campos aleatorios mientras que la sección 4 trata el tema de la simulación condicional. Se presentan ejemplos en todas las diferentes secciones para ilustrar los temas que se discuten.

2.1 Definiciones

Formalmente, una cópula es una función $C(\cdot)$ en el cubo n -dimensional, es decir: $C : [0,1]^n \rightarrow [0,1]$ que acopla una función de distribución multivariada $F_{x_1, \dots, x_n}(x_1, \dots, x_n)$ con sus funciones de distribución marginales $F_{x_1}(x_1), \dots, F_{x_n}(x_n)$. La expresión matemática de una cópula es (Sklar, 1959):

$$F_{x_1, \dots, x_n}(x_1, \dots, x_n) = C(F_{x_1}(x_1), \dots, F_{x_n}(x_n)) \quad (2.1)$$

Si todas las F_{x_1}, \dots, F_{x_n} son continuas, entonces $C(\cdot)$ es única. Además, si: $F_{x_1}^{-1}, \dots, F_{x_n}^{-1}$ son los inversos de las funciones de distribución de F_{x_1}, \dots, F_{x_n} , entonces:

$$C(u_1 = F_{x_1}(x_1), \dots, u_n = F_{x_n}(x_n)) = F_{x_1, \dots, x_n}(F_{x_1}^{-1}(u_1), \dots, F_{x_n}^{-1}(u_n)) \quad (2.2)$$

Debido a que todas las: $U_1 = F_{x_1}(x_1), \dots, U_n = F_{x_n}(x_n)$ son variables aleatorias uniformes en el intervalo $[0,1]$ (Rosenblatt, 1952), una cópula se puede interpretar entonces como una función de distribución multivariada con funciones de distribución marginales uniformes.

Se pueden derivar varias propiedades de las cópulas (Nelsen, 2006). Por ejemplo, si $C(\cdot)$ es absolutamente continua, de la definición: $f_x(x) = d / dx(F_x(x))$; se muestra que la densidad $c(\cdot)$ de la cópula es:

$$c(u_1, \dots, u_n) = \frac{f_{x_1, \dots, x_n}(F_{x_1}^{-1}(u_1), \dots, F_{x_n}^{-1}(u_n))}{\prod_{i=1}^n f_{x_i}(F_{x_i}^{-1}(u_i))} \quad (2.3)$$

donde: $f_{x_1, \dots, x_n}(\cdot)$ se refiere a la función de densidad de probabilidad multivariada correspondiente a $F_{x_1, \dots, x_n}(\cdot)$ y $f_{x_i}(\cdot)$ es la función de densidad de probabilidad marginal correspondiente a $F_{x_i}(\cdot)$ con: $i = 1, \dots, n$.

Además, de la definición: $f_{x|y}(x|y) = f_{x,y}(x,y) / f_y(y)$; se muestra que la densidad de la cópula condicional es:

$$c(u_0 | u_1 = F_x(x_1), \dots, u_n = F_x(x_n)) = \frac{f_{x_0, x_1, \dots, x_n}(x_0, x_1, \dots, x_n)}{\prod_{i=1}^n f_{x_i}(x_i)} \frac{1}{c(u_1, \dots, u_n)} \quad (2.4)$$

Las cópulas son de interés como funciones aleatorias en modelos de variabilidad espacial porque expresan la dependencia sin la influencia de sus distribuciones marginales y por el hecho de que las cópulas son invariantes a transformaciones monótonas estrictamente crecientes. Por ejemplo, si $Y(\mathbf{x}) = \ln(K_s(\mathbf{x}))$, entonces $K_s(\mathbf{x})$ y $Y(\mathbf{x})$ comparten la misma cópula. Como resultado, medidas de correlación y asimetría expresadas solo en términos de sus cópulas son también invariantes.

La inspección de la ec. 2.2 muestra que las cópulas tienen por lo menos las siguientes propiedades:

- Las cópulas contienen información valiosa del tipo de dependencia que existe entre variables aleatorias.
- Las cópulas expresan dependencia entre variables aleatorias en la manera más pura o elemental, sin la influencia del tipo de distribuciones marginales.
- Las cópulas pueden ser construidas a partir de cualquier tipo de distribuciones marginales continuas.
- Las cópulas sólo dependen de la jerarquía de las variables que no cambia en transformaciones monótonas estrictamente crecientes, tales como las que involucran logaritmos.

Para ilustrar estas propiedades, considérese el siguiente ejemplo. La Fig. 2.1(a) muestra el diagrama de dispersión de un par de vectores aleatorios gaussianos correlacionados (X_1, X_2) con coeficiente de correlación de *Pearson* $\rho = 0.65$. Las funciones de distribución de cada vector se muestran a la izquierda (Fig. 2.1b) y debajo (Fig. 2.1(c)) de la Fig. 2.1(a), respectivamente. Después de extraer los valores de las probabilidades de cada función de distribución, se obtienen las funciones de distribución uniformes de cada vector aleatorio que se muestran en las Figs. 2.1(d),(e). Cada par $(u_1, u_2) \in [0, 1]$ en tales funciones uniformes de los vectores (U_1, U_2) , se grafica sobre el cuadrado unitario $[0, 1]^2$ según se ilustra en la Fig. 2.1(f). Esta gráfica muestra el conjunto de puntos de una cópula bivariada. Las distribuciones uniformes de cada vector aleatorio se pueden utilizar entonces para generar un par de vectores aleatorios (Z_1, Z_2) cada uno de ellos teniendo cualquier tipo de distribución marginal. Por ejemplo, los vectores (U_1, U_2) se utilizan en las Figs. 2.1(f),(g) para generar los vectores (Z_1, Z_2) a partir de la función de distribución *t* de *Student* y de la distribución gamma, respectivamente.

La cópula bivariada expresa la dependencia entre variables aleatorias independientemente del tipo de distribuciones marginales. En efecto, para construir la cópula bivariada se utilizó un par

de vectores aleatorios gaussianos pero podrían haberse utilizado en forma análoga cualquier otro tipo de distribuciones continuas para los vectores (X_1, X_2) en la Fig. 2.1(a). Además, bajo cualquier transformación monótona estrictamente creciente de las variables en los vectores (X_1, X_2) la cópula habría permanecido invariante debido a que la transformación $(X_1, X_2) \rightarrow (U_1, U_2)$ es en efecto monótona.

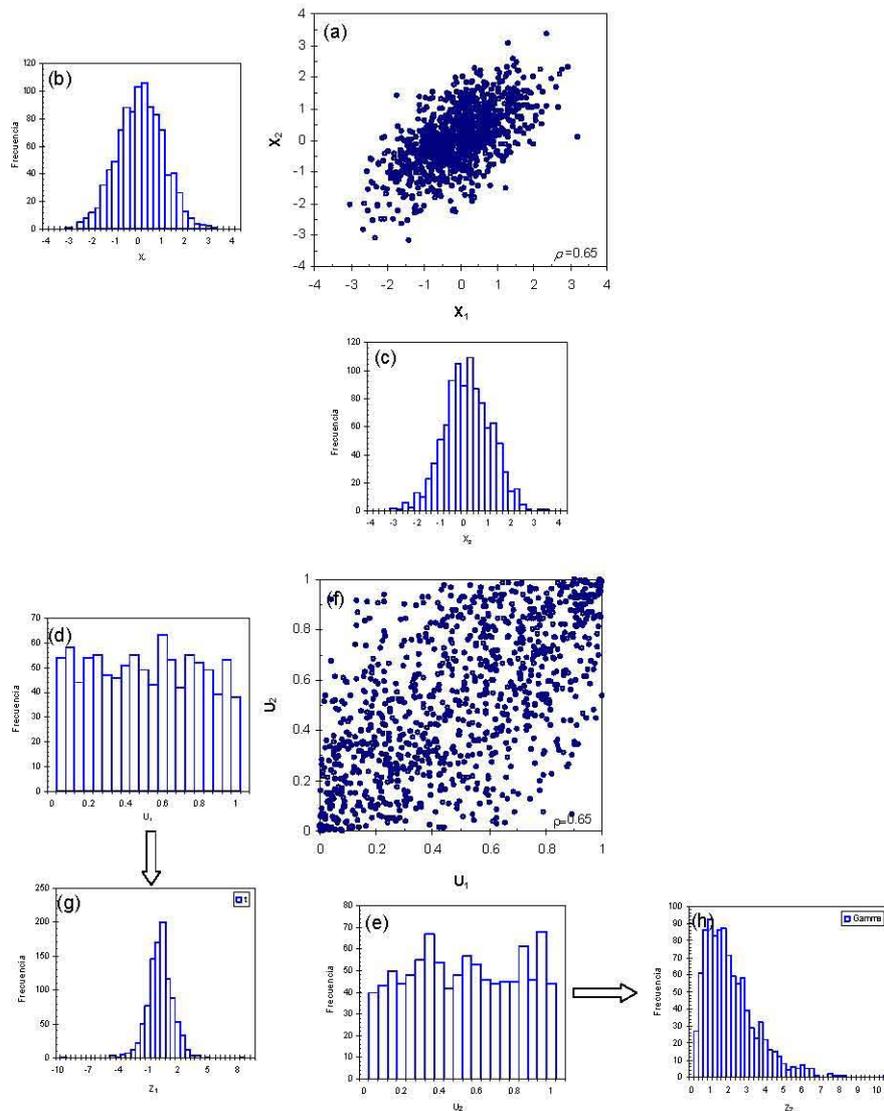


Fig. 2.1 Construcción de una cópula bivariada usando un par de vectores aleatorios gaussianos correlacionados. a) Diagrama de dispersión de los vectores gaussianos correlacionados. b) y c) Funciones de distribución de los vectores aleatorios. d) y e) Valores de la probabilidad de las funciones de distribución en b) y c), respectivamente. f) La cópula bivariada de los vectores aleatorios (X_1, X_2) . g) y h) Las funciones de distribución t de Student y Gamma que se obtienen con los valores de probabilidad en d) y e), respectivamente.

Comentario:

- El coeficiente de correlación de los vectores aleatorios con funciones de distribución t de Student y gamma del ejemplo anterior ya no es $\rho=0.65$. En efecto, el coeficiente de Pearson depende del tipo de distribuciones marginales de los vectores (Z_1, Z_2)

(Hoeffding, 1940; Lehmann, 1966). Por lo tanto, al modelar variabilidad espacial mediante cópulas las medidas invariantes de dependencia son más adecuadas, como se explicará más adelante en la sección 2.3.1.

2.2 Modelado espacial

En esta sección se presenta el concepto de cópula espacial bivariada. También se describe un método para construir cópulas espaciales bivariadas a partir de una muestra. Una vez que se determinan las cópulas espaciales de una muestra, se tiene que adoptar una cópula multivariada para modelar la variabilidad espacial. En esta sección se discute este tema. Para fines ilustrativos, se explica primero la manera en que se construye la cópula gaussiana multivariada. Después se describe en detalle el modelo de cópula multivariada adoptado en esta investigación para representar la heterogeneidad de los campos de conductividades hidráulicas.

2.2.1 Cópulas espaciales bivariadas

Considérese un campo aleatorio estrictamente estacionario $\{Z(\mathbf{x}) | \mathbf{x} \in S\}$, donde $S \in \mathcal{R}^n$, es el dominio de interés y $n=1,2$ o 3 es el espacio dimensional. Los experimentos disponibles sobre S se interpretan como una realización del campo aleatorio bajo la hipótesis de ergodicidad. Sean F_z las funciones de distribución marginales del campo aleatorio. La estacionariedad del campo asegura que las F_z sean las mismas en cada ubicación $\mathbf{x} \in S$; por tanto $F_z = F_1 = F_2 = \dots = F_n$. En forma análoga a una descripción con funciones de autocovarianza, las *cópulas espaciales bivariadas* se pueden utilizar para describir variabilidad espacial; esto es, la cópula bivariada C_s en dos ubicaciones cualesquiera separadas por un vector \mathbf{h} , se puede escribir a través del teorema de Sklar como (Bárdossy, 2006):

$$\begin{aligned} C_s(\mathbf{h}; u_1, u_2) &= P[Z(\mathbf{x}) \leq z_1, Z(\mathbf{x} + \mathbf{h}) \leq z_2] \\ &= C(P[Z(\mathbf{x}) \leq z_1], P[Z(\mathbf{x} + \mathbf{h}) \leq z_2]) \\ &= C(F_z(Z(\mathbf{x})), F_z(Z(\mathbf{x} + \mathbf{h}))) \end{aligned} \quad (2.5)$$

Por lo tanto, la cópula resulta una función del vector de separación \mathbf{h} o de la distancia $|\mathbf{h}|$. Note que u_1 y u_2 son percentiles de $Z(\mathbf{x})$ y $Z(\mathbf{x} + \mathbf{h})$, respectivamente. Para una \mathbf{h} o $|\mathbf{h}|$ dada, la cópula espacial C_s , describe entonces la dependencia espacial entre los percentiles u_1, u_2 de pares de variables aleatorias.

2.2.2 Densidades de cópulas empíricas bivariadas

Para determinar las *cópulas empíricas bivariadas* de una muestra $\{z(\mathbf{x}_1), z(\mathbf{x}_2), \dots, z(\mathbf{x}_n)\}$, se obtiene primero la función de distribución empírica $\hat{F}_n(z)$. Luego, para cualquier vector dado \mathbf{h} , el conjunto de pares de valores $S(\mathbf{h})$, de la función de distribución empírica del parámetro de interés, se obtiene por medio de (Bárdossy, 2006):

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

$S(\mathbf{h})$ es entonces un conjunto de puntos en el cuadrado unitario $[0,1]^2$. Note que $S(\mathbf{h})$ es por definición simétrica con respecto al eje principal $u_1 = u_2$ del cuadrado unitario; esto es: si $(u_1, u_2) \in S(\mathbf{h})$, entonces $(u_2, u_1) \in S(\mathbf{h})$.

La densidad de la cópula empírica bivariada \hat{c} de $S(\mathbf{h})$ se puede estimar por medio de (p. ej. Nelsen, 2006; Bárdossy, 2006):

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

donde η_{ij} , es la frecuencia empírica correspondiente a los valores en una malla regular ($m \times m$) con coordenadas (i,j) , donde $i,j=1,\dots,m$. Para un par dado (i,j) , la frecuencia empírica η_{ij} es igual a la cardinalidad (numero de elementos) del siguiente conjunto:

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

con $|\cdot|$ indicando la cardinalidad del conjunto. De esta manera se puede estimar la densidad de la cópula empírica bivariada $\hat{c}(\mathbf{h}, u_1, u_2)$ para un vector dado \mathbf{h} .

Para ilustrar estos conceptos, considérese el siguiente ejemplo. La Fig. 2.2 muestra tres realizaciones de tres campos aleatorios diferentes que comparten distribuciones univariadas gaussianas. El histograma y las funciones de correlación (correlogramas) de los pares en todas direcciones de cada realización se muestran en la parte alta de la Fig. 2.2. Las imágenes (a) y (b) son realizaciones de un campo aleatorio multi-gaussiano con funciones de autocovarianza isotropas exponencial y gaussiana, respectivamente. La escala de correlación en ambas funciones es $\alpha=60$ unidades. La imagen (c) es una realización de un campo aleatorio generado a base de transformaciones introducido por Vögel, 2002 (véase también Zinn y Harvey, 2003; Knudby y Carrera, 2005; Nowak *et al.*, 2008). Tales transformaciones imponen una dependencia no multi-gaussiana en las realizaciones tal que la continuidad en los valores grandes o pequeños es mayor. La función de autocovarianza del campo aleatorio en este caso es también exponencial e isotropa con escala de correlación $\alpha=60$ unidades.

La estructura de variabilidad espacial de cada realización se examina a través de la evaluación de los gráficos de las densidades empíricas de sus cópulas bivariadas a diferentes distancias. Sólo se consideran vectores horizontales. Los gráficos se muestran por columnas debajo de cada realización en la Fig. 2.2.

Las densidades de las cópulas empíricas de la realización no multi-gaussiana no son simétricas (Fig. 2.2(c)). Estas muestran por el contrario una asociación más fuerte en los percentiles inferiores que en los superiores. Según se puede observar al inspeccionar la realización, tal asimetría impone una continuidad mayor en los valores grandes. Es importante mencionar que aunque la inspección visual de la realización (c) sugiere cierta conectividad de los valores grandes, no se puede decir nada de la conectividad al examinar las densidades de la cópula. Intuitivamente, la conectividad se puede asociar a la amplia dispersión de los valores de densidades en los percentiles superiores. Sin embargo, en su estado actual, las cópulas no proporcionan información alguna con respecto a la conectividad. Alternativamente, se pueden utilizar funciones de conectividad para este propósito (sección 2.4.2).

El ejemplo anterior permite verificar las siguientes ventajas de las densidades de las cópulas bivariadas sobre los variogramas indicadores para examinar dependencia entre variables aleatorias (Bárdossy, 2006; Bárdossy y Li, 2008):

- 1) Las densidades de las cópulas favorecen el manejo conjunto de la estructura de dependencia en todos los percentiles de las variables en un mismo gráfico, para una distancia dada.
- 2) Las diferencias en los tipos de asociación entre variables se identifican fácil y rápidamente mediante la forma de la densidad de la cópula.

- 3) Los variogramas y covarianzas indicadoras no son necesarios para examinar la dependencia en diferentes percentiles. Con las densidades de las cópulas se logra esta tarea a un menor costo de inferencia.

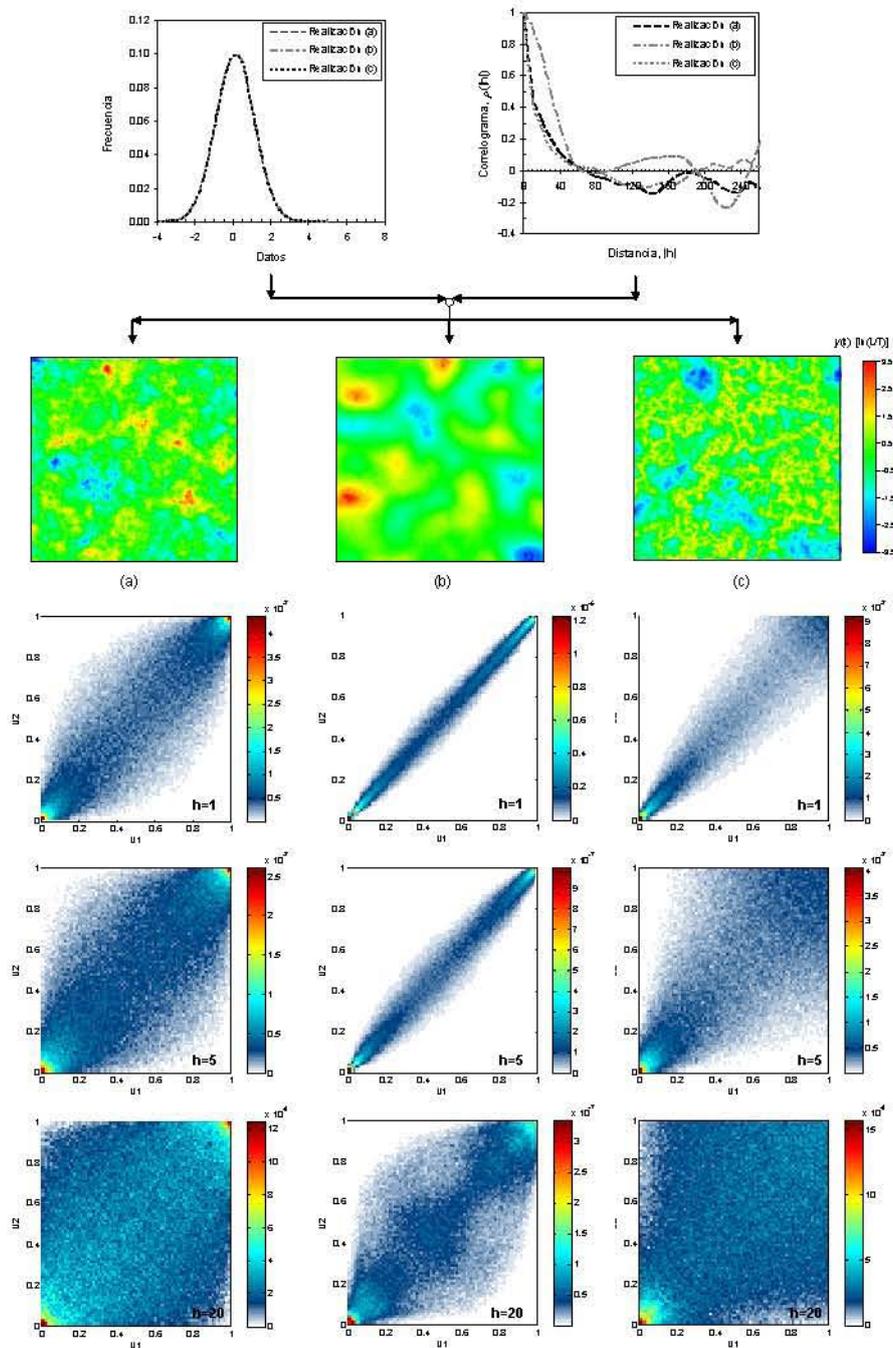


Fig. 2.2 Realizaciones representativas de diferentes campos aleatorios y sus correspondientes gráficos de densidades de cópulas empíricas. a) Realización de un campo aleatorio gaussiano con función de auto-covarianza exponencial. b) Realización de un campo aleatorio gaussiano con función de auto-covarianza gaussiana. c) Realización de un campo aleatorio no gaussiano con función de auto-covarianza exponencial. Los histogramas y correlogramas experimentales de cada realización se muestran en la parte alta de la figura.

Comentario:

- Para modelar variabilidad espacial las cópulas empíricas tienen que ser reemplazadas por cópulas teóricas, de la misma manera en que los variogramas teóricos se ajustan a los experimentales para asegurar que la matriz de varianzas-covarianzas del campo sea positivo definida (Chilès y Delfiner, 1999). Además, para describir la estructura de dependencia multivariada del campo aleatorio se requiere de una cópula multivariada teórica. La construcción de tal cópula es el tema de la siguiente sección.

2.2.3 La cópula gaussiana multivariada

La cópula gaussiana multivariada se puede formular fácilmente. En efecto, si $H(\cdot) = \Phi_{\Gamma}(\cdot)$ en la ec. 2.2 es la función de distribución multivariada gaussiana con media cero y matriz de correlación Γ y si $F_1 = \dots = F_n = \Phi$ es la distribución gaussiana estándar, entonces la cópula gaussiana multivariada $C_{\Gamma}^G(\cdot)$ está dada por:

$$C_{\Gamma}^G(u_1, \dots, u_n) = \Phi_{\Gamma}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)) \quad (2.9)$$

Obsérvese que en este caso la cópula gaussiana multivariada se parametriza por completo a través de la matriz de correlación Γ . Vale la pena mencionar que en el caso bivariado la matriz de correlación sólo contiene el coeficiente de correlación entre dos variables aleatorias.

Se dispone de una expresión analítica para la densidad de la cópula gaussiana (p. ej. Bárdossy, 2006). Con esta expresión se pueden construir fácilmente gráficos de densidades de cópulas gaussianas. Como un ejemplo de lo anterior, la Fig. 2.3 despliega gráficos de densidades de cópulas bivariadas gaussianas para diferentes coeficientes de correlación ($\rho=0.95, 0.85$ y 0.45). Obsérvese su forma típica y su propiedad de simetría.

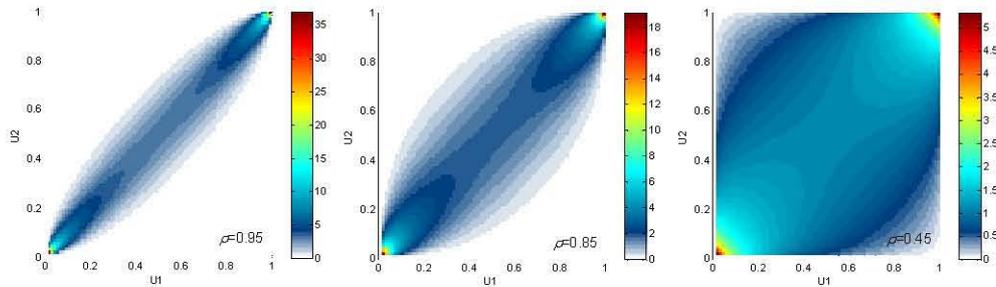


Fig. 2.3 Gráficos de densidades de cópulas teóricas gaussianas bivariadas para diferentes coeficientes de correlación.

2.2.4 La cópula multivariada V-transformada

La cópula multivariada V -transformada se obtiene por medio de una transformación no monótona de un campo aleatorio multi-gaussiano $G(\mathbf{x})$ con media cero, varianza unitaria y matriz de correlación Γ tal que (Bárdossy y Li, 2008; Li y Bardossy, 2009):

$$Y(\mathbf{x}) = \begin{cases} k(G(\mathbf{x}) - m)^{\alpha} & \text{si } G(\mathbf{x}) \geq m \\ m - G(\mathbf{x}) & \text{si } G(\mathbf{x}) < m \end{cases} \quad (2.10)$$

donde k es una constante positiva; m y α son números reales arbitrarios. Note que si $m=0$, $k=1$ y $\alpha=1$, entonces $Y(\mathbf{x})=|G(\mathbf{x})|$. En este caso, la transformación en la ec. 2.10 es equivalente a la función aleatoria basada en transformaciones utilizada para generar la realización (c) de la Fig. 2.3. Cuando $k=1$ y $\alpha=1$ la transformación en la ec. 2.10 conduce a la distribución multivariada *chi-cuadrada* no centrada. Además, el efecto de la transformación no lineal desaparece cuando $m \rightarrow \pm\infty$ y la cópula resultante converge hacia la cópula gaussiana.

El efecto de la transformación en la ec. 2.10 conduce a un campo aleatorio $Y(\mathbf{x})$ tal que sus funciones de distribución marginales son idénticas y están dadas por:

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

La cópula V -transformada multivariada con parámetros $\lambda=\{k,m,\alpha\}$ se puede escribir como:

$$C_{\lambda\Gamma}^V(u_1, \dots, u_n) = H_{\lambda\Gamma}(F_Y^{-1}(u_1), \dots, F_Y^{-1}(u_n)) \quad (2.12)$$

donde $H_{\lambda\Gamma}(\cdot)$ es la función de distribución multivariada de $Y(\mathbf{x})$ (Apéndice C).

Para $n=2$, la cópula V -transformada bivariada es:

$$C_{\lambda\Gamma}^V(u_1, u_2) = \Phi_{\Gamma}[(y_i/k)^{1/\alpha} + m, (y_j/k)^{1/\alpha} + m] - \Phi_{\Gamma}[m - y_i, (y_j/k)^{1/\alpha} + m] \\ - \Phi_{\Gamma}[(y_i/k)^{1/\alpha} + m, m - y_j] + \Phi_{\Gamma}[m - y_i, m - y_j] \quad (2.13)$$

donde $\Phi_{\Gamma}(\cdot)$ se utiliza para representar la distribución normal bivariada con marginales normales estandarizadas y matriz de correlación Γ .

Se pueden representar diferentes estructuras de dependencia asimétricas con la cópula V -transformada modificando simplemente el conjunto de parámetros λ de la cópula y la matriz de correlación Γ . La Fig. 2.4 presenta gráficos de densidades de cópulas para diferentes parámetros calculados con la ec. 2.4. Obsérvese la amplia flexibilidad de la cópula V -transformada para representar asimetría en la dependencia.

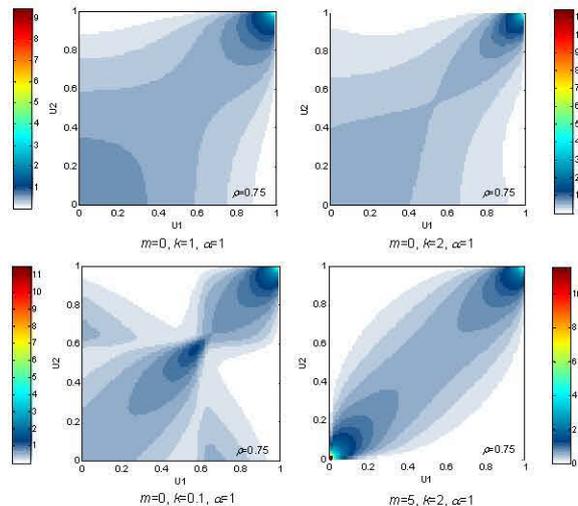


Fig. 2.4 Gráficos de densidades de cópulas teóricas V -transformadas bivariadas para diferentes parámetros m , k y α

La dependencia bivariada de campos aleatorios empleando cópulas está controlada por las cópulas bivariadas y la dependencia multivariada de tales campos estará controlada de manera explícita por la cópula multivariada.

Comentario:

- Las estructuras de dependencia de la Fig. 2.4 exhiben una asociación más fuerte en los percentiles superiores. La estructura de dependencia opuesta se obtiene tomando el complemento a la unidad de cada valor de u , es decir, calculando la densidad de la cópula $c(1-u_1, 1-u_2)$.
- A pesar de su amplia flexibilidad, la cópula V -transformada multivariada podría no ajustarse bien a un conjunto de observaciones. En tal caso se deberán considerar otros modelos de cópulas. Li y Bárdossy (2009) proponen la denominada *cópula máxima* que fue construida en forma semejante a la cópula V -transformada pero en este caso la transformación involucra los máximos de dos campos aleatorios gaussianos.

2.3 Simulación no condicional

Una realización no condicional empleando la cópula V -transformada se puede generar fácilmente aplicando simplemente la V -transformación en la ec. 2.10 a una realización multigaussiana. Sin embargo, por el efecto de la transformación, la matriz de correlación de $G(\mathbf{x})$ no se mantiene en $Y(\mathbf{x})$. Por lo tanto, se tiene que encontrar un modelo de correlación apropiado para $G(\mathbf{x})$ de tal manera que se logre la función de correlación preescrita para $Y(\mathbf{x})$ después de la transformación. El conjunto de parámetros k , m y α de la cópula controlarán la asimetría de la dependencia imponiendo asociaciones fuertes y/o débiles en las colas. La asimetría en la dependencia espacial se puede cuantificar utilizando funciones de asimetría. Los valores de la probabilidad de la distribución marginal de $Y(\mathbf{x})$ se pueden utilizar para imponer una distribución marginal de cualquier tipo al campo. Los detalles de estos temas se explican a continuación.

2.3.1 La función de correlación

Medidas de dependencia alternativas al coeficiente de correlación de *Pearson* son el coeficiente de correlación ρ de *Spearman* (Hoeffding, 1940; Quesada-Molina, 1992):

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

y el coeficiente de correlación τ de *Kendall* (Hoeffding, 1948):

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

ambos coeficientes dependen de los valores (u,v) asociados a las variables aleatorias de la cópula bivariada $C(u,v)$ y ambos son invariantes ante transformaciones monótonas tales como $Y(\mathbf{x}) = \ln(K_s(\mathbf{x}))$. En este caso, tanto $K_s(\mathbf{x})$ como $Y(\mathbf{x})$ compartirán funciones de correlación idénticas expresadas en términos de la ρ de *Spearman* o de la τ de *Kendall*. Los correlogramas de $K_s(\mathbf{x})$ y $Y(\mathbf{x})$, por otra parte, en efecto no son los mismos.

De las ecs. 2.14 y 2.15 resulta evidente que tanto el coeficiente de correlación ρ de *Spearman* como el τ de *Kendall* se pueden obtener sólo en términos de la cópula bivariada $C(u,v)$. Cada coeficiente ρ de *Pearson* de la cópula bivariada $C(u,v)$ se puede relacionar con alguna distancia en el correlograma de $G(\mathbf{x})$. Por lo tanto, calculando ρ_s para diferentes ρ , la curva $\rho_s - \rho$ se puede

utilizar para determinar la función de correlación prescrita para $Y(\mathbf{x})$, dado un modelo de correlación apropiado para $G(\mathbf{x})$. En otras palabras, una función de correlación prescrita para $Y(\mathbf{x})$ se puede obtener en términos del coeficiente ρ de *Spearman* siguiendo las instrucciones:

- 1) Ajustar una curva a la relación $\rho_s - \rho$.
- 2) Sustituir los valores de ρ_s en la curva ajustada para obtener los valores de *Pearson* en las mismas distancias de ρ_s . Esta es la curva $p-h$.
- 3) Ajustar una curva a la curva $p-h$ previa utilizando una combinación lineal de funciones de correlación derivadas de modelos válidos de funciones de autocovarianza.

Vale la pena mencionar que la función de correlación prescrita para $Y(\mathbf{x})$ en términos del coeficiente ρ de *Spearman* se obtiene como la media de un conjunto de realizaciones y que tal curva es sólo el mejor ajuste, usualmente, por mínimos cuadrados.

2.3.2 La asimetría de la dependencia

En un campo aleatorio no multi-gaussiano, la dependencia en los percentiles superiores e inferiores de las distribuciones marginales del campo puede ser diferente, es decir, puede ser asimétrica. En la literatura científica se pueden encontrar algunas medidas de asimetría (Haslauer *et al.*, 2008; Li y Bárdossy, 2009; Manner, 2010). Éstas son útiles para cuantificar la asimetría en función de la distancia, en forma análoga a las funciones de correlación, recolectando pares de valores (u,v) de la distribución marginal del campo separadas a una misma distancia. En esta sección se utiliza el concepto de correlación de excedencia para estos fines.

La correlación de excedencia se define como la correlación condicional en un nivel de excedencia c . Manner (2010) sugiere una definición de este concepto basada en el coeficiente τ de *Kendall* y obtiene las siguientes expresiones analíticas para evaluar la asimetría en un nivel de excedencia dado por la mediana $c=0.5$:

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

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

donde $\bar{\tau}$ y τ^+ se utilizan para representar la correlación de los valores por debajo y por encima de la mediana, respectivamente. La asimetría de la dependencia queda dada entonces por la diferencia $|\bar{\tau} - \tau^+|$.

Por ejemplo, la Fig. 2.5 muestra la asimetría de una cópula V -transformada teórica $C(u,v)$ con parámetros $m=0$, $k=2.5$ y $\alpha=0.25$ en función de una secuencia de valores de τ . Tal cópula se implementó en el programa proporcionado por Manner (2010) para evaluar las ecs. 2.16 y 2.17. Se puede observar que la dependencia espacial del campo transformado $Y(\mathbf{x})$ será en efecto asimétrica, con una correlación más fuerte en los valores por encima de la mediana que para los valores debajo de la mediana. Note que la asimetría opuesta se obtiene simplemente haciendo $Y(\mathbf{x})=-Y(\mathbf{x})$, y que se mantendrá la distribución marginal y la función de correlación debido a que tal transformación es monótona.

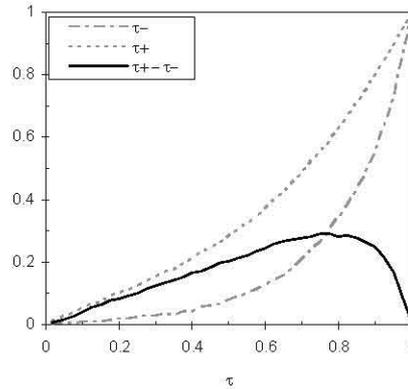


Fig. 2.5 Funciones de correlación condicional de *Kendall* para valores debajo de la mediana $t(0.5)$ y arriba de la mediana $t^+(0.5)$ para una cópula *V*-transformada con parámetros $m=0.0$, $k=2.5$ y $\alpha=0.25$.

2.3.3 La distribución marginal

Los valores (u,v) de la distribución marginal de $Y(\mathbf{x})$ se pueden utilizar para imponer una distribución marginal de cualquier tipo al campo por medio de una transformación “sin memoria” tal como: $Y(\mathbf{x})=\Phi^{-1}(u=F_Y(y(\mathbf{x})))$ donde $\Phi(\cdot)$ es la distribución gaussiana univariada. Por lo tanto, un campo aleatorio $Y(\mathbf{x})$ puede tener distribuciones marginales gaussianas, la función de correlación preescrita del campo de conductividades $K_s(\mathbf{x})$ y una dependencia no multi-gaussiana controlada por los parámetros m , v y α de la *V*-transformación. Además, haciendo: $Y'(\mathbf{x})=-Y(\mathbf{x})$ ambos campos aleatorios pueden compartir primeros momentos pero exhibir la asimetría opuesta.

2.3.4 Ejemplos ilustrativos

La distribución del logaritmo natural de la conductividad hidráulica $Y(\mathbf{x})=\ln(K_s(\mathbf{x}))$ en un dominio triangular se interpreta como una realización de un campo aleatorio no multi-gaussiano con funciones de distribución marginales gaussianas. La media de este campo es cero y la varianza es unitaria. Por tanto, la distribución de las conductividades hidráulicas saturadas tiene una media geométrica $K_G=1.0$ m/día y un coeficiente de variación $CV=131\%$. Se supone que el campo de conductividades hidráulicas es estacionario en el sentido estricto con función de correlación de *Spearman* dada por:

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

donde $\rho(\mathbf{h})$ es un correlograma esférico con escala de correlación $\alpha=20$ m (Deutsch y Journel, 1992). \mathbf{h} es el vector de separación. La ecuación anterior simplemente expresa la función de correlación de *Spearman* en términos de la función de correlación de *Pearson* (Krusal, 1954) porque, estrictamente hablando, no hay funciones de correlación en términos del coeficiente de correlación de *Spearman*. Note que las diferencias en ambos miembros de la ecuación son muy pequeñas para fines prácticos.

Se supone que la asimetría de la estructura de dependencia del campo está descrita por medio de la cópula *V*-transformada con parámetros $m=0.0$, $k=2.5$ y $\alpha=0.25$. La función de correlación de *Spearman* preescrita en la ec. 2.18 para el campo $Y(\mathbf{x})$ se obtiene con un modelo de correlación anidado que incluye las covarianzas gaussianas y esféricas. Las funciones (pesos) de contribución de varianza son: $w_1=0.90$, $w_2=0.10$ y las escalas de correlación isotropas son: $a_1=21.5$ m y $a_2=20.0$ m, respectivamente. Se generan 10 realizaciones multi-gaussianas no condicionales sobre un dominio cuadrado de 102.5m por lado, discretizado uniformemente con 41x41

elementos utilizando el programa SGSIM (Deutsch y Journel, 1992). Las realizaciones se transforman posteriormente con la cópula V -transformada.

La Fig. 2.6 muestra las funciones de correlación de *Spearman* en la dirección horizontal de las 10 realizaciones. Con fines comparativos, también se muestra la función de correlación teórica. Posteriormente se selecciona una realización representativa y sólo los valores de la sección triangular superior de la imagen se consideran en lo subsecuente. Una segunda realización se genera a partir de la primera haciendo: $Y(\mathbf{x})=-Y(\mathbf{x})$. Por lo tanto, ambas realizaciones comparten primeros momentos pero exhiben la asimetría no multi-gaussiana opuesta.

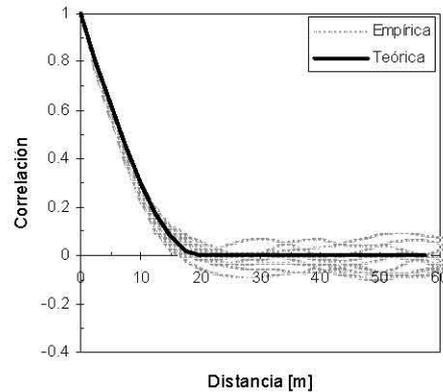


Fig. 2.6 Funciones de correlación de *Spearman* de diez realizaciones no condicionales, no multi-gaussianas (líneas interrumpidas) y la función de correlación teórica de un modelo esférico (línea sólida) (Deutsch y Journel, 1998).

Las realizaciones del campo aleatorio sobre el dominio triangular antes mencionado se muestran en la Fig. 2.7 y sus descriptores estadísticos en la Fig. 2.8. La Fig. 2.7(a) muestra valores pequeños aislados dentro de una estructura espacial continua de valores grandes. La Fig. 2.7(b) muestra la asimetría opuesta. La Fig. 2.8(b) indica una ligera anisotropía con una continuidad mayor en la dirección horizontal que en la vertical. Sin embargo, la estructura de dependencia de las realizaciones no se puede explicar por medio de tal anisotropía. Se debe enfatizar que la asimetría en la dependencia (Fig. 2.8(c)) es la que genera tal estructura, porque las zonas más continuas en las realizaciones están claramente asociadas a los valores grandes (Fig. 2.7(a)) o pequeños (Fig. 2.7(b)).

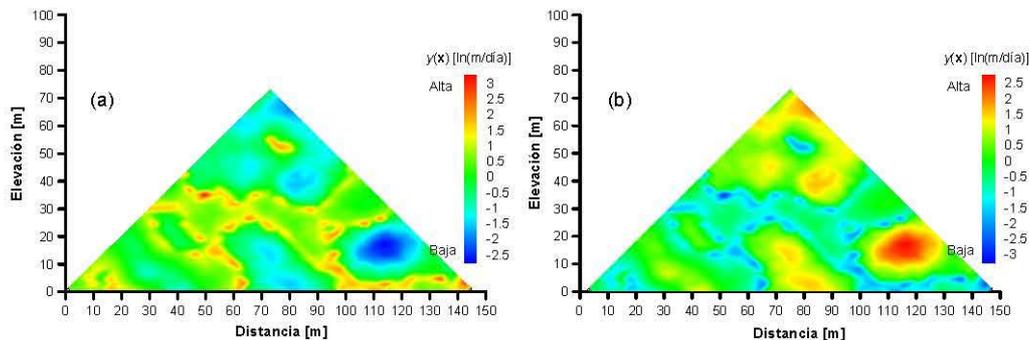


Fig. 2.7 Realizaciones representativas del campo aleatorio del logaritmo natural de la conductividad $Y(\mathbf{x})=\ln(K_s(\mathbf{x}))$. a) Valores grandes mejor estructurados. b) Valores pequeños mejor estructurados.

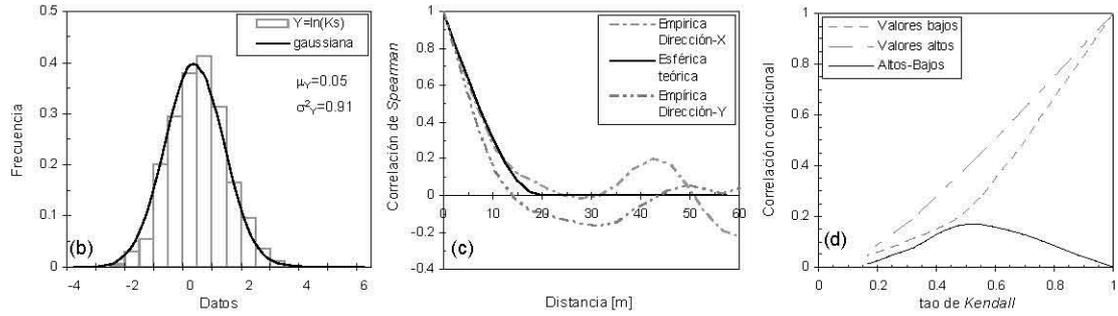


Fig. 2.8 Descriptores estadísticos de las realizaciones en la Fig. 2.7. a) Histogramas. b) Funciones de correlación experimentales en las direcciones horizontal y vertical. c) Funciones de correlación condicional de *Kendall* en la dirección horizontal (Manner, 2010).

2.4 Simulación condicional

En esta sección se propone un algoritmo para simular campos aleatorios condicionales por medio de cópulas. Para validar el simulador, se analizan las cópulas empíricas espaciales y funciones de conectividad de un conjunto de realizaciones simuladas.

2.4.1 Exposición del problema

Considérese el condicionamiento del campo aleatorio $Z(\mathbf{x})$ en N localidades i dado un conjunto α con n observaciones. La función de distribución N -variada acumulativa condicional (CCDF) de $Z(\mathbf{x})$ se puede escribir:

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

La ecuación anterior es la expresión general de una CCDF. Ella muestra que el valor simulado en la ubicación i donde $i=1, \dots, N$ se puede generar por muestreo a partir de la CCDF en etapas sucesivas que involucran una CCDF en cada ubicación i con un nivel de condicionamiento incremental (Devroye, 1986); es decir:

$$\begin{aligned} F_1(\mathbf{x}, z_1) &= P(Z(\mathbf{x}_1) \leq z(\mathbf{x}_1) | Z(\mathbf{x}_\alpha) = z(\mathbf{x}_\alpha), \alpha=1, \dots, n) \\ F_2(\mathbf{x}, z_2) &= P(Z(\mathbf{x}_2) \leq z(\mathbf{x}_2) | Z(\mathbf{x}_\alpha) = z(\mathbf{x}_\alpha), \alpha=1, \dots, n+1) \\ F_3(\mathbf{x}, z_3) &= P(Z(\mathbf{x}_3) \leq z(\mathbf{x}_3) | Z(\mathbf{x}_\alpha) = z(\mathbf{x}_\alpha), \alpha=1, \dots, n+2) \\ &\vdots \\ F_N(\mathbf{x}, z_N) &= P(Z(\mathbf{x}_N) \leq z(\mathbf{x}_N) | Z(\mathbf{x}_\alpha) = z(\mathbf{x}_\alpha), \alpha=1, \dots, n+N-1) \end{aligned} \quad (2.20)$$

Las ecuaciones 2.20 describen de hecho un proceso de simulación muy versátil conocido como método de simulación secuencial (Devroye, 1983). Este enfoque presenta las siguientes características principales:

- 1) Es independiente del tipo de función aleatoria utilizada para establecer la secuencia en la ec. Eq. 2.20.
- 2) Sólo requiere la determinación de la i -ésima CCDF univariada con $i=1, \dots, N$ en cada ubicación por simular.
- 3) Las CCDF univariadas pueden ser de cualquier tipo.

Estas propiedades permiten formular fácilmente un esquema de simulación basado en cópulas sin hipótesis adicionales respecto a la forma de la CCDF univariada o su monotonicidad (como en el caso de los formalismos de indicadores múltiples y de representaciones iso-factoriales). Además, en un modelo de cópulas tanto las distribuciones bivariadas como las multivariadas están definidas. Este no es el caso de las representaciones iso-factoriales que sólo se fundamentan en la distribución bivariada. Todos estos argumentos parecen proporcionar bases suficientes para generar campos aleatorios por medio de cópulas adoptando un enfoque de simulación secuencial.

Comentario:

- En el contexto de los métodos de simulación de campos aleatorios, algunos autores buscan que los descriptores estadísticos de las realizaciones (histograma y función de autocovarianza experimental) sean reproducidos de manera exacta. Argumentan que si se especifican ciertos descriptores estadísticos, entonces se tienen que reproducir exactamente. Sin embargo, se debe reconocer que en la práctica tales descriptores se infieren a partir de un número reducido de muestras por tanto no se pueden considerar estimadores exactos de las estadísticas de la población. Las diferencias entre los parámetros estadísticos de las realizaciones y las especificadas, llamadas *fluctuaciones ergódicas* (Deutsch y Journel, 1992), son más pequeñas conforme aumentan las dimensiones relativas del campo con respecto a la escala de correlación. Tales fluctuaciones ergódicas proporcionan de hecho una evaluación más conservadora de la incertidumbre en análisis con el método de Monte Carlo.

2.4.2 Proceso de simulación

En términos de cópulas, el enfoque de simulación secuencial se puede formular como sigue (Vázquez y Auvinet, 2014):

$$\begin{aligned}
 F_1(\mathbf{x}, z_1) &= C_{\mathbf{x}|n}(u_1 = F(Z(\mathbf{x}_1) \leq z(\mathbf{x}_1))) | u_\alpha = F(z(\mathbf{x}_\alpha)), \alpha = 1, \dots, n) \\
 F_2(\mathbf{x}, z_2) &= C_{\mathbf{x}|n}(u_2 = F(Z(\mathbf{x}_2) \leq z(\mathbf{x}_2))) | u_\alpha = F(z(\mathbf{x}_\alpha)), \alpha = 1, \dots, n+1) \\
 F_3(\mathbf{x}, z_3) &= C_{\mathbf{x}|n}(u_3 = F(Z(\mathbf{x}_3) \leq z(\mathbf{x}_3))) | u_\alpha = F(z(\mathbf{x}_\alpha)), \alpha = 1, \dots, n+2) \\
 &\vdots \\
 &\vdots \\
 F_N(\mathbf{x}, z_N) &= C_{\mathbf{x}|n}(u_N = F(Z(\mathbf{x}_N) \leq z(\mathbf{x}_N))) | u_\alpha = F(z(\mathbf{x}_\alpha)), \alpha = 1, \dots, n+N-1) \quad (2.21)
 \end{aligned}$$

donde $F(\cdot)$ es la función de distribución marginal y $C_{\mathbf{x}|n}(\cdot)$ es la cópula condicional (Apéndice C). El proceso de simulación se restringe a observaciones locales \mathbf{x}_i , para $i=1, \dots, n$ cercanas al nodo por simular debido al efecto “de pantalla” o “de sombra”. Los valores u_α incluyen tanto observaciones como valores simulados en etapas anteriores. El proceso de simulación se realiza visitando las localidades no observadas de la malla siguiendo una trayectoria aleatoria. El proceso de simulación se completa después de visitar todos los nodos de la malla.

2.4.3 Algoritmo

La simulación condicional de campos aleatorios empleando cópulas se puede efectuar implementando el siguiente algoritmo. Por razones ilustrativas se considera el modelo de cópula V -transformada multivariada.

- 1) Obtener los valores de la función de distribución acumulativa (CDF) $u_i \in [0,1]$, con $i=1, \dots, n$ empleando la función de distribución acumulativa marginal empírica del campo aleatorio $Z(\mathbf{x})$, tales que $u_i = \hat{F}(z_i)$.
- 2) Obtener los valores teóricos z_i^j , utilizando los valores u_i de la etapa anterior y la función de distribución univariada del modelo de cópula, por ejemplo, el de la cópula V -transformada haciendo $z_i^j = F_Y^{-1}(u_i)$ con $i=1, \dots, n$.
- 3) Preparar cada nodo por similar j para $j=1, \dots, N$ sobre una trayectoria aleatoria l .
- 4) Hacer $j=1$.
- 5) Elegir n observaciones cercanas al nodo por similar j y obtener la cópula condicional para un intervalo de valores de $u_i \in [0,1]$ utilizando (Apéndice C):

$$C_{\mathbf{x}|n}(u | u_1 = F_Y(z_1^j), \dots, u_n = F_Y(z_n^j)) \quad (2.22)$$

- 6) Muestrear con el método de Monte Carlo el valor u_j^* utilizando la cópula condicional:

$$u_j^* = C_{\mathbf{x}|n}^{-1}(p | u_1 = F_Y(z_1^j), \dots, u_n = F_Y(z_n^j)) \quad (2.23)$$

donde p es una variable aleatoria uniforme $p \in [0,1]$.

- 7) Asignar el valor muestreado u_j^* al correspondiente nodo j . Ahora, éste puede ser considerado en la vecindad de los nodos subsecuentes por similar.
- 8) Hacer $j=j+1$.
- 9) Si $j \leq N$, obtener $z_j^j = F_Y^{-1}(u_j^*)$ y seguir los pasos 5 a 7. De lo contrario ir al paso 10.
- 10) Obtener el valor simulado z_j^l para cada ubicación j de la realización l empleando la función de distribución marginal empírica:

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

Múltiples realizaciones independientes se obtienen visitando los nodos con distintas trayectorias aleatorias l . Obsérvese que el primer nodo por simular en cada realización se condiciona a n observaciones cercanas. Además, se puede observar en el paso 6 que el valor simulado $u_j^* \in [0,1]$. Por lo tanto, la distribución marginal del campo aleatorio simulado es la distribución uniforme. Luego, utilizando la función de distribución empírica en el paso 10 se obtiene la distribución marginal original del campo observado.

Las estrategias de búsqueda para seleccionar los valores cercanos en la vecindad del nodo por simular son idénticas a las que utiliza el programa SGSIM de la librería GSLIB (Deutsch y Journel, 1992). Para mostrar algunos resultados que proporciona el simulador, considérense los siguientes ejemplos.

2.4.3 Ejemplos ilustrativos

El objetivo es generar realizaciones de un campo aleatorio $Z(\mathbf{x})$ con funciones de distribución marginales gaussianas y estructura de correlación dada por una función de correlación de *Spearman* tal que: $\rho_s(\mathbf{h}) = 6/\pi \cdot \sin^{-1}(\rho(\mathbf{h})/2)$ siendo $\rho(\mathbf{h})$ un correlograma exponencial con escala de correlación isotrópica $a=10$ unidades. Interesa generar un conjunto de realizaciones para diferentes parámetros $\lambda = \{m, k, \alpha\}$ de la cópula V -transformada (tabla 2.1), pero cumpliendo con la distribución marginal y función de correlación de *Spearman* preescritas.

Debido a que las realizaciones del campo deben satisfacer la función de correlación de *Spearman* preescrita, la estructura de correlación Γ de la cópula V -transformada multivariada $C_{\lambda, \Gamma}^V$ se tiene que parametrizar como se discutió en la sección 2.3. La función preescrita se obtuvo en cada caso con los modelos de correlación anidados que se indican en la Tabla 2.1.

Tabla 2.1 Modelo de correlación anidado para cada caso analizado. Las contribuciones de la varianza y escalas de correlación corresponden a las funciones de auto-covarianza gaussiana, esférica y exponencial, respectivamente.

Caso	Parámetros de la cópula			Contribución de varianzas			Escalas de correlación 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

¹Las escalas de correlación son isotrópicas y están dadas en cualquier unidad de longitud consistente.

Se generaron 15 realizaciones de cada campo aleatorio indicado en la Tabla 2.1. La simulación se realizó sobre una malla cuadrada de 75 unidades por lado con nodos regularmente espaciados a cada 1 unidad. Se seleccionó la segunda realización de cada conjunto. Éstas se muestran en la Fig. 2.9(b). Los histogramas y funciones de correlación de cada realización se muestran en la parte alta de dicha figura (Fig. 2.9(a)). Se examinaron las densidades de las cópulas espaciales de las realizaciones seleccionadas en diferentes distancias pero sólo aquellas que corresponden a la distancia $|h|=1$ unidad, es decir, donde $\rho_s(1) \approx 0.74$, se muestran en la Fig. 2.9 con fines ilustrativos. Para obtener una referencia visual, a la izquierda de las densidades de las cópulas experimentales se presentan sus correspondientes densidades teóricas.

Como se puede observar en la Fig. 2.9(a), los histogramas de las realizaciones son muy cercanos a la densidad gaussiana. Esto significa que la distribución de valores en las realizaciones es muy cercana a la distribución uniforme, como se esperaba. Además, la función de correlación experimental de Spearman de cada realización sigue aproximadamente la correspondiente función teórica prescrita (Fig. 2.9(a)). Como se observa por inspección visual de las realizaciones, la continuidad de los percentiles superiores parece mayor que la continuidad de los percentiles inferiores excepto en el caso 3, es decir, cuando: $m=5$, $k=2$ y $\alpha=1$. En este caso, los percentiles superiores e inferiores aparecen como cúmulos aislados. En efecto, este caso es aproximadamente equivalente a un campo aleatorio gaussiano.

Las densidades de las cópulas empíricas bivariadas muestran claramente las características de la dependencia. En todos los casos las densidades empíricas se asemejan muy bien a sus correspondientes densidades teóricas. La asimetría que se observa en las densidades empíricas confirma lo que inicialmente sugería la inspección visual de las realizaciones; esto es, una continuidad mayor en los percentiles superiores. El método propuesto parece por tanto proporcionar resultados satisfactorios en el sentido de que las densidades de las cópulas bivariadas de las realizaciones están bastante bien aproximadas.

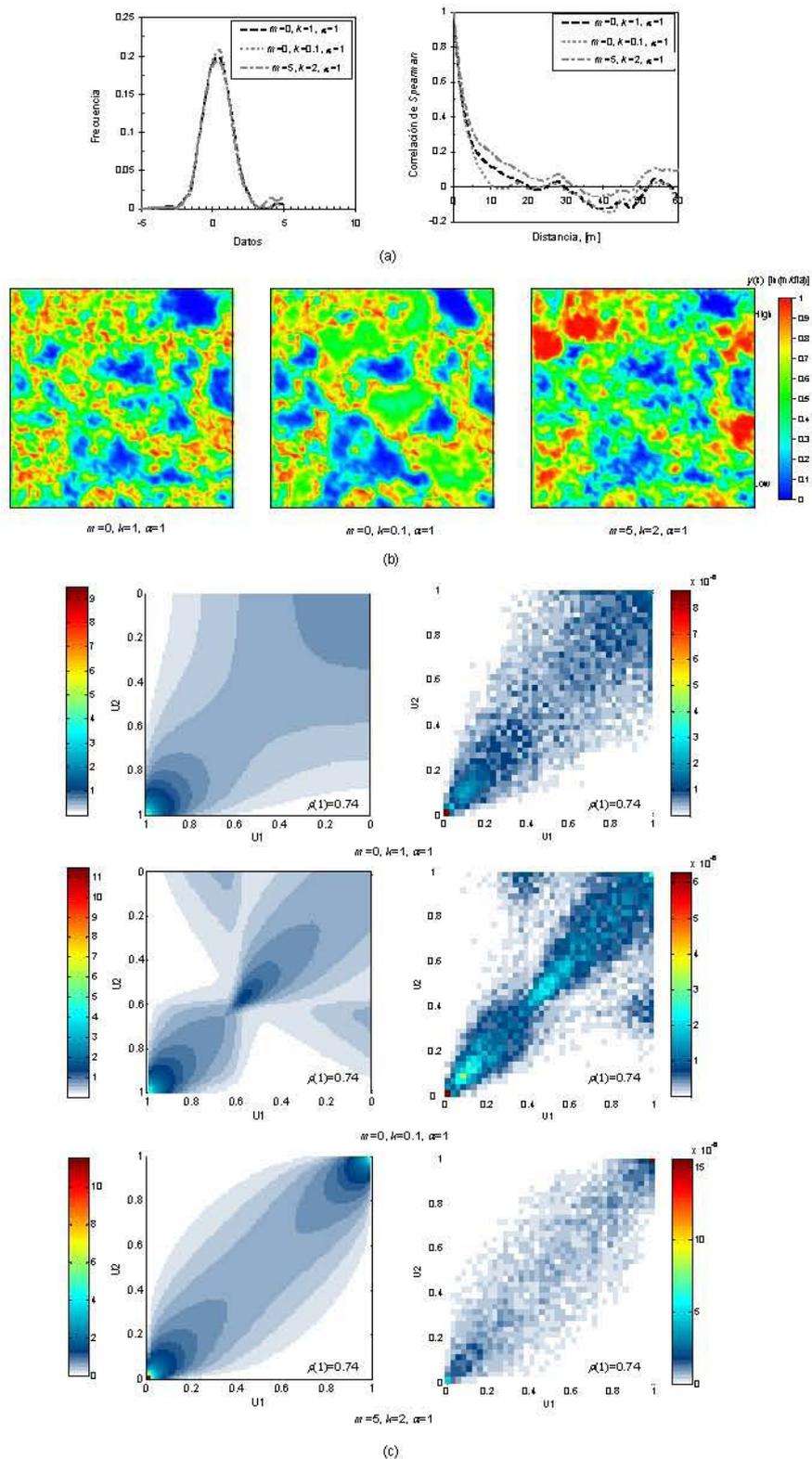


Fig. 2.9 Realizaciones representativas de campos aleatorios generadas por medio de cópulas utilizando la técnica de simulación secuencial. Todos los campos aleatorios comparten funciones de distribución gaussianas y funciones de correlación de *Spearman* idénticas sin embargo difieren en la asimetría que imponen los parámetros m , k y α de la cópula. a) Histogramas y funciones de correlación empíricas en la dirección horizontal. b) Realizaciones de los campos aleatorios con los parámetros de la cópula que se indican debajo de ellas. c) Densidades de las cópulas empíricas (derecha) y las correspondientes versiones teóricas (izquierda).

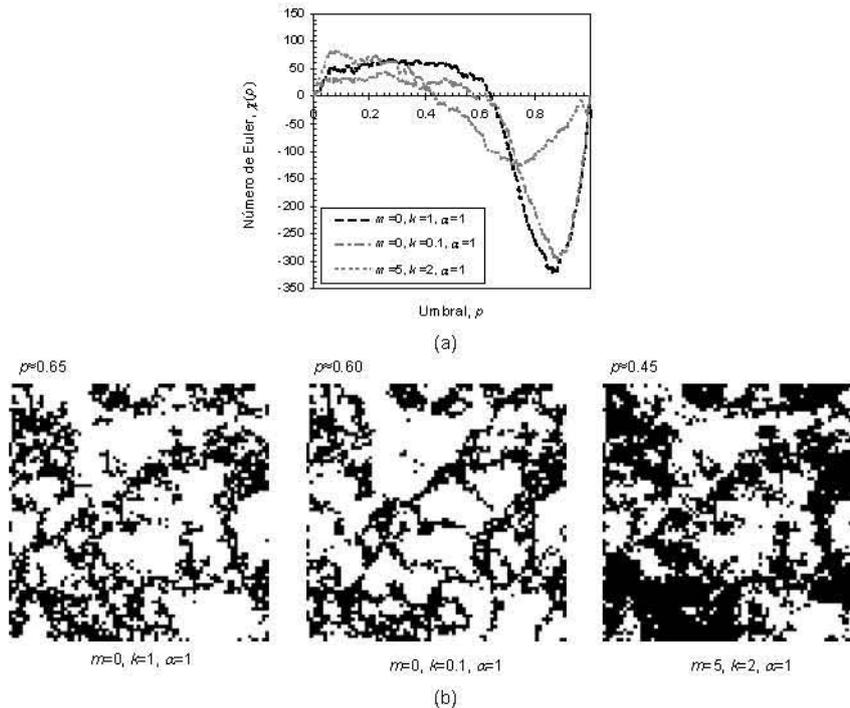


Fig. 2.10 Número de Euler como una función del percentil p y las correspondientes imágenes binarias en $\chi(p)=0$. a) Funciones de Euler de las realizaciones simuladas con la cópula V -transformada cuyos parámetros se indican debajo de cada figura. b) Imágenes binarias en $\chi(p)=0$.

Para explorar las características de conectividad de las realizaciones, se construyeron funciones de conectividad calculando el número de Euler en diferentes umbrales (Mecke y Wagner, 1991; Vögel, 2002). Los umbrales corresponden con una serie de percentiles. En cada umbral se obtiene una imagen binaria separando los valores menores (fase blanca) y mayores (fase negra) que el umbral. Después se obtiene el número de Euler en cada umbral (Fig. 2.10(a)). El conjunto de valores así obtenidos se conoce como función de conectividad. Un valor de la función igual a cero significa que el campo comienza a estar bien conectado con quizás una fase continua que recorre dos lados opuestos del campo (véanse las imágenes binarias de la Fig. 2.10(b)). El valor mínimo de la función corresponde con el umbral en el que la imagen se encuentra intensamente conectada.

Las realizaciones con dependencia asimétrica claramente están más intensamente conectadas que la realización con dependencia simétrica (Fig. 2.10(a)). Las estructuras más intensamente conectadas aparecen en los percentiles cercanos a uno. El umbral de la conectividad ($\chi(p)=0$) en estos casos ocurre en los percentiles superiores a la mediana y la fase conectada es más amplia (Fig. 2.10(b)). Por lo tanto, se puede argumentar que el algoritmo propuesto ciertamente genera realizaciones cuyas funciones de conectividad se desvían de las funciones de conectividad de las realizaciones multi-gaussianas (capítulo 1).

La mayor conectividad de las realizaciones asimétricas se puede asociar con una dispersión mayor de las densidades en los percentiles superiores de las cópulas empíricas bivariadas de la Fig. 2.9(c). Sería muy interesante formular una medida de conectividad en términos de la densidad de la cópula bivariada.

2.5 Conclusiones

A lo largo de este capítulo se introdujeron algunos conceptos básicos de las cópulas. Se hizo énfasis en conceptos tales como dependencia y asimetría. El proceso de modelación espacial

mediante cópulas se trató con cierto detalle. Se propuso un algoritmo para la simulación de campos aleatorios por medio de cópulas. Se simularon campos aleatorios con dicho algoritmo y se examinaron las densidades de las cópulas empíricas espaciales de las realizaciones. Se encontraron resultados satisfactorios en términos de la reproducción de las correspondientes densidades teóricas.

Características de la estructura de variabilidad espacial tales como dependencia y asimetría fueron descritas cuantitativamente por medio de escalares individuales que sólo dependen de las cópulas bivariadas. Expresando dichos escalares en función de la distancia se construyeron funciones de dependencia y asimetría. Estas funciones y los gráficos de densidades de cópulas bivariadas se utilizaron para examinar y visualizar la estructura de variabilidad espacial de realizaciones de campos aleatorios simulados. También se determinaron las funciones de conectividad de las realizaciones en términos del número de Euler.

El modelado de la variabilidad espacial por medio de cópulas difiere de ciertos enfoques geoestadísticos más tradicionales como por ejemplo el formalismo de indicadores múltiples en que el primero utiliza la ley de probabilidad bivariada completa y no sólo una aproximación basada en covarianzas indicadoras.

Debe reconocerse sin embargo que la simulación de campos aleatorios empleando cópulas puede resultar muy demandante computacionalmente en mallas con un gran número de nodos. En general el tiempo de cálculo necesario para simular un campo aleatorio no multi-gaussiano es considerablemente mayor al que se necesita para simular uno multi-gaussiano. El algoritmo aquí presentado se podría optimizar para reducir el tiempo de cálculo.

CAPÍTULO 3

SOLUCIÓN AL PROBLEMA INVERSO CON UN FILTRO DE KALMAN ENSAMBLADO MODIFICADO

El Filtro de Kalman Ensamblado (EnKF) es una técnica utilizada en ciencia y en ingeniería para actualizar (en un sentido bayesiano) estimaciones de variables físicas consistentemente con el fenómeno físico en estudio y también con las mediciones disponibles en ubicaciones y tiempos específicos. El método fue desarrollado originalmente por Evensen (1994) para aplicaciones de la Geofísica, pero en la última década ha sido transferido y adaptado para enfrentar problemas de estimación en diversas disciplinas donde EnKFs se acoplan a modelos físicos para actualizar estimaciones de variables de estado, de parámetros del modelo físico o ambos (Evensen, 2003; Ghanem y Ferro, 2006; Butala *et al.*, 2009). La estimación de parámetros de un modelo físico también se conoce en ingeniería como identificación de parámetros y modelación inversa.

La relación funcional entre las variables de estado y los parámetros en el enfoque EnKF se representa mediante las ecuaciones de la dinámica del sistema, que frecuentemente se escriben en la forma de alguna clase de ecuaciones diferenciales parciales estocásticas. El método EnKF se puede resumir en dos etapas principales que incluyen una etapa de pronóstico y una etapa de actualización. La etapa de pronóstico utiliza el método de Monte Carlo para aproximar la función de densidad de probabilidad (PDF) de la respuesta del sistema mediante un conjunto (ensamble) de realizaciones. En la etapa de actualización se hace una aproximación de la PDF *a posteriori* del parámetro que se va a estimar mediante el condicionamiento de cada realización *a priori* a las historias de observaciones disponibles, empleando una técnica de estimación lineal. Debido a que el EnKF es de hecho un método de Monte Carlo, es apropiado incluso para complejos sistemas no lineales, pero debido a la etapa de actualización lineal, el método solo es exacto cuando las variables aleatorias involucradas son conjuntamente gaussianas

Para aplicaciones del método EnKF en el contexto de los campos aleatorios, Bertino *et al.* (2003) han propuesto la aplicación de técnicas de transformación univariadas antes de la etapa de actualización para generar un espacio geométrico en el que, al menos, las distribuciones marginales sean gaussianas. La transformación gaussiana es de hecho un procedimiento estándar en aplicaciones geoestadísticas conocido como *transformada normal estandarizada* (Deutsch y Journel, 1996). Aun cuando investigaciones recientes en el campo del flujo de agua en medios porosos aleatorios han encontrado que esta variante del EnKF tiene un mejor

desempeño que el original y que también es más conveniente para el manejo de distribuciones no gaussianas de los parámetros (Zhu *et al.*, 2011; Shöniger *et al.*, 2012; Xu *et al.*, 2013), el método sigue siendo subóptimo y en esta alternativa se transfieren algunas deficiencias asociadas a las técnicas de estimación lineal como por ejemplo el uso de la función de covarianza como el único descriptor de una dependencia espacial evolvente (Vázquez y Auvinet, 2015).

En la etapa de actualización del enfoque EnKF aplicado a los campos aleatorios, en lugar de modelar la función de distribución acumulativa condicional de N variables (CCDF), se establece una CCDF univariada en cada una de las N ubicaciones que se van a estimar, tomando en cuenta únicamente las historias de observaciones disponibles. Esta estrategia se presta para que el esquema sea más eficiente a nivel computacional y adecuado para problemas de grandes dimensiones. Además, dicho esquema permite la incorporación de una técnica de simulación conocida como “*p-field*” (Srivastava, 1992; Froidevaux, 1993) sin alterar la eficiencia computacional del método. En la simulación “*p-field*”, se toman muestras de las CCDFs locales utilizando números aleatorios (valores p) autocorrelacionados y distribuidos uniformemente, de tal manera que los valores simulados resultantes se aproximen a las CCDF locales a lo largo del conjunto de realizaciones.

En este capítulo se muestra cómo se puede implementar fácilmente en el contexto del método EnKF una etapa de actualización basada en la técnica de simulación “*p-field*”. Para desarrollar este enfoque, primero se determinan con el método EnKF las CCDFs locales del campo aleatorio que se va a estimar y luego se generan realizaciones de estas distribuciones mediante la técnica de simulación “*p-field*”. Se espera que el esquema propuesto, que aquí se designa mediante las siglas pf-EnKF, sea más flexible que el EnKF en el sentido de que en cada etapa de actualización el primero permite, al menos, dos opciones diferentes para generar los campos *a posteriori*. Una primera alternativa consiste en utilizar como campos p las realizaciones del campo aleatorio *a priori*. En este caso, ambos enfoques son equivalentes y se espera que las diferencias entre ellos sólo se observen en las estadísticas de las realizaciones simuladas sin ninguna mejora anticipada. Sin embargo, una alternativa más interesante consiste en actualizar primero los campos *a priori* con el EnKF y luego utilizarlos como campos p en el pf-EnKF. En este caso, la estructura de correlación de los campos p estará dada por la función de autocovarianza condicional por lo que la etapa de actualización con la técnica “*p-field*” podría refinar la aproximación del EnKF introduciendo fluctuaciones en torno de las estimaciones medias. El desempeño del esquema propuesto bajo esta última opción de los campos p se evaluará con respecto a diferentes criterios cuantitativos.

Una formulación alternativa del procedimiento EnKF basada en soluciones disponibles de la teoría de estimación lineal aplicada a funciones aleatorias se utiliza para incorporar la metodología propuesta. Para ilustrar los efectos de la etapa de actualización del pf-EnKF y para comparar los resultados con los del EnKF, en la estimación de conductividades y cargas hidráulicas, se considera un problema de flujo monofásico, unidimensional, en un medio poroso continuo y aleatorio.

3.1 Ecuaciones de flujo de aguas subterráneas

En el siguiente análisis se considera el modelo dinámico que describe el flujo de un fluido en un medio poroso, unidimensional, completamente saturado, con conductividad hidráulica espacialmente variable:

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

sujeto a las condiciones iniciales y de frontera:

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

donde, H es la carga hidráulica [L] en el dominio Ω , \mathbf{x} es la coordenada espacial, ($\mathbf{x}=(x_1, x_2, x_3)$ [L], donde x_3 representa la coordenada vertical que es positiva hacia arriba), $K_s(\mathbf{x})$ es la conductividad hidráulica saturada [L/T], S_s es el almacenamiento específico [L^{-1}], h_0 representa la carga inicial y h_1 la carga prescrita en la frontera Γ_D de Dirichlet.

Aun cuando las condiciones iniciales y de frontera se consideran constantes deterministas, la conductividad hidráulica $K_s(\mathbf{x})$ se considera un campo aleatorio; por lo tanto, la ecuación 3.1 se convierte en una ecuación diferencial estocástica y la respuesta del flujo H también será un campo aleatorio. Para obtener la respuesta dinámica del modelo que satisface las ecuaciones 3.1 y 3.2 se utiliza el método de Monte Carlo. Primeramente se generan realizaciones independientes de un campo aleatorio *a priori* de conductividades y se resuelven numéricamente las ecuaciones de flujo con los valores deterministas de cada realización utilizando una versión modificada de un código de elementos finitos (Smith y Griffiths, 2004). Las estadísticas de las variables dependientes del flujo tales como las funciones de autocovarianza y covarianza cruzada entre conductividades y cargas hidráulicas se determinan con el ensamble de realizaciones utilizando formulas bien establecidas de estadística descriptiva, en cada tiempo en el que el campo aleatorio *a priori* es actualizado por medio del siguiente proceso de condicionamiento.

3.2 Simulación de campos aleatorios espacio-temporales

En esta sección se presenta una formulación alternativa del método EnKF. Adicionalmente, se muestra cómo se puede implementar fácilmente dentro del contexto del EnKF una etapa de actualización basada en la técnica de simulación “*p-field*”.

3.2.1 Definiciones

Sea $Y_t(\mathbf{x})$, la colección de n variables aleatorias continuas del logaritmo natural de la conductividad hidráulica saturada K_s es decir, $Y_t = \ln(K_s)$ referidas a las ubicaciones espaciales \mathbf{x} en la región Ω_x con $\Omega_x \subset \mathbb{R}^d$ donde $d=1,2$ o 3 y en los tiempos discretos $t \in \{0,1,2,\dots\}$. Este campo aleatorio espacio-temporal del logaritmo natural de la conductividad hidráulica con dimensión n se escribe como:

$$\{Y_t(\mathbf{x}) : \mathbf{x} \in \Omega_x, t = 0, 1, 2, \dots\} \quad (3.3)$$

Defina $H_t(\boldsymbol{\chi})$, la colección de N variables aleatorias continuas de la carga hidráulica referidas a las ubicaciones espaciales $\boldsymbol{\chi}$ en la región Ω_χ con $\Omega_\chi \subset \mathbb{R}^d$ donde $d=1,2$ o 3 y en los tiempos discretos $t \in \{0,1,2,\dots\}$. Este campo aleatorio espacio-temporal de la carga hidráulica con dimensión N se escribe por medio de:

$$\{H_t(\boldsymbol{\chi}) : \boldsymbol{\chi} \in \Omega_\chi, t = 0, 1, 2, \dots\} \quad (3.4)$$

La función de distribución acumulativa (CDF) multivariada de $H_t(\boldsymbol{\chi})$ en el tiempo particular t dependerá de las distribuciones multivariadas en todos los tiempos previos. Sin embargo, la evolución de dicha función se puede determinar suponiendo que el modelo dinámico de la ecs. 3.1-3.2 se comporta como un proceso markoviano de primer orden, es decir, $P[H_t(\boldsymbol{\chi})|H_{t-1}(\boldsymbol{\chi}), H_{t-2}(\boldsymbol{\chi}), H_{t-3}(\boldsymbol{\chi}), \dots] = P[H_t(\boldsymbol{\chi})|H_{t-1}(\boldsymbol{\chi})]$. En consecuencia, solo el pasado más reciente determina la

CDF condicional multivariada de $H_t(\boldsymbol{\chi})$ dado todo el pasado. Esta evolución simplificada de $H_t(\boldsymbol{\chi})$ está dada por:

$$H_t(\boldsymbol{\chi}) = \mathfrak{S}(H_{t-1}(\boldsymbol{\chi}), Y_{t-1}(\mathbf{x})) \quad (3.5)$$

donde $\mathfrak{S}(\cdot)$ representa las ecuaciones de la dinámica del modelo. Observe en la ec. 3.5 que para avanzar $H_t(\boldsymbol{\chi})$ hacia el primer tiempo, es decir $H_1(\boldsymbol{\chi})$, se tienen que especificar las condiciones iniciales $H_0(\boldsymbol{\chi})$, el campo aleatorio de parámetros $Y_0(\mathbf{x})$ y las condiciones de frontera.

Supóngase ahora que se toman sin error N_h observaciones de $H_t(\boldsymbol{\chi})$ en las ubicaciones $\boldsymbol{\chi}_\alpha$ donde $\alpha \in \{1, \dots, N_h\}$ y en los tiempos discretos $t \in \{0, 1, 2, \dots\}$. El conjunto de historias de observaciones disponibles se puede representar entonces por medio de:

$$\{H_t(\boldsymbol{\chi}_\alpha) = h_{\alpha,t}, \alpha = 1, 2, \dots, N_h, t = 1, 2, 3, \dots\} \quad (3.6)$$

Un campo aleatorio que se hace específico a un conjunto de observaciones directas y/o indirectas en sus ubicaciones exactas se denomina condicional. Los campos aleatorios $Y_t(\mathbf{x})$ y $H_t(\boldsymbol{\chi})$ dado el conjunto de observaciones $H_t(\boldsymbol{\chi}_\alpha)$ se escriben como:

$$\{Y_t(\mathbf{x}) | H_t(\boldsymbol{\chi}_\alpha): \mathbf{x} \in \Omega_x, \boldsymbol{\chi}_\alpha \in \Omega_\chi, \alpha = 1, 2, \dots, N_h, t = 1, 2, 3, \dots\} \quad (3.7)$$

$$\{H_t(\boldsymbol{\chi}) | H_t(\boldsymbol{\chi}_\alpha): \boldsymbol{\chi}_\alpha \in \Omega_\chi, \alpha = 1, 2, \dots, N_h, t = 1, 2, 3, \dots\} \quad (3.8)$$

El análisis de los campos aleatorios de las ecuaciones 3.7 y 3.8 requiere la determinación de todas las funciones de distribuciones acumulativas (CCDFs) marginales y todas las CCDFs de n -variables (y de N -variables) en cada ubicación \mathbf{x} (o $\boldsymbol{\chi}$) en el tiempo particular t . Esta tarea se simplifica sustituyendo cada CCDF n -variada (y cada CCDF N -variada) con una CCDF univariada en cada una de las n (o N) ubicaciones espaciales en el tiempo particular t . Los campos condicionales de las ecs. 3.7 y 3.8 se pueden escribir entonces:

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

$$\{H_t(\boldsymbol{\chi}_i) | H_t(\boldsymbol{\chi}_\alpha): \boldsymbol{\chi}_\alpha \in \Omega_\chi, i = 1, 2, \dots, N, \alpha = 1, 2, \dots, N_h, t = 1, 2, 3, \dots\} \quad (3.10)$$

Suponiendo que la relación conjunta de $Y(\mathbf{x})$ y $H(\boldsymbol{\chi})$ es multi-gaussiana (esto es, si se supone un modelo multi-gaussiano), entonces el campo aleatorio condicional $Y_t(\mathbf{x}_i) | H_t(\boldsymbol{\chi}_\alpha)$ se puede escribir como la suma de dos campos aleatorios independientes (Journel y Huijbregts, 1978):

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

y ambos campos aleatorios en la ec. 3.11 se pueden calcular con la técnica del Kriging:

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

$$Y_s^*(\mathbf{x}_i) = \sum_{i=1}^n \sum_{\alpha=1}^{N_h} \lambda_i(\boldsymbol{\chi}_\alpha) H_t(\boldsymbol{\chi}_\alpha)$$

Substituyendo las ecs. 3.12 en la ec. 3.11 se obtiene:

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

donde $\lambda_i(\boldsymbol{\chi}_\alpha)$ son funciones de ponderación que representan la importancia relativa de las observaciones $H_t(\boldsymbol{\chi}_\alpha)$ en la estimación del valor de $Y_t(\mathbf{x}_i)$. Las funciones ponderadoras son las soluciones de los siguientes sistemas de ecuaciones lineales:

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

donde $C_H(\boldsymbol{\chi}_\alpha, \boldsymbol{\chi}_j; t, t)$ representa las funciones de autocovarianza espacio-temporal entre valores del logaritmo natural de la conductividad hidráulica y $C_{YH}(\mathbf{x}_i, \boldsymbol{\chi}_\alpha; t-1, t)$ representa las funciones de covarianza cruzada espacio-temporales entre valores del logaritmo natural de la conductividad hidráulica y la carga hidráulica. Para que la ec. 3.14 tenga solución única, las funciones de covarianza espacio-temporales involucradas tienen que ser positivo-definidas. Se pueden escribir ecuaciones semejantes a la 3.13 y a la 3.14 para el campo condicional $H_t(\boldsymbol{\chi}_{\alpha_i}) | H_t(\boldsymbol{\chi}_\alpha)$ (Vázquez y Auvinet, 2015).

La determinación de las funciones de covarianza espacio-temporales en la ec. 3.14 sin embargo está lejos de ser un problema trivial debido a que efectos no lineales asociados al problema de flujo de agua impactan las relaciones estadísticas entre $Y_t(\mathbf{x})$ y $H_t(\boldsymbol{\chi})$. A pesar de que algunos modelos teóricos generales de funciones de covarianza espacio-temporales podrían adoptarse para cuantificar las dependencias, no es del todo clara la conveniencia de tales modelos para enfrentar problemas de flujo de agua en medios porosos con propiedades no multi-gaussianas. Por ello, para resolver la ec. 3.5 se utiliza el método de Monte Carlo y las funciones de covarianza necesarias se determinan con el ensamble de realizaciones. Esta es la esencia de los métodos ensamblados. Las funciones de covarianza que se obtienen de esta manera son en efecto empíricas, sin embargo, como un promedio de varias realizaciones estas funciones conducen a matrices positivas definidas y pueden utilizarse directamente sin ser modeladas.

Vale la pena mencionar que la ec. 3.13 se puede interpretar como un mecanismo bayessiano en el sentido de que el campo aleatorio condicional $Y_t(\mathbf{x})$ depende del conocimiento *a priori* que se compone del campo aleatorio $Y_{t-1}(\mathbf{x})$, las funciones de covarianza y las observaciones.

3.2.2 El método EnKF

El Filtro de Kalman Ensamblado (EnKF) utiliza el método de Monte Carlo para aproximar las distribuciones marginales de $H_t(\boldsymbol{\chi})$ suponiendo que el modelo de flujo de agua es un proceso markoviano de primer orden. Posteriormente simula las distribuciones de $Y_t(\mathbf{x}_i)$ condicionando cada realización de $Y_{t-1}(\mathbf{x}_i)$ a las observaciones de $H_t(\boldsymbol{\chi}_\alpha)$ suponiendo un modelo multi-gaussiano. Esta etapa de actualización se puede escribir:

$$\mathbf{U}_t^u = \mathbf{U}_0 + \mathbf{K}_t [\mathbf{Z}_t - \mathbf{H}_t^f] \quad (3.15)$$

donde $U_t^u = [\hat{y}_t(\mathbf{x}_1), \hat{y}_t(\mathbf{x}_2), \dots, \hat{y}_t(\mathbf{x}_n)]$ es un vector de dimensión n que contiene realizaciones actualizadas del logaritmo natural de la conductividad hidráulica, $U_0 = [y_{t-1}(\mathbf{x}_1), y_{t-1}(\mathbf{x}_2), \dots, y_{t-1}(\mathbf{x}_n)]$ es un vector de dimensión n con realizaciones simuladas del logaritmo natural de la conductividad hidráulica (realizaciones *a priori*), $Z_t = [h_{1,t}, h_{2,t}, \dots, h_{N_h,t}]$ es el vector de las observaciones de dimensión N_h , y $H_t^f = [h_t(\boldsymbol{\chi}_1), h_t(\boldsymbol{\chi}_2), \dots, h_t(\boldsymbol{\chi}_{N_h})]$ es un vector reducido de estados pronosticados (realizaciones de la carga hidráulica en las ubicaciones de las observaciones) de dimensión N_h .

La matriz K_t ($n \times N_h$) es la denominada “ganancia de Kalman”, que se ensambla después de obtener las funciones de ponderación asociadas a cada ubicación \mathbf{x}_i en el tiempo t , como:

$$K_t = \begin{bmatrix} \lambda_1(\mathbf{x}_1) & \lambda_1(\mathbf{x}_2) & \dots & \lambda_1(\mathbf{x}_{N_h}) \\ \vdots & \vdots & & \vdots \\ \lambda_n(\mathbf{x}_1) & \lambda_n(\mathbf{x}_2) & \dots & \lambda_n(\mathbf{x}_{N_h}) \end{bmatrix} \quad (3.16)$$

La etapa de actualización del método EnKF se repite en el siguiente tiempo para el que haya observaciones disponibles, pero el nuevo campo aleatorio *a priori* del logaritmo natural de la conductividad hidráulica es el campo *a posteriori*. De esta manera, la etapa de actualización se efectúa secuencialmente incorporando observaciones del pasado y las observaciones en el tiempo presente en el proceso de inferencia.

3.2.3 El método EnKF modificado

En esta variante del método EnKF la etapa de actualización se realiza dentro de un espacio transformado en el que las distribuciones marginales de los campos aleatorios involucrados son gaussianas con media cero y varianza uno. Al finalizar la etapa de actualización las distribuciones de las variables son retransformadas a sus distribuciones originales. Las etapas adicionales del método EnKF modificado se explican en seguida.

3.2.3.1 Transformaciones gaussianas

Para realizar las transformaciones gaussianas, se establecen estadísticamente, con el conjunto de realizaciones, CDFs del logaritmo natural de la conductividad hidráulica en las ubicaciones \mathbf{x} , es decir, $F_Y(y; \mathbf{x}_i)$ para $i=1, 2, \dots, n$ y CDFs de cargas hidráulicas en las ubicaciones $\boldsymbol{\chi}$, es decir, $F_H(h; \boldsymbol{\chi}_j)$ para $j=1, 2, \dots, N$. Luego se aplica la transformación gaussiana G^{-1} (con media cero y varianza unitaria) a cada caso, es decir, $y' = G^{-1}[F_Y(y; \mathbf{x}_i)]$ y $h' = G^{-1}[F_H(h; \boldsymbol{\chi}_j)]$, donde y' y h' son los valores gaussianos transformados. Adicionalmente se aplica una estandarización de las observaciones en cada ubicación $\boldsymbol{\chi}_\alpha$ para $\alpha=1, \dots, N_h$. La función que relaciona y con y' o h con h' en el plano cartesiano x - y se denomina *función de anamorfosis gaussiana* (Fig. 3.1(a)).

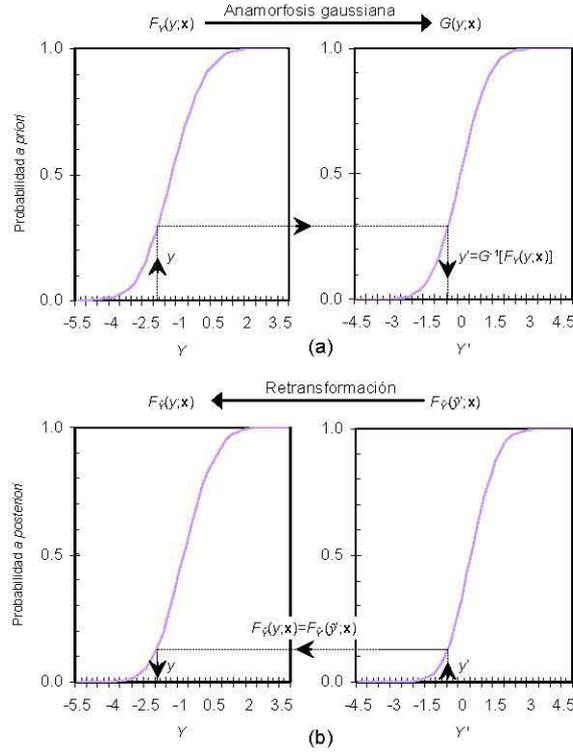


Fig. 3.1 Ilustración de la etapa de transformación del EnKF y el pf-EnKF. a) Proceso de transformación gaussiana. b) Proceso de retransformación.

3.2.3.2 Etapa de actualización

Después de realizar todas las transformaciones descritas en la sección anterior, la etapa de actualización del método EnKF modificado se puede escribir ahora como:

$$\tilde{\mathbf{U}}_t^u = \tilde{\mathbf{U}}_0 + \tilde{\mathbf{K}}_t [\tilde{\mathbf{Z}}_t - \tilde{\mathbf{H}}_t^f] \quad (3.17)$$

donde $\tilde{\mathbf{U}}_t^u = [\hat{y}'_t(\mathbf{x}_1), \hat{y}'_t(\mathbf{x}_2), \dots, \hat{y}'_t(\mathbf{x}_n)]$ es un vector de dimensión n que contiene realizaciones actualizadas del logaritmo natural de la conductividad hidráulica, $\tilde{\mathbf{U}}_0 = [y'_{t-1}(\mathbf{x}_1), y'_{t-1}(\mathbf{x}_2), \dots, y'_{t-1}(\mathbf{x}_n)]$ es un vector de dimensión n con realizaciones simuladas del logaritmo natural de la conductividad hidráulica (realizaciones *a priori*), $\tilde{\mathbf{Z}}_t = [h'_{1,t}, h'_{2,t}, \dots, h'_{N_h,t}]$ es el vector de las observaciones de dimensión N_h , y $\tilde{\mathbf{H}}_t^f = [h'(\boldsymbol{\chi}_1), h'(\boldsymbol{\chi}_2), \dots, h'(\boldsymbol{\chi}_{N_h})]$ es un vector reducido de estados pronosticados (realizaciones de la carga hidráulica en las ubicaciones de las observaciones) de dimensión N_h .

La matriz $\tilde{\mathbf{K}}_t$ ($n \times N_h$) se ensambla después de obtener las funciones de ponderación asociadas a cada ubicación \mathbf{x}_i en el tiempo de actualización t utilizando los valores transformados. La etapa de actualización del método EnKF modificado también se realiza secuencialmente.

3.2.3.3 Retransformación

Los valores condicionales que corresponden a las variables originales se pueden obtener por medio de un proceso de retransformación en el que la CCDF del logaritmo natural de la conductividad hidráulica retransformada en cada ubicación \mathbf{x} , es decir, $F_{\tilde{Y}}(\tilde{y};\mathbf{x})$, se obtiene

haciendo $F_{\hat{y}}(y; \mathbf{x}) = F_{\hat{y}'}(\hat{y}'; \mathbf{x})$. Esto significa que el valor de la CCDF de la variable original y (ver la sección 3.2.3.1) se identifica con el valor de la CCDF en su valor correspondiente de transformación gaussiana \hat{y}' . Dicho proceso se ilustra en la Fig. 3.1(b).

3.2.4 Método propuesto basado en la técnica de simulación “p-field”

Además de la etapa de actualización del método EnKF modificado presentado anteriormente, se puede llevar a cabo una etapa de actualización basada en la técnica de simulación “p-field” como se propone en este capítulo. Esta extensión del método EnKF se denomina aquí método pf-EnKF y se presenta en los siguientes párrafos.

Una CCDF de valores del logaritmo natural de la conductividad en cada ubicación \mathbf{x} , es decir, $F_{\hat{y}}(\hat{y}'; \mathbf{x}_i)$ para $i=1,2,\dots,n$ se establecen estadísticamente utilizando los conjuntos de realizaciones actualizadas con el método EnKF modificado. Muestras de valores del logaritmo natural de la conductividad hidráulica en cada ubicación \mathbf{x} , es decir, $\hat{y}'_i(\mathbf{x}_i)$ se pueden obtener a partir de las distribuciones condicionales mediante la técnica de simulación “p-field” (Srivastava, 1992; Froidevaux, 1993):

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

donde $p_{t-1}(\mathbf{x}_i)$ son los denominados campos de probabilidades (campos p), del campo de valores del logaritmo natural de la conductividad hidráulica. Los campos p se pueden establecer con los valores de las funciones de distribución acumulativas empíricas de cada realización antes de la etapa de actualización. Por ejemplo, si la etapa de actualización está en el tiempo $t=1$, entonces los campos p se forman con las realizaciones en el tiempo $t=0$. El proceso de realización descrito por la ecuación 3.16 se ilustra en la Fig. 3.2.

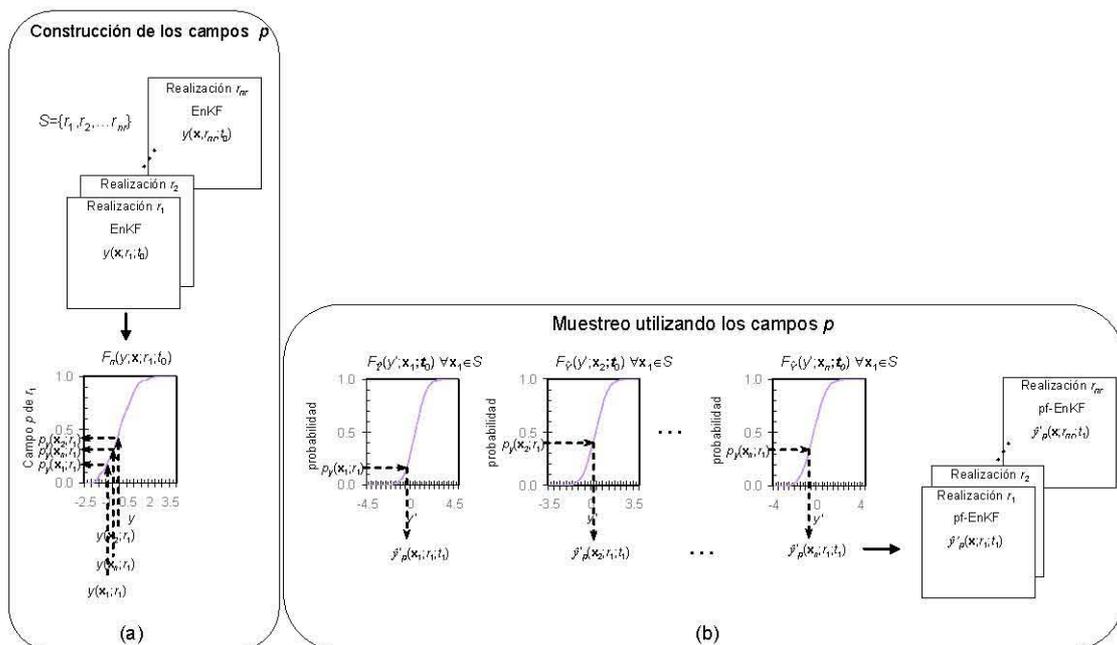


Fig. 3.2 Ilustración del proceso de simulación descrito por la ecuación 3.16. a) Generación de los campos p . b) Muestreo de las funciones de distribución condicionales utilizando los campos p .

La estructura de correlación de los campos simulados actualizados $\hat{y}'_i(\mathbf{x})$ es tomada en cuenta por la estructura de correlación de los campos $p_{t-1}(\mathbf{x})$ (Goovaerst, 1997). El proceso de retransformación de los valores simulados $\hat{y}'_i(\mathbf{x}_i)$ se realiza según lo explicado en la sección 3.2.3.3.

3.3 Descripción del experimento numérico

Para explorar el desempeño del método propuesto (pf-EnKF), se considera un acuífero hipotético, completamente saturado, unidimensional, vertical y heterogéneo. El acuífero tiene 40 m de profundidad y está subdividido con 80 elementos finitos, cada uno de los cuales tiene una longitud de 0.5 m. A cada elemento finito e_i para $i=1, \dots, 80$ se le asigna un valor del logaritmo natural de la conductividad hidráulica $y^{ref}(\mathbf{x}_i)$ de acuerdo al siguiente procedimiento. Primeramente, se simula un campo multi-gaussiano $g(\mathbf{x})$ con función de autocovarianza exponencial y escala de correlación $\alpha_x=2.5$ m utilizando una versión modificada del generador de campos aleatorios SGSIM (Deutsch y Journel, 1996) (Fig. 3.3(a)). En segundo lugar, se aplica la V -transformación al campo $g(\mathbf{x})$ (Bárdossy y Li, 2008):

$$v(\mathbf{x}) = \begin{cases} k(g(\mathbf{x})-m)^\alpha & \text{si } g(\mathbf{x}) \geq m \\ (m-g(\mathbf{x})) & \text{si } g(\mathbf{x}) < m \end{cases} \quad (3.19)$$

con parámetros elegidos arbitrariamente $m=0$, $k=1$ y $\alpha=2$ para obtener el campo transformado $v(\mathbf{x})$ mostrado en la Fig. 3.3(b). En tercer lugar, se impone una distribución gaussiana al campo $v(\mathbf{x})$ haciendo $y'=G^{-1}[F_V(v)]$, donde $F_V(v)$ es la CDF empírica del campo $v(\mathbf{x})$ y G la CDF gaussiana teórica. Finalmente, dicho campo $y'(\mathbf{x})$ se modula para obtener un campo $y^{ref}(\mathbf{x})$ normalmente distribuido con valor medio $\mu_Y=-1.654$ y varianza $\sigma_Y^2=1$ haciendo: $y^{ref}(\mathbf{x})=\mu_Y+y'(\mathbf{x})\sigma_Y$. Se supone que cada uno de estos valores es constante dentro de su elemento finito e_i . Este campo del logaritmo natural de la conductividad hidráulica, que se muestra en la Fig. 3.3(c), se considera un “verdadero estado de la naturaleza” y se denomina el *acuífero de referencia*.

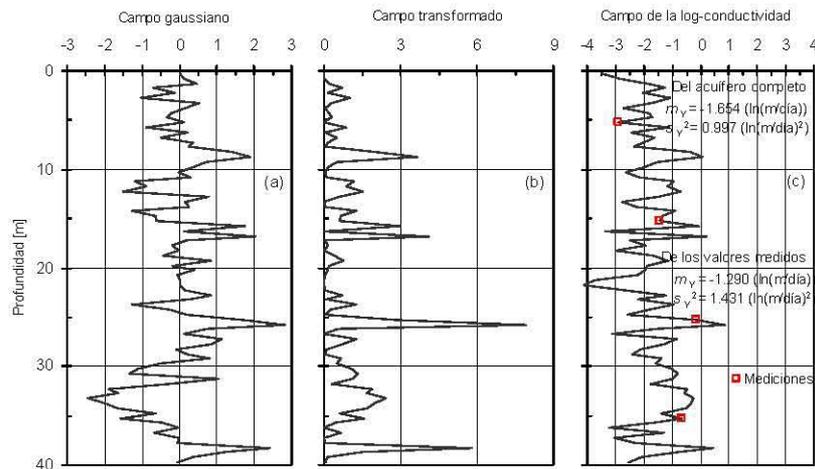


Fig. 3.3 Campos unidimensionales. a) Campo gaussiano inicial. b) Campo después de aplicar la V -transformación con parámetros $m=0$, $k=1$ y $\alpha=2$ al campo gaussiano inicial. c) Campo final del logaritmo natural de la conductividad después de imponerle una distribución normal con valor esperado $\mu_Y=-1.654$ y varianza $\sigma_Y^2=1$ al campo V -transformado. También se reportan las estadísticas de los valores muestreados (cuadrados vacíos).

Deben mencionarse varias propiedades interesantes de la V -transformación. En primer lugar, la densidad simétrica del campo gaussiano $g(\mathbf{x})$ es transformada en una densidad asimétrica a través de los parámetros m , k y α . En segundo lugar, la función de autocovarianza empírica de $g(\mathbf{x})$ no se preserva en $v(\mathbf{x})$ debido a que la V -transformación no es monótona (Fig. 3.4). En tercer lugar, la correlación espacial de $v(\mathbf{x})$ es más fuerte para los valores mayores que la mediana que para los valores menores que la mediana; es decir, la correlación espacial de $v(\mathbf{x})$ es asimétrica. Esta última característica del campo se mantiene después de imponerle la función de distribución gaussiana (normal) porque la transformación gaussiana es monótona (Deutsch y Journel, 1996). Debido a que los valores en $y^{ref}(\mathbf{x})$ tienen la distribución normal, los valores $k_s(\mathbf{x})=\exp(y^{ref}(\mathbf{x}))$ tienen la distribución lognormal con valor esperado $\mu_{k_s}=0.315$ m/día y coeficiente de variación $CV_{k_s}=1.31$. Los campos de conductividades con distribuciones marginales lognormales y estructuras de correlación asimétricas se consideran más representativos de los acuíferos naturales (Gómez-Hernández y Wen, 1998; Journel y Zhang, 2006), como se discutió en el capítulo 1.

Utilizando el acuífero de conductividades de referencia, se generan respuestas de cargas de agua subterránea mediante la solución de una condición de flujo transitorio con elementos finitos. En $t=0$ días, la distribución inicial de las cargas es hidrostática. En $t \geq 0$ días, la carga hidráulica decrece con el tiempo a razón de 0.15 m por día durante 150 días en la frontera inferior. Para los fines de este ejemplo numérico, se supone que la distribución de las cargas en $t=90$ días es la condición inicial (designada como $t=0$ días en la Fig. 3.5 y en lo sucesivo). Adicionalmente, se supone que las respuestas de cargas de agua subterránea están disponibles en los tiempos $t=3$, $t=18$ y $t=60$ días en las dos ubicaciones indicadas en la Fig. 3.5. De esta forma se generan dos historias con tres valores de la carga hidráulica cada una de ellas. Estas variables indirectas e informativas de la conductividad hidráulica del acuífero se consideran *observaciones piezométricas transitorias disponibles*. En cada uno de estos tres tiempos, se lleva a cabo el proceso de actualización de los dos esquemas, EnKF y pf-EnKF. El coeficiente de almacenamiento se supone igual a 0.001 en todo el acuífero.

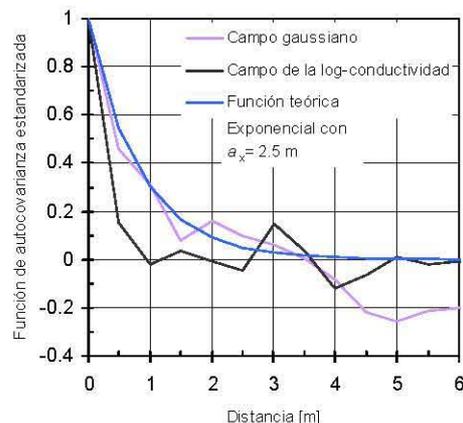


Fig. 3.4 Funciones de autocovarianza estandarizada del campo gaussiano inicial y del logaritmo natural de la conductividad del campo de referencia. También se muestra una función exponencial teórica con fines comparativos.

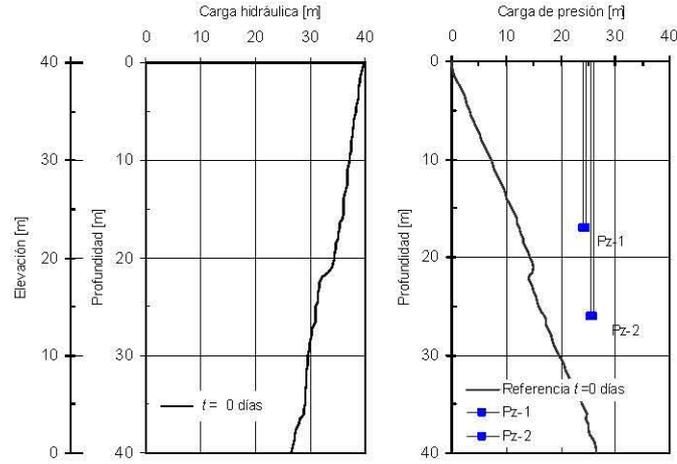


Fig. 3.5 Perfil de cargas hidráulicas en el acuífero de referencia al tiempo $t=0$ días. También se indican con cuadrados sólidos las profundidades de las puntas de dos piezómetros (Pz-1 y Pz-2).

En las cuatro ubicaciones que se indican en la Fig. 3.3(c) se hace un muestreo del acuífero de referencia $y^{ref}(\mathbf{x})$ y los valores se consideran *mediciones directas del logaritmo natural de la conductividad hidráulica*. La media y la varianza del conjunto de valores muestreados se reportan en la misma figura. Observe que estas estadísticas sobreestiman el valor medio y la varianza del logaritmo natural de la conductividad hidráulica del acuífero de referencia. Note además en la Fig. 3.4 que una función de autocovarianza exponencial con escala de correlación $\alpha_x=2.5$ m también sobreestima la escala de correlación del acuífero de referencia. Para modelar una situación en la que las estadísticas del acuífero sólo se conocen *a priori* de manera aproximada, la media y varianza de la muestra así como la función de autocovarianza antes mencionada se utilizan para simular dos mil realizaciones condicionales del logaritmo natural de la conductividad hidráulica. Este ensamble de realizaciones representa una aproximación inicial al acuífero de referencia. El número de realizaciones simuladas asegura la estabilidad de las siguientes dos medidas de error, de acuerdo con algunas cálculos preliminares.

Para evaluar el desempeño tanto del filtro EnKF como del filtro pf-EnKF en la estimación del logaritmo natural de la conductividad y la carga hidráulica, se consideran la raíz cuadrada del error cuadrático medio (RMSE) y la medida de dispersión SPREAD. La RMSE se evalúa por medio de:

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

donde n es el número de conductividades en la región de flujo; $y^*(\mathbf{x}_i)$ es el valor medio del logaritmo natural de la conductividad hidráulica estimada en la ubicación \mathbf{x}_i y $y^{ref}(\mathbf{x}_i)$ es el logaritmo natural de la conductividad hidráulica de referencia también en la ubicación \mathbf{x}_i .

La dispersión (SPREAD) se calcula como:

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

donde, $s_{en}^2(x_i)$ es la varianza de la estimación del logaritmo natural de la conductividad hidráulica en la ubicación x_i calculada estadísticamente con el conjunto de realizaciones.

La RMSE es una medida de la diferencia de las medias de los campos estimados y de referencia y SPREAD es una medida de la dispersión del campo estimado en torno al campo de referencia. Por lo tanto, pueden considerarse como medidas de exactitud y precisión de las estimaciones, respectivamente.

3.4 Resultados y discusión

En esta sección se discuten las características de las realizaciones de la media condicional de los campos del logaritmo natural de la conductividad hidráulica y de la carga hidráulica. Los resultados del EnKF se comparan con los resultados del pf-EnKF. Antes de analizar los efectos del condicionamiento se analizan los efectos de la transformación gaussiana.

3.4.1 Efectos de la transformación gaussiana

La Fig. 3.6 ilustra el efecto de la transformación gaussiana de las cargas hidráulicas en un nodo seleccionado arbitrariamente antes de la primera etapa de actualización. En general, la forma de las distribuciones locales dependerá de la ubicación del nodo dentro de la región de flujo y de las condiciones de frontera del problema en cuestión. En cualquier caso, las funciones de distribución locales se podrán transformar en distribuciones gaussianas construyendo funciones locales de anamorfosis gaussianas en forma numérica, como se ha explicado. Por ejemplo, observe en la Fig. 3.6(a) que aunque los valores originales exhiben una distribución sesgada, la variable transformada se vuelve simétrica en torno a la media, mostrando la bien conocida forma de campana después de la anamorfosis gaussiana (Figs. 3.6(b),(c)).

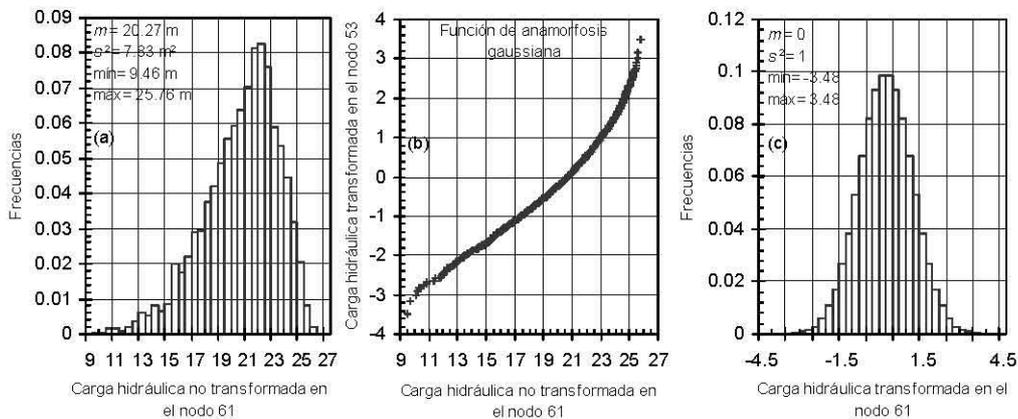


Fig. 3.6 Transformación gaussiana de las cargas hidráulicas en el nodo 61. a) Histograma de las cargas hidráulicas no transformadas. b) Función de la anamorfosis gaussiana (con media cero y varianza unitaria). c) Histograma de cargas hidráulicas después de la anamorfosis gaussiana.

La Fig. 3.7 representa la relación entre valores del logaritmo natural de la conductividad hidráulica y valores de la carga hidráulica en ubicaciones seleccionadas arbitrariamente, antes (Fig. 3.7(a)) y después (Fig. 3.7(b)) de aplicar las respectivas transformaciones gaussianas. Debido a que la transformación gaussiana es monótona, no modifica las características bivariadas de la dependencia como por ejemplo la estructura de correlación en diferentes percentiles (Deutsch y Journel, 1996; Chilès y Delfiner, 1999). Sin embargo, el coeficiente de correlación lineal de *Pearson*, que depende del tipo de distribuciones marginales de las variables aleatorias, podría ser diferente antes y después de las transformaciones. En el caso particular de

las variables en las ubicaciones indicadas en la Fig. 3.7, se observa en la esquina superior izquierda de cada figura que el coeficiente de correlación de *Pearson* presenta prácticamente el mismo valor antes y después de las transformaciones. Por lo tanto, el efecto implícito de pseudo-linealización asociado con la anamorfosis gaussiana reportado por Shöniger *et al.* (2012) debe ser considerado como dependiente de la aplicación.

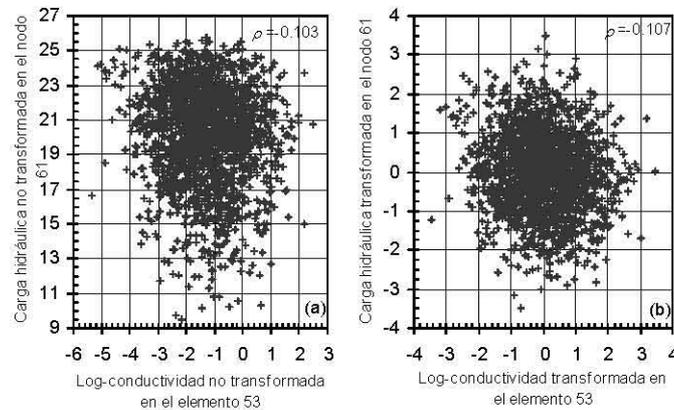


Fig. 3.7 Relación entre el logaritmo natural de la conductividad y la carga hidráulica en dos ubicaciones seleccionadas arbitrariamente (ρ es el coeficiente de correlación de *Pearson*). a) Antes de la transformación de ambas variables. b) Después de la transformación gaussiana de ambas variables.

3.4.2 Efectos del condicionamiento basado sólo en las log-conductividades

Primero se analiza el impacto de condicionar las realizaciones del logaritmo natural de la conductividad hidráulica sólo a las mediciones directas. Las Figs. 3.8(a),(b) muestran comparaciones entre el campo de referencia y la media del conjunto de realizaciones condicionales del logaritmo natural de la conductividad hidráulica que se obtienen con los filtros EnKF y pf-EnKF, respectivamente. Al contrastar ambos campos condicionales, se observa que el filtro pf-EnKF produce, hasta cierto grado, más variabilidad entre mediciones que el filtro EnKF. Al observar los valores RMSE y SPREAD que se muestran en la esquina inferior derecha de las figuras se puede establecer que el filtro EnKF es más exacto y que el pf-EnKF es más preciso.

Las Figs. 3.9(a),(b) reproducen los perfiles de desviación estándar (incertidumbre) calculados con las realizaciones condicionales del logaritmo natural de la conductividad hidráulica de los filtros EnKF y pf-EnKF, respectivamente. Observe que el efecto del condicionamiento es reducir, en promedio, la incertidumbre *a priori* y colapsarla a cero en las ubicaciones de las mediciones. Los perfiles de ambos filtros son muy similares.

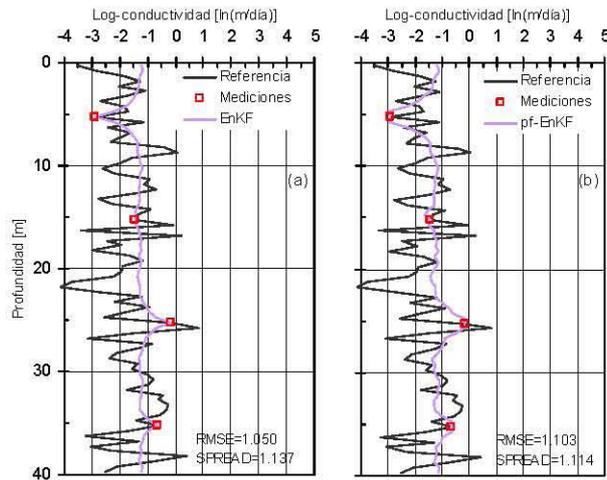


Fig. 3.8 Campos del logaritmo natural de la conductividad condicionados a mediciones de esta misma variable. a) Con el método EnKF. b) Con el método pf-EnKF. También se muestra el campo de referencia.

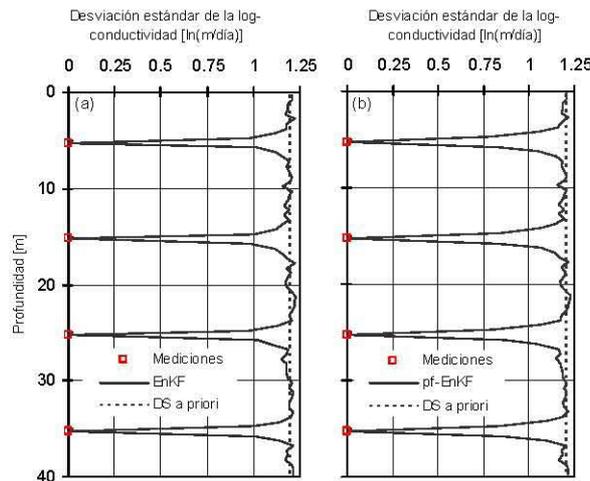


Fig. 3.9 Perfiles de la desviación estándar condicional del logaritmo natural de la conductividad hidráulica con respecto a la profundidad para diferentes tiempos (los cuadrados vacíos indican las ubicaciones de los valores conocidos). a) De los campos del EnKF. b) De los campos del pf-EnKF.

3.4.3 Efectos del condicionamiento basado en las log-conductividades y en las cargas transitorias

Ahora se examina el impacto adicional de condicionar las realizaciones del logaritmo natural de la conductividad hidráulica a las cargas hidráulicas transitorias. En las Figs. 3.10(a),(b),(c) se muestran las comparaciones entre el campo de referencia y la media del conjunto de realizaciones condicionales que se obtienen con el método EnKF en los tiempos $t=3$ días, $t=18$ días y $t=60$ días, respectivamente. Las Figs. 3.11(a),(b),(c) muestran las mismas comparaciones pero ahora la media del campo condicional del logaritmo natural de la conductividad hidráulica se obtiene con las realizaciones el método pf-EnKF.

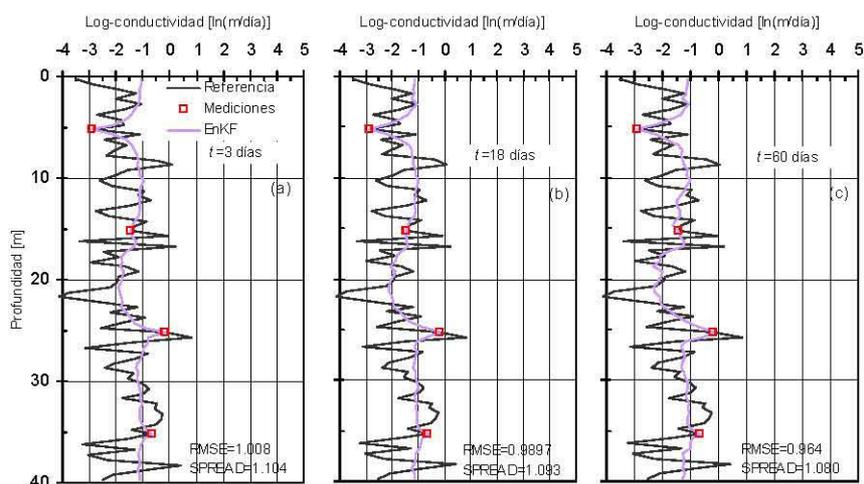


Fig. 3.10 Campos del logaritmo natural de la conductividad condicionados a historias de cargas hidráulicas con el método EnKF. También se muestra el campo de referencia. a) Al tiempo $t=3$ días. b) Al tiempo $t=18$ días. c) Al tiempo $t=60$ días.

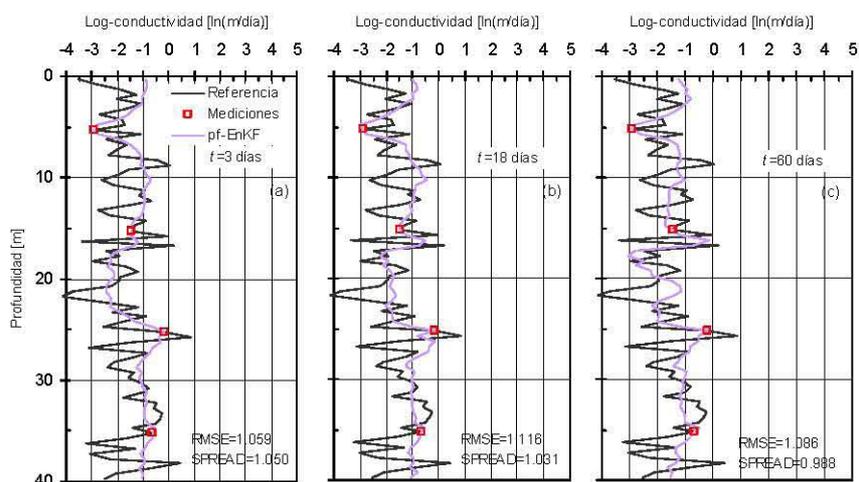


Fig. 3.11 Campos del logaritmo natural de la conductividad condicionados a historias de cargas hidráulicas con el método pf-EnKF. También se muestra el campo de referencia. a) Al tiempo $t=3$ días. b) Al tiempo $t=18$ días. c) Al tiempo $t=60$ días.

Al examinar las medidas RMSE y SPREAD indicadas en cada figura (esquina inferior derecha), se puede establecer que el EnKF produce resultados más exactos que el pf-EnKF en todos los tiempos, aunque las estimaciones del pf-EnKF son más precisas. Por ejemplo, los valores RMSE y SPREAD para la media del campo condicional del EnKF en el tiempo $t=60$ son 0.964 y 1.080, respectivamente, en tanto que los valores de tales medidas para la media del campo condicional del pf-EnKF son 1.086 y 0.988, respectivamente. La mayor exactitud del método EnKF puede atribuirse a que el pf-EnKF produce mayores fluctuaciones entre mediciones que no siguen cercanamente la variabilidad del acuífero de referencia. No obstante, deberá recordarse que los valores RMSE y SPREAD sólo miden la calidad de la estimación local, es decir, no indican nada acerca de la calidad de la estimación multivariada. Para explorar la calidad de la estimación multivariada, en el capítulo 4 se examinarán las cópulas empíricas bivariadas.

Los perfiles de desviación estándar calculados con las realizaciones condicionales del logaritmo natural de la conductividad hidráulica del EnKF y del pf-EnKF se reportan en las figuras 3.12(a),(b), respectivamente. En términos generales, la incertidumbre decrece en ambos casos a medida que se utilizan las observaciones de cargas de agua subterránea en más tiempos, excepto en las ubicaciones de las mediciones directas en donde la incertidumbre es cero en todos los tiempos. En esas figuras se ilustra que la incertidumbre se vuelve más pequeña en torno a las profundidades de 17 m y 26 m (en donde se localizan las puntas de los dos piezómetros) que en otras profundidades y que esta reducción es más significativa en el perfil del pf-EnKF que en el perfil del EnKF.

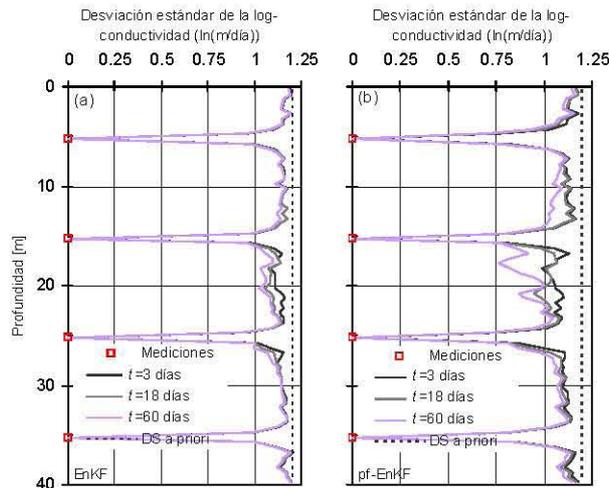


Fig. 3.12 Perfiles de la desviación estándar condicional del logaritmo natural de la conductividad hidráulica con respecto a la profundidad para diferentes tiempos (los cuadrados vacíos indican las ubicaciones de los valores conocidos). a) De los campos del EnKF. b) De los campos del pf-EnKF.

La Fig. 3.13 muestra las distribuciones de frecuencias de valores del logaritmo natural de la conductividad hidráulica tanto del campo de referencia como del ensamble de realizaciones *a priori* y de los ensambles *a posteriori* al final del proceso de condicionamiento. Recordemos que el valor medio de la conductividad del acuífero de referencia estaba sobreestimada por el campo aleatorio *a priori*, por ello la función de distribución de dicho campo se ubica a la derecha de la función de distribución del acuífero. Al examinar las distribuciones de los campos *a posteriori* se observa que muestran algunos rasgos de la distribución de referencia (como algunos de los “picos” de ambas ramas) y que están desplazados ligeramente hacia la izquierda de la distribución del campo aleatorio *a priori*. Esto indica el intento del proceso de condicionamiento de ambos filtros para llevar sus distribuciones *a priori* hacia la distribución de referencia. Observe que las distribuciones *a posteriori* de ambos filtros son muy similares.

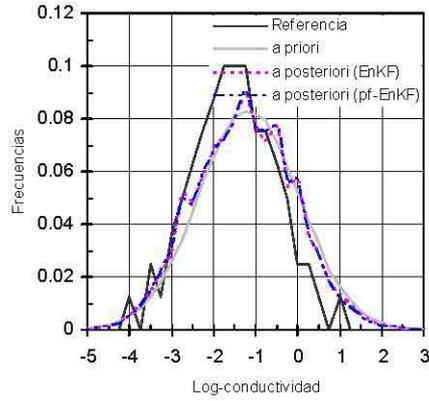


Fig. 3.13 Distribución de frecuencias del logaritmo natural de la conductividad del campo de referencia, del conjunto de realizaciones *a priori* y de los conjuntos de realizaciones *a posteriori* del EnKF y pf-EnKF al tiempo $t=60$ días.

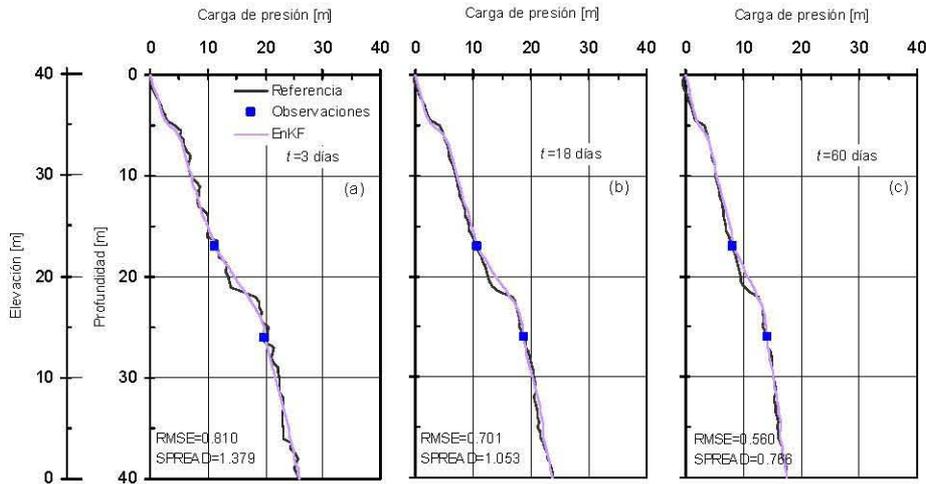


Fig. 3.14 Campos condicionales de la carga de presión que se obtienen con el método EnKF. También se muestra el campo de referencia. a) Al tiempo $t=3$ días. b) Al tiempo $t=18$ días. c) Al tiempo $t=60$ días.

Ahora se revisan los campos condicionales de cargas de presión. Las Figs. 3.14(a),(b) y (c) muestran comparaciones entre los perfiles de cargas de referencia y los perfiles de valores medios que se calculan con las realizaciones condicionales de la carga hidráulica del filtro EnKF en los tiempos $t=3$, $t=18$ y $t=60$ días, respectivamente. Se ve que dichos perfiles se acercan más a los de referencia a medida que se toman en cuenta un mayor número de observaciones de los registros de cargas piezométricas en la etapa de actualización. El mismo comportamiento se observa en las Figs. 3.15(a),(b),(c) donde se muestran comparaciones de los perfiles de cargas de referencia con los perfiles medios que se obtienen con las realizaciones condicionales de la carga hidráulica del filtro pf-EnKF en los tiempos mencionados anteriormente, excepto que en este caso el filtro pf-EnKF introduce mayores fluctuaciones entre las observaciones que se desvían más de los valores de referencia. En consecuencia, las estimaciones del pf-EnKF son menos exactas que las del EnKF. Esto se puede verificar al observar los valores de la medida RMSE que se indican en las figuras. Los perfiles de cargas de presión estimados con el pf-EnKF aún así corresponden bien con los perfiles de referencia.

También se puede destacar, a partir de los valores de la medida SPREAD en los tiempos especificados, que las estimaciones del pf-EnKF son más precisas.

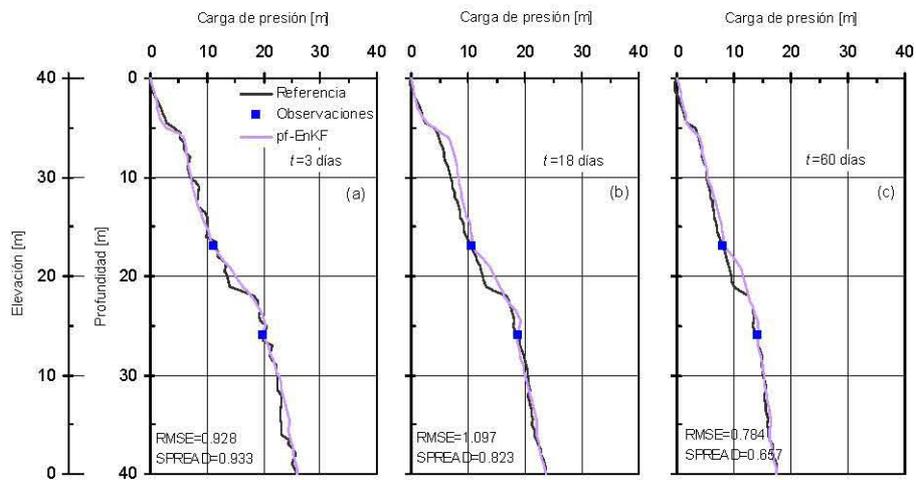


Fig. 3.15 Campos condicionales de la carga de presión que se obtienen con el método pf-EnKF. También se muestra el campo de referencia. a) Al tiempo $t=3$ días. b) Al tiempo $t=18$ días. c) Al tiempo $t=60$ días.

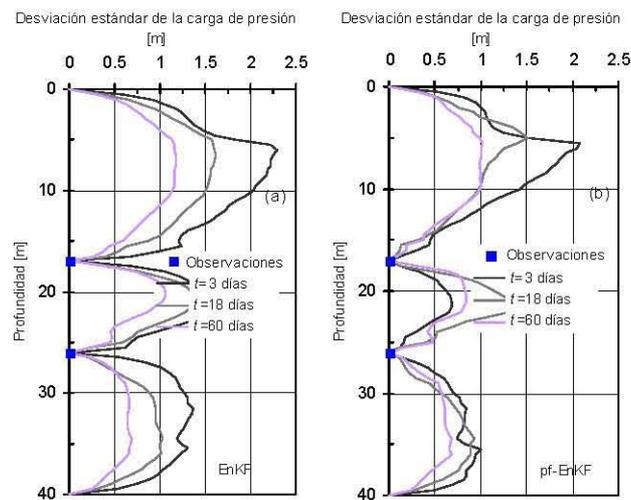


Fig. 3.16 Perfiles de la desviación estándar condicional de la carga de presión con respecto a la profundidad para diferentes tiempos (los cuadrados sólidos indican las ubicaciones de dos observaciones). a) De los campos del EnKF. b) De los campos del pf-EnKF.

Las Figs. 3.16(a),(b) muestran los perfiles de desviación estándar condicional de cargas de presión en los tiempos $t=3$, $t=18$ y $t=60$ días cuantificados con las realizaciones del EnKF y del pf-EnKF, respectivamente. Como se esperaba, la incertidumbre es nula en las profundidades donde se registra la carga de presión y en las fronteras superior e inferior donde se prescriben las cargas hidráulicas. Como se ilustra en ambas figuras, la desviación estándar disminuye en términos generales con el tiempo, pero localmente aumenta en algunas profundidades. Este comportamiento es evidente en el perfil calculado con los campos del pf-EnKF. Dicho comportamiento puede ser explicado por la naturaleza numérica de las funciones de covarianza calculadas en los esquemas EnKF. Como lo explican Xu *et al.* (2013), las funciones de

covarianza calculadas numéricamente producen estimaciones de esta que fluctúan en torno a cero en distancias para las que deberían ser cero. Una forma de superar este problema es mediante el uso de técnicas de localización de covarianza.

3.5 Conclusiones

En este capítulo se describieron los detalles de la implementación de una etapa de actualización dentro del procedimiento EnKF basada en la técnica de simulación “*p-field*”. Esta extensión del método EnKF se designó con el nombre de método pf-EnKF. Para ilustrar los efectos y para comparar los resultados del método propuesto frente a los del EnKF, se consideró un problema de flujo monofásico, unidimensional en un medio poroso aleatorio continuo.

Los métodos EnKF y pf-EnKF involucran transformaciones de las funciones de distribución locales en funciones de distribución gaussianas estandarizadas antes de las etapas de actualización. De esta forma, cada etapa de actualización se realiza dentro de un espacio geométrico de distribuciones gaussianas univariadas. Esto le proporciona a los filtros mayor estabilidad durante las etapas de actualización y la posibilidad de utilizar distribuciones no gaussianas para los parámetros del modelo dinámico. También ambos filtros requieren la retransformación de las distribuciones condicionales locales a sus distribuciones originales después de las etapas de actualización.

El resultado más notable de la metodología propuesta es que induce mayores fluctuaciones entre las observaciones en las realizaciones condicionales. Se encontró, sin embargo, que la incitación de mayores fluctuaciones en torno a las estimaciones del valor medio no proporcionó mejores resultados (en términos de exactitud local) en la estimación del logaritmo natural de la conductividad hidráulica y tampoco en la estimación de las cargas hidráulicas. Estas conclusiones pueden explicarse por el hecho de que las fluctuaciones no necesariamente siguieron más de cerca la variabilidad de los correspondientes campos de referencia. Se requiere investigación adicional para evaluar por completo el desempeño del filtro propuesto. El capítulo 4 explorará el desempeño de la metodología propuesta sobre la reproducción de las cópulas espaciales de un campo de referencia de conductividades en 2D.

CAPÍTULO 4

CARACTERIZACIÓN DE CAMPOS HETEROGÉNEOS DE LA CONDUCTIVIDAD HIDRÁULICA

Desde el inicio de este estudio se estableció que la conductividad hidráulica de las formaciones porosas es muy sensible a variaciones de estructura interna. Se mencionó que incluso en un mismo tipo de suelo, la conductividad hidráulica puede variar en varios órdenes de magnitud. Además, se señaló que el número de mediciones disponibles para determinar la conductividad hidráulica de un sitio específico generalmente es limitado. Se ha mencionado también que las mediciones disponibles muestran que, para fines prácticos, la heterogeneidad de la conductividad hidráulica en las formaciones porosas se puede caracterizar mejor de manera probabilista.

Diversos autores han aceptado que la conductividad hidráulica de las formaciones porosas naturales se puede interpretar como una realización de un campo aleatorio (Dagan, 1989; Gelhar, 1993; Zhang, 2002; Rubin, 2003). Los parámetros descriptivos del campo aleatorio se pueden obtener a partir de una realización misma suponiendo que el campo es ergódico (Chilès y Delfiner, 1999). Estimaciones de la conductividad hidráulica de la formación porosa en ubicaciones no observadas y una medida de la incertidumbre de la estimación se pueden obtener por medio de técnicas geoestadísticas de estimación o de simulación (Chilès y Delfiner, 1999).

Un campo aleatorio que se hace específico a un conjunto de mediciones directas de la conductividad hidráulica y/o a un conjunto de observaciones indirectas, como es el caso de las cargas hidráulicas, tomando en cuenta sus ubicaciones particulares, se denomina campo aleatorio condicional. En la práctica, los campos aleatorios condicionales con frecuencia son deseables debido a que proporcionan estimaciones específicas al sitio de la conductividad hidráulica de la formación porosa y debido a que el condicionamiento reduce, en promedio, las incertidumbres *a priori*.

Como ya se ha mencionado, los campos aleatorios condicionales se pueden generar por medio de técnicas de estimación o de simulación (Deutsch y Journel, 1996). En el segundo caso, el reto consiste en integrar la información disponible en un modelo de incertidumbre *a priori* a fin de obtener estimaciones *a posteriori* de la conductividad hidráulica de los medios porosos. En la literatura científica se pueden encontrar diferentes esquemas para afrontar este problema bajo

condiciones de flujo establecido (RamaRao *et al.*, 1995; Gómez-Hernández *et al.*, 1997; Janssen *et al.*, 2006; Hendricks y Gómez-Hernández, 2002) y bajo condiciones de flujo transitorio (Hendricks-Franssen *et al.*, 1999; Zhu y Yeh, 2005; Capilla y Llopis-Albert, 2009). En la última década, los Filtros de Kalman Ensamblados han captado una mayor atención para realizar la simulación estocástica de campos aleatorios condicionados a observaciones dinámicas debido a la facilidad de su implementación y un tiempo de computación aceptable en el manejo de problemas de grandes dimensiones. El método se basa en un esquema de trabajo del tipo Monte Carlo y permite que el modelo físico se incorpore al proceso de caracterización. De esta manera, también se pueden efectuar predicciones con el modelo de flujo de agua subterránea.

En este capítulo se emplean el filtro EnKF modificado y el filtro propuesto pf-EnKF, que se presentaron en el capítulo 3, para realizar la caracterización de un acuífero hipotético no homogéneo con respecto a su conductividad hidráulica. La distribución de las conductividades en el acuífero se genera por medio de la simulación de un campo aleatorio. Un campo simulado, en contraposición a uno real, se conoce completamente y es posible por tanto evaluar el desempeño de ambas técnicas en la recuperación de algunos de sus descriptores estadísticos relevantes, tales como: histograma, funciones de correlación y cópulas bivariadas. Para la generación del campo simulado se utiliza la cópula normal V -transformada. El acuífero hipotético se utiliza para generar historias de la carga hidráulica simulando una prueba de bombeo. Para efectuar el proceso de condicionamiento de ambos filtros, se utiliza un subconjunto de tales historias las cuales están asociadas a ciertas ubicaciones. Para evaluar la calidad de la caracterización, se utilizan diferentes criterios cuantitativos.

4.1 Exposición del problema

En el siguiente análisis se trata el flujo de agua subterránea en un medio poroso continuo, bidimensional y completamente saturado con conductividad hidráulica espacialmente variable descrito por medio de la ecuación:

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

sujeto a:

$$\text{condición inicial:} \quad H|_{t=0} = h_0 \quad (4.2)$$

y condiciones de frontera:

$$\text{Carga de presión constante:} \quad H|_{\Gamma_D} = h_D \quad (4.3)$$

$$\text{Sin flujo normal (descarga específica):} \quad \mathbf{q}(\mathbf{x}, t) \cdot \mathbf{n}(\mathbf{x}) = 0 \quad \mathbf{x} \in \Gamma_N \quad (4.4)$$

donde: ∇ es el operador de gradiente; H es la carga hidráulica [L], h_0 es la carga inicial en la región Ω [L], h_D es la carga prescrita en los segmentos de frontera de Dirichlet Γ_D [L]; $Q(\mathbf{x})$ es la tasa de bombeo por unidad de volumen del acuífero [1/T]; $K_s(\mathbf{x})$ es la conductividad hidráulica [L/T], $\mathbf{q}(\mathbf{x}, t)$ es el flujo (flux) prescrito a través de segmentos de frontera de Neumann Γ_N , $\mathbf{n}(\mathbf{x})$ es un vector saliente normal a la frontera Γ_N y S_s es el almacenamiento específico [L⁻¹].

Debido a que la conductividad hidráulica $K_s(\mathbf{x})$ es un campo aleatorio, la ecuación 4.1 se convierte en una ecuación diferencial estocástica y la respuesta del flujo H también será un campo aleatorio en un tiempo específico. Para obtener la respuesta de la dinámica del modelo que satisface las ecuaciones 4.1 a 4.4 en el contexto de los métodos EnKF, en este estudio se utiliza el método de Monte Carlo. Primeramente, se generan realizaciones independientes de un

campo aleatorio *a priori* de conductividades hidráulicas y las ecuaciones de flujo se resuelven numéricamente con los valores deterministas de cada realización utilizando una versión modificada de un código de elementos finitos (Smith y Griffiths, 2004). La esperanza y varianza de la respuesta del flujo así como las funciones de autocovarianza y varianza cruzada entre conductividades y cargas hidráulicas se determinan con los ensambles de realizaciones por medio de fórmulas bien establecidas de estadística descriptiva, en cada tiempo en que los campos aleatorios *a priori* son actualizados con los métodos EnKF y pf-EnKF.

4.2 Descripción del experimento numérico

4.2.1 Distribución espacial de las conductividades

El acuífero es un dominio cuadrado con lado de 1289 m, subdividido en 64x64 elementos finitos, cada uno de los cuales tiene una longitud de 20 m (Fig. 4.1). A cada elemento finito e_i para $i=1,\dots,4096$ se le asigna un valor del logaritmo natural de la conductividad hidráulica $y^{ref}(\mathbf{x}_i)$ de acuerdo al siguiente procedimiento. En primer lugar se simula un campo aleatorio multi-gaussiano no condicional con función de autocovarianza exponencial y escala de correlación isótropa $\alpha_x=\alpha_y=160$ m utilizando una versión modificada del generador de campos aleatorios SGSIM (Deutsch y Journel, 1996). En segundo lugar, la V -transformación (Bárdossy, 2006):

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

con parámetros arbitrariamente elegidos $m=5$, $k=2$ y $\alpha=1$ se aplica a una realización $g(\mathbf{x})$ del campo aleatorio multi-gaussiano para obtener el campo transformado $v(\mathbf{x})$ de la ecuación 4.5. En tercer lugar, la distribución gaussiana se impone al campo $v(\mathbf{x})$ como $y'=G^{-1}[F_V(v)]$, donde $F_V(v)$ es la CDF empírica del campo $v(\mathbf{x})$ y $G[\cdot]$ la CDF gaussiana teórica. Finalmente, se modula dicho campo $y'(\mathbf{x})$ para obtener un campo y distribuido normalmente con valor medio $\mu_y=1.223$ y desviación estándar $\sigma_y=1$ como: $y^{ref}(\mathbf{x})=\mu_y+y'(\mathbf{x})\sigma_y$. Se supone que cada uno de estos valores es constante dentro de su elemento finito e_i . Vale la pena mencionar que debido a que $y^{ref}(\mathbf{x})$ tiene una distribución normal, $k_s(\mathbf{x})=\exp(y^{ref}(\mathbf{x}))$ tiene una distribución log-normal.

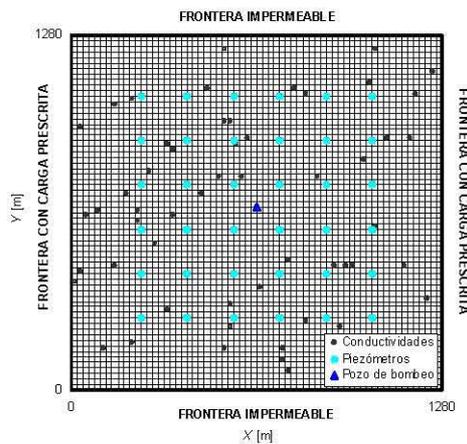


Fig. 4.1 Dimensiones de la región de flujo. Se indican las ubicaciones de las mediciones de la conductividad (círculos vacíos) y de las observaciones de la carga hidráulica (círculos llenos). El triángulo señala la ubicación del pozo de bombeo. Las condiciones de frontera del problema al tiempo $t \geq 0$ se indican en los lados de la región de flujo.

Debido a que los parámetros estadísticos de un campo aleatorio simulado se reproducen con un ensamble de realizaciones, las realizaciones individuales pueden mostrar estadísticas ligeramente diferentes, aunque todas conservarán las mismas características de la dependencia espacial. Estas características están dadas por el modelo de cópula con los parámetros ya mencionados. La gráfica de densidad bivariada de este modelo teórico de cópula se exhibe en la Fig. 4.2. Observe que este modelo de cópula se aproxima mucho al modelo de cópula gaussiana. Por lo tanto, los resultados de este estudio están limitados a aquellos casos de acuíferos no homogéneos con estructura espacial de conductividades con dependencia simétrica. El campo del logaritmo natural de la conductividad hidráulica resultante $y^{ref}(\mathbf{x})$ antes mencionado se considera un “verdadero estado de la naturaleza” y en este estudio se designa como el campo de referencia. Este representa la distribución de las conductividades en un acuífero hipotético. Sus descriptores estadísticos se mostrarán posteriormente (tabla 4.1).

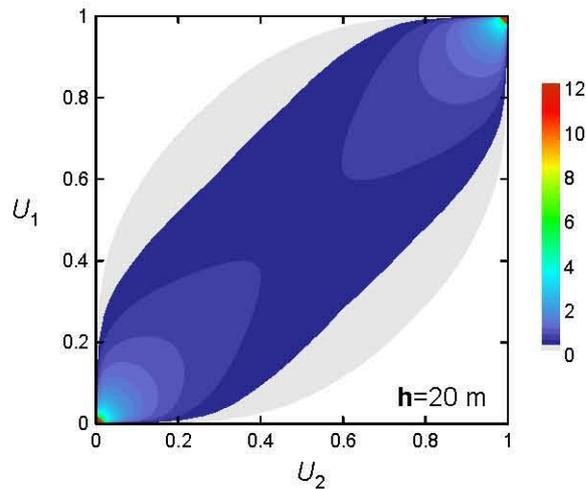


Fig. 4.2 Modelo teórico de densidad de cópula bivariada V -transformada con parámetros $m=5.0$, $k=2.0$ y $\alpha=1.0$ utilizado para generar el campo de referencia (acuífero hipotético).

4.2.2 Historias de la carga hidráulica

El campo de referencia del acuífero sintético se utiliza para generar respuestas de cargas de agua subterránea resolviendo una condición de flujo transitorio con elementos finitos. Los lados norte y sur del acuífero se consideran fronteras impermeables, en tanto que los lados este y oeste son fronteras de Dirichlet con cargas hidráulicas prescritas de 200 m y 198 m, respectivamente. Al tiempo $t=0$ días, la carga hidráulica inicial es de 200 m en cada nodo y para $t>0$ días se bombea desde el acuífero un gasto de $150 \text{ m}^3/\text{día}$ con un pozo ubicado en $X=640$ m y $Y=640$ m hasta que el flujo alcanza un estado uniforme en aproximadamente $t=5$ días. Este período de tiempo se subdivide en 20 etapas de tiempo de igual tamaño a intervalos de $\Delta t=0.25$ días. Para los fines de este ejemplo numérico se supone que las respuestas de cargas de agua subterránea están disponibles en los tiempos $t=0.25$, $t=0.75$, $t=1.5$, $t=2.5$, $t=3.75$ y $t=5$ días en las 36 ubicaciones indicadas con círculos llenos en la Fig. 4.1. De esta forma se generan 36 historias con seis respuestas de carga hidráulica. Estas variables indirectas e informativas de la conductividad hidráulica del acuífero se consideran observaciones piezométricas transitorias disponibles. El proceso de actualización tanto del esquema EnKF como del esquema pf-EnKF se efectúa en cada uno de los seis tiempos mencionados anteriormente; en los demás tiempos, sólo se efectúa la etapa de pronóstico. El coeficiente de almacenamiento se supone que es igual a 0.0001 en todo el acuífero.

4.2.3 Definición del ensamble de realizaciones *a priori*

El campo de referencia se muestrea en las ubicaciones marcadas con círculos vacíos en la Fig. 4.1. Tales valores se consideran mediciones directas de conductividad. La media y la varianza de estas muestras, así como la distribución de frecuencias, se reportan en la Fig. 4.3(a). En la Fig. 4.3(b) se muestran las funciones de correlación empíricas de *Spearman* de las muestras en las direcciones X , Y y de las muestras en todas direcciones. En esta figura se puede observar que es posible acomodar un modelo teórico común de correlación a todas las funciones, ya que el campo de referencia proviene de un campo aleatorio isótropo. Los parámetros estadísticos de media, varianza y función de correlación que se muestran en las Figs. 4.3(a),(b) se utilizan para parametrizar un campo aleatorio multi-gaussiano y generar dos mil realizaciones no condicionales con un simulador modificado de campos aleatorios SGSIM (Deutsch y Journel, 1996). Tal conjunto de realizaciones representa una aproximación inicial de la conductividad hidráulica del acuífero de referencia. De esta forma, se modela una situación práctica en la que los parámetros descriptivos del acuífero se estiman en forma aproximada *a priori*, en el sentido de que difieren de las estadísticas reales del campo de referencia (tabla 4.1). El número de realizaciones simuladas del campo *a priori* asegura la estabilidad de las medidas de error que se presentan a continuación, de acuerdo a algunas simulaciones preliminares.

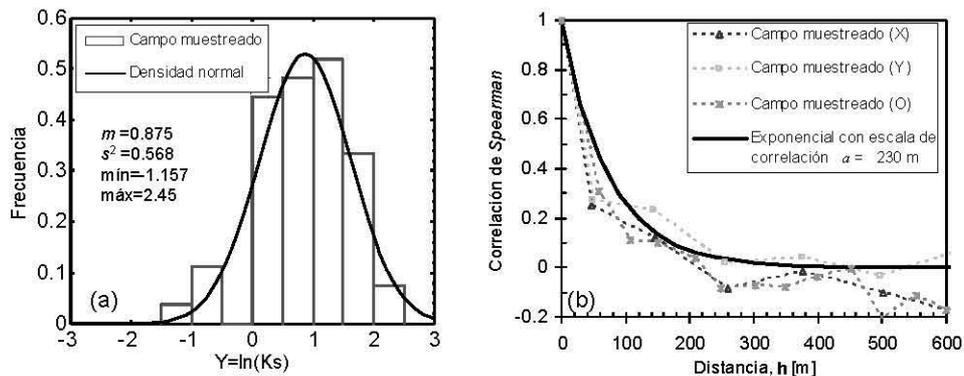


Fig. 4.3 Parámetros estadísticos de las muestras tomadas del campo de referencia. a) Histograma del logaritmo natural de la conductividad del campo muestreado. La línea sólida gruesa representa la distribución normal ajustada con la media y la varianza que se indican en la figura. b) Funciones de correlación empíricas del campo muestreado en las direcciones X y Y . También se indica la función empírica omni-direccional (O). La línea sólida gruesa representa el modelo teórico ajustado de auto-correlación correspondiente a la función exponencial con la escala de correlación a que se indica.

4.3 Criterios para evaluar el error

El promedio en cada ubicación en todo el conjunto de realizaciones condicionales de cualquiera de los filtros en un tiempo específico proporciona una estimación del valor esperado condicional del campo aleatorio en la ubicación espacio-temporal. Al mapa de tales valores se denomina *campo medio* del logaritmo natural de la conductividad hidráulica y puede ser visto como un mapa estimado de “mínima varianza” del campo aleatorio correspondiente.

Para evaluar cuantitativamente la exactitud de la estimación se consideran tres medidas de error: la raíz cuadrada del error cuadrático medio (RMSE), el error absoluto medio (MAE) y el error lineal en el espacio de probabilidades (LEPS).

El RMSE ($L1$) se evalúa con:

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

donde $y^{ref}(\mathbf{x}_i)$ es el logaritmo natural de la conductividad hidráulica de referencia en la ubicación \mathbf{x}_i ; $y^*(\mathbf{x}_i)$ es el promedio estimado del logaritmo natural de la conductividad hidráulica en la ubicación \mathbf{x}_i y n es el número de valores del logaritmo natural de la conductividad hidráulica en la región del flujo. El RMSE es, por lo tanto, una medida de la dispersión del campo estimado en torno al campo de referencia.

El MAE ($L2$) se calcula por medio de:

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

El MAE es una medida de la diferencia de las medias del campo estimado y el de referencia.

El LEPS ($L3$) se calcula mediante la fórmula (Bárdossy y Li, 2008):

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

donde $F(y^{ref}(\mathbf{x}_i))$ es el valor de la función de distribución empírica del campo de referencia asociada al logaritmo natural de la conductividad hidráulica en la ubicación \mathbf{x}_i ; $F(y^*(\mathbf{x}_i))$ es el valor de la función de distribución empírica del campo estimado de valores medios asociada al logaritmo natural de la conductividad hidráulica en la misma ubicación \mathbf{x}_i . A diferencia del RMSE y el MAE, el LEPS no se ve afectado por las unidades de las variables. Varía desde cero hasta uno, representando una estimación perfecta y una imperfecta, respectivamente.

Para evaluar la varianza de la estimación (precisión) se determinan tres medidas de error adicionales: la *SPREAD*, la desviación absoluta promedio (AAD) y la varianza total de la estimación (Var_{tot}) (promediada espacialmente).

La *SPREAD* ($L4$) se calcula como:

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

donde $s_{en}^2(\mathbf{x}_i)$ es la varianza de la estimación en la ubicación \mathbf{x}_i obtenida estadísticamente en todo el ensamble de realizaciones. *SPREAD* es por tanto una medida de la dispersión del campo estimado en torno al campo de referencia.

La *AAD* ($L5$) está definida como (Zhou *et al.*, 2011):

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

donde $m^*(\mathbf{x}_i)$ es el promedio de valores estimados en la ubicación \mathbf{x}_i que se obtiene estadísticamente en todo el ensamble de realizaciones y $y^*(\mathbf{x}_{ij})$ es el valor estimado del logaritmo natural de la conductividad hidráulica en la ubicación \mathbf{x}_i de la realización j .

La Var_{tot} (L6) se define como (Shöniger *et al.*, 2012):

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

El valor de Var_{tot} con frecuencia se interpreta como una medida de la información absorbida en la estimación espacial.

Como se mencionó anteriormente, los criterios L1 a L3 y L4 a L6 son medidas de la exactitud y la precisión locales de las estimaciones, respectivamente. En términos generales, entre menores sean los valores de tales normas, las estimaciones serán más exactas y precisas localmente. Sin embargo, estas medidas no proporcionan ninguna información acerca de la bondad de la estimación en términos de la reproducción de los patrones de variabilidad espacial en el campo de referencia. Se pueden analizar varias características de la variabilidad espacial (asimetría, entropía, conectividad, etc.), pero en este estudio el análisis se limita a evaluar la reproducción de las funciones de correlación empíricas de *Spearman* y de las cópulas empíricas bivariadas del campo de referencia.

Para evaluar la reproducción de las funciones de correlación de *Spearman* se determinan escalas de correlación:

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

donde $\rho_s(\mathbf{h})$ es la función de correlación empírica de *Spearman* asociada a un vector \mathbf{h} . Las escalas de correlación del promedio de las funciones de correlación de *Spearman* de todas las funciones de los campos condicionales se comparan con las escalas de correlación del campo de referencia. Entre menores sean las diferencias entre ambas escalas, más aproximada será la reproducción de la función de correlación de *Spearman*.

La función de correlación de *Spearman* es comparable al semivariograma o correlograma. Expresa el coeficiente de correlación de *Spearman* entre pares de valores (cópulas bivariadas) para varios vectores \mathbf{h} . La distancia donde la función llega a cero es, de hecho, la escala de correlación (rango) definida en geoestadística. Las funciones de correlación empíricas de *Spearman* se pueden calcular por medio de:

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

donde $R(\mathbf{x}_i)$ y $R(\mathbf{x}_j)$ son valores de las funciones de distribución empíricas asociadas al logaritmo natural de la conductividad hidráulica en las ubicaciones \mathbf{x}_i y \mathbf{x}_j , respectivamente; $\sigma_{R(\mathbf{x}_i)}$ y $\sigma_{R(\mathbf{x}_j)}$ son desviaciones estándar de dichos valores y $n(\mathbf{h})$ es el número de pares de valores separados aproximadamente por el vector \mathbf{h} .

Para verificar la reproducción de las distribuciones bivariadas se calculan las distancias de Kolmogorov:

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

donde $\hat{C}(u_1, u_2)$ es la cópula empírica media de todas las cópulas de los campos condicionales y $C(u_1, u_2)$ es la cópula del campo de referencia. Estas cópulas se determinan para diferentes vectores \mathbf{h} .

4.4 Resultados y discusiones

Esta sección muestra que vale la pena tomar en cuenta las observaciones de la carga hidráulica del agua subterránea en modelos de incertidumbre *a priori* para revelar información acerca de la conductividad hidráulica de su campo fuente. Este hecho se ilustra con los resultados de la evolución de las estimaciones de la media condicional y de su incertidumbre, así como con los resultados de la reproducción de la variabilidad espacial del denominado campo de referencia, una vez que se completa el proceso de condicionamiento ($t=5$ días).

4.4.1 Estimaciones de la media condicional y de su incertidumbre

En el experimento que se estableció en la sección anterior, al final de cada etapa de actualización, se obtiene con cada filtro, un ensamble que consta de dos mil realizaciones del logaritmo natural de la conductividad hidráulica condicionadas a las cargas hidráulicas. Cada conjunto de realizaciones se utiliza por separado para calcular la media y la desviación estándar (incertidumbre) condicionales en cada ubicación.

La Fig. 4.4 muestra la gráfica de los campos medios del logaritmo natural de la conductividad hidráulica estimados con las realizaciones condicionales del EnKF y del pf-EnKF después de integrar las observaciones de las cargas de agua subterránea hasta los tiempos $t=0.75$ días y $t=2.5$ días. Con fines comparativos, también se ilustra el campo medio del logaritmo natural de la conductividad hidráulica que se obtiene con el ensamble de realizaciones *a priori*.

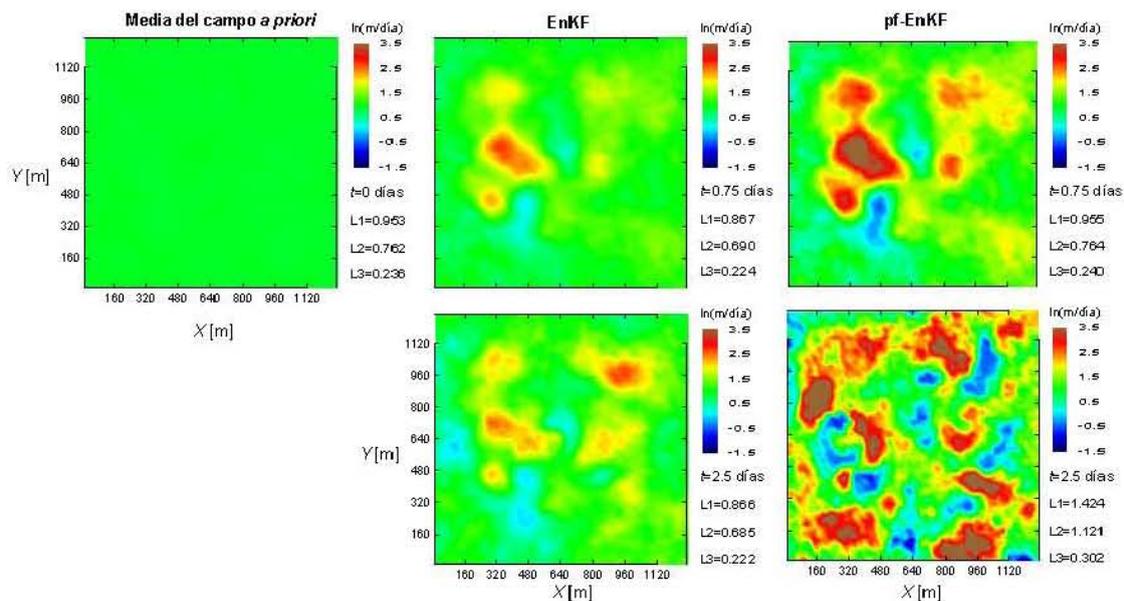


Fig. 4.4 La media del campo del logaritmo natural de la conductividad estimado con el método EnKF (al centro) y con el método pf-EnKF (derecha) en los tiempos $t=0.75$ días y $t=2.5$ días. Se muestra además la media del campo *a priori* del logaritmo natural de la conductividad con fines comparativos (izquierda). Los valores numéricos de las medidas de exactitud local de la estimación se indican a la derecha de los campos estimados. $L1=RMSE$ (raíz cuadrada del error cuadrático medio), $L2=MAE$ (valor absoluto del error medio) y $L3=LEPS$ (error lineal en el espacio de las probabilidades).

Debido a que el campo aleatorio *a priori* no es condicional, los valores locales promediados con todo el ensamble de realizaciones se aproximan a la esperanza del campo aleatorio *a priori*, en consecuencia, el campo medio *a priori* parece ser homogéneo. Sin embargo, la información proporcionada por las respuestas de cargas de agua subterránea modifica gradualmente la distribución *a priori* del logaritmo natural de las conductividades hidráulicas. Observe que los valores de todas las medidas de error reportados a la derecha de los campos medios del EnKF indican que tales campos se vuelven más próximos al campo de referencia que el campo medio *a priori*. Esto no ocurre con el campo medio del pf-EnKF. De hecho, las normas indican mayores errores que los errores *a priori*. No obstante, se mostrará más adelante que las estimaciones del pf-EnKF reproducen mejor los parámetros estadísticos del campo de referencia que las estimaciones *a priori*, a pesar del comportamiento de las normas que se reportan en la figura 4.4.

Una vez completado el proceso de condicionamiento, es decir, después de integrar las seis respuestas de cargas de agua subterránea de cada piezómetro, las medias de los campos condicionales del logaritmo natural de la conductividad hidráulica alcanzan los patrones finales que se muestran en la Fig. 4.5. Es evidente que la distribución de las conductividades observadas en los tiempos anteriores (Fig. 4.4) presenta una evolución y se revelan más detalles del campo de referencia. Las medidas de exactitud local reportadas a la derecha de los campos indican que la estimación del EnKF es más exacta que la del pf-EnKF. Sin embargo, la distribución de las conductividades del segundo parece estar más próxima a la estructura espacial del campo de referencia. Esto se evaluará cuantitativamente más adelante.

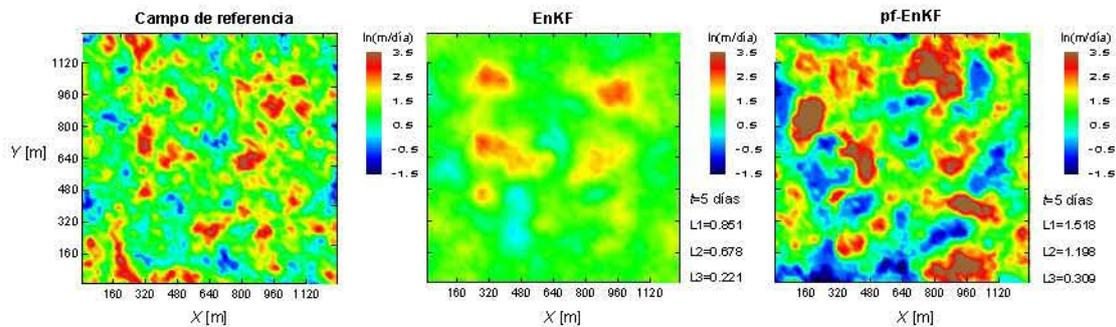


Fig. 4.5 La media del campo del logaritmo natural de la conductividad estimado con el método EnKF (al centro) y con el método pf-EnKF (derecha) en el tiempo $t=5$ días. El campo de referencia se muestra con fines comparativos (izquierda). Los valores numéricos de las medidas de exactitud local de la estimación se indican a la derecha de los campos estimados. El significado de $L1$, $L2$ y $L3$ se explica en la Fig. 4.4.

A medida que las respuestas de las cargas de agua subterránea se integran en el proceso de condicionamiento, también ocurre una evolución de las varianzas locales *a priori*. La Fig. 4.6 muestra la evolución de la desviación estándar (incertidumbre) del campo del logaritmo natural de las conductividades hidráulicas. En las imágenes se puede ver que la desviación estándar global decrece con el tiempo. Observando las medidas de incertidumbre reportadas a la derecha de cada imagen, se encuentra que esta reducción es más significativa en las estimaciones del pf-EnKF que en las estimaciones del EnKF. Por lo tanto, el pf-EnKF es más eficiente que el EnKF, pero esta eficiencia no corresponde con una mayor exactitud local, como se mencionó anteriormente. La evolución de la desviación estándar en el caso general dependerá de varios factores: error en la aproximación inicial, número y ubicación de piezómetros, condiciones de frontera, etc. y el pf-EnKF parece proporcionar estimaciones más precisas, aunque dichas estimaciones no necesariamente serán más exactas localmente.

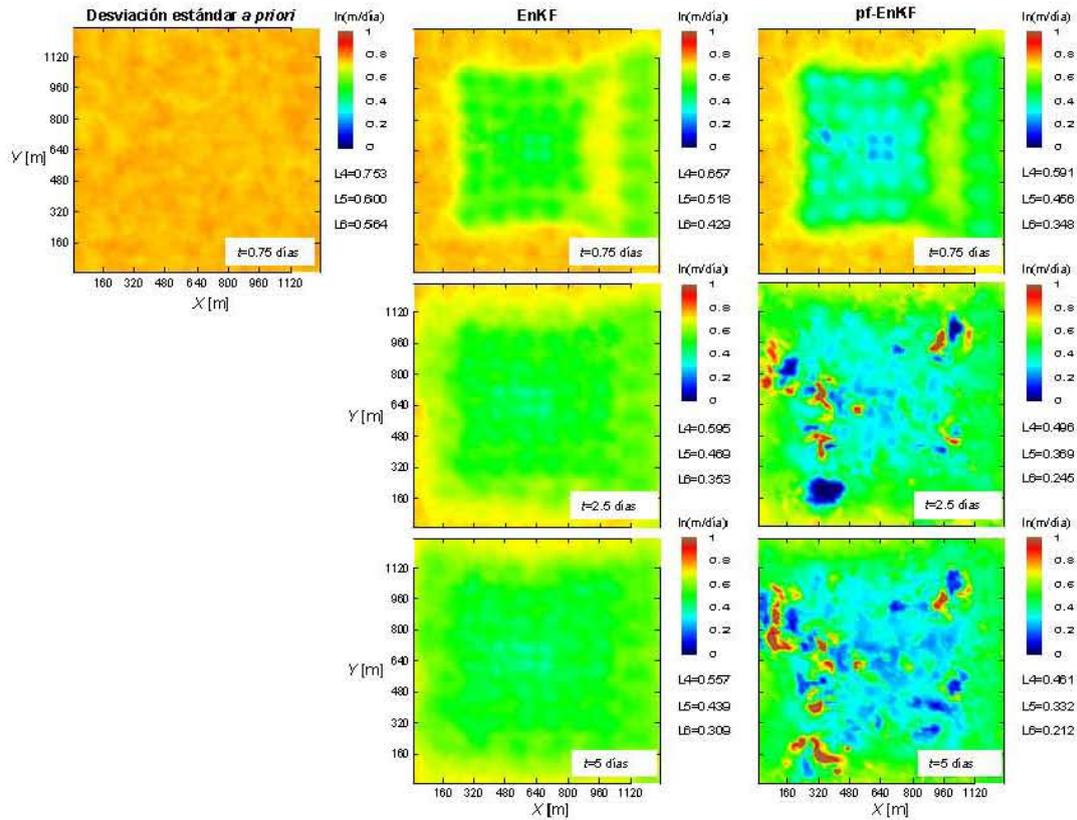


Fig. 4.6 Evolución de la desviación estándar de la estimación del logaritmo natural de la conductividad en los tiempos $t=0.75$ días, $t=2.5$ días y $t=5$ días calculada con las estimaciones del EnKF (izquierda) y pf-EnKF (derecha). Las mediciones de la incertidumbre de la estimación se reportan a la derecha de cada imagen. $L1=SPREAD$, $L2=AAD$ (Valor absoluto de la desviación media) y $L4=Var_{tot}$ (Varianza total de la predicción).

Las Figs. 4.7(a),(b) muestran las distribuciones de frecuencias del logaritmo natural de las conductividades hidráulicas del campo de referencia, del ensamble *a priori* y de los ensambles finales *a posteriori* de los campos del EnKF y pf-EnKF, respectivamente. La media y la varianza globales de cada conjunto de realizaciones se utilizan para ajustar una función de distribución normal teórica a cada distribución de frecuencias. Observe que la función de distribución *a priori* se localiza a la izquierda de la función de distribución del campo de referencia, indicando que la función de distribución *a priori* subestima la media global del logaritmo natural de la conductividad hidráulica del campo de referencia. También se observa que las funciones de distribución *a posteriori* están desplazadas hacia la derecha de la función de distribución *a priori*. Esto indica el intento del proceso de condicionamiento por llevar la función de distribución *a priori* hacia la función de distribución del campo de referencia. Vale la pena mencionar que la función de distribución *a posteriori* del pf-EnKF abarca mejor las frecuencias de valores extremos del campo de referencia, pero su varianza es mayor que la varianza del campo de referencia.

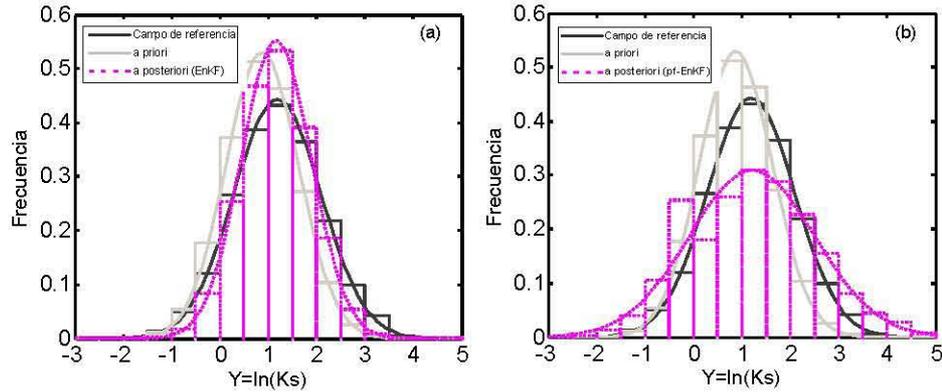


Fig. 4.7 Distribución de frecuencias del logaritmo natural de la conductividad. a) Del campo de referencia, del conjunto de realizaciones *a priori* y del conjunto de realizaciones *a posteriori* del EnKF al tiempo $t=5$ días. b) Del campo de referencia, del conjunto de realizaciones *a priori* y del conjunto de realizaciones *a posteriori* del pf-EnKF al tiempo $t=5$ días.

4.4.2 Reproducción de la variabilidad espacial

Las funciones de correlación empíricas de *Spearman* (por rangos) de un conjunto de realizaciones condicionales al tiempo $t=5$ días se comparan contra las del campo de referencia para la dirección X en la Fig. 4.8(a) y para la dirección Y en la Fig. 4.8(b). El promedio de las funciones de correlación de *Spearman* de las dos mil realizaciones condicionales del EnKF y del pf-EnKF también se ilustran en las figuras. El promedio de las funciones de correlación de ambos filtros exhiben una correlación más fuerte que la correlación del campo de referencia en ambas direcciones, particularmente en distancias cortas. Esto puede explicarse por el desempeño de la técnica de interpolación lineal de ambos filtros. En efecto, esta técnica utiliza una sola función de autocovarianza como único descriptor de variabilidad espacial, optimizando así la interpolación en torno a los valores medios, pero sin tomar en cuenta la interpolación en percentiles específicos, haciendo que la interpolación sea más burda. No obstante, los promedios de las funciones de correlación de las realizaciones condicionales de ambos filtros se aproximan mejor a la función de correlación del campo de referencia que el promedio de las funciones de correlación del campo *a priori*. Por lo tanto, las respuestas de cargas de agua subterránea ciertamente ayudan a identificar la variabilidad espacial de su campo fuente.

Ahora se examinarán los resultados de la reproducción de las cópulas empíricas del campo de referencia. En la Fig. 4.9 se muestran las gráficas de densidades de cópulas empíricas (ECD) para una distancia de 20 m en la dirección X . La ECD del campo de referencia se muestra con fines comparativos en la Fig. 4.9(a). Las Figs. 4.9(b),(c) son ECD de los campos medios del EnKF y del pf-EnKF, respectivamente. Observe que las mayores densidades de estas cópulas corresponden bien con las mayores densidades de la ECD del campo de referencia, aunque las ECD de los campos medios son más estrechas. Esto indica que la correlación de los campos medios es más fuerte que la correlación del campo de referencia en esa distancia. La correlación del campo medio del pf-EnKF es ligeramente más fuerte que la del campo medio del EnKF, pero la forma de su cópula se asemeja mejor a la densidad de la cópula del campo de referencia. La mejor reproducción de esta densidad de la cópula se confirmará cuantitativamente más adelante. Las Figs. 4.9(d),(e) ilustran la ECD media de un ensemble de dos mil densidades de cópulas, cada una de las cuales se ha establecido con los valores de cada realización condicional del EnKF y pf-EnKF, respectivamente. Es evidente que la forma de esta ECD está bastante mejor definida que la ECD de los campos medios. Observe que estas graficas hacen más evidente lo que se ha mencionado con anterioridad en este párrafo.

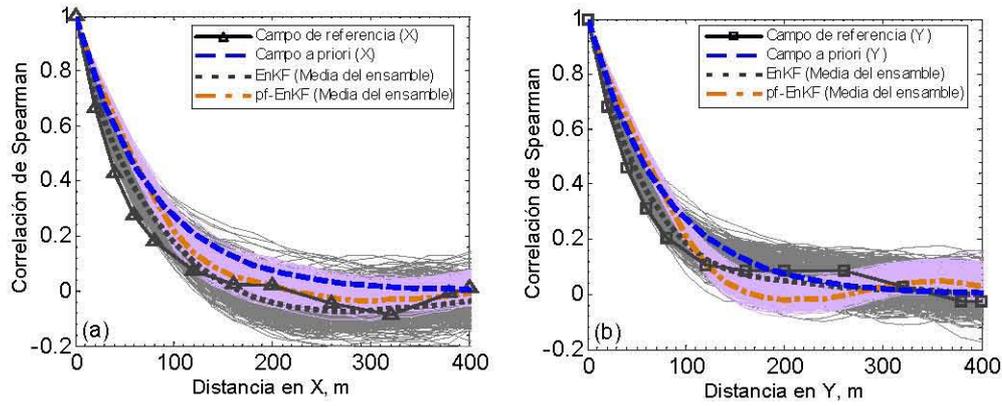


Fig. 4.8 Funciones de correlación empíricas del campo de referencia (líneas con símbolos), de un conjunto de realizaciones condicionales al tiempo $t=5$ días (líneas delgadas) y del promedio de dos mil funciones de correlación al tiempo $t=5$ días (líneas gruesas). a) En la dirección X . b) En la dirección Y .

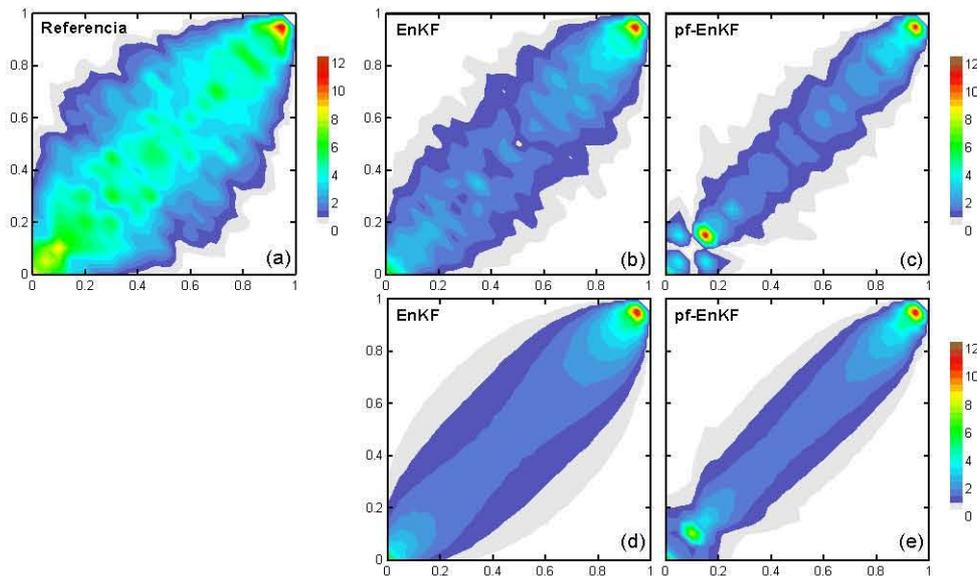


Fig. 4.9 Densidades de la cópula bivariada empírica a una distancia de 20 m en la dirección X . a) Del campo de referencia. b) y c) De la media del campo estimado al tiempo $t=5$ días. d) y e) Del promedio de dos mil densidades de cópulas que se obtienen de las realizaciones condicionales al tiempo $t=5$ días.

La Fig. 4.10 muestra las comparaciones entre las cópulas del campo de referencia y las cópulas medias de un ensemble de dos mil cópulas. Las comparaciones son representativas de las cópulas en la dirección X para distancias de 20 m, 40 m y 80 m. Para evaluar cuantitativamente la bondad de la reproducción de las cópulas del campo de referencia, el valor de la distancia de Kolmogorov ($D1$) se indica en la esquina inferior izquierda de cada figura. Con base en estos valores numéricos, se puede decir que los campos del pf-EnKF reproducen mejor las cópulas del campo de referencia que los campos del EnKF. Por lo tanto, la estructura espacial bivariada del campo de referencia parece ser mejor captada por el pf-EnKF.

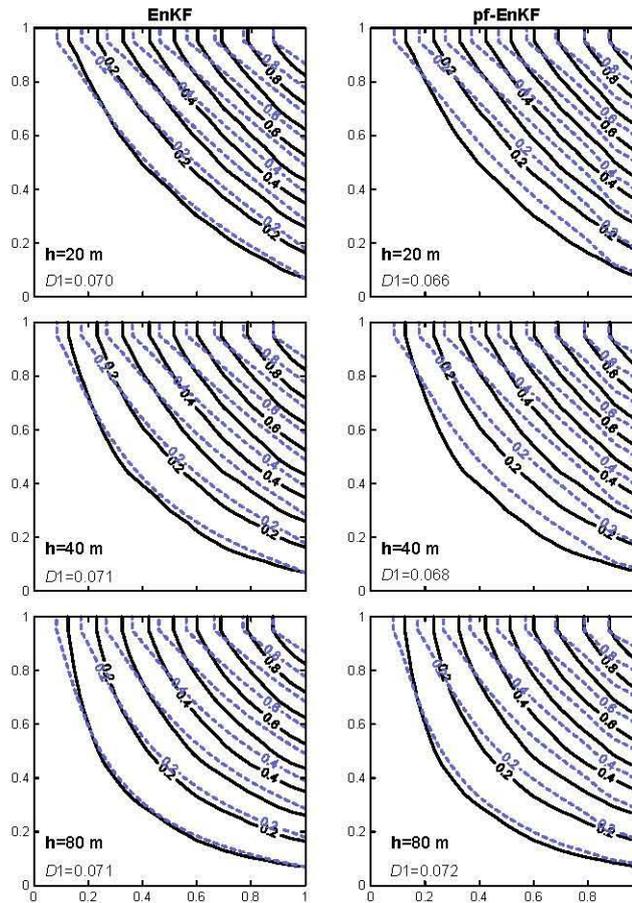


Fig. 4.10 Cópulas empíricas bivariadas para tres diferentes distancias en la dirección X al tiempo $t=5$ días. En la primera columna se muestran las cópulas medias de las realizaciones condicionales del EnKF. En la segunda columna se reportan las cópulas medias de las realizaciones condicionales del pf-EnKF. Los valores en la parte inferior izquierda de cada figura son las distancias de Kolmogorov ($D1$).

En la Fig. 4.11(a) se muestran los valores de los criterios $D1$ entre las cópulas del campo de referencia y las cópulas medias de un ensemble de dos mil cópulas con respecto a varias distancias en la dirección X . En la Fig. 4.11(b) se muestran los mismos resultados, pero ahora de las cópulas en la dirección Y . Es claro que los campos del pf-EnKF reproducen mejor las cópulas del campo de referencia en distancias cortas en la dirección X y en todas las distancias en la dirección Y . En distancias largas en la dirección X , las distribuciones bivariadas del campo de referencia se aproximan mejor mediante los campos del EnKF.

4.4.2 Resumen de los resultados

Finalmente, los parámetros estadísticos de los campos estimados se comparan con los del campo de referencia en la tabla 4.1. En términos generales, la media global del campo de referencia es captada bastante bien por ambos filtros, aunque su varianza es subestimada por el EnKF y sobreestimada por el pf-EnKF. Las escalas de correlación del campo de referencia son sobreestimadas por ambos filtros. Observe en la tabla 4.1 que ninguno de los dos filtros se desempeña mejor de manera consistente, pero ambos proporcionan mejores estimaciones de la estructura espacial del campo de referencia que las estimaciones *a priori*. Por lo tanto, parece ser conveniente considerar ambos filtros en las aplicaciones.

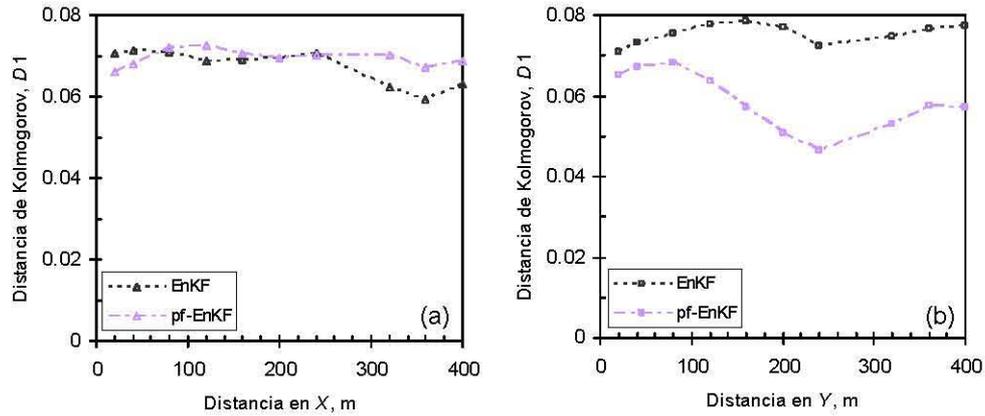


Fig. 4.11 Promedio de la distancia de Kolmogorov ($D1$) calculada con dos mil cópulas obtenidas de las realizaciones condicionales del logaritmo natural de la conductividad hidráulica al tiempo $t=5$ días para diez diferentes separaciones. a) En la dirección X . b) En la dirección Y .

Tabla 4.1 Comparación entre los parámetros estadísticos de los campos de referencia y de los campos *a priori* y estimados después de completar el proceso de condicionamiento ($t=5$ días).

	Campo de referencia	Campos <i>a priori</i>	Campos estimados (EnKF)	Campos estimados (pf-EnKF)
Media, $\ln(Ks(m/día))$	1.19	0.87	1.16	1.20
Desv. estándar, $\ln(Ks(m/día))$	0.90	0.75	0.72	1.29
Escala de correlación en dir. X , m	147.5	230.0	162.1	196.5
Escala de correlación en dir. Y , m	190.0	230.0	188.0	188.2

4.5 Conclusiones

En este capítulo se ilustraron los beneficios de la integración de respuestas de cargas de agua subterránea en un modelo de incertidumbre *a priori* para caracterizar acuíferos no homogéneos con respecto a su conductividad hidráulica utilizando el método EnKF y el método propuesto pf-EnKF presentados en el capítulo 3. Para tales fines ilustrativos se consideró un acuífero hipotético, llamado campo de referencia, en el que la distribución de conductividades se generó por medio de la simulación de un campo aleatorio con la cópula normal V -transformada. Se supuso que el modelo de incertidumbre *a priori* es multi-gaussiano y sus parámetros descriptivos se definieron por medio de un muestreo del campo de referencia. Las respuestas de las cargas de agua subterránea se generaron a partir de la solución de un problema de flujo de agua transitorio con las permeabilidades del campo de referencia. Un conjunto de historias de cargas registradas en ubicaciones especificadas fue integrado con el proceso de condicionamiento de ambos filtros.

Se evaluó la calidad de la reproducción de las estadísticas univariadas y multivariadas del campo de referencia con diferentes criterios cuantitativos. Se encontró que ninguno de los dos filtros se desempeñaba mejor de manera consistente, pero los resultados mostrados por ellos demostraron que las respuestas de cargas de agua subterránea fueron capaces de revelar información acerca de la variabilidad espacial de la distribución de conductividades en el campo de referencia. El proceso de condicionamiento de ambos filtros mejoró las estimaciones *a priori*.

El campo de referencia utilizado en este estudio fue, en efecto, sólo una idealización de un acuífero real que seguramente será más complejo y además tridimensional; no obstante, las técnicas numéricas que se presentaron ofrecerán aproximaciones útiles en situaciones prácticas.

CAPÍTULO 5

DETECCIÓN DE TRAYECTORIAS PREFERENCIALES DE FILTRACIÓN EN EL NÚCLEO DE UNA PRESA TRIANGULAR

Filtración preferencial es un término utilizado para describir movimientos concentrados de un fluido dentro de la región de flujo a velocidades de filtración contrastantemente elevadas. Este tipo de filtraciones ocurren en medios heterogéneos porque el fluido se mueve más rápidamente a lo largo de trayectorias con menor resistencia. Los caminos preferenciales menos resistentes están asociados a zonas continuas con conductividades hidráulicas más altas que las de zonas circundantes. En las estructuras térreas la presencia de zonas continuas con conductividades altas tiene diferentes orígenes (ICOLD, 1995, 1997, 2004; Fell y Wan, 2005). Pueden ser el resultado de alguno o algunos de los siguientes eventos: agrietamiento del suelo, erosión interna en desarrollo y zonas permeables continuas que contienen materiales gruesos y/o pobremente compactados, por mencionar algunos.

Para observar trayectorias preferenciales de filtración en el núcleo de una presa de tierra, en principio basta caracterizar la distribución de conductividades y analizar velocidades de filtración para una condición de flujo específica. Sin embargo, este no es un problema trivial porque, en la práctica, sólo se dispone de un número reducido de mediciones directas de la conductividad hidráulica. Además, la distribución de conductividades en terraplenes no depende únicamente de las propiedades intrínsecas de los materiales y de la calidad de compactación, también de eventos como los mencionados anteriormente.

Para caracterizar campos de conductividades hidráulicas, su variabilidad espacial se interpreta comúnmente como una realización de un campo aleatorio (Dagan, 1989; Gelhar, 1993; Zhang, 2002; Rubin, 2003). La descripción del campo aleatorio se obtiene con base en las observaciones de la realización misma suponiendo que éste es *ergódico* (Deutsch y Journel, 1992; Chilés y Delfiner, 1999). Posteriormente, estimaciones y la incertidumbre asociada a éstas en las ubicaciones no observadas se obtienen por medio de técnicas geoestadísticas, ya sean de estimación o de simulación (p. ej. Journel y Huijbregts, 1978; Deutsch y Journel, 1992; Chilés y Delfiner, 1999).

El modelo de campo aleatorio multi-gaussiano ha sido ampliamente utilizado para representar variaciones espaciales de la conductividad hidráulica de los suelos (Fenton y Griffiths, 1996, 1997; Gui *et al.*, 2000; López-Acosta y Auvinet, 2003, 2004; Ahmed, 2009). En la práctica, este modelo se justifica comúnmente sólo con base en una distribución univariada lognormal de conductividades (p. ej. Freeze, 1975; Hoeksema y Kitanidis, 1985; Fogg, 1986; Woodbury y Zudicky, 1986; Gelhar, 1993; López-Acosta y Auvinet, 2011). Las densidades de mayor orden, esto es, la bi-variada, tri-variada, n -variada son aceptadas “a siegas”. En otras palabras, la decisión de considerar un modelo multi-gaussiano no está basada en los datos, simplemente tiene la ventaja de su simplicidad matemática y facilidad de inferencia.

El fenómeno de trayectorias preferenciales de filtración sugiere que la variabilidad espacial de conductividades posee características de dependencia no multi-gaussiana (cap. 1). En efecto, se ha reconocido que en campos aleatorios multi-gaussianos las trayectorias más continuas tienen conductividades alrededor de la media (Silliman y Wright, 1988). Esto muestra que en los campos aleatorios multi-gaussianos no ocurren trayectorias de valores extremos conectados (Journel y Alabert, 1989; Journel y Deutsch, 1993; Gómez-Hernández y Wen, 1998; Journel y Zhang, 2006). Sin embargo, las trayectorias preferenciales de filtración son el resultado de la manera en que las conductividades altas y bajas están espacialmente interconectadas en la región de flujo (Fogg, 1986; Bradbury y Muldoon, 1990; Webb y Anderson, 1996; Teles *et al.*, 2004; Zappa *et al.*, 2006; Sánchez-Vila *et al.*, 2006). Por lo tanto, para representar satisfactoriamente la continuidad de los valores extremos en los campos de conductividades hidráulicas se deben considerar campos aleatorios con dependencia no multi-gaussiana.

Campos aleatorios de conductividades con características de dependencia no multi-gaussiana tales como una continuidad mayor para los valores grandes y/o pequeños, se pueden generar utilizando diferentes enfoques (Journel y Huijbregts, 1978; Journel, 1983; Yamazaki y Shinozuka, 1988; Christakos, 1990; Grigoriu, 1998; Sánchez-Vila *et al.*, 1996; Popescu *et al.*, 1998; Chilès y Delfiner, 1999; Emery, 2008). En la práctica, la descripción de estos modelos no siempre es una tarea sencilla debido a la escasez o inexistencia de mediciones directas de la conductividad. Por lo tanto, es necesario considerar modelos multi-gaussianos *a priori* la mayoría de las veces. Estos modelos sin embargo se pueden mejorar integrando observaciones indirectas de la conductividad hidráulica.

En efecto, las observaciones de la carga hidráulica se pueden relacionar con la distribución real de conductividades resolviendo el problema inverso. Una solución inversa requiere que el campo de cargas hidráulicas que se obtiene del modelo de flujo de agua coincida con las observaciones de la carga hidráulica misma en las ubicaciones observadas. En la literatura científica se encuentran diferentes técnicas de inversión pero la simulación estocástica de campos aleatorios de conductividades hidráulicas condicionales a variables dinámicas tales como las cargas piezométricas ha recibido una atención considerable (p. ej. Gómez-Hernández *et al.*, 1997; Zhu y Yeh, 2005; Alcolea *et al.*, 2006; Capilla y Llopis-Albert, 2009).

El propósito de este capítulo es caracterizar un campo no multi-gaussiano que simula zonas continuas de conductividades hidráulicas altas dentro de una presa de tierra hipotética. La caracterización de este tipo de campos es de gran interés en la práctica porque permitiría detectar trayectorias preferenciales de filtración por inspección visual del campo asociado de velocidades. Para este propósito se obtiene una solución al problema estocástico inverso por medio del filtro de Kalman Ensamblado modificado (EnKF) presentado en el capítulo 3. Los resultados se comparan con los que se obtienen con el filtro propuesto (pf-EnKF). En los ejemplos presentados en este capítulo, la distribución de conductividades *a priori* en la región de flujo se representa por medio de un conjunto de realizaciones multi-gaussianas no condicionales. De esta manera se modelará una situación tal que la presencia de zonas continuas con conductividades altas se ignora *a priori* y la única fuente de información acerca de la distribución real de conductividades será proporcionada únicamente por historias de cargas piezométricas.

La metodología se ilustra con un campo sintético de referencia que describe la distribución del logaritmo natural de la conductividad hidráulica saturada en el núcleo de una presa triangular (Harr, 1962; Poluvarinova-Kochina, 1962; Bear *et al.*, 1968). Se supone que dicho campo representa un “estado de la naturaleza” en el que son evidentes las zonas continuas con conductividades altas. Para generar historias de cargas hidráulicas de referencia se resuelve una condición de flujo transitorio simulando un abatimiento “rápido” del embalse y después se obtienen muestras del campo de cargas hidráulicas en ubicaciones seleccionadas con intervalos de tiempo especificados. El conjunto de realizaciones de la conductividad condicionales a las observaciones de la carga hidráulica se utilizan para pronosticar campos de velocidades. Las trayectorias preferenciales de filtración se identifican por inspección visual de estos campos. El error de las predicciones también será analizado.

El capítulo se organiza como sigue. El problema se expone en la sección 5.1; los campos de referencia tanto de conductividades como de cargas hidráulicas se describen en la sección 5.2 y los métodos de inversión estocástica se presentan en la sección 5.3; los resultados se discuten en la sección 5.4, antes de concluir el estudio en la sección 5.5.

5.1 Exposición del problema

El estudio se realiza en el núcleo interno de una presa de enrocamiento que se idealiza con una sección transversal triangular. Se supone que el nivel de agua coincide con la cresta como en Harr (1962) (Fig. 5.1). La filtración en secciones triangulares ha sido considerada por varios autores (Poluvarinova-Kochina, 1962; Bear *et al.*, 1968; Collins, 1971). En estos casos sólo se analiza la filtración en el núcleo interno, debido a que las pérdidas de carga hidráulica en los taludes de roca son despreciables para fines prácticos.

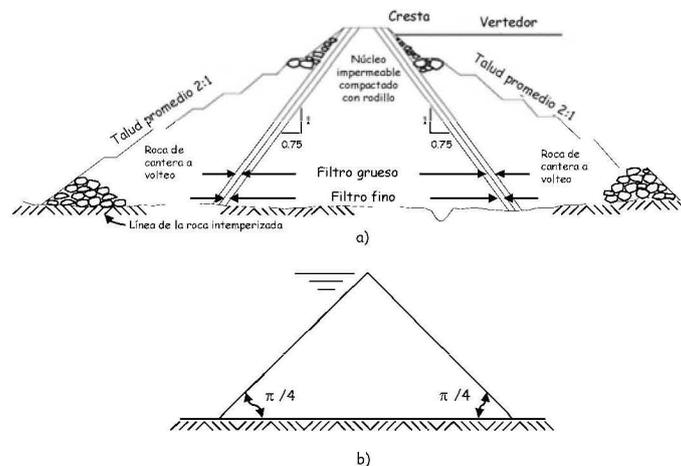


Fig. 5.1 Presa de enrocamiento y la idealización de la sección transversal de su núcleo interno (Harr, 1962). Presa de enrocamiento. b) Idealización.

En el siguiente análisis, se supone que el flujo de agua a través del núcleo de la presa está descrito por la siguiente ecuación de continuidad y ley de Darcy (p. ej. Bear, 1972; Freeze y Cherry, 1979; de Marsily, 1986):

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

Sujeta a:

$$\text{Condición inicial; } H_{t=0}=h_0 \quad (5.2)$$

y condiciones de frontera:

$$\text{Carga hidráulica constante; } H_{\Gamma_D} = h_D \quad (5.3)$$

En la frontera impermeable se cumple;

$$\mathbf{q}(\mathbf{x}, t) \cdot \mathbf{n}(\mathbf{x}) = 0 \quad \mathbf{x} \in \Gamma_N \quad (5.4)$$

$$\text{En la superficie de filtración; } H_{\Gamma_D} = z_D \quad (5.5)$$

donde: H es la carga hidráulica [L], h_0 es la carga inicial en el dominio Ω [L], h_D es la carga prescrita en los segmentos de frontera de Dirichlet Γ_D [L]; $K_s(\mathbf{x})$ es la conductividad hidráulica [L/T], $\mathbf{q}(\mathbf{x}, t)$ es la descarga específica a través de los segmentos de frontera de Neumann Γ_N , $\mathbf{n}(\mathbf{x})$ es un vector normal saliente perpendicular a la frontera Γ_N y S_s es el almacenamiento específico [L⁻¹]. En este estudio, el logaritmo natural de la conductividad hidráulica se considera un campo aleatorio, mientras que el almacenamiento específico se trata como una constante determinista. La solución numérica del problema de filtración descrito por las ecuaciones 5.1-5.5 se obtiene por medio de elementos finitos (Istok, 1989; Smith y Griffiths, 2004).

5.2 Descripción del experimento numérico

5.2.1 Campo de log-conductividades hidráulicas

Se supone que la distribución del logaritmo natural de la conductividad hidráulica saturada en el núcleo $Y(\mathbf{x})=\ln(K_s(\mathbf{x}))$ se puede describir por medio de un campo aleatorio no multi-gaussiano con funciones de distribución marginales gaussianas. La media de dicho campo aleatorio es -1.193 y su varianza es unitaria. Por lo tanto, la distribución de conductividades saturadas tiene una media de $K_s=0.5$ m/día y un coeficiente de variación $CV=131\%$. Se supone además que el campo aleatorio es estrictamente estacionario caracterizado por una función de correlación de *Spearman* dada por un correlograma esférico con escala de correlación isótropa $\alpha=15$ m (Deutsch y Journel, 1992; Genest y Favre, 2007). Además, se considera que el campo aleatorio presenta una estructura de dependencia asimétrica que se puede describir por medio de la cópula *V*-transformada (Bárdossy y Li, 2008; Li y Bárdossy, 2009) con parámetros $m=0.0$, $k=2.5$ y $\alpha=0.25$.

Con base en la información mencionada anteriormente, la cópula *V*-transformada se utiliza para generar un conjunto de realizaciones no condicionales del logaritmo natural de la conductividad sobre un dominio cuadrado de 102.5 m por lado dividido uniformemente con 41x41 elementos cuadrados. Se selecciona una realización representativa de dicho conjunto y los valores sobre el núcleo de la presa se asignan a la malla de elementos finitos suponiendo que cada valor permanece constante dentro de su elemento. Las características de la malla de elementos finitos se explican más adelante. Se supone que esta realización representa un “estado de la naturaleza” en el que están presentes zonas continuas con conductividades altas en el núcleo. Otra realización del conjunto mostraría en efecto una distribución diferente del logaritmo natural de la conductividad pero preservaría las mismas características de la estructura de variabilidad espacial.

La realización representativa y sus descriptores estadísticos se muestran en la Fig. 5.2. Para cuantificar la asimetría del campo se considera una definición de la correlación de excedencia basada en el coeficiente *tau* de *Kendall* (Manner, 2010). La línea sólida en la Fig. 5.2(d)

representa la asimetría de la dependencia espacial. Las zonas más continuas y más permeables en la realización tienen conductividades $K_s(x)$ alrededor de 20 veces más altas que el valor medio. Esta característica es consistente con un estudio reciente de la distribución espacial de conductividades en el núcleo de una presa de tierra real en donde se ha detectado filtración preferencial (Smith y Konrad, 2011). Además, se cree que la escala de correlación especificada es razonable porque dentro de las prácticas de construcción actuales no se espera una continuidad importante del logaritmo natural de la conductividad en la sección transversal del núcleo, ya que el material se compacta en capas a lo largo de la sección longitudinal.

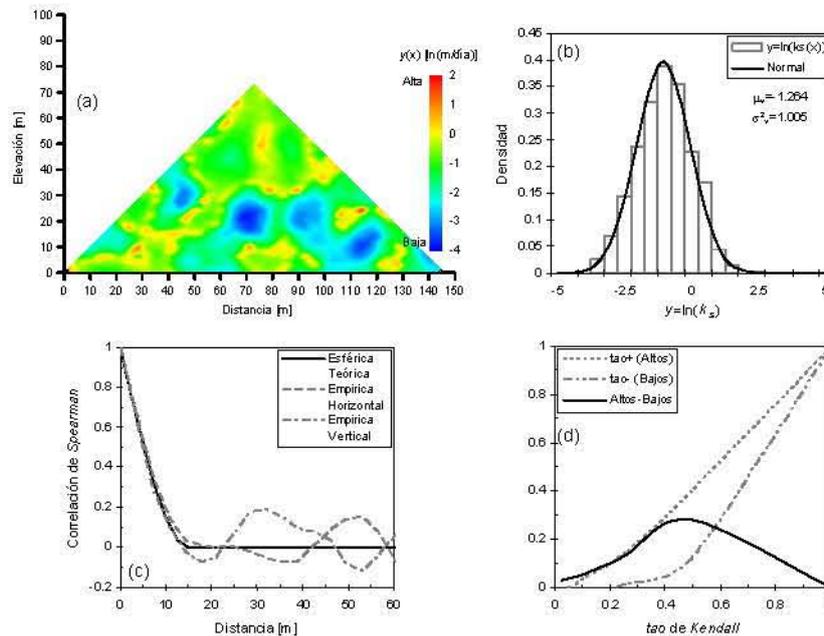


Fig. 5.2 Realización del campo aleatorio del logaritmo natural de la conductividad hidráulica generada por medio de la V -transformación con parámetros $m=0.0$, $k=2.5$ y $\alpha=0.25$. Se muestran además sus descriptores estadísticos. a) Realización de log-conductividades $Y(x)=\ln(K_s(x))$. b) Histograma. c) Funciones de correlación en las direcciones horizontal y vertical. d) Funciones de correlación condicional de Kendall (Manner, 2010) en la dirección horizontal.

5.2.2 Historias de cargas hidráulicas

Para generar historias de cargas hidráulicas, se resuelve una condición de flujo transitorio en el núcleo prescribiendo las cargas hidráulicas en los taludes de aguas arriba y aguas abajo y suponiendo una frontera impermeable en la base. El núcleo de la presa se subdivide en 861 elementos finitos isoparamétricos de 2.5×2.5 m por lado (Fig. 5.3). A estos se les asigna un valor del campo del logaritmo natural de la conductividad de la Fig. 5.2(a), como se ha explicado. La condición de flujo transitorio se modela de la siguiente manera. En $t=0$ días, la carga inicial está en régimen establecido (Fig. 5.4(a)). En $t>0$ días, se prescribe un abatimiento en el talud de aguas arriba a razón constante de 0.81 m/día hasta que alcanza la elevación mostrada en la Fig. 5.3. Esta condición de flujo transitorio se modela suponiendo la saturación completa del núcleo durante el abatimiento del embalse (abatimiento “rápido”) (Fig. 5.4(b)).

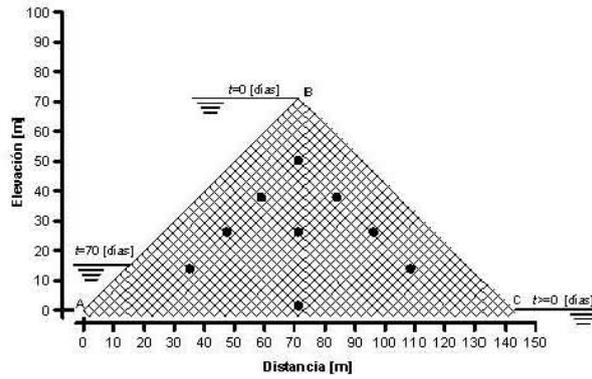


Fig. 5.3 La región de flujo discretizada y las ubicaciones de nueve piezómetros (círculos rellenos).

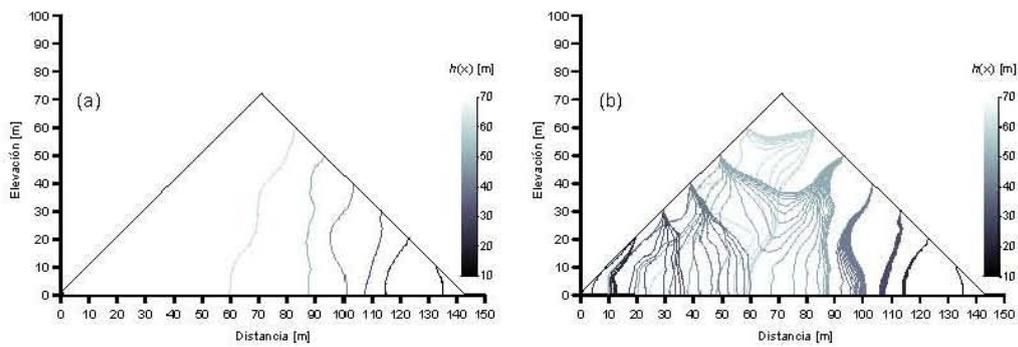


Fig. 5.4 Campos de referencia de la carga hidráulica [m]. a) En flujo establecido. b) En flujo transitorio después de un “vaciado rápido” del embalse, es decir, suponiendo que el material permanece saturado.

El abatimiento prescrito se alcanza en $t=70$ días. Este periodo se selecciona como la duración de la simulación y se subdivide en 40 tiempos de 1.75 días cada uno. Las cargas piezométricas se observan en las nueve ubicaciones mostradas en la Fig. 5.3, en los tiempos $t=t_1$, $t=t_2$ y de ahí en adelante en intervalos de tiempo igualmente espaciados de 3.5 días. Por lo tanto, cada conjunto de datos considerado en la inversión estocástica consiste de 21 observaciones. La historia de cada piezómetro se muestra en la Fig. 5.5. Se supone que las observaciones están exentas de errores.

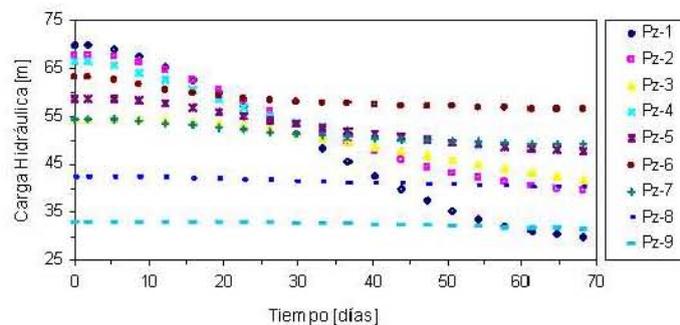


Fig. 5.5 Historias piezométricas.

5.2.3 Ensamblajes iniciales de log-conductividades

Se generan dos conjuntos con mil realizaciones no condicionales del campo aleatorio del logaritmo natural de la conductividad cada uno utilizando una versión modificada del programa SGSIM (Deutsch y Journel, 1992). Los descriptores estadísticos de cada ensamble se muestran en la Tabla 5.1. El tamaño del ensamble se selecciona con base en la estabilidad de los resultados numéricos a diferentes medidas de error (capítulo 3). Observe que en ambos casos considerados en la Tabla 5.1 se ignora la asimetría real del campo de referencia, ya que las realizaciones son multi-gaussianas. En otras palabras, se supone que la presencia de zonas continuas con conductividades altas se desconoce *a priori*. En el segundo caso se supone además una estructura de variabilidad espacial menos heterogénea y menos continua que la del campo de referencia. También vale la pena mencionar que en ambos casos mostrados en la Tabla 5.1 sólo las historias piezométricas por sí mismas modificarán la estructura de variabilidad espacial *a priori* del logaritmo natural de la conductividad dentro del núcleo; es decir, se considera que las mediciones directas de la conductividad no están disponibles o bien, son inexistentes.

Tabla 5.1 Estadísticas de los conjuntos iniciales de realizaciones no condicionales del logaritmo natural de la conductividad $Y(\mathbf{x})=\ln(K_r(\mathbf{x}))$ utilizados en el modelación inversa.

Caso	Media del ensamble	Varianza del ensamble	Escala de correlación	Asimetría
1	-1.193	1.0	a=15 m	ninguna
2	-1.193	0.8	a=13 m	ninguna

¹Esférico con escala de correlación isotropa a

5.2.4 Implementación numérica

El modelo de inversión estocástica se programa en lenguaje de programación FORTRAN para LINUX y se ejecuta en el cluster HPC “Tonatiuh” del Instituto de Ingeniería de la UNAM.

5.3 Campos de la conductividad condicionales a historias de cargas hidráulicas

En el siguiente análisis, se considera un conjunto de n_r realizaciones del campo aleatorio *a priori* del logaritmo natural de la conductividad hidráulica en un tiempo particular $t=0$, es decir, $Y_0(\mathbf{x})$. La dimensión de este campo aleatorio es n . Cada realización de dicho campo aleatorio se utiliza para resolver las ecuaciones de flujo de agua del modelo de filtración y se obtiene un conjunto de realizaciones del campo aleatorio de la carga hidráulica en el tiempo $t=1$, es decir, $H_1(\boldsymbol{\chi})$. La dimensión de este campo aleatorio es N .

Una vez que todas las distribuciones marginales de $Y_0(\mathbf{x})$ y $H_1(\boldsymbol{\chi})$ han sido transformadas numéricamente en distribuciones gaussianas de acuerdo con el procedimiento descrito en el capítulo 3, se puede definir, para cada realización en el ensamble, un vector de dimensión n que consiste de valores *a priori* del logaritmo natural de la conductividad, es decir, $\tilde{\mathbf{U}}_0 = [y'_{t-1}(\mathbf{x}_1), y'_{t-1}(\mathbf{x}_2), \dots, y'_{t-1}(\mathbf{x}_n)]$, un vector de dimensión N_h formado con los valores de las observaciones, es decir, $\tilde{\mathbf{Z}}_t = [h'_{1,t}, h'_{2,t}, \dots, h'_{N_h,t}]$ y un vector reducido de dimensión N_h que incluye estados pronosticados (valores de la carga hidráulica en las ubicaciones de las observaciones), es decir, $\tilde{\mathbf{H}}_t^f = [h'(\boldsymbol{\chi}_1), h'(\boldsymbol{\chi}_2), \dots, h'(\boldsymbol{\chi}_{N_h})]$.

Una realización actualizada del logaritmo natural de la conductividad en el tiempo $t=1$, es decir, $\tilde{\mathbf{U}}_t^u = [\hat{y}'(\mathbf{x}_1), \hat{y}'(\mathbf{x}_2), \dots, \hat{y}'(\mathbf{x}_n)]$ se puede determinar por medio de (Vázquez *et al.*, 2014):

$$\tilde{\mathbf{U}}_t^u = \tilde{\mathbf{U}}_0 + \tilde{\mathbf{K}}_t [\tilde{\mathbf{Z}}_t - \tilde{\mathbf{H}}_t^f] \quad (5.6)$$

donde \mathbf{K}_t es una matriz de tamaño $n \times N_h$ que contiene las funciones ponderadoras $\lambda_i(\boldsymbol{\chi}_\alpha)$ que se obtienen de las soluciones de los siguientes sistemas de ecuaciones:

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

donde $C_H(s, \tau)$ son funciones de autocovarianza entre las cargas piezométricas en $s=(\boldsymbol{\chi}_\alpha, \boldsymbol{\chi}_j)$ para $j=1, \dots, N$ y $\tau=t=1$; $C_{Y'H}(s, \tau)$ son funciones de covarianzas cruzadas entre el logaritmo natural de la conductividad y las cargas hidráulicas en $s=(\mathbf{x}, \boldsymbol{\chi}_j)$ y $\tau=(t=0, t=1)$. Las funciones se determinan estadísticamente con el ensamble de realizaciones.

Después de obtener las funciones ponderadoras $\lambda_i(\boldsymbol{\chi}_\alpha)$ para cada ubicación \mathbf{x} , la matriz \mathbf{K}_t se puede ensamblar como:

$$\mathbf{K}_t = \begin{bmatrix} \lambda_1(\boldsymbol{\chi}_1) & \lambda_1(\boldsymbol{\chi}_2) & \dots & \lambda_1(\boldsymbol{\chi}_{N_h}) \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_n(\boldsymbol{\chi}_1) & \lambda_n(\boldsymbol{\chi}_2) & \dots & \lambda_n(\boldsymbol{\chi}_{N_h}) \end{bmatrix} \quad (5.8)$$

Además de la etapa de actualización previamente descrita del método EnKF, se puede realizar una etapa de actualización basada en la técnica de simulación p -field como se propuso en el capítulo 3. Esta extensión del método EnKF se denomina de aquí en adelante método pf-EnKF.

Una CCDF de valores del logaritmo natural de la conductividad en cada ubicación \mathbf{x} , es decir, $F_{\hat{Y}}(\hat{y}; \mathbf{x})$ se establecen estadísticamente utilizando los conjuntos de realizaciones actualizadas correspondientes. Muestras de valores del logaritmo natural de la conductividad hidráulica en cada ubicación \mathbf{x} , es decir, $\hat{y}'_i(\mathbf{x})$ se pueden obtener a partir de sus distribuciones condicionales respectivas mediante la técnica de simulación " p -field" (Srivastava, 1992; Froidevaux, 1993):

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

donde $p_0(\mathbf{x})$ son los denominados campos de probabilidades (campos p), del campo de valores del logaritmo natural de la conductividad hidráulica. Los campos p se pueden establecer con los valores de las funciones de distribución acumulativas empíricas de cada realización antes de la etapa de actualización según se explicó en el capítulo 3.

Después de retransformar cada CCDF local del campo aleatorio actualizado del logaritmo natural de la conductividad siguiendo el procedimiento descrito en el capítulo 3, se puede formar un vector $\mathbf{Y}_t^u(\hat{y}_1(\mathbf{x}), r)$ de dimensión n que contiene n_r realizaciones $\hat{y}_1(\mathbf{x})$ del logaritmo natural de la conductividad. Tal vector contiene un ensamble de realizaciones del campo a posteriori del logaritmo natural de la conductividad en el tiempo $t=1$, es decir, $Y_1(\mathbf{x})$. La media condicional $\hat{\mu}_{Y|H}(\mathbf{x}; t_1)$ y varianza condicional $\hat{\sigma}_{Y|H}^2(\mathbf{x}; t_1)$ de tal campo aleatorio se pueden estimar en forma estadística con el conjunto de realizaciones en \mathbf{Y}_t^u :

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

donde $m(\mathbf{x}; t_1)$ es la media aritmética en \mathbf{x} al tiempo t_1 ; n_r es el número de realizaciones en el ensamble.

El procedimiento descrito en los párrafos anteriores se repite en el siguiente tiempo en que se disponga de observaciones pero el nuevo ensamble de realizaciones del logaritmo natural de la conductividad será el conjunto de realizaciones *a posteriori* en el tiempo $t=1$, es decir, $Y_1(\mathbf{x})$.

5.4 Resultados y discusiones

Los resultados en esta sección muestran que la media de los campos de velocidades que se obtienen de las realizaciones del logaritmo natural de la conductividad condicionadas por historias de cargas hidráulicas proporcionan información útil acerca de la existencia y ubicación de trayectorias preferenciales de filtración en el núcleo, a pesar del conocimiento incompleto *a priori* de los parámetros estadísticos del campo de referencia del logaritmo natural de la conductividad. En esta sección también se analizan los intervalos de confianza y el error de las estimaciones.

5.4.1 Predicción de los campos de velocidades

En los experimentos de este capítulo, al final de cada etapa de actualización se obtiene en cada caso analizado un ensamble de 1000 realizaciones del logaritmo natural de la conductividad hidráulica condicionadas por las observaciones piezométricas. El conjunto de realizaciones al final del proceso de condicionamiento (en $t=70$ días) se utiliza en un esquema de Monte Carlo para pronosticar velocidades de filtración y cuantificar la incertidumbre de la predicción. En el método de Monte Carlo, el problema directo se resuelve numéricamente por medio de elementos finitos respetando las condiciones de frontera de las ecuaciones 5.1 a 5.5 y utilizando los valores de la conductividad de cada una de las realizaciones condicionales. La media y varianza del campo de velocidades de descarga se determinan estadísticamente con el conjunto de realizaciones.

En la siguiente discusión, sólo se utiliza la media del campo de velocidades al final del abatimiento (en $t=70$ días) para detectar, por inspección visual, trayectorias preferenciales de filtración en el núcleo y para comparar las ubicaciones de tales trayectorias con las que ocurren en el campo de velocidades asociado al campo de conductividades de referencia. Tanto el EnKF como el pf-EnKF se emplean con este propósito.

En la Fig. 5.6(a) se muestra el campo de velocidades asociado al campo de conductividades de referencia al final del abatimiento (en $t=70$ días). Las trayectorias preferenciales de filtración en la figura corresponden claramente con las zonas continuas de valores grandes del logaritmo natural de la conductividad en el campo de referencia de la Fig. 5.2(a). El análisis del campo de velocidades es por tanto útil para detectar trayectorias preferenciales de filtración. El campo de velocidades de referencia se ilustra con el propósito de comparar la media de los campos de velocidades que se obtienen de las realizaciones condicionales de la conductividad en ese tiempo particular.

Las figuras 5.6(b) y (c) muestran la media de los campos de velocidades correspondientes al ensamble de realizaciones condicionales del logaritmo natural de la conductividad del caso 1 que se obtienen con el EnKF y el pf-EnKF, respectivamente, después de integrar secuencialmente las nueve historias piezométricas hasta el tiempo $t=70$ días. Se puede destacar que en tales figuras los campos de velocidades muestran trayectorias preferenciales de filtración en el núcleo (zonas más claras). Esta observación se mantiene también en los campo de velocidades del caso 2 (Fig. 5.6(d) y (e)). Como se puede observar en todas las figuras antes mencionadas los resultados de ambos filtros son muy similares en ambos casos analizados.

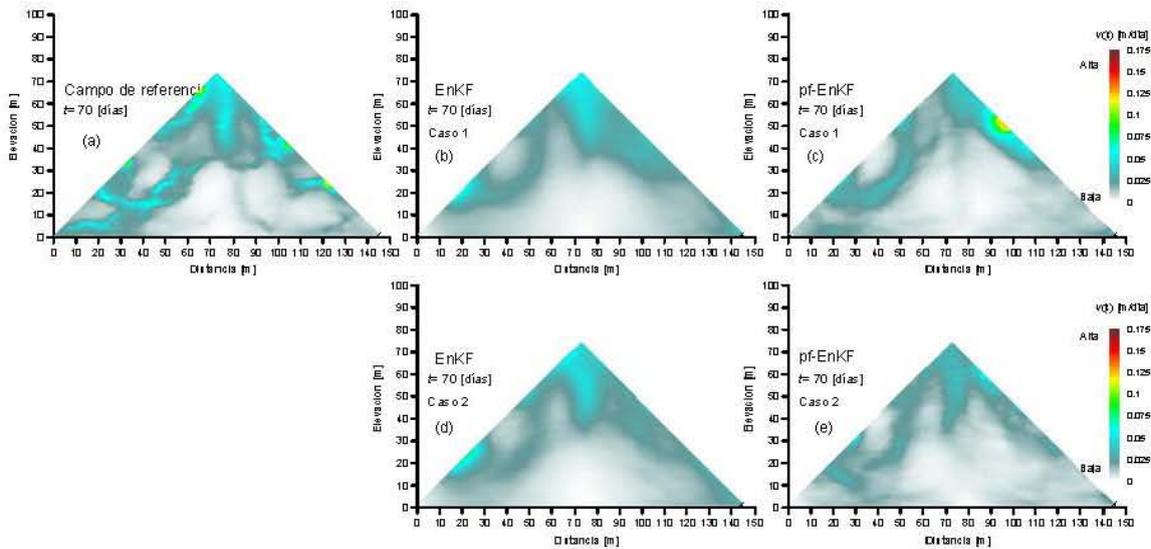


Fig. 5.6 Campos de velocidades promedio [m/día] que se obtienen de las realizaciones de la conductividad condicionadas a nueve historias piezométricas. También se muestra el campo de velocidades de referencia con fines comparativos.

5.4.2 Intervalos de confianza

También se considera en el análisis la confianza del campo de velocidades pronosticado después de completar el proceso de condicionamiento (en $t=70$ días). Se determinan intervalos de confianza con la media estimada más y menos una desviación estándar. A continuación se discuten los resultados en dos secciones de control. Una de las secciones se ubica a lo largo del talud de aguas arriba y la otra está en el eje central del núcleo.

El intervalo de confianza de las velocidades a lo largo del talud de aguas arriba que se obtiene para el caso 1 con las realizaciones condicionales del EnKF muestra que en general las velocidades de referencia están incluidas en el intervalo, excepto en algunas elevaciones con velocidades extremas que corresponden con trayectorias de filtración que dejan la superficie de aguas arriba del núcleo (Fig. 5.7(a)). En el eje central del núcleo las velocidades de referencia están bien acotadas por el intervalo de confianza (Fig. 5.7(b)). Como se puede observar en las Figs. 5.7(c) y (d), los intervalos de confianza calculados con las realizaciones condicionales del pf-EnKF son más estrechos que los del EnKF y fallan al incluir velocidades de referencia en varias elevaciones de las secciones de control.

Para el caso 2, los intervalos de confianza que se obtienen con las realizaciones condicionales del EnKF muestran perfiles más irregulares que se desvían más de las velocidades de referencia que en el caso 1. Este comportamiento podría tener una explicación en la aproximación inicial más pobre que se supuso para el caso 2. Sin embargo, el intervalo de confianza estimado aun así es representativo de las velocidades de referencia (Fig. 5.8(a) y (b)). Los intervalos de confianza

del pf-EnKF son más estrechos y fallan al incluir velocidades de referencia en varias elevaciones. Por lo tanto, a pesar de que los campos de velocidades de la Fig. 5.6 parecen muy semejantes, los campos de velocidades del EnKF están más próximos al campo de velocidades de referencia que los del pf-EnKF. Esta conclusión se revisará en la siguiente sección por medio de un análisis del error de la estimación del campo de velocidades de referencia.

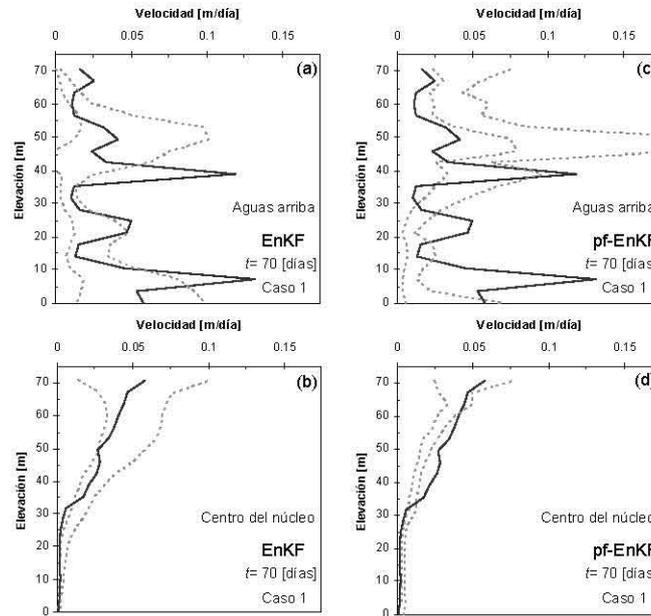


Fig. 5.7 Caso 1. Intervalos de confianza de las velocidades pronosticadas en el tiempo $t=70$ días (líneas interrumpidas). Las velocidades de referencia se indican con líneas sólidas. Intervalos de confianza del EnKF (izquierda). Intervalos de confianza del pf-EnKF (derecha). a) y b) Perfiles a lo largo del talud de aguas arriba. b) y d) Perfiles a lo largo del eje central del núcleo.

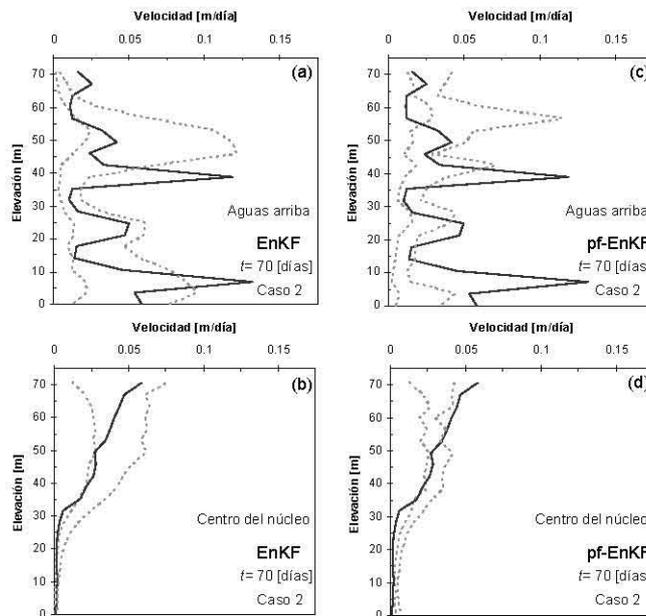


Fig. 5.8 Caso 2. Intervalos de confianza de las velocidades pronosticadas en el tiempo $t=70$ días (líneas interrumpidas). Las velocidades de referencia se indican con líneas sólidas. Intervalos de confianza del EnKF (izquierda). Intervalos de confianza del pf-EnKF (derecha). a) y b) Perfiles a lo largo del talud de aguas arriba. b) y d) Perfiles a lo largo del eje central del núcleo.

5.4.3 Validación cruzada

En esta sección se realiza la validación cruzada entre el campo de velocidades estimado y el de referencia. Para complementar el análisis, se consideran también las validaciones cruzadas del campo estimado del logaritmo natural de la conductividad y del campo estimado de cargas hidráulicas. Además de dos medidas bien conocidas de la *bondad-del-ajuste*, a saber, la media del error absoluto (MAE) y la raíz cuadrada del error cuadrático medio (RMSE), también se evalúa el así llamado error lineal en el espacio de la probabilidad (LEPS). Este último se define como (Bárdossy y Li, 2008):

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

donde $|\cdot|$ significa valor absoluto; $F(y^{\text{ref}}(\mathbf{x}_i))$ es el valor de la función de distribución acumulativa (CDF) correspondiente a la ubicación \mathbf{x}_i del campo de referencia; $F(y^*(\mathbf{x}_i))$ es el valor de la CDF también en la ubicación \mathbf{x}_i del campo estimado. Los campos $y^*(\mathbf{x}_i)$ y $y^{\text{ref}}(\mathbf{x}_i)$ representan, intercambiamente, los campos estimados y de referencia, respectivamente, del logaritmo natural de la conductividad, carga hidráulica y velocidad de descarga.

Los resultados de la validación cruzada de cada caso analizado se muestran en la Tabla 5.2. La validación cruzada se realiza después de completar el proceso de condicionamiento (en $t=70$ días). El LEPS es más fácil de comparar que el MAE o el RMSE porque el primero no se ve afectado por las unidades de las variables. El LEPS varía de cero a uno indicando una estimación perfecta o imperfecta, respectivamente.

Tabla 5.2 Resultados de la validación cruzada de la media estimada del campo del logaritmo natural de la conductividad $Y(\mathbf{x})$ y de las predicciones de la media de los campos de carga hidráulica $H(\mathbf{x})$ y velocidad $V(\mathbf{x})$ después de completar el proceso de condicionamiento ($t=70$ días). La media de los campos de la carga hidráulica y velocidad se obtuvieron estadísticamente a partir de los resultados de la solución del problema directo usando los campos estimados de la conductividad hidráulica después de completar el proceso de condicionamiento. Se indican los resultados de cada caso analizado.

Caso	Campo	EnKF			pf-EnKF		
		RMSE	MAE	LEPS	RMSE	MAE	LEPS
1	$y(\mathbf{x})$	0.887	0.674	0.211	0.933	0.715	0.207
	$h(\mathbf{x})$	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
2	$y(\mathbf{x})$	0.897	0.688	0.215	1.066	0.827	0.247
	$h(\mathbf{x})$	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

En ambos casos analizados, la mejor estimación en términos del LEPS corresponde al campo de cargas hidráulicas. La estimación del logaritmo natural de la conductividad presenta el error más grande en ambos casos. Estos resultados ilustran el bien conocido “*efecto de suavizado*” de las ecuaciones de flujo de agua en suelos, el cual establece que comportamientos semejantes de la carga hidráulica se pueden producir con diferentes campos de la conductividad hidráulica (Delhomme, 1979; Chilès y Delfiner, 1999). El LEPS sugiere que se deben esperar errores similares en la estimación del campo del logaritmo natural de la conductividad y en el campo de velocidades, aunque se puede presentar un efecto de compensación estadística al estimar velocidades. El LEPS indica además que, en términos generales, el EnKF proporciona mejores estimaciones que el pf-EnKF, excepto en la estimación del logaritmo natural de la

conductividad del caso 1. Los resultados de las otras medidas muestran tendencias muy similares.

5.5 Conclusiones

En este capítulo se presentó un enfoque estocástico de modelación inversa para enfrentar el problema de la detección de trayectorias preferenciales de filtración en presas de tierra. Se supuso que un campo simulado con evidentes zonas continuas de conductividades hidráulicas altas representaba la distribución real de conductividades en el núcleo de una presa de tierra hipotética. Tal campo de conductividades se utilizó para generar historias de cargas piezométricas resolviendo una condición de flujo transitorio por vaciado “rápido” del embalse. Se obtuvieron soluciones al problema estocástico inverso con los métodos EnKF y pf-EnKF. Ambos métodos proporcionaron realizaciones del logaritmo natural de la conductividad hidráulica condicionadas por cargas piezométricas en tiempos especificados.

Los resultados de ambos filtros mostraron que aun en el caso de un conocimiento *a priori* incompleto de los parámetros estadísticos del campo de conductividades de referencia así como en ausencia de mediciones directas de la conductividad, es posible identificar las características principales de la distribución espacial de conductividades en el núcleo por medio de la integración de un conjunto limitado de observaciones de la carga hidráulica en el modelo multi-gaussiano *a priori*. Los promedios de los campos de velocidades asociados a las caracterizaciones fueron capaces de proporcionar información útil acerca de la existencia y localización de trayectorias preferenciales de filtración en el núcleo.

La variabilidad espacial de la conductividad hidráulica en estructuras térreas es en efecto tridimensional y además durante su periodo de vida útil ocurren trayectorias de filtración más elaboradas que un simple abatimiento “rápido” del embalse. La metodología propuesta se puede extender fácil y rápidamente para enfrentar tales condiciones.

CONCLUSIONES GENERALES Y PERSPECTIVAS

Resumen

El objetivo central de esta tesis fue el desarrollo de herramientas numéricas para caracterizar masas térricas heterogéneas en relación con sus conductividades hidráulicas desde una perspectiva de análisis probabilista. Se desarrollaron dos algoritmos con este propósito, uno basado en la llamada cópula V -transformada y otro basado en el método del filtro de Kalman ensamblado (EnKF). Los algoritmos se validaron a través de experimentos numéricos.

El documento se dedicó a discutir primero el concepto de dependencia que se define en el contexto de los campos aleatorios. Esta discusión puso en evidencia la necesidad de considerar campos aleatorios con dependencia no multi-gaussiana para representar incertidumbre debida a la variabilidad espacial de la conductividad hidráulica en las masas de suelo. El modelo de cópula V -transformada se utilizó con este propósito.

Se propuso una técnica de inversión basada en el filtro de Kalman ensamblado. Tanto la técnica original (EnKF) como la propuesta (pf-EnKF) se utilizaron para simular campos aleatorios de la conductividad hidráulica condicionales a historias de cargas hidráulicas. En un experimento numérico, las realizaciones condicionales de la conductividad hidráulica se utilizaron para pronosticar velocidades de filtración en la sección transversal del núcleo interno de una presa de tierra idealizada para identificar trayectorias preferenciales de filtración.

Descubrimientos y discusiones

Los experimentos numéricos realizados en esta investigación han mostrado que:

- El modelo de la copula V -transformada es ciertamente una herramienta muy flexible para representar estructuras heterogéneas de la conductividad hidráulica de los suelos, en el sentido de que los mapas generados con este modelo exhiben características realistas de la conductividad de los suelos tales como asimetría y conectividad. El modelo se recomienda por tanto para estimaciones y simulaciones en aplicaciones de la geoestadística.
- En los experimentos que tuvieron el propósito de caracterizar conductividades hidráulicas heterogéneas, se obtuvieron mejores resultados en términos de exactitud

local con el filtro original (EnKF) que con el propuesto (pf-EnKF), pero este último capturó mejor las cópulas empíricas bivariadas del campo de referencia. Lo deseable sería desarrollar métodos que no sólo sean globalmente más exactos sino que además sean localmente más exactos que los métodos originales. Sin embargo, este reto se deja para futuras investigaciones. El método propuesto proporcionó estimaciones más exactas del histograma y funciones de correlación del campo de referencia que las estimaciones *a priori*. Se encontró que tanto el filtro original como el propuesto proporcionan resultados útiles aun en campos de conductividades con dependencia no multi-gaussiana.

- Definitivamente vale la pena tomar en cuenta historias de cargas hidráulicas en procesos de caracterización de la conductividad hidráulica en masas de suelo. Todos los experimentos mostraron que las estimaciones de la conductividad hidráulica mejoran al integrar en la tarea de caracterización respuestas de la carga hidráulica.
- Debido a que las herramientas desarrolladas están basadas en el método de simulación de Monte Carlo, su ejecución puede resultar muy demandante computacionalmente en mallas con un gran número de nodos. Se recomienda el uso de cómputo de alto desempeño para su ejecución.

Recomendaciones para futuras investigaciones

- La calidad de las simulaciones generadas con los modelos basados en el filtro de Kalman ensamblado se podría mejorar incorporando, en la etapa de actualización, técnicas de estimación no lineales tales como la técnica del Kriging indicador.
- Los modelos basados en el filtro de Kalman ensamblado se podrían modificar para analizar trayectorias de filtración más elaboradas que ocurren en presas de tierra tales como la del primer llenado del embalse.
- Se recomienda investigar con más detalle el efecto de la aproximación inicial en el desempeño de las dos técnicas de inversión basadas en el filtro de Kalman ensamblado.
- En los experimentos numéricos se utilizaron por separado los modelos de cópula y filtro de Kalman ensamblado, pero en las aplicaciones se podrían utilizar en forma combinada; por ejemplo, con el modelo de cópula se podrían simular primero campos aleatorios de la conductividad hidráulica condicionales a mediciones disponibles de la conductividad hidráulica misma; posteriormente, estos campos se podrían condicionar a historias de cargas hidráulicas con el filtro de Kalman ensamblado.
- Las herramientas basadas en el filtro de Kalman ensamblado que se desarrollaron en esta investigación permitieron estimar cargas hidráulicas en diferentes ubicaciones y tiempos tomando en cuenta historias de la carga hidráulica misma. En estos casos se consideraron medios porosos indeformables. En un modelo que tome en cuenta la deformabilidad del medio se podrían incorporar en forma análoga historias de asentamientos para pronosticar hundimientos en masas de suelo sometidas a extracción de agua por bombeo y estimar la incertidumbre de la predicción.

BIBLIOGRAFÍA

- Ababou, R., Sagar, B. y Wittmeyer, G. (1992). "Testing procedures for spatially distributed models", *Adv. Water Resour.*, 15, pp.181-198.
- Abramovitz, M. y 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. y 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. y 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. y 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. y Li, J. (2008). "Geostatistical interpolation using copulas", *Water Resour. Res.*, 44, W07412 (1 of 15).
- Bárdossy, A. y 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. y Irmay, S. (1968). "Physical principles of water percolation and seepage", UNESCO.
- Bertino, L., Evensen, G. y Wackernagel, H. (2003). "Sequential Data Assimilation Techniques in Oceanography", *International Statistical Review*, 71(2), pp. 223-241.
- Bradbury, K. R. y 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. y 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. y Kamalabadi, F. (2009). "Tomographic Imaging of dynamic objects with the ensemble Kalman filter", *IEEE Transactions on Image Processing*, 18, pp. 1573–1587.
- Capilla, J. y 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. y George, E. (1992). "Explaining the Gibbs sampling", *American Statistical Association*, 46(3), pp. 167-174.
- Chilès, C. y 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. y 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. y Cockerham, P. (1994). "Practical considerations in the application of simulated annealing to stochastic simulation", *Math. Geology*, 26(1), pp. 67-82.
- Deutsch, C. V. y Journel A. G. (1992). "GSLIB, Geostatistical Library and User's Guide", Oxford University Press, Oxford, N. Y., USA.
- Deutsch, C. y 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. y 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. y 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. y Cherry, J. (1979). "Groundwater", New Jersey, Prentice Hall.
- Frind, E., Muhammad, D. y 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. y 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. y 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. y 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. y 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. y Spanos, P. (1991). "Stochastic finite elements: A spectral approach", Springer, Berlin.
- Gómez-Hernández, J. Sahuquillo, A. y 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. y 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. y Fenton, G. A. (1993). "Seepage beneath water retaining structures founded on spatially random soil", *Geotechnique*, 43(4), pp.577-587.
- Griffiths, D. V. y 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. y 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. y 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. y 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. y 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. y 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. y Virgin, B. (1997). "Local percolation probabilities for a natural sandstone", *Physica A*, 241, pp. 105–110.
- Hoeffding, W. (1940). "Masstabinvariante Korrelationstheorie", *Schriften des Mathematischen 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. y 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., y 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., y 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. y Alabert, F. (1989). "Non-Gaussian data expansion in the earth sciences", *Terra Nova*, 1, pp. 123-134.
- Journel, A. y Deutsch, C. (1993). "Entropy and spatial disorder", *Math. Geology*, 25(3), pp. 329-355.
- Journel, A. G. y Huijbregts, C. (1978). "Mining geostatistics", *Academic Press*, San Diego, Calif., 600 pp.
- Journel, A. y Posa, D. (1990). "Characteristic behavior and order relations for indicator variograms", *Math. Geology*, 22(8), pp. 1011-1025.
- Journel, A. y 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. y 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. y 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. y 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. y Bárdossy, A. (2009). "Stochastic simulation of hydraulic conductivity fields using non Gaussian dependence", *American Geophysical Union*, Fall Meeting, Abstract #H54B-06.
- López-Acosta, N. y 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. y 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). "Incertidumbre en el análisis del flujo de agua en suelos", *Tesis de Doctorado en Ingeniería*, DEPEFI, UNAM.
- López-Acosta, N. y Auvinet, G. (2011). "Uncertainty in analyses of one-dimensional steady-state seepage through random porous media", *Probabilistic Eng. Mech.*, 26(3), pp. 501-510.
- Lu, Z. y 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. y Wagner, H. (1991). "Euler characteristic and related measures for random geometric sets", *Journal of Statistical Physics*, 64(3-4), pp. 843-850.
- Mood, A. y 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. y 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. y Auvinet, G. A. (2013). "Elemento finito estocástico en ingeniería geotécnica. Enfoque espectral", *Ingeniería Investigación y Tecnología*, XIV(1), pp. 11-22.
- Poeter, E. y 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. y 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. y 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. y 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. y Girardi, J. (1996). "Scale effects in transmissivity", *Journal of Hydrology*, 183(1), pp. 1-22.
- Sánchez-Vila, X. Guadagnini, A. y Carrera, J. (2006). "Representative hydraulic conductivities in saturated groundwater flow", *Rev. Geophys.*, 44(RG3002), pp. 1-46.
- Sakamoto, S. y 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. y 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. y 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. y 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. y Griffiths, D. (2004). "Programming the finite element method", England, John Wiley and Sons Ltd.
- Smith, M. y 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. y 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.
- Trincheró, P., Sánchez-Vila, X. y 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. y 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. y Auvinet, G. (2014). Simulación de campo aleatorios con dependencia no multi-gaussiana empleando cópulas, *Investigación y Tecnología*.
- Vázquez, F., Auvinet, G. y Vermeer, P. (2014). Stochastic inverse modeling applied to earth dams for detecting preferential seepage paths, NUMGE, Delft, The Netherlands.
- Vázquez, F. y Auvinet, G. (2015). Simulación de campos aleatorios espacio-temporales utilizando un filtro de Kalman modificado, *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. y 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. y 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. y 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., y Shinozuka, M. (1988). “Digital generation of non-Gaussian stochastic fields”, *J. Eng. Mech.*, 114(7), pp. 1183–1197.
- Yamazaki, F., Shinozuka, M. y 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. y 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. y Yeh, T. (2005). “Characterization of aquifer heterogeneity using transient hydraulic tomography”, *Water Resour. Res.*, 41, W07028, doi:10.1029/2004WR003790.
- Zinn, B. y 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).

APÉNDICE A

CONCEPTOS BÁSICOS DE LA FILTRACIÓN EN MEDIOS ALEATORIOS SATURADOS

A.1 Ley de Darcy

La *ley de Darcy* fue derivada en forma experimental y se limitó al flujo unidimensional. Cuando el flujo es tridimensional, una posible generalización de esta es (Bear, 1972):

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

donde; \mathbf{q} es un vector de *descargas específicas* con componentes q_x , q_y , q_z en las direcciones de las coordenadas cartesianas x , y , z , respectivamente, y \mathbf{J} es el *gradiente hidráulico* con componentes en las direcciones x , y , z , respectivamente. El flujo descrito por la ley de Darcy es irrotacional y laminar. La ecuación A.1 permanece válida en flujo tridimensional en medios isotrópicos heterogéneos. En tales casos: $K_s = K_s(x, y, z)$. El coeficiente de proporcionalidad K_s se denomina *conductividad hidráulica*. En medios isotrópicos, este es un escalar (dimensiones L/T). Cuando éste se expresa solamente en términos de la matriz sólida, se utiliza el término *permeabilidad* o *permeabilidad intrínseca* (dimensiones L^2). Bajo ciertas condiciones, la permeabilidad puede variar con el tiempo. Los hundimientos por consolidación son fenómenos asociados con cambios en la permeabilidad con respecto al tiempo. En el presente trabajo, el término permeabilidad o conductividad hidráulica se utiliza equivalentemente para referirse al concepto de conductividad hidráulica y solo se consideran análisis de filtración unidimensionales y bidimensionales.

A.2 Tensores de la conductividad hidráulica

Las formaciones naturales de suelo y estructuras térreas son, de hecho, anisótropas con respecto a la conductividad hidráulica; es decir, la magnitud de esta propiedad depende de la dirección. La permeabilidad y conductividad hidráulica en estos casos son, de hecho, *tensores simétricos de segundo orden*; por lo tanto, se necesitan seis experimentos independientes para

determinarlos por completo. En la práctica, se desconocen las direcciones principales de anisotropía tanto en la dirección del flujo como en la dirección del gradiente. Por lo tanto, es necesario suponer tres ejes principales de tal manera que solo sean necesarios tres experimentos independientes. En ciertos medios porosos, la complejidad del problema se puede reducir aún más suponiendo solo dos direcciones principales, por ejemplo, x - y y z . En tales casos, cualquier dirección en el plano x - y es una dirección principal y así solo dos experimentos independientes son necesarios.

En la práctica, las mediciones de la conductividad hidráulica son con frecuencia escasas y sus correspondientes mediciones direccionales prácticamente nunca se determinan. Por esta razón, los análisis geoestadísticos de tensores de la conductividad hidráulica son prácticamente imposibles de efectuar; sin embargo los valores escalares de la conductividad hidráulica se pueden utilizar para representar conductividades hidráulicas anisótropas por medio de *elipses de anisotropía*. Se puede definir una elipse de anisotropía de tal manera que describa una continuidad mayor en la dirección horizontal que en la vertical. La continuidad se define en geoestadística por medio de lo que se denomina *escala de correlación* de los variogramas o funciones de covarianza direccionales. Los experimentos numéricos presentados en esta tesis se podrían extender para incorporar elipses de anisotropía en futuras investigaciones.

A.3 Velocidad de filtración

La conductividad hidráulica sea un escalar o un tensor es una característica macroscópica del medio poroso. Así, $v=q/A$ es una *velocidad de filtración* y no la velocidad media de flujo dentro de los poros. A una escala microscópica el flujo es gobernado por las ecuaciones de *Navier-Stokes*. Estas ecuaciones son válidas a la escala de los poros y están sujetas a complicadas condiciones de frontera. La solución de estas ecuaciones requeriría una descripción detallada de la geometría del espacio poroso. En este trabajo solo se considera filtración a una escala macroscópica.

A.4 Efectos de Escala

La conductividad hidráulica de los medios porosos depende de la escala de observación. Dagan (1989) definió tres escalas diferentes: la escala *macroscópica* de las muestras de suelo que se ensayan en laboratorio, la escala *megascópica* que involucra un acuífero completo y a varios pozos de bombeo y por último la *escala de los bloques* en modelos del tipo elemento finito (valor definido dentro de un elemento de una malla de elementos finitos). Se pueden definir diferentes tensores de conductividades a estas escalas. La razón por la cual tal efecto de escala aparece se debe al hecho de que la conductividad hidráulica no es una *variable aditiva*. Por ejemplo, dos valores en serie de la conductividad hidráulica se combinan por medio de un promedio armónico. El contenido de agua y la porosidad son ejemplos de variables aditivas ya que ambas cantidades se combinan por medio de simples medias aritméticas.

Al pasar de la escala microscópica a la macroscópica de un espécimen de laboratorio, la heterogeneidad del espacio poroso es promediada y aparece un valor único de la conductividad hidráulica en la ley de Darcy. Una pregunta interesante que surge es si una nueva ley de Darcy aparecería al pasar de la escala macroscópica a la megascópica, de tal manera que haga posible caracterizar globalmente una formación con un valor escalar individual “promediado” de la conductividad hidráulica. Esta pregunta conduce al concepto de *conductividad efectiva*.

APÉNDICE B

CONDUCTIVIDAD EFECTIVA Y CONDUCTIVIDAD APARENTE

B.1 Conductividad hidráulica efectiva

La *conductividad hidráulica efectiva* K_{eff} , se define como el tensor de segundo orden que relaciona los valores esperados de la descarga específica q y carga hidráulica φ a través de la ley generalizada de Darcy:

$$E\{q\} = -K_{eff} E\{\nabla \varphi\} \quad (B.1)$$

donde las esperanzas representan el promedio de todos los campos posibles de descargas específicas y cargas hidráulicas que podrían obtenerse con un conjunto de realizaciones de la conductividad hidráulica.

En la literatura se dispone de solo dos resultados exactos del valor de K_{eff} . El primero dice que K_{eff} está limitado por la *media armónica* $K_h = (E[K^{-1}])^{-1}$ y la *media aritmética* $K_a = E[K]$ de las conductividades locales (Matheron; 1967):

$$(E[K^{-1}])^{-1} \leq K_{eff} \leq E[K] \quad (B.2)$$

El segundo resultado exacto también se debe a Matheron (1967) quien demostró que en un dominio infinito bidimensional, K_{eff} es igual a la *media geométrica* $K_G = \exp(E[Y])$, toda vez que Y sea un campo aleatorio multi-gaussiano con función de correlación isótropa.

El término conductividad efectiva se utiliza únicamente cuando el valor K_{eff} es constante en todo el dominio. En estos casos, se considera una propiedad característica del medio. La presencia de fronteras o de fuentes/sumideros provocaría que K_{eff} varíe en el espacio. En tales casos, se utiliza el término conductividad *pseudo-efectiva*, el cual enfatiza su naturaleza local (e.g. Sánchez-Vila, 1997).

B.2 Conductividad equivalente

La *conductividad equivalente* se define como la conductividad de un medio homogéneo ficticio que conlleva la misma descarga que el medio heterogéneo real para una caída de carga hidráulica determinada. Cuando esta se utiliza para reemplazar a un bloque heterogéneo por uno homogéneo equivalente en simulaciones numéricas, también se denomina *conductividad escalada* o *conductividad de bloque*. En este caso, los promedios del conjunto de realizaciones se reemplazan por promedios espaciales (de volúmenes):

$$\bar{q} = -K_b \overline{\nabla \phi} \quad (\text{B.3})$$

donde:

$$\bar{q} = \frac{1}{V} \int_V q dV \quad \overline{\nabla \phi} = \frac{1}{V} \int_V \nabla \phi dV \quad (\text{B.4})$$

Cuando el volumen promediado es muy grande, con un tamaño representativo tal que sea varias veces más grande que el tamaño del volumen dado por la escala de correlación del parámetro heterogéneo (aleatorio), el volumen comprende todas las escalas de heterogeneidad. En tal caso, se dice que la función aleatoria que representa al medio heterogéneo es *ergódica*. *Ergodicidad* implica entonces que los parámetros equivalentes y efectivos (o pseudo efectivos) sean idénticos.

APÉNDICE C

ALGUNAS DEFINICIONES Y PROPIEDADES DE LAS CÓPULAS MULTIVARIADAS

C.1 Cópulas multivariadas

Una cópula n -variada (n -cópula) es un mapeo C del hipercubo unitario $[0,1]^n$ sobre el intervalo unitario $[0,1]$; es decir:

$$C : [0,1]^n \rightarrow [0,1] \quad (\text{C.1})$$

con las siguientes propiedades:

- 1) El rango de $C(u_1, \dots, u_n)$ es el intervalo unitario $[0,1]$;
- 2) $C(u_1, \dots, 0, \dots, u_n) = 0$ para cualquier $u_i = 0$, donde $i = 1, 2, \dots, n$;
- 3) $C(1, \dots, u_i, \dots, 1) = u_i$, para toda $u_i \in [0,1]$;
- 4) $C(u_1, \dots, u_n)$ es n -creciente en el sentido de que para cualquier hipercubo n -dimensional en el cubo unitario la correspondiente medida asignada por C no tiene que ser negativa. Bárdossy (2006) escribe dicha condición por medio de:

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

si $0 \leq u_i \leq u_i + \Delta_i \leq 1$ y $i = \sum_{k=0}^{n-1} j_k 2^k$

(C.2)

A partir de las propiedades descritas antes, se entiende que C no es decreciente en cada variable y que si es uniformemente continua. La existencia (casi en todas partes) de todas las derivadas parciales es además una propiedad útil de las cópulas (Nelsen 2006).

C.2 Cópulas y distribuciones multivariadas

Las funciones de distribución multivariadas y las funciones de distribución unidimensionales están ligadas una con la otra a través de una cópula. Considérese una distribución n -variada $F_Z \in F(F(z_1), \dots, F(z_n))$ con j^{th} marginales uni-variadas $F(z_j)$. Por medio de una versión multidimensional del teorema de Sklar (1959) (e.g. Nelsen, 2006), se muestra que hay una cópula asociada $C : [0,1]^n \rightarrow [0,1]$ que satisface:

$$\begin{aligned} F_Z(z_1, \dots, z_n) &= C(F(z_1), \dots, F(z_n)) \\ &= P\{Z_1 \leq z_1, \dots, Z_n \leq z_n\} \\ &= C(P\{Z_1 \leq z_1\}, \dots, P\{Z_n \leq z_n\}) \end{aligned} \quad (\text{C.3})$$

Si la distribución es continua, entonces la cópula es única. Por el contrario, si $F_Z(\cdot)$ es una función de distribución n -variada continua con marginales uni-variadas $F(z_1), \dots, F(z_n)$ y funciones inversas $F^{-1}(z_1), \dots, F^{-1}(z_n)$ entonces:

$$C(u) = C(u_1, \dots, u_n) = F_Z(F^{-1}(u_1), \dots, F^{-1}(u_n)) \quad (\text{C.4})$$

C.3 Densidades de cópulas

Recordando que la derivada de una función de distribución de probabilidad $F_Z(z)$ es igual a la función de densidad de probabilidad $f_Z(z)$, diferenciando la ecuación C.4 con respecto a todas las variables, la *densidad de la cópula multivariada* se obtiene como sigue:

$$\begin{aligned} f_Z(z_1, \dots, z_n) &= \frac{\partial^n C(F(z_1), \dots, F(z_n))}{\partial F(z_1), \dots, \partial F(z_n)} \prod_{i=1}^n f_i(z_i) \\ &= c(F(z_1), \dots, F(z_n)) \cdot \prod_{i=1}^n f_i(z_i) \end{aligned} \quad (\text{C.5})$$

donde $f_Z(z_1, \dots, z_n)$ es la función de densidad de probabilidad conjunta. Entonces, definiendo $u_1 = F(z_1), \dots, u_n = F(z_n)$, la densidad de la cópula multivariada se puede expresar por medio de:

$$c(u_1, \dots, u_n) = \frac{\partial^n C(u_1, \dots, u_n)}{\partial u_1, \dots, \partial u_n} \quad (\text{C.6})$$

Por lo tanto, la densidad de la cópula multivariada es:

$$c(F(z_1), \dots, F(z_n)) = \frac{f_Z(z_1, \dots, z_n)}{\prod_{i=1}^n f_i(z_i)} \quad (C.7)$$

C.4 Densidades de cópulas condicionales

La densidad de la cópula multivariada condicional puede expresarse como (Bárdossy and Li, 2008; Salvadori *et al.*, 2007):

$$c(u | u_1 = F(z_1), \dots, u_n = F(z_n)) = \frac{f_{n+1}(z, z_1, \dots, z_n)}{f_1(z) f_1(z_1) \dots f_1(z_n)} \frac{1}{c_n(u_1, u_2, \dots, u_n)} \\ \propto \frac{f_{n+1}(z, z_1, \dots, z_n)}{f_1(z) f_1(z_1) \dots f_1(z_n)} \quad (C.8)$$

la cual expresa la relación entre la función de densidad condicional conjunta $f_{n+1}(z, z_1, \dots, z_n)$ de las n -condiciones más uno, dividida por el producto de las correspondientes $n+1$ densidades unidimensionales $f_1(z_i)$; $i = 1, 2, \dots, n$. La variable $n+1$ es una variable u en el punto condicionado. La densidad de la cópula condicional se puede ver entonces como la distribución condicional conjunta de n variables aleatorias uniformes condicionales. El valor u en el punto condicionado se puede obtener evaluando la densidad de la cópula condicional en el intervalo completo de valores de $u \in [0, 1]$.

Es importante notar que la ec. C.8 define la densidad de la cópula condicional. Para fines de interpolación o simulación, el valor en cualquier punto no observado deberá generarse de la cópula condicional la cual es equivalente a la función de distribución condicional. Tal cópula condicional se puede calcular numéricamente integrando la densidad de la cópula condicional. Utilizando un esquema de diferencias finitas, la ecuación final que resulta para la cópula condicional se puede escribir como:

$$C(u^* | u_1, \dots, u_n) \propto \sum_{u=0}^{u^*} c(u | u_1 = F(z_1), \dots, u_n = F(z_n)) \Delta u \quad (C.9)$$

para Δu , pequeña. Más explícitamente:

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

APÉNDICE D

FUNCIONES MULTIVARIADAS DE LA CÓPULA V -TRANSFORMADA

D.1 La función de distribución multivariada

Bárdossy y Li (2008), Li y Bárdossy (2009) y Bárdossy y Pegram, 2009 escriben la función de distribución multivariada del campo aleatorio $V(\mathbf{x})$ en la ec. 2.7 del capítulo 2 como (ver nota al pie)¹:

$$\begin{aligned} H_n(x_1, \dots, x_n) &= P\{X_1 \leq x_1, \dots, X_n \leq x_n\} \\ &= \sum_{i=0}^{2^n-1} (-1)^{\sum_{j=1}^n i_j} \Phi(\zeta_i + \bar{m}) \end{aligned} \quad (\text{D.1})$$

donde:

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

$i_j = 0$ o 1 ; y

$$i = \sum_{j=1}^n i_j 2^{j-1} \quad (\text{D.3})$$

con:

$$b = \begin{cases} -1 & \text{if } (-1)^{i_j} = -1 \\ \frac{1}{v} & \text{if } (-1)^{i_j} = 1 \end{cases} \quad (\text{D.4})$$

y

¹ La variable X se usa en este apéndice en lugar de la variable $Y(\mathbf{x})$ de la ec. 2.7 del capítulo 2 para propiciar que el lector siga más fácilmente la formulación original.

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

Todos los elementos en el vector m se suponen iguales².

D.1.1 Expansión de la serie en la ec. D.1 para el caso bivariable

Para evaluar la ec. D.1 se tienen que determinar los coeficientes i de la ec. D.3. Después, con los coeficientes b y a de la ec. D.4 se podrán construir los vectores ζ_i^T de la ec. D.2, los cuales a su vez serán diferentes dependiendo si la variable transformada x_i es positiva o negativa y dependiendo del signo de cada termino en ζ_i^T , el cual está determinado por los súper índices i 's en $(-1)^{i_1}, \dots, (-1)^{i_n}$. Ahora, para obtener las i 's en la ec. D.1, las i_j 's tendrán que hacerse 0 o 1 con tal de cumplir ambos lados de la igualdad. En el caso bivariable la secuencia se escribe como sigue:

Para $n = 2$:

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

Expandiendo la ecuación anterior, se tiene:

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

En la ec. D.1, i varía desde 0 hasta 3. Entonces, se puede escribir:

$$\begin{aligned} 0 &= 0_1 2^0 + 0_2 2^1 \rightarrow 0 = 0_1 + 0_2 2 \rightarrow i_j = 0_1 = 0; i_j = 0_2 = 0 \\ 1 &= 1_1 2^0 + 1_2 2^1 \rightarrow 1 = 1_1 + 1_2 2 \rightarrow i_j = 1_1 = 1; i_j = 1_2 = 0 \\ 2 &= 2_1 2^0 + 2_2 2^1 \rightarrow 2 = 2_1 + 2_2 2 \rightarrow i_j = 2_1 = 0; i_j = 2_2 = 1 \\ 3 &= 3_1 2^0 + 3_2 2^1 \rightarrow 3 = 3_1 + 3_2 2 \rightarrow i_j = 3_1 = 1; i_j = 3_2 = 1 \end{aligned}$$

Haciendo las i_j 's ya sea 0 o 1, se pueden determinar los signos frente a las x 's así como los valores de las b 's y a 's. Por ejemplo:

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

² Note que, por consistencia, la ecuación $i = \sum_{j=0}^{n-1} i_j 2^j$, en el artículo de Bárdossy and Li (2008) se reemplaza aquí por medio de la ec. D.3. Además, el súper índice i , de la ec. D.1 en dicho artículo se corrige aquí escribiendo $\sum_{j=1}^n i_j$ para lograr la correcta secuencia de signos en la expansión.

$$\begin{aligned}
(-1)^{1_1} = -1 &\rightarrow b = -1, a = 1 && ; && (-1)^{1_2} = 1 &\rightarrow b = \frac{1}{v}, a = \frac{1}{v} \\
(-1)^{2_1} = 1 &\rightarrow b = \frac{1}{v}, a = \frac{1}{\alpha} && ; && (-1)^{2_2} = -1 &\rightarrow b = -1, a = 1 \\
(-1)^{3_1} = -1 &\rightarrow b = -1, a = 1 && ; && (-1)^{3_2} = -1 &\rightarrow b = -1, a = 1
\end{aligned}$$

Después, los vectores ζ_i^T se pueden construir por medio de:

$$\begin{aligned}
\zeta_0^T &= \left[\left(\frac{x_1}{v} \right)^{1/\alpha} ; \left(\frac{x_2}{v} \right)^{1/\alpha} \right] \\
\zeta_1^T &= \left[-x_1 ; \left(\frac{x_2}{v} \right)^{1/\alpha} \right] \\
\zeta_2^T &= \left[\left(\frac{x_1}{v} \right)^{1/\alpha} ; -x_2 \right] \\
\zeta_3^T &= [-x_1 ; -x_2]
\end{aligned}$$

Finalmente, para $n=2$ la expansión de la ec. D.1 resulta:

$$\begin{aligned}
H_2(x_1, x_2) &= P\{X_1 \leq x_1, X_2 \leq x_2\} \\
&= \Phi(\zeta_0 + \bar{m}) - \Phi(\zeta_1 + \bar{m}) - \Phi(\zeta_2 + \bar{m}) + \Phi(\zeta_3 + \bar{m})
\end{aligned} \tag{D.6}$$

Una etapa primordial es por tanto la determinación de la secuencia de 0's y 1's de las i_j 's los cuales a su vez determinan los signos frente a los valores de las x 's dentro de los vectores ζ_i^T . Tal secuencia se presenta en la siguiente tabla para el caso $n=3$.

Table E.1 Combinaciones binarias de i_j y los signos que resultan para $n=3$.

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 La función de densidad multivariada

La correspondiente función de densidad multivariada de X_j se obtiene diferenciando la ec. D.1 (Bárdossy and Li, 2008; Li and Bárdossy, 2009; Bárdossy and Pegram, 2009):

$$h_n(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Gamma^{-1}|^{\frac{1}{2}}} \sum_{i=0}^{2^n-1} \frac{1}{(\alpha \cdot v)^{n-\sum_{j=0}^{n-1} i_j}} \left(\left(\frac{x_p}{v} \right)^{\frac{1}{\alpha}-1} \right)^{n-\sum_{j=0}^{n-1} i_j} \exp\left(-\frac{1}{2} (\zeta_i + \bar{m})^T \Gamma^{-1} (\zeta_i + \bar{m})\right) \quad (D.7)$$

donde $|\Gamma^{-1}|$, denota el determinante de la matriz inversa de coeficientes de correlación de Pearson del campo aleatorio normal por transformar Y , T es la transpuesta de los vectores $(\zeta_i + \bar{m})$ y x_p es el producto de las variables x 's de la secuencia que se ejemplifica en seguida.

D.2.1 Expansión de la ec. D.7 para el caso bivariado

Para $n = 2$, la expansión de la ec. D.7 es:

$$\begin{aligned} h_2(x_1, x_2) = & \frac{1}{2\pi |\Gamma^{-1}|^{\frac{1}{2}}} \cdot \left\{ \frac{1}{(\alpha v)^2} \left(\frac{x_1 x_2}{v^2} \right)^{\frac{1}{\alpha}-1} \exp\left(-\frac{1}{2} (\zeta_0 + \bar{m})^T \Gamma^{-1} (\zeta_0 + \bar{m})\right) + \dots \right. \\ & \dots + \frac{1}{(\alpha v)} \left(\frac{x_1}{v} \right)^{\frac{1}{\alpha}-1} \exp\left(-\frac{1}{2} (\zeta_1 + \bar{m})^T \Gamma^{-1} (\zeta_1 + \bar{m})\right) + \dots \\ & \dots + \frac{1}{(\alpha v)} \left(\frac{x_2}{v} \right)^{\frac{1}{\alpha}-1} \exp\left(-\frac{1}{2} (\zeta_2 + \bar{m})^T \Gamma^{-1} (\zeta_2 + \bar{m})\right) + \dots \\ & \left. \dots + \exp\left(-\frac{1}{2} (\zeta_3 + \bar{m})^T \Gamma^{-1} (\zeta_3 + \bar{m})\right) \right\} \end{aligned}$$

APÉNDICE E

ECUACIONES DE ALGUNAS FUNCIONES DE CONECTIVIDAD

E.1 La función de conectividad y la función de separación entre 2 clusters

Además de la función de Euler utilizada en los capítulos 1 y 2, la conectividad de las realizaciones simuladas se puede cuantificar a través de funciones de conectividad. Aunque no existe una definición matemática única de este concepto (Renard & Allard, 2013), la llamada función de conectividad de Allard (1994) también se emplea en las aplicaciones. La función de conectividad de un campo aleatorio estacionario se define como:

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

La función de conectividad, por tanto, expresa la probabilidad de que un valor en la posición \mathbf{x} de la fase X esté conectado “ \leftrightarrow ” con otro valor de X en la posición $\mathbf{x} + \mathbf{h}$; donde X es el subconjunto de Ω (región de flujo) con Ω en $\mathfrak{R}^{1,2,3}$ tal que: $I(\mathbf{x}; z) = 1$; siendo: $I(\mathbf{x}; z)$ una variable aleatoria binaria tal que: $I(\mathbf{x}; z) = 1$ si $Z(\mathbf{x}) \leq z$ o $I(\mathbf{x}; z) = 0$ si $Z(\mathbf{x}) > z$.

En la ecuación (E.1), $\{\mathbf{x} \leftrightarrow \mathbf{x} + \mathbf{h}\}$ necesita que $\{Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h}) \in X\}$, por lo tanto, la probabilidad incondicional $P[\mathbf{x} \leftrightarrow \mathbf{x} + \mathbf{h}]$ puede ser descompuesta como (Torquato *et al.*, 1988):

$$\begin{aligned} P[\mathbf{x} \leftrightarrow \mathbf{x} + \mathbf{h}] &= P[\mathbf{x} \leftrightarrow \mathbf{x} + \mathbf{h} \mid \mathbf{x}, \mathbf{x} + \mathbf{h} \in X] P[\mathbf{x}, \mathbf{x} + \mathbf{h} \in X] \\ \tau_2(\mathbf{h}; z) &= \tau_1(\mathbf{h}; z) K_I(\mathbf{h}) \end{aligned} \quad (\text{E.2})$$

donde: $K_I(\mathbf{h})$ es la función de covarianza no centrada de X . La ecuación (E.2) se puede interpretar como la “función de separación entre dos clusters” (two-point cluster function).

Recordando que: $K_I(\mathbf{h}) = P[Z(\mathbf{x}) \leq z_1; Z(\mathbf{x} + \mathbf{h}) \leq z_2]$; la ecuación (E.2) se puede escribir en términos de cópulas como:

$$\tau_2(\mathbf{h}; z) = \tau_1(\mathbf{h}; z) C(u_1 = F_{z_1}(z_1), u_2 = F_{z_2}(z_2)) \quad (\text{E.3})$$

donde: $C(u_1, u_2)$ es la copula bivariada.

Para evaluar la distancia promedio en la que dos puntos de la región de flujo Ω se encuentran conectados, se puede calcular la escala integral de la ecuación (E.3):

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