



NATURE DU DOCUMENT : <b>RAPPORT FINAL</b>	Identification : <b>B RP 0FSS 98-001/A</b>  Code (facultatif) :
--	--

ANDRA002 - Ind B

Émetteur : ANDRA - DS/HG	Repère support/Secrétaire : BRP0FSS98001.DOC	Date d'origine : 29/12/98	Page : 1/245
-----------------------------	---	------------------------------	-----------------

**ANALYSE BIBLIOGRAPHIQUE DES  
METHODES GEOSTATISTIQUES  
ADAPTEES A LA  
CARACTERISATION DES  
AQUIFERES**

Titulaire : <b>FSS</b> 1956 Avenue Salengro 93370 CHAVILLE	N° du marché : 005377SHG
	Référence du titulaire : Approbation ANDRA Nom et visa : E. FILLION

Ce document est la propriété de l'ANDRA et ne peut être reproduit ou communiqué sans son autorisation

Ind. : A	Dates : 29/12/98	Nom et visa Rédacteur : J. GOMEZ-HERNANDEZ M. GARCIA	Nom et visa Vérificateur : M. GARCIA	Nom et visa Approbateur : M. GARCIA
-------------	---------------------	--	---	--



Révisions		MODIFICATIONS
Ind.	Date	
A	29/12/98	EMISSION INITIALE



## Résumé

La géostatistique est une discipline adaptée à la modélisation de variables ou de phénomènes structurés dans l'espace. Ceci concerne en particulier les propriétés des roches dans les formations géologiques. Ce type de variables présente en effet une continuité spatiale qui se traduit par le fait que deux mesures, effectuées en deux points distants, de la variable considérée, vont avoir tendance à être d'autant plus semblables que les points de mesure sont proches.

Pour modéliser des propriétés du sous-sol, l'information disponible se résume souvent à des données directes, obtenues en laboratoire sur carottes ou in situ à partir de sondages, et de données indirectes, de type géophysique ou autre, qui ne renseignent pas directement sur les propriétés mais présentent certaines corrélations avec elles. Le problème de modélisation de la répartition spatiale de propriétés du sous-sol devient alors un problème d'interpolation de valeurs à partir de données de différents types. Il s'agit de générer des modèles de propriété qui honorent l'ensemble des données directes disponibles mais aussi reproduisent les caractéristiques statistiques qui décrivent la continuité spatiale des propriétés et leurs corrélations avec les données indirectes. Les aspects liés à l'évaluation de l'incertitude sur les modèles obtenus doivent être abordés également.

La géostatistique fournit une panoplie d'outils qui permettent d'atteindre ces objectifs soit par estimation (interpolation de valeurs) soit par simulation stochastique (génération d'une multitude de réalisations équiprobables). Appliquée à la caractérisation d'aquifères, elle offre une grande variété de méthodes appropriées pour modéliser la répartition spatiale des propriétés nécessaires à la simulation de phénomènes d'écoulement et de transport en milieu poreux. Ces propriétés incluent en particulier la perméabilité et la porosité.

L'objet de cette étude était de réaliser une analyse bibliographique des méthodes géostatistiques adaptées à la caractérisation d'aquifères dans la perspective d'évaluer les temps de migration de radionucléides dans des aquifères situés à proximité de stockages souterrains de déchets radioactifs. Pour appréhender les enjeux liés à ce type d'application et pouvoir envisager une démarche géostatistique systématique, le site de l'Est a été choisi par l'ANDRA comme référence et deux scénarios d'impact d'un stockage sur un aquifère ont été définis. Le premier scénario envisage la mise en communication directe entre le stockage et l'aquifère par un puits qui traverse l'ensemble (cas d'une source ponctuelle de radionucléides). Le second porte sur la migration verticale de radionucléides depuis le stockage jusqu'à l'aquifère (cas d'une source diffuse de radionucléides).

Une réflexion préalable, fondée sur les données disponibles de ce site et les scénarios d'impact proposés, a été menée pour établir les grandes lignes d'une démarche géostatistique qui répondrait aux objectifs recherchés et permettrait d'exploiter au mieux les informations disponibles sur un site. La démarche décrit l'ensemble des étapes à franchir depuis l'analyse des données jusqu'à la simulation des écoulements et du transport en passant par la modélisation structurale de l'aquifère, la simulation des types de faciès et des propriétés, et le changement d'échelle souvent nécessaire pour pouvoir faire tourner un simulateur d'écoulement et de transport à partir d'un modèle géologique de grande résolution. Cette réflexion a aussi confirmé le besoin

de méthodes géostatistiques capables d'intégrer une information hétérogène constituée en particulier :

- de données directes (propriétés de la roche) et indirectes (mesures géophysiques) corrélées de manière plus ou moins complexe,
- de mesures piézométriques (problème inverse),
- de résultats d'interprétation d'essais hydrogéologiques (problème inverse également).

L'étude bibliographique proprement dite devait s'intéresser à la fois aux aspects théoriques des méthodes géostatistiques et, dans la mesure du possible, à leurs performances par des tests comparatifs sur des exemples choisis. Elle a porté sur les trois familles de méthodes suivantes :

1. méthodes d'estimation (interpolation),
2. méthodes de simulation stochastique,
3. méthodes de résolution de problèmes inverses (intégration de données d'écoulement).

Le cas d'étude utilisé pour évaluer les méthodes d'estimation et de simulation stochastique est un cas réel, tiré de données topographiques, comprenant deux variables corrélées. La première peut être assimilée à une perméabilité et représente la variable directe qui est supposée avoir été échantillonnée en un nombre réduit de points. La seconde s'apparente à une information indirecte de type géophysique connue partout, en chaque point d'estimation. La solution exacte de la variable directe (perméabilité) est connue également et sert de référence pour comparer les résultats obtenus avec les différentes méthodes géostatistiques.

Pour évaluer comparativement les performances des méthodes de simulation stochastique, plusieurs problèmes d'écoulement et de transport de radionucléides, depuis une ligne source jusqu'à des points cibles, ont été introduits et simulés numériquement sur le champ de perméabilité de référence et sur les champs simulés de manière stochastique. Cet exercice a permis de traduire l'incertitude sur les champs de perméabilité simulés en une incertitude sur des positions et des temps d'arrivée de radionucléides. Les incertitudes ainsi calculées varient selon la méthode de simulation stochastique employée. Pour ce type d'application, les méthodes les mieux adaptées sont alors celles qui donnent les réponses en transport (temps et position) qui encadrent le mieux les réponses obtenues sur la solution de référence.

L'étude bibliographique a également porté sur les méthodes de changement d'échelle dont l'importance est reconnue unanimement dans la mise en œuvre d'une démarche géostatistique destinée à fournir les données d'entrée à des simulateurs d'écoulement et de transport en milieu poreux.

## Sommaire

Liste des figures .....	9
Liste des tableaux .....	13
Présentation de l'étude .....	15
1. APPROCHE GEOSTATISTIQUE : DISCUSSIONS AUTOUR DES DONNEES du SITE de l'Est ...	17
1.1 PRESENTATION DU PROBLEME .....	19
1.2 PRESENTATION DU SITE .....	20
1.3 LA DEMARCHE GEOSTATISTIQUE .....	21
1.4 APPLICATION AU SITE DE L'EST .....	23
1.4.1 Résumé des étapes de la démarche géostatistique .....	23
1.4.2 Modélisation structurale de l'aquifère .....	24
1.4.3 Simulation des faciès .....	25
1.4.4 Simulation des propriétés de la roche (perméabilité, porosité) .....	27
1.4.5 Changement d'échelle : remaillage et calcul de propriétés moyennes .....	33
1.4.6 Vers la modélisation des écoulements et du transport de radionucléides .....	37
2. MODELISATION STATISTIQUE DU TRANSPORT DE RADIONUCLEIDES : INTEGRATION D'INFORMATIONS PAR UNE APPROCHE GEOSTATISTIQUE .....	39
2.1 INTRODUCTION .....	41
2.1.1 Analyse de l'incertitude .....	41
2.1.2 Méthodes statistiques .....	43
2.1.3 Typologie de l'information .....	45
2.1.4 Interpolation spatiale .....	46
2.2 ALGORITHME D'INTERPOLATION .....	49
2.2.1 Régression traditionnelle .....	49
2.2.2 Le modèle (paradigme) de krigeage .....	49
2.2.3 Krigeage avec dérive externe .....	58
2.2.4 Cokrigeage .....	59
2.2.5 Cokrigeage co-localisé .....	62
2.2.6 Krigeage d'indicatrice .....	64
2.2.7 Cokrigeage d'indicatrice .....	67
2.2.8 Estimation par champ de probabilités .....	72
2.3 ESTIMATION : EXEMPLE D'APPLICATION .....	73
2.3.1 Introduction .....	73
2.3.2 Analyse des résultats .....	84
2.3.3 Evaluation finale .....	98
2.4 ALGORITHMES DE REPRESENTATION STOCHASTIQUE .....	99
2.4.1 Introduction .....	99
2.4.2 Simulation séquentielle .....	100
2.4.3 Simulation par recuit simulé (simulated annealing) .....	104
2.4.4 Simulation par champ de probabilités .....	106
2.4.5 Algorithmes booléens .....	108
2.4.6 Simulation par gaussiennes tronquées .....	109
2.5 SIMULATION : EXEMPLE D'APPLICATION .....	111
2.5.1 Principe de l'exercice .....	111
2.5.2 Analyse des résultats .....	146
CONCLUSIONS SOMMAIRES SUR LES METHODES D'ESTIMATION ET DE SIMULATION STOCHASTIQUE .....	149
REFERENCES BIBLIOGRAPHIQUES .....	152
3. PROBLEME INVERSE : RESOLUTION PAR UNE APPROCHE GEOSTATISTIQUE .....	156
3.1 INTRODUCTION .....	158
3.2 LINEAR METHODS .....	160
3.2.1 The linearized cokriging approach .....	160
3.2.2 The linearized semi-analytical approach .....	164

3.3	ITERATIVE NON-LINEAR METHODS .....	169
3.3.1	The Fast Fourier Transform Method .....	169
3.3.2	The iterative cokriging-like technique.....	172
3.4	ITERATIVE NUMERICAL NON-LINEAR METHODS.....	176
3.4.1	The zonation procedure of Carrera and Neuman .....	176
3.4.2	The Sequential Self Calibrated Method.....	181
3.4.3	The Pilot Point Method.....	185
3.4.4	The Markov Chain Monte Carlo Method .....	189
3.5	THE FRACTAL SIMULATION METHOD.....	192
	MAIN REFERENCES ON INVERSE MODELLING .....	194
	OTHER REFERRED LITERATURE .....	195
4.	CHANGEMENT D'ECHELLE DE PROPRIETES D'ECOULEMENT ET DE TRANSPORT .....	198
4.1	PERMEABILITES DE GRANDE ECHELLE EN MILIEU HETEROGENE .....	200
4.1.1	Introduction.....	200
4.1.2	Perméabilité de bloc.....	202
4.1.3	Evolution des méthodes de calcul de la perméabilité équivalente.....	203
4.1.4	Conclusions sommaires sur le calcul de perméabilités équivalentes.....	225
	REFERENCES BIBLIOGRAPHIQUES SUR LA PERMEABILITE EQUIVALENTE .....	227
4.2	PARAMETRES DE TRANSPORT DE GRANDE ECHELLE EN MILIEU HETEROGENE .....	231
	REFERENCE BIBLIOGRAPHIQUE.....	231
	ANNEXE 1 : VARIANCE D'UNE COMBINAISON LINEAIRE PONDEREE DE VARIABLES ALEATOIRES .....	233
	ANNEXE 2 : BREF APERÇU DU MODELE DE MARKOV.....	237
	ANNEXE 3 : PUBLICATION SUR LE MODELE DE MARKOV-BAYES.....	241
	ANNEXE 4 : PUBLICATION SUR LA THEORIE ET LA PRATIQUE DE LA SIMULATION SEQUENTIELLE.....	243

## LISTE DES FIGURES

Figure 2-1 : Exemple de diagramme de dispersion sur lequel la relation (régression) n'est pas linéaire. Les distributions conditionnelles de Z changent de forme, d'aplatissement et de variance selon les classes de V. Les histogrammes correspondent aux points à l'intérieur des bandes grisées pour V entre 200 et 400, à gauche, et pour V entre 600 et 800 à droite.....	48
Figure 2-2 : Base de données de Walker Lake. En haut : champ de la variable U (variable principale). En bas : champ de la variable V (variable secondaire). La même échelle de gris a été utilisée pour U et V.....	74
Figure 2-3 : Information disponible. L'image du haut montre l'emplacement des 30 données dures régulièrement espacées. L'image du bas indique la position des 780 données douces provenant d'une campagne sismique. ....	75
Figure 2-4 : Données de calibrage. L'image du haut montre l'emplacement des 224 données de calibrage et celle du bas le diagramme de dispersion correspondant. Les données de calibrage, plus nombreuses ici que les données dures, n'interviennent pas directement dans l'estimation de la variable principale. Elles servent uniquement à calculer les paramètres statistiques qui quantifient la corrélation entre les deux variables.....	76
Figure 2-5 : Variogrammes expérimentaux et modèles calés. Par colonne, les variogrammes dans les direction N180 (direction de plus grande continuité) et N72E (direction de plus petite continuité). Par ligne, les variogrammes de U, croisés et de V. ....	78
Figure 2-6 : Variogrammes d'indicateur expérimentaux et les modèles calés pour les seuils $z_1 = 0,46$ , $z_2 = 5,87$ et $z_3 = 19,02$ . Les variogrammes de gauche correspondent à la direction de plus grande continuité (N180) et ceux de droite à la direction de plus petite continuité (N72E). Les modèles calés sont représentés par une ligne continue et les variogrammes expérimentaux par une ligne discontinue. ....	81
Figure 2-7 : Variogrammes d'indicateur expérimentaux et les modèles calés pour les seuils $z_4 = 33,4$ , $z_5 = 61,52$ et $z_6 = 135,93$ . Les variogrammes de gauche correspondent à la direction de plus grande continuité (N180) et ceux de droite à la direction de plus petite continuité (N72E). Les modèles calés sont représentés par une ligne continue et les variogrammes expérimentaux par une ligne discontinue. ....	82
Figure 2-8 : Variogrammes d'indicateur expérimentaux et les modèles calés pour les seuils $z_7 = 226,03$ , $z_8 = 428,27$ et $z_9 = 758,96$ . Les variogrammes de gauche correspondent à la direction de plus grande continuité (N180) et ceux de droite à la direction de plus petite continuité (N72E). Les modèles calés sont représentés par une ligne continue et les variogrammes expérimentaux par une ligne discontinue. ....	83
Figure 2-9 : Champ interpolé de la variable U par krigeage simple avec moyenne globale tirée des 30 données dures. Dans cette estimation, l'information liée à la variable secondaire n'intervient pas.....	86
Figure 2-10 : Champ de la variable secondaire interpolé par krigeage simple des 780 données disponibles. Il constitue l'information douce pour toutes les méthodes d'interpolation qui nécessitent de connaître la valeur de la variable secondaire en chaque point d'estimation de la variable principale. ....	86
Figure 2-11 : Champ de la variable principale estimé par régression linéaire.....	87
Figure 2-12 : Champ de la variable principale estimé par krigeage avec dérive externe.....	87
Figure 2-13 : Champ de la variable principale estimé par cokrigeage complet .....	89
Figure 2-14 : Champ de la variable principale estimé par cokrigeage co-localisé .....	89
Figure 2-15 : Champ de la variable principale estimé par krigeage d'indicateur.....	91
Figure 2-16 : Champ de la variable principale estimé par cokrigeage d'indicateur selon un modèle de Markov-Bayes .....	91
Figure 2-17 : Champ de la variable principale estimé par champ de probabilités.....	92
Figure 2-18 : Histogrammes montrant les distributions de résidus pour chacune des méthodes d'estimation ...	95
Figure 2-19 : Diagrammes de dispersion des valeurs vraies et estimées de U .....	97
Figure 2-20 : Champ de transmissivité d'origine (haut) et champ piézométrique de référence (bas).....	113
Figure 2-21 : Champ de transmissivité d'origine (haut) et champ piézométrique de référence (bas) sur lesquels ont été portées les trajectoires de 20 particules pour le scénario 1.....	114
Figure 2-22 : Champ de transmissivité d'origine (haut) et champ piézométrique de référence (bas) sur lesquels ont été portées les trajectoires de 20 particules pour le scénario 2.....	115
Figure 2-23 : Champ de transmissivité d'origine (haut) et champ piézométrique de référence (bas) sur lesquels ont été portées les trajectoires de 20 particules pour le scénario 3.....	116
Figure 2-24 : Cas du scénario 1. A gauche, deux champs de transmissivité générés par simulation gaussienne séquentielle. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules. ....	117

Figure 2-25 : Cas du scénario 1. A gauche, deux champs de transmissivité générés par simulation gaussienne séquentielle par cokrigage co-localisé suivant un modèle markovien. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules	118
Figure 2-26 : Cas du scénario 1. A gauche, deux champs de transmissivité générés par simulation séquentielle d'indicatrice. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.	119
Figure 2-27 : Cas du scénario 1. A gauche, deux champs de transmissivité générés par simulation séquentielle par cokrigage d'indicatrice suivant un modèle markovien. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.	120
Figure 2-28 : Cas du scénario 1. A gauche, deux champs de transmissivité générés par simulation par champ de probabilités. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.	121
Figure 2-29 : Cas du scénario 1. A gauche, deux champs de transmissivité générés par simulation par recuit simulé. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.	122
Figure 2-30 : Cas du scénario 2. A gauche, deux champs de transmissivité générés par simulation gaussienne séquentielle. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.	123
Figure 2-31 : Cas du scénario 2. A gauche, deux champs de transmissivité générés par simulation gaussienne séquentielle par cokrigage co-localisé suivant un modèle markovien. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules	124
Figure 2-32 : Cas du scénario 2. A gauche, deux champs de transmissivité générés par simulation séquentielle d'indicatrice. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.	125
Figure 2-33 : Cas du scénario 2. A gauche, deux champs de transmissivité générés par simulation séquentielle par cokrigage d'indicatrice suivant un modèle markovien. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.	126
Figure 2-34 : Cas du scénario 2. A gauche, deux champs de transmissivité générés par simulation par champ de probabilités. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.	127
Figure 2-35 : Cas du scénario 2. A gauche, deux champs de transmissivité générés par simulation par recuit simulé. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.	128
Figure 2-36 : Cas du scénario 3. A gauche, deux champs de transmissivité générés par simulation gaussienne séquentielle. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.	129
Figure 2-37 : Cas du scénario 3. A gauche, deux champs de transmissivité générés par simulation gaussienne séquentielle par cokrigage co-localisé suivant un modèle markovien. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules	130
Figure 2-38 : Cas du scénario 3. A gauche, deux champs de transmissivité générés par simulation séquentielle d'indicatrice. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.	131
Figure 2-39 : Cas du scénario 3. A gauche, deux champs de transmissivité générés par simulation séquentielle par cokrigage d'indicatrice suivant un modèle markovien. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.	132
Figure 2-40 : Cas du scénario 3. A gauche, deux champs de transmissivité générés par simulation par champ de probabilités. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.	133
Figure 2-41 : Cas du scénario 3. A gauche, deux champs de transmissivité générés par simulation recuit simulé. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.	134
Figure 2-42 : Courbes des temps d'arrivée, sur la frontière Est, pour chaque méthode pour le scénario 1. La courbe en trait continu correspond au champ de transmissivité de référence, la courbe en pointillés aux temps	135
Figure 2-43 : Courbes des temps d'arrivée, sur la frontière Est, pour chaque méthode pour le scénario 2. La courbe en trait continu correspond au champ de transmissivité de référence, la courbe en pointillés aux	

temps moyens sur l'ensemble des réalisations et la courbe en tirets les temps extrêmes pour des probabilités de 5 et 90% (cf. note 1, p 75).....	136
Figure 2-44 : Courbes des temps d'arrivée, sur la frontière Est, pour chaque méthode pour le scénario 3. La courbe en trait continu correspond au champ de transmissivité de référence, la courbe en pointillés aux temps .....	137
Figure 2-45 : Histogrammes des positions d'arrivée, sur la frontière Est, pour le champ de transmissivité de référence et pour les 100 réalisations générées par chaque méthode de simulation pour le scénario 1...	138
Figure 2-46 : Histogrammes des positions d'arrivée, sur la frontière Est, pour le champ de transmissivité de référence et pour les 100 réalisations générées par chaque méthode de simulation pour le scénario 2...	139
Figure 2-47 : Histogrammes des positions d'arrivée, sur la frontière Est, pour le champ de transmissivité de référence et pour les 100 réalisations générées par chaque méthode de simulation pour le scénario 3...	140
Figure 2-48 : Histogrammes des temps d'arrivée, sur la frontière Est, pour le champ de transmissivité de référence et pour les 100 réalisations générées par chaque méthode de simulation pour le scénario 1...	141
Figure 2-49 : Histogrammes des temps d'arrivée, sur la frontière Est, pour le champ de transmissivité de référence et pour les 100 réalisations générées par chaque méthode de simulation pour le scénario 2...	142
Figure 2-50 : Histogrammes des temps d'arrivée, sur la frontière Est, pour le champ de transmissivité de référence et pour les 100 réalisations générées par chaque méthode de simulation pour le scénario 3...	143
Figure 4-1 : Processus de passage à grande échelle. a) Grille avec une résolution équivalente à l'échelle de la mesure et pouvant contenir plusieurs millions de cellules . b) Grille à l'échelle de la simulation numérique utilisée pour résoudre le problème d'écoulement en milieu poreux. Le calcul de valeurs de perméabilité de bloc en b) à partir des valeurs correspondantes en a) est l'objet du passage à grande échelle.....	201
Figure 4-2 : Conditions aux limites utilisées dans la méthode de résolution numérique simple. Ces conditions aux limites sont utilisées pour déterminer le terme $K_{V,xx}$ du tenseur diagonal de perméabilité de bloc. Après rotation à $90^\circ$ , les mêmes conditions aux limites servent à calculer le terme $K_{V,yy}$ .....	208
Figure 4-3 : Conditions aux limites employées dans l'approche par résolution numérique avec voisinage. Pour chaque type de conditions aux limites, l'écoulement est résolu à l'échelle de la mesure à l'intérieure de l'aire contenant le bloc et son voisinage. La taille de ce dernier est définie arbitrairement comme une demi taille de bloc dans chaque direction.....	211
Figure 4-4 : Grille élastique typique obtenue avec la méthode de Garcia (1990) avec pour objectif de minimiser les hétérogénéités de perméabilité de cellule à l'intérieur de chaque bloc. La grille à l'échelle de la mesure contient 200 par 150 cellules. La grille à l'échelle numérique est constituée de 20 par 30 blocs. Les cellules sombres indiquent de grandes valeurs de perméabilité. ....	219
Figure 4-5 : Grille élastique adaptée à une variable de flux pour un écoulement stationnaire provoqué par un puit injecteur et un producteur.....	221
Figure 4-6 : Approche en deux étapes contre approche directe pour le calcul de perméabilités de bloc. La ligne du haut décrit l'approche en deux étapes, celle du bas l'approche directe.....	223



## LISTE DES TABLEAUX

Tableau 2-1 : Paramètres du modèle de co-régionalisation pour la base de données de Walker Lake. $C_0$ est l'effet de pépite, $C_1$ le poids de la première structure imbriquée, $C_2$ le poids de la seconde structure, $a_1$ la portée de la première structure qui est isotrope, $\theta$ l'angle (mesuré dans le sens inverse des aiguilles d'une montre) de l'axe de plus petite continuité de la seconde structure avec l'horizontale (axe x), $a_2$ et $b_2$ les portées de corrélation de la seconde structure dans les directions de plus petite et plus grande continuité. .....	77
Tableau 2-2 : Paramètres des modèles calés sur les variogrammes expérimentaux en fonction du seuil. ....	80
Tableau 2-3 : Comparaison des distributions de valeurs vraies et estimées de $U$ par les 8 méthodes utilisées... ..	93
Tableau 2-4 : Déciles pour le champ de référence de $U$ et ceux estimés par les différentes méthodes.....	93
Tableau 2-5 : Comparaison des distributions de résidus pour les 8 méthodes d'estimation utilisées.....	96
Tableau 2-6 : Statistiques sur les positions d'arrivée pour les trois scénarios et chacune des méthodes de simulation stochastique. $N^\circ$ est le nombre de données, $m$ la moyenne, $\sigma$ l'écart-type, $CV$ le coefficient de variation, $max$ la valeur maximale, $Q75$ le quartile supérieur, $M$ la médiane, $Q25$ le quartile inférieur et $min$ la valeur minimale.....	144
Tableau 2-7 : Statistiques sur les temps d'arrivée pour les trois scénarios et chacune des méthodes de simulation stochastique. $N^\circ$ est le nombre de données, $m$ la moyenne, $\sigma$ l'écart-type, $CV$ le coefficient de variation, $max$ la valeur maximale, $Q75$ le quartile supérieur, $M$ la médiane, $Q25$ le quartile inférieur et $min$ la valeur minimale. ....	145



## PRESENTATION DE L'ETUDE

Cette étude a pour objet de réaliser une analyse bibliographique des méthodes géostatistiques adaptées à la caractérisation d'aquifères. Les modèles d'aquifères ainsi réalisés sont destinés à des études de sûreté liées au stockage souterrain de déchets radioactifs dans des formations peu perméables recouvertes par un aquifère (zone plus perméable). Les deux scénarios suivants sont considérés dans l'étude.

1. Un puits, foré à travers l'aquifère, atteint le stockage et crée ainsi une connexion directe le long de laquelle les radionucléides sont libres de se déplacer, depuis le stockage jusqu'à l'aquifère. Le puits peut alors être assimilé à une source ponctuelle de radionucléides dans l'aquifère. Ce scénario, qui doit être adapté aux spécificités des sites étudiés par l'ANDRA, s'apparente à celui étudié par Sandia National Laboratories pour le site américain de WIPP.
2. Il est supposé que des radionucléides migrent verticalement depuis le stockage jusqu'à l'aquifère sus-jacent. Cette migration crée une source diffuse de radionucléides au mur de l'aquifère.

Tout l'intérêt de la géostatistique est alors de pouvoir exploiter la variété des données disponibles pour générer des modèles de la répartition spatiale des propriétés de la roche dans l'aquifère (faciès, perméabilité, porosité, paramètres de transport). Le choix des méthodes géostatistiques dépend de la finalité de l'étude, du type de données disponibles et des corrélations qui peuvent exister entre les différentes informations ou variables étudiées. Généralement, des méthodes stochastiques, comme celles utilisées en simulation conditionnelle, sont préférées car elles permettent d'honorer à la fois des données et des modèles statistiques établis préalablement (histogrammes univariés ou multivariés, corrélations entre variables, corrélations spatiales).

Pour cette analyse bibliographique, les méthodes géostatistiques qui seront étudiées doivent répondre à un ou plusieurs des objectifs suivants :

- prise en compte à la fois d'informations directes (propriétés de la roche) et indirectes (attributs géophysiques) corrélées linéairement ou non,
- prise en compte de données piézométriques (problème inverse),
- prise en compte de résultats d'interprétation d'essais hydrogéologiques (problème inverse).

Font partie également de l'analyse bibliographique les techniques dites de changement d'échelle (*upscaling* en anglais) qui permettent de convertir un modèle géostatistique de haute résolution, pour lequel les propriétés de la roche ont été simulées sur une grille fine, en un modèle maillé beaucoup plus grossier et destiné à la simulation du transport.

A titre d'illustration, des résultats d'évaluations comparatives de méthodes géostatistiques sont présentés et discutés dans cette étude. Ce travail comprend également une analyse des méthodes géostatistiques les mieux adaptées aux études de site de l'ANDRA, c'est-à-dire pour des

informations (types et quantités) et des objectifs donnés. Les données du site de l'Est ont été retenues par l'ANDRA pour cet exercice.

Les résultats de cette étude proviennent de différents travaux menés par les personnes suivantes qui ont participé à l'étude :

- Michel Garcia (FSS International),
- Jaime Gómez-Hernández (consultant associé à FSS International et Professeur de géostatistique et d'hydrogéologie à l'Université Polytechnique de Valence — UPV — Espagne),
- Eduardo Cassiraga (en thèse à l'UPV avec des travaux sur l'intégration d'informations secondaires)
- Harrie-Jan Hendricks Franssen (en thèse à l'UPV avec des travaux sur les problèmes inverses).

Le rapport se divise en quatre parties comme suit.

1. Application : discussions autour des données du site de l'Est
2. Modélisation statistique du transport de radionucléides : intégration d'informations par une approche géostatistique
3. Problème inverse : résolution par une approche géostatistique
4. Changement d'échelle de propriétés d'écoulement et de transport

**1. APPROCHE GEOSTATISTIQUE :  
DISCUSSIONS AUTOUR DES DONNEES  
DU SITE DE L'EST**



## 1.1 Présentation du problème

La conception d'un stockage souterrain de déchets radioactifs s'appuie sur des études de sûreté dans lesquelles l'impact du stockage sur la géosphère ou sur les populations est estimé par rapport à différents scénarios. Dans la plupart de ces scénarios, le transport de radionucléides, à travers le milieu poreux ou fracturé que constitue le sous-sol, est une étape essentielle avant d'atteindre la géosphère et les individus.

Parmi tous les scénarios possibles, il en est deux qui présentent un intérêt particulier en raison du caractère aléatoire qu'ils revêtent. Ils se placent tous les deux dans le cas d'une formation hôte peu perméable et recouverte d'un aquifère par lequel des radionucléides pourraient transiter. Ces scénarios s'énoncent comme suit.

1. Un puits, foré à travers l'aquifère, atteint le stockage et crée ainsi une connexion directe le long de laquelle les radionucléides sont libres de se déplacer, depuis le stockage jusqu'à l'aquifère. Le puits peut alors être assimilé à une source ponctuelle de radionucléides dans l'aquifère. Ce scénario s'apparente à celui étudié par Sandia National Laboratories pour le site américain de WIPP.
2. Des radionucléides migrent verticalement depuis le stockage jusqu'à l'aquifère sus-jacent. Cette migration crée une source diffuse de radionucléides au mur de l'aquifère.

Pour ces deux scénarios, on souhaite évaluer le parcours des radionucléides depuis la source (ponctuelle ou diffuse) et le temps de parcours jusqu'à des points cibles (par ex. puits de pompage, exutoires naturels), ainsi que l'évolution au cours du temps de la concentration en radionucléides en ces points cibles. Compte tenu des nombreuses incertitudes d'ordre géologique (hétérogénéités du milieu), hydrogéologique (état actuel et évolutions futures des conditions aux limites de l'aquifère<sup>1</sup>) et de stockage (comportement des colis et des cavités de stockage), une approche stochastique est nécessaire pour quantifier l'incertitude sur le transport de radionucléides. C'est dans ce cadre que peut intervenir la géostatistique et les nombreuses méthodes qui sont présentées dans les chapitres suivants.

Dans ce chapitre, l'objectif est de présenter l'ensemble des étapes qui permettent de mener à bien une telle étude géostatistique. Afin d'illustrer la démarche, des données réelles sur un site potentiel étudié par l'ANDRA (site de l'EST) ont été utilisées. Pour ce cas d'étude particulier, des suggestions sont faites sur les approches géostatistiques qui permettraient d'exploiter au mieux les informations disponibles.

---

<sup>1</sup> Les conditions aux limites incluent les sources et exutoires extérieurs (par ex. cours d'eau, zones d'affleurement), les infiltrations et évaporations surfaciques aux endroits où la nappe est libre, et la drainance aux endroits où elle est confinée.

## 1.2 Présentation du site

Le site de l'EST a été choisi pour détailler l'approche géostatistique qui conviendrait pour modéliser un exemple de site de laboratoire ANDRA. Le site a été très largement décrit dans de nombreux documents référencés ("D.A.I.E – Site EST – Mémoire", "Recherches préliminaires à l'implantation des laboratoires de recherche souterrains – Bilan des travaux de janvier 1994 à mars 1996", rapport B RP O CGG 97-002/A intitulé "Inversion stratigraphique – Est France 97", rapport D RP 0ANT 97-058/A). On retiendra ici les quelques points suivants.

1. La formation hôte sélectionnée pour un stockage potentiel correspond aux Argilites du Callovo-oxfordien. Cette formation, d'une épaisseur moyenne de 130 m à l'endroit envisagé, est située sous l'Oxfordien calcaire qui constitue l'aquifère sus-jacent que l'on souhaiterait modéliser.
2. Cet aquifère est d'une épaisseur de 300 m au droit de l'emplacement prévu pour le laboratoire mais l'épaisseur cumulée des niveaux producteurs est plus faible. Cet emplacement se situe à quelques kilomètres à l'ouest d'une zone d'affleurement de l'Oxfordien calcaire où de nombreux puits de pompage existent le long de la Meuse en amont du site.
3. Cinq **sondages d'exploration**, réalisés par l'ANDRA à partir de 3 plates-formes, renseignent sur l'Oxfordien calcaire. Deux des plates-formes regroupent les sondages HTM102 et EST107, pour la première, et EST103 et EST104 pour la seconde. La dernière plate-forme ne comprend que le sondage MSE101. Sur tous ces sondages, des diagraphies renseignent sur la lithologie ainsi que sur la porosité. Des tests hydrauliques ont été réalisés dans l'aquifère sur un sondage de chaque plate-forme (MSE101, EST103 et HTM102). Pour chacun de ces sondages, plusieurs tests, de pompage ou par choc hydraulique, ont été effectués dans l'aquifère pour des intervalles différents entre obturateurs. L'interprétation de ces tests renseigne sur la perméabilité du milieu par intervalle. Deux de ces sondages (HTM102 et MSE101) ont été convertis en piézomètres d'où les variations de charge de la nappe sont suivies.
4. D'autres informations hydrogéologiques relatives à l'Oxfordien calcaire peuvent être tirées des **forages d'alimentation en eau** situés le long de la Meuse. Ces forages renseignent sur la perméabilité locale (données de production) et sur le niveau piézométrique de la nappe.
5. D'autres données proviennent de 68 **forages pétroliers** anciens sur lesquels des diagraphies avaient été enregistrées. Ces forages sont pour la plupart très éloignés du site (plus de 20 km) vers le nord-ouest. Si les diagraphies sont pertinentes et recouvrent l'Oxfordien calcaire, ces données constituent une information secondaire précieuse pour la caractérisation préalable de l'aquifère sachant que peu de données existent à l'endroit du site.
6. Des **données sismiques** existent également sur le site. Elles se composent de 1300 km de profils de sismique réflexion (2D)<sup>2</sup> dont 980 km ont été réinterprétés pour l'ANDRA. A cela se rajoutent 14,9 km de profils sismiques réalisés par l'ANDRA. Ces profils interprétés fournissent des temps de parcours pour différents marqueurs (y compris le toit et le mur de l'aquifère) et une impédance qui devrait être corrélée aux lithofaciès et à la porosité.

---

<sup>2</sup> Ces profils datent de campagnes d'exploration pétrolière anciennes.

### 1.3 La démarche géostatistique

La modélisation d'un aquifère passe par plusieurs étapes dans lesquelles différentes caractéristiques du milieu sont abordées. Dans ses grandes lignes, la modélisation comprend tout ou partie des étapes suivantes.

**1. Modélisation structurale : topographie du toit ou du mur de l'aquifère, épaisseur de l'aquifère.**

Une approche déterministe, du type estimation (interpolation), est généralement employée. Cette étape intervient dans tous les problèmes tridimensionnels (3D). En 2D, à moins de s'intéresser uniquement à des écoulements (**pas de transport**) à densité constante en nappe confinée ou à une nappe libre dont la hauteur saturée est grande, l'épaisseur de l'aquifère doit également être modélisée. En 2D, le mur de l'aquifère est aussi nécessaire pour modéliser des effets de densité (phénomènes de gravité).

**2. Modélisation d'unités géologiques ou de types de faciès : vers une discrimination des propriétés pétrophysiques du milieu à partir de la connaissance de la répartition spatiale des roches.**

L'incertitude sur cette répartition spatiale pouvant être importante, une approche stochastique est souvent requise. Le choix de l'approche dépend des objectifs recherchés (modélisation d'objets géologiques ou de répartitions spatiales de faciès) et de l'information disponible.

**3. Modélisation des propriétés pétrophysiques (par ex. porosité, perméabilité, paramètres de transport) : vers un remplissage des corps géologiques ou des types de faciès par des propriétés du milieu poreux.**

Cette fois encore, une approche stochastique est nécessaire pour mieux appréhender l'incertitude, pour générer des représentations réalistes de la répartition spatiale des propriétés et pour s'orienter vers des analyses de risque ou de sûreté. Plusieurs étapes peuvent alors être utiles pour intégrer l'ensemble des données disponibles, informations directes et secondaires, et informations physiques relatives aux écoulements (problème inverse).

**4. Changement d'échelle : passage d'un modèle géostatistique construit avec une grande résolution (grilles contenant jusqu'à plusieurs millions de blocs) à un modèle remaillé plus grossièrement (quelques milliers à dizaines de milliers d'éléments) destiné à la simulation des écoulements et du transport.**

Cette étape comprend la construction du maillage plus grossier et le calcul des propriétés moyennes (effectives) des éléments. Le type de maillage à générer dépend du simulateur d'écoulement et de transport employé (simulateur par différences finies standard, différences finies généralisées avec volumes de contrôle, éléments finis avec volumes de contrôle, éléments finis mixtes, etc.).

Dans chacune de ces étapes, l'**analyse exploratoire des données** est un prélude indispensable pour déduire de l'ensemble des informations disponibles, de nature quantitative ou qualitative, les modèles statistiques inhérents à toute approche probabiliste ou stochastique dont la géostatistique fait partie. Ces modèles incluent les modèles univariés (distributions) relatives à chacune des variables estimées ou simulées (variables structurales, types de faciès, propriétés pétrophysiques), les modèles bi ou multivariés entre variables primaires et secondaires et

l'établissement entre elles de corrélations qui peuvent être de nature linéaire ou non-linéaire, les modèles de variographie qui déterminent la continuité (ou corrélation) spatiale de chacune des variables du problème. Cet aspect de la démarche géostatistique n'est pas davantage détaillé dans la suite.

## 1.4 Application au site de l'EST

### 1.4.1 Résumé des étapes de la démarche géostatistique

La modélisation de l'aquifère a pour objet de simuler des scénarios autour d'un stockage souterrain de déchets radioactifs à savoir la mise en connexion du stockage et de l'aquifère par un puits qui court-circuite la formation hôte ou la migration verticale dans la formation hôte de radionucléides vers l'aquifère. Dans un premier temps, ces deux problèmes peuvent être simulés avec un modèle bidimensionnel d'écoulement et de transport pour représenter l'aquifère. Si, comme cela semble être le cas, le milieu est fortement hétérogène verticalement (grandes variations de propriétés sur la hauteur de l'aquifère), le passage à un modèle tridimensionnel devient inévitable.

La démarche géostatistique et l'utilisation des données dépendent de la dimension du problème. En particulier, s'il est envisageable en 3D de modéliser la répartition spatiale de types de faciès, et par conséquent d'utiliser l'information qui s'y rapporte telle que l'impédance, il n'en est pas de même en 2D où une notion de faciès moyen pris sur l'épaisseur de l'aquifère n'a pas de sens. En revanche, l'impédance peut servir en 2D directement comme une information secondaire sur la transmissivité.

Pour l'étude du site de l'EST, selon le choix d'un modèle bi ou tridimensionnel, les séquences de traitement possibles sont les suivantes.

#### *Étapes de construction d'un modèle bidimensionnel d'aquifère*

- 1) Estimation de l'épaisseur de l'aquifère.
- 2) Simulations stochastiques jointes (co-simulation) de la porosité et de la transmissivité ou de la perméabilité. Résolution simultanée ou consécutive du ou des problèmes inverses.
- 3) Si nécessaire et justifié, estimation ou simulation de paramètres de transport (par ex. dispersivité).
- 4) Si nécessaire, processus de changement d'échelle (choix et construction d'un maillage grossier et calcul de propriétés équivalentes — perméabilité, porosité — affectées aux éléments).

#### *Étapes de construction d'un modèle tridimensionnel d'aquifère*

- 1) Estimation du toit ou du mur de l'aquifère et de son épaisseur.
- 2) Simulation stochastique des types de faciès.
- 3) Simulations stochastiques jointes de la porosité et de la perméabilité. Résolution simultanée ou consécutive du ou des problèmes inverses.
- 4) Si nécessaire, simulation stochastique de paramètres de transport.
- 5) Si nécessaire, processus de changement d'échelle (choix et construction d'un maillage grossier et calcul de propriétés équivalentes — perméabilité, porosité — affectées aux éléments).

## 1.4.2 Modélisation structurale de l'aquifère

### *Objet*

Pour traiter le problème de transport de radionucléides, il est indispensable de connaître l'épaisseur de l'aquifère pour déduire, à partir d'un champ de transmissivité ou de perméabilité et d'un champ de pression, des variables d'écoulement telles que le flux et la vitesse. Perméabilité et épaisseur de l'aquifère sont utiles à la fois dans les problèmes bi et tridimensionnels. Pour un problème 3D, la topographie du mur de l'aquifère peut aussi être un élément d'information important s'il influe sur la migration de radionucléides. Tel est le cas si l'élévation du mur de l'aquifère détermine le temps d'arrivée de radionucléides depuis un stockage jusqu'à l'aquifère ou si un phénomène de gravité doit être pris en compte dans le transport<sup>3</sup>. Ce dernier cas (effet de densité) s'applique aussi en 2D; il faut alors connaître la géométrie de l'aquifère.

### *Données disponibles*

Pour le site de l'EST, l'information disponible sur la topographie du mur et du toit de l'aquifère (et donc sur l'épaisseur) provient de plusieurs sources.

- **Données de puits** : cotes relevées sur les sondages traversant l'aquifère. En nombre très réduit, ces données constituent l'information précise sur l'emplacement de l'aquifère.
- **Données sismiques** : marqueurs de toit et de mur de l'aquifère. Beaucoup plus abondantes que les précédentes, ces données indiquent, avec une certaine précision (généralement quelques mètres), la cote du toit et du mur de l'aquifère.
- **Autres données de type géologique** : de nature qualitative, ces données reflètent l'avis expert de géologues et servent dans le choix ou l'établissement de modèles statistiques sur la structure de l'aquifère ou dans l'énoncé de certaines hypothèses.

### *Méthodes préconisées*

Il est deux raisons qui justifient d'avoir recours à l'estimation (interpolation) plutôt qu'à la simulation pour générer un modèle structural d'aquifère.

1. La topographie et l'épaisseur des aquifères varient généralement en douceur (grandes portées de corrélation spatiale).
2. L'erreur commise sur le modèle a relativement peu d'importance sur la modélisation des écoulements. En d'autres termes, il ne faut pas s'attendre à des différences significatives dans les résultats d'écoulement pour des modèles structuraux équivalents (équiprobables) qui seraient obtenus par une approche stochastique.

---

<sup>3</sup> Un tel phénomène est à prendre en compte s'il influe sur la trajectoire des radionucléides, soit directement lorsque la force de gravitation à laquelle sont soumis les radionucléides est d'un même ordre de grandeur que les autres forces (capillaire et de viscosité), soit à travers la ségrégation de fluides comme pour les écoulements diphasiques où les radionucléides seraient transportés par la phase gaz générée à partir du stockage, soit encore s'il existe des variations de la densité de l'eau dans la formation.

Parmi les méthodes géostatistiques d'estimation capables d'intégrer une information secondaire, la plupart pourraient être utilisées ici. Certaines sont cependant mieux adaptées ou d'utilisation plus facile. Elles sont données ci-après.

- **Krigeage universel**

A utiliser si les données secondaires disponibles sont insuffisantes ou mal corrélées aux variables structurales de l'aquifère, et s'il peut être établi, à partir des quelques données disponibles ou sur l'avis de géologues, que les variations de grande échelle du toit ou du mur de l'aquifère ou de son épaisseur ont des tendances qui peuvent se décrire facilement à partir de lois polynomiales simples (par ex. variations autour d'un plan moyen incliné).

- **Krigeage avec dérive externe**

A utiliser s'il peut être établi une corrélation entre un marqueur sismique et une variable structurale (élévation du toit ou du mur, épaisseur) et si ce marqueur décrit les variations à grande échelle de la variable (variations douces du marqueur) sans qu'il ne soit possible, à partir du faible nombre de données directes disponibles, de quantifier convenablement la corrélation. Si l'information secondaire (sismique) n'est pas présente partout, il est nécessaire, dans un premier temps, d'interpoler l'information secondaire à l'ensemble du domaine et d'utiliser le champ de valeurs interpolées comme dérive externe pour l'estimation de la variable structurale.

- **Cokrigeage co-localisé**

A utiliser s'il a pu être établi et quantifié une corrélation linéaire entre l'attribut sismique et la variable structurale. Comme précédemment, si l'information secondaire n'existe pas partout, une première étape obligatoire consiste à l'interpoler.

- **Cokrigeage d'indicatrice selon un modèle de markov-bayes**

A utiliser si la corrélation entre l'attribut sismique et la variable structurale est non-linéaire ou partielle, c'est-à-dire si la précision de l'information secondaire, sur la variable primaire, varie selon sa valeur ou l'intervalle de valeurs auquel elle appartient.

### 1.4.3 Simulation des faciès

#### *Objet*

La modélisation de types de faciès ou de roches se justifie lorsque les propriétés du milieu poreux varient considérablement d'un faciès à l'autre et que l'information disponible permet de s'intéresser à eux. L'objectif n'est alors pas de modéliser les faciès à la seule fin de connaître ou de comprendre leur répartition spatiale, mais de générer un modèle intermédiaire (de faciès) à partir duquel les propriétés pétrophysiques du terrain pourront être décrites avec plus de précision et de réalisme. Comme cela a été évoqué plus haut (voir §1.4.1), la modélisation de faciès n'est envisageable que pour des modèles tridimensionnels d'aquifère.

Les conditions requises pour s'intéresser aux faciès sont donc les suivantes.

- Problème tridimensionnel.
- Possibilité de discrimination de populations de propriétés fondée sur une certaine classification des terrains (types ou groupes de faciès ou de roches).

- Disponibilité d'une information relative aux types de faciès ou de roche.

### *Données disponibles*

Dans le cas du site de l'EST, trois types d'information renseignent sur les faciès.

- **Coupes lithologiques** sur carottes (information directe).
- **Diagraphies** de puits (information secondaire relativement précise).
- **Impédances** (information secondaire moins précise mais plus largement échantillonnée).

Ces informations montrent une classification très riche des faciès en une dizaine de formations régionales décomposées chacune en une vingtaine ou plus de faciès. Si cette classification présente un intérêt géologique certain pour la compréhension de la genèse de l'aquifère, elle est trop complexe pour être modélisée par une approche géostatistique ou par toute autre approche, et quand bien même cela serait possible, une telle précision du modèle serait inutile dans le seul but de modéliser ensuite les propriétés de la roche puis les écoulements et le transport.

La simplification de cette classification des faciès doit être une des tâches principales de l'analyse exploratoire des données. Elle consistera à regrouper les faciès en familles sur des critères de propriétés et de variabilité spatiale équivalentes. Le nombre final de familles de faciès ne devrait pas dépasser la dizaine.

### *Méthodes préconisées*

Compte tenu du fait que les types de faciès ou de roche sont des variables catégoriques, que leur continuité spatiale peut être complexe et différente d'un type à l'autre, et que l'incertitude est grande, une approche par simulation stochastique est nécessaire pour modéliser leur répartition spatiale. Une telle approche conduit à générer un ensemble de réalisations (images) dites équiprobables, toutes similaires en apparence mais différentes dans le détail, et qui décrivent autant d'états (ou scénarios) possibles de la répartition spatiale des faciès. Le nombre de réalisations nécessaires pour décrire l'éventail des états possibles dépend de la complexité et de l'incertitude sur le phénomène simulé. Il est courant de générer 100 à 1000 réalisations pour échantillonner convenablement l'espace des états possibles.

Parmi les méthodes de simulation stochastique décrites dans la suite, une approche par indicatrices est mieux indiquée car elle permet de traiter directement des variables catégoriques. Pour exploiter la ou les informations secondaires disponibles, une méthode par indicatrices du type markov-bayes convient parfaitement. Elle peut intégrer des données secondaires de nature et donc de précision différente (diagraphies et impédances), et ce quelle que soit la complexité de la corrélation entre les faciès et les variables secondaires. Dans les cas favorables (formation sédimentaire avec une séquence stratigraphique de faciès identique partout), une méthode par gaussiennes tronquées conviendrait également.

Il est à noter que pour le site de l'EST, l'impédance n'est disponible qu'en un nombre réduit de profils (sismique 2D). Il pourrait donc être envisagé de l'interpoler sur l'ensemble du domaine comme cela a été évoqué précédemment pour les marqueurs sismiques (voir §1.4.2, modèle structural). Pour que l'interpolation se justifie, il faudrait que la variabilité spatiale de

l'impédance soit douce (grande portée de corrélation pour justifier d'une approche par estimation) et qu'elle renseigne donc sur des variations à grande échelle des faciès. Ce dernier point n'a bien entendu aucun sens lorsque l'on parle de variables catégoriques comme les faciès. L'unique solution est alors d'avoir recours à la simulation stochastique pour générer des champs d'impédance qui ensuite sont utilisés comme information secondaire dans la simulation des faciès. Pour que l'approche soit cohérente, il faut générer une nouvelle réalisation de l'impédance pour chaque nouvelle réalisation de faciès. Un tel traitement peut être long et coûteux en temps de calcul. Cette difficulté à générer une information secondaire exhaustive nous conduit ici à écarter les méthodes de simulation qui requièrent une telle information.

Une autre approche, envisageable pour conditionner ou plus précisément contraindre les simulations de faciès à partir des diagraphies et des impédances qui constituent des informations secondaires de nature différente et non uniformément réparties, est le recuit simulé. Il permet en effet de produire des réalisations qui satisfont à un ensemble de critères tant statistiques (reproduction d'histogrammes et de variogrammes) que liés aux données (directes, corrélations avec des variables secondaires). Si d'un point de vue performance le recuit simulé n'est pas nécessairement plus rapide que la simulation séquentielle d'indicateurs, et si son fondement probabiliste est moins bien établi, il reste un moyen unique de satisfaire à un ensemble de critères que la simulation séquentielle peut, dans certains cas, ne pas pouvoir atteindre.

#### **1.4.4 Simulation des propriétés de la roche (perméabilité, porosité<sup>4</sup>)**

##### ***Objet***

Les propriétés de la roche restent l'objectif principal de la modélisation de l'aquifère. Elles conditionnent directement les phénomènes d'écoulement et de transport de radionucléides qui, du point de vue hydrogéologique, sont essentiels lors de l'étude d'un stockage. Dans un premier temps, la modélisation porterait sur les propriétés d'écoulement qui sont importantes (perméabilité, porosité, etc.). Dans un deuxième temps, il serait utile d'analyser également le transport réactif qui requiert des paramètres supplémentaires et parfois une formulation différente des équations de transport couplées avec les équations d'équilibre chimiques. L'aspect d'interpolation ou de simulation de paramètres de transport tels que la dispersivité n'est pas abordé dans la suite. En effet, les phénomènes de dispersion les plus importants proviennent de la dispersion hydrodynamique produite par les hétérogénéités spatiales. Cette dispersion se calcule au moment du changement d'échelle entre une grille géostatistique fine et un maillage plus grossier destiné au simulateur d'écoulement et de transport.

##### ***Données disponibles***

Concernant le site de l'EST, les données disponibles sur les propriétés du milieu dépendent de la dimension choisie pour étudier le problème d'écoulement et de transport. Deux cas se distinguent comme suit.

#### **Modélisation 2D de l'aquifère**

---

<sup>4</sup> Egalement le coefficient d'emmagasinement dans la partie où la nappe est confinée et la porosité de drainage dans celle où elle est libre.

En 2D, les propriétés à simuler sont la transmissivité de l'aquifère<sup>5</sup> et la porosité moyenne sur l'épaisseur de l'aquifère. Les informations disponibles sont alors de six ordres (on se place ici dans le cas particulier du site de l'EST pour lequel aucune mesure de porosité ou de perméabilité sur carotte n'a été effectuée).

- **Données de diagraphie :** déduction de porosités moyennes sur la hauteur de l'aquifère; bien que mesurée aux puits, cette information présente une imprécision inhérente aux diagraphies qui ne mesurent pas directement la porosité mais des attributs bien corrélés à la porosité (par ex. mesures par neutrons).
- **Interprétations d'essais de puits :** selon le type d'essai (sur l'Oxfordien calcaire des essais de pompage et des essais par choc hydraulique ont été réalisés), il peut être déduit une transmissivité ou, de manière équivalente, une perméabilité, et, pour les essais par choc hydraulique (essais en écoulement transitoire), un coefficient d'emmagasinement. Les valeurs de transmissivité ou de perméabilité sont des valeurs moyennes sur des intervalles qui recouvrent tout ou partie de la hauteur de l'aquifère à l'endroit du puits. Elles constituent donc une information directe, bien que parfois incertaine<sup>6</sup>, pour la simulation de champs de perméabilité. Si les volumes d'investigation, auxquels sont attachées les transmissivités déduites des essais, sont beaucoup plus grands que les cellules de la grille géostatistique, ces valeurs ne sont pas utilisées pour conditionner des cellules uniques mais pour contraindre la perméabilité moyenne à l'intérieur des volumes d'investigation des essais. Concernant le coefficient d'emmagasinement, il est lié à la porosité et peut donc être utilisé pour contraindre l'intervalle de valeurs possibles de porosité à l'endroit du puits. Si des données de diagraphie existent sur le puits, l'information déduite du coefficient d'emmagasinement vient en complément ou pour valider la distribution de porosité tirée des diagraphies.
- **Données piézométriques :** en tant que mesures de la charge hydraulique en un ou plusieurs puits d'observation, ces données renseignent, pour des conditions aux limites données, sur la répartition spatiale de la transmissivité (problème inverse). Les niveaux piézométriques peuvent être relativement stationnaires (cas du site de l'EST) ou révéler des variations saisonnières (phénomènes transitoires). Dans les deux cas, il est souhaitable de générer des champs de perméabilité qui reproduisent les niveaux piézométriques mesurés ou leur historique. Si les conditions aux limites sont mal connues (conditions qui peuvent varier au cours du temps), il est nécessaire de quantifier cette incertitude et de la prendre en compte dans la simulation de la perméabilité. Il en est de même des infiltrations, des évaporations ou de la drainance potentielle de l'aquifère.

---

<sup>5</sup> Si, comme prévu, la première étape conduit à construire un modèle structural — d'épaisseur — de l'aquifère, il peut être plus intéressant de modéliser la perméabilité horizontale de l'aquifère — perméabilité moyenne sur toute la hauteur — si elle varie indépendamment de l'épaisseur de l'aquifère. En effet, en procédant de la sorte, la variabilité spatiale de la perméabilité, alors différente de celle de la transmissivité, peut mieux être reproduite.

<sup>6</sup> Il est courant de déduire d'un essai de puits une valeur de perméabilité affectée d'une erreur. Si cette incertitude sur les données a des conséquences significatives sur le problème d'écoulement et de transport à traiter, elle peut aussi être intégrée au processus de simulation stochastique.

- **Données d'essai de puits :** lors d'un essai, les mesures de débit et de charge hydraulique relevées au cours du temps, sur un ou plusieurs puits de sollicitation ou d'observation, constituent une information partielle sur les écoulements provoqués par l'essai (et qui se superposent aux écoulements régionaux). Ces écoulements dépendent à la fois des conditions aux limites qui, dans le cas présent, sont bien connues, de la répartition spatiale de la perméabilité (ou de la transmissivité) dans un voisinage plus ou moins immédiat des puits de sollicitation, et du coefficient d'emmagasinement de la nappe dans ce même voisinage (et donc de la porosité moyenne). Il s'agit donc à nouveau d'un problème inverse à intégrer à la simulation de la perméabilité.
- **Données géophysiques :** il a été dit plus haut (voir §1.4.1) que les impédances déduites des profils sismiques ne peuvent pas être utilisées en 2D pour modéliser la répartition spatiale de types de faciès ou de roche. La porosité et la perméabilité sont cependant très fortement corrélées aux faciès. S'il n'est pas possible d'utiliser l'impédance pour déduire un faciès moyen, qui n'a pas de sens, sur la hauteur de l'aquifère, elle peut être utilisée comme information secondaire sur la perméabilité ou sur la porosité moyenne. La manière de corréler cette information géophysique aux propriétés du milieu poreux demande une analyse détaillée des données.
- **Informations d'ordre géologique :** il est bon de générer des champs de perméabilité et de porosité qui sont conformes aux interprétations que les géologues ont pu rapporter de leur travail sur le terrain. Par exemple, pour le site de l'EST, il est dit dans les rapports d'étude de l'ANDRA que l'Oxfordien calcaire a une bonne perméabilité en nappe libre à l'affleurement et que la perméabilité diminue avec l'approfondissement de la formation et le recouvrement par les marnes kimméridgiennes. Une telle information est utile dans l'analyse des données car elle indique des voies d'exploration pour l'établissement de corrélations entre la perméabilité et des informations secondaires (en particulier des données sismiques ou structurales). A défaut d'information secondaire, elle peut être utilisée directement pour spécifier un modèle de dérive de la perméabilité.

### Modélisation 3D de l'aquifère

En 3D, les notions de propriétés moyennes sur la hauteur de l'aquifère ou de transmissivité disparaissent. Elles donnent place à des notions de perméabilité et porosité ponctuelles ou à l'échelle de blocs.

Les sources d'information décrites plus haut, pour un modèle 2D, restent valables en 3D. Leur utilisation peut être cependant différente. A ces informations s'en ajoutent d'autres comme celles relatives aux faciès qui ne peuvent pas être exploités en 2D. Les données disponibles pour un modèle 3D sont résumées ci-après.

- **Données de diagraphie :** déduction de porosités (moyennes) le long des puits<sup>7</sup>. Le problème d'imprécision de la mesure décrit en 2D s'applique encore ici. S'il est envisagé d'incorporer une incertitude sur ces valeurs de porosité dans le processus stochastique, il

---

<sup>7</sup> La grille employée pour la simulation des propriétés, et plus particulièrement le nombre de nœuds en Z, détermine la taille des intervalles sur lesquels doit être moyennée la porosité.

peut être nécessaire de corréliser entre elles les perturbations (aléatoires) apportées à chaque donnée.

- **Interprétations d'essais de puits :** déduction d'une perméabilité<sup>8</sup> et, pour les essais par choc hydraulique, d'un coefficient d'emménagement lié à la porosité. Les valeurs de perméabilité et de porosité ainsi déduites correspondent cependant à des valeurs moyennes à une échelle (volume d'investigation de l'essai) généralement supérieure à celle définie par la grille de simulation. Selon le cas, ces données peuvent être prises comme données ponctuelles (c'est-à-dire à une échelle équivalente à celle de la grille) ou comme données moyennes sur des volumes plus grands. Dans le premier cas, l'information intervient directement dans la simulation de la perméabilité comme une donnée directe. Dans le deuxième cas, la perméabilité moyenne doit être équivalente à celle que donnerait la répartition spatiale des valeurs de perméabilité simulées à l'échelle fine si l'essai devait être simulé numériquement. Ceci revient alors à traiter un problème inverse.
- **Données piézométriques :** utilisation identique au cas bidimensionnel.
- **Données d'essai de puits :** utilisation identique au cas bidimensionnel.
- **Simulations de faciès :** utilisation des simulations réalisées à l'étape 2 (§1.4.3). Tout l'intérêt des faciès est de pouvoir mieux discriminer des populations de propriétés. Les simulations de faciès sont générées à cette fin et constituent donc ici l'information secondaire principale.
- **Données géophysiques :** les simulations de faciès capitalisent déjà l'information géophysique. Comme la perméabilité et la porosité sont mieux corrélées aux types de faciès qu'à l'impédance, l'information géophysique n'a ici plus grande utilité.
- **Informations d'ordre géologique :** l'utilité de ces informations dans l'analyse exploratoire des données demeure au passage au 3D. En outre, si certaines interprétations géologiques permettent d'établir des modèles de dérive, ils devront être intégrés à la simulation des propriétés du milieu.

### *Méthodes préconisées*

La perméabilité et la porosité sont deux attributs fortement corrélés. Il est donc vraisemblable qu'ils soient simulés conjointement. Le choix d'une méthode de simulation dépend alors des aspects suivants :

---

<sup>8</sup> Bien que cela ne soit pas le cas ici, il est à noter que certains essais, mettant en œuvre des écoulements sphériques à partir de lanternes de petite taille, peuvent conduire à évaluer le rapport des perméabilités horizontale et verticale, c'est-à-dire le coefficient d'anisotropie de la perméabilité. A partir de la simulation de la perméabilité horizontale, le champ de perméabilité verticale se déduit alors par multiplication des valeurs de perméabilité horizontale par le coefficient d'anisotropie. Si plusieurs coefficients d'anisotropie ont pu être calculés sur différents puits, une première étape consiste à interpoler (par ex. estimation par krigeage ordinaire) ces valeurs aux nœuds de la grille de simulation. La perméabilité verticale se déduit alors, en chaque nœud de la grille, par multiplication de la perméabilité horizontale simulée par le coefficient d'anisotropie estimé. L'étape d'estimation des coefficients d'anisotropie se justifie ici par le fait que ces coefficients montrent généralement des variations douces dans l'espace (grande continuité spatiale).

- dimension du problème (2D ou 3D),
- parmi la multitude d'informations disponibles, la sélection de données qui contribuent véritablement à renseigner sur la perméabilité ou sur la porosité, sans être redondantes avec d'autres informations ou de moins bonne qualité,
- types de corrélation (linéaires, non-linéaires, partielles) qui ont pu être établies entre propriétés et informations secondaires,
- résolution simultanée ou a posteriori du ou des problèmes inverses.

Certaines méthodes de résolution de problèmes inverses (voir §3) sont conçues pour pouvoir simuler un champ de perméabilité en une seule étape à partir de données directes de perméabilité et de données d'écoulement. A trop vouloir intégrer d'informations d'un coup, ces méthodes tendent à produire des champs de perméabilité qui n'honorent pas pleinement toutes les données. En outre, et bien qu'il soit envisageable de les améliorer, ces méthodes ne prévoient pas d'intégrer d'informations secondaires, informations qui peuvent être corrélées de façon plus ou moins complexes à la perméabilité<sup>9</sup>.

Une alternative est de simuler les champs de perméabilité en deux étapes. Dans un premier temps, les propriétés sont simulées indépendamment des données d'écoulement mais en intégrant l'ensemble des données primaires et secondaires. Dans un deuxième temps, les simulations sont améliorées en résolvant le ou les problèmes inverses. Cette démarche permet de mieux intégrer toute une variété de données, de signification et d'échelles différentes, dans le processus de simulation stochastique de la perméabilité et de la porosité. Il permet en outre de mieux distinguer l'apport de chaque type d'information (statique et dynamique), de mettre à jour facilement le modèle avec de nouvelles données (en particulier des données d'écoulement) et plus généralement de mieux conditionner ou contraindre la simulation<sup>10</sup>.

Si on se place dans cette optique, il reste à choisir une méthode de simulation initiale pour exploiter l'ensemble des données à l'exception de celles d'écoulement.

### **Modélisation 2D de l'aquifère**

Si on considère qu'entre les données géophysiques, le modèle structural d'épaisseur et les interprétations d'ordre géologique, il a pu être tiré une information secondaire fiable ou un modèle de dérive, la méthode de simulation doit permettre de l'intégrer. Dans le cas le plus simple, pour des corrélations linéaires entre attributs primaire et secondaire, et pour un modèle unique de variogramme pour décrire la corrélation spatiale indépendamment de classes de pro-

---

<sup>9</sup> Il est à noter que l'intégration d'une information secondaire a été rendue possible très récemment dans la méthode d'auto-calibrage. Cette amélioration n'a pas pu être présentée dans ce rapport mais a fait l'objet d'une publication récente : Capilla, J. E., Rodrigo, J., et Gomez-Hernandez, J. J., "Worth of secondary data compared to piezometric data for the probabilistic assessment of radionuclide migration", *Stochastic Hydrology and Hydraulics*, vol 12, no 3, pp 171-190, 1998. Ces mêmes ont soumis également un autre papier intitulé "Simulation of non-gaussian transmissivity fields honoring piezometric data and integrating soft and secondary information" qui sera publié en 1999 dans la revue *Mathematical Geology*.

<sup>10</sup> Ces objectifs sont précisément ceux qui ont motivé les cas tests 3 et 4 dans l'exercice du WIPP pour Sandia Laboratories.

priétés<sup>11</sup>, une méthode de simulation de type multigaussien peut être employée. S'il est souhaitable de distinguer des classes de propriétés (par ex. des classes de perméabilité faible, moyenne et forte), une méthode par indicatrices est beaucoup mieux adaptée. Si différentes informations secondaires sont à prendre en compte (diagraphies, impédances, épaisseur, autre dérive), la forme de corrélation avec les propriétés ou l'incertitude pouvant varier d'une information à l'autre, une approche par indicatrices de type markov-bayes présente des avantages certains. Pour certaines données, une alternative pour traiter l'incertitude est de tirer aléatoirement, avant chaque simulation, des valeurs à partir de distributions qui décrivent l'incertitude sur ces données<sup>12</sup>. Ces tirages peuvent être indépendants d'une donnée à l'autre ou corrélés.

### **Modélisation 3D de l'aquifère**

Pour le 3D, l'objectif premier consiste à exploiter au mieux l'information sur les faciès. Quelle que soit la méthode de simulation employée, il est fréquent, pour chaque réalisation, de simuler les propriétés partout autant de fois qu'il y a de types de faciès, puis de combiner les simulations en ne gardant en chaque nœud que la valeur de propriété obtenue sur la simulation qui a le bon type de faciès. Pour une simulation à l'intérieur d'un type de faciès donné, ne sont pris en compte que les données relatives à ce type de faciès et les modèles statistiques qui s'y rapportent (distributions a priori, variogrammes).

Pour le site de l'EST, si l'information disponible permet visiblement d'affecter des données de porosité aux types de faciès, l'exercice semble plus compliqué avec les perméabilités pour lesquelles on ne dispose pas de données à petite échelle. Les perméabilités tirées des essais de puits portent en général sur plusieurs types de faciès en même temps. La solution peut alors être de simuler d'abord la porosité puis d'en déduire la perméabilité. Une telle approche justifie de traiter les problèmes inverses dans une étape ultérieure.

Pour le choix de la méthode de simulation il faut tenir compte encore de deux autres particularités de ce cas d'étude. D'une part, les données, tant de perméabilité que de porosité, présentent une incertitude<sup>13</sup>. D'autre part, il peut être utile d'intégrer d'autres informations secondaires (par ex. un modèle de dérive pour tendre à générer des perméabilités plus faibles pour des profondeurs plus grandes du toit de l'aquifère). Selon la complexité des incertitudes et des relations avec l'information secondaire, une méthode de simulation plutôt qu'une autre sera préférée. A performances (et cohérences) égales, la préférence sera donnée à la méthode la plus simple (par ex. entre une méthode multigaussienne ou par indicatrices, la première sera choisie). Si la complexité du problème est grande, le choix se portera davantage sur une méthode plus générale comme par indicatrices du type markov-bayes.

---

<sup>11</sup> Lorsque la perméabilité dépend d'entités géologiques particulières (par ex. fracturation locale, présence de karsts), la continuité spatiale peut être différente pour des valeurs de perméabilité faibles, moyennes et fortes. Un modèle unique de variogramme ne permet pas alors de décrire les diverses structurations spatiales de l'attribut simulé (voir Gomez Hernandez et Journel, 1993).

<sup>12</sup> Voir par exemple les résultats de l'étude réalisée par Intera pour ANDRA (rapport B RP OINA 96.016/A).

<sup>13</sup> Pour le site de l'EST, comme aucune mesure sur carotte n'existe pour les calcaires de l'Oxfordien, les données de perméabilité sont déduites d'essais de puits et les données de porosité de diagraphies. Dans les deux cas, l'incertitude sur la mesure peut ne pas être sans conséquence sur la modélisation de l'aquifère.

### 1.4.5 Changement d'échelle : remaillage et calcul de propriétés moyennes

Les étapes décrites aux paragraphes précédents avaient pour objet de construire un modèle d'aquifère qui honore l'ensemble des données disponibles et reproduise un ensemble de caractéristiques statistiques (modèles statistiques *a priori* inférés des données ou d'informations autres) qui font du modèle une représentation réaliste de l'aquifère. Dans l'approche stochastique qui s'avère nécessaire à plusieurs étapes de la modélisation, ce n'est pas un modèle unique d'aquifère qui est construit mais une infinité de modèles ou chaque modèle est une réalisation possible (équiprobable) de la réalité. Les modèles d'aquifère ainsi produits ne sont pas une fin en soi. Ils sont destinés à la simulation de phénomènes d'écoulement et de transport qui permettront de quantifier la migration de radionucléides selon différents scénarios d'évolution du stockage souterrain de déchets radioactifs.

Bien que cela ne soit pas requis, les programmes géostatistiques d'estimation ou de simulation stochastique travaillent généralement sur des grilles régulières<sup>14</sup> pour augmenter les performances des algorithmes de calcul<sup>15</sup>. Le choix de la grille dépend alors de l'échelle de représentativité des données<sup>16</sup>, de la continuité spatiale des variables estimées ou simulées<sup>17</sup> (types de faciès, propriétés) et de contraintes informatiques (mémoire de la machine, temps de calcul).

Dans la pratique, les grilles employées contiennent plusieurs dizaines de milliers à plusieurs millions voire dizaines de millions de blocs. A titre d'exemple, pour le site de l'EST, le domaine à mailler représente environ une superficie de  $50 \times 70 \text{ km}^2$  sur une hauteur de 300 m. En misant sur des portées de corrélation supérieures à 300 m horizontalement et à 30 m verticalement pour toutes les variables, des tailles de bloc de  $100 \text{ m} \times 100 \text{ m} \times 10 \text{ m}$  conviendraient et correspondraient à une grille de  $50 \times 70 \times 30 = 105\,000$  blocs. En diminuant par deux les tailles de bloc pour se ramener à  $50 \text{ m} \times 50 \text{ m} \times 5 \text{ m}$ , le nombre total de blocs de la grille est multiplié par 8 pour atteindre 840 000.

---

<sup>14</sup> On entend ici par grilles régulières toute une variété de grilles définies simplement à partir d'un nombre de nœuds (ou blocs) dans trois directions principales et une transformation géométrique simple. Cette catégorie de grilles inclut les grilles cartésiennes (remplissage d'un domaine rectangulaire), les grilles structurales (grilles cartésiennes projetées sur le toit ou le mur d'une séquence géologique), les grilles stratigraphiques (grille remplissant une séquence géologique définie par sa carte d'épaisseur ou par son toit et son mur) et les grilles curvilignes (par ex. grilles remplissant un chenal et dont les directions sont alignées ou perpendiculaires à la direction — curviligne — du chenal).

<sup>15</sup> Les performances sont la seule justification de l'utilisation de grilles régulières tant que l'on s'intéresse à l'estimation ou à la simulation de valeurs en des points du domaine. Si, comme cela est souvent le cas, les valeurs aux points sont assimilées à des valeurs de blocs centrés sur les points, la taille des blocs doit être conforme au support de l'information et constante partout. Si tel n'est pas le cas, une fonction aléatoire différente de la variable estimée ou simulée doit être définie pour chaque taille de bloc qui aura sa propre distribution de valeurs (en particulier une moyenne et une variance spécifiques) et sa propre portée de corrélation. Des relations analytiques simples entre moyennes, variances et portées de corrélation, à toutes les échelles, peuvent alors être établies si le changement d'échelle de la variable considérée suit une loi linéaire.

<sup>16</sup> Mesures sur carottes, données tirées de diagraphie, propriétés déduites d'essais de puits.

<sup>17</sup> Il existe des règles empiriques comme celle de choisir un espacement entre les nœuds inférieur ou égal au tiers de la portée de corrélation de la variable étudiée.

De telles résolutions sont très supérieures à ce que la plupart des codes de calcul d'écoulement et de transport peuvent accepter en entrée<sup>18</sup>. Il est alors nécessaire de remailler le domaine avec moins d'éléments de plus grande taille. Chaque élément de la grille grossière contient plusieurs blocs de la grille fine utilisée pour le modèle stochastique. Si les blocs de la grille fine situés à l'intérieur d'un élément du maillage grossier n'ont pas tous les mêmes valeurs de propriétés, des propriétés moyennes doivent être calculées et affectées à l'élément. Ces étapes de construction d'un maillage grossier et de calcul de propriétés moyennes d'éléments grossiers constituent le changement d'échelle (*up-scaling* en anglais) du modèle d'aquifère. Dans une approche stochastique, l'opération doit être répétée pour chaque réalisation de l'aquifère. A ce titre au moins, elle doit être automatisée.

Le changement d'échelle comprend donc deux phases successives :

1. construction du maillage grossier,
2. calcul de propriétés moyennes d'éléments,

avec comme objectifs :

- réduire le nombre d'éléments (ou blocs) par rapport à la grille fine,
- conserver, au niveau du maillage grossier, les caractéristiques principales qui font du modèle stochastique fin une représentation réaliste de l'aquifère étudié. Ces caractéristiques incluent la variabilité (contrastes) de propriétés<sup>19</sup> et leur continuité spatiale et dépendent de l'utilisation du modèle maillé<sup>20</sup>.

### ***Choix d'un type de maillage grossier***

Le choix du maillage grossier dépend pour beaucoup du simulateur d'écoulement et de transport utilisé, ou plus précisément du schéma numérique sur lequel est fondé le simulateur. Pour un simulateur par différences finies, le maillage sera un maillage structuré (grille) constitué d'éléments hexaédriques (boîtes), alors que pour un simulateur fondé sur les éléments finis n'importe quel maillage structuré ou non structuré (par ex. constitué de tétraèdres ou de prismes à base triangulaire) fera l'affaire. Pour plus d'informations sur le choix d'un maillage, sa signification et son utilisation, le lecteur pourra se reporter à la partie du rapport consacrée au changement d'échelle (§4).

---

<sup>18</sup> Il est rare qu'un code de transport tourne sur des maillages de plus de quelques dizaines de milliers d'éléments.

<sup>19</sup> En passant d'un modèle fin à un modèle grossier d'aquifère, le calcul nécessaire de propriétés moyennes tend à lisser les valeurs de propriétés du modèle à l'échelle grossière et à diminuer les contrastes. Dans des cas extrêmes où, par exemple, la perméabilité prend uniquement des valeurs très faibles (matrice rocheuse) ou très fortes (chenaux, karsts), le changement d'échelle peut conduire à générer des éléments grossiers affectés d'une perméabilité moyenne, ni faible ni forte, **qui n'a aucune réalité physique**. Si ces valeurs moyennes permettent de satisfaire à des critères de flux moyens équivalents à l'échelle des éléments grossiers, elles conduisent à des champs de vitesse erronés.

<sup>20</sup> Les valeurs moyennes de perméabilité d'éléments grossiers ne seront pas les mêmes si on souhaite reproduire des flux moyens ou des voies d'écoulement rapide.

### *Construction du maillage grossier*

La première phase de construction du maillage joue un rôle essentiel sur la qualité du modèle grossier d'aquifère généré. En effet, s'il est toujours possible de calculer des propriétés moyennes d'éléments, et si ces propriétés satisfont à certains critères qui se résument à la conservation de flux pour des écoulements permanents à l'échelle de l'élément grossier, les propriétés moyennes des éléments sont d'autant plus proches des propriétés réelles du milieu que les éléments tendent à être homogènes. En d'autres termes, les propriétés moyennes d'un élément grossier seront plus convenables et réalistes si l'élément contient des blocs de la grille fine dont les propriétés sont relativement identiques.

Dans la phase de construction du maillage grossier, il sera donc recherché la génération d'éléments les plus homogènes possibles. Cette construction intervient dans un processus de maillage adaptatif au cours duquel les hétérogénéités de l'aquifère doivent être d'une certaine manière reconnues.

Pour un problème d'écoulement et transport, toutes les hétérogénéités du milieu poreux ne participent pas, cependant, au phénomène étudié, ou n'ont pas une influence identique sur ce phénomène. Par exemple, dans le cas du premier scénario où un puits met en connexion le stockage avec l'aquifère, s'il existe dans l'aquifère une zone fortement perméable, relativement éloignée du puits et non connectée à celui-ci<sup>21</sup>, cette zone aura très peu d'influence sur la migration des radionucléides et ne nécessite donc pas d'être modélisée avec précision. Compte tenu du nombre (crédit) limité d'éléments du maillage grossier, il est primordial de reconnaître au mieux les hétérogénéités qui agissent sur le transport au dépens de celles qui ont une influence moindre.

Dans un cas d'étude comme celui du site de l'EST, avec comme objectifs la simulation des deux scénarios décrits au paragraphe §1.1, la position des conditions aux limites pour l'écoulement et le transport est bien connue et invariante au cours du temps. Il est donc souhaitable de pouvoir en tenir compte pour construire le maillage grossier et l'adapter aux hétérogénéités importantes. Différentes variables d'écoulement ou de transport, calculées sur la grille fine, peuvent être utilisées pour l'adaptation du maillage. Une liste, non-exhaustive, est donnée ci-après.

- **Variable de connectivité** : variable booléenne (valeur 0 ou 1) qui détermine si un nœud de la grille fine est connecté ou non à une ou plusieurs conditions aux limites (pour l'écoulement, le transport ou les deux) et à des points cibles. La connectivité se calcule à partir des valeurs de perméabilité simulées sur la grille fine pour un seuil de perméabilité donné (tous les nœuds connectés doivent avoir une perméabilité supérieure au seuil fixé). Dans le cas le plus simple, elle porte sur tous les nœuds connectés à la source de radionucléides, ponctuelle ou diffuse selon le scénario traité. Lorsque les points cibles ne sont pas clairement établis ou peuvent varier au cours du temps, il est difficile de faire mieux. Si tel n'est pas le cas, pour mieux préciser les zones d'intérêt, la connexion peut porter à la fois

---

<sup>21</sup> La notion de connectivité est importante dans les problèmes d'écoulement et de transport. Elle signifie ici qu'il existe, entre deux points ou régions de l'aquifère (par ex. le puits qui met en connexion le stockage et l'aquifère, et la zone de forte perméabilité mentionnée dans le texte), un chemin continu le long duquel la perméabilité n'est jamais inférieure à un seuil donné.

sur la source et sur un ou plusieurs points cibles. Un nœud de la grille est alors mis à un 1 s'il existe un chemin continu de forte perméabilité qui relie la source à un point cible en passant par le nœud. Le critère d'adaptation est alors de rendre les éléments du maillage grossier les plus homogènes possibles, c'est-à-dire de positionner chacun au mieux pour qu'il soit entièrement à l'intérieur d'une zone de 0 ou de 1.

- **Variable d'écoulement** : il s'agit généralement de calculer un écoulement stationnaire sur la grille fine<sup>22</sup> pour des conditions aux limites conformes au problème étudié. Ce calcul fournit un champ de potentiel hydraulique à partir duquel peuvent être déduites d'autres variables (par ex. flux moyens à travers l'élément, vitesse moyenne dans l'élément). Pour une variable de flux, le critère d'adaptation du maillage sera une minimisation du flux cumulé dans chaque élément (vers des éléments plus petits où les flux sont plus grands). Par rapport aux deux scénarios à modéliser, l'inconvénient de cette approche est que la source de radionucléides ne constitue pas une condition aux limites pour les écoulements. La solution stationnaire est fondée sur des conditions aux limites régionales et éventuellement des puits de pompage (points cibles) et peut donc faire apparaître des voies d'écoulement sans influence sur la migration de radionucléides mais qui seraient à mailler plus finement.
- **Variable de transport** : il s'agit ici de calculer une solution de transport stationnaire (concentrations) à partir d'un champ de vitesse stationnaire. Contrairement au cas précédent, la source de radionucléides intervient dans la simulation. Le critère d'adaptation peut alors être fondé sur la minimisation du flux massique de radionucléides cumulé dans chaque élément (vers des éléments plus petits où les flux massiques sont plus grands).
- **Trajectoires de radionucléides** : il s'agit de calculer une multitude de fois la trajectoire aléatoire de radionucléides depuis la source pour un champ de vitesse stationnaire. Un compteur affecté à chaque bloc de la grille fine comptabilise le nombre de passages. Le maillage est alors adapté à la variable *compteur* en minimisant le nombre cumulé de passages dans chaque élément (vers des éléments plus petits où les passages de radionucléides sont plus nombreux).

### *Calcul des propriétés moyennes des éléments du maillage*

En pratique, même si le maillage est construit *intelligemment* avec une approche adaptative, la complexité de la répartition spatiale des propriétés, par rapport la résolution du maillage grossier, ne permet pas de rendre tous les éléments du maillage homogènes. Au mieux, ceux, situés dans la ou les régions où le phénomène de transport est prépondérant, seront relativement homogènes. Pour les autres, beaucoup seront hétérogènes et devront être affectés de propriétés moyennes.

La complexité du calcul des propriétés moyennes des éléments du maillage grossier varie d'une propriété à l'autre. Pour des variables additives telles que la porosité, une moyenne arithmétique suffit. Pour des paramètres de transport tels que la dispersivité, il existe des méthodes

---

<sup>22</sup> Malgré le grand nombre de blocs, une telle simulation est réalisable. Le champ de potentiel hydraulique est alors calculé par une méthode itérative moins coûteuse en espace mémoire (par ex. gradient conjugué). Il est à noter que la précision du calcul importe peu ici. Il s'agit simplement d'évaluer une solution d'écoulement qui montre les voies préférentielles d'écoulement et sert à l'adaptation du maillage.

pour évaluer la dispersion hydrodynamique à l'intérieur d'un élément en fonction des hétérogénéités internes et de la taille de l'élément.

Concernant la perméabilité, le choix d'une méthode de calcul de propriétés moyennes des éléments dépend du type de maillage et du simulateur d'écoulement et de transport utilisés. Pour un type de maillage donné, il peut exister plusieurs méthodes de calcul de perméabilités moyennes. Ces méthodes se distinguent par leurs performances en vitesse de calcul et par leur capacité à satisfaire à des conditions de flux moyens équivalents pour des écoulements stationnaires à l'échelle des éléments grossiers. S'il est prévu de traiter ainsi un nombre important de réalisations du champ de perméabilité de l'aquifère, la préférence sera donnée à une méthode rapide quitte à perdre un peu dans la qualité des perméabilités moyennes par rapport à des critères de flux. Pour une adaptation convenable du maillage, quelle que soit la propriété considérée (perméabilité, porosité, etc.), les éléments les plus hétérogènes du maillage seront aussi les plus éloignés des phénomènes de migration de radionucléides sur lesquels ils auront très peu d'effet.

#### **1.4.6 Vers la modélisation des écoulements et du transport de radionucléides**

Toute la démarche géostatistique décrite dans les paragraphes précédents a pour but de produire un modèle géologique réaliste de l'aquifère. Ce modèle n'est cependant pas une fin en soi. Il doit servir à alimenter un simulateur d'écoulement et de transport destiné à prédire la migration de radionucléides. Le modèle géologique est constitué en partie de champs de propriétés (perméabilité, porosité, faciès, etc.) qui doivent satisfaire à un certain nombre de critères :

- honorer les données aux points de mesure,
- reproduire certaines caractéristiques statistiques (distribution globale, continuité spatiale, corrélation avec d'autres variables ou avec des informations secondaires de type géophysique, modèles de dérive),
- redonner, lorsque cela s'applique et pour des conditions aux limites convenablement choisies, des réponses en écoulement identiques à celles observées (mesures piézométriques, résultats en débit et pression d'essais de puits).

Compte tenu des nombreuses incertitudes sur la répartition spatiale des faciès et des propriétés du milieu, le modèle géologique obtenu n'est pas unique. Ces incertitudes, combinées au besoin de satisfaire aux critères précédents, justifient une approche stochastique qui se traduit par un recours à de la simulation stochastique pour générer les champs de propriétés. Les champs ainsi obtenus sont des représentations réalistes de l'état de la nature. A ce titre, ils conviennent pour modéliser des phénomènes physiques d'écoulement et de transport<sup>23</sup>.

---

<sup>23</sup> Comme cela est largement décrit et illustré dans le Chapitre 2, l'utilisation de méthodes d'estimation (interpolation par krigeage ou par toute autre technique), pour calculer la répartition spatiale de propriétés d'un aquifère, conduisent à des champs très lissés et donc très peu contrastés. Ceci peut se traduire en particulier par une sous-estimation des voies d'écoulement les plus perméables ou par une surestimation de la perméabilité de faciès de faible conductivité hydraulique qui jouent un rôle de barrières aux écoulements. L'utilisation de

Tout le travail consiste alors à traduire les incertitudes d'ordre géologique en des incertitudes sur la migration des radionucléides, c'est-à-dire en des incertitudes sur les trajectoires des radionucléides et sur les temps de parcours depuis des points sources jusqu'à des points cibles.

La quantification de ce type d'incertitudes doit s'intégrer dans une approche stochastique globale qui comprend le traitement géostatistique et la simulation de l'écoulement et du transport. Cette approche peut être résumée comme suit.

1. Simulation stochastique, à l'échelle des données, des champs de propriétés et si nécessaire de faciès (étapes décrites aux paragraphes §1.4.3 et §1.4.4). Le nombre de réalisations<sup>24</sup> à produire doit être suffisamment grand pour décrire entièrement la variabilité possible des champs de propriétés. Il est courant de générer 100 à plusieurs centaines de réalisations.
2. Pour chaque réalisation, changement d'échelle pour passer à des maillages plus grossiers mieux adaptés aux simulateurs d'écoulement et de transport (étape décrite au paragraphe §1.4.5).
3. Pour chaque réalisation, simulation de l'écoulement et du transport sur le maillage grossier issu du changement d'échelle.

Un traitement statistique des résultats de transport obtenus sur l'ensemble des réalisations fournit alors les distributions de probabilité (histogrammes) sur les temps de parcours entre des points sources et cibles, et sur les proportions de radionucléides qui atteignent différents points cibles. De ces distributions de probabilité peuvent être déduits des intervalles de confiance et des estimateurs "optimaux" (par ex. la moyenne ou la médiane).

Il est à noter que si le problème d'écoulement et de transport à simuler numériquement est tel que les temps de calcul restent longs malgré le changement d'échelle, une étape intermédiaire, entre les étapes 1 et 2 ou 2 et 3, consiste à classer les réalisations selon des critères d'écoulement<sup>25</sup> ou de connectivité<sup>26</sup> et de choisir parmi les réalisations ainsi classées un nombre réduit d'entre elles pour décrire les variations d'écoulement et de transport possibles. La simulation complète de l'écoulement et du transport se fait alors sur les quelques réalisations sélectionnées.

---

champs estimés pour simuler la migration de radionucléides tend alors à produire des temps de parcours, entre points sources et cibles, supérieurs à la réalité (effet de la sous-estimation des perméabilités fortes) et des trajectoires moins déviées (effet du lissage qui filtre les perméabilités faibles et fortes). Ces champs ne permettent pas non plus de quantifier l'incertitude sur les résultats d'écoulement et de transport.

<sup>24</sup> Une réalisation consiste ici en un ensemble de champs de propriétés (faciès, perméabilité, porosité, etc.) simulés conjointement.

<sup>25</sup> A l'aide d'un simulateur d'écoulement simplifié et donc plus rapide.

<sup>26</sup> Il peut s'agir de déterminer, pour chaque champ de perméabilité simulé sur une grille, la proportion de nœuds de la grille connectés à un point source, et éventuellement à un point cible aussi, par un chemin continu le long duquel la perméabilité est toujours supérieure à un seuil de perméabilité choisi (le calcul peut être répété avec des seuils différents pour mieux décrire la connectivité).

**2. MODELISATION STATISTIQUE DU  
TRANSPORT DE RADIONUCLEIDES :  
INTEGRATION D'INFORMATIONS PAR  
UNE APPROCHE GEOSTATISTIQUE**



## 2.1 Introduction

### 2.1.1 Analyse de l'incertitude

L'évaluation des performances d'un stockage de déchets radioactifs exige de connaître le comportement des écoulements et du transport de contaminants dans le milieu géologique avoisinant. Ce problème requiert une bonne description du sous-sol en termes d'architecture géologique, de répartition spatiale des hétérogénéités, de zones d'écoulement préférentiel et de propriétés et d'interactions entre la roche et les fluides; la solution passe alors généralement par le développement d'un *modèle numérique* qui permet de prédire les réponses du système.

En hydrogéologie, l'implémentation d'un modèle numérique suppose de connaître, d'une part, la structure géologique du milieu (par ex. la localisation des limites entre les différentes unités géologiques et les conditions aux limites d'écoulement et de transport), et, d'autre part, les paramètres descriptifs des phénomènes d'écoulement et de transport de masse dans ce milieu (conductivité hydraulique, porosité, dispersivité, coefficient de retard, etc.). Il est reconnu que tous les systèmes hydrogéologiques sont naturellement complexes et fortement hétérogènes (voir par exemple Freeze, 1975, Gelhar, 1984 et 1993, ou Dagan, 1989), ce qui rend difficile, sinon impossible, une description adéquate, par des lois déterministes, de la structure géologique et des paramètres qui décrivent les écoulements et le transport. En ce sens, l'ingénieur qui construit un modèle doit admettre que toute prédiction faite à partir d'un modèle mathématique est essentiellement incertaine, et qu'une telle incertitude fait partie intégrante des sciences de la terre (voir Gómez-Hernández, 1997).

Dans le cas de l'évaluation de l'emplacement d'un stockage de déchets radioactifs, Eisenberg *et al.* (1987) suggèrent cinq types d'incertitude à prendre en compte.

1. Incertitude due aux erreurs aléatoires et systématiques commises sur les mesures réalisées.
2. Incertitude due aux hétérogénéités spatiales des paramètres géologiques.
3. Incertitude inhérente à la détermination d'un modèle conceptuel.
4. Incertitude relative à la modélisation des processus physico-chimiques.
5. Incertitude liée aux états futurs de la nature.

L'incertitude sur les mesures réalisées découle d'une ou plusieurs des causes suivantes :

- les limites de précision, de sensibilité et d'exactitude des instruments utilisés in situ et en laboratoire, et les limites de précision et de validité des modèles d'interprétation des grandeurs mesurées,
- les écarts à l'idéalité des conditions de mesure et d'évaluation des échantillons de sol ou d'eau,

- l'altération possible des échantillons au cours du processus de collecte, de transport et de mesure.

Un moyen de prendre en compte ce type d'incertitude se trouve dans les techniques standard de propagation des erreurs.

L'incertitude relative aux hétérogénéités spatiales est le résultat des variations considérables des propriétés hydrogéologiques du sous-sol d'un endroit à un autre du terrain. La conductivité hydraulique peut par exemple varier de plusieurs ordres de grandeur. La caractérisation appropriée d'une telle hétérogénéité nécessite une grande quantité d'échantillons qui, pour des raisons économiques et pratiques, ne peuvent être réalisés. Ce manque de données introduit une incertitude qui doit être quantifiée. Un des moyens de palier à ce manque est de considérer d'autres types d'informations telles que des mesures géophysiques. Ces informations, dites secondaires, ne peuvent être incorporées sans un processus adéquat de calibrage qui, à son tour, introduira une autre composante d'incertitude inhérente, par exemple, aux relations qui peuvent exister entre les paramètres enregistrés sur un sondage géophysique et les paramètres hydrogéologiques recherchés. Le traitement de ce type d'incertitude est l'objet premier de cette partie du rapport et sera très largement discuté dans les chapitres suivants.

L'incertitude sur le modèle conceptuel provient du besoin de l'ingénieur de simplifier le problème afin de le rendre mathématiquement traitable et compréhensible. Tout modèle conceptuel est une version simplifiée de la réalité fondée sur une série d'hypothèses explicites et implicites sur la nature des processus et des systèmes modélisés, hypothèses qui introduisent une erreur dans les prédictions obtenues. Un moyen de quantifier l'incertitude relative au modèle conceptuel est de comparer des prédictions obtenues à partir de modèles conceptuels différents (par ex. en faisant varier l'étendue du domaine étudié ou les conditions aux limites appliquées). Les variations de prédiction ainsi observées peuvent alors être considérées comme une mesure de l'incertitude entre les différents modèles. Il est à noter que le nombre de modèles conceptuels à évaluer reste souvent limité. Si, après avoir fait varier certaines hypothèses (par ex. le choix de l'étendue du domaine étudié ou de conditions aux limites), il apparaît que l'incertitude sur le modèle conceptuel est grande, elle peut être réduite considérablement en incorporant, dans l'élaboration du modèle, des caractéristiques supplémentaires pour lever les hypothèses les plus pénalisantes.

L'incertitude sur les lois qui régissent les processus physico-chimiques liés au phénomène de transport rentre dans la catégorie précédente d'incertitude sur le modèle conceptuel. La possibilité d'inclure ces processus dans la modélisation est de toute première importance pour l'évaluation de la migration de radionucléides. Les processus d'écoulement d'eau souterraine, de transport de masse et de réactions chimiques avec l'eau, sont relativement bien connus quand ils se produisent sous certaines conditions bien contrôlées. Il existe en revanche une grande incertitude sur l'effet combiné (couplage) de ces processus sur le phénomène de transport de masse, en particulier lorsqu'ils sont soumis à une configuration géométrique compliquée comme celle d'un système de fractures. Une manière d'évaluer ce type d'incertitude est de comparer les différences entre les prédictions réalisées avec des modèles qui intègrent tout ou partie de ces processus, et celles déduites de modèles qui n'en tiennent pas compte.

Il existe enfin une source d'incertitude liée aux états futurs de la nature, lesquels sont vitaux lorsqu'il s'agit d'évaluer des projets d'ingénierie sur de très longues durées comme celles ren-

contrées pour le stockage de déchets radioactifs. Ce type d'incertitude découle de l'impossibilité de connaître convenablement les conditions dans lesquelles un stockage opérera dans un futur plus ou moins proche. Un moyen de quantifier cette incertitude consiste à traduire les résultats de l'évaluation par une fonction de distribution de probabilité qui inclut, d'une certaine manière, les incertitudes relatives aux divers états futurs possibles de la nature.

En résumé, on peut dire que le processus de caractérisation d'un milieu hydrogéologique quelconque vise à réduire les incertitudes provenant de différentes sources. L'analyse d'incertitude consiste à évaluer le comportement d'un système donné à partir de l'incertitude sur un ou plusieurs facteurs associés à ce système (Brandstetter et Buxton, 1987).

Dans cette partie, nous nous intéresserons uniquement à l'incertitude due à l'hétérogénéité spatiale de paramètres qui s'utilisent dans les modèles numériques d'écoulement et de transport de contaminants. Ceci suppose que le modèle conceptuel convienne au problème.

### 2.1.2 Méthodes statistiques

Pour modéliser l'incertitude sur les paramètres d'un modèle numérique comme ceux d'écoulement et de transport de masse, il est fait recours à l'utilisation de *méthodes statistiques*. Dans un schéma statistique, la variabilité spatiale des paramètres d'entrée, dont en particulier la conductivité hydraulique, est modélisée comme une fonction aléatoire spatiale (régionalisée) caractérisée par une fonction de répartition multivariée (cf. Matheron, 1973, Freeze, 1975, Journel et Huijbregts, 1978, et Dagan, 1986). Les équations aux dérivées partielles qui gouvernent l'écoulement et le transport deviennent des équations aux dérivées partielles statistiques et les variables dépendantes obtenues par résolution de ces équations, hauteur piézométrique, vitesse d'écoulement, concentration en masse ou temps de parcours, deviennent également des fonctions aléatoires affectées de leurs propres distributions de probabilité. L'incertitude sur les paramètres d'entrée se propage aux variables de sortie ou aux prédictions à travers les équations aux dérivées partielles (fonctions de transfert) sous-jacentes aux modèles implémentés.

Les méthodes statistiques fournissent deux formes de caractérisation des milieux hydrogéologiques par estimation ou par simulation. Les algorithmes d'estimation (ou interpolation) produisent une solution (image) unique qui consiste en un champ de valeurs interpolées de la variable considérée. En un certain nombre de points (ou nœuds) de ce champ, il est estimé une valeur conformément à un certain critère d'optimisation. Par exemple, le krigeage cherche à minimiser l'erreur quadratique moyenne commise dans chaque estimation. Cette erreur minimale est considérée comme une mesure de l'incertitude sur l'estimation produite, et la carte construite à partir de ces erreurs est indicative de l'incertitude relative sur le champ estimé. Les champs obtenus par n'importe quel algorithme d'estimation ont la particularité d'être lissés (variations douces des valeurs estimées entre les points de données) et ne reflètent pas la variabilité que le paramètre estimé peut avoir en réalité; chaque valeur estimée est alors cohérente avec les données mais pas avec les valeurs estimées aux autres nœuds. Ceci fait que les techniques d'interpolation ne sont pas les mieux adaptées pour résoudre certains problèmes réels. Tel est le cas de la migration de contaminant dans le sous-sol.

Les techniques de simulation (ou représentations stochastiques) cherchent à générer des réalisations multiples de l'attribut modélisé. Toutes les réalisations ont alors la même chance d'exister dans le sens où elles reproduisent toutes les données et les formes de corrélations spatiales

observées. L'information dite *dure* (ou précise) y est reproduite exactement et l'information dite *douce* (incertaine ou incomplète) avec un degré de tolérance qui est déterminé par la précision de l'information. En générant plusieurs réalisations du même processus, les différences entre chaque réalisation fournissent une mesure quantitative et visuelle de l'incertitude sur le paramètre modélisé. Pour autant que toutes les réalisations reproduisent les données et le degré de variabilité imposé, n'importe laquelle d'entre elles peut correspondre à la réalité.

La simulation ou représentation stochastique, dans le cadre d'un processus de Monte Carlo, est la méthode numérique la plus communément employée en hydrogéologie pour réaliser une analyse probabiliste de l'écoulement et du transport de masse. L'utilisation d'une méthode par Monte Carlo signifie que la répartition spatiale incertaine d'un attribut donné; par exemple la conductivité hydraulique, est modélisée au moyen d'un grand nombre de réalisations équiprobables. Chacune d'entre elles possède des propriétés statistiques qui déterminent son degré de variabilité et de continuité spatiale. Les équations qui gouvernent l'écoulement d'eau en milieu poreux et le transport de masse sont résolues numériquement sur chacune de ces réalisations qui vont donner des réponses généralement différentes. Les fonctions de distribution de probabilité des variables de réponse, comme les hauteurs piézométriques, les vitesses d'écoulement ou les temps de parcours, sont construites à partir de l'analyse statistique des valeurs obtenues sur les différentes réalisations.

Ces fonctions de distribution constituent de vrais modèles d'incertitude (Gómez-Hernández, 1991, 1993a et b) des paramètres impliqués dans la description du comportement de l'écoulement d'eau et la migration de contaminants dans le sous-sol, et peuvent être utilisés dans un schéma d'analyse de décision ou d'optimisation pour évaluer la convenance de différentes conceptions d'ouvrages (Massmann et Freeze, 1987a, b et 1989; Wagner et Gorelick, 1987 et 1989; Freeze *et al.*, 1990 et 1992; Massmann *et al.*, 1991; Sperling *et al.*, 1992; James et Freeze, 1993; Rautman et Istok, 1996; Rosen et Gustafson, 1996; Istok et Rautman, 1996; Wen *et al.*, 1996; Freeze et McWhorther, 1997).

Appliqués à l'analyse de performance de stockages de déchets radioactifs, les modèles d'incertitude permettent de comparer directement des projets concurrents de la façon suivante. Les agences de contrôle établissent de manière quantitative un ensemble de conditions à satisfaire concernant, par exemple, le temps de parcours du flux souterrain de contaminants et le taux maximal de radionucléides que peut recevoir la biosphère. Il est fréquent, en particulier, que soit exigé que le temps de parcours n'excède pas un seuil donné avec une certaine marge de sécurité (probabilité). Si, à partir du modèle d'incertitude (fonction de distribution), il apparaît que le seuil de temps fixé ne peut être garanti avec la probabilité exigée, la performance du projet de stockage étudié peut être jugée insuffisante.

Dans le but d'obtenir des distributions de probabilité (histogrammes) aussi précises et exactes que possible sur certaines variables d'écoulement et de transport (flux, temps de parcours, etc.), il est important d'inclure toute l'information pertinente dans la construction du modèle de fonctions aléatoires des paramètres utiles à la description géologique de l'aquifère (champs de perméabilité, de porosité, etc.). Le *conditionnement* aux données disponibles est la technique la plus puissante pour la construction de distributions de probabilité plus précises. Par conditionnement il faut comprendre que les réalisations générées par simulation stochastique reproduisent les données dont on dispose.

### 2.1.3 Typologie de l'information

Dans la perspective de faire tourner un modèle numérique, la caractérisation d'un milieu géologique suppose d'affecter des valeurs à un ou plusieurs paramètres en des points où les mesures physiques sont manquantes. Supposons qu'il s'agisse de cartographier la porosité d'un milieu géologique donné. Il est possible initialement de compter sur une série de mesures directes de porosité qui proviennent d'essais effectués in situ ou de tests en laboratoire. Dans les deux cas, les mesures nécessitent de réaliser des perforations dans le terrain, ce qui implique, d'une part, des coûts d'investigation élevés et, d'autre part, l'altération du sous-sol. Les mesures directes de paramètres hydrogéologiques sont souvent insuffisantes et particulièrement dispersées. Du point de vue du modèle mathématique du problème de caractérisation, l'information directe (ou principale) est considérée comme une *information précise ou dure* (Journel, 1983), c'est-à-dire comme une information sur laquelle l'incertitude peut être jugée comme négligeable.

En complément de l'information directe de porosité, il est courant de disposer d'un autre type d'information comme, par exemple, une carte de vitesses ou d'amplitudes sismiques provenant de l'inversion d'une tomographie entre sondages. Cette information est dite indirecte (ou secondaire) car elle ne fournit pas une mesure de porosité mais celle d'un attribut, comme la vitesse ou l'amplitude sismique, que l'on espère pouvoir corrélérer à la porosité. Etant donné la nature indirecte de cette information, on l'appelle aussi *information imprécise ou douce* (Journel, 1983). Il s'agit alors d'une information dont les valeurs doivent être accompagnées d'une mesure qui spécifie son degré de confiance.

Ainsi, quand on parle d'information précise ou imprécise, il n'est pas fait allusion à la qualité de la mesure effectuée. La distinction concerne deux ensembles de données. Le premier contient des mesures directes sur l'attribut étudié. Le second porte sur un paramètre quelconque lié indirectement à cet attribut, et ce paramètre doit pouvoir être traité de manière appropriée pour pouvoir être incorporé dans le processus d'estimation de la variable primaire. L'erreur sur la détermination d'une mesure de vitesse sismique peut être de 1%, tandis que la perméabilité déduite d'un essai de pompage peut présenter une erreur de 15 à 20%. L'incertitude qui existe, cependant, autour de la relation entre perméabilité et vitesse sismique, au moment d'estimer la variable primaire en fonction de la secondaire, fait que la confiance assignée à la perméabilité déduite de la mesure de vitesse est moindre que celle octroyée à la mesure directe de perméabilité.

L'intérêt de considérer une information secondaire réside dans le fait que cette information est généralement disponible en un plus grand nombre de points que l'information hydrogéologique directe, et dans le fait qu'il n'est pas nécessaire de détruire le milieu pour l'obtenir. Une campagne sismique 3D, par exemple, peut fournir une information exhaustive, sur l'ensemble du domaine étudié, pour une ou plusieurs variables secondaires.

Outre son taux de fiabilité, une autre caractéristique de l'information secondaire est l'échelle à laquelle elle a été obtenue ou se rapporte. Les mesures directes par test, les plus chères et les plus rares, sont effectuées à une échelle très petite de l'ordre de quelques centimètres cubes. La géophysique de puits fournit des mesures dont la représentativité se limite aux régions immédiatement adjacentes à la paroi du trou. L'interprétation géologique des données de puits peut donner une information sur la continuité latérale des niveaux rencontrés le long du forage. La tomographie et les méthodes sismiques traditionnelles renseignent sur les régions entre puits.

Les essais de pompage peuvent donner des valeurs de perméabilité et d'emménagement pour des volumes très grands.

Dans l'exemple de la cartographie de la porosité introduit plus haut, les données sismiques sont représentatives d'une superficie qui peut être de plusieurs mètres carrés, alors que les mesures de puits correspondantes sont inférieures au mètre carré. Les enregistrements sismiques couvrent, avec une grande densité, les régions entre puits, alors que les enregistrements aux puits ne sont disponibles qu'à l'endroit des puits et ne sont représentatifs que des terrains situés au voisinage immédiat des trous. Etant donné que la structure spatiale de la vitesse sismique est assez similaire à celle de la porosité (conséquence de la bonne corrélation en général entre ces deux variables), l'incorporation de cette information dans le processus d'estimation produit de meilleurs résultats.

L'affectation de valeurs de propriétés, aux endroits où elles ne sont pas connues, doit par conséquent pouvoir intégrer des informations de qualité variable et représentatives d'échelles différentes. En ce sens, la *géostatistique* fournit une série de méthodologies et d'algorithmes qui permettent un traitement systématique et réel de toute l'information disponible dans le but de modéliser des systèmes géologiques. L'utilisation de ces outils nécessite de travailler dans le cadre d'un modèle probabiliste ou statistique. Ceci fait l'objet des paragraphes suivants.

#### **2.1.4 Interpolation spatiale**

Le problème qui se pose est celui de l'interpolation spatiale d'une variable à partir de données de différents types et donc de différentes fiabilités, données qui proviennent de diverses sources et échelles d'information. L'idée clé consiste à utiliser l'ensemble des données disponibles pour élaborer un modèle de corrélation spatiale entre les variables auxquelles on s'intéresse. Ce modèle est exporté vers les zones où la variable principale doit être déduite à partir des données, généralement dispersées, disponibles sur cette variable et de l'information, beaucoup plus abondante mais aussi moins précise, sur une ou plusieurs variables secondaires. L'efficacité de cette procédure dépend autant de l'information extraite des données que de la représentativité du modèle de corrélation spatiale à exporter.

Dans le processus d'intégration de l'information, le calibrage des divers types de données est une étape essentielle. Le calibrage signifie que les données dures (précises) sont prises comme référence tandis que les données douces (incertaines) doivent être calibrées par rapport aux dures. Pour réaliser ce calibrage, il est souhaitable de disposer d'un ensemble de données constitué à la fois de mesures de la variable principale et de valeurs de la ou des variables secondaires mesurées aux mêmes endroits que la variable principale ou à proximité. Tous les paramètres qui définissent le modèle de corrélation spatiale dont il a été fait référence sont extraits de ce groupe de données (coefficients de corrélation, covariances, variogrammes, etc.). Dans le cas le plus simple, un modèle de régression linéaire suffit pour quantifier la relation entre deux variables ou plus. Les paramètres statistiques correspondants peuvent alors être extraits à partir de quelques données.

Par exemple, le diagramme de dispersion<sup>27</sup> de la Figure 2-1 représente graphiquement les valeurs correspondants à 960 paires de calibrage. A partir de ce diagramme, on peut voir que la relation entre les deux variables n'est pas linéaire mais aussi que la dispersion verticale des valeurs, dispersion relative à la variable principale, est différente selon la valeur de la variable secondaire considérée. Les valeurs faibles des deux variables sont mieux corrélées que les valeurs élevées. Un modèle d'estimation de  $Z$  à partir de  $V$  doit être capable d'incorporer la dispersion ainsi identifiée, c'est-à-dire non seulement une mesure de l'emplacement moyen du nuage de points mais aussi une mesure de la dispersion ou de la variance du nuage.

Quand les distributions de  $Z$  conditionnelles aux valeurs de  $V$  sont très différentes pour des valeurs distinctes de  $V$ , il peut être intéressant d'exporter non seulement la moyenne et la variance de chaque distribution, mais aussi la forme de celle-ci. On s'intéresse alors à modéliser la distribution complète de la variable à estimer conditionnée par l'information secondaire. Cette distribution s'exprime alors comme :

$$Prob\{Z \leq z_k | V \in (v_1, v_2)\}$$

où  $Z$  représente la variable principale,  $z_k$  une valeur de celle-ci,  $V$  la variable secondaire, et  $(v_1, v_2)$  un intervalle de variation qui détermine une classe de  $V$ . Cette expression représente l'information dont on dispose sur la variable principale quand la variable secondaire appartient à l'intervalle donné.

Les probabilités associées à l'expression précédente peuvent également se déduire à partir d'un ensemble de données de calibrage. Dans un schéma d'analyse bayésien (voir Box et Tiao, 1973), ces valeurs définissent une fonction de probabilité préalable (a priori) calibrée à partir d'un minimum d'informations (les données de calibrage). L'objectif de l'interpolation spatiale est d'améliorer, en chaque point d'estimation, cette fonction initiale en incorporant les mesures, tant dures que douces, disponibles sur les variables considérées à l'endroit du point d'estimation et dans son voisinage. Ceci donne comme résultat, en chaque point d'estimation, une fonction de probabilité a posteriori qui dépend de toute l'information disponible. Elle s'écrit :

$$Prob\{Z(u) \leq z_k | \text{toute l'information disponible}\}.$$

Une telle fonction constitue un modèle d'incertitude de la variable principale. Si on dispose de modèles d'incertitude en chaque point à estimer, il est possible d'obtenir des cartes de risque qui représentent divers niveaux d'intérêt comme par exemple, pour un estimateur supposé optimal de la variable, la probabilité que la valeur de la variable soit au-dessous ou au-dessus d'un certain seuil, ou encore l'erreur commise dans les estimations (voir Gómez-Hernández 1991, 1993 a et b).

Les méthodes d'interpolation (ou estimation) et de représentation stochastique sont décrites plus en détail dans la suite. Cette présentation est illustrée par un cas pratique qui permet d'évaluer le comportement des méthodes d'estimation. Pour les méthodes de représentation stochastique, l'exercice est plus compliqué et rivalise avec le processus d'analyse de l'incer-

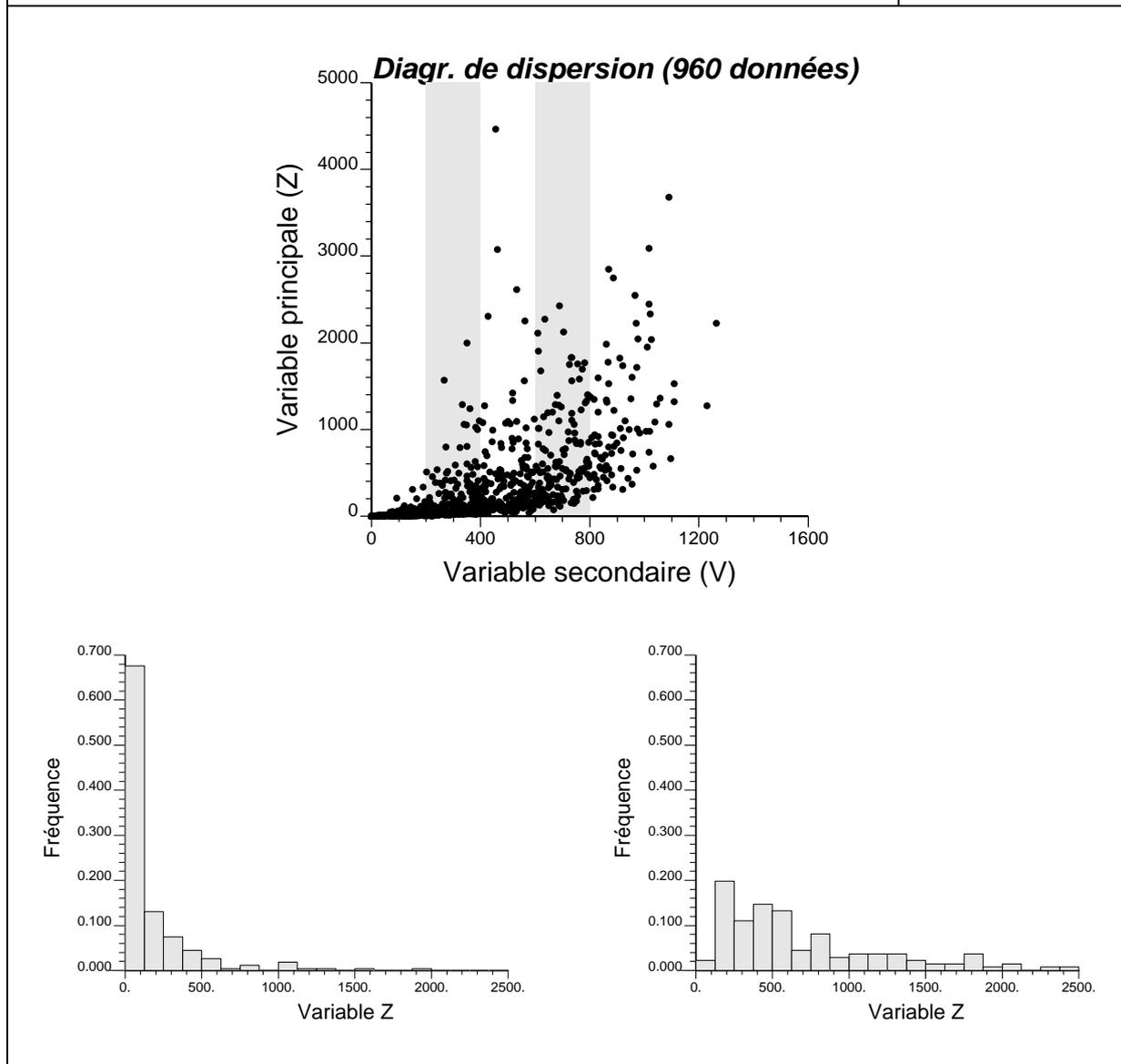
---

<sup>27</sup>Un diagramme de dispersion est un graphique x-y sur lequel la coordonnée x correspond aux valeurs d'une des variables et la coordonnée y aux valeurs de l'autre variable. Il constitue la forme la plus courante de représentation graphique de données bivariées.

ti+tude sur des variables telles que les temps de parcours et la position d'arrivée de particules représentant les radionucléides dans un scénario déterminé.

Exemple de diagramme de dispersion sur lequel la relation (régression) n'est pas linéaire. Les distributions conditionnelles de Z changent de forme, d'aplatissement et de variance selon les classes de V. Les histogrammes correspondant aux points à l'intérieur des bandes grisées pour V entre 200 et 400, à gauche, et pour V entre 600 et 800 à droite.

Figure 2-1



## 2.2 Algorithme d'interpolation

### 2.2.1 Régression traditionnelle

La prédiction la plus simple d'une variable en fonction d'une autre est la régression linéaire. Dans ce modèle, on suppose que la dépendance entre variables peut être décrite de façon adéquate par une ligne rectiligne (droite), c'est-à-dire par une équation du type :

$$z_2 = az_1 + b$$

où  $z_1$  et  $z_2$  sont les valeurs des variables et où  $a$  et  $b$  sont les coefficients de la droite à déterminer. Une technique pour estimer les paramètres  $a$  et  $b$  du modèle est celle par moindres carrés. Elle donne comme résultat :

$$a = \rho \frac{\sigma_{z_2}}{\sigma_{z_1}}$$

$$b = m_{z_2} - am_{z_1}$$

où  $\rho$  est le coefficient de corrélation linéaire entre les variables  $z_1$  et  $z_2$ ;  $\sigma_{z_1}$  est l'écart-type de la variable  $z_1$ ;  $\sigma_{z_2}$  est celui de la variable  $z_2$ ;  $m_{z_2}$  et  $m_{z_1}$  sont les valeurs moyennes correspondantes. Cette méthode est extrêmement simple et présente certaines limitations qui sont commentées dans la suite. En chaque point où l'on souhaite estimer la variable principale, il faut compter sur une mesure de la variable secondaire. Les valeurs estimées sont fonction de cette valeur seule, alors qu'il est plus raisonnable de penser qu'une variable corrélée dans l'espace dépend également des valeurs mesurées dans son voisinage. Avec cette technique, il n'existe aucun moyen ni de quantifier l'incertitude sur les estimations faites, ni de prendre en compte la corrélation spatiale des données. La droite ajustée ne reproduit pas, en général, les données disponibles. En dépit de ses inconvénients, cette technique a été et reste encore très largement utilisée pour l'intégration et la corrélation de paramètres hydrogéologiques et géophysiques (voir l'état des lieux réalisé par Cassiraga et Gómez-Hernández, 1994). Ceci explique pourquoi il a été décidé de présenter et commenter brièvement ici cette technique.

### 2.2.2 Le modèle (paradigme) de krigeage

Le krigeage (voir par exemple Journel et Huijbregts, 1978, Isaaks et Srivastava, 1989, Govaerts, 1997) est le nom générique donné à un ensemble de techniques statistiques d'interpolation qui appartiennent à la famille des algorithmes (méthodes) de régression par moindres carrés généralisés. L'idée de base du krigeage est d'estimer la valeur inconnue d'un attribut  $z$ , évalué en un point de coordonnées  $\mathbf{u}$ , comme étant une combinaison linéaire de  $n$  valeurs connues de  $z$  en des points de coordonnées  $\mathbf{u}_\alpha$  avec  $\alpha=1, \dots, n$ . Ceci s'écrit :

$$z^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_\alpha z(\mathbf{u}_\alpha),$$

où le symbole \* indique qu'il s'agit d'un estimateur et où les coefficients  $\lambda_{\alpha}$  avec  $\alpha=1, \dots, n$ , sont les coefficients de pondération inconnus a priori<sup>1</sup>. Pour connaître les valeurs des coefficients  $\lambda_{\alpha}$  on a recours à un modèle probabiliste qui consiste à admettre que les valeurs que peut prendre la variable  $z$  à l'intérieur du domaine d'étude sont déterminées par une fonction aléatoire que l'on notera  $Z(\mathbf{u})$ <sup>2</sup>. L'expression de l'estimateur s'écrit alors :

$$E\ 2-2 \quad Z^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha})$$

Si on définit l'erreur  $R$  commise dans l'estimation de l'attribut  $z$  au point de coordonnées  $\mathbf{u}$  comme :

$$E\ 2-3 \quad R(\mathbf{u}) = Z^*(\mathbf{u}) - Z(\mathbf{u}),$$

où  $Z^*(\mathbf{u})$  et  $Z(\mathbf{u})$  sont les variables aléatoires qui représentent respectivement la valeur estimée et la valeur vraie (inconnue), cette erreur  $R(\mathbf{u})$  est également une variable aléatoire. A partir de l'expression E 2-3, on peut définir l'erreur moyenne  $m_R(\mathbf{u})$  commise dans l'estimation comme :

$$E\ 2-4 \quad m_R(\mathbf{u}) = E\{Z^*(\mathbf{u}) - Z(\mathbf{u})\}$$

et la variance de l'erreur  $\sigma_R^2(\mathbf{u})$ , appelée aussi variance d'estimation, comme :

$$E\ 2-5 \quad \sigma_R^2(\mathbf{u}) = \text{Var}\{Z^*(\mathbf{u}) - Z(\mathbf{u})\}.$$

Les objectifs du krigeage sont d'obtenir les valeurs des coefficients  $\lambda_{\alpha}$  de manière à ce que l'estimateur trouvé soit non biaisé (erreur moyenne nulle) et que l'erreur d'estimation commise soit minimale. Mathématiquement, cela peut se traduire par le problème suivant :

$$E\ 2-6 \quad \text{minimiser } \sigma_R^2(\mathbf{u}) = \text{Var}\{Z^*(\mathbf{u}) - Z(\mathbf{u})\}$$

soumis à la condition que :

$$E\ 2-7 \quad m_R(\mathbf{u}) = E\{Z^*(\mathbf{u}) - Z(\mathbf{u})\} = 0.$$

---

<sup>1</sup>A proprement parler, le nombre de données et les coefficients de pondération sont fonction de l'emplacement du point à estimer. La notation correcte serait  $n(\mathbf{u})$  et  $\lambda_{\alpha}(\mathbf{u})$  bien que pour simplifier ces paramètres soient remplacés respectivement par  $n$  et  $\lambda_{\alpha}$ .

<sup>2</sup> La notion de fonction aléatoire signifie ici un ensemble de variables aléatoires, chaque variable étant affectée à un point de l'espace, dont les relations de dépendance entre elles sont spécifiées par certains mécanismes probabilistes (covariances).

Le paragraphe suivant analyse en détail la résolution de ce problème dans le cas du krigeage ordinaire.

### 2.2.2.1 Krigeage ordinaire

#### Condition de non-biais de l'estimateur

En substituant l'équation E 2-2 dans l'équation E 2-3, l'expression de l'erreur d'estimation commise prend la forme suivante :

$$R(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha}) - Z(\mathbf{u}).$$

La condition de non-biais de l'estimateur (équation E 2-7) s'écrit alors comme

$$E\{R(\mathbf{u})\} = E\left\{\sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha}) - Z(\mathbf{u})\right\} = \sum_{\alpha=1}^n \lambda_{\alpha} E\{Z(\mathbf{u}_{\alpha})\} - E\{Z(\mathbf{u})\} = 0.$$

En supposant que la fonction aléatoire est stationnaire, c'est-à-dire que sa moyenne est la même partout, soit encore que

$$E\{Z(\mathbf{u})\} = m(\mathbf{u}) = \text{constante}, \forall \mathbf{u},$$

l'espérance mathématique (la moyenne) de l'erreur peut s'exprimer comme :

$$E\{R(\mathbf{u})\} = \sum_{\alpha=1}^n \lambda_{\alpha} E\{Z\} - E\{Z\} = 0$$

d'où on déduit que, pour que l'erreur moyenne soit nulle,  $E\{R(\mathbf{u})\} = 0$ , les coefficients de pondération  $\lambda_{\alpha}$  doivent satisfaire à la condition de non-biais :

$$\text{E 2-8} \quad \sum_{\alpha=1}^n \lambda_{\alpha} = 1,$$

c'est-à-dire qu'ils doivent sommer à 1 en chaque point d'estimation.

#### Condition de variance de l'erreur minimale

Cette condition suppose d'établir pour  $\sigma_R^2$  une expression qui ensuite doit être minimisée au moyen d'une analyse différentielle. L'équation E 2-3 peut être considérée comme une combinaison linéaire des deux variables aléatoires  $Z^*(\mathbf{u})$  et  $Z(\mathbf{u})$  dont les coefficients de pondération sont égaux à 1 et -1. La variance d'une combinaison linéaire de variables aléatoires (voir Annexe A) est donnée par l'expression suivante :

$$E 2-9 \quad \text{Var}\left\{\sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha})\right\} = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C\{Z(\mathbf{u}_{\alpha})Z(\mathbf{u}_{\beta})\}$$

où  $C\{Z(\mathbf{u}_{\alpha})Z(\mathbf{u}_{\beta})\}$  est la covariance entre les variables aléatoires affectées aux points de coordonnées  $\mathbf{u}_{\alpha}$  et  $\mathbf{u}_{\beta}$ . Dans l'hypothèse d'une fonction aléatoire  $Z(\mathbf{u})$  stationnaire, la covariance entre les deux variables aléatoires  $Z(\mathbf{u}_{\alpha})$  et  $Z(\mathbf{u}_{\beta})$  ne dépend que de leur distance relative et peut s'écrire aussi  $C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta})$ . Ces deux notations seront utilisées indifféremment dans la suite.

Appliquée à l'erreur d'estimation, l'expression précédente de la variance donne la relation suivante :

$$\begin{aligned} \text{Var}\{R(\mathbf{u})\} &= C\{Z^*(\mathbf{u})Z^*(\mathbf{u})\} - C\{Z^*(\mathbf{u})Z(\mathbf{u})\} - C\{Z(\mathbf{u})Z^*(\mathbf{u})\} + C\{Z(\mathbf{u})Z(\mathbf{u})\} \\ &= C\{Z^*(\mathbf{u})Z^*(\mathbf{u})\} - 2C\{Z^*(\mathbf{u})Z(\mathbf{u})\} + C\{Z(\mathbf{u})Z(\mathbf{u})\}. \end{aligned}$$

On peut analyser les trois termes de la partie de droite de l'équation précédente. Le premier terme est la covariance de  $Z^*(\mathbf{u})$  avec elle-même, c'est-à-dire la variance  $\text{Var}\{Z^*(\mathbf{u})\}$ . Comme, par définition,

$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha})$$

l'équation E 2-9 s'applique et donne

$$\text{Var}\{Z^*(\mathbf{u})\} = \text{Var}\left\{\sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha})\right\} = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C\{Z(\mathbf{u}_{\alpha})Z(\mathbf{u}_{\beta})\}$$

$$E 2-10 \quad = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}).$$

Le second terme peut s'écrire comme :

$$\begin{aligned} C\{Z^*(\mathbf{u})Z(\mathbf{u})\} &= C\left\{\left(\sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha})\right)Z(\mathbf{u})\right\} \\ &= E\left\{\sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha})Z(\mathbf{u})\right\} - E\left\{\sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha})\right\}E\{Z(\mathbf{u})\} \\ &= \sum_{\alpha=1}^n \lambda_{\alpha} E\{Z(\mathbf{u}_{\alpha})Z(\mathbf{u})\} - \sum_{\alpha=1}^n \lambda_{\alpha} E\{Z(\mathbf{u}_{\alpha})\}E\{Z(\mathbf{u})\} \\ &= \sum_{\alpha=1}^n \lambda_{\alpha} C\{Z(\mathbf{u}_{\alpha})Z(\mathbf{u})\} \end{aligned}$$

$$E 2-11 \quad = \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{u}_{\alpha} - \mathbf{u}).$$

Le troisième terme,  $C\{Z(\mathbf{u})Z(\mathbf{u})\}$ , est la covariance de la variable aléatoire  $Z(\mathbf{u})$  avec elle-même, c'est-à-dire la variance de  $Z(\mathbf{u})$  qui peut s'exprimer comme :

$$E 2-12 \quad C\{Z(\mathbf{u})Z(\mathbf{u})\} = \sigma^2.$$

Si on remplace par leur nouvelle expression les trois termes de droite de l'équation de la variance de l'erreur  $\sigma_R^2$ , on obtient :

$$E 2-13 \quad Var\{R(\mathbf{u})\} = \sigma_R^2 = \sigma^2 + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{u}_{\alpha} - \mathbf{u}).$$

L'étape suivante est de trouver les coefficients  $\lambda_{\alpha}$  qui minimisent cette variance et satisfont à la condition (contrainte) de non-biais donnée par l'équation E 2-8. Pour cela, on utilise la technique des multiplicateurs de Lagrange qui permet de transformer un problème de minimisation sous contrainte en un problème sans contrainte. L'expression à minimiser devient alors :

$$E 2-14 \quad \sigma_R^2 = \sigma^2 + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{u}_{\alpha} - \mathbf{u}) + 2\mu \underbrace{\sum_{\alpha=1}^n \lambda_{\alpha} - 1}_0,$$

où  $\mu$  est le paramètre de Lagrange associé à la contrainte de non-biais. Il est à noter que le dernier terme de cette équation n'altère pas l'expression de la variance de l'erreur car il vaut 0. Cette expression de  $\sigma_R^2$  atteint son minimum pour les valeurs de  $\lambda_{\alpha}$  et de  $\mu$  qui annulent les  $n+1$  dérivées partielles de  $\sigma_R^2$  par rapport aux  $n$  coefficients  $\lambda_{\alpha}$  et au paramètre de Lagrange  $\mu$ .

Il peut être vérifié facilement que la dérivée partielle par rapport  $\mu$  donne, lorsqu'on l'égalise à 0, la condition de non-biais (équation E 2-8).

Pour ce qui concerne les dérivées partielles par rapport aux coefficients  $\lambda_{\alpha}$ , elles sont toutes identiques. Seule la dérivée partielle par rapport à  $\lambda_1$  est donc développée en détail dans la suite. Le premier terme de l'expression E 2-14 est constant et ne dépend donc pas de  $\lambda_1$ ; sa dérivée par rapport à  $\lambda_1$  donne logiquement 0. Dans le deuxième terme, si on développe la somme, la dérivée partielle par rapport à  $\lambda_1$  ne conserve que les éléments de la somme où  $\lambda_1$  intervient. On obtient :

$$\begin{aligned} \frac{\partial \left( \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) \right)}{\partial \lambda_1} &= \frac{\partial \left( \lambda_1^2 C(\mathbf{u}_1 - \mathbf{u}_1) + 2\lambda_1 \sum_{\beta=2}^n \lambda_{\beta} C(\mathbf{u}_1 - \mathbf{u}_{\beta}) \right)}{\partial \lambda_1} \\ &= 2\lambda_1 C(\mathbf{u}_1 - \mathbf{u}_1) + 2 \sum_{\beta=2}^n \lambda_{\beta} C(\mathbf{u}_1 - \mathbf{u}_{\beta}) \end{aligned}$$

$$E\ 2-15 \quad = 2 \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{u}_1 - \mathbf{u}_{\beta}).$$

Il en est de même avec le troisième terme dans lequel un seul élément de la somme est fonction de  $\lambda_1$ . La dérivée partielle par rapport à  $\lambda_1$  donne :

$$E\ 2-16 \quad \frac{\partial \left( \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{u}_{\alpha} - \mathbf{u}) \right)}{\partial \lambda_1} = \frac{\partial (\lambda_1 C(\mathbf{u}_1 - \mathbf{u}))}{\partial \lambda_1} = C(\mathbf{u}_1 - \mathbf{u}).$$

Quant au dernier terme, la dérivée partielle par rapport à  $\lambda_1$  vaut  $\mu$  :

$$E\ 2-17 \quad \frac{\partial \left( \mu \sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right)}{\partial \lambda_1} = \frac{\partial (\mu \lambda_1)}{\partial \lambda_1} = 1.$$

On en déduit l'expression finale de la dérivée partielle de  $\sigma_R^2$  par rapport à  $\lambda_1$  :

$$\frac{\partial \sigma_R^2}{\partial \lambda_1} = 2 \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{u}_1 - \mathbf{u}_{\beta}) - 2C(\mathbf{u}_1 - \mathbf{u}) + 2\mu.$$

En posant cette dérivée partielle égale à 0 et en réordonnant les termes, on obtient l'équation suivante tirée de la dérivée partielle par rapport à  $\lambda_1$  :

$$\sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{u}_1 - \mathbf{u}_{\beta}) + \mu = C(\mathbf{u}_1 - \mathbf{u}).$$

Une expression identique à la précédente peut être déduite de la dérivée partielle par rapport à chacun des coefficients  $\lambda_{\alpha}$ . On trouve ainsi :

$$\begin{aligned} \frac{\partial \sigma_R^2}{\partial \lambda_1} = 0 &\Rightarrow \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{u}_1 - \mathbf{u}_{\beta}) + \mu = C(\mathbf{u}_1 - \mathbf{u}) \\ &\vdots \\ \frac{\partial \sigma_R^2}{\partial \lambda_{\alpha}} = 0 &\Rightarrow \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) + \mu = C(\mathbf{u}_{\alpha} - \mathbf{u}) \\ &\vdots \\ \frac{\partial \sigma_R^2}{\partial \lambda_n} = 0 &\Rightarrow \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{u}_n - \mathbf{u}_{\beta}) + \mu = C(\mathbf{u}_n - \mathbf{u}). \end{aligned}$$

Les  $n$  équations précédentes, ajoutées à celle obtenue en dérivant par rapport au multiplicateur de Lagrange  $\mu$  pour introduire la condition de non-biais, constituent un système de  $n+1$  équations linéaires appelé *système de krigeage ordinaire*. Ce système s'écrit, de manière plus compacte, comme suit :

$$E\ 2-18 \quad \begin{cases} \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) + \mu = C(\mathbf{u}_{\alpha} - \mathbf{u}), \alpha = 1, \dots, n \\ \sum_{\beta=1}^n \lambda_{\beta} = 1 \end{cases} .$$

Sa forme matricielle est :

$$E\ 2-19 \quad \underbrace{\begin{pmatrix} C(\mathbf{u}_1 - \mathbf{u}_1) \dots C(\mathbf{u}_1 - \mathbf{u}_n) 1 \\ \vdots \vdots \\ C(\mathbf{u}_n - \mathbf{u}_1) \dots C(\mathbf{u}_n - \mathbf{u}_n) 1 \\ 1 \quad \dots \quad 1 \quad 0 \end{pmatrix}}_C \cdot \underbrace{\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix}}_{\lambda} = \underbrace{\begin{pmatrix} C(\mathbf{u}_1 - \mathbf{u}) \\ \vdots \\ C(\mathbf{u}_n - \mathbf{u}) \\ 1 \end{pmatrix}}_D$$

d'où on tire la solution du système sous la forme :

$$\lambda = C^{-1} \cdot D.$$

La matrice  $C$  détermine la corrélation entre les données qui interviennent dans l'estimation. Plus des données sont proches, plus grande est leur corrélation. Dans le processus de krigeage, cette corrélation permet de tenir compte de la redondance de l'information, c'est-à-dire de la possibilité que certaines régions de l'espace soient mieux échantillonnées que d'autres et que donc certaines données (isolées) soient plus significatives que d'autres (regroupées). Concrètement, si deux données sont situées à égale distance d'un point d'estimation mais que l'une d'entre elles est isolée alors que l'autre est proche d'une troisième donnée, la résolution du système de krigeage attribuera un poids plus grand à la donnée isolée, qui est la seule à fournir une information locale sur une partie de l'espace, qu'aux deux autres données qui renseignent sur la même région. **Cette prise en compte de la redondance de l'information, en plus de l'utilisation d'une distance statistique (covariance ou variogramme), est ce qui différencie les méthodes de krigeage des autres méthodes d'interpolation.**

De la même façon, le vecteur  $D$  exprime la corrélation entre les données et le point d'estimation. Il définit ainsi la continuité spatiale entre les positions des données et celle du point à estimer. Le système de krigeage ainsi obtenu requiert les valeurs des  $n^2+n$  covariances de la matrice  $C$  et du vecteur  $D$ . En pratique, ces covariances sont définies à partir d'un modèle de covariance convenablement choisi. Ce modèle consiste en une fonction dite de covariance, et notée  $C(\mathbf{h})$ , qui dépend uniquement du vecteur distance  $\mathbf{h}$  entre deux points pour décrire leur corrélation.

Après avoir calculé les valeurs des coefficients  $\lambda_{\alpha}$ , on peut déduire la valeur de l'estimateur (équation E 2-1) et celle de la variance d'estimation (équation E 2-13). Pour la variance, il existe cependant une façon plus simple de la calculer. Elle consiste à multiplier chacune des  $n$  premières équations du système E 2-18 par  $\lambda_{\alpha}$ . On obtient les  $n$  équations suivantes :

$$\lambda_{\alpha} \left[ \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) + \mu \right] = \lambda_{\alpha} C(\mathbf{u}_{\alpha} - \mathbf{u}), \alpha = 1, \dots, n,$$

dont la somme donne :

$$\sum_{\alpha=1}^n \lambda_{\alpha} \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) + \sum_{\alpha=1}^n \lambda_{\alpha} \mu = \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{u}_{\alpha} - \mathbf{u})$$

qui peut encore s'écrire :

$$\sum_{\alpha=1}^n \lambda_{\alpha} \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) = \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{u}_{\alpha} - \mathbf{u}) - \sum_{\alpha=1}^n \lambda_{\alpha} \mu .$$

Sachant que la somme des coefficients  $\lambda_{\alpha}$  vaut 1, l'expression précédente se simplifie et devient :

$$\sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) = \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{u}_{\alpha} - \mathbf{u}) - \mu .$$

En substituant le terme de gauche par celui de droite dans l'équation E 2-13, on obtient l'expression suivante de la *variance de l'erreur par krigeage ordinaire* :

E 2-20

$$\sigma_{KO}^2(\mathbf{u}) = \sigma^2 - \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{u}_{\alpha} - \mathbf{u}) - \mu .$$

Ce paragraphe a décrit en détail, pour le krigeage ordinaire, l'ensemble des calculs qui conduisent à établir le système d'équations linéaires de krigeage, à le résoudre et à déduire l'estimateur et la variance d'estimation. Pour les autres méthodes de krigeage qui sont présentées dans la suite, des développements analogues permettent de déduire les expressions mathématiques correspondantes. On se contentera dans les paragraphes suivants de donner les résultats de ces calculs, c'est-à-dire l'expression de l'estimateur et de sa variance, et le système d'équations à partir duquel s'obtiennent les pondérateurs affectés aux données.

### 2.2.2.2 *Krigeage simple*

L'*estimateur par krigeage simple* a pour expression :

$$z_{KS}^*(\mathbf{u}) - m = \sum_{\alpha=1}^n \lambda_{\alpha} [z(\mathbf{u}_{\alpha}) - m],$$

où  $m$ , qui est égal à la valeur moyenne de la variable aléatoire  $Z$ , est supposée connue et constante partout dans le domaine d'étude. En pratique, cette moyenne est assimilée à la valeur moyenne calculée sur l'ensemble des données.

Comme pour le krigeage ordinaire, les pondérateurs  $\lambda_{\alpha}$  sont déterminés en imposant que l'estimateur soit non biaisé et que la variance de l'erreur soit minimale, ce qui conduit au système d'équations linéaires suivant, appelé *système de krigeage simple*,

$$\sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) = C(\mathbf{u}_{\alpha} - \mathbf{u}), \alpha = 1, \dots, n ,$$

où  $C(\mathbf{h})$  est la covariance associée à la variable aléatoire  $Z$  pour un vecteur de séparation  $\mathbf{h}$ . Dans ce système, le terme de gauche contient les covariances entre les données (redondance de l'information) et le terme de droite les covariances entre les données et le point d'estimation (corrélations avec les données). Il est à noter que, dans le cas du krigeage simple, la condition de non-biais est nécessairement vérifiée car la moyenne des variables aléatoires  $Z_{KS}^*(\mathbf{u}) - m$  et  $Z(\mathbf{u}) - m$  est nulle et ce quelles que soient les valeurs des pondérateurs  $\lambda_\alpha$ .

La variance de l'erreur (ou variance d'estimation) par krigeage simple devient :

$$\sigma_{KS}^2(\mathbf{u}) = \sigma^2 - \sum_{\alpha=1}^n \lambda_\alpha C(\mathbf{u}_\alpha - \mathbf{u})$$

où  $\sigma^2$  est la variance globale tirée des données, soit encore  $C(0)$ . Les valeurs de  $\sigma_{KS}^2$  donne une quantification, au moins relative, de l'incertitude sur les valeurs estimées<sup>3</sup>.

### **Krigeage simple avec moyenne locale (variable)**

L'estimateur par krigeage simple réclame une valeur pour la moyenne  $m$  qui est supposée être identique partout (stationnarité de la moyenne). Si on dispose d'une information sur une variable secondaire très largement échantillonnée et corrélée d'une manière ou d'une autre à la variable principale, il est possible de remplacer la moyenne  $m$ , constante partout, par une moyenne  $m(\mathbf{u})$  variable dans l'espace et déduite de l'information secondaire. Un moyen de réaliser cette inférence est de diviser l'intervalle de variation de l'attribut secondaire en  $K$  classes et de calculer, pour chaque classe, la moyenne des valeurs de la variable principale qui ont été mesurées à des endroits où la variable secondaire appartient à la classe considérée. Ceci conduit à calculer  $K$  valeurs moyennes de la variable principale, une pour chaque classe. Ce calibrage peut se faire à partir d'un diagramme de dispersion comme celui de la Figure 2-1. En chaque point d'estimation  $\mathbf{u}$ , la moyenne locale  $m(\mathbf{u})$  est déduite de la valeur co-localisée de la variable secondaire. On parle alors de krigeage simple avec moyenne locale (variable) et l'estimateur s'écrit :

$$z_{KS}^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_\alpha [z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)].$$

#### **2.2.2.3 Krigeage universel**

Le krigeage universel, appelé aussi plus judicieusement krigeage avec modèle de dérive, suppose une variation spatiale de la moyenne locale selon une loi polynomiale fonction des coordonnées. L'ordre du polynôme est donné (choisi préalablement) mais ses coefficients sont inconnus. En général, cette dérive prend la forme suivante :

$$m(\mathbf{u}) = \sum_{l=0}^L a_l f_l(\mathbf{u})$$

---

<sup>3</sup> Une quantification relative de l'incertitude signifie ici que, aux endroits où la variance est plus grande, il est peut être déduit que l'incertitude est plus grande aussi.

où  $m(\mathbf{u})$  est la moyenne au point de coordonnées  $\mathbf{u}$ , les coefficients  $a_l$ ,  $l=0, \dots, L$ , sont les coefficients inconnus de la forme polynomiale, et les fonctions  $f_l(\mathbf{u})$  sont les termes de chaque ordre du polynôme (monômes). Par exemple, en deux dimensions, un modèle linéaire prend la forme suivante :

$$m(\mathbf{u}) = a_0 + a_1x + a_2y$$

et un modèle quadratique

$$m(\mathbf{u}) = a_0 + a_1x + a_2y + a_3x^2 + a_4y^2 + a_5xy.$$

L'estimateur par krigeage universel s'écrit simplement :

$$z_{KU}^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_{\alpha} z(\mathbf{u}_{\alpha}).$$

La condition de non-biais sur cet estimateur conduit à imposer  $L+1$  contraintes que doivent satisfaire les pondérateurs  $\lambda_{\alpha}$  dans la minimisation de l'erreur. Ceci conduit au *système de krigeage universel* dont l'expression est la suivante :

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) + \sum_{l=0}^L \mu_l f_l(\mathbf{u}_{\alpha}) = C(\mathbf{u}_{\alpha} - \mathbf{u}), \alpha = 1, \dots, n \\ \sum_{\beta=1}^n \lambda_{\beta} f_l(\mathbf{u}_{\beta}) = f_l(\mathbf{u}), l = 0, \dots, L \end{cases}$$

La limitation de la forme polynomiale à l'ordre zéro équivaut au krigeage ordinaire qui correspond à un modèle de moyenne  $m(x, y) = a_0$ , c'est-à-dire à une moyenne constante mais inconnue. La *variance de l'erreur par krigeage universel* a pour expression :

$$\sigma_{KU}^2(\mathbf{u}) = \sigma^2 - \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{u}_{\alpha} - \mathbf{u}) - \sum_{l=0}^L \mu_l f_l(\mathbf{u}_{\alpha}), \alpha = 1, \dots, n.$$

### 2.2.3 Krigeage avec dérive externe

Le krigeage avec dérive externe est une extension du krigeage universel (voir par exemple Journel et Rossi, 1989, Deutsch, 1991). Il revient à considérer une fonction de dérive  $f_1(\mathbf{u})$ , définie partout, à partir d'une variable secondaire  $z_2(\mathbf{u})$  qui varie en douceur dans l'espace. Il est indispensable de disposer d'une valeur de variable secondaire en chaque point d'estimation. Si tel n'est pas le cas, une solution consiste à kriger (estimer) préalablement la variable secondaire pour en déduire une valeur partout.

En général, il est adopté une relation linéaire entre la variable secondaire  $z_2(\mathbf{u})$  et la moyenne  $m_1(\mathbf{u})$  de la variable primaire :

$$m_1(\mathbf{u}) = a_0 + a_1 z_2(\mathbf{u}).$$

L'estimateur par krigeage avec dérive externe s'écrit alors :

$$z_{KDE}^*(\mathbf{u}) = \sum_{\alpha=1}^{n_1} \lambda_{\alpha} z_1(\mathbf{u}_{\alpha}),$$

où  $n_1$  est le nombre de données, correspondant à la variable primaire  $z_1(\mathbf{u})$ , à prendre en compte en chaque point d'estimation. Les coefficients  $\lambda_{\alpha}$  de cet estimateur s'obtiennent en résolvant le système de krigeage avec dérive externe :

$$\left\{ \begin{array}{l} \sum_{\beta=1}^{n_1} \lambda_{\beta} C_1(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) + \mu_0 + \mu_1 z_2(\mathbf{u}_{\alpha}) = C_1(\mathbf{u}_{\alpha} - \mathbf{u}), \alpha = 1, \dots, n_1 \\ \sum_{\beta=1}^{n_1} \lambda_{\beta} = 1 \\ \sum_{\beta=1}^{n_1} \lambda_{\beta} z_2(\mathbf{u}_{\beta}) = z_2(\mathbf{u}) \end{array} \right. .$$

Les deux dernières équations proviennent de la condition de non-biais à imposer à l'estimateur. Il est à noter que ce système d'équations utilise uniquement la covariance de la variable principale. L'expression de la variance de l'erreur par krigeage avec dérive externe est la suivante :

$$\sigma_{KDE}^2(\mathbf{u}) = \sigma^2 - \sum_{\alpha=1}^{n_1} \lambda_{\alpha} C_1(\mathbf{u}_{\alpha} - \mathbf{u}) - \mu_0 - \mu_1 z_2(\mathbf{u}_{\alpha}), \alpha = 1, \dots, n_1 .$$

Le champ  $z_1(\mathbf{u})$  ainsi estimé est, par construction, semblable à celui de la variable secondaire. Néanmoins, cette méthode n'intègre pas de corrélation croisée entre les deux variables; la variable secondaire ne renseigne que sur la dérive de la variable primaire.

Par exemple, s'il peut être justifié de penser que les temps de parcours sismiques déterminent la dérive locale de la porosité, mais qu'il n'est pas possible, à partir de données de porosité rares et dispersées, d'établir de corrélation entre les deux variables, le krigeage avec dérive externe est un moyen d'incorporer dans le processus d'estimation cette information qui demeure utile.

## 2.2.4 Cokrigeage

Les techniques de krigeage présentées précédemment se distinguent entre elles par la manière qu'elles ont d'incorporer une surface moyenne dans un problème d'interpolation spatiale. Concernant l'incorporation d'informations secondaires, le krigeage simple avec moyenne locale comme le krigeage avec dérive externe requièrent une information secondaire disponible en chaque point d'estimation. Dans bien des cas, l'information secondaire n'existe pas partout de façon exhaustive. La technique de cokrigeage permet d'incorporer une ou plusieurs variables secondaires dans l'estimation d'un attribut principal quand les variables secondaires ne sont pas connues partout à l'intérieur du domaine d'étude.

D'un point de vue algorithmique, le cokrigeage ne se distingue pas du krigeage (voir Myers, 1982, 1984 et 1985, Isaaks et Srivastava, 1989, Goovaerts, 1997). Sa nouveauté est de per-

mettre aux données secondaires de participer directement à l'estimation de la variable principale. Ceci signifie que l'information secondaire se traite comme une co-variable. Du point de vue de l'incorporation d'attributs externes, le cokrigeage fait un meilleur usage de l'information car, dans l'estimation de la variable principale, il tient compte à la fois de la valeur de la variable secondaire et de ses formes de variation spatiale. Le niveau de co-dépendance linéaire entre les variables primaire et secondaire est contenu dans les covariances croisées. A la différence du krigeage avec dérive, la variable externe n'a pas besoin d'être disponible en chaque point où la variable principale est à estimer. La forme générale que prend l'estimateur par cokrigeage d'une variable  $z_1$  en un point de coordonnées  $(\mathbf{u})$  est la suivante :

$$E\ 2-21 \quad z_1^*(\mathbf{u}) - m_1(\mathbf{u}) = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} [z_1(\mathbf{u}_{\alpha_1}) - m_1(\mathbf{u}_{\alpha_1})] + \sum_{i=2}^{N_v} \sum_{\alpha_i=1}^{n_i} \lambda_{\alpha_i} [z_i(\mathbf{u}_{\alpha_i}) - m_i(\mathbf{u}_{\alpha_i})]$$

où

- $z_1(\mathbf{u}_{\alpha_1})$ ,  $\alpha_1 = 1, \dots, n_1$ , sont les  $n_1$  données principales,
- $z_i(\mathbf{u}_{\alpha_i})$ ,  $\alpha_i = 1, \dots, n_i$ ,  $i = 2, \dots, N_v$ , sont les données secondaires qui correspondent à  $N_v - 1$  variables secondaires chacune renseignée par  $n_i$  données,
- $m_1$  et  $m_i$  sont respectivement les moyennes de la variable principale et de la variable secondaire  $z_i$  ( $i = 2, \dots, N_v$ ),
- les coefficients  $\lambda_{\alpha_1}$  et  $\lambda_{\alpha_i}$  sont les coefficients de pondération des données à calculer en chaque point d'estimation.

Dans cette expression, l'emplacement et le nombre de données peuvent être différents d'une variable à l'autre, variables secondaires et principale confondues. Par mesure de simplification, sauf cas particuliers clairement spécifiés, il ne sera considéré dans la suite qu'une seule variable secondaire. Comme pour le krigeage, les coefficients de pondération dans l'équation E 2-21 s'obtiennent, en chaque point d'estimation, par résolution du système d'équations linéaires qui provient de la minimisation de la variance de l'erreur (équation E 2-5) soumise à la condition de non-biais (équation E 2-7). Sont décrites brièvement dans la suite les diverses formes de cokrigeage.

#### 2.2.4.1 Cokrigeage simple

Lorsque l'on dispose de données à la fois sur une variable principale  $z_1$  et sur une variable secondaire  $z_2$ , l'estimateur par cokrigeage simple de la variable principale prend la forme suivante :

$$z_{CKS}^*(\mathbf{u}) - m_1 = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} [z_1(\mathbf{u}_{\alpha_1}) - m_1] + \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2} [z_2(\mathbf{u}_{\alpha_2}) - m_2]$$

où  $m_1$  et  $n_1$  sont la moyenne et le nombre de données de la variable principale  $z_1$ , et  $m_2$  et  $n_2$  ceux de la variable secondaire. Les poids  $\lambda_{\alpha}$  se calculent par résolution du système d'équations linéaires ou *système de cokrigeage simple* :

$$\begin{cases}
 \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} C_1(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_1}) + \sum_{\beta_2=1}^{n_2} \lambda_{\beta_2} C_{12}(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_2}) \\
 \qquad \qquad \qquad = C_1(\mathbf{u}_{\alpha_1} - \mathbf{u}), \alpha_1 = 1, \dots, n_1 \\
 \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} C_{12}(\mathbf{u}_{\alpha_2} - \mathbf{u}_{\beta_1}) + \sum_{\beta_2=1}^{n_2} \lambda_{\beta_2} C_2(\mathbf{u}_{\alpha_2} - \mathbf{u}_{\beta_2}) \\
 \qquad \qquad \qquad = C_{12}(\mathbf{u}_{\alpha_2} - \mathbf{u}), \alpha_2 = 1, \dots, n_2
 \end{cases}$$

E 2-22

où intervient la covariance de la variable primaire  $C_1(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_1})$  mais aussi celle de la variable secondaire  $C_2(\mathbf{u}_{\alpha_2} - \mathbf{u}_{\beta_2})$  et la covariance croisée  $C_{12}(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_2})$ . Comme dans le krigeage simple, la méthode de cokrigeage simple nécessite de connaître les moyennes des variables considérées. L'expression de la *variance de l'erreur (ou variance d'estimation) par cokrigeage simple* est la suivante :

$$\sigma_{CKS}^2(\mathbf{u}) = \sigma_1^2 - \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} C_1(\mathbf{u}_{\alpha_1} - \mathbf{u}) - \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2} C_{12}(\mathbf{u}_{\alpha_2} - \mathbf{u}).$$

#### 2.2.4.2 Cokrigeage ordinaire

L'estimateur par cokrigeage ordinaire est donné par l'équation suivante :

$$z_{CKO}^*(\mathbf{u}) = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} z_1(\mathbf{u}_{\alpha_1}) + \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2} z_2(\mathbf{u}_{\alpha_2})$$

où les coefficients de pondération proviennent de la résolution du système d'équations suivant ou *système de cokrigeage ordinaire* :

$$\begin{cases}
 \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} C_1(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_1}) + \sum_{\beta_2=1}^{n_2} \lambda_{\beta_2} C_{12}(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_2}) + \mu_1 \\
 \qquad \qquad \qquad = C_1(\mathbf{u}_{\alpha_1} - \mathbf{u}), \alpha_1 = 1, \dots, n_1 \\
 \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} C_{12}(\mathbf{u}_{\alpha_2} - \mathbf{u}_{\beta_1}) + \sum_{\beta_2=1}^{n_2} \lambda_{\beta_2} C_2(\mathbf{u}_{\alpha_2} - \mathbf{u}_{\beta_2}) + \mu_2 \\
 \qquad \qquad \qquad = C_{12}(\mathbf{u}_{\alpha_2} - \mathbf{u}), \alpha_2 = 1, \dots, n_2 \\
 \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} = 1 \\
 \sum_{\beta_2=1}^{n_2} \lambda_{\beta_2} = 0
 \end{cases}$$

E 2-23

où les deux dernières équations sont des contraintes suffisantes, bien que non nécessaires, pour garantir une estimation sans biais. Les deux variables  $\mu_1$  et  $\mu_2$  sont les paramètres de Lagrange introduits dans la minimisation pour tenir compte des deux contraintes. L'expression de la *variance de l'erreur par cokrigeage ordinaire* est la suivante :

$$\sigma_{CKO}^2(\mathbf{u}) = \sigma_1^2 - \mu_1 - \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} C_1(\mathbf{u}_{\alpha_1} - \mathbf{u}) - \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2} C_{12}(\mathbf{u}_{\alpha_2} - \mathbf{u}).$$

### 2.2.5 Cokrigage co-localisé

La résolution d'un système de cokrigage complet, comme ceux présentés aux paragraphes précédents (systèmes d'équations E 2-22 et E 2-23), suppose de connaître les fonctions de covariance des variables principale et secondaire et la covariance croisée entre elles (cas où d'une variable secondaire unique). Ceci signifie davantage d'effort en terme de modélisation et de temps de calcul machine par rapport aux techniques de krigeage qui elles ne nécessitent que la fonction de covariance de la variable principale. Très souvent, cet effort supplémentaire constitue la raison principale pour ne pas utiliser le cokrigage.

Lorsque l'information secondaire est échantillonnée exhaustivement, comme cela est fréquent avec des données sismiques, il existe une difficulté supplémentaire dans l'implémentation du cokrigage complet. Cette difficulté se manifeste à travers l'instabilité des matrices du système de cokrigage à résoudre. L'instabilité est causée par la proximité extrême et la forte auto-corrélation qui existent en général pour les données douces, et les grandes distances et la plus faible corrélation des données dures.

Ces raisons, plus le fait que s'il existe au point d'estimation une donnée secondaire co-localisée celle-ci tend à atténuer l'influence des données voisines (effet d'écran), nous conduisent à présenter l'alternative du cokrigage co-localisé (voir Xu *et al.*, 1992, Journel *et al.*, 1992, et Almeida et Journel, 1994). Il s'agit d'un cokrigage qui, comme information externe pour estimer la variable primaire au point de coordonnées  $\mathbf{u}$ , ne retient que la donnée secondaire localisée en ce point (donnée co-localisée). L'*estimateur par cokrigage co-localisé simple* de la variable principale s'écrit :

$$z_{CKCS}^*(\mathbf{u}) - m_1 = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} [z_1(\mathbf{u}_{\alpha_1}) - m_1] + \lambda_2 [z_2(\mathbf{u}) - m_2]$$

et le système de cokrigage co-localisé simple qui fournit les coefficients de pondération est :

$$\begin{cases} \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} C_1(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_1}) + \lambda_2 C_{12}(\mathbf{u}_{\alpha_1} - \mathbf{u}) = C_1(\mathbf{u}_{\alpha_1} - \mathbf{u}), \alpha_1 = 1, \dots, n_1 \\ \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} C_{12}(\mathbf{u} - \mathbf{u}_{\beta_1}) + \lambda_2 C_2(0) = C_{12}(0) \end{cases}.$$

Un système tel que le précédent est identique à celui du cokrigage simple complet où le nombre de données secondaires  $n_2$  est égal 1. Même ainsi, la méthode nécessite le calcul des covariances croisées entre variables  $C_{12}(\mathbf{h})$ , où  $\mathbf{h}$  est le vecteur de séparation entre les points considérés<sup>4</sup>. L'utilisation d'un modèle markovien décharge l'utilisateur de cette tâche.

<sup>4</sup> C'est-à-dire, dans l'équation ci-dessus, le point d'estimation  $\mathbf{u}$  pour la variable secondaire, les points de coordonnées  $\mathbf{u}_{\alpha}$  des données pour la variable principale.

### 2.2.5.1 Cokrigage co-localisé selon un modèle markovien

Le modèle markovien stipule que la donnée dure  $z_1(\mathbf{u})$  "occulte" (effet d'écran) l'influence de tout autre donnée dure  $z_1(\mathbf{u}+\mathbf{h})$  sur la donnée secondaire co-localisée  $z_2(\mathbf{u})$ , c'est-à-dire :

$$E\{Z_2(\mathbf{u})|Z_1(\mathbf{u})=z, Z_1(\mathbf{u}+\mathbf{h})=z'\} = E\{Z_2(\mathbf{u})|Z_1(\mathbf{u})=z\}, \forall \mathbf{h}, z.$$

En d'autres termes, la relation E 2-24 signifie que la dépendance de la variable secondaire sur la variable primaire est limitée à la donnée primaire co-localisée. A partir de cette hypothèse, on peut démontrer (voir Annexe 2) que les covariances croisées  $C_{12}(\mathbf{h})=C_{21}(\mathbf{h})$ , du système de cokrigage, peuvent s'écrire en fonction de la covariance des données dures comme :

$$C_{12}(\mathbf{h}) = \frac{C_{12}(0)}{C_1(0)} C_1(\mathbf{h}), \forall \mathbf{h}$$

ou, en termes de corrélogramme,

$$\rho_{12}(\mathbf{h}) = \rho_{12}(0)\rho_1(\mathbf{h}), \forall \mathbf{h}.$$

Dans cette dernière expression,  $\rho_1(\mathbf{h})$  est le corrélogramme de la variable principale,  $\rho_{12}(\mathbf{h})$  est le corrélogramme croisé entre les variables principale et secondaire et  $\rho_{12}(0)$  est le coefficient de corrélation entre ces deux variables. Ceci revient à dire que, selon un modèle de Markov, la covariance croisée  $C_{12}(\mathbf{h})$  se calcule par une remise à l'échelle de la covariance de la variable principale. Le facteur d'échelle est le rapport entre  $C_{12}(0)$  et  $C_1(0)$ .

Dans sa forme standard, l'estimateur par cokrigage co-localisé selon un modèle de Markov s'écrit de la façon suivante :

$$\frac{z_{CKCS}^*(\mathbf{u}) - m_1}{\sigma_1} = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} \frac{[z_1(\mathbf{u}_{\alpha_1}) - m_1]}{\sigma_1} + \lambda_2 \frac{[z_2(\mathbf{u}) - m_2]}{\sigma_1}$$

où les pondérateurs sont solution du système de cokrigage co-localisé selon un modèle de Markov :

$$\begin{cases} \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} \rho_1(\mathbf{u}_{\alpha_1} - \mathbf{u}_{\beta_1}) + \lambda_2 \rho_{12}(0) \rho_1(\mathbf{u}_{\alpha_1} - \mathbf{u}) = \rho_1(\mathbf{u}_{\alpha_1} - \mathbf{u}), \alpha_1 = 1, \dots, n_1 \\ \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} \rho_{12}(0) \rho_1(\mathbf{u} - \mathbf{u}_{\beta_1}) + \lambda_2 \rho_2(0) = \rho_{12}(0) \end{cases}$$

Dans ce système, si  $\rho_{12}(0) = 0$  ( $Z_1$  et  $Z_2$  indépendantes), il s'en suit que  $\lambda_2 = 0$  et la donnée secondaire est ignorée. Si  $\rho_{12}(0) = 1$  ( $Z_1$  et  $Z_2$  parfaitement corrélée linéairement), le système se transforme en un système de krigeage simple, avec  $(n + 1)$  données, dont la solution est

$\lambda_2 = 1$  et  $\lambda_{\beta_1} = 0$  pour tout  $\beta_1$ ; l'estimateur est alors égal à la valeur (remise à l'échelle) de la variable secondaire.

L'algorithme de cokrigeage co-localisé markovien nécessite le même effort d'implémentation que le krigeage avec dérive externe mais il présente l'avantage important de permettre à la variable secondaire de participer directement à l'estimation de la variable principale. En outre, et à la différence du krigeage avec dérive externe, le modèle de cokrigeage co-localisé se fonde sur un paramètre de calibrage qui est le coefficient de corrélation, ce qui permet de valider le choix de ce modèle à partir des données (il suffit alors d'établir l'existence d'une corrélation linéaire entre les deux variables). Il requiert, cependant, que la variable secondaire soit connue en chaque point d'estimation de la variable primaire.

## 2.2.6 Krigeage d'indicatrice

Les méthodes vues jusqu'à présent fournissent comme résultat une valeur estimée de la variable principale en chaque point considéré. La valeur estimée se calcule en fonction des valeurs que les variables principale et secondaire prennent dans le voisinage du point d'estimation; le mode de calcul dépend alors de la technique adoptée. Nous nous intéresserons, dans la suite, à un ensemble de techniques qui fournissent non pas une estimation de la valeur de la variable modélisée mais une estimation de la probabilité que cette valeur dépasse un seuil donné. Comme nous le verrons, ces techniques permettent de déduire en tout point la distribution complète des valeurs de la variable. Une telle distribution, qui se construit également en considérant, dans un certain voisinage du point, l'information locale sur les deux variables, constitue un modèle d'incertitude sur l'attribut principal. A partir de ce modèle, il est alors possible de choisir une valeur estimée affectée au point.

Ces méthodes ne travaillent pas directement sur la variable considérée mais sur des variables transformées sous forme d'indicatrices. La variable indicatrice  $I(\mathbf{u}; z_k)$ , associée à la variable  $Z(\mathbf{u})$  pour un seuil  $z_k$  en un point de coordonnées  $\mathbf{u}$ , se définit conformément à la fonction suivante :

$$\text{E 2-25} \quad I(\mathbf{u}; z_k) = \begin{cases} 1 & \text{si } Z(\mathbf{u}) \leq z_k \\ 0 & \text{si } Z(\mathbf{u}) > z_k \end{cases} .$$

Dans le krigeage d'indicatrice (voir Journel, 1983, 1986, 1989, Alabert, 1987, Isaaks et Srivastava, 1989, Gómez-Hernández, 1991, 1993a, Goovaerts, 1997), l'intervalle de variation de la variable étudiée se décompose en  $K+1$  classes à partir de  $K$  seuils.

Par exemple, considérons une variable dont l'intervalle de variation est divisé en trois classes à partir de deux seuils de 25 et 50. Au point de coordonnées  $\mathbf{u}$ , supposons qu'il existe une donnée dont la valeur de l'attribut est 43. Les fonctions d'indicatrice (équation E 2-25) au point  $\mathbf{u}$  prennent alors les valeurs suivantes :  $i(\mathbf{u}; 25) = 0$  et  $i(\mathbf{u}; 50) = 1$ . Si cette opération est répétée pour chaque donnée, la transformation de la variable originale, dans un formalisme par indicatrices, conduit à déterminer un ensemble de vecteurs, un pour chaque donnée. Pour une donnée située au point de coordonnées  $\mathbf{u}$ , le vecteur correspondant est composé des valeurs des  $K$  variables binaires  $i(\mathbf{u}; z_k)$ ,  $k=1, \dots, K$ .

Choisissons de réaliser un krigeage ordinaire des variables indicatrices, pour chaque seuil, de manière à obtenir au point de coordonnées  $\mathbf{u}$  l'estimateur par krigeage d'indicatrice suivant :

$$i^*(\mathbf{u}; z_k) = \sum_{\alpha=1}^n \lambda_{\alpha}(\mathbf{u}_{\alpha}; z_k) i(\mathbf{u}_{\alpha}; z_k).$$

Pour déduire les coefficients  $\lambda_{\alpha}$ , qui dans le cas le plus général dépendent à la fois des coordonnées du point d'estimation et du seuil considéré, il est nécessaire de résoudre le système d'équations suivant appelé *système de krigeage d'indicatrice* :

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta}(\mathbf{u}_{\beta}; z_k) C_I(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}; z_k) + \mu(\mathbf{u}; z_k) = C_I(\mathbf{u}_{\alpha} - \mathbf{u}; z_k), \alpha = 1, \dots, n \\ \sum_{\beta=1}^n \lambda_{\beta}(\mathbf{u}_{\beta}; z_k) = 1 \end{cases}$$

où

- $C_I(\mathbf{h}; z_k)$ , avec  $\mathbf{h}$  égal au vecteur de séparation entre points, est la fonction de covariance de la variable indicatrice; dans le cas le plus général, cette fonction est différente pour chaque seuil considéré;
- $\mu$  est le paramètre de Lagrange.

Le concept de covariance d'indicatrice est identique à celui de covariance classique, si ce n'est qu'il est calculé pour des variables indicatrices.

La valeur estimée  $i^*(\mathbf{u}; z_k)$ , généralement comprise entre 0 et 1, est en tant que tel un mauvais estimateur de la variable indicatrice qui elle ne peut prendre que les valeurs 0 ou 1. Cependant, lorsque l'on interprète cette valeur comme celle de l'espérance mathématique (la moyenne) de la variable indicatrice, soit  $E\{i(\mathbf{u}; z_k)\}$ , il s'en suit que  $i^*(\mathbf{u}; z_k)$  est le meilleur estimateur linéaire de la fonction cumulative de distribution conditionnelle, c'est-à-dire de la probabilité conditionnelle, par rapport aux données, que la variable soit inférieure au seuil  $z_k$ .

Ceci est la conséquence de la définition même de la variable indicatrice  $I(\mathbf{u}; z)$  dont la valeur moyenne est égale à la fonction cumulative de distribution, appelée aussi plus brièvement fonction de distribution et notée  $F(z)$  :

$$\begin{aligned} E\{I(\mathbf{u}; z)\} &= 1 \cdot \text{Prob}[Z(\mathbf{u}) \leq z] + 0 \cdot \text{Prob}[Z(\mathbf{u}) > z] \\ &= \text{Prob}[Z(\mathbf{u}) \leq z] = F(z) \end{aligned}$$

De cette manière, la relation suivante est vérifiée :

$$F^*[\mathbf{u}; z_k | (n)] = \text{Prob}^*[Z(\mathbf{u}) \leq z_k | (n)] = \sum_{\alpha=1}^n \lambda_{\alpha}(\mathbf{u}_{\alpha}; z_k) i(\mathbf{u}_{\alpha}; z_k)$$

où  $F^*[\cdot]$  est l'estimateur de la fonction de distribution conditionnelle par rapport aux  $n$  données qui interviennent dans l'estimation.

Le krigeage de la variable indicatrice pour chaque seuil  $z_k$ ,  $k=1,\dots,K$ , fournit une estimation de la fonction de distribution de la variable étudiée au point de coordonnées  $\mathbf{u}$ . Plus exactement, il fournit des valeurs discrètes de la fonction de distribution (conformément au nombre de seuils sélectionnés), valeurs qui une fois interpolées et extrapolées convenablement donnent une fonction continue. A partir de cette fonction de distribution, il peut alors être choisi comme estimateur  $z^*(\mathbf{u})$  de la variable  $Z(\mathbf{u})$  la valeur moyenne, la valeur médiane ou toute autre valeur de quantile.

Le krigeage d'indicatrice présente un avantage important sur les techniques de krigeage classiques : il est capable d'intégrer des informations imprécises ou douces. Le formalisme par indicatrices permet de codifier de manière générale et simple les données qui proviennent de différentes sources et sont associées à des degrés de confiance variables. Pour comprendre cet aspect du krigeage d'indicatrice, voyons l'exemple suivant.

On suppose que l'intervalle de variation de la variable originale à estimer est  $[0 ; 100]$  (cas par exemple d'une porosité en pourcentage). On décide de décomposer (discrétiser) cet intervalle en  $K = 10$  classes au moyen de 9 seuils, par exemple :

$$\{z_k, k=1,\dots,9\} = \{5 ; 10 ; 20 ; 30 ; 40 ; 60 ; 80 ; 90 ; 95\}.$$

On dispose de trois données qui proviennent de sources différentes.

1. Une mesure de porosité de 25% obtenue en laboratoire sur carotte; cette mesure peut être considérée comme exacte et constitue donc une donnée dure (incertitude nulle ou négligeable sur la valeur).
2. Une donnée calibrée par régression linéaire, à partir d'une information géophysique (par exemple une mesure de conductivité ou de résistivité électrique); elle indique que la porosité, à l'endroit de la mesure géophysique, doit appartenir à l'intervalle  $[15 ; 35]$ . Ce type de donnée s'appelle *donnée d'intervalle*.
3. En un point où on ne dispose d'aucune mesure, mais où on connaît le type de terrain, il est adopté une distribution a priori (empirique) de la porosité qui provient d'un avis expert ou de données répertoriées. Cette distribution détermine les quantiles suivants, correspondant aux seuils définis précédemment :

$$\{P(Z \leq z_k), k=1,\dots,9\} = \{0 ; 0 ; 0,3 ; 0,5 ; 0,7 ; 1 ; 1 ; 1 ; 1\}.$$

Chaque quantile exprime la probabilité que la valeur de porosité soit inférieure ou égale au seuil correspondant. Par exemple, pour un seuil de 20% (troisième valeur), la probabilité est de 0,3.

En appliquant la définition de la fonction d'indicatrice (équation E 2-25), chaque donnée se transforme en un vecteur de variables indicatrices comme suit.

1. Donnée dure :  $z = 25 \rightarrow \{0 ; 0 ; 0 ; 1 ; 1 ; 1 ; 1 ; 1 ; 1\}$ .
2. Donnée d'intervalle :  $z \in [15,35] \rightarrow \{0 ; 0 ; ? ; ? ; 1 ; 1 ; 1 ; 1 ; 1\}$ .
3. Distribution a priori :  $\rightarrow \{0 ; 0 ; 0,3 ; 0,5 ; 0,7 ; 1 ; 1 ; 1 ; 1\}$ .

Ainsi, la donnée dure est représentée par un vecteur complet d'indicatrices dures. Le vecteur est complet car la valeur exacte de la donnée est connue. Les indicatrices sont dites dures car elles prennent les valeurs 0 ou 1.

La donnée d'intervalle est décrite par un vecteur incomplet d'indicatrices dures. Le vecteur est incomplet car les valeurs des variables indicatrices sont connues à l'extérieur de l'intervalle (seuil inférieur à la borne inférieure ou supérieur à la borne supérieure) mais pas à l'intérieur de celui-ci.

Pour la donnée de type distribution a priori, la transformation en indicatrices n'affecte pas la donnée qui est déjà définie en terme de probabilité d'être inférieure à chaque seuil. Le vecteur est complet et constitué de données "douces" avec des valeurs entre 0 et 1.

De cette manière, toute l'information disponible peut se codifier en vecteurs d'indicatrices que peut recevoir l'algorithme de krigeage d'indicatrice pour estimer ici la porosité. Lorsque l'information est traduite selon le formalisme d'indicatrices, il n'existe plus de différence entre les données dures et douces d'un point de vue algorithmique. Pour le système de krigeage d'indicatrice, toute l'information est sous forme de vecteurs dans lesquels certaines valeurs peuvent être manquantes lorsque le vecteur est incomplet.

### 2.2.7 Cokrigeage d'indicatrice

Dans un modèle tel que celui présenté précédemment, il n'est pas possible de distinguer la forme de continuité spatiale des données dures de celle des données douces. Dans quelques applications, cette distinction entre les continuités spatiales des indicatrices correspondant aux données dures et douces peut être importante. Dans l'exemple du paragraphe précédent, la rareté des données de puits fait qu'il est peu réaliste de vouloir inférer des variations de porosité dans la région entre puits à partir de cette information unique.

Se pose alors le besoin d'un algorithme qui, en plus de considérer des données de types différents, fait une distinction entre leurs formes de continuité spatiale. Une solution consiste à considérer un *estimateur par cokrigeage d'indicatrice* tel que celui donné ci-dessous :

$$i^*(\mathbf{u}; z_k) = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1}(\mathbf{u}_{\alpha_1}; z_k) i(\mathbf{u}_{\alpha_1}; z_k) + \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2}(\mathbf{u}_{\alpha_2}; z_k) i(\mathbf{u}_{\alpha_2}; z_k) .$$

Dans cet estimateur, on distingue l'influence des  $n_1$  données dures de celle des  $n_2$  données douces codifiées sous le formalisme d'indicatrice.

La résolution du système de cokrigeage, correspondant à cet estimateur, requiert un calcul laborieux d'au moins deux auto-covariances et une covariance croisée pour chaque seuil, dans

le cas où une seule variable secondaire est utilisée, et plus encore si plusieurs variables secondaires interviennent. Comme pour le cokrigage classique, l'expérience a montré que la nécessité de caler plusieurs modèles de covariance conduit en pratique à une très faible utilisation du cokrigage. Pour éviter cette difficulté, la solution est d'avoir recours à un modèle markovien qui permet d'alléger le calcul des fonctions de covariance.

### 2.2.7.1 Cokrigage d'indicatrice suivant un modèle de Markov-Bayes

Comme pour le krigeage d'indicatrice, dans le cokrigage d'indicatrice (Journel et Zhu, 1990, Zhu, 1991, Zhu et Journel, 1992), l'information locale, dure, douce ou globale, disponible en tout point<sup>5</sup>, peut être codifiée comme une fonction de distribution a priori de l'attribut modélisé. Cette distribution a priori locale est ensuite améliorée (mise à jour) en utilisant les données dures et douces avoisinantes, ce qui conduit à construire une distribution dite a posteriori à partir de laquelle une valeur peut être déduite comme estimateur de la variable étudiée au point considéré. Le modèle de Markov-Bayes est un moyen de déduire une telle distribution a posteriori, conditionnelle à la fois par rapport aux données dures et douces.

Dans un schéma bayésien, toutes les données, dures et douces, peuvent se traduire sous la forme de distributions locales a priori qui s'écrivent comme suit :

$$\text{Prob}\{Z(\mathbf{u}) \leq z_k | \text{information locale en } \mathbf{u}\}$$

où  $Z(\mathbf{u})$  est la fonction aléatoire qui représente la variable principale étudiée au point d'estimation de coordonnées  $\mathbf{u}$  et  $z_k$ ,  $k=1, \dots, K$ , désigne les seuils pour lesquels la probabilité est connue ou calculée. Conformément au formalisme par indicatrices vu précédemment, ces distributions a priori peuvent être représentées par des vecteurs complets ou incomplets d'indicatrices dures (valeurs de 0 ou 1) ou douces (valeurs entre 0 et 1). On notera dans la suite  $i(\mathbf{u}; z_k)$  les termes d'un vecteur d'indicatrices dures et  $y(\mathbf{u}; z_k)$  ceux d'un vecteur d'indicatrices douces.

L'information locale désigne alors l'information disponible en  $\mathbf{u}$  et peut prendre une des formes suivantes.

- S'il existe une donnée dure  $z(\mathbf{u})$ , elle se caractérise par un vecteur complet d'indicatrices dures et se traite comme une fonction de distribution de variance nulle, c'est-à-dire :

$$i(\mathbf{u}; z_k) = \begin{cases} 0, \forall z \leq z_k(\mathbf{u}) \\ 1, \forall z > z_k(\mathbf{u}) \end{cases}$$

---

<sup>5</sup> Comme cela a été illustré par un exemple au paragraphe §2.2.6, l'information locale peut se limiter à la connaissance du type de terrain (ou à tout autre information déterminant des régions stationnaires statistiquement) auquel est associé une distribution a priori de l'attribut modélisé. Si une telle information est absente également, l'information locale se résume à une distribution a priori **globale** qui s'applique à l'ensemble du domaine modélisé dont on suppose alors, nécessairement, qu'il est stationnaire.

- S'il existe une donnée d'intervalle qui indique que  $z(\mathbf{u})$  appartient à l'intervalle  $]a, b]$ , elle est codée comme un vecteur incomplet d'indicatrices dures qui correspond à une fonction de distribution incomplète, soit :

$$i(\mathbf{u}; z_k) = \begin{cases} 0, \forall z \leq a \\ \text{indéfinie}, \forall z \in ]a; b] \\ 1, \forall z > b \end{cases} .$$

- S'il existe une donnée secondaire  $v(\mathbf{u})$ , une fonction de distribution locale doit être calibrée pour représenter la distribution a priori de la variable principale conditionnelle par rapport à l'information secondaire locale, c'est-à-dire :

$$y(\mathbf{u}; z_k) = \text{Prob}\{Z(\mathbf{u}) \leq z_k | v(\mathbf{u})\} \in [0;1].$$

Ce calibrage se fait à partir de l'ensemble des paires de données disponibles (données primaire et secondaire mesurées en un même point). L'intervalle de variation de chaque variable est divisé en plusieurs classes et pour chaque seuil de la variable primaire sont calculées les proportions de données suivantes :

$$\text{proportion}\{Z \leq z_k | V \in v_l\}, l = 1, \dots, L$$

où  $z_k$  et  $v_l$  désignent les seuils respectifs considérés pour la variable primaire  $Z$  et la variable secondaire  $V$ . Ces proportions se calculent en comptant simplement le nombre de mesures de  $Z$  inférieures à chaque seuil  $z_k$  pour chaque intervalle de  $V^6$ . Ce nombre est alors divisé par le nombre total de mesures dans l'intervalle de  $V$  considéré.

- Enfin, s'il n'existe au point  $\mathbf{u}$  aucune information d'un des types précédents, l'information locale se réduit à une fonction de distribution globale  $F(z)$  affectée à l'ensemble du domaine et connue au moins pour chacun des seuils  $z_k$  choisis pour décrire la variable principale. Cette distribution globale fournit pour chaque seuil la probabilité moyenne d'avoir une valeur  $z$  inférieure ou égale au seuil (c'est-à-dire, sur l'ensemble des valeurs de  $z$  dans le domaine, la proportion de valeurs  $z \leq$  au seuil  $z_k$ ).

En résumé, pour chaque point d'estimation  $\mathbf{u}$  à l'intérieur du domaine d'étude, il peut être défini une distribution a priori locale sous la forme d'un vecteur de valeurs de probabilité. La dimension du vecteur est égale au nombre  $K$  de seuils choisis pour discrétiser la variable principale. Selon le type d'information locale disponible, tout ou partie des termes du vecteur sont définis. Une distinction est faite entre les indicatrices dures (valeurs de 0 ou 1) qui proviennent d'une information précise mais parfois partielle (données d'intervalle) et les indicatrices douces (valeurs continues entre 0 et 1) qui se déduisent d'une variable secondaire. Cette distinction permet de considérer une continuité spatiale (ou auto-covariance) différente pour ces deux types d'indicatrices qui découlent d'informations d'origine et de signification différentes (informations primaires et secondaires).

---

<sup>6</sup> Il est à noter que dans certains cas, pour éviter les biais d'échantillonnage (par ex. regroupement de données dans certaines zones), les données doivent être pondérées, un poids plus grand étant attribué aux données les plus isolées. L'opération qui consiste à calculer ces poids s'appelle *dégroupement* (*declustering* en anglais).

Il est important de garder à l'esprit que la distribution a priori ainsi obtenue en chaque point d'estimation dépend au plus de l'information locale au point d'estimation. La distribution dite a posteriori se calcule alors, toujours en chaque point d'estimation, à partir des données avoisinantes, c'est-à-dire à partir des distributions affectées aux points d'échantillonnage ou de mesure proches (données dures, d'intervalle ou secondaires). Ce calcul fournit une distribution de probabilités qui est généralement différente de la distribution a priori et s'écrit sous la forme générale suivante :

$$\text{Prob}\{Z(\mathbf{u}) \leq z_k \mid \text{toute l'information disponible au point } \mathbf{u} \text{ et dans son voisinage}\}.$$

Les distributions locales a priori  $i(\mathbf{u}; z_k)$  ou  $y(\mathbf{u}; z_k)$ , associées à des données dures ou douces, peuvent être interprétées comme des réalisations de fonctions aléatoires  $I(\mathbf{u}; z_k)$  ou  $Y(\mathbf{u}; z_k)$ . L'estimateur par cokrigage, à utiliser pour améliorer en tout point d'estimation  $\mathbf{u}$  les valeurs de la distribution de probabilité a priori pour chaque seuil  $z_k$ , s'écrit alors :

$$\text{Prob}\{Z(\mathbf{u}) \leq z_k \mid (n)\}^* = \lambda_0 F(z_k) + \sum_{\alpha=1}^{n_1+n_2} \lambda_\alpha i(\mathbf{u}_\alpha; z_k) + \sum_{\beta=1}^{n_3} \nu_\beta y(\mathbf{u}'_\beta; z_k)$$

où,

- $F(z_k)$  = distribution globale a priori,
- $i(\mathbf{u}_\alpha; z_k)$  = distributions a priori générées à partir des  $n_1 + n_2$  données dures situées aux points de mesure  $\mathbf{u}_\alpha$ ,
- $y(\mathbf{u}'_\beta; z_k)$  = distributions a priori générées à partir des  $n_3$  données douces situées aux emplacements  $\mathbf{u}'_\beta$ .

Les pondérateurs  $\lambda_0$ ,  $\lambda_\alpha$  et  $\nu_\beta$  proviennent de la résolution d'un système d'équations linéaires semblable à ceux présentés dans la partie consacrée au cokrigage traditionnel (voir §2.2.4).

Il est à noter que, dans sa forme la plus courante, le schéma bayésien suppose que les fonctions aléatoires  $I(\mathbf{u}; z)$  et  $Y(\mathbf{u}; z)$  aient la même espérance (moyenne) qui est donnée par la fonction de distribution globale  $F(z)$  :

$$E\{I(\mathbf{u}; z)\} = E\{Y(\mathbf{u}; z)\} = F(z) \text{ connue.}$$

L'estimateur précédent est alors un estimateur par cokrigage simple (voir §2.2.4.1) à la condition d'attribuer à  $\lambda_0$  la valeur suivante qui assure le non biais de l'estimation :

$$\lambda_0 = 1 - \sum_{\alpha=1}^{n_1+n_2} \lambda_\alpha + \sum_{\beta=1}^{n_3} \nu_\beta .$$

Comme pour tout estimateur par cokrigage, la régression précédente requiert, pour chaque seuil  $z_k$ , la spécification d'au moins trois modèles de covariances (cas d'une unique variable secondaire) : un pour chaque fonction  $I(\mathbf{u}; z_k)$  et  $Y(\mathbf{u}; z_k)$  et un pour leur covariance croisée. Une hypothèse de type markovien, similaire à celle introduite pour le cokrigage co-localisé, réduit le calcul des covariance à la modélisation de celles correspondant aux données dures. L'auto-covariance de l'information douce et la covariance croisée entre données de type différent s'écrivent comme une fonction de la covariance de l'information dure et de quelques paramètres déduits de l'ensemble des données de calibrage (voir plus haut le cas d'une information locale de type secondaire). Cette hypothèse conduit aux relations suivantes :

$$C_{IY}(\mathbf{h}; z_k) = B(z_k) \cdot C_I(\mathbf{h}; z_k), \forall \mathbf{h}$$

$$C_Y(\mathbf{h}; z_k) = B^2(z_k) \cdot C_I(\mathbf{h}; z_k), \forall \mathbf{h} > 0$$

$$C_Y(0; z_k) = V_c^2(z_k) + V_f^2(z_k)$$

avec :

$$C_I(\mathbf{h}; z_k) = Cov\{I(\mathbf{u}; z_k), I(\mathbf{u} + \mathbf{h}; z_k)\}$$

$$m^{(1)}(z_k) = E\{Y(\mathbf{u}; z_k) | I(\mathbf{u}; z_k) = 1\} \in [0, 1]$$

$$m^{(0)}(z_k) = E\{Y(\mathbf{u}; z_k) | I(\mathbf{u}; z_k) = 0\} \in [0, 1]$$

$$B(z_k) = m^{(1)}(z_k) - m^{(0)}(z_k) \in [-1, +1]$$

$$\sigma^{2(1)}(z_k) = Var\{Y(\mathbf{u}; z_k) | I(\mathbf{u}; z_k) = 1\}$$

$$\sigma^{2(0)}(z_k) = Var\{Y(\mathbf{u}; z_k) | I(\mathbf{u}; z_k) = 0\}$$

$$F(z_k) = prob\{Z(\mathbf{u}) \leq z_k\} = E\{I(\mathbf{u}; z_k)\}$$

$$V_c^2(z_k) = F(z_k)[1 - F(z_k)]B^2(z_k)$$

$$V_f^2(z_k) = F(z_k)\sigma^{2(1)}(z_k) + [1 - F(z_k)]\sigma^{2(0)}(z_k)$$

où,

- $C_I$ ,  $C_Y$  et  $C_{IY}$  = auto-covariances et covariances croisées,
- $\mathbf{h}$  = vecteur de séparation entre points considérés,
- $z_k$  = valeur (seuil) de la variable principale  $Z$ .

Les autres termes sont des paramètres tirés des données de calibrage. Le lecteur intéressé trouvera les démonstrations complètes de toutes ces expressions dans Zhu (1991). L'annexe 3 contient une publication sur le sujet.

Si le modèle de Markov-Bayes demande un même effort de calcul que celui requis par le krigeage d'indicateur, il présente sur ce dernier deux avantages certains. D'une part, il distingue les auto-covariances des données dures et douces. D'autre part, à travers le cokrigeage d'indicateur, il est possible d'améliorer les distributions locales a priori calculées à partir de l'information locale (donnée d'intervalle ou douce), en fonction des données situées dans le voisinage du point d'estimation.

### **2.2.8 Estimation par champ de probabilités**

Comme son nom l'indique, l'estimation par champ de probabilité (Bourgault et Journel, 1995) est fondée sur le concept de champ de probabilités (Srivastava, 1992, Froidevaux, 1992). Ce concept est présenté en détail dans la partie consacrée aux techniques de représentation stochastique. En résumé, la simulation par champs de probabilités génère, en chaque nœud de la grille à simuler, des valeurs de probabilité qui sont utilisées pour échantillonner les distributions locales de probabilité conditionnelle de la variable principale. Dans cette approche, le calcul des probabilités est découplé du calcul des distributions locales (a posteriori) de probabilité. Toute technique de simulation non conditionnelle (par ex. la simulation gaussienne séquentielle) peut être employée pour produire le champ de probabilité qui doit avoir comme propriété de reproduire la même covariance (corrélation spatiale) que celle tirée des données primaires après transformation uniforme. Quant aux distributions locales de probabilité, elles se calculent par ailleurs à l'aide de techniques traditionnelles d'estimation (par ex. krigeage d'indicateur).

Appliquée à de l'estimation, l'approche consiste à utiliser comme champ de probabilités la transformation uniforme d'une variable secondaire connue exhaustivement, plutôt que de générer par une méthode stochastique non conditionnelle un ou plusieurs champs de probabilités. Comme le champ de la variable secondaire est unique, l'approche produit un unique champ estimé de la variable primaire. La principale caractéristique de cette méthode d'estimation est d'éliminer l'effet de lissage propre aux techniques traditionnelles de krigeage ou de cokrigeage (et plus généralement aux techniques d'interpolation) pour générer un champ estimé qui présente une forme de continuité spatiale proche de celle de la variable secondaire utilisée.

Cette méthode suppose de connaître en tout point d'estimation la valeur de la variable secondaire. Si cette condition n'est pas vérifiée, une solution consiste à kriger les données secondaires et à utiliser le champ ainsi interpolé comme champ de la variable secondaire. Le résultat obtenu dépend alors fortement du choix de la méthode d'interpolation de la variable secondaire.

## 2.3 Estimation : exemple d'application

### 2.3.1 Introduction

Pour évaluer le comportement et les performances des techniques d'estimation décrites dans les paragraphes précédents, un exemple synthétique a été utilisé. Il représente un aquifère rectangulaire, de 26 par 30 kilomètres, échantillonné par 78000 points correspondant à une grille régulière de  $260 \times 300$  nœuds.

Les données d'origine proviennent d'élévations digitalisées d'une région du Nevada (Etats-Unis) appelée Walker Lake (Isaaks et Srivastava, 1989, Gomez-Hernandez, 1992). A ce titre, il peut être admis que ces données révèlent une forme de continuité spatiale susceptible d'être rencontrée dans la nature.

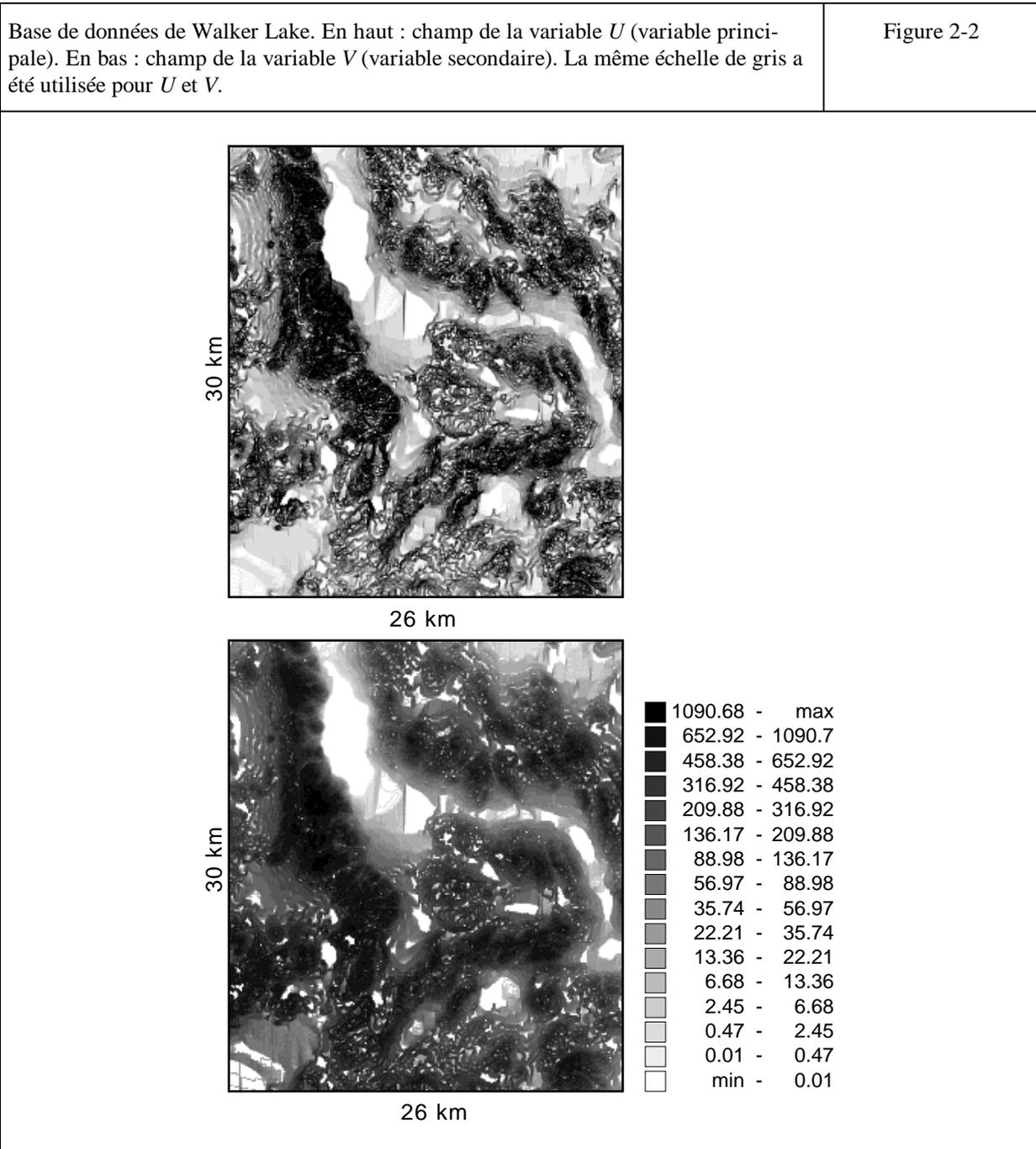
Dans leur forme actuelle, les données de Walker Lake comprennent deux variables  $U$  et  $V$  qui sont connues exhaustivement à l'endroit des 78000 points d'échantillonnage. Ces données exhaustives déterminent la solution de référence concernant l'aquifère. Elles fournissent également l'information partielle, décrite plus bas, nécessaire à cet exercice d'évaluation.

Dans cette identification à un aquifère, la variable  $U$ , variable principale, représente la perméabilité. La variable  $V$ , variable secondaire, correspond à un attribut géophysique, de type amplitude sismique, corrélé à la perméabilité. La Figure 2-2 montre graphiquement les valeurs des deux variables. La carte du haut concerne la variable  $U$  et celle du bas la variable  $V$ . La même échelle de gris a été employée pour les deux cartes afin de montrer que, malgré la grande similarité de forme de continuité spatiale des deux variables, la variable secondaire montre moins de variabilité que la variable principale, pour un coefficient de corrélation linéaire entre les deux égal à 0,65.

L'information partielle sur le champ de la variable principale consiste en 30 données régulièrement réparties sur l'aquifère comme le montre le schéma du haut de la Figure 2-3. Ces données sont assimilées à des mesures directes de perméabilité à l'endroit de sondages (mesures sur échantillons ou interprétations d'essais de puits); elles constituent les données dures de l'exercice. L'information partielle sur le champ de la variable secondaire se compose de 780 données (douces) que l'on suppose provenir de l'interprétation d'une campagne de prospection sismique 3D. L'emplacement des ces données, régulièrement réparties également, apparaît dans le schéma du bas de la Figure 2-3. L'ensemble de ces données dures et douces constitue l'information disponible de cet exercice d'évaluation de méthodes d'estimation.

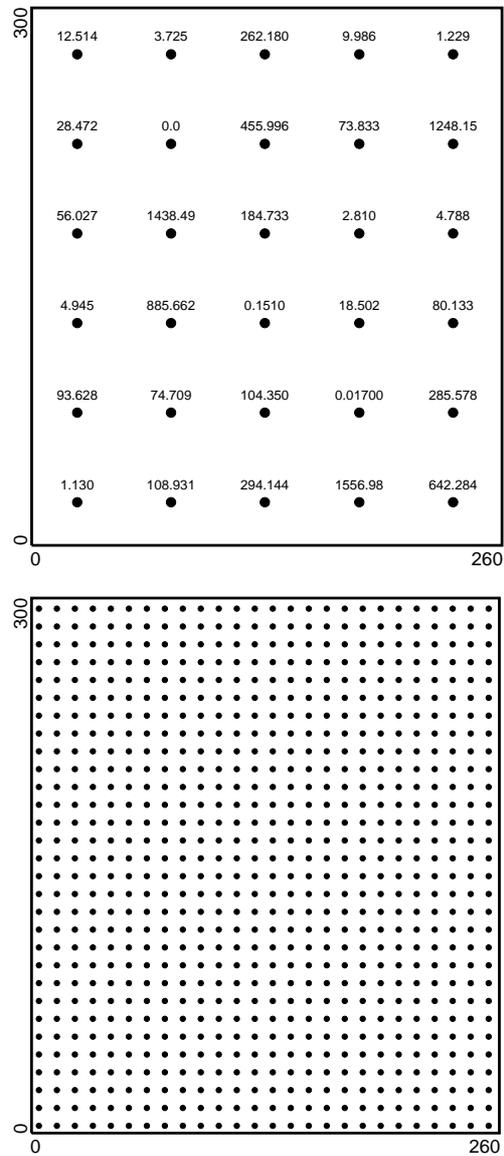
Pour l'estimation des paramètres statistiques qui quantifient la corrélation spatiale des variables  $U$  et  $V$  et la corrélation entre elles, on dispose d'un ensemble de données de calibrage constitué de 224 paires de valeurs. L'emplacement de ces données a été déterminé aléatoirement et les valeurs proviennent des champs de référence (information exhaustive) de la Figure 2-2. L'emplacement de ces paires de données est visualisé dans le schéma du haut de la Figure 2-4. Ces données de calibrage n'interviennent pas dans les estimations de la variable principale présentées dans la suite; elles sont utilisées uniquement pour obtenir des paramètres de corrélation fiables pour chaque algorithme de calcul. En effet, l'objectif n'est pas ici d'évaluer l'incidence de l'information disponible sur les résultats d'une estimation, mais de comparer entre elles des

méthodes d'estimation différentes. Cette étude comparative n'a de sens que si les paramètres d'entrée des modèles sont convenablement choisis. Certains de ces paramètres sont calibrés à partir du diagramme de dispersion constitué des 224 paires de valeurs et présenté en bas de la Figure 2-4.



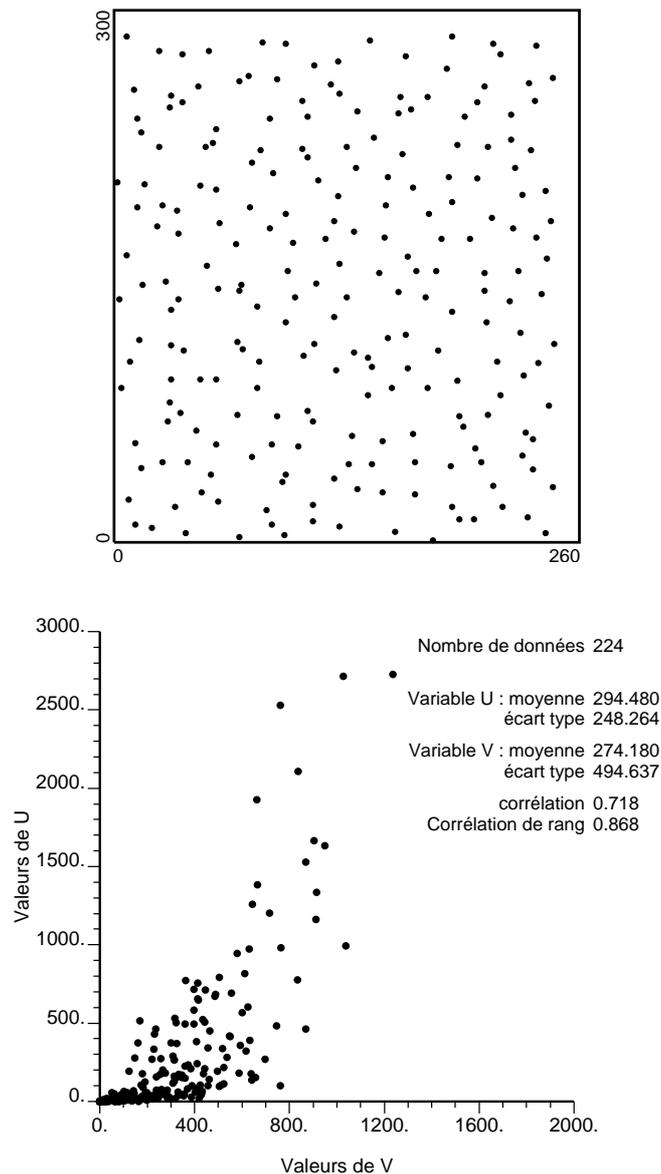
Information disponible. L'image du haut montre l'emplacement des 30 données dures régulièrement espacées. L'image du bas indique la position des 780 données douces provenant d'une campagne sismique.

Figure 2-3



Données de calibrage. L'image du haut montre l'emplacement des 224 données de calibrage et celle du bas le diagramme de dispersion correspondant. Les données de calibrage, plus nombreuses ici que les données dures, n'interviennent pas directement dans l'estimation de la variable principale. Elles servent uniquement à calculer les paramètres statistiques qui quantifient la corrélation entre les deux variables.

Figure 2-4



La Figure 2-5 présente les variogrammes obtenus à partir des données de calibrage pour les variables principale et secondaire et le variogramme croisé entre elles. Le variogramme est une fonction vectorielle qui mesure la distance structurale entre deux points en fonction de la distance euclidienne. C'est une fonction croissante qui se stabilise en général à une valeur dite de palier au-delà d'une distance appelée portée. La portée est la distance à partir de laquelle la corrélation spatiale entre deux variables disparaît; elle varie généralement avec la direction dans laquelle est mesurée la corrélation. La pente du variogramme est plus faible dans la direction de

plus grande continuité, ce qui signifie une perte plus lente de la corrélation avec la distance. A l'inverse, la pente est plus importante dans la direction de plus petite continuité le long de laquelle la corrélation disparaît plus rapidement.

L'expression du modèle de variogramme anisotrope choisi pour cet exercice d'évaluation est le suivant :

$$\gamma(\mathbf{h}) = C_0 + C_1 \cdot Sph \left( \sqrt{\left(\frac{h_x}{a_1}\right)^2 + \left(\frac{h_y}{b_1}\right)^2} \right) + C_2 \cdot Sph \left( \sqrt{\left(\frac{h_x}{a_2}\right)^2 + \left(\frac{h_y}{b_2}\right)^2} \right).$$

Ce modèle est composé de trois structures imbriquées qui sont les suivantes :

1. effet de pépite dont le palier vaut  $C_0$ ,
2. structure sphérique isotrope ( $a_1 = b_1$ ) de portée égale à 20 et de palier  $C_1$ ,
3. structure sphérique anisotrope de portée  $a_2 = 40$  dans la direction  $h_x$  de plus petite continuité (N72E), de portée  $b_2 = 100$  dans la direction  $h_y$  de plus grande continuité (N18O) et de palier  $C_2$ .

Le modèle de variogramme sphérique s'écrit :

$$Sph(|\mathbf{h}|) = \begin{cases} 1,5|\mathbf{h}| - 0,5|\mathbf{h}|^3, & \text{pour } |\mathbf{h}| < 1 \\ 1 & \text{pour } |\mathbf{h}| \geq 1 \end{cases}$$

où  $|\mathbf{h}| = \sqrt{h_x^2 + h_y^2}$ . Le Tableau 2-1 fournit les paramètres du modèle de co-régionalisation linéaire calé sur les données de calibrage des variables  $U$  et  $V$ .

Tableau 2-1 : Paramètres du modèle de co-régionalisation pour la base de données de Walker Lake.  $C_0$  est l'effet de pépite,  $C_1$  le poids de la première structure imbriquée,  $C_2$  le poids de la seconde structure,  $a_1$  la portée de la première structure qui est isotrope,  $\theta$  l'angle (mesuré dans le sens inverse des aiguilles d'une montre) de l'axe de plus petite continuité de la seconde structure avec l'horizontale (axe x),  $a_2$  et  $b_2$  les portées de corrélation de la seconde structure dans les directions de plus petite et plus grande continuité.

	$C_0$	$C_1$	$a_1$	$C_2$	$\theta$	$a_2$	$b_2$
$U$	43400	68200	20	198400	18	40	100
$V$	9800	9100	20	51100	18	40	100
$U - V$	15400	18700	20	75900	18	40	100

Variogrammes expérimentaux et modèles calés. Par colonne, les variogrammes dans les directions N180 (direction de plus grande continuité) et N72E (direction de plus petite continuité). Par ligne, les variogrammes de U, croisés et de V.

Figure 2-5

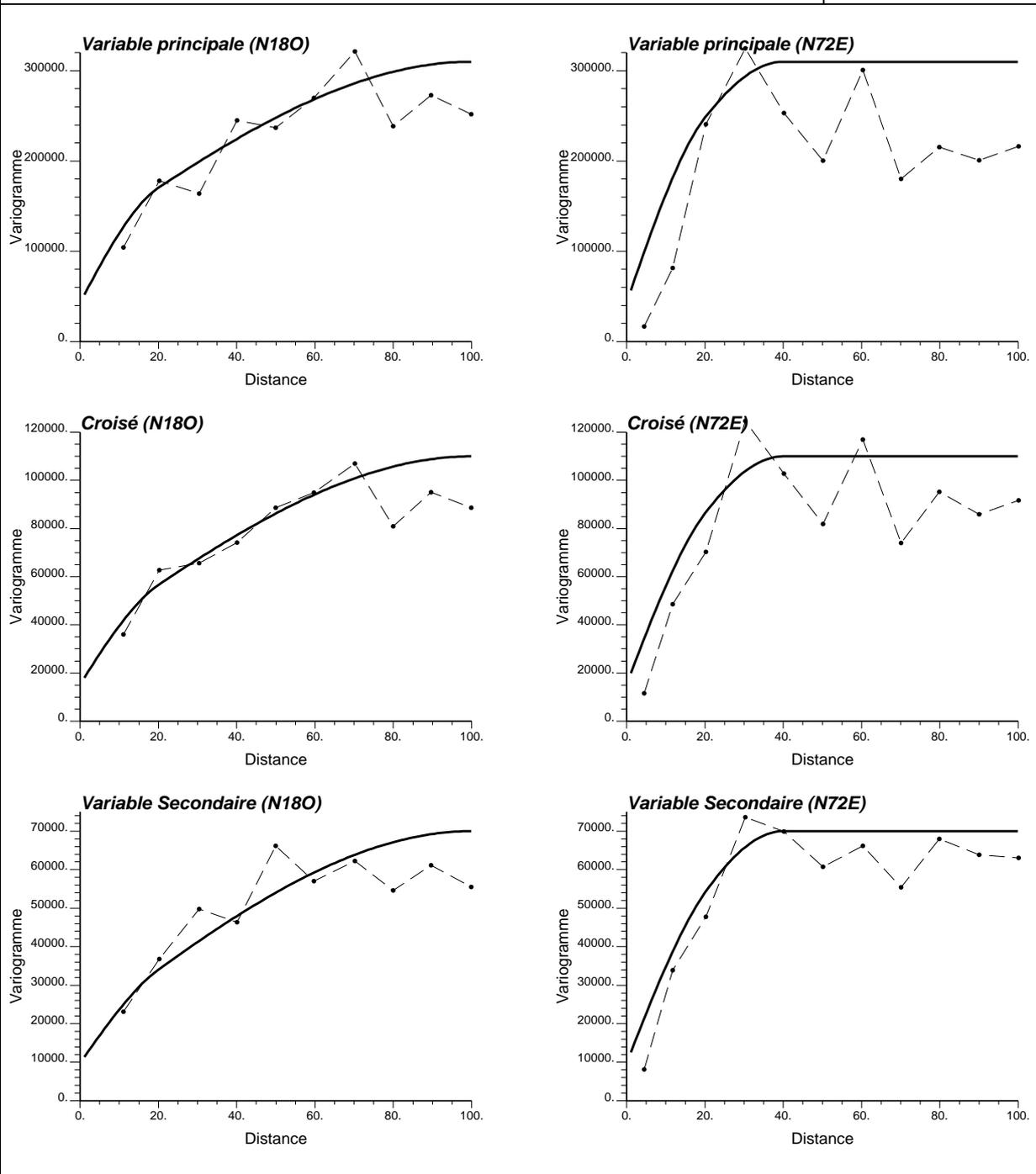


Figure 2-6, Figure 2-7 et Figure 2-8 montrent les variogrammes d'indicateur expérimentaux et normalisés (palier = 1) et les modèles ajustés pour 9 seuils choisis. Ces variogrammes sont calculés à partir des 224 valeurs de  $U$  (variable principale), provenant des données de calibrage, après transformation de ces valeurs par les fonctions indicatrices liées aux seuils. Les 9 seuils correspondent aux déciles de cet ensemble de données. Leurs valeurs sont : 0,46, 5,87, 19,02, 33,4, 61,52, 135,93, 226,03, 428,27 et 758,96.

L'expression des variogrammes d'indicateur normalisés  $\Gamma_I(\mathbf{h}; z_k)$  est la suivante :

$$\Gamma_I(\mathbf{h}; z_k) = \frac{\gamma_I(\mathbf{h}; z_k)}{C(z_k)}$$

où  $\gamma_I(\mathbf{h}; z_k)$  est le variogramme d'indicateur standard,  $C(z_k) = F(z_k)[1 - F(z_k)]$  est le palier théorique du variogramme d'indicateur (= variance de la variable indicatrice) et  $F(z_k)$  est la moyenne de la variable indicatrice associée au seuil  $z_k$ . L'expression du modèle anisotrope calé pour chaque seuil est le suivant :

$$\gamma_I(\mathbf{h}; z_k) = C(z_k) \cdot \left[ C_0(z_k) + C_1(z_k) \cdot Sph \left( \sqrt{\left( \frac{h_x}{a_1(z_k)} \right)^2 + \left( \frac{h_y}{b_1(z_k)} \right)^2} \right) + C_2(z_k) \cdot Sph \left( \sqrt{\left( \frac{h_x}{a_2(z_k)} \right)^2 + \left( \frac{h_y}{b_2(z_k)} \right)^2} \right) \right]$$

où  $C_0$ ,  $C_1$  et  $C_2$  sont les valeurs des paliers des trois structures imbriquées dans le modèle. Ces valeurs sont exprimées en proportions relatives par rapport au palier total  $F(z_k)[1 - F(z_k)]$  de sorte que :

$$C_0(z_k) + C_1(z_k) + C_2(z_k) = 1 \text{ pour chaque seuil } z_k.$$

Les trois structures imbriquées sont :

1. effet de pépite dont le palier est identique pour tous les seuils et vaut 0,1,
2. structure sphérique isotrope ( $a_1(z_k) = b_1(z_k)$ ) avec une portée qui varie de 20 à 30 m selon le seuil,
3. structure sphérique anisotrope avec une portée  $a_2(z_k)$  inférieure dans la direction  $h_x$  et une portée  $b_2(z_k)$  supérieure dans la direction  $h_y$  de plus grande continuité.

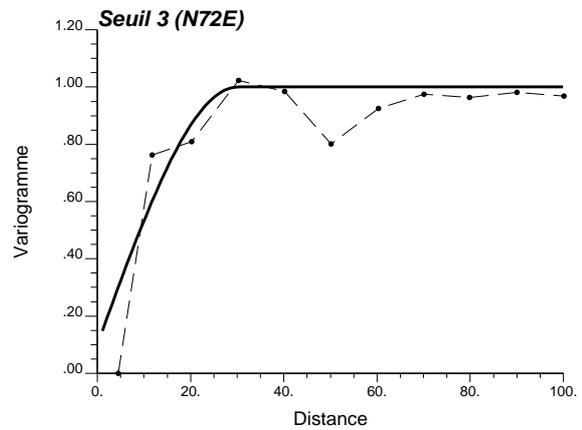
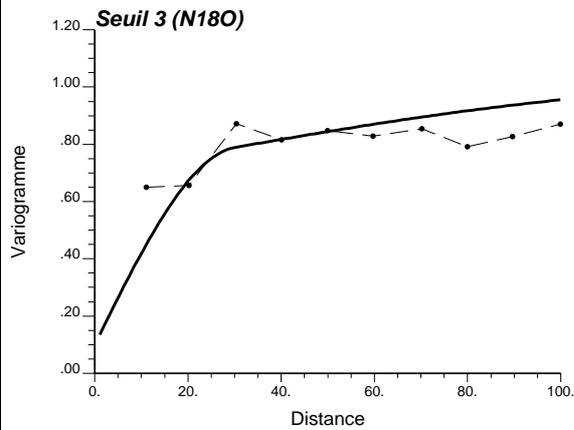
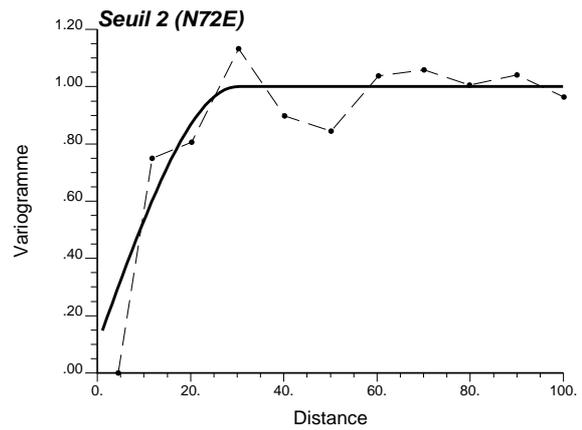
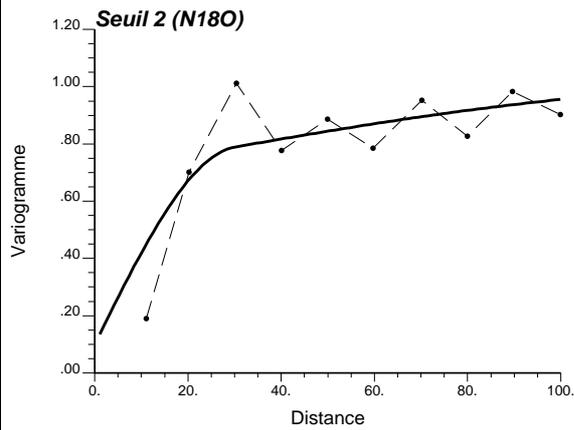
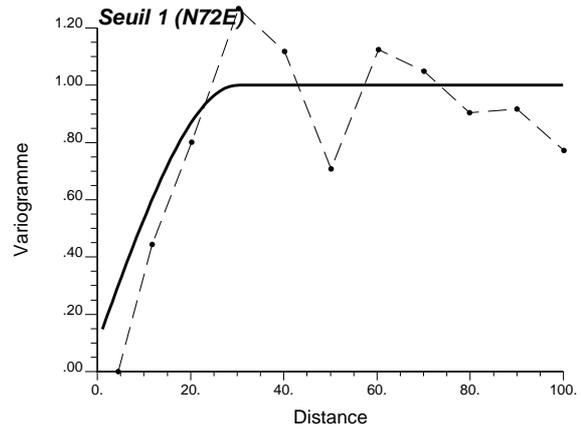
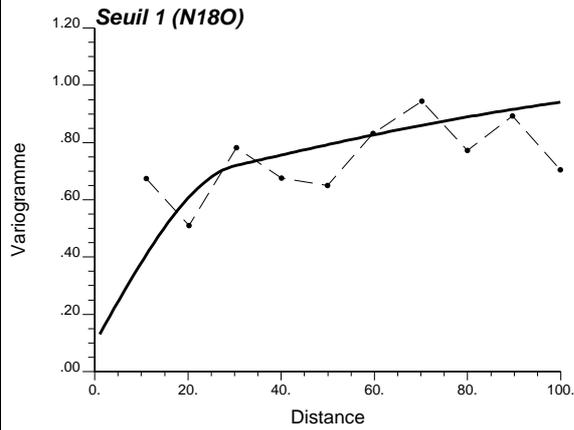
Les directions de continuité maximale et minimale et l'expression du modèle de variogramme sphérique sont les mêmes que pour le modèle de variogramme de la variable  $U$  non transformée par indicatrice. Le Tableau 2-2 regroupe les paramètres choisis pour caler les modèles de variogramme pour chaque seuil  $z_k$ .

Tableau 2-2 : Paramètres des modèles calés sur les variogrammes expérimentaux en fonction du seuil.

$k$	$z_k$	$C_0(z_k)$	$C_1(z_k)$	$a_1(z_k)$	$C_2(z_k)$	$\theta(z_k)$	$a_2(z_k)$	$b_2(z_k)$
1	0,46	0,1	0,5	30	0,4	18	30	150
2	5,87	0,1	0,6	30	0,3	18	30	150
3	19,02	0,1	0,6	30	0,3	18	30	150
4	33,40	0,1	0,7	30	0,2	18	30	120
5	61,52	0,1	0,55	20	0,35	18	20	90
6	135,93	0,1	0,55	20	0,35	18	20	80
7	226,03	0,1	0,55	20	0,35	18	20	80
8	428,27	0,1	0,65	25	0,25	18	25	80
9	758,96	0,1	0,65	30	0,25	18	30	90

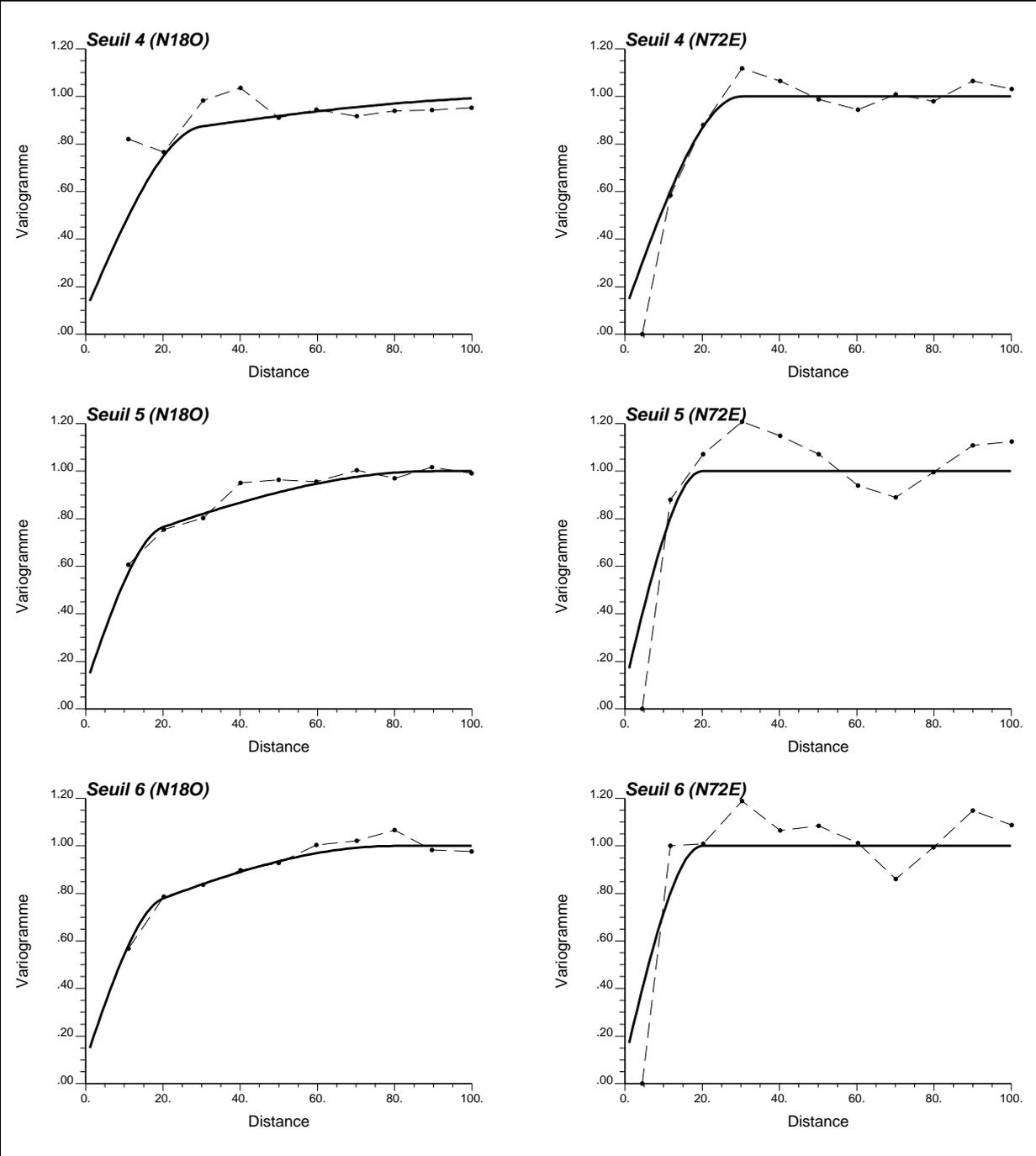
Variogrammes d'indicateur expérimentaux et les modèles calés pour les seuils  $z_1 = 0,46$ ,  $z_2 = 5,87$  et  $z_3 = 19,02$ . Les variogrammes de gauche correspondent à la direction de plus grande continuité (N180) et ceux de droite à la direction de plus petite continuité (N72E). Les modèles calés sont représentés par une ligne continue et les variogrammes expérimentaux par une ligne discontinue.

Figure 2-6



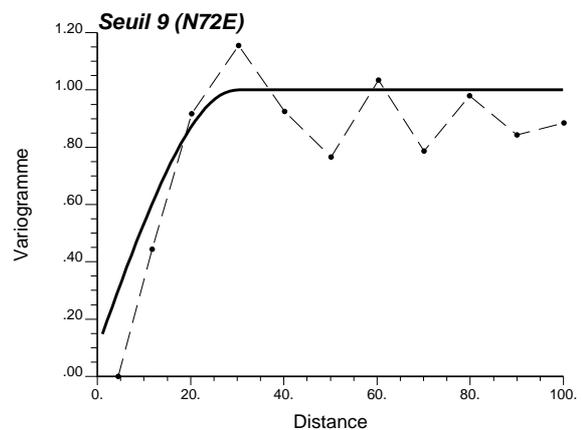
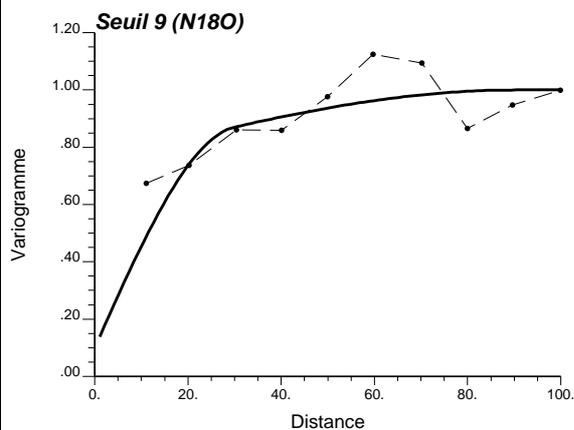
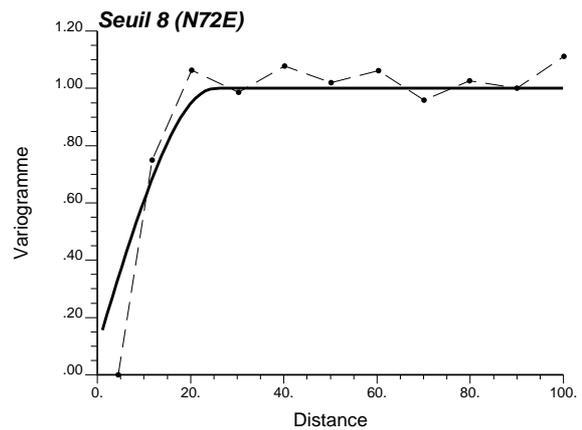
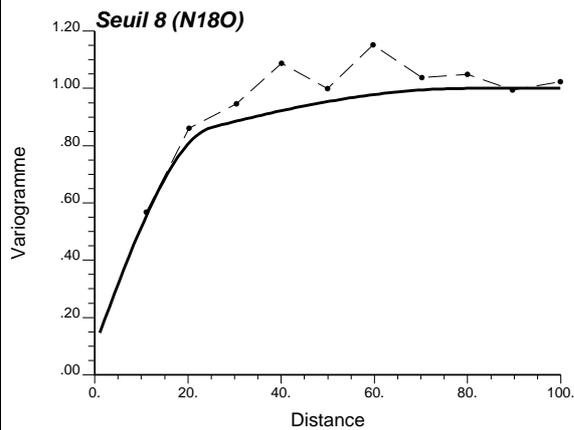
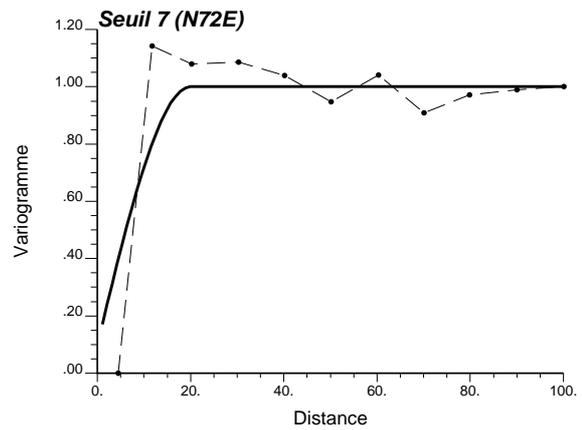
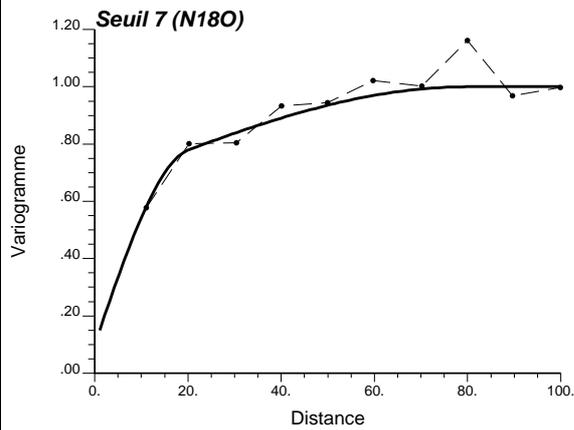
Variogrammes d'indicatrice expérimentaux et les modèles calés pour les seuils  $z_4 = 33,4$ ,  $z_5 = 61,52$  et  $z_6 = 135,93$ . Les variogrammes de gauche correspondent à la direction de plus grande continuité (N180) et ceux de droite à la direction de plus petite continuité (N72E). Les modèles calés sont représentés par une ligne continue et les variogrammes expérimentaux par une ligne discontinue.

Figure 2-7



Variogrammes d'indicateur expérimentaux et les modèles calés pour les seuils  $z_7 = 226,03$ ,  $z_8 = 428,27$  et  $z_9 = 758,96$ . Les variogrammes de gauche correspondent à la direction de plus grande continuité (N180) et ceux de droite à la direction de plus petite continuité (N72E). Les modèles calés sont représentés par une ligne continue et les variogrammes expérimentaux par une ligne discontinue.

Figure 2-8



## 2.3.2 Analyse des résultats

Le but de l'exercice est d'obtenir un champ estimé de la variable principale  $U$  en considérant toute l'information disponible à savoir les 30 données dures et les 780 données douces. Pour évaluer les champs interpolés, les tests suivants ont été réalisés :

- comparaisons purement visuelles avec le champ de référence de la Figure 2-2 (image du haut) et avec le champ interpolé uniquement à partir des données dures sans incorporer les l'information secondaire (voir plus loin Figure 2-9),
- description univariée des valeurs estimées,
- description univariée des résidus<sup>1</sup> de chaque champ estimé,
- description bivariée des valeurs estimées et de référence.

### 2.3.2.1 Analyse visuelle

#### Krigeage simple (KS)

La solution d'interpolation la plus simple est d'éliminer les données douces et d'interpoler les 30 données de la variable principale. La Figure 2-9 montre le résultat du krigeage simple des 30 données dures. Cette carte présente une résolution pauvre et un contraste faible en comparaison avec le champ de référence de la variable principale (haut de la Figure 2-2). L'interpolation ainsi produite ne tient pas compte de l'information secondaire et n'a été réalisée que dans le seul but d'être comparée aux résultats obtenus lorsque cette information est incorporée dans l'estimation.

#### Régression linéaire (RL)

Un calcul de régression linéaire réalisé sur les 224 données de calibrage donne comme droite de régression :

$$U = 1,42 V - 145,32.$$

Le calcul de la variable principale avec cette relation nécessite de connaître la valeur de  $V$  en chaque point d'estimation (c'est-à-dire en chacun des 78000 nœuds de la grille utilisée pour la solution de référence). Pour cela, les 780 données secondaires ont été interpolées par krigeage simple sur l'ensemble du domaine d'étude. Le résultat est illustré à la Figure 2-10. Le champ de  $U$ , estimé par régression linéaire, s'obtient alors par transformation des valeurs du champ secondaire de la Figure 2-10 par la droite de régression calée sur les données. Le résultat ainsi produit est montré à la Figure 2-11. Cette méthode est simple mais ne prend pas en compte la dépendance spatiale des variables. La qualité du résultat dépend du degré de corrélation li-

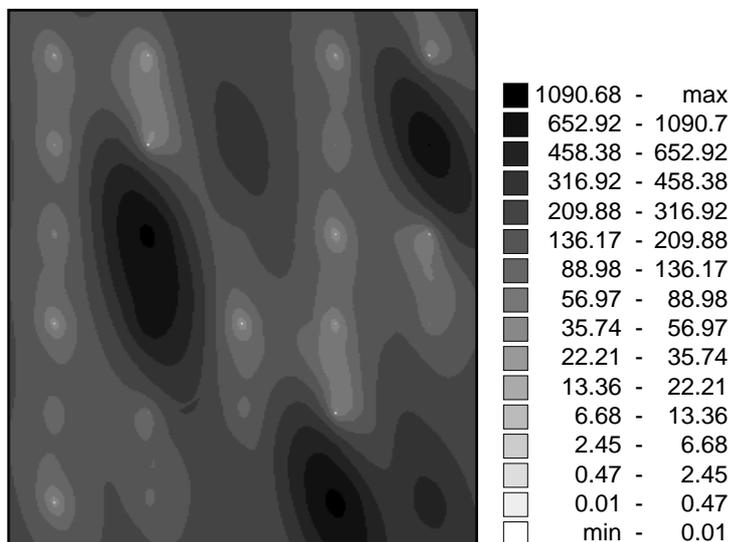
---

<sup>1</sup> En chaque point, le résidu  $r$  de l'estimation est calculé comme la différence entre la valeur estimée  $\hat{e}$  et la valeur vraie  $e$ , soit  $r = \hat{e} - e$ .

néaire entre les variables primaire et secondaire. Comme cela apparaît sur les figures, cette estimation revient à effectuer un changement d'échelle pour passer d'une variable à l'autre.

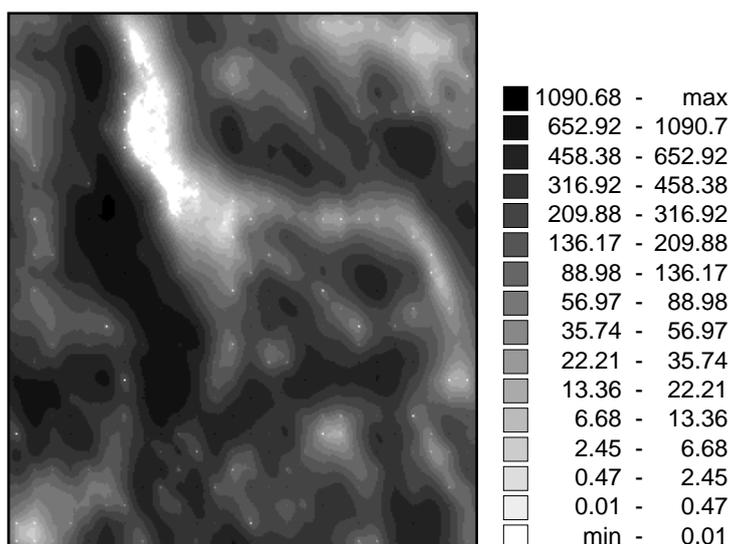
Champ interpolé de la variable U par krigéage simple avec moyenne globale tirée des 30 données dures. Dans cette estimation, l'information liée à la variable secondaire n'intervient pas.

Figure 2-9



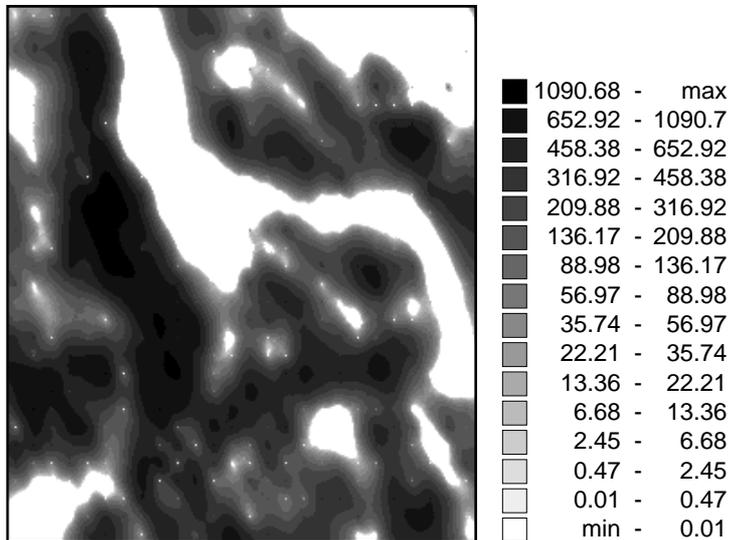
Champ de la variable secondaire interpolé par krigéage simple des 780 données disponibles. Il constitue l'information douce pour toutes les méthodes d'interpolation qui nécessitent de connaître la valeur de la variable secondaire en chaque point d'estimation de la variable principale.

Figure 2-10



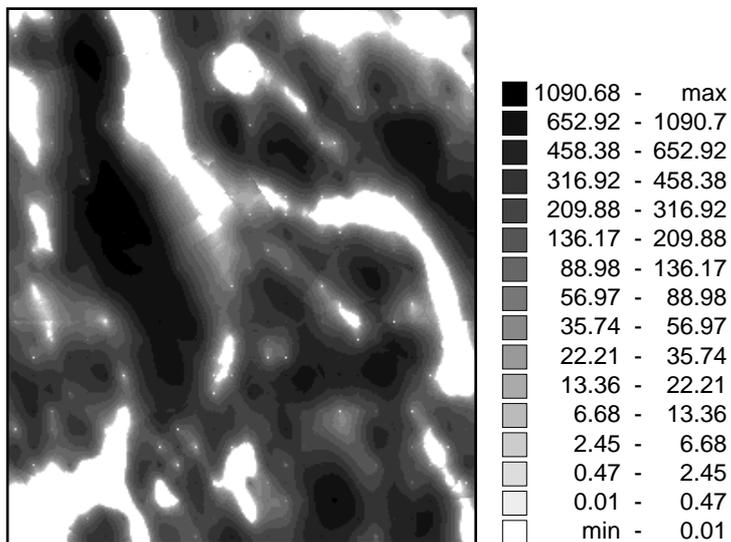
Champ de la variable principale estimé par régression linéaire.

Figure 2-11



Champ de la variable principale estimé par krigeage avec dérive externe

Figure 2-12



### **Krigeage avec dérive externe (KDE)**

Cette méthode s'applique quand il y a de bonnes raisons de croire que la variable secondaire détermine la variation (tendance) à grande échelle de la variable principale. Dans le cas présent, le champ de la Figure 2-10, qui provient de l'interpolation de l'information secondaire (douce), peut être un bon indicateur de la tendance de la variable principale. En utilisant ce champ comme dérive externe, on obtient pour  $U$  le champ estimé de la Figure 2-12. Comme pour la régression linéaire, cette méthode requiert de connaître en chaque point d'estimation la valeur de la variable secondaire. Le champ de la variable secondaire peut être interpolé, comme dans le cas de  $V$ , dans la mesure où le résultat n'est pas affecté par une trop grande rareté de l'information douce. Il est à noter que cette méthode n'exploite pas la corrélation spatiale entre les variables considérées.

### **Cokrigeage complet (CK)**

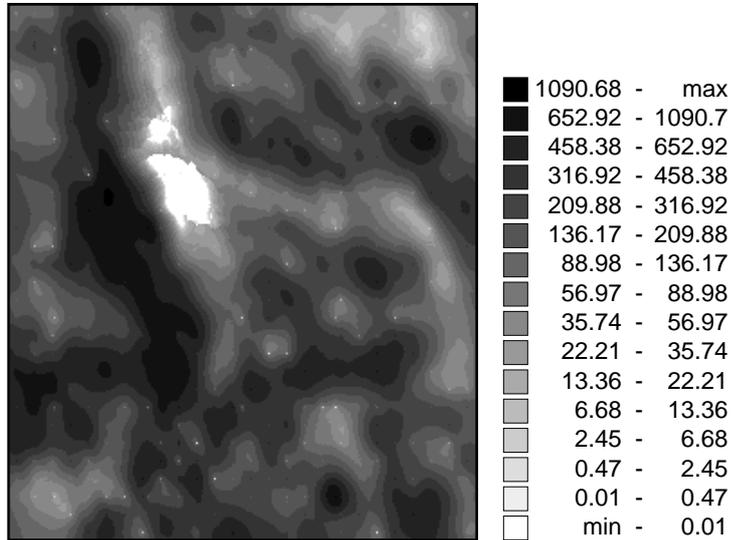
Toute l'information est traitée ici sous la forme de co-variables, ce qui signifie que la méthode d'estimation utilise l'auto-corrélation spatiale des variables principale et secondaire et la corrélation croisée de ces variables pour calculer la variable principale. Le champ estimé ainsi produit est visualisé sur la Figure 2-13. Parmi les méthodes dont les résultats ont été présentés jusqu'ici, la méthode de cokrigeage complet est celle qui incorpore le plus d'information sur la corrélation entre les variables. En contrepartie, elle nécessite un plus grand effort de modélisation de variogrammes et le temps de calcul de l'estimation peut devenir très élevé si plusieurs variables secondaires sont prises en compte simultanément.

### **Cokrigeage co-localisé (CC)**

Cette méthode incorpore l'information secondaire en tant que co-variable, c'est-à-dire qu'elle exploite à la fois l'information sur la continuité spatiale de la variable principale et celle concernant la corrélation linéaire ponctuelle (co-localisée) entre les variables principale et secondaire. Contrairement à la méthode précédente, elle suppose de connaître partout, en chaque point d'estimation, la valeur de la variable secondaire. La Figure 2-14 montre le champ estimé de  $U$  par cokrigeage co-localisé. L'intérêt premier de cette méthode réside dans son plus faible coût de calcul surtout pour traiter plusieurs variables secondaires.

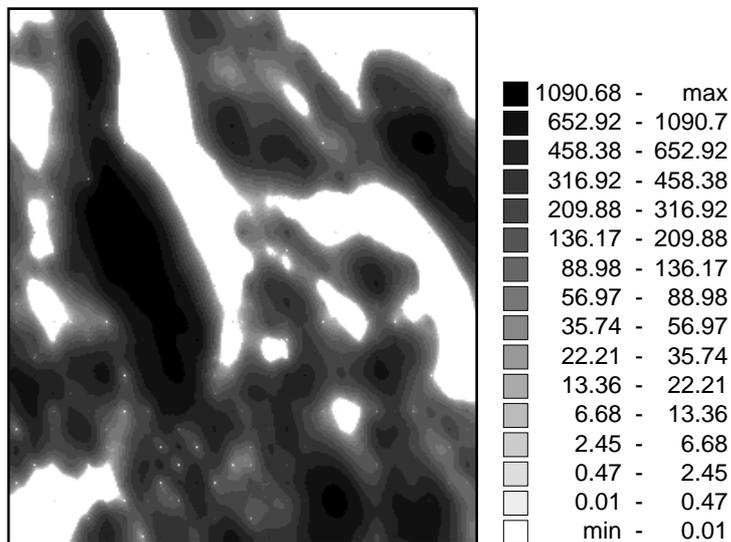
Champ de la variable principale estimé par cokrigage complet

Figure 2-13



Champ de la variable principale estimé par cokrigage co-localisé

Figure 2-14



### **Krigeage d'indicatrice (KI)**

Comme cela a été détaillé au paragraphe §2.2.6 consacré au krigeage d'indicatrice, cette méthode ne fournit pas une estimation de la valeur de la variable modélisée mais estime la valeur de la probabilité que cette variable soit inférieure à un seuil déterminé. Si l'opération est répétées pour plusieurs seuils judicieusement choisis, ces valeurs de probabilité, interpolées et extrapolées, déterminent en chaque point d'estimation la fonction de distribution locale à partir de laquelle une valeur estimée de la variable considérée peut être déduite (généralement la moyenne ou la médiane). Dans le cas présent, la valeur moyenne a été choisie comme estimateur et le résultat est présenté à la Figure 2-15.

Visuellement, l'aspect de ce champ est le pire de tous ceux vus jusqu'à présent. Pour cette estimation, les données secondaires ont été codifiées comme données d'intervalle à partir du diagramme de dispersion des données de calibrage. A cette fin, le domaine de variation de la variable secondaire a été divisé en considérant les déciles. A l'intérieur de chaque intervalle ainsi défini pour la variable secondaire, il a été déterminé les valeurs minimale et maximale entre lesquelles varie la variable principale. Pour chaque donnée secondaire (douce), on sait alors que la variable principale ne peut pas être en dessous de la borne inférieure de l'intervalle correspondant ni au-dessus de la borne supérieure, mais on ne connaît pas sa valeur exacte à l'intérieur de l'intervalle. Bien entendu, moins grande est la dispersion du nuage que forment les données de calibrage, plus étroits sont les intervalles de calibrage ce qui restreint davantage la variabilité de la variable principale et améliore la qualité de l'information secondaire incorporée. Après calibrage, les données secondaires sont transformées en vecteurs d'indicatrices généralement incomplets<sup>2</sup> et incorporées dans le processus de krigeage au même titre que les données dures. Ceci introduit une grande incertitude sur les données tirées d'une information secondaire car la valeur exacte de la variable principale à l'intérieur des limites de chaque intervalle reste inconnue. Néanmoins, le krigeage d'indicatrice se comporte mieux que les méthodes précédentes quand l'information secondaire n'est pas échantillonnée abondamment.

### **Cokrigeage d'indicatrice avec un modèle de Markov-Bayes (MB)**

Comme pour le krigeage d'indicatrice, cette méthode permet de calculer en chaque point d'estimation la distribution locale de la variable principale à partir de laquelle peut être déduite une valeur estimée. Pour cet exercice, la valeur médiane a été choisie. La Figure 2-16 montre le champ de  $U$  obtenu. En comparant cette figure à la précédente, on peut observer l'impact sur l'estimation du calcul de distributions locales à partir d'indicatrices distinctes pour les variables principale et secondaire.

### **Champ de probabilités (CP)**

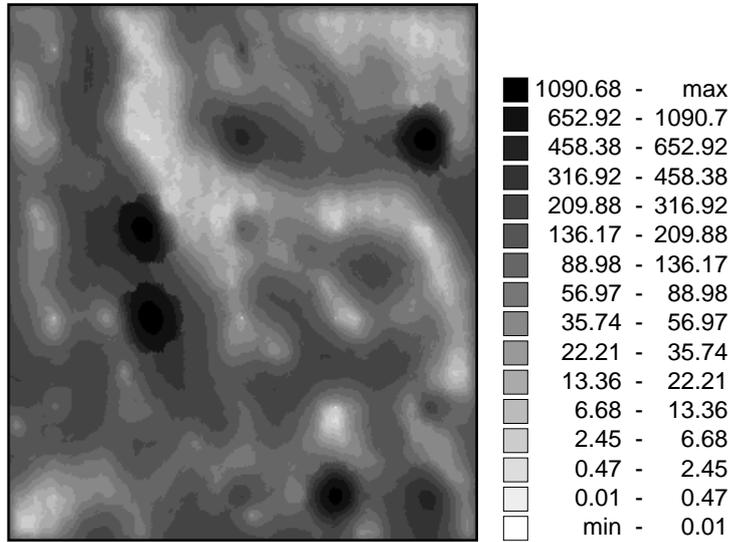
Dans la Figure 2-17, on peut voir le champ interpolé obtenu en appliquant par extension la méthode de simulation par champ de probabilités implémentée par Bourgault et Journel (1995). Si on dispose d'une information secondaire exhaustive et des distributions locales de probabilité, cet algorithme est le plus efficace en temps de calcul.

---

<sup>2</sup> Pour chaque donnée secondaire, seules sont définies les indicatrices dont le seuil est hors de l'intervalle de calibrage correspondant.

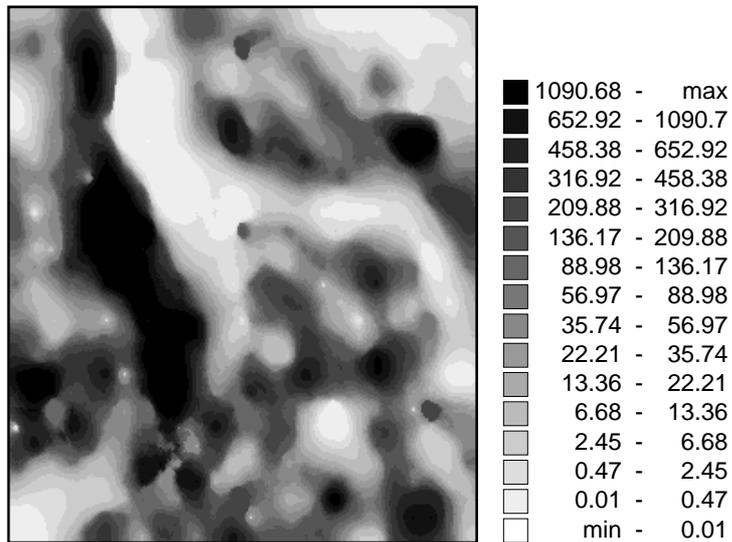
Champ de la variable principale estimé par krigeage d'indicatrice

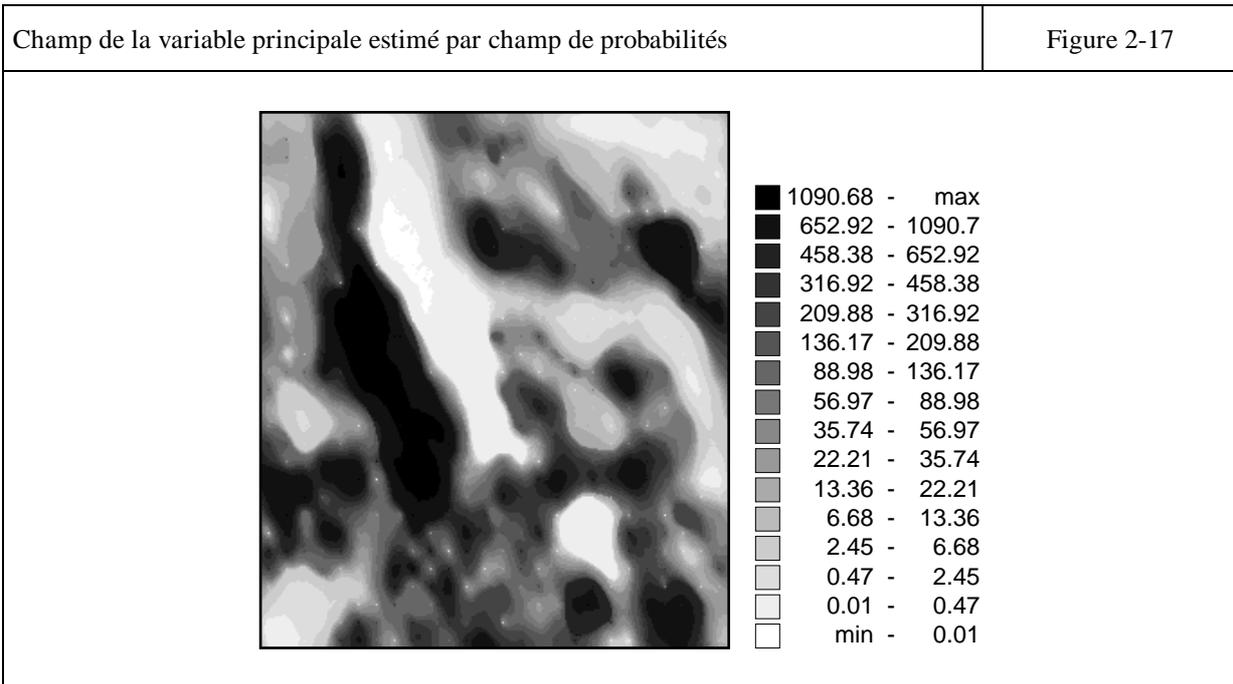
Figure 2-15



Champ de la variable principale estimé par cokrigeage d'indicatrice selon un modèle de Markov-Bayes

Figure 2-16





### 2.3.2.2 Analyse univariée des valeurs estimées

Les distributions univariées des valeurs estimées et vraies (de référence) sont résumées dans le Tableau 2-3. Les déciles sont regroupés dans le Tableau 2-4. Ces résultats suscitent les commentaires suivants.

1. La moyenne est bien reproduite par toutes les méthodes à l'exception de celle par krigeage d'indicatrice.
2. La variance des champs estimés est plus petite que celle de la solution de référence pour toutes les méthodes excepté pour celles par cokrigeage d'indicatrice et par champ de probabilités. Il est à noter qu'une diminution de la variabilité de valeurs estimées est typique des méthodes d'interpolation qui tendent à lisser les valeurs estimées en combinant des données de valeurs diverses pour construire les estimateurs.
3. Le coefficient de variation, la médiane et les quartiles sont en revanche mieux reproduits par le krigeage d'indicatrice suivi par le champ de probabilités.
4. Ces deux méthodes présentent également comme avantage sur les autres de mieux estimer la queue inférieure de la distribution de valeurs. En particulier, dans cet exemple, les méthodes paramétriques (RL, KDE, CK, CC) tendent à sous-estimer les valeurs minimales et donnent même en plusieurs points d'estimation des valeurs négatives<sup>3</sup>. Cette meilleure re-

<sup>3</sup> Il convient de préciser que pour cet exercice on s'est autorisé de conserver les valeurs négatives. Dans un cas d'étude réel, il est indispensable de vérifier si physiquement les valeurs estimées sont cohérentes avec l'attribut modélisé. Par exemple, pour l'interpolation de variables telles que la conductivité hydraulique, la porosité ou les précipitations, il est évident qu'une valeur négative manque de sens physique et suppose de remettre en

présentation de la queue de la distribution s'observe également à partir du Tableau 2-4 où les 4 premiers déciles des champs estimés par krigeage d'indicatrice et par champ de probabilités sont nettement plus proches des valeurs de références que ceux obtenus avec les autres méthodes.

5. Concernant le coefficient de corrélation, il est presque identique pour toutes les méthodes qui incorporent l'information secondaire.
6. Si on s'intéresse plus particulièrement aux résultats du Tableau 2-4, on constate que les déciles des champs estimés diffèrent assez des valeurs de référence, et que l'écart est accentué pour les méthodes paramétriques et par krigeage d'indicatrice.

En résumé, si on se fonde sur les résultats de l'analyse univariée des valeurs estimées, le cokrigeage d'indicatrice selon un modèle de Markov-Bayes ressort comme étant la méthode qui fournit le champ estimé le plus proche de la solution de référence.

Tableau 2-3 : Comparaison des distributions de valeurs vraies et estimées de  $U$  par les 8 méthodes utilisées

	Réf.	KS	RL	KDE	CK	CC	KI	MB	CP
$n$	780	780	780	780	780	780	780	780	780
$m$	258	265	258	271	277	267	154	251	281
$\sigma$	427	176	357	341	238	384	169	437	446
$CV$	1.66	0.67	1.38	1.26	0.86	1.44	1.10	1.75	1.59
$min$	0	46	-145	-284	-36	-256	1.05	0.06	0
$Q_1$	7.3	154	-45	0.23	76	-34	51	6.39	2.51
$M$	62	208	166	197	218	172	121	49	52
$Q_3$	341	308	489	462	434	483	206	288	457
$max$	3177	1201	1739	1939	1289	1898	1467	1893	3000
$\rho$	1.00	0.30	0.65	0.71	0.65	0.64	0.47	0.57	0.61

Tableau 2-4 : Déciles pour le champ de référence de  $U$  et ceux estimés par les différentes méthodes

$q$	Réf.	KS	RL	KDE	CK	CC	KI	MB	CP
1	0.22	120.01	-130.75	-72.97	25.45	-135.53	16.87	0.92	0.17
2	3.76	143.98	-84.23	-20.08	58.67	-75.85	39.25	3.11	1.04
3	11.23	162.21	-10.57	26.12	99.08	14.97	59.68	8.98	4.54
4	26.15	187.62	68.29	93.91	154.24	80.54	94.87	22.81	18.85
5	62.24	207.67	166.15	197.37	217.93	172.45	120.69	48.90	51.94
6	129.20	238.32	294.64	299.73	291.71	287.71	150.97	123.33	144.92
7	274.84	278.95	426.14	406.57	383.53	417.35	187.80	202.49	296.59
8	444.12	347.19	559.58	526.76	478.70	558.82	221.29	363.39	610.22
9	752.77	486.42	764.96	733.86	618.18	772.66	2789.84	776.41	819.03

question ou de corriger la méthode d'estimation employée. Il n'en serait pas de même avec le logarithme de la perméabilité ou celui de la transmissivité qui peut prendre des valeurs négatives.

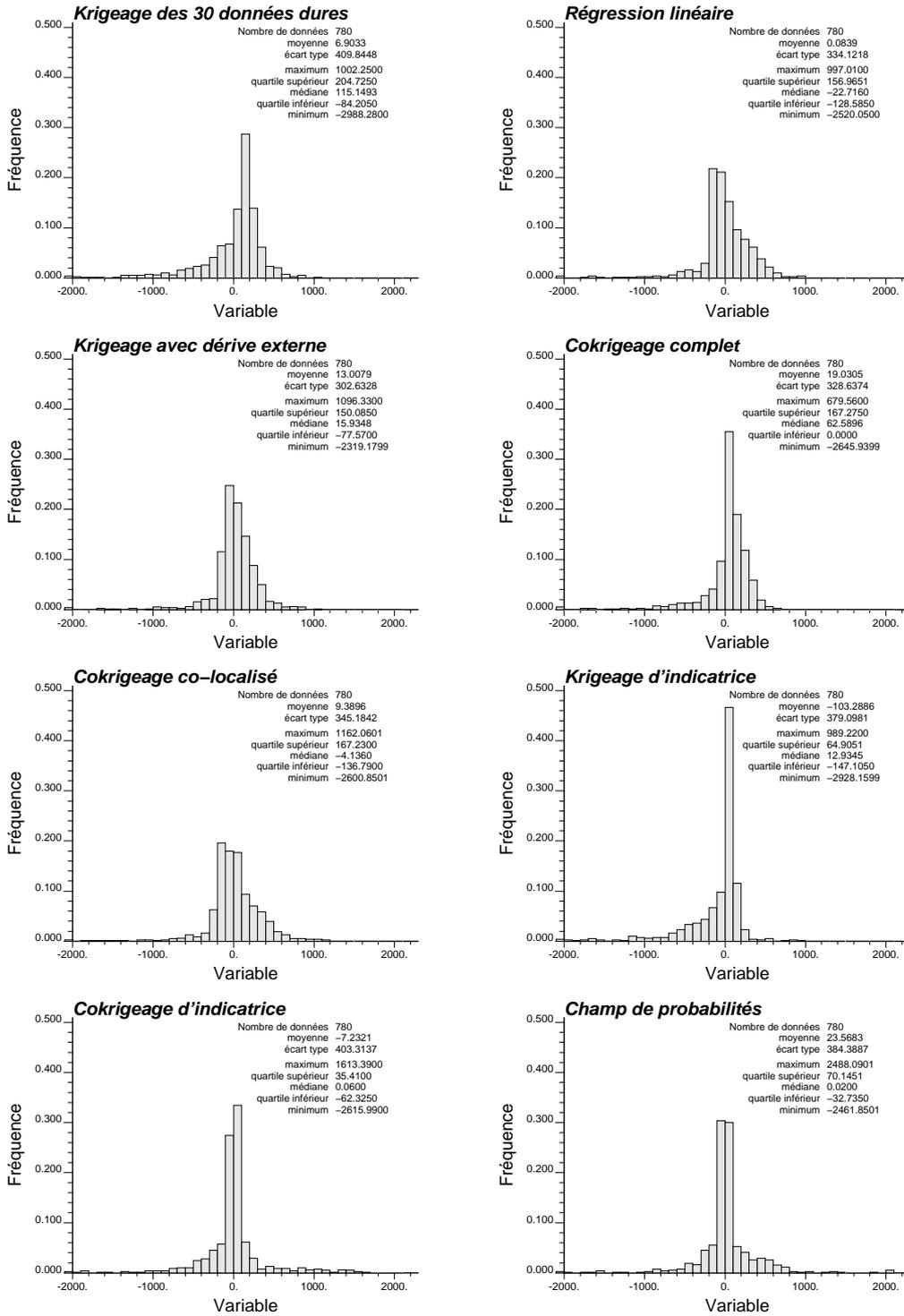
### 2.3.2.3 *Analyse univariée des résidus*

La Figure 2-18 montre graphiquement les histogrammes des résidus (valeur estimée – valeur de référence en chaque point d'estimation) dont les statistiques sont résumées au Tableau 2-5. Tels que les résidus sont définis, les distributions qui ont une moyenne négative reflètent une tendance générale à la sous-estimation, dans la mesure où l'histogramme des erreurs présente une queue pour les valeurs basses. Cet effet s'apprécie beaucoup avec le krigeage d'indicatrice. Idéalement, on souhaiterait que la moyenne des erreurs soit nulle, ce qui signifie que les surestimations compensent les sous-estimations. Si, en plus de la moyenne, la médiane est proche de zéro, la distribution tend à être symétrique et les surestimations et sous-estimations sont symétriques en magnitudes. Comme on l'observe à partir du Tableau 2-5, il existe des écarts entre les moyennes et les médianes pour toutes les méthodes d'estimation employées. De ce point de vue, le krigeage avec dérive externe et le cokrigeage d'indicatrice sont les deux méthodes qui se comportent le mieux.

Une autre caractéristique souhaitable des distributions de résidus est une dispersion des valeurs aussi petite que possible. De ce point de vue, on peut constater que toutes les méthodes d'estimation qui incorporent l'information secondaire produisent une variance des résidus inférieure à celle issue du krigeage simple des 20 données principales.

Histogrammes montrant les distributions de résidus pour chacune des méthodes d'estimation

Figure 2-18



Pour résumer conjointement le biais et la dispersion des résidus pour chaque champ estimé, l'erreur absolue moyenne (*EAM*) et l'erreur quadratique moyenne (*EQM*) ont été calculées. Elles se définissent comme suit :

$$EAM = \frac{1}{n} \sum_{i=1}^n |r_i|$$

$$EQM = \frac{1}{n} \sum_{i=1}^n r_i^2$$

où les  $r_i$  représentent les résidus et  $n$  le nombre de valeurs estimées (ou points d'estimation) pour chaque champ. Les résultats sont présentés dans le Tableau 2-5. On remarque que les erreurs absolues moyennes sont semblables pour toutes les méthodes d'estimation qui, d'une manière ou d'une autre, prennent en compte l'information secondaire. En revanche, dans le cas de l'erreur quadratique moyenne, on peut distinguer parmi ces méthodes d'estimation une différence sensible au détriment des algorithmes non paramétriques (KI, MB, CP).

Tableau 2-5 : Comparaison des distributions de résidus pour les 8 méthodes d'estimation utilisées

	KS	RL	KDE	CK	CC	KI	MB	CP
$n$	780	780	780	780	780	780	780	780
$m$	6.90	0.08	13.00	19.03	9.39	-103.29	-7.23	23.57
$\sigma$	409.84	334.12	302.63	328.64	345.18	379.10	403.31	384.39
$M$	115.15	-22.12	15.93	62.59	-4.14	12.93	0.06	0.02
<i>EAM</i>	272.06	211.32	183.80	190.822	221.44	201.48	200.28	189.73
<i>ECM</i>	168021	111637	91755	108364	119240	154384	162714	148310

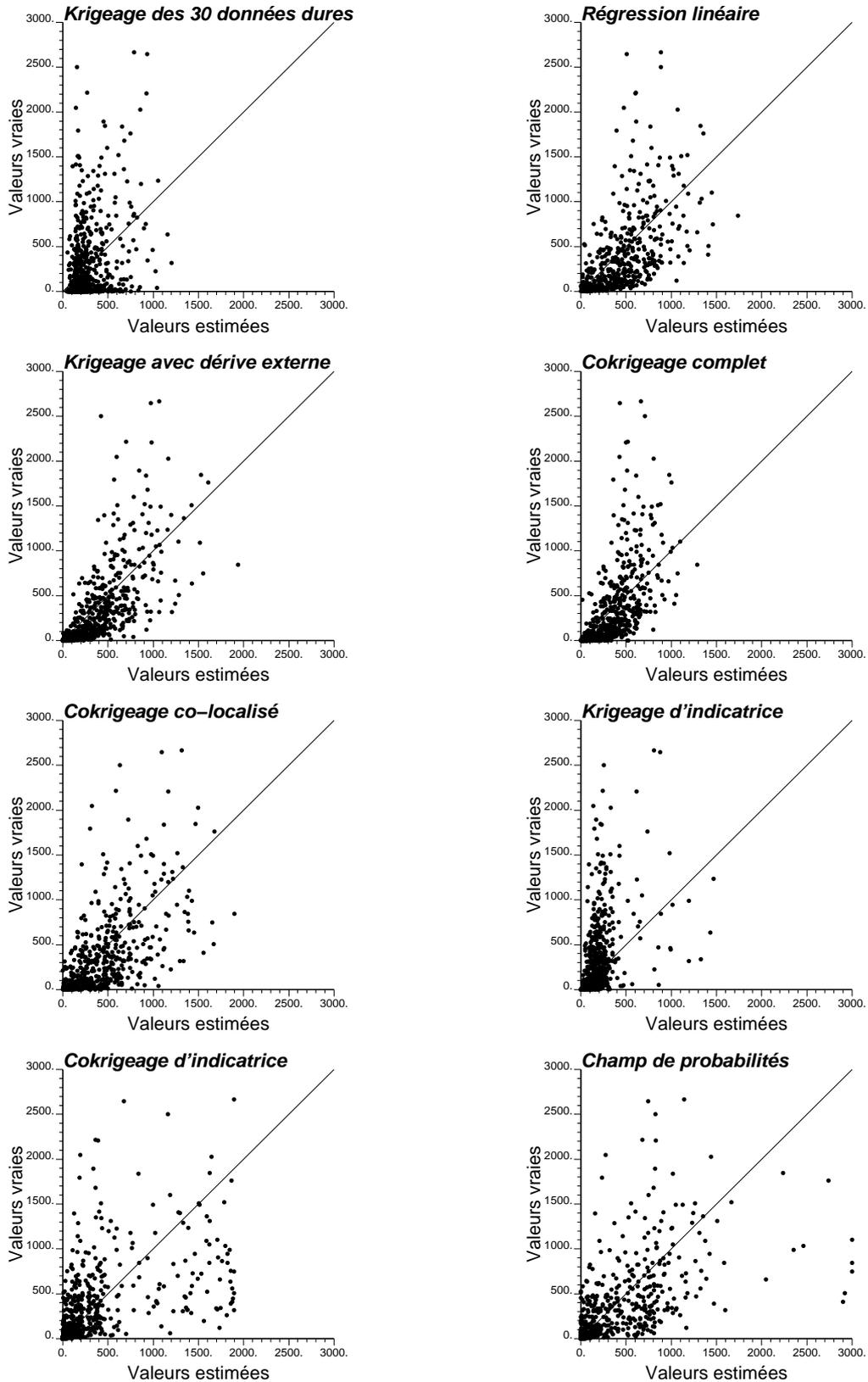
#### 2.3.2.4 Analyse bivariée des valeurs estimées et de référence

Les diagrammes de dispersion des valeurs vraies et estimées de  $U$  pour les 780 emplacements des données douces sont représentés graphiquement à la Figure 2-19. Ces diagrammes fournissent une information complémentaire sur le comportement des méthodes d'estimation employées. On s'attendrait à ce que les valeurs estimées et vraies définissent un nuage de points répartis le long d'une droite linéaire (droite à 45° si les deux axes ont même unité).

En réalité, la première observation est qu'il existe une grande dispersion des points pour toutes les méthodes. A l'exception du cokrigeage d'indicatrice et du champ de probabilité, on constate également la tendance générale à une sous-estimation de  $U$ , sous-estimation qui se caractérise ici par un éloignement des points, par rapport à la droite à 45°, en direction de l'axe des ordonnées. On remarque aussi sur ces diagrammes que le krigage d'indicatrice ne se comporte pas aussi bien que les autres méthodes.

Diagrammes de dispersion des valeurs vraies et estimées de  $U$

Figure 2-19



### 2.3.3 Evaluation finale

Conformément à notre expérience et à l'analyse précédente, il est possible d'émettre les conclusions suivantes.

- La prise en compte de l'information secondaire améliore considérablement le contraste des champs estimés. Pour s'en rendre compte, il suffit de comparer le champ de la Figure 2-9, obtenu avec les 30 données dures uniquement, aux autres champs estimés qui intègrent l'information secondaire.
- Toutes les méthodes, à l'exception du cokrigage d'indicateur et de l'estimation par champ de probabilités, produisent des champs où la variabilité des valeurs estimées est moindre que celle des valeurs vraies. Cet effet de lissage est commun à toutes les méthodes où l'estimateur se construit directement comme une combinaison (fonction) de données.
- La méthode fondée sur les champs de probabilités fonctionne très bien. La qualité du champ estimé dépend cependant de l'existence ou non d'une information secondaire exhaustive. Si cette information est réduite, ce qui suppose que dans une première étape elle soit interpolée, la qualité du résultat peut être affectée.
- La méthode de krigeage d'indicateur produit un champ estimé de qualité moyenne. Ceci est dû au calibrage sommaire des données où les données douces sont traduites sous forme de données d'intervalle. Le champ ainsi obtenu reste cependant meilleur que celui estimé sans intégrer l'information secondaire.
- D'un point de vue performance de calcul, on peut dire que, si on dispose d'une information secondaire exhaustive, le cokrigage co-localisé est la méthode la plus recommandée.
- Si, en plus d'une information secondaire exhaustive, on dispose en chaque point d'estimation de la fonction de distribution locale de la variable modélisée, la méthode d'estimation par champ de probabilités, selon la méthodologie implémentée par Bourgault et Journel (1995), donne de bons résultats.
- Bien que le choix d'une méthode d'estimation dépende du type de problème à modéliser, de la quantité d'information disponible et des caractéristiques statistiques des variables en jeu, on peut véritablement conclure ici que le cokrigage d'indicateur suivant un modèle de Markov-Bayes est la méthode qui produit le meilleur champ estimé. Il est utile de noter que cette méthode a été conçue à l'origine comme technique de simulation stochastique et adaptée ensuite comme méthode d'estimation dans le cadre de cette étude comparative. La théorie et la pratique de la simulation de champs aléatoires par cokrigage d'indicateur sont commentées dans les parties de ce rapport consacrées aux techniques de représentation stochastique.

## 2.4 Algorithmes de représentation stochastique

### 2.4.1 Introduction

Les chapitres précédents ont montré que les méthodes d'estimation fournissent comme résultat une représentation unique de l'attribut interpolé. Ceci signifie que la distribution spatiale, inconnue, de l'attribut  $z$  considéré est représentée par un ensemble unique de valeurs estimées  $\{z^*(\mathbf{u}), \mathbf{u} \in D\}$ , où  $D$  représente le domaine d'étude.

Les méthodes d'interpolation permettent d'obtenir des valeurs estimées en fonction de l'information directe située au voisinage de chaque point d'estimation et, dans certain cas, d'une information secondaire. Sous certaines conditions, ces méthodes fournissent également une mesure de l'incertitude locale. Selon la méthode employée, l'estimation tend cependant à produire des champs lissés où la variabilité spatiale de l'attribut modélisé est atténuée. En général, sans l'apport d'une information secondaire, les valeurs faibles sont surestimées et les valeurs fortes sont sous-estimées. En outre, plus la zone à estimer se trouve loin des données, plus l'effet de lissage est prononcé.

Dans les applications qui sont sensibles aux valeurs extrêmes (fortes ou faibles) et à leur continuité spatiale, l'utilisation de techniques d'interpolation doit absolument être évitée. Ceci concerne les phénomènes de migration de contaminants ou de radionucléides en milieu poreux pour lesquels les zones de fortes et faibles valeurs de conductivité hydraulique (chemins préférentiels et barrières aux écoulements) contrôlent véritablement le transport de masse. Comme solution à ces problèmes, il est nécessaire de recourir à des techniques qui génèrent des champs où le niveau de variabilité spatiale de l'attribut modélisé est identique à celui observé dans la réalité. Ces techniques se regroupent sous le concept de *simulation ou représentation stochastique*.

On appelle ici simulation stochastique le processus de génération de champs multiples d'une variable  $z$  modélisée comme une fonction aléatoire  $Z(\mathbf{u})$ . Ces champs simulés honorent toute l'information disponible et reproduisent le modèle de continuité spatiale utilisé pour décrire  $Z(\mathbf{u})$ .

Dans ce chapitre, on s'intéressera aux trois catégories de simulation suivantes :

1. simulation séquentielle,
2. simulation par recuit simulé,
3. simulation par champ de probabilités.

Les techniques décrites dans la suite ne sont pas forcément les seules à exister dans leur catégorie. Toutefois, de par notre expérience, ces techniques peuvent être considérées comme celles qui sont les plus utilisées en pratique actuellement pour la génération de réalisations multiples d'une ou plusieurs variables aléatoires.

Notre attention portera plus particulièrement sur l'approche par simulation séquentielle qui, dès son origine, est apparue comme une alternative à des techniques plus traditionnelles comme celle des bandes tournantes (Journel et Huijbregts, 1978). Dowd (1992) a comparé la technique des bandes tournantes à la simulation séquentielle en manifestant une certaine méfiance à l'égard de cette dernière. Notons cependant que son étude fut simplement descriptive et qu'une évaluation plus rigoureuse aurait dû être réalisée. L'annexe 4 présente un travail qui, dans un certain sens, prétend compléter celui de P. Dowd et montre que la simulation séquentielle est une approche robuste et fiable pour la représentation stochastique de variables soumises à une répartition spatiale. Ce chapitre s'achève avec une présentation des fondements théoriques des méthodes de simulation par recuit simulé et par champ de probabilités, et par une brève évocation des méthodes booléennes.

## 2.4.2 Simulation séquentielle

L'algorithme de simulation séquentielle peut se résumer comme suit.

1. Définir un chemin aléatoire qui visite tous les nœuds (points) à simuler.
2. Pour chaque nœud à simuler, déterminer la fonction (cumulative) de distribution, conditionnelle par rapport à l'ensemble des données disponibles de tout type (données principales et secondaires) et par rapport aux valeurs qui ont été simulées précédemment en d'autres nœuds. Au cours du processus de simulation, le nombre de valeurs déjà simulées, à prendre en compte pour le calcul de la fonction de distribution, augmente, ce qui entraîne également une augmentation du temps de calcul. Pour cette raison, la fonction de distribution est prise conditionnelle par rapport uniquement aux données et valeurs déjà simulées les plus proches du nœud considéré.
3. Tirer aléatoirement, dans la fonction de distribution calculée à l'étape précédente, une valeur de la variable et l'affecter au nœud simulé; cette valeur s'ajoute à l'ensemble des données disponibles pour simuler les nœuds suivants.
4. Retourner à l'étape 2 jusqu'à ce que tous les nœuds aient été simulés.
5. Répéter ce processus à partir d'un autre chemin aléatoire pour générer une autre réalisation.

La détermination des fonctions de distribution conditionnelles peut se faire indifféremment à l'aide d'une approche paramétrique ou non paramétrique. Dans le premier cas, la fonction de distribution conditionnelle correspond à un modèle connu dont il faut calculer les paramètres en chaque nœud simulé. Par exemple, si la fonction de distribution suit un modèle multigaussien, les fonctions de distributions conditionnelles sont gaussiennes et leur calcul requiert uniquement de connaître la moyenne et la variance de l'attribut principal. Dans le second cas, la fonction de distribution se détermine à partir du calcul de quelques points de la courbe de distribution (calcul discret de la probabilité pour un certain nombre de seuils convenablement choisis). La simulation non paramétrique utilise des variables indicatrices. Dans bien des problèmes, elle se révèle beaucoup plus puissante et souple que les approches paramétriques.

### 2.4.2.1 Simulation gaussienne séquentielle

Si on adopte un modèle multigaussien pour décrire la dépendance spatiale des variables, on peut démontrer que les fonctions de distribution conditionnelles, dont il est question à l'étape 2 de l'algorithme de simulation, sont gaussiennes. Il suffit alors de connaître la moyenne et la variance conditionnelles pour les décrire complètement.

Parmi les méthodes de simulation gaussienne séquentielle, on s'intéresse dans la suite à celles qui permettent d'intégrer une information secondaire. Il s'agit de celles fondées sur le cokrigage complet et sur le cokrigage co-localisé avec un modèle de Markov.

Etant donné que les variables principale et secondaire(s) ne suivent généralement pas une distribution normale, une transformation préalable des données est souvent nécessaire pour utiliser la simulation gaussienne. Dans le cas de deux variables aléatoires, les variables transformées peuvent s'écrire :

$$\begin{aligned} Y_1(\mathbf{u}) &= \phi_1[Z_1(\mathbf{u})] \\ Y_2(\mathbf{u}) &= \phi_2[Z_2(\mathbf{u})] \end{aligned}$$

où  $Z_1$  et  $Z_2$  sont les variables principale et secondaire,  $\mathbf{u}$  le vecteur de coordonnées spatiales et  $\phi_1$  et  $\phi_2$  représentent les fonctions qui transforment  $Z_1$  et  $Z_2$  en variables  $Y_1$  et  $Y_2$  dont les distributions sont normales, réduites et centrées (moyenne nulle et variance de 1). Ces fonctions se déduisent empiriquement à partir des données (Journel et Huijbregts, 1978, p. 478). Une fois la simulation terminée, une transformation inverse des valeurs simulées fournit le résultat dans les variables d'origine.

#### **Simulation gaussienne séquentielle par cokrigage complet**

Le cokrigage complet de  $Y_1(\mathbf{u})$ , en considérant à la fois les données primaires  $y_1(\mathbf{u}_{\alpha_1})$ ,  $\alpha_1=1, \dots, n_1$ , et secondaires  $y_2(\mathbf{u}_{\alpha_2})$ ,  $\alpha_2=1, \dots, n_2$ , fournit la moyenne et la variance de la distribution conditionnelle de  $Y_1(\mathbf{u})$ . La moyenne est donnée par l'estimateur de cokrigage simple  $y_1^*(\mathbf{u})$  et la variance par la variance de cokrigage simple  $\sigma_{CKS}^2(\mathbf{u})$ , soit :

$$\begin{aligned} y_1^*(\mathbf{u}) &= \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} y_1(\mathbf{u}_{\alpha_1}) + \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2} y_2(\mathbf{u}_{\alpha_2}) \\ \sigma_{CKS}^2(\mathbf{u}) &= 1 - \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} S_1(\mathbf{u}_{\alpha_1} - \mathbf{u}) - \sum_{\alpha_2=1}^{n_2} \lambda_{\alpha_2} S_{12}(\mathbf{u}_{\alpha_2} - \mathbf{u}) \end{aligned}$$

où  $S_1(\mathbf{u}_{\alpha_1} - \mathbf{u})$  est la covariance<sup>1</sup> de la variable transformée  $Y_1(\mathbf{u})$  et  $S_{12}(\mathbf{u}_{\alpha_2} - \mathbf{u})$  est la covariance croisée entre  $Y_1(\mathbf{u})$  et  $Y_2(\mathbf{u})$ . Les  $(n_1+n_2)$  coefficients  $\lambda_{\alpha}$  s'obtiennent en résolvant le système de cokrigage simple correspondant. L'estimateur précédent nécessite de modéliser

---

<sup>1</sup> La lettre  $S$  est utilisée à la place de  $C$  pour distinguer la covariance de la variable d'origine  $Z$  de celle de la variable transformée  $Y$ .

les auto-covariances et les covariances croisées entre variables et de résoudre, pour chaque nœud simulé, un système de  $(n_1+n_2)$  équations<sup>2</sup>. L'alternative du cokrigage co-localisé suivant un modèle markovien réduit la charge de calcul.

### **Simulation gaussienne séquentielle par cokrigage co-localisé suivant un modèle markovien**

Cette méthode est identique à la précédente à la différence que seule la donnée secondaire co-localisée est prise en compte pour simuler chaque nœud. Comme cela a été décrit dans la partie consacrée à l'estimation par cokrigage co-localisé selon un modèle de Markov (§2.2.5.1), une hypothèse de type markovien réduit l'effort nécessaire de modélisation et de calcul des covariances.

La moyenne et la variance des fonctions de distribution gaussiennes conditionnelles s'écrivent comme suit :

$$y_1^*(\mathbf{u}) = \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} y_1(\mathbf{u}_{\alpha_1}) + \lambda_2 y_2(\mathbf{u})$$

$$\sigma_{CKS}^2(\mathbf{u}) = 1 - \sum_{\alpha_1=1}^{n_1} \lambda_{\alpha_1} S_1(\mathbf{u}_{\alpha_1} - \mathbf{u}) - \lambda_2 S_{12}(0).$$

Cette méthode nécessite de résoudre un système de  $(n_1+1)$  équations et de connaître la fonction de covariance de la variable principale. Les fonctions de covariance secondaire et croisée se déduisent alors par un simple changement d'échelle de la covariance principale (voir §2.2.5.1, p 63).

### **Simulation conjointe des diverses variables**

Une extension des méthodes précédentes est la co-simulation séquentielle dans laquelle les diverses variables du problème sont simulées conjointement. Les variables sont généralement corrélées entre elles<sup>3</sup> comme par exemple dans le cas de la perméabilité horizontale, de la perméabilité verticale et de la porosité d'un aquifère.

La co-simulation gaussienne séquentielle avec cokrigage complet est décrite par Gómez-Hernández et Journel (1993c) et par Verly (1993). La technique est identique à celle développée pour une variable unique. La seule différence est que, au lieu de simuler une valeur en chaque nœud, la simulation porte sur un vecteur de  $Q$  termes où chaque terme  $q$ ,  $q=1, \dots, Q$ , désigne une variable principale. En chaque nœud simulé, la simulation de chacune des variables principales est conditionnelle par rapport aux données disponibles les plus proches suivantes :

- celles de la variable principale considérée,
- celles des autres variables principales,

<sup>2</sup> En réalité, si  $n_1$  et  $n_2$  désignent les nombres maximaux de données de chaque type que l'on s'autorise à intégrer dans l'estimation de la moyenne, le système à résoudre en chaque nœud simulé peut contenir de 0 à  $(n_1+n_2)$  équations.

<sup>3</sup> Sans quoi il n'y a aucun intérêt particulier à vouloir les simuler en même temps.

- les données secondaires,
- les valeurs simulées précédemment en d'autres nœuds.

Cette abondance de données, ajoutée à l'inférence des auto-covariances et des covariances croisées entre toutes les variables, entraîne nécessairement un coût de calcul supplémentaire par rapport à la simulation séquentielle d'une variable unique.

Almeida (1993a et b) et Almeida et Journel (1994) ont développé un algorithme de co-simulation séquentielle avec une information secondaire co-localisée et suivant un modèle de Markov. Pour chaque nœud simulé, ce modèle réduit le nombre de données incorporées en considérant uniquement la valeur co-localisée pour chaque variable secondaire et pour chaque variable principale déjà simulée. Ce modèle réduit également, de manière significative, le nombre de covariances à modéliser et donc à calculer au cours du processus de simulation. En effet, les auto-covariances et les covariances croisées concernant des variables secondaires peuvent toutes s'exprimer comme une fonction d'une des covariances de variable principale.

Il est à noter que le processus de co-simulation impose une hiérarchie entre les variables principales pour déterminer l'ordre dans lequel elles doivent être simulées.

#### *2.4.2.2 Simulation séquentielle d'indicatrice*

La simulation séquentielle d'indicatrice n'adopte aucun modèle statistique a priori contrairement à la simulation gaussienne séquentielle qui se fonde sur un modèle multigaussien. La première étape de la simulation consiste à codifier toute l'information disponible, y compris les données douces, en terme de valeurs d'indicatrices. Comme cela a été décrit au paragraphe §2.2.6 consacré au krigeage d'indicatrice, cette méthode, dans ses différentes variantes, fournit un estimateur non paramétrique de la fonction (cumulative) de distribution de la variable principale en chaque nœud estimé ou simulé; cette fonction de distribution peut alors être directement employée dans l'algorithme de simulation séquentielle pour tirer aléatoirement la valeur de la variable.

Le krigeage d'indicatrice comme le cokrigeage suivant un formalisme de Markov-Bayes peuvent s'utiliser dans la simulation séquentielle. La différence d'une méthode à l'autre réside dans la manière d'intégrer l'information douce dans le système de krigeage ou de cokrigeage. Dans le cas du krigeage d'indicatrice (cf. §2.2.6), il n'est pas possible de distinguer la forme de continuité spatiale de l'information douce de celle de l'information dure. Dans le cas du cokrigeage d'indicatrice avec ou sans une hypothèse de type Markov-Bayes (cf. §2.2.7), les données douces (secondaires) rentrent dans l'expression de l'estimateur. Ceci permet non seulement de distinguer des variables principales et secondaires entre elles mais aussi d'améliorer (mettre à jour) les fonctions de distribution estimées a priori, à partir de l'information douce, en les rendant conditionnelles par rapport aux données avoisinantes.

### 2.4.3 Simulation par recuit simulé (simulated annealing)

La simulation par recuit simulé (Deutsch, 1992, Deutsch et Journel, 1991 et 1992b, Deutsch et Cockerman, 1994) est un outil capable de compléter les techniques de simulation stochastique plus traditionnelles. Son intérêt majeur réside dans sa capacité à imposer certaines formes de continuité spatiale qu'il est difficile de caractériser au moyen de fonctions de covariance. Son intérêt se trouve aussi dans la facilité, au moins théorique, que la méthode procure pour incorporer des informations qui proviennent de différentes sources. Ces avantages, qui rendent la méthode très flexible, sont possibles grâce à un algorithme simple mais très demandeur de temps de calcul informatique. Pour cette raison, le recuit simulé n'est pas recommandé pour traiter les problèmes pour lesquels il existe d'autres méthodes capables d'atteindre les mêmes objectifs.

Le recuit simulé se fonde sur une méthode conçue initialement pour résoudre des problèmes d'optimisation où il s'agit de trouver le ou les arrangements optimaux d'un système constitué d'un grand nombre de composants (ou variables). Dans le contexte de la modélisation stochastique qui nous intéresse, les composants s'identifient à la variable principale simulée en chaque nœud d'une grille qui constitue la représentation discrète du champ modélisé. Le caractère optimal de l'arrangement des valeurs de la variable affectées aux nœuds peut se mesurer en évaluant comment ces valeurs aux nœuds reproduisent la forme de corrélation spatiale (covariance) que l'on souhaite imposer au modèle. Comme il existe plusieurs arrangements capables d'atteindre cet objectif, la recherche d'une solution optimale revient à générer une réalisation particulière de l'attribut modélisé. Ainsi, le problème initial de représentation stochastique est transformé en un problème d'optimisation qui est résolu par recuit simulé, méthode qui appartient à la famille des méthodes dites de relaxation stochastique.

La particularité des méthodes de relaxation stochastique est de perturber itérativement une solution "candidate" du problème et, à chaque itération, d'accepter ou non la perturbation opérée en fonction d'une certaine règle de décision. Cette règle dépend en partie de l'amélioration des propriétés statistiques de la solution perturbée par rapport aux propriétés voulues.

Pour illustrer l'application de cette méthode, on considère une variable catégorique (qualitative)  $Z(\mathbf{u})$  qui peut prendre les valeurs  $k$  avec  $k = 1, \dots, K$ . Le problème consiste à générer, suivant une répartition spatiale donnée, des valeurs de  $Z(\mathbf{u})$  aux  $N$  nœuds d'une grille. Il s'agit donc de simuler des valeurs  $z(\mathbf{u}_i)$ ,  $i = 1, \dots, N$ , qui préservent une certaine corrélation spatiale. Cette dernière peut être donnée, par exemple, par l'histogramme bivarié suivant :

$$f(\mathbf{h}; k_1, k_2) = \text{Prob}\{Z(\mathbf{u}) \in \text{à la catégorie } k_1, Z(\mathbf{u}+\mathbf{h}) \in \text{à la catégorie } k_2\},$$

$$\text{avec } k_1, k_2 = 1, \dots, K.$$

Cet histogramme est supposé connu pour les vecteurs de séparation  $\mathbf{h} = \mathbf{h}_1, \dots, \mathbf{h}_L$ . Une fonction objectif possible, qui mesure l'écart entre la corrélation spatiale d'une réalisation et celle souhaitée, peut être :

$$O = \sum_{l=1}^L \sum_{k_1=1}^K \sum_{k_2=1}^K \left[ f(\mathbf{h}_l; k_1, k_2)_{référence} - f(\mathbf{h}_l; k_1, k_2)_{réalisation} \right]^2.$$

En utilisant cette fonction, l'algorithme général du recuit simulé peut se résumer par les étapes suivantes.

1. Etablir les termes de référence de la fonction objectif, c'est-à-dire les valeurs de l'histogramme bivarié que l'on souhaite reproduire :

$$f(\mathbf{h}_l; k_1, k_2)_{référence}, l = 1, \dots, L; k_1 = 1, \dots, K; k_2 = 1, \dots, K.$$

2. Générer une réalisation initiale  $z(\mathbf{u}_i)$ ,  $i = 1, \dots, N$ , qui reproduit au moins les fréquences relatives de chacune des catégories auxquelles peut appartenir  $Z(\mathbf{u})^4$ . Cette solution initiale peut être obtenue, par exemple, en tirant aléatoirement les valeurs aux nœuds à partir de la distribution globale de  $Z(\mathbf{u})$ .

3. Calculer les termes de la fonction objectif pour la réalisation initiale, c'est-à-dire :

$$f(\mathbf{h}_l; k_1, k_2)_{réalisation}, l = 1, \dots, L; k_1 = 1, \dots, K; k_2 = 1, \dots, K.$$

4. Evaluer la fonction objectif  $O$ .
5. Sélectionner deux nœuds au hasard et permuter leurs valeurs. Cette permutation conduit à une nouvelle réalisation  $z'(\mathbf{u}_i)$ ,  $i = 1, \dots, N$ .
6. Mettre à jour tous les termes de la fonction objectif pour la nouvelle réalisation et réévaluer la fonction objectif pour obtenir la valeur  $O'$  affectée à cette réalisation.
7. La permutation, et par conséquent la nouvelle réalisation, est acceptée systématiquement si la fonction objectif décroît, c'est-à-dire si  $O' < O$ , et avec une certaine probabilité si la fonction objectif augmente. La seconde possibilité d'acceptation fait qu'une fraction des permutations, qui ne contribuent pas à réduire la fonction objectif, est acceptée de manière à éviter que le processus d'optimisation s'arrête sur un minimum local. La probabilité d'acceptation va en diminuant à mesure que la simulation progresse pour que diminue, à chaque étape, la fraction de permutations de ce type qui sont acceptées. C'est cette possibilité, d'accepter des perturbations qui ne contribuent pas à diminuer la fonction objectif, qui a donné à cette méthode son nom de recuit simulé.

---

<sup>4</sup> C'est-à-dire l'histogramme global,  $f(k)$ , qui donne la proportion de valeurs de chaque catégorie.

8. Si la fonction objectif atteint une valeur proche de zéro, la réalisation obtenue est gardée comme solution du problème. Dans le cas contraire, le processus est recommencé à partir de l'étape 5 avec une nouvelle permutation.

Cet algorithme peut connaître de nombreuses variantes selon la fonction objectif considérée, la manière de générer une réalisation initiale, le mécanisme de perturbation employé ou la règle d'acceptation choisie.

La technique du recuit simulé décrite précédemment peut être adaptée également pour tenir compte d'une information secondaire. Cette adaptation passe par une modification de la fonction objectif qui doit pouvoir intégrer l'information secondaire et sa relation avec la variable principale dans le processus d'optimisation. Par exemple, si on possède un ensemble de données de calibrage, il est envisageable d'obtenir une information sur la corrélation entre les variables principale et secondaire à partir du diagramme de dispersion correspondant. Si on résume cette corrélation par un coefficient de corrélation linéaire  $\rho_{calibrage}$ , que l'on souhaite voir reproduit par les réalisations de la variable principale, il suffit de rajouter à la fonction objectif précédente le terme suivant :

$$O_c = [\rho_{calibrage} - \rho_{réalisation}]^2.$$

De la même manière, on peut prendre en compte tout autre type d'information secondaire dans la mesure où on peut exprimer sa relation avec la variable principale sous une forme qui s'intègre dans la fonction objectif. Néanmoins, le choix d'une fonction objectif trop complexe peut rendre la méthode inapplicable d'un point de vue informatique.

Comme application intéressante de la méthode, il peut être noté celle de Deutsch et Journel (1992b) qui proposent un schéma d'approximation par recuit simulé pour inclure comme information secondaire des perméabilités moyennes (effectives) déduites d'essais de puits dans la simulation de perméabilités à une échelle fine.

L'inconvénient majeur de cette approche est que, en voulant traiter itérativement un problème d'optimisation, la convergence vers une solution optimale peut être lente et donc réduire l'efficacité en temps de calcul de la méthode. Il existe des implémentations qui accélèrent la convergence mais qui, en contrepartie, peuvent produire des solutions biaisées. En choisissant convenablement la réalisation initiale, ce biais peut être cependant minimisé.

#### **2.4.4 Simulation par champ de probabilités**

Froidevaux (1992) et Srivastava (1992 et 1994) ont présenté une technique de simulation conditionnelle, connue sous le nom de simulation par champ de probabilités, capable d'incorporer différents types d'information. Cette méthode présente deux caractéristiques originales. La première concerne la dissociation entre les tâches d'estimation de la fonction de distribution conditionnelle locale, en chaque point à simuler, et la génération des réalisations équiprobables

de l'attribut<sup>5</sup>. La seconde réside dans la simplicité d'intégration d'informations de différents types.

On considère un attribut  $Z$  à simuler aux nœuds d'une grille à partir d'une série de mesures. Au préalable, il peut être déterminé en chaque nœud la fonction de distribution conditionnelle locale selon un modèle de fonction aléatoire qui dépend de la méthode de calcul choisie parmi celles décrites aux paragraphes précédents (méthodes paramétriques ou non paramétriques). Concrètement, on calcule en chaque nœud de coordonnées  $\mathbf{u}_i$  la fonction (cumulative) de distribution conditionnelle :

$$F(\mathbf{u}_i; z|(n)) = \text{Prob}\{Z(\mathbf{u}_i) \leq z|(n)\},$$

où  $(n)$  désigne l'ensemble des informations locales, situées dans un certain voisinage du nœud, utilisé pour le conditionnement.

On considère à présent une réalisation quelconque de l'attribut, soit  $\{z_s(\mathbf{u}_i), i = 1, \dots, N\}$ . Etant donné que les fonctions de distribution locales sont définies de manière unique, il existe, entre les valeurs simulées de la réalisation et les valeurs de probabilité associées, une relation biunivoque qui s'exprime à travers les distributions locales comme :

$$\{z_s(\mathbf{u}_i), i = 1, \dots, N\} \Leftrightarrow \{p_s(\mathbf{u}_i), i = 1, \dots, N\}$$

avec  $p_s(\mathbf{u}_i) = F(\mathbf{u}_i; z_s(\mathbf{u}_i)|(n))$  où les probabilités locales  $p_s(\mathbf{u}_i)$  peuvent être interprétées comme la réalisation d'une fonction aléatoire  $P(\mathbf{u})$ . Cette interprétation offre la possibilité de générer une réalisation du champ de probabilités à partir de laquelle, en utilisant les fonctions de distribution locales calculées préalablement, on obtient les valeurs de l'attribut à modéliser.

L'intérêt de cette procédure est double. D'une part, les réalisations des champs de probabilité n'ont pas à être conditionnelles. D'autre part, la fonction  $F(\mathbf{u}; z|(n))$ , calculée une fois seulement pour l'ensemble des réalisations de l'attribut, peut s'obtenir avec les méthodes d'estimation les plus élaborées sans soucis de performance.

Pour la génération des champs de probabilités, les hypothèses suivantes sont admises.

1. En tant que champ de probabilité,  $P(\mathbf{u})$  suit une loi de distribution uniforme.
2. Entre la transformée uniforme  $U(Z(\mathbf{u}))$  et  $P(\mathbf{u})$ , il existe une "conservation des caractéristiques majeures" qui permet d'assimiler la covariance du champ de probabilités à celle de la transformée uniforme des données disponibles, c'est-à-dire :

$$C_P(\mathbf{h}) \approx C_U(\mathbf{h})$$

avec,

---

<sup>5</sup> Ceci est à comparer à la simulation séquentielle qui consiste à calculer séquentiellement, en chaque point à simuler, la fonction de distribution locale conditionnelle par rapport aux données et aux valeurs déjà simulées aux étapes précédentes.

$$U(\mathbf{u}) = \text{Prob}\{Z < z(\mathbf{u})\}.$$

Ce point est le plus intuitif de l'algorithme et il n'en existe pas de justification théorique. Dans Froidevaux (1992), ces hypothèses sont vérifiées expérimentalement.

L'implémentation pratique de la méthode de simulation par champ de probabilités peut se résumer comme suit.

1. Définir une grille qui recouvre le domaine d'intérêt.
2. Déterminer en chaque nœud de la grille la fonction (cumulative) de distribution conditionnelle locale  $F(\mathbf{u}, z)$  de l'attribut à simuler.
3. Calculer et modéliser la fonction de covariance de la transformée uniforme de  $Z(\mathbf{u})$ . Il est alors admis que :

$$C_P(\mathbf{h}) \approx C_U(\mathbf{h})$$

4. Générer aux nœuds de la grille une simulation non conditionnelle de  $P(\mathbf{u})$  selon une distribution uniforme et une covariance  $C_P(\mathbf{h})$ .
5. En chaque nœud, calculer la valeur  $z_s(\mathbf{u})$  de l'attribut à partir de la fonction de distribution locale  $F(\mathbf{u}, z)$  et de la probabilité  $p_s(\mathbf{u})$  affectée au nœud, c'est-à-dire :

$$z_s(\mathbf{u}) = F^{-1}[\mathbf{u}, p_s(\mathbf{u})].$$

6. Répéter les étapes 4 et 5 pour obtenir d'autres réalisations équiprobables de l'attribut  $Z$ .

### 2.4.5 Algorithmes booléens

Les méthodes booléennes constituent une autre catégorie parmi les méthodes de simulation stochastique. L'objectif de ces méthodes est de remplir l'espace à simuler avec des objets géométriques dont les paramètres géométriques suivent certaines lois de distribution. Un exemple classique de modèle booléen est celui des réseaux de fractures pour la modélisation d'écoulements dans des massifs granitiques. Les objets sont des fractures planes de forme elliptique et les paramètres géométriques qui les définissent sont les coordonnées du centre, l'orientation du grand axe et les longueurs des deux axes principaux.

Bien que leur fondement géométrique rende les modèles booléens attractifs, ces derniers sont d'une mise en œuvre complexe (il est difficile d'inférer statistiquement les distributions de probabilité des paramètres géométriques) et sont très difficiles à conditionner sur des données existantes.

## 2.4.6 Simulation par gaussiennes tronquées

Parmi les méthodes décrites précédemment, celles susceptibles de pouvoir simuler une variable catégorique sont la simulation séquentielle d'indicatrice qui offre un cadre général et la simulation booléenne dans la mesure où les catégories peuvent être associées à des objets géométriques<sup>6</sup>.

Une autre méthode, conçue spécifiquement pour simuler des variables catégoriques et plus particulièrement des types de faciès, mérite également d'être mentionnée ici. Il s'agit de la simulation par gaussiennes tronquées (voir par exemple Matheron *et al.*, 1987, de Fouquet *et al.*, 1989, Xu et Journel, 1993).

Pour ce type de simulation, les données d'entrée sont :

- une répartition spatiale des proportions locales de classes de la variable (appliquée à la simulation de faciès, il s'agit de distributions de probabilité locales des faciès simulés<sup>7</sup>),
- un variogramme pour la simulation d'une variable aléatoire gaussienne.

Les cartes de proportions locales sont transformées en cartes de seuils (cartes de *troncation*) qui sont utilisées pour transformer (tronquer) le champ gaussien simulé (simulation de la variable gaussienne) en carte de faciès<sup>8</sup>.

Les avantages de la méthode sont les suivants :

- grande vitesse d'exécution,
- possibilité d'incorporation de tout type d'information dans la construction des cartes de proportions (par ex. en utilisant du krigeage d'indicatrice).

Elle présente cependant aussi des inconvénients :

- forme de continuité spatiale contrôlée par une covariance unique<sup>9</sup>,

---

<sup>6</sup> Il est à noter aussi la simulation par recuit simulé qui pourrait être adaptée à une variable catégorique.

<sup>7</sup> Ces distributions locales donnent, en tout point de l'espace, la probabilité de trouver, au point considéré, un faciès particulier.

<sup>8</sup> Il est à noter que si les cartes de proportions locales étaient conservées telles quelles (c'est-à-dire sans transformation en cartes de seuils) et que le champ Gaussien était remplacé par le champ de probabilités utilisé pour la simulation (chaque valeur simulée de la variable Gaussienne est nécessairement associée à une valeur de probabilité tirée aléatoirement), une lecture directe à partir des proportions locales, en utilisant les probabilités, permettrait de déterminer les faciès. Cette variante correspondrait à une approche par champ de probabilités.

<sup>9</sup> Par exemple, il peut être difficile d'introduire des anisotropies différentes pour deux types de faciès. Une solution consisterait à prendre en compte les anisotropies variables lors de la construction des cartes de proportions locales de faciès qui pourraient être calculées par krigeage d'indicatrices. Cette solution n'est cependant pas entièrement satisfaisante car elle ne garantit pas de reproduire convenablement la continuité spatiale de chacun des faciès simulés.

- nécessité d'avoir verticalement une séquence (stratigraphique) de faciès dont l'ordre est figé<sup>10</sup>.

En pratique, pour simuler des faciès répondant aux spécificités précédentes, les bons ou mauvais résultats de la méthode par gaussiennes tronquées dépendent pour beaucoup de l'effort consacré à la construction des cartes de proportions locales et à la modélisation du variogramme du champ gaussien utilisé pour produire les faciès simulés. Si cette étape préparatoire est accomplie convenablement, la méthode semble donner des résultats aussi bons que la simulation séquentielle d'indicatrices (voir par ex. Xu et Journel 1993). Dans son principe, la méthode par gaussiennes tronquées reste cependant **équivalente** à une approche par champs de probabilités.

Concernant le logiciel Heresim, dans lequel est implémentée la simulation par gaussiennes tronquées, il est à noter que, sauf amélioration du code, le seul type de covariance (variogramme) autorisé pour simuler la variable gaussienne est un type exponentiel factorisé (produit de termes indépendants pour décrire la covariance dans le plan horizontal et dans la direction verticale). Ce type de modèle de covariance rend le programme de simulation beaucoup plus rapide mais limite la précision et les possibilités de simulation (il est en particulier mal adapté à la modélisation de stratifications obliques).

---

<sup>10</sup> Par conséquent, ce type de simulation conviendrait uniquement pour modéliser des formations sédimentaires.

## 2.5 Simulation : exemple d'application

### 2.5.1 Principe de l'exercice

Pour évaluer conjointement le comportement des méthodes de simulation, un exercice a été réalisé dans lequel sont modélisés l'écoulement et le transport de masse pour l'aquifère qui a été décrit dans la partie consacrée à l'évaluation des méthodes d'estimation. Les méthodes de simulation testées sont les suivantes :

1. simulation gaussienne séquentielle de la variable principale (SGSIM = sequential gaussian simulation),
2. simulation gaussienne séquentielle de deux variables (principale et secondaire) en utilisant le cokrigage co-localisé sous un modèle markovien (CC-SIM = collocated cokriging),
3. simulation séquentielle d'indicateur en utilisant le krigeage d'indicateur (SISIM = sequential indicator simulation),
4. simulation séquentielle d'indicateur en utilisant le cokrigage d'indicateur avec un modèle de Markov-Bayes (MBSIM = markov-bayes simulation),
5. simulation par champs de probabilités (PFSIM = probability field simulation),
6. simulation par recuit simulé (SASIM = simulated annealing simulation).

La méthodologie suivie fut identique à celle décrite par Gómez-Hernández et Wen (1992) pour quantifier l'incertitude sur l'estimation du temps et de la position d'arrivée d'un contaminant qui chemine à travers un aquifère.

Le problème d'écoulement et de transport a d'abord été résolu pour le champ de transmissivité de référence (schéma du haut de la Figure 2-20). Les conditions aux limites prises en comptes sont : potentiels hydrauliques imposés sur les frontières Ouest et Est, avec des valeurs respectives de 30 et 4 mètres, et flux nul à travers les frontières du Nord et du Sud. La carte des niveaux piézométriques correspondants se trouve au bas de la Figure 2-20. On observe que ces conditions aux limites provoquent un écoulement dirigé de l'Ouest vers l'Est.

Pour le problème de transport, on se place dans le cas hypothétique d'un stockage souterrain de déchets radioactifs situé dans l'aquifère et à partir duquel s'échappent un millier de radionucléides. Le transport par convection a été modélisé pour les trois scénarios suivants qui se distinguent par l'emplacement et la taille de la zone d'émission des radionucléides.

- Scénario 1 : émission à partir d'un segment de 2 km de longueur dont le centre, situé sur le bord Ouest de l'aquifère, a pour coordonnée verticale  $y = 26,5$  km.
- Scénario 2 : émission à partir d'un segment de 4 km de longueur dont le centre, situé sur le bord Ouest de l'aquifère, a pour coordonnée verticale  $y = 12$  km.

- Scénario 3 : émission à partir d'un segment de 1 km de longueur dont les extrémités, situées à l'intérieur de l'aquifère, ont pour coordonnées ( $x = 5,5$  km ;  $y = 17$  km) et ( $x = 6,5$  km ;  $y = 17$  km).

Les trajectoires suivies par les 20 premières particules émises, sur un total de 1000, sont montrées sur la Figure 2-21 (scénario 1), sur la Figure 2-22 (scénario 2) et sur la Figure 2-23 (scénario 3). Sur chacune de ces figures, les trajectoires sont visualisées avec en fond le champ de transmissivité de référence (schéma du haut) et celui des niveaux piézométriques (schéma du bas).

Les simulations stochastiques des champs de transmissivité ont été réalisées à partir des données, principales et secondaires, et des modèles statistiques décrits dans la partie consacrée à l'exercice d'application des méthodes d'estimation (§2.3.1, p 73). La méthodologie suivie, pour étudier l'incertitude sur la prédiction des temps et des positions d'arrivées des particules, a été la même pour tous les scénarios et pour toutes les méthodes de simulation testées. Elle se résume comme suit.

1. Générer 100 simulations du champ de transmissivité. Ces simulations sont des réalisations équiprobables qui honorent toutes l'ensemble des informations dures et douces (exceptées celles produites par la simulation gaussienne séquentielle qui ne permet pas d'incorporer une information douce) et les formes de continuité spatiale modélisées à partir des données de calibrage.
2. Sur chacun des champs de transmissivité obtenus à l'étape précédente, et pour les conditions aux limites d'écoulement décrites plus haut, résoudre l'équation de diffusivité en régime permanent et en déduire le champ piézométrique.
3. Pour chacun des champs piézométriques calculés à l'étape précédente, résoudre l'équation de transport (transport par convection uniquement) et en déduire les trajectoires des particules.
4. Analyser statistiquement les résultats de transport obtenus. Cette analyse consiste à déterminer, pour l'ensemble des 100 réalisations générées par chaque méthode de simulation stochastique, la courbe moyenne des temps d'arrivée, les courbes de temps d'arrivée extrêmes, correspondant à des probabilités de 5 et 90%<sup>1</sup>, et les histogrammes des positions et des temps d'arrivée.

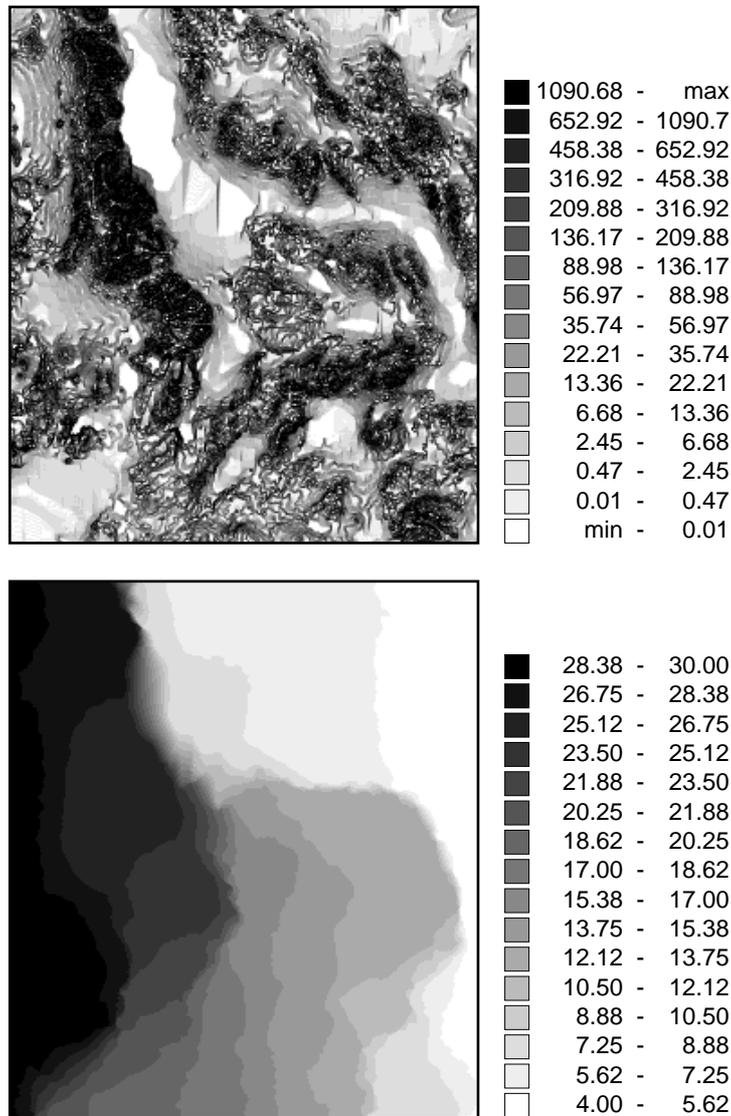
Les figures 10.5 à 10.22 montrent, à gauche, deux champs de transmissivité simulés, et à droite les champs piézométriques correspondants qui proviennent de la résolution du modèle d'écoulement pour chaque scénario et pour chaque méthode de simulation stochastique testée. Sur tous ces champs, les trajectoires suivies par 20 particules qui s'échappent du stockage ont été représentées. Les figures 10.5 à 10.10 concernent le scénario 1, les figures 10.11 à 10.16 le scénario 2 et les figures 10.17 à 10.22 le scénario 3.

---

<sup>1</sup> Pour 1000 particules simulées, il s'agit des temps d'arrivée mis par la 50<sup>ème</sup> et par la 900<sup>ème</sup> particule la plus rapide pour chacune des 100 réalisations du champ de transmissivité générées par une méthode donnée.

Champ de transmissivité d'origine (haut) et champ piézométrique de référence (bas)

Figure 2-20

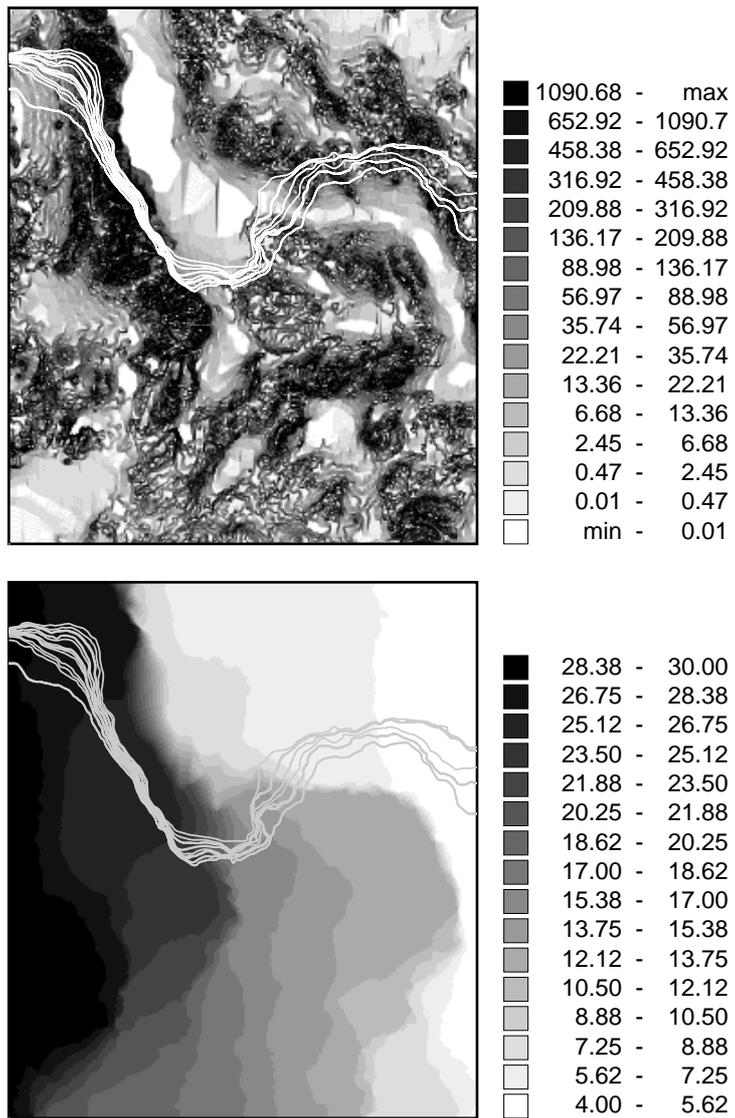


Les figures 10.23, 10.24 et 10.25 fournissent les courbes des temps d'arrivée moyens et extrêmes, correspondant à des probabilités de 5 et 90%, pour chacune des méthodes testées et pour chaque scénario. Est également représentée sur ces figures la courbe des temps d'arrivée obtenus sur le champ de transmissivité de référence.

Les figures 10.26, 10.27 et 10.28 montrent les histogrammes de la position d'arrivée correspondant au champ de transmissivité de référence et à chacune des méthodes de simulation stochastique pour chaque scénario étudié. De la même manière, les figures 10.29, 10.30 et 10.31 fournissent les histogrammes des temps d'arrivée. Le Tableau 2-6 résume les statistiques sur les positions d'arrivée et le Tableau 2-7 celles sur les temps.

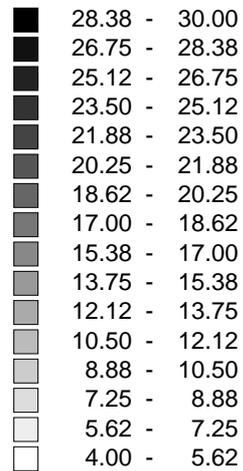
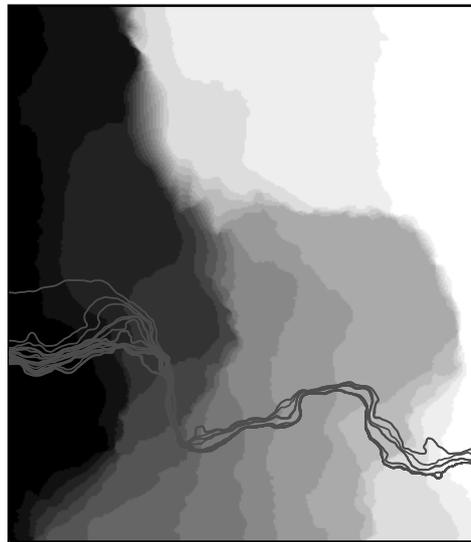
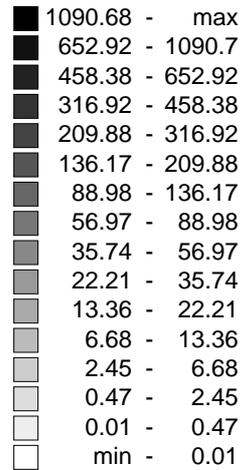
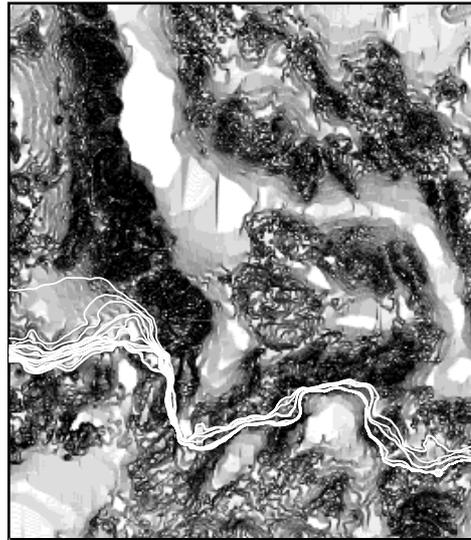
Champ de transmissivité d'origine (haut) et champ piézométrique de référence (bas) sur lesquels ont été portées les trajectoires de 20 particules pour le scénario 1.

Figure 2-21



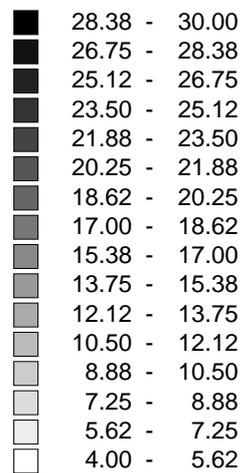
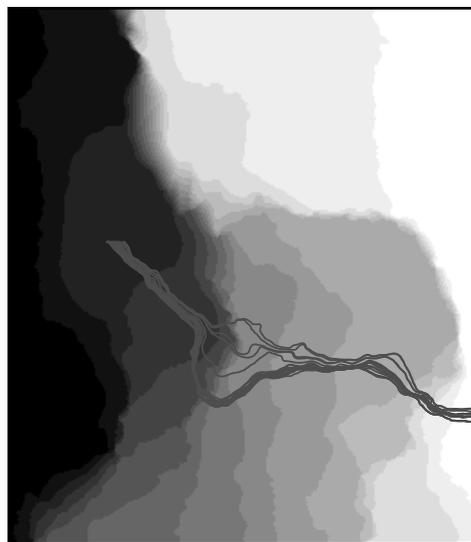
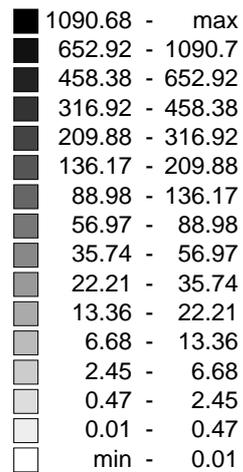
Champ de transmissivité d'origine (haut) et champ piézométrique de référence (bas) sur lesquels ont été portées les trajectoires de 20 particules pour le scénario 2.

Figure 2-22



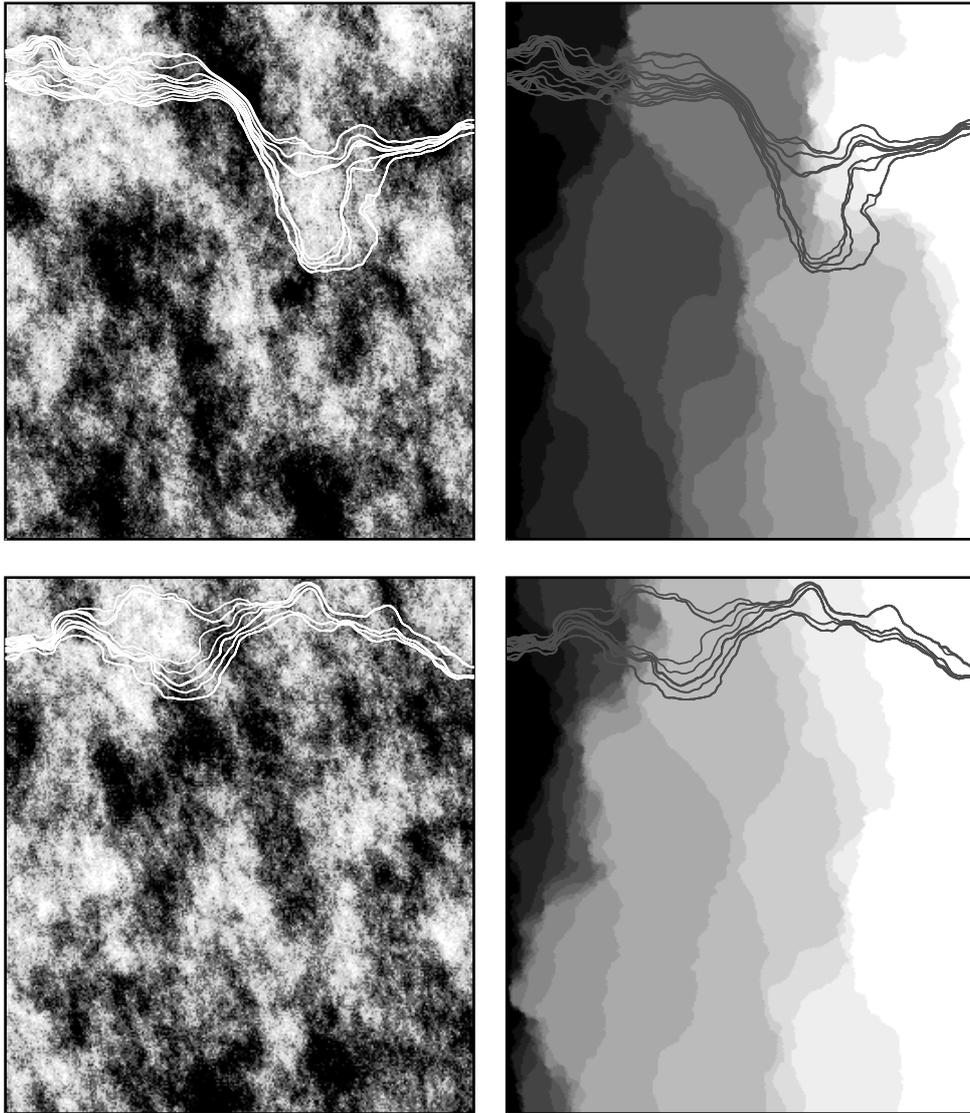
Champ de transmissivité d'origine (haut) et champ piézométrique de référence (bas) sur lesquels ont été portées les trajectoires de 20 particules pour le scénario 3.

Figure 2-23



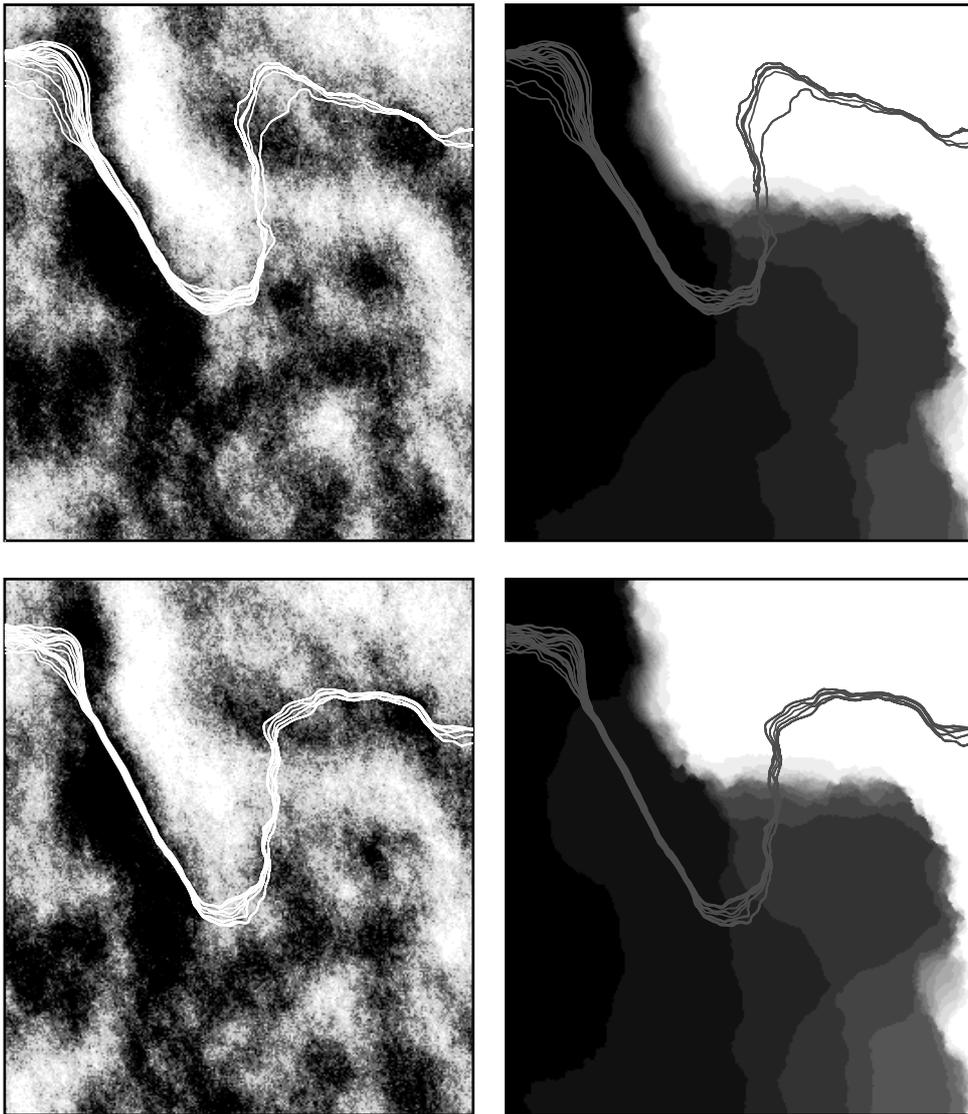
Cas du scénario 1. A gauche, deux champs de transmissivité générés par simulation gaussienne séquentielle. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-24



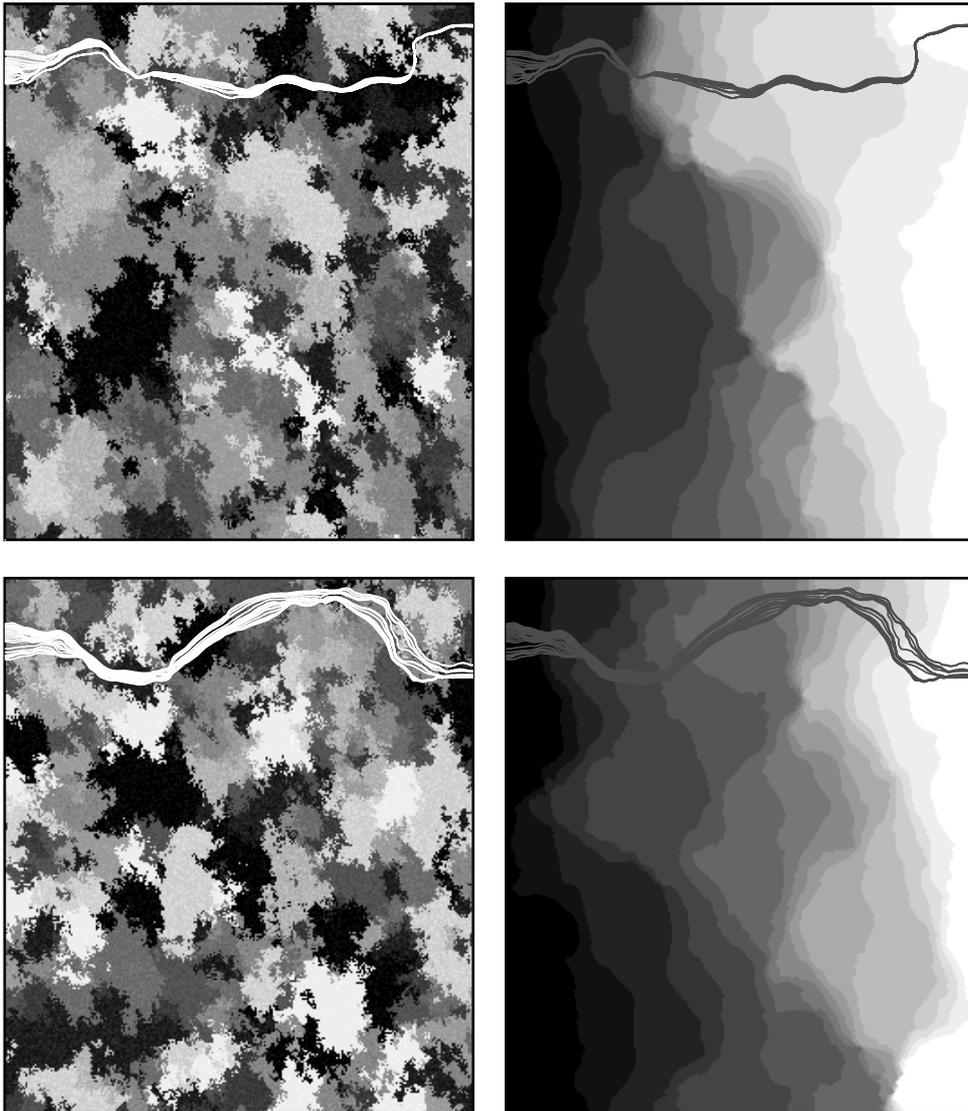
Cas du scénario 1. A gauche, deux champs de transmissivité générés par simulation gaussienne séquentielle par cokrigage co-localisé suivant un modèle markovien. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-25



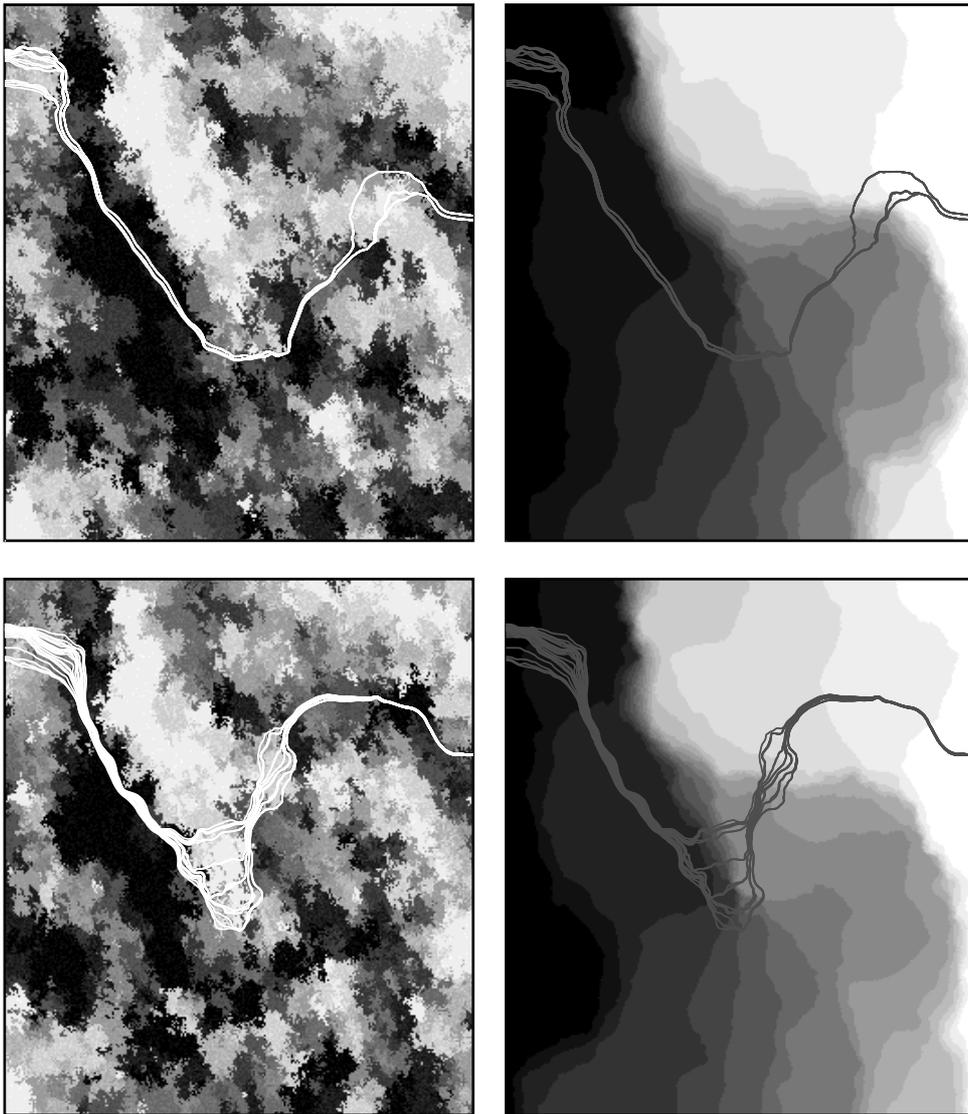
Cas du scénario 1. A gauche, deux champs de transmissivité générés par simulation séquentielle d'indicateur. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-26



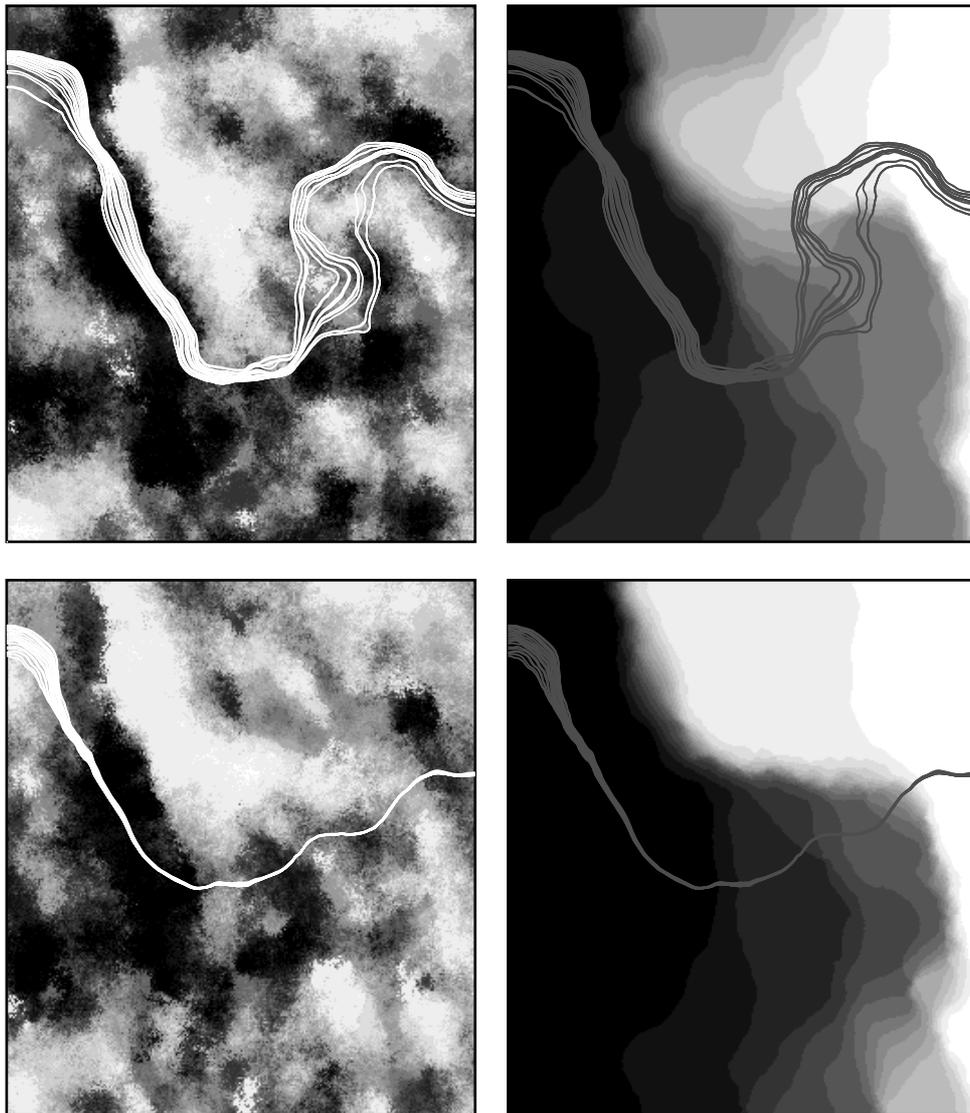
Cas du scénario 1. A gauche, deux champs de transmissivité générés par simulation séquentielle par cokrigage d'indicateur suivant un modèle markovien. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-27



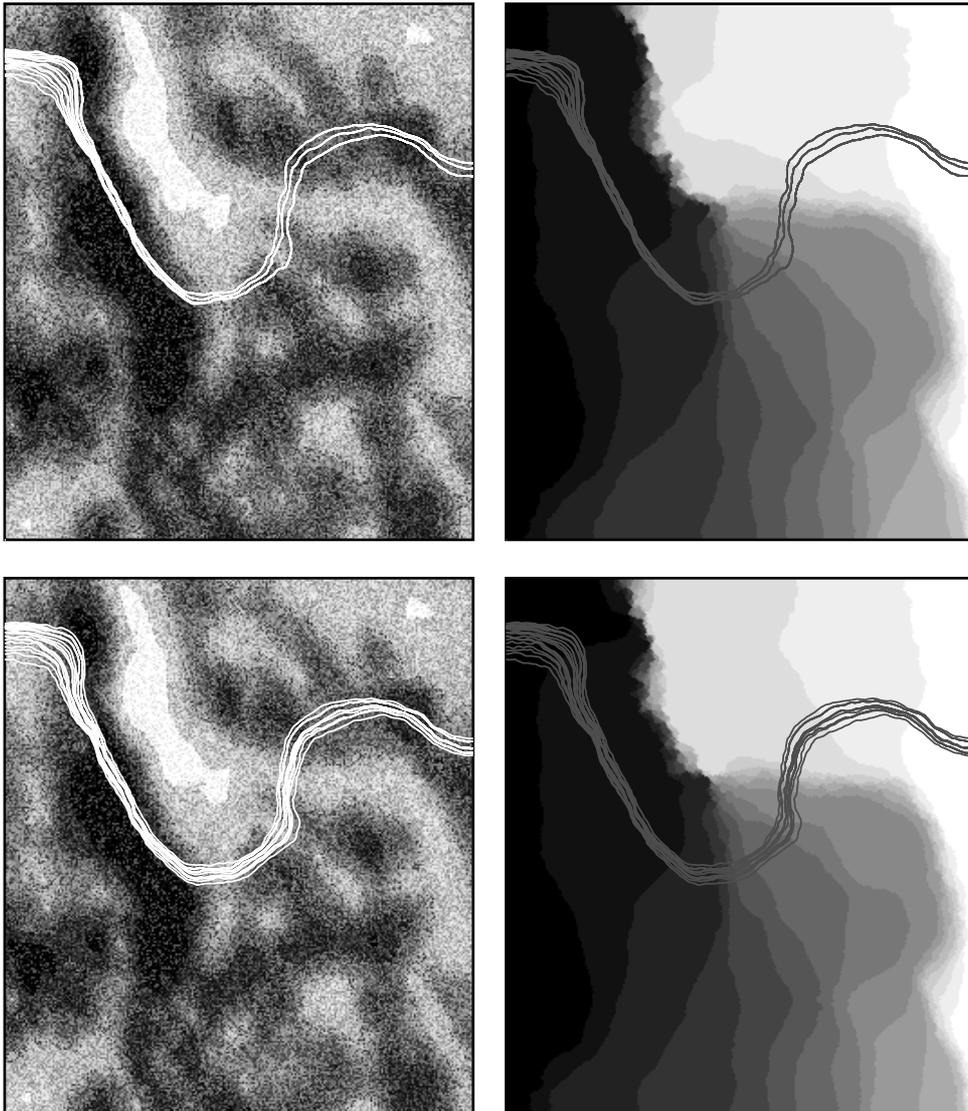
Cas du scénario 1. A gauche, deux champs de transmissivité générés par simulation par champ de probabilités. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-28



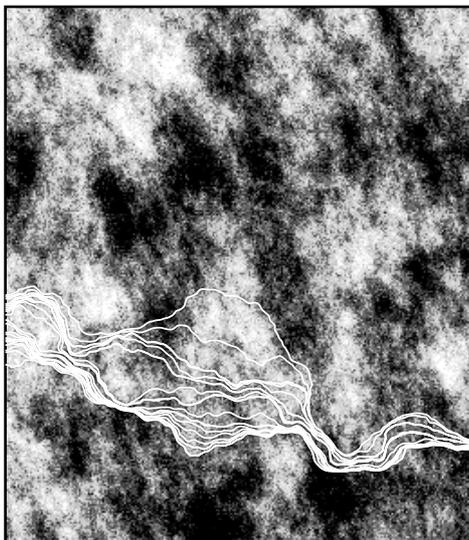
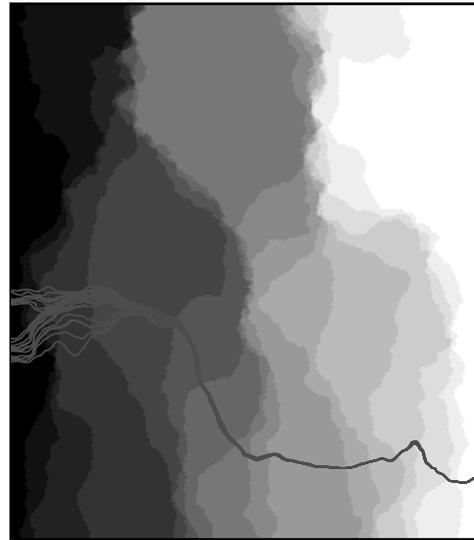
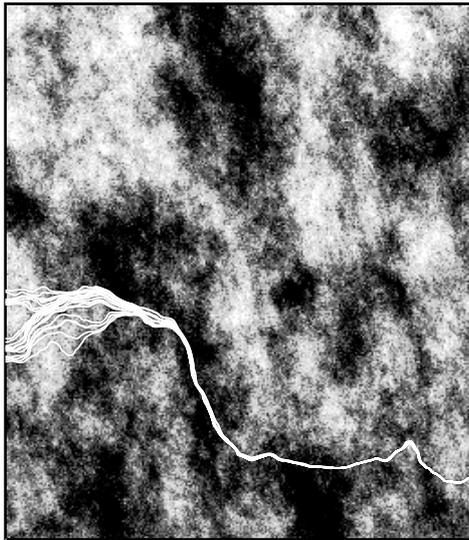
Cas du scénario 1. A gauche, deux champs de transmissivité générés par simulation par recuit simulé. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-29



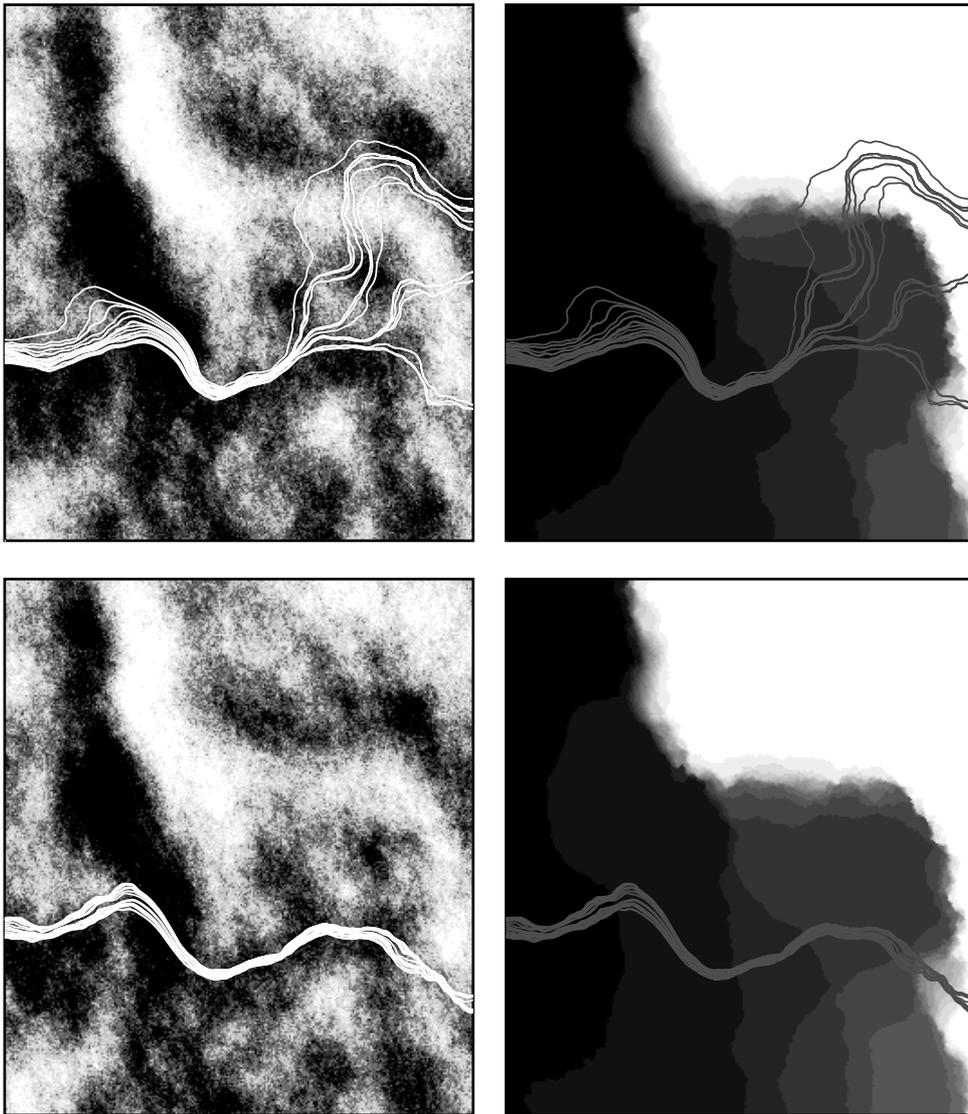
Cas du scénario 2. A gauche, deux champs de transmissivité générés par simulation gaussienne séquentielle. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-30



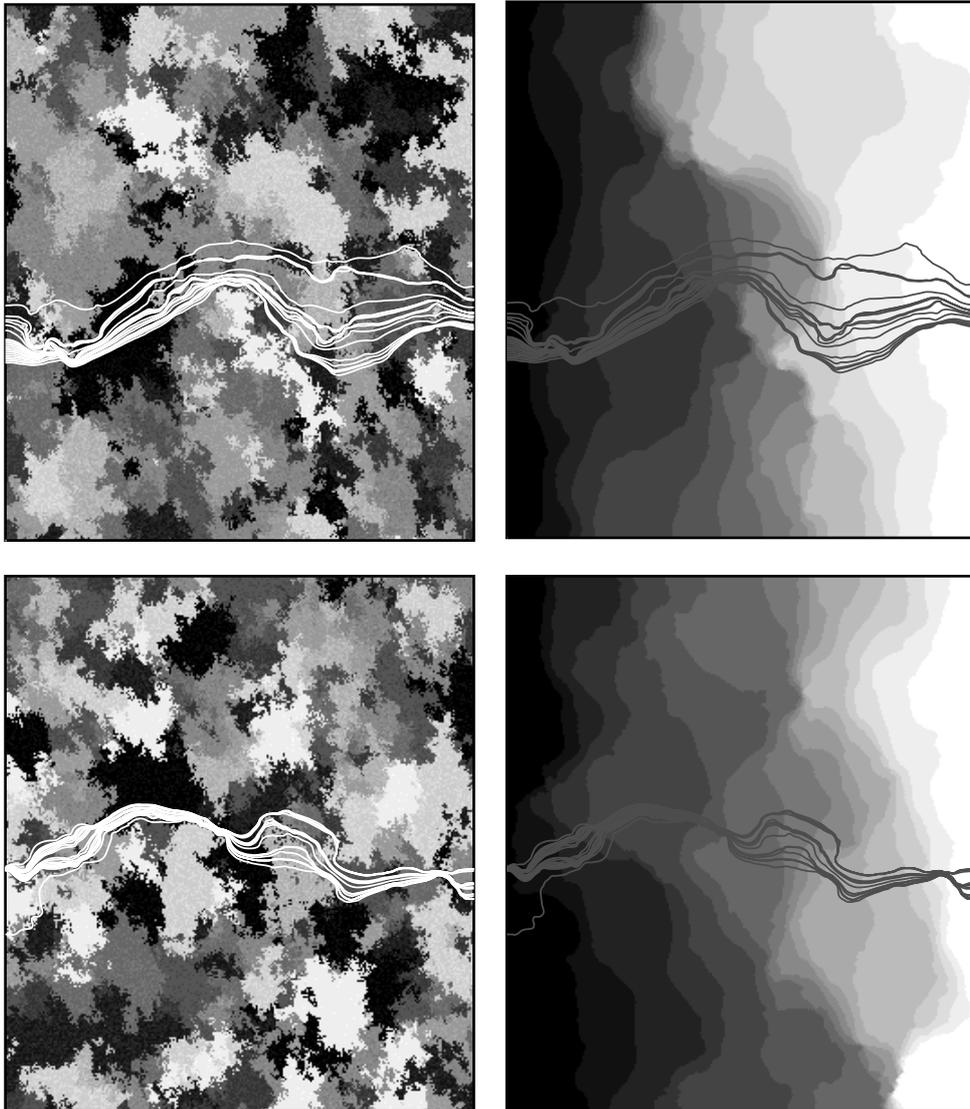
Cas du scénario 2. A gauche, deux champs de transmissivité générés par simulation gaussienne séquentielle par cokrigage co-localisé suivant un modèle markovien. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-31



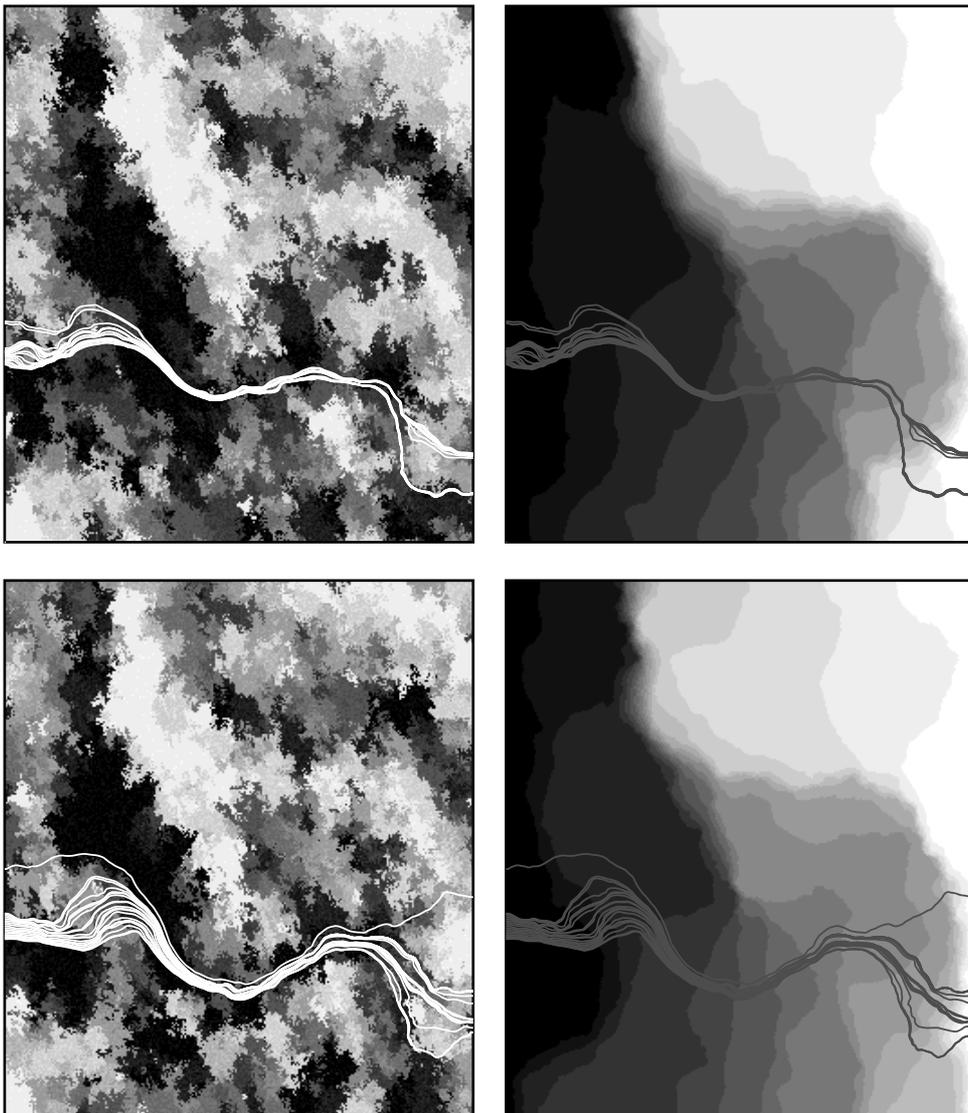
Cas du scénario 2. A gauche, deux champs de transmissivité générés par simulation séquentielle d'indicateur. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-32



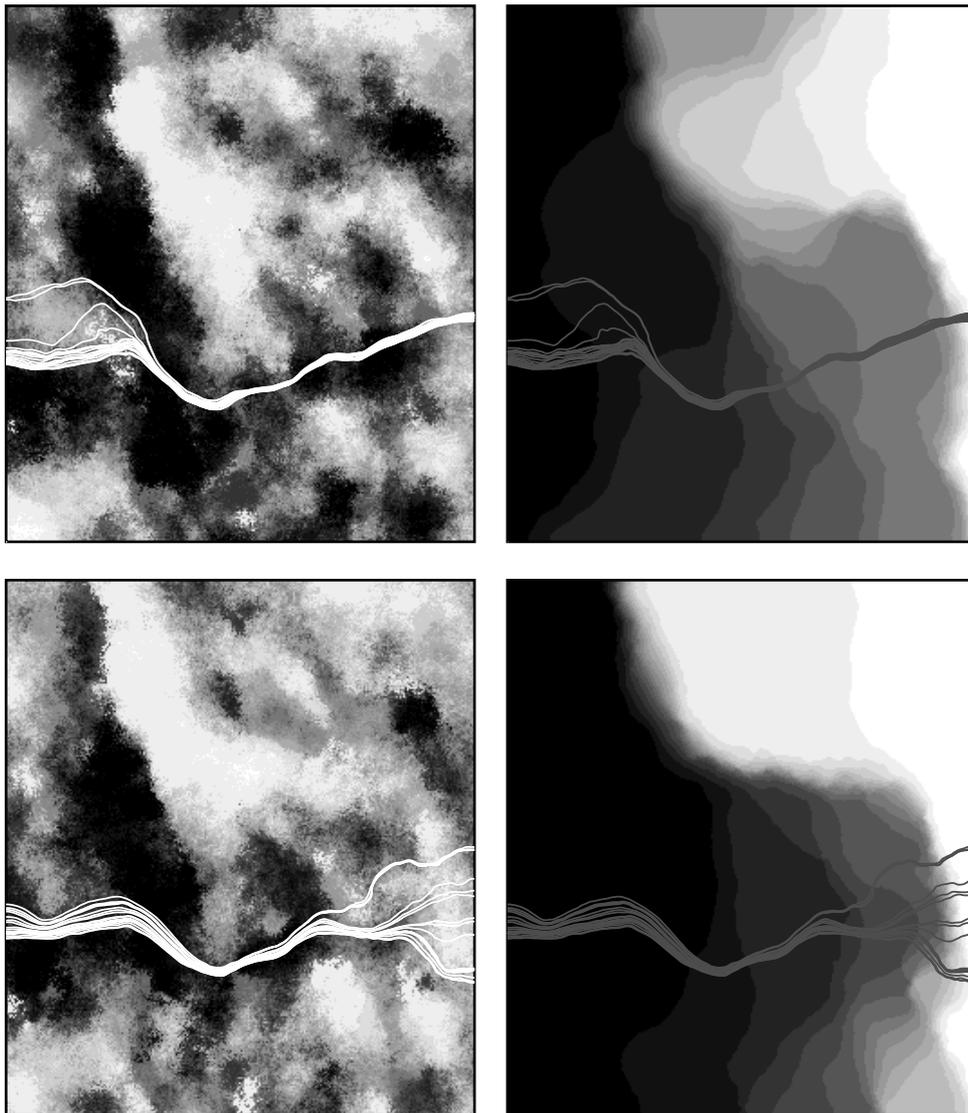
Cas du scénario 2. A gauche, deux champs de transmissivité générés par simulation séquentielle par cokrigage d'indicateur suivant un modèle markovien. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-33



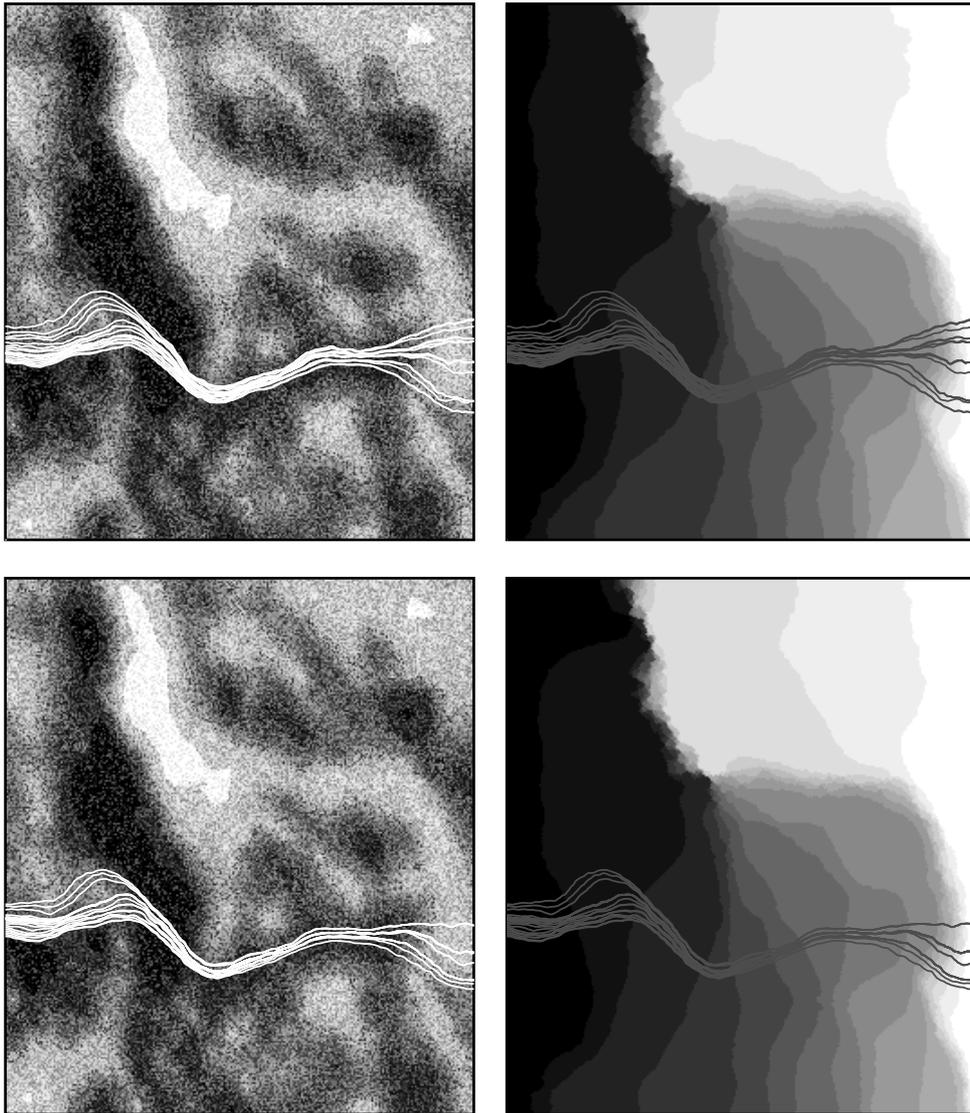
Cas du scénario 2. A gauche, deux champs de transmissivité générés par simulation par champ de probabilités. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-34



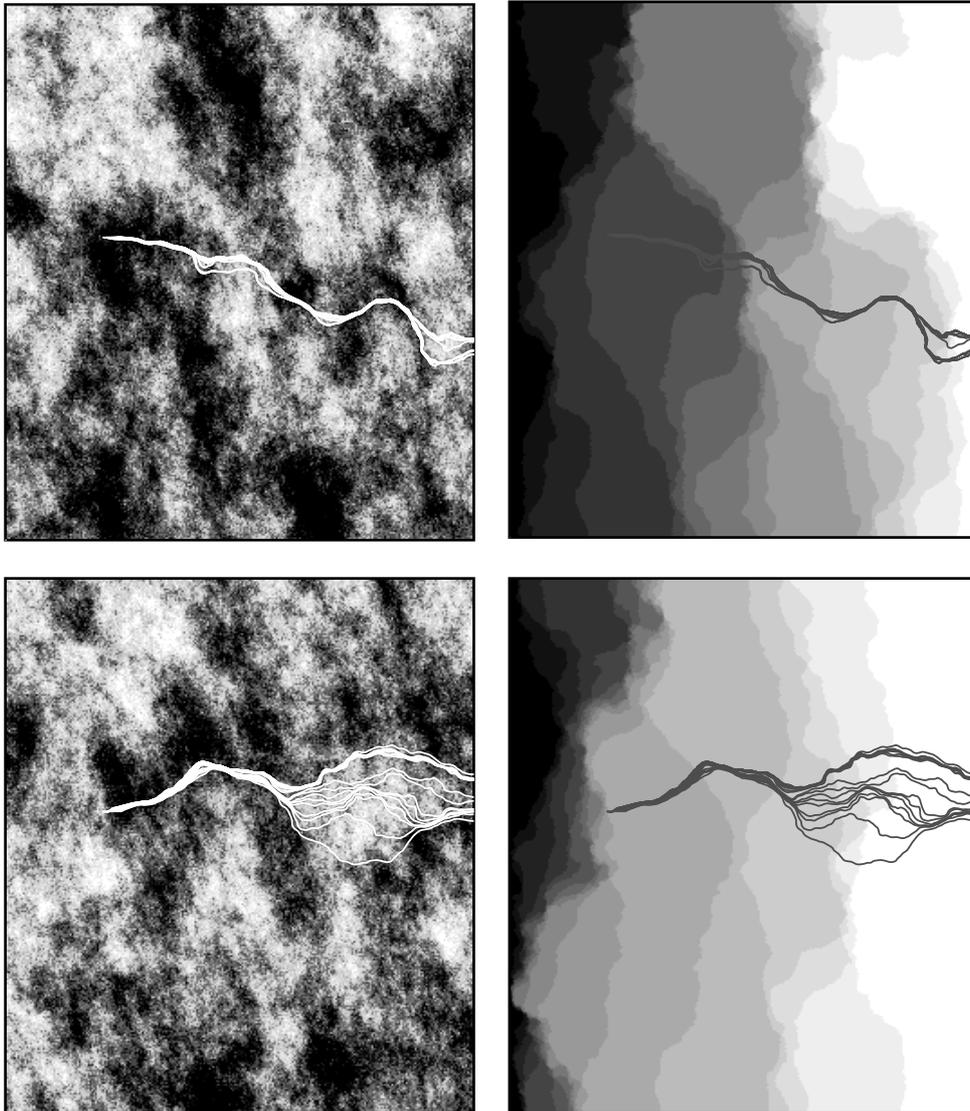
Cas du scénario 2. A gauche, deux champs de transmissivité générés par simulation par recuit simulé. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-35



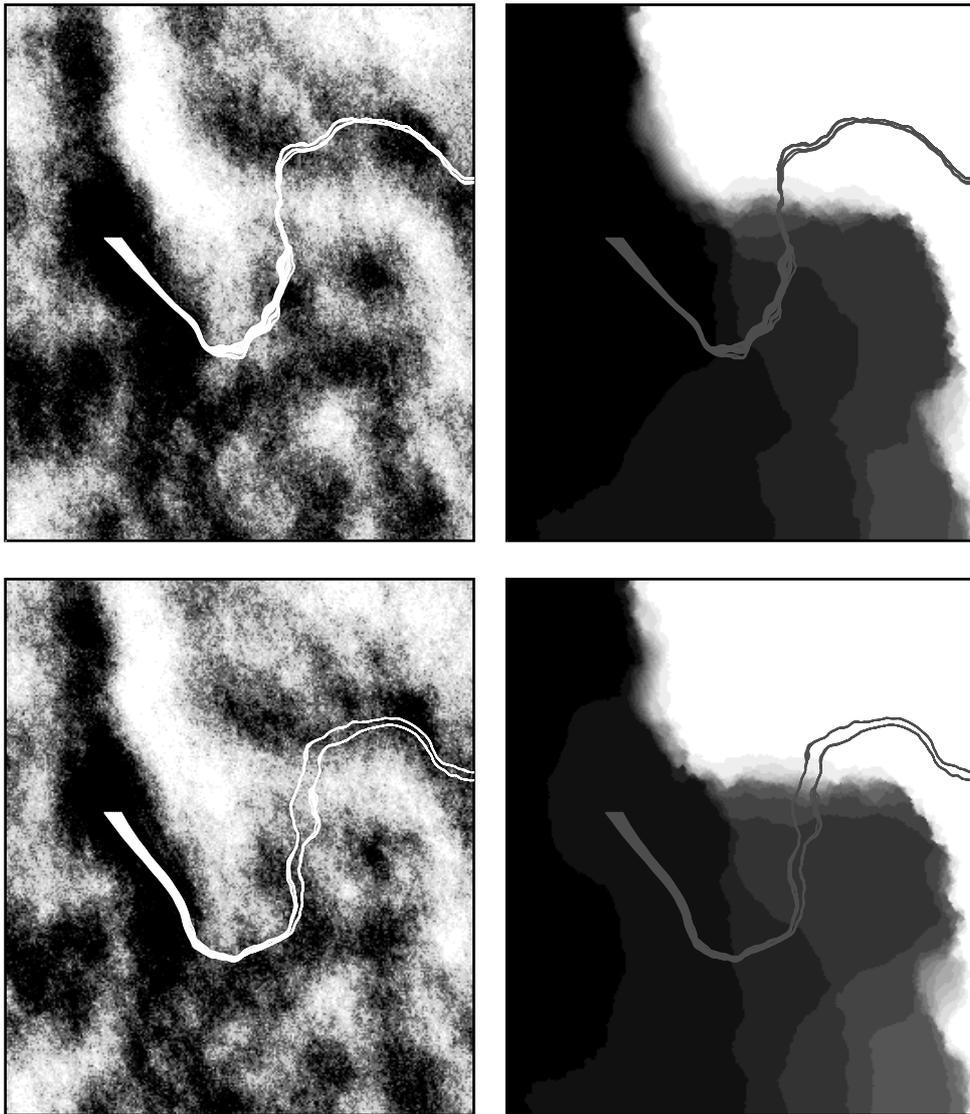
Cas du scénario 3. A gauche, deux champs de transmissivité générés par simulation gaussienne séquentielle. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-36



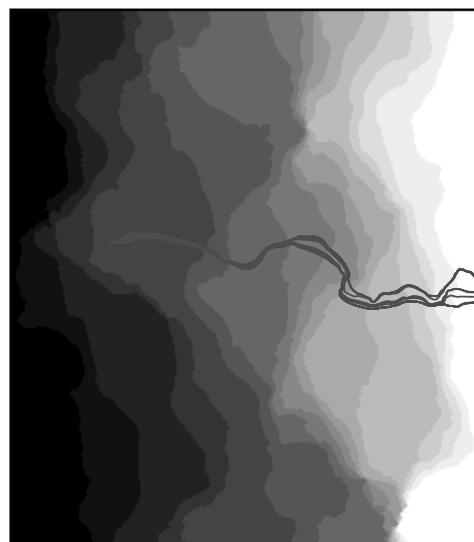
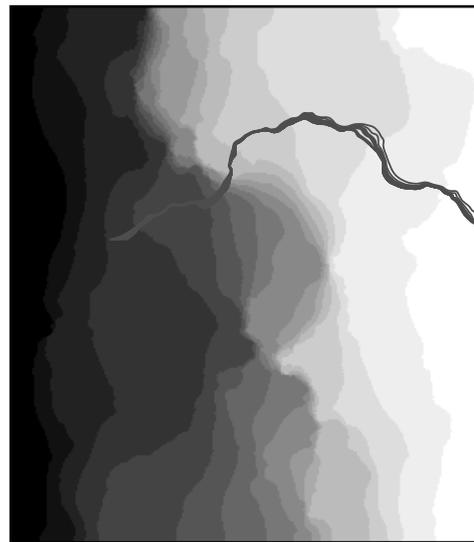
Cas du scénario 3. A gauche, deux champs de transmissivité générés par simulation gaussienne séquentielle par cokrigage co-localisé suivant un modèle markovien. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-37



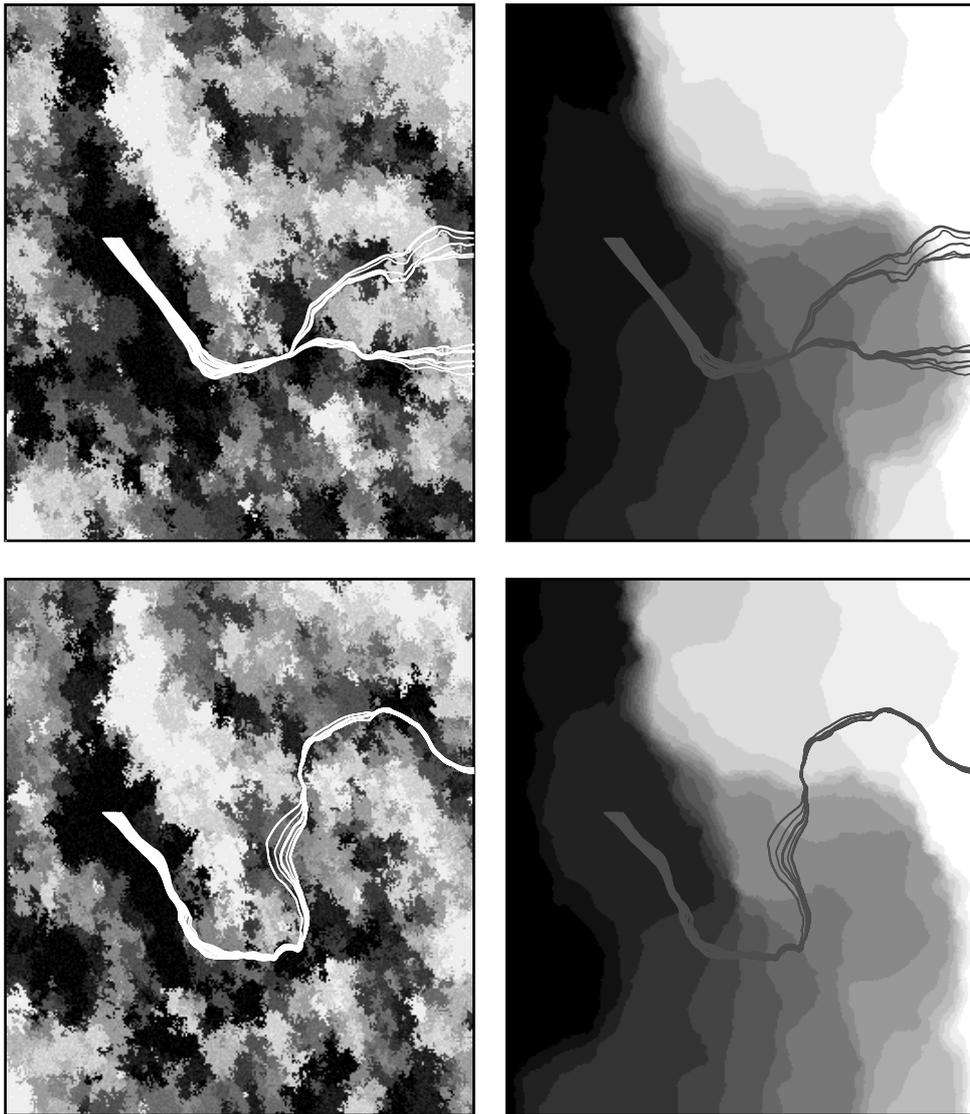
Cas du scénario 3. A gauche, deux champs de transmissivité générés par simulation séquentielle d'indicateur. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-38



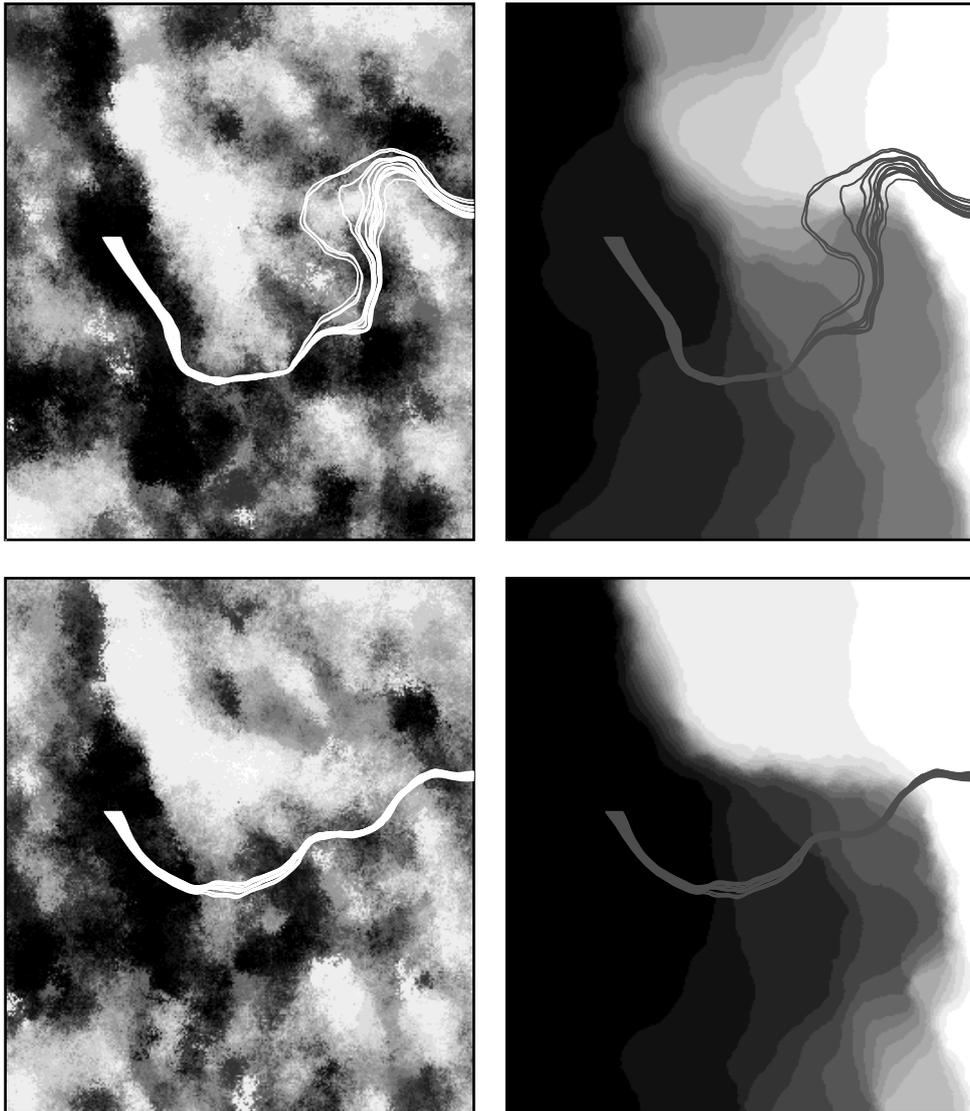
Cas du scénario 3. A gauche, deux champs de transmissivité générés par simulation séquentielle par cokrigage d'indicateur suivant un modèle markovien. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-39



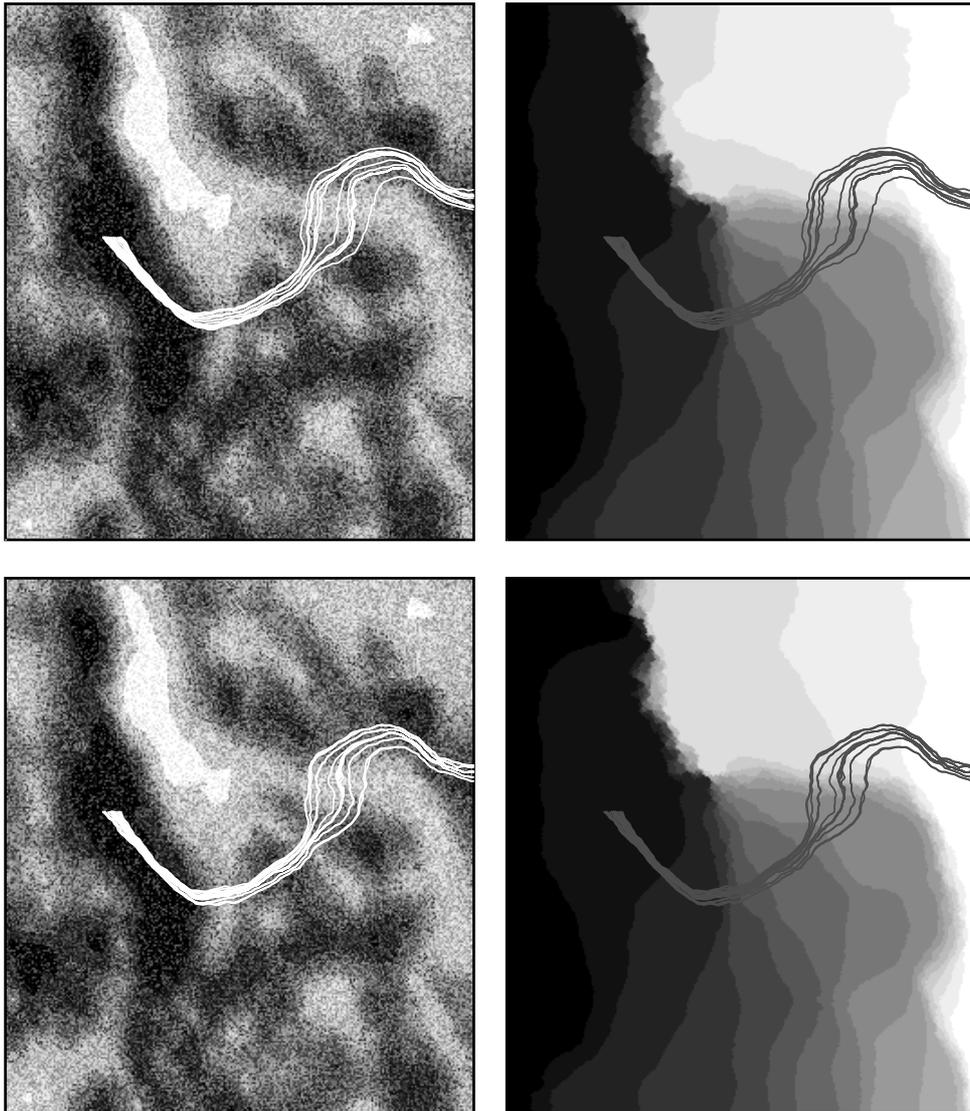
Cas du scénario 3. A gauche, deux champs de transmissivité générés par simulation par champ de probabilités. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-40



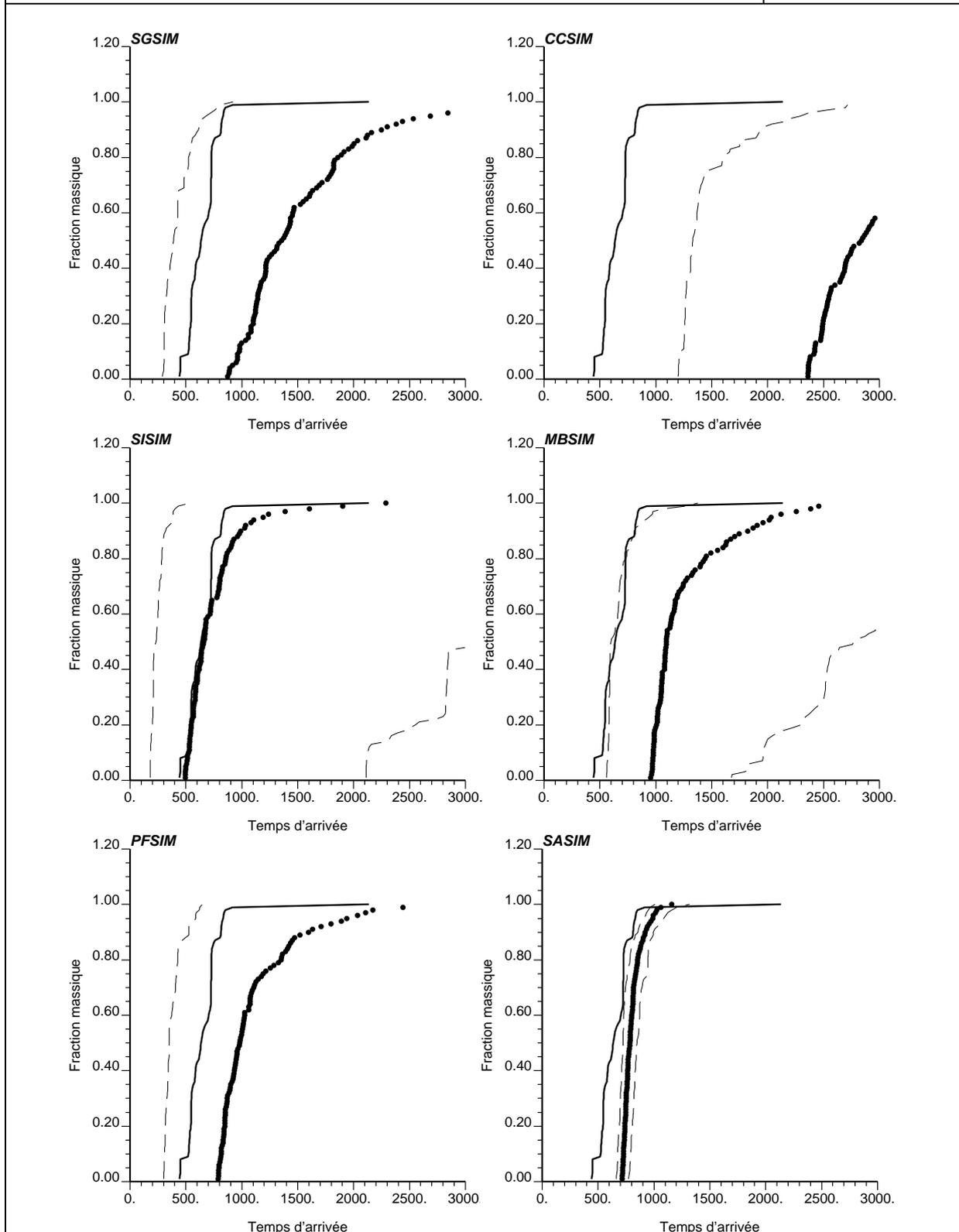
Cas du scénario 3. A gauche, deux champs de transmissivité générés par simulation recuit simulé. A droite, les champs piézométriques correspondants. Sur tous les champs, sont représentées les trajectoires de 20 particules.

Figure 2-41



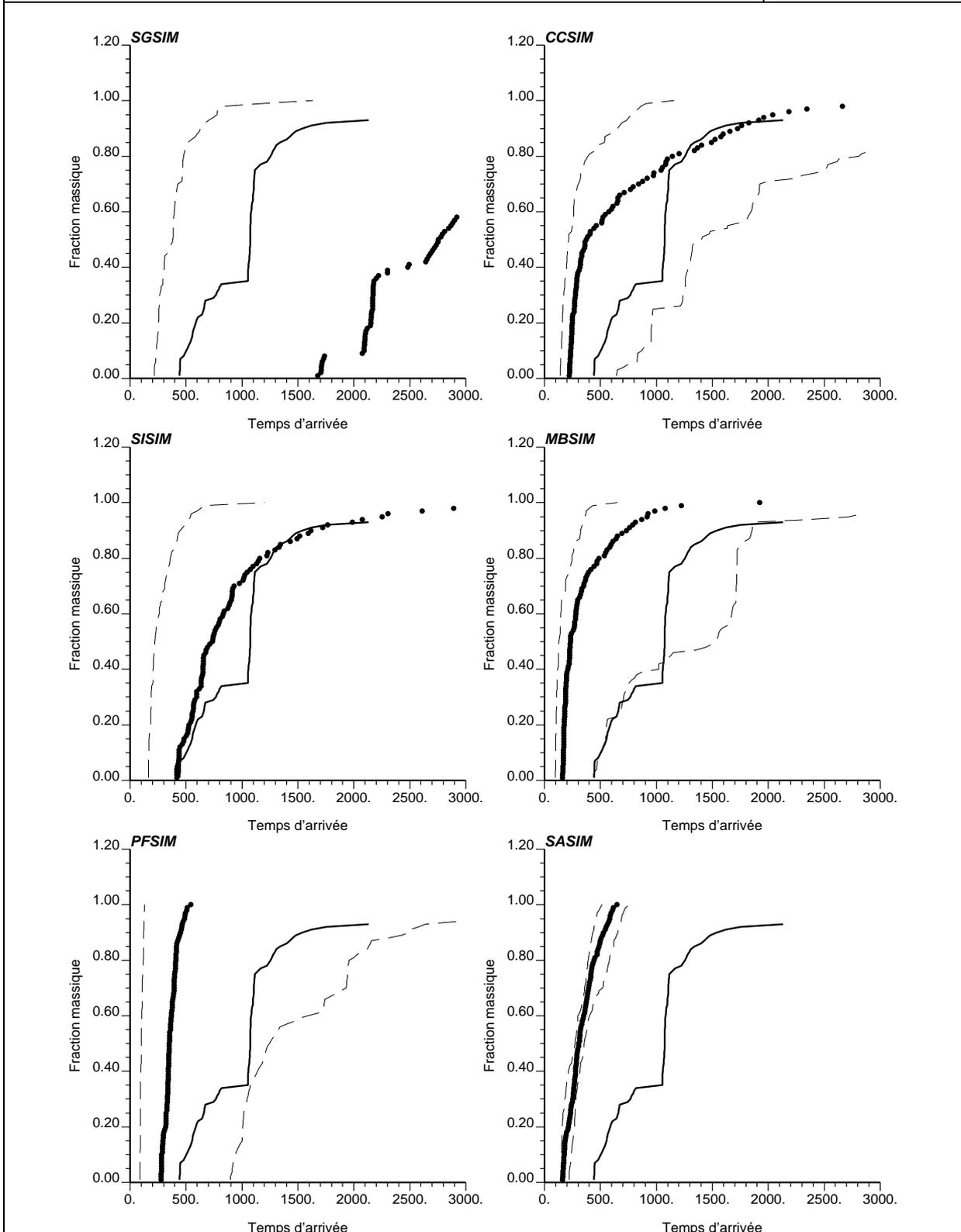
Courbes des temps d'arrivée, sur la frontière Est, pour chaque méthode pour le scénario 1. La courbe en trait continu correspond au champ de transmissivité de référence, la courbe en pointillés aux temps moyens sur l'ensemble des réalisations et la courbe en tirets les temps extrêmes pour des probabilités de 5 et 90% (cf. note 1, p 112).

Figure 2-42



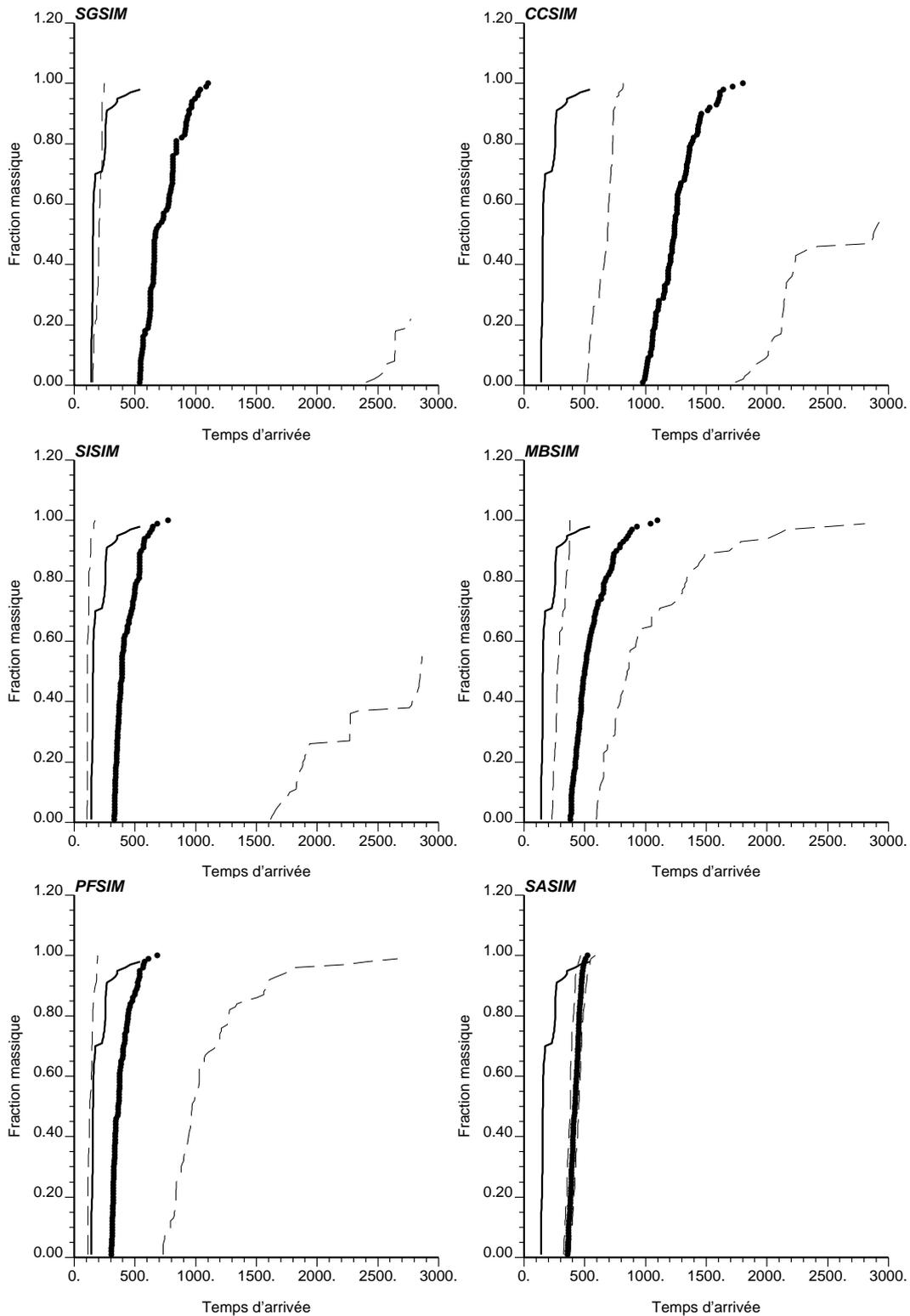
Courbes des temps d'arrivée, sur la frontière Est, pour chaque méthode pour le scénario 2. La courbe en trait continu correspond au champ de transmissivité de référence, la courbe en pointillés aux temps moyens sur l'ensemble des réalisations et la courbe en tirets les temps extrêmes pour des probabilités de 5 et 90% (cf. note 1, p 112).

Figure 2-43



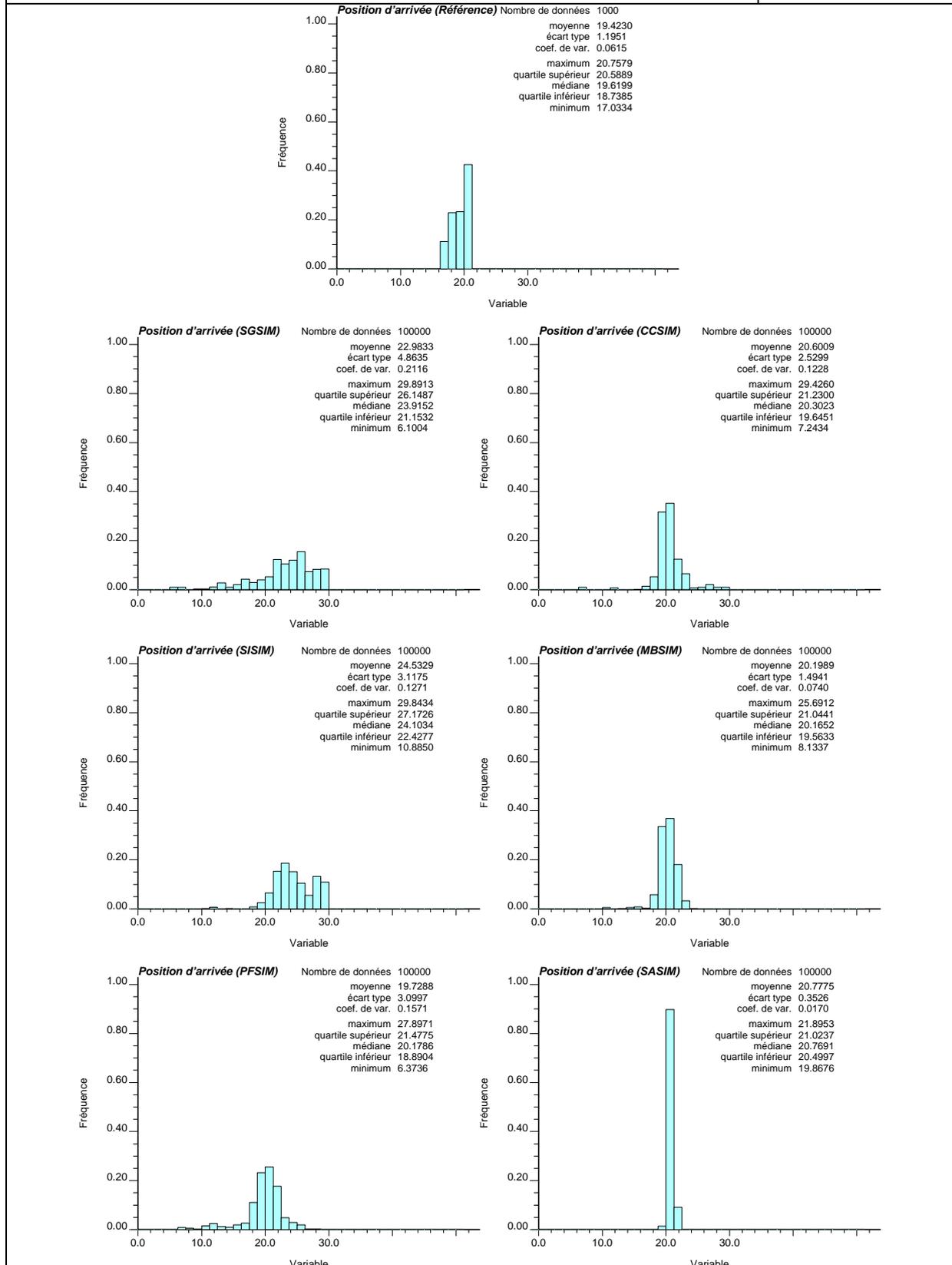
Courbes des temps d'arrivée, sur la frontière Est, pour chaque méthode pour le scénario 3. La courbe en trait continu correspond au champ de transmissivité de référence, la courbe en pointillés aux temps moyens sur l'ensemble des réalisations et la courbe en tirets les temps extrêmes pour des probabilités de 5 et 90% (cf. note 1, p 112).

Figure 2-44



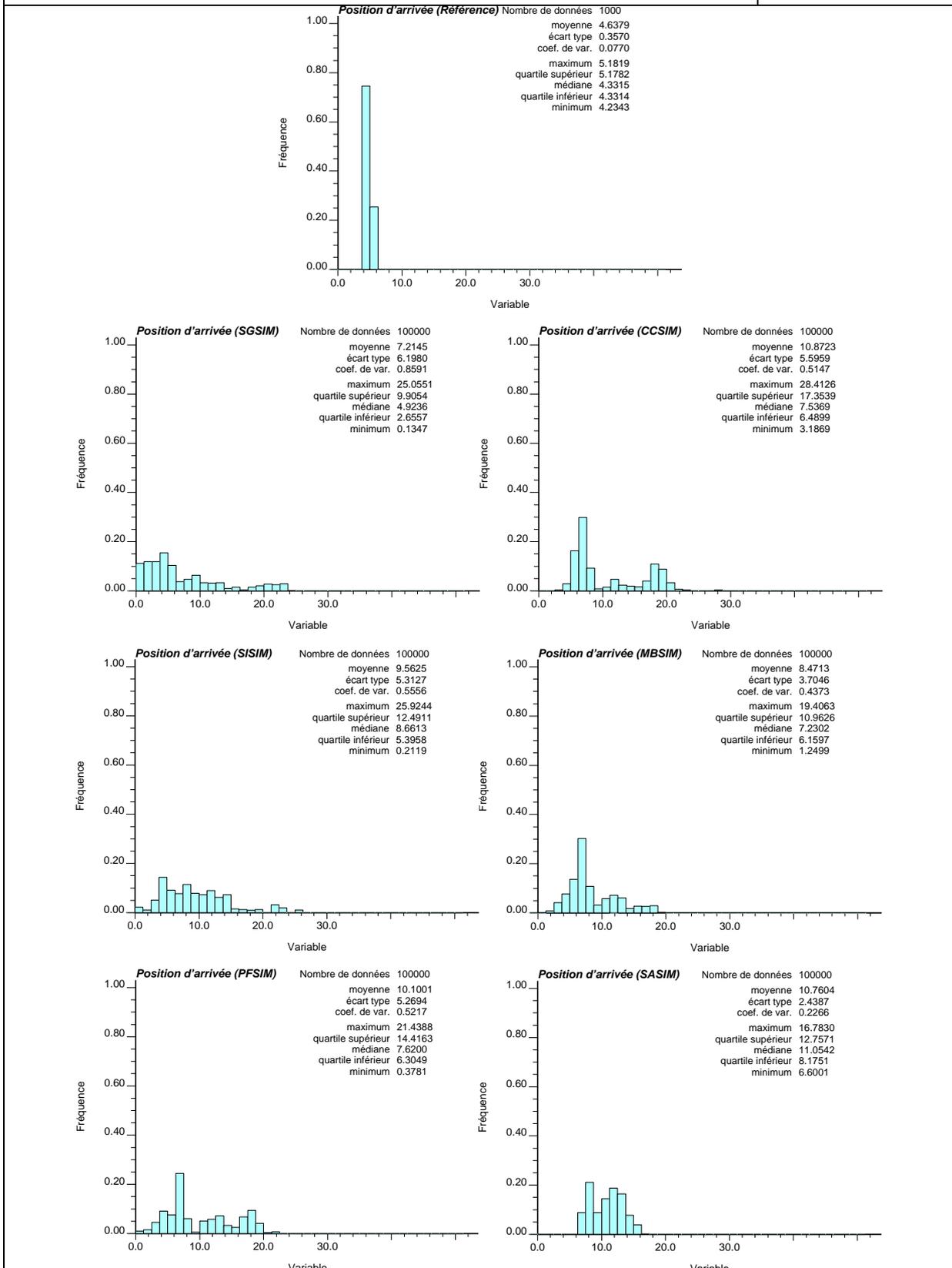
Histogrammes des positions d'arrivée, sur la frontière Est, pour le champ de transmissivité de référence et pour les 100 réalisations générées par chaque méthode de simulation pour le scénario 1.

Figure 2-45



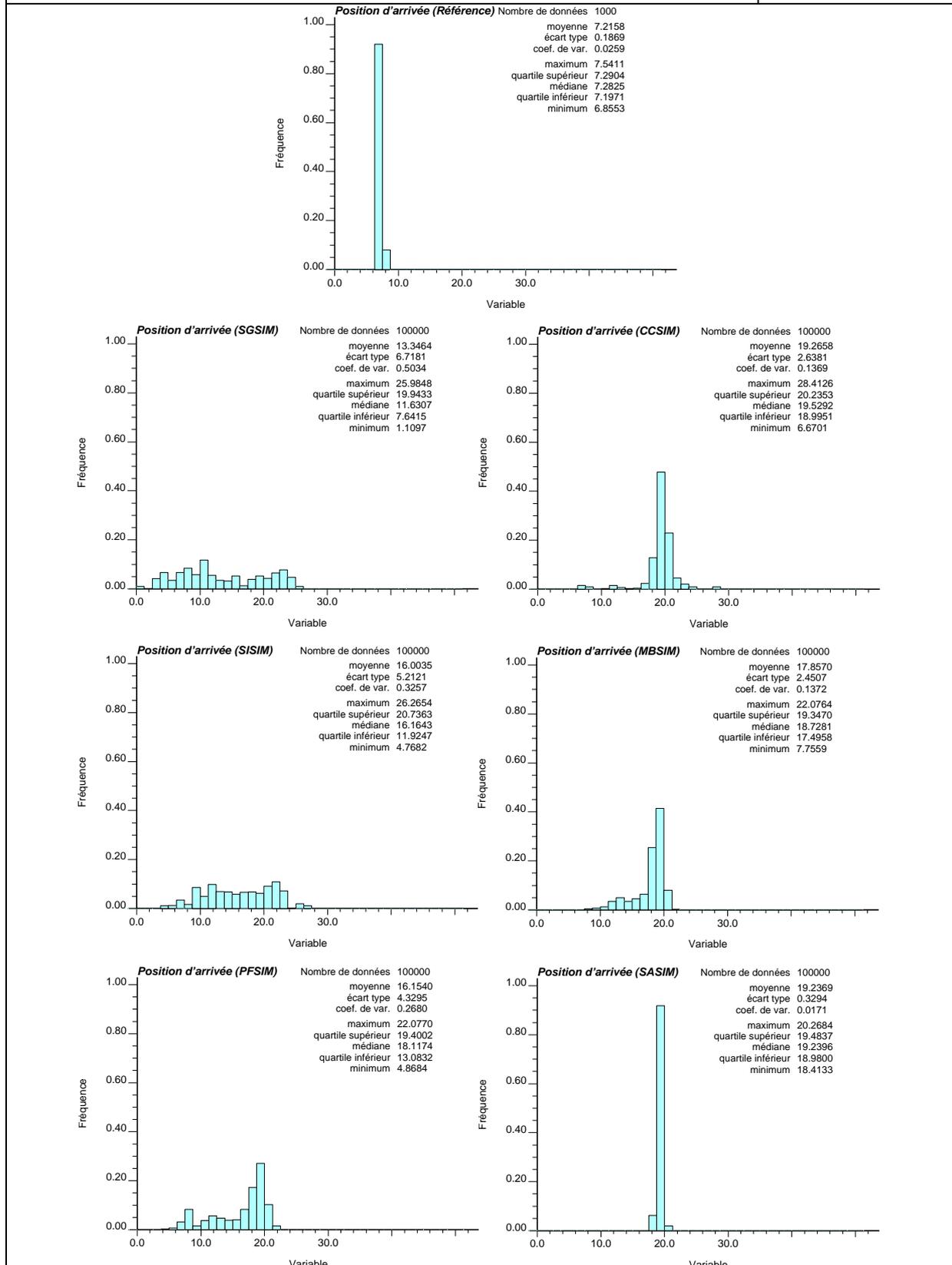
Histogrammes des positions d'arrivée, sur la frontière Est, pour le champ de transmissivité de référence et pour les 100 réalisations générées par chaque méthode de simulation pour le scénario 2.

Figure 2-46



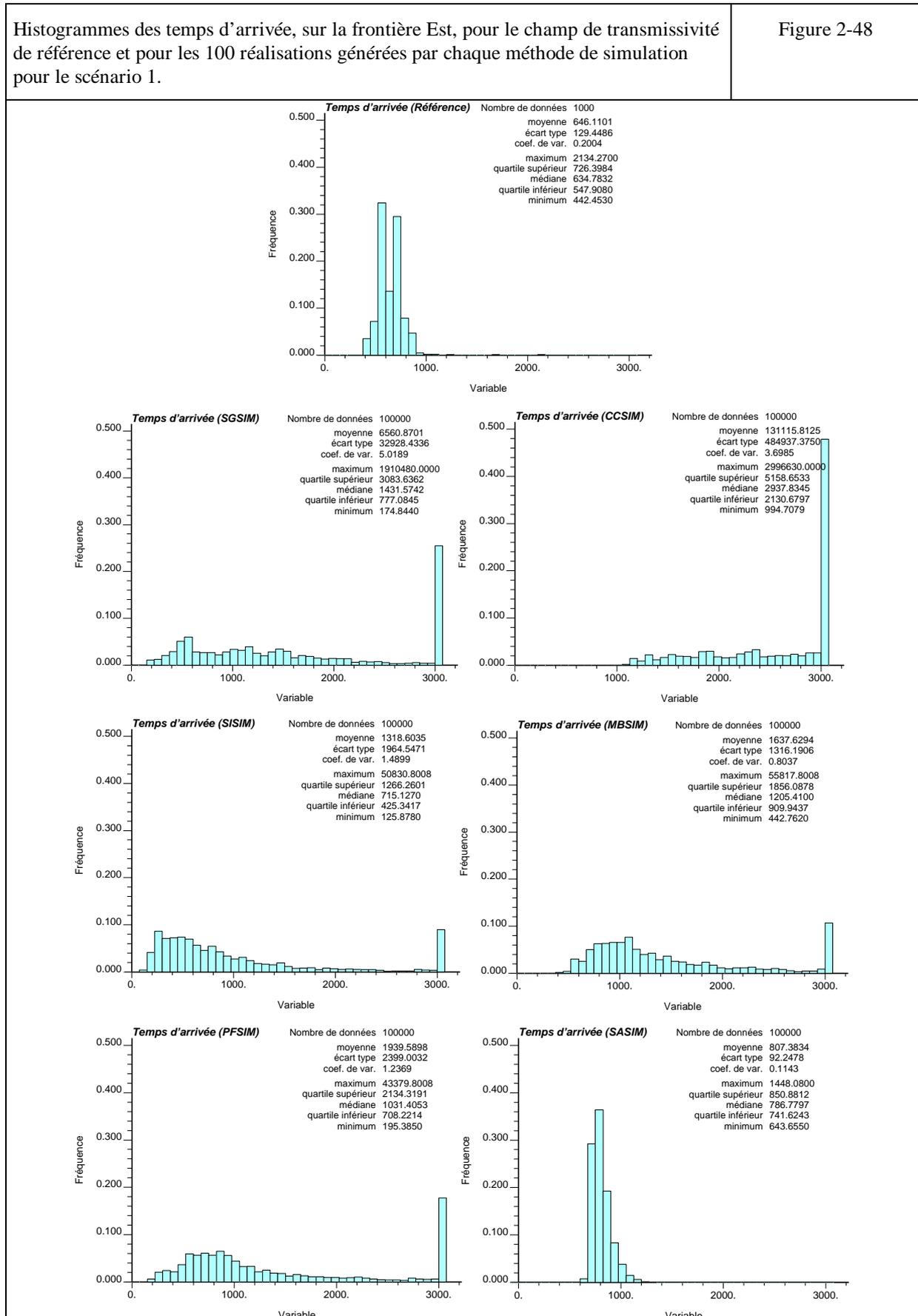
Histogrammes des positions d'arrivée, sur la frontière Est, pour le champ de transmissivité de référence et pour les 100 réalisations générées par chaque méthode de simulation pour le scénario 3.

Figure 2-47



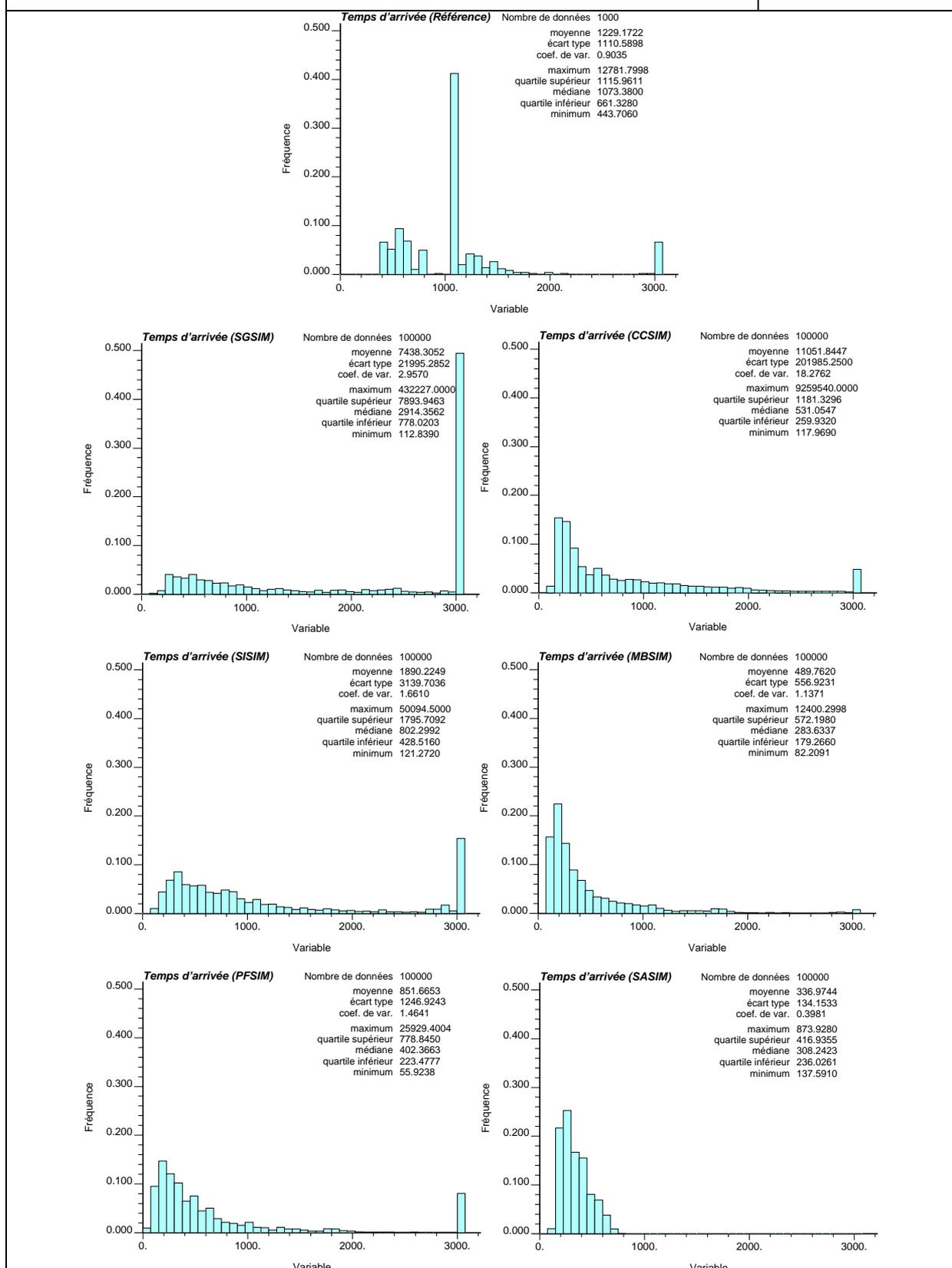
Histogrammes des temps d'arrivée, sur la frontière Est, pour le champ de transmissivité de référence et pour les 100 réalisations générées par chaque méthode de simulation pour le scénario 1.

Figure 2-48



Histogrammes des temps d'arrivée, sur la frontière Est, pour le champ de transmissivité de référence et pour les 100 réalisations générées par chaque méthode de simulation pour le scénario 2.

Figure 2-49



Histogrammes des temps d'arrivée, sur la frontière Est, pour le champ de transmissivité de référence et pour les 100 réalisations générées par chaque méthode de simulation pour le scénario 3.

Figure 2-50

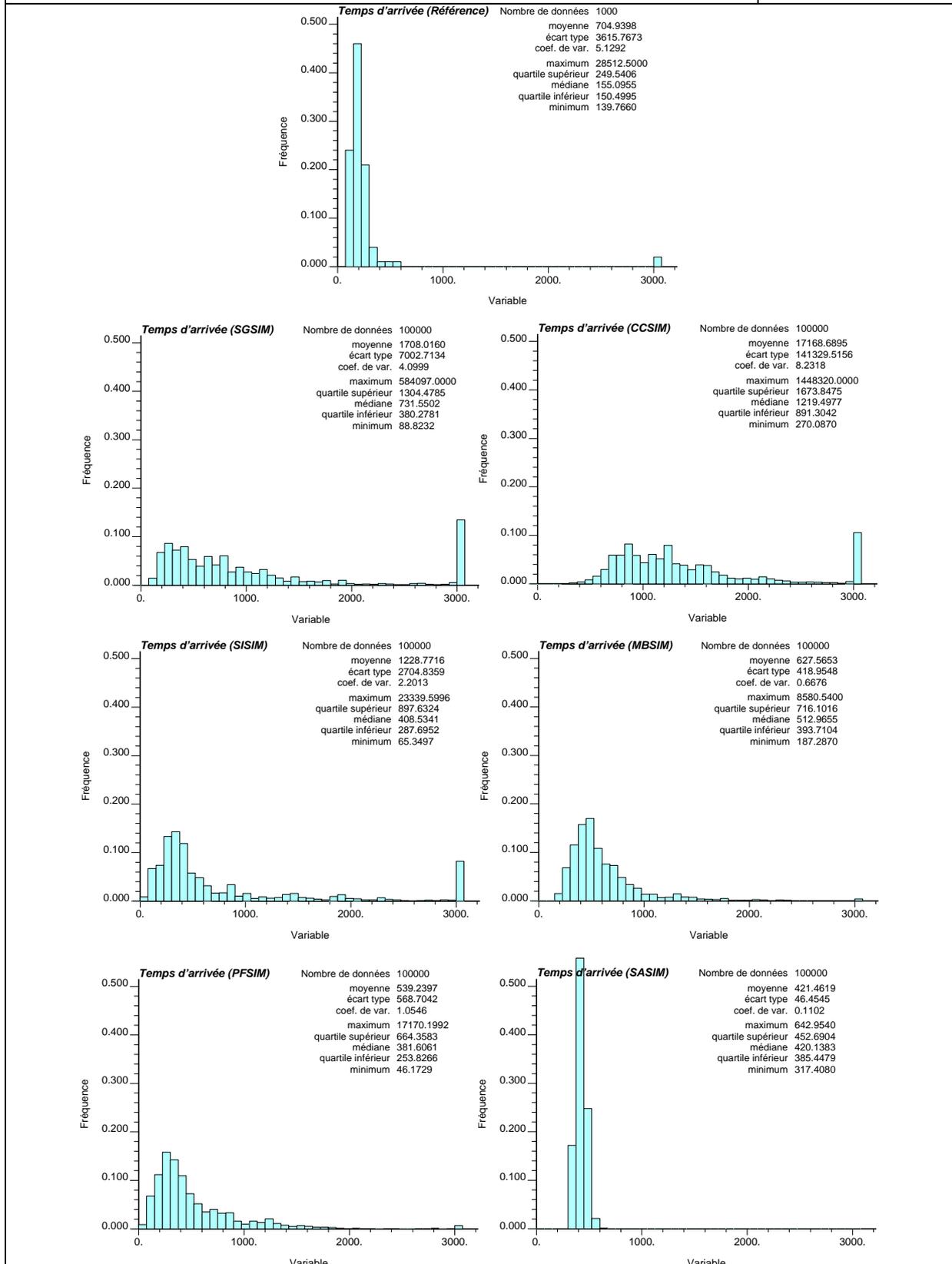


Tableau 2-6 : Statistiques sur les positions d'arrivée pour les trois scénarios et chacune des méthodes de simulation stochastique.  $N^\circ$  est le nombre de données,  $m$  la moyenne,  $\sigma$  l'écart-type,  $CV$  le coefficient de variation,  $max$  la valeur maximale,  $Q75$  le quartile supérieur,  $M$  la médiane,  $Q25$  le quartile inférieur et  $min$  la valeur minimale.

POSITION D'ARRIVEE							
Scénario n°1							
	Réf	SGSIM	CCSIM	SISIM	MBSIM	PFSIM	SASIM
$N^\circ$	1000	$10^5$	$10^5$	$10^5$	$10^5$	$10^5$	$10^5$
$m$	19.42	22.98	20.60	24.53	20.19	19.72	20.77
$\sigma$	1.195	4.863	2.529	3.117	1.494	3.099	0.352
$CV$	0.061	0.211	0.122	0.127	0.074	0.157	0.017
$max$	20.75	29.89	29.42	29.84	25.69	27.89	21.89
$Q75$	20.58	26.14	21.23	27.17	21.04	21.47	21.02
$M$	19.62	23.91	20.30	24.10	20.16	20.17	20.76
$Q25$	18.73	21.15	19.64	22.42	19.56	18.89	20.49
$min$	17.03	6.100	7.243	10.88	8.133	6.373	19.86
Scénario n°2							
	Réf	SGSIM	CCSIM	SISIM	MBSIM	PFSIM	SASIM
$N^\circ$	1000	$10^5$	$10^5$	$10^5$	$10^5$	$10^5$	$10^5$
$m$	4.637	7.214	10.87	9.562	8.471	10.10	10.76
$\sigma$	0.357	6.198	5.595	5.312	3.704	5.269	2.438
$CV$	0.077	0.859	0.514	0.555	0.437	0.521	0.226
$max$	5.181	25.05	28.41	25.92	19.40	21.43	16.78
$Q75$	5.178	9.905	17.35	12.49	10.96	14.41	12.75
$M$	4.331	4.923	7.536	8.661	7.230	7.620	11.05
$Q25$	4.331	2.655	6.489	5.395	6.159	6.304	8.175
$min$	4.234	0.134	3.186	0.211	1.249	0.378	6.600
Scénario n°3							
	Réf	SGSIM	CCSIM	SISIM	MBSIM	PFSIM	SASIM
$N^\circ$	1000	$10^5$	$10^5$	$10^5$	$10^5$	$10^5$	$10^5$
$m$	7.215	13.34	19.26	16.00	17.85	16.15	19.23
$\sigma$	0.186	6.718	2.638	5.212	2.450	4.329	0.329
$CV$	0.025	0.503	0.136	0.325	0.137	0.268	0.017
$max$	7.541	25.98	28.41	26.26	22.07	22.07	20.26
$Q75$	7.290	19.94	20.23	20.73	19.34	19.40	19.48
$M$	7.282	11.63	19.52	16.16	18.72	18.11	19.23
$Q25$	7.197	7.641	18.99	11.92	17.49	13.08	18.98
$min$	6.855	1.109	6.670	4.768	7.755	4.868	18.41

Tableau 2-7 : Statistiques sur les temps d'arrivée pour les trois scénarios et chacune des méthodes de simulation stochastique.  $N^\circ$  est le nombre de données,  $m$  la moyenne,  $\sigma$  l'écart-type,  $CV$  le coefficient de variation,  $max$  la valeur maximale,  $Q75$  le quartile supérieur,  $M$  la médiane,  $Q25$  le quartile inférieur et  $min$  la valeur minimale.

TEMPS D'ARRIVEE							
Scénario n°1							
	Réf	SGSIM	CCSIM	SISIM	MBSIM	PFSIM	SASIM
$N^\circ$	1000	$10^5$	$10^5$	$10^5$	$10^5$	$10^5$	$10^5$
$m$	646	6560	131115	1318	1637	1939	807
$\sigma$	129	32928	484937	1964	1316	2399	92
$CV$	0.20	5.01	3.69	1.48	0.80	1.23	0.11
$max$	2134	$1.9 \cdot 10^6$	$3.0 \cdot 10^6$	50830	55817	43379	1448
$Q75$	726	3083	5158	1266	1856	2134	850
$M$	634	1431	2937	715	1205	1031	786
$Q25$	547	777	2130	425	909	708	741
$min$	442	174	994	125	442	195	643
Scénario n°2							
	Réf	SGSIM	CCSIM	SISIM	MBSIM	PFSIM	SASIM
$N^\circ$	1000	$10^5$	$10^5$	$10^5$	$10^5$	$10^5$	$10^5$
$m$	1229	7438	11051	1890	489	851	336
$\sigma$	1110	21995	201985	3139	556	1246	134
$CV$	0.90	2.95	18.2	1.66	1.13	1.46	0.39
$max$	12781	432227	$9.2 \cdot 10^6$	50094	12400	25959	873
$Q75$	1115	7893	1181	1795	572	778	416
$M$	1073	2914	531	802	283	402	308
$Q25$	661	778	259	428	179	223	236
$min$	443	112	117	121	82	55	137
Scénario n°3							
	Réf	SGSIM	CCSIM	SISIM	MBSIM	PFSIM	SASIM
$N^\circ$	1000	$10^5$	$10^5$	$10^5$	$10^5$	$10^5$	$10^5$
$m$	704	1708	17168	1228	627	539	421
$\sigma$	3615	7002	141329	2704	418	568	46
$CV$	5.12	4.09	8.23	2.20	0.66	1.05	0.11
$max$	28512	584097	$1.4 \cdot 10^6$	23339	8580	17170	642
$Q75$	249	1304	1673	897	716	664	452
$M$	155	731	1219	408	512	381	420
$Q25$	150	380	891	287	393	253	385
$min$	139	88	270	65	187	46	317

## 2.5.2 Analyse des résultats

### 2.5.2.1 Simulation gaussienne séquentielle (SGSIM)

La simulation de la variable principale, avec un modèle multigaussien pour décrire sa dépendance spatiale, a été réalisée à partir de la seule information concernant la variable principale, soit les 30 données dures. Il est à noter, cependant, que la simulation gaussienne séquentielle peut être utilisée également pour simuler une variable principale unique en intégrant une ou plusieurs variables secondaires, ou pour simuler conjointement plusieurs variables principales (co-simulation). Dans le cas présent, l'objectif était d'évaluer, entre autres choses, l'apport d'une information secondaire sur la simulation d'un champ de transmissivité et par suite son incidence sur la modélisation de problèmes de transport de particules. Ceci explique la raison de cette simulation sans information secondaire dont les résultats seront comparés à ceux obtenus en intégrant de différentes manières cette information.

Deux réalisations du champ de transmissivité et les champs piézométriques obtenus dessus sont présentés à la Figure 2-24 (scénario 1), à la Figure 2-30 (scénario 2) et à la Figure 2-36 (scénario 3). Sur chaque image, les trajectoires de 20 particules ont été représentées également.

### 2.5.2.2 Simulation gaussienne séquentielle par cokrigage co-localisé suivant un modèle markovien (CC-SIM)

Cette simulation gaussienne de l'attribut principal a été réalisée en considérant, en plus des 30 données dures (méthode précédente), la donnée secondaire co-localisée au point de simulation. Étant donné que l'information secondaire n'est pas exhaustive, on utilise pour cette méthode le champ secondaire estimé à partir des 780 données secondaires, champ qui avait été créé et déjà utilisé pour le cokrigage co-localisé dans de l'exercice d'évaluation de méthodes d'estimation (voir §2.3.1).

Deux réalisations du champ de transmissivité et les champs piézométriques obtenus dessus sont présentés à la Figure 2-25 (scénario 1), à la Figure 2-31 (scénario 2) et à la Figure 2-37 (scénario 3). Si on compare ces figures à celles obtenues précédemment pour la simulation gaussienne séquentielle (Figure 2-24, Figure 2-30 et Figure 2-36), on constate l'influence importante de l'information secondaire. Les champs de transmissivité simulés avec une information secondaire ressemblent davantage au champ de référence (Figure 2-20), ce qui naturellement a des répercussions sur les champs piézométriques qui sont davantage conformes. Plus précisément, les caractéristiques à grande échelle sont mieux décrites lorsque l'information secondaire est prise en compte.

L'incertitude sur la prédiction des positions et des temps d'arrivée est moindre pour les scénarios 2 et 3 comme on peut l'observer sur les courbes de temps d'arrivée (Figure 2-42 à Figure 2-44). Si, en général, les deux méthodes donnent des courbes de temps moyens qui s'écartent dans le même sens par rapport à la courbe de référence (même tendance du biais), l'écart entre les courbes extrêmes (pour des probabilités de 5 et 90%), qui définissent l'enveloppe des temps, est beaucoup plus réduit lorsque l'information secondaire est prise en compte. Pour le scénario 1, l'intégration de l'information secondaire détériore la courbe des temps d'arrivée

des particules (Figure 2-42). Ces temps sont plus grands que ceux obtenus avec les autres méthodes, ce qui laisse supposer que les voies d'écoulement préférentielles sont moins bien reproduites dans le cas présent. Plus intéressantes encore sont les trajectoires des particules. Lorsque seule l'information principale est considérée, la position d'arrivée des particules peut varier considérablement d'une réalisation à l'autre du champ de transmissivité. Tel n'est pas le cas lorsque l'information secondaire est intégrée, ce qui conduit à une diminution considérable de l'intervalle de variation des positions d'arrivée. Ce phénomène s'apprécie clairement à partir des histogrammes de la position d'arrivée de la Figure 2-45 (scénario 1), de la Figure 2-46 (scénario 2) et de la Figure 2-47 (scénario 3), et du Tableau 2-6. Pour tous les scénarios et toutes les méthodes, l'incorporation de l'information secondaire réduit l'écart-type sur la position d'arrivée. A moindre échelle, cette tendance apparaît également avec les temps d'arrivée (Tableau 2-7). Il est à noter que pour le scénario 3, aucune des méthodes de simulation stochastique évaluées ne permet de reproduire la moyenne de la position d'arrivée.

### **2.5.2.3 Simulation séquentielle d'indicatrice (SISIM)**

La simulation séquentielle d'indicatrice intègre l'ensemble de l'information disponible. Les données douces y ont été codifiées sous la forme de données d'intervalle tirées du diagramme de dispersion des données de calibrage comme cela avait été déjà fait pour l'évaluation des méthodes d'estimation (voir §2.3.1). La Figure 2-26 (scénario 1), la Figure 2-32 (scénario 2) et la Figure 2-38 (scénario 3) illustrent la méthode avec deux réalisations du champ de transmissivité sur lesquelles ont été calculés le champ piézométrique et le transport de particules pour chacun des scénarios étudiés.

Cette méthode ne réussit pas non plus à reproduire correctement la variabilité spatiale à grande échelle de la transmissivité, ce qui affecte les trajectoires de particules et par conséquent les temps et les positions d'arrivée. Toutefois, la concordance des courbes de temps d'arrivée de référence et de temps d'arrivée moyens obtenus avec cette méthode est très bonne comme cela apparaît sur la Figure 2-42 à la Figure 2-44.

L'incertitude sur les temps d'arrivée, représentée par l'écart entre les courbes extrêmes, est supérieure à celle des autres méthodes. Cette tendance, qui existe également pour les positions d'arrivée, est confirmée par les histogrammes de la Figure 2-45 à la Figure 2-50, le Tableau 2-6 et le Tableau 2-7. On constate en particulier que l'écart-type sur la position d'arrivée est le plus élevé des méthodes qui incorporent l'information secondaire pour les scénarios 1 et 3 et parmi les plus grands pour le scénario 2.

### **2.5.2.4 Simulation séquentielle par cokrigage d'indicatrice suivant un modèle de Markov-Bayes (MBSIM)**

Comme cela a été dit à propos de l'estimation, le cokrigage d'indicatrice suivant un modèle de Markov-Bayes est la méthode qui permet une meilleure intégration de l'information secondaire pour un effort de modélisation et de calcul raisonnable. Deux réalisations générées par cette méthode, pour chacun des scénarios, sont présentées à la Figure 2-27 (scénario 1), la Figure 2-33 (scénario 2) et la Figure 2-39 (scénario 3) avec les champs piézométriques calculés à partir de ces réalisations et les trajectoires de 20 particules. On peut voir que la reproduction

du champ de transmissivité est très bonne, ce qui agit favorablement sur les trajectoires des particules et sur les temps d'arrivée. Les courbes extrêmes de la Figure 2-42 à la Figure 2-44, pour des probabilités de 5 et 90%, délimitent, autour de la courbe moyenne, un intervalle plus étroit que ceux vus précédemment. Le Tableau 2-6 montre également que les statistiques globales, sur l'ensemble des 100 réalisations, sont de celles qui se rapprochent le plus des statistiques de référence.

#### ***2.5.2.5 Simulation par champ de probabilités (PF-SIM)***

La simulation par champ de probabilités a été implémentée en utilisant comme fonctions de distribution locales conditionnelles par rapport à l'information disponible, celles estimées avec la méthode de cokrigage d'indicateur suivant un modèle de Markov-Bayes. La méthode se comporte relativement bien pour les temps d'arrivée moyens mais est moins satisfaisante lorsque l'on regarde les courbes de temps extrêmes (Figure 2-42 à Figure 2-44). Ce résultat moyen se vérifie également à partir des histogrammes et des statistiques sur les positions et les temps d'arrivée (Figure 2-45 à Figure 2-50, Tableau 2-6 et Tableau 2-7). L'attrait pour cette méthode se trouve dans ses performances de calcul qui en font la méthode la plus rapide en temps CPU.

#### ***2.5.2.6 Simulation par recuit simulé***

Une technique de recuit simulé a été utilisée pour générer des champs de transmissivité conditionnés par les 30 données dures, la forme de continuité spatiale (modèle de variogramme), le coefficient de corrélation entre les variables principale et secondaire et l'histogramme bivarié tiré des données de calibrage. Étant donné que le recuit simulé est une méthode d'optimisation et que l'information secondaire est très largement échantillonnée et présente une bonne corrélation avec l'attribut principal, les champs simulés sont très semblables entre eux, ce qui conduit à des courbes de temps extrêmes très proches de la courbe moyenne (Figure 2-42 à Figure 2-44). Les valeurs fortes des temps d'arrivée sont assez mal reproduites et considérablement sous-estimées.

## Conclusions sommaires sur les méthodes d'estimation et de simulation stochastique

Les chapitres 2.2 à 2.5 ont été consacrés aux méthodes de représentation statistique les plus employées actuellement pour intégrer différents types d'information. Les méthodes disponibles se classent en deux catégories selon que l'on cherche à produire une ou plusieurs représentations de l'attribut à modéliser. Il s'agit dans un cas de méthodes d'estimation et dans l'autre de méthodes de simulation stochastique.

Parmi les méthodes d'estimation, à l'exception du krigeage simple, ordinaire et universel, toutes permettent d'intégrer une ou plusieurs variables secondaires. La régression traditionnelle est la plus simple de ces méthodes mais elle reste un moyen très limité pour incorporer l'information qui provient d'une source secondaire. Elle se comporte convenablement quand l'information secondaire est disponible partout et se corrèle bien, de façon linéaire, avec la variable principale.

Le krigeage avec dérive externe permet d'inclure dans le champ estimé une tendance (à grande échelle) extraite d'une variable secondaire. Les valeurs de la variable secondaire n'interviennent pas, cependant, directement dans l'estimation de l'attribut principal. En outre, cette technique suppose implicitement que la forme de continuité spatiale de la variable secondaire est représentative de celle de la variable principale, équivalence qu'il est souvent difficile d'établir.

Le cokrigeage, dans toutes ses formes, fournit un estimateur dans lequel rentre séparément chaque type d'information. D'une part, la variable secondaire intervient directement dans le calcul de la variable principale estimée. D'autre part, il est possible de prendre en compte l'auto-corrélation (ou l'auto-covariance) de chaque variable et la corrélation croisée entre elles. Dans son implémentation traditionnelle, le coût de calcul du cokrigeage est élevé. Il existe cependant des variantes qui permettent de réduire le calcul des covariances nécessaires au système d'équations à résoudre en chaque point d'estimation. Le cokrigeage complet intègre l'ensemble des données disponibles situées à l'intérieur d'un certain voisinage autour du point d'estimation. Il n'en est pas de même du cokrigeage co-localisé qui ne prend en compte, comme information secondaire, que la donnée située au point d'estimation, ce qui réduit considérablement la taille du système d'équations à résoudre et le nombre de covariances à calculer. S'il est fait de plus l'hypothèse d'un modèle de type markovien, seule la covariance de la variable principale est nécessaire à partir de laquelle se déduisent simplement celle de la variable secondaire et la covariance croisée. Il s'en suit un effort réduit de modélisation des covariances en plus du gain sur le temps de calcul de l'estimation.

L'introduction de la fonction indicatrice permet de codifier, de manière simple et générale, différents types de données. Au moyen du krigeage d'indicatrice, il peut être déduit non pas une unique valeur estimée mais un modèle d'incertitude complet qui fournit la fonction cumulative de distribution locale de l'attribut étudié. A partir de ces distributions locales, il peut être choisi comme estimateur de l'attribut la moyenne, la médiane ou tout autre quantile.

Le modèle de Markov-Bayes, appliqué aux variables indicatrices, fournit un estimateur par cokrigeage capable d'intégrer une information secondaire quelconque. L'hypothèse de type

markovien, discutée précédemment, permet de réduire le nombre de modèles de covariance nécessaires, de telle sorte que, en pratique, le nombre de covariances à calculer est le même que pour le krigeage d'indicateur standard. En revanche, le cokrigeage complet offre la possibilité de séparer les formes de continuité spatiale des différents types d'information, ce qui permet d'améliorer, en fonction de l'information avoisinante, les fonctions de distribution locales a priori.

Toutes les méthodes d'interpolation qui fournissent un modèle d'incertitude (fonctions de distribution locales) peuvent être utilisées pour la simulation stochastique. Il s'agit seulement de générer diverses réalisations à partir des fonctions de distribution calculées en chaque nœud d'une grille. La simulation séquentielle se révèle être une des méthodes de simulation les plus souples et performantes.

L'hypothèse d'un modèle multigaussien, pour représenter la distribution d'un attribut modélisé, conduit à la simulation gaussienne séquentielle. Comme cela a été montré, cette méthode de simulation est la plus simple compte tenu de la simplicité de la fonction de distribution utilisée pour décrire les variables en jeu. En adoptant un estimateur de cokrigeage co-localisé avec un modèle de Markov, la méthode se simplifie encore davantage. La simulation gaussienne d'une seule variable s'étend facilement à plusieurs, avantage qui en fait une méthode particulièrement appropriée pour traiter les problèmes d'incorporation d'attributs secondaires multiples.

Pour une approche non paramétrique, le krigeage d'indicateur comme le cokrigeage co-localisé suivant un modèle de Markov-Bayes peuvent être employés pour la simulation.

La simulation par champ de probabilités sépare les tâches de construction des modèles d'incertitude et de génération des réalisations équiprobables. Les performances en temps de calcul de cette technique la rendent très attractive dès lors que l'information disponible est suffisante pour pouvoir estimer convenablement le modèle d'incertitude en tout point de simulation.

La simulation par recuit simulé est une autre technique capable d'incorporer une information secondaire. La condition à cela est que l'information puisse s'exprimer comme la composante d'une fonction objectif à minimiser.

Sur la base des résultats obtenus dans l'exemple traité par simulation stochastique, il peut être conclu que les méthodes qui se comportent le mieux sont la simulation par cokrigeage d'indicateur (MBSIM) et la simulation par champ de probabilités (PFSIM). La supériorité de ces méthodes se vérifie généralement en pratique sur d'autres cas d'étude.

**Note :** Tous les programmes d'estimation et de simulation utilisés dans les deux exercices d'évaluation proviennent de la librairie GSLIB (Deutsch et Journel, 1992a).

Il est à noter également que cette présentation de méthodes d'interpolation et de simulation n'avait pas pour ambition de couvrir l'ensemble des méthodes existantes mais de s'intéresser à celles qui présentent un intérêt potentiel pour la modélisation d'aquifères ou qui ont une utilité pédagogique. Parmi les méthodes qui n'ont pas été abordées, certaines se fondent sur une approche par krigeage. On peut citer par exemple le krigeage disjonctif en tant que méthode d'interpolation complexe qui, fondée en partie sur des distributions de type multigaussien, pré-

sente de nombreux défauts (voir l'analyse détaillée de Gómez-Hernández, 1990). Une telle méthode, qui n'apporte aucune amélioration valable par rapport aux approches par krigeage d'indicatrice, est au demeurant très peu employée dans la pratique.

## Références bibliographiques

1. Alabert, F. G. (1987). Stochastic imaging of spatial distributions using hard and soft data. Master's thesis, Stanford University, Branner Earth Sciences Library.
2. Almeida, A. S. (1993a). Joint simulation of multiple variables with a Markov-type coregionalization model. Dans Stanford Center of Reservoir Forecasting (SCRF) report 6, May 1993.
3. Almeida, A. S. (1993b). *Joint simulation of multiple variables with a Markov-type coregionalization model*. Ph. D. dissertation, Stanford University.
4. Almeida, A. S. et Journel, A. G. (1994). Joint simulation of multiple variables with a Markov-type coregionalization model. *Mathematical Geology*, 26(5), 565-588.
5. Bourgault, G., Journel, A., Lesch, S. M., Rhodes, J. D. et Corwin, D. L. (1995). Geostatistical analysis of a soil salinity data set. Dans *Application of GIS to the Modeling of Non-Point Source Pollutants in the Vadose Zone*, Proceedings of the 1995 Bouyoucos Conference, May 1-3, Riverside, CA.
6. Bourgault, G. et Journel, A. (1995). Unsmoothed estimation with dense secondary information using probability field technique. Dans *Computer Applications in the Mineral Industry*, Hani Sabri Mitri de., CAMI'95, Montreal, october 22-25, 95-102.
7. Box, G. E. P. et Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Addison-Wesley Publishing Company.
8. Brandstetter, A. et Buxton, B. E. (1987). The role of geostatistical, sensitivity, and uncertainty analysis in performance assessment. Dans Proceedings of the Conference on *Geostatistical, sensitivity, and uncertainty methods for ground-water flow and radionuclide transport modeling*, Buxton, B. E. (ed.), San Francisco, California, September 15-17, 89-110.
9. Cassiraga, E. F. y Gómez Hernández, J. J. (1994). Relación entre parámetros geofísicos e hidrogeológicos. Una revisión de literatura. Publicación técnica 11/94 ENRESA.
10. Dagan, G. (1986). Stochastic theory of groundwater flow and transport: Pore to laboratory, laboratory to formation, and formation to regional scale. *Water Resources Research*, 22(9), 120S-134S.
11. Dagan, G. (1989). *Flow and Transport in Porous Formations*. Springer-Verlag, New York.
12. de Fouquet, C., Beucher, H., Galli, A. et Ravenne, C. (1989). Conditional simulation of random sets – Application to an argillaceous sandstone reservoir, dans *Geostatistics*, Armstrong, M. (ed.), Kluwer Academic Publ., vol. 2, 517-530.
13. Deutsch, C. V. (1991). The relationship between universal kriging, kriging with an external drift, and cokriging. Dans Stanford Center of Reservoir Forecasting (SCRF) report 4, May 1991.
14. Deutsch, C. V. (1992). *Annealing techniques applied to reservoir modeling and the integration of geological and engineering (well test) data*. Ph. D. dissertation, Stanford University.
15. Deutsch, C. V. et Journel, A. G. (1991). The application of simulated annealing to stochastic reservoir modeling. Dans Stanford Center of Reservoir Forecasting (SCRF) report 4, May 1991.
16. Deutsch, C. V. et Journel, A. G. (1992). Annealing technics applied to the integration of geological and engineering data. Dans Stanford Center of Reservoir Forecasting (SCRF) report 5, May 1992.

17. Deutsch, C. V. et Cockerham, P. W. (1994). Practical considerations in the application of simulated annealing to stochastic simulation. *Mathematical Geology*, 26(1), 67-82.
18. Dowd, P. A. (1992). A review of recent developments in geostatistics. *Computers & Geosciences*, 17 (10), 1481-1500.
19. Eisenberg, N. A., Rickertsen, L. D. et Voss, C. (1987). Performance assessment, site characterization, and sensitivity and uncertainty methods: Their necessary association for licensing. Dans Proceedings of the Conference on *Geostatistical, sensitivity, and uncertainty methods for ground-water flow and radionuclide transport modeling*, Buxton, B. E. (ed.), San Francisco, California, September 15-17, 9-38.
20. Freeze, R. A. (1975). A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. *Water Resources Research*, 11(5), 725-741.
21. Freeze, R. A., Massmann, J., Smith, J. L., Sperling, T. et James, B. R. (1990). Hydrogeological decision analysis, 1, A framework. *Ground Water*, 28 (5), 738-766.
22. Freeze, R. A., James, B. R., Massmann, J., Sperling, T. et Smith, J. L. (1992). Hydrogeological decision analysis, 4, The concept of data worth and its use in the development of site investigation strategies. *Ground Water*, 30 (4), 574-588.
23. Freeze, R. A. et McWhorter, D. B. (1997). A framework for assessing risk reduction due to DNAPL mass removal from low-permeability soils. *Ground Water*, 35 (1), 111-123.
24. Froidevaux, R. (1992). Probability field simulation. Dans *Geostatistics Troia'92*, proceedings of the Fourth Geostatistics Congress, A. Soares (ed.), Kluwer Academic Publishers, Dordrecht, Holland.
25. Gelhar, L. W. (1984). Stochastic analysis of flow in heterogeneous media. Dans *Fundamentals of Transport Phenomena in Porous Media*, J. Bear and M. Y. Corapcioglu (eds.), 675-720.
26. Gelhar, L. W. (1993). *Stochastic Subsurface Hydrology*. Prentice-Hall, Inc., Englewood Cliffs.
27. Gómez-Hernández, J. J. (1990). Disjunctive kriging: the untold story, Technical report, Stanford University, 68 p.
28. Gómez-Hernández, J. J. (1991). Geoestadística para el análisis de riesgos. Una introducción a la geoestadística no paramétrica. Publicación técnica 04/91 ENRESA.
29. Gómez-Hernández, J. J. (1992). Descripción de la base de datos de Walker Lake. Publicación técnica 13/92 ENRESA.
30. Gómez-Hernández, J. J. (1993a). Programas complementarios para el análisis estocástico del transporte de radionucleidos. Publicación técnica 08/93 ENRESA.
31. Gómez-Hernández, J. J. (1993b). Métodos para la evaluación de la incertidumbre espacial. En *Primeras Jornadas de I + D en la Gestión de Residuos Radiactivos*, 21-23 abril, 1993, tomo II, publicado por ENRESA, 145-157.
32. Gómez-Hernández, J. J. (1997). Uncertainty in groundwater modeling. Dans proceedings of *Precision Agricultural: spatial and temporal variability of environmental quality*. Wiley, Chichester (Ciba Foundation Symposium 210), Wageningen, Netherland, 221-230.
33. Gómez-Hernández, J. J. et Journel, A. J. (1993). Joint sequential simulation of multiGaussian fields. Dans *Geostatistics Troia '92*, A. Soares (ed.), Kluwer Academic Publ., vol. 1, 85-94.
34. Gómez-Hernández, J. J. et Wen, X.-H. (1994). Probabilistic assessment of travel times i, groundwater modeling. *Stochastic Hydrology and Hydraulics*, 8 (1), 19-55.
35. Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press.

36. Isaaks, E. et Srivastava, R. (1989). *An introduction to applied geostatistics*. Oxford University Press, New York.
37. Istok, J. D. et Rautman, C. A. (1996). Probabilistic assessment of ground-water contamination: 2. Results of case study. *Ground Water*, 34 (6), 1050-1064.
38. James, B. R. et Freeze, R. A. (1993). The worth of data in predicting aquitard continuity in hydrogeological design. *Water Resources Research*, 29(7), 2049-2065.
39. Journel, A. G. (1983). Non-parametric estimation of spatial distributions. *Mathematical Geology*, 15( 3), 445-468.
40. Journel, A. G. (1986). Constrained interpolation and qualitative information. The soft kriging approach. *Mathematical Geology*, 18( 3), 269-286.
41. Journel, A. G. (1989). *Fundamentals of geostatistics in five lessons*. Short course in geology, vol. 8, American Geophysical Union.
42. Journel, A. G. et Huijbregts, C. J. (1978). *Mining Geostatistics*, Academic Press.
43. Journel, A. G. et Rossi, M. E. (1989). When do we need a trend model?. *Mathematical Geology*, 21( 7), 715-739.
44. Journel, A. G. et Zhu, H. (1990). Integrating soft seismic data: Markov-Bayes updating, an alternative to cokriging and traditional regression. Dans Stanford Center of Reservoir Forecasting (SCRF) report 3, May 1990.
45. Journel, A. G., Xu, W. et Tran, T. (1992). Integrating seismic data in reservoir modeling: the collocated cokriging alternative. Dans Stanford Center of Reservoir Forecasting (SCRF) report 5, May 1992.
46. Massmann, J. et Freeze, R. A. (1987a). Groundwater contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy, 1, Methodology. *Water Resources Research*, 23(2), 351-367.
47. Massmann, J. et Freeze, R. A. (1987b). Groundwater contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy, 2, Results. *Water Resources Research*, 23(2), 368-380.
48. Massmann, J. et Freeze, R. A. (1989). Updating random hydraulic conductivity fields: A two-step procedure. *Water Resources Research*, 25(7), 1763-1765.
49. Massmann, J., Freeze, R. A., Smith, J. L., Sperling, T. et James, B. R. (1991). Hydrogeological decision analysis, 2, Applications to groundwater contamination. *Ground Water*, 29 (4), 536-548.
50. Matheron, G. (1973). The intrinsic random function and their applications. *Adv. Appl. Prob.*, 5, 439-468.
51. Matheron, G., Beucher, H., de Fouquet, C., Galli, A. Guerillot, D. et Ravenne, C. (1987). Conditional simulation of the geometry of fluvio-deltaic reservoirs, *SPE paper # 16753*.
52. Myers, D. E. (1982). Matrix formulation of co-kriging. *Mathematical Geology*, 14(3), 249- 257.
53. Myers, D. E. (1984). Cokriging-New developments. Dans *Geostatistics for Natural Resources Characterization*, G. Verly et al. (eds.), 295-305, NATO ASI Series, Reidel, Dordrecht, Holland.
54. Myers, D. E. (1985). Cokriging: Methods and alternatives. Dans *The Role of Data in Scientific Progress*, P. S. Glaeser (ed.), Elsevier Science, New York, 425-428.
55. Rautman, C. A. et Istok, J. D. (1996). Probabilistic assessment of ground-water contamination: 1. Geostatistical framework. *Ground Water*, 34 (5), 899-909.
56. Rosen, L. et Gustafson, G. (1996). A Bayesian Markov geostatistical model for estimation of hydrogeological properties. *Ground Water*, 34 (5), 865-875.

57. Sperling, T., Freeze, R. A., Massmann, J., Smith, J. L., et James, B. R. (1992). Hydrogeological decision analysis, 3, Applications to design of a groundwater control system at an open pit mine. *Ground Water*, 30 (2), 376-389.
58. Srivastava, R. M. (1992). Reservoir characterization with probability field simulation. *SPE paper # 24753* presented at the 1992 Annual Technical Conference of the Society of Petroleum engineers, SPE, Richardson, Texas.
59. Srivastava, R. M. (1994). The visualization of spatial uncertainty. Dans Stanford Center of Reservoir Forecasting (SCRF) report 7, May 1994 (only the abstract).
60. Verly, G. W. (1993). Sequential Gaussian cosimulation: A simulation method integrating several types of information. Dans *Geostatistics Troia '92*, A. Soares (ed.), Kluwer Academic Publ., vol. 1, 543-554.
61. Wagner, B. J. et Gorelick, S. M. (1987). Optimal groundwater quality management under parameter uncertainty. *Water Resources Research*, 23(7), 1162-1174.
62. Wagner, B. J. et Gorelick, S. M. (1989). Reliable aquifer remediation in the presence of spatial variable hydraulic conductivity: From data to design. *Water Resources Research*, 25(10), 2211-2225.
63. Xu, W., Tran, T., Srivastava, R. M. et Journel, A. (1992). Integrating seismic data in reservoir modeling: The collocated cokriging alternative. *SPE paper # 24742*.
64. Xu, W. et Journel, A. G. (1993), GTSIM: Gaussian truncated simulations of lithofacies. Dans Stanford Center for Reservoir Forecasting (SCRF), report 6, May 1993.
65. Zhu, H. (1991). *Modeling mixture of spatial distributions with integration of soft data*. Ph. D. dissertation, Stanford University, p 250.
66. Zhu, H. et Journel, A. (1992). Formatting and integrating soft data: stochastic imaging via the Markov-Bayes algorithm. Dans *Geostatistics Troia' 92*, A. Soares (ed.), Kluwer Academic Publ., 1-12.

**3. PROBLEME INVERSE :  
RESOLUTION PAR UNE APPROCHE GEOSTATISTIQUE**



### 3.1 Introduction

Pour simuler numériquement des écoulements en milieu poreux, plusieurs paramètres doivent être définis : la taille du domaine étudié (il sera fait dans la suite l'hypothèse que le domaine est bidimensionnel<sup>1</sup>), la répartition spatiale de la transmissivité à l'intérieur du domaine et les débits d'alimentation et de drainage de la nappe, sources ponctuelles ou réparties situées sur la superficie du domaine. Pour des écoulements transitoires, il est aussi nécessaire de connaître le coefficient d'emmagasinement, les conditions initiales et l'historique au cours du temps des débits d'alimentation et de drainage. Dans la pratique, quel que soit le paramètre considéré, le nombre de mesures disponibles reste limité et l'ingénieur hydrogéologue se voit confronté à un problème d'incertitude sur des paramètres estimés. Cette incertitude n'est pas sans conséquences importantes sur l'état et le comportement du système hydrogéologique à modéliser tant d'un point de vue de l'écoulement que du transport.

La caractérisation de la répartition spatiale de la transmissivité suscite un grand intérêt car l'incertitude sur un champ de transmissivité hétérogène est prépondérante dans l'évaluation de l'incertitude sur les prédictions de l'écoulement dans un aquifère. Il existe également un intérêt croissant dans l'utilisation de mesures piézométriques pour contraindre la modélisation de champs de transmissivité. La raison est que l'acquisition de données piézométriques est moins onéreuse que celle de données de transmissivité (ou de perméabilité) et que le nombre de données disponibles est souvent plus grand. Comme la relation entre transmissivité et charge hydraulique est non linéaire, l'utilisation de mesures piézométriques dans l'estimation de transmissivités n'est pas triviale. Des procédures de type essai-et-erreur ont été utilisées dans le passé pour produire des répartitions spatiales de la transmissivité qui reproduisent les transmissivités et les niveaux piézométriques observés. Des méthodes automatiques sont proposées depuis plusieurs dizaines d'années pour résoudre ce problème connu sous le nom de problème inverse d'écoulement en milieu poreux. La notion de problème inverse fait référence ici à l'estimation de propriétés telles que la transmissivité, présentes dans l'équation qui régit les écoulements en milieu poreux (équation de diffusivité), à partir de mesures sur des variables d'écoulements telles que la charge hydraulique.

Les méthodes inverses passées en revue dans ce chapitre sont regroupées par catégories comme cela est expliqué dans la suite. Le paragraphe §3.2 est consacré aux méthodes qui se fondent sur une linéarisation de l'équation de diffusivité de manière à établir une relation linéaire entre la charge hydraulique et la transmissivité et parfois les débits d'alimentation ou de drainage de la nappe. Les variables sont traitées selon une approche stochastique à l'aide d'un modèle de fonctions aléatoires qui détermine les caractéristiques probabilistes des champs de transmissivité, de charge hydraulique et éventuellement de débits d'alimentation et de drainage de la nappe. Ce modèle repose sur quelques paramètres<sup>2</sup> dont les valeurs sont estimées par une méthode de maximum de vraisemblance<sup>3</sup>. Une fois le modèle de fonctions aléatoires établi, les

---

<sup>1</sup> La plupart des méthodes décrites dans ce chapitre ont été développées en 2D. Les possibilités d'extension au 3D et les difficultés attendues seront discutées au cas par cas dans la suite.

<sup>2</sup> Ces paramètres sont ceux qui entrent dans les lois de probabilité affectées aux variables (par ex. moyenne et variance pour des lois normales), dans les modèles de covariance (par ex. la portée de corrélation) et dans d'autres caractéristiques structurales telles qu'une dérive externe dans le cas de variables non stationnaires.

<sup>3</sup> L'appellation "maximum de vraisemblance" (qui en anglais se dit "maximum likelihood") est donnée en statistiques à une catégorie bien connue de méthodes destinées à déterminer les paramètres d'une fonction que l'on

variables d'entrée et sortie du problème d'écoulement (charge hydraulique, transmissivité, débit) sont calculées en chaque point du domaine à l'aide de la théorie de l'estimation linéaire. Seront discutées dans la suite la méthode de cokrigage linéarisé de Kitanadis et coauteurs (Kitanidis et Vomvoris, 1983, Hoeksema et Kitanadis, 1984) et la méthode linéarisée semi-analytique de Dagan et Rubin (Dagan, 1985, Rubin et Dagan, 1987a, Rubin et Dagan, 1987b, Dagan et Rubin, 1988).

Le paragraphe §3.3 porte sur les méthodes qui, par une approche itérative, intègrent la relation non linéaire entre transmissivité et charge hydraulique pour mieux décrire le modèle de fonctions aléatoires. En dehors de leur nature itérative, ces méthodes partagent, avec celles décrites précédemment, la même philosophie : après avoir déterminé le modèle de fonctions aléatoires, les variables d'entrée et de sortie sont obtenues par estimation linéaire. Rentrent dans cette catégorie la méthode par transformée de Fourier rapide de Gutjahr et coauteurs (Gutjahr et Wilson, 1989, Gutjahr *et al.*, 1994) et la méthode itérative de type cokrigage de Yeh et coauteurs (Yeh *et al.*, 1995 ; Yeh *et al.*, 1996).

Le paragraphe §3.4 présente les méthodes qui prennent en compte la relation non linéaire entre charge hydraulique et transmissivité par résolution numérique de l'équation d'écoulement dans un processus itératif au cours duquel l'équation est résolue un grand nombre de fois. Ces méthodes se différencient entre elles par la manière de formuler le problème d'optimisation. Seront présentées la procédure de *zonation* de Carrera et Neuman (Carrera et Neuman, 1986 a, Carrera et Neuman, 1986b, Carrera et Neuman, 1986c), la méthode séquentielle auto-calibrée de Gómez-Hernández, Capilla et Sahuquillo (Gómez-Hernández *et al.*, 1997, Capilla *et al.*, 1995), la méthode de points pilotes de Marsily, LaVenue et coauteurs (LaVenue *et al.*, 1995, RamaRao *et al.*, 1995) et la méthode de Monte Carlo par chaîne de Markov d'Oliver et coauteurs (Oliver *et al.*, 1996a, Oliver *et al.*, 1996b).

Le paragraphe §3.5 décrit une méthode différente de toutes les précédentes et connue sous le nom de méthode de simulation par fractal de Grindrod et Impey (1991). Cette méthode consiste simplement à calibrer les conditions aux limites pour recaler les charges hydrauliques en fonction du champ de transmissivité. Les charges hydrauliques ne sont cependant pas utilisées pour le calcul des transmissivités.

---

cherche à caler sur des données en minimisant une certaine erreur. L'aspect probabiliste intervient dans des considérations quelque peu intuitives (non démontrées et non démontrables comme beaucoup de techniques en statistiques) entre deux notions de probabilité qui sont mises en correspondance. Etant donné une fonction paramétrique qui relie des variables mesurées expérimentalement, il s'agit de : 1) la probabilité, pour des paramètres choisis de la fonction, de décrire convenablement un ensemble de données, 2) à l'inverse, de la probabilité, pour un ensemble de données, de pouvoir exister compte tenu d'un choix de paramètres. La méthode des moindres carrés entre par exemple dans cette catégorie.

## 3.2 Linear methods

### 3.2.1 The linearized cokriging approach

#### Methodology

##### *Introduction*

The linearized cokriging approach, developed by Kitanidis and others (Kitanidis and Vomvoris, 1983; Hoeksema and Kitanidis, 1984), is the first one to use geostatistics in the inverse modeling of groundwater flow. The two seminal papers that present the theoretical basis of the method are discussed. In the first paper (Kitanidis and Vomvoris, 1983) the approach is applied for 1-D steady groundwater flow, and, in the second paper (Hoeksema and Kitanidis, 1984) an extension is made to 2-D steady groundwater flow. The inversion can be extended to the case in which sinks and sources are present. Boundary conditions may also be subject to calibration.

In the 70's an increasing amount of papers was devoted to the statistical treatment of the groundwater flow inverse problem. The linearized cokriging approach treats the inverse problem geostatistically; the random transmissivity field and the random hydraulic head field are related by a linearization of the steady groundwater flow equation. In addition, cokriging is used to estimate hydrogeologic parameters (transmissivities and possibly boundary conditions and/or leakage rates) on a grid which covers the aquifer.

##### *The steps in the calibration process*

In any geostatistical approach, the spatial variability of the parameters is mathematically described using a random function model. A random function is a set of inter-related random variables, each variable corresponding to a point in space. The variability of the aquifer parameters at small scales is rather erratic and it is neither possible nor useful to describe it in a deterministic way. It is more expedient to model the parameters in a probabilistic fashion. More details on the geostatistical approach and the random field representation can be found in numerous works (e.g. Chapter 2 of this report, Journel and Huybregts, 1978; Isaaks and Srivastava, 1989; Goovaerts, 1997). The geostatistical approach was applied before in other studies in groundwater hydrology, but not in inverse modeling. Kitanidis and Vomvoris (1983) state that the main advantage of their application of the geostatistical approach to inverse modeling is that it avoids the problem of large dimensionality by drastically reducing the number of independent parameters to be estimated.

#### *1. Identification of the structure of the parameter field - structure selection.*

A geostatistical model is adopted for the random fields of transmissivity, hydraulic head and eventually recharge. Adopting a geostatistical model means that decisions are made on the functional representation of the variogram (whether it is exponential, spherical, or otherwise), stationarity or non-stationarity and the drift (if a non-stationary model is selected). These decisions leave as only parameters to be determined those defining the variograms and those defining the constant mean (stationary model) or the drift function (non-stationary model). Kitanidis and Vomvoris (1983) state that the parameter identification

should be repeated for a number of random function models and possibly for a number of conceptual models on the system behavior (for instance whether recharge should be considered or not) until the most appropriate is found. Attention is usually restricted to models that have been chosen on the basis of their performance in past applications. The aim of the structure selection step is to choose a model in agreement with the data and as simple as possible (the parsimony principle). Kitanidis and Vomvoris state that model selection criteria like Akaike's criterion could serve as a guide, although the model selection has a multiobjective nature, although we have our doubts on this respect due to the strict underlying assumptions that are needed to come up with the information criterion. All available prior information should be taken into account in the selection of the most appropriate model for the structure of the hydrogeologic parameter field.

2. *Identification of the structure of the parameter field - maximum likelihood estimation of structural parameters.*

For the selected model the parameters defining the random function model for the transmissivity field, the hydraulic head field and eventually the recharge field are estimated simultaneously by maximum likelihood. The likelihood function is written in terms of the vector of mean values (for transmissivity, hydraulic head and possibly recharge, at each node in the domain) and the (cross) covariance matrices among these parameters. In turn, these values are given as a function of a few structural parameters, such as, constant means or drift coefficients (if a drift is assumed for the mean) or the correlation ranges for the covariances. It is assumed that the joint probability function of the measurements of all observable quantities (both parameters and state variables) is multi-Gaussian. The normality assumption is consistent with first-order analysis, where the hydraulic head is a linear function of multi-Gaussian log transmissivity. The hydraulic head mean values, the hydraulic head covariance and the cross-covariance log transmissivity-hydraulic head which appear in the likelihood expression are obtained from the transmissivity mean value and the transmissivity covariance by a first-order linearization of the groundwater flow equation. This first-order linearization is considered to be valid up to variances of  $\log_{10}$  transmissivity of 1.0. The linearized groundwater flow equation is solved by analytical (1-D) or numerical (2-D) methods. The structural parameters defining mean and covariance functions are estimated by maximum likelihood so that the probability that the given measurements would occur is maximized. Gauss-Newton optimization is used to estimate, in an iterative fashion, the structural parameters.

3. *Identification of the structure of the parameter field - model validation and diagnostic checking.*

The validity of the assumed geostatistical structure is tested by the following procedure. The error covariance matrix for the estimated geostatistical parameters is estimated. In addition, the significance of the parameter estimates and the consistency between model and data can be tested statistically. On the basis of these statistical tests the hypothesized structure is accepted or modified (Kitanidis and Vomvoris, 1983). It is unfortunate that all these tests have to rely on the prior hypothesis of joint multi-Gaussianity of the different parameters; hypothesis that cannot be verified and is stronger than the consistency that it is being tested.

#### 4. *Estimates of hydrogeological parameters by kriging.*

Once the physical model and the geostatistical structure have been decided upon and its structural parameters have been determined by maximum likelihood, point estimates of the hydrogeologic parameters can be obtained by cokriging. The estimation of log transmissivity at a certain location is a weighted sum of point measurements of log transmissivity and hydraulic head. The weights are obtained by solving a linear system of equations (cokriging system) so that the log transmissivity estimate is unbiased and the estimation error variance is minimum. In the estimation procedure the covariance matrices of transmissivity and hydraulic head and the cross covariance matrix of transmissivity-hydraulic head are used. These covariance matrices had been obtained using the first order approximation of the groundwater flow equation, as described in step 2.

#### *Numerical Experiments*

Kitanidis and Vomvoris present rather simple 1-D synthetic examples with steady state flow, without sinks and sources. A log transmissivity field is generated and hydraulic head is solved for this field. In addition, some transmissivity and head samples are taken from the reference fields. These measurements could be corrupted with white noise to simulate measurement error. It was found that the calibrated transmissivities vary in a very smooth way in space, especially if the measurements contain error. It is found that the final estimates are robust with respect to the geostatistical assumptions. Kitanidis and Vomvoris were optimistic about the usefulness of the inverse problem in groundwater flow modeling and they state that an advantage of the method is that smooth solutions are obtained which do not attempt to restore details which cannot be actually extracted from available data (Kitanidis and Vomvoris, 1983). Nowadays, this is regarded as a shortcoming since it has been shown that proper modelling of mass transport can only be accomplished in heterogeneous fields displaying the type of spatial variability observed in the field.

In a second paper Hoeksema and Kitanidis (1984) present a comparable 2-D synthetic study. Again, a log transmissivity field is generated and the hydraulic head is solved for this field. It is shown that head measurements improve the characterization of the log transmissivity field. The head measurements could be closely reproduced if the estimated log transmissivity field was used in a numerical flow model. The study also shows that adequate results could be obtained with a coarse discretization of the aquifer.

Hoeksema and Kitanidis (1989) apply the methodology to a hypothetical 2-D aquifer to predict not only transmissivities and hydraulic heads, but also seepage velocities. The aquifer is 80 km by 60 km and discretized into 40 by 30 square grid blocks of 5 km. Constant head values are prescribed in the boundaries. Transmissivities are generated over the grid according to an exponential variogram with range of 30 km and variance 0.6. The solution of the flow equation in the transmissivity field generated yields a reference head field. Then, both the transmissivity and the head fields are sampled at 15 locations. The transmissivity samples are corrupted with some noise. The linearized cokriging approach is applied for the estimation of the transmissivity and head fields using only transmissivity data and using both transmissivity and head data. The results show the reduction in estimation variance due to the use of head measurements in the inverse procedure.

### **Some important characteristics of the method**

The geostatistical approach for inverse modelling explicitly handles the uncertainty associated with spatially variable phenomena. In addition, the linearized cokriging approach circumvents the ill-posedness of the inverse problem by using just a few parameters in the inversion procedure, which are the parameters describing the geostatistical structure of the log transmissivity random function. Once these parameters are known the transmissivity and the heads can be determined on any grid by the use of the standard cokriging equations.

It should be noted that the linearized cokriging approach has a limited field of application. The (cross) covariance matrices and the estimated means for hydraulic head are obtained by a first-order linearization of the groundwater flow equation. This linearization is valid for  $\log_{10}$  transmissivity variances below 1.0, and in practice  $\log_{10}$  transmissivity variances above 1.0 are common. The use of the transmissivity field obtained by this approach as input to a groundwater flow solver (which does not linearize the groundwater flow equation) will yield a deviation between the measured and simulated heads. This deviation will grow with the magnitude of the log transmissivity variance.

As most of the methods revised in this Chapter, it was originally aimed at determining a unique single best solution to the inverse groundwater flow problem. As a by-product, the maximum likelihood approach provides a lower bound of the error covariance matrix that could be used for the generation of multiple solutions to the inverse groundwater flow problem; simulations of the errors can be made to be added to the «optimal» transmissivity field producing simulations of the transmissivity field around the optimum. However, since the covariance matrix is just an approximation of the real covariance matrix, and, in any case, the covariance only captures the linear relationship between two variables, the solution of the flow equation using the transmissivity fields generated in this way will be only approximations of the measured heads.

The linearized cokriging approach as presented by Kitanidis and co-workers seems not to be very flexible for extensions to transient groundwater flow, and for the simultaneous estimation of a different parameters, such as transmissivity, storativity or recharge. Sun and Yeh (1992) extended the linearized cokriging approach to transient groundwater flow, making use of adjoint state equations.

An advantage of the linearized cokriging method is that the CPU time usage is small compared to other methods.

More comments on the linearized cokriging approach can be found in section §3.4.1, in the discussion of the characteristics of the zonation procedure.

### **Application of the linearized cokriging approach to the Jordan aquifer in Iowa**

Hoeksema and Kitanidis (1984) present an application of the linearized cokriging approach to a real case study. They apply the linearized cokriging approach to the Jordan aquifer of Iowa, USA. The model is for 2-D steady state flow in a leaky confined aquifer and with prescribed head boundary conditions. A total of 31 steady-state head and 56 log transmissivity measurements are available. The case study illustrates the iterative approach in the model selection: the conceptual models considered are groundwater flow without leakage, groundwater flow with

leakage and groundwater flow with leakage, discarding two head measurements. This last model yields results that pass the multi-Gaussian model validation and diagnostic checking tests. The conclusion was that there is leakage and that two head measurements are unreliable.

### 3.2.2 The linearized semi-analytical approach

#### Methodology

##### *Introduction*

The work by Dagan and Rubin (Dagan, 1985; Rubin and Dagan, 1987a; Rubin and Dagan, 1987b; Dagan and Rubin, 1988) is closely linked to the work of Kitanidis and others. The approach which is used to calibrate the transmissivity field to hydraulic head data does not differ much from the linearized cokriging approach. A main difference is that, in the linearized semi-analytical approach, a Bayesian updating procedure is used instead of cokriging. Four papers illustrate the development of the linearized semi-analytical approach. The first paper (Dagan, 1985) presents the methodology in a general framework. In the second paper (Rubin and Dagan, 1987a), the method is extended for the estimation of a constant, but random, effective recharge after relaxing some of the assumptions. The third paper (Rubin and Dagan, 1987b) illustrates the approach using a case study. The fourth paper is an extension to incorporate transient head measurements and to the estimation of the storativity (Dagan and Rubin, 1988). Dagan and Rubin were the first to solve the inverse problem for time-dependent flow in an analytical framework, but Carrera and Neuman (1986a, 1986b) defined a numerically-based procedure to solve this problem some years before.

##### *The steps in the calibration process*

The steps to be taken in the linearized semi-analytical approach are similar to the ones which have to be taken in the linearized cokriging approach. However, the authors pay less attention to the first and last step regarding the structure selection and focus on the second step, the maximum likelihood procedure.

##### *1. Maximum likelihood estimation of structural parameters.*

The starting point is steady state or transient state flow with the presence of recharge in a 2-D domain at which some transmissivity measurements and hydraulic head measurements have been taken. The aim is to estimate a vector of structural parameters which describe the random transmissivity field jointly with the random head field. Generally these parameters are the average log transmissivity, the log transmissivity variance and the integral scale of transmissivity. In order to estimate the parameters a number of assumptions are made (Rubin and Dagan, 1987a):

- (i) The variance of  $\log_{10}$  transmissivity is smaller than one so that the groundwater flow equation can be linearized by a first-order small perturbation analysis.
- (ii) Log transmissivity is normally distributed with constant mean and isotropic exponential covariance.
- (iii) The average head shows a quadratic trend in space.

In case of transient groundwater flow the following additional assumptions are made (Dagan and Rubin, 1988):

- (iv) Storativity is constant in the spatial domain and recharge is constant in space, but variable in time.
- (v) The transient groundwater flow equation is split in two parts. One part corresponds to the flow field after the effects of the pumping wells are filtered out, and the other part is related exclusively to the pumping wells. The part which is related exclusively to the pumping wells is expressed by a generalization of the Theis elementary solutions for each well in an unbounded domain and for time depending pumping rate. The measured transmissivity or, if a transmissivity measurement at the well is not available, the effective transmissivity value is used in the generalization of the Theis elementary solution.
- (vi) The remaining part of the solution, with the effect of the pumping wells filtered out, is related to the recharge and to the boundary conditions. It is assumed that the expected head value has a slowly varying quadratic trend in time. In addition, a quasi-steady state representation is adopted, describing the transient groundwater flow as a series of (different) steady states in time. The average head component is related directly to the recharge and indirectly to the boundary conditions, while some other terms in which time derivatives appear are neglected.
- (vii) The residual head (the head minus the expected head after filtering out the effects of the pumping wells) is assumed to satisfy a Poisson equation and the time derivative is neglected for the same reasons as mentioned before.
- (viii) The effect of the presence of boundaries is taken into account in an indirect, approximate way. First, the justification for adopting the presentation of the head as a gradually, in space and time, varying function, results partly from the boundary conditions. Second, the covariances of the transmissivity and hydraulic head values at two different points are taken equal to those pertaining to an unbounded domain. This holds if the points are sufficiently far from the boundaries. In particular, it is considered that, in a band of one integral scale from the boundary, this assumption does not hold. In the work of Dagan (1985) an even more restrictive assumption was made: the flow domain was assumed infinite.

The linearization of the groundwater flow equation and the assumption that log transmissivity is normally distributed with constant mean and exponential covariance make it possible to derive, by a small perturbation approach (Green's function techniques), the head covariance and the cross covariance between transmissivity and head. Dagan (1985, section 3) made these derivations for a uniform head gradient and Rubin and Dagan (1987a, section 2) made derivations for the more general case of a quadratic trend in head values, so that also recharge can be identified. Finally, Dagan and Rubin (1988, section 5) derived the covariances for the case of time-dependent (quasi-steady) groundwater flow.

Once the head covariance and the cross-covariance head-log transmissivity are derived, the likelihood function can be written in terms of the structural parameters that will be subject to identification using a maximum likelihood approach.

2. *Some spatial invariant parameters can be estimated.*

The constant storativity, the effective recharge can be incorporated in the likelihood expression and estimated, together with their variances, during the maximum likelihood procedure. The effective recharge would be a function of time in case of transient groundwater flow. If recharge and storativity are the only quantities of interest the calculations can be concluded at this point.

3. *Estimates of hydrogeological parameters by Bayesian updating.*

After the structural parameters are determined, the values for the hydrogeological parameters and their estimation variances are by a Bayesian updating procedure. It is assumed that log transmissivity, hydraulic head and recharge are multivariate normal distributed. The updated expected transmissivity value (also referred to as the conditional expected transmissivity) is the sum of the unconditional transmissivity plus a weighted linear combination of transmissivity residuals and a weighted linear combination of head residuals. The weights are solutions, in this special case, of a cokriging system. It is considered that these weights are also random variables, because the estimation of the vector of structural parameters (see step 1) is associated with uncertainty. A linearization technique is used to expand the weights as a sum of the estimated weights plus a weighted linear combination of the errors of estimation of structural parameters. As a result, in the expression of the conditional transmissivity covariance an extra error term appears associated with the uncertainty in the estimation of the structural parameters.

**Some important characteristics of the method**

The same comments which were made in the discussion of the linearized cokriging approach also hold for the linearized semi-analytical approach and are repeated below. The differences between the methods are minor.

In principle, the linearized semi-analytical approach poses the inverse modelling of groundwater flow in a more general framework than the linearized cokriging approach. The Bayesian updating approach is a more general framework than the cokriging approach to obtain the conditional moments from the unconditional moments for, say, transmissivity. The Bayesian updating is not limited to Gaussian processes and produces the complete characterization of the variable of interest, not just the expected value and the covariance. However, in a reply, Kitanidis (1986) states that for other than multi-normal distributed variables the integration of the equations involved in the Bayesian updating equations can only be achieved numerically with a prohibitive computational cost. In addition, Kitanidis argues that the data seldom include sufficient information to determine the finer details of the joint probability density function of the properties of interest. Finally, he states that even if the complete probability density function of transmissivity is known no general and computationally efficient method is known to derive the probability density function of the piezometric head. Another difference between the linearized cokriging procedure and the linearized semi-analytical approach is that, in the linearized semi-analytical procedure, an analytical solution to the linearize equation is used, while Kitanidis and others in their later works use a numerical procedure to solve the linearized groundwater flow equation and to obtain the covariance matrices (Hoeksema and Kitanidis, 1984). Analytical solutions can only be obtained for simple geometries, boundary conditions

and input distributions. On the contrary, numerical solutions can account for complex geometries, boundary conditions and inputs.

It should be noted that the linearized semi-analytical approach has a limited field of application, like the semi-analytical cokriging approach. The (cross) covariance matrices and the estimated means are obtained by a first-order linearization of the groundwater flow equation. This linearization is valid for  $\log_{10}$  transmissivity variances below 1.0. However, in practice  $\log_{10}$  transmissivity variances above 1.0 are common. In addition, the effect of the boundaries is neglected. Rubin and Dagan (1987a) use more restrictive assumptions with respect to the boundary conditions than Kitanidis and others in order to solve the linearized groundwater flow equation. In particular, Rubin and Dagan state that the assumptions made to solve the equation are violated for the zone which is within one integral scale of the boundary. Finally, and most importantly, the use of the transmissivity field obtained as input to a groundwater flow solver will yield a deviation between the measured and the simulated heads. This deviation will grow as the log transmissivity variance increases.

The method determines a unique single best solution of the inverse groundwater flow problem. The disadvantage of a single best solution of the inverse groundwater flow problem is that it is always smooth and it can never represent reality. However, given that the entire statistical structure of the variables of interest is also determined, it is possible to use their covariance matrix to simulate multiple solutions to the inverse groundwater flow problem. The big disadvantage is that, in any case, the solution to the groundwater flow equation in the generated fields will not honor the measured heads.

The linearized semi-analytical approach seems not to be very flexible for extensions to complex domains, three-dimensional flow, simultaneous estimation of a number of possibly spatially variable parameters or the combined inverse modelling of groundwater flow and mass transport. The extension Dagan and Rubin (1988) made to transient groundwater flow is a quasi steady-state approach. However, others extended the method successfully to more complex situations. Sun and Yeh extended the linearized approach to transient groundwater flow, making use of adjoint state equations (Sun and Yeh, 1992). They found that considering transient groundwater flow yielded much better results than the quasi steady-state approach of Dagan and Rubin (1988). Grenier et al (1996) extended the method of Dagan and Rubin to cases of non uniform groundwater flow and non stationarity of the mean log transmissivity by a scale separation approach. In the scale separation approach the hypothesis of locally uniform flow is made and conditioning is limited to neighboring data points for a spatial covariance model derived from scale-separation considerations. However, the methodology is not suited for the high non-uniformities encountered near wells.

An advantage of the linearized cokriging method is that it is not CPU intensive method and that it is conceptually simple.

### **The application of the linearized semi-analytical approach in two case studies**

We will discuss two applications of the linearized semi-analytical approach. One is taken from Rubin and Dagan (1987b) and is about the inverse modelling of steady state groundwater flow with the presence of constant recharge. The other study deals with transient groundwater flow with the presence of time-variable recharge (Dagan and Rubin, 1988).

Rubin and Dagan (1987b) deal with the inverse modelling of groundwater flow at the Avra Valley aquifer in Arizona (USA). A total of 106 transmissivity values and 98 water heads are available from the aquifer. The aquifer is divided into two regions, north and south, because the transmissivity in the northern part is significantly higher on the average than that of the southern part. The spatially variable log transmissivity and the effective recharge are determined separately, and simultaneously, for the two subregions of the aquifer. The maximum likelihood procedure results in the estimation of the vector of structural parameters and its associated estimation error covariance matrix. It is found that incorporating head measurements reduces the uncertainty associated with the estimation of the vector of structural parameters. The maximum likelihood procedure is also carried out for only transmissivity data or only hydraulic head data. It is found that the use of both transmissivity and hydraulic head data results in the lowest estimation variances. The method also indicates that some recharge may exist, although of a small magnitude.

Dagan and Rubin (1988) deal with the inverse modelling of transient groundwater flow on a section of Israel's coastal aquifer. The aquifer is of 120 km length and 12-20 km width and the case study focuses on a subarea of 7 by 7 km, about 9 km east of the coastline. The region is an area with intensive pumping in summertime and artificial recharge through wells in wintertime. The modelling was carried out for the period of October 1970 until April 1971. In total 5 transmissivity measurements were available and 40 hydraulic head measurements were taken each month. There is a total of 29 pumping wells in the region. The drawdown induced by pumping at the wells has been computed using the monthly pumping and injection records. These drawdowns are subtracted from the observed head. By means of the maximum likelihood method the structural parameters are estimated. It is found that values for the structural parameters are strongly correlated over time. The estimation of the structural parameters and the estimation of transmissivity and storativity show (1) that the first-order approximation of the flow equation was valid for this aquifer, (2) that the quasi-steady approach was valid (the estimates for the structural parameters show little changes between months), (3) that the maximum likelihood estimations of the structural parameters have a relatively small variance, and (4) that the maximum likelihood estimates for consecutive months display a variation of the order of the spread indicated by the variance of estimation. In addition, it is found that the storativity is a parameter of low spatial variability. The recharge was estimated by different methods. The estimates of recharge are close to estimates made by multiseasonal and regional water balances and estimates based on texture analysis. It was found that the head measurements improve the characterization of the transmissivity field dramatically; with only five transmissivity measurements it was hardly possible to characterize the transmissivity field while with the help of the hydraulic head data a more detailed identification was possible.

Graham and Tankersley (1994) derive closed-form expressions for the unconditional head covariance, the cross-covariance between head and log-transmissivity and the cross-covariance between head and recharge. These covariances are derived from a linearization of the groundwater flow equation using spectral transform techniques and assuming an infinite aquifer. (In addition, it is assumed that the recharge has a zero mean and that is uncorrelated with transmissivity.) Once the covariances are known, the method proceeds as described earlier by applying cokriging to the data. The difference with the method of Rubin and Dagan is the treatment of recharge. Whereas Rubin and Dagan estimate a spatial average of the recharge, Graham and Tankersley produce a smooth spatially variable estimate of recharge.

Graham and Neff (1994) apply the method presented by Graham and Tankersley (1994) to the Upper Floridan Aquifer in Northeastern Florida. A subarea of the aquifer in which the assumption of a constant mean hydraulic head gradient could be made was selected. The following parameters: mean log transmissivity, log transmissivity variance, log transmissivity correlation scale, mean recharge, recharge variance, recharge correlation scale, mean head gradient and head variance, are estimated from the data. The results of the estimation procedure show that the estimation head variance is clearly lower in the case in which both head and transmissivity data are used than in the unconditional case. However, the transmissivity data have little contribution to this reduction. The estimation log transmissivity variance is also clearly reduced by the use of head data. In this particular case, this is mostly due to the large amount of head data available compared to the few transmissivity measurements. It was also concluded that considering recharge in the estimation process had a large impact in the resulting estimated log transmissivity field. On the other hand, the resulting estimate map of recharge is successful in identifying an area of spring recharge and to capture the major features of the known areal recharge patterns.

### **3.3 Iterative non-linear methods**

#### **3.3.1 The Fast Fourier Transform Method**

##### **Methodology**

###### *Introduction*

The Fast Fourier Transform Method was originally a method to solve the inverse groundwater flow problem by relating the random head field and the random transmissivity field by linearizing the groundwater flow equation in the Fourier space. The (cross) covariances are found by a spectral representation. As such, the method is very similar to the ones presented in section §3.2, since it is based on a linearization of the flow equation. Later this method has been extended to account for the non-linear relationship between transmissivity and hydraulic head in the computation of the head covariance and the head-transmissivity cross-covariance using an iterative approach. Moreover, the method does not calibrate a single best smoothed transmissivity field, but a simulated transmissivity field which resembles the spatial variability as observed in the field. Therefore the method is also referred to as an iterative co-conditional simulation approach. We focus in the presentation of the method by Gutjahr et al. (1994). However, in this paper, Gutjahr and co-workers devote few comments to the iterative aspect of the approach. Later Yeh and others (Yeh et al., 1995; Yeh et al., 1996) discuss the iterative approach and present its mathematical basis. Section §3.3.2 will present the discussions by Yeh and co-workers. The Fast Fourier Transform Method was not the first method to take into account the nonlinear relationship between transmissivity and hydraulic head neither the first method aiming to produce multiple realizations of co-conditional transmissivity fields. We discuss the Fast Fourier Transform now because it is a direct modification of the methods presented before in section §3.2.

### *The steps in the calibration process*

1. *Estimate covariances.* The authors consider 2-D steady state groundwater flow in an unbounded domain. The objective is to simulate a transmissivity field conditional to both transmissivity and hydraulic head data. It is assumed that transmissivity and hydraulic head are second-order stationary random fields. The groundwater flow equation is linearized by a small-perturbation analysis. Then the spectral representation algorithm is applied to get a relationship between the spectral components of transmissivity and head, from which it is extremely easy to derive the spectral density of hydraulic head and the cross-spectral density of hydraulic head and transmissivity. Finally, the auto and cross-covariances are obtained by using the two-dimensional inverse Fourier transform (Gutjahr and Wilson, 1989). Thus, the (cross) covariances are obtained through a linearization approach, like in the linearized cokriging method and the linearized semi-analytical method.
2. *Co-conditional simulation.* The Fast Fourier Transform algorithm is applied to generate unconditioned transmissivity and head fields with the (cross) covariance functions determined in step 1 (Gutjahr et al., 1994). Conditioning is understood as reproduction of the transmissivity data and as reproduction of the linear relationship between head and log-transmissivity established before. In order to apply this step it has to be assumed that transmissivity and hydraulic head are jointly multivariate normal distributed. The conditional realizations of transmissivity and hydraulic head are found by the sum of the unconditional transmissivity and hydraulic head fields plus a weighted linear combination of differences between measured and simulated data (both transmissivity and heads). The weights are determined by solving a cokriging system. Until now, the procedure to determine transmissivity statistics conditioned to transmissivity and hydraulic head data is very similar to the methods presented in sections §3.2.1 and §3.2.2. The main and very important difference with the methods in section §3.2 is that the authors focus on simulated fields instead of estimated fields.
3. *Solve the groundwater flow equation.* The resulting transmissivity field is used, together with boundary conditions, to solve the groundwater flow equation for hydraulic head. The simulated heads will differ from the measured heads, especially if the log transmissivity variance is large. The new head field satisfies the continuity equation. The head field based on the linearized relation between transmissivity and head does not satisfy this relation.
4. *Apply again cokriging.* The way the pseudo-conditional transmissivity field generated in step 2 and the deviation between the generated heads and the measured heads are used is not clear from the paper. It is most likely that the groundwater flow equation is linearized again around the new parameters and the (cross) covariances re-evaluated. Then return to step 2.
5. *Stopping criterion.* The calibration process is stopped if the head field is sufficiently close to the measured head data.

### **Numerical experiments**

Gutjahr and others test the methodology with numerical experiments.

The experiment is carried out in a square 2-D aquifer of side equal to 6.3 log-transmissivity integral scale. An synthetic transmissivity field and the steady-head solution of the groundwater flow equation is used as reference. Several transmissivity fields with a mean  $\log_{10}$  transmissivity of -3 and different variances are considered. Prescribed head boundaries are imposed on the left and right boundaries of the system so that a uniform head gradient parallel to the x-axis is created. The upper and lower boundary of the system are impermeable. In total 30 transmissivity and 30 head samples are taken from the reference fields which are then used in the conditioning procedure. Different scenarios are studied and for each scenario 100 conditioned fields are generated. In order to compare the conditioned fields, particles were tracked (no diffusion) from the center of the field to the boundary, and travel paths and travel time distributions were computed. Four cases with varying number of conditioning data are studied: (i) no conditioning, (ii) 30 transmissivity values only, (iii) 30 head values only and (iv) 30 head and 30 transmissivity values. Three different cases with respect to the  $\log_{10}$  transmissivity variance are studied: (i) a variance of 0.25, (ii) a variance of 1.0 and (iii) a variance of 2.0.

For the case of a  $\log_{10}$  transmissivity field with variance 1.0 it is found that more conditioning data (head and transmissivity) result in less variance in the arrival times of the particles. However, using 30 head values and no transmissivity values the variance in the travel times was hardly reduced. It is concluded that transmissivity data are most influential variance reduction of arrival times estimates. With respect to the arrival locations of the particles it is found that using 30 head conditioning data (and no transmissivity values) yields a notable reduction in the variance of the arrival location, while using 30 conditioning transmissivity data (and no head values) hardly reduces this variance. It seems that hydraulic head data reduce mainly the uncertainty in the travel paths. For the  $\log_{10}$  transmissivity variance of 2.0 the iterative approach reduces the uncertainty in the travel times and travel paths (compared are the results without using an iterative approach and the results after the 4<sup>th</sup> iteration). Furthermore, the simulated head values after the 4<sup>th</sup> iteration are very close to the measured head values. The conditional distributions obtained for the case of a  $\log_{10}$  transmissivity variance of 2.0 by the iterative approach are compared with the «true» distributions. The «true» distributions are obtained by brute force after generating a very large number of transmissivity fields conditioned only to transmissivity measurement and selecting those fields in which the solution of the groundwater flow equation reproduces closely the measured heads. For each of these transmissivity fields the groundwater flow in the domain has been solved and the fields for which the measured head values are closely reproduced are maintained. These fields are used to calculate at several locations the distributions of hydraulic head and transmissivity and compare them with the distributions of head and transmissivity for the iteratively conditioned fields. It is found that after the 4<sup>th</sup> iteration the conditional distributions approach very well the true ones in most cases.

### **Some important characteristics of the method**

The comments below refer to the advantages the Fast Fourier Transform Method has as compared with the linearized methods from section §3.2.

The Fast Fourier Transform Method (FFTM) is an attempt to circumvent the problem of the first-order linearization of the groundwater flow equation. The linearization based on a first-order approximation is only valid if the unconditional variance of  $\log_{10}$  transmissivity is less than 1.0. Gutjahr and others (1994) start with a procedure very similar to the linearized cokriging method, but update later iteratively the transmissivity field with the aim to match the meas-

ured heads. The iterative approach results finally in a transmissivity field in which the solution of the groundwater flow equation reproduces the measured heads. The iterative approach allows to handle cases with a variance of  $\log_{10}$  transmissivity above 1.0.

Gutjahr and others (1994) give little information about the iterative approach so that it is difficult to determine to what extent the proposed methodology avoids the problems associated with the linearization of the groundwater flow equation. For example, the authors do not make clear whether they update the hydraulic head covariance matrix, the hydraulic head- transmissivity cross-covariance matrix or the transmissivity covariance matrix at each iteration. Iteratively updated or not, in all cases these (cross) covariance matrices are estimated by a first-order approximation of the groundwater flow equation.

Gutjahr and others use a simulation approach for solving the inverse groundwater flow problem. Although they do not state it explicitly, their aim is to determine a sufficient number of equally likely transmissivity simulations conditioned to transmissivity and head measurements. This has a clear advantage compared with a single best solution of the inverse groundwater flow problem. By means of a solution in the form of an ensemble of transmissivity fields the probability density functions of different input variables of the flow and transport equations are obtained and in addition it is possible to obtain the probability density functions of output variables like particle travel times or arrival positions.

The Fast Fourier Transform Method seems moderately flexible for extensions to more complex domains, three-dimensional flow, simultaneous estimation of a number of possibly spatial variable parameters or the combined inverse modelling of groundwater flow and mass transport. A problem is that, like in the presented linearized methods in section §3.2, the (cross) covariance matrices have to be determined by a linearization of the groundwater flow equation. However, the method is more flexible than the linearized methods because if these covariance matrices can be computed the method can handle more complex situations using the iterative approach. Recently an extension to handle transient groundwater flow has been made (Hughson, 1997).

### **3.3.2 The iterative cokriging-like technique**

#### **Methodology**

##### *Introduction*

The second method we discuss which incorporates the nonlinear relationship between transmissivity and hydraulic head fields is the iterative cokriging-like technique (ICT) by Yeh and others. We discuss the method as presented in two recently published papers (Yeh et al., 1995; Yeh et al., 1996). The method starts like the linearized cokriging approach. In addition, the linear estimator is improved in an iterative manner. The two papers differ in the method in which the iterative update is made. We discuss both of the methods, with emphasis in the method presented in the later paper. This method is very similar to the iterative method of section §3.3.1 (the Fast Fourier Transform Method), but with a better discussion of the iterative approach used.

##### *The steps in the calibration process*

The first step in ICT is a straight application of the linearized cokriging method with no iterations. However, ICT continues with an iterative process which improves the characterization of the transmissivity field. The two articles formulate two different iterative updating processes.

1. «Classical» cokriging. As in section §3.2.1 and §3.3.1. A very short summary is given here. A calibrated transmissivity field is obtained by a cokriged estimate of transmissivities and heads. The necessary (cross) covariance matrices are obtained from a linearization of the groundwater flow equation by a first order approximation. This is the philosophy not only of the linearized cokriging approach (see section §3.2.1), but also of the Fast Fourier Transform Method (see section §3.3.1). The way the (cross) covariances are obtained varies. The perturbation-spectral analysis used in the Fast Fourier Transform Method to obtain the (cross) covariance matrices is applied by Yeh et al., 1995. A numerical method is used by Yeh et al., 1996.
2. *Solve the groundwater flow equation.* The resulting (smooth) transmissivity field is used in a finite elements steady state groundwater flow solver. (The field used here is obtained by cokriging, as opposed to the Fast Fourier Method which starts from a simulated field.) Notice that now the groundwater flow equation is solved and not a linearized version of it. The two articles use two different approaches to solve the groundwater flow equation:
  - (i) The observed head values at the sample locations are placed as internal constant-head boundary conditions. Therefore, the simulated head values are equal to the measured head values. The measured head values in this case have no influence on the calibration of the transmissivity field since they are always reproduced exactly. (Yeh et al., 1995).
  - (ii) The flow equation is solved without setting constant-head boundary conditions at the grid cells corresponding to the head sample locations (Yeh et al., 1996).
3. *Apply cokriging again.* The two approaches mentioned in step 2 are followed by the application of cokriging, but the procedure differs.
  - (i) Cokriging is used again, but this time the transmissivity estimate is not only function of the measured transmissivity and head data, but also of all the head data resulting from the solution of the flow equation. It results in a new cokriged transmissivity field, which is again used as input to the groundwater flow equation (Yeh et al., 1995).
  - (ii) The updated transmissivity at a certain location is the transmissivity value from the pervious iteration plus a weighted linear combination of the head residuals (a head residual is the difference between the cokriging head estimate and the measured head value at a head measurement location). The weights are determined by solving again a cokriging system. For solving the cokriging system updated estimates of the conditional transmissivity covariance matrix, the conditional hydraulic head covariance matrix and the conditional head-transmissivity cross covariance matrix are needed. The conditional transmissivity covariance matrix is easy to update. The other two covariances are updated by approximating the residual heads by a first-order Taylor series. The sensitivity matrix which appears in the resulting expression is estimated

by an adjoint state sensitivity method subject to boundary conditions. The covariance matrices are updated at each iteration (Yeh et al., 1996).

4. *Stopping criterion.* The two different iterative approaches, as described in step 2 and step 3 use also different stopping criteria.
- If the absolute change in the head value between two succeeding iterations decreases below a specified tolerance at all grid cells, the calibration process is terminated.
  - If the absolute difference in the variance of the estimated transmissivity field between two successive iterations is smaller than a prescribed tolerance the iteration stops.

### **Numerical experiments**

In Yeh et al. (1996) a synthetic study is presented to test the iterative cokriging-like technique. The results obtained from the iterative approach are compared with the results from the non-iterative approach. Synthetic heterogeneous transmissivity fields are generated by a multiGaussian random field generator with an exponential anisotropic variogram. These transmissivity fields are considered as reference fields. The 2-D steady state groundwater flow equation with a constant-rate pumping well, no flux conditions at the upper and lower boundaries and prescribed heads at the left and right boundaries is solved for these fields.

First a deterministic inverse problem is solved. Yeh et al. (1996) use the term deterministic in the sense that the inverse problem has a unique solution; the head values at all the grid nodes of the flow solver are known and all the transmissivity values near the left boundary (one value for each streamline) are known. The solution of the deterministic inverse problem taking head and transmissivity values from a reference transmissivity field with a low variance ( $\log_{10}$  transmissivity variance = 0.38) resembles for both the iterative and non-iterative approach very well the «real» transmissivity field. However, even in this case the non-iterative approach gives a too smooth transmissivity field, whereas the iterative approach results in a transmissivity field with a spatial variability closer to the «true» field. The solution of the deterministic inverse problem for a reference transmissivity field with high variance ( $\log_{10}$  transmissivity variance = 3.01) results in a very smooth calibrated transmissivity field for the non-iterative approach and a calibrated transmissivity field for the iterative case which is again closer to the «real» transmissivity field. It is shown that due to the linear approximation of the groundwater flow equation in the non-iterative approach the updated transmissivity field is of bad quality, in spite of the fact that the heads at all the grid cells are known.

Also a stochastic inverse problem is solved, which means that the hydraulic head values are known on a limited number of locations. In total 30 error-free head measurements and 5 error-free transmissivity measurements are taken from the «true» fields. It is found that the iteratively updated transmissivity field reveals better the structures of the «true» transmissivity field than the transmissivity field obtained by the non-iterative (classical co-kriging) approach. Comparing the updated transmissivity field by the non-iterative approach to the updated transmissivity field by the iterative approach, the latter is closer to the «real» field than the former and it is less smooth. However, the iteratively updated transmissivity field is still smoother than the «true» field due to the fact that the updating is achieved by starting from a smooth kriged field and adding smooth (kriged) perturbation surfaces.

It should be commented that in the synthetic experiments conducted (not only for this method but in almost all cases) the reference log-transmissivity fields are multiGaussian. Therefore, the methods are bound to work, since a key decision in the development of the algorithms is that the log-transmissivity fields are multiGaussian. However, in practice, frequently the spatial distribution of transmissivities may be non-multiGaussian and it is very common that not enough data are available to conclude whether the spatial auto-correlation pattern of transmissivities is multiGaussian or not (Gómez-Hernández and Wen, 1998).

### **Some important characteristics of the method**

In this discussion we focus on the second version of the method (Yeh et al., 1996). The use of the measured heads as boundary conditions in the solution of the groundwater flow equation does not seem very appropriate.

The iterative cokriging-like technique is an attempt to circumvent the problem of the first-order linearization of the groundwater flow equation. The linearization based on a first-order approximation is only valid if the unconditional variance of transmissivity is less than 1.0. Yeh and others use as a starting point the cokriging approach and apart from the efforts made to take into account the nonlinear relation between transmissivity and hydraulic head by an additional iterative approach nothing new is added as compared to the linearized cokriging approach. The newest variant of the method (Yeh et al., 1996) has the same philosophy as the linear methods of section §3.2 in the sense that the resulting iterative estimator of transmissivity is unbiased with minimal variance. Notice that other methods, developed earlier, like the zonation procedure of Carrera and Neuman (see section §3.4.1), the Sequential Self Calibrated Method (see section §3.4.2) and the Pilot Point Method (see section §3.4.3) already took into account the nonlinear relation between transmissivity and hydraulic head.

The final estimated transmissivity fields is conditional to the transmissivity measurements and close to the hydraulic head measurements, although the reproduction of the hydraulic head measurements is not explicitly stated in the calibration process. In spite of this, the transmissivity field will in general honor the measured hydraulic heads reasonably, also for transmissivity variances above 1.0. The updating of the transmissivities is based on a first-order approximation of the hydraulic head covariance matrix and the hydraulic head-transmissivity cross-covariance matrix. In addition, the transmissivity covariance matrix at later iterations is also influenced by the linearizations, because in the updating equation of the transmissivity covariance matrix the hydraulic head covariance matrix and the cross covariance matrix appear. The fact that the (cross) covariance matrices are updated using a first-order approximation means that the quality of the updated transmissivity fields is worse if the transmissivity variance is higher; especially the estimation of the auto-correlation pattern of the transmissivity field is expected to become biased.

The main difference between the method presented by Yeh et al. (1996) and the Fast Fourier Transform Method (see section §3.3.1) is that Yeh and co-workers produce single best estimates of the inverse groundwater flow equation as opposed to Gutjahr and co-workers who produce equally likely realizations. The methodology to achieve their goal is virtually the same, but the Fast Fourier Transform Method should be deemed as superior since the transmissivity fields generated with this method resemble the spatial variability of transmissivity as observed in the field and modeled by a covariance function.

Like the Fast Fourier Transform Method, the iterative cokriging-like method seems moderately flexible for extensions to more complex groundwater flow and possibly mass transport problems, for the same reasons as given in section §3.3.1.

### 3.4 Iterative numerical non-linear methods

#### 3.4.1 The zonation procedure of Carrera and Neuman

##### Methodology

###### *Introduction*

The theoretical basis of the zonation procedure, a method for the inverse modelling of groundwater flow developed by Carrera and Neuman, has been published in a series of three papers (Carrera and Neuman, 1986a, Carrera and Neuman, 1986b, Carrera and Neuman, 1986c). We will explain the method by means of these three papers, although later the method has been extended for the inverse modelling of coupled groundwater flow and mass transport (e.g. Medina, 1993, Medina and Carrera, 1996).

In the method developed by Carrera and Neuman for solving the inverse groundwater flow problem, the exact formulation of the steady or transient groundwater flow equation is used, without linearizing it. The inverse problem solution is obtained by an iterative fashion, in which the groundwater flow equation is solved numerous times. The most important characteristic of the method is that aquifer parameters are estimated for a limited number of zones which partition the aquifer. Different parameters can be estimated for each zone such as, values and directions of principal hydraulic conductivities in anisotropic media, specific storage, boundary conditions, or recharge rates.

###### *The steps in the calibration process*

The method of Carrera and Neuman can be summarized by the following six steps (Carrera and Neuman, 1986b):

1. *Choose an initial set of parameters.*

The following model parameter values have to be initialized and may be subject to calibration: values and directions of the principal components of hydraulic conductivities, specific storage, boundary conditions and recharge rates. The covariance matrix of all model parameters has to be provided, too. Arguing that prior errors in model parameters can occur due to different causes and calling upon the central limit theorem, the authors conclude that the errors must be Gaussian with zero mean. The parameter covariance matrix  $C_p$  is block diagonal, its diagonal components being the covariance matrices  $C_i$  for the different parameters; it is assumed that prior estimates of the various types of parameters are mutually uncorrelated. The covariance matrix of an individual parameter  $C_i$  is the product of a known or unknown positive scalar and a known symmetric positive-definite matrix  $V_i$ .

The covariance matrix of the prior head errors  $\mathbf{C}_h$  has to be given, too. Carrera and Neuman use again the central limit theorem to state that prior head errors are Gaussian with zero mean; the reason being that there are a large number of contributing factors to the head errors (Carrera and Neuman, 1986a). The covariance matrix of the prior head errors  $\mathbf{C}_h$  is the product of a positive, possibly unknown, scalar and a symmetric positive definite matrix  $\mathbf{V}_h$ , which may be unknown up to one or more parameters. The authors state that errors at different well locations are normally weakly correlated and that this correlation is not likely to change with time. It means that the head errors show weak spatial auto-correlation. This auto-correlation structure is considered invariant in time. The head errors may show temporal correlation, due to systematic components of the prior head errors.

The prior errors in the model parameters and the prior head errors are considered independent.

2. *Solve the groundwater flow equation.*

The groundwater flow equation (steady or transient) is solved by the finite elements method. The input parameter values necessary to solve the groundwater flow equation are, for the first iteration, the prior estimates, for the subsequent iterations the updated values.

3. *Compute the gradient.*

During the calibration, all the unknown parameters will be estimated; the likelihood that, given the observations, the estimated parameter values are true, given the observations A likelihood function is formulated that results in an objective function to be minimized that penalizes the deviation between the predicted heads and the measured ones (weighted by the matrix  $\mathbf{V}_h$ ) and the deviation between the updated parameter values and their prior estimates (weighted by the matrices  $\mathbf{V}_i$ ).

The objective function value can be calculated after the groundwater flow equation is solved. Then, the gradient of the objective function with respect to the model parameters is determined. This requires, for each parameter, the sensitivity of the model fit with respect to the perturbation of that particular parameter. Carrera and Neuman apply the adjoint state method to calculate the gradient in an efficient way (Carrera and Neuman, 1986b, see also Carrera et al., 1991). The adjoint state method makes it possible to calculate the gradient of the objective function without having to calculate the sensitivity matrix (the Jacobian matrix), which may result in important savings in CPU time in case the number of measurement spatial locations is less than the number of model parameters. (The adjoint state equations, like the groundwater flow equations, are solved by finite elements.)

4. *Compute the updating direction.*

Non linear optimization algorithms are used to calculate the updating direction. Carrera and Neuman alternate three different optimization algorithms during the calibration process, namely Fletcher-Reeves Conjugate Gradients, Broyden and Fletcher-Powell-Davidon. It was found that alternating different optimization algorithms yields a faster convergence (Carrera and Neuman, 1986b) than using any of this algorithms independently. Fletcher-Reeves Conjugate Gradients does not use an estimation of second order deriva-

tives of the objective function with respect to the model parameters. The other two algorithms make use of the second order derivatives and apply different methods to estimate the Hessian matrix.

Once the updating direction is known, the objective function is minimized in the direction of the updating direction by Newton's method.

5. *Update the parameter estimates.*

The one-dimensional minimization by Newton's method in the updating direction provides the step size that multiplied by the gradient in the updating direction yields a vector that added to the previous iteration estimates results in the updated parameter estimates.

6. *Check for Convergence.*

Several criteria are used to decide whether to stop the calibration process or not. If none of the criteria is met a new iteration is carried out (return to step 2), otherwise the procedure stops. The stopping criteria are:

- (i) the gradient norm is less than a prescribed minimum.
- (ii) the ratio of the norm of the gradient at the end of the iteration to the gradient norm at the beginning of the iteration is smaller than a prescribed value.
- (iii) the objective function value decreases below a user defined value.
- (iv) the relative change in objective function value falls below a prescribed value.

Carrera and Neuman rely most on criterion (ii), although the initial norm of the gradient is dependent on the initial parameter estimates. With respect to criterion (iv) they state that there are cases that the relative change in objective function becomes very small and, in spite of this, parameter values change significantly later (Carrera and Neuman, 1986b).

*Model selection criteria*

Steps 1 till 6 summarize the calibration process once a given zonation is selected and a given covariance model is adopted. This model selection is a key step in the potential success of the calibration process and for this reason, Carrera and Neuman suggest the analysis of alternative models and the application of a model selection criteria to decide which model is most adequate. They take advantage of the model structure identification work carried out in the context of Box-Jenkins time series analysis and consider the Akaike Information Criterion, the Akaike Maximum Entropy Principle, the Hannan Criterion and the Kashyap Criterion. The four criteria are based on the principle of parsimony in that, everything being equal, the model with the smallest number of parameters is chosen. In general, the criteria balance the model fit (e.g. the differences in observed and simulated heads) and the number of parameters to achieve this fit. The application of the model selection criteria results, in general, in a zonation of the aquifer in a small number of zones, for each of which the parameter values are estimated. The zonation is such that if the number of zones would be augmented the model selection criterion would give a less favorable result because the increase of the number of parameters is not compensated by a sufficient improvement in the model fit.

## Numerical experiments

Two examples of application of the zonation procedure for the inverse modelling of groundwater flow are given by Carrera and Neuman (1986c). A synthetic example, and a real case study. We will give some results from the synthetic study. The paper demonstrates, however, that the method can also handle successfully real world situations.

In the synthetic example a 2-D groundwater flow problem in a square aquifer is considered. A mix of different boundary conditions are used: no flow boundaries at the top and right, a prescribed head boundary at the bottom and a prescribed flow boundary at the left. The aquifer is divided in nine constant transmissivity zones. Two wells are pumped from the beginning of the simulation period. The steady state flow problem prior to pumping and the transient flow problem during 1000 days are solved by a finite elements code. In total 18 head measurements are taken from the «true» head field. These head measurements are corrupted by white Gaussian noise. Nine sets of head measurements are considered: three sets from three different scenarios; the scenarios differ in the degree in the amount of noise added to the head measurements. The noise has zero mean and variable standard deviations of 0.01 m, 0.10 m and 1.0 m. The following conclusions can be drawn (Carrera and Neuman, 1986c):

- In the case that transmissivities are estimated from steady state data, without prior information on transmissivity, the solutions are somewhat unstable. The results are better for head data which are less corrupted by noise than for head data which are corrupted by noise with a standard deviation of 1.0 m.
- Including prior information on transmissivities improves the estimates of transmissivity. The final estimates lie, in general, between the prior estimates and the estimates without prior information.
- The error covariance matrix is a good representation of the actual errors. It is also illustrated how this matrix could be used to analyze the relative worth of data. Including transient data together with steady state data reduces all the eigenvalues of the error covariance matrix by about half order of magnitude in both cases indicating a reduction of the uncertainty on the estimates.
- Carrera and Neuman illustrate how the error covariance matrix can be used for monitoring network design both in space and time. The determinant, the condition number and the individual eigenvalues of the matrix are useful criteria to compare monitoring networks.
- In another experiment, ten different conceptual models are considered, each with a different transmissivity zonation pattern. The aforementioned model selection criteria are used to choose the best parameter zonation pattern among a number of given alternatives. Carrera and Neuman find that the Kashyap criterion is the most appropriate among the four criteria tested, because it recognizes that if the head data are corrupted severely with white noise there is little justification for an elaborate model with a lot of parameters.
- The method has also been applied successfully to estimate simultaneously different kind of parameters, like transmissivities, areal recharge, flow across the left boundary and head at the bottom boundary.

### **Some important characteristics of the method**

The work of Carrera and Neuman addressed various new issues. It was novel not to linearize the groundwater flow equation for the solution of the inverse problem and successively improve the estimates of the aquifer parameters by an iterative approach. Furthermore, they used new approaches in the context of groundwater flow inverse modelling, like the adjoint state approach for the calculation of the gradient (a more detailed discussion can be found in Carrera et al., 1991) and the alternating of different optimization algorithms in determining the updating direction from the calculated gradient.

The inverse groundwater flow problem is solved by taking into account the nonlinear relationship between hydraulic heads and transmissivities. The method can therefore also be applied to transmissivity fields which have a  $\log_{10}$  variance above one; the method is not limited to transmissivity fields with a small variance as the methods presented in section §3.2.

In order to circumvent the ill-posedness of inverse problems and to achieve a meaningful solution the authors propose the estimation of model parameters on a limited number of zones which divide the aquifer. A disadvantage of this is that the spatial representation of the aquifer parameters is too smooth. Moreover, it is somewhat artificial to assign constant parameter values to large zones. Notice that for example in the linearized cokriging approach (see section §3.2.1) and the iterative cokriging-like technique (see section §3.3.2), in which also a single best smooth solution is determined, the calibrated transmissivities vary gradually instead of step-wise throughout space. In the linearized cokriging approach and the iterative cokriging-like technique a series of structural parameters which describe the cross correlated random fields of hydraulic head and transmissivity are calibrated. Carrera and Neuman calibrate the spatially variable values of transmissivity directly, so that, in order to get a well-defined parameter estimation problem, the number of zones at which parameter values are estimated has to be limited. Carrera and Neuman state that the number of zones in which to divide the aquifer may be determined by model selection criteria. Furthermore, the authors stress the importance of including prior information about the parameters in the estimation criterion to reduce the degree of ill-posedness of the problem.

In a comparison made by Carrera and Glorioso (1991) of the linearized cokriging approach (see section §3.2.1) and the zonation procedure it is suggested that the methods are basically the same, except at the stage in which the groundwater flow equation is linearized and because the zonation procedure continues the calibration by an iterative approach. According to Carrera and Glorioso (1991) the linearized cokriging approach yields the same results as the zonation procedure after the first iteration. Subsequent iterations result in an improved calibration of the aquifer parameters. In a reply, Kitanidis states that the linearized cokriging approach does yield the same results as the first iteration of the zonation procedure. According to him, the linearized cokriging approach yields unbiased estimates with minimum variance and a reliable estimation of the error covariance matrix. The zonation procedure, on the contrary, yields biased estimates of the aquifer parameters and a not accurate estimation of the error covariance matrix, especially in case of a fine discretization of the aquifer. Kitanidis states that this bias is due to the parameter estimation employed procedure in which the covariance parameter estimation is detached from the estimation of the spatial variable (Kitanidis, 1996).

The zonation method determines a unique single best solution to the inverse groundwater flow problem. See the sections §3.2.1 and §3.2.2 for the disadvantages associated with a single best solution to the inverse problem. However, as all methods that provide an estimation error covariance, their single best estimate together with the estimation error covariance can be used to generate realizations with expected value equal to the estimate and covariance equal to the error covariance.

The method is flexible to handle complex groundwater flow situations with mixed boundary conditions and non rectangular domains due to its numerical nature. The method allows to estimate a series of aquifer parameters and has already been extended to the combined inverse modelling of groundwater flow and mass transport (Medina, 1993; Medina and Carrera, 1996).

### 3.4.2 The Sequential Self Calibrated Method

#### Methodology

##### *Introduction*

The self calibrated method (SCM) was first outlined by Sahuquillo et al (1992). An extended description together with applications was presented in a series of three papers by Gómez-Hernández et al. (1997) and Capilla et al. (1997, 1998). This discussion focuses in the latter three papers. Some extensions of the method to transient groundwater flow and to the calibration of storativities, which have been made recently, are also discussed.

The SCM can be considered as the first inverse stochastic simulation method aiming to the direct generation of realizations of transmissivity fields conditional to both transmissivity and piezometric head measurements. Besides, conditioning to piezometric head is achieved through the solution of the groundwater flow equation, not through any linearization. The difference with respect to the zonation procedure of Carrera and Neuman is in the parameterization of the transmissivity spatial distribution of each realization. In the SCM, the spatially variable transmissivity values are parameterized as the sum of a seed transmissivity field obtained by standard geostatistical techniques plus a perturbation of the entire field which is expressed as a linear function of the perturbations at a selected number of locations referred to as master blocks.

##### *The steps in the calibration process*

#### 1. *Generate a seed field.*

A seed transmissivity field is generated which honors the transmissivity data and displays the spatial variability observed in the field and modeled by a variogram. At this stage, prior estimates of other parameters, like boundary conditions and interior recharge or leakage rates, are made. This prior estimates can be subject of updating during the calibration process if deemed necessary. In case of transient groundwater flow equation, also prior estimates of storativity and time-dependent recharge or leakage rates have to be made. It is also possible to generate a seed storativity field (possibly jointly correlated with the transmissivity field), honoring the storativity data and displaying the spatial variability observed in the field as modeled by a variogram (and possibly a cross-variogram).

2. *Solve the groundwater flow equation.*

The steady or transient groundwater flow equation is solved by the finite differences method using the parameter values of the current iteration for the transmissivity field, boundary conditions, time-dependent recharge rates and the storativity field.

3. *Evaluate the penalty function.*

In the self calibrated method an objective function similar to the one used by Carrera and Neuman is employed. The objective function is the sum of squared differences between measured and simulated heads, weighted by a head error matrix, plus a sum of squared deviations between the updated parameter values and the prior estimated parameter values, weighted by the covariance matrix of the corresponding parameter. Weights are assigned to the two terms of the objective function. It is common to repeat the calibration procedure (steps 1 to 6) for different weights. The objective function value can be calculated after the groundwater flow equation is solved.

4. *Compute the updating direction.*

In addition, a first-order approximation of the solution of the groundwater flow equation in the updated transmissivity field (and eventually updated storativity field, updated boundary conditions and updated estimates of recharge rates) is obtained by a first order Taylor series expansion around the head solution in the transmissivity field from the previous iteration. Notice that the linearization is made after solving the full groundwater flow equation and calculating the objective function value and its only purpose is to define an updating direction.

Two different procedures are available to calculate the updating direction. In both procedures the transmissivity update is expressed as a linear function of the updated values at a selected number of master blocks spread over the aquifer in a regular way; therefore, the number of parameters to optimize is equal to the number of master blocks. One of the procedures is based on the computation of the sensitivity coefficients required to compute the gradient of the objective function, the other procedure uses the adjoint-state formulation to compute directly the gradient of the objective function by-passing the computation of the sensitivity coefficients.

5. *Update the parameter estimates.*

The perturbations at the master blocks, which are the subject of the optimization, are determined by the product of the updating direction and the step size. These perturbations are then interpolated to all the grid cells by ordinary kriging. The updated transmissivity field is the sum of the transmissivity field at the end of the previous iteration plus the perturbation field. In a similar way an updated storativity field (sum of the storativity field at the end of the previous iteration plus the storativity perturbation field), updated boundary conditions (sum of the boundary conditions at the end of the previous iteration and calculated perturbations) and updated leakage or recharge rates (a sum of rates at the end of the previous iteration and calculated perturbations) can be obtained.

## 6. *Check for Convergence.*

A number of criteria is used to decide whether to stop the calibration process or not. If none of the criteria is met, the procedure continues back to step 2, using the updated parameter estimates as the seeds for the next iteration. The stopping criteria are:

- (i) the objective function value decreases below a user defined value.
- (ii) the relative change in objective function value falls below a prescribed value.
- (iii) the number of iterations reaches a user defined maximum number of iterations.

## **Numerical Experiments**

In Capilla et al. (1997) a number of controlled numerical experiments are presented. In all cases, a 2-D rectangular domain (60 x 60 grid cells) under steady state flow without sources or sinks is considered. Heads are prescribed along all four boundaries. The reference  $\log_{10}$  transmissivity field has zero mean, unit variance and isotropic correlation distance of one third of the domain size. Uncertainty in boundary conditions is not considered. The reference heterogeneous transmissivity field is sampled at a few locations for transmissivity and head values which are used as conditioning data for the generation of transmissivity fields. In the base case, the reference log transmissivity field and the reference steady state head field are sampled at 30 locations and a total of 100 master blocks are laid out on a regular grid. With respect to this base case some variations are studied, such as sensitivities to the log transmissivity variance, the number of conditioning data, or the number of master blocks. For each scenario 10 transmissivity realizations are generated conditional to transmissivity and head measurements. The main findings are:

- The transmissivity fields are conditional to the steady state head data, since the head data are reproduced by the solution of the groundwater flow equation in the resulting fields. It is found that conditioning to head data, in addition to transmissivity data, improves the overall reproduction of the reference head field and at the same time reduces its uncertainty, with respect to the case in which only transmissivity data are used for conditioning. Similarly, conditioning to head data also improves the transmissivity estimates. Satisfactory results are also obtained for a reference field with  $\log_{10}$  transmissivity variance as high as 2.0. It is found that, there is a threshold on the number of head data beyond which it is not worth to collect any additional head data.
- When the variogram of the seed field departs from the model variogram used for its generation (due to ergodic fluctuations) the updating procedure corrects the seed field so that the final update field has a variogram which is closer to the model one.
- From a sensitivity analysis on the number of master blocks it is concluded that at least 3 master blocks per range in each direction are needed in order for the algorithm to work.

Wen et al. (1996) present another synthetic study which demonstrates the worth of piezometric head data versus transmissivity data. A synthetic 2-D reference field of 30 km by 30 km which has been discretized into 60 by 60 square cells of 500 m is used. The reference  $\log_{10}$  transmissivity field had a mean of  $-0.6$ , a variance of  $0.75$ , and an isotropic spherical variogram with a

range of 8000 m. The steady state groundwater flow equation is solved by finite difference using prescribed boundary conditions, resulting in the reference head field. Then, the advective travel paths and travel times of four particles are computed. To determine the relative impact that measurements of transmissivity and head would have in the advective travel predictions, 64 different sampling configurations were considered with different combinations of head and log transmissivity data sampled from the reference fields. For each of the 64 sampling configurations, 250 conditional realizations of transmissivity were generated. The main conclusions are, the best predictions are obtained with the largest number of data of both type, transmissivity data have a larger impact than head data in the prediction of travel times, whereas the opposite is true for the prediction of travel paths. It was also found that when no or little data are available, head measurement are more valuable than transmissivity measurements; however, there is a threshold number of head measurement beyond which additional measurements are virtually irrelevant.

### **Some important characteristics of the method**

Like all the methods discussed in section §3.4 this method, the SCM is not limited to transmissivity fields with small variances because no first-order approximation is made to establish the dependence between piezometric head and transmissivity. The full groundwater flow equation is iteratively solved numerous times in order to match the measured heads. As such, if the final updated parameters are used as input to the groundwater flow equation the measured heads are reproduced. On the contrary, for the methods presented in section §3.2 this is in general not the case.

The SCM aims to the generation of an ensemble of conditionally simulated transmissivity fields (and eventually an ensemble of conditionally simulated storativity fields, boundary conditions and recharge) calibrated (conditioned) to steady state and transient state head measurements. This has a clear advantage compared with a single best solution of the inverse groundwater flow problem, particularly for later analysis of mass transport. Using the ensemble of transmissivity fields generated, it is possible to compute not only the probability density functions of the input parameters, but also the probability density functions of output variables such as travel times or solute concentrations at given control locations.

The method is, due to its numerical nature, flexible to handle complex groundwater flow situations with mixed boundary conditions and non rectangular domains. The method has already been extended to the inverse modeling of three dimensional groundwater flow. Further extensions are possible with the only limitation being the computer resources because the method is very CPU intensive.

### **Application of the methodology to the WIPP-site**

The Waste Isolation Pilot Plant (WIPP) is a potential location to host a repository for the disposal of radioactive waste. The approval, if granted, relies on performance assessment analyses; a key component is evaluating the potential for radionuclide transport from the underground repository to the accessible environment by groundwater. The WIPP repository is located approximately 650 meters below the land surface in the lower part of a predominantly halite formation. In addition, it is thought that the Culebra dolomite formation is the principal pathway for transport of radionuclides to the accessible environment in case of escape of par-

ticles from the repository. Due to some secondary processes like halite dissolution, subsidence, and calcium sulfate hydration, fluid density is spatially variable. Also a strong spatial heterogeneity in transmissivity exists. The thickness of the Culebra Dolomite at the WIPP site is on average 7.7 m.

Capilla et al. (1988) assume confined 2-D steady state flow in the Culebra formation without sinks and sources. Along all boundaries, prescribed head values are set which were obtained from preliminary reports by SANDIA National Laboratories. The conditioning data are 33 steady state head measurements, 37 transmissivity measurements, and 31 water density measurements. Groundwater flow and advective transport is analyzed. Particles are released from the hypothetical repository and arrival times, and arrival locations, to the southern boundary of the model area are computed.

A total of 300 log transmissivity seed fields are generated, conditional to the transmissivity measurements and displaying the spatial variability observed in the field and modeled by a spherical variogram. The variogram is modeled as the sum of three nested structures, which take into account a strong zonal anisotropy. For each of the fields the groundwater flow equation is solved by finite differences using 43 by 61 square blocks of 500 m by 500 m. The 300 transmissivity fields are then calibrated to steady state heads. The most important results are:

- The SMC can be applied to a real case with strong spatial heterogeneity.
- Conditioning the transmissivity fields to steady state heads results in a reduction of the ensemble variance of both transmissivity and hydraulic head. Also uncertainty about the travel paths and arrival times of the released particles reduces noticeably.
- When the boundary conditions are calibrated together with the transmissivities, the reproduction of the measured heads improves. The side effect is a reduction of the travel times due to the systematic reduction of the overall gradient across the formation.

### 3.4.3 The Pilot Point Method

#### Methodology

##### *Introduction*

Originally the Pilot Point method was formulated as a standard inverse groundwater flow problem yielding a single «best» map of conductivities (de Marsily et al., 1984). However, we will discuss the newest version of the Pilot Point method (RamaRao et al., 1995; LaVenue et al., 1995) which pursues the generation of multiple realizations like the Self Calibrated Method (see section §3.4.2).

The aim of the method is to generate a large ensemble of equally likely transmissivity fields, each of which is in agreement with all the measured data and, in this respect, can be considered a plausible model of the true aquifer transmissivity field. The methodology can handle both steady and transient groundwater flow, in 1-D, 2-D or 3-D, with the possibility of spatially variable fluid density. In the discussion of the method we will focus on the 2-D problem (the authors do not present 1-D or 3-D examples).

*The steps in the calibration process*

The method, which is implemented in the GRASP-INV code, consists of the following steps (these steps are repeated for each simulated transmissivity field (RamaRao et al., 1995):

1. *Simulation of a transmissivity field.* The turning bands simulation method is used to generate a transmissivity field which is conditional to the transmissivity measurements.
2. *Solution of the groundwater flow equation.* The steady or transient groundwater flow equation is solved for the transmissivity field with given boundary conditions by finite differences. A fully implicit scheme is used to solve the transient groundwater flow equation. It is possible to handle cases with spatially variable fluid density.
3. *The objective function.* The model fit criterion is a weighted sum of the squared deviations between the computed and measured pressures at each location and time where they are available. During the calibration process the fit between computed and measured pressures improves. In case of transient groundwater flow first the inverse problem for steady groundwater flow is solved (so that the solution of the flow equation is close to the measured steady pressure values) and afterwards the transient flow problem. The steady state calibration should not be degraded during the course of transient calibration and therefore different weights can be given to the steady part of the objective function and the transient part.
4. *Calculation of sensitivity coefficients for the objective function.* In case the model fit between computed and measured pressures is not good enough, the sensitivity coefficients of the objective function with respect to a change of the transmissivity in each grid block are calculated. The objective is to determine the grid blocks with the highest sensitivities. These sensitivity coefficients are obtained through adjoint sensitivity analysis. Because of the considerable amount of CPU time needed to compute them, the authors developed an ad hoc method that does not require their exact computation at each grid block.
5. *Pilot Point Location.* A main characteristic of the method is the use of pilot points. A pilot point is defined by its spatial location and by the transmissivity value assigned to it. The pilot points are placed at locations where their potential for reducing the objective function is the highest, that is, the locations where the absolute value of the sensitivity coefficients is the largest. The authors found from computer experiments that using one or three pilot points in the optimization yields very similar results, because normally the grid cells with the highest absolute values for the sensitivity coefficients are located next to each other.
6. *Optimization and updating.* The authors distinguish between two optimization levels. At the inner optimization level, the transmissivity values at the pilot point locations are determined through an iterative non-linear optimization procedure, that alternates Fletcher Reeves algorithm, the Broyden algorithm and the Davidon-Fletcher-Powell algorithm. Once the optimal transmissivities at the pilot point locations are determined, the transmissivity field is updated using kriging and a new iteration of the outer optimization level starts. (Back to step 4 to recompute the sensitivity coefficients and determine new locations for the pilot points.)

7. *Convergence criteria.* At the inner iteration level, during which the transmissivities at the pilot point locations are calculated, a total of five convergence criteria are used to decide whether to stop the optimization process:

- (i) the objective function value drops below a prescribed minimum.
- (ii) the number of iterations equals a prescribed maximum number of inner iterations.
- (iii) the ratio of the norm of the gradient to the initial norm decreases below a prescribed value.
- (iv) the gradient norm is less than a prescribed minimum
- (v) the relative change in objective function falls below a prescribed value.

At the outer iteration level two convergence criteria are used, which determine whether the calibration process as a whole is finished. These convergence criteria are:

- (i) the objective function value drops below a prescribed minimum.
- (ii) the number of iterations equals a prescribed maximum number of outer iterations.

### **Some important characteristics of the method**

Many of the comments given about the Sequential Self Calibrated Method also hold for the Pilot Point Method; the method can handle cases with a variance of the  $\log_{10}$  transmissivity field above 1.0 and generates an ensemble of equally likely solutions to the groundwater flow inverse problem. See section §3.4.2 for more detailed comments about these advantages.

The main differences between the Pilot Point Method and the Self Calibrated Method stand from their different initial aims. The Pilot Point Method was originally devised for estimation purposes, in this respect the concept of pilot point was very appropriate: the single point that should be modified in a kriging map that would produce the best approximation to the measured piezometric heads. This concept can be applied and has been extended into a simulation context. However, in a simulation context, the idea of modifying sequentially an initially heterogeneous field is suboptimal. In this respect, the Self Calibrated Method, which was devised for stochastic simulation, does not focus on optimal location of pilot points but on a simple way to parameterize a perturbation to an initial seed field that can be easily optimized and that will ensure that the final updated field does not depart dramatically from the original seed field. The master blocks in the Self Calibrated Method and the pilot points in the Pilot Point Method have the same purpose, parameterize a perturbation of a transmissivity field. In the Self Calibrated Method, the master block perturbations cover the whole formation and are computed all at once, whereas, in the Pilot Point Method, the pilot point values are computed sequentially and are located at the locations with higher sensitivity coefficients. The authors of the Sequential Self Calibrated Method state that at least 3 master blocks per correlation range are necessary (Capilla et al., 1997). On the contrary, the authors of the Pilot Point Method state that the use of just one pilot point in each inner iteration is optimal (RamaRao et al., 1995).

The method is, due to its numerical approach, flexible to handle complex groundwater flow situations with mixed boundary conditions and non rectangular domains. The method can already handle variable density groundwater flow, estimation of different aquifer parameters and

three-dimensional groundwater flow. Further extensions are possible with main limitation the CPU time because the method is very CPU intensive.

### **Application of the methodology to the WIPP-site**

The methodology has been applied for the inverse modelling of groundwater flow at the WIPP-site (LaVenue et al., 1995). Before we already presented a brief site description of the site (see section §3.4.2).

The authors assume a 2-D flow system, because of the existence of low permeability beds above and below the Culebra formation. The northwestern boundary of the modelling domain is partly a no flow boundary and along the other boundaries prescribed heads are imposed. The prescribed head values are obtained by kriging the observed heads. The variogram and the measured transmissivities are used to generate conditional simulations of transmissivity. The calibration of the transmissivity field is constrained to the kriged transmissivity plus or minus three times the kriging standard deviation.

In total 70 transmissivity fields are calibrated to steady state heads and transient state heads. The transient state heads come from three different pumping tests, all performed in 1988. An analysis of the resulting transmissivity fields after calibrating to steady state heads shows that the root mean squared error (RMSE) between calculated and observed steady state heads is between 1.5 and 2.5 m for most of the fields. It is found that if the initial mismatch between simulated and measured heads is larger, more pilot points have to be placed to calibrate the transmissivity field.

The calibrated transmissivity fields have been used in the simulation of non reactive mass transport. Particles are released from a point in the Culebra formation which coincides with the centroid of the waste panels and are transported to the southern WIPP site boundary. For each transmissivity field the travel time to this boundary and the travel path are determined. Travel times were also computed for a set of unconditional transmissivity field and for a set of transmissivity fields conditional only to the transmissivity values. It is found that the more conditioning data are used, the less is the uncertainty in travel time estimates and travel paths. Comparing the impact of using additional data, it is evident how the uncertainty in the advective transport predictions is progressively reduced as additional types of data are used, starting with transmissivity data, then steady state head data, and finally, transient state head data.

The uncertainty on the boundary conditions has not been subject of investigation in this study. However, the authors expect that considering uncertainty in the boundary conditions would not change the end results, because the large drawdowns that have been produced in the Culebra are much more significant during the calibration process than the uncertainty in the boundary conditions.

### 3.4.4 The Markov Chain Monte Carlo Method

#### Methodology

##### *Introduction*

The Markov Chain Monte Carlo Method (MCMC) (Oliver et al., 1996a; Oliver et al., 1996b) has the same objectives as the methods we have discussed in sections §3.4.2 and §3.4.3; that is, to generate an ensemble of equally likely transmissivity/conductivity fields conditional to hydraulic head data, transmissivity/conductivity data and possibly other prior information. In earlier work, Oliver and others (Oliver, 1994; Chu et al., 1995; Oliver, 1996) attempted the generation of equally likely transmissivity fields by adding error to the smooth single best solution of the inverse groundwater flow problem. However, the two papers we discuss here start with simulations of transmissivity/conductivity (not solutions to the inverse problem) and each of them is then conditioned to hydraulic head data. MCMC determines the ensemble of realizations by an optimization approach different from the non linear optimization algorithms used in the Sequential Self Calibrated Method and the Pilot Point Method. MCMC can handle 1-D, 2-D and 3-D steady and transient groundwater flow.

##### *The steps in the calibration process*

The authors do not describe, in a systematic way, the steps taken in the MCMC for the generation of the conditional transmissivity fields. From the information in the two papers by Oliver et al. (1996a, 1996b), they can be summarized as follows:

1. *Simulation of a transmissivity field.* A transmissivity field, conditional to transmissivity measurements, is generated by a Markov chain method. The conditional simulation of a transmissivity field is approached as a minimization problem, with the weighted sum of the misfit of the variogram of the simulated field with respect to the objective variogram as the objective function. As a result, the simulated transmissivity field will reproduce the hard transmissivity data and the model variogram. This simulation technique is based on the same grounds as the simulated annealing techniques described by Deutsch and Journel (1992). Oliver et al (1996b) propose a computationally efficient algorithm for the generation of the fields.
2. *Solution of the groundwater flow equation.* The full steady or transient groundwater flow equation is solved by finite differences for the transmissivity field with given boundary conditions.
3. *The objective function.* The objective function to be minimized is given by the sum of the weighted squared deviations between measured and simulated heads.
4. *Calculation of sensitivity coefficients.* The sensitivities of the model responses for changes in the transmissivities in the domain are calculated. The calculation of the sensitivity coefficients is based on a linearization of the groundwater flow equation by a Taylor series expansion.

5. *A transition is made.* The Markov chain method is used again to propose a transition of the transmissivity field. The information of the sensitivities is used to propose transitions with a high probability of acceptance. If the new transmissivity field results in a reduction of the objective function value, the transition is accepted, if not, the transition is accepted or rejected according to a given probability law. Oliver et al (1996b) consider two kind of transitions: global ones, in which new values of the transmissivity are drawn at all locations and local ones, in which just the permeability at a single location is altered. Global transitions yield a series of independent realizations of the transmissivity field, while local transitions result in a series of inter-correlated realizations. It is easier and computationally more efficient to generate and evaluate local transitions, however, it is not easy to determine how many local transitions are necessary to obtain independent realizations. Oliver et al (1996b) apply suggestions from the literature to determine the number of local transitions to generate independent realizations.
6. *Convergence criteria.* Oliver et al (1996a, 1996b) do not specify explicitly their criteria to stop the perturbation of the transmissivity field. They refer to «a sufficiently close reproduction of the heads» and a «very small convergence rate».

### **Numerical experiments**

Oliver et al (1996b) consider 2-D transient groundwater flow in an area of 457 m by 457 m discretized into 225 square grid blocks. The four edges are no flow boundaries. A pumping well is located in the center of the domain. Two  $\log_{10}$  transmissivity fields are generated, each with a mean of 3.4 but with variances of 0.25 and 1.0. The 2-D spatial continuity is described by an isotropic spherical variogram with a range of 183 m. The groundwater flow equation is solved for the two fields and piezometric head samples are obtained at five locations. At each measurement location ten samples in time are taken the start of the pumping. Equally likely transmissivity fields (solutions of the groundwater flow inverse problem) are generated according three methods. In the first method, local transitions are made without using sensitivity information from the groundwater flow model to propose these transitions. In the second and third methods, the sensitivity coefficients are used to improve the acceptance probability of the transitions. The second method uses local transitions and the third one uses global transitions. The three methods are evaluated by the number of independent realizations of transmissivity conditional to hydraulic head data obtained in a certain amount of time.

For the reference transmissivity field with the lowest variance it is found that the third method yields the best results in the sense that the largest number of independent realizations (namely 200) is generated in a certain amount of time. The second method yields less favorable results, 1200 local transitions are needed to generate independent realizations (only 31 independent realizations are generated). The first method yielded no conditional realizations: not considering the sensitivity information coming from the groundwater flow equation prevented the method to generated realizations reproducing the measured hydraulic heads.

For the case of the reference  $\log_{10}$  transmissivity field with a variance of 1.0 the results are quite different. Again the first method yielded no conditional realizations, but, this time, the second method yielded more independent realizations than the third one (14 versus 8). For both methods, the number of independent realizations was smaller as compared with the case of a variance of 0.25. It seems that for slightly nonlinear problems, global transitions are the

best to obtain the largest number of independent realizations of the transmissivity field. However, for highly nonlinear problems, local transitions are the most efficient to obtain the largest number of independent realizations, because it is difficult to determine a global perturbation yielding a transmissivity field that matches the hydraulic head data. A series of local transitions, on the contrary, is more likely to produce a field matching the hydraulic head data.

### **Some important characteristics of the method**

MCMC like the Sequential Self Calibrated Method and the Pilot Point Method, is not limited to a small variance of transmissivity, and it is flexible enough to handle complex geometries and boundary conditions. See sections §3.4.2 and §3.4.3 for more details. Below we will focus on the differences between MCMC and the Sequential Self Calibrated Method and the Pilot Point Method.

In the MCMC the initial transmissivity field is generated with a variogram very close to the model variogram. The other methods generate an initial transmissivity field which, due to ergodic fluctuations, does not necessarily reproduce the variogram, if the domain size is not much larger than the correlation length. As such, we expect that the ensemble of initial transmissivity fields shows less dispersion for MCMC than for Sequential Self Calibrated or Pilot Point. In addition, during the calibration of the transmissivity fields to the hydraulic head data, MCMC ensures that the variogram of transmissivities is not altered. The self calibrated method and the pilot point method do not constrain the transmissivity variogram during the calibration of the fields to the head data, and, in this respect, they can be utilized as variogram identification algorithms.

The optimization method used in MCMC is very different from the method used in the other two methods. While the other two methods use standard non linear optimization algorithms, MCMC uses simulated annealing. The use of sensitivity coefficients to determine the transitions from one transmissivity field to the next increases the acceptance probability of any given transitions. Oliver et al (1996a) state that although the proposed transitions are based on the posterior linearization of the groundwater flow equation, the «true» distribution of transmissivity fields (conditioned to hydraulic head data) is sampled as long as the «correct» acceptance criterion for transitions is used.

The numerical examples presented by Oliver et al. (1996b) indicate that MCMC is extremely CPU intensive, much more than the Pilot Point Method and the Sequential Self Calibrated Method. The proposals of Oliver et al (1996b) result in an important reduction of costs for the application of the simulated annealing method as optimization tool in the inverse modelling of groundwater flow; however, the method is still not efficient enough as compared to the more «classical» non linear optimization algorithms.

## 3.5 The Fractal Simulation Method

### Methodology

#### *Introduction*

The Fractal Simulation method by Grindrod and Impey (1991) is different from all the discussed methods in sections §3.2 §3.3 and §3.4 because the hydraulic head data are not used to improve the estimation of the transmissivity field. The transmissivity field reproduces the hydraulic head values by adjusting the heads on the boundaries of an artificially large field that enclosed the area of interest. The description that follows is based on the information contained in the paper by Zimmerman et al. (1996).

#### *Steps in the calibration process*

Consider a steady state groundwater flow problem in a 2-D domain with prescribed head boundaries.

1. *Simulation of a fractal transmissivity field.* A spatially variable log transmissivity field is generated upon the assumption that the spatial variability of the log transmissivity is best represented by a fractal process. The experimental variogram of the log transmissivity data is fitted by a power-law (fractal) variogram. The two parameters which appear in the expression of the fractal variogram are computed using maximum likelihood estimation. A fractal log transmissivity field is generated using the Fast Fourier Transform Method (Gutjahr and Wilson, 1989) over a field much larger than the area of interest. The field is conditioned to transmissivity data through a linear superposition of unconditioned fields, where the coefficients are determined so that the transmissivity data is honored and the variance of the final transmissivity field is as close as possible to that of the data.
2. *Solution of the groundwater flow equation.* The steady groundwater flow equation is solved for the generated transmissivity field with prescribed head boundary conditions. It is not specified which numerical method is used to solve the groundwater flow equation.
3. *The calibration process.* The calibration of the transmissivity field to the measured heads is achieved by modifying the prescribed heads at the boundaries. The perturbations at the boundaries are computed using an optimization procedure. The calibration to the heads does not modify the transmissivity field.

### Some important characteristics of the method

Fundamental in the Fractal Simulation approach is the representation of the transmissivity field by a fractal process. Although this could be acceptable in some instances, does not necessarily have to hold true in all cases.

In the Fractal Simulation Method the head data are not used to improve the characterization of the transmissivity field, as is done by all other methods. The measured heads are calibrated by adjusting the prescribed heads at the boundaries. It is a disadvantage that the method does not

use the hydraulic head data to update the transmissivity field. Some other methods calibrate the boundary conditions and the transmissivity field jointly.

The Fractal Simulation Method is not limited to transmissivity fields with a small variance, since it does not make any small variability assumption. Furthermore, the method aims to calibrate a series of equally likely transmissivity fields. We already explained before the benefits of obtaining a series of equally likely transmissivity fields as compared to a single best calibrated transmissivity field.

### **Practical applications of the method**

Applications of the Fractal Simulation Method are not available to us. However, it is known that the calibration process may yield unrealistic hydraulic head values on the boundaries in order to fit the head data. The method also participated in the comparison of seven geostatistically-based inverse approaches to estimate transmissivities for modeling advective transport (Zimmerman et al., 1996).

## Main references on inverse modelling

1. Capilla, J., Gómez-Hernández, J.J. and Sahuquillo, A., 1997, Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data. 2. Demonstration on a synthetic aquifer: *Journal of Hydrology*, 203, 175-188.
2. Capilla, J., Gómez-Hernández, J.J. and Sahuquillo, A., 1998, Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data. 3. Application to the Culebra formation at the waste isolation pilot plan (WIPP), New Mexico, USA: Accepted for publication in *Journal of Hydrology*.
3. Carrera, J. and Glorioso, L., 1991, On geostatistical formulations of the groundwater flow inverse problem: *Advances in Water Resources Research*, 14(5), 272-283.
4. Carrera, J. and Neuman, S.P., 1986a, Estimation of aquifer parameters under transient and steady state conditions: 1. Maximum likelihood method incorporating prior information: *Water Resources Research*, 22(2), 199-210.
5. Carrera, J. and Neuman, S.P., 1986b, Estimation of aquifer parameters under transient and steady state conditions: 2. Uniqueness, stability and solution algorithms: *Water Resources Research*, 22(2), 211-227.
6. Carrera, J. and Neuman, S.P., 1986c, Estimation of aquifer parameters under transient and steady state conditions: 3. Application to synthetic and field data: *Water Resources Research*, 22(2), 228-242.
7. Dagan, G., 1985, Stochastic modeling of groundwater flow by unconditional and conditional probabilities: The inverse problem: *Water Resources Research*, 21(1), 65-72.
8. Dagan, G. and Rubin, Y., 1988, Stochastic identification of recharge, transmissivity, and storativity in aquifer transient flow: a quasi-steady approach: *Water Resources Research*, 24(10), 1698-1710.
9. Deutsch, C. V. and A. G. Journel, 1992, *GSLIB, Geostatistical Software Library and User's Guide*, Oxford University Press, 369 pp.
10. Gómez-Hernández, J.J., Sahuquillo, A. and Capilla, J.E., 1997, Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data. 1. Theory: *Journal of Hydrology*, 203, 162-174.
11. Graham, W. D. and Tankersley, C. D., 1994, Optimal estimation of spatially variable recharge and transmissivity fields under steady-state groundwater flow. Part 1. Theory: *Journal of Hydrology*, 157, 247-266.
12. Graham, W. D. and Neff, C. R., 1994, Optimal estimation of spatially variable recharge and transmissivity fields under steady-state groundwater flow. Part 2. Case study: *Journal of Hydrology*, 157, 267-285.
13. Grindrod, P. and Impey, M.D., 1991, *Fractal field simulations of tracer migration within the WIPP Culebra dolomite*. Report. Intera Information Tech..
14. Gutjahr, A. and Wilson, J., 1989, Co-kriging for stochastic flow models: *Transport in Porous Media*, 4(6), 585-598.
15. Gutjahr, A., Bullard, B., Hatch, S. and Hughson, L., 1994, Joint conditional simulations and the spectral approach for flow modeling: *Stochastic Hydrology and Hydraulics*, 8, 79-108.
16. Hoeksema, R.J. and Kitanidis, P.K., 1984, An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling: *Water Resources Research*, 20(7), 1003-1020.

17. Hoeksema, R.J. and Kitanidis, P.K., 1989, Prediction of transmissivities, heads and seepage velocities using mathematical modeling and geostatistics: *Advances in Water Resources*, 12(2), 90-102.
18. Kitanidis, P.K., 1986, Comment on «stochastic modeling of groundwater flow by unconditional and conditional probabilities: the inverse problem» by Gedeon Dagan: *Water Resources Research*, 22(6), 984-986.
19. Kitanidis, P.K., 1996, On the geostatistical approach to the inverse problem: *Advances in Water Resources Research*, 19(6), 333-342.
20. Kitanidis, P.K. and Vomvoris, E.G., 1983, A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations: *Water Resources Research*, 19(3), 677-690.
21. LaVenue, A.M., RamaRao, B.S., de Marsily, G. and Marietta, M.G., 1995, Pilot point methodology to automated calibration of an ensemble of conditionally simulated transmissivity fields. 2. Application: *Water Resources Research*, 31(3), 495-516.
22. Oliver, D.S., He, N., Reynolds, A.C., 1996a, Conditioning permeability fields to pressure data: 5<sup>th</sup> European Conference on the Mathematics of Oil Recovery.
23. Oliver, D.S., Cunha, L.B., Reynolds, A.C., 1996b, Markov chain Monte Carlo methods for conditioning a permeability field to pressure data: *Mathematical Geology*, 29(1), 61-91.
24. RamaRao, B.S., LaVenue, A.M., de Marsily, G. and Marietta, M.G., 1995, Pilot point methodology to automated calibration of an ensemble of conditionally simulated transmissivity fields. 1. Theory and computational experiments: *Water Resources Research*, 31(3), 474-493.
25. Rubin, Y. and Dagan, G., 1987a, Stochastic identification of transmissivity and effective recharge in steady groundwater flow: 1. Theory: *Water Resources Research*, 23(7), 1185-1192.
26. Rubin, Y. and Dagan, G., 1987b, Stochastic identification of transmissivity and effective recharge in steady groundwater flow: 2. Case study: *Water Resources Research*, 23(7), 1193-1200.
27. Sahuquillo, A., Capilla, J.E., Gómez-Hernández, J.J. and Andreu, J., 1992, Conditional simulation of transmissivity fields honoring piezometric data, in *Hydraulic Engineering Software IV, Fluid Flow Modeling*, W.R. Blain and E. Cabrera (Eds.), pp. 201-214.
28. Wen, X.-H., Gómez-Hernández, J. J., Capilla, J. E. and Sahuquillo, A., 1996, Significance of conditioning to piezometric head data for predictions of mass transport in groundwater modeling: *Mathematical Geology*, 28(7), 951-968.
29. Yeh, T.-C., Gutjahr, A.L., Jin, M., 1995, An iterative cokriging-like technique for groundwater flow modeling: *Ground Water*, 33(1), 33-41.
30. Yeh, T.-C., Jin, J.M., Hanna, S., 1996, An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields: *Water Resources Research*, 32(1), 85-92.

## Other referred literature

31. Carrera, J., Navarrina, F., Vives, L., Heredia, J. and Medina, A., 1990, Computational aspects of the inverse problem, in *Computational Methods in Subsurface Hydrology*. G. Gambolati and others (Eds.), 373-383, Springer-Verlag.
32. Chu, L., Reynolds, A.C. and Oliver, D.S., 1995, Computation of sensitivity coefficients for conditioning the permeability field to well-test procedure: *In Situ*, 19(2), 179-223.

33. Gómez-Hernández, J.J. and Wen, X.-H., 1998, To be or not to be MultiGaussian? A reflection on stochastic hydrogeology: *Advances in Water Resources Research*, 21(1), 47-61.
34. Goovaerts, P., 1997, *Geostatistics for natural resources evaluation*. 483 p. Oxford University Press.
35. Grenier, C., Mouche, E. and Dewiere, L., 1996, Stochastic modeling of groundwater flow by the method of conditional probabilities: theoretical study and numerical evaluation, in *Calibration and reliability in groundwater modeling* (Proceedings of the ModelCARE'96 Conference), IAHS Publ. no. 237, p. 95-104.
36. Hughson, D.L., 1997, *A geostatistical inverse method for conditioning randomly heterogeneous transmissivity fields on transient hydraulic head data*. PhD Thesis. 220 p. New Mexico Institute of Mining and Technology.
37. Isaaks, E.H. and Srivastava, R.M., 1989, *An Introduction to Applied Geostatistics*. 561 p. Oxford University Press.
38. Journel, A.G. and Huijbregts, C., 1978, *Mining Geostatistics*. ??? p. Academic Press.
39. Marsily, G. de, Lavedan, C., Boucher, M. and Fasanino, G., 1984, Interpretation of inference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model, in *Geostatistics for natural resources characterization*, Verly, G., David, M., Journel, A.G. and Marechal, A. (Eds.), NATO ASI Ser., Ser. C, 182, p. 831-849.
40. Medina, A., 1993, *Estimación conjunta de parámetros de las ecuaciones de flujo y transporte*. PhD Thesis. Universidad Politécnica de Catalunya.
41. Medina, A. and Carrera, J., 1996, Coupled estimation of flow and solute transport parameters: *Water Resources Research*, 32(10), 3063-3076.
42. Oliver, D.S., 1994, The incorporation of transient pressure data into reservoir characterization: *In Situ*, 18(3), 243-275.
43. Oliver, D.S., 1996, Multiple realizations of the permeability field from well test data: *Society of Petroleum Engineering Journal*, 1(2), 145-154.
44. Sun, N.-Z. and Yeh, W. W.-G., 1992, A stochastic inverse solution for transient groundwater flow: parameter identification and reliability analysis: *Water Resources Research*, 28(12), 3269-3280.
45. Zimmerman, D.A., de Marsily, G., Gotway, C.A., Marietta, M.G., Axness, C.L., Beauheim, R., Bras, R., Carrera, J., Dagan, G., Davies, P.B., Gallegos, D.P., Galli, A., Gómez-Hernández, J.J., Gorelick, S.M., Grindrod, P., Gutjahr, A.L., Kitanidis, P.K., LaVenue, A.M., McLaughlin, D., Neuman, S.P., RamaRao, B.S., Ravenne, C. and Rubin, Y., 1996, A comparison of seven geostatistically-based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow: To be submitted to *Water Resources Research*.



**4. CHANGEMENT D'ECHELLE DE PROPRIETES  
D'ECOULEMENT ET DE TRANSPORT**



## 4.1 Perméabilités de grande échelle en milieu hétérogène

### 4.1.1 Introduction

Un domaine d'étude important en simulation numérique d'écoulement souterrain et de transport de masse est lié au problème d'échelle. Les données sont collectées à une échelle différente de celle utilisée pour discrétiser (mailler) l'aquifère dans un modèle numérique et souvent plus petite. Par exemple, les mesures de conductivité hydraulique sur un champ (par ex. à partir de tests sur carottes ou d'essais hydrogéologiques in situ) sont représentatives de la perméabilité sur un rayon d'investigation de quelques centimètres à quelques mètres alors que les modèles numériques d'écoulement en milieu poreux requièrent des perméabilités affectées à des mailles de quelques dizaines à quelques centaines de mètres de côté.

Ce chapitre est consacré aux différentes approches qui, depuis plus de trente ans, ont été utilisées pour transformer une description détaillée de la variabilité spatiale de conductivités hydrauliques en une description plus grossière comme cela est décrit sur la Figure 4-1. On fera référence dans la suite à l'échelle correspondant à la description détaillée de la Figure 4-1a comme échelle de la mesure (ou échelle fine) et à l'échelle correspondant à la description grossière de la Figure 4-1b comme échelle numérique (ou grande échelle). On utilisera les termes de perméabilité de cellule et de perméabilité de bloc (ou perméabilité équivalente) pour distinguer les valeurs de perméabilité à l'échelle de la mesure et à l'échelle numérique. La manière d'affecter des valeurs de perméabilité aux cellules d'une grille à l'échelle de la mesure, à partir d'un ensemble de données directes et indirectes, a fait l'objet des chapitres consacrés à l'estimation (§2.2 et §2.3) et à la simulation stochastique (§2.4 et §2.5). Ces aspects ne seront donc pas abordés davantage ici. On supposera simplement que des perméabilités de cellule ont été calculées préalablement par une méthode (géo)statistique appropriée. Le problème auquel on s'intéressera est donc de déterminer, à partir d'une ou plusieurs grilles contenant des perméabilités à l'échelle de la mesure, les perméabilités de bloc d'une grille à l'échelle numérique. La grille affectée de propriétés de bloc est alors destinée à un simulateur d'écoulement en milieu poreux.

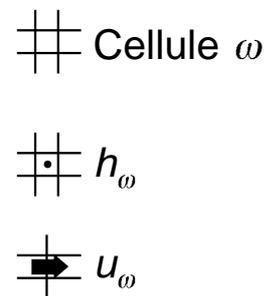
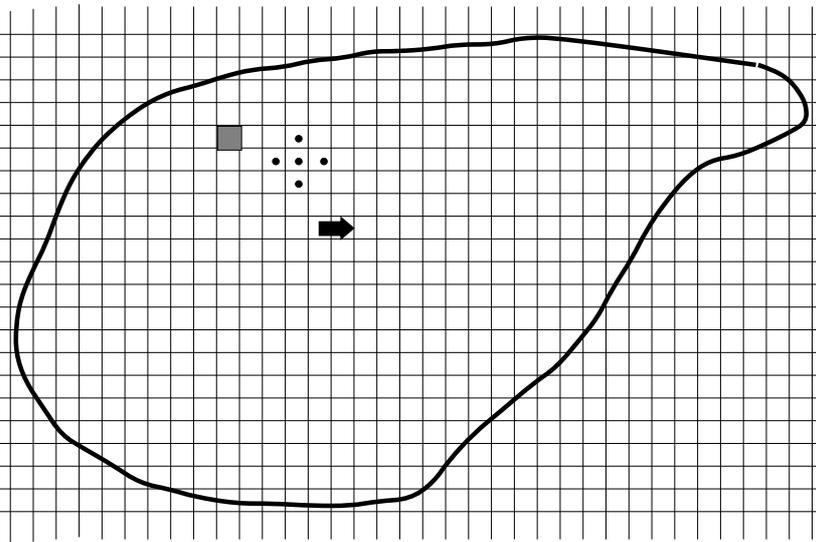
Certains auteurs utilisent le terme de perméabilité effective pour parler de la perméabilité de bloc. La perméabilité effective ne sera pas employée dans ce sens ici pour éviter la confusion avec sa signification particulière en hydrogéologie stochastique. La perméabilité effective définie en hydrogéologie stochastique est une propriété du modèle de fonction aléatoire adopté pour décrire la variabilité spatiale de la perméabilité. En tant que telle, elle ne détermine pas une propriété de bloc mais une propriété de point qui représente une certaine valeur moyenne calculée à partir d'un ensemble de réalisations possibles de la perméabilité au point considéré. Ce n'est que pour certains modèles de fonctions aléatoires que la perméabilité effective est constante partout dans l'espace (et peut donc être affectée à un milieu homogène équivalent) et coïncide avec la perméabilité d'un bloc unique infini (Matheron, 1967, Dagan, 1979, 1982a,b, 1989). Dans ce cadre, les perméabilités effectives peuvent s'utiliser pour vérifier partiellement la validité de techniques de calcul de la perméabilité équivalente. Bien que les perméabilités effectives ont fait l'objet de nombreuses études (par ex., Matheron, 1967, Dagan, 1979, 1982a, 1989, Gutjahr et al., 1978, Gelhar et Axness, 1983, Poley, 1988, Ababou et Wood, 1990, Naff, 1991, Neuman et Orr, 1993) elles ne sont pas abordées ici. Les lecteurs intéressés pourront se reporter sur le travail de Neuman et Orr (1993), qui présentent les résultats les plus généraux,

ou celui de Renard (1997) qui s'y intéresse comme moyen de validation des techniques de calcul de la perméabilité équivalente.

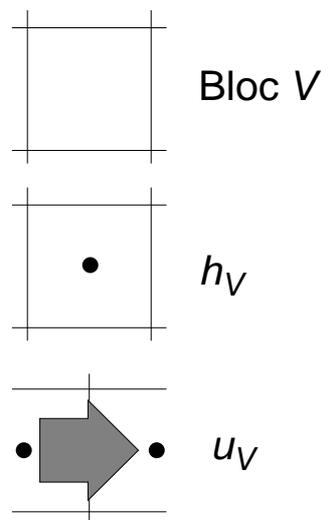
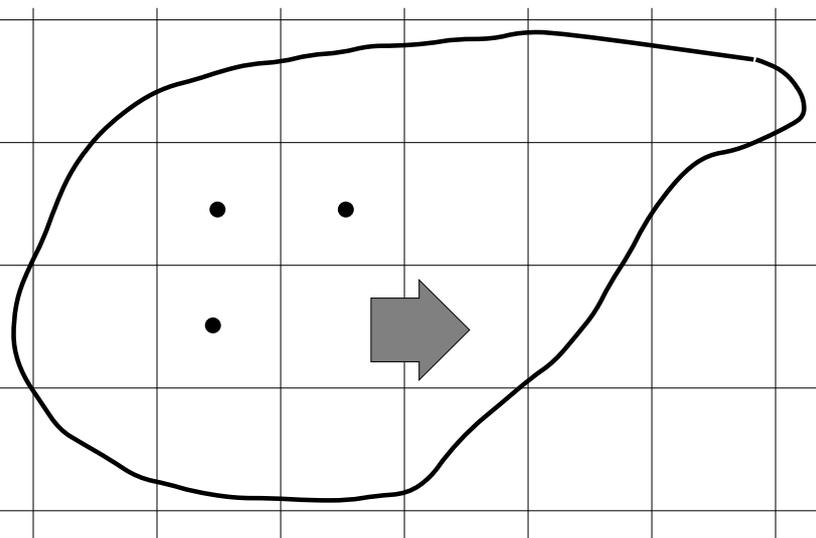
Processus de passage à grande échelle. a) Grille avec une résolution conforme à l'échelle de la mesure et pouvant contenir plusieurs millions de cellules . b) Grille à l'échelle de la simulation numérique utilisée pour résoudre le problème d'écoulement en milieu poreux. Le calcul de valeurs de perméabilité de bloc en b) à partir des valeurs correspondantes en a) est l'objet du passage à grande échelle.

Figure 4-1

a) Echelle de la mesure



b) Echelle numérique



Ce chapitre se divise en trois parties. La première définit la perméabilité de bloc. La deuxième, la plus importante, passe en revue, par ordre de complexité croissante, les méthodes de calcul de la perméabilité équivalente. Ne sont abordées que les méthodes destinées à des milieux saturés pour des écoulements monophasiques. Pour le calcul de perméabilités équivalentes dans des milieux non saturés, le lecteur intéressé pourra se reporter aux travaux de Yeh *et al.* (1985), Mantoglou et Gelhar (1987) et Russo (1992) ou pour des écoulements polyphasiques à ceux de Haldorsen (1986), Muggeridge (1991) et Fayers et Hewett (1992). La troisième partie contient la liste des références bibliographiques sur le sujet.

### 4.1.2 Perméabilité de bloc

L'utilisation de qualificatifs tels que équivalente, de bloc, effective et de grande échelle appliqués à la perméabilité, plutôt que le terme plus simple de perméabilité moyenne, reflète le fait que la perméabilité n'est pas une propriété additive des milieux poreux. Il est bien connu, par analogie à la conductivité électrique, que la perméabilité équivalente d'un groupe de cellules disposées en série est égale à leur moyenne harmonique, tandis que pour des cellules disposées en parallèle elle est donnée par leur moyenne arithmétique. Cardwell et Parsons (1945) ont montré par ailleurs que la perméabilité équivalente d'un bloc hétérogène est compris entre la moyenne arithmétique (borne supérieure) et la moyenne harmonique (borne inférieure) des perméabilités des cellules situées dans le bloc.

#### 4.1.2.1 Définition

Conformément à Rubin et Gómez-Hernández (1990), la perméabilité de bloc, notée  $\mathbf{K}_V$  dans la suite, se définit, par extension de la loi de Darcy à l'échelle numérique, comme :

$$E\ 4-1 \quad \frac{1}{V} \int_V \mathbf{u}_\omega dv = -\mathbf{K}_V \left( \frac{1}{V} \int_V \nabla h_\omega dv \right)$$

où  $V$  représente le volume du bloc,  $\mathbf{u}_\omega$  le vecteur vitesse de Darcy à l'échelle de la mesure et  $\nabla h_\omega$  le gradient de charge hydraulique à l'échelle de la mesure également. L'intégration sur les vecteurs s'effectue par intégration des termes. L'équation précédente signifie que la perméabilité de bloc est un tenseur qui relie la vitesse de Darcy moyenne à l'intérieur du bloc au gradient de charge hydraulique moyen à l'intérieur du bloc.

Les perméabilités de cellule  $\mathbf{K}_\omega$  sont implicites dans l'équation E 4-1 car  $\mathbf{u}_\omega$  et  $\nabla h_\omega$  sont reliés, à l'échelle de la mesure, à travers par  $\mathbf{K}_\omega$  par la relation suivante :

$$\mathbf{u}_\omega = -\mathbf{K}_\omega \nabla h_\omega.$$

Il est à noter que, par définition,  $\mathbf{K}_V$  est, de la même manière que  $\mathbf{K}_\omega$ , un tenseur.

Dans les différentes approches de calcul de la perméabilité équivalente décrites dans la suite, la définition E 4-1 n'est pas toujours adoptée explicitement. Dans certains cas, cette formulation

est implicite, dans d'autres ce sont des définitions ou critères différents qui s'appliquent. Ces autres définitions seront précisées plus loin avec la présentation des approches qui s'y rapportent. Dans tous les cas, l'objet de la procédure de calcul de la perméabilité à grande échelle est d'obtenir, à l'échelle numérique, une description de la variabilité spatiale de la perméabilité qui reproduit certains comportements moyens du champ de perméabilité à l'échelle de la mesure.

### 4.1.3 Evolution des méthodes de calcul de la perméabilité équivalente

Les méthodes de calcul de la perméabilité équivalente sont classées dans la suite en quatre groupes qui sont décrits successivement. Les groupes sont définis comme suit.

- Méthodes locales : elles supposent que le tenseur de perméabilité de bloc est intrinsèque, c'est-à-dire qu'il ne dépend que des perméabilités de cellule à l'intérieur du bloc.
- Méthodes non locales : dans ces méthodes, la perméabilité de bloc ne dépend pas seulement des perméabilités de cellule à l'intérieur du bloc mais également des conditions aux limites d'écoulement nécessaires pour déterminer  $\mathbf{u}_\omega$  et  $\nabla h_\omega$  dans l'équation E 4-1.
- Méthodes dans lesquelles la définition de la géométrie des blocs fait partie du passage à grande échelle.
- Ensemble de méthodes pour la génération directe de grilles à l'échelle numérique à partir des données à l'échelle de la mesure.

#### 4.1.3.1 Méthodes locales

Les méthodes locales partent de l'hypothèse que la perméabilité de bloc est intrinsèque et dépend donc uniquement des perméabilités de cellule à l'intérieur du bloc. Ceci revient à assimiler la perméabilité de bloc à la propriété d'un matériau. Les méthodes locales sont la généralisation naturelle de résultats mono-dimensionnels pour lesquels il est connu que la perméabilité de bloc est égale à la moyenne harmonique des perméabilités hétérogènes quelles que soient les conditions d'écoulement. Dans les méthodes locales, la perméabilité de bloc est une fonction explicite des perméabilités de cellule. Les techniques de ce type décrites dans la suite reposent sur l'hypothèse que le tenseur de perméabilité de cellule  $\mathbf{K}_\omega$  se réduit à un scalaire  $K_\omega$ . Une exception est la méthode récente de renormalisation simplifiée de Renard (1997) qui tient compte d'une anisotropie de perméabilité (axes d'anisotropie alignés sur la grille de propriétés) et semble pouvoir s'appliquer dans certains cas de faible anisotropie. Certaines méthodes, par moyenne simple ou par renormalisation, supposent également que le tenseur de perméabilité de bloc  $\mathbf{K}_V$  est aussi un scalaire  $K_V$ . Si la première hypothèse (sur les cellules) peut se justifier, il n'en est pas de même de la seconde car la nature tensorielle des perméabilités de bloc est en partie la conséquence de la corrélation spatiale anisotrope des perméabilités de cellule (Lake, 1988), même dans le cas de perméabilités de cellule scalaires.

#### Moyenne simple

Pour des écoulements 1-D, la perméabilité de bloc est égale à la moyenne harmonique des perméabilités de cellule dans le bloc. Pour des écoulements 2D, il existe un résultat important de la théorie des paramètres effectifs (Matheron, 1967). Ce résultat s'applique aux blocs infinis constitués de perméabilités hétérogènes dont la corrélation spatiale est isotrope et dont la loi de distribution est la même que pour l'inverse de la perméabilité (comme cela est le cas si la perméabilité suit une loi log-normale). Matheron (1967) montre que, sous ces conditions, la perméabilité de bloc est égale à la moyenne géométrique. Ce résultat, qui en principe ne s'applique qu'aux blocs infinis, a été très largement appliqué, sans grande justification, à des blocs finis dans la littérature en hydrogéologie et dans le domaine pétrolier (par ex. Dagan, 1982b, 1985, Clifton et Neuman, 1982, Hoeksema et Kitanidis, 1984, 1985).

Warren et Price (1961) ont conclu, à partir de tests numériques en 3D, que la moyenne géométrique est la plus appropriée pour l'estimation de perméabilités de bloc à partir de valeurs de cellule non corrélées dans l'espace. Bouwer (1969) a utilisé des simulations analogues pour arriver à la même conclusion.

Grindheim (1990) et Durlowski (1992) ont exploré l'utilisation de méthodes de moyenne simple pour calculer la perméabilité de bloc en 2D pour différentes répartitions spatiales des perméabilités de cellule (non corrélées et corrélées, statistiquement isotropes et anisotropes, répartition binaire sable-argile). Parmi les moyennes évaluées, on trouve les moyennes arithmétique, géométrique, harmonique, arithmétique-harmonique et harmonique-arithmétique. En comparant les résultats d'écoulement sur la grille à l'échelle de la mesure (grille fine) et sur la grille à l'échelle numérique (grille grossière), les deux auteurs sont arrivés à la conclusion qu'il n'existe aucune moyenne simple valide pour toutes les formations hétérogènes.

Gómez-Hernández et Wen (1994) ont montré que, en 2D, la moyenne géométrique donne de bons estimateurs de la perméabilité de bloc tant que la variabilité spatiale des perméabilités de cellule ne révèle pas une forte anisotropie.

### **Moyenne de puissance**

Cardwell et Parsons (1945) ont montré que les perméabilités de bloc sont nécessairement comprises entre les moyennes arithmétique et harmonique des perméabilités de cellule dans le bloc. A partir de ce résultat, Journel *et al.* (1986) ont proposé d'utiliser une moyenne de puissance pour le calcul de la perméabilité équivalente de bloc :

$$K_V = \left( \frac{1}{V} \int_V K_\omega^p dV \right)^{1/p}.$$

En faisant varier  $p$  entre -1 et 1, la moyenne de puissance varie entre la moyenne harmonique et la moyenne arithmétique, en passant par la moyenne géométrique pour  $p = 0$ . Ces auteurs défendent l'idée que l'exposant  $p$  dépend de la spécificité des hétérogénéités spatiales des perméabilités de cellules et peut donc être déduit par calibrage à partir des résultats de simulation des écoulements à l'échelle de la mesure (grille fine) sur un sous-domaine constitué de quelques blocs dont on cherche la perméabilité équivalente. La méthode a été appliquée avec succès par Deutsch (1989), Gómez-Hernández et Gorelick (1989), Bachu et Cuthiell (1990), Desbarats et Dimitrakopoulos (1990) et Desbarats (1992b). On peut retenir de ces applications les résultats suivants.

La valeur de l'exposant  $p$  est bien spécifique à chaque cas et dépend du type d'hétérogénéité dans le bloc, de la forme et de la taille de celui-ci et des conditions d'écoulement dans le bloc.

Pour des perméabilités de cellule anisotropes statistiquement, l'exposant  $p$  dépend de la direction et peut être utilisé pour déterminer les termes d'un tenseur de bloc sachant que l'on connaît déjà les directions principales d'anisotropie.

Il n'existe pas de manière simple de prédire la valeur de  $p$  sans avoir recours à des résultats de simulations à l'échelle de la mesure.

L'intérêt pour cette méthode se trouve dans sa simplicité, l'intervalle borné des valeurs du seul paramètre du modèle qui est l'exposant  $p$  et son application indifféremment aux problèmes 2-D et 3-D.

Lachassagne et al. (1989) se sont livrés à un exercice intéressant pour vérifier si le choix d'une méthode de résolution des équations d'écoulement à l'échelle numérique devait être pris en compte lors de la détermination de l'exposant  $p$ . Leur résultat est quelque peu dérangeant car il montre que, pour caler le flux total à travers un grand bloc hétérogène soumis à une différence de charge donnée, pour une loi log-normale de distribution des perméabilités de cellule et une covariance isotrope, la perméabilité équivalente dépend du simulateur d'écoulement utilisé par différences finies ou par éléments finis.

### **Renormalisation**

La méthode dite de renormalisation dans l'espace réel, développée initialement pour l'étude de phénomènes critiques en physique tels que la percolation (Kirkpatrick, 1973, Wilson, 1979, Shah et Ottino, 1986), a été appliquée pour la première fois au calcul de perméabilités de bloc par King (1989). La méthode se fonde sur un calcul de perméabilités équivalentes de très petits blocs (constitués de  $2 \times 2$  cellules ou sous-blocs en 2-D) et sur un algorithme itératif qui successivement calcule la perméabilité équivalente de blocs de plus en plus grands en combinant à chaque étape un nombre réduit de blocs de taille inférieure et ce jusqu'à atteindre la taille de bloc finale. La méthode est extrêmement rapide et n'est pas limitée par la taille du domaine ou la variance des perméabilités de cellule. La vitesse provient de l'expression analytique simple de la perméabilité équivalente de blocs constitués uniquement de  $2 \times 2$  cellules, expression empruntée aux lois de l'électricité pour un réseau de  $2 \times 2$  résistances.

Initialement, King (1989) a appliqué la renormalisation à des grilles 2-D de valeurs de perméabilité non corrélées dans l'espace. Plus tard, Shah et Ottino (1986) ont étendu la méthode au 3-D et Mohanty et Sharma (1990) à des champs corrélés.

La méthode donne d'excellents résultats (en comparant les simulations d'écoulements à l'échelle de la mesure et à celle des blocs) pour des champs de perméabilité isotropes statistiquement et log-normaux. Pour des milieux fortement anisotropes (par ex. formations sable-argile), la méthode ne se comporte pas bien à cause de la faible résolution des étapes successives de changement d'échelle à la frontière des argiles.

Williams (1989) a développé une méthode par grandes cellules fondée sur la renormalisation pour le calcul de perméabilités de bloc de milieux fortement anisotropes et constitués d'une fraction de matériaux non perméables. Depuis, Gautier et Nøttinger (1993) ont donné à la renormalisation une forme vectorielle qui, en introduisant des conditions aux limites périodiques, améliore les résultats mais ne permet toujours pas de traiter les cas fortement anisotropes. Renard (1997) a proposé une variante pour traiter les cas anisotropes mais avec une faible variance pour les valeurs de perméabilité de cellule.

Le net avantage de la renormalisation est sa vitesse. L'inconvénient majeur reste que, de par l'analogie à des réseaux de  $2 \times 2$  résistances électriques en 2-D, les conditions aux limites particulières appliquées aux côtés de chaque bloc de  $2 \times 2$  cellules peuvent être irréalistes même pour les méthodes les plus avancées comme celle de Gautier et Nøttinger (1993). Malick (1995) a démontré que lorsque ces conditions aux limites, qui peuvent être irréalistes, sont appliquées répétitivement au cours de la renormalisation, elles peuvent induire de grandes erreurs sur la perméabilité estimée du bloc final. Il a comparé les perméabilités de bloc obtenues par renormalisation (standard) aux perméabilités de bloc calculées à l'aide de la relation E 4-1 après résolution de l'équation d'écoulement à l'échelle de la mesure. L'erreur peut atteindre 70% pour une distribution log-normale des perméabilités et un coefficient de variation de 3. Elle monte jusqu'à 200% pour une formation sable-argile avec une proportion d'argile de 50%.

### **Tubes de courant**

La méthode des tubes de courant (Haldorsen et Lake 1984, Begg et King, 1985, Haldorsen et Chang, 1986, Begg et Carter, 1987, Begg *et al.*, 1985, 1989) a été conçue spécialement pour calculer la perméabilité de bloc de formations sable-argile, c'est-à-dire pour des formations dans lesquelles les argiles sont distribuées en corps homogènes dans une matrice de sable avec un fort contraste de perméabilités entre les argiles et les sables de sorte que les argiles peuvent être perçues comme des barrières aux écoulements verticaux. L'épaisseur des argiles est supposée être négligeable et leur direction être uniforme (par ex. la direction horizontale). Dans la direction parallèle aux argiles, leur épaisseur négligeable fait que la perméabilité de bloc est égale à la perméabilité de la matrice sableuse. Dans la direction perpendiculaire, les argiles agissent comme des barrières qui augmentent la tortuosité des lignes d'écoulement proportionnellement à la longueur et à la fréquence des lentilles d'argileuses. La perméabilité de bloc perpendiculairement aux argiles est liée aux longueurs de lignes d'écoulement à travers un facteur de tortuosité. Begg et King (1985) et Begg *et al.* (1985, 1989) ont déduit une expression de la perméabilité verticale de bloc comme une fonction de la proportion d'argile et des statistiques sur la longueur et l'épaisseur des argiles. Cette relation permet de déterminer une perméabilité verticale de bloc sans connaître la géométrie exacte des argiles dans le bloc.

La méthode des tubes de courant a été largement employée dans l'industrie pétrolière et s'est révélée capable de fournir de bons résultats. Desbarats (1987) a montré, cependant, que la méthode tend à sous-estimer la perméabilité verticale quand la densité des argiles est grande et que donc les voies d'écoulement deviennent très tortueuses.

### **Anisotropie de l'écoulement**

Parmi les méthodes présentées précédemment, celles par moyenne de puissance et par tubes de courant admettent des perméabilités de bloc tensorielles. Dans l'approche par moyenne de puissance, les directions principales de la perméabilité de bloc sont supposées être parallèles aux côtés du bloc (il en est de même avec les variantes de la méthode de renormalisation qui fournissent un tenseur de perméabilité de bloc). Une valeur différente de l'exposant  $p$  est alors calculée pour chaque terme diagonal du tenseur de perméabilité de bloc. Dans l'approche par tubes de courant, les directions principales de la perméabilité de bloc sont supposées être parallèle et orthogonale à l'orientation des lentilles argileuses. Dans certains cas, cependant, il n'est pas possible d'établir avec certitude les directions principales du tenseur de perméabilité de bloc, ni de faire l'hypothèse que les directions principales sont identiques pour tous les blocs.

Kasap et Lake (1989) se sont intéressés au calcul de tenseurs de perméabilité de bloc pour une stratification oblique à l'échelle de la mesure. Ceci correspond à un des cas où les directions principales du tenseur de perméabilité de bloc sont inconnues a priori et peuvent changer d'un bloc à l'autre. Ils ont développé une méthode analytique de calcul des tenseurs de perméabilité de bloc pour traiter les cas de perméabilité anisotrope à l'échelle de la mesure. Sous l'hypothèse que les lignes d'écoulement sont parallèles, leur méthode est fondée sur l'utilisation récursive des valeurs de perméabilité de bloc obtenues pour un bloc composé de deux couches homogènes mais de perméabilité anisotrope. Les équations qui déterminent les termes du tenseur de perméabilité de bloc ont été obtenues et validées à l'aide de simulations numériques.

#### **4.1.3.2 Méthodes non locales**

Bien que les méthodes locales permettent de déduire des perméabilités de bloc qui sont liées explicitement aux perméabilités de cellule dans le bloc, le fait que les valeurs ainsi obtenues peuvent dépendre de conditions aux limites implicites indique que les perméabilités de bloc ne sont pas locales, c'est-à-dire qu'elles ne sont pas intrinsèques au bloc considéré mais dépendent des conditions d'écoulement dans le bloc et par conséquent des conditions aux limites.

Dans les méthodes non locales discutées ci-après, la perméabilité de bloc n'est pas donnée comme une expression explicite des perméabilités de cellule dans le bloc; à la place, l'équation E 4-1 est appliquée directement après avoir déterminé les vecteurs vitesse de Darcy  $\mathbf{u}_\omega$  et les gradients de charge  $\nabla h_\omega$  par résolution de l'équation d'écoulement à l'échelle de la mesure. Les méthodes diffèrent entre elles dans le choix de l'étendue du domaine dans lequel est résolu numériquement l'écoulement à l'échelle de la mesure et dans les conditions aux limites employées pour le calcul de la solution. Toutes ces méthodes sont dites par résolution numérique ou Laplacienne (*Laplacian* en anglais) étant sous-entendu qu'elles se fondent sur la résolution de l'équation de Laplace (équation de diffusivité en régime permanent sans terme de source).

#### **Résolution numérique simple**

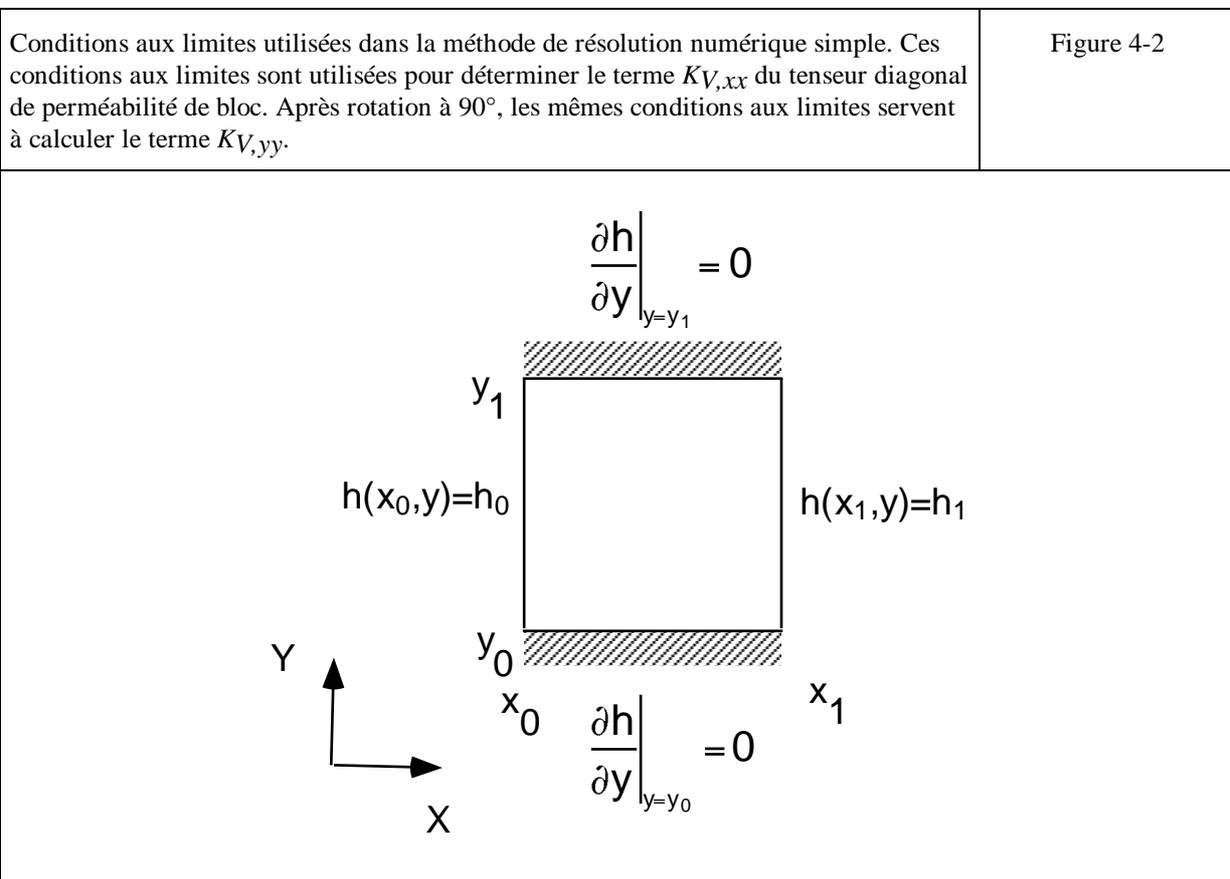
Dans la résolution numérique la plus simple, le bloc, pour lequel la perméabilité équivalente est calculée, est isolé des autres blocs de l'aquifère. Les directions principales du tenseur de perméabilité de bloc sont supposées être parallèles aux côtés du bloc et les termes diagonaux du tenseur sont calculés indépendamment par résolution numérique de l'équation d'écoulement

avec pour conditions aux limites des charges imposées sur les faces du bloc perpendiculaires à la direction principale considérée et des flux nuls sur les autres (frontières imperméables). La Figure 4-2 montre le problème d'écoulement à résoudre pour déterminer le terme en  $x$  du tenseur diagonal de perméabilité de bloc en 2-D. Un problème similaire doit être résolu pour trouver le terme en  $y$  après avoir effectué une rotation de  $90^\circ$  des conditions aux limites. L'application de l'équation E 4-1 à la solution de ces deux problèmes fournit une expression très simple des termes de la perméabilité de bloc comme le rapport entre la vitesse moyenne, prise sur n'importe quelle section perpendiculaire à la direction principale, et le gradient de charge entre les deux faces opposées où la charge est imposée. (Gómez-Hernández, 1991, Sánchez-Vila, 1995). Pour l'exemple 2-D de la Figure 4-2, l'expression du terme en  $x$  du tenseur diagonal de perméabilité de bloc est :

$$K_{V,xx} = -\left(\frac{Q}{y_1 - y_0}\right) \left/ \left(\frac{h_1 - h_0}{x_1 - x_0}\right)\right.$$

où  $Q$  est le flux total à travers n'importe quelle section perpendiculaire à l'axe  $x$  (et donc le rapport  $Q/(y_1 - y_0) =$  vitesse de Darcy moyenne),  $y_1 - y_0$  est la largeur du bloc,  $h_1 - h_0$  est la différence de charge hydraulique entre les faces opposées dans la direction de l'écoulement et  $x_1 - x_0$  est la longueur du bloc.

Warren et Price (1961) ont été les premiers à utiliser cette approche. Ils ont analysé la répartition aléatoire de perméabilités de cellule dans un bloc pour trouver que la moyenne géométrique était une valeur approchée de la perméabilité de bloc.



Les méthodes par résolution numérique sont supposées apporter une nette amélioration dans l'estimation de perméabilités de bloc par rapport aux méthodes locales décrites précédemment. Toutefois, dans l'approche par résolution numérique simple, il peut être critiqué les deux points suivants.

1. L'hypothèse concernant les directions principales du tenseur de perméabilité de bloc qui doivent être parallèles aux côtés des blocs.
2. Les conditions aux limites, utilisées pour résoudre l'équation d'écoulement à l'échelle de la mesure, car elles peuvent être différentes de celles qui existent en réalité à l'échelle du bloc.

### **Résolution numérique simple étendue**

Etant donné que la résolution numérique simple discutée précédemment nécessite de simuler un petit problème d'écoulement pour chaque bloc de la grille, Holden *et al.* (1989) ont essayé d'accélérer le méthode par une approche itérative. Si une approche itérative est employée pour trouver la solution de l'équation d'écoulement dans chaque bloc, il peut être calculé, à chaque pas d'itération, des valeurs approchées des termes du tenseur de perméabilité de bloc à partir des valeurs de  $\mathbf{u}_\omega$  et  $h_\omega$  déduites de l'itération précédente. L'idée fondamentale est de stopper le processus itératif avant que le calcul de la solution de l'équation d'écoulement n'ait convergé. Elle se fonde sur la conjecture (hypothèse) que le calcul des termes du tenseur de perméabilité de bloc converge plus rapidement que celui de la répartition spatiale des charges hydrauliques dans le bloc. Dans leur exemple de blocs rectangulaires en 2-D, les auteurs ont trouvé que la convergence des tenseurs de perméabilité de bloc vers une solution stable était obtenue en beaucoup moins d'itérations que pour la convergence des charges hydrauliques. Holden et Lia (1992) ont étendue davantage la méthode au calcul de tenseurs complets de perméabilité de bloc en 3-D.

### **Résolution numérique par calcul de la solution d'écoulement à l'échelle de la mesure sur l'ensemble de l'aquifère**

Les conditions aux limites utilisées dans la résolution numérique simple sont fixées arbitrairement pour faciliter le calcul de perméabilités directionnelles de bloc dans des directions parallèles aux côtés du bloc. Les perméabilités équivalentes ainsi obtenues étant fonction des conditions aux limites appliquées, il serait préférable d'utiliser des conditions aux limites les plus proches possibles de celles qui peuvent exister autour du bloc lorsque l'équation d'écoulement est résolue à l'échelle de la mesure sur l'ensemble de l'aquifère. Une telle approche remettrait en cause cependant le principe du changement d'échelle dont la vocation première est de fournir une description grossière de l'aquifère avec des perméabilités de bloc pour éviter les coûts prohibitifs de simulation des écoulements à l'échelle de la mesure.

Il existe un cas favorable qui justifie le calcul d'une solution globale de l'équation d'écoulement à l'échelle de la mesure pour obtenir des perméabilités de bloc : l'utilisation des perméabilités de bloc pour la simulation d'écoulements transitoires et en particulier d'écoulements polyphasiques. L'apparition de stations de travail de plus en plus rapides, les ressources informatiques disponibles pour les ingénieurs de réservoir et leur intérêt dans des prédictions plus précises et fiables par simulation d'écoulements polyphasiques ont suscité le

développement d'un ensemble de méthodes de calcul de perméabilités de bloc (saturé) à partir de la répartition spatiale dans le bloc des variables d'écoulement  $\mathbf{u}_\omega$  et  $\nabla h_\omega$  obtenues par résolution de l'équation d'écoulement (monophasique en régime permanent) à l'échelle de la mesure sur l'ensemble du réservoir. Beaucoup de ces méthodes sont détaillées et discutées dans la littérature pétrolière.

Cette quête pour des perméabilités de bloc plus précises découle du constat fait alors que des directions principales de perméabilité de bloc parallèles aux côtés du bloc n'étaient pas justifiables. Il fallait donc développer des méthodes capables de déduire des tenseurs de perméabilité de bloc avec des directions principales arbitraires.

White (1987) et White et Horne (1987) ont été les premiers à proposer une méthode pour déterminer des tenseurs complets de perméabilité de bloc et les premiers à suggérer l'utilisation de l'équation E 4-1 avec des valeurs de  $\mathbf{u}_\omega$  et de  $\nabla h_\omega$  issues d'une solution de l'équation d'écoulement à l'échelle de la mesure sur l'ensemble de l'aquifère. Pour cet approche en 2-D, leur suggestion pour déduire la perméabilité de bloc est de résoudre pour chaque bloc l'équation suivante :

$$\begin{pmatrix} \overline{u_{\omega,x}} \\ \overline{u_{\omega,y}} \end{pmatrix} = - \begin{pmatrix} K_{V,xx} & K_{V,xy} \\ K_{V,yx} & K_{V,yy} \end{pmatrix} \begin{pmatrix} \overline{\nabla h_{\omega,x}} \\ \overline{\nabla h_{\omega,y}} \end{pmatrix}$$

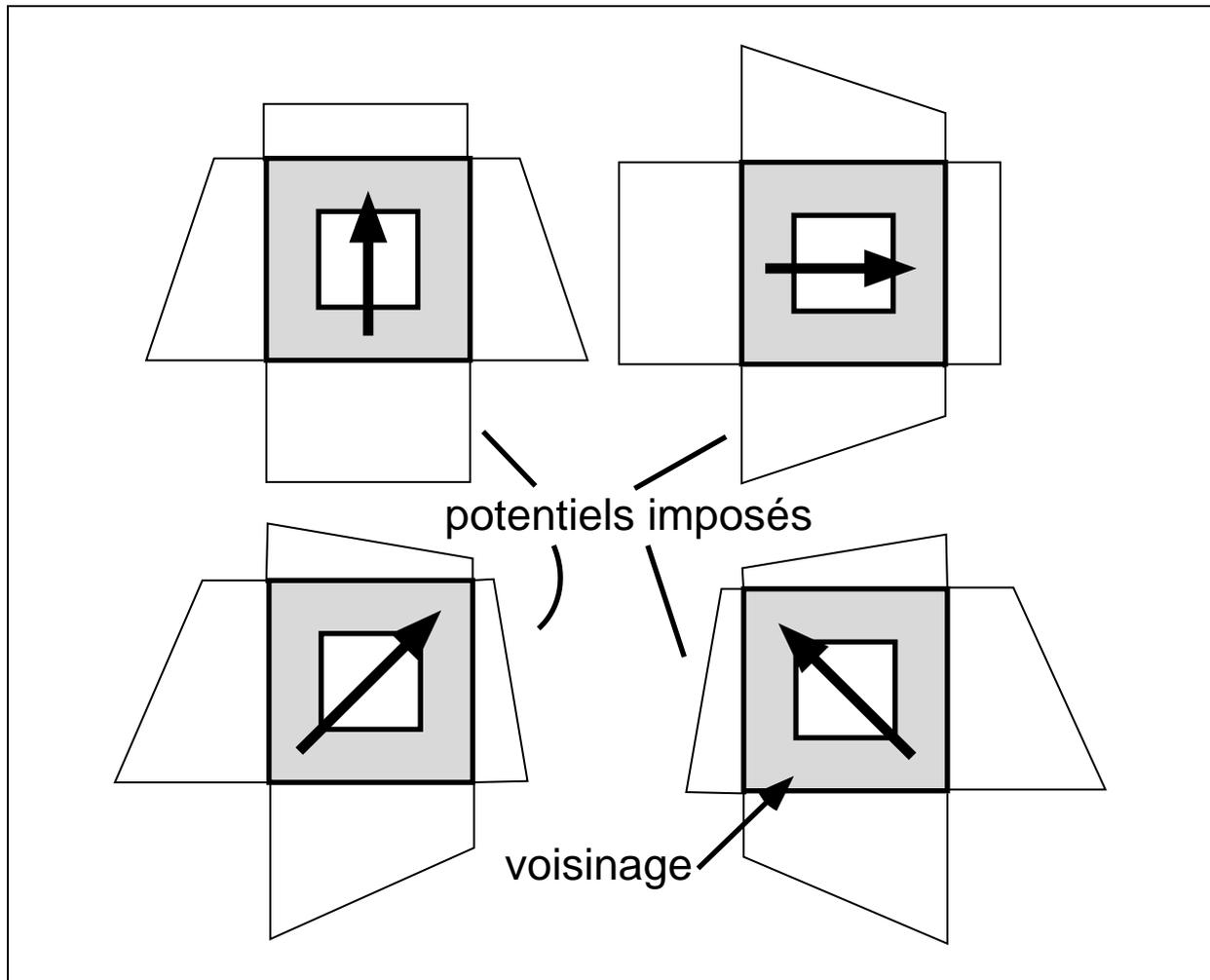
où la barre au-dessus des termes  $u_{\omega,x}$ ,  $u_{\omega,y}$ ,  $\nabla h_{\omega,x}$  et  $\nabla h_{\omega,y}$  des vecteurs  $\mathbf{u}_\omega$  et  $\nabla h_\omega$  signifie la valeur moyenne à l'intérieur du bloc (moyenne spatiale) et où  $K_{V,xx}$ ,  $K_{V,xy}$ ,  $K_{V,yx}$  et  $K_{V,yy}$  sont les termes du tenseur de perméabilité de bloc  $\mathbf{K}_V$ . Le problème avec cette approche est que le système d'équations linéaires à résoudre contient deux équations pour quatre inconnues (les termes du tenseur), ce qui rend la perméabilité de bloc indéterminée. White et Horne ont contourné le problème en suggérant de résoudre à l'échelle de la mesure l'équation d'écoulement  $n$  fois pour des conditions aux limites, appliquées à l'ensemble de l'aquifère, différentes. Le système (sur-déterminé) des  $2n$  équations ainsi obtenues est résolu par moindres carrés en minimisant l'écart sur les flux. Ce besoin d'utiliser plusieurs conditions aux limites libère le calcul des tenseurs de perméabilité de bloc du choix d'un écoulement régional spécifique; il s'en suit des perméabilités de bloc qui conviennent bien dans différents scénarios d'écoulement.

L'approche de White et Horne est capable de fournir des tenseurs de perméabilité de bloc complets et non-diagonaux et de prendre en compte des conditions aux limites d'écoulement à l'échelle de l'aquifère. L'inconvénient majeur se trouve dans l'effort de calcul nécessaire qui est dissuasif pour modéliser de grands aquifères (plusieurs millions ou dizaines de millions de cellules) en particulier en 3-D.

Pickup *et al.* (1992) considèrent qu'il serait plus convenable d'essayer de déterminer les tenseurs de perméabilité de bloc pour des conditions aux limites spécifiques qui sont représentatives de celles que l'on peut trouver à l'échelle de l'aquifère, plutôt que de calculer des tenseurs qui sont supposés être adaptés à une grande variété de conditions d'écoulement. Leur argument en faveur d'une telle approche mieux ciblée est qu'elle devrait conduire à des per-

méabilités de bloc plus précises pour un scénario d'écoulement donné. L'inconvénient est que les perméabilités équivalentes doivent être recalculées si les conditions aux limites de l'aquifère changent. Après résolution de l'équation d'écoulement à l'échelle de la mesure sur l'ensemble de l'aquifère, les conditions aux limites aux faces (ou arêtes) de chaque bloc sont déterminées. Ce sont elles qui devraient être utilisées pour résoudre l'équation d'écoulement dans le bloc et, par application de l'équation E 4-1, pour donner le tenseur de perméabilité de bloc. Comme cela a été mentionné plus haut, le problème est qu'il est nécessaire de calculer la solution d'écoulement pour au moins deux conditions aux limites différentes pour pouvoir lever l'indétermination sur les quatre termes du tenseur de perméabilité de bloc en 2-D. Pickup *et al.* ont ainsi proposé de résoudre pour chaque bloc deux problèmes d'écoulement avec des conditions aux limites obtenues par perturbation des conditions aux limites correctes qui s'appliquent sur le bloc (résultat de la simulation préalable de l'écoulement sur l'ensemble de l'aquifère). La perturbation doit être suffisamment grande pour entraîner des écoulements différents dans le bloc, mais suffisamment faible pour ne pas trop s'écarter des conditions correctes aux limites de bloc. L'inconvénient majeur de cette approche se trouve dans la grande sensibilité des résultats de perméabilité de bloc à l'amplitude de la perturbation et dans la difficulté à choisir une perturbation appropriée.

Conditions aux limites employées dans l'approche par résolution numérique avec voisinage. Pour chaque type de conditions aux limites, l'écoulement est résolu à l'échelle de la mesure à l'intérieure de l'aire contenant le bloc et son voisinage. La taille de ce dernier est définie arbitrairement comme une demi taille de bloc dans chaque direction.	Figure 4-3
---	------------



Une autre approche, fondée sur une solution de l'équation d'écoulement à l'échelle de la mesure pour des conditions aux limites globales, a été développée par Yamada (1995). Elle sera discutée plus loin dans la partie consacré aux méthodes dans lesquelles la définition de la géométrie des blocs fait partie du passage à grande échelle.

### **Résolution numérique avec voisinage (peau)**

Pour réduire le temps de calcul dans la méthode de White et Horne, Gómez-Hernández (1990, 1991) a présenté une autre approche par résolution numérique inspirée de celle de White et Horne (1987). Sa méthode produit également des tenseurs complets de perméabilité de bloc et tente d'imposer des conditions aux limites réalistes aux frontières des blocs pour en déduire la répartition spatiale des vitesses et des charges dans les blocs. Au lieu de résoudre l'équation d'écoulement à l'échelle de la mesure sur l'ensemble de l'aquifère, la résolution s'effectue localement sur une aire (un volume en 3-D) comprenant le bloc et un certain voisinage (peau) autour de celui-ci. La taille du voisinage a été fixée arbitrairement à une moitié de taille de bloc dans chaque direction. À l'identique de White et Horne (1987), pour pouvoir déterminer tous les termes du tenseur de perméabilité de bloc, l'écoulement doit être simulé pour au moins deux types de conditions aux limites appliquées aux frontières extérieures du voisinage. Dans l'approche de Gómez-Hernández, quatre types de conditions aux limites sont employés (Figure 4-3) pour forcer l'écoulement parallèlement aux côtés du bloc et diagonalement. L'auteur a

montré que cette méthode fonctionnait bien pour toute une variété de formations hétérogènes (isotrope, anisotrope et binaire sable-argile). Holden et Lia (1992) ont étendu la méthode au 3D.

### **Résolution numérique avec conditions aux limites périodiques**

Ni White et Horne (1987) ni Gómez-Hernández (1990, 1991) ne contraignent le tenseur de perméabilité de bloc à être symétrique et défini positif. Physiquement, il est impératif d'avoir des tenseurs de perméabilité définis positifs pour que l'eau ne se déplace pas dans le sens des charges hydrauliques croissantes. Pour ce qui concerne la symétrie des tenseurs de perméabilité de bloc, le sujet reste ouvert à discussion.

Durlofsky et Chung (1990) et Durlofsky (1991) ont présenté une approche qui produit toujours des tenseurs de perméabilité de bloc symétriques et définis positifs. La nouveauté de leur approche réside dans la répétition dans l'espace du bloc dont on cherche la perméabilité équivalente. Leur méthode se fonde sur l'hypothèse que les hétérogénéités spatiales des perméabilités (à l'échelle de la mesure) se produisent à deux échelles, une variabilité à grande échelle, qui définit les grandes tendances de variation spatiale de la perméabilité, et une variabilité à petite échelle qui est périodique dans l'espace. La périodicité de cette description de l'aquifère entraîne une correspondance biunivoque entre les charges et les flux (ou les vitesses) sur des faces opposées de bloc. L'équation d'écoulement est résolue à l'intérieur de chaque bloc pour deux directions perpendiculaires du gradient de charge d'ensemble (gradient moyen dans le bloc). L'équation E 4-1 est établie pour chaque solution d'écoulement, ce qui conduit à un système de quatre équations à quatre inconnues. Durlofski a montré que la solution de ce système d'équations produit toujours un tenseur symétrique défini positif<sup>1</sup>. Appliquée à de vrais milieux périodiques, pour lesquels on connaît l'expression analytique exacte du tenseur de perméabilité de bloc  $\mathbf{K}V^2$ , la méthode reproduit les résultats théoriques. Pour les milieux non périodiques, la méthode reste théoriquement valide si la taille de bloc est très supérieure aux dimensions des hétérogénéités à l'échelle de la mesure (portée de corrélation spatiale). Des comparaisons réalisées par Pickup *et al.* (1992, 1994) ont montré que l'approche de Durlofski est très précise et robuste même dans des situations où la périodicité des conditions aux limites ne s'applique pas strictement.

### **Écoulements non parallèles**

A travers les résultats de certains des travaux discutés précédemment, il était devenu évident que la perméabilité de bloc dépendait des conditions aux limites d'écoulement au bloc. Toutes les méthodes par résolution numérique décrites jusqu'à présent reposent sur des conditions aux limites qui imposent des écoulements parallèles à travers les blocs. Desbarats (1992a) a étudié le problème de la perméabilité équivalente pour des écoulements radiaux. Plus précisément, il a déterminé la perméabilité équivalente, sous forme scalaire, d'un bloc hétérogène soumis à des conditions d'écoulement radial. L'écoulement radial est provoqué par une source à débit constant située au centre du bloc (puits producteur ou injecteur) et des potentiels hydrauliques constants imposés aux frontières du bloc. À l'aide d'une approche numérique empirique, il est

<sup>1</sup> C'est-à-dire un tenseur dont les termes vérifient les deux relations suivantes :

$$K_{V,xy} = K_{V,yx} \text{ et } K_{V,xx} K_{V,yy} - K_{V,xy} K_{V,yx} > 0.$$

<sup>2</sup> Ceci suppose cependant que les tailles de bloc sont des multiples de la période dans chaque direction de bloc.

arrivé à la conclusion que, en 2-D, une moyenne géométrique des perméabilités de cellule, avec des poids proportionnels à l'inverse du carré de la distance à la source (puits), donne de bons résultats pour des variances faibles à moyennes du log de la perméabilité. Desbarats (1993) a étudié aussi le problème de la perméabilité équivalente pour des blocs contenant deux puits (configuration en dipôle). Pour ce cas, il a établi que la perméabilité entre puits est la moyenne harmonique des perméabilités moyennées sur des régions circulaires centrées sur chaque puits.

### Méthodes analytiques

Rubin et Gómez-Hernández (1990) ont établi et validé numériquement une expression analytique de la perméabilité de bloc en 2-D pour un bloc plongé dans un aquifère infini hétérogène avec des conditions de potentielles imposées à l'infini. Les perméabilités de cellule sont supposées être scalaires et présenter une corrélation spatiale isotrope. Comme conséquence de l'hypothèse précédente, les perméabilités de bloc sont supposés être scalaires également. Dans ces conditions, la perméabilité de bloc peut s'écrire :

$$E 4-2 \quad K_V = K_g \left[ 1 + \overline{Y'} + \frac{\overline{Y'^2}}{2} + \frac{1}{j_a} (\overline{Y'j'_x} - \overline{Y'} \overline{j'_x}) \right]$$

où  $K_g$  est la moyenne géométrique de la perméabilité calculée sur l'ensemble de l'aquifère, la barre supérieure indique la moyenne spatiale des valeurs dans le bloc de la variable correspondante,  $V$  est le volume du bloc,  $Y'$  est la fluctuation autour de la valeur moyenne de  $Y = \ln(K\omega)$ ,  $j_a$  est l'amplitude du gradient moyen de charge (qui est supposé être parallèle à l'axe  $x$ ) et  $j'_x$  est la fluctuation de la composante moyenne en  $x$  du gradient  $\nabla h_\omega$ . Cette expression, qui est valide uniquement pour des faibles variances de  $Y$ , montre clairement la relation qui peut exister entre la perméabilité de bloc et les conditions d'écoulement dans le bloc à travers les termes  $j'_x$ .

Sáez *et al.* (1989) ont tiré également une expression analytique de la perméabilité de bloc en utilisant la méthode des échelles multiples. Ils considèrent deux échelles d'observation. À l'échelle la plus grande, il est supposé que la petite échelle d'hétérogénéité est indétectable et que les perméabilités varient doucement dans l'espace. À la plus petite échelle, l'hétérogénéité du milieu est apparente et les perméabilités varient de manière abrupte et erratique dans l'espace. En tout point, la perméabilité s'exprime comme la somme d'une valeur à grande échelle et d'une perturbation due aux hétérogénéités de petite échelle. À l'identique, la charge hydraulique s'écrit comme la somme d'une valeur de grande échelle et d'une perturbation. La perméabilité d'un bloc dont la taille est très supérieure à la petite échelle d'hétérogénéité est définie par un tenseur  $\mathbf{K}_V$  pour lequel le terme de grande échelle de la charge hydraulique, noté  $h^{(0)}$ , satisfait à l'équation de conservation suivante :

$$\nabla(\mathbf{K}_V \cdot h^{(0)}) = 0.$$

Sáez *et al.* (1989) ont résolu l'équation ci-dessus dans le cas d'un aquifère périodique avec une période unitaire égale à la taille des blocs pour lesquels la perméabilité équivalente est recherchée. Il ont obtenu pour la perméabilité de bloc l'expression suivante :

$$E 4-3 \quad \mathbf{K}_V = \mathbf{K}_A + t .$$

Cette expression indique que la perméabilité de bloc est la somme de la moyenne arithmétique  $\mathbf{K}_A$  et d'un tenseur de tortuosité défini par la moyenne spatiale suivante :

$$t = \frac{1}{V} \int_V \mathbf{K}_\omega \cdot \nabla g dV$$

où la fonction vectorielle  $\mathbf{g}$  satisfait au problème avec conditions aux limites :

$$E 4-4 \quad \nabla(\mathbf{K}_\omega \cdot \nabla g) = -(\nabla \mathbf{K}_\omega), \text{ avec } \mathbf{g} \text{ périodique dans le bloc unitaire.}$$

L'inconvénient de la solution analytique de la perméabilité de bloc donnée par Sáez *et al.* est qu'elle est limitée à des blocs beaucoup plus grands que l'échelle de variabilité erratique des perméabilités de cellule. En outre, elle impose au milieu poreux perméable d'être périodique avec une période est égale à la taille des blocs.

### Méthodes des moments

Kitanidis (1990) a utilisé une définition de la perméabilité de bloc différente de celle donnée par l'équation E 4-1. Sa définition repose sur la méthode des moments introduite initialement par Aris (1956) et généralisée ensuite par Brenner (1980). Dans la méthode des moments, au lieu d'essayer de trouver une perméabilité de bloc qui relie la vitesse (de Darcy) moyenne au gradient moyen de charge hydraulique, l'objectif est de trouver une valeur de bloc qui cale les moments spatiaux de la charge hydraulique dans le milieu hétérogène comme cela est décrit dans la suite. Cette définition nécessite la solution d'un problème transitoire contrairement à la définition de l'équation E 4-1 qui repose sur la solution d'un problème d'écoulement en régime permanent. L'application de la méthode des moments se comprend mieux sur un exemple. On se place dans le cas 2-D d'un aquifère infini de perméabilité constante et dans lequel une impulsion ponctuelle de charge est provoquée au temps 0 au point d'origine. La solution de la dissipation transitoire de l'impulsion de charge est donnée par une loi gaussienne (en fonction de la distance à l'origine pour un temps donné) :

$$h(\mathbf{x}, t) = (2\pi)^{-3/2} |2\mathbf{D}t|^{-1/2} \exp(-\mathbf{x}\mathbf{D}^{-1}\mathbf{x} / 4t)$$

où  $t$  est le temps,  $\mathbf{x}$  est le vecteur position,  $\mathbf{D}=\mathbf{K}/S$  est appelé le tenseur de diffusivité et  $S$  le coefficient d'emmagasinement spécifique. Le moment d'ordre 2 de la fonction  $h(\mathbf{x}, t)$  par rapport à l'origine est le moment d'inertie qui s'écrit :

$$\Delta(t) = \int \mathbf{x} \cdot \mathbf{x} h(\mathbf{x}, t) d\mathbf{x} = 2\mathbf{D}t .$$

Il est à noter que la moitié de la dérivée par rapport au temps du moment est donnée par  $(1/2)d\Delta(t)/dt=\mathbf{D}$ . Dans la méthode des moments, l'évolution de l'impulsion de charge à

l'origine est résolue dans le block hétérogène et la vitesse de variation (dérivée) du moment d'ordre 2 est assimilée au tenseur de diffusivité d'un bloc homogène de perméabilité équivalente.

Pour résoudre le problème transitoire d'impulsion de charge, certaines hypothèses sont nécessaires. D'abord, l'aquifère est supposé être périodique avec une période beaucoup plus grande que l'échelle de variabilité de  $\mathbf{K}_\omega$ . Ensuite, si les moments existent à deux échelles (comme pour Sáez *et al.*, 1989), seuls les moments à l'échelle la plus grande sont considérés. Enfin, les écoulements sont supposés varier lentement. Sous ces hypothèses, l'expression obtenue pour  $\mathbf{K}_V$  est très semblable à celle de Sáez *et al.* (1989). Le tenseur de perméabilité de bloc s'écrit comme la somme de sa moyenne arithmétique dans l'espace et d'un terme d'intégrale qui équivaut au terme de tortuosité de l'équation E 4-3. Ce terme d'intégrale s'exprime comme une fonction  $g$  qui doit satisfaire à un problème équivalent à celui donné par l'équation E 4-4.

Dykaar et Kitanidis (1993) ont proposé une méthode, appelée méthode numérique spectrale, de résolution numérique efficace du problème E 4-4.

Cette méthode présente les mêmes inconvénients que celle de Sáez *et al.* (1989) : elle est limitée à de grands blocs et requiert l'hypothèse d'un milieu périodique dont la période est égale à la taille de bloc.

### **Energie de dissipation**

Indelman et Dagan (1993a, b) et Bøe (1994) ont employé un concept de dissipation d'énergie pour déterminer des tenseurs de perméabilité de bloc. L'énergie de dissipation est définie comme l'énergie par unité de temps nécessaire pour forcer le fluide à traverser le bloc en tant que milieu poreux. Une condition nécessaire que doit vérifier la perméabilité de bloc est que l'énergie de dissipation dépensée dans le bloc homogène équivalent doit être la même que celle dépensée dans le bloc hétérogène. Bøe (1994) a montré que, avec cette définition et des conditions aux limites périodiques appliquées sur les faces du bloc, les perméabilités de bloc calculées par la méthode coïncident avec celles de Durlofsky (1991). Indelman et Dagan (1993a, b) se sont davantage intéressés à obtenir les propriétés statistiques des blocs dont on cherche la perméabilité équivalente plutôt que de regarder une à une les relations entre chaque blocs et les perméabilités de cellule à l'intérieur. Leur approche est commentée plus loin dans la partie consacrée au calcul direct des perméabilités de bloc.

#### **4.1.3.3 Géométrie des blocs**

Toutes les méthodes mentionnées précédemment calculent la perméabilité équivalente de blocs rectangulaires mais aucune n'a été étendue au cas de blocs non rectangulaires. Beaucoup d'efforts ont été consacrés au calcul de perméabilités équivalentes sans s'intéresser à la géométrie des blocs. Il est pourtant bon de ne pas oublier que le calcul de perméabilités équivalentes n'est pas une fin en soi mais a pour objet de déterminer des valeurs de bloc pour la simulation de phénomènes d'écoulement en milieu poreux (mono ou polyphasique) et parfois de transport. Les méthodes par résolution numérique les plus élaborées fournissent déjà, pour une géométrie de bloc donnée, le meilleur estimateur de la perméabilité de bloc qu'il est possible

d'obtenir. Pour améliorer ces estimateurs, la seule solution est d'optimiser la définition géométrique des blocs.

Aussi bonne soit la méthode employée, le calcul de perméabilités équivalentes conduit toujours à moyenniser des valeurs de perméabilité pour résoudre un problème d'écoulement, ce qui conduit souvent à filtrer certains détails de l'écoulement à l'intérieur de blocs hétérogènes. Il peut alors être tentant de concevoir une méthode pour définir la géométrie des blocs de manière à minimiser ou au moins réduire l'effet de moyenne, inhérent au calcul de perméabilités équivalentes, aux endroits où les particularités de l'écoulement, dues aux hétérogénéités, ont le plus de conséquence sur les objectifs finaux de la modélisation. Par exemple, si l'objectif final est de simuler le transport de masse et que le temps d'arrivée du corps transporté est le paramètre le plus critique de l'étude, une définition convenable de la géométrie des blocs devrait délimiter les chemins ou chenaux de plus forte perméabilité de manière à réduire l'impact qu'aurait sur la prédiction du temps d'arrivée une mauvaise résolution de la grille dans ces zones d'écoulement préférentiel (voir par ex. Durlofsky *et al.*, 1994a,b).

Si la construction de grilles pour la simulation d'écoulement est un sujet d'étude de longue date, l'intérêt pour des méthodes de génération de grilles ou de maillages en association avec le calcul de perméabilités équivalentes est relativement récent. Quatre de ces techniques sont décrites dans la suite.

### **Grilles élastiques**

Garcia *et al.* (1990, 1992) et Garcia (1991) ont introduit le concept de grille "élastique". Leur méthode part d'une grille initiale (généralement une grille rectangulaire uniforme) et déplace les sommets des blocs avec pour objectif de minimiser l'hétérogénéité des blocs par rapport à un ou plusieurs attributs dont les valeurs de cellule sont connues (par ex. perméabilité, porosité, type de faciès). A chaque arête connectant deux sommets de bloc (arête de bloc) est attaché un coefficient d'élasticité qui est défini comme une fonction des variances de bloc des attributs considérés pour les blocs adjacents à l'arête (c'est-à-dire ceux possédant l'arête)<sup>3</sup>. L'énergie potentielle de chaque arête est proportionnelle au carré de sa longueur et à son coefficient d'élasticité. L'énergie potentielle de l'ensemble de la grille est la somme des énergies potentielles d'arête. Pour des valeurs de coefficient d'élasticité données, l'état d'équilibre de la grille est celui qui minimise l'énergie potentielle de la grille. Pour des conditions aux limites convenables, qui consistent à attacher certains des sommets de bloc à des frontières du domaine (frontières extérieures et intérieures), la résolution du problème de minimisation de l'énergie potentielle de la grille est immédiate et fournit la position d'équilibre des sommets de bloc. Tout l'intérêt de la méthode est alors de pouvoir, par un procédé itératif, mettre à jour progressivement les coefficients d'élasticité des arêtes en fonction de l'hétérogénéité des blocs avoisinants dont la géométrie est corrigée également à chaque itération. Le procédé itératif s'arrête lorsque la grille converge vers une solution où il n'est plus possible d'améliorer globalement la qualité des blocs par rapport au critère choisi (choix d'attributs et des paramètres de la fonction de calcul des coefficients d'élasticité). Cette approche conduit à générer des grilles

---

<sup>3</sup> Les variances sont calculées pour chaque bloc à partir des valeurs situées à l'intérieur du bloc. Pour un attribut catégorique tel que le faciès ou le type de roche, l'hétérogénéité du bloc n'est pas mesurée statistiquement par une variance, qui n'aurait aucune signification, mais par la proportion de la population (valeur) la plus fréquente dans le bloc.

irrégulières dont les blocs sont faiblement à moyennement hétérogènes<sup>4</sup> (Figure 4-4); des méthodes simples de calcul de perméabilités équivalentes donnent alors de bon résultats sur ces blocs améliorés. La méthode a été généralisée au 3-D par Garcia (1994).

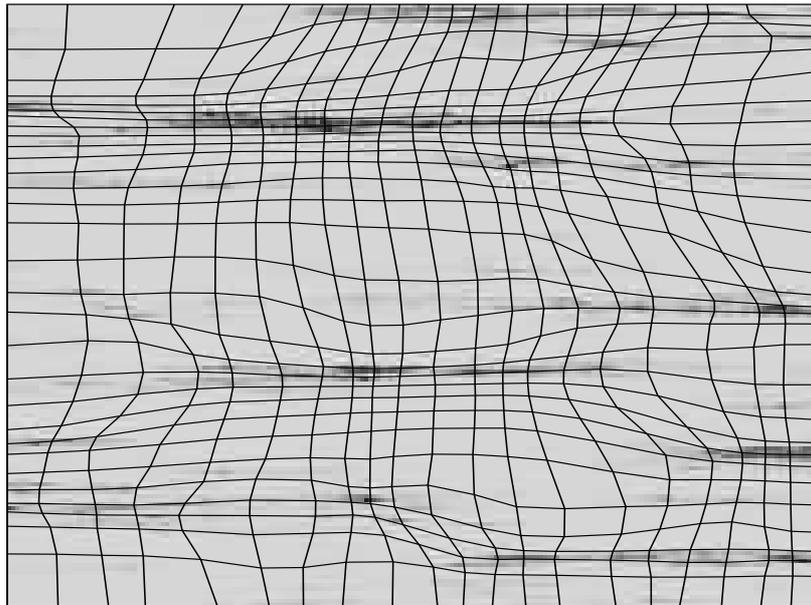
Farmer *et al.* (1990, 1991) ont développé une méthode similaire dans ses objectifs à celle de Garcia mais dans laquelle l'optimisation de la grille est résolue en partie par une technique de recuit simulé. Plus récemment, Robey (1995) a introduit une approche par grille élastique dans laquelle l'objectif à atteindre pour la grille finale est un minimum local d'hétérogénéité de bloc (contrairement au minimum global employé dans l'approche de Garcia).

---

<sup>4</sup> La quantification de l'hétérogénéité d'un bloc est très fortement liée à la mesure employée (variance, intervalle inter-quartiles, covariances pour des espacements donnés) et reste une notion très subjective qui dépend non seulement de la variabilité des valeurs dans le bloc mais aussi de la taille du bloc (c'est-à-dire du nombre de valeurs dans le bloc). Dans le cas qui nous intéresse ici du calcul de perméabilités équivalentes, l'amélioration de l'hétérogénéité d'un bloc devrait se juger sur des critères qui déterminent si les valeurs de perméabilité dans le bloc rentrent toutes dans une même catégorie (par ex. faiblement, moyennement ou fortement perméable) et si la perméabilité équivalente tend à devenir une propriété intrinsèque indépendante des conditions d'écoulement.

Grille élastique typique obtenue avec la méthode de Garcia (1990) avec pour objectif de minimiser les hétérogénéités de perméabilité de cellule à l'intérieur de chaque bloc. La grille à l'échelle de la mesure contient 200 par 150 cellules. La grille à l'échelle numérique est constituée de 20 par 30 blocs. Les cellules sombres indiquent de grandes valeurs de perméabilité.

Figure 4-4



### **Grilles élastiques appliquées à des variables d'écoulement**

Les utilisations mentionnées précédemment des grilles élastiques ont porté sur des attributs pétrophysiques sans s'intéresser au problème d'écoulement à l'exception de la prise en compte de la position des puits (sources). Tran (1995) et Tran et Journel (1995) ont réalisé que la minimisation de la variance des perméabilités de cellule dans les blocs ne suffit pas à garantir que les voies d'écoulement rapide (chemins fortement perméables) sont correctement reconnues lors de la construction de la grille et préservées au niveau du modèle de grande échelle. Ceci justifie de s'intéresser à des variables d'écoulement pour définir la géométrie des blocs. Tran et Journel ont proposé une extension de la méthode par résolution numérique de White et Horne (1987). Après résolution de l'équation d'écoulement en régime permanent à l'échelle de la mesure (grille fine) sur l'ensemble de l'aquifère et pour des conditions aux limites et des sources (puits) convenablement choisies, la géométrie des blocs est déterminée par la méthode de grille élastique de Garcia (1990, 1992, 1994) appliquée cette fois à des valeurs de ligne de courant. Le coefficient d'élasticité de chaque segment reliant deux sommets de bloc devient alors une fonction de la variance des valeurs de lignes de courant de chaque bloc adjacent. De cette manière, il peut être assuré que la densité de lignes de courant est relativement identique à l'intérieur de chaque bloc. En conséquence, de petits blocs sont placés aux endroits où les vitesses d'écoulement sont grandes et des grands blocs aux endroits où les vitesses sont faibles. Du point de vue de la simulation de phénomènes d'écoulement et de transport à l'échelle de l'aquifère, l'approche de Tran et Journel donne de meilleurs résultats que celle prévue initialement par Garcia.

L'adaptation de grilles élastiques à des variables d'écoulement a été reprise par Garcia *et al.* (1997) dans le cadre d'un projet européen de démonstration des grilles élastiques sur des jeux de données réels fournis par quatre compagnies pétrolières (voir Garcia *et al.*, 1997). Sur chacun des jeux de données, l'exercice a consisté à simuler un ou plusieurs historiques de production avec des modèles d'écoulement polyphasique de type black-oil (3 phases, 3 composants) ou compositionnels (3 phases, nombre quelconque de composants). Chaque historique a été simulé sur une grille uniforme régulière et sur des grilles élastiques adaptées soit à des hétérogénéités (perméabilité, porosité ou type de roche) soit à des variables d'écoulement. Dans certains cas, il a été possible également de simuler l'historique à l'échelle de la mesure (grille fine) de manière à obtenir une solution de référence utile pour comparer les résultats. De la même manière que Tran et Journel, l'approche consiste à simuler préalablement un problème d'écoulement monophasique en régime permanent à l'échelle de la mesure sur l'ensemble du réservoir pour des conditions aux limites et des sources conformes à l'historique de production à recalcr ensuite. Contrairement à Tran et Journel, qui utilisent comme variable d'écoulement des lignes de courant qui ne peuvent pas être calculées en 3-D, la variable d'écoulement utilisée est le flux moyen calculé à l'échelle de la mesure pour chaque cellule. Le critère d'hétérogénéité<sup>5</sup> affecté aux blocs se fonde soit sur la variance des flux soit sur leur moyenne à l'intérieur des blocs<sup>6</sup>. Les résultats ont montré qu'il n'est pas suffisant d'adapter des grilles uniquement sur des hétérogénéités sans tenir compte des conditions d'écoulement (en particulier de la position des puits) car les hétérogénéités peuvent être situées à des endroits où les écoulements sont faibles ou sans conséquence sur le comportement des puits. Pire encore, en affinant inutilement la grille dans certaines régions au détriment d'autres dont le voisinage des puits<sup>7</sup>, l'utilisation de grilles adaptées uniquement sur des hétérogénéités peut conduire à détériorer considérablement les résultats des simulations d'écoulement polyphasique à l'échelle du gisement (grilles grossières) comparés à ceux obtenus avec une grille régulière de même dimension. L'adaptation des grilles sur les flux convient davantage mais, en favorisant la distorsion des blocs, elles montrent les limites de beaucoup de simulateurs par différences finies qui sont prévues théoriquement pour travailler sur des grilles rectangulaires et sont souvent incapables de prendre en compte convenablement des grilles déformées.

---

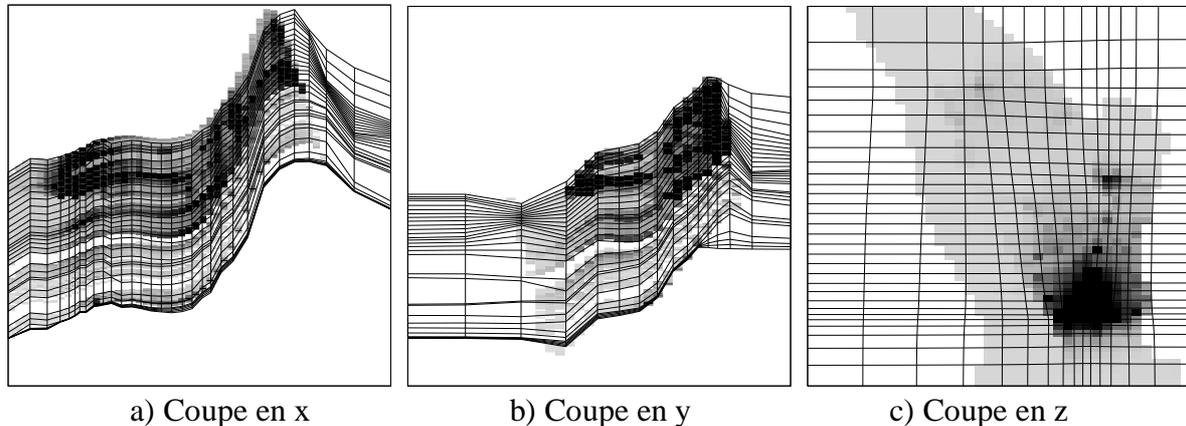
<sup>5</sup> Il est alors plus convenable de parler de critère de qualité d'un bloc que de critère d'hétérogénéité.

<sup>6</sup> L'utilisation de la variance des flux a tendance à produire des blocs à l'intérieur desquels les flux sont relativement homogènes. La taille des blocs ne dépend pas alors de l'amplitude de ces flux dans le bloc. L'utilisation de la moyenne permet de conditionner directement la taille des blocs à l'amplitude des flux. En cherchant à minimiser le flux moyen à l'intérieur des blocs (cette moyenne peut être pondérée par la taille du bloc), des blocs plus petits sont placés aux endroits où les flux sont plus grands.

<sup>7</sup> Le nombre total de blocs étant fixé, si certaines régions contiennent davantage de blocs, d'autres en auront nécessairement moins.

Grille élastique adaptée à une variable de flux pour un écoulement stationnaire provoqué par un puit injecteur et un producteur.

Figure 4-5



a) Coupe en x

b) Coupe en y

c) Coupe en z

### **Approche simplifiée de grilles élastiques appliquées à des variables d'écoulement**

Les approches précédentes, bien que supérieures à celle suivie à l'origine par Garcia (1990), présentent le même inconvénient que celui mentionné pour la méthode de White et Horne (1987) : l'équation d'écoulement en régime permanent doit être résolue à l'échelle de la mesure sur l'ensemble de l'aquifère pour calculer les valeurs de variables d'écoulement dans les cellules à l'intérieur des blocs. Pour de grands aquifères, les temps de calcul peuvent devenir prohibitifs. Pour contourner cette difficulté, Wen (1996) a implémenté une variante de la méthode de Tran qui ne requiert pas de solution d'écoulement à l'échelle de la mesure. Son approche est itérative et se fonde sur une mise à jour simultanée, à chaque itération, des perméabilités de bloc et de la géométrie des blocs. Sa méthode commence par adapter une grille élastique à l'aide de la méthode Garcia (1990) dans son implémentation initiale, c'est-à-dire par minimisation de l'hétérogénéité des perméabilités de cellule à l'intérieur des blocs. Dans un deuxième temps, une version de la méthode par résolution numérique de Durlofsky (1991) étendue à des blocs irréguliers est utilisée pour déterminer des tenseurs de perméabilité de bloc. L'équation d'écoulement est alors résolue à l'échelle numérique (et non pas à l'échelle de la mesure) à partir des géométries et des perméabilités de bloc calculées à l'étape précédente. De la solution d'écoulement sont déduites les vitesses d'écoulement en chaque cellule de l'aquifère (échelle de la mesure) sur lesquelles est appliquée à nouveau la méthode d'adaptation de grilles élastiques de Garcia pour redéfinir (corriger) la géométrie des blocs. Pour cette adaptation de la grille, l'énergie potentielle des arêtes est rendue proportionnelle à la variance des vitesses d'écoulement des blocs adjacents à la place de la variance des perméabilités de cellules. A partir de la nouvelle géométrie des blocs, le processus itératif est répété avec le calcul de perméabilités de bloc et la résolution à nouveau de l'écoulement à l'échelle numérique. Ce processus itératif s'arrête lorsque ni la géométrie des blocs ni leurs perméabilités équivalentes ne sont modifiés significativement d'une itération à la suivante. Cette méthode donne des résultats aussi bons que ceux de Tran (1995) pour une fraction du coût de calcul. Son inconvénient majeur réside dans le fait qu'il n'existe pas de démonstration théorique de la convergence de la méthode bien que, sur tous les cas où elle a été testée, la convergence a toujours été obtenue.

### **Discrétisation dans l'espace des fonctions de courant et de potentiel**

Yamada (1995) a abordé le problème de la perméabilité équivalente sous un angle différent. Cette approche reconnaît que les conditions aux limites employées dans la résolution numérique simple (Figure 4-2) peuvent s'écarter des conditions aux limites qui existent en réalité sur le périmètre des blocs. Toutefois, au lieu de chercher à affiner les conditions aux limites pour les rendre plus proches de celles qu'aurait le bloc s'il était plongé dans l'aquifère, Yamada cherche à déterminer des géométries de bloc pour lesquelles les conditions aux limites de la Figure 4-2 s'appliquent. Ces géométries peuvent être trouvées facilement si la construction de la grille s'effectue dans l'espace des fonctions de courant et de potentiel.

Après résolution de l'équation d'écoulement à l'échelle de la mesure, les résultats sont tracés dans l'espace des fonctions de courant et de potentiel. Dans cet espace, chaque bloc est défini avec une forme rectangulaire, ce qui apporte la garantie que sur deux des faces opposées du bloc la fonction de courant est constante (arêtes alignées sur des lignes de courant) et que sur les deux autres faces opposées le potentiel est constant (arêtes alignées sur des lignes de potentiel). Avec cette définition, l'utilisation de conditions de flux nul sur deux des faces du bloc et de potentiel imposé sur les autres se justifie. En outre, comme l'écoulement ne peut pas traverser des lignes de courant, il est inutile de calculer le terme du tenseur de perméabilité de bloc dans la direction perpendiculaire aux lignes de courant (frontière à flux nul). Yamada a trouvé que les simulations d'écoulement réalisées sur des grilles ainsi construites reproduisent avec succès le champ de potentiel pour un écoulement monophasique et donnent des approximations raisonnables des champs de potentiel et de saturation pour des problèmes diphasiques.

Bien qu'attractive, cette méthode présente les trois inconvénients majeurs suivants.

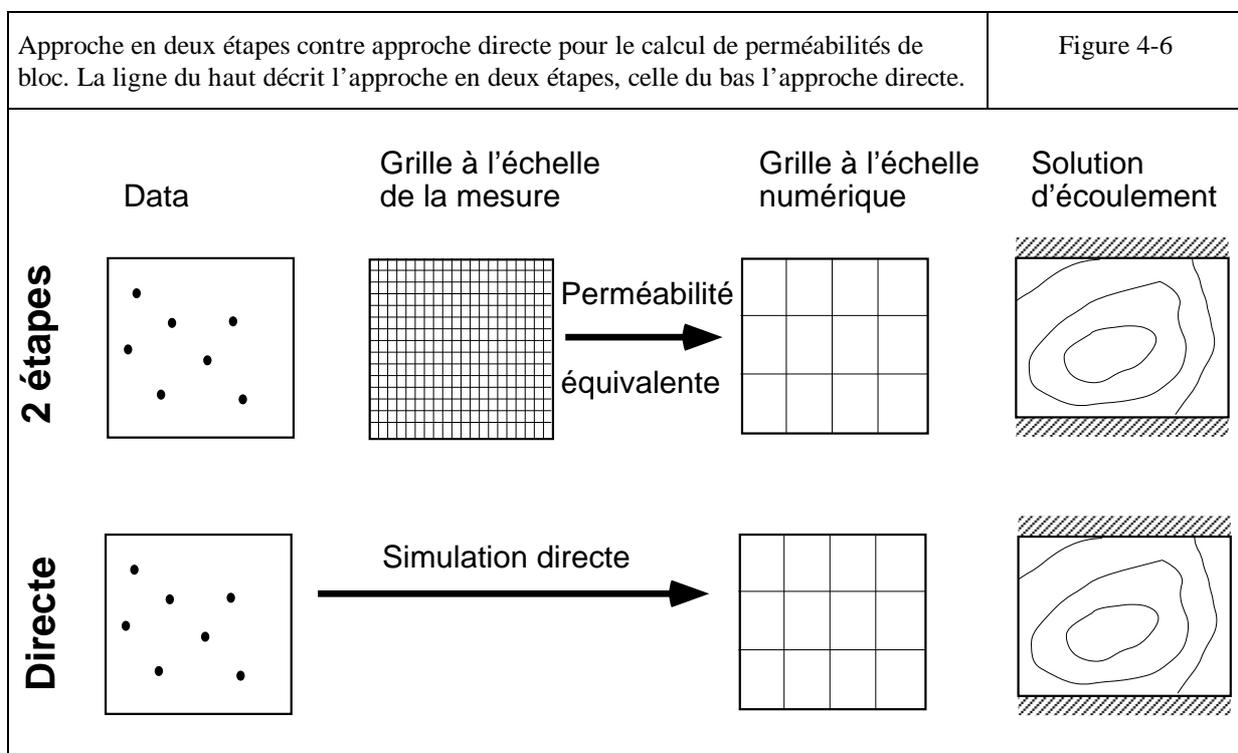
1. Conçue en 2-D, son extension au 3-D est loin d'être triviale compte tenu de la très grande difficulté qui existe à calculer des lignes de courant en 3-D.
2. La construction des blocs dans l'espace des fonctions de courant et de potentiel suppose qu'en chaque point de l'espace ne passe qu'une seule ligne de courant. Cette condition n'est pas vérifiée à l'endroit des points sources (position des puits).
3. Elle nécessite une solution d'écoulement à l'échelle de la mesure sur l'ensemble de l'aquifère.

#### ***4.1.3.4 Calcul direct de perméabilités de bloc***

En considérant à nouveau que l'objectif du calcul de perméabilités équivalentes est de fournir des valeurs aux blocs pour la simulation d'écoulement à une échelle numérique, certains auteurs sont arrivés à la conclusion que la génération de grilles à l'échelle de la mesure (Figure 4-1a), dans le seul but d'effectuer un changement d'échelle pour déduire un modèle de perméabilités de bloc, était un gaspillage d'efforts. La question que se sont posés ces auteurs est la suivante : est-ce que les perméabilités de bloc peuvent être générées directement à partir d'un nombre réduit de données à l'échelle de la mesure, sans avoir besoin de passer par une étape

intermédiaire coûteuse de calcul de valeurs à l'échelle de la mesure, mais en obtenant des caractéristiques de perméabilité de blocs identiques à celles observées avec les autres méthodes ?

La Figure 4-6 illustre ce concept. Sur la ligne du haut, l'approche dite en deux étapes (Gómez-Hernández, 1991) est décrite : à partir d'un ensemble de données (mesures), une série de réalisations de la perméabilité sont générées à l'échelle de la mesure; un passage à une échelle numérique est opéré sur chacune de ces réalisations et le problème d'écoulement est résolu sur chacun des modèles de grand échelle ainsi obtenus pour produire une série de réalisation de la répartition spatiale possible des charges hydrauliques dans l'aquifère. Sur la ligne du bas, l'approche par simulation directe est représentée : les points de départ et d'arrivée sont les mêmes que précédemment, la seule différence est que les valeurs de bloc sont calculées directement à partir des mesures.



### Perméabilités de bloc scalaires

Rubin et Gómez-Hernández (1990) ont été les premiers à aborder ce problème analytiquement. Avec comme point de départ le résultat analytique de leur méthode de calcul de perméabilités équivalentes (équation E 4-2), ils ont calculés ce qui devait être la valeur moyenne et la covariance des perméabilités de bloc, ainsi que la covariance croisée entre les valeurs de perméabilité de cellule et de bloc. Avec ces résultats statistiques et l'hypothèse que les perméabilités de bloc conservent une loi multi-log-normale comme cela est supposé pour les perméabilités de cellules, il ont procédé, à l'aide d'une approche géostatistique standard, au calcul de perméabilités de bloc conditionnelles par rapport aux données à l'échelle de la mesure. Leur méthode est limitée par les hypothèses utilisées pour déduire l'équation E 4-2, c'est-à-dire une variabilité spatiale isotrope des perméabilités de cellule, une faible variabilité du log des valeurs de perméabilité et une perméabilité de bloc scalaire.

### **Images de calibrage**

Pour contourner les limitations de l'approche précédente, Gómez-Hernández (1990, 1991) a proposé de déduire les covariances croisées entre perméabilités de bloc et de cellule à l'aide d'images de calibrage. L'idée consiste à générer au départ une grille de calibrage à l'échelle de la mesure et d'utiliser cette grille pour appliquer une quelconque des méthodes de calcul de perméabilités équivalentes décrites dans les paragraphes précédents de manière à générer des tenseurs de perméabilité de bloc. A ce stade, deux réalisations sont modèles de perméabilité sont disponibles, un à l'échelle de la mesure avec des valeurs de perméabilité de cellule et une à l'échelle numérique avec des tenseurs de perméabilité de bloc. Ces deux modèles permettent de déduire expérimentalement les covariances et covariances croisées entre la perméabilité (scalaire) de cellule et les termes du tenseur de perméabilité de bloc. Après avoir inféré ainsi l'ensemble des covariances nécessaires, il devient possible de générer directement les termes des tenseurs de perméabilité de bloc sans avoir besoin de passer à nouveau par un changement d'échelle.

L'ennui avec cette méthode est le trop grand nombre d'auto-covariances et de covariances croisées à calculer. En 2-D, le nombre de variables impliquées, en supposant que le tenseur de perméabilité est symétrique, est de 4 : une pour la perméabilité de cellule et trois pour le tenseur de perméabilité de bloc. Il s'en suit un nombre de covariances à modéliser de 9. En 3-D, le nombre de variables monte à 7 (1 + 6) et celui de covariance à 27 !

Pour réduire le nombre de covariances, Tran (1995) a proposé de calculer les auto-covariances et les covariances croisées uniquement pour les termes du tenseur de perméabilité de bloc. De plus, en supposant que les directions principales du tenseur de perméabilité sont parallèles aux côtés du bloc, il a réduit le nombre de variables à 2 en 2-D et le nombre de covariances à 3. The conditionnement par rapport aux données à l'échelle de la mesure (pour lesquelles la covariance croisée entre cellules et blocs serait nécessaire) est résolu par une approche heuristique.

### **Régularisation**

Desbarats (1989, 1992b) a réussi à déduire la valeur moyenne et la variance de perméabilités de bloc lorsque ces dernières sont obtenues par moyenne de puissance. Il a utilisé ensemble une approximation linéaire du log de la perméabilité générée par moyenne de puissance avec des techniques de régularisation standard (voir par ex. Journel et Huijbregts, 1978) pour calculer la moyenne et la variance de valeurs de bloc. Desbarats n'a pas cherché alors à poursuivre pour calculer la covariance entre valeurs de cellule et de bloc ou entre blocs. Il aurait été pourtant simple d'étendre ses résultats pour déterminer ces covariances qui, une fois connues, permettraient de générer directement les perméabilités de bloc, conditionnelles par rapport aux données de perméabilité de cellule, d'une manière simple par co-simulation stochastique de variables multiples. Les limitations de cette approche sont dues à celles propres à l'utilisation de la méthode des moyennes de puissance pour calculer des perméabilités de bloc et à la linéarisation du log de la perméabilité.

### **Dissipation d'énergie**

A partir d'une définition de l'énergie de dissipation dans le bloc, Indelman et Dagan (1993 a, b) ont déduit la moyenne et la covariance de perméabilités de bloc. Ils ont établi qu'il n'était pas

suffisant d'imposer la préservation de l'énergie de dissipation dans le calcul des perméabilités de bloc pour déduire une relation entre la perméabilité de bloc et les perméabilités de cellule à l'intérieur du bloc. Ils ont montré cependant qu'il était nécessaire d'avoir l'égalité, au sens statistique, des énergies de dissipation à l'échelle de la cellule et du bloc. Sur ces prémices, ils ont procédé au calcul de la moyenne et de la covariance des termes des tenseurs 3-D symétriques de perméabilité équivalente. Dans le cas le plus général, ils sont arrivés à l'expression de 6 moyennes géométriques et de 15 covariances croisées du log de la perméabilité. Ces expressions sont complexes et les auteurs précisent que l'existence et l'unicité de la solution dans le cas le plus général n'est pas prouvée. Il est à noter cependant que, sur les cas spécifiques qu'ils ont étudiés, la variabilité des perméabilités de bloc présente de bonnes caractéristiques. Parmi leurs conclusions, ils ont établi que, excepté pour des blocs de forme circulaire ou sphérique, le tenseur de perméabilité de bloc est toujours anisotrope. Par conséquent, dans toute méthode numérique, il est indispensable de pouvoir définir comme perméabilité de bloc un tenseur de covariance anisotrope car en simulation d'écoulement les blocs ne sont jamais circulaires ou sphériques. Dans leurs papiers, les auteurs décrivent pas à pas la procédure de génération directe de perméabilités de bloc. Sánchez-Vila *et al.* (1995) ont montré que la méthode de Rubin et Gómez-Hernández (1990) produit des perméabilités de bloc qui sont statistiquement très proches de celles d'Indelman et Dagan (1993 a, b).

### **Inconvénients du calcul direct de perméabilités de bloc**

La génération directe de perméabilités de bloc est très attractive car elle concentre les efforts dans la partie qui intéresse la personne en charge de la simulation des écoulements à savoir le calcul des perméabilités de bloc de la grille de simulation. La limitation première de ce type de méthodes est qu'elle s'applique uniquement à des blocs de taille uniforme et ne peut donc pas profiter de méthodes d'adaptation de la géométrie blocs. La raison est que des tailles différentes de bloc doivent être associées à des variables aléatoires différentes pour la perméabilité de bloc. L'introduction de tailles multiples de bloc auraient pour conséquence de multiplier le nombre de variables aléatoires à prendre en compte et le nombre de covariances à modéliser. Seules les approches analytiques de Rubin et Gómez-Hernández (1990) et Indelman et Dagan (1993 a, b), qui fournissent des expressions analytiques des perméabilités de bloc, pourraient s'appliquer à des tailles de bloc multiples.

#### **4.1.4 Conclusions sommaires sur le calcul de perméabilités équivalentes**

Ce chapitre a passé en revue l'essentiel des nombreuses méthodes existantes de calcul de la perméabilité équivalente. Un des résultats importants à retenir est que les perméabilités de bloc ne sont généralement pas propriétés intrinsèques des blocs mais dépendent des conditions d'écoulement dans leur voisinage. Selon les cas à modéliser et les contraintes de l'étude (contraintes de temps, limitation de machine, simulateur d'écoulement disponible), une méthode peut mieux être adaptée qu'une autre. Pour une approche stochastique dans laquelle l'objectif est de quantifier l'impact de l'incertitude géologique sur certaines caractéristiques d'écoulement ou de transport (par ex. des temps d'arrivée), une approche par génération directe de la perméabilité de bloc présente des avantages certains. Ces méthodes demandent cependant à être testées davantage avant de pouvoir être utilisées d'une manière systématique.

Chaque fois que cela est possible, il est cependant préférable d'avoir recours à des méthodes de résolution numérique qui, dans tous les cas, donnent de meilleurs résultats. Lorsque le milieu

présente de forts contrastes de perméabilité et que les corps de plus grande perméabilité révèlent une bonne continuité spatiale (typiquement une portée supérieure ou égale à la taille moyenne d'un bloc à l'échelle numérique), il est impératif de définir préalablement la géométrie des blocs qui convient le mieux pour reproduire les contrastes de perméabilité à l'échelle numérique. Une approche par grilles adaptatives couplée à une résolution numérique des perméabilités équivalentes est alors fortement recommandée.

## Références bibliographiques sur la perméabilité équivalente

1. Ababou, R. et Wood, E. F., 1990. Comment on "Effective groundwater model parameter values: Influence of spatial variability of hydraulic conductivity, leakance, and recharge" by J. J. Gómez-Hernández et S. M. Gorelick. *Water Resour. Res.*, 26(8), 1843-1846.
2. Aris, R., 1956. On the dispersion of a solute in a fluid flowing through a tube. *Proc. R. Soc. London. Ser. A*, 235, 67-77.
3. Bachu, S. et Cuthiell, D., 1990. Effects of core-scale heterogeneity on steady state and transient fluid flow in porous media: Numerical analysis. *Water Resour. Res.*, 26(5), 683-874.
4. Begg, S. H. et King, P. R., 1985. Modelling the effects of shales on reservoir performance: Calculation of effective vertical permeability. Paper SPE 13529.
5. Begg, S. H., Chang, D. M. et Haldorsen, H. H., 1985. A simple statistical method for calculating the effective vertical permeability of a reservoir containing discontinuous shales. Paper SPE 14271.
6. Begg, S. H. et Carter, R. R., 1987. Assigning effective values to simulator grid-block parameter for heterogeneous reservoir. Paper SPE 16754.
7. Begg, S. H., Carter, R. R. et Dranfield, P., 1989. Assigning effective values to simulator gridblock parameters for heterogeneous reservoirs. *SPE Reservoir Engineering*, 455-463, November, 1989.
8. Boe, O., 1994. Analysis of an upscaling method based on conservation of dissipation. *Transport in Porous media*, 17(1), 77-86.
9. Bouwer, H., 1969. Planing and interpreting soil permeability measurements. *J. of the Irrigation and Drainage Division of the A.S.C.E.*, IR(3), 391-402.
10. Brenner, H., 1980. Dispersion resulting from flow through spatially periodic porous media. *Philos. Trans. R. Soc. London, Ser. A*, 297 (1498), 81-133.
11. Cardwell, W. T. et Parsons, R. L., 1945. Averaging permeability of heterogeneous oil sands. *Trans. Am. Inst. Min. Metall. Pet. Eng.*, 160, 34-42.
12. Cliton, M. P. et Neuman, S. P., 1982. Effect of kriging and inverse modelling on conditional simulation of the Avra Valley aquifer in Southern Arizona. *Water Resour. Res.*, 18(4), 1215-1234.
13. Dagan, G., 1979. Models of groundwater flow in statistically homogeneous porous formations. *Water Resour. Res.*, 15(1), 47-63.
14. Dagan, G., 1982a. Stochastic modeling of groundwater flow by unconditional and conditional probabilities, 1. Conditional simulation and the direct problem. *Water Resour. Res.*, 18(4), 813-833.
15. Dagan, G., 1982b. Analysis of flow through heterogeneous random aquifers, 2. Unsteady flow in confined formations. *Water Resour. Res.*, 18, 1571-1585.
16. Dagan, G., 1985. Stochastic modeling of groundwater flow by unconditional and conditional probabilities, 2: The inverse problem. *Water Resour. Res.*, 21(1), 65-72.
17. Dagan, G., 1989. *Flow and transport in porous formations*. Springer-Verlag, New York.
18. Desbarats, A. J., 1987. Numerical estimation of effective permeability in sand-shale formation. *Water Resour. Res.*, 23(2), 273-286.
19. Desbarats, A. J., 1989. Support effect and the spatial averaging of transport properties. *Math. Geology*, 21(3), 383-389.
20. Desbarats, A. J., 1992a. Spatial averaging of transmissivity in heterogeneous fields with flow toward a well. *Water Resour. Res.*, 28(3), 757-767.
21. Desbarats, A. J., 1992b. Spatial averaging of hydraulic conductivity in three-dimensional heterogeneous porous media. *Math. Geology*, 24(3), 249-267.
22. Desbarats, A. J., 1993. Geostatistical analysis of interwell transmissivity in heterogeneous aquifers. *Water Resour. Res.*, 29(4), 1239-1246.
23. Desbarats, A. J. et Dimitrakopoulos, R., 1990. Geostatistical modeling of transmissivity for two-dimensional reservoir studies. *SPE Formation Evaluation, Dic.*, 437-443.
24. Deutsch, C. 1989. Calculating effective absolute permeability in sandstone/shale sequences. *SPE Formation Evaluation, Sept.*, 343-348.

25. Deutsch, C. et Journel, A. G., 1992. *GSLIB: Geostatistical Software Library and User's Guide*. Oxford University Press, New York, 340p.
26. Durlofsky, L. J et Chung, E. Y., 1990. Effective permeability of heterogeneous reservoir regions. 2nd European Conference on the Mathematics of Oil Recovery, Paris, 1990, 57-64.
27. Durlofsky, L. J., 1991. Numerical calculation of equivalent grid block permeability tensors for heterogeneous porous media. *Water Resour. Res.*, 27(5), 699-708.
28. Durlofsky, L. J., 1992. Representation of grid block permeability in coarse scale models of randomly heterogeneous porous media. *Water Resour. Res.*, 28(7), 1791-1800.
29. Durlofsky, L. J., Jones, R. C. et Milliken, W. J., 1994a. A new method for the scale up of displacement processes in heterogeneous reservoirs. 4th European Conference on the Mathematics of Oil Recovery, Roros, Norway, June, 1994.
30. Durlofsky, L. J., Milliken, W. J., Dehghani, K et Jones, R. C., 1994b. Application of a new scale up methodology to the simulation of displacement process in heterogeneous reservoirs. SPE International Petroleum Conference & Exhibition of Mexico, Veracruz, Oct. 1994.
31. Dykaar, B. B. et Kitanidis, P. K., 1992a. Determination of the effective hydraulic conductivity for heterogeneous porous media using a numerical spectral approach, 1. Method. *Water Resour. Res.*, 28(4), 1155-1166.
32. Dykaar, B. B. et Kitanidis, P. K., 1992b. Determination of the effective hydraulic conductivity for heterogeneous porous media using a numerical spectral approach, 2. Method. *Water Resour. Res.*, 28(4), 1167-1178.
33. Farmer, C. L., Heath, D. E. et Moody, R. O., 1991. A Global optimization approach to grid generation. 11th SPE Symposium on Reservoir Simulation, Anaheim, CA, Feb. 1991, 341-350.
34. Fayers, F. J. et Hewett, T. A., 1992. A review of current trends in petroleum reservoir description and assessment of the impacts on oil recovery. *Adv. in Water Resources*, 15(4), 341-365.
35. Garcia, M. H., Journel, A. G. et Aziz, K., 1990. An automatic grid generation and adjustment method for modeling reservoir heterogeneities, Report 3, Stanford Center for Reservoir Forecasting, Stanford.
36. Garcia, M., 1991, "Génération Automatique de Maillage Appliquée aux Domaines Complexes du Sous-sol", Rapport de Thèse, Ecole Nationale Supérieure des Mines de Paris, CGES, 204 p.
37. Garcia, M. H., Journel, A. G. et Aziz, K., 1992. Automatic grid generation for modeling reservoir heterogeneities. *SPE Reservoir Engineering*, 278-284.
38. Garcia, M., *3D-DEGA Software Version 1.0, Reference Manual & User's Guide*, FSS International r&d, janvier 1994, 138 p.
39. Garcia, M., Hadj Hassen, F. et Sonier, F., 1997. "Maillage Adaptatif : vers une Reconnaissance d'Entités Volumiques et l'Intégration de Contraintes Géométriques", présenté aux Journées Modélisation du Sous-Sol, Paris, 3-4 Feb. 1997.
40. Gelhar, L. W. et Axness, C. L., 1982. Three-dimensional stochastic analysis of macrodispersion in aquifers. *Water Resour. Res.*, 19(1), 161-180.
41. Gómez-Hernández, J. J. et Gorelick, S. M., 1989. Effective groundwater model parameter values: Inference of spatial variability of hydraulic conductivity, leakance, and recharge. *Water Resour. Res.*, 25(3), 405-419.
42. Gómez-Hernández, J. J., 1990. Simulation of block permeability conditioned upon data measured at a different scale. In "ModelCARE 90: Calibration and Reliability in Groundwater Modelling", IAHS Publ. no. 195, 407-416.
43. Gómez-Hernández, J. J., 1991. A Stochastic approach to the simulation of block conductivity fields conditioned upon data measured at a smaller scale. Ph.D. thesis, Stanford University.
44. Gómez-Hernández, J. J. et Wen, X. H., 1994. Probabilistic assessment of travel times in groundwater modeling. In press, *J. Stochastic Hydrology and Hydraulics*, 8(1), 19-55.
45. Grindheim, A. O., 1990. An evaluation of homogenization techniques for estimating effective absolute block permeability by use of a stochastic reservoir description simulator. Master thesis, Norwegian Institute of Technology (NTH).

46. Gutjahr, A. L., Gelhar, L. W., Bakr, A. A. et MacMillan, J. R., 1978. Stochastic analysis of spatial variability in subsurface flows, 2. Evaluation and application. *Water Resour. Res.*, 14(5), 953-959.
47. Haldorsen, H. H. et Lake, L. W., 1984. A new approach of shales management in field scale simulation models. *Soc. of Petrol. Eng. Jour.*, 24 (Aug.), 447-457.
48. Haldorsen, H. H., 1986. Simulation parameters assignment and the problem of scale in reservoir engineering. In Lake, L. W. and Carroll, H. B. (Eds.), *Reservoir Characterization*, Academic Press, 445-485.
49. Haldorsen, H. H. et Chang, D. M., 1986. Notes on stochastic shales; from outcrop to simulation model. In Lake, L. W. and Carroll, H. B. (Eds.), "Reservoir Characterization", Academic Press, 445-485.
50. Hoeksema, R. J. et Kitanidis, P. K., 1984. An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling. *Water Resour. Res.*, 20(7), 1003-1020.
51. Hoeksema, R. J. et Kitanidis, P. K., 1985. Comparison of Gaussian conditional mean and kriging estimation in the geostatistical solution of the inverse problem. *Water Resour. Res.*, 21(6), 825-836.
52. Holden, L., Hoiberg, J. et Lia, O., 1989. Homogenization of absolute permeability. Norwegian Computer Centre (NCC) Report.
53. Holden, L. et Lia, O., 1992. A tensor estimator for the homogenization of absolute permeability. *Transport in porous media*, 8, 37-46.
54. Indelman, P et Dagan, G., 1993a. Upscaling of permeability of anisotropic heterogeneous formations, 1. The general framework. *Water Resour. Res.*, 29(4), 917-923.
55. Indelman, P et Dagan, G., 1993b. Upscaling of permeability of anisotropic heterogeneous formations, 2. General structure and small perturbation analysis. *Water Resour. Res.*, 29(4), 925-933.
56. Journel, A., G., Deutsch, C., et Desbarats, A. J., 1986. Power averaging for block effective permeability. SPE 15128.
57. Journel, A. G. et Huijbregts, C. J., 1978, *Mining Geostatistics*, Academic Press, London.
58. Kasap, E. et Lake, L. W., 1989. An analytical method to calculate the effective permeability tensor of a grid block and its application in an outcrop study. SPE 18434.
59. King, P. R., 1989. The use of renormalization for calculating effective permeability. *Transport in Porous Media*, 4(1), 37-58.
60. Kirkpatrick, S., 1973. Percolation and conduction. *Reviews of Modern Physics*, 45(4), 574-588.
61. Kitanidis, P. K., 1990. Effective hydraulic conductivity for gradually varying flow. *Water Resour. Res.*, 26(6), 1197-1208.
62. Lachassagne, P., Ledoux, E., et de Marsily, G., 1989, Evaluation of hydrogeological parameters in heterogeneous porous media. In Sahuquillo, A., Andreu, J., and O'Donnell (Eds.), *Groundwater Management: Quantity and Quality*, IAHS Publication No. 188, IAHS.
63. Lake, L. W., 1988. The origins of anisotropy. *Journal of Petroleum Technology*, 395- 396.
64. Malick, K. M., 1995. Boundary effects in the successive upscaling of absolute permeability. Master thesis, Stanford University.
65. Mantoglou, A. et Gelhar, L. W., 1987. Effective hydraulic conductivity of transient unsaturated flow in stratified soils. *Water Resour. Res.*, 23(1), 57-67.
66. Matheron, G., 1967. *Éléments pour une théorie des milieux*. Mason et Cie.
67. Mohanty, S. et Sharma, M. M., 1990. A recursive method for estimating single and multiphase permeabilities. SPE paper 20477 presented at the 65th Annual Technical Conference and Exhibition in New Orleans, Sept. 23-26, 1990.
68. Muggeridge, A. H., 1991. Generation of effective relative permeabilities from detailed simulation of flow in heterogeneous porous media. In "Reservoir Characterization II", edited by Lake, L. W., Carroll, H. B. and Wesson, T. C., Academic Press, San Diego, pp197-225.
69. Naff, R. L., 1991. Radial flow in heterogeneous porous media: An analysis of specific discharge. *Water Resour. Res.*, 27(3), 307-316.

70. Neuman, S. P. et Orr, S., 1993. Prediction of steady state flow in nonuniform geologic media by conditional moments: Exact nonlocal formalism, effective conductivities, and weak approximation. *Water Resour. Res.*, 29(2), 341-364.
71. Pickup, G. E., J. L. Jensen, P. S. Ringrose et K. S. Sorbie, 1992. A method for calculating permeability tensors using perturbed boundary conditions. In 3rd European Conf. on the Mathematics of Oil Recovery, Delft, June, 1992.
72. Pickup, G. E., Ringrose, P. S., Jensen, J. L. et Sorbie, K. S., 1994. Permeability tensors for sedimentary structures. *Math. Geology*, 26(2), 227-250.
73. Poley, A. D., 1988. Effective permeability and dispersion in locally heterogeneous aquifers. *Water Resour. Res.*, 24(11), 1921-1926.
74. Renard, P., 1997, "Modélisation des écoulements en milieux poreux hétérogènes, calcul des perméabilités équivalentes", Rapport de Thèse, Ecole Nationale Supérieure des Mines de Paris, CIG, 246 p.
75. Robey, T. H., 1995. An adaptive grid technique for minimizing heterogeneity of cells or elements. *Math. Geology*, 27(6), 709-729.
76. Rubin, Y. et Gómez-Hernández, J. J., 1990. A stochastic approach to the problem of upscaling of conductivity in disordered media: Theory and unconditional numerical simulations. *Water Resour. Res.*, 26(4), 691-701.
77. Russo, D., 1992. Upscaling of hydraulic conductivity in partially saturated heterogeneous porous formation. *Water Resour. Res.*, 28(2), 397-409.
78. Sáez, A., Otero, C. J. et Rusinek, I., 1989. The effective homogeneous behavior of heterogeneous porous media. *Transport in Porous Media*, 4, 213-238.
79. Sánchez-Vila, X., 1995. On the geostatistical formulations of the groundwater flow and solute transport equations. Ph.D. thesis, Universitat Politècnica de Catalunya, Barcelona, Spain.
80. Sánchez-Vila, X., Girardi, J. P. et Carrera, J., 1995. A synthesis of approaches to upscaling of hydraulic conductivity. *Water Resour. Res.*, 31(4), 867-882.
81. Shah, N. et Ottino, J. M., 1986 Effective transport properties of disordered, multiphase composites: Application of real-space renormalization group theory. *Chem. Eng. Sci.*, 41(2), 283-296.
82. Tran, T. T., 1995, Stochastic simulation of permeability field and their scale-up for flow modeling. Ph. D. thesis, Stanford University.
83. Tran, T. T. et Journel, A. G., 1995. Automatic generation of corner-point-geometry flow simulation grids from detailed geostatistical descriptions. *Adv. of Water Resour.*, in press.
84. Warren, J. E. et Price, H. S., 1961. Flow in heterogeneous porous media. *Society of Petroleum Engineering Journal*, 1: 153-169.
85. Wen, X. H., 1996. Stochastic simulation of groundwater flow and mass transport in heterogeneous formations: Conditioning and problem of scales. Ph.D. thesis, Universidad Politécnica de Valencia, Spain, In preparation.
86. White, C. D., 1987. Representation of heterogeneity for numerical reservoir simulation. Ph.D. thesis, Stanford University.
87. White, C. D. et Horne, R. N., 1987. Computing absolute transmissivity in the presence of fine-scale heterogeneity. Paper SPE 16011, 209-221.
88. Williams, J. K., 1989. Simple renormalization schemes for calculating effective properties of heterogeneous reservoirs. In 1st European Conf. on the Mathematics of oil recovery, Cambridge, July 1989, 281-298.
89. Wilson, K. G., 1979. Problems in physics with many scales length. *Sci. Am.* 241 (August), 158-179.
90. Yamada, T., 1995. A dissipation-based coarse grid system and its application to the scale-up of two phase problems. Ph.D. thesis, Stanford University.
91. Yeh, T. C., Gelhar, L. W. et Gutjahr, A. L., 1985. Stochastic analysis of unsaturated flow in heterogeneous soils, 1, Statistically isotropic media. *Water Resour. Res.*, 21(4), 447-456.

## **4.2 Paramètres de transport de grande échelle en milieu hétérogène**

Il est inhabituel pour ne pas dire exceptionnel de chercher à modéliser à l'échelle de la mesure des paramètres de transport tels que les coefficients de dispersion pour ensuite calculer des valeurs équivalentes de bloc. Une approche possible serait de modéliser de manière stochastique à l'échelle de la mesure les lithofaciès ou des types de roche d'un aquifère et, à la condition de pouvoir affecter des propriétés de transport spécifiques à chaque type de roche, de déduire des valeurs de ces paramètres aux cellules. L'étape suivante consisterait alors à passer à des blocs à une échelle numérique et si les blocs sont hétérogènes de calculer des valeurs équivalentes aux blocs.

La seule publication qui a été trouvée sur le sujet est de Dagan.

### **Référence bibliographique**

1. Dagan, G., 1993. Upscaling of dispersion coefficients in transport through heterogeneous formations, présenté au GXG meeting, Sandia National Laboratories, mars 1993.



**ANNEXE 1 :**  
**VARIANCE D'UNE COMBINAISON LINEAIRE PONDEREE**  
**DE VARIABLES ALEATOIRES**



Soit une variable aléatoire  $Z^*(\mathbf{u})$  définie comme une combinaison linéaire de  $n$  variables aléatoires  $Z(\mathbf{u}_\alpha)$  avec  $\alpha = 1, \dots, n$  :

$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_\alpha Z(\mathbf{u}_\alpha).$$

On cherche à calculer sa variance qui s'écrit comme suit :

$$\text{Var}\{Z^*(\mathbf{u})\} = \text{Var}\left\{\sum_{\alpha=1}^n \lambda_\alpha Z(\mathbf{u}_\alpha)\right\}.$$

Conformément à la définition de la variance, on obtient :

$$\begin{aligned} \text{Var}\left\{\sum_{\alpha=1}^n \lambda_\alpha Z(\mathbf{u}_\alpha)\right\} &= E\left\{\left[\sum_{\alpha=1}^n \lambda_\alpha Z(\mathbf{u}_\alpha)\right]^2\right\} - E\left\{\left[\sum_{\alpha=1}^n \lambda_\alpha Z(\mathbf{u}_\alpha)\right]\right\}^2 \\ &= E\left\{\sum_{\alpha=1}^n \lambda_\alpha Z(\mathbf{u}_\alpha) \cdot \sum_{\beta=1}^n \lambda_\beta Z(\mathbf{u}_\beta)\right\} - E\left\{\sum_{\alpha=1}^n \lambda_\alpha Z(\mathbf{u}_\alpha)\right\} E\left\{\sum_{\beta=1}^n \lambda_\beta Z(\mathbf{u}_\beta)\right\} \\ &= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_\alpha \lambda_\beta E\{Z(\mathbf{u}_\alpha)Z(\mathbf{u}_\beta)\} - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_\alpha \lambda_\beta E\{Z(\mathbf{u}_\alpha)\} E\{Z(\mathbf{u}_\beta)\} \\ &= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_\alpha \lambda_\beta \{E[Z(\mathbf{u}_\alpha)Z(\mathbf{u}_\beta)] - E[Z(\mathbf{u}_\alpha)]E\{Z(\mathbf{u}_\beta)\}\} \\ &= \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_\alpha \lambda_\beta C\{Z(\mathbf{u}_\alpha)Z(\mathbf{u}_\beta)\} \end{aligned}$$

d'où on tire

$$\text{Var}\left\{\sum_{\alpha=1}^n \lambda_\alpha Z(\mathbf{u}_\alpha)\right\} = \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_\alpha \lambda_\beta C\{Z(\mathbf{u}_\alpha)Z(\mathbf{u}_\beta)\}.$$



**ANNEXE 2 :**  
**BREF APERÇU DU MODELE DE MARKOV**



On considère deux fonctions aléatoires stationnaires  $Z_1(\mathbf{u})$  et  $Z_2(\mathbf{u})$ . Sans perdre en généralité, on peut supposer que ces fonctions aléatoires ont une moyenne nulle et une variance unitaire. Leurs covariances peuvent alors s'exprimer comme suit :

$$\begin{aligned}\rho_1(\mathbf{h}) &= E\{Z_1(\mathbf{u})Z_1(\mathbf{u} + \mathbf{h})\} \\ \rho_2(\mathbf{h}) &= E\{Z_2(\mathbf{u})Z_2(\mathbf{u} + \mathbf{h})\} \\ \rho_{12}(\mathbf{h}) &= \rho_{21}(\mathbf{h}) = E\{Z_2(\mathbf{u})Z_1(\mathbf{u} + \mathbf{h})\}.\end{aligned}$$

On se place à présent dans le cas où les conditions suivantes sont vérifiées :

$$(A2-1) \quad E\{Z_2(\mathbf{u}) \mid Z_1(\mathbf{u}) = z\} = \rho_{12}(0) z$$

$$(A2-2) \quad E\{Z_2(\mathbf{u}) \mid Z_1(\mathbf{u}) = z, Z_1(\mathbf{u} + \mathbf{h}) = z'\} = E\{Z_2(\mathbf{u}) \mid Z_1(\mathbf{u}) = z\}, \forall \mathbf{h}, z$$

La première condition signifie que la régression de  $Z_2$  par rapport à  $Z_1$  est linéaire, condition qui est vérifiée si la distribution jointe de  $Z_1(\mathbf{u})$  et  $Z_2(\mathbf{u})$  est normale. La seconde condition s'interprète de la manière suivante : le donnée co-localisée  $Z_1(\mathbf{u}) = z$  "occulte" l'influence de toute autre donnée dure  $Z_1(\mathbf{u} + \mathbf{h}) = z'$  sur la donnée secondaire co-localisée de  $Z_2(\mathbf{u})$ .

**Théorème :** à partir des conditions A2-1 et A2-2, l'expression suivante peut être déduire pour la covariance croisée :

$$\rho_{12}(\mathbf{h}) = \rho_{12}(0)\rho_1(\mathbf{h}), \forall \mathbf{h}.$$

**Démonstration :** Soit  $f_{\mathbf{h}}(z, z')$  la fonction bivariée de distribution de probabilité des deux variables aléatoires  $Z_1(\mathbf{u})$  et  $Z_1(\mathbf{u} + \mathbf{h})$ . La covariance croisée  $\rho_{12}(\mathbf{h})$  de ces variables s'écrit :

$$\rho_{12}(\mathbf{h}) = E\{Z_2(\mathbf{u})Z_1(\mathbf{u} + \mathbf{h})\}.$$

Si on développe on obtient :

$$\begin{aligned}\rho_{12}(\mathbf{h}) &= \iint E\{Z_2(\mathbf{u})Z_1(\mathbf{u} + \mathbf{h}) \mid Z_1(\mathbf{u}) = z, Z_1(\mathbf{u} + \mathbf{h}) = z'\} f_{\mathbf{h}}(z, z') dz dz' \\ &= \iint z' E\{Z_2(\mathbf{u}) \mid Z_1(\mathbf{u}) = z, Z_1(\mathbf{u} + \mathbf{h}) = z'\} f_{\mathbf{h}}(z, z') dz dz' \\ &= \iint z' E\{Z_2(\mathbf{u}) \mid Z_1(\mathbf{u}) = z\} f_{\mathbf{h}}(z, z') dz dz', \text{ en utilisant la condition A2-1} \\ &= \rho_{12}(0) \iint z z' f_{\mathbf{h}}(z, z') dz dz', \text{ en utilisant la condition A2-2} \\ &= \rho_{12}(0) \cdot \rho_1(\mathbf{h}), \text{ en utilisant la définition de } \rho_1(\mathbf{h})\end{aligned}$$



**ANNEXE 3 :**  
**PUBLICATION SUR LE MODELE DE MARKOV-BAYES**



**ANNEXE 4 :**  
**PUBLICATION SUR LA THEORIE ET LA PRATIQUE**  
**DE LA SIMULATION SEQUENTIELLE**



