

**Titre:** Caractérisation stochastique des hétérogénéités  
hydrostratigraphiques appliquée à la modélisation hydrogéologique régionale  
**Title:**

**Auteur:** Nicolas Benoit  
**Author:**

**Date:** 2020

**Type:** Mémoire ou thèse / Dissertation or Thesis

**Référence:** Benoit, N. (2020). Caractérisation stochastique des hétérogénéités  
hydrostratigraphiques appliquée à la modélisation hydrogéologique régionale  
**Citation:** [Thèse de doctorat, Polytechnique Montréal]. PolyPublie.  
<https://publications.polymtl.ca/5294/>

 **Document en libre accès dans PolyPublie**  
Open Access document in PolyPublie

**URL de PolyPublie:**  
PolyPublie URL: <https://publications.polymtl.ca/5294/>

**Directeurs de recherche:** Denis Marcotte, & John W. Molson  
**Advisors:**

**Programme:** Génie minéral  
**Program:**

**POLYTECHNIQUE MONTRÉAL**

affiliée à l'Université de Montréal

**Caractérisation stochastique des hétérogénéités hydrostratigraphiques  
appliquée à la modélisation hydrogéologique régionale**

**NICOLAS BENOIT**

Département de génie civil, géologique et des mines

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*  
Génie minéral

Juin 2020

**POLYTECHNIQUE MONTRÉAL**

affiliée à l'Université de Montréal

Cette thèse intitulée :

**Caractérisation stochastique des hétérogénéités hydrostratigraphiques  
appliquée à la modélisation hydrogéologique régionale**

présentée par **Nicolas BENOIT**

en vue de l'obtention du diplôme de *Philosophiæ Doctor*  
a été dûment acceptée par le jury d'examen constitué de :

**James GOULET**, président

**Denis MARCOTTE**, membre et directeur de recherche

**John MOLSON**, membre et codirecteur de recherche

**Gabriel FABIEN-OUELLET**, membre

**Thomas HERMANS**, membre externe

## DÉDICACE

*À Mélanie, Alice et Marie-Jeanne,  
vous m'inspirez à me dépasser....*



## REMERCIEMENTS

Je remercie mon directeur de thèse Denis Marcotte pour d'abord avoir cru à ce projet. Denis m'a toujours témoigné une grande confiance et un support constant. Travailler avec Denis c'est agréable, efficace et très inspirant. Merci à mon codirecteur John Molson pour les nombreuses discussions constructives sur la modélisation hydrogéologique, pour son intérêt dans mon projet ainsi que son soutien avec ses logiciels de modélisation.

Un grand merci à la Commission géologique du Canada (CGC), plus particulièrement Andrée Bolduc directrice à la division de Québec et Yves Michaud chef du Programme géoscientifique des eaux souterraines, de m'avoir donné l'occasion de faire cette thèse à travers mes activités de recherche à la CGC. Je remercie Hazen Russell qui a accepté d'intégrer cette thèse dans son projet de recherche sur les eaux souterraines du sud de l'Ontario. Merci à Ryan Post du Nottawasaga Conservation Authority et Andy Bajc de l'Ontario Geological Survey pour leur support et les données fournies.

Je remercie mon collègue et ami Miro Nastev qui m'a toujours encouragé à faire une thèse et qui me soutient professionnellement depuis de nombreuses années. Merci à mon bon ami Alexandre Boucher pour ses encouragements et ses idées. Un merci spécial à mes amis de Running Blind, notre groupe rock avec qui je me détends à la basse chaque lundi soir !

Évidement, un très grand merci à ma conjointe Mélanie et mes merveilleuses filles Alice et Marie-Jeanne, pour leur amour, leurs grands encouragements et soutien moral durant toutes ces années.

## RÉSUMÉ

L'une des principales sources d'incertitude des modèles hydrogéologiques porte sur les hétérogénéités hydrostratigraphiques. Ces dernières jouent un rôle majeur sur l'écoulement de l'eau souterraine et le transport des contaminants. Pourtant la caractérisation hydrogéologique inclut très rarement l'évaluation de l'effet de l'incertitude des hétérogénéités. Cette situation s'explique en partie par la difficulté à reproduire, de manière satisfaisante, ces caractéristiques physiques en modèles équivalents. Peu ou pas de méthodologie complète et efficace permet d'effectuer la modélisation hydrogéologique régionale dans un cadre stochastique tout en quantifiant l'incertitude des hétérogénéités inter et intra unités. La difficulté de simuler efficacement les hétérogénéités inter unités de ce type de contexte est un frein à la modélisation hydrogéologique stochastique régionale. Également, la difficulté associée à la paramétrisation stochastique régionale intra unités (conductivité hydraulique ou  $K$ ) de ces modèles représente aussi un obstacle. Ces difficultés constituent une problématique majeure quand vient le besoin de quantifier adéquatement l'incertitude des modèles.

Les méthodes de simulation géostatistique moderne, telles que la méthode plurigaussienne et multipoint, permettent de produire des modèles équivalents des hétérogénéités inter unités, appelés réalisations. Malgré leurs récents développements, ces méthodes conservent différentes lacunes telles que le manque de réalisme géologique dans certains contextes. Par exemple, la méthode plurigaussienne ne peut facilement incorporer des transitions asymétriques entre unités, de même que la méthode multipoint qui éprouve aussi des difficultés avec la non-stationnarité ainsi que les forts contrastes d'épaisseurs. Par contre, la méthode MCP peut s'affranchir de ces difficultés notamment grâce au transiogramme et sa propriété de forçage 0/1 des probabilités.

Les hétérogénéités intra unités des modèles hydrogéologiques sont définies à partir des données de  $K$  scalaire à une échelle locale. Cette échelle n'est généralement pas représentative de la taille des éléments d'un modèle hydrogéologique régional. La caractérisation stochastique de la  $K$  équivalente (tenseur- $K$ ) à l'échelle de l'élément est un défi en soi. Elle nécessite plusieurs étapes incluant des contraintes à respecter dont la préservation de la structure spatiale et la contrainte d'inégalité  $K_{\text{verticale}} \leq K_{\text{horizontale}}$  des tenseurs ainsi que les corrélations inter composantes. À cela s'ajoute, la mise à l'échelle de la  $K$  qui tient compte de l'effet des hétérogénéités locale régionalisée, la régionalisation des composantes corrélées des tenseurs- $K$  des différentes unités ainsi que leur calage avec les variables d'état (p. ex. charge hydraulique) dans un contexte multivariable contraint. Une méthodologie complète devrait également four-

nir des réalisations significativement différentes afin de quantifier adéquatement l'incertitude associée aux tenseurs- $K$ .

L'objectif général de la thèse consiste à développer une démarche complète de caractérisation stochastique de l'incertitude des hétérogénéités inter et intra unités hydrostratigraphiques appliquées à la modélisation hydrogéologique régionale. Plus spécifiquement, la thèse vise à développer une méthodologie de simulation géostatistique des unités hydrostratigraphiques (UHS) afin de fournir des modèles équivalents d'aspect réaliste pour un contexte de sédimentation directionnelle (asymétrie) et à développer une méthodologie de caractérisation stochastique, de mise à l'échelle et de calage des tenseurs- $K$  équivalents à partir de la  $K$  scalaire tout en préservant la structure spatiale des tenseurs, la contrainte d'inégalité ainsi que les corrélations inter composantes.

La méthodologie globale s'oriente autour de trois étapes principales : i) la compilation et la préparation des données, ii) la simulation de modèles hydrostratigraphiques directionnels et iii) la caractérisation stochastique des tenseurs- $K$ . La première étape a porté sur la compilation et le traitement des données (hydrostratigraphie et  $K$ ) pour les développements méthodologiques et les différents tests méthodologiques appliqués sur le sous-bassin Innisfil Creek (comté de Simcoe, Ontario). La deuxième étape est basée sur le développement d'une démarche de simulation par la méthode MCP afin de générer des modèles équivalents dans un contexte hydrostratigraphique 3D complexe dont la séquence des unités est ordonnée verticalement. La méthodologie proposée inclut l'estimation des probabilités bivariées par transformée de Fourier à partir du modèle déterministe et des critères de qualité de simulation. La méthode est testée pour montrer sa capacité à quantifier la variabilité locale de la solution et son incertitude. La troisième étape fournit une méthodologie complète de caractérisation stochastique des hétérogénéités intra UHS qui inclut : la mise à l'échelle d'hétérogénéités locales ( $K$ ) en tenseur- $K$  équivalent, l'utilisation du modèle linéaire de corégionalisation pour la simulation des composantes des tenseurs- $K$  équiprobables, la préservation des corrélations entre les composantes des tenseurs- $K$  par nouvelle transformation gaussienne bivariée et l'utilisation de la méthode de déformation graduelle (GDM) pour caler les tenseurs- $K$  en fonction des observations de charge hydraulique.

Les résultats montrent la capacité de la méthode MCP à simuler efficacement un contexte hydrostratigraphique régional 3D complexe avec de multiples unités. De plus, les résultats montrent la capacité de la méthode à reproduire l'ordination verticale des unités et tout ça malgré la forte hypothèse d'indépendance conditionnelle. Les résultats des simulations conditionnelles et non conditionnelles pour le cas synthétique montrent clairement l'absence de tout biais substantiel. Dans l'exemple plus complexe du contexte réel, le biais est moins

important pour le cas non conditionnel et pour l'échantillonnage aléatoire des données que pour le biais observé dans les données dû à l'échantillonnage préférentiel. La surreprésentation de données conditionnantes n'a pas d'impact significatif sur les proportions moyennes simulées mais l'ajout de données supplémentaires réduit la variabilité. Soulignons que la méthode MCP fournit de bons résultats malgré une image d'entraînement fortement non stationnaire. Avec une autre méthode (PGS ou MPS), rien n'aurait garanti un tel succès dans le contexte étudié.

Les résultats de la méthodologie de caractérisation de la  $K$  montrent que la nouvelle méthode proposée fournit des modèles équivalents calés des tenseurs- $K$ . La méthode de mise à l'échelle de la  $K$  scalaire en tenseur- $K$  équivalent tient compte de la structure des hétérogénéités locales et des propriétés statistiques de chacune des UHS. La régionalisation des composantes des tenseurs- $K$  effectuée par simulation géostatistique à partir de modèles linéaires de corrélation (LCM) rend le temps de calcul raisonnable. La simulation reproduit aussi, pour chacune des unités, les caractéristiques statistiques des différentes composantes de  $K$ . Pour rétablir les corrélations non linéaires des tenseurs- $K$  non-gaussiens ainsi que les histogrammes des composantes souhaités, une nouvelle transformation bivariée est appliquée sur les réalisations. Les résultats du calage des tenseurs- $K$  montrent que la méthode des déformations graduelles permet le calage multivariable contraint tout en préservant la covariance spatiale et croisée de UHS. Il a été démontré que la méthode conserve les relations non linéaires entre les composantes du tenseur- $K$  grâce à l'intégration de la transformation bivariée. L'approche proposée a ainsi permis de préserver les propriétés spatiales des tenseurs ainsi que les relations entre les composantes. Les résultats démontrent que la GDM s'applique à un contexte multivariable contraint et pas seulement au contexte univariable. Soulignons que les inversions convergent rapidement en termes du nombre d'itérations grâce à l'optimisation du paramètre de déformation de la GDM. Il a aussi été montré que la méthode permet de caler la recharge séparément ou simultanément avec les tenseurs- $K$ . Les exemples de calage montrent qu'il est avantageux d'inclure la recharge stochastique dans l'inversion. Par ailleurs, il a été observé qu'un nombre élevé d'observations de charge n'améliore pas significativement les résultats. Une comparaison, portant sur la définition de zones de captage autour de deux puits de pompage, entre la méthode GDM proposée et la méthode d'inversion PEST montrent des tenseurs- $K$  géologiquement plus réalistes avec GDM que ceux obtenus avec PEST.

L'ensemble des résultats permet de conclure que la méthode MCP simule efficacement un contexte régional complexe d'une succession stratigraphique directionnelle d'un bassin sédimentaire glaciaire. La méthode prend efficacement en charge la forte non-stationnarité spatiale des unités ainsi que d'imposantes contraintes comme la déposition verticale ordonnée en tenant compte de l'asymétrie stratigraphique à partir des probabilités de transition entre les unités. La méthode MCP s'est révélée sans biais dans le cas non conditionnel. La méthodolo-

gie proposée de caractérisation des tenseurs- $K$  permet le calage individuel ou simultané des tenseurs- $K$  de différentes UHS et/ou de recharge. La méthodologie inclut une mise à l'échelle des conductivités hydrauliques quasi ponctuelles vers un tenseur équivalent en tenant compte de la structure spatiale à l'échelle locale. Bien que la mise à l'échelle soit intensive en temps de calcul, il a été démontré qu'il est possible de limiter au minimum le nombre d'éléments à mettre à l'échelle grâce à la régionalisation des tenseurs- $K$  par simulation géostatistique.

La thèse a apporté quelques contributions originales. D'abord, il s'agit d'une première application de la méthode MCP pour la simulation 3D. À ma connaissance, aucune autre méthode de simulation stochastique ne semble être en mesure de solutionner avec autant de satisfaction un contexte hydrostratigraphique complexe 3D comme celui du comté de Simcoe. La méthode des déformations graduelles permet le calage efficace des composantes des tenseurs- $K$  avec préservation des corrélations non linéaires des composantes et de leurs fonctions de covariance. À ma connaissance, il s'agit de la première approche utilisant la GDM pour un contexte multivariable avec contraintes. Enfin, cette thèse contribue à l'avancement des connaissances et des techniques de la caractérisation stochastique des hétérogénéités inter et intra hydrostratigraphiques. Dans une plus grande perspective, la thèse contribue à fournir des outils adaptés pour mieux quantifier l'incertitude associée aux hétérogénéités présente dans les modèles hydrogéologiques régionaux.

## ABSTRACT

One of the main sources of uncertainty in hydrogeological models relates to hydrostratigraphic heterogeneities. They play a major role in quantification of groundwater dynamics and mass transport. Yet hydrogeological characterization rarely includes the assessment of the effect of uncertainty and spatial variability of hydrostratigraphic units and hydrogeologic parameters. This is partly explained by the difficulty of reproducing these physical characteristics satisfactorily in equivalent models. There is at present little or no comprehensive and efficient methodology that allows regional hydrogeological modelling to be carried out in a stochastic framework while quantifying the uncertainty of inter- and intra-unit heterogeneities. The difficulty to effectively simulate inter-unit heterogeneities under such conditions is a hindrance to stochastic regional hydrogeological modelling. As well, the difficulty associated with the regional stochastic intra-unit parameterization (hydraulic conductivity or  $K$ ) of these models further aggravates the problem. These difficulties constitute a major issue when it comes to the need to adequately quantify model uncertainty. Modern geostatistical simulation methods, such as plurigaussian and multipoint simulations, make it possible to produce equivalent models of inter-unit heterogeneities, referred to as realizations. Despite their recent developments, these methods retain various shortcomings such as the lack of geological rigour in certain settings. For example, the plurigaussian method cannot easily incorporate asymmetrical transitions between units, whereas the multipoint method has difficulties with non-stationarity and strong thickness contrasts. On the other hand, the MCP (Markov-type categorical prediction) method has the capacity to overcome these difficulties thanks to the transiogram and its 0/1 probability forcing property.

The intra-unit heterogeneities of the hydrogeological models are defined from scalar  $K$  data at a local scale. This scale is generally not representative of the size of the elements used in regional hydrogeological models. Stochastic characterization of the equivalent  $K$  ( $K$ -tensor) at the element scale is a challenge in itself. It requires several steps including constraints to be respected such as the preservation of the spatial structure, the natural constraint  $K_{vertical} \leq K_{horizontal}$ , and the inter-component correlations. In addition, the upscaling of  $K$  takes into account the effect of local heterogeneities, the regionalization of the correlated components of the  $K$ -tensors of the different units and their calibration with state variables (e.g., hydraulic head) in a constrained multivariate context. A comprehensive methodology should also provide significantly different realizations in order to adequately quantify the uncertainty associated with the  $K$ -tensors.

The general objective of this thesis is to develop a comprehensive approach to stochastic characterization of the uncertainty of inter- and intra-hydrostratigraphic unit heterogeneities applied to regional hydrogeological modelling. More specifically, the thesis aims to develop a methodology for geostatistical simulation of hydrostratigraphic units (HSU) in order to provide equivalent models of realistic appearance for a directional sedimentation context (asymmetry) and to develop a methodology for stochastic characterization, upscaling and calibration of equivalent  $K$ -tensors from the scalar  $K$  while preserving the spatial structure of the tensors, the inequality constraint as well as the inter-component correlations.

The overall approach revolves around three main steps: (i) data compilation and preparation, (ii) simulation of directional hydrostratigraphic models, and (iii) stochastic characterization of  $K$ -tensors. The first step focuses on data compilation and processing (hydrostratigraphy and  $K$ ) for methodological developments and various methodological tests applied to the Innisfil Creek sub-watershed (Simcoe County, Ontario). The second stage is based on the development of a simulation approach using the MCP method to generate equivalent models in a complex 3D hydrostratigraphic context with a vertically ordered sequence of units. The proposed methodology includes the estimation of bivariate probabilities by Fourier transform from the deterministic model and the simulation quality criteria. The method is tested to show its ability to quantify the local variability of the solution and its uncertainty. The third step provides a complete methodology for stochastic characterization of intra-HSU heterogeneities which includes: upscaling of local heterogeneities ( $K$ ) into equivalent  $K$ -tensors, use of the linear co-regionalization model for simulation of equiprobable  $K$ -tensor components, preservation of correlations between  $K$ -tensor components by new bivariate Gaussian transformation and use of the Gradual Deformation Method (GDM) to calibrate the  $K$ -tensors according to hydraulic head observations.

The results show the ability of the MCP method to effectively simulate a complex 3D regional hydrostratigraphic with multiple units. Furthermore, the results show the ability of the method to reproduce the vertical ordering of the units, despite the strong assumption of conditional independence. The results of the conditional and unconditional simulations for the synthetic case clearly show the absence of any substantial bias. In the more complex example of field hydrogeological settings, the bias is less important in the unconditional case and in the random sampling case than the bias observed in the data due to preferential sampling. Over-representation of conditioned data does not have a significant impact on the mean simulated proportions, but the addition of supplementary data reduces variability. The MCP method provides good results despite a highly non-stationary training image. With another method (PGS or MPS), nothing would have guaranteed such success in the studied context.

The results of the  $K$  characterization methodology show that the presented original method provides equivalent calibrated models of the  $K$ -tensors. The scalar  $K$  upscaling method to equivalent  $K$ -tensor takes into account the structure of local heterogeneities and the statistical properties of each of the HSU. The regionalization of the  $K$ -tensor components by geostatistical simulation using Linear Co-Regionalization Models (LCMs) keeps the computation time reasonable. The simulation also reproduces, for each of the units, the statistical characteristics of the different  $K$  components. To restore the non-linear correlations of the non-Gaussian  $K$ -tensors as well as the histograms of the desired components, a new bivariate transformation is applied to the realizations. The results of the calibration of the  $K$ -tensors show that the method of gradual deformations allows the multivariate constrained calibration while preserving the spatial and cross-covariance of HSU. The method has been shown to preserve the non-linear relationships between the  $K$ -tensor components through the integration of the bivariate transformation. The proposed approach has thus preserved the spatial properties of the tensors and the relationships between the components. The results show that the GDM applies to a constrained multivariate context and not only to the univariate context. It should be noted that the inversions converge rapidly in terms of the number of iterations thanks to the optimization of the deformation parameter of the GDM. It has also been shown that the method allows to set the recharge separately or simultaneously with the  $K$ -tensors. The calibration examples show that it is advantageous to include stochastic recharge in the inversion. It has also been observed that a high number of head observations does not significantly improve the results. A comparison of the definition of capture zones around two pumping wells between the proposed GDM method and the PEST inversion method shows geologically more realistic  $K$ -tensors with GDM than those obtained with PEST.

The overall results suggest that the LCM method effectively simulates a complex regional context of a directional stratigraphic succession in a glacial sedimentary basin. The method effectively supports the strong spatial non-stationarity of the units as well as imposing constraints such as ordered vertical deposition by taking into account stratigraphic asymmetry from the transition probabilities between the units. The MCP method proved to be unbiased in the unconditional case. The proposed  $K$ -tensor characterization methodology allows individual or simultaneous calibration of  $K$ -tensors from different HSU and/or recharge units. The methodology includes an upscaling of quasi-point hydraulic conductivities to an equivalent tensor taking into account the spatial structure at the local scale. Although the upscaling is computationally intensive, it has been shown that it is possible to minimize the number of elements to be scaled by regionalizing the  $K$ -tensors by geostatistical simulation.

The thesis provided substantial original contributions. This is a first application of the MCP



method for 3D simulations. To the author's knowledge, no other stochastic simulation method seems to be able to solve as satisfactorily a 3D hydrostratigraphic context as complex as that of Simcoe County. The gradual deformation method allows for efficient calibration of the  $K$ -tensor components while preserving the non-linear correlations of the components and their covariance functions. To the author's knowledge, this is also the first time where GDM has been used in a multivariate context with constraints. Finally, this thesis contributes to the advancement of knowledge and techniques for stochastic characterization of inter- and intra-hydrostratigraphic heterogeneities. In a broader perspective, the thesis contributes original methods and tools adapted to better quantify the uncertainty associated with heterogeneities present in regional hydrogeological models.

## TABLE DES MATIÈRES

DÉDICACE . . . . .	iii
REMERCIEMENTS . . . . .	iv
RÉSUMÉ . . . . .	v
ABSTRACT . . . . .	ix
TABLE DES MATIÈRES . . . . .	xiii
LISTE DES TABLEAUX . . . . .	xvii
LISTE DES FIGURES . . . . .	xviii
LISTE DES SIGLES ET ABRÉVIATIONS . . . . .	xxiii
LISTE DES ANNEXES . . . . .	xxv
CHAPITRE 1 INTRODUCTION . . . . .	1
1.1 Mise en contexte . . . . .	1
1.2 Problématique générale . . . . .	2
1.3 Originalité du projet . . . . .	3
1.4 Hypothèses de recherche . . . . .	3
1.5 Objectifs . . . . .	4
1.5.1 Objectif général . . . . .	4
1.5.2 Les objectifs spécifiques . . . . .	4
1.6 Contenu de la thèse . . . . .	4
1.7 Organisation de la thèse . . . . .	6
CHAPITRE 2 REVUE DE LITTÉRATURE . . . . .	9
2.1 Simulation géostatistique de modèles hydrostratigraphiques . . . . .	9
2.1.1 Simulation plurigaussienne (PGS) . . . . .	11
2.1.2 Simulation multipoint (MPS) . . . . .	14
2.1.3 Prédiction catégorique de type Markovienne (MCP) . . . . .	17
2.2 Caractérisation stochastique de la conductivité hydraulique des modèles hydrogéologiques . . . . .	20

2.2.1	Estimation de la conductivité hydraulique . . . . .	21
2.2.2	Mise à l'échelle de la conductivité hydraulique en tenseur équivalent . . . . .	22
2.2.3	Régionalisation de la conductivité hydraulique . . . . .	25
2.2.4	Calage du tenseur de conductivité hydraulique . . . . .	27
2.3	Remarques sur la caractérisation hydrostratigraphique régionale et son incertitude sur la modélisation hydrogéologique . . . . .	31
CHAPITRE 3 CONTEXTE HYDROGÉOLOGIQUE DU COMTÉ DE SIMCOE SUD		
	EN ONTARIO . . . . .	33
3.1	Géologie . . . . .	33
3.2	Hydrostratigraphie . . . . .	35
3.3	Hydrogéologie . . . . .	36
3.3.1	Potentiel aquifère . . . . .	36
3.3.2	Données . . . . .	37
CHAPITRE 4 ARTICLE 1 : DIRECTIONAL HYDROSTRATIGRAPHIC UNITS SIMULATION USING MCP ALGORITHM . . . . .		
4.1	abstract . . . . .	45
4.2	Introduction . . . . .	46
4.3	MCP simulation methodology . . . . .	47
4.3.1	MCP definition . . . . .	48
4.3.2	MCP simulation . . . . .	48
4.3.3	Illustration of MCP . . . . .	51
4.3.4	MCP variability in the Gaussian case . . . . .	53
4.3.5	Sensitivity to neighborhood selection . . . . .	55
4.4	Case study - MCP simulation of a Basin glacial stratigraphy . . . . .	57
4.4.1	Hydrostratigraphic setting : Simcoe County area . . . . .	57
4.4.2	Simulation scenarios and parameters . . . . .	60
4.4.3	Statistics of the different scenarios . . . . .	75
4.4.4	Illustration of local proportions variability . . . . .	75
4.5	Discussion . . . . .	76
4.6	Conclusion . . . . .	79
CHAPITRE 5 ARTICLE 2 : STOCHASTIC CORRELATED HYDRAULIC CONDUCTIVITY TENSOR CALIBRATION USING GRADUAL DEFORMATION . . . . .		
5.1	Abstract . . . . .	80
5.2	Introduction . . . . .	81

5.3	Methodology . . . . .	82
5.3.1	Hydraulic block-conductivity tensor simulation . . . . .	82
5.3.2	Bivariate transform . . . . .	85
5.3.3	Gradual deformation method . . . . .	85
5.4	Study area . . . . .	87
5.4.1	Hydrostratigraphic model . . . . .	87
5.4.2	Hydraulic conductivity . . . . .	88
5.4.3	Groundwater flow model . . . . .	92
5.5	Model calibration . . . . .	93
5.5.1	The reference model . . . . .	94
5.5.2	Random effects . . . . .	94
5.5.3	Recharge effects . . . . .	97
5.5.4	Influence of number of head data . . . . .	99
5.6	Capture zones . . . . .	100
5.6.1	PEST calibration . . . . .	102
5.6.2	Well capture zone delineation and uncertainty . . . . .	104
5.7	Discussion . . . . .	105
5.8	Conclusion . . . . .	106
CHAPITRE 6	DISCUSSION GÉNÉRALE . . . . .	107
6.1	Caractérisation stochastique des hétérogénéités inter unités hydrostratigraphiques . . . . .	107
6.1.1	Hypothèse de l'indépendance conditionnelle . . . . .	107
6.1.2	Modèle déterministe comme image d'entraînement . . . . .	108
6.1.3	Comparaison de la méthode MCP avec la méthode MPS . . . . .	110
6.1.4	Autres remarques sur la méthode MCP . . . . .	112
6.2	Caractérisation stochastique des hétérogénéités intra unités hydrostratigraphiques ( $K$ ) . . . . .	113
6.2.1	Mise à l'échelle d'hétérogénéités locales ( $K$ ) . . . . .	114
6.2.2	Simulation du tenseur- $K$ avec modèles linéaires de corégionalisation . . . . .	114
6.2.3	Préservation des corrélations inter composants de tenseurs- $K$ . . . . .	114
6.2.4	Utilisation de la GDM pour caler les tenseurs- $K$ . . . . .	115
6.2.5	Comparaison du calage par GDM et PEST . . . . .	117
6.2.6	Autres remarques sur la méthodologie . . . . .	118
6.3	Limites des approches proposées . . . . .	118
CHAPITRE 7	CONCLUSION ET RECOMMANDATION . . . . .	120

RÉFÉRENCES . . . . .	124
ANNEXES . . . . .	151

## LISTE DES TABLEAUX

2.1	Résumé de méthodes de mise à l'échelle des paramètres . . . . .	24
2.2	Régionalisation des paramètres hydrauliques . . . . .	26
2.3	Synthèse des méthodes d'inversion indirecte . . . . .	29
3.1	Classification hydrostratigraphique du comté de Simcoe Sud . . . . .	36
3.2	Méthodes de prédictions de la conductivité hydraulique et critères d'ap- plication ( $C_u$ est le coefficient d'uniformité) . . . . .	39
4.1	Statistics C1 and C2 (100 realizations) . . . . .	53
4.2	Statistics of 100 realizations for each scenario. See Eq. 4.7 for definition of C1 and C2 ( <sup>1</sup> random sampling of TI) . . . . .	75
5.1	Upscaled HSU block-scale $\log(K_{xx})$ components (m/s) . . . . .	90
5.2	Modelled Gaussian variogram structures and linear model of coregio- nalisation values of $\ln(K)$ . The nugget effect $\delta_{zx}$ is symmetric to $\delta_{xz}$ and $\delta_{yx}$ is symmetric to the $\delta_{xy}$ component. Range parameters ( $a_x$ , $a_y$ and $a_z$ ) are in block units of 200x200x1 meters . . . . .	92
5.3	Initial $K$ values ( $\overline{\log(K_{xx0})}$ and $\overline{\log(K_{zz0})}$ ) and calibrated values ( $\overline{\log(K_{xxF})}$ and $\overline{\log(K_{zzF})}$ ) (m/s) from PEST inversion and from GDM inversion (reference model) . . . . .	104

## LISTE DES FIGURES

1.1	Organisation des principales étapes de la thèse . . . . .	7
2.1	Exemple de règle de lithotype en 2D et 3D ( $y_1$ , $y_2$ et $y_3$ sont des variables gaussiennes). . . . .	12
2.2	Impact de la variation spatiale des proportions des matériaux géologiques sur les lithotypes où chacun des drapeaux correspond à un contexte différent du domaine ( $y_1$ et $y_2$ sont des variables gaussiennes). . . . .	12
2.3	Effet de la structure des variogrammes gaussiens (haut) et exponentiels (bas) sur les champs gaussiens (gauche) et le domaine géologique correspondant (droite). . . . .	13
2.4	Image d'entraînement non stationnaire d'un delta fluvial inspirée d'Honarkhah [1] (gauche) ; b) Réalisation SNESIM sans tenir compte de la non-stationnarité (droite). . . . .	15
2.5	Exemple de fonctions de probabilité bivariable 3D calculées par transformée de Fourier [2] ; le $h = 0$ se trouve au point d'intersection des trois plans. . . . .	19
2.6	Modélisation hydrogéologique . . . . .	28
3.1	Localisation de la zone d'étude du programme de caractérisation géologique de l'OGS et du sous-bassin Innisfil Creek, comté de Simcoe Sud . . . . .	34
3.2	Modèle hydrostratigraphique déterministe du sous-bassin versant d'Innisfil Creek ; unités décrites au tableau 3.1 (modèle adapté de Bajc, communication personnelle, 29 février 2016). . . . .	38
3.3	Boxplot des proportions d'hydrofaciès (argile, silt et sable) et histogramme du $\ln(K)$ estimé à partir des granulométries pour chaque UHS. Sur l'histogramme, les polygones colorés (argile=bleu ; silt=cyan et sable=jaune) correspondent aux valeurs typiques de $\ln(K)$ pour chacun des 3 hydrofaciès [3]. Le titre en haut de la boxplot est l'UHS avec les proportions moyennes de chaque hydrofaciès. Les titres sur l'histogramme sont la moyenne et la variance de la distribution modélisée. . . . .	42
3.4	Comparaison des Boxplots de $\ln(K)$ estimé par analyses granulométriques et $\ln(K)$ provenant des essais de perméabilité (étoiles noires avec un cercle rouge). Note : les étoiles noires seules signifient une corrélation incertaine avec l'unité. . . . .	43

3.5	Comparaison des Boxplots de $\ln(K)$ estimé par test de capacité et $\ln(K)$ provenant des essais de perméabilité (étoiles noires avec un cercle rouge). Note : Les étoiles noires seules signifient une corrélation incertaine avec l'unité. . . . .	44
4.1	From left to right — top row : Synthetic TI, 4 NCS realizations, NCS dissimilarity — middle row : 1% TI as data, four CS1 realizations, CS1 dissimilarity — bottom row : 20% TI as data, four CS2 realizations, CS2 dissimilarity. Based on 100 realizations, dissimilarities expressed as %. . . . .	52
4.2	Boxplots of unit proportions (in %) for NCS (left), CS1 (middle) and CS2 (right); TI proportions as *. Based on 100 realizations of the synthetic model. . . . .	53
4.3	Experimental bivariate probability as a function of cell distance for MCP realizations (grey) and 95% confidence interval for multiGaussian simulation (blue). Facies ordered F1, F2 and F3 from left to right and top to bottom. . . . .	54
4.4	Top left C1 and C2 as a function of nmax using dmax infinite and ndata=10; Top right C1 and C2 as a function of dmax with nmax=10 and ndata=10; Bottom left C1 and C2 as a function of ndata with nmax=10 and dmax infinite; Bottom right C1 and C2 as a function of number of realizations with nmax=10 and dmax infinite. . . . .	56
4.5	Top - Location map of the deterministic TI case study in South Simcoe County; Bottom - 3D view of the deterministic TI (vertical exaggeration 15x). . . . .	59
4.6	Boxplot of HSU thickness (m) within the TI. . . . .	60
4.7	Data for the four conditional simulations, from top to bottom : — CS1 : 176 picks and 2 boreholes — CS2 : 44 pseudo-data (random sampling of TI) — CS3 : CS1 data+pseudo-data (10% of bedrock top and all surficial HSU) — CS4 : CS1 data + pseudo-data (0.3% of cells from TI). . . . .	61
4.8	NCS : non-conditional simulation scenario; Cross-section ( $y=3$ ) of two realizations (a and b); Dissimilarity c); Boxplots of HSU proportions, 100 realizations d). . . . .	63
4.9	Boxplot of unit thickness (m) with NCS (TI mean and maximum thicknesses indicated by dot and star respectively). . . . .	64



4.10	CS1a : scenario with HD (2 boreholes and 176 picks); Cross-section ( $y=3$ ) of two realizations (a and b); Dissimilarity c); Boxplots of HSU proportions, 100 realizations d). . . . .	66
4.11	CS1b : scenario with HD (2 boreholes and 176 double picks); Cross-section ( $y=3$ ) of two realizations (a and b); Dissimilarity c); Boxplots of HSU proportions, 100 realizations d). . . . .	67
4.12	CS1b : four realizations of the stratigraphically lowest HSU : bedrock, AFF1, ATE1, AFD4. . . . .	68
4.13	CS2 : 44 pseudo-data (random sampling of TI). Cross-section ( $y=3$ ) of two realizations (a and b); Dissimilarity c); Boxplots of HSU proportions, 100 realizations d). . . . .	70
4.14	CS3 : scenario with HD (as CS1) and pseudo-data (37 bedrock cells and 324 surficial cells). Cross-section ( $y=3$ ) of two realizations (a and b); Dissimilarity c); Boxplots of HSU proportions, 100 realizations d). . . . .	72
4.15	CS4 : scenario with HD (as CS1) and 143 cells obtained from random sampling of TI). Cross-section ( $y=3$ ) of two realizations (a and b); Dissimilarity c); Boxplots of HSU proportions, 100 realizations d). . . . .	74
4.16	CS1a : Local HSU proportions within 25x25 moving windows over central vertical section. First column : local mean in TI, second column : mean of proportions (over 100 realizations); third column : standard deviation of the local proportions across realizations. From top row to bottom row : three most abundant units, ATB2, ATD1 and ATE1. . . . .	76
5.1	$K$ upscaling workflow from $K$ local-scale realization to 3D block $K$ -tensors. . . . .	84
5.2	3D histogram examples of $\log(K)$ for the reference bivariate cumulative density function (left) and simulated bivariate cumulative density function from the LCM after applying the bivariate transformation using Eq. 5.1 (right). . . . .	86
5.3	a) South Simcoe County location in southern Ontario, Canada; b) Valley partly filled by glaciolacustrine sediments located in Innisfil creek watershed lowland; c) Cross-section of the valley with model location (black rectangle); d) Simplified 3D numerical model used in the study (4 km x 6 km x 60 m). . . . .	87

5.4	a) ATB2 geologic material proportions (top) and corresponding statistical distribution of $\log(K)$ range values. Colors represents material $\log(K)$ where clay is in blue, silt in cyan and sand in yellow ; b) Boxplot of point-scale $K$ values for each unit. . . . .	89
5.5	Scatter plot illustrating the relationship between $\log(K_{xx})$ and $\log(K_{zz})$ components for : a) AFB1, b) ATB1, c) AFB2 and d) ATC1. Blue hollow dots represent the reference values obtained from upscaling, whereas the red dots are examples of bivariate transformation of simulated values. . . . .	91
5.6	a) True model parameters : a) simulated hydraulic conductivity of $\log(K_{xx})$ component, b) simulated recharge field in mm/year, c) head field in meters and d) 59 head observations randomly selected from the 3D simulated head field which refer to as “true observations”. . . . .	93
5.7	<b>Initial condition :</b> a) recharge rate (mm/y), b) initial hydraulic conductivity ( $\log(K_{xx})$ ), c) simulated head field(m) ; <b>Calibrated condition :</b> d) calibrated recharge (mm/y), e) hydraulic conductivity ( $\log(K_{xx})$ ), f) simulated head field (m) ; g) Initial (top) and calibrated (bottom) histogram of the difference between true and simulated heads, h) scatter plot of simulated vs true head observations (blue=initial ; grey=intermediate ; red=final) . . . . .	95
5.8	$K$ -tensor variograms and cross-variograms of $\ln(K_{xx})$ and $\ln(K_{zz})$ for 3 principal directions (distances are in block units) : a) x-direction, b) y-direction, c) z-direction. For each direction, $\gamma_{xx}$ , $\gamma_{xz}$ and $\gamma_{zz}$ variograms in reading order. Variograms of calibrated fields (black), of the theoretical models (red), and of the realizations used in GDM (grey). . . . .	96
5.9	Evolution of the GDM with iterations for : a) OF, b) deformation coefficient $t$ , c) weight $\cos(t)$ , and d) OF difference (in $\log_{10}$ ). . . . .	97
5.10	Four final hydraulic conductivity fields in $\log(K_{xx})$ obtained with different seeds. . . . .	98
5.11	Boxplots of $K_{xx}$ (a) and $K_{zz}$ (b) components of the model tensors before ( $K_{xx0}$ and $K_{zz0}$ ) and after calibration for the four realizations ( $K_{xx1}$ to $K_{xx4}$ and $K_{zz1}$ to $K_{zz4}$ ). . . . .	99

5.12	Recharge effect on $\log(K_{xx})$ field obtained by GDM for 3 different scenarios : stochastic recharge (top row) ; uniform recharge (middle row) and a single fixed realization recharge (bottom row). Columns are : recharge rate in mm/y (left), calibrated $\log(K_{xx})$ field (middle) and the difference between the calibrated and the true $\log(K_{xx})$ field (right column). . . . .	100
5.13	Recharge scenario $K$ -tensor variograms and cross-variograms of $\ln(K_{xx})$ and $\ln(K_{zz})$ for 3 principal directions (distances are in block units) : a) x-direction, b) y-direction, c) z-direction. For each direction, $\gamma_{xx}$ , $\gamma_{xz}$ and $\gamma_{zz}$ variograms in reading order. Variograms of calibrated fields for the different scenarios : stochastic recharge (black), uniform recharge (blue), single realization recharge (green) and true recharge (cyan) . .	101
5.14	MAE (eq. 5.4) boxplot of 300 experimental variograms of calibrated models vs the true model variogram for each recharge scenario . . . .	102
5.15	RMSE and correlation variation with the number of observation wells used into the calibration . . . . .	102
5.16	PEST inversion results : observed hydraulic head vs initial (blue) and calibrated (red) simulated head on left ; hydraulic head difference for the initial (right top) and the calibrated (right bottom) models . . . .	103
5.17	Capture probability contours for $P = 0.5$ (light grey) and $P = 0.1$ (dark grey) at 100 years simulation : plan view on left and cross-section view (y=3000) on right (well casings are black and screens are grey) ; capture probability contours with PEST (red) and true model (blue) . . . .	105
6.1	Image d'entraînement et quatre réalisations MPS-motifs montrant des inversions d'UHS. . . . .	111
6.2	Réalisations MPS-DS montrant des artefacts dans la périphérie et la quasi absence de variabilité dans la partie centrale. . . . .	112
6.3	Comparaison de valeurs du $\log(K_{xx})$ et $\log(K_{zz})$ de l'unité AFB1 obtenues par la transformation proposée avec celles transformées par PPMT selon : a) un diagramme binaire, b) leurs histogrammes respectifs, c) et leurs distributions bivariées estimées par densité de noyau. . . .	116
A.1	Domaine stationnaire (gauche) et non-stationnaire (droite). . . . .	152

## LISTE DES SIGLES ET ABRÉVIATIONS

AF	Aquifères
AFB1	Dépôts grossiers d'eau peu profonde aquifère
AFB2	Dépôts fluvioglaciaires équivalents à ORM aquifère
AFD1	Formation de Thorncliffe aquifère partie 1
AFD2	Formation de Thorncliffe aquifère partie 1
AFD3	Formation de Thorncliffe aquifère partie 1
AFD4	Formation de Thorncliffe aquifère partie 4
AFF1	Sédiments glaciaires anciens aquifère
ASM	Méthode de segmentation automatique
AT	Aquitards
ATB2	Sédiments fins glaciolacustres d'eau profonde aquitard
ATC1	Till de Newmarket aquitard Diamicton compact
ATD1	Formation de Thorncliffe aquitard partie 1
ATD2	Formation de Thorncliffe aquitard partie 2
ATD3	Formation de Thorncliffe aquitard partie 3
ATE1	Sédiments glaciaires anciens aquitard Diamicton
ATG1	Sédiments glaciaires anciens aquitard Diamicton
ATI	Image d'entraînement Auxiliaire
biCDF	Fonction de densité bivariable cumulative
BME	Entropie maximale bayésienne
C1	Divergence de Kullback-Leibler
C2	Étendue moyenne des proportions d'UHS
CAMC	Autorités de conservation de la moraine d'Oak Ridges
CPU	Unité centrale de traitement
CS	Simulation conditionnelle
dmax	Distance du rayon de recherche
EPDS	Simulation par bases de données de motifs enrichies
FFTMA	Transformée rapide de Fourier
FILTERSIM	Algorithme géostatistique multipoint basé sur des motifs
$\gamma_{xx}$	variogram de la variable $x$
$\gamma_{xz}$	cross-variogram des variables $xy$
$\gamma_{zz}$	variogram de la variable $z$
GDM	Méthode des déformations graduelles

GTS	Simulation de gaussienne tronquée
HD	Données principales
IMPALA	Algorithme parallélisé basé sur une approche de type liste
$K$	Conductivité hydraulique
$K_{eq}$	Conductivité hydraulique équivalente
$K_{horizontale}$	Composante horizontale du tenseur- $K$
$K_{verticale}$	Composante verticale du tenseur- $K$
$K_{xx}$	Composante $xx$ du tenseur- $K$
$K_{yy}$	Composante $yy$ du tenseur- $K$
$K_{zz}$	Composante $zz$ du tenseur- $K$
$K_a$	Moyenne arithmétique de la conductivité hydraulique
$K_h$	Moyenne harmonique de la conductivité hydraulique
MAE	Erreur moyenne absolue
MCP	Prédiction Markovienne catégorique
MPS	Simulation multipoint
NCS	Simulation non conditionnelle
ndata	Nombre de données
nmax	Nombre maximal de voisins
nmin	Nombre minimal de voisins
OF	Fonction objective
OGS	Commission géologique de l'Ontario
ORM	Moraine d'Oak Ridges
PD	Pseudo-données
PEST	Parameter ESTimation algorithm
PGS	Simulation plurigaussienne
PPMT	Transformation multivariée par recherche de projection
RMSE	Erreur quadratique moyenne
tenseur- $K$	Tenseur de la conductivité hydraulique
SIMPAT	Simulation par motifs
SSM	Méthode de similitude spatiale
T-PROG	Simulation de la probabilité de transition
TBS	Simulation par bandes tournantes
TI	Image d'entraînement
UHS	Unité hydrostratigraphique

## LISTE DES ANNEXES

Annexe A	Notions de stationnarité . . . . .	151
Annexe B	SIMULATIONS GÉOSTATISTIQUES DU TENSEUR COMPLET DE CONDUCTIVITÉ HYDRAULIQUE DE BLOC 3D EN TENANT COMPTE DE LA VARIABILITÉ INTERNE À L'ÉCHELLE LOCALE . . . . .	153
Annexe C	Pseudocodes des principaux algorithmes . . . . .	154

## CHAPITRE 1 INTRODUCTION

### 1.1 Mise en contexte

Les experts en hydrogéologie doivent de plus en plus être en mesure d'évaluer le risque environnemental associé au développement d'activités anthropiques notamment lorsqu'il s'agit de l'élaboration d'études d'impact sur l'environnement des grands projets industriels. Plusieurs pays industrialisés ont une réglementation environnementale qui impose des standards minimums d'évaluation du risque depuis plusieurs années [4]. Malgré de telles dispositions, l'évaluation du risque est généralement effectuée de manière qualitative notamment à cause de la difficulté à quantifier l'incertitude (composante principale du risque). Une explication possible réside dans la complexité à quantifier les paramètres d'entrée des modèles hydrogéologiques et les incertitudes associées [5]. En hydrogéologie, l'élaboration des modèles numériques est généralement de type déterministe basé sur l'opinion d'experts, c'est-à-dire le meilleur modèle possible selon les connaissances. Cependant, ce type de modèle ne permet pas de quantifier l'incertitude [6, 7].

Les modèles physiques d'écoulement et de transport sont des outils numériques qui s'intègrent bien dans une approche stochastique pour quantifier l'incertitude [8]. L'hydrogéologie stochastique consiste à combiner les concepts et les outils physiques et probabilistes [7]. En réalité, cette pratique est plutôt rare chez les professionnels en hydrogéologie [9, 10]. Pour arriver plus facilement à quantifier l'incertitude, les professionnels ont besoin de méthodologies simples et efficaces qui répondent à leurs besoins. Des méthodes qui permettent l'élaboration de modèles équivalents, c'est-à-dire de modèles qui respectent les mêmes observations et qui peuvent être utilisés pour différents scénarios d'écoulement ou de transport. Un ensemble de modèles permet ainsi de représenter l'incertitude. Évidemment, cette incertitude repose sur la qualité du modèle et des simplifications supposées.

L'une des principales sources d'incertitude des modèles hydrogéologiques porte sur les hétérogénéités hydrostratigraphiques. Ces hétérogénéités jouent un rôle majeur sur l'écoulement de l'eau souterraine et le transport de composés chimiques [11]. L'incertitude se situe à deux niveaux, soit celle associée aux unités hydrostratigraphiques elles-mêmes, leurs épaisseurs et leurs distributions spatiales, ainsi que celle associée à leurs propriétés hydrauliques telle que la conductivité hydraulique ( $K$ ). Même s'il ne s'agit pas de l'unique source d'incertitude, les hétérogénéités inter et intra unités hydrostratigraphiques ont potentiellement des effets non négligeables sur les résultats de modélisation. Pourtant la caractérisation hydrogéologique inclut très rarement l'évaluation de l'effet de l'incertitude des hétérogénéités hydrostratigra-

phiques. Cette situation s’explique en partie par la difficulté à reproduire ces caractéristiques physiques en modèles équivalents qui soient satisfaisants.

Les méthodes de simulation géostatistique permettent de produire un ensemble de modèles équivalents appelés réalisations. Ces dernières caractérisent la variabilité spatiale et statistique des caractéristiques physiques des modèles, soient les hétérogénéités inter et intra unités (hydrostratigraphie et  $K$ ). Bien que les méthodes récentes de simulation permettent une caractérisation plus réaliste des modèles hydrogéologiques, plusieurs difficultés demeurent.

## 1.2 Problématique générale

Les récents développements des algorithmes de simulation géostatistique tels que la méthode plurigaussienne (PGS), multipoint (MPS) et la prédiction catégorique de type Markovienne (MCP) ont grandement amélioré l’aspect des réalisations simulées notamment grâce à l’intégration des connaissances géologiques. Ces dernières sont prises en compte dans la simulation par l’intégration de règles topologiques complexes ou des images d’entraînement qui reproduisent ainsi des structures complexes connectées ou non. Malgré que les algorithmes soient de plus en plus performants, une lacune souvent mentionnée demeure le manque de réalisme géologique des réalisations [12–14]. Chaque algorithme a ses propres particularités et le choix de ce dernier doit se faire en fonction d’objectifs bien précis. Par exemple, pour simuler un contexte géologique des chenaux, la méthode MPS est reconnue comme étant très performante. Pour d’autres contextes, le choix de la méthode peut être beaucoup moins évident. En effet, il est difficile de prévoir la performance des différentes méthodes dans un contexte hydrogéologique régional où l’hydrostratigraphie est généralement représentée par un empilement stratigraphique directionnel composé de nombreuses unités hydrostratigraphiques (UHS). L’absence de méthode de simulation efficace pour ce type de contexte est un frein à la modélisation hydrogéologique stochastique régionale. Cela empêche notamment de quantifier l’incertitude inter unités.

La paramétrisation stochastique régionale de la conductivité hydraulique de ces modèles est également un enjeu pour quantifier adéquatement l’incertitude intra unités. Cette étape demeure délicate et difficile à réaliser notamment lorsqu’il faut tenir compte du changement de support entre les données et le maillage (éléments finis) d’un modèle numérique d’écoulement. De plus, plusieurs défis s’ajoutent à cette problématique, notamment :

- caractériser la structure des hétérogénéités à petite échelle (données) ;
- estimer le tenseur- $K$  équivalent représentatif de ces hétérogénéités ;
- définir des tenseurs- $K$  équivalents sur un très grand nombre d’éléments ;
- caler les champs de tenseurs- $K$  tout en préservant la structure spatiale et les corréla-



tions des composantes des tenseurs- $K$ .

Actuellement, aucune méthodologie complète et efficace ne semble permettre d'effectuer la modélisation hydrogéologique régionale dans un cadre stochastique tout en quantifiant l'incertitude des hétérogénéités inter et intra unités. Une telle méthodologie ajoute une base solide pour mieux quantifier le risque et ainsi aider les gestionnaires à prendre des décisions plus éclairées.

### 1.3 Originalité du projet

La plupart des études hydrogéologiques régionales reposent sur des modèles déterministes qui ne permettent pas de quantifier leur incertitude contrairement aux modèles stochastiques. Une part importante de cette incertitude provient des caractéristiques physiques des modèles (hydrostratigraphie et  $K$ ). Ce projet porte sur le développement d'une méthodologie complète de caractérisation stochastique des hétérogénéités hydrostratigraphiques (inter et intra unités) appliquée à la modélisation hydrogéologique régionale. Le projet propose, pour une première fois, une méthodologie de simulation géostatistique adaptée au contexte régional de sédimentation directionnelle des UHS en 3D. Cette méthodologie, pratique et efficace, permet de produire des modèles équivalents pouvant être utilisés avec un simulateur d'écoulement. Les modèles permettent de quantifier l'incertitude inter UHS. Le projet propose également une méthodologie de caractérisation de l'incertitude associée aux hétérogénéités intra UHS. Cette méthodologie permet aussi de caler efficacement des tenseurs- $K$  stochastiques à partir de  $K$  scalaires simulées à l'échelle locale. Peu ou pas d'études hydrogéologiques régionales offrent une démarche aussi complète de caractérisation régionale stochastique du tenseur- $K$  qui tient compte de la corrélation spatiale à l'échelle de l'échantillon. La caractérisation intra UHS est complémentaire à la caractérisation inter UHS. Globalement, cette thèse contribue à l'avancement des connaissances et des techniques de la caractérisation stochastique des hétérogénéités inter et intra hydrostratigraphiques. Dans une plus grande perspective, la thèse contribue à fournir des outils adaptés pour mieux quantifier l'incertitude associée aux hétérogénéités présente dans les modèles hydrogéologiques régionaux. Ces outils peuvent notamment contribuer à mieux évaluer le risque environnemental des grands projets.

### 1.4 Hypothèses de recherche

Les hypothèses de recherche de la thèse sont :

- les méthodes de simulation géostatistique basées sur une approche gaussienne ne permettent pas la simulation directionnelle des UHS ;

- les méthodes de simulation de type bayésienne qui utilisent la probabilité de transition permettent la simulation directionnelle de UHS ;
- la mise à l'échelle d'un nombre limité de tenseurs- $K$  permet de définir la covariance spatiale 3D de chacune des composantes pour une régionalisation rapide par simulation géostatistique ;
- la méthode des déformations graduelles s'applique à un contexte multivariable contraint permettant ainsi le calage du tenseur- $K$  de différentes UHS tout en préservant la structure de covariance spatiale, la contrainte d'inégalité  $K_{verticale} \leq K_{horizontale}$  ainsi que les corrélations inter composantes des tenseurs.

## 1.5 Objectifs

### 1.5.1 Objectif général

L'objectif général de la thèse est de développer une démarche complète de caractérisation stochastique de l'incertitude des hétérogénéités inter et intra unités hydrostratigraphiques appliquée à la modélisation hydrogéologique régionale.

### 1.5.2 Les objectifs spécifiques

L'objectif général est divisé en deux objectifs spécifiques principaux :

- Développer une méthodologie de simulation géostatistique des UHS afin de fournir des modèles hydrostratigraphiques équivalents d'aspect réaliste pour un contexte de sédimentation directionnelle ;
- Développer une méthodologie de caractérisation stochastique, de mise à l'échelle et de calage des tenseurs- $K$  équivalents à partir de la  $K$  scalaire tout en préservant la structure spatiale des tenseurs, la contrainte d'inégalité  $K_{verticale} \leq K_{horizontale}$  ainsi que les corrélations inter composantes.

## 1.6 Contenu de la thèse

La thèse comprend sept chapitres et trois annexes. Deux chapitres sont tirés d'articles scientifiques dont un est publié et l'autre est sous révision des pairs.

Le chapitre 2 fait la revue de la littérature des deux thèmes principaux de la thèse soit :

- une méthodologie de simulation géostatistique pouvant potentiellement permettre de simuler des modèles hydrostratigraphiques équivalents avec ordination des unités ;

- une méthodologie de caractérisation et de calage des tenseurs- $K$  simulés à partir de  $K$  scalaire.

Le chapitre se termine par quelques remarques générales sur la caractérisation hydrostratigraphique régionale et son incertitude sur la modélisation hydrogéologique.

Le chapitre 3 fait la synthèse de la géologie, de l’hydrostratigraphie et de l’hydrogéologie de sous-bassin versant d’Innisfil Creek ainsi que de la compilation et du traitement des données. Toutes les méthodes développées dans la thèse sont testées sur le contexte hydrostratigraphique de ce bassin.

Le chapitre 4 porte sur le premier objectif spécifique de la thèse. Ce chapitre a été publié dans la revue scientifique *Stochastic Environmental Research and Risk Assessment* (2017). Il présente une méthodologie de simulation directionnelle d’UHS à partir de l’algorithme MCP (méthode de prédiction catégorique de type Markovienne ou MCP) [15]. Ces travaux ont aussi été présentés lors de conférences [16, 17]. En résumé, les résultats ont montré que la méthode MCP permet de simuler des unités hydrostratigraphiques contraintes par un modèle déterministe présentant un ordre stratigraphique directionnel. Le succès de la méthode passe principalement par la propriété de forcer des probabilités zéro. Cette propriété fournit une contrainte efficace pour la simulation directionnelle. L’exemple du cas réel tiré du sous-bassin d’Innisfil Creek montre la capacité de la méthode MCP à simuler des unités ordonnées verticalement d’un système hydrostratigraphique régional 3D contenant 11 UHS de proportions variées. La méthode fournit des modèles équiprobables d’aspects réalistes. Les résultats montrent que la méthode MCP est non biaisée pour les proportions d’UHS dans le cas non conditionnel et s’avère robuste dans les cas conditionnels. La méthode s’avère donc très pratique pour générer des modèles équivalents dans un contexte hydrostratigraphique 3D complexe ainsi que pour quantifier l’incertitude des unités.

Le chapitre 5 présente un article scientifique soumis en décembre 2019 au *Journal of Hydrology*. Ce chapitre porte sur le deuxième objectif spécifique de la thèse. Il présente le développement d’une méthodologie de caractérisation, de mise à l’échelle et de calage des composantes du tenseur de conductivité hydraulique [18]. Les résultats montrent que la nouvelle méthode proposée permet de calculer et de modéliser, à partir d’un nombre restreint de blocs mis à l’échelle, la covariance et les corrélations entre les composantes du tenseur- $K$ . De plus les résultats montrent que le calage des tenseurs- $K$  par la méthode des déformations graduelles permet de préserver la covariance spatiale et croisée de UHS. Il a été démontré que la méthode préserve les relations non linéaires entre les composantes du tenseur- $K$  par l’intégration d’une transformation bivariée. Ainsi, l’ensemble des réalisations calibrées permet de quantifier l’incertitude associée aux tenseurs- $K$  des modèles d’écoulement. Des exemples

d'application ont aussi été présentés dans un article de conférence [19, 20]. Ces exemples portent sur la simulation du transport d'âge moyen et la définition d'aires de captage de cours d'eau pour différentes réalisations des tenseurs- $K$  (voir annexe B).

La thèse se termine par une réflexion sur l'ensemble des résultats (chapitre 6) et une conclusion (chapitre 7) fait la synthèse des points forts de la thèse et propose quelques perspectives de développement futur. Noter que le lecteur trouvera les pseudocodes des principaux algorithmes des chapitres 4 et 5 à l'annexe C.

## 1.7 Organisation de la thèse

L'organisation de la thèse s'articule autour trois étapes principales (figure 1.1). La première étape porte sur la compilation et la préparation des données, la deuxième sur la simulation de modèles hydrostratigraphiques directionnels et la troisième sur la caractérisation, la mise à l'échelle et le calage des tenseurs- $K$ .

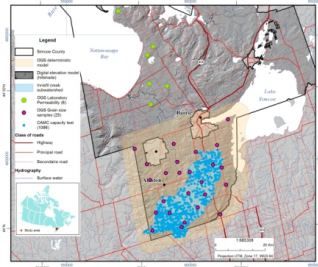
Ces étapes ont été définies à partir de la problématique, des hypothèses et des objectifs de recherche décrits précédemment. Elles ont été précisées à partir des lacunes identifiées lors de la revue de la littérature. Le reste de la section décrit chacune de ces étapes et résume succinctement la méthodologie appliquée.

La première étape a consisté à compiler et traiter les données nécessaires pour les développements méthodologiques et les différents tests sur le sous-bassin Innisfil Creek. Cette étape a permis d'acquérir les connaissances de base sur l'hydrostratigraphie et l'hydrogéologie du sous-bassin et de préparer un modèle numérique de l'hydrostratigraphie. Ce modèle déterministe a servi d'image d'entraînement pour la méthode MCP. Cette étape a aussi permis de compiler une base de données de conductivités hydrauliques provenant de différentes sources. Les données de conductivités hydrauliques, estimées à partir d'échantillons et des essais dans conditions de laboratoire ou de terrain ont servi à définir la distribution statistique de  $K$  de chacune des UHS ainsi que la structure spatiale à l'échelle locale.

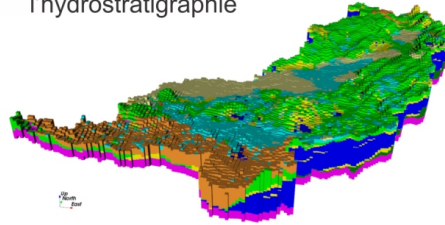
La deuxième étape a porté sur le développement de la méthodologie de simulation directionnelle des UHS par de la méthode MCP. Cette étape a permis de vérifier l'hypothèse selon laquelle la méthode MCP peut générer des modèles équivalents dans un contexte hydrostratigraphique 3D complexe dont la séquence des unités est ordonnée verticalement. La méthode MCP a été revue en détail et un algorithme a été développé pour effectuer des simulations 3D. La méthodologie proposée inclut l'estimation des probabilités bivariées (transiogramme) par transformée de Fourier à partir du modèle déterministe de l'étape 1. L'algorithme a été testé sur un cas synthétique à partir duquel la sensibilité de la méthode

## 1. Compilation et préparation des données

Compilation de données



Modèle déterministe de l'hydrostratigraphie



AFB1  
AFB2  
AFB3  
AFB4  
AFD1  
AFD2  
AFD3  
AFD4  
ATE1  
ATE2  
ATE3  
ATE4  
ATE5  
ATE6  
ATE7  
ATE8  
ATE9  
ATE10  
ATE11  
ATE12  
ATE13  
ATE14  
ATE15  
ATE16  
ATE17  
ATE18  
ATE19  
ATE20  
ATE21  
ATE22  
ATE23  
ATE24  
ATE25  
ATE26  
ATE27  
ATE28  
ATE29  
ATE30  
ATE31  
ATE32  
ATE33  
ATE34  
ATE35  
ATE36  
ATE37  
ATE38  
ATE39  
ATE40  
ATE41  
ATE42  
ATE43  
ATE44  
ATE45  
ATE46  
ATE47  
ATE48  
ATE49  
ATE50  
ATE51  
ATE52  
ATE53  
ATE54  
ATE55  
ATE56  
ATE57  
ATE58  
ATE59  
ATE60  
ATE61  
ATE62  
ATE63  
ATE64  
ATE65  
ATE66  
ATE67  
ATE68  
ATE69  
ATE70  
ATE71  
ATE72  
ATE73  
ATE74  
ATE75  
ATE76  
ATE77  
ATE78  
ATE79  
ATE80  
ATE81  
ATE82  
ATE83  
ATE84  
ATE85  
ATE86  
ATE87  
ATE88  
ATE89  
ATE90  
ATE91  
ATE92  
ATE93  
ATE94  
ATE95  
ATE96  
ATE97  
ATE98  
ATE99  
ATE100

Base de données interprétées



## 2 Développement de la méthodologie de simulation directionnelle d'UHS par MCP

Développement de l'algorithme MCP pour la simulation directionnelle 3D

$$p_{i_0|i_1, \dots, i_n}^* = \frac{p_{i_0} \prod_{k=1}^n p_{i_k|i_0}(h_{0k})}{\sum_{i_0=1}^I p_{i_0} \prod_{k=1}^n p_{i_k|i_0}(h_{0k})}$$

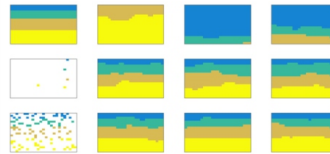
**Propriétés de forcer 0/1**

Interdit:  $p_{i_k|i_0}(h_{0k}) = 0$

Obligatoire:  $p_{i_k|i_0}(h_{0k}) = 1$

Méthodologie testée sur:

Cas synthétique

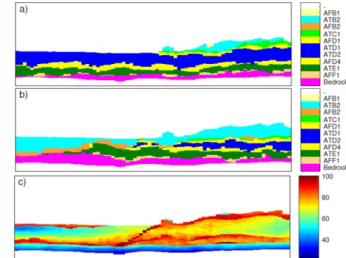


Critère de validation

$$C1 = \sum_{i=1}^I (pTI_i \times \log(pTI_i / pMCP_i))$$

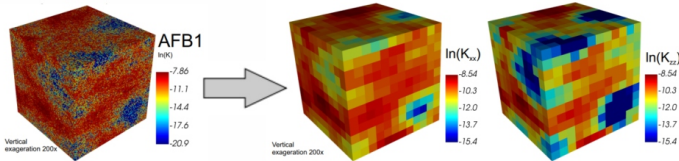
$$C2 = \frac{1}{I} \sum_{i=1}^I (q_{95,i} - q_{5,i})$$

Cas réel du sous-bassin d'Innisfil Creek

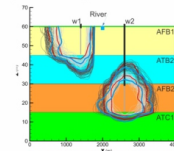


## 3. Développement de la méthodologie de caractérisation et de calage stochastique des composantes du tenseur-K

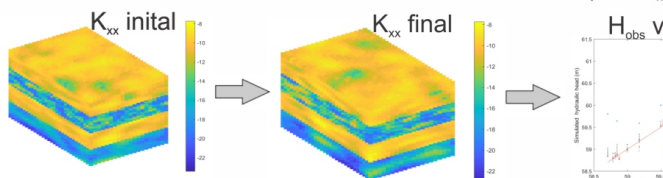
Mise à l'échelle de K au tenseur-K



Application: aire d'alimentation stochastique de puits de pompage



Calibration du tenseur-K (ex.:  $K_{xx}$ )



Structure de covariance

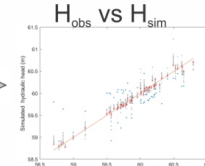
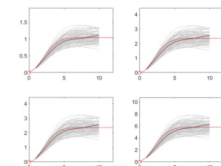


Figure 1.1 Organisation des principales étapes de la thèse

a été évaluée et des critères de qualité de simulation ont été proposés. La méthode a ensuite été testée en 3D pour le contexte hydrostratigraphique du sous-bassin d’Innisfil Creek pour plusieurs scénarios de simulation conditionnelle et non conditionnelle. La méthode a aussi été testée pour montrer sa capacité à quantifier la variabilité locale de la solution et son incertitude. Le succès de la méthode MCP repose sur sa propriété de forçage de la probabilité 0 et 1 fournissant ainsi la contrainte recherchée pour la simulation directionnelle. La mise au point de l’algorithme MCP a permis d’illustrer la capacité de la méthode à simuler des unités ordonnées verticalement, c’est-à-dire de reproduire une caractéristique importante des dépôts sédimentaires soit l’ordre directionnel des unités hydrostratigraphiques (UHS). Des exemples ont montré que la méthode est robuste au nombre de données conditionnelles, leur abondance respective ainsi que la non-stationnarité des unités.

La troisième étape porte sur le développement de la méthodologie de caractérisation, de mise à l’échelle et de calage stochastique des composantes du tenseur- $K$ . La méthodologie a permis de transformer des valeurs de  $K$  scalaire (quasi point) vers une échelle adéquate (élément) pour être utilisées avec un simulateur d’écoulement. La mise à l’échelle prend en compte l’effet des hétérogénéités à l’intérieur de l’élément. Cette opération nécessite l’utilisation d’un simulateur d’écoulement afin de tenir compte de la grande variabilité statistique et spatiale de la  $K$  scalaire et ainsi définir un tenseur- $K$  équivalent. Pour pallier au temps de calcul imposant que nécessite la mise à l’échelle à partir d’un modèle d’écoulement, la démarche combine la mise à l’échelle [21] et la simulation géostatistique par bande tournante. La démarche tient compte des corrélations inter composantes et les contraintes d’inégalité observées (ex. :  $K_{horizontale} \geq K_{verticale}$ ). Enfin, la méthode des déformations graduelles (GDM) permet de caler chacune des composantes stochastiques des tenseurs- $K$  tout en préservant la covariance spatiale, la covariance croisée et les relations non linéaires entre les composantes. Le calage préserve aussi la relation bivariable entre la  $K_{horizontal}$  et  $K_{vertical}$  des éléments. À ma connaissance, il s’agit de la première approche utilisant de GDM pour un contexte multivariable contraint. Les exemples d’application montrent qu’un ensemble de réalisations permet de quantifier l’incertitude des tenseurs- $K$ , mais aussi l’effet de cette incertitude sur l’écoulement.

## CHAPITRE 2 REVUE DE LITTÉRATURE

La revue de la littérature porte sur la caractérisation stochastique des hétérogénéités inter et intra UHS des modèles hydrogéologiques régionaux. Tel que décrit en introduction, l'objectif est de trouver les meilleures solutions pour quantifier l'incertitude associée à ces hétérogénéités ainsi qu'identifier les lacunes scientifiques à combler. La caractérisation des hétérogénéités inter unité est étudiée à travers les méthodes de simulation géostatistique de modèles (section 2.1) tandis que la caractérisation des hétérogénéités intra unité l'est à travers la régionalisation du tenseur de conductivité hydraulique (section 2.2). La revue de la littérature se termine par quelques remarques générales sur la caractérisation et l'effet de l'incertitude sur la modélisation hydrogéologique (section 2.3).

### 2.1 Simulation géostatistique de modèles hydrostratigraphiques

Le modèle hydrostratigraphique constitue la base du modèle hydrogéologique. Souvent, la quantité limitée de forages de qualité dont disposent les hydrogéologues pour étudier un système hydrostratigraphique complexe ne suffit pas pour bien modéliser les caractéristiques critiques des UHS, telles leurs proportions, leurs distributions spatiales et la connectivité inter unités. Ces caractéristiques sont essentielles puisqu'elles contrôlent les réponses issues de simulation de l'écoulement et du transport de l'eau souterraine [11, 22]. Cette situation laisse place à un niveau d'incertitude variable en fonction de la complexité du contexte, de la qualité et de la quantité d'information disponible. Malgré la présence connue de cette incertitude, la plupart des modèles hydrogéologiques sont fondés sur un modèle déterministe.

Peu utilisé en hydrogéologie, les méthodes de simulation géostatistique permettent de proposer différents modèles hydrostratigraphiques équivalents appelés réalisations [23]. Une méthode de simulation efficace devrait permettre de générer des réalisations qui respectent les données, les proportions globales des UHS et leurs structures spatiales. L'ensemble des réalisations permet de quantifier l'incertitude du modèle. De plus, chacune des réalisations peut-être utilisée comme une entrée à un simulateur d'écoulement à partir de laquelle une réponse à un stimulus spécifique peut être calculée. Ainsi, la distribution des réponses de  $n$  modèles décrit l'incertitude associée à la réponse réelle du système pour un même stimulus physique [24]. Pour être utile, l'approche exige que chaque réalisation soit compatible avec les caractéristiques naturelles pertinentes que l'on s'attend à trouver dans la séquence hydrostratigraphique. Par conséquent, pour les modèles complexes, ces derniers doivent inclure les connaissances géologiques disponibles tout en préservant la variabilité correspondante

aux connaissances incomplètes. Pour certains contextes, ces contraintes peuvent être difficiles à atteindre. Le choix de la méthode doit donc être orienté en fonction des contraintes du contexte à simuler. Il existe plusieurs méthodes géostatistiques qui permettent de simuler des UHS, les plus connues sont les méthodes de simulation suivantes : indicatrices [25–27], gaussienne tronquée [28], plurigaussienne ou PGS [29, 30] et multipoint ou MPS [31–34]. Au cours des dernières années, les développements ont plutôt porté sur la méthode MPS [35–41] et PGS [42–45]. Toutefois, une autre méthode moins connue, mais fort prometteuse et qui mérite de lui porter attention s’ajoute à cette sélection, il s’agit de la méthode de prédiction catégorique de type Markovienne ou MCP [46].

Les sous-sections qui suivent portent uniquement sur les méthodes qui semblent les plus appropriées pour la simulation de contexte hydrostratigraphique complexe, soit la méthode PGS (section 2.1.1), MPS (section 2.1.2) et MCP (section 2.1.3). Cette sélection est basée sur trois considérations principales soient, la possibilité de : i) tenir compte des connaissances géologiques, ii) fonctionner dans un domaine hétérogène statistiquement et iii) être en mesure de simuler une stratigraphie directionnelle prédéfinie telle que généralement observée dans les modèles régionaux (section 3.2).

L’hétérogénéité statistique des domaines demande l’utilisation de modèles non stationnaires globalement ou du moins robustes à la non-stationnarité. Bien que la stationnarité soit une propriété des modèles et non des domaines étudiés, le terme stationnarité est ici utilisé pour décrire également l’homogénéité statistique du domaine. Le traitement de la non-stationnarité est une considération importante afin de produire des modèles hydrostratigraphiques d’aspect plus réaliste et pour potentiellement mieux définir l’incertitude liée aux hétérogénéités géologiques. Un domaine stationnaire est rarement rencontré dans les environnements géologiques, car la non-stationnarité est presque une condition intrinsèque à la géologie [47]. Il est donc important de pouvoir intégrer cette notion de manière adéquate et efficace lors des simulations géostatistiques des modèles hydrostratigraphiques. L’annexe A présente quelques exemples en lien avec la stationnarité. Outre la non-stationnarité, la stratigraphie directionnelle verticale des unités présentes dans les environnements sédimentaires régionaux ajoute un niveau de complexité supplémentaire. Par exemple, dans les stratigraphies glaciaires non déformées, ces dernières sont représentées par la loi de superposition [48]. Cette simple contrainte semble difficile à intégrer dans les méthodes de simulation géostatistiques de variables catégoriques traditionnelles. Le fondement gaussien des méthodes géostatistiques semble être la principale cause. Ainsi, produire des réalisations de modèles équiprobables sous contrainte d’orientation directionnelle des unités représente un défi et certainement une lacune importante à combler. Il est donc proposé dans cette section d’examiner plus en détail les méthodes de simulation : PGS, MPS et MCP en portant une attention particulière au traitement de la



non-stationnarité et de la problématique de simulation directionnelle des unités.

### 2.1.1 Simulation plurigaussienne (PGS)

La méthode de simulation de type plurigaussienne est une généralisation de la méthode de simulation de gaussienne tronquée (GTS). Les fondements théoriques de la méthode ont été établis respectivement par Matheron et al. [28] et par Le Loc'h et Galli [49]. Comparativement à la méthode GTS, la méthode PGS permet une grande flexibilité dans la reproduction complexe des unités et constitue l'une des méthodes les plus adaptées pour la simulation d'unités [30, 50–52]. Elle permet d'inclure des contraintes topologiques (relation de contact) tout en reproduisant les proportions des unités et la corrélation spatiale des structures géologiques [53]. À titre d'exemple, elle permet d'exclure le contact entre deux unités. La méthode de PGS a été appliquée avec succès au cours des dernières années notamment dans l'industrie des mines et du pétrole [50]. On retrouve plusieurs exemples dans la littérature d'utilisation de la méthode PSG, par exemple : définir la géométrie de différents types de gisements [54–57] ; effectuer la caractérisation hydrogéologique des hétérogénéités des aquifères [58] ; modéliser le transport réactif dans des aquifères contaminés [59] ; évaluer l'impact des hétérogénéités sur les charges hydrauliques [60] ; prédire le contenu en glace du glacier Bonnard en Suisse [61].

Le concept de la PGS consiste à produire deux ou plusieurs champs gaussiens à partir de technique de simulation multigaussienne et d'appliquer à ces champs un diagramme de troncation (lithotype) représentant les différentes unités (Figure 2.1). Chaque élément du lithotype représente une unité, la position des différentes unités représente leur relation relative et la taille de l'élément unité indique sa proportion relative. L'élaboration du lithotype est une étape importante qui repose sur les données disponibles (calcul de probabilité de transition) et le modèle conceptuel de secteur à l'étude. L'estimation des règles de lithotype à partir des données ou d'un modèle conceptuel n'est pas intuitive pour un géologue [46, 62].

Les données géologiques (forages, affleurements, cartes, etc.) permettent de calculer les proportions de chacune des unités. Évidemment, ces proportions ne sont pas nécessairement stationnaires dans l'espace. Pour tenir compte de cette non-stationnarité, l'approche classique est de générer des courbes de proportion horizontale et verticale des unités. Les courbes de proportion sont définies à partir d'un niveau de référence et sont comptées sur la base de la présence de l'unité pondérée par son épaisseur [28, 63, 64]. Une fois que l'ordre des unités est établi, il existe une relation directe entre les proportions et les seuils des variables gaussiennes. Ainsi les seuils formant le lithotype sont ajustés localement aux proportions ciblées tout en préservant la position relative des unités (ex. : Figure 2.2). Prendre note que le contrôle des proportions demande un calcul précis des seuils [50].

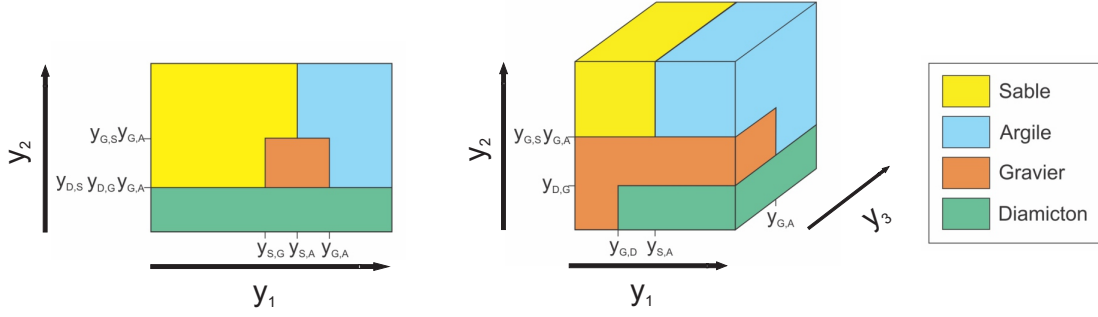


Figure 2.1 Exemple de règle de lithotype en 2D et 3D ( $y_1$ ,  $y_2$  et  $y_3$  sont des variables gaussiennes).

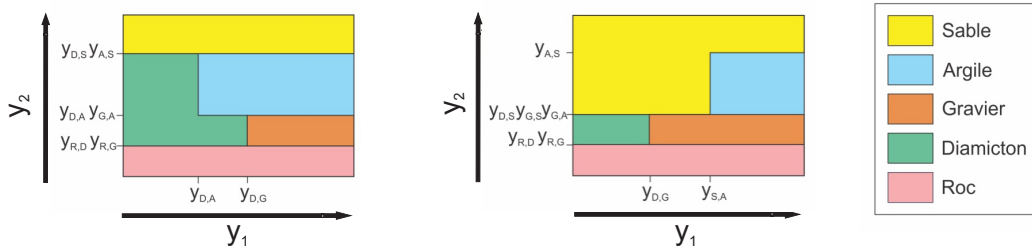


Figure 2.2 Impact de la variation spatiale des proportions des matériaux géologiques sur les lithotypes où chacun des drapeaux correspond à un contexte différent du domaine ( $y_1$  et  $y_2$  sont des variables gaussiennes).

Le processus de simulation par PGS se divise en trois grandes étapes : l'inférence du variogramme des variables gaussiennes, le conditionnement des données et la simulation des unités. Les variogrammes ajustés de la fonction aléatoire gaussienne permettent ensuite de générer les champs multigaussiens. La structure des variogrammes définit l'allure des contacts qui seront produits entre les unités (figure 2.3). Par exemple, la forme des contacts sera plus lisse pour une structure de variogramme gaussienne qu'une structure sphérique ou exponentielle.

Une fois l'inférence du variogramme des variables gaussiennes terminée, il faut effectuer le conditionnement des données. On peut imposer les contraintes d'inégalité locale lors de la simulation par l'utilisation de l'échantillonneur de Gibbs [50]. L'échantillonneur de Gibbs a pour fonction de créer des données gaussiennes conditionnées. Il ne reste plus qu'à simuler ces données gaussiennes à l'aide d'une méthode de simulation standard de variable continue (ex. : bandes tournantes) et de tronquer les valeurs simulées selon la règle de lithotype.

La méthode de simulation plurigaussienne a fait ses preuves dans le cas stationnaire. Au cours des dernières années, plusieurs auteurs ont contribué au développement de la méthode de PGS et au traitement de la non-stationnarité. À titre d'exemple, des algorithmes ont

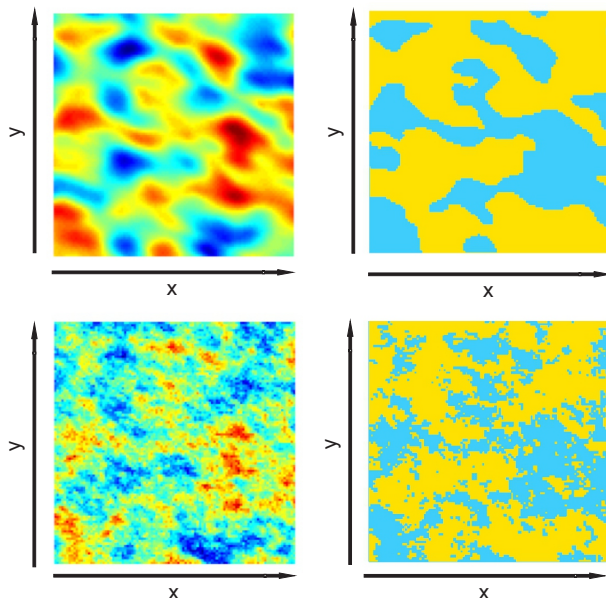


Figure 2.3 Effet de la structure des variogrammes gaussiens (haut) et exponentiels (bas) sur les champs gaussiens (gauche) et le domaine géologique correspondant (droite).

été adaptés pour : i) permettre de cosimuler une plurigaussienne avec une variable continue [65] ; ii) remplacer la règle de lithotype par un diagramme non paramétrique en utilisant des attributs sismiques comme l'impédance acoustique et le ratio de Poisson [62] ; iii) effectuer une simulation qui joint les unités avec la diagenèse ou avec des teneurs [66–68] ; iv) étendre le modèle plurigaussien utilisé pour représenter les variables catégoriques avec des catégories mutuellement exclusives [69] ; v) tenir compte de la relation entre la chronologie et les contacts [53] ; vi) optimiser la reproduction du variogramme à l'aide d'un espace multivariable [44]. Toutefois, des problématiques théoriques et pratiques persistent pour le cas non stationnaire [70]. Récemment, Madani et Emery [71] ont proposé l'utilisation de fonction aléatoire intrinsèque d'ordre  $k$  avec incrément gaussien pour traiter une forme de non-stationnarité.

Un inconvénient majeur de la méthode plurigaussienne est qu'elle ne peut facilement incorporer des transitions asymétriques entre unités. Récemment, il a été montré qu'une certaine asymétrie peut être introduite avec la méthode PSG en utilisant des variables gaussiennes corrélées et spatialement retardées [72]. Il est possible de produire des réalisations de simulation directionnelle pour 3 ou 4 unités [73]. Cependant, on peut facilement anticiper que cette méthode ne peut fonctionner lorsque plusieurs faciès sont présents comme dans le cas d'Innisfiell Creek (11 HSU). En raison de ces difficultés de tenir compte de l'asymétrie directionnelle et du contexte non-stationnaire, la méthode PGS n'a pas été retenue.

### 2.1.2 Simulation multipoint (MPS)

La méthode simulation multipoint (MPS) fait partie de la famille des algorithmes de simulation géostatistique qui permet de reproduire l'agencement spatial de motifs (structures géologiques et unités) à partir d'une image d'entraînement (TI) conditionnel ou non à des données [74]. Les fondements de la méthode MPS ont été développés par Guardiano et Srivastava [75]. Strebel [76] a réussi la première implantation efficace d'un algorithme MPS [33]. Cet algorithme permet de simuler un pixel à la fois, mais il est également possible de simuler un ensemble de pixels à la fois [34]. Même s'il existe plusieurs variantes, le principe de base des algorithmes MPS demeure le même. Il s'agit de trouver, à travers une TI, les probabilités conditionnelles d'un événement quelconque pour ensuite tirer une valeur de la distribution conditionnelle qui sera appliquée sur la grille de simulation. L'évènement est défini sur la grille par une fenêtre de recherche prédéfinie et comprend les données et cellules déjà simulées. La TI représente la configuration spatiale (2D ou 3D) de motifs ou objets géologiques. La relation entre ces objets remplace ainsi le variogramme utilisé par la plupart des méthodes géostatistiques. Une image haute résolution comme une image sismique ou une bathymétrie peut servir de TI et fournir des statistiques a priori et des contraintes architecturales [77]. La stratégie de développement d'une TI varie selon les données disponibles et l'objectif de la simulation. Elle peut être basée sur les données [47, 78], sur un modèle conceptuel, sur un modèle déterministe ou même être stochastique [79]. Mentionnons que différents algorithmes permettent de produire des TI [80, 81].

La MPS est très utile dans le contexte où les données ne permettent pas d'inférer les statistiques nécessaires à la modélisation. De manière générale, les résultats obtenus montrent que les contacts entre les unités sont bien reproduits. Toutefois, il est difficile par cette méthode de reproduire de forts contrastes d'épaisseur dans les formations géologiques (ex. : structure composée de plusieurs lits très minces dans une formation épaisse).

La recherche portant sur la méthode MPS est très active, car plusieurs questions pratiques et fondamentales restent ouvertes. Mariethoz et Lefebvre [82] ont récemment établi une liste de points à étudier dont voici quelques exemples :

1. améliorer l'efficacité du temps de calcul ;
2. augmenter le degré de satisfaction de la qualité des modèles résultants ;
3. améliorer la définition des structures allongées et connectées ;
4. établir des ponts entre MPS et les algorithmes de synthèse de texture ;
5. améliorer le traitement de la non-stationnarité.

Le traitement de la non-stationnarité est très important pour le contexte hydrostratigra-

phique du sous-bassin Innisfil Creek. Même avec seulement deux unités, ce traitement par la méthode MPS est effectivement complexe. La figure 2.4a, inspiré d’Honarkhah et Caers [83], montre un exemple de l’utilisation d’une image d’entraînement à deux unités (delta fluviale) dont la non-stationnarité n’est pas directement prise en compte par l’algorithme utilisé (figure 2.4b).

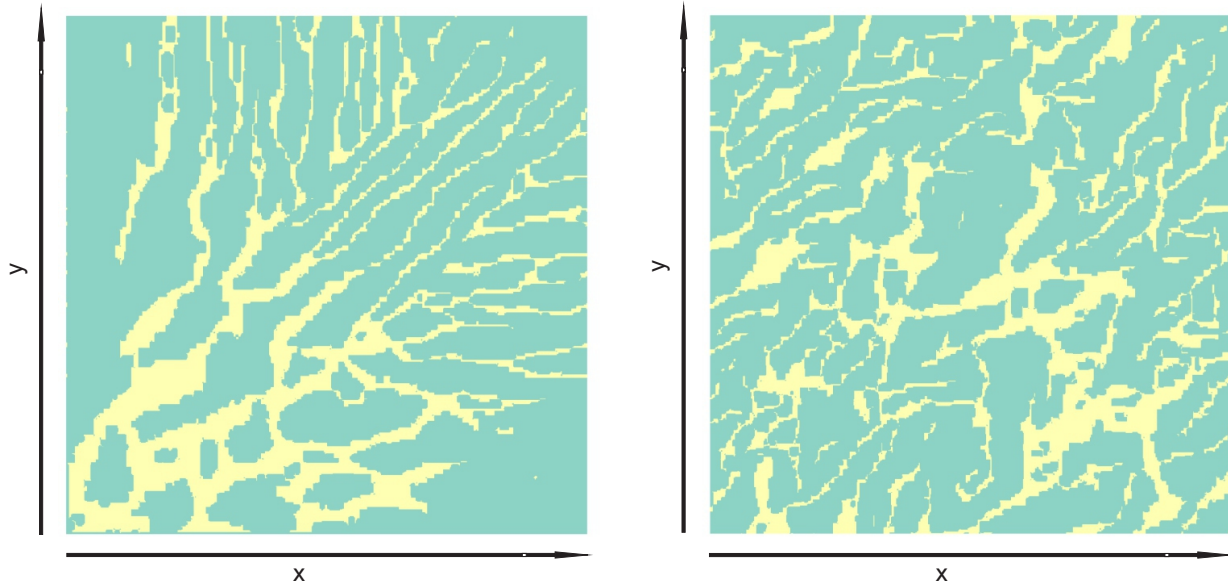


Figure 2.4 Image d’entraînement non stationnaire d’un delta fluviale inspirée d’Honarkhah [1] (gauche) ; b) Réalisation SNESIM sans tenir compte de la non-stationnarité (droite).

Plusieurs auteurs se sont intéressés au traitement de la non-stationnarité de la méthode MPS en traitant cette problématique de différentes manières [35, 83–90]. L’approche du champ de probabilité permet de tenir compte de la non-stationnarité par l’utilisation d’une carte de probabilité d’occurrence des unités. Selon [74], le principal inconvénient de cette approche porte sur l’ajustement et le calage des probabilités. De plus, la reproduction des tendances par cette technique est exigeante. Une approche traditionnelle en MPS est d’introduire la non-stationnarité par le concept de rotation et d’affinité [91]. Strebelle et Zhang [92] ont montré qu’appliquer ces transformations directement sur la TI améliore les résultats et la demande en calcul (CPU). Cette approche demande toutefois de connaître les cartes de rotation et d’affinité en plus de la TI ce qui rend la méthode peu pratique. Une TI complexe peut aussi être divisée en zone stationnaire où chacune des parties correspondra à une zone particulière de la grille de simulation [35, 86]. Cette approche est toutefois laborieuse puisqu’il est difficile de définir ces zones d’autant plus qu’il faut fournir une base de données de motifs pour chacune des zones. Wu et al. [78] présentent une application de cette technique.

Une alternative aux approches précédentes est de coupler à la TI une ou plusieurs variables auxiliaires continues de faible variance spatiale appelées image d'entraînement auxiliaire ou champ de dérive [90]. Pour plusieurs, il s'agit de la meilleure technique pour des applications réelles [47, 74, 90]. Plusieurs techniques issues de cette famille comme IMPALA [89] ont été commercialisées (Jewelsuite, Isatis, SKUA-GOCAD) ou brevetées (Direct Sampling). Jha et al. [77] ainsi que [93] présentent une application du Direct Sampling avec une variable continue (Bathymétrie). Hocker [94] présente un exemple probant d'une MPS avec variable auxiliaire. Il existe quelques variantes à cette approche. Par exemple, [35] propose d'utiliser une TI large constituée de zones stationnaires afin d'y extraire les statistiques de chacune (« feature based methods »). Mariethoz et Kelly [95] proposent plutôt d'utiliser une TI simple et de la transformer en utilisant des statistiques d'ordre faible sur l'ensemble d'une région pour créer des motifs à grande échelle. Bien que performante, cette approche peut s'avérer difficile d'application sur des cas réels. Bien que maintenant incontournables [40, 83], les premiers algorithmes à utiliser la notion de distance sont SIMPAT et FILTERSIM [91, 96]. Avec l'ajout de techniques particulières, telles que SSM (spatial-similarity method) et ASM (automatic segmentation method), il est peu nécessaire de se préoccuper de la stationnarité de l'image d'entraînement [83]. Récemment, [40] ont proposé une méthode de simulation par motif basée sur l'utilisation d'image d'entraînement auxiliaire (ATI) générée de manière objective et un calcul de distance (EPDS - enriched pattern databases simulation). La méthode proposée permet d'obtenir des résultats de très bonne qualité reproduisant les données à un haut taux de succès ainsi que la continuité entre les motifs. Rezaee et Marcotte [97] montrent aussi, pour des cas synthétiques, qu'une TI stationnaire peut générer des réalisations non stationnaires conditionnelles. Les résultats conservent la texture fournie par la TI en suivant les tendances identifiées dans les cartes de probabilité obtenues à partir de données secondaires. Ce résultat a été reproduit pour l'étude de cas réel du champ de méthane hydraté de Mallik. Mentionnons qu'il existe plusieurs autres algorithmes (SIMPAT, Patchwork Simulation) de simulation basés sur les motifs [96, 98].

Dans le contexte hydrostratigraphique du sous-bassin versant d'Innisfil Creek (section 3.2), la méthode multipoint risque d'éprouver des difficultés en raison de la non-stationnarité et surtout de l'asymétrie des UHS ainsi que les forts contrastes d'épaisseurs. Par un exemple, si l'algorithme commence la simulation à partir d'une surface connue et progresse de haut en bas, deux cas de figure sont possibles. Le premier cas, où la fenêtre conditionnante est petite selon la verticale. Dans ce cas, il y aura peu de contrôle sur les épaisseurs qui seront simulées pour chaque unité, car il sera difficile de voir les contacts dans la fenêtre (sans parler de la continuité latérale). Le deuxième cas, où la fenêtre conditionnante est grande selon la verticale, assez peu de variabilité sera observée sur les épaisseurs, car les contacts seront

présents dans cette fenêtre. Pour illustrer ces deux scénarios, supposons (en 1D) que le modèle conceptuel selon la verticale est 5 pixels rouges, suivi de 5 pixels noirs et de 5 pixels bleus. On impose le 1<sup>er</sup> pixel rouge. Si l'on prend une fenêtre de 2 pixels, dont le 1<sup>er</sup> est connu rouge, le 2<sup>e</sup> pixel sera rouge avec probabilité 4/5, noir avec probabilité 1/5 et bleu avec probabilité 0/5. Supposons que rouge est tiré pour le 2<sup>e</sup> pixel. En répétant, on aura à nouveau une probabilité de 4/5 que le 3<sup>e</sup> pixel soit rouge sachant que le 2<sup>e</sup> est rouge et ainsi de suite. Ainsi, il pourrait arriver que l'on simule par exemple 10 pixels rouges de suite, ce qui revient à tirer 9 fois le rouge et dont la probabilité  $p$  est  $(4/5)^9 = 0.13$ . Il y aurait donc beaucoup de variabilité sur les épaisseurs de l'unité rouge et des autres unités. Par contre, si l'on prend plus de pixels consécutifs pour conditionner, par exemple 6 pixels alors quand on simule les pixels 7 à 12 étant donné les pixels 1 à 6 (5 rouges et 1 noir), on aura la probabilité 1 de choisir 4 pixels noirs et 2 pixels bleus étant donné l'information fournie par le modèle conceptuel. Comme on peut le constater, avec une grande fenêtre on risque de simplement reproduire un modèle conceptuel sans variabilité et avec une petite fenêtre, on risque d'avoir trop de variabilité et de produire des situations peu réalistes. La problématique décrite dans cet exemple peut être évitée par l'utilisation combinée d'un chemin de simulation aléatoire, d'une approche multi-grille et de l'ajout de variables auxiliaires. Toutefois, la définition de ces variables demeure compliquée particulièrement en 3D et rien ne garantit des résultats satisfaisants. La méthode multipoint ne semble pas appropriée pour la simulation en contexte hydrostratigraphique à sédimentation directionnelle tel que présent dans le sous-bassin d'Innisfil Creek.

### 2.1.3 Prédiction catégorique de type Markovienne (MCP)

La méthode de prédiction catégorique de type Markovienne (MCP), introduite par Allard et al. [99], permet de prédire spatialement des variables catégoriques comme des UHS. La méthode repose sur une approximation de la formule de Bayes. Elle calcule la probabilité conditionnelle qu'une unité  $i$  soit observée à une localisation donnée  $x_0$  étant donné les unités  $i_1, \dots, i_n$  localisées à  $x_1, \dots, x_n$ . Voici le développement du calcul de la probabilité conditionnelle :

$$p_{i_0|i_1, \dots, i_n} = \frac{p_{i_0, i_1, \dots, i_n}}{p_{i_1, \dots, i_n}} = \frac{p_{i_0, i_1, \dots, i_n}}{\sum_{i_0=1}^I p_{i_0, i_1, \dots, i_n}} = \frac{p_{i_0} p_{i_1, \dots, i_n|i_0}}{\sum_{i_0=1}^I p_{i_0} p_{i_1, \dots, i_n|i_0}} \quad (2.1)$$

La probabilité conditionnelle  $p_{i_1, \dots, i_n|i_0}$  de l'équation 2.1 est approximée par les probabilités univariées et bivariées. Le calcul de ces probabilités est approximé par l'équation suivante :

$$p_{i_1, \dots, i_n | i_0}^* = \prod_{k=1}^n p_{i_k | i_0}(h_{0k}) \quad (2.2)$$

L'idée fondamentale de la méthode est de remplacer la distribution conjointe  $(i_1, \dots, i_n | i_0)$  par le produit  $(i_1 | i_0 \times i_2 | i_0 \dots)$ . Dans l'approche MCP, seules les probabilités bivariées impliquant le point à estimer et chaque point de données sont prises en compte. Les probabilités bivariées entre les paires de points de données sont ignorées. On formule donc une hypothèse d'indépendance conditionnelle puisque  $i_k$  est indépendant de  $i_{k'}$  sachant  $i_0$  pour tout  $k'$  différent de  $k$ . Finalement, l'estimation de la probabilité conditionnelle par MCP se calcule à l'aide de l'équation suivante :

$$p_{i_0 | i_1, \dots, i_n}^* = \frac{p_{i_0} \prod_{k=1}^n p_{i_k | i_0}(h_{0k})}{\sum_{i_0=1}^I p_{i_0} \prod_{k=1}^n p_{i_k | i_0}(h_{0k})} \quad (2.3)$$

où  $p_{i_0}$  est la probabilité univariée de l'unité  $i$  à  $x_0$  et le terme  $p_{i_k | i_0}(h_{0k})$  représente le transiogramme. Allard et al. [46] soulèvent une propriété forte intéressante du transiogramme. En effet, si pour un datum donné, le transiogramme est 0 ou 1, l'équation de probabilité bivariable conditionnelle  $p_{i_1, \dots, i_n | i_0}^* = \sum_{k=1}^n p_{i_k | i_0}(h_{0k})$  sera respectivement égale à 0 ou 1. Ainsi, lors des simulations lorsque  $p_{i_k | i_0}(h_{0k}) = 0$  les transitions d'unités sont interdites et lorsque  $p_{i_k | i_0}(h_{0k}) = 1$  les transitions sont obligatoires. Cette propriété représente un atout majeur surtout pour simuler des séquences spécifiques des unités directionnelles comme dans du contexte hydrostratigraphique d'Innisfil Creek.

Les fonctions de probabilité bivariable servant à calculer les probabilités conditionnelles des unités sont estimées à partir des données, d'une image d'entraînement ou d'un modèle existant. La figure 2.5 montre un exemple de fonctions de probabilité bivariable 3D entre l'unité ATB2 (numéro 2) et les 11 unités du modèle déterministe du sous-bassin d'Innisfil Creek (figure 3.2). Ces probabilités ont été calculées par transformée de Fourier rapide [2]. La figure 2.5 montre entre autres de fortes asymétries spatiales des probabilités entre l'unité ATB2 (2) et les unités AFB1 (1), ATD1 (6) et ATD2 (8) par exemple.

Dans MCP, le calcul des probabilités bivariées implique le point à simuler et les données avoisinantes. À chaque point de la grille de simulation, le nombre de points de données et leur position spatiale influencent le résultat de l'unité simulée. Les points de données utilisés sont sélectionnés à partir des paramètres de simulation. Ces paramètres sont le nombre maximal de voisins (nmax) et le rayon de recherche (dmax). Ces deux paramètres ainsi que le nombre de données (ndata) contrôlent les résultats de simulation. Bien qu'il s'agit de paramètres typiques pour des simulations géostatistiques, il est important de souligner que MCP semble



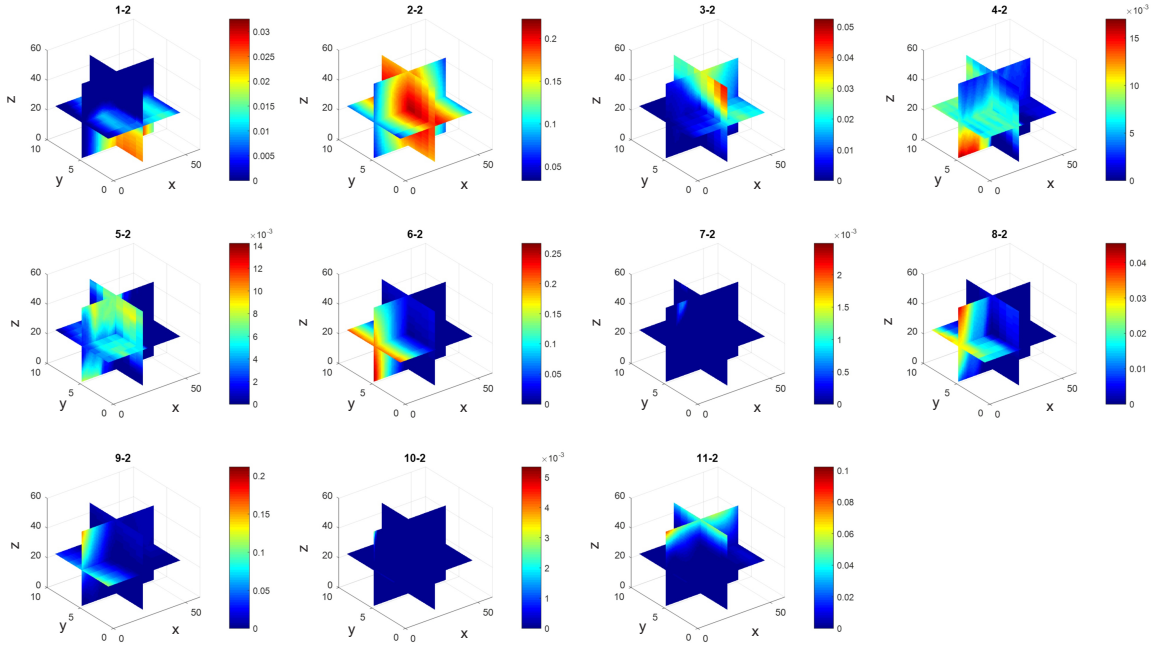


Figure 2.5 Exemple de fonctions de probabilité bivariable 3D calculées par transformée de Fourier [2] ; le  $h = 0$  se trouve au point d'intersection des trois plans.

mieux fonctionner avec un  $n_{\max}$  faible ( $\leq 5$ ) [99]. Le paramètre  $d_{\max}$  doit être assez grand pour éviter de tirer directement de la marginale, mais pas trop grand non plus pour éviter d'augmenter significativement le temps de calcul. L'assignation d'un nombre minimum de voisins ( $n_{\min}$ ) permet d'éviter le tirage de la marginale ( $n_{\min}=0$  par défaut). Lorsque  $n_{\min}$  n'est pas atteint, le point à simuler est déplacé à la fin de la séquence de simulation.

En résumé, la méthode compile la distribution conditionnelle locale en tenant compte des observations à l'intérieur du voisinage prédéterminé. Ainsi la MCP évalue la probabilité que l'unité  $i$  soit observée à un endroit donné  $x_0$  du domaine compte tenu des observations disponibles. La méthode garantit de calculer, à une localisation donnée, des probabilités entre 0 et 1 avec une somme égale à 1 sans post-traitement contrairement à la simulation par indicatrice. Ainsi, si la probabilité bivariable à une localisation donnée considérant le voisinage est de 0, alors la probabilité conditionnelle sera aussi de 0. Cette propriété permet de reproduire des séquences [99].

Avant de terminer cette section, mentionnons que la simulation de variable catégorique basée sur des probabilités de transition comme MCP a initialement été développée pour estimer des variables continues comme la méthode de l'entropie maximale bayésienne (BME) [100–104].

Par la suite, elle a été adaptée pour l'estimation de variables catégorielles [105–108]. BME permet aussi de reproduire des structures géologiques complexes qui tiennent compte des asymétries dans les probabilités bivariées. Cependant, les calculs explosent pour des modèles à multiples unités et multiples points. Des approches plus simples sur le plan des calculs ont par la suite été proposées [99, 109–111]. Toutefois, Allard et al. [99] ont démontré que MCP fait une bonne approximation de BME et constitue ainsi un substitut efficace à cette méthode de calcul intensif. Mentionnons aussi que la méthode de simulation T-PROG [112], basée sur les probabilités de transition, permet aussi de bien reproduire des structures géologiques et des asymétries entre des unités. Toutefois, T-PROG rencontre des problèmes similaires à ceux observés avec la simulation d'indicatrice notamment pour la somme de probabilités supérieure à 1 pour laquelle des corrections doivent être apportées [99]. La somme des probabilités avec la méthode MCP est toujours de 1. La méthode MCP comporte un autre avantage majeur, la non-stationnarité est indirectement prise en compte par le transiogramme et aussi grâce aux données conditionnantes. En effet, le transiogramme permet de préserver l'ordre des unités et les données conditionnantes contraignent beaucoup leur position. Sans données conditionnantes, il est possible de respecter l'ordre, mais en surface on pourrait partir de n'importe quelle unité. Le transiogramme estime la probabilité de transition spatiale de l'unité et a l'avantage de pouvoir considérer l'asymétrie entre unités [112]. La MCP devrait donc permettre de simuler des modèles hydrostratigraphiques conceptuels de déposition ordonnée. Cependant, la méthode n'a jamais été testée sur un système complexe avec plus de trois unités, ni en 3D et ni sous un contrôle directionnel fort des unités. Il semble donc que la méthode MCP soit suffisamment prometteuse pour produire des modèles équivalents du contexte hydrostratigraphique du sous-bassin versant Innisfil Creek et quantifier l'incertitude de ces modèles.

## **2.2 Caractérisation stochastique de la conductivité hydraulique des modèles hydrogéologiques**

En modélisation hydrogéologique, l'étape qui suit le développement d'un modèle hydrostratigraphique numérique est l'attribution des propriétés hydrauliques [113]. À l'échelle régionale, cette étape est complexe et dépend beaucoup de la quantité et de la distribution des données disponibles [114, 115]. Les principales propriétés hydrauliques, à définir dans un modèle d'écoulement, sont la conductivité hydraulique, le coefficient d'emménagement et la porosité. La conductivité hydraulique est généralement la propriété montrant le plus d'influence (sensibilité) sur les résultats d'écoulement tandis que la porosité joue un rôle important sur le calcul du champ de vitesse pour évaluer le transport [11]. Ces deux paramètres sont liés

par la loi de Hagen-Poiseuille [116] et il est donc nécessaire d'en connaître au moins un. Par exemple, Mariethoz et al. [58] utilisent les simulations non conditionnelles pour simuler la porosité et dérive la conductivité hydraulique à partir de cette loi. Contrairement à  $K$ , le coefficient d'emménagement est un paramètre généralement peu sensible et intervient seulement dans la solution du régime transitoire. Siirila-Woodburn et al. [117] montrent comment une caractérisation précise de la distribution de  $K$  peut être importante pour la modélisation particulièrement pour les unités hydrogéologiques présentant des valeurs extrêmes de  $K$  rarement observées. Les structures de perméabilité extrême jouent un rôle clé lors de l'écoulement et le transport à petite échelle. À titre d'exemple, un horizon continu de faibles valeurs de  $K$  peut restreindre la percolation verticale et ainsi diminuer la migration d'un composé quelconque et affecter le niveau des mélanges [118].

Dans cette section, on s'intéresse aux étapes de la caractérisation de la conductivité hydraulique des systèmes hydrostratigraphiques régionaux. Un tel système est généralement composé de nombreuses UHS où chacune des unités inclut typiquement un à trois hydrofaciès en proportions variables. Ces derniers sont habituellement représentés par une  $K$  scalaire qui est estimée à l'échelle locale (exemple : estimation sur un support quasi-ponctuel à partir d'une analyse granulométrique, section 2.2.1). Un ensemble suffisant de valeurs de  $K$ , pour une unité donnée, permet de définir sa structure de covariance spatiale 3D et ainsi définir les hétérogénéités à l'échelle locale. Toutefois, le volume des éléments finis du modèle d'écoulement régional correspond généralement à un support beaucoup plus grand que celui des données. Ainsi, pour être représentative de ce support, une mise à l'échelle de la  $K$  scalaire doit être effectuée vers un tenseur- $K$  équivalent qui prend en compte la variabilité locale (section 2.2.2). La mise à l'échelle peut rapidement s'avérer une étape de calcul imposante. La régionalisation des tenseurs- $K$  par estimation ou simulation géostatistique peut être une solution à cette problématique (section 2.2.3). La dernière étape consiste à caler les composantes des tenseurs- $K$  sur une variable d'état du modèle (section 2.2.4). Les sous-sections suivantes décrivent les résultats de la revue de la littérature de chacune des étapes de la caractérisation stochastique régionale de la  $K$  hydraulique.

### 2.2.1 Estimation de la conductivité hydraulique

La conductivité hydraulique scalaire s'estime à partir d'essais hydrauliques in situ (pompage, slug tests, traceur, ect.), d'essais en laboratoires ou d'analyses granulométriques (formules empiriques) [119]. À l'échelle régionale, les données nécessaires pour caractériser un modèle complet sont normalement insuffisantes en quantité et en qualité. Les études de cas présentés dans cette thèse utilisent l'estimation de la  $K$  à partir des analyses granulométriques puisqu'il

s'agit des seules données de qualité suffisantes disponibles (voir section 3.3.2). De plus, ces données sont bien corrélées aux UHS et elles montrent une excellente résolution verticale qui permet de définir la structure de covariance de la  $K$  selon la direction.

Il existe de nombreuses méthodes empiriques pour évaluer la  $K$  à partir de la courbe granulométrique [120]. Parmi celles-ci, il y a notamment la méthode de Sauerbrei [121], Chapuis [122], Beyer [123], Hazen [124], Navfac [125], Kozeny-Carman [126] et USBR [127]. Le choix de l'une ou l'autre des méthodes dépend surtout des limites de celles-ci à définir l'étendue possible des valeurs de  $K$  et de leurs critères d'application respectifs. Pour une hydrostratigraphie complexe comme celle d'Innisfil Creek constituée d'un large éventail de valeur de  $K$  pour chacune des unités aquifères ou aquitards, la méthode Sauerbrei est celle qui répond le mieux aux critères d'application et celle qui fournit les résultats les plus satisfaisants. La méthode de Sauerbrei produit des résultats réalistes aussi bien pour les unités aquifères qu'aquitards fournissant aussi une grande variance sur l'estimation de  $K$ . La méthode est donc flexible et permet d'estimer la  $K$  pour chaque type de distribution granulométrique.

Le calcul de la  $K$  scalaire selon la méthode de Sauerbrei s'estime à partir de la relation suivante :

$$K = \beta \frac{g}{v} \frac{n^3}{(1-n)^2} d_{17}^2 \quad (2.4)$$

où  $K$  est la conductivité hydraulique estimée,  $\beta$  une constante ( $3.75 \times 10^{-3}$ ),  $g$  la constante gravitationnelle,  $v$  la viscosité de l'eau,  $n$  la porosité et  $d_{17}^2$  est le diamètre moyen des grains de la fraction correspondante.

Les valeurs de  $K$ , estimées à partir de la granulométrie, sont à l'échelle de l'échantillon. Cette échelle, dite locale, est beaucoup trop petite pour être représentative des éléments d'un modèle hydrogéologique régional. Il est donc essentiel d'établir une valeur de  $K$  équivalente à l'échelle de l'élément. Dans un modèle d'écoulement régional 3D, la  $K$  équivalente est généralement définie par un tenseur selon au moins trois directions orthogonales principales suivant une orientation donnée [128]. Lorsque ces composantes suivent les axes de coordonnées du modèle ( $K_{xx}$ ,  $K_{yy}$  et  $K_{zz}$ ), les composantes intermédiaires ( $K_{xy}$ ,  $K_{yz}$  et  $K_{zx}$ ) sont négligeables. Pour définir le tenseur- $K$  équivalent à partir de valeurs scalaires, il est nécessaire d'effectuer un changement d'échelle qui tient compte de l'effet des hétérogénéités de la  $K$  locale.

### 2.2.2 Mise à l'échelle de la conductivité hydraulique en tenseur équivalent

Il est préférable de définir la  $K$  sur une grille de dimension similaire au support des données. Lorsque ce support est significativement plus petit que la résolution des éléments d'un modèle d'écoulement, il est nécessaire d'effectuer une mise à l'échelle des  $K$  [8, 129]. Les va-

leurs de  $K$  attribuées aux éléments doivent donc être représentatives de leur échelle. Dans certaines circonstances, la mise à l'échelle permet aussi de réduire le temps de calcul. Notamment lorsque le nombre d'éléments devient un enjeu. Il est ainsi possible d'utiliser des éléments de plus grande dimension permettant de réduire le temps de calcul du simulateur d'écoulement [130–132]. La littérature abonde sur les méthodes du calcul de mise à l'échelle en hydrogéologie [21, 133–137]. Bien qu'il existe plusieurs approches pour calculer la  $K$  équivalente, aucune d'entre elles n'est entièrement satisfaisante [138]. Pour tenir compte des hétérogénéités à l'échelle locale, la mise à l'échelle doit inclure l'anisotropie 3D de  $K$  [133]. Cela revient à calculer la conductivité équivalente ( $K_{eq}$ ) du tenseur ( $K_{xx}$ ,  $K_{yy}$  et  $K_{zz}$ ). Pour y arriver, les approches intéressantes se regroupent en deux catégories principales, les méthodes algébriques et déterministes [136].

La conductivité équivalente peut être calculée algébriquement par la moyenne arithmétique ( $K_a$ ) ou harmonique ( $K_h$ ) selon la connaissance du milieu, soit la géométrie des unités et la direction d'écoulement. Par exemple, lorsque l'écoulement se fait parallèlement à la stratigraphie on calcul  $K_a$  et  $K_h$  perpendiculairement. La  $K_{eq}$  est bornée par  $K_h$  et  $K_a$  [139]. La moyenne géométrique se trouve entre ces bornes et elle peut être utilisée lors de l'absence de stratification. Le problème avec cette approche, c'est qu'en réalité l'agencement des structures est plus complexe.

Pour des structures stratigraphiques complexes comme celui du sous-bassin Innisfil Creek, la  $K$  équivalente doit être calculée numériquement par une méthode déterministe à partir d'un modèle d'écoulement par égalité des flux ou des énergies de dissipation selon une approche locale ou non locale. Un point important pour cette approche consiste à inférer la bonne fonction de covariance à l'échelle locale pour prendre en compte la structure des hétérogénéités locales. Cette fonction est estimée pour chacune des unités au niveau du support des données [140].

La mise à l'échelle par l'approche locale estime la  $K_{eq}$  d'une maille grossière (bloc) à partir des valeurs de  $K$  sur une maille fine à l'intérieur du bloc. Par la suite, il suffit d'imposer des conditions frontières adéquates et de calculer numériquement le  $K_{eq}$  dans chacune des directions [136]. En 2D et selon une direction d'écoulement donnée, la  $K_{eq}$  est obtenue par l'équation de Darcy :

$$K_{eq} = -\frac{Q}{A} \frac{\Delta l}{\Delta h} \quad (2.5)$$

où  $Q$  est le débit d'écoulement,  $A$  l'aire de la section traversée par le fluide,  $\Delta l$  la distance entre deux faces de charge constante,  $\Delta h$  la différence de charge entre les deux faces.

La méthode locale ne prend pas en compte l'influence des hétérogénéités des blocs avoisinants,

lesquelles influencent les flux aux limites du bloc à mettre à l'échelle pouvant ainsi mener à une estimation moins fiable dans le cas de modèles complexes. S'ajoute à cela, la sensibilité de la méthode aux effets de bordures causés par les conditions frontières aux limites du bloc. La méthode non locale permet de s'affranchir de ces problématiques simplement en augmentant la dimension de la maille fine à l'extérieur du bloc. Selon [129], cette approche reproduit mieux les connectivités comme les chenaux puisqu'elle est moins sensible aux conditions frontières. Par contre, la méthode peut rapidement devenir très lourde puisqu'elle nécessite de résoudre le modèle de plus grande taille à l'échelle fine. La méthode est aussi plus efficace en faisant varier plusieurs fois les conditions frontières [21]. Le tableau 2.1 résume les avantages et inconvénients des méthodes de mise à l'échelle des paramètres hydrauliques.

Tableau 2.1 Résumé de méthodes de mise à l'échelle des paramètres

Méthodes	Avantages	Inconvénients	Références
Paramètre effectif	Une valeur unique pour tout le champ	Une valeur unique pour un champ ne donne pas des résultats réalistes notamment pour le transport.	[113, 129]
Paramètre équivalent : méthodes algébriques	Calcul simple de moyenne arithmétique, géométrique ou harmonique.	On doit connaître, la direction d'écoulement par rapport à la configuration des unités.	[129, 137]
Paramètre équivalent : méthodes déterministes	Se calculent numériquement avec un modèle d'écoulement par égalité des flux	Ne représente pas les zones de connectivité de valeurs extrêmes (conductivité hydraulique) ; choix des conditions frontières n'est pas toujours évident.	[129, 137]

Malgré les améliorations des méthodes de mise à l'échelle, l'incertitude des hétérogénéités à l'échelle locale n'est pas totalement prise en compte par le modèle [11]. Il est donc important de souligner que cette opération risque à un certain niveau d'homogénéiser le milieu [138]. Le calcul de la  $K_{eq}$  est donc complexe puisqu'il dépend notamment des conditions d'écoulement, mais aussi de la connaissance des structures des hétérogénéités [134]. À cet effet, Renard et Allard [141] soulèvent la difficulté du tenseur équivalent à bien préserver l'effet des structures de connectivité à petite échelle. Malgré ces difficultés, la mise à l'échelle de la  $K$  s'avère une opération nécessaire pour rendre l'exécution de la modélisation régionale fonctionnelle en termes de temps de calcul (dimension raisonnable du maillage).

Dans le cadre de cette thèse, on s'intéresse à inclure l'effet des hétérogénéités locales dans

le modèle régional. Le contexte de modélisation régionale, du sous-bassin Innisfil Creek, où chaque élément comprend un niveau important d'hétérogénéité interne favorise l'utilisation de la  $K_{eq}$  calculée numériquement par la méthode déterministe non locale. Ce choix est fait malgré la lourdeur de cette méthode puisqu'elle représente la meilleure méthode dans ce contexte. Toutefois, le principal défi est de réduire le temps des calculs. Pour y arriver, il est proposé de combiner cette approche numérique de mise à l'échelle avec une méthode de simulation géostatistique afin de rendre le temps de calcul fonctionnel (section 2.2.3). Ce choix repose sur l'hypothèse qu'il existe une fonction de covariance des composantes des tenseurs- $K$ . Si tel est le cas, il est possible de réduire au minimum le nombre de blocs à mettre à l'échelle. De plus, une telle approche fournit des réalisations équivalentes du tenseur. Cela permet de caractériser l'incertitude des tenseurs- $K$ .

### 2.2.3 Régionalisation de la conductivité hydraulique

La  $K$  scalaire ou le tenseur- $K$  de milieux hétérogènes sont caractérisés par une variabilité statistique et spatiale. Pour une unité donnée, la variabilité peut s'étendre sur plusieurs ordres de grandeur et être plus au moins continue dans l'espace. La variabilité statistique de la  $K$  influence le choix de l'approche proposée pour régionaliser les valeurs. Évidemment, l'approche la plus simple consiste à estimer une conductivité hydraulique unique par unité et d'attribuer une valeur moyenne ou typique à chacune [113]. Cette approche est généralement utilisée lorsque les données sont peu abondantes ou présentent une faible variabilité [142]. Lorsque la variabilité statistique devient non-négligeable, la régionalisation de la  $K$  s'impose. La régionalisation repose sur la structure spatiale de la  $K$  et elle doit se faire sur un support identique par estimation ou par simulation [143]. Notons que les étapes de base pour régionaliser la  $K$  scalaire ou une composante du tenseur- $K$  demeurent les mêmes. Toutefois, pour le tenseur une difficulté s'ajoute, celle de préserver les relations inter composantes telles que l'imposition de contraintes d'inégalités (ex. :  $K_{verticale} \leq K_{horizontale}$ ) et la préservation des corrélations inter composantes.

L'estimation par krigeage permet de discrétiser spatialement la  $K$  selon la structure spatiale du variogramme. Le krigeage a le grand avantage d'assurer la variance d'estimation minimale. Toutefois, dans le cas qui nous intéresse, le krigeage ne convient pas pour différentes raisons. La raison principale, c'est que le krigeage ne permet pas de produire des modèles équivalents de la  $K$  ce qui rend la quantification de l'incertitude difficile. Aussi le krigeage n'assure pas la reproduction des variogrammes et entraîne généralement une diminution de la variabilité du champ de  $K$  [144]. La méthode ne convient pas aux objectifs de la thèse.

Les méthodes de simulation géostatistique multigaussienne sont souvent utilisées en hydrogéo-

logie stochastique pour définir la variabilité spatiale de la conductivité hydraulique [137, 145]. De plus, les réalisations ont l'avantage, par rapport à l'estimation, de reproduire les fonctions de covariance spatiale ainsi que la distribution statistique cible [143]. Autre avantage, à l'échelle régionale, il n'est pas nécessaire de conditionner les simulations puisque les données sont sur support différent. Par contre, l'hypothèse multigaussienne sous-jacente peut être problématique pour la reproduction de structures de valeurs extrêmes comme des chenaux [141]. Le tableau 2.2 résume quelques avantages et inconvénients des méthodes de régionalisation des paramètres hydrauliques.

Tableau 2.2 Régionalisation des paramètres hydrauliques

Méthodes	Avantages	Inconvénients	Références
Attribuer une valeur moyenne ou typique	Simple	Ne permet pas de modéliser le transport adéquatement si la variabilité interne des unités est importante.	[113]
Krigeage	Permet de discrétiser spatialement le paramètre à une échelle fine conditionnelle aux données selon leur structure spatiale définie par le variogramme.	Ne permet pas de produire des modèles équivalents, n'assure pas la reproduction des variogrammes et entraîne une diminution de la variabilité.	[143, 144]
Simulation géostatistique	Produit des champs stochastiques, reproduit le variogramme ainsi que la distribution statistique des données.	L'hypothèse multigaussienne sous-jacente peut être problématique pour la reproduction de structures de valeurs extrêmes comme des chenaux.	[137, 141]

Pour une étude régionale comme celle du sous-bassin d'Innisfil Creek, la simulation géostatistique des composantes du tenseur- $K$  s'avère très intéressante. Cela permet de réduire le nombre de blocs à mettre à l'échelle, de produire des versions équivalentes des champs de tenseurs- $K$  et de quantifier leur incertitude. Évidemment, ce choix repose sur l'hypothèse que de telles fonctions existent à l'échelle du bloc. Parmi les différentes méthodes de simulation, la méthode des bandes tournantes est très intéressante puisque cette méthode est efficace pour produire de nombreuses réalisations de très grands champs [2].

L'objectif de produire des champs de tenseurs- $K$  équiprobables passe donc par les simulations géostatistiques. Ces dernières sont bien adaptées pour être utilisées avec un simulateur d'écoulement. Toutefois, rien n'assure la reproduction adéquate des variables d'état comme



la charge hydraulique. Ainsi il est nécessaire de perturber les réalisations afin d’assurer un calage satisfaisant.

#### 2.2.4 Calage du tenseur de conductivité hydraulique

La régionalisation du tenseur- $K$  n’assure pas la reproduction de la charge hydraulique ou autres variables d’état. En réalité, c’est quasi improbable de reproduire la charge hydraulique sans une étape de calage. Donc pour améliorer la cohérence entre les tenseurs- $K$  simulés et la charge, il est préférable d’optimiser l’erreur entre la charge mesurée et la charge simulée. La charge simulée est obtenue par modélisation hydrogéologique à partir d’un simulateur d’écoulement.

La modélisation hydrogéologique peut se faire de manière directe ou inverse [8]. La résolution du problème direct est essentiellement utilisée pour faire des prédictions à partir d’un modèle préalablement calé. Elle est aussi utilisée dans les étapes de résolution du problème inverse de type indirecte (figure 2.6).

Selon [146], le problème inverse est la seule solution pour le calage adéquat d’un modèle et pour la validation de la  $K$ . Les méthodes d’inversion ont passablement évolué depuis l’émergence de l’utilisation des modèles numériques comme en témoignent les nombreux articles synthèses [8, 147–150]. L’objectif général du problème inverse est de déterminer les paramètres inconnus d’un modèle. Le problème inverse peut être résolu directement ou indirectement. L’inversion par une méthode directe est presque abandonnée de nos jours [8]. Elle est peu utilisée notamment en 3D parce qu’elle requiert le champ de charges à tous les noeuds du modèle [151, 152]. Toutefois, si l’on dispose d’un modèle d’écoulement préliminaire, cette approche s’avère intéressante [153]. Brouwer et al. [154] présentent une méthode semi-directe intéressante basée sur l’imposition d’une double contrainte qui permet de calculer la conductivité hydraulique par le ratio des champs de flux sur le champ de gradient. Toutefois, la tendance actuelle de la résolution du problème inverse est orientée vers la méthode indirecte comme en témoignent les nombreuses approches [155–168]. Le problème inverse par méthodes indirectes demande de résoudre le problème direct à plusieurs reprises contrairement à la résolution du problème inverse directe. Notons que la forme la plus simple du problème inverse indirect consiste à ajuster manuellement les paramètres (ex. : conductivité hydraulique) jusqu’à la meilleure reproduction d’une variable d’état (ex. : charge hydraulique).

La principale difficulté du problème inverse vient de la non-unicité de la solution, il s’agit d’un problème mal posé [169]. Il existe plusieurs possibilités pour réduire l’espace des solutions telles que réduire le nombre de paramètres à déterminer, réduire l’espace des valeurs des paramètres par de l’information a priori, réduire les fluctuations lors de l’optimisation en

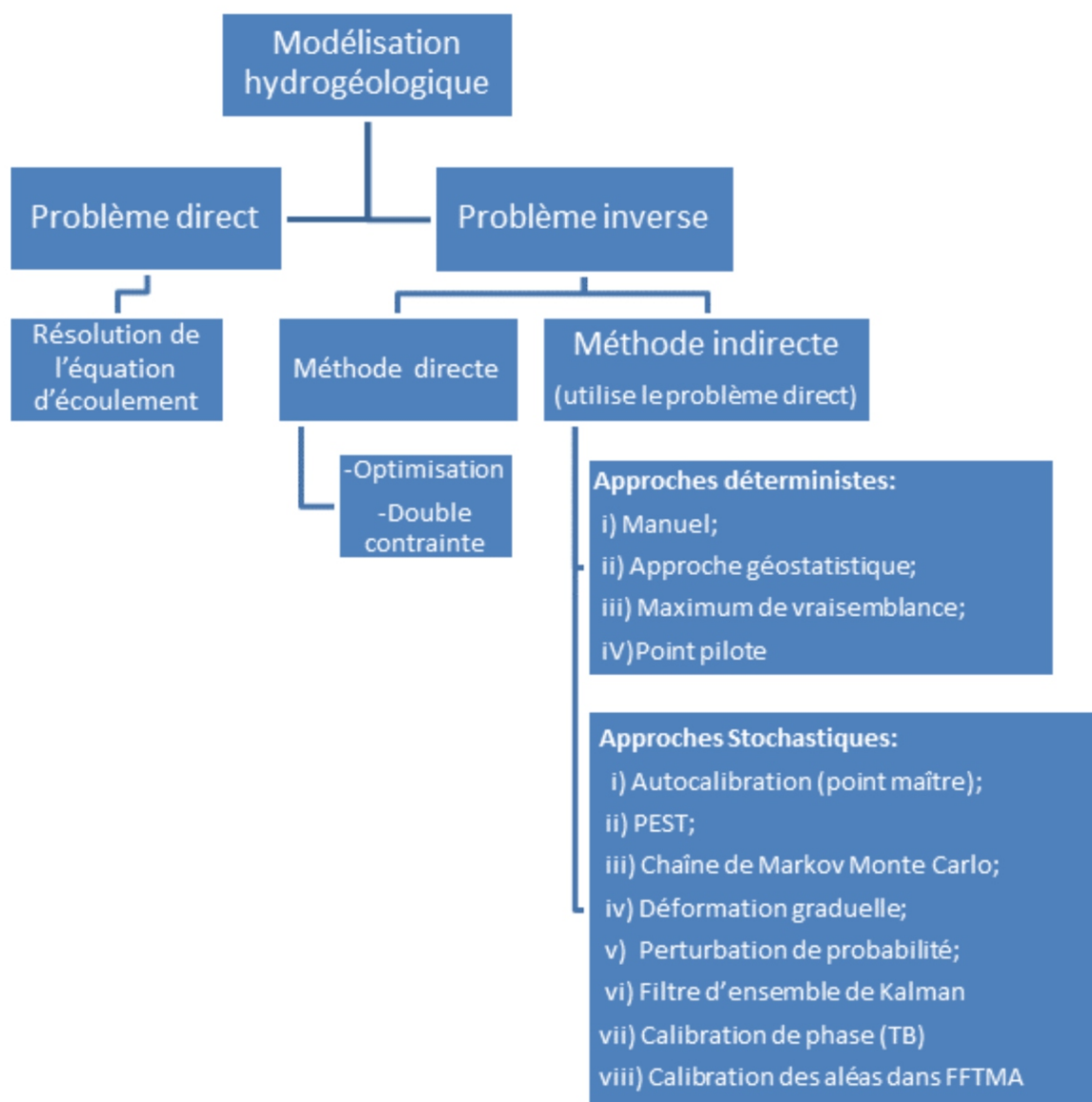


Figure 2.6 Modélisation hydrogéologique

imposant un terme de régularisation et minimiser la non-linéarité dans l'équation du modèle à résoudre [8]. L'autre problème majeur à considérer avec l'inversion est le temps de calcul surtout pour la méthode indirecte. Le changement d'échelle, la parallélisation et des méthodes de résolution efficace comme des techniques de noyau plutôt que la minimisation de gradient sont des solutions envisageables [8]. Chacune des méthodes d'inversion indirecte comporte ses propres avantages et lacunes et aucune d'entre elles n'est entièrement satisfaisante (tableau 2.3).

Tableau 2.3 Synthèse des méthodes d'inversion indirecte

Types	Approches	Méthodes	Avantages	Inconvénients	Référence
Linéariser l'équation différentielle d'écoulement	Minimisations d'une fonction objective de l'écart entre variables d'état mesuré et simulé	Approche géostatistique	Réduire le nombre d'inconnus atténuant la non-unicité de la solution ; faible temps de calcul	Valide seulement pour des valeurs (ex. : $\log K$ ) avec variance faible ; distribution finale lisse ; reproduction inexacte des valeurs de charge mesurées.	[170]
Déterministe		Maximum de vraisemblance	Reproduit bien les observations de charge	distribution lisse du paramètre ; discontinuité entre zones	[137, 171, 172]
Stochastique		Point pilote	Introduit une structure spatiale dans la distribution du paramètre	distribution lisse du paramètre ; instabilité et produit des artefacts ; fluctuation du paramètre limité à la longueur de corrélation	
		Autocalibration (Point maître)	Approche stochastique ; permet de caler plusieurs réalisations équiprobables et évaluer l'incertitude	Krigeage global de la perturbation à chaque point maître	
		Estimation non linéaire de modèle indépendant	Distribution commerciale (PEST)	Champ lisse particulièrement avec la régularisation ;	[164]
	Échantillonnage d'une distribution de probabilité multivariable	Chaîne de Markov Monte-Carlo	Produit plusieurs champs équiprobables ; préserve la structure a priori ; permet les textures non multigaussiennes	Très demandant numériquement ; nombre élevé d'itérations	
	Filtre d'assimilation de données sur la minimisation de la covariance à postériori	Filtre d'ensemble de Kalman	Intégration d'information en temps réel sans modification de l'algorithme ; permet les textures non multigaussiennes	Complexe ; ajout d'une 4e dimension (temps) ; sous-estimation de la variabilité ;	[131, 173]
	Combinaison linéaire successive de paires de réalisations	Déformation graduelle	Préserve la structure a priori ; simple ; permet les textures non multigaussiennes	Taux de convergence ; problème si le modèle a priori est faux	[160, 174]
	Simulation séquentielle	Perturbation de probabilité	Préserve la structure a priori ; permet texture non multigaussienne	problème si le modèle a priori est faux ; basé sur zone	[166]
	Simulation séquentielle	Calibration de phase	Reproduction exacte des données ; fonctionne sur grille irrégulière ; consomme peu de mémoire	La sélection de la phase influence un peu la structure de la corrélation spatiale	[168]
	Simulation séquentielle	Calibration des aléas dans FFTMA	Reproduction exacte des données ; Préserve la structure de covariance spatiale ; perturbation locale ou globale du champ ; meilleure reproduction de l'asymétrie spatiale	Variabilité réduite par l'augmentation du nombre de réalisations ; fonctionne sur grille régulière uniquement ; limité par la mémoire pour des modèles larges en 3D	[167]

Le choix de la méthode indirecte doit donc se faire sur l’objectif de l’inversion, les données disponibles et l’hypothèse que l’on veut imposer. Le tableau 2.3 montre bien des avantages et inconvénients de chacune des méthodes. Pour choisir une méthode, voici quelques exemples de questions à se poser :

1. Est-ce que quantifier l’incertitude est nécessaire (approche stochastique ou non) ?
2. Peut-on accepter l’hypothèse multigaussienne ?
3. Est-ce important de préserver la structure a priori ?
4. Quelle précision recherche-t-on par rapport à la reproduction des variables d’état ?
5. Quel est le temps de calcul que l’on juge raisonnable ?

Selon Zhou et al. [8] le « meilleur » modèle permet de prendre en compte plusieurs variables d’état (ex. : charge et flux) gouvernées par le modèle d’écoulement, il n’est pas limité aux réalisations multigaussiennes, il permet de pondérer l’information a priori et il permet de réaliser plusieurs réalisations de manière efficace. Parmi ces conditions, l’hypothèse multigaussienne dépend essentiellement du choix de l’approche utilisée. En effet, certaines méthodes reposent sur cette hypothèse, telles que la méthode déformation graduelle (GDM) qui combine des champs multigaussiens. D’autres approches incorporent des caractéristiques non gaussiennes dans les réalisations ou lors de la calibration [159, 161, 167, 168, 175]. Toutefois, la GDM se démarque aussi des autres par sa simplicité. Aussi, la méthode est intéressante puisqu’elle combine l’utilisation des gradients et la perturbation du champ simulé de manière itérative lors de l’optimisation de la fonction objectif [176]. De plus, d’autres avantages de la GDM résident dans sa flexibilité à générer des textures variables et sa vitesse rapide de convergence ou du moins de stabilisation [8, 160, 177]. Le calage que l’algorithme cherche à imposer peut venir altérer les caractéristiques multigaussiennes du champ, mais cela devrait dépendre assez fortement de la compatibilité ou non de ces contraintes avec un autre champ multigaussien. Le calage par charges hydrauliques, tel qu’utilisé dans cette thèse, ne devrait pas ou peu affecter l’aspect multigaussien.

Pour la GDM, le fait d’atteindre rapidement un optimum local avec peu d’itérations est primordial considérant le temps d’exécution d’un simulateur d’écoulement. Dans sa plus simple expression, la GDM telle qu’introduite par [160], permet l’optimisation globale d’une fonction aléatoire gaussienne univariable à partir d’une seconde fonction de moyenne et covariance identique. L’équation utilisée pour la déformation est :

$$Y - m = (Y_1 - m) \cos(\pi\theta) + (Y_2 - m) \sin(\pi\theta) \quad (2.6)$$

où  $Y_1$  et  $Y_2$  sont deux fonctions aléatoires gaussiennes indépendantes,  $m$  est la moyenne de ces

fonctions,  $\pi\theta$  est le paramètre de déformation qui varie de  $-\pi$  à  $\pi$  et  $Y$  la réalisation déformée. Lorsque  $\pi\theta = 0$ ,  $Y = Y_1$  et lorsque  $\theta = 0.5$ ,  $Y = Y_2$ . Cette équation s'applique aux réalisations de simulation géostatistique. Dans le contexte du sous-bassin d'Innisfil Creek, la GDM est forte intéressante, car elle permet de générer des champs équivalents calés des composantes du tenseur. Il reste cependant à valider si l'algorithme est en mesure de converger vers des solutions significativement différentes à partir de champs initiaux différents. Cela permettrait de rencontrer l'un des objectifs de la thèse de quantifier l'incertitude associée aux tenseurs- $K$ . Il semble que la GDM a toujours été utilisée en contexte univariable seulement. Dans cette thèse l'approche utilisée sera adaptée et testée pour la première fois pour un contexte multivariable contraint.

### **2.3 Remarques sur la caractérisation hydrostratigraphique régionale et son incertitude sur la modélisation hydrogéologique**

Les modèles hydrogéologiques comportent un niveau variable d'incertitude qui s'accumule avec la complexité de l'étude. Rubin [178] énumère quelques raisons qui caractérisent cette incertitude : la variabilité spatiale, le nombre et la configuration des données ; la perte d'information par la conceptualisation du modèle et les effets numériques comme la taille de la grille. Ainsi, plusieurs sources d'incertitude sont présentes dont celles relatives aux caractéristiques physiques et hydrogéologiques des sédiments (modèle conceptuel,  $K$ , etc.) [179–185]. D'autres sources d'incertitudes sont également à prendre en considération comme le choix de la résolution du modèle (échelle) où s'expriment ces caractéristiques. S'ajoutent à cela, les modèles de changement d'échelle, l'estimation de la recharge, les données d'observation (ex. : charge hydraulique) et les conditions frontières par exemple. Dans cette thèse, on s'intéresse surtout aux caractéristiques physiques pour lesquelles il y a deux types d'hétérogénéités principales, celles inter UHS et celles à intra UHS (hydrofaciès représentés par la  $K$ ). L'approche proposée consiste à modéliser ces deux niveaux types d'hétérogénéités séparément : d'abord les UHS (chapitre 4) ensuite les tenseurs- $K$  dans chaque UHS (chapitre 5) et finalement quantifier leur incertitude respective.

La modélisation stochastique permet de définir un intervalle de confiance des prédictions qui reflète l'incertitude prise en compte. Un bon modèle stochastique est celui dont l'incertitude diminue avec l'ajout de données [178]. Le modèle devrait ainsi converger vers la réalité. Souvent, l'incertitude des modèles provient du manque d'observations des caractéristiques réelles. De plus l'incertitude, est intimement liée aux modèles choisis pour représenter les deux types d'hétérogénéités. Lorsque l'ajout de données augmente l'incertitude, cela signifie que le modèle conceptuel est probablement à revoir. Ce dernier peut s'avérer plus compliqué

que prévu. Un bon modèle doit aussi reproduire les données observées et inclure les connaissances géologiques comme l'ordre connu des UHS ou des connaissances sur la structure particulière (ex. : vallées enfouies). Entre deux modèles mathématiques permettant de reproduire ces informations, on devrait choisir le modèle montrant le plus de variabilité. Les modèles sont définis pour l'essentiel par des histogrammes pour la  $K$  ou par des proportions pour les UHS ainsi qu'une fonction de structure telle que le transiogramme pour MCP et les variogrammes pour  $K$  qui sont d'une part estimée et supposée connues par la suite pour produire des réalisations. Par ailleurs, certains auteurs [186–188] suggèrent d'inclure un méta modèle pour inclure l'incertitude sur les paramètres de la fonction de structure (approche Bayésienne). À première vue, calculer la vraisemblance dans le contexte du sous-bassin d'Innisfil Creek semble irréalisable. Toutefois, l'approche Bayésienne serait à considérer pour des travaux futurs. Pour terminer, rappelons que l'intérêt de produire des modèles hydrogéologiques équivalents est de caractériser quantitativement l'incertitude associée à leurs hétérogénéités hydrostratigraphiques.

## CHAPITRE 3    CONTEXTE HYDROGÉOLOGIQUE DU COMTÉ DE SIMCOE SUD EN ONTARIO

Ce chapitre fait la synthèse des connaissances hydrogéologiques du comté de Simcoe Sud en Ontario (Canada). La synthèse se limite aux besoins de la thèse. Elle inclut une description sommaire du contexte géologique (section 3.1) et hydrostratigraphique (section 3.2) ainsi que les données hydrogéologiques de base pour la modélisation (section 3.3). Les développements méthodologiques de la thèse (chapitres 4 et 5) ont été testés dans le contexte hydrostratigraphique du sous-bassin versant d’Innisfil Creek situé dans le comté de Simcoe Sud (figure 3.1). La région de Simcoe Sud est localisée au nord de la péninsule de l’Ontario, en bordure sud-est de la baie Georgienne. En 2010, un vaste programme de caractérisation géologique et hydrostratigraphique a été développé par l’«Ontario Geological Survey» nommé OGS dans cette région. Les travaux ont été complétés en 2019 [189–194]. La zone d’étude de l’OGS est approximativement calquée sur le découpage municipal du secteur sud de Simcoe et couvre une superficie approximative de  $2333 \text{ km}^2$  (figure 3.1).

### 3.1 Géologie

Dans la région d’étude, la géologie du roc est composée de formations datant du Paléozoïque surmonté d’une épaisse couche de sédiments quaternaires représentant plusieurs épisodes glaciaires formant ainsi un agencement structural complexe marqué par plusieurs séquences d’érosion. Le socle rocheux est composé d’unités sédimentaires pseudo-horizontales datant de l’Ordovicien et du Silurien [195]. Ces roches sont principalement composées de shales et calcaires.

L’ensemble de la région est recouvert de sédiments glaciaires pouvant atteindre près de 190 m d’épaisseur. Dans la partie des basses-terres, on retrouve des sédiments glaciolacustres de remplissage superposés de sédiments fluviaux le long de certains cours d’eau. Des sédiments fluvioglaciaires recouvrent les hautes-terres et les vallées. Le long de la rivière Nottawasaga, la plaine de sable glaciolacustre a été incisée créant ainsi des surfaces de suintement importantes. Adjacent aux sédiments glaciolacustres, on retrouve plusieurs drumlins [189].

Les principales unités lithostratigraphiques sont décrites dans les travaux de l’OGS [189, 191, 196–198]. De la base au sommet, la géologie se divise en une séquence composée de cinq unités principales : les sédiments glaciaires anciens, la formation de Thorncliffe, le till de Newmarket, une séquence équivalente à celle de la moraine de Oak Ridges et la séquence de

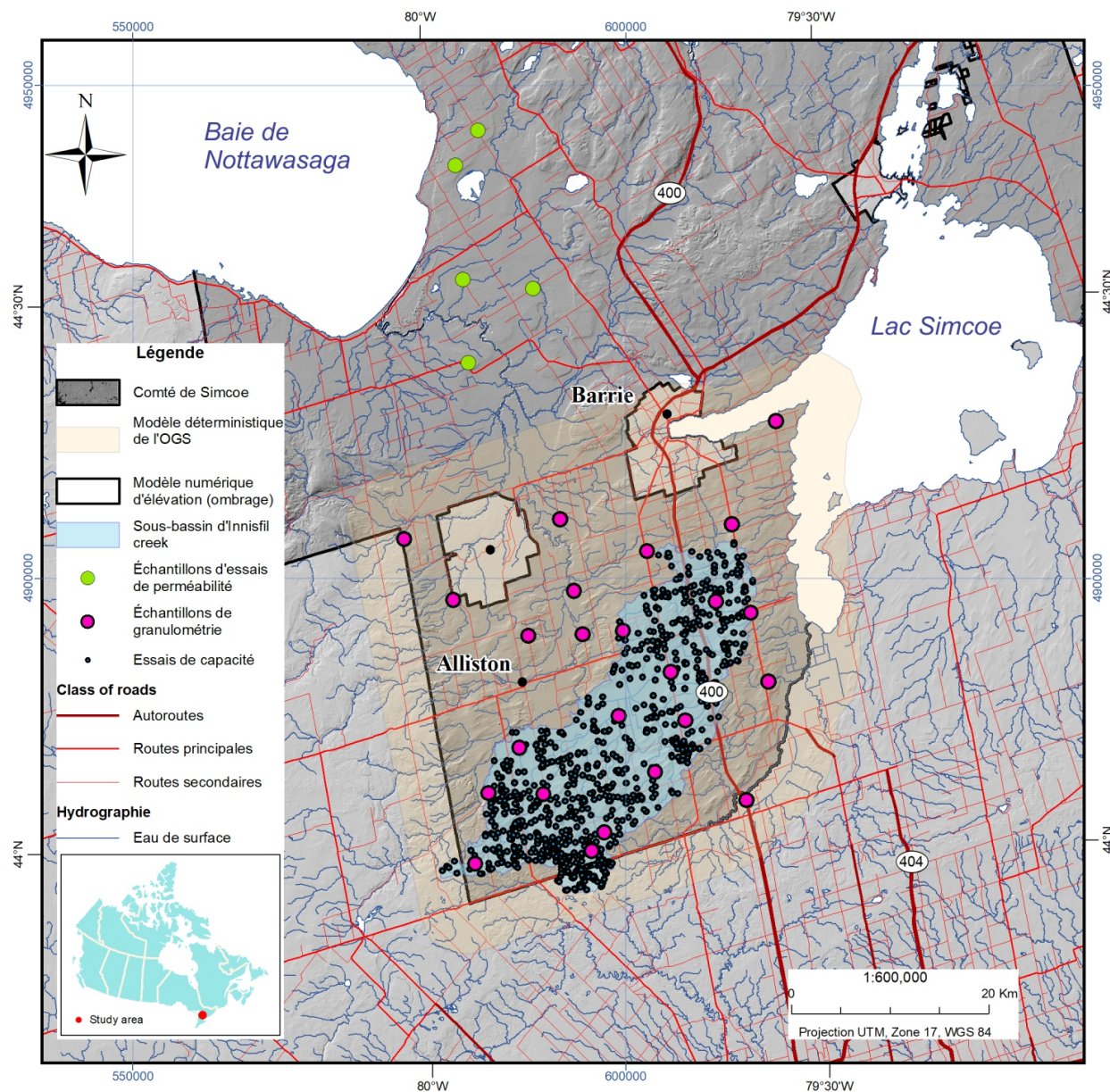


Figure 3.1 Localisation de la zone d'étude du programme de caractérisation géologique de l'OGS et du sous-bassin Innisfil Creek, comté de Simcoe Sud

sédiments récents de remplissage de vallée. Les sédiments glaciaires anciens sont essentiellement composés de matériaux à texture variable. De la base au sommet, ils regroupent le till de York, la formation de Don, les formations inférieures et supérieures de Scarborough et de la formation de Sunnybrook. Le toit du Sunnybrook correspond à une surface de discordance importante avec la formation de Thorncliffe qui surmonte les sédiments glaciaires anciens. La formation sableuse et silteuse de Thorncliffe est généralement présente sous le



till de Newmarket.

Le till de Newmarket est une unité très dense d'origine glaciaire et essentiellement composée de silt sableux avec cailloux [199]. Le poids des glaciers et la précipitation intra pores de la calcite explique la surconsolidation de cette unité [200]. Le toit du Newmarket est représenté par une surface d'érosion majeure d'écoulement sous-glaciaire de forte énergie. Au-dessus de ce till, les sédiments fluvioglaciaires sont équivalents à ceux de la moraine de Oak Ridges [201,202]. Ces derniers sont généralement composés de silt, sable et gravier selon des proportions et des configurations variables [202,203]. Les sédiments postglaciaires, localisés dans les basses-terres, consistent en une épaisse séquence glaciolacustre de remplissage de vallée reposant majoritairement sur diamicton de la formation de Newmarket, mais parfois aussi sur la formation de Thorncliffe ou directement sur le roc. La séquence glaciolacustre est localement recouverte par des sédiments deltaïques sableux graveleux.

### 3.2 Hydrostratigraphie

La classification hydrostratigraphique du comté de Simcoe est tirée de [192]. On y retrouve un total de 15 unités hydrostratigraphiques (UHS). Noter qu'une UHS représente un regroupement d'unités géologiques successives ayant des propriétés hydrauliques similaires.

La classification hydrostratigraphique de la région de Simcoe Sud est présentée au tableau 3.1. Les deux premières lettres du code de l'UHS réfèrent aux aquifères (AF) sableux et graveleux et aux aquitards (AT) silteux et argileux.

La base de l'hydrostratigraphie représente la partie supérieure du socle rocheux où la fracturation contrôle l'écoulement. Le roc est suivi successivement de l'aquitard ATG1, de l'aquifère AFF1 et de l'aquitard ATE1 dont les âges ne sont pas connus. Ils sont identifiés comme des sédiments glaciaires anciens. ATE1 repose au-dessus de ATG1 qui est plus grossier. Ce dernier est parfois séparé par des dépôts stratifiés de sable et gravier (AFF1). Les trois unités au-dessus du roc forment les dépôts glaciaires anciens. Ensuite, se présente une succession de 7 UHS provenant de la formation de Thorncliffe divisée par une alternance de 4 aquifères (AFD4, AFD3, AFD2, AFD1) et de 3 aquitards silteux argileux (ATD3, ATD2, ATD1). L'unité AFD4 de sable et gravier est en discordance avec l'unité ATE1 (diamicton à texture fine). Le till de Newmarket est une unité (ATC1) omniprésente représentant un aquitard régional parfois érodé localement. Il est suivi par les sédiments fluvioglaciaires de type aquifère (AFB2) d'âge équivalent à la moraine d'Oak Ridges (ORM) et dont la géomorphologie est de forme de chenaux d'érosion. Cet UHS comprend des sédiments sableux graveleux à la base suivie d'une séquence de silt et d'argile (granoclassement inverse) et il se termine par

Tableau 3.1 Classification hydrostratigraphique du comté de Simcoe Sud

Unité hydrostratigraphique	Unité lithostratigraphique	Unité hydrogéologique	Lithologie
AFB1	Dépôts grossiers d'eau peu profonde	aquifère	Sable et gravier
ATB2	Sédiments fins glaciolacustres d'eau profonde	aquitard	Silt et argile
AFB2	Dépôts fluvioglaciaires équivalents à ORM	aquifère	Sable et gravier
ATC1	Till de Newmarket	aquitard	Diamicton sableux silteux compact
AFD1	Formation de Thorncliffe	aquifère	Sable et sable graveleux
ATD1	Formation de Thorncliffe	aquitard	Silt et argile
AFD2	Formation de Thorncliffe	aquifère	Sable et sable graveleux
ATD2	Formation de Thorncliffe	aquitard	Silt et argile
AFD3	Formation de Thorncliffe	aquifère	Sable et sable graveleux
ATD3	Formation de Thorncliffe	aquitard	Silt et argile
AFD4	Formation de Thorncliffe	aquifère	Sable et silt
ATE1	Sédiments glaciaires anciens	aquitard	Diamicton silteux argileux
AFF1	Sédiments glaciaires anciens	aquifère	Sable et gravier
ATG1	Sédiments glaciaires anciens	aquitard	Diamicton silteux sableux
Socle rocheux	calcaire, siltstone et shale	aquifère	Socle rocheux

un diamicton. Le AFB2 est superposé d'un aquitard formé des sédiments fins glaciolacustres (ATB2) déposés en eau profonde au-delà du retrait de la glace des lacs glaciaires Schomberg et Algonquin. Le ATB2 est superposé par des dépôts de sable et gravier (AFB1) déposé en eau peu profonde (régression). La dernière unité (ATB1) n'est observée qu'à deux endroits et est sujette à réinterprétation.

### 3.3 Hydrogéologie

Cette section décrit d'abord le potentiel aquifère du contexte hydrostratigraphique du secteur de Simcoe Sud (section 3.3.1). Le reste de la section présente les données hydrogéologiques utilisées dans la thèse pour les besoins de modélisation hydrogéologique (section 3.3.2).

#### 3.3.1 Potentiel aquifère

Le potentiel hydrogéologique des différents aquifères est très variable. Les formations calcaires dans la région d'étude présentent généralement un bon potentiel aquifère dans la zone d'altération des 5 premiers mètres à partir du toit du roc [204]. Cette unité comporte un potentiel karstique dont le rôle sur l'écoulement de l'eau souterraine est toutefois mal connu dans la région. L'épaisseur importante des formations quaternaires et leur potentiel aquifère élevé pour plusieurs d'entre elles font que le socle rocheux représente actuellement une source d'approvisionnement moins attrayante. Les unités sous-jacentes à la formation de Thorncliffe représentent des aquifères à potentiel limité compte tenu de la mauvaise qualité de l'eau souterraine (présence de méthane). La formation de Thorncliffe est parfois considérée comme aquifère lorsque le sable domine et comme aquitard lorsque le silt domine. Pris dans son en-

semble, le Thorncliffe peut montrer un potentiel aquifère significatif lorsque l'épaisseur totale avoisine 10 m ou plus. Lorsque le till de Newmarket est érodé (ou très mince), le Thorncliffe se trouve par endroit en lien direct avec les aquifères sus-jacents comme les sédiments de chenaux enfouis [201]. Cet aquifère peut fournir jusqu'à 100 L/min pour certains puits municipaux [204].

La composition du till de Newmarket et ses propriétés hydrauliques indiquent qu'il s'agit d'un aquitard. Lorsque présent, ce till limite la recharge des aquifères inférieurs tout en réduisant leur vulnérabilité à la contamination. L'érosion complète de ce till provoque l'effet inverse permettant un lien hydraulique direct avec les unités sous-jacentes favorisant l'écoulement vertical. Les conductivités hydrauliques du till de Newmarket varient de l'ordre de  $10^{-11}$  à  $10^{-9}$  m/s avec un potentiel de recharge maximale de 30 à 40 mm/an.

La fonte du glacier Laurentien a produit un large volume d'eau de fonte redistribuant les sédiments glaciaires. Simcoe fut le théâtre de récessions glaciaires produisant des lacs proglaciaires. Le lac Algonquin a ainsi formé dans la partie des basses-terres des aquifères peu profonds à bon potentiel. Notons que les besoins en eau dans la région sud de Simcoe viennent exclusivement de l'eau souterraine à l'exception des municipalités de Alliston et Beeton.

D'un point de vue de la modélisation, il est important de souligner que la méthode de simulation de UHS doit permettre de tenir compte de l'asymétrie verticale, de la classification hydrostratigraphique présentée, afin de respecter l'ordination des UHS. En d'autres mots, la séquence présentée au tableau 3.1 doit être préservée verticalement puisqu'elle est construite suivant une chronologie temporelle. Toutefois, l'absence d'une ou plusieurs unités (discordance) est permise, mais aucune permutation verticale d'UHS ne peut être faite. Latéralement, ces contraintes sont relâchées.

### 3.3.2 Données

L'ensemble des données collectées ont servi à tester les différentes méthodologies développées dans la thèse. Les données proviennent principalement de base de données et de publications existantes. Des échantillons de terrain ont aussi été fournis par l'OGS. Toutes les données ont été traitées et validées pour une utilisation efficace. La description des données porte sur l'hydrostratigraphie (section 3.3.2) et la conductivité hydrauliques (section 3.3.2).

#### Données hydrostratigraphiques

Les données hydrostratigraphiques numériques ont été fournies par l'OGS (Bajc, 29 février 2016, communication personnelle). Ces données incluent les surfaces de l'élévation du toit de

chacune des unités ainsi que des observations hydrostratigraphiques ayant servi à générer les surfaces (forages, points de contrôle ou cartographie de surface). Ces données contenaient les informations nécessaires à l'élaboration du modèle hydrostratigraphique déterministe 3D du sous-bassin versant d'Innisfil Creek (figure 3.2). Noter que ce modèle n'a pas été mis à jour depuis la publication de la version finale [192] puisque c'est celui-ci qui a été utilisé pour les articles de la thèse. Les données et le modèle hydrostratigraphique ont servi à produire les tests du cas réel de l'article MCP (chapitre 4).

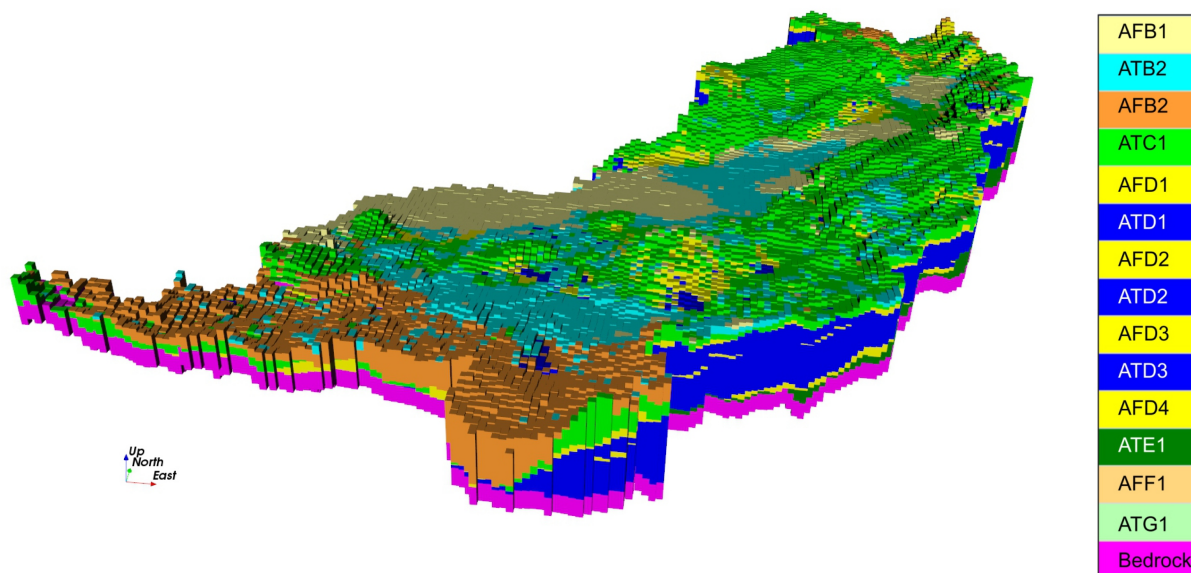


Figure 3.2 Modèle hydrostratigraphique déterministe du sous-bassin versant d'Innisfil Creek ; unités décrites au tableau 3.1 (modèle adapté de Bajc, communication personnelle, 29 février 2016).

## Données de conductivité hydraulique

L'acquisition de données de conductivité hydraulique avait pour objectif de définir la variabilité de conductivité hydraulique de chacune des UHS du modèle hydrostratigraphique. Les données utilisées pour l'estimation de la  $K$  proviennent d'analyses granulométriques [197], d'essais en laboratoire sur des échantillons de forage [205] et d'essais de capacité spécifique disponible à partir des informations de la base de données fournie par les autorités de conservation de la moraine d'Oak Ridges de l'Ontario ou CAMC [206]. La distribution spatiale de ces données est présentée à la figure 3.1.

Les données de granulométrie ont été prélevées par l'OGS [197,207]. Ces données proviennent de 26 forages de haute résolution localisés dans la région de Simcoe Sud. Les échantillons

(n=1694) prélevés le long des forages sont d'excellente qualité et quantité abondante. Des échantillons sont disponibles pour toutes les unités et ils sont tous corrélés à l'hydrostratigraphie du modèle 3D. Les sédiments ont été recueillis majoritairement en échantillons de 10 cm.

Plusieurs méthodes empiriques ont été testées pour évaluer la  $K$  à partir de la courbe granulométrique. Le tableau 3.2 résume les critères d'application des méthodes testées.

Tableau 3.2 Méthodes de prédictions de la conductivité hydraulique et critères d'application ( $C_u$  est le coefficient d'uniformité)

Méthodes de prédiction	Critères d'applications
Hazen [124]	$C_u \leq 5$
	$0,1 \text{ mm} \leq D_{10} \leq 3 \text{ mm}$
	$e \approx e_{max}$ (sol lâche)
NAVFAC DM7 [125,208]	$2 \leq C_u \leq 12$
	$D_{10}/D_5 \leq 1,4 \text{ mm}$
	$0,3 \leq e \leq 0,7$
	$0,06 \text{ mm} \leq D_{17} \leq 1,5 \text{ mm}$
Chapuis et al. [122]	$0,03 \text{ mm} \leq D_{10} \leq 3 \text{ mm}$
	$0,3 \leq e \leq 0,7$
	Sol naturel
Beyer [123]	$0,06 \text{ mm} \leq D_{10} \leq 0,6 \text{ mm}$
	$1 \leq C_u \leq 20$
Kozeny-Carman [126]	Grains de sable grossier
Sauerbrei [121]	Sable et argile sableuse
	$D_{17} \leq 0,5 \text{ mm}$
U.S Bureau of Reclamation USBR [127]	Grains de sable moyen
	$C_u < 5$

Basées sur ces critères ainsi que sur les données disponibles, les méthodes de Hazen, Kozeny et USBR ont été rejetées, car peu d'échantillons répondaient aux critères qualitatifs. La méthode NAVFAC a été rejetée, car aucun échantillon ne répondait au critère quantitatif. Les méthodes de Chapuis et Beyer ont été rejetées, car beaucoup d'échantillons de type aquitard ne répondaient aux critères quantitatifs. Finalement, la méthode de Sauerbrei a été retenue, car la grande majorité des échantillons répondaient à ses critères et les résultats obtenus sont cohérents avec les valeurs anticipées telles que décrits dans la suite de la section. La méthode de Sauerbrei est présentée à la section 2.2.1.

La distribution statistique des valeurs estimées de  $K$  pour chacun de UHS correspond bien aux plages de valeurs typiques de chacun des hydrofaciès (argile, du silt et du sable) ainsi qu'à leurs proportions respectives (Figure 3.3). Ces résultats ont été comparés aux valeurs

de  $K$  obtenues à partir d'essais de perméabilité en laboratoire. Ces essais ont été effectués à partir d'échantillons fournis par l'OGS [209] qui provenaient de Simcoe centre puisque qu'aucun échantillon n'était disponible dans la région de Simcoe Sud. En effet, les forages dans le secteur Simcoe Sud avaient été effectués bien avant le début de la thèse. Les tests de perméabilité ont été effectués à Polytechnique Montréal sur un total de 52 échantillons provenant de 6 forages [205]. Les échantillons étaient principalement composés du sable non consolidé, du silt compact ou de till très compact. Au total, 27 échantillons ont été soumis à un essai de charge constante et à un essai de charge variable effectué selon la norme ASTM [210] avec un perméamètre à paroi rigide. En raison de leur faible conductivité hydraulique, un perméamètre à paroi souple a été utilisé pour effectuer des essais de charges variables sur les échantillons de silt et de till conformément à la norme ASTM [211]. Les conductivités hydrauliques obtenues varient de  $1,5 \times 10^{-6}$  à  $1,2 \times 10^{-4}$  m/s pour les échantillons de sable non consolidé, et de  $9,2 \times 10^{-11}$  à  $1,2 \times 10^{-9}$  m/s pour les échantillons de silt et de till compact. On peut observer que les mesures de perméabilité en laboratoire (étoiles noires) sont en assez bon accord avec les valeurs de  $K$  provenant des analyses granulométriques (figure 3.4) malgré que les faciès des unités dans le centre du comté de Simcoe diffèrent en partie de ceux observés au sud (Bajc, 2019, communication personnelle).

Des valeurs de conductivité hydraulique ont aussi été estimées à partir de données de capacité spécifique [212] extraites de la base de données fournie par le CAMC [206]. Un programme MATLAB proposé par [213] a été modifié pour calculer  $K$  pour un grand volume de données. La valeur de  $K$  estimée par les données de capacité spécifiques est obtenue par la relation itérative suivante :

$$T_{i+1} = \frac{Q}{4\pi i(s_t - s_w)} \left[ \ln \left( \frac{2.25 T_i t}{r_s^2 S} \right) + 2s_p \right] \quad (3.1)$$

où  $T_{i=1}$  est la transmissivité,  $T_i$  la valeur de transmissivité précédente,  $Q$  le taux de pompage,  $s_t$  le rabattement total,  $s_w$  le rabattement dû à la perte de charges au puits,  $t$  le temps de pompage,  $r$  le rayon du puits,  $S$  le coefficient d'emmagasinement de l'aquifère ou la capacité spécifique et  $s_p$  un facteur de pénétration partielle. Le coefficient d'emmagasinement de l'aquifère est basé sur les valeurs de la littérature [214] et le coefficient  $s_p$  est calculé en utilisant l'épaisseur de l'aquifère et la longueur de la crépine du puits. La valeur finale de  $T_{i=1}$  est convertie en  $K_s$  en la divisant par l'épaisseur de l'aquifère.

Au total, 4210 puits étaient disponibles dans la région de Simcoe Sud. Après validation, 1062 puits ont été conservés pour une estimation de  $K$ . Les puits rejetés présentaient de l'information incomplète ou erronée (e.g. mauvaise localisation ou paramètres manquants pour l'estimation de  $K$ ). La figure 3.5 résume les valeurs de  $K$  (indiquées avec les boxplots) obtenues à partir de tests de capacité spécifiques.

L'analyse de ces résultats indique d'abord que les valeurs de  $K$  provenant des tests de capacité spécifique sont biaisées vers les valeurs de perméabilité élevées pour les unités aquitard (ex. : ATD1, ATD2, ATE1, ATG1). En effet, les boxplots de ces unités diffèrent significativement des essais de perméabilité en laboratoire. Les résultats montrent qu'une valeur de  $\ln(K)$  moyenne pour les sédiments aquifères ou les unités aquitards se situe entre  $10^{-5}$  et  $10^{-4}$  m/s. Comme les unités aquitards sont peu perméables, ces valeurs de  $K$  n'ont pas de sens. De nombreuses raisons peuvent expliquer ces résultats. L'association de l'UHS est incertaine puisque ces données proviennent d'une base de données publique de qualité variable. Toutefois, la principale raison vient probablement du fait que le but des forages présent dans la base de données porte sur la mise en place d'un puits fournissant un volume d'eau suffisant pour un approvisionnement. Le puits ne fournissant pas ou peu d'eau ne sont probablement pas rapportés. Il peut aussi y avoir un phénomène d'échelle puisque la conductivité horizontale sur l'ensemble de la zone crépinée sera similaire à la moyenne arithmétique pondérée des conductivités des unités traversées, une faible épaisseur d'unité conductrice venant ainsi tirer fortement la conductivité de l'ensemble à la hausse. En fait, on ne peut pas s'attendre à un puits de captage au sein d'un aquitard.

En conclusion, les résultats de  $K$  issus des analyses granulométriques sont plus réalistes pour l'aquitard. Leur résolution verticale permet aussi de définir la structure de covariance de  $K$  verticale. En contrepartie, la faible résolution horizontale ne le permet pas.

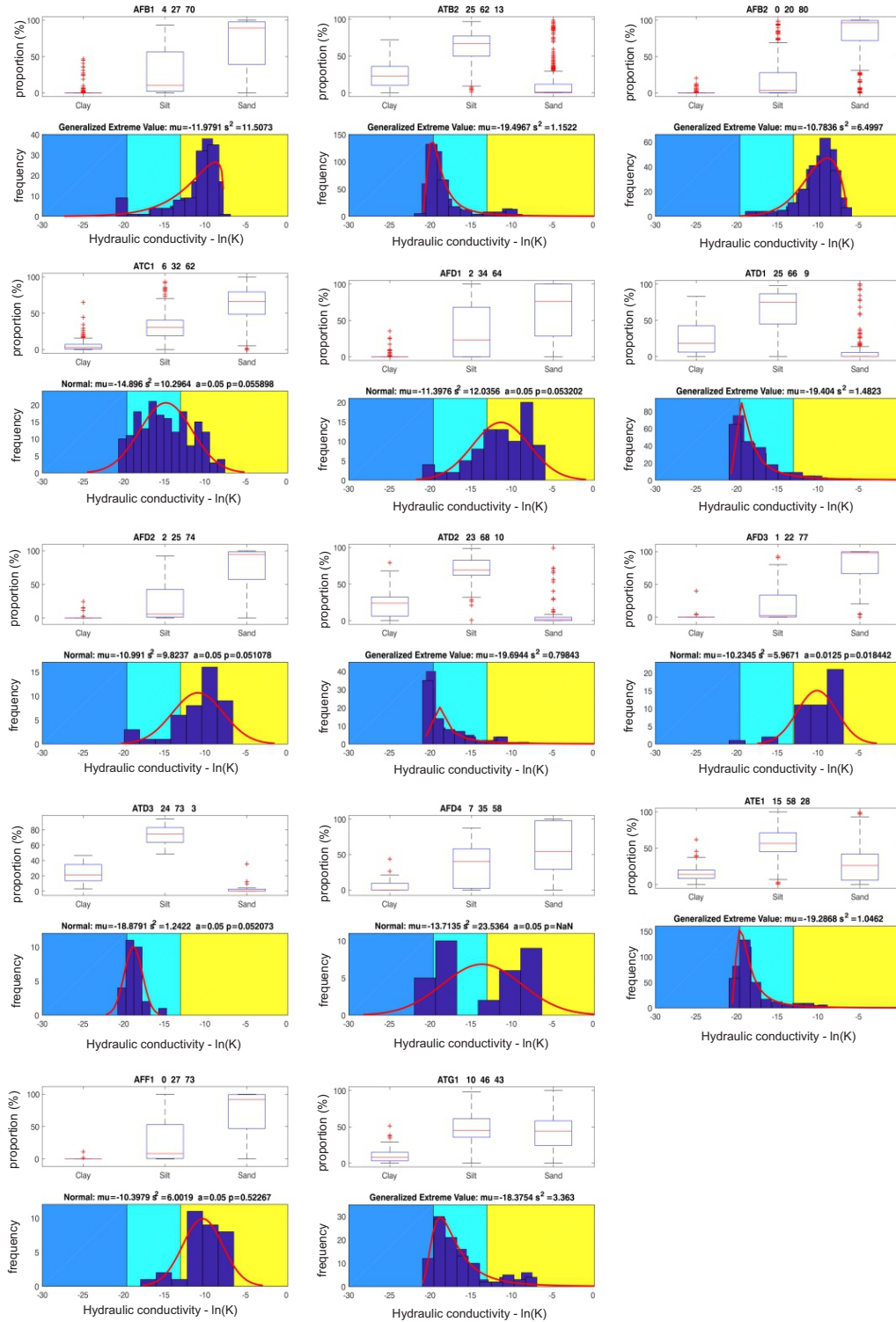


Figure 3.3 Boxplot des proportions d'hydrofaciès (argile, silt et sable) et histogramme du  $\ln(K)$  estimé à partir des granulométries pour chaque UHS. Sur l'histogramme, les polygones colorés (argile=bleu ; silt=cyan et sable=jaune) correspondent aux valeurs typiques de  $\ln(K)$  pour chacun des 3 hydrofaciès [3]. Le titre en haut de la boxplot est l'UHS avec les proportions moyennes de chaque hydrofaciès. Les titres sur l'histogramme sont la moyenne et la variance de la distribution modélisée.



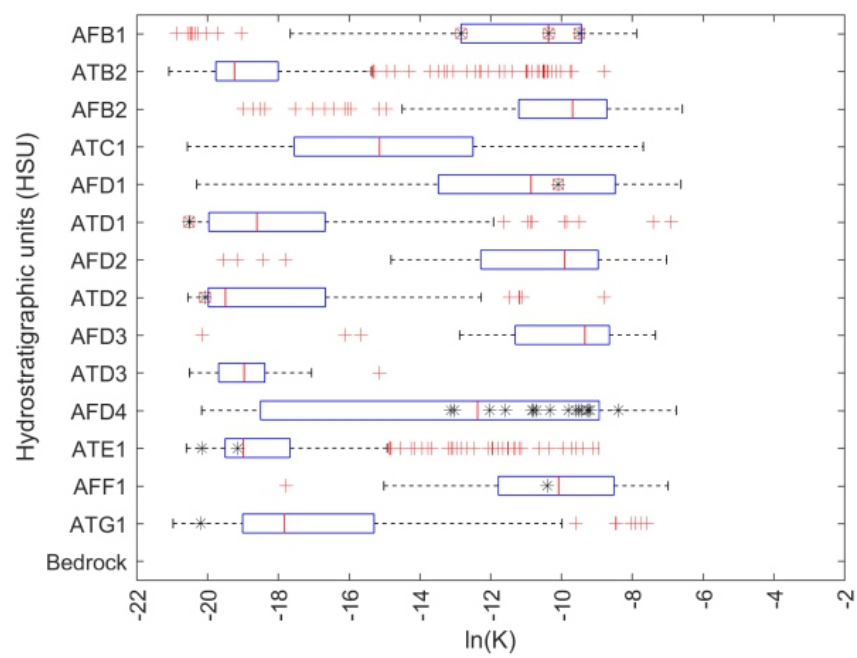


Figure 3.4 Comparaison des Boxplots de  $\ln(K)$  estimé par analyses granulométriques et  $\ln(K)$  provenant des essais de perméabilité (étoiles noires avec un cercle rouge). Note : les étoiles noires seules signifient une corrélation incertaine avec l'unité.

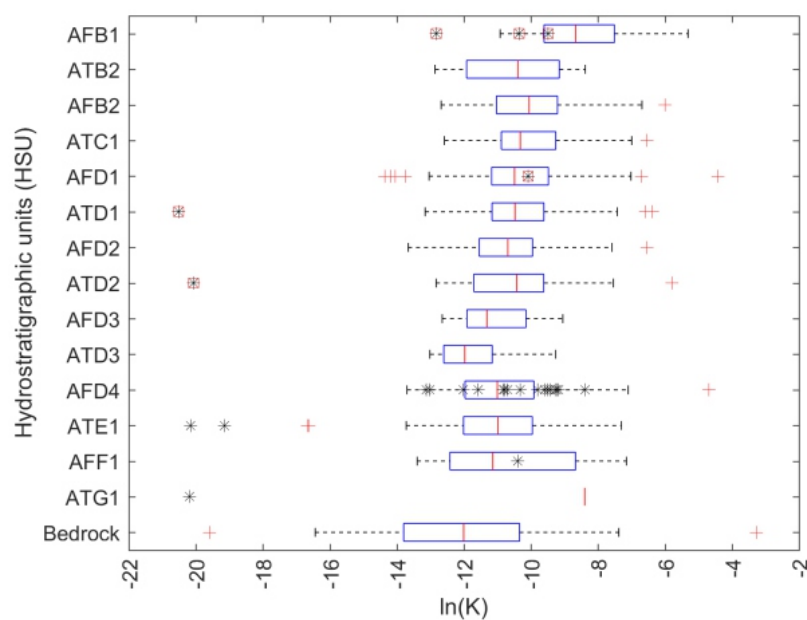


Figure 3.5 Comparaison des Boxplots de  $\ln(K)$  estimé par test de capacité et  $\ln(K)$  provenant des essais de perméabilité (étoiles noires avec un cercle rouge). Note : Les étoiles noires seules signifient une corrélation incertaine avec l'unité.

## CHAPITRE 4    ARTICLE 1 : DIRECTIONAL HYDROSTRATIGRAPHIC UNITS SIMULATION USING MCP ALGORITHM

**Nicolas Benoit**<sup>12</sup>, **Denis Marcotte**<sup>2</sup>, **Alexandre Boucher**<sup>3</sup>, **Dimitri D’Or**<sup>4</sup>, **Andy Bajc**<sup>5</sup> and **Hassan Rezaee**<sup>2</sup>

Stochastic Environmental Research and Risk Assessment (2018) 32 1435 1455

DOI 10.1007/s00477-017-1506-9

Published online : 26 December 2017

@ Springer Berlin Heidelberg 2017

### 4.1 abstract

Understanding the geological uncertainty of hydrostratigraphic models is important for risk assessment in hydrogeology. An important feature of sedimentary deposits is the directional ordering of hydrostratigraphic units (HSU). Geostatistical simulation methods propose efficient algorithm for assessing HSU uncertainty. Among different geostatistical methods to simulate categorical data, Bayesian Maximum Entropy method (BME) and its simplified version Markov-type Categorical Prediction (MCP) present interesting features. In particular, the zero-forcing property of BME and MCP can provide a valuable constrain on directional properties. We illustrate the ability of MCP to simulate vertically ordered units. A regional hydrostratigraphic system with 11 HSU and different abundances is used. The transitional deterministic model of this system presents lateral variations and vertical ordering. The set of 66 (11x12/2) bivariate probability functions is directly calculated on the deterministic model with Fast Fourier Transform (FFT). Despite the trends present in the deterministic model, MCP is unbiased for the HSU proportions in the non-conditional case. In the conditional

---

1. N. Benoit, Geological Survey of Canada, Natural Resources Canada, 490, rue de la Couronne, Quebec, QC, G1K 9A9, Canada, e-mail : nicolas.benoit@canada.ca

2. D. Marcotte, H. Rezaee, Department of Civil, Geological and Mining Engineering, Polytechnique Montreal, Montreal, Canada

3. A. Boucher, Advanced Resources and Risk Technology LLC, Denver, USA

4. D. D’Or, Ephesia Consult, Sombrefe, Belgium

5. A. Bajc, Ontario Geological Survey, Sudbury, Canada

cases, MCP proved robust to datasets over-representing some HSU. The inter-realizations variability is shown to closely follow the amount and quality of data provided. Our results with different conditioning datasets show that MCP replicates adequately the directional units arrangement. Thus, MCP appears to be a practical method for generating stochastic models in a 3D hydrostratigraphic context.

## 4.2 Introduction

The limited sampling available to study complex geological systems are not sufficient to properly characterize critical features, such as the proportions and connectivity of units, that control groundwater flow simulation responses [11,22]. These attributes are essential to assess the response of a hydrogeological system to external stimuli. Hence, adequate characterization of uncertainty about unit proportions and connectivity is of paramount importance for hydrogeological modeling and prediction of aquifer responses [215].

Most groundwater flow models are based on deterministic models that lack any quantification of uncertainty in the geometry, connectivity, and relative proportions of the geological units. Geostatistical simulation methods aim to propose different hydrogeological models called realizations [216]. Each realization is typically an input to a flow simulator from which a response to a specified stimulus is computed. The distribution of modeled responses describes the uncertainty about the real system response to a similar physical stimulus [24]. To be useful, the approach requires each hydrogeological model to be consistent with the relevant natural features expected to be found in the subsurface. Moreover, the ensemble of models should provide a fair assessment of uncertainty. Hence, the models must include all the geological knowledge available while preserving the variability due to incomplete knowledge.

An important common feature in sedimentary deposits is the directional ordering of stratigraphical geological units. For example, in undeformed glacial stratigraphies, this is represented by the law of superposition. This constraint is difficult to incorporate in traditional categorical geostatistical simulation methods like PluriGaussian Simulation (PGS).

The PGS method of [50] with independent Gaussians does not allow asymmetrical ordering, as all transitions are symmetric by construction. Some asymmetry can be introduced using correlated and spatially delayed Gaussians [72]. However, this approach does not allow to fully control the transition probabilities in presence of numerous units.

The multiple-point statistical methods (MPS) have two main variants : either pixel- (e.g. [47, 85]) or patch-based (e.g. [96, 217]). Both are based on the use of one or many training

images that synthesize geological knowledge. Although directionality can be present in a training image, it might be difficult to strictly enforce transitional features into realizations, especially for the patch-based variant. Furthermore, multipoint method used with training image that have low pattern repeatability like the ones seen in sedimentary environments with directional trends [218] may generate realizations lacking variability.

Categorical simulation (units or facies) based on transition probabilities like Bayesian Maximum Entropy (BME) [105,107,108] can reproduce complex geological structures that account for asymmetries in the bivariate probabilities. However, the computations explode for multiple facies and multiple points. Computationally simpler approach like Li's algorithm [109–111] or Markov Category Prediction method (MCP) [99] also enable to account for asymmetries in bivariate probabilities. [99] demonstrated that MCP constitutes a good approximation to BME. However, MCP has never been tested on a complex system with more than 3 units and strong directional control.

The main objective of this study is to demonstrate the ability of MCP to simulate units with directional ordering. After reviewing the MCP approach and its properties, effect of conditioning data is discussed and illustrated with a simple synthetic example. Then, a complex real 3D deterministic model, representing a sub-domain of the Simcoe County hydrostratigraphic system in south-central Ontario, is used as training image for the application of MCP. Various simulation scenarios using different types and levels of information are compared and discussed.

### 4.3 MCP simulation methodology

The MCP approach was introduced by [99] as an efficient substitute for the computationally intensive Bayesian Maximum Entropy (BME) method. The BME, initially developed to estimate continuous variables [100–104], was later adapted for the estimation of categorical variables [105, 106]. In the BME approach for categorical variables, the parameters of the joint discrete distributions are estimated so as to match simultaneously the bivariate probabilities between all pairs. By contrast, in the MCP approach only the bivariate probabilities involving the point to estimate and each data point are considered. The bivariate probabilities between pairs of data points are ignored. Allard et al. [99] showed this corresponds assuming a conditional independence hypothesis : the joint category distributions at locations  $\mathbf{x}_i$ ,  $i = 1...n$  are considered independent once the category at estimation location  $\mathbf{x}_0$  is known.

### 4.3.1 MCP definition

The conditional probability to observe the category  $i_0$  at location  $\mathbf{x}_0$  given the categories  $i_1, \dots, i_n$  observed at locations  $\mathbf{x}_1, \dots, \mathbf{x}_n$  is by definition :

$$p_{i_0|i_1, \dots, i_n} = \frac{p_{i_0, i_1, \dots, i_n}}{p_{i_1, \dots, i_n}} = \frac{p_{i_0} p_{i_1, \dots, i_n|i_0}}{\sum_{i_0=1}^I p_{i_0} p_{i_1, \dots, i_n|i_0}} \quad (4.1)$$

where  $I$  is the number of categories.

In the BME approach, one seeks to estimate directly the above joint distributions subject to a series of constraints aimed at recovering imposed univariate and bivariate distributions. In practice, BME is limited to use of small neighborhood due to the difficulty to estimate high-dimensional joint distributions.

In the MCP method, one assumes a conditional independence hypothesis such that :

$$p_{i_1, \dots, i_n|i_0} = \prod_{k=1}^n p_{i_k|i_0} \quad (4.2)$$

which enables to write Eq. 4.1 as :

$$p_{i_0|i_1, \dots, i_n}^{MCP} = \frac{p_{i_0} \prod_{k=1}^n p_{i_k|i_0}}{\sum_{i_0=1}^I p_{i_0} \prod_{k=1}^n p_{i_k|i_0}} = \frac{p_{i_0}^{1-n} \prod_{k=1}^n p_{i_k, i_0}}{\sum_{i_0=1}^I p_{i_0}^{1-n} \prod_{k=1}^n p_{i_k, i_0}} \quad (4.3)$$

Hence, only the bivariate distributions involving the estimated point  $\mathbf{x}_0$  are required in the MCP approach. It was shown by [99] through a series of examples that, despite the rather strong conditional independence hypothesis, differences observed between the BME and the MCP results were negligible.

One interesting feature of MCP, called the zero-forcing property, is the proper integration of 0/1 probabilities. As soon as  $p_{i_k|i_0=j} = 0$  for any  $k$ , one has  $p_{i_0=j|i_1, \dots, i_n}^{MCP} = 0$ . Similarly, when  $p_{i_k|i_0=j} = 1$  one has  $p_{i_0=j|i_1, \dots, i_n}^{MCP} = 1$ . Therefore, directional sequence of categories can be easily reproduced by MCP as noted by [99]. It makes the method attractive to simulate stratigraphic models with naturally ordered units arrangement.

### 4.3.2 MCP simulation

We apply the MCP approach for the particular case where a training image (TI) is available either in the form of a conceptual model or a deterministic one. In both cases, we assume the basic grid cell is the same size in the TI and the simulated field. Moreover, we assume that all hard data are observed directly at the cell scale or are deemed representative of this

scale.

## Bivariate probabilities

The first step in MCP is to estimate the bivariate probabilities. These are computed directly from the TI for all separation vectors. For efficient computation, we use the Fast Fourier Transform algorithm (FFT) [2]. The bivariate probabilities are computed simultaneously by FFT for each variable pair and every separation vector. During sequential simulation, the separation vectors between the point to simulate and the known points (data or previously simulated) in the neighborhood are computed and used to read directly the corresponding bivariate probabilities. We stress that because everything is defined at a cell scale there is no modeling of bivariate probabilities required nor any smoothing. We simply read directly the experimental probabilities for the MCP computations.

As an example, suppose 11 HSU are available in a 3D setting. One wants to simulate the HSU at cell  $\mathbf{x}_0 = (50, 50, 5)$ . Also assume that HSU 5 and 7 were observed at cells  $\mathbf{x}_1 = (49, 52, 4)$  and  $\mathbf{x}_2 = (55, 46, 7)$  respectively. These two informed cells define the neighborhood of  $\mathbf{x}_0$ . Then, the two separation vectors are  $(1, -2, 1)$  and  $(-5, 4, -2)$ . To compute Eq. 4.3, we read directly the 11 values (for  $i_0 = 1 \dots 11$ ) of  $p_{i_0, i_1=5}(1, -2, 1)$  and the 11 values of  $p_{i_0, i_2=7}(-5, 4, -1)$ .

The bivariate probability of observing categories  $i$  at location  $\mathbf{x}$  and  $j$  at location  $\mathbf{x} + \mathbf{h}$  is simply equal to the non-centered cross-indicator covariance  $P(I(\mathbf{x}), J(\mathbf{x} + \mathbf{h})) = E[I(\mathbf{x})J(\mathbf{x} + \mathbf{h})]$  where  $I(\mathbf{x})$  is the indicator variable that takes value one if HSU  $i$  is observed at point  $\mathbf{x}$  and value zero otherwise. Similarly  $J(\mathbf{x} + \mathbf{h})$  takes value one if HSU  $j$  is observed at point  $\mathbf{x} + \mathbf{h}$ . The expectation is estimated by  $\frac{1}{N(\mathbf{h})} \sum_{\mathbf{x}} I(\mathbf{x})J(\mathbf{x} + \mathbf{h})$  over the  $N(\mathbf{h})$  pair of points with separation vector  $\mathbf{h}$ . As described in [2], for data on a (possibly incomplete) regular grid, this can be computed by FFT using the following equations :

$$N(\mathbf{h}) = \left[ F^{-1} \left( \overline{F(\mathbf{K})} F(\mathbf{K}) \right) \right] (\mathbf{h}) \quad (4.4)$$

$$\sum_{\mathbf{x}} I(\mathbf{x})J(\mathbf{x} + \mathbf{h}) = \left[ F^{-1} \left( \overline{F(\mathbf{I})} F(\mathbf{J}) \right) \right] (\mathbf{h}) \quad (4.5)$$

where  $F$  represents the FFT,  $F^{-1}$  the FFT-inverse, matrix  $\mathbf{K}$  takes the value one at all points of the deterministic model and zero outside, matrices  $\mathbf{I}$  and  $\mathbf{J}$  take the value one at locations where respectively categories  $i$  and  $j$  are observed and zero elsewhere. The matrices  $\mathbf{I}$  and  $\mathbf{J}$  are extended by zero-padding to nullify periodic repetitions implicit in FFT. The over-line indicates complex conjugate. Note that the FFT computes the indicator cross-covariances simultaneously for all lag-distances  $\mathbf{h}$ . The fast computation in Eq. 4.5 is repeated separately for each pair of categories.

## Data and pseudo-data

Conditioning to hard data (HD) is straightforward in MCP as opposed to some other categorical simulation methods like object-based, process-based or patch-based variants of MPS. The HD simply populate the corresponding cells and are never changed during the simulation. In the absence of HD, the algorithm starts either by drawing directly from the marginal distributions or, preferably, by imposing a pseudo-data (PD). Pseudo-data can be provided, as examples, by surficial maps or bedrock topography interpretations. Another possible source of pseudo-data can be a deterministic model, such as a geologist-interpreted CAD model. Taking a small proportion of the deterministic model as pseudo-data helps to reproduce the main characteristics and constraints interpreted by the geologist whilst keeping variability in the simulated models. However, one should refrain from taking too much pseudo-data as this will reduce variability in the realizations.

## Simulation algorithm

The simulation proceeds following the sequential simulation framework where the conditional distributions are computed using Eq. 4.3. The simulated value at the current cell is obtained by a random draw from the category conditional probabilities. The use of a multigrid approach [219] helps to better reproduce the long range structural characteristics described by the TI. The neighborhood search parameters include : extent and anisotropy of the search, number of neighbors to use and search by octant. The effect of these choices can be important (see section 4.3.5).

In the particular case of highly ordered arrangement of units like the ones considered in the next sections, it is advisable to avoid drawing from the marginal distributions as nothing guarantees that the desired order of units will be preserved. A simple modification of the simulation path alleviates this problem by re-simulating at the end of path any location that does meet a minimum number of data found in the neighborhood search.

## Assessing quality of realizations

Two statistics are used to measure the quality of the realizations : the Kullback-Leibler divergence (C1) from [220] and the average spread of HSU proportions (C2) where the spread for a given HSU is measured by the difference between the proportion quantiles 0.95 and 0.05 among the  $n_r$  realizations. Hence,



$$C1 = \sum_{i=1}^I \left( pTI_i \times \log \left( pTI_i / \overline{pMCP_i} \right) \right) \quad (4.6)$$

$$C2 = \frac{1}{I} \sum_{i=1}^I (q_{95,i} - q_{5,i}) \quad (4.7)$$

where  $I$  is the number of HSU,  $pTI_i$  represents the HSU $_i$  proportion in the TI,  $\overline{pMCP_i}$  is the average proportion of HSU $_i$  in the  $n_r$  realizations and  $q_{5,i}$  and  $q_{95,i}$  are the 0.05 and 0.95 quantiles of distribution obtained with the  $n_r$  realizations for HSU $_i$  proportion. The first criterion measures similarity of the distribution of HSU proportions in realizations compared to the TI proportions. Criterion C2 measures the variability of the simulated HSU proportions among the different realizations.

The above criteria were used to assess the impact of various choices for the control parameters. We generally seek to favor small C1 and large C2 values. These two objectives tend to oppose each other. For example, taking a larger proportion of pseudo-data from a deterministic model favors the reproduction of the proportions globally but also in each individual realization which reduces the inter-realization variability of HSU proportions.

In addition to the global statistics C1 and C2, we also computed dissimilarity map between the model used as TI and the different realizations using :

$$d(\mathbf{x}) = \frac{1}{n_r} \sum_{r=1}^{n_r} I_{S_r, TI}(\mathbf{x}) \quad (4.8)$$

where  $n_r$  is the number of realizations, and  $I_{S_r, TI}(\mathbf{x})$  is an indicator variable taking value zero when the simulated unit at location  $\mathbf{x}$  coincides with unit in TI, and value one otherwise. The dissimilarity map provides a spatial representation of where the uncertainty is large with respect to the deterministic model.

### 4.3.3 Illustration of MCP

A synthetic example based on a simple deterministic model is used to illustrate typical results obtained by MCP. The deterministic model used as TI consists of four horizontal hydrogeological units with increasing proportions from top to bottom (respectively 15%, 20%, 30% and 35%).

Figure 4.1 presents four different realizations obtained using three initial scenarios : non-conditional simulation (NCS), conditional simulation with 1% data extracted from the TI

(CS1) and conditional simulation with 20% data extracted from the TI (CS2). Simulations are performed on a 2D 25x20 grid. The minimum and maximum number of neighbors are set to 0 and 5. The search is done by quadrant with a maximum of 2 samples per quadrant.

Based on the 0/1 forcing property of MCP, the sequence of units present in the TI is respected in each realization. As expected, the dissimilarity decreases with the addition of data and the dissimilarity is stronger close to units interfaces.

Figure 4.2 shows the boxplots of the units proportions obtained from 100 realizations of NCS, CS1 and CS2 for the synthetic model. The TI proportions are well recovered for each unit, in all cases. For NCS, the mean proportions obtained from the realizations are not statistically different from proportions in the TI (test on equality of proportions), indicating the absence of bias of the method. A similar test cannot be applied for cases CS1 and CS2 due to the correlation between realizations induced by the conditioning data. However, the boxplots strongly suggest absence of bias of the method. As expected, variability of unit proportions in the realizations decreases with the amount of conditioning data. The expected variability of MCP conditional simulation is studied more in detail in section 4.3.4.

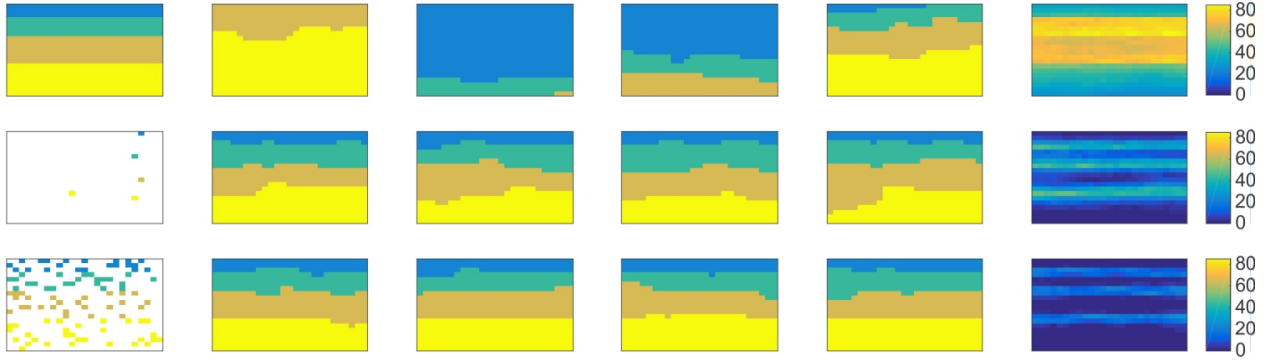


Figure 4.1 From left to right — top row : Synthetic TI, 4 NCS realizations, NCS dissimilarity — middle row : 1% TI as data, four CS1 realizations, CS1 dissimilarity — bottom row : 20% TI as data, four CS2 realizations, CS2 dissimilarity. Based on 100 realizations, dissimilarities expressed as %.

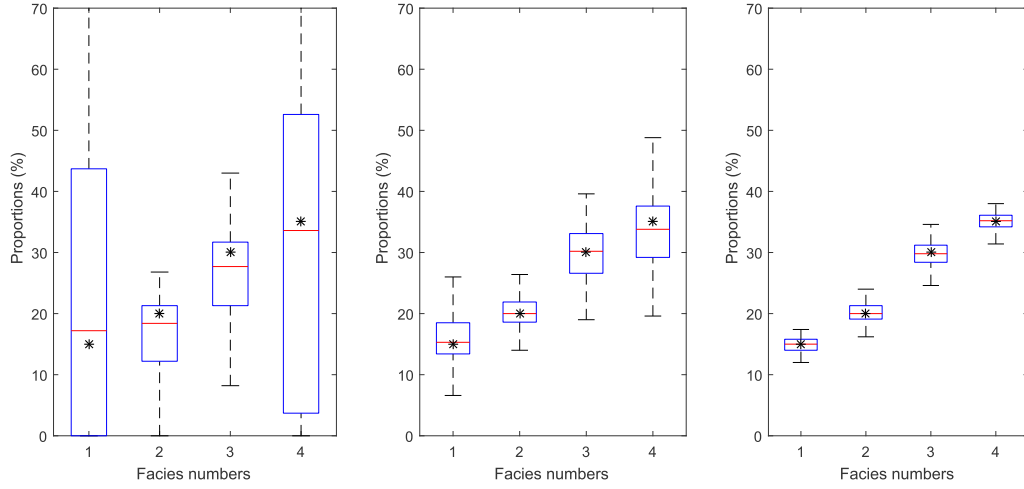


Figure 4.2 Boxplots of unit proportions (in %) for NCS (left), CS1 (middle) and CS2 (right) ; TI proportions as \*. Based on 100 realizations of the synthetic model.

Table 4.1 presents statistics C1 and C2 (see Eq. 4.7), for the different scenarios.

Tableau 4.1 Statistics C1 and C2 (100 realizations)

Simulation	Description	C1	C2
NCS	Unconditional	3.05	58.8
CS1	HD (1% TI)	0.06	15.5
CS2	HD (20% TI)	0.01	5.4

The NCS presents the highest values for C1, and C2. Adding a few data randomly (CS1 vs NCS) decreases strongly both statistics values. As expected, adding more data reduces further C2 (CS2 vs CS1).

From this synthetic example, we conclude that the MCP is unbiased. Its 0/1 forcing property is determinant to enforce directional units sequence. The variability of the resulting realizations is controlled by the number of data available. The realizations appear as different versions of the deterministic model, especially in presence of conditioning data.

#### 4.3.4 MCP variability in the Gaussian case

To illustrates expected variability of MCP conditional simulation in a spatial context, we compare the variability of its transitional bivariate probability functions with the multiGaussian case.

We first perform conditional simulation of 100 realizations of a Gaussian variable using

FFTMA algorithm [221] on a 2D simulation grid (250x250 cells) using a cubic covariance function with an isotropic range representing 1/4 of the field size. An initial realization was sampled at 251 cells (0.4% of the total number of cells) to provide the conditioning data. Gaussian realizations are post-conditioned by kriging and then truncated using thresholds -0.67 and 0.52 to define three units with theoretical proportions of 0.25, 0.45 and 0.30 respectively.

For each truncated conditional realization we calculated the experimental non-centered cross-indicator covariance that estimates  $E[I(\mathbf{x})J(\mathbf{x} + \mathbf{h})]$ . At each lag, the 100 experimental non-centered covariances and cross-covariances were sorted and used to define 95% confidence interval of bivariate probabilities (blue lines in Fig. 4.3) for the Gaussian case.

Next, we calculated for each lag distance the theoretical bivariate probabilities from the biGaussian law using the same covariance model and unit proportions [216, p. 104]. We used the theoretical bivariate probabilities for MCP simulation with search radii equal to field size and using a maximum of 5 neighbors. From MCP realizations, we calculated the experimental bivariate probabilities for the MCP case (grey lines in Fig. 4.3).

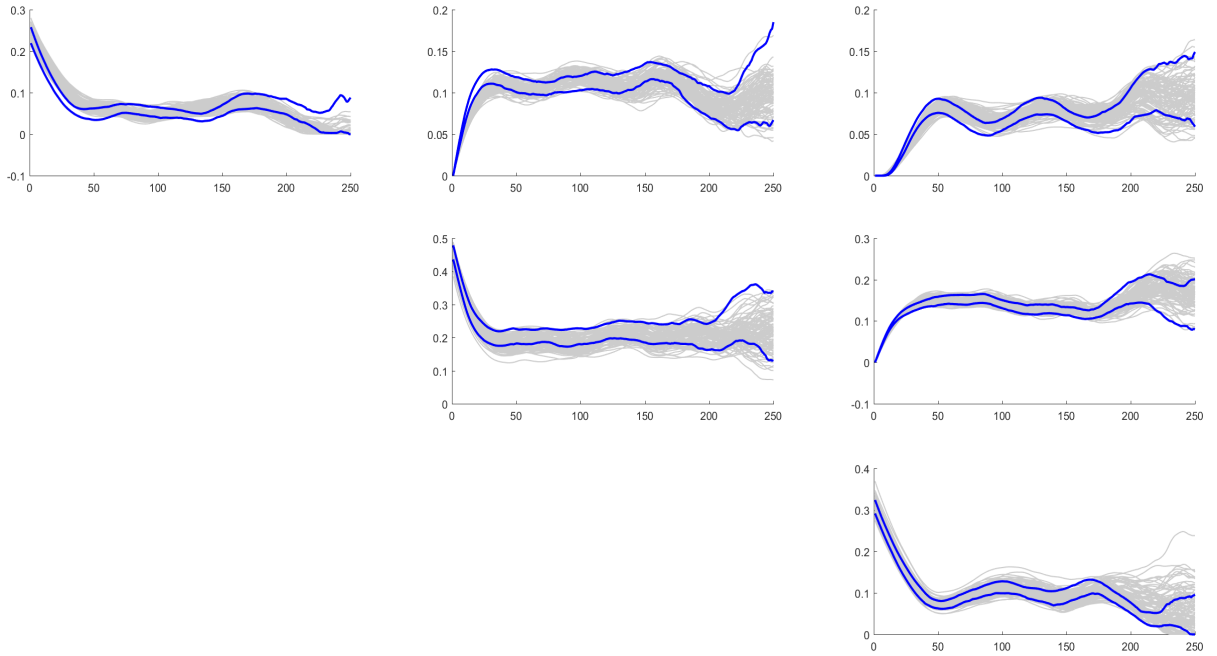


Figure 4.3 Experimental bivariate probability as a function of cell distance for MCP realizations (grey) and 95% confidence interval for multiGaussian simulation (blue). Facies ordered F1, F2 and F3 from left to right and top to bottom.

Figure 4.3 shows a rather good agreement of MCP bivariate probabilities with those obtained in the Gaussian case. There is no strong bias for probabilities and the variability of the probabilities match well the confidence intervals obtained from exact Gaussian simulations. This example illustrates that MCP is able to reproduce variability comparable to the one obtained from an exact Gaussian method in the conditional Gaussian case. Admittedly, results can vary with the number of conditioning points used.

#### 4.3.5 Sensitivity to neighborhood selection

To assess the effect of choices related to the main simulation parameters (number of neighbors and search radius), a sensitivity analysis is performed on the synthetic case shown in Fig. 4.1 based on C1 and C2 criterion values. One hundred conditional realizations with 10 HD were run. In the first test, the search radius (dmax) covers the full extent of the domain and maximal number of neighbors (nmax) varies (Fig. 4.4-top left). In the second test, nmax is set at 10 with increasing dmax (Fig. 4.4-top right). In the third test, nmax is set at 10, dmax covers the full extent of the domain and number of HD increases (ndata). In the fourth test, nmax is set at 10, dmax covers the full extent of the domain, the number of HD is 10 (ndata) and number of realizations increases. In these four tests no other constraints are applied. Hence, multigrid, correction for inversion, search by quadrant, and minimal number of neighbor are all deactivated.

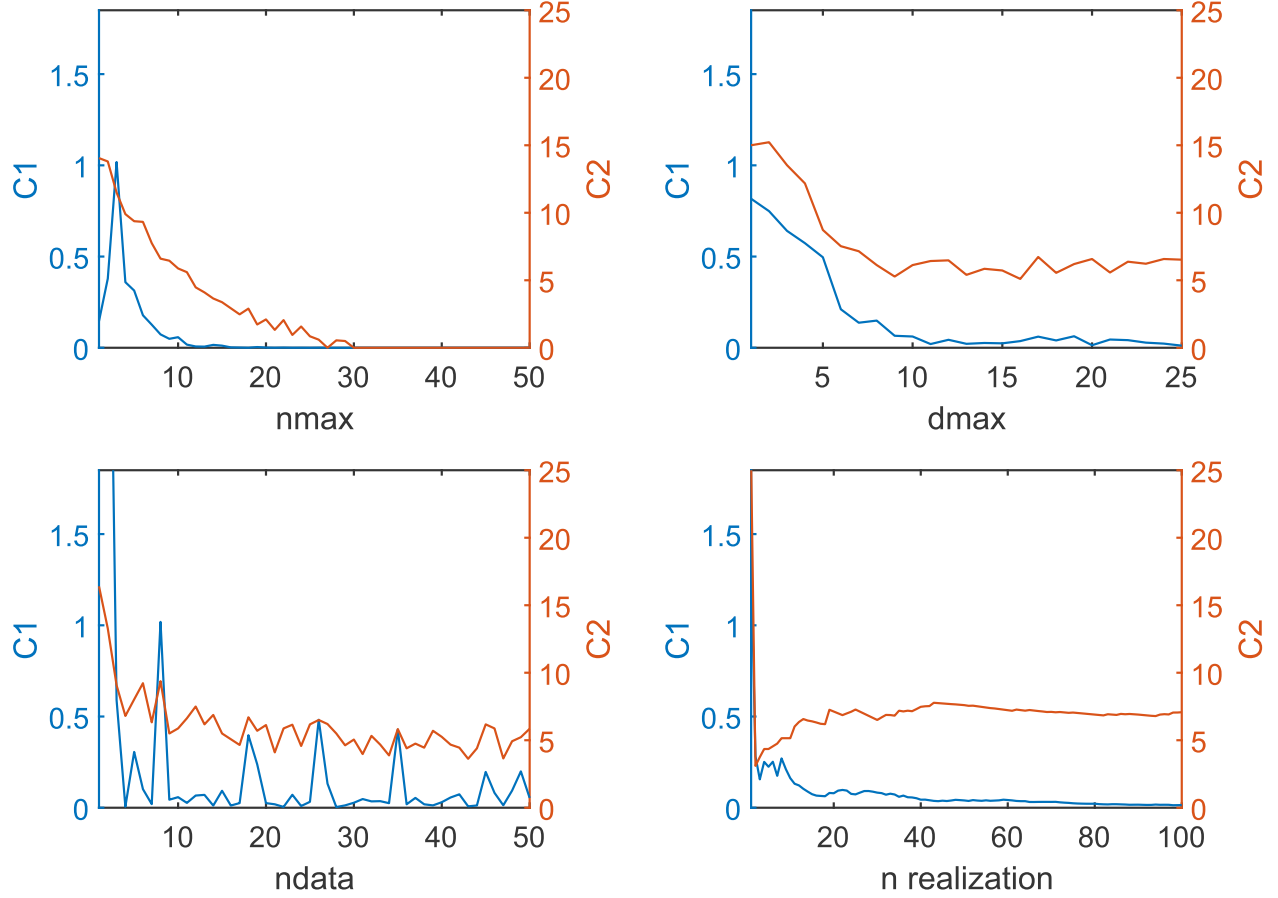


Figure 4.4 Top left C1 and C2 as a function of  $n_{\max}$  using  $d_{\max}$  infinite and  $n_{\text{data}}=10$ ; Top right C1 and C2 as a function of  $d_{\max}$  with  $n_{\max}=10$  and  $n_{\text{data}}=10$ ; Bottom left C1 and C2 as a function of  $n_{\text{data}}$  with  $n_{\max}=10$  and  $d_{\max}$  infinite; Bottom right C1 and C2 as a function of number of realizations with  $n_{\max}=10$  and  $d_{\max}$  infinite.

Figure 4.4-top left shows that C1 and C2 decrease to zero as  $n_{\max}$  increases. With numerous neighbors, the data event becomes so constraining that a single category at simulation point is observed in the TI (C2 goes to zero). On the opposite, a lack of neighbors does not bring enough constraints to reproduce correctly the TI proportions, so C1 is high. For this particular example, values of  $n_{\max}$  between say 5-20 could be envisaged as they allow adequate reproduction of TI proportions without eliminating totally the variability.

Figure 4.4-top right shows globally fluctuations around the levels observed for C1 and C2 in the Fig. 4.4-top left for  $n_{\max}=10$ , except for small  $d_{\max}$  where both C1 and C2 are higher. When  $d_{\max}$  is small, it is likely that less than  $n_{\max}$  neighbors are found, so the simulation behaves like with a smaller  $n_{\max}$ . This indicates that results are rather robust to the choice of  $d_{\max}$ . As soon as  $d_{\max}$  is not set too small, one can expect a range of values for  $n_{\max}$  where C1 is simultaneously small and C2 not close to zero. This range is expected to be case

dependent, so the selection of an acceptable  $n_{max}$  may require a few trials.

Figure 4.4-bottom left shows that C1 and C2 results are robust to the number of HD initially available as soon as  $n_{data} \geq 10$ . C1 is close to zero and C2 decreases slowly with  $n_{data}$ . So the variability obtained is more directly related to the number of neighbors used in the simulation ( $n_{max}$ ) than the number of initially available HD ( $n_{data}$ ). This result is rather expected as each simulated cell is added sequentially to the list of available data. As the initial HD represent only a small portion of cells present in the simulated field, the neighborhood becomes quickly composed mostly of previously simulated points.

Figure 4.4-bottom right shows convergence of C1 to zero and stabilization of C2 as the number of realizations increases. Note that in this case for each realization a new drawing of HD is done so as to eliminate the effect of any particular drawing.

#### 4.4 Case study - MCP simulation of a Basin glacial stratigraphy

We apply the MCP method to a complex deterministic model comprising 11 HSU with strong vertical ordering. After introducing the study area and the deterministic model, five scenarios for simulation by MCP are considered and their results compared.

##### 4.4.1 Hydrostratigraphic setting : Simcoe County area

The 3D TI represents a small portion of a larger deterministic model describing the regional hydrostratigraphy of South Simcoe County in South-central Ontario, Canada (Fig. 4.5). The deterministic model and the accompanying data were provided by the Ontario Geological Survey.

The 11 HSU show proportions varying from 0.25% to 27.28%. The first two letters of the HSU label refer to sand and gravel aquifers (AF) or silty and clayey aquitards (AT). From bottom to top, the hydrostratigraphy comprises : Ordovician limestone bedrock (23.24%) and pre-Middle Wisconsin stratified sand and gravel aquifers (AFF1 ; 0.25%) and aquitards ATE1 (12.19%). The overlying Thorncliffe Formation consists of interbedded aquifers (AFD4 : 2.00% ; AFD1 : 2.90%) and aquitards (ATD2 : 1.43%, ATD1 ; 27.28%). Stratified deposits of AFD4 unconformably overlies ATE1 which in this area occurs as a fine textured diamicton. Newmarket till (ATC1 ; 4.00%) caps the sedimentary sequence and acts as a regional aquitard. The till is eroded locally and has been traced down into broad valley systems interpreted to be tunnel channels. Glaciofluvial sand and gravel aquifers (AFB2 ; 2.45%) occur at the base of the tunnel channels and are equivalent in age to early Oak Ridges moraine deposits. AFB2 includes gravelly and sandy sediments at the base and is capped by a sequence of glaciolacus-

trine silt, clay and subordinate sand (aquitard ATB2 ; 22.27%). Shallow regressional deposits of sand and gravel (AFB1 ; 1.99%) completes the stratigraphic sequence. The HSU thickness is highly variable (Fig. 4.6). Each HSU is composed of one or more successive stratigraphic sub-units (hydrofacies) with similar hydrogeologic properties determined at reservoir scale.



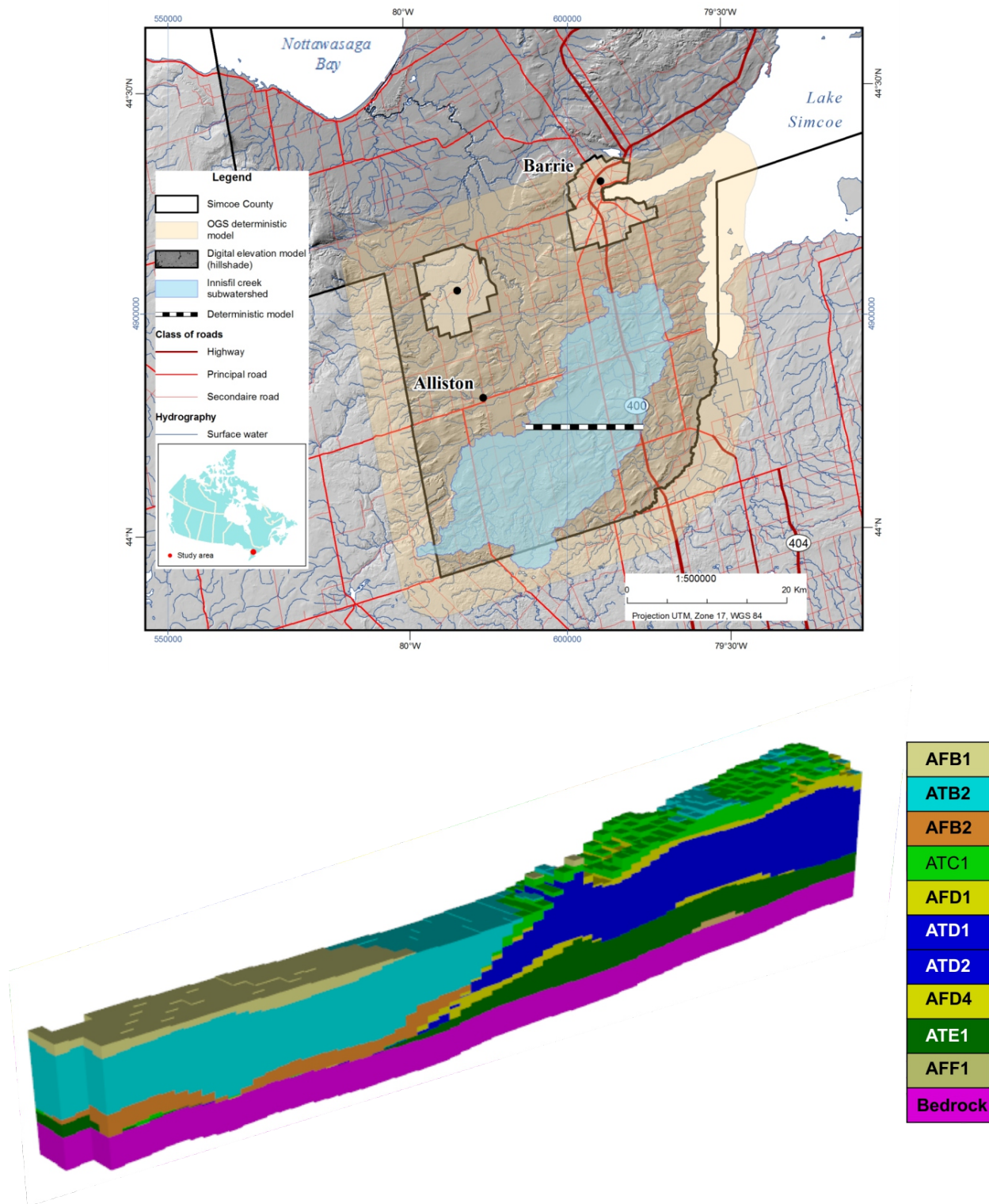


Figure 4.5 Top - Location map of the deterministic TI case study in South Simcoe County; Bottom - 3D view of the deterministic TI (vertical exaggeration 15x).

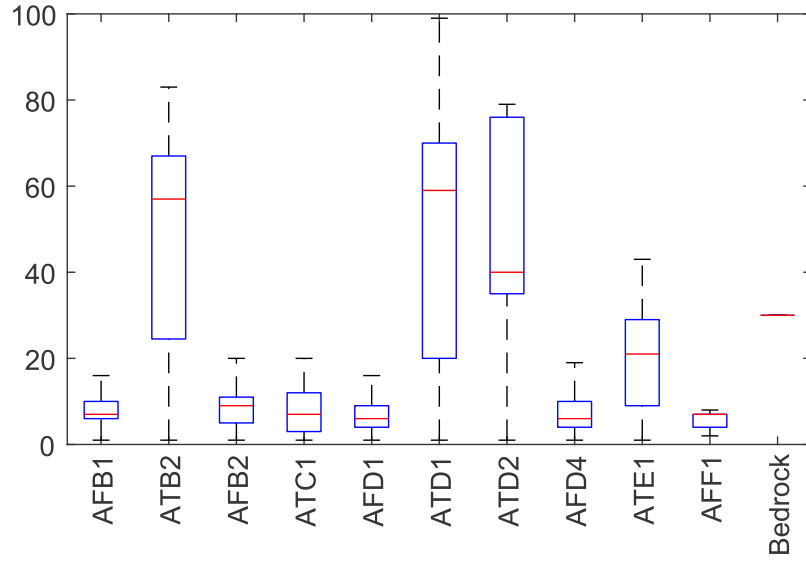


Figure 4.6 Boxplot of HSU thickness (m) within the TI.

#### 4.4.2 Simulation scenarios and parameters

Five different scenarios are compared. The first scenario is a non-conditional simulation (NCS). It uses only the TI in Fig. 4.5. The four other scenarios (CS1 to CS4) are conditional to information depicted in Fig. 4.7 and described in Table 4.2.

Two types of HD are used. The first one comes from continuously sampled high quality boreholes. The second type represents picks (3D points) interpreted by the geologist from low quality borehole logs from public domain database. The picks are used to control unit geometry. Each pick corresponds to the top of a certain HSU. The HSU immediately above a pick is considered unknown although it must obviously be a younger unit. In addition to HD, pseudo-data are used in some of the scenarios. The sources of pseudo-data are : i) the hydrostratigraphic units derived from surficial geology mapping, ii) the bedrock interpolated surface and iii) random points selected within the deterministic model. In the latter case, a new drawing is done for each realization.

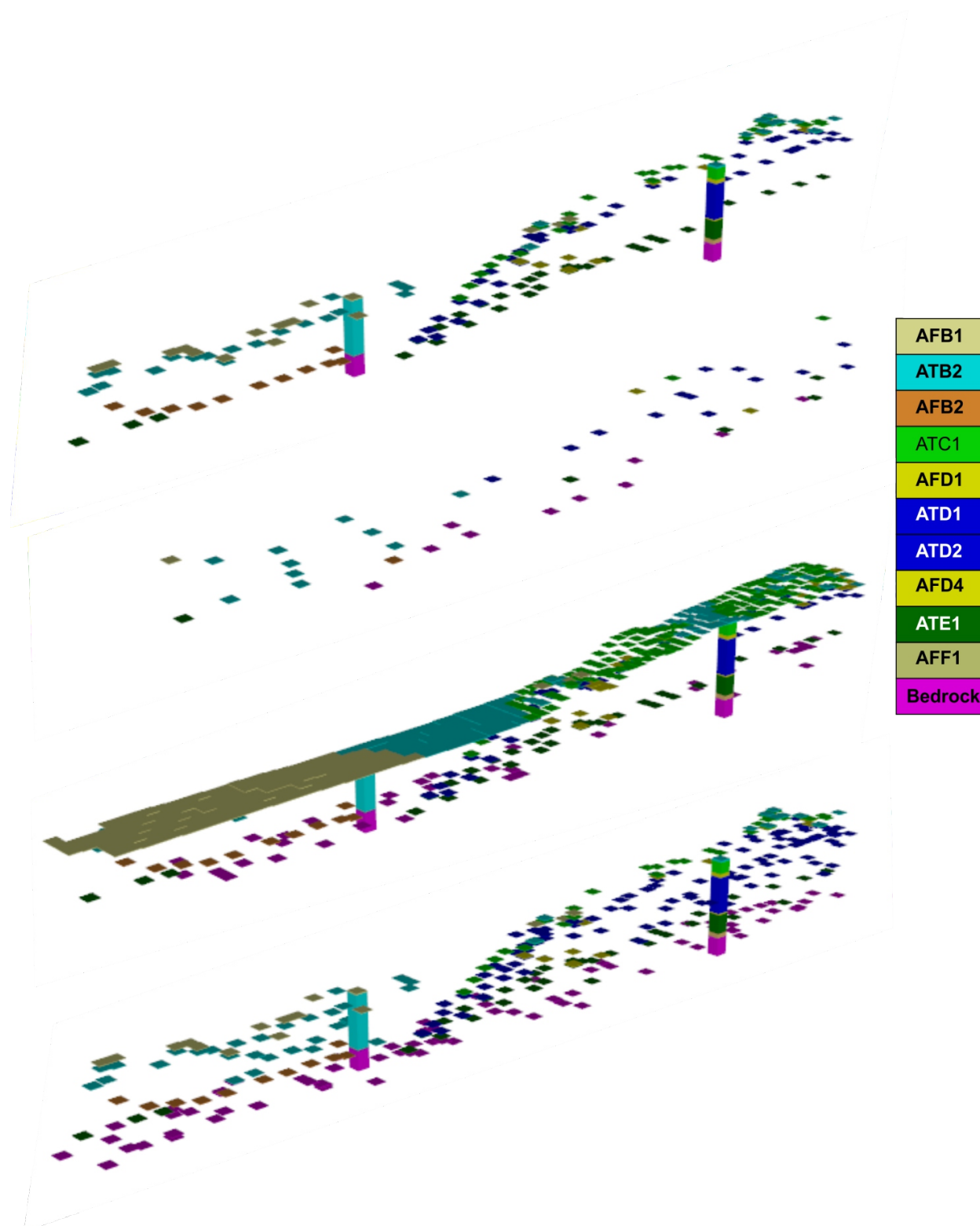


Figure 4.7 Data for the four conditional simulations, from top to bottom : — CS1 : 176 picks and 2 boreholes — CS2 : 44 pseudo-data (random sampling of TI) — CS3 : CS1 data + pseudo-data (10% of bedrock top and all surficial HSU) — CS4 : CS1 data + pseudo-data (0.3% of cells from TI).

The simulation grid is regular. Each cell is 200 x 200 x 1 meters. The deterministic model (TI) and simulated field are defined over the same grid of 76 x 5 x 193 cells. The bivariate probabilities are calculated over the entire 3D TI.

One hundred realizations are produced for each scenario. MCP is applied with search radii of 10x5x25 cells and a maximum of 5 neighbors with at most two neighbors per octant. A multigrid procedure is activated comprising 9 levels. All simulation results are compared to the deterministic model considered as a smoothed interpretation of the ground truth.

### **Non-conditional simulation (NCS)**

The NCS scenario illustrates the capacity of MCP to preserve HSU relationships found in the TI (Fig. 4.8). The directional HSU sequence is respected (Fig. 4.8b) due to the zero-forcing property. The spatial variability of the HSU is larger close to HSU contacts (Fig. 4.8c). The TI HSU proportions are well reproduced (Fig. 4.8d) by the MCP realizations, even for units representing very low proportions. The proportion variability is large and approximately proportional to the TI proportions as expected for a non-conditional case.

The differences between the TI proportions and the mean proportions over the realizations are statistically non-significant for 9 of the 11 HSU (we have  $|\overline{pMCP_i} - pTI_i| / \sqrt{pTI_i(1 - pTI_i)/n} < z_{0.05/2} = 1.96$ ), including the four most abundant HSU. The only two HSU showing significant differences, AFD4 and AFF1, are among the least abundant ones.

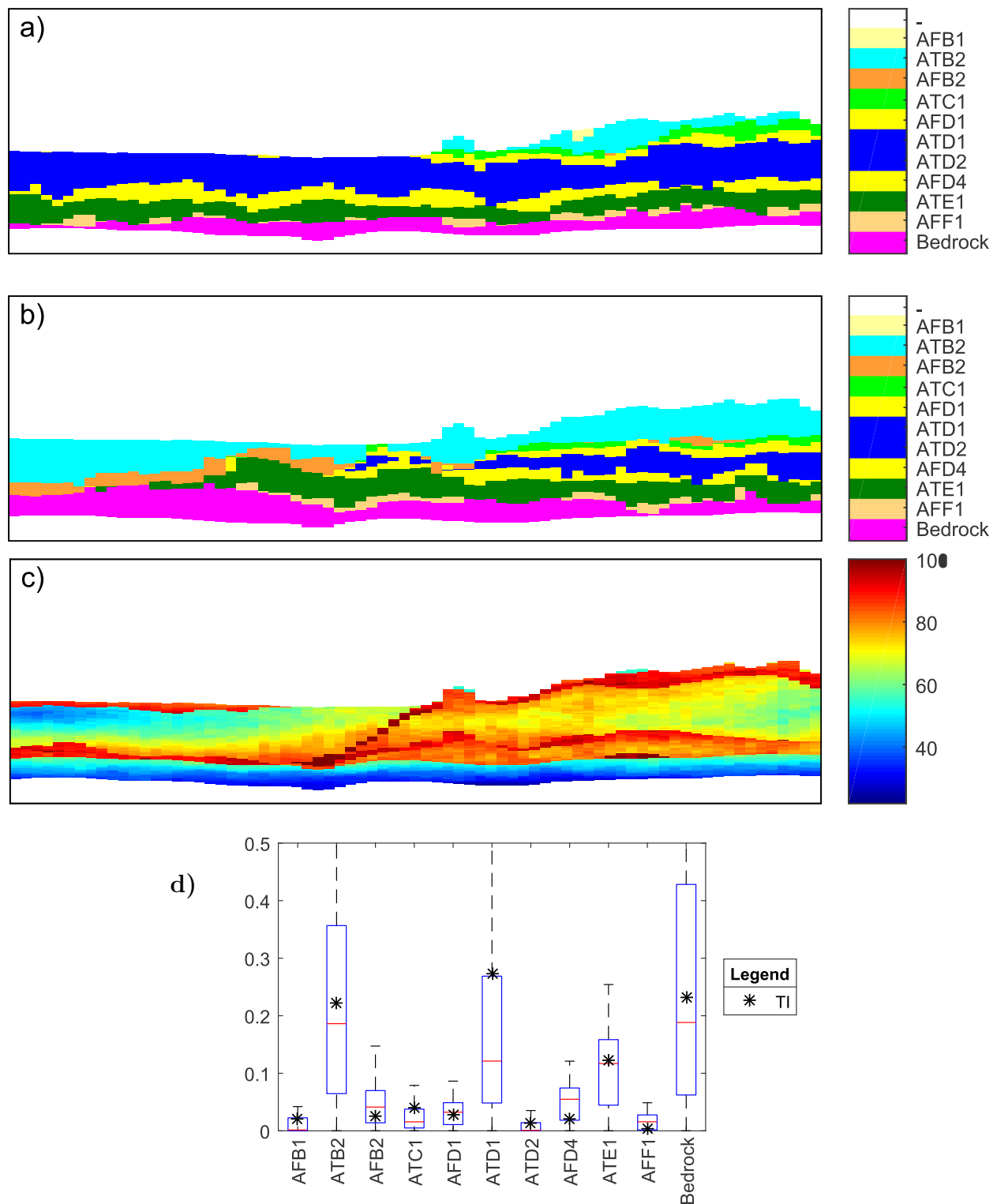


Figure 4.8 NCS : non-conditional simulation scenario ; Cross-section ( $y=3$ ) of two realizations (a and b) ; Dissimilarity c) ; Boxplots of HSU proportions, 100 realizations d).

Fig. 4.9 shows boxplots of HSU thicknesses measured over all vertical scan lines for the 100 realizations. The thickness distributions compares well to the mean and maximum thicknesses observed in the 3D TI for the different HSU.

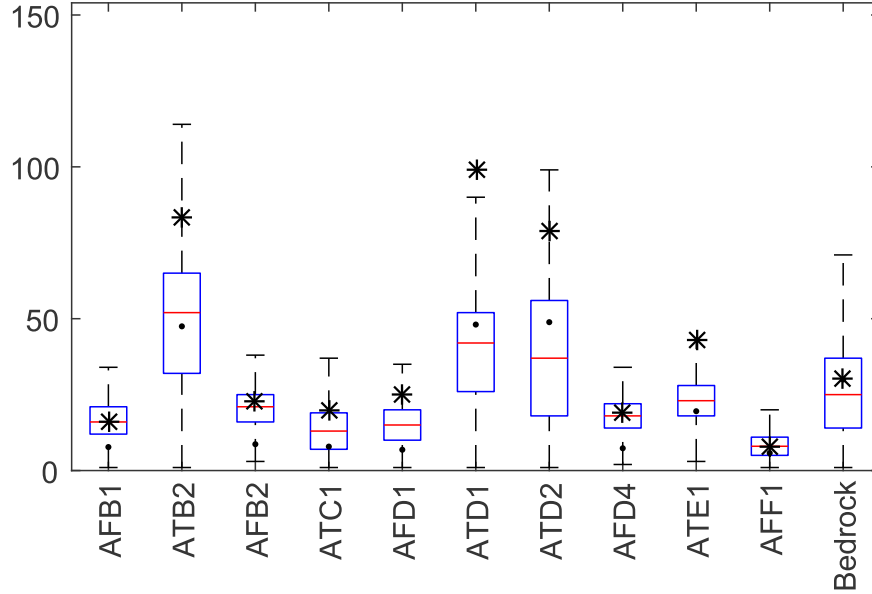


Figure 4.9 Boxplot of unit thickness (m) with NCS (TI mean and maximum thicknesses indicated by dot and star respectively).

### Conditional simulation with HD : CS1

Data used for CS1 scenarios are 176 picks (CS1a) or 176 double picks (i.e. in all 352 picks obtained by adding a pick just above the 176 original picks) (CS1b) and HSU extracted from two high quality continuous drilling [197]. A single pick represents the top of a HSU while a double pick represents the top of a HSU and the bottom of the overlying HSU as identified from the TI. Data locations are illustrated in Fig. 4.7–top.

Figures 4.10 and 4.11 show the strong effect of conditioning data on MCP realizations. Compared to NCS, adding a few data helps reproduce the main TI structures. Uncertainty is larger close to HSU contacts (Fig. 4.10c). HSU proportions variability is greatly reduced (Fig. 4.10d) compared to NCS. The differences between scenarios CS1a and CS1b are subtle, mainly visible in the leftmost part of the sections and on the dissimilarity maps where CS1a shows more variability with respect to the TI. Moreover the abundance of AFD4 is slightly better reproduced with CS1b than with CS1a and the HSU units appear laterally a bit more continuous.

We note that many HSU have mean proportions in the realizations different from the TI.

For the three most abundant HSU (ATB2, ATD1 and bedrock) the mean proportions are less in CS1 realizations than in TI. Considering only the picks (i.e. excluding borehole data), the ATB2, ATD1 and bedrock HSU were also less abundant than in the TI (respectively 17.1%—22.5%, 23.5%—27.1%, 0.5%—24% for ATB2, ATD1 and bedrock picks—TI). Hence, in the initial steps of MCP a simulated point remote from the boreholes will find fewer ATB2, ATD1 and bedrock data in its neighborhood than present in the TI. This could explain the observed disparity between MCP and TI HSU proportions.

Figure 4.12 shows four realizations obtained with CS1b of AFD4, ATE1, AFF1 and Bedrock HSU sub-ensemble. Notice variations in thickness of ATE1 and AFF1 HSU across the realizations. ATE1 has many direct contacts with the bedrock in the fourth realization but almost none in the second one. This could illustrate erosional events or local discontinuities. The 0/1 forcing capability of MCP preserves nevertheless the upward depositional sequence.

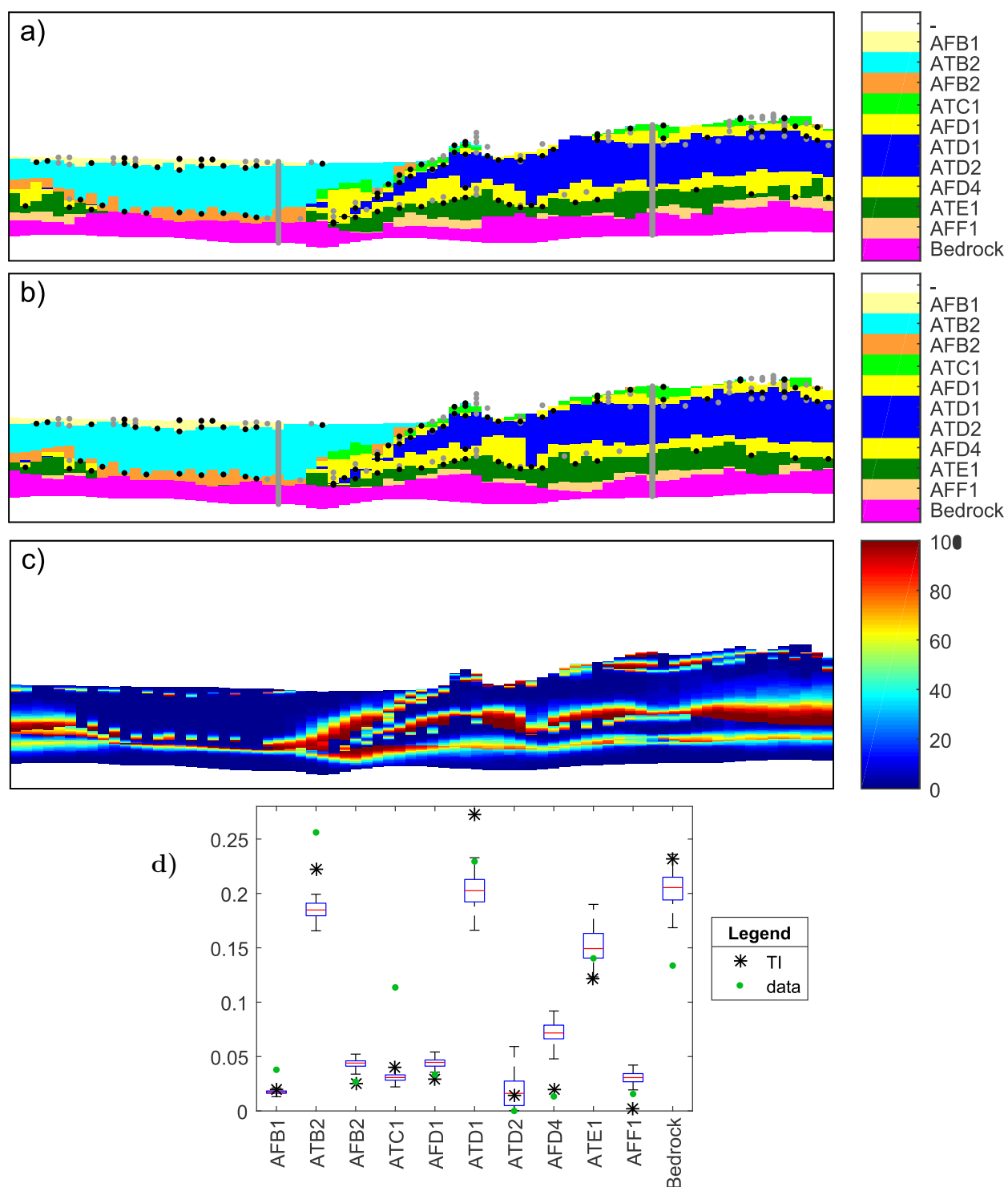


Figure 4.10 CS1a : scenario with HD (2 boreholes and 176 picks) ; Cross-section ( $y=3$ ) of two realizations (a and b) ; Dissimilarity c) ; Boxplots of HSU proportions, 100 realizations d).



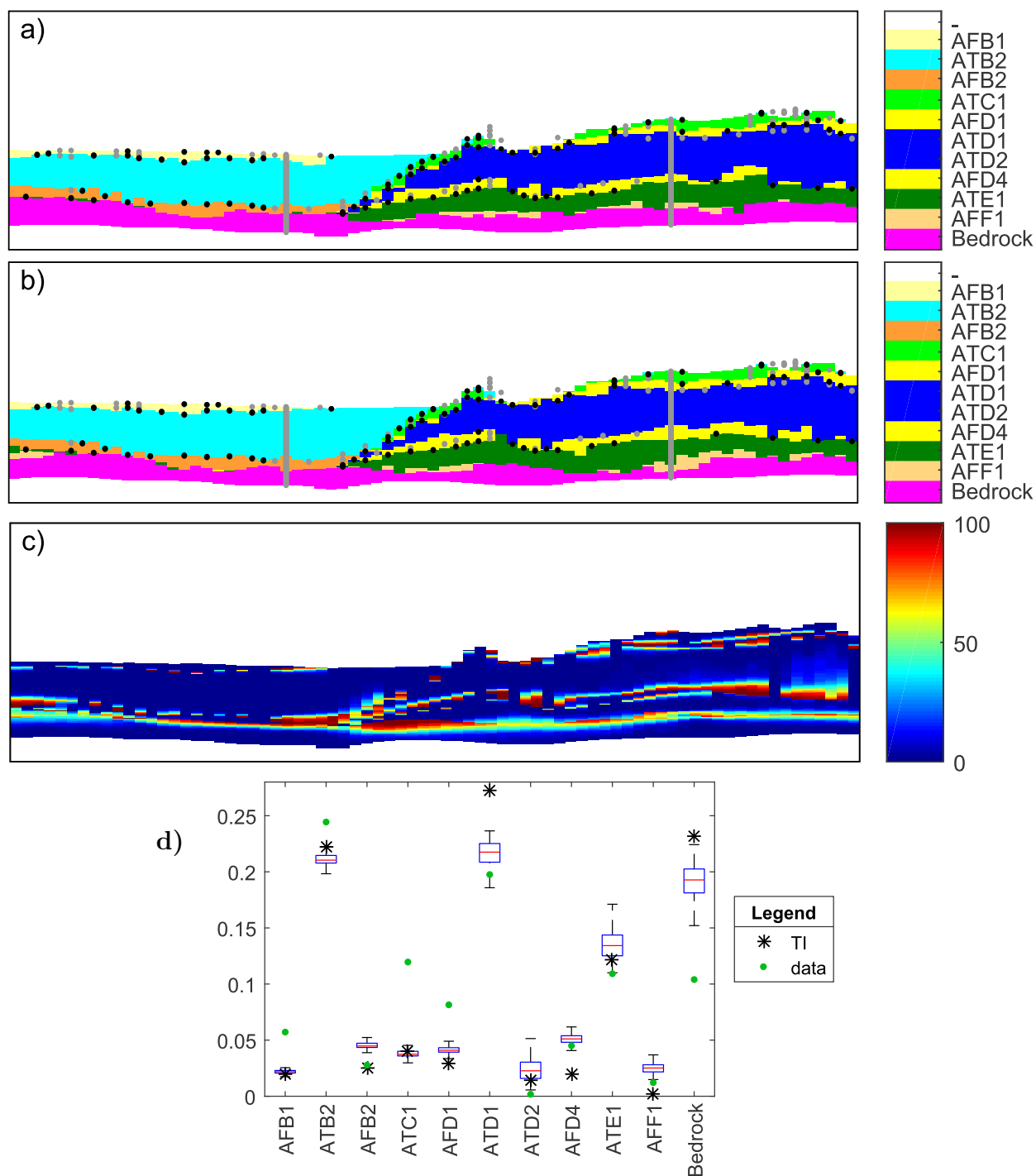


Figure 4.11 CS1b : scenario with HD (2 boreholes and 176 double picks) ; Cross-section ( $y=3$ ) of two realizations (a and b) ; Dissimilarity c) ; Boxplots of HSU proportions, 100 realizations d).

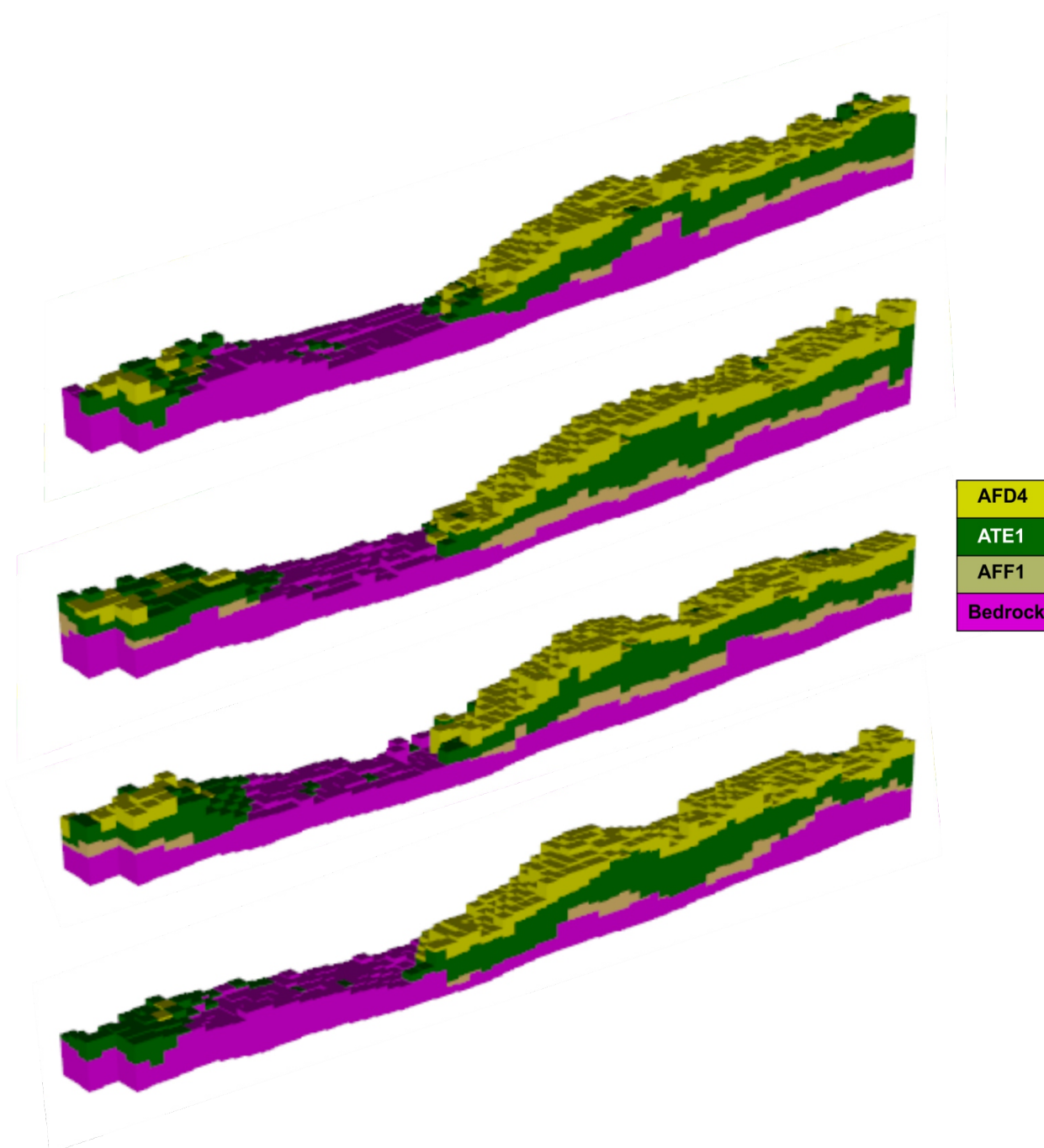


Figure 4.12 CS1b : four realizations of the stratigraphically lowest HSU : bedrock, AFF1, ATE1, AFD4.

### Conditional simulation with pseudo-data : CS2

In this scenario we do not use the available HD. Instead, 44 pseudo-data are extracted from the deterministic model by random sampling. Hence, the HSU proportions in pseudo-data and TI are similar. The comparison of results for the "balanced" case CS2 with the "unbalanced" case CS1a aims to assess the influence of differences between data and TI HSU proportions on global HSU proportions obtained in realizations.

Figure 4.13 shows the results for CS2. One can still observe differences between HSU proportions in the simulation and TI for AFB2, ATD1, AFD4, AFF1 and Bedrock, but the differences are globally reduced compared to CS1. Having conditioning data more evenly spread in the TI helps reduce the observed discrepancies in HSU proportions. However, it is not clear whether or not the observed proportion differences are statistically significant for both CS1 and CS2 when compared to proportions in TI.

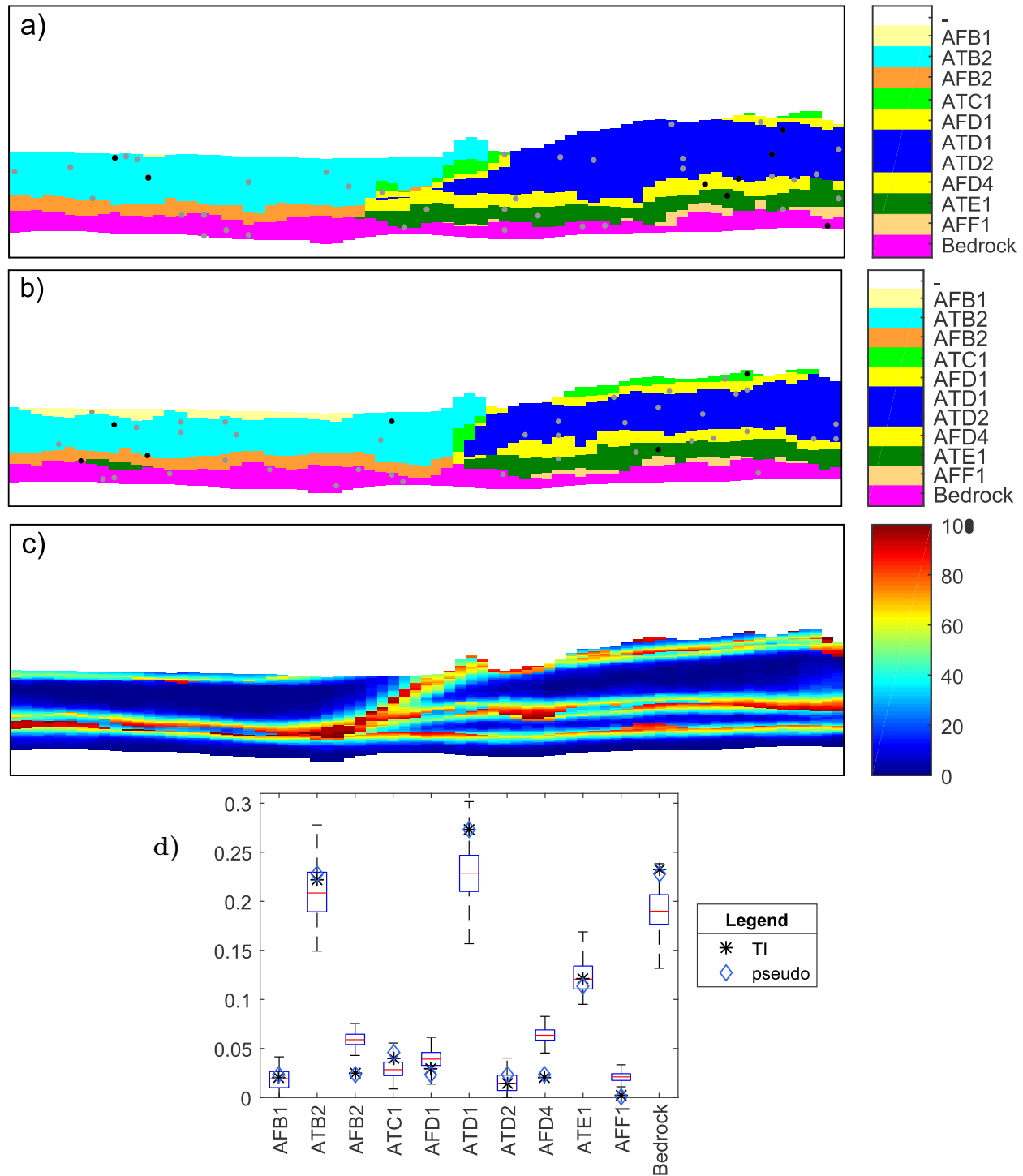


Figure 4.13 CS2 : 44 pseudo-data (random sampling of TI). Cross-section ( $y=3$ ) of two realizations (a and b); Dissimilarity c); Boxplots of HSU proportions, 100 realizations d).

### **Conditional simulation with data and pseudo-data : CS3**

One source of commonly available geological information is the surficial geology for which reliable maps exist. A second source is the bedrock topography which can be interpolated reliably given the availability of all the boreholes archived in databases. These two sources can be used to provide pseudo-data that help constrain the simulated models.

Figure 4.14 presents the results of conditional simulation using pseudo-data from surficial geology (all 324 cells) and bedrock (37 cells) in addition to the HD set (see Fig. 4.7c). The most striking effect of the added pseudo-data is the reduction of variability observed on the boxplots compared to NCS and CS2.

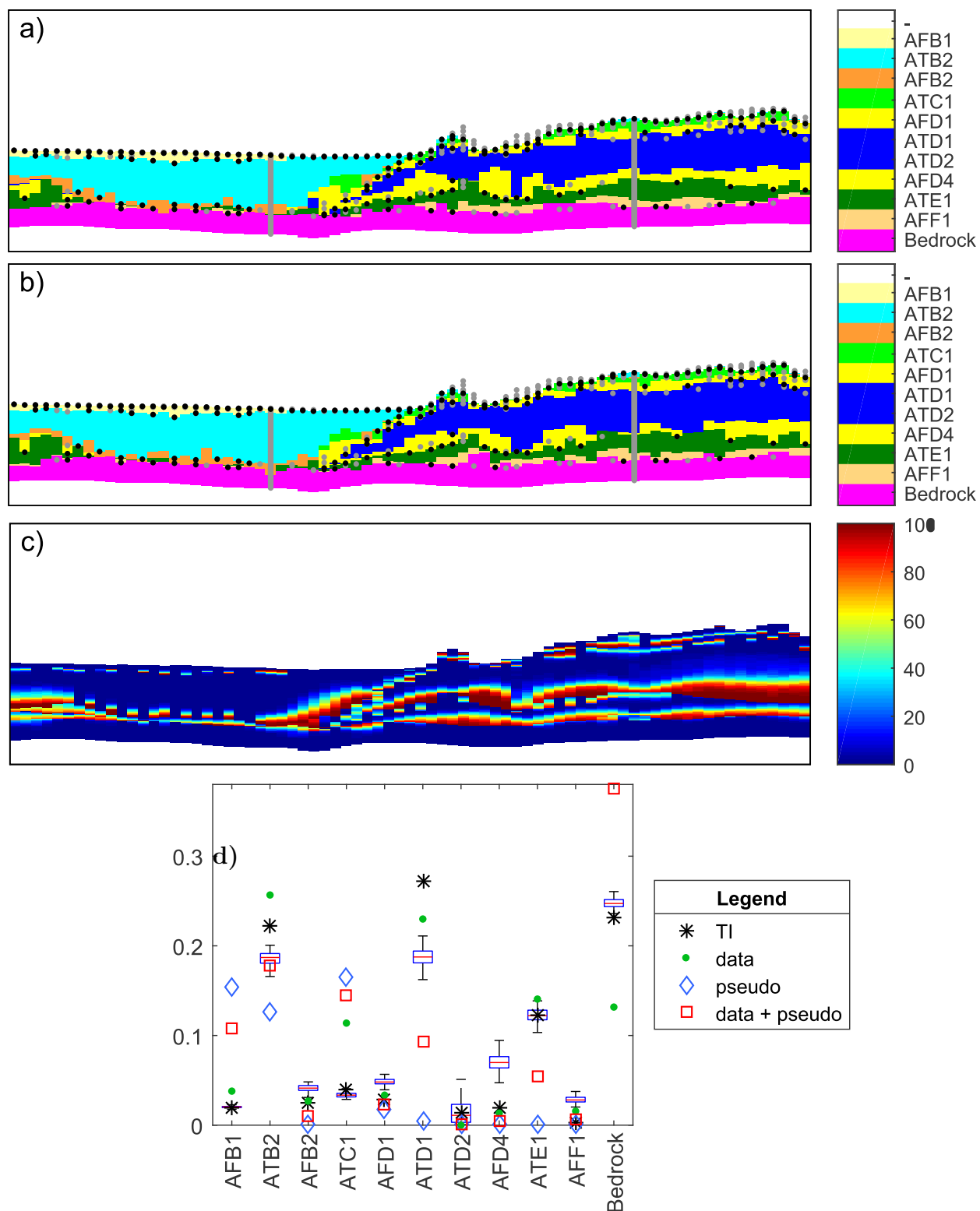


Figure 4.14 CS3 : scenario with HD (as CS1) and pseudo-data (37 bedrock cells and 324 surficial cells). Cross-section ( $y=3$ ) of two realizations (a and b) ; Dissimilarity c) ; Boxplots of HSU proportions, 100 realizations d).

### **Conditional simulation with HD and pseudo-data from TI : CS4**

As indicated before and as used in CS2, the deterministic model can itself be a source of pseudo-data. The added pseudo-data from the TI helps to control the HSU proportions in the simulated models. However, their number must be kept small to preserve variability of the realizations.

Figure 4.15 presents an example using a small proportion of the TI (0.3%) obtained from a random sampling that adds 143 pseudo-data to the available HD. As with CS3, the main effect of added pseudo-data is to diminish strongly the variability of proportions among realizations.

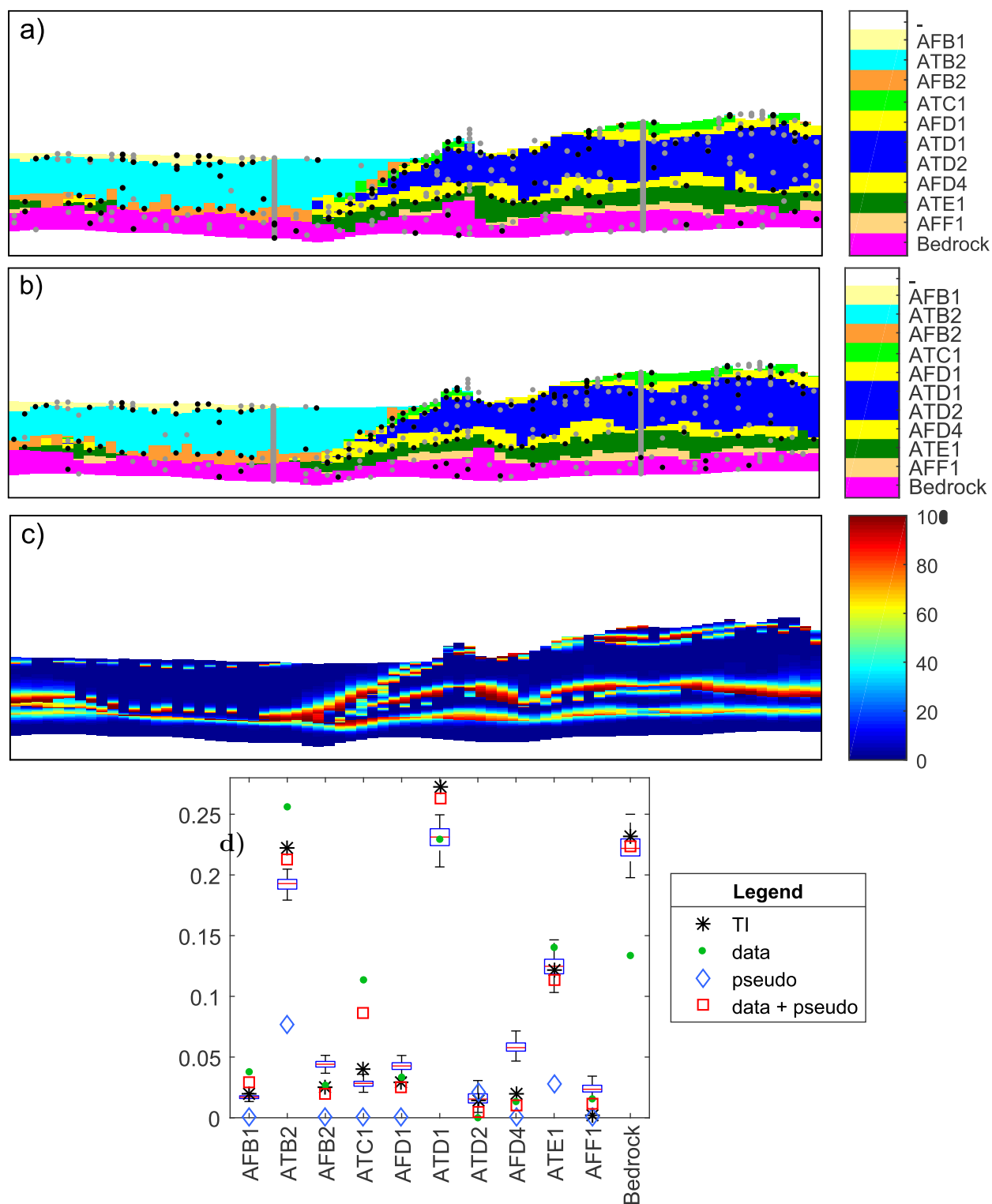


Figure 4.15 CS4 : scenario with HD (as CS1) and 143 cells obtained from random sampling of TI). Cross-section ( $y=3$ ) of two realizations (a and b) ; Dissimilarity c) ; Boxplots of HSU proportions, 100 realizations d).



#### 4.4.3 Statistics of the different scenarios

The criteria C1 and C2 are shown in Table 4.2 for the five different scenarios. The NCS scenario has the most variability, it presents the largest C2 values. Adding data (CS1) or pseudo-data (CS3) unevenly distributed in the simulation field affects slightly the reproduction of HSU proportions compared to a more balanced sampling as shown by the higher C1 values for CS1 and CS3 compared to NCS, CS2 and CS4. Adding more data reduces the variability as indicated by statistic C2 lower for CS4 compared to CS2 and CS1, and lower for CS3 than for CS1.

Tableau 4.2 Statistics of 100 realizations for each scenario. See Eq. 4.7 for definition of C1 and C2 (<sup>1</sup>random sampling of TI)

<b>Simulation</b>	<b>Description</b>	<b>C1</b>	<b>C2</b>
NCS	Unconditional	5.8	25.3
CS1a	HD (2 boreholes + 176 picks)	7.9	3.1
CS1b	HD (2 boreholes + 176 double picks)	5.1	2.4
CS2	PD (44 cells <sup>1</sup> )	5.8	4.7
CS3	HD (CS1) + PD (37 bedrock <sup>1</sup> + all 324 surficial cells)	7.5	2.3
CS4	HD (CS1) + PD (143 cells <sup>1</sup> )	4.9	1.8

#### 4.4.4 Illustration of local proportions variability

Maps of local proportions are computed over the TI and the realizations to help visualizing the reproduction of local trends (Fig. 4.16). Maps of average and standard deviations of the local proportions over the 100 realization are also included. The maps represent the three most abundant units : ATB2, ATD1 and ATE1 for scenario CS1a. The results show that TI local proportions are well reproduced by MCP realizations and variability occurs close to HSU locations identified in the deterministic model.

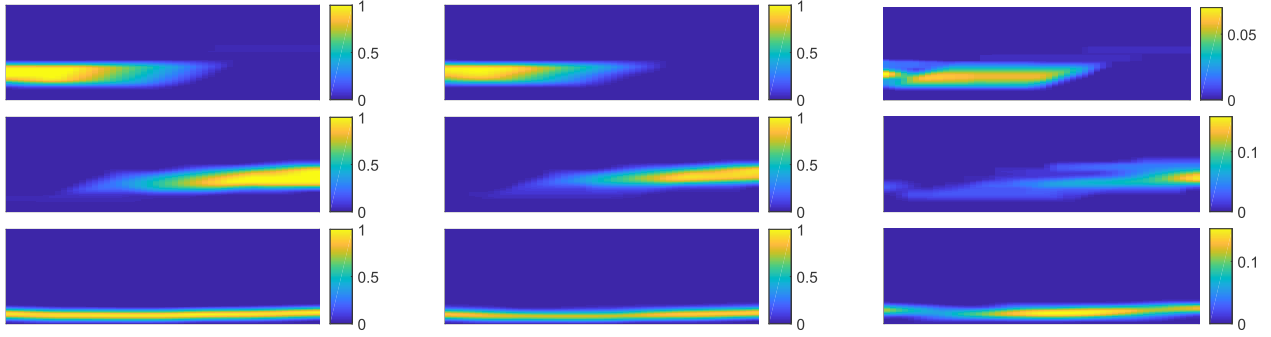


Figure 4.16 CS1a : Local HSU proportions within 25x25 moving windows over central vertical section. First column : local mean in TI, second column : mean of proportions (over 100 realizations) ; third column : standard deviation of the local proportions across realizations. From top row to bottom row : three most abundant units, ATB2, ATD1 and ATE1.

## 4.5 Discussion

MCP has been proposed recently as an efficient alternative to the computationally intensive BME for the simulation of fields showing complex unit arrangements and non-symmetric transition probabilities as are often found in sedimentary environments. It is based on the rather strong conditional independence hypothesis that assumes the unit at data points are independent once the unit at the estimation point is known. This assumption is more likely to be valid when working with small neighborhoods and conditioning data points evenly dispersed around the estimation point. Hence, we selected a localized octant search with a maximum of two points per octant and a maximum of 5 points over all octants.

Results by [99] on a series of small datasets show no significant differences between BME and MCP. The same authors provide theoretical justifications for the use of MCP. However, to our knowledge, the method has never been tested on a complex model based on real geological data such as Simcoe hydrostratigraphic system which comprises 11 different HSU having quite different proportions and a clear directional control.

The 0/1 forcing property of MCP is determinant in reproducing the directional trends. This property can deal with transitional cases, like those presented in the synthetic and Simcoe County examples, without the need to impose directional trend with auxiliary fields as with PGS. Also, it is not necessary to segment the studied field into homogeneous subdomains. The field can be handled globally at once by MCP, which greatly facilitates the analysis.

The most practical and influential aspect in the application of MCP is the definition of the neighborhood. We obtained better results visually and as measured by the statistics C1 and C2 when using octant search and a multigrid approach. The number of neighbors and extent

of the search should be large enough to avoid unit inversions in the simulated field, but small enough to favor variability between realizations. Good compromises between proportional reproduction and variability were obtained with five neighbors for both the synthetic and the Simcoe County test cases. Using more neighbors generally decreased too much C2. D’Or et al. [103] reported that five neighbors was enough in BME to stabilize the C1 criterion. Our findings in this 3D case confirm those of [99] that the number of neighbors need not to be large provided the neighbors are well spread around the simulated point.

Despite the care employed in the definition of the neighborhood, unit inversions can still seldom occur due to the peculiarities of the available HD and previously simulated points found in the neighborhood. Inversions can be corrected on the fly in the sequential simulation. The procedure consists to verify whether the candidate HSU at a simulated cell satisfies the vertical hydrostratigraphy depicted by the deterministic model. It compares the simulated cell with conditioning data and previously simulated cells in the same column. When in conflict with the hydrostratigraphy, the candidate HSU is discarded and the cell returned at the end of the list of remaining cells to simulate. In the Simcoe County test case, less than 0.3% of the points presented unit inversions, a small number considering that 11 units are present.

When the assumption of conditional independence is valid, MCP is unbiased by construction as one is then drawing sequentially from the true conditional distributions. When the conditional independence assumption does not hold completely then one is not drawing from the exact conditional distributions and nothing firm about the bias can be stated. However, our experimental results for the synthetic case (Fig. 2) clearly show absence of any substantial bias in both conditional and unconditional situations. In the more complex Simcoe test case example, Figs 6d, 8d, 9d, 11d, 12d, and 13d indicate that the bias is less for the unconditional case (Fig. 6d) and for the random sampling of conditioning data (Figs 11d and 13d), then for cases where a preferential sampling of HD is present (Figs 8d, 9d, 12d). However, even in the cases of preferential sampling the bias between TI proportions and simulated proportions remains smaller (Fig. 8d) or even much smaller (Fig. 9d, and 12d) than the bias observed in the preferential sampling (see Fig. 12d). The over-representation of bedrock and surficial HSU in CS3 compared to CS1 had no significant impact on the average simulated proportions. The additional data did reduce variability as expected (C2 smaller) without further distorting the average proportions (C1 comparable). Admittedly, whether these good results about bias and robustness to preferential sampling can be extrapolated to other different complex cases remains to be verified.

In the Simcoe County test case, the bivariate probabilities required for MCP were obtained

by direct computation on the deterministic model. This model summarizes all the geological knowledge and available data in the area. This justified the use of TI as a source of pseudo-data to increase control of the unit proportions. Proportions of the different HSU in the deterministic model were assumed representative of the ground truth. In other applications or in earlier stages of application, the deterministic model could be unavailable. In these cases, a simpler conceptual model can be used as TI to provide the bivariate probabilities. However, no pseudo-data should be sampled directly from the conceptual model as the unit proportions and locations in the conceptual model are only loosely known. Pseudo-data from surficial geology or bedrock can still be used however, but one would be prudent to verify the robustness of simulated unit proportions relative to the quantity of pseudo-data added.

One reviewer, quoting the usual practice in MPS studies to borrow the TI from an external source to represent texture/structure, raised an important concern about the idea of using a deterministic model as the source for the TI. Most published MPS studies were done in the petroleum domain where the TI represents essentially borrowed analogous textures from external sources (e.g. satellite images, object-based simulation or process-based simulations). The textures are then conditioned to HD and modulated in space by combining with soft data coming from seismic survey. In other fields of study, such external sources and soft data do not exist. In mining for example each deposit is unique and complex and seismic data are usually not available or simply cannot identify mineralized zones. Possible sources for TIs could be in open pits the blast holes on a few benches located close to the area to simulate. The TI is then expected to be also close to the “target field” [222]. Another source in mining is a deterministic (geological) model [40, 218]. In regional hydrogeology, the quasi-universal practice is to elaborate a deterministic model integrating all available information : sedimentary environment and geological context, known stratigraphy, water ages, geomorphology, surficial geology, few sampled “quality” boreholes, the low quality boreholes registered in governmental data bases, pumping and tracer tests, borehole permeability tests and sample size grading, and few geophysical data sometimes available in part of the area. These soft informations are not of the same nature as seismic data in petroleum studies. They are complex, elusive, often rare and with very partial coverage, and represent a variety of supports, so they cannot be combined directly with a TI representing merely the texture, as in the petroleum case with seismic data. The available information have to be incorporated in the deterministic model thanks to the geologist’s knowledge. Despite all the expertise involved in the design of the deterministic model, it nevertheless remains a quite idealized and smoothed interpretation of the real field. If the stratigraphy and global proportions of each HSU may be well represented in the deterministic model, the HSU vertical thicknesses are likely much more variable than assumed. Similarly, lateral continuity of HSU is probably exaggerated

in the deterministic model. As these thickness and lateral variations might influence significantly the flow response of the model, it is important to consider alternate models around the deterministic one to be able to assess flow uncertainty. This was the goal pursued in this paper.

## 4.6 Conclusion

This study has applied the MCP simulation method to both a synthetic and to the complex stratigraphic succession within a glacial sedimentary basin. The MCP simulation method has succeeded at simulating the complex depositional system involving many units and a clear directional trend inducing asymmetry between units transition probabilities. The method was shown to be unbiased in the non-conditional case. The different realizations obtained by MCP propose different field models around the overly smoothed deterministic model. The alternative models are essential inputs to assess uncertainty on groundwater flow and transport problems.

## Acknowledgments

Constructive comments from two anonymous reviewers were helpful improving the manuscript. In particular, one reviewer suggested to us the idea of comparing MCP to Gaussian simulator as in section 4.3.4. The authors thank A. Bolduc, Y. Michaud and H. Russell for their support from the Groundwater Geoscience Program, Geological Survey of Canada, Natural Resources Canada. Research was partly financed by NSERC (RGPIN-2015-06653).

## Addendum

Pour la figure 4.3, l'axe  $x$  est la distance, l'axe  $y$  est une probabilité.

Pour la figure 4.5, les axes de la figure du haut sont en UTM zone 17 WGS84, celle du bas est une vue 3D dans un repère  $x,y,z$ ,  $z$  étant l'élévation.

Pour la figure 4.6 l'axe  $Y$  est une épaisseur en mètre.

Pour les figures 4.7 et 4.12 la vue 3D est dans un repère  $x,y,z$ ,  $z$  étant l'élévation.

Pour les figures 4.8d, 4.9, 4.10d, 4.11d, 4.13d, 4.14d et 4.15d l'axe  $Y$  est une proportion.

Pour les figures 4.8, 4.10, 4.11, 4.13 4.14 et 4.15 (a à c) la vue 2D est dans un repère  $x,z$  où  $z$  est l'élévation.

Pour la figure 4.16 la vue 2D est dans un repère  $x,z$  où  $z$  est l'élévation. L'échelle de couleur est une probabilité.

# CHAPITRE 5    ARTICLE 2 : STOCHASTIC CORRELATED HYDRAULIC CONDUCTIVITY TENSOR CALIBRATION USING GRADUAL DEFORMATION

Nicolas Benoit<sup>1,2</sup>, Denis Marcotte<sup>2</sup>, and John W. Molson<sup>3</sup>

Journal of Hydrology (2020) (in press)

Submitted online : November 27, 2019

@ Elsevier Amsterdam, Netherlands 2020

## 5.1 Abstract

Quasi-point hydraulic properties ( $K$ ) measured locally under laboratory or field conditions need to be upscaled to block-scale  $K$ -tensors for use in flow simulators. The upscaled model also needs to be calibrated to hydraulic head observations. The calibration must preserve spatial covariance, cross-covariance and non-linear relations between tensor components. We apply a new upscaling method that allows to compute and model the covariance between block  $K$ -tensor components. We use a gradual deformation method for calibration of simulated  $K$ -tensor fields to measured head data. Our method incorporates a new bivariate transform that preserves the non-linear relations between  $K$ -tensor components. The ensemble of calibrated realizations allows quantification of uncertainty of groundwater flow models. A comparison with PEST on a test case defining capture zones for water supply wells shows that our method calibrates better to measured heads than PEST, provides more realistic  $K$ -tensors and results in larger capture zones.

---

1. N. Benoit, Geological Survey of Canada, Natural Resources Canada, 490, rue de la Couronne, Quebec, QC, G1K 9A9, Canada, e-mail : nicolas.benoit@canada.ca

2. D. Marcotte, H. Rezaee, Department of Civil, Geological and Mining Engineering, Polytechnique Montreal, Montreal, Canada

3. J.W. Molson, Département de géologie et de génie géologique, Faculté des sciences et de génie, Université Laval, Québec, Canada

## 5.2 Introduction

The calibration of groundwater flow models by solving the inverse problem is the critical part of a standard modelling workflow [223]. The objective of the inverse problem is to determine the values of the unknown parameters and their ranges (e.g., hydraulic conductivity) ensuring a good reproduction of the state variables (e.g. hydraulic head).

Inversion methods have evolved significantly over the past few decades [8, 147–150]. Inversion by the direct method has been almost abandoned, mainly since it requires the 3D field of hydraulic heads determined at all model nodes [8, 151, 152]. This approach is still valuable when a preliminary flow model is available as a starting point [224]. On the other hand, [154] presented an interesting semi-direct method based on the imposition of a double constraint which allows to calculate the hydraulic conductivity as the ratio of the flux and gradient fields. The current trend of solving the inverse problem is oriented towards the indirect method as shown by a considerable number of studies reported in the literature (e.g. [155–159, 161–166, 225], among many others).

In certain cases the inverse approach is the only option for simultaneously providing an adequate model fit to pressure observations and realistic  $K$  fields [146]. The main difficulty with the inverse problem resides in the non-uniqueness of the solution. It is then referred to as an ill-posed problem [226], for which several possibilities exist to reduce the solution space. For example, one can reduce the number of parameters to be determined, impose constraints on the range of the values of the parameters using *a priori* information, control the fluctuations during the optimization by adding a term of regularization and transform the model so as to minimize the model non-linearity [8].

A different perspective adopted in stochastic approaches is to cope with the non-uniqueness of the solution by proposing a series of calibrated possible solutions. The calibrated solutions reflect the uncertainty about the field state given the current information. However, model calibration for each realization can be tedious for complex models with a spatially varying  $K$ -tensor. Hence, efficient calibration methods must be used. One such method is the gradual deformation method (GDM) [225, 227]. The GDM has been applied to heterogeneous transmissivity fields in 2D in a stochastic context [228, 229], with non-Gaussian distributions [229, 230] and for history matching in 2D [231, 232]. To our knowledge, GDM has never been applied to 3D stochastic  $K$ -tensor fields.

To simplify the analysis and without real loss of generality, we assume the HSU (hydrostratigraphic unit) model itself is considered deterministic and we apply GDM only for calibration of the  $K$ -tensors defined at the block-scale within each HSU. If desired, uncertainty on the

HSU spatial distribution can be accounted for by calibrating the  $K$ -tensors on one realization at a time from a series of realizations obtained by a facies simulation method (e.g. see [15]).  $K$ -tensor components are often strongly and possibly non-linearly cross-correlated. The preservation of these correlations is important. We propose a practical method to maintain the desired cross-correlation between tensor components, within each HSU while applying the GDM. The approach uses a linear model of coregionalization in Gaussian space. A bivariate transform of the Gaussian LCM realizations restores the desired zero-lag non-linear correlations and component histograms.

The paper is structured as follows. First we present the method used to upscale scalar point hydraulic conductivity to 3D block tensors of conductivity. The new method allows to assess the spatial covariance and cross-covariance of Gaussian transformed  $K$ -tensor components. We then describe the simulation of 3D tensors respecting the spatial auto and cross-correlation of the components combined with a bivariate transform aimed at maintaining the non-linear relationships observed between tensor components. We proceed with a brief review of the GDM approach before applying the methodology to a simplified case study. Calibration results are compared to those obtained with PEST. Well capture zones defined with PEST and our method are compared. We complete the paper with discussions and conclusions.

### 5.3 Methodology

The proposed methodology aims to characterize and calibrate a groundwater flow model with correlated  $K$ -tensors. The calibration is performed using GDM. The approach mainly consists of optimizing the tensor components based on observation data (hydraulic head) using a set of  $K$ -tensor realizations. The methodology is divided into three main sections : i) hydraulic block-conductivity tensor simulation, ii) bivariate transform, and iii) calibration with GDM.

#### 5.3.1 Hydraulic block-conductivity tensor simulation

To define the stochastic  $K$ -tensor fields used with the GDM, a multi-step method is proposed herein. It consists of upscaling the spatial distribution of scalar  $K$  from local-scale values (quasi-point) to a block 3D  $K$ -tensor. This is repeated for each HSU. The approach comprises four main steps : (i) local-scale simulation of the scalar  $K$  field ; (ii) upscaling of quasi-point realizations into 3D block  $K$ -tensors using a block-by-block finite element flow simulator ; (iii) definition of the spatial covariance and cross-covariance of the block  $K$ -tensors ; and (iv) direct geostatistical simulation of the block  $K$ -tensors for regional-scale modelling.



### Local-scale simulation of the scalar $K$

The local-scale  $K$  values can be estimated from different sources such as grain size analyses, laboratory tests on borehole samples, specific capacity tests or pumping tests. For geostatistical processing, it is important to use consistent data, preferably from the same type (similar support) to avoid scale issues and to ensure similar levels of measurement error and spatial uncertainty.

For each HSU, the  $K$  values can span several orders of magnitude. The spatial structure of the quasi-point  $K$  is therefore important to assess for block upscaling.  $K$  values within a given HSU are spatially structured according to depositional processes, while the covariance functions describe the spatial structures and are used in the geostatistical simulations. A different point-scalar  $K$  covariance model is determined for each unit.

A series of scalar point-support realizations over local-scale fields are generated for each unit using a non-conditional turning band simulation (TBS) algorithm [233] over a fine grid (see Section 5.4.2). Non-conditional simulation is used as the objective here is simply to determine block-scale  $K$ -tensor covariance functions. The simulations are constrained by the observed quasi-point  $K$  marginal distributions for each unit.

### Upscaling of quasi-point realizations into 3D block $K$ -tensors

For regional 3D numerical modelling, the  $K$  field is represented by block-scale finite elements. The quasi-point  $K$  simulation must therefore be upscaled to equivalent block  $K$ -tensors accounting for the structural connectivity of hydrofacies within each block [140]. Each quasi-point field realization is solved for the governing flow equations along the diagonals and the three principal directions to get a series of block-conductivity tensors following the approach described in [19]. This approach works on sets of 27 blocks ( $3 \times 3 \times 3$ ) corresponding to the central block for which the tensor is computed, with the other blocks forming a skin to diminish the influence of the applied boundary equations. As in [234] the 3D hydraulic conductivity tensors for each block were obtained by forcing flow conditions along 8 different directions using different sets of prescribed head conditions. The mean heads and flux along the principal directions for the central block are then used to compute its  $K$ -tensor. The complete upscaling workflow is presented in Fig. 5.1.

The Saltflow simulator [235] was used to obtain the 3D velocity and head fields. Mean 3D components of flux ( $\bar{q}_x, \bar{q}_y, \bar{q}_z$ ) and gradients ( $\nabla \bar{h}_x, \nabla \bar{h}_y, \nabla \bar{h}_z$ ) were calculated and used with Darcy's law to solve the full 3D  $K$ -tensor.

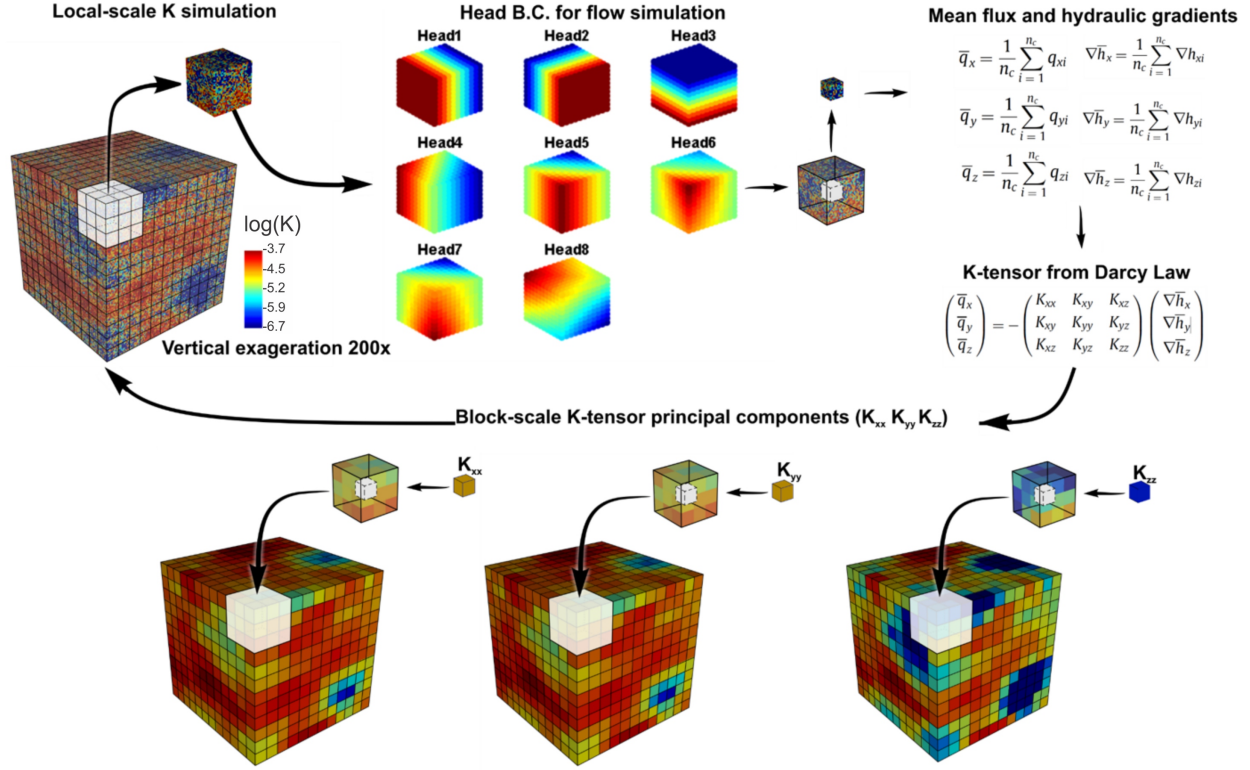


Figure 5.1  $K$  upscaling workflow from  $K$  local-scale realization to 3D block  $K$ -tensors.

### Spatial covariance and cross-covariance of the block $K$ -tensors

The upscaled block  $K$ -tensors are non-Gaussian and present non-linear relationships between tensor components. Hence, they are first Gaussian transformed to determine the linear coregionalization model (LCM) at the block-scale.

The spatial covariance and cross-covariance of the Gaussian transformed block  $K$ -tensor components are computed and modelled by a LCM. To simplify the modelling, components of relatively negligible amplitude are discarded and components uncorrelated to the others are treated separately. The off-diagonal tensor components, being negligible with respect to diagonal terms, were set to 0. Among the diagonal terms,  $K_{xx}$  appeared highly correlated to  $K_{yy}$ . To further simplify the model, and because this makes sense geologically, we assumed  $K_{xx} = K_{yy}$ . So finally the LCM describes two variables,  $K_{horizontal}$  and  $K_{vertical}$ .

### Geostatistical simulation of the block $K$ -tensors

The last step consists of simulating the 3D  $K$ -tensor for the hydrostratigraphic model. Complete fields of Gaussian-transformed block tensors are directly simulated using the LCMs

models of the previous step.

Each realization of block-tensor components is back-transformed using a bivariate transform to preserve the non-linear relationship observed between  $K_{horizontal}$  and  $K_{vertical}$  (see Section 5.3.2).

### 5.3.2 Bivariate transform

The bivariate transformation uses as reference the  $K$ -tensor bivariate distribution for  $(K_{xx}, K_{zz})$  obtained in the initial upscaling phase of the point-support  $K$  scalar fields. The various pairs  $(K_{xx}, K_{zz})$  obtained during this phase identify a bivariate cumulative density function (biCDF). Similarly, the biGaussian simulation of the LCM defines a biCDF for  $(Y_{xx}, Y_{zz})$ . To each simulated couple  $(Y_{xx}, Y_{zz})$  one associates a couple  $(K_{xx}, K_{zz})$  by assuming equality of both biCDFs following Eq. 5.1 :

$$(Y_{xx}, Y_{zz}) \rightarrow F_{biGaussian}(Y_{xx}, Y_{zz}) \stackrel{assumed}{\equiv} F_{upscaled}(K_{xx}, K_{zz}) \rightarrow (K_{xx}, K_{zz}) \quad (5.1)$$

Figure 5.2 compares the original bivariate distributions from upscaling (left) and those obtained from LCM after applying Eq. 5.1. The transform was defined by computing  $F(Y_{xx})$ ,  $F(Y_{zz})$  and  $F(Y_{xx}, Y_{zz})$  on one hand and  $F(K_{xx})$ ,  $F(K_{zz})$  and  $F(K_{xx}, K_{zz})$  on the other hand and then by measuring the Euclidean distance between the triplets. The closest triplet  $(F(K_{xx}), F(K_{zz}), F(K_{xx}, K_{zz}))$  to a simulated triplet  $(F(Y_{xx}), F(Y_{zz}), F(Y_{xx}, Y_{zz}))$  was selected for mapping  $Y_{xx} \rightarrow K_{xx}$  and  $Y_{zz} \rightarrow K_{zz}$ . To introduce more randomness in the bivariate transform, a small number of closest points to  $F_{biGaussian}(Y_{xx}, Y_{zz})$  in  $F_{upscaled}(K_{xx}, K_{zz})$  are considered for possible mapping and one is randomly selected. Note that the transform so defined is not reversible. Here we did not have to inverse the transform because we used only unconditional simulations. For conditional simulations, invertibility is required. In that case, projection pursuit multivariate transform [236] could be used, but then the inequality  $K_{zz} \leq K_{zz}$  is not guaranteed contrary to our approach.

### 5.3.3 Gradual deformation method

The gradual deformation method was introduced by Hu [225] as an efficient way to gradually modify realizations of Gaussian fields while preserving their spatial structure and variability. It linearly combines a current field  $(Y_{cur}^{i-1})$  with a new realization  $(Y_{new}^i)$  to form an updated field  $(Y_{cur}^i)$ . The updated field becomes the current field and the process is repeated over many iterations until no further significant improvement can be obtained. The updated field

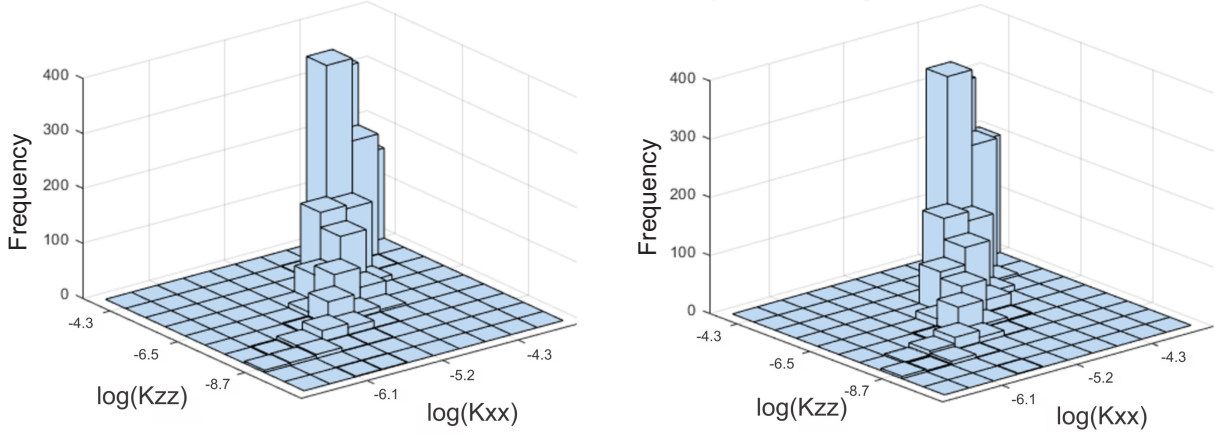


Figure 5.2 3D histogram examples of  $\log(K)$  for the reference bivariate cumulative density function (left) and simulated bivariate cumulative density function from the LCM after applying the bivariate transformation using Eq. 5.1 (right).

is calculated with Equation 5.2 :

$$Y_{cur}^i = Y_{cur}^{i-1} \cos(t_i) + Y_{new}^i \sin(t_i) \quad (5.2)$$

where  $t_i \in [-\pi, \pi]$  is the (scalar) parameter to optimize at iteration  $i$ , and  $Y_{new}^i$  and  $Y_{cur}^0$  are independent realizations with the (same) desired spatial correlation structure. Here, the  $t_i$  parameter is optimized by a combination of golden search method [237] and parabolic interpolation [238] which requires a small number of iterations to reach a value close to the optimum. The main advantage of GDM is that it preserves, by construction, the spatial structure of the fields at every iteration. Its main disadvantage is that it tends to quickly reach a local minimum [167, 225, 239] from which it can be difficult to escape.

Parameter  $t_i$  appearing in Eq. 5.2 is chosen to optimize the selected objective function (OF). In this study, we use the mean absolute errors between simulated ( $h_{sim}$ ) and observed ( $h_{obs}$ ) hydraulic heads, following Equation 5.3 :

$$OF(k) = \frac{1}{n} \sum_{i=1}^n |h_{sim_i} - h_{obs_i}| \quad (5.3)$$

where  $k$  is the iteration index and  $n$  is the number of head observations. Note that at each iteration and for each different trial value for parameter  $t_i$ , the bivariate transform of the current field has to be applied (see Eq. 5.1) and the flow simulator has to be run to get the simulated heads.

## 5.4 Study area

The proposed methodology is applied on a simplified groundwater flow model representing a typical regional valley-fill environment. In this section, we present the hydrostratigraphy of a valley-fill conceptual model, the characterization of the  $K$ -tensor by HSU, the numerical groundwater model and the GDM results of the model calibration.

### 5.4.1 Hydrostratigraphic model

The considered study area is located in south Simcoe County, southern Ontario, Canada, and represents a valley partly filled with glaciolacustrine sediments (Fig. 5.3).

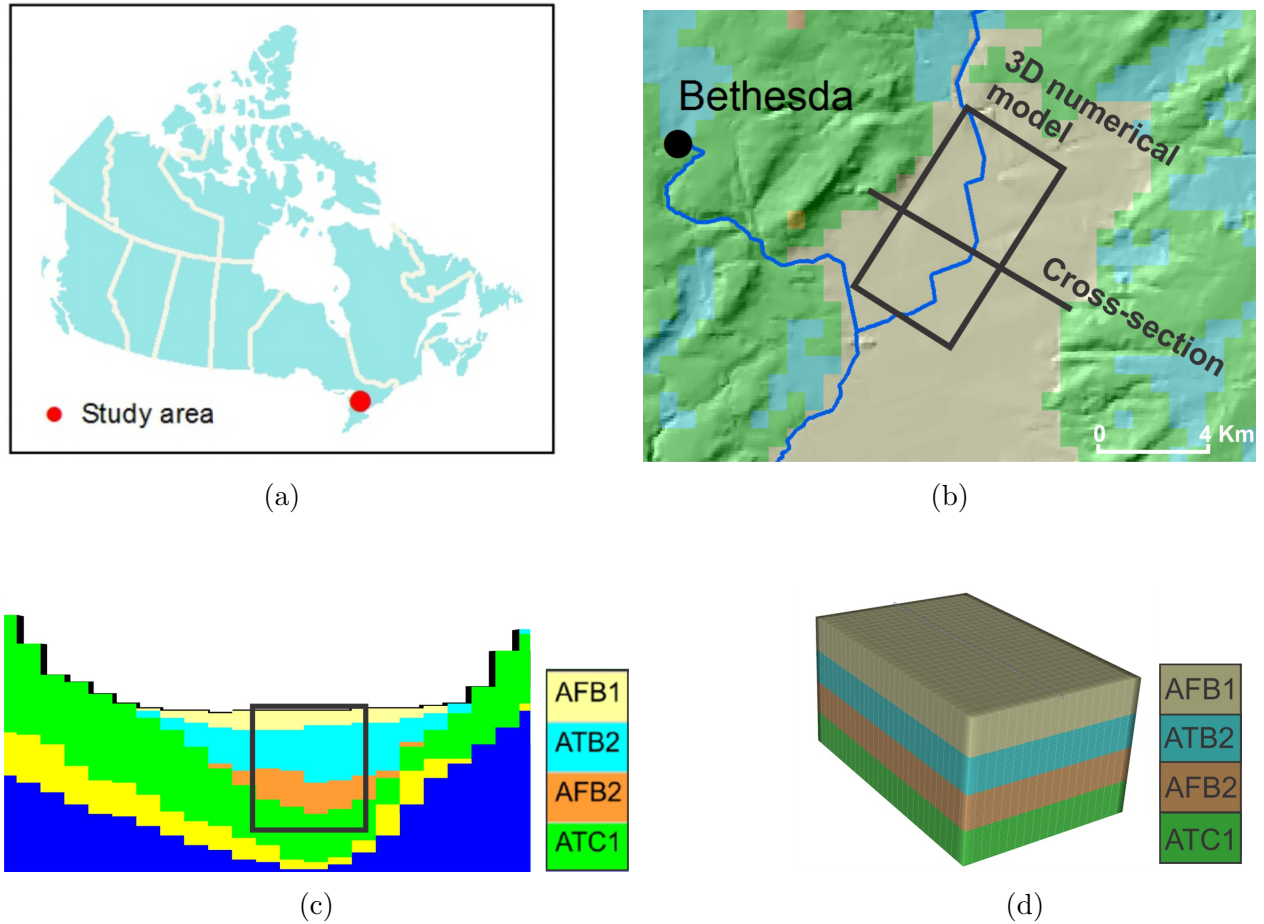


Figure 5.3 a) South Simcoe County location in southern Ontario, Canada ; b) Valley partly filled by glaciolacustrine sediments located in Innisfil creek watershed lowland ; c) Cross-section of the valley with model location (black rectangle) ; d) Simplified 3D numerical model used in the study (4 km x 6 km x 60 m).

The groundwater flow system is controlled by Innisfil Creek flowing in the central portion of the valley (Fig. 5.3c). The simplified 3D numerical model contains four HSUs with a sequence from the base to the top (Fig. 5.3d) : Newmarket till acting as a regional continuous aquitard (ATC1), a glaciofluvial sand and gravel aquifer (AFB2), a fine-grained glaciolacustrine aquitard consisting of silt, clay and subordinate fine sand (ATB2), and a shallow regressional coarse-grained glaciolacustrine aquifer of sand and gravel (AFB1).

The model is divided into a regular block grid with elements of 200x200x1 meters. Each HSU block is strongly heterogeneous with specific statistical and spatial distribution of hydraulic conductivities (Section 5.4.2).

### 5.4.2 Hydraulic conductivity

An HSU block is composed of various proportions of geologic materials (clay, silt and sand). For the modelling point of view, the material proportions are grouped into hydrofacies or simply represented by a hydraulic conductivity ( $K$ ) distribution.

The  $K$  data were collected at a point-scale under field or laboratory conditions. For regional modelling purposes, these data needed to be regionalized and upscaled from point to block-scale. Inner block heterogeneity was then represented by the block  $K$ -tensor. These steps are summarized in Section 5.3.1 and described in more detail in [19, 20].

The characterization of  $K$  values for the groundwater flow model included the point-scale characterization (Section 5.4.2),  $K$  upscaling from the local-scale 3D field to block-scale tensor fields, and the definition of the geostatistical model of each HSU (Section 5.4.2). The latter models are used for stochastic assessment of the HSU  $K$ -tensor needed for GDM application.

#### Point-scale to block-scale hydraulic conductivities

Several data (grain size, specific capacity and laboratory tests) were collected per unit to characterize the variability of the HSU hydraulic conductivity. Only the  $K$  data estimated from grain size samples were retained, as it was the only dataset with satisfactory spatial distribution and density to model the vertical covariance structure of  $K$  HSUs. Sauerbrei's method was applied for the estimation of the  $K$  values from grain size data [240]. An example for ATB2 material distribution and corresponding  $K$  values is shown in Fig. 5.4a. The box-plots in Fig. 5.4b show the distribution of the units with typical value ranges for two aquifer units (AFB1 and AFB2) and two aquitards (ATB1 and ATC1). All units except ATC1 clearly show non-Gaussian marginal distributions.

The  $K$  values were estimated from grain size samples which have a high resolution in the

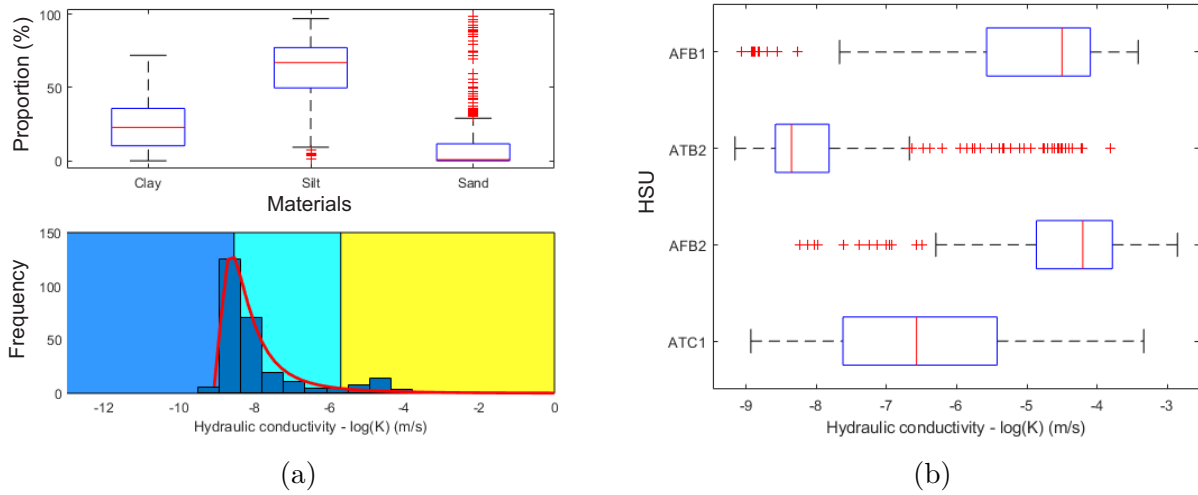


Figure 5.4 a) ATB2 geologic material proportions (top) and corresponding statistical distribution of  $\log(K)$  range values. Colors represents material  $\log(K)$  where clay is in blue, silt in cyan and sand in yellow; b) Boxplot of point-scale  $K$  values for each unit.

borehole vertical directions. The covariance models all have spherical structures with a nugget effect representing 21% to 49% of the total variance. The scalar conductivity is assumed isotropic horizontally and anisotropic along the vertical with a much smaller range along the vertical (2.3 to 7 m). Horizontal variograms could not be modeled based on these grain size samples due to the large distances between the boreholes. Instead, a global horizontal range of about 1800 m was estimated from conductivity values based on relatively dense specific capacity tests from Ontario well record information. This range suggests that a maximum anisotropy ratio between the horizontal range ( $ah$ ) and the vertical range ( $av$ ) of  $ah/av = 300$  can be expected.

Unconditional realizations of scalar  $K$  were then simulated on a grid of  $120 \times 120 \times 120$  points with grid spacing of  $20 \text{ m} \times 20 \text{ m} \times 0.1 \text{ m}$ . The grid resolution was deemed sufficient to represent small-scale variability considering the much larger correlation ranges of  $K$  measurements obtained from quasi-point samples and the 10 times larger size of blocks required for numerical modelling. The upscaling described in Section 5.3.1 was applied to obtain block values on a coarser grid of  $12 \times 12 \times 12$  blocks of size  $200 \text{ m} \times 200 \text{ m} \times 1 \text{ m}$ . The resulting block  $K$ -tensors are used in the flow simulator, which integrate the small-scale variability at the block-scale.

A correction had to be made to the  $K$ -tensors of the ATC1 unit. Since the Newmarket till is a very compact unit, its initial  $K$  values derived from the grain size samples completely ignored

the compaction effect. Instead, laboratory permeability tests were conducted on unreworked core samples from the boreholes, which conserved the compaction state. Results from the permeability tests indicate lower  $K$  values by about two orders of magnitude than those from the grain size analyses [205] and, consequently, ATC1  $K$ -tensor values were modified by this factor.

The developed HSU  $K$ -tensors show a significant variability with a succession of aquifer and aquitard value ranges (e.g.  $K_{x_{ref}}$ , Table 5.1).

Tableau 5.1 Upscaled HSU block-scale  $\log(K_{xx})$  components (m/s)

HSU	min	max	mean	$\sigma^2$
AFB1	-6.7	-3.71	-4.47	0.19
ATB2	-8.82	-4.26	-6.63	1.38
AFB2	-5.86	-3.36	-4.35	0.16
ATC1*	-10.18	-4.81	-7.6	0.99

Geostatistical modelling was performed on the HSU block-scale  $K$ -tensors to determine the LCM spatial covariance and cross-covariance structures of the components used in the block simulations.

The upscaling reveals a very high correlation between  $K_{xx}$  and  $K_{yy}$ , therefore they were assumed equal. Moreover, the off-diagonal tensor components were small enough to be neglected hereafter.

### Block-scale hydraulic conductivity tensor modelling

The geostatistical structures of the  $K$ -tensor were determined only for the principal components of the groundwater flow model ( $K_{xx}$ ,  $K_{yy}$ ,  $K_{zz}$ ) oriented in the principal direction of the model axes. In such case, the cross components are practically negligible and were set to zero. In addition to the very strong correlation between the horizontal components of the reference  $K$ -tensor ( $K_{yy}$  set equal to  $K_{xx}$ ), there is also a strong correlation, not necessarily linear, between horizontal and vertical components ( $K_{xx}$  and  $K_{zz}$ ) for all units (Fig. 5.5).

The covariance modelling showed that the correlated  $\log(K_{xx})$  and  $\log(K_{zz})$  components have similar covariance structures and are thus admissible for linear models of coregionalisation (LMC). The structures, variance and covariance of the LCM are unique for each HSU (Table 5.2).

The LCM models were further used to simulate the stochastic block fields of the  $K$ -tensor principal components. This step involves the bivariate transform described by Eq. 5.1 to



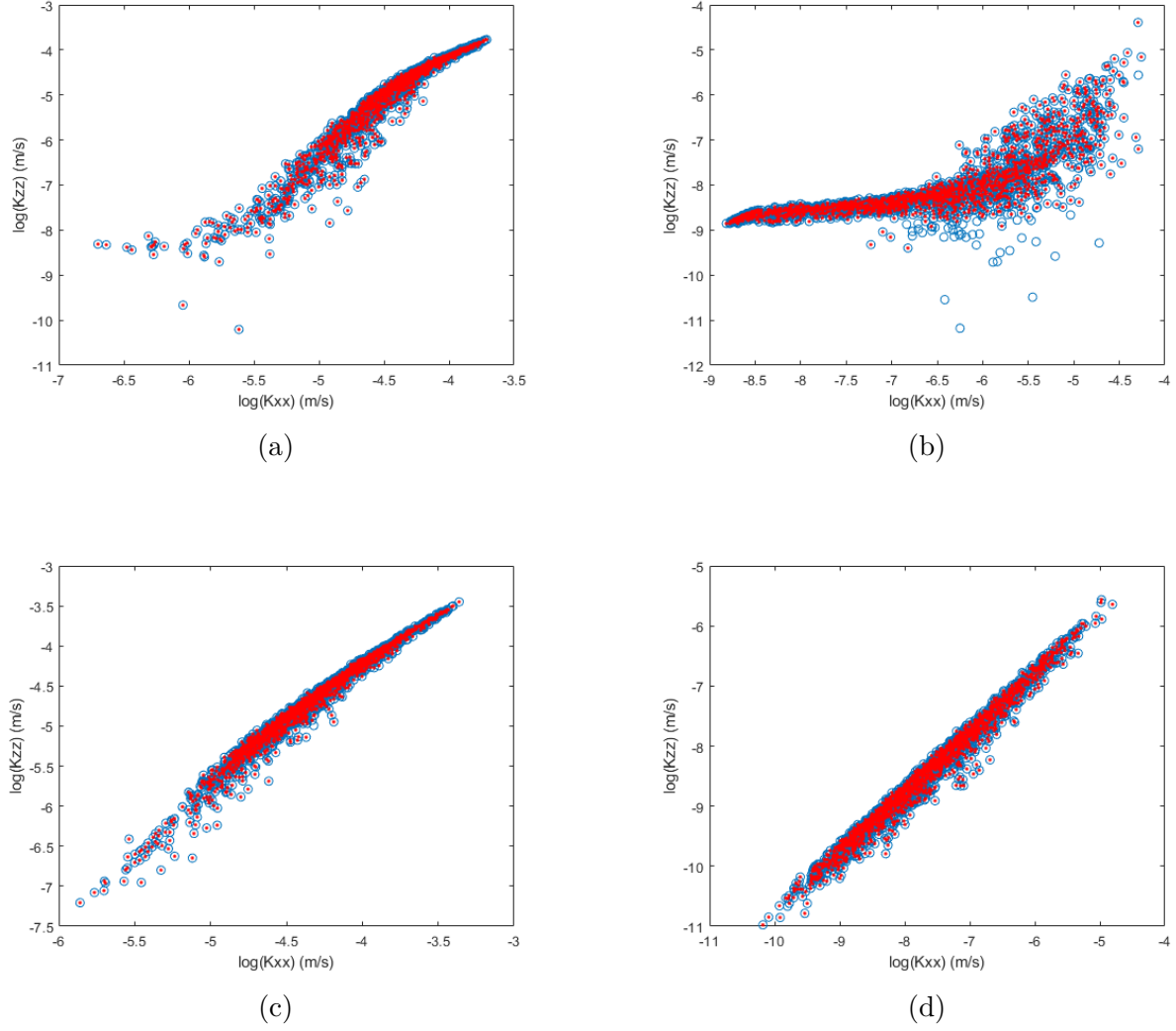


Figure 5.5 Scatter plot illustrating the relationship between  $\log(K_{xx})$  and  $\log(K_{zz})$  components for : a) AFB1, b) ATB1, c) AFB2 and d) ATC1. Blue hollow dots represent the reference values obtained from upscaling, whereas the red dots are examples of bivariate transformation of simulated values.

restore the nonlinear correlations between the tensor components observed in HSU AFB1 and ATB1. The resulting realizations were used for calibration using the GDM (see Section 5.5). Note that all realizations are unconditional as no observations of the block conductivity tensor are available at the block-scale.

Tableau 5.2 Modelled Gaussian variogram structures and linear model of coregionalisation values of  $\ln(K)$ . The nugget effect  $\delta_{zx}$  is symmetric to  $\delta_{xz}$  and  $\delta_{yx}$  is symmetric to the  $\delta_{xy}$  component. Range parameters ( $a_x$ ,  $a_y$  and  $a_z$ ) are in block units of 200x200x1 meters

HSU	$a_{xx}$	$a_{yy}$	$a_{zz}$	$\delta_{xx}$	$\delta_{xz}$	$\delta_{zz}$	$\sigma_{xx}^2$	$\sigma_{xz}^2$	$\sigma_{zz}^2$
AFB1	3.29	2.89	2.08	0.04	0.01	0.20	1.00	2.36	5.58
ATB2	1.85	1.85	1.15	0.30	0.30	0.30	7.40	3.46	3.00
AFB2	4.04	4.04	2.89	0.05	0.05	0.06	1.00	1.50	2.30
ATC1	5.20	4.04	2.31	0.20	0.20	0.20	5.20	5.15	5.20

### 5.4.3 Groundwater flow model

The 3D groundwater flow model was developed with the Saltflow code [235]. The considered study area was 4 km (horizontal x-direction) by 6 km (horizontal y-direction) by 60 m (vertical z-direction). The model domain contains 36,000 isoparametric 3D finite elements (20x30x60) and 39,711 nodes. The stream flowing in the central portion is represented by a constant head boundary condition. The values along the stream vary linearly between 57 to 59 meters over a thickness of one meter. The stream actually represents a major drainage boundary for the infiltrated precipitation water (recharge) across the study area.

In the model, the HSUs are represented by the principal components of the  $K$ -tensor realizations. There is one single value for each component per block. The component spatial distributions were simulated with the turning band algorithm [233], whereas the correlated components were simulated using a LCM model (Table 5.2). A model referred to as the “true model” was taken as one of the unconditional realizations (Figure 5.6). This model is used in the flow simulator to get "true" head values that one will seek to calibrate.

The average recharge rate in the model was assumed at 25 mm/y, which represents a fairly low rate for glaciolacustrine sediments in southern Ontario. Such a low value is justified by the location of the study area, which is restrained to a relatively narrow corridor along the stream that serves rather as a discharge zone for regional groundwater flow. In addition, intense agriculture activities regularly include trench drainage systems, which capture considerable amounts of infiltrated water. For development purposes, the recharge rate was simulated using the turning band method with a small variance, i.e., small spatial variability around the mean value, which allows for a GDM of the  $K$ -tensor combined with the recharge rate (Figure 5.6b). Figure 5.6c shows the head distribution of the true model. To test the accuracy of the calibration procedure (Section 5.5), 59 head observations were randomly selected from the 3D simulated head field covering each of the model units (5.6d). Since no field observations of the water levels are considered herein, these values will represent field observations referred

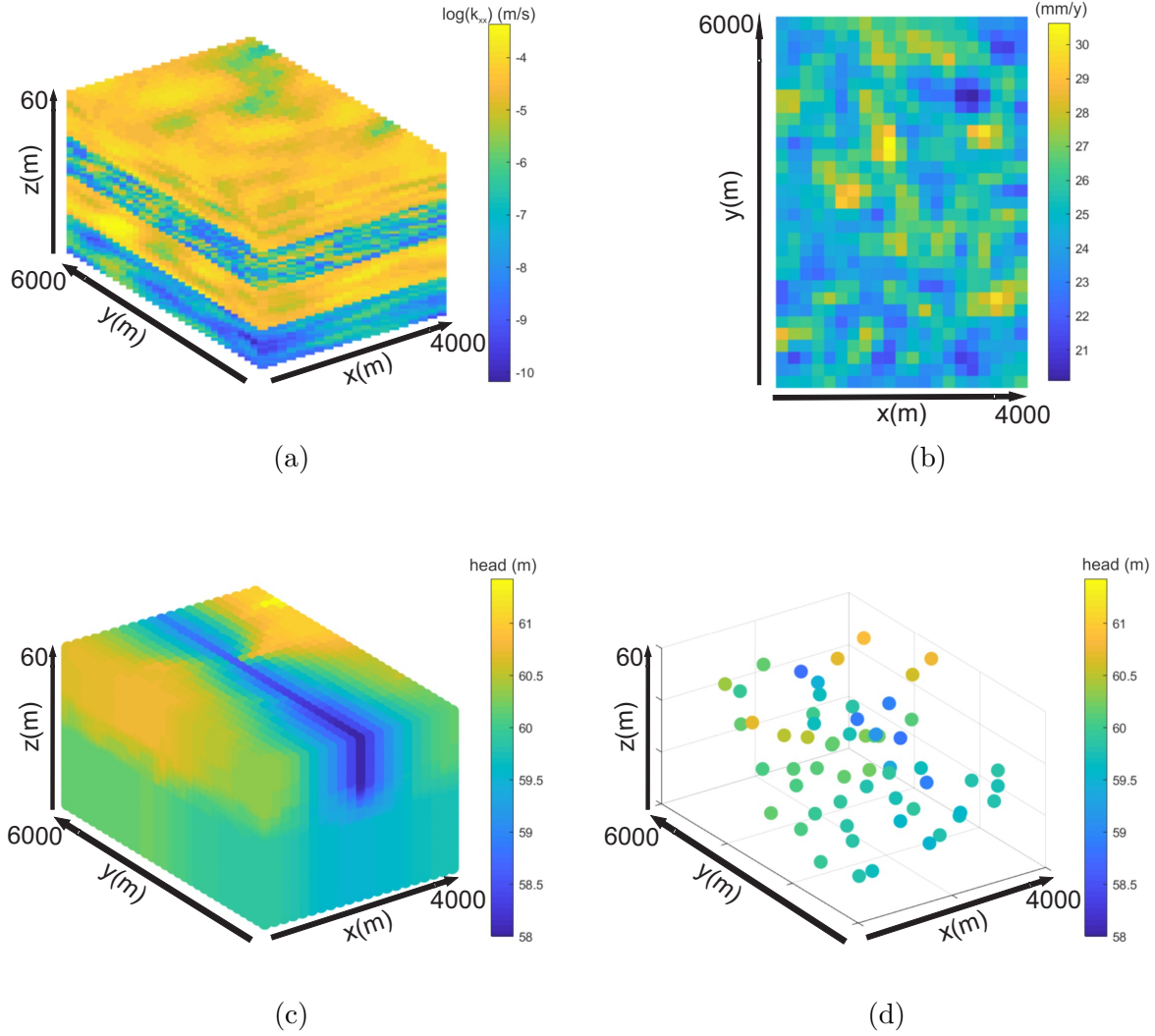


Figure 5.6 a) True model parameters : a) simulated hydraulic conductivity of  $\log(K_{xx})$  component, b) simulated recharge field in mm/year, c) head field in meters and d) 59 head observations randomly selected from the 3D simulated head field which refer to as “true observations”.

to as “true observations”, that will be the target values to be attained by the GDM scenarios.

## 5.5 Model calibration

In this section, the true model  $K$ -tensor fields and recharge rate fields are considered as unknown. To test the proposed GDM methodology, the sets of new  $K$ -tensors and recharge realizations, optimized by the calibrated groundwater flow model, were compared against the head observation data of the true model (Fig. 5.6).

Among all the results presented in this section, a reference model (Section 5.5.1) was determined around which three scenarios are compared in order to better understand the performance of calibration by the proposed methodology. The reference model is considered to best fit the true observations. The different scenarios evaluate : (i) the random effects of equivalent realizations on the variability of the solution (Section 5.5.2), (ii) recharge effects on the head results (Section 5.5.3), and (iii) the influence of the number of head data observation (Section 5.5.4).

### 5.5.1 The reference model

The reference model represents the base case or the so called "best" model compared to the unknown true model. Therefore, the results of the other scenarios are compared to this reference model. The calibration of the reference model was conducted using 100 realizations of the  $K$ -tensor and recharge rate (Fig. 5.7).

The results of the GDM show an acceptable calibration for the reference model. The head difference at observation points is considerably reduced as shown in Fig. 5.7g, where the calibrated reference model adequately reproduces the true head observations.

Figure 5.7h shows the simulated vs true head observations for initial field (blue), intermediate GDM field (grey) to the final GDM field (red). The spatial distribution of the calibrated recharge,  $K_{xx}$  and head fields (Figs 5.7d, 5.7e and 5.7f, respectively) compare well with the initial structures (Figs 5.7a, 5.7b and 5.7c) and the true model fields (Fig. 5.6a, 5.6b and 5.6c). The reference model also correctly reproduced the variograms and cross-variograms in different directions, e.g. AFB1 unit depicted in Fig. 5.8.

Most of the improvement in the OF brought by the GDM occurs in the first 10 realizations (Figure 5.9). For the remaining 90 realizations (from 10 to 100), a relatively small OF reduction occurs and the weight  $\cos(t)$  becomes very close to 1, indicating no further significant modification affects the calibrated field.

### 5.5.2 Random effects

GDM was applied similarly with different initial seeds. The seed controls the set of  $K$ -tensor and recharge realizations available to the GDM. Everything else was kept identical.

Figure 5.10 show four different final calibrated  $\log(K_{xx})$  fields. The quality of calibration on the head field is similar for the four fields. Although the structures globally also appear similar, the spatial distribution of values is significantly different. This shows that the proposed methodology can generate many different models able to calibrate the head fields. In absence

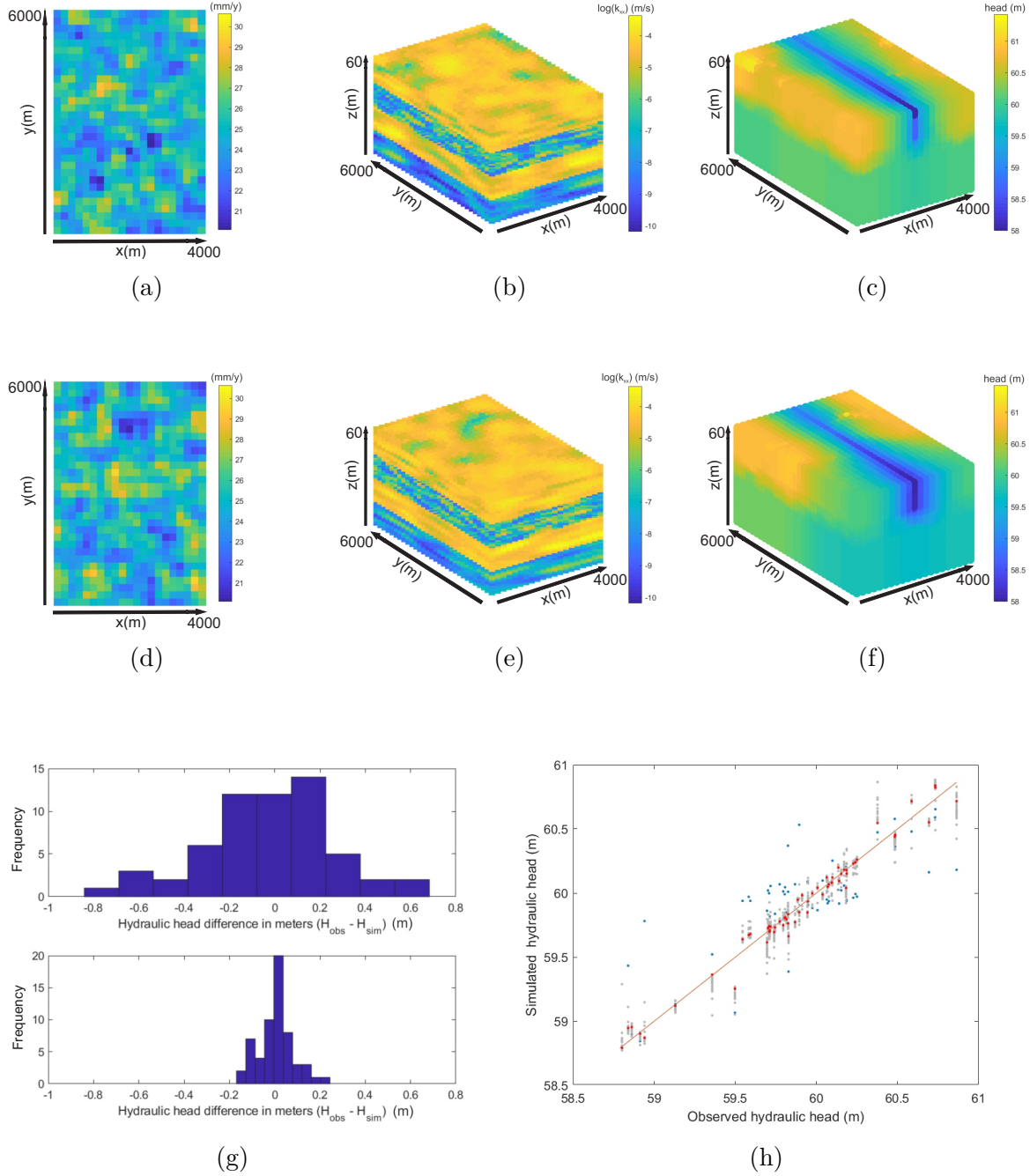


Figure 5.7 **Initial condition** : a) recharge rate (mm/y), b) initial hydraulic conductivity ( $\log(K_{xx})$ ), c) simulated head field(m); **Calibrated condition** : d) calibrated recharge (mm/y), e) hydraulic conductivity ( $\log(K_{xx})$ ), f) simulated head field (m) ; g) Initial (top) and calibrated (bottom) histogram of the difference between true and simulated heads, h) scatter plot of simulated vs true head observations (blue=initial ; grey=intermediate ; red=final)

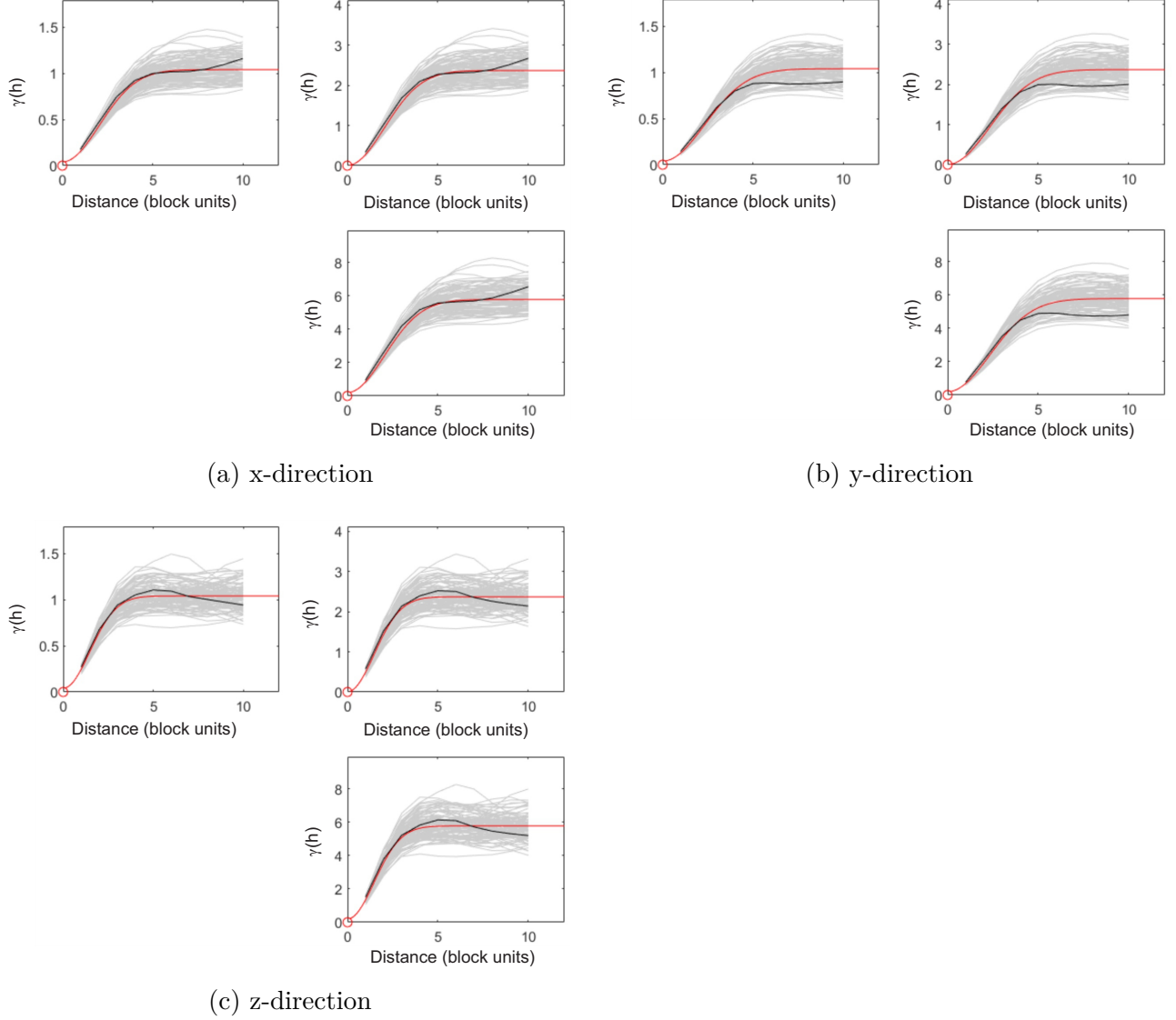


Figure 5.8  $K$ -tensor variograms and cross-variograms of  $\ln(K_{xx})$  and  $\ln(K_{zz})$  for 3 principal directions (distances are in block units) : a) x-direction, b) y-direction, c) z-direction. For each direction,  $\gamma_{xx}$ ,  $\gamma_{xz}$  and  $\gamma_{zz}$  variograms in reading order. Variograms of calibrated fields (black), of the theoretical models (red), and of the realizations used in GDM (grey).

of other constraining information, these fields appear equivalent and can thus be used to assess the uncertainty about the real field.

Figure 5.11 shows the boxplots of  $K_{xx}$  (a) and  $K_{zz}$  (b) components of the model tensors before ( $K_{xx0}$  and  $K_{zz0}$ ) and after the calibration for the four realizations ( $K_{xx1}$  to  $K_{xx4}$  and  $K_{zz1}$  to  $K_{zz4}$ ). For each observation point located at a model node, the corresponding  $K$  value is determined as the average of the eight connected elements. It can be observed that

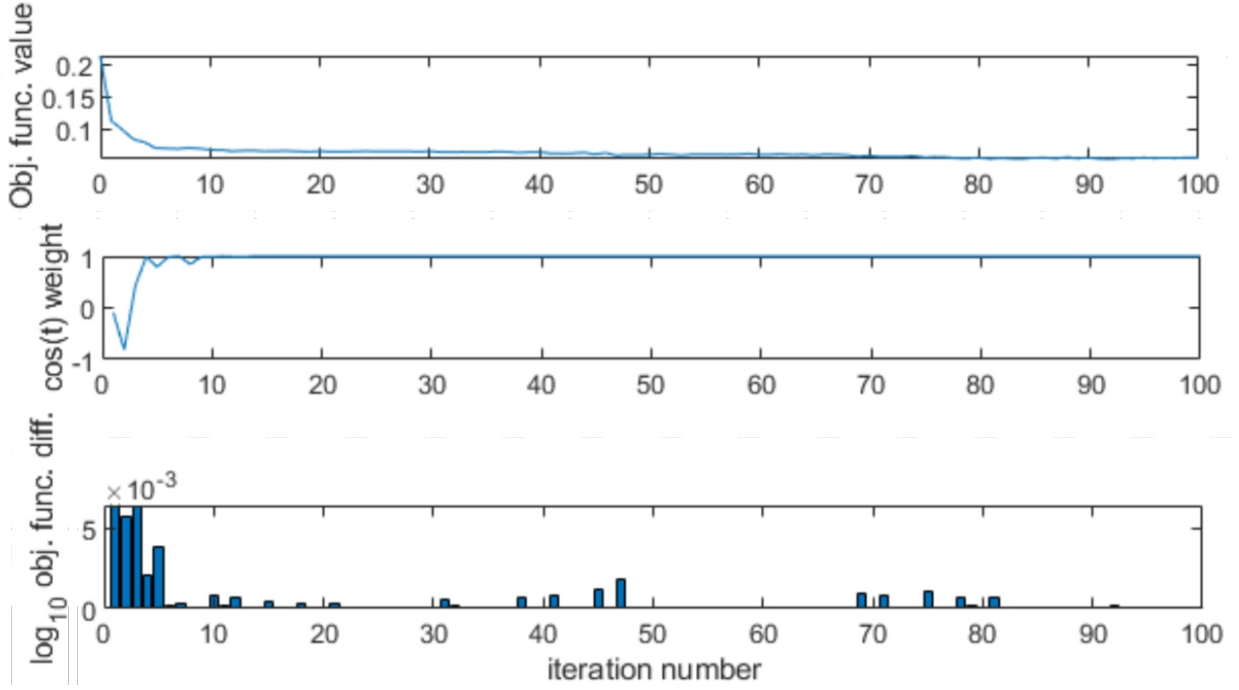


Figure 5.9 Evolution of the GDM with iterations for : a) OF, b) deformation coefficient  $t$ , c) weight  $\cos(t)$ , and d) OF difference (in  $\log_{10}$ ).

the calibration has no significant effect on the local variability of the  $K$ -tensors components. Although the hydraulic head value is measured locally, it is influenced by the upstream and downstream  $K$  values even at great distances, as well as by the recharge flux, which in this case is also stochastic.

### 5.5.3 Recharge effects

The uncertainty of the recharge rate affects the GDM solutions as well. To illustrate these effects, different recharge scenarios are considered (different recharge for each GDM step, constant recharge over all cells, unique recharge realization for all GDM steps). The  $K$ -tensor realizations used in the GDM remain the same for all recharge scenarios. Figure 5.12 illustrates the studied recharge scenarios, the corresponding calibrated  $\log(K_{xx})$  field and the difference between the calibrated and the true  $\log(K_{xx})$  field.

The results indicate that the recharge scenario impacts the calibrated  $K$ -tensors. The calibrated fields of the various scenarios show a different spatial distribution of  $\log(K_{xx})$ . In addition, the recharge scenario affects the variogram of the calibrated  $K_{xx}$  component. Figure 5.13 shows the variograms of the calibrated fields for  $K_{xx}$  and  $K_{zz}$  components as well as the

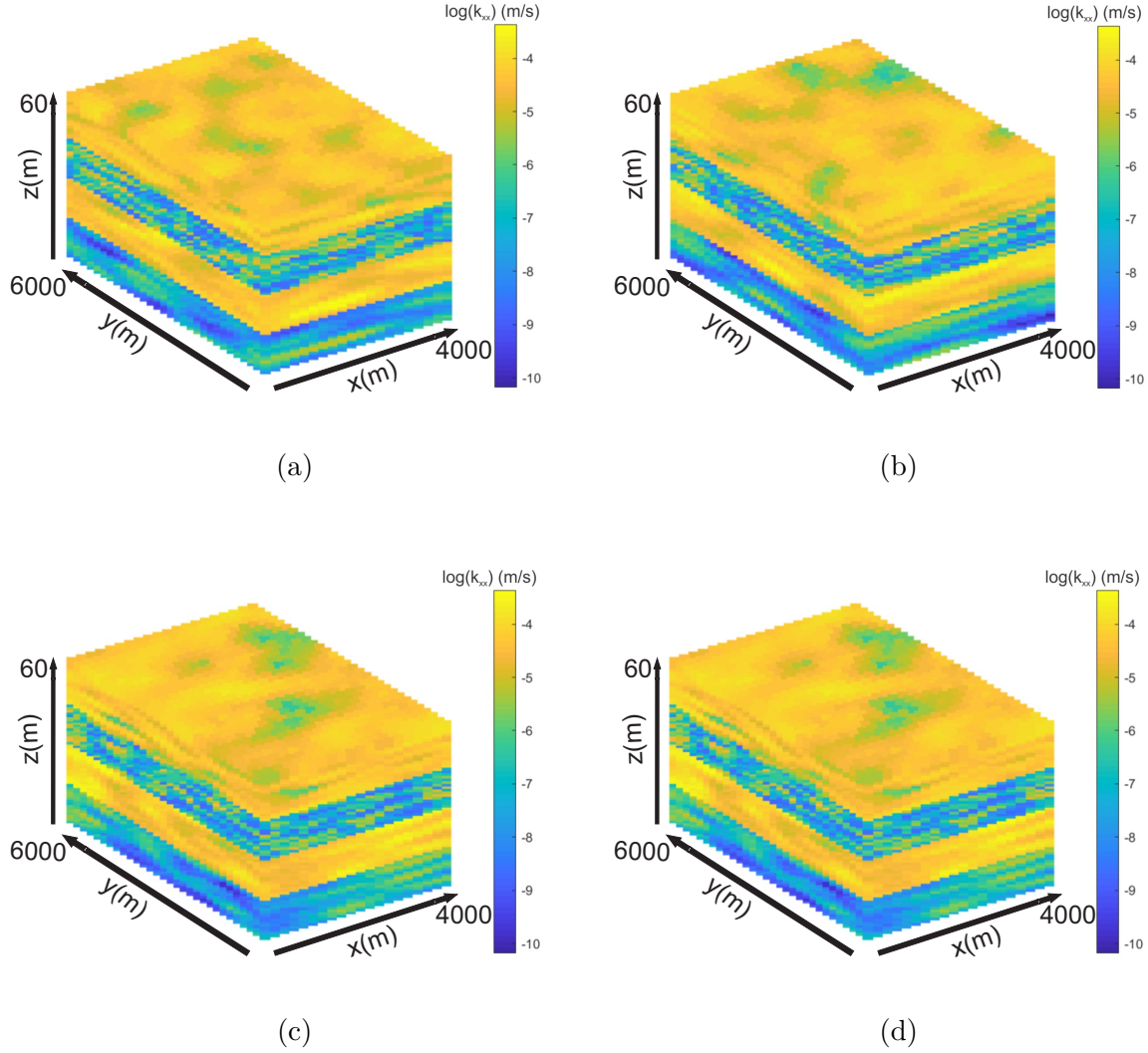


Figure 5.10 Four final hydraulic conductivity fields in  $\log(K_{xx})$  obtained with different seeds.

cross-variograms of these components of each of the three recharge scenarios and for the true model in the three main directions.

The quality of the fit obtained with each recharge scenario is assessed using the MAE between the  $K$ -tensor variograms of the true and calibrated fields under the different recharge scenarios :

$$MAE_i = \sum_h w(h) \left| \frac{\gamma_t(h) - \gamma_i(h)}{\gamma_t(h)} \right| \quad (5.4)$$

where  $h$  is the lag distance,  $\gamma_t(h)$  and  $\gamma_i(h)$  are respectively the true and calibrated experimental variogram at lag  $h$  for a given realization and tensor component, and  $w(h)$  is the



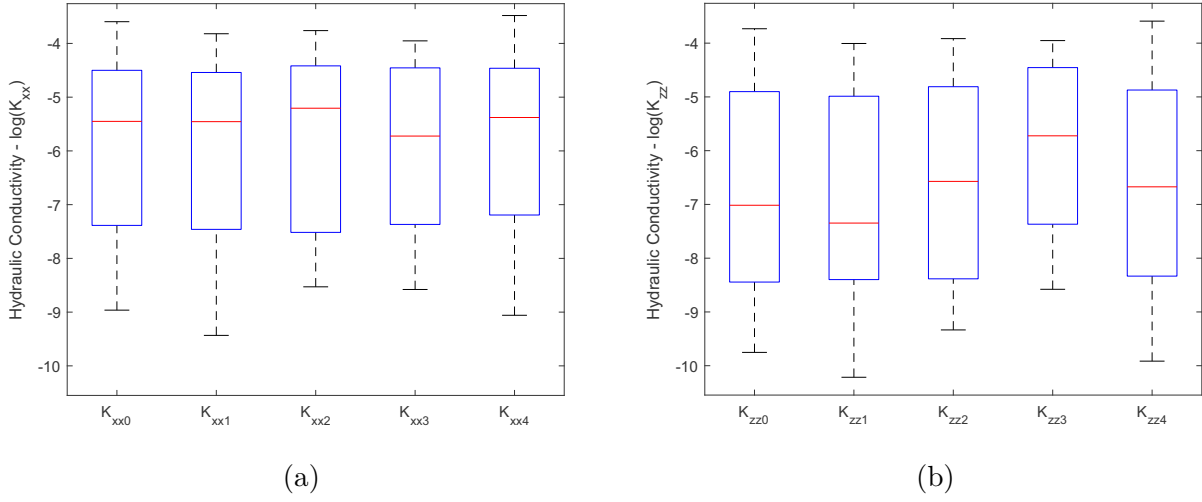


Figure 5.11 Boxplots of  $K_{xx}$  (a) and  $K_{zz}$  (b) components of the model tensors before ( $K_{xx0}$  and  $K_{zz0}$ ) and after calibration for the four realizations ( $K_{xx1}$  to  $K_{xx4}$  and  $K_{zz1}$  to  $K_{zz4}$ ).

weight factor applied to lag  $h$ . The sum of weights is one which is defined as  $w(h) = \frac{N(h)/h^2}{\sum_h N(h)/h^2}$  where  $N(h)$  is the number of pairs per lag. Fig. 5.14 shows the boxplot of the 300 MAEs (100 calibrated fields  $\times$  3 tensor components) under the three different recharge scenarios.

The visual comparison of boxplots indicates that the stochastic scenario best reproduces the conductivity structure of the true field. The median of the MAE (and mean) computed with Eq. 5.4 is least for the stochastic scenario indicating a better variogram reproduction with this approach.

#### 5.5.4 Influence of number of head data

In this section, calibration is done on various numbers of head observations with 59, 30, 15, 11, 7 and only 3 observations. The heads obtained with the calibrated models are compared over a dense 3D grid to the true head field by computing mean square error (RMSE) and correlation.

Figure 5.15 shows that the applied method calibrates well over the entire domain under the condition that there is a sufficient number of observation wells. The GDM converges towards acceptable error and correlation values with only several head measurements (about 10 in this case). Correlation and RMSE steadily improves with increasing number of data.

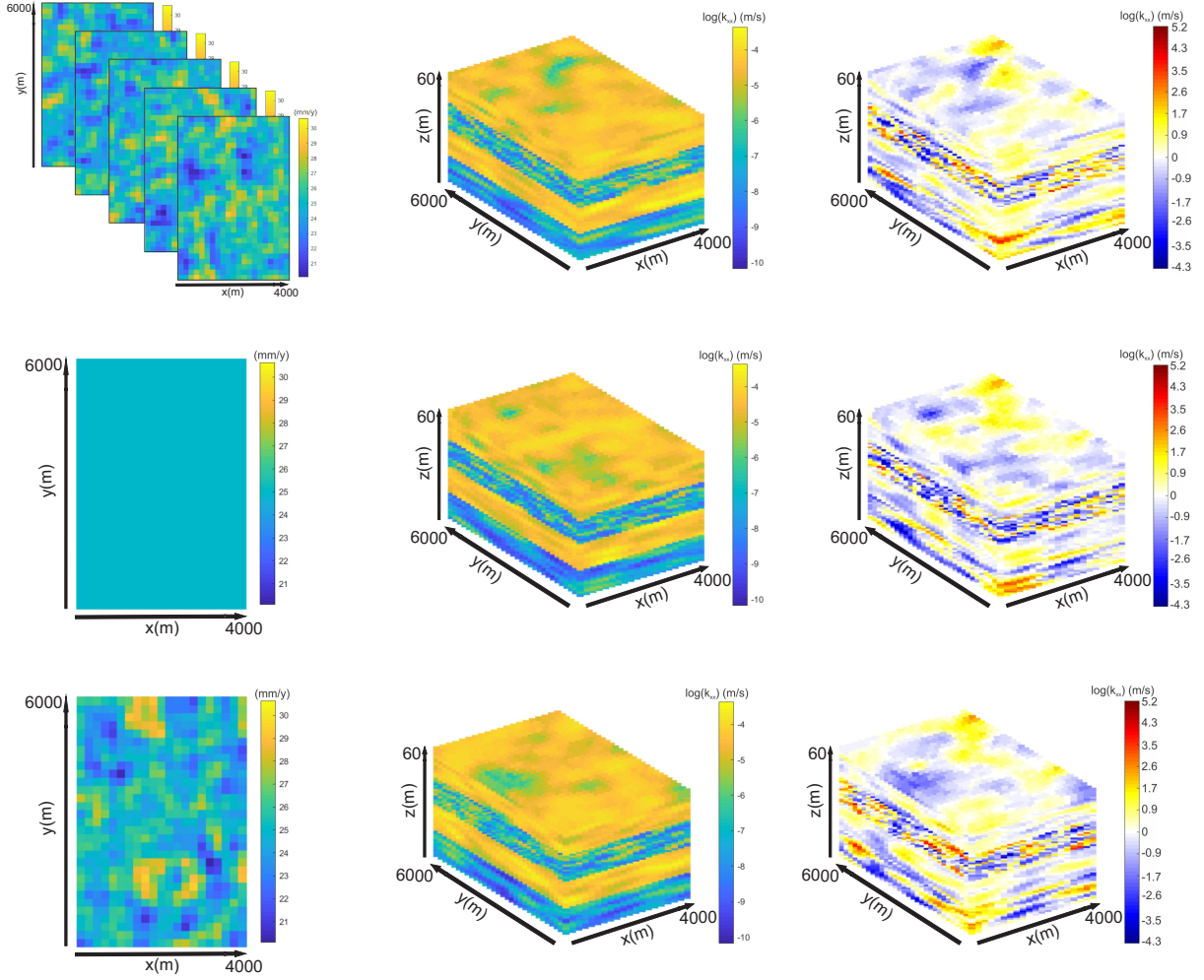


Figure 5.12 Recharge effect on  $\log(K_{xx})$  field obtained by GDM for 3 different scenarios : stochastic recharge (top row) ; uniform recharge (middle row) and a single fixed realization recharge (bottom row). Columns are : recharge rate in mm/y (left), calibrated  $\log(K_{xx})$  field (middle) and the difference between the calibrated and the true  $\log(K_{xx})$  field (right column).

## 5.6 Capture zones

The proposed methodology was used to illustrate the uncertainty assessment for the delineation of well capture zones using the calibrated fields. The different capture zones by GDM are compared to those determined using the commonly used PEST inversion algorithm [241].

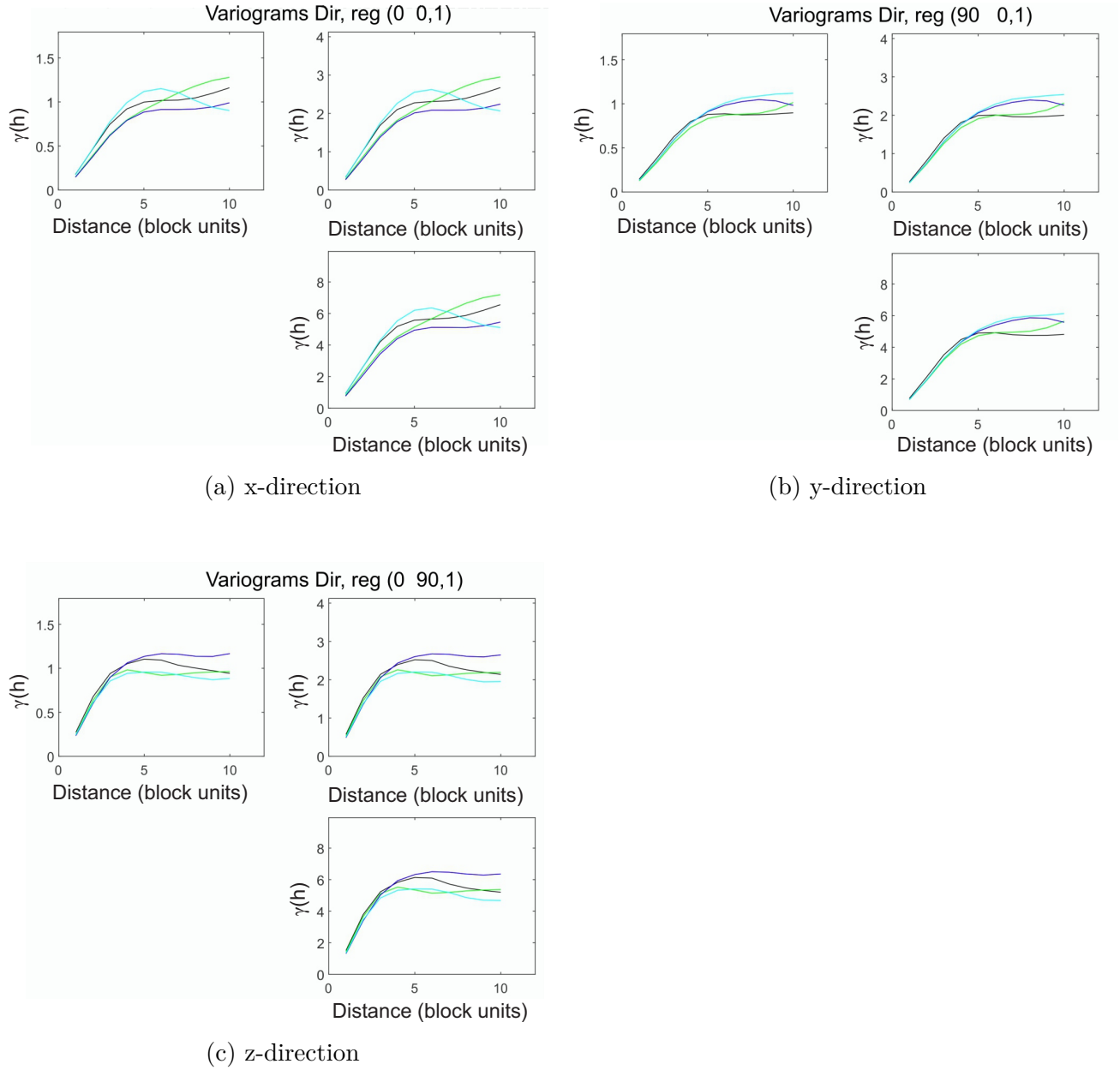


Figure 5.13 Recharge scenario  $K$ -tensor variograms and cross-variograms of  $\ln(K_{xx})$  and  $\ln(K_{zz})$  for 3 principal directions (distances are in block units) : a) x-direction, b) y-direction, c) z-direction. For each direction,  $\gamma_{xx}$ ,  $\gamma_{xz}$  and  $\gamma_{zz}$  variograms in reading order. Variograms of calibrated fields for the different scenarios : stochastic recharge (black), uniform recharge (blue), single realization recharge (green) and true recharge (cyan)

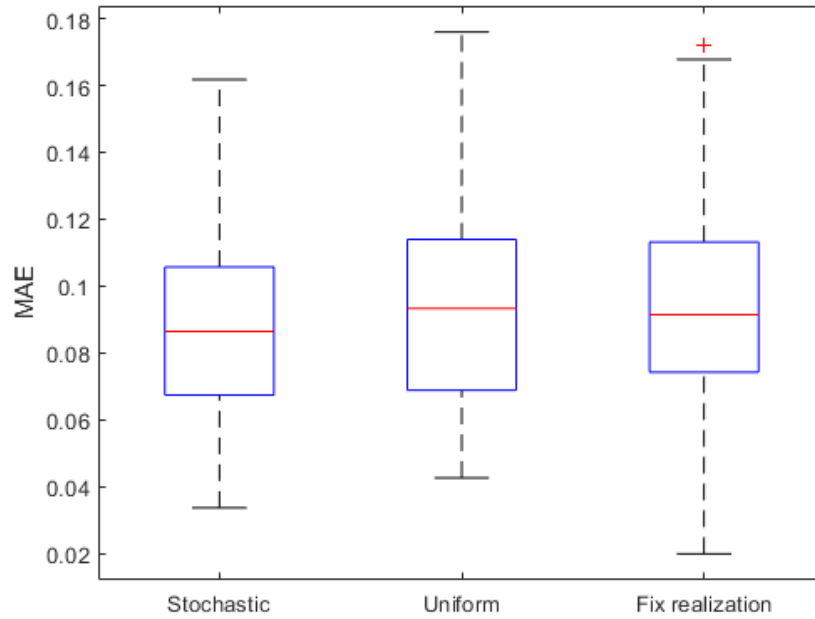


Figure 5.14 MAE (eq. 5.4) boxplot of 300 experimental variograms of calibrated models vs the true model variogram for each recharge scenario

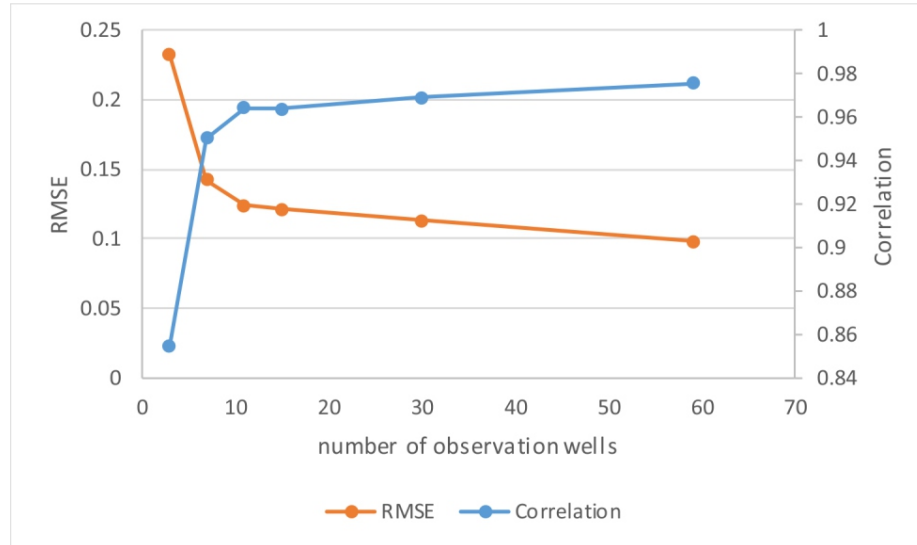


Figure 5.15 RMSE and correlation variation with the number of observation wells used into the calibration

### 5.6.1 PEST calibration

The PEST inversion method is widely used in hydrogeological modelling [242]. In this section, the calibration of the reference model (see Section 5.5.1) is done using PEST. The calibration

with PEST was carried out zonally, i.e. using one homogeneous  $K$ -tensor per HSU. The initial values of the tensor components were set equal to the average values of the corresponding HSU in the reference model. The PEST inversion was constrained by the minimum and maximum values found in the reference model in each HSU. The results presented in Fig. 5.16 show that the PEST algorithm calibrates well using the hydraulic head observations, similar to the GDM model (compare with Fig. 5.9).

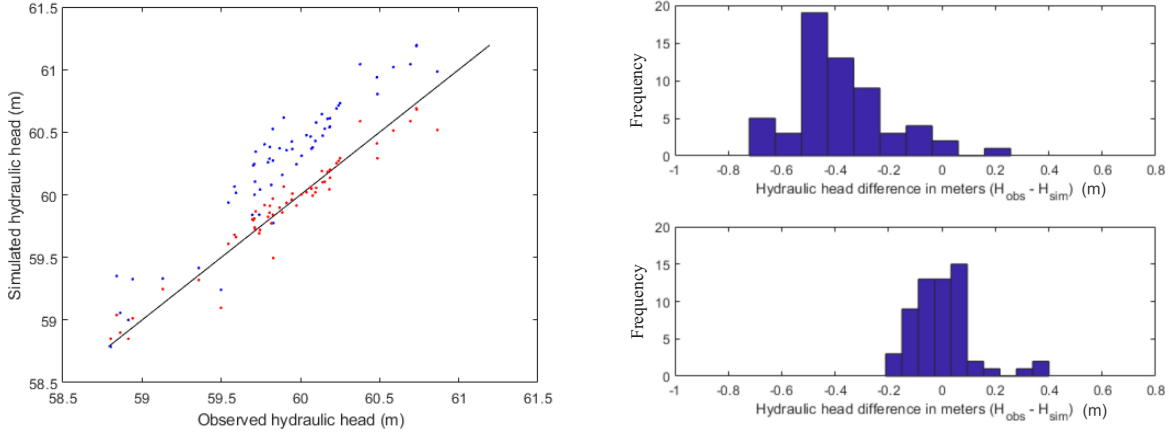


Figure 5.16 PEST inversion results : observed hydraulic head vs initial (blue) and calibrated (red) simulated head on left; hydraulic head difference for the initial (right top) and the calibrated (right bottom) models

It is interesting to note that the hydraulic head field obtained by PEST with the initial  $K$ -tensor generally overestimates the hydraulic head observations. Since PEST was used zonally, the algorithm had to adjust the mean tensor component values to calibrate. Table 5.3 compares the initial and calibrated values after PEST and GDM inversion. For inversion with PEST there is approximately one order of magnitude difference between the initial and calibrated tensors for ATB2 and AFB2. The AFB2 values are clearly inadequate after the PEST calibration as calibrated  $\log(K_{xx}) < \log(K_{zz})$  by an order of magnitude, something geologically unacceptable for this type of environment. Moreover, after PEST inversion the ATB2 aquitard and the AFB2 aquifer show approximately the same value for  $\log(K_{xx})$ , which is also geologically unrealistic. On the contrary, the mean conductivities of calibrated fields by GDM are equal to the initial means.

Hence, despite good calibration to head data, the PEST-calibrated field is not acceptable based on geological considerations. Multiple zones within each HSU or the pilot point method within PEST could have been used. None of these approaches would allow reproducing the spatial correlations and complex relationship between tensor components, therefore they were

not attempted.

Tableau 5.3 Initial  $K$  values ( $\overline{\log(K_{xx0})}$  and  $\overline{\log(K_{zz0})}$ ) and calibrated values ( $\overline{\log(K_{xxF})}$  and  $\overline{\log(K_{zzF})}$ ) (m/s) from PEST inversion and from GDM inversion (reference model)

Algorithm	HSU	$\overline{\log(K_{xx0})}$	$\overline{\log(K_{zz0})}$	$\overline{\log(K_{xxF})}$	$\overline{\log(K_{zzF})}$
PEST	AFB1	-4.5	-5.2	-4.3	-5.8
	ATB2	-6.9	-8.2	-6.0	-9.5
	AFB2	-4.4	-4.8	-5.8	-4.8
	ATC1	-7.6	-8.4	-7.4	-8.4
GDM	AFB1	-4.5	-5.2	-4.5	-5.2
	ATB2	-6.9	-8.2	-6.9	-8.2
	AFB2	-4.4	-4.8	-4.4	-4.8
	ATC1	-7.6	-8.4	-7.6	-8.4

### 5.6.2 Well capture zone delineation and uncertainty

To illustrate the benefit of the proposed approach, 100 equivalent flow models were calibrated using GDM to assess the uncertainty regarding delineation of well capture zones. For this purpose, two pumping wells are added to the model, one located in the surficial aquifer (AFB1) and one in the confined aquifer (AFB2). Well w1 is located at coordinates  $x=1400$  and  $y=3000$  and well w2 at  $x=2600$  and  $y=3000$ . The well screens extend vertically from 46 to 59 m (w1) and 16 to 29 m (w2).

To delineate the well capture zones, the capture probability concept was applied [243]. The approach consists of quantifying a capture probability using the backward mode of the advective-dispersive transport model by inverting the groundwater velocity field [244–247]. In this approach the macrodispersion term simulates the local-scale uncertainties within elements [248]. A capture probability boundary condition of  $P = 1$  is applied at well locations. Transient transport simulations provide the probability field which decreases up-gradient.

Figure 5.17 shows capture probability contours of the 100 calibrated models for probabilities  $P = 0.5$  (light grey) and  $P = 0.1$  (dark grey) after 100 years of simulation. For comparison purposes, the capture probability contours (red) are also determined for the calibrated model with PEST and the true model (blue).

Wells w1 and w2 have separate capture zones. Well w2 shows a capture probability  $P > 0.1$  from the PEST inversion that only slightly exceeds the capture probability  $P > 0.5$  of the GDM models. In this example, the stochastic approach appears more conservative than PEST as the capture zones are wider for the same capture probability level. Both approaches show in plan view asymmetric capture zones for w1 and symmetric zones for w2. For w1, the river

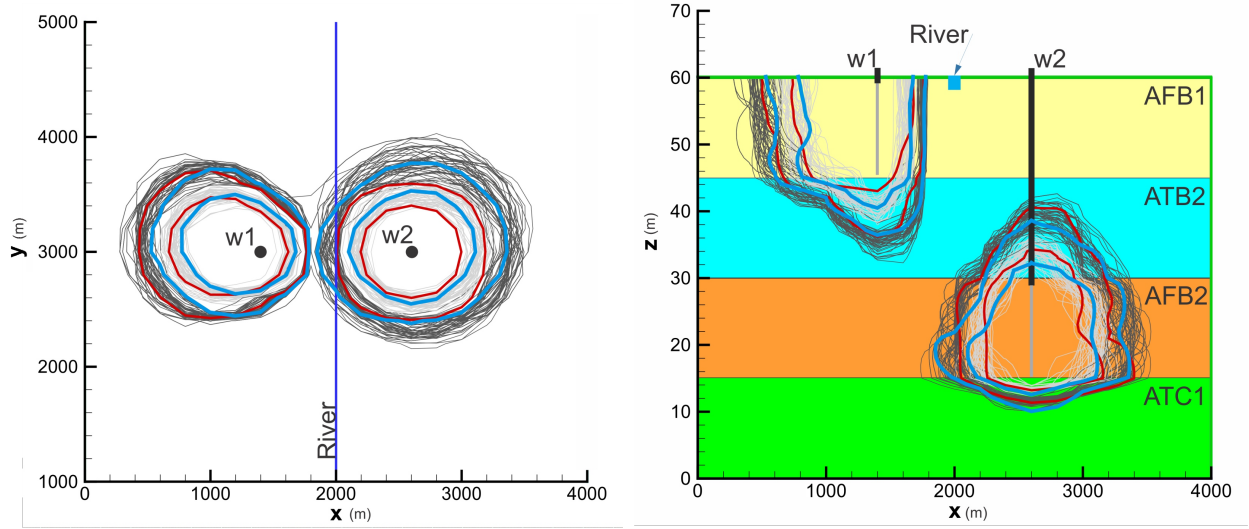


Figure 5.17 Capture probability contours for  $P = 0.5$  (light grey) and  $P = 0.1$  (dark grey) at 100 years simulation : plan view on left and cross-section view ( $y=3000$ ) on right (well casings are black and screens are grey) ; capture probability contours with PEST (red) and true model (blue)

flows north-south and acts as a discharge zone, so the regional groundwater flow is toward the river explaining the capture zone asymmetry. On the contrary, w2 is not hydraulically connected to the river due to the presence of the aquitard ATB2. The capture zones for the true model (blue) differ significantly from those obtained by PEST (red) and are everywhere well within the interval of capture zones obtained by the stochastic approach using GDM (light gray for  $p=0.5$  and dark gray for  $P=0.1$ ).

## 5.7 Discussion

Our approach includes upscaling of the quasi-point conductivity to  $K$ -tensor conductivity and a bivariate transformation aimed at preserving the non-linear relationship between tensor components. The results clearly show that GDM is able to calibrate a highly heterogeneous 3D multi-layer regional groundwater flow system while preserving the desired spatial direct and cross-correlation of the  $K$ -tensor. The computation time of this indirect inversion method is dominated by the number of calls to the groundwater flow simulator. This time can be kept reasonable by allowing only a few line search steps to find parameter  $t$  in Eq. 5.2 and by asking only a small number of GDM iterations, justified by the fact that the most significant improvements to the OF occur in the few first iterations (see Fig. 5.9).

The method allows to generate numerous equally well calibrated models which can be used to assess the uncertainty about the real field. Our results indicate that the recharge scenario impacts the calibrated  $K$ -tensors and that the stochastic recharge scenario produced calibrated fields with  $K$ -tensor component variograms closer to variograms of the true model.

One interesting result was that the entire head field was well approximated after calibrating on only a small number of head data ( $n=10$ ). Using more head data barely enhanced the approximation to the entire true head field.

The calibrated head fields obtained by GDM and by PEST were compared for the definition of capture zones defined around two pumping wells. The  $K$ -tensor for the different HSUs appeared geologically more realistic by GDM than the calibrated  $K$ -tensor obtained with PEST. Moreover, the stochastic approach leads to larger, possibly more conservative capture zones than with a single PEST-calibrated model.

## 5.8 Conclusion

In this study the GDM approach was applied to calibrate a groundwater flow model of a valley-fill hydrostratigraphic succession within a glacial sedimentary basin. The GDM successfully calibrated the groundwater flow models with stochastic correlated  $K$ -tensors for each HSU obtained with a linear model of coregionalisation. During the simulation, the covariance structure and the tensor component nonlinear correlations were preserved thanks to a bivariate transform. Using different  $K$ -tensor realizations, GDM calibrates alternative models. The latter are essential inputs to assess uncertainty of groundwater flow and transport problems.

## Acknowledgments

We are grateful to Miroslav Nastev and two anonymous reviewers for their constructive comments which helped to improve the manuscript. The authors thank the support from the Groundwater Geoscience Program, Geological Survey of Canada, Natural Resources Canada. This research was partly financed by NSERC (RGPIN-2015-06653).



## CHAPITRE 6 DISCUSSION GÉNÉRALE

Les hétérogénéités hydrostratigraphiques inter et intra unités constituent des sources importantes d'incertitude des modèles hydrogéologiques. La caractérisation géostatistique des hétérogénéités permet de proposer des modèles équivalents et ainsi de quantifier leurs incertitudes. Les résultats présentés dans les chapitres précédents montrent que la méthodologie proposée permet cette caractérisation tant au niveau de l'architecture spatiale des UHS (chapitre 4) que de leur conductivité hydraulique (chapitre 5). Les méthodes proposées fournissent les données nécessaires pour alimenter des modèles d'écoulement équivalents afin de quantifier l'incertitude touchant l'écoulement de l'eau et le transport. L'application des développements se concentre sur les besoins de la modélisation régionale mais ils pourraient aussi bien s'appliquer à la modélisation à d'autres échelles. La discussion de ce chapitre porte sur la caractérisation stochastique des hétérogénéités inter UHS (section 6.1) et intra unités (section 6.2) ainsi que sur les limites des approches proposées (section 6.3).

### 6.1 Caractérisation stochastique des hétérogénéités inter unités hydrostratigraphiques

Les résultats du chapitre 4 exposent la capacité de la méthode MCP à simuler efficacement un contexte hydrostratigraphique régional. Il s'agit d'une première application de la méthode en 3D et sur un système réellement complexe avec de multiples unités (11). À cela s'ajoute la démonstration de sa capacité à reproduire l'ordination verticale des unités de la séquence sédimentaire et tout ça malgré la forte hypothèse d'indépendance conditionnelle de la méthode. Les sous-sections qui suivent discutent des particularités observées avec la méthode MCP qui portent sur l'hypothèse de l'indépendance conditionnelle, l'utilisation d'un modèle déterministe comme TI, la comparaison de la méthode MCP avec la méthode MPS et quelques autres remarques sur la méthode.

#### 6.1.1 Hypothèse de l'indépendance conditionnelle

Lorsque l'hypothèse de l'indépendance conditionnelle est valable, le MCP est non biaisé par construction, car la méthode tire séquentiellement de distributions conditionnelles exactes. Lorsque l'hypothèse de l'indépendance conditionnelle ne tient pas complètement, la méthode ne tire pas des distributions conditionnelles exactes et rien de ferme sur le biais ne peut être affirmé. Cependant, les résultats expérimentaux pour le cas synthétique montrent clairement

l'absence de tout biais substantiel dans les situations conditionnelles et non conditionnelles. Dans l'exemple plus complexe du contexte réel de Simcoe, les résultats indiquent que le biais est moins important pour le cas non conditionnel et pour l'échantillonnage aléatoire des données que pour les cas où un échantillonnage préférentiel de données est présent. Cependant, même dans les cas d'échantillonnage préférentiel, le biais entre les proportions de la TI et les proportions simulées reste plus faible ou même beaucoup plus faible que le biais observé dans les données dû à l'échantillonnage préférentiel. La surreprésentation de données comme données conditionnantes (surface du roc et UHS superficielle) n'a pas eu d'impact significatif sur les proportions moyennes simulées. Comme prévu, les données supplémentaires ont toutefois réduit la variabilité, mais sans fausser davantage les proportions moyennes. Certes, il resterait à vérifier si ces bons résultats concernant le biais et la robustesse à l'échantillonnage préférentiel peuvent être extrapolés à d'autres cas complexes.

### 6.1.2 Modèle déterministe comme image d'entraînement

Une TI doit refléter l'état des connaissances d'un contexte donné, par exemple la géologie. Pour différentes TI, il existe tout un continuum allant d'une connaissance vague de la texture/structure du domaine à une connaissance parfaite, mais certes inaccessible. Le développement d'une TI dépend essentiellement de l'information disponible.

Dans le domaine pétrolier, la TI représente essentiellement des textures analogues empruntées à des sources externes (exemples : images satellites ou simulations basées sur des objets/processus). Les textures sont ensuite modulées dans l'espace en les combinant avec des données secondaires provenant d'études sismiques et elles sont conditionnées avec une quantité restreinte de données primaires. La plupart des études MPS sont réalisées avec ce type de TI. Un examinateur de l'article MCP a soulevé une préoccupation concernant l'idée d'utiliser un modèle déterministe comme source pour la TI. Cependant, dans d'autres domaines, les données mentionnées précédemment ne sont tout simplement pas disponibles. Par exemple, dans un contexte d'exploitation minière, chaque gisement est unique et complexe et les données sismiques ne sont généralement pas disponibles ou ne peuvent tout simplement pas être utilisées pour identifier les zones minéralisées. Dans un contexte de mines à ciel ouvert, les sources de données possibles pour créer une TI sont les trous de dynamitage situés à proximité de la zone à simuler. On s'attend alors à ce que la TI soit également proche du champ cible [222]. Boucher et al. [218] utilisent aussi un modèle géologique de type déterministe comme TI dans un contexte d'exploitation minière.

En hydrogéologie régionale, la pratique quasi universelle consiste à élaborer un modèle déterministe intégrant toutes les informations disponibles : environnement sédimentaire et contexte

géologique, stratigraphie connue, âges des sédiments et de l'eau, géomorphologie, géologie de surface, quelques forages de qualité échantillonnés, plusieurs forages de moindre qualité provenant de bases de données publiques, tests de pompage et de traceur et données géophysiques parfois disponibles dans une partie de la région. Il existe des exemples d'utilisation de données géophysiques via MPS [249, 250]. En général, les informations disponibles ne sont pas de même nature que les données sismiques des études pétrolières. Elles sont complexes avec une couverture variable et souvent très réduite, sur des supports variés, parfois seulement formulés comme des connaissances d'experts subjectives, de sorte qu'elles ne peuvent pas être intégrées directement avec une TI représentant simplement la texture, comme dans le cas du domaine pétrolier. Généralement, elles doivent être incorporées dans le modèle déterministe grâce à l'expertise du géologue. Malgré toute l'expertise impliquée dans l'élaboration du modèle déterministe, celui-ci reste néanmoins une interprétation assez idéalisée et lissée du terrain réel. Soulignons donc que le modèle déterministe n'est pas le champ réel. Si la stratigraphie et les proportions globales de chaque UHS sont probablement bien représentées dans le modèle déterministe, les épaisseurs verticales des UHS sont probablement beaucoup plus variables que ce qui est supposé dans le modèle déterministe. De même, la continuité latérale de l'UHS est probablement exagérée dans le modèle déterministe. Comme ces variations d'épaisseur et latérales peuvent influencer de manière significative la réponse du modèle aux flux, il est important de considérer d'autres modèles autour du modèle déterministe pour pouvoir évaluer l'incertitude du flux hydraulique et du transport.

En somme, l'utilisation d'un modèle déterministe comme TI ajoute une difficulté supplémentaire puisqu'il s'agit d'une interprétation idéalisée faite sur des données limitées en quantité et en distribution spatiale. Le modèle déterministe ne représente pas la vérité, il s'agit plutôt d'une interprétation idéalisée et lissée de la réalité. On peut s'attendre à ce que les proportions et la stratigraphie verticale soient bien représentées dans le modèle déterministe, mais on ne peut pas s'attendre à ce que les statistiques d'ordre supérieur (e.g. statistiques multipoints) et la variabilité locale soient représentées fidèlement par le modèle déterministe. Une telle TI est donc probablement plus lisse que le champ réel. Par conséquent, il n'est pas souhaitable de reproduire très étroitement les statistiques en points multiples de la TI. Dans cette étude, la méthode MCP est appliquée avec une seule TI. Il n'est donc pas possible obtenir un "vrai" paramètre de variabilité local ou global de celle-ci. Il s'agit d'une limite commune à toutes les méthodes basées sur une seule TI. Pour éviter de sous-estimer l'incertitude (absence de l'inférence d'un hyper-modèle), il semble judicieux de faire des choix qui maximisent le critère de variance C2 en assurant simultanément un C1 faible et sans artefacts évidents tels que des inversions d'unités ou des discontinuités irréalistes. Rappelons que le critère C1 mesure la différence entre les proportions moyennes des UHS de la TI et les proportions des réalisations

et C2 mesure la variabilité des proportions des UHS simulées entre les différentes réalisations (section 4.3.2).

### 6.1.3 Comparaison de la méthode MCP avec la méthode MPS

Dans l'exemple d'application pour le contexte de Simcoe, il est intéressant de constater que la méthode MCP fournit de bons résultats malgré une TI fortement non stationnaire. Un tel contexte aurait demandé beaucoup plus d'effort pour arriver à des résultats similaires avec une autre méthode telles que PGS ou MPS. De plus, rien n'aurait garanti que ces méthodes respecteraient l'ordre stratigraphique des UHS. À titre d'exemple, le reste de la section établit une comparaison entre la méthode MCP et la méthode MPS par motifs (MPS-motifs) ainsi qu'avec la méthode "direct sampling" (DS).

Pour la MPS-motifs, deux problèmes se posent. Le premier est que la méthode de "quilting" en 3D proposée par Mahmud et al. [251] n'est pas aussi efficace que celle développée par Efros et Freeman [252] pour le 2D. Il est donc assez difficile d'éviter les discontinuités dans le cadre du 3D. Des tests indiquent que celles-ci seraient très présentes. Le deuxième problème est que la variabilité basée sur la MPS-motifs, lorsqu'appliquée en 2D, est faible en raison de la non-stationnarité de la TI. La MPS-motifs a tendance à reproduire, en grande partie, des copies exactes de la TI, car les répétitions d'événements de données similaires dans la TI sont rares, voire uniques. La donnée force, d'une certaine manière, à choisir le même motif ou un motif très similaire pour de nombreuses réalisations, réduisant ainsi la variabilité. Enfin, il n'y a pas de moyen évident d'empêcher le collage complet d'un motif qui crée des inversions d'unité par rapport à d'autres motifs voisins. La figure 6.1 montre, sur quatre réalisations distinctes, la difficulté de la méthode avec les inversions d'unités indiquées avec les flèches rouges.

Par rapport à ces résultats, l'approche DS de Mariethoz et al. [47] donnent des résultats légèrement meilleurs. Noter que les tests ont demandé de modifier le code DS original pour tester les inversions d'unités. La comparaison de MCP à DS conduit aux observations suivantes. D'abord, plus d'artefacts sont obtenus avec DS, en particulier à proximité des cellules périphériques. Les critères de qualité C1 et C2 étaient tous deux plus faibles avec DS qu'avec MCP. Ce résultat est un bon point pour C1, mais un moins bon pour C2, surtout si l'on considère que les artefacts périphériques sont responsables d'une part importante de la variabilité observée. Une valeur moindre de C2 indique une diminution de la variabilité des proportions des UHS parmi l'ensemble des réalisations réduisant ainsi l'étendue de l'incertitude. Il est donc possible de sous-estimer l'incertitude, ce qui peut entraîner un impact potentiel sur l'évaluation du risque.

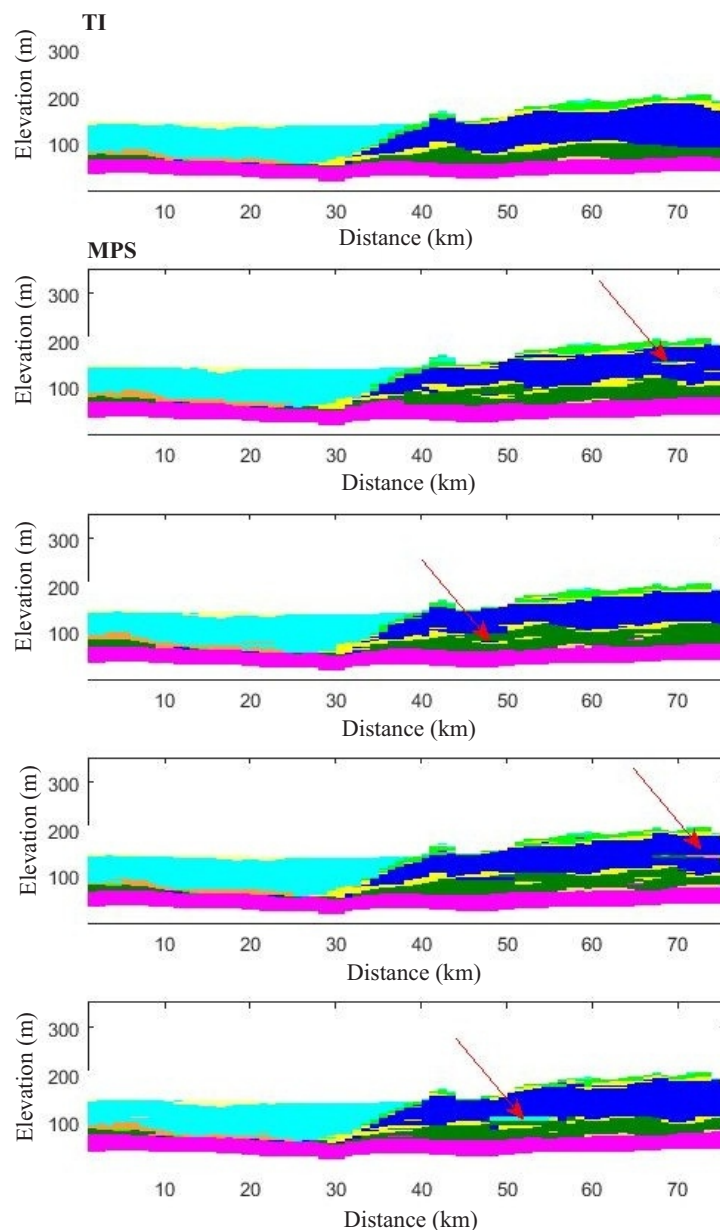


Figure 6.1 Image d'entraînement et quatre réalisations MPS-motifs montrant des inversions d'UHS.

Le contrôle d'inversion de la stratigraphie des UHS a également été activé dix fois plus fréquemment, soit 3% des points simulés avec MPS contre 0,3% des points simulés avec MCP. Enfin, le temps de calcul de DS était environ 60 fois plus lent qu'avec MCP. Pour ce point, MCP à l'avantage de ne pas avoir à scanner de liste d'événements de données, car les probabilités sont directement lues dans les matrices de probabilités bivariées. La figure 6.2 montre trois réalisations obtenues par DS. On peut noter les artefacts sur les côtés gauche

et droit et sur la base. Au centre du modèle, on constate le peu de variations.

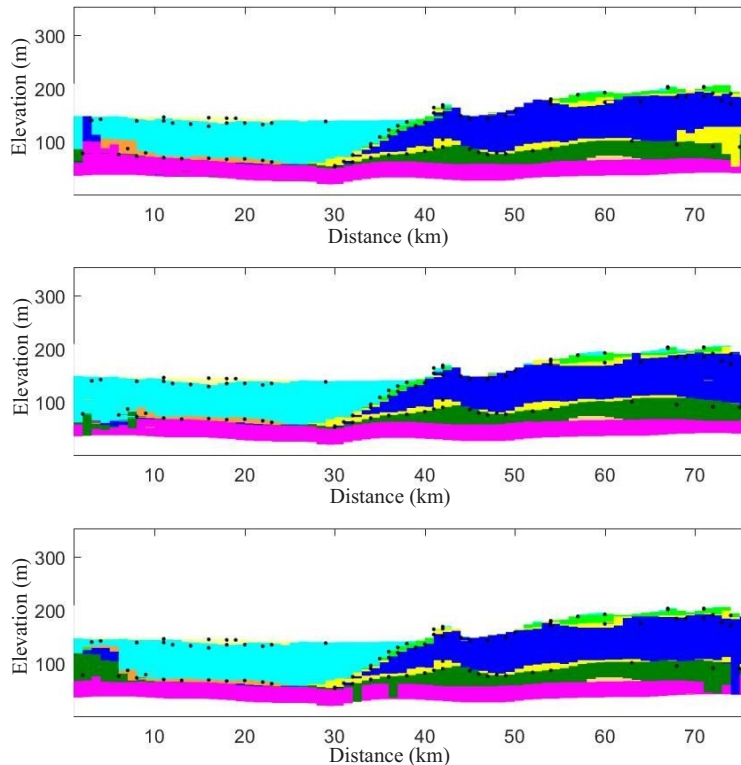


Figure 6.2 Réalisations MPS-DS montrant des artefacts dans la périphérie et la quasi absence de variabilité dans la partie centrale.

#### 6.1.4 Autres remarques sur la méthode MCP

La méthodologie de simulation directionnelle 3D à partir de la méthode MCP rencontre ainsi l'un des objectifs spécifiques de la thèse. Les développements ont permis de proposer des modèles hydrostratigraphiques équivalents complexes d'aspect réaliste et de quantifier leur incertitude. L'approche permet le conditionnement des données d'observation, la reproduction de la cartographie de surface, inclut les connaissances du géologue en intégrant les structures et les caractéristiques du modèle conceptuel de manière satisfaisante. Ces résultats sont d'autant plus satisfaisants que la programmation et l'application de la méthode sont relativement simples comparées à d'autres méthodes basées sur les probabilités de transition, comme la méthode BME [105, 107, 253]. Comme pour la méthode BME, la méthode MCP a l'avantage de forcer la probabilité 0 ou 1 et ainsi contraindre la direction de simulation des unités. La MCP a toutefois l'avantage d'être simple et non limitée par le nombre d'unités. De plus, contrairement à d'autres méthodes de simulation géostatistique, MCP s'accommode des tendances (non-stationnarité) sans la nécessiter de fournir des variables auxiliaires. Pour

traiter la non-stationnarité avec PGS, comme celle du comté de Simcoe, il aurait été nécessaire d'imposer une tendance directionnelle avec des champs auxiliaires. De plus, il aurait probablement fallu segmenter le domaine étudié en sous-domaines homogènes. La propriété de forçage 0/1 de MCP permet de se passer de tout ça, ce qui facilite grandement l'analyse.

Bien que l'on ne connaisse pas les proportions exactes de chaque UHS dans le cas réel, il a été démontré que MCP permettait de créer des modèles alternatifs autour du modèle déterministe considéré comme le plus près de la réalité pour les proportions des UHS et leur agencement dans l'espace. Les modèles alternatifs ainsi créés sont sans biais par rapport au modèle déterministe et permettent de proposer des scénarios quantifiant l'incertitude relative au modèle déterministe actuel et conditionnelle à celui-ci et aux données conditionnantes. Certes, il reste à vérifier si ces bons résultats peuvent être extrapolés à d'autres cas complexes pour notamment comparer les résultats en matière de biais et de robustesse à l'échantillonnage préférentiel. Il faut retenir que la MCP est affectée par la configuration du voisinage des données ainsi que le choix de la méthode de recherche. Ce dernier affecte non seulement la variabilité des solutions, mais aussi la possibilité de rares inversions d'unité simulées lesquelles peuvent toutefois être facilement corrigées. Terminons par mentionner que les modèles équivalents générés par MCP peuvent directement être utilisés avec un simulateur d'écoulement ou de transport et ainsi quantifier les effets associés à leurs incertitudes.

## 6.2 Caractérisation stochastique des hétérogénéités intra unités hydrostratigraphiques ( $K$ )

Les résultats du chapitre 5 portent sur caractérisation stochastique et le calage de la conductivité hydraulique. Il s'agit d'une méthodologie qui inclut, en deux temps, la mise à l'échelle et la régionalisation du tenseur- $K$  à partir de simulation ponctuelle de la  $K$  locale. À cela s'ajoute une nouvelle transformation bivariée qui permet la préservation des relations non linéaires entre les composantes des tenseurs- $K$ . L'approche permet la calibration de tenseurs- $K$  structurés spatialement. La structure spatiale réelle des tenseurs- $K$  ne peut être connue de manière exacte. Toutefois, elle est modélisée par les variogrammes 3D de chaque UHS. Il a démontré que la méthode GDM permettait de créer des modèles alternatifs préservant cette structure modélisée ainsi que les relations entre les composantes. Ces modèles alternatifs permettent de proposer des scénarios quantifiant l'incertitude relative des structures modélisées. Les sous-sections suivantes discutent des particularités observées avec la méthodologie proposée de caractérisation stochastique des hétérogénéités intra UHS soit : i) la mise à l'échelle d'hétérogénéités locales ( $K$  quasi point) stochastiques en tenseur- $K$  équivalent (bloc), ii) l'utilisation du modèle linéaire de corégionalisation pour la simulation des composantes

des tenseurs- $K$  équiprobables, iii) la préservation des corrélations entre les composantes des tenseurs- $K$  par nouvelle transformation gaussienne bivariée, iv) l'utilisation de la méthode de déformation graduelle (GDM) pour caler les tenseurs- $K$  en fonction des observations de charge hydraulique, v) la comparaison du calage par GDM et PEST pour caler les tenseurs- $K$  et définir des zones de captage et vi) quelques autres remarques sur la méthodologie.

### 6.2.1 Mise à l'échelle d'hétérogénéités locales ( $K$ )

La méthode de mise à l'échelle de la  $K$  scalaire en tenseur- $K$  équivalent tient compte de la structure des hétérogénéités locales et des propriétés statistiques de chacune des UHS, car ces hétérogénéités sont présentes dans le champ simulé de la  $K$  scalaire. L'opération est coûteuse en temps de calcul, car elle nécessite plusieurs appels du simulateur d'écoulement. Il est important d'utiliser un très grand champ de  $K$  scalaire pour éviter les effets des conditions frontières et aussi assurer la continuité avec les blocs voisins. Dans le cas d'un modèle régional comptant des centaines de milliers ou des millions de blocs, cette méthode s'avère trop lourde. Donc, le défi était de limiter le nombre de blocs à mettre à l'échelle. L'idée adoptée est de simuler directement les tenseurs- $K$  par simulation géostatistique (non-conditionnelle) basée sur un modèle linéaire de corégionalisation.

### 6.2.2 Simulation du tenseur- $K$ avec modèles linéaires de corégionalisation

La régionalisation des composantes des tenseurs- $K$  en champs équivalents est faite par simulation géostatistique à partir de modèles linéaires de corégionalisation (LCM). En plus de rendre le temps de calcul raisonnable, ces simulations permettent la production des champs équivalents de chacune des composantes des tenseurs- $K$ . Aussi pour chacune des unités, les caractéristiques statistiques des composantes de  $K$  sont reproduites. Il est important de retenir que les tenseurs- $K$  mis à échelle étaient non-gaussiens et qu'ils présentaient des relations non linéaires entre les composantes des tenseurs. Comme l'approche proposée utilise un LCM dans l'espace gaussien, les tenseurs ont dû d'abord subir une transformation gaussienne pour déterminer le modèle LCM. Pour rétablir les corrélations non linéaires à distance zéro ainsi que les histogrammes des composantes souhaités, une nouvelle transformation bivariée des réalisations du LCM gaussien, discutée à la section suivante, a été ajoutée à la démarche.

### 6.2.3 Préservation des corrélations inter composants de tenseurs- $K$

Les corrélations non linéaires des composants de tenseurs- $K$  sont reproduites grâce à une nouvelle transformation bivariable à partir des modèles de référence de chacune des UHS.



La préservation des corrélations entre les composants s'applique aux résultats de simulation LCM, mais aussi à chacune des itérations de la GDM. La raison est simple, durant la déformation la relation doit demeurer non-linéaire entre les composantes du tenseur et GDM n'assure pas la préservation de cette relation.

La transformation bivariable proposée accomplit sa tâche et les résultats obtenus sont très satisfaisants (chapitre 5). Toutefois, cette transformation comporte tout de même quelques inconvénients. D'abord, la transformation inverse n'est pas possible. À noter toutefois que la transformation inverse n'est jamais requise dans la simulation LCM ou dans le GDM car des simulations non-conditionnelles sont produites. De plus, la méthode procure des résultats correspondants à des valeurs de référence. Il y a donc dépendance à ces valeurs et à leur variabilité. Pour éviter ces inconvénients, il serait aussi possible d'envisager d'utiliser d'autres approches pour la transformation bivariable.

La revue de la littérature n'aborde pas la discussion des méthodes de transformation. Cependant, il est intéressant de soulever qu'une simple transformation gaussienne des deux variables ( $K_{xx}$  et  $K_{zz}$ ) aurait suffi à éliminer la non-linéarité lors des simulations. Toutefois, elle n'aurait pas assuré de préserver la relation dure recherchée où  $K_{zz} \leq K_{xx}$ . Il existe d'autres approches pour la transformation qui aurait pu être considérée comme la "projection poursuit multivariate transform (PPMT)" [236]. L'avantage de cette méthode est que la transformation inverse redonne exactement les mêmes valeurs contrairement à l'approche utilisée où les valeurs inversées ne sont pas tout à fait exactes. Quelques tests indiquent que globalement les distributions sont un peu mieux reproduites (figure 6.3) qu'avec la méthode proposée au chapitre 5. Il faut toutefois noter que PPMT ne garantit pas la reproduction de certaines relations dures comme  $K_{zz} \leq K_{xx}$  contrairement à la transformation bivariable utilisée. À titre d'exemple, une transformation PPMT sur une réalisation comprenant 21% (1893/9000) de valeurs  $K_{zz} > K_{xx}$  montre, après transformation, que 2.1% (192/9000) des  $K_{zz}$  demeurent supérieures à  $K_{xx}$ .

#### 6.2.4 Utilisation de la GDM pour caler les tenseurs- $K$

Le calage d'un modèle hydrogéologique a pour objectifs de perturber des propriétés (tenseurs- $K$ ) d'un modèle afin de bien reproduire des observations (charge hydraulique). Le défi était de caler un modèle représentant un système régional d'écoulement des eaux souterraines dans un modèle 3D multicouches très hétérogènes. De plus, le calage devait préserver les covariances spatiales et les corrélations non linéaires entre les composantes (section 6.2.3). Pour le cas considéré et sujet aux hypothèses adoptées et aux limitations décrites à la section 6.3, les résultats indiquent que l'hypothèse selon laquelle la GDM serait un bon choix s'est avérée

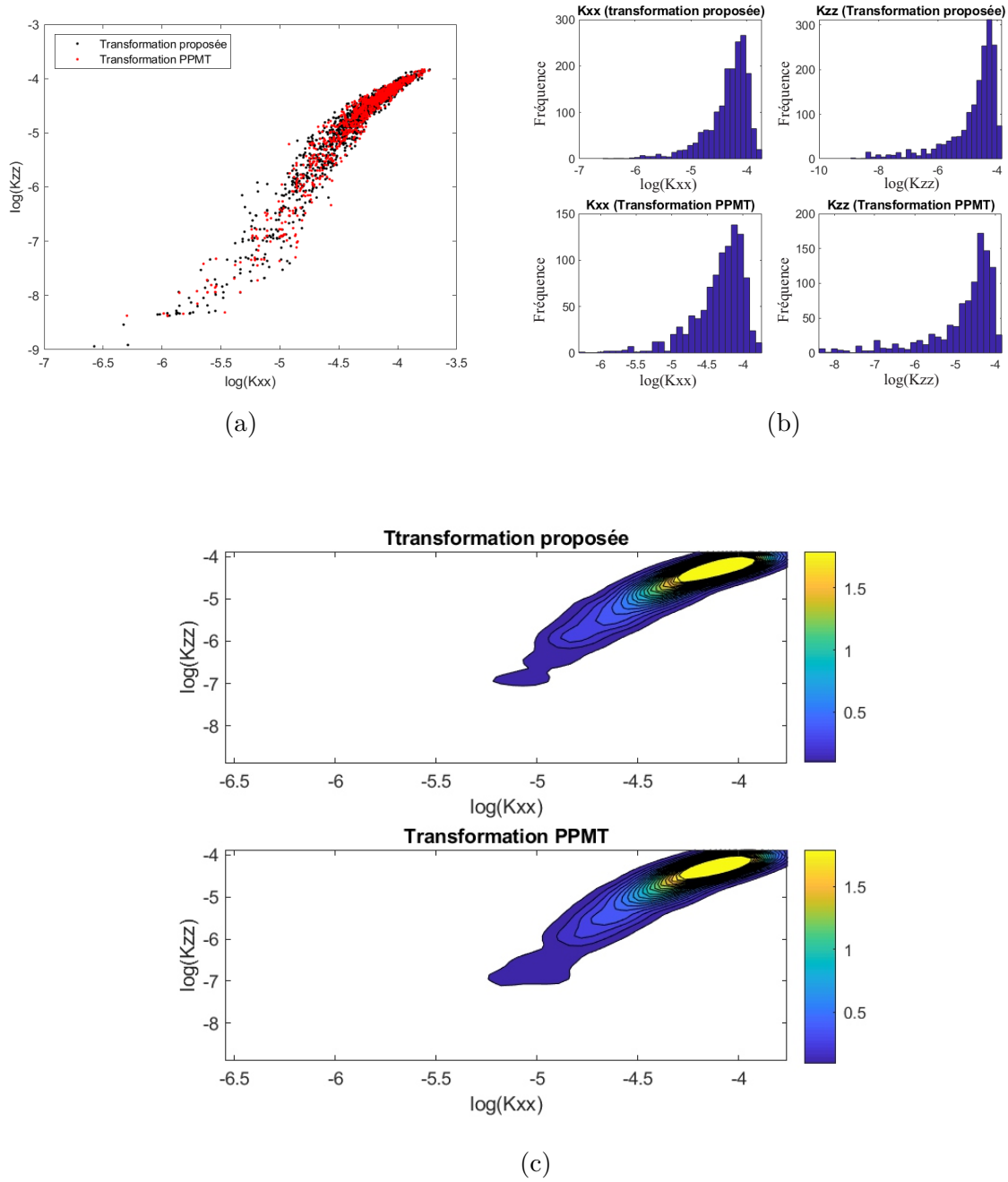


Figure 6.3 Comparaison de valeurs du  $\log(K_{xx})$  et  $\log(K_{zz})$  de l'unité AFB1 obtenues par la transformation proposée avec celles transformées par PPMT selon : a) un diagramme binaire, b) leurs histogrammes respectifs, c) et leurs distributions bivariées estimées par densité de noyau.

juste. En effet, l'approche proposée a permis de préserver les propriétés spatiales des ten-

seurs ainsi que les relations entre les composantes. Après calage, les champs de chacune des composantes du tenseur respectent les mêmes caractéristiques que les champs simulés. Les résultats de l'approche développée démontrent que la GDM s'applique à un contexte multivariable contraint et pas seulement au contexte univariable. Évidemment, l'inconvénient de la méthode réside dans sa lenteur attribuable aux nombreux appels du simulateur d'écoulement (méthode d'inversion indirecte). Malgré cela, les inversions convergent rapidement en termes du nombre d'itérations grâce à l'optimisation du paramètre de déformation  $t$  de l'équation 5.2.

Il a aussi été montré que la méthode permet aussi de caler la recharge séparément ou simultanément avec les tenseurs- $K$ . Les tests ont permis d'observer que la recharge a un impact sur le résultat des tenseurs- $K$  calibrés. À cet effet, l'utilisation d'une recharge stochastique a produit des champs calibrés avec des variogrammes plus proches des variogrammes du modèle cible et ce pour les différentes composantes du tenseur- $K$ . Donc, lorsque c'est possible, il est avantageux d'inclure la recharge dans la calibration.

Autre point intéressant, selon les différents tests, un nombre élevé d'observations de charge ( $n > 10$ ) n'améliore pas significativement les résultats. Ce point est à considérer lorsque vient le temps de planifier l'acquisition ou l'ajout de données. Aussi, la méthode produit des champs calés de tenseurs- $K$  différents selon le germe aléatoire utilisé. Il est ainsi possible de caler  $n$  modèles de tenseurs- $K$  équivalents. Cet ensemble de modèles permet de quantifier l'incertitude associée à la variabilité des tenseurs- $K$  calés.

Pour terminer, rappelons que la portée des énoncées de cette section est valide pour le cas considéré et sujet aux hypothèses adoptées et aux limitations décrites à la section 6.3.

### 6.2.5 Comparaison du calage par GDM et PEST

Les résultats GDM ont été comparés avec la méthode d'inversion PEST. La comparaison est basée sur la définition de zones de captage autour de deux puits de pompage. Il a été observé que les tenseurs- $K$  pour les différentes UHS sont géologiquement plus réalistes avec GDM que ceux obtenus avec PEST. Autre point intéressant, l'aspect stochastique de GDM conduit à des zones de captage plus grandes et possiblement plus conservatrices qu'avec un seul modèle calé par PEST.

Comme dans le cas de la comparaison MCP et MPS, il faut mentionner ici que les comparaisons décrites dans cette section sont délicates. En effet, PEST a été appliquée avec des déformations par unités (zone) comparativement à GDM dont la déformation est appliquée à chacun des éléments. Il est possible que d'accroître le nombre de zones dans PEST eût

permis d'améliorer les résultats.

Pour terminer, soulignons que l'approche proposée est indiquée pour quantifier l'incertitude des hétérogénéités intra unités et fournir des modèles équivalents pour simuler l'écoulement et le transport. Pour de futurs développements, il serait intéressant de combiner les approches proposées dans cette thèse (MCP et GDM). Cela permettrait notamment de comparer l'effet individuel et combiné de l'incertitude associée aux hétérogénéités inter et intra unités sur l'écoulement et le transport.

### 6.2.6 Autres remarques sur la méthodologie

L'évaluation de l'incertitude par inversion indirecte à partir de plusieurs modèles équivalents est un processus qui prend beaucoup de temps de calcul. Hermans et al. [254] proposent d'éviter le processus d'inversion en utilisant une approche Bayésienne appelée "Bayesian Evidential Learning". Pour ce faire, ils proposent d'utiliser plusieurs modèles équivalents sans se soucier du calage. Une première application avec données réelles dans un contexte hydrogéologique montre le potentiel prometteur de cette méthode [255].

Une autre difficulté pour les méthode d'inversion indirecte porte sur le conditionnement de données. Avec la méthode MCP, les données de faciès sont implicitement conditionnées, car elle sont intégrées à la simulation. Dans le cas de la caractérisation de la  $K$  des modèles régionaux, le conditionnement n'est pas nécessaire, car la mise à l'échelle fournit des tenseurs- $K$  qui sont à une échelle beaucoup plus grande que les données. Ce qui justifie d'ailleurs l'utilisation de simulations non conditionnelles des tenseurs.

## 6.3 Limites des approches proposées

L'interprétation des résultats associée aux approches proposées dans cette thèse comporte certaines limites. D'abord, il est important de souligner que l'objectif principal de la modélisation stochastique consiste à produire des modèles équivalents qui reflètent les données et les connaissances utilisées. Ces modèles n'ont pas la prétention de représenter une réalité inconnue. Comme pour un modèle déterministe, ils sont utiles pour effectuer des prédictions. L'ensemble des modèles équivalents permet de fournir une approximation de l'incertitude des données et des connaissances modélisées. Évidemment, rien ne garantit la validité des modèles lorsque de l'on ajoute de l'information supplémentaire. L'intégration de la nouvelle information peut demander de redéfinir les modèles. Dans une approche stochastique, l'ajout de nouvelles informations devrait réduire l'espace des solutions et ainsi l'incertitude relative du modèle. Il est important de mentionner que la notion d'incertitude dans cette thèse fait

référence à l'incertitude des modèles, pas celle de la réalité inconnue. Par exemple, l'ajout de données avec la méthode MCP réduit la variabilité (incertitude) des réalisations, mais cela ne veut pas dire que l'on représente mieux une réalité inconnue. En effet, on pourrait très bien avoir un très grand nombre de données se traduisant par de très faibles variations entre les réalisations, mais si la TI est complètement erronée, ces réalisations ne seraient pas précises. Donc la variabilité des réalisations reflète l'incertitude du modèle et non pas celle d'une réalité. Noter qu'il pourrait être intéressant d'évaluer la capacité de la méthode MCP à prédire de manière probabiliste un ensemble de données non conditionnantes. Cela permettrait d'évaluer la robustesse des réalisations à l'ajout d'information complémentaire. En complément des critères C1 et C2, l'utilisation d'une validation croisée permettrait, par exemple, d'évaluer plus en profondeur la performance de la simulation pour mieux juger la qualité des réalisations. Toutefois, le nombre de données disponibles n'était pas jugé suffisant ici pour permettre une évaluation fiable de la performance du modèle basé sur une validation croisée.

Les résultats du calage des tenseurs- $K$  et de la recharge par GDM portent sur un modèle d'écoulement qui utilise des observations synthétiques de charges hydrauliques. Bien que la simulation des tenseurs- $K$  est basée sur de vraies données de  $K$ , le modèle d'éléments finis demeure une version simplifiée des conditions hydrogéologiques des Basse-Terre du bassin Innisfil Creek notamment les conditions frontières du modèle. Par exemple, un gradient régional aurait pu être ajouté pour plus de réalisme. Fort probablement que l'ajout d'un gradient régional aurait déformé davantage les aires d'alimentation de l'exemple présenté à la section 5.6.2. Même si les résultats proviennent d'un modèle simplifié, la démonstration montre tout de même que la méthode proposée fournit les données nécessaires pour caractériser l'incertitude des modèles équivalents utilisés. Pour un calage plus robuste des tenseurs- $K$ , il pourrait être intéressant de montrer la performance de la méthodologie proposée à prédire des données non utilisées lors du calage.

Rappelons en terminant que l'incertitude des simulations MCP ou celle associée aux tenseurs- $K$  calés est représentative du modèle, des données et des connaissances utilisées, en particulier le modèle déterministe ayant servi à calculer les transiogrammes et les identifications des unités. Toute modification substantielle au modèle déterministe et toute réinterprétation des données demanderait de reprendre la simulation permettant la modélisation de l'incertitude.

## CHAPITRE 7 CONCLUSION ET RECOMMANDATION

Les hydrogéologues utilisent en général un seul modèle 3D, considéré comme le "meilleur modèle possible" basé sur les connaissances, les données disponibles et le jugement professionnel. Pour qu'un expert soit en mesure de prendre des décisions éclairées sur la base d'un modèle, la connaissance de l'incertitude relative à ce modèle est nécessaire. Pourtant les méthodes de simulations géostatistiques, par exemple, permettent de proposer des modèles équivalents et ainsi quantifier l'incertitude. Ces méthodes présentent toutefois certaines difficultés telles que générer des modèles d'aspects suffisamment réalistes, respecter les connaissances présentes dans le modèle conceptuel stratigraphique ainsi que reproduire les données dynamiques d'un modèle d'écoulement. Dans le cas particulier, mais non exclusif des modèles régionaux, les modèles conceptuels comprennent généralement une stratification directionnelle verticale fixe des unités qui est difficile à respecter par les méthodes de simulation géostatistique sans manipulation complexe ou l'ajout de variables. À cela s'ajoute, la caractérisation régionale interne des unités hydrostratigraphiques à partir de la conductivité hydraulique. Cette caractérisation est également difficile puisqu'elle demande un changement d'échelle qui doit tenir compte de la variabilité spatiale à petite échelle vers un tenseur- $K$  équivalent à l'échelle des éléments des modèles. De plus, ces tenseurs doivent caler les observations hydrauliques lorsqu'utilisés avec un simulateur d'écoulement. Donc l'objectif principal de la thèse était de développer une démarche complète de caractérisation stochastique des hétérogénéités hydrostratigraphiques appliquée à la modélisation hydrogéologique régionale.

La méthodologie développée permet de générer des modèles équivalents de l'hydrostratigraphie et des tenseurs- $K$  ainsi que quantifier leur incertitude respective. Ces modèles permettent également de quantifier l'effet de l'incertitude sur l'écoulement et le transport de l'eau souterraine. Les diverses étapes du développement méthodologique ont été appliquées et validées avec succès au contexte hydrostratigraphique du sous-bassin versant d'Innisfil Creek. La caractérisation stochastique efficace des hétérogénéités inter unités hydrostratigraphiques et de leur incertitude a été atteinte à partir de la méthode MCP. Les résultats ont montré que l'approche proposée est efficace et permet de produire des modèles hydrostratigraphiques équivalents d'aspect réaliste qui tiennent compte de la sédimentation directionnelle des unités typiques des modèles à l'échelle régionale. L'approche de caractérisation stochastique des hétérogénéités inter unités est quant à elle basée sur le développement d'une méthodologie multi étapes. Cette dernière consiste, d'une part, à mettre à l'échelle la  $K$  quasi ponctuelle des données vers un tenseur- $K$  équivalent à l'échelle de l'élément. Ensuite, les structures spatiales des tenseurs et leurs corrélations inter composantes sont définies et régionalisées

à partir d'un nombre restreint d'éléments mis à l'échelle. La dernière étape consiste à caler les composantes des tenseurs- $K$  tout en préservant leurs structures spatiales, la contrainte d'inégalité  $K_{verticale} \leq K_{horizontale}$  ainsi que les corrélations.

Certaines caractéristiques particulières aux méthodes utilisées ont aidé significativement à atteindre les objectifs de la thèse. Soulignons quelques-unes de ces caractéristiques. Premièrement, la méthodologie de simulation hydrostratigraphique proposée (MCP) permet d'intégrer, via les probabilités bivariées, les connaissances géologiques par l'entremise d'une image d'entraînement complexe (conceptuelle ou déterministe) laquelle ne semble pas limitée par le nombre d'unités ni par les variations importantes de leurs proportions. La simulation MCP permet de produire un ensemble  $n$  de modèles équivalents permettant ainsi de quantifier l'incertitude spatiale des hétérogénéités des unités hydrostratigraphiques. Une caractéristique pratique de la méthode MCP est qu'elle ne nécessite aucun effort pour le conditionnement des données. En effet, ces dernières sont implicitement conditionnées, car elles sont directement intégrées à la simulation. De plus, la propriété de forçage 0/1 des probabilités combinées à la simulation séquentielle prévient l'inversion des unités telle que dictée par l'image d'entraînement. D'autre part, la méthodologie proposée de caractérisation des tenseurs- $K$ , à partir de  $K$  scalaire, permet également de produire un ensemble  $n$  de modèles équivalents des tenseurs calés sur la charge hydraulique. Évidemment, le calage pourrait s'étendre à d'autres variables d'état comme le débit de base de rivières. Le calage par la méthode des déformations graduelles permet de quantifier l'incertitude des tenseurs- $K$  et leur effet sur la modélisation hydrogéologique. Il est important de souligner que l'incertitude des résultats est représentative du modèle, des données et des connaissances utilisées et non pas d'une réalité inconnue. La portée des conclusions est valide pour le cas considéré et sujet aux hypothèses adoptées et aux limitations décrites à la section 6.3.

Parmi l'ensemble des résultats obtenus durant la thèse, voici quelques points forts à retenir. D'abord, il s'agit d'une première application de la méthode MCP pour la simulation 3D. La méthode MCP simule efficacement un contexte régional complexe d'une succession stratigraphique d'un bassin sédimentaire glaciaire. Clairement, la méthode prend en charge la forte non-stationnarité spatiale des unités ainsi que d'imposantes contraintes comme la déposition verticale ordonnée en tenant compte de l'asymétrie stratigraphique à partir des probabilités de transition entre les unités. À ma connaissance, aucune autre méthode de simulation stochastique ne semble être en mesure de solutionner avec autant de satisfaction un contexte hydrostratigraphique complexe 3D comme celui du sous-bassin Innisfil Creek. La méthode MCP s'est révélée sans biais dans le cas non conditionnel même lorsque de fortes tendances sont présentes dans la TI. La méthodologie proposée de caractérisation des tenseurs- $K$  permet le calage individuel ou simultané des tenseurs- $K$  de différentes UHS et/ou de recharge.

La méthodologie inclut une mise à l'échelle des conductivités hydrauliques quasi ponctuelles vers un tenseur équivalent en tenant compte de la structure spatiale à l'échelle locale. La mise à l'échelle est intensive en temps de calcul, mais il a été démontré les tenseurs- $K$  pouvaient inclure une structure de covariance spatiale permettant de limiter le nombre de blocs à mettre à l'échelle grâce à la régionalisation par simulation géostatistique. La méthode des déformations graduelles permet le calage efficace des composantes des tenseurs- $K$  avec préservation des corrélations non linéaires des composantes et de leurs fonctions de covariance. À ma connaissance, il s'agit de la première approche utilisant la GDM pour un contexte multivariable contraint. La stratégie de sélection automatique du coefficient de déformation  $t$  permet d'accélérer le calage.

Pour faire suite à ces dernières conclusions, voici quelques recommandations et quelques perspectives potentielles de travaux futurs. Puisque la thèse était orientée sur le développement méthodologique, la méthode MCP a été testée uniquement pour une petite partie du bassin versant d'Innisfil Creek. Toutefois, il serait intéressant de tester la méthode sur le bassin complet et caractériser l'effet de l'incertitude sur l'écoulement et le transport pour différents scénarios d'intérêt. Aussi, il serait profitable de tester la simulation MCP dans un contexte différent de celui de Simcoe et comparer les observations et les conclusions de la thèse. Dans certains contextes, notamment lorsqu'un modèle déterministe n'est pas disponible, l'image d'entraînement peut comporter une large part d'incertitude. Il serait donc intéressant de tester l'influence du choix de la TI sur la performance de la méthode, les résultats et la variabilité des réalisations. Il pourrait même s'avérer avantageux d'utiliser une combinaison de plusieurs TI pour un même problème pour mieux caractériser l'incertitude et la variabilité des modèles équivalents. D'un point de vue pratique, l'algorithme MCP développé fonctionne sur des grilles régulières dont les éléments sont identiques pour notamment accélérer les temps de calcul et faciliter la recherche de voisins lors de la simulation. Toutefois, le maillage classique des modèles d'éléments finis est généralement basé sur des prismes triangulaires de dimensions variées. Il serait utile de proposer un algorithme de simulation MCP rapide pour la simulation sur grilles non régulières pour satisfaire ces modèles. D'un point de vue fondamental la notion de biais, lorsque l'hypothèse de l'indépendance conditionnelle de la méthode MCP ne tient pas complètement, reste à préciser. Il serait intéressant d'analyser plus en profondeur cet aspect pour définir plus précisément l'effet du conditionnement sur le biais et ainsi vérifier si les bons résultats concernant le biais et la robustesse à l'échantillonnage préférentiel peuvent être extrapolés à d'autres cas complexes.

Par ailleurs, les calages des modèles d'écoulement et/ou de transport sont à la fois influencés par la distribution des UHS et celle des tenseurs- $K$ . Dans cette optique, il serait intéressant de combiner les simulations MCP au calage des tenseurs- $K$  simulés et de comparer l'effet



de l'incertitude associée aux hétérogénéités inter unités versus celles intra unités et celles combinant les deux. Pour élever l'ensemble de la méthodologie à un autre niveau, il reste à définir des métriques pour synthétiser les résultats de l'approche de modélisation stochastique vers des indicateurs de risque. Ces indicateurs pourraient aider concrètement les experts du domaine dans la prise de décision relative aux impacts directs, indirects ou cumulatifs pour des problématiques hydrogéologiques données. Ainsi, l'approche pourrait être utilisée pour évaluer le risque associé à des problématiques classiques, exemples, la caractérisation de la zone de rabattement d'une mine à ciel ouvert, l'impact du rabattement sur l'eau de surface, la définition de l'aire d'alimentation d'un captage et les effets cumulatifs sur la dynamique d'écoulement ou sur la migration de contaminant.

## RÉFÉRENCES

- [1] M. Honarkhah, “Stochastic simulation of patterns using distance-based pattern modeling,” PhD thesis, Stanford University, Stanford, USA, 2011.
- [2] D. Marcotte, “Fast variogram computation with FFT,” *Computers & Geosciences*, vol. 22, n°. 10, p. 1175–1186, dec 1996. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S009830049600026X>
- [3] P. A. Domenico et F. W. Schwartz, *Physical and Chemical Hydrogeology*, 2<sup>e</sup> éd. New York, NY : Wiley, 1997. [En ligne]. Disponible : <http://www.crcnetbase.com/doi/book/10.1201/9781420058925>
- [4] P. Renard, “Stochastic hydrogeology : What professionals really need?” *Ground Water*, vol. 45, n°. 5, p. 531–541, sep 2007. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1111/j.1745-6584.2007.00340.x/abstract>
- [5] D. M. Tartakovsky, “Assessment and management of risk in subsurface hydrology : A review and perspective,” *Advances in Water Resources*, vol. 51, p. 247–260, jan 2013. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0309170812000917>
- [6] S. P. Neuman, “Stochastic groundwater models in practice,” *Stochastic Environmental Research and Risk Assessment*, vol. 18, n°. 4, p. 268–270, août 2004. [En ligne]. Disponible : <https://doi.org/10.1007/s00477-004-0192-6>
- [7] Y. Rubin, “Stochastic hydrogeology – challenges and misconceptions,” *Stochastic Environmental Research and Risk Assessment*, vol. 18, n°. 4, p. 280–281, août 2004. [En ligne]. Disponible : <https://doi.org/10.1007/s00477-004-0193-5>
- [8] H. Zhou, J. J. Gómez-Hernández et L. Li, “Inverse methods in hydrogeology : Evolution and recent trends,” *Advances in Water Resources*, vol. 63, p. 22–37, jan 2014. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0309170813002017>
- [9] G. Dagan, “An overview of stochastic modeling of groundwater flow and transport : From theory to applications,” *Eos, Transactions American Geophysical Union*, vol. 83, n°. 53, p. 621–625, 2002. [En ligne]. Disponible : <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002EO000421>
- [10] Y. Rubin, C.-F. Chang, J. Chen, K. Cucchi, B. Harken, F. Heße et H. Savoy, “Stochastic hydrogeology’s biggest hurdles analyzed and its big blind spot,” *Hydrology and Earth System Sciences*, vol. 22, n°. 11, p. 5675–5695, nov 2018. [En ligne]. Disponible : <https://www.hydrol-earth-syst-sci.net/22/5675/2018/>

- [11] J. C. Refsgaard, S. Christensen, T. O. Sonnenborg, D. Seifert, A. L. Højberg et L. Trolborg, “Review of strategies for handling geological uncertainty in groundwater flow and transport modeling,” *Advances in Water Resources*, vol. 36, p. 36–50, feb 2012. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0309170811000625>
- [12] H. A. Michael, H. Li, A. Boucher, T. Sun, J. Caers et S. M. Gorelick, “Combining geologic-process models and geostatistics for conditional simulation of 3-D subsurface heterogeneity,” *Water Resources Research*, vol. 46, n°. 5, p. W05527, may 2010. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1029/2009WR008414/abstract>
- [13] A. Walgenwitz, D. Allard et P. Biver, “Flexible objects - a way to generate more realistic object-based simulations,” dans *Petroleum Geostatistics 2015*. European Association of Geoscientists & Engineers, sep 2015, p. cp-456-00 027. [En ligne]. Disponible : <https://www.earthdoc.org/content/papers/10.3997/2214-4609.201413610>
- [14] X. Emery, J. M. Ortiz et A. M. Cáceres, “Geostatistical modelling of rock type domains with spatially varying proportions : Application to a porphyry copper deposit,” *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 108, n°. 5, p. 284–292, may 2008. [En ligne]. Disponible : [http://www.scielo.org.za/scielo.php?script=sci\\_abstract&pid=S2225-62532008000500005&lng=en&nrm=iso&tlng=en](http://www.scielo.org.za/scielo.php?script=sci_abstract&pid=S2225-62532008000500005&lng=en&nrm=iso&tlng=en)
- [15] N. Benoit, D. Marcotte, A. Boucher, D. D’Or, A. Bajc et H. Rezaee, “Directional hydrostratigraphic units simulation using MCP algorithm,” *Stochastic Environmental Research and Risk Assessment*, vol. 32, n°. 5, p. 1435–1455, dec 2017.
- [16] N. Benoit, D. Marcotte, A. Boucher et A. F. Bajc, “Directional geostatistical simulation for regional hydrostratigraphic units uncertainty characterization, Innisfil Creek sub-watershed, Ontario,” Natural Resources Canada, Geological Survey of Canada, Rapport technique Open File 8212, 2017. [En ligne]. Disponible : <https://doi.org/10.4095/299758>
- [17] N. Benoit, D. Marcotte, A. Boucher, D. D’Or et A. Bajc, “Regional hydrostratigraphic unit modelling using directional MCP simulation,” communication présentée à National Groundwater Association, NGWA, Nashville, USA, avr. 2017. [En ligne]. Disponible : <https://ngwa.confex.com/ngwa/2017gws/webprogram/Paper11875.html>
- [18] N. Benoit, M. D. Marcotte et J. Molson, “Stochastic correlated hydraulic conductivity tensor calibration using gradual deformation,” *Journal of Hydrology*, nov. 2019, soumis pour publication.

- [19] N. Benoit, D. Marcotte, J. W. Molson et P. Pasquier, “Geostatistical simulations of the full 3D block hydraulic conductivity tensor considering inner local-scale variability,” dans *12th International Conference on Geostatistics for Environmental Applications*, 2018. [En ligne]. Disponible : <https://2018.geoenvia.org/>
- [20] N. Benoit, D. Marcotte, J. Molson, P. Pasquier et R. Mulligan, “Characterization of full 3D hydraulic conductivity tensors of hydrostratigraphic units applied to Innisfil Creek watershed, Ontario,” Natural Resources Canada, Geological Survey of Canada, Rapport technique, 2018. [En ligne]. Disponible : <https://doi.org/10.4095/306472>
- [21] H. Zhou, L. Li et J. Jaime Gómez-Hernández, “Three-dimensional hydraulic conductivity upscaling in groundwater modeling,” *Computers & Geosciences*, vol. 36, n°. 10, p. 1224–1235, 2010. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0098300410001718>
- [22] A. Jakeman, O. Barreteau, R. Hunt, J.-D. Rinaudo et A. Ross, *Integrated Groundwater Management - Concepts, Approaches and Challenges*. Springer International Publishing, 2016, n°. 1. [En ligne]. Disponible : <http://www.springer.com/us/book/9783319235752>
- [23] J.-P. Chilès et P. Delfiner, *Geostatistics*. Wiley-Blackwell, 2012. [En ligne]. Disponible : [https://www.ebook.de/de/product/15356462/jean\\_paul\\_chil\\_egrave\\_s\\_geostatistics.html](https://www.ebook.de/de/product/15356462/jean_paul_chil_egrave_s_geostatistics.html)
- [24] Y. Zhou et W. Li, “A review of regional groundwater flow modeling,” *Geoscience Frontiers*, vol. 2, n°. 2, p. 205–214, apr 2011. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S167498711100020X>
- [25] F. G. Alabert, “Stochastic imaging of spatial distributions using hard and soft data,” PhD thesis, Stanford University, Stanford, USA, 1987.
- [26] A. G. Journel, *Fundamentals of Geostatistics in Five Lessons*. American Geophysical Union, 1989, p. 2–10. [En ligne]. Disponible : <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/9781118667606.ch0>
- [27] A. G. Journel et E. H. Isaaks, “Conditional indicator simulation : Application to a Saskatchewan uranium deposit,” *Journal of the International Association for Mathematical Geology*, vol. 16, n°. 7, p. 685–718, oct 1984. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/BF01033030>
- [28] G. Matheron, H. Beucher, A. Galli, D. Guérillot et C. Ravenne, “Conditional simulation of the geometry of fluvio-deltaic reservoirs,” dans *62nd Annual Technical Conference and Exhibition of the Society of Petroleum Engineers*, ser. SPE Paper 16753, 1987, p. 591–599.

- [29] G. L. Loc'h, H. Beucher, A. Galli et B. Doligez, "Improvement in the truncated Gaussian method : Combining several Gaussian functions," dans *ECMOR IV - 4th European Conference on the Mathematics of Oil Recovery*. EAGE Publications BV, jun 1994.
- [30] X. Emery, "Simulation of geological domains using the pluriGaussian model : New developments and computer programs," *Computers & Geosciences*, vol. 33, p. 1189–1201, 2007.
- [31] T. L. Chugunova et L. Y. Hu, "Multiple-point simulations constrained by continuous auxiliary data," *Mathematical Geosciences*, vol. 40, n<sup>o</sup>. 2, p. 133–146, jan 2008.
- [32] F. B. Guardiano et R. M. Srivastava, "Multivariate Geostatistics : Beyond Bivariate Moments," dans *Geostatistics Tróia '92 : Volume 1*, ser. Quantitative Geology and Geostatistics, A. Soares, édit. Dordrecht : Springer Netherlands, 1993, p. 133–144. [En ligne]. Disponible : [https://doi.org/10.1007/978-94-011-1739-5\\_12](https://doi.org/10.1007/978-94-011-1739-5_12)
- [33] S. Strebelle, "Conditional simulation of complex geological structures using multiple-point statistics," *Mathematical Geology*, vol. 34, n<sup>o</sup>. 1, p. 1–21, 2002.
- [34] T. Zhang, P. Switzer et A. Journel, "Filter-based classification of training image patterns for spatial simulation," *Mathematical Geology*, vol. 38, n<sup>o</sup>. 1, p. 63–80, jan 2006.
- [35] A. Boucher, "Considering complex training images with search tree partitioning," *Computers & Geosciences*, vol. 35, n<sup>o</sup>. 6, p. 1151–1158, jun 2009. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0098300408002185>
- [36] C. Faucher, A. Saucier et D. Marcotte, "A new patchwork simulation method with control of the local-mean histogram," *Stochastic Environmental Research and Risk Assessment*, vol. 27, n<sup>o</sup>. 1, p. 253–273, jan 2013. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s00477-012-0586-9>
- [37] M. Honarkhah et J. Caers, "Stochastic simulation of patterns using distance-based pattern modeling," *Mathematical Geosciences*, vol. 42, n<sup>o</sup>. 5, p. 487–517, jul 2010. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11004-010-9276-7>
- [38] G. Mariethoz, "A general parallelization strategy for random path based geostatistical simulation methods," *Computers & Geosciences*, vol. 36, n<sup>o</sup>. 7, p. 953–958, jul 2010. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0098300410000774>
- [39] O. Peredo et J. M. Ortiz, "Parallel implementation of simulated annealing to reproduce multiple-point statistics," *Computers & Geosciences*, vol. 37, n<sup>o</sup>. 8, p. 1110–1121, aug 2011. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0098300410003407>

- [40] H. Rezaee, D. Marcotte, P. Tahmasebi et A. Saucier, “Multiple-point geostatistical simulation using enriched pattern databases,” *Stochastic Environmental Research and Risk Assessment*, p. 1–21, oct 2014. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s00477-014-0964-6>
- [41] P. Tahmasebi, A. Hezarkhani et M. Sahimi, “Multiple-point geostatistical modeling based on the cross-correlation functions,” *Computational Geosciences*, vol. 16, n°. 3, p. 779–797, jun 2012. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s10596-012-9287-1>
- [42] A. Astrakova et D. S. Oliver, “Conditioning Truncated Pluri-Gaussian Models to Facies Observations in Ensemble-Kalman-Based Data Assimilation,” *Mathematical Geosciences*, p. 1–23, apr 2014. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11004-014-9532-3>
- [43] H. Beucher, J. Rivoirard, I. Cojan et F. Ors, “Contribution of the process-based modeling to pluri-Gaussian simulation in the case of meandering channelized reservoirs,” AAPG 2012 Annual Convention & Exhibition, Long Beach, USA, apr 2012. [En ligne]. Disponible : <https://hal-mines-paristech.archives-ouvertes.fr/hal-00777050>
- [44] J. L. Deutsch et C. V. Deutsch, “A multidimensional scaling approach to enforce reproduction of transition probabilities in truncated pluriGaussian simulation,” *Stochastic Environmental Research and Risk Assessment*, vol. 28, n°. 3, p. 707–716, mar 2014. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s00477-013-0783-1>
- [45] T. Mannseth, “Relation Between Level Set and Truncated Pluri-Gaussian Methodologies for Facies Representation,” *Mathematical Geosciences*, vol. 46, n°. 6, p. 711–731, aug 2014. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11004-013-9507-9>
- [46] D. Allard, A. Comunian et P. Renard, “Probability aggregation methods in geoscience,” *Mathematical Geosciences*, vol. 44, n°. 5, p. 545–581, juill. 2012.
- [47] G. Mariethoz, P. Renard et J. Straubhaar, “The direct sampling method to perform multiple-point geostatistical simulations,” *Water Resources Research*, vol. 46, n°. 11, p. W11536, nov 2010. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1029/2008WR007621/abstract>
- [48] W. K. Hamblin et E. H. Christiansen, *Earth’s dynamic systems.*, 6<sup>e</sup> éd. Upper Saddle River, NJ : Prentice Hall, 2001.

- [49] G. Le Loc'h et A. Galli, "Improvement in the truncated Gaussian method : Combining several Gaussian functions," dans *Ecmor 4, 4th European Conference on the Mathematics of Oil Recovery*, 1994.
- [50] M. Armstrong, A. Galli, H. Beucher, G. Loc'h, D. Renard, B. Doligez, R. Eschard et F. Geffroy, *PluriGaussian Simulations in Geosciences*. Springer Berlin Heidelberg, 2011. [En ligne]. Disponible : <http://www.springer.com/earth+sciences+and+geography/book/978-3-642-19606-5>
- [51] P. A. Dowd, E. Pardo-Igúzquiza et C. Xu, "Plurigau : a computer program for simulating spatial facies using the truncated pluriGaussian method," *Computers & Geosciences*, vol. 29, n°. 2, p. 123–141, mar 2003. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0098300402000705>
- [52] C. Xu, P. A. Dowd, K. V. Mardia et R. J. Fowell, "A flexible true pluriGaussian code for spatial facies simulations," *Computers & Geosciences*, vol. 32, n°. 10, p. 1629–1645, dec 2006. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0098300406000550>
- [53] N. Madani et X. Emery, "Simulation of geo-domains accounting for chronology and contact relationships : application to the Río Blanco copper deposit," *Stochastic Environmental Research and Risk Assessment*, p. 1–19, dec 2014. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s00477-014-0997-x>
- [54] T. Skvortsova, H. Beucher, M. Armstrong, J. Forkes, A. Thwaites et R. Turner, "Simulating the geometry of a granite-hosted uranium orebody," dans *Geostatistics Rio 2000*, ser. Quantitative Geology and Geostatistics, M. Armstrong, C. Bettini, N. Champigny, A. Galli et A. Remacre, édit. Springer Netherlands, jan 2002, n°. 12, p. 85–99. [En ligne]. Disponible : [http://link.springer.com/chapter/10.1007/978-94-017-1701-4\\_7](http://link.springer.com/chapter/10.1007/978-94-017-1701-4_7)
- [55] T. Y. Yunsel et A. Ersoy, "Geological modeling of gold deposit based on grade domaining using pluriGaussian simulation technique," *Natural Resources Research*, vol. 20, n°. 4, p. 231–249, dec 2011. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11053-011-9150-4>
- [56] —, "Geological modeling of rock type domains in the Balya (Turkey) lead-zinc deposit using pluriGaussian simulation," *Central European Journal of Geosciences*, vol. 5, n°. 1, p. 77–89, mar 2013. [En ligne]. Disponible : <http://link.springer.com/article/10.2478/s13533-012-0113-z>
- [57] H. Talebi, O. Asghari et X. Emery, "Simulation of the lately injected dykes in an Iranian porphyry copper deposit using the pluriGaussian model," *Arabian Journal*

- of *Geosciences*, vol. 7, n°. 7, p. 2771–2780, jul 2014. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s12517-013-0911-8>
- [58] G. Mariethoz, P. Renard, F. Cornaton et O. Jaquet, “Truncated pluriGaussian simulations to characterize aquifer heterogeneity,” *Ground Water*, vol. 47, n°. 1, p. 13–24, jan 2009. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1111/j.1745-6584.2008.00489.x/abstract>
- [59] C. Cherubini, C. Giasi, F. Misci et N. Pastore, “Application of truncated pluriGaussian method for the reactive transport modeling of a contaminated aquifer,” dans *4th IASME/WSEAS International Conference on Water Resources, Hydraulics and Hydrology*, Cambridge, England, feb 2009. [En ligne]. Disponible : [http://www.academia.edu/228621/Application\\_of\\_truncated\\_plurigaussian\\_method\\_for\\_the\\_reactive\\_transport\\_modeling\\_of\\_a\\_contaminated\\_aquifer](http://www.academia.edu/228621/Application_of_truncated_plurigaussian_method_for_the_reactive_transport_modeling_of_a_contaminated_aquifer)
- [60] R. Perulero Serrano, L. Guadagnini, M. Riva, M. Giudici et A. Guadagnini, “Impact of two geostatistical hydro-facies simulation strategies on head statistics under non-uniform groundwater flow,” *Journal of Hydrology*, vol. 508, p. 343–355, jan 2014. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0022169413008251>
- [61] N. Jeannée, E. Bardou, C. Faucheux et P. Ornstein, “Geostatistical assessment of ice content distribution within the glacier Bonnard,” *Mathematical Geosciences*, vol. 45, n°. 5, p. 591–599, jul 2013. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11004-013-9465-2>
- [62] D. Allard, D. D’Or, P. Biver et R. Froidevaux, “Non-parametric diagrams for pluri-Gaussian simulations of lithologies,” dans *Ninth International Geostatistics Congress*, vol. 1115, Oslo, Norway, 2012.
- [63] C. Ravenne, A. Galli, B. Doligez, H. Beucher et R. Eschard, “Quantification of facies relationships via proportion curves,” dans *Geostatistics Rio 2000*, ser. Quantitative Geology and Geostatistics, M. Armstrong, C. Bettini, N. Champigny, A. Galli et A. Remacre, édit. Springer Netherlands, jan 2002, n°. 12, p. 19–39. [En ligne]. Disponible : [http://link.springer.com/chapter/10.1007/978-94-017-1701-4\\_3](http://link.springer.com/chapter/10.1007/978-94-017-1701-4_3)
- [64] C. Ravenne, “Stratigraphy and oil : a review - part 2 : Characterization of reservoirs and sequence stratigraphy : Quantification and modeling,” *Oil & Gas Science and Technology*, vol. 57, n°. 4, p. 311–340, jul 2002. [En ligne]. Disponible : <http://ogst.ifpenergiesnouvelles.fr/10.2516/ogst:2002021>
- [65] X. Emery et D. A. Silva, “Conditional co-simulation of continuous and categorical variables for geostatistical applications,” *Computers & Geosciences*, vol. 35, n°. 6, p.



- 1234–1246, jun 2009. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0098300408002859>
- [66] M. Barbier, “Hétérogénéités multi-échelles sédimento-diagénétiques et structurales de la Formation Carbonatée Madison (Mississippien, Wyoming, USA) : implications réservoirs,” PhD thesis, Aix-Marseille Université, 2012.
- [67] B. Doligez, H. Beucher, M. Pontiggia, A. Ortenzi et A. Mariani, “Comparison of methodologies and geostatistical approaches for diagenesis quantification,” dans *AAPG Convention*, jun 2009, p. Article #40492, <http://www.searchanddiscovery.com/author.shtml#d>. [En ligne]. Disponible : <https://hal-mines-paristech.archives-ouvertes.fr/hal-00566939>
- [68] M. Maleki et X. Emery, “Joint simulation of grade and rock type in a stratabound copper deposit,” *Mathematical Geosciences*, p. 1–25, sep 2014. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11004-014-9556-8>
- [69] X. Emery et I. Gálvez, “A PluriGaussian Model for Simulating Regionalized Compositions,” dans *Geostatistics Oslo 2012*, ser. Quantitative Geology and Geostatistics, P. Abrahamsen, R. Hauge et O. Kolbjørnsen, édit. Springer Netherlands, jan 2012, n°. 17, p. 39–50. [En ligne]. Disponible : [http://link.springer.com/chapter/10.1007/978-94-007-4153-9\\_4](http://link.springer.com/chapter/10.1007/978-94-007-4153-9_4)
- [70] N. Madaniesfahani, “PluriGaussian simulation of non-stationary categorical variables and its application to ore body modeling,” Thèse de doctorat, Universidad de Chile, Santiago, Chile, 2016. [En ligne]. Disponible : <http://repositorio.uchile.cl/handle/2250/139158>
- [71] N. Madani et X. Emery, “PluriGaussian modeling of geological domains based on the truncation of non-stationary Gaussian random fields,” *Stochastic Environmental Research and Risk Assessment*, vol. 31, n°. 4, p. 893–913, may 2017. [En ligne]. Disponible : <https://link.springer.com/article/10.1007/s00477-016-1365-9>
- [72] T. L. Blévec, O. Dubrule, C. M. John et G. J. Hampson, “Modelling asymmetrical facies successions using pluri-Gaussian simulations,” dans *Geostatistics Valencia 2016*, ser. Quantitative Geology and Geostatistics. Springer, Cham, 2017, p. 59–75. [En ligne]. Disponible : [https://link.springer.com/chapter/10.1007/978-3-319-46819-8\\_4](https://link.springer.com/chapter/10.1007/978-3-319-46819-8_4)
- [73] T. Le Blévec, O. Dubrule, C. M. John et G. J. Hampson, “Geostatistical modelling of cyclic and rhythmic facies architectures,” *Mathematical Geosciences*, vol. 50, n°. 6, p. 609–637, aug 2018. [En ligne]. Disponible : <https://doi.org/10.1007/s11004-018-9737-y>
- [74] A. Boucher, “Strategies for modeling with multiple-point simulation algorithms,” dans *Gussow Geoscience Conference*, Banf, Canada, 2011.

- [75] F. B. Guardiano et R. M. Srivastava, “Multivariate geostatistics : Beyond bivariate moments,” dans *Geostatistics Tróia '92*, ser. Quantitative Geology and Geostatistics, A. Soares, édit. Springer Netherlands, jan 1993, n°. 5, p. 133–144. [En ligne]. Disponible : [http://link.springer.com/chapter/10.1007/978-94-011-1739-5\\_12](http://link.springer.com/chapter/10.1007/978-94-011-1739-5_12)
- [76] S. Strebelle, “Sequential simulation drawing structures from training images,” PhD thesis, Stanford University, 2000.
- [77] S. K. Jha, G. Mariethoz et B. F. J. Kelly, “Bathymetry fusion using multiple-point geostatistics : Novelty and challenges in representing non-stationary bedforms,” *Environmental Modelling & Software*, vol. 50, p. 66–76, dec 2013. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S1364815213001874>
- [78] J. Wu, A. Boucher et T. Zhang, “A SGeMS code for pattern simulation of continuous and categorical variables : FILTERSIM,” *Computers & Geosciences*, vol. 34, n°. 12, p. 1863–1876, dec 2008. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0098300408000642>
- [79] S. Suzuki et J. Caers, “A distance-based prior model parameterization for constraining solutions of spatial inverse problems,” *Mathematical Geosciences*, vol. 40, n°. 4, p. 445–469, may 2008. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11004-008-9154-8>
- [80] A. Boucher, R. Gupta, J. Caers et A. Satija, “Tetris : a training image generator for SGeMS,” paper presented at Proceedings of 23nd SCRF annual affiliates meeting, Stanford University, Stanford, USA, may 2010.
- [81] A. Maharaja, “TiGenerator : Object-based training image generator,” *Computers & Geosciences*, vol. 34, n°. 12, p. 1753–1761, dec 2008. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0098300408001362>
- [82] G. Mariethoz et S. Lefebvre, “Bridges between multiple-point geostatistics and texture synthesis : Review and guidelines for future research,” *Computers & Geosciences*, vol. 66, p. 66–80, may 2014. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0098300414000041>
- [83] M. Honarkhah et J. Caers, “Direct pattern-based simulation of non-stationary geostatistical models,” *Mathematical Geosciences*, vol. 44, n°. 6, p. 651–672, aug 2012. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11004-012-9413-6>
- [84] A. G. Journel, “Combining knowledge from diverse sources : An alternative to traditional data independence hypotheses,” *Mathematical Geology*, vol. 34, n°. 5, p. 573–596, jul 2002. [En ligne]. Disponible : <http://link.springer.com/article/10.1023/A%3A1016047012594>

- [85] S. Strebelle, “Conditional simulation of complex geological structures using multiple-point statistics,” *Mathematical Geology*, vol. 34, n<sup>o</sup>. 1, p. 1–21, jan 2002. [En ligne]. Disponible : <http://link.springer.com/article/10.1023/A%3A1014009426274>
- [86] L. M. d. Vries, J. Carrera, O. Falivene, O. Gratacós et L. J. Slooten, “Application of multiple point geostatistics to non-stationary images,” *Mathematical Geosciences*, vol. 41, n<sup>o</sup>. 1, p. 29–42, jan 2009. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11004-008-9188-y>
- [87] E. Meerschman, M. Van Meirvenne, G. Mariethoz, M. M. Islam, P. De Smedt, E. Van De Vijver et T. Saey, “Using bivariate multiple-point statistics and proximal soil sensor data to map fossil ice-wedge polygons,” *Geoderma*, vol. 213, p. 571–577, jan 2014. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S001670611300030X>
- [88] P. W. Mirowski, D. M. Tetzlaff, R. C. Davies, D. S. McCormick, N. Williams et C. Signer, “Stationarity scores on training images for multipoint geostatistics,” *Mathematical Geosciences*, vol. 41, n<sup>o</sup>. 4, p. 447–474, may 2009. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11004-008-9194-0>
- [89] J. Straubhaar, P. Renard, G. Mariethoz, R. Froidevaux et O. Besson, “An improved parallel multiple-point algorithm using a list approach,” *Mathematical Geosciences*, vol. 43, n<sup>o</sup>. 3, p. 305–328, apr 2011. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11004-011-9328-7>
- [90] T. L. Chugunova et L. Y. Hu, “Multiple-point simulations constrained by continuous auxiliary data,” *Mathematical Geosciences*, vol. 40, n<sup>o</sup>. 2, p. 133–146, feb 2008. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11004-007-9142-4>
- [91] T. Zhang, P. Switzer et A. Journel, “Filter-based classification of training image patterns for spatial simulation,” *Mathematical Geology*, vol. 38, n<sup>o</sup>. 1, p. 63–80, jan 2006. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11004-005-9004-x>
- [92] S. Strebelle et T. Zhang, “Non-stationary multiple-point geostatistical models,” dans *Geostatistics Banff 2004*, ser. Quantitative Geology and Geostatistics, O. Leuangthong et C. V. Deutsch, édit. Springer Netherlands, 2005, p. 235–244. [En ligne]. Disponible : [https://doi.org/10.1007/978-1-4020-3610-1\\_24](https://doi.org/10.1007/978-1-4020-3610-1_24)
- [93] E. Meerschman, G. Pirot, G. Mariethoz, J. Straubhaar, M. Van Meirvenne et P. Renard, “A practical guide to performing multiple-point statistical simulations with the direct sampling algorithm,” *Computers & Geosciences*, vol. 52, p. 307–324, mar 2013. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0098300412003299>

- [94] C. Höcker, “More realistic facies patterns using MPS modeling with auxiliary variables,” dans *Gussow Geoscience Conference*, Banf, Canada, 2011.
- [95] G. Mariethoz et B. F. J. Kelly, “Modeling complex geological structures with elementary training images and transform-invariant distances,” *Water Resources Research*, vol. 47, n°. 7, p. W07527, jul 2011. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1029/2011WR010412/abstract>
- [96] G. B. Arpat et J. Caers, “Conditional simulation with patterns,” *Mathematical Geology*, vol. 39, n°. 2, p. 177–203, feb 2007.
- [97] H. Rezaee et D. Marcotte, “Integration of multiple soft data sets in MPS thru multinomial logistic regression : a case study of gas hydrates,” *Stochastic Environmental Research and Risk Assessment*, p. 1–19, jun 2016. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s00477-016-1277-8>
- [98] A. El Ouassini, A. Saucier, D. Marcotte et B. D. Favis, “A patchwork approach to stochastic simulation : A route towards the analysis of morphology in multiphase systems,” *Chaos, Solitons & Fractals*, vol. 36, n°. 2, p. 418–436, apr 2008. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0960077906006680>
- [99] D. Allard, D. D’Or et R. Froidevaux, “An efficient maximum entropy approach for categorical variable prediction,” *European Journal of Soil Science*, vol. 62, p. 381–393, 2011.
- [100] G. Christakos, *Random field models in earth sciences*. San Diego : Academic Press, 1992.
- [101] M. L. Serre et G. Christakos, “Modern geostatistics : computational BME analysis in the light of uncertain physical knowledge – the equus beds study,” *Stochastic Environmental Research and Risk Assessment*, vol. 13, n°. 1, p. 1–26, apr 1999. [En ligne]. Disponible : <https://link.springer.com/article/10.1007/s004770050029>
- [102] M. L. Serre, G. Christakos, H. Li et C. T. Miller, “A BME solution of the inverse problem for saturated groundwater flow,” *Stochastic Environmental Research and Risk Assessment*, vol. 17, n°. 6, p. 354–369, dec 2003. [En ligne]. Disponible : <https://link.springer.com/article/10.1007/s00477-003-0156-2>
- [103] D. D’Or, P. Bogaert et G. Christakos, “Application of the BME approach to soil texture mapping,” *Stochastic Environmental Research and Risk Assessment*, vol. 15, n°. 1, p. 87–100, mar 2001. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s004770000057>
- [104] T. G. Orton et R. M. Lark, “Accounting for the uncertainty in the local mean in spatial prediction by bayesian maximum entropy,” *Stochastic Environmental Research*

- and Risk Assessment*, vol. 21, n°. 6, p. 773–784, nov 2007. [En ligne]. Disponible : <https://link.springer.com/article/10.1007/s00477-006-0089-7>
- [105] P. Bogaert et D. D’Or, “Estimating soil properties from thematic soil maps : the Bayesian Maximum Entropy approach,” *Soil Science Society of America Journal*, vol. 66, n°. 5, p. 1492–1500, 2002.
- [106] D. D’Or et P. Bogaert, “Spatial prediction of categorical variables with the Bayesian Maximum Entropy approach : the Ooypolder case study,” *European Journal of Soil Science*, vol. 55, n°. 4, p. 763–775, 2004.
- [107] P. Bogaert, “Spatial prediction of categorical variables : the Bayesian maximum entropy approach,” *Stochastic Environmental Research and Risk Assessment*, vol. 16, n°. 6, p. 425–448, 2002. [En ligne]. Disponible : <https://link.springer.com/article/10.1007/s00477-002-0114-4>
- [108] D. D’Or, “Spatial prediction of soil properties, the Bayesian maximum entropy approach,” PhD thesis, Université catholique de Louvain, 2003.
- [109] W. Li, C. Zhang, J. E. Burt, A.-X. Zhu et J. Feyen, “Two-dimensional Markov chain simulation of soil type spatial distribution,” *Soil Science Society of America Journal*, vol. 68, n°. 5, p. 1479–1490, sep 2004. [En ligne]. Disponible : <https://dl.sciencesocieties.org/publications/sssaj/abstracts/68/5/1479>
- [110] W. Li, “Markov chain random fields for estimation of categorical variables,” *Mathematical Geology*, vol. 39, n°. 3, p. 321–335, jun 2007. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11004-007-9081-0>
- [111] W. Li et C. Zhang, “A random-path Markov chain algorithm for simulating categorical soil variables from random point samples,” *Soil Science Society of America Journal*, vol. 71, n°. 3, p. 656–668, may 2007. [En ligne]. Disponible : <https://dl.sciencesocieties.org/publications/sssaj/abstracts/71/3/656>
- [112] S. F. Carle et G. E. Fogg, “T-PROGS : Transition probability-based indicator geostatistics,” *Mathematical Geology*, vol. 28, n°. 4, p. 453–476, 1996. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/BF02083656>
- [113] M. P. Anderson et W. W. Woessner, “Chapter 9 - documenting and reporting you modeling study,” dans *Applied Groundwater Modeling*. Academic Press, 2002, p. 275–285. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/B9780080886947500146>
- [114] N. Benoit et D. Paradis, “Three dimensional groundwater flow model of the Nanoose-Deep Bay area, Nanaimo Lowland, British Columbia,” Natural Resources Canada, Geological Survey of Canada, Rapport technique, 2015.

- [115] E. Hudon-Gagnon, R. Chesnaux, P. A. Cousineau et A. Rouleau, “A hydrostratigraphic simplification approach to build 3D groundwater flow numerical models : example of a quaternary deltaic deposit aquifer,” *Environmental Earth Sciences*, vol. 74, n°. 6, p. 4671–4683, apr 2015. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s12665-015-4439-y>
- [116] L. A. Buryakovsky, G. V. Chilingar et F. Aminzadeh, “CHAPTER 10 - mathematical models in oil and gas exploration and production (static geologic systems),” dans *Petroleum Geology of the South Caspian Basin*, L. A. Buryakovsky, G. V. Chilingar et F. Aminzadeh, édit. Gulf Professional Publishing, jan 2001, p. 248–346. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/B978088415342950014X>
- [117] E. R. Siirila-Woodburn et R. M. Maxwell, “A heterogeneity model comparison of highly resolved statistically anisotropic aquifers,” *Advances in Water Resources*, vol. 75, p. 53–66, jan 2015. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0309170814002139>
- [118] A. D. Woodbury et E. A. Sudicky, “The geostatistical characteristics of the Borden aquifer,” *Water Resources Research*, vol. 27, n°. 4, p. 533–546, apr 1991. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1029/90WR02545/abstract>
- [119] D. Todd et L. Mays, *Hydrology Groundwater Hydrology*, ser. Wiley Series in Hydrology (Civil Engineering). John Wiley & Sons, Inc., jul 2004. [En ligne]. Disponible : <https://www.wiley.com/en-ca/Groundwater+Hydrology%2C+3rd+Edition-p-9780471059370>
- [120] J. Rosas, O. Lopez, T. M. Missimer, K. M. Coulibaly, A. H. Dehwah, K. Sesler, L. R. Lujan et D. Mantilla, “Determination of hydraulic conductivity from grain-size distribution for different depositional environments,” *Groundwater*, vol. 52, n°. 3, p. 399–413, may 2014. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1111/gwat.12078/abstract>
- [121] M. Vuković et A. Soro, *Determination of hydraulic conductivity of porous media from grain-size composition*. Water Resources Pubns, 1992.
- [122] R. P. Chapuis, V. Dallaire, D. Marcotte, M. Chouteau, N. Acevedo et F. Gagnon, “Evaluating the hydraulic conductivity at three different scales within an unconfined sand aquifer at Lachenaie, Quebec,” *Canadian Geotechnical Journal*, vol. 42, n°. 4, p. 1212–1220, aug 2005. [En ligne]. Disponible : <http://www.nrcresearchpress.com/doi/abs/10.1139/t05-045>

- [123] W. Beyer, “Zur bestimmung der wasserdurchlässigkeit von kiesen und sanden aus der kornverteilungskurve,” *Wasserwirtschaft-Wassertechnik (WWT)*, vol. 14, n°. 6, p. 165–168, 1964.
- [124] A. Hazen, “Some physical properties of sands and gravels, with special reference to their use in filtration.” Massachusetts State Board of Health, Rapport technique 24th Annual Report, 1892.
- [125] D. Navfac, “Design manual-soil mechanics, foundations, and earth structures,” *US Government Printing Office, Washington, DC*, 1974.
- [126] M. Wyllie et G. Gardner, “The generalized Kozeny-Carman equation,” *World oil*, vol. 146, n°. 4, p. 121–128, 1958.
- [127] W. Craeger, J. Justin et J. Hinds, *Engineering for dams*. Wiley & Sons, 1947, vol. 2.
- [128] H.-J. G. Diersch, *FEFLOW - Finite Element Modeling of Flow, Mass and Heat Transport in Porous and Fractured Media*. Springer, 2014. [En ligne]. Disponible : <http://www.springer.com/earth+sciences+and+geography/hydrogeology/book/978-3-642-38738-8>
- [129] M. Ravalec-Dupin et L. Hu, “Gradual deformation of boolean simulations,” dans *Geostatistics Banff 2004*, ser. Quantitative Geology and Geostatistics, O. Leuangthong et C. Deutsch, édit. Springer Netherlands, jan 2005, vol. 14, p. 939–948.
- [130] T. T. Tran, X.-H. Wen et R. A. Behrens, “Efficient conditioning of 3D fine-scale reservoir model to multiphase production data using streamline-based coarse-scale inversion and geostatistical downscaling,” *SPE Journal*, vol. 6, n°. 4, p. 364–374, dec 2001. [En ligne]. Disponible : <http://www.onepetro.org/doi/10.2118/74708-PA>
- [131] L. Li, H. Zhou, J. J. Gómez-Hernández et H.-J. Hendricks Franssen, “Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble Kalman filter,” *Journal of Hydrology*, vol. 428–429, p. 152–169, 2012.
- [132] I. Myrseth, J. Sætrom et H. Omre, “Resampling the ensemble Kalman filter,” *Computers & Geosciences*, vol. 55, p. 44–53, jun 2013. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S009830041200204X>
- [133] L. Li, H. Zhou et J. J. Gómez-Hernández, “A comparative study of three-dimensional hydraulic conductivity upscaling at the macro-dispersion experiment (MADE) site, Columbus Air Force Base, Mississippi (USA),” *Journal of Hydrology*, vol. 404, n°. 3, p. 278–293, jul 2011. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0022169411002976>

- [134] P. Renard et G. de Marsily, “Calculating equivalent permeability : a review,” *Advances in Water Resources*, vol. 20, n°. 5, p. 253–278, oct 1997. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0309170896000504>
- [135] X. Sanchez-Vila, A. Guadagnini et J. Carrera, “Representative hydraulic conductivities in saturated groundwater flow,” *Reviews of Geophysics*, vol. 44, n°. 3, p. RG3002, sep 2006. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1029/2005RG000169/abstract>
- [136] X.-H. Wen et J. J. Gómez-Hernández, “Upscaling hydraulic conductivities in heterogeneous media : An overview,” *Journal of Hydrology*, vol. 183, n°. 1, p. ix–xxxii, aug 1996. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0022169496800308>
- [137] J. J. Gómez-Hernández et X.-H. Wen, “To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology,” *Advances in Water Resources*, vol. 21, n°. 1, p. 47–61, feb 1998. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0309170896000310>
- [138] M. L. Ravalec-Dupin, *Inverse stochastic modeling of flow in porous media*. Editions Technip, sep 2005.
- [139] O. Wiener, “Die theorie des mischkorpers fur das feld der station-aren stromung,” *Ges. Wiss.*, vol. 32, n°. 6, p. 509–604, 1912.
- [140] J. J. Gomez-Hernandez et A. G. Journel, “Stochastic characterization of grid-block permeabilities : from point values to block tensors,” dans *ECMOR II - 2nd European Conference on the Mathematics of Oil Recovery*. EAGE Publications BV, sep 1990.
- [141] P. Renard et D. Allard, “Connectivity metrics for subsurface flow and transport,” *Advances in Water Resources*, vol. 51, p. 168–196, jan 2013. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0309170811002223>
- [142] N. Benoit, J. W. Molson, M. Y. Brun Koné et M. Nastev, “Modèle hydrogéologique 3d du bassin versant de la rivière chaudière,” Natural Resources Canada, Geological Survey of Canada, Geological Survey of Canada, Rapport technique Open File 7734, 2015. [En ligne]. Disponible : <http://geoscan.nrcan.gc.ca/starweb/geoscan/servlet.starweb>
- [143] J.-P. Chilès et P. Delfiner, *Geostatistics : Modeling Spatial Uncertainty*. John Wiley & Sons, Inc., 1999. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1002/9780470316993.ch/summary>
- [144] T. Ouellon, R. Lefebvre, D. Marcotte, A. Boutin, V. Blais et M. Parent, “Hydraulic conductivity heterogeneity of a local deltaic aquifer system from the



- kriged 3D distribution of hydrofacies from borehole logs, Valcartier, Canada,” *Journal of Hydrology*, vol. 351, n°. 1, p. 71–86, mar 2008. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0022169407007111>
- [145] M. M. Meerschaert, M. Dogan, R. L. Van Dam, D. W. Hyndman et D. A. Benson, “Hydraulic conductivity fields : Gaussian or not ?” *Water resources research*, vol. 49, n°. 8, aug 2013. [En ligne]. Disponible : <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3885184/>
- [146] E. P. Poeter et M. C. Hill, “Inverse models : A necessary next step in ground-water modeling,” *Groundwater*, vol. 35, p. 250–260, 1997.
- [147] J. Carrera, A. Alcolea, A. Medina, J. Hidalgo et L. J. Slooten, “Inverse problem in hydrogeology,” *Hydrogeology Journal*, vol. 13, n°. 1, p. 206–222, mar 2005. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s10040-004-0404-7>
- [148] G. de Marsily, J. P. Delhomme, A. Coudrain-Ribstein et A. M. Lavenue, “Four decades of inverse problems in hydrogeology,” dans *Theory, modeling, and field investigation in hydrogeology : a special volume in honor of Shlomo P. Neumans 60th birthday*. Geological Society of America, 2000.
- [149] D. Oliver et Y. Chen, “Recent progress on reservoir history matching : a review,” *Computational Geosciences*, vol. 15, p. 185–221, jan 2011.
- [150] W. W.-G. Yeh, “Review of parameter identification procedures in groundwater hydrology : The inverse problem,” *Water Resources Research*, vol. 22, n°. 2, p. 95–108, feb 1986. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1029/WR022i002p00095/abstract>
- [151] P. Pasquier et D. Marcotte, “Steady- and transient-state inversion in hydrogeology by successive flux estimation,” *Advances in Water Resources*, vol. 29, n°. 12, p. 1934–1952, déc. 2006. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0309170806000297>
- [152] G. Ponzini et A. Lozej, “Identification of aquifer transmissivities : The comparison model method,” *Water Resources Research*, vol. 18, n°. 3, p. 597–622, jun 1982. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1029/WR018i003p00597/abstract>
- [153] M. Rivest, D. Marcotte et P. Pasquier, “Including conceptual model information when kriging hydraulic heads,” dans *geoENV VI – Geostatistics for Environmental Applications*, ser. Quantitative Geology and Geostatistics, A. Soares, M. J. Pereira et R. Dimitrakopoulos, édit. Springer Netherlands, jan 2008, n°. 15, p. 141–151. [En ligne]. Disponible : [http://link.springer.com/chapter/10.1007/978-1-4020-6448-7\\_12](http://link.springer.com/chapter/10.1007/978-1-4020-6448-7_12)

- [154] G. K. Brouwer, P. A. Fokker, F. Wilschut et W. Zijl, “A direct inverse model to determine permeability fields from pressure and flow rate measurements,” *Mathematical Geosciences*, vol. 40, n°. 8, p. 907–920, jul 2008. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11004-008-9183-3>
- [155] A. Alcolea, J. Carrera et A. Medina, “Pilot points method incorporating prior information for solving the groundwater flow inverse problem,” *Advances in Water Resources*, vol. 29, n°. 11, p. 1678–1689, nov 2006. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0309170805002976>
- [156] J. Caers, “Geostatistical history matching under training-image based geological model constraints,” dans *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2002. [En ligne]. Disponible : <http://www.onepetro.org/doi/10.2118/77429-MS>
- [157] J. Carrera et S. P. Neuman, “Estimation of aquifer parameters under transient and steady state conditions : 1. maximum likelihood method incorporating prior information,” *Water Resources Research*, vol. 22, n°. 2, p. 199–210, feb 1986. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1029/WR022i002p00199/abstract>
- [158] G. Evensen, “Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics,” *Journal of Geophysical Research : Oceans*, vol. 99, p. 10 143–10 162, may 1994. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1029/94JC00572/abstract>
- [159] —, *Data Assimilation*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2009. [En ligne]. Disponible : <http://link.springer.com/10.1007/978-3-642-03711-5>
- [160] L. Hu, “Gradual deformation and iterative calibration of Gaussian-related stochastic models,” *Mathematical Geology*, vol. 32, n°. 1, p. 87–108, jan 2000.
- [161] J. Jaime Gómez-Hernández, A. Sahuquillo et J. Capilla, “Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data—i. theory,” *Journal of Hydrology*, vol. 203, n°. 1, p. 162–174, dec 1997. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S002216949700098X>
- [162] P. K. Kitanidis, “Quasi-linear geostatistical theory for inversing,” *Water Resources Research*, vol. 31, n°. 10, p. 2411–2419, oct 1995. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1029/95WR01945/abstract>
- [163] G. de Marsily, G. Lavedau, M. Boucher et G. Fasanino, “Interpretation of interference test in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model,” dans *Geostatistics for Natural Resources Characterization, Pro-*

- ceedings of the NATO Advanced Study Institute*, G. Verly, M. David, A. Journel et A. Marechal, édit. Dordrecht, Holland, 1984.
- [164] D. S. Oliver, L. B. Cunha et A. C. Reynolds, “Markov chain Monte Carlo methods for conditioning a permeability field to pressure data,” *Mathematical Geology*, vol. 29, n°. 1, p. 61–91, mar 1997. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/BF02769620>
- [165] B. S. RamaRao, A. M. LaVenue, G. de Marsily et M. G. Marietta, “Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields : 1. theory and computational experiments,” *Water Resources Research*, vol. 31, n°. 3, p. 475–493, mar 1995. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1029/94WR02258/abstract>
- [166] B. Todd Hoffman et J. Caers, “Regional probability perturbations for history matching,” *Journal of Petroleum Science and Engineering*, vol. 46, n°. 1, p. 53–71, feb 2005. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0920410504001780>
- [167] D. Lauzon et D. Marcotte, “Calibration of random fields by FFTMA-SA,” *Computers & Geosciences*, vol. 127, p. 99–110, jun 2019.
- [168] —, “Calibration of random fields by a sequential spectral turning bands method,” *Computers & Geosciences*, vol. 135, p. 104390, févr. 2020. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0098300419306752>
- [169] J. Hadamard, *Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques ; leçons professées à l’Université Yale*. Hermann et cie, 1932.
- [170] P. K. Kitanidis et E. G. Vomvoris, “A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations,” *Water Resources Research*, vol. 19, n°. 3, p. 677–690, jun 1983. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1029/WR019i003p00677/abstract>
- [171] J. Doherty, “Ground water model calibration using pilot points and regularization,” *Ground water*, vol. 41, n°. 2, p. 170–177, 2003, \_eprint : <https://ngwa.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-6584.2003.tb02580.x>. [En ligne]. Disponible : <https://ngwa.onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-6584.2003.tb02580.x>
- [172] —, “PEST - model-independent parameter estimation,” Watermark Numerical Computing, Rapport technique, 2002. [En ligne]. Disponible : <http://www.pesthomepage.org/Downloads.php>

- [173] M. Ferraresi, E. Todini et R. Vignoli, “A solution to the inverse problem in groundwater hydrology based on Kalman filtering,” *Journal of Hydrology*, vol. 175, n°. 1, p. 567–581, feb 1996. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0022169496800254>
- [174] M. Le Ravalec Dupin et B. Noetinger, “Optimization with the gradual deformation method,” *Mathematical Geology*, vol. 34, n°. 2, p. 125–142, feb 2002.
- [175] S. Hörning et A. Bárdossy, “Phase annealing for the conditional simulation of spatial random fields,” *Computers & Geosciences*, vol. 112, p. 101–111, mar 2018. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S009830041730403X>
- [176] L. Y. Hu et M. Le Ravalec-Dupin, “Elements for an integrated geostatistical modeling of heterogeneous reservoirs,” *Oil & Gas Science and Technology*, vol. 59, n°. 2, p. 141–155, mar 2004. [En ligne]. Disponible : <http://ogst.ifpenergiesnouvelles.fr/10.2516/ogst:2004011>
- [177] S. Suzuki, “Effect of structure geometry on history matching faulted reservoirs : Stanford center for reservoir forecasting,” Stanford University, Rapport technique 14, 2004.
- [178] Y. Rubin, *Applied Stochastic Hydrogeology*. New York, NY : Oxford University Press, apr 2003.
- [179] J. Wu et X. Zeng, “Review of the uncertainty analysis of groundwater numerical simulation,” *Chinese Science Bulletin*, vol. 58, n°. 25, p. 3044–3052, sept. 2013. [En ligne]. Disponible : <https://doi.org/10.1007/s11434-013-5950-8>
- [180] J. Bredehoeft, “The conceptualization model problem—surprise,” *Hydrogeology Journal*, vol. 13, n°. 1, p. 37–46, mars 2005. [En ligne]. Disponible : <https://doi.org/10.1007/s10040-004-0430-5>
- [181] S. P. Neuman, “Maximum likelihood bayesian averaging of uncertain model predictions,” *Stochastic Environmental Research and Risk Assessment*, vol. 17, n°. 5, p. 291–305, nov 2003. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s00477-003-0151-7>
- [182] L. Trolldborg, J. C. Refsgaard, K. H. Jensen et P. Engesgaard, “The importance of alternative conceptual models for simulation of concentrations in a multi-aquifer system,” *Hydrogeology Journal*, vol. 15, n°. 5, p. 843–860, août 2007. [En ligne]. Disponible : <https://doi.org/10.1007/s10040-007-0192-y>
- [183] A. L. Højberg et J. C. Refsgaard, “Model uncertainty–parameter uncertainty versus conceptual models,” *Water Science and Technology : A Journal of the International Association on Water Pollution Research*, vol. 52, n°. 6, p. 177–186, 2005.

- [184] D. Seifert, T. O. Sonnenborg, P. Scharling et K. Hinsby, “Use of alternative conceptual models to assess the impact of a buried valley on groundwater vulnerability,” *Hydrogeology Journal*, vol. 16, n<sup>o</sup>. 4, p. 659–674, juin 2008. [En ligne]. Disponible : <https://doi.org/10.1007/s10040-007-0252-3>
- [185] S. Christensen, “On the strategy of estimating regional-scale transmissivity fields,” *Ground Water*, vol. 35, n<sup>o</sup>. 1, p. 131–139, jan 1997.
- [186] M. J. Asher, B. F. W. Croke, A. J. Jakeman et L. J. M. Peeters, “A review of surrogate models and their application to groundwater modeling,” *Water Resources Research*, vol. 51, n<sup>o</sup>. 8, p. 5957–5973, aug 2015. [En ligne]. Disponible : <https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2015WR016967>
- [187] S. Razavi, B. A. Tolson et D. H. Burn, “Review of surrogate modeling in water resources,” *Water Resources Research*, vol. 48, n<sup>o</sup>. 7, jul 2012. [En ligne]. Disponible : <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011WR011527>
- [188] H. A. Jensen, C. Esse, V. Araya et C. Papadimitriou, “Implementation of an adaptive meta-model for bayesian finite element model updating in time domain,” *Reliability Engineering & System Safety*, vol. 160, p. 174–190, apr 2017. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0951832016303131>
- [189] Bajc, A.F. et D. Rainsford, “Three-dimensional mapping of quaternary deposits in the southern part of the county of Simcoe, Southern Ontario; in summary of field work and other activities 2010,” Ontario Geological Survey, Open File Report 6260, 2010.
- [190] A. Bajc, F. Brunton, A. Burt, S. Hamilton, A. Marich, R. Mulligan et E. Priebe, “An overview of the Ontario geological survey’s groundwater initiative,” *Geological Survey of Canada, Open File*, vol. 8022, p. 1–1, 2016.
- [191] A. Bajc, R. Mulligan, A.-M. Pugin et D. Rainsford, “33. project unit 10-026. an update on the final phase of subsurface data collection in the southern part of the county of Simcoe, Southern Ontario,” Ontario Geological Survey, Rapport technique, 2014.
- [192] A. Bajc, R. Mulligan et J. Dodge, “Three-dimensional mapping of surficial deposits in the southern part of the County of Simcoe, south-central Ontario,” Ontario Geological Survey, Groundwater Resources Study 18, 2019.
- [193] R. P. M. Mulligan, “Quaternary landform and sediment analysis of the Alliston area (southern Simcoe County), Ontario, Canada,” PhD thesis, McMaster University, Hamilton, Canada, oct 2013. [En ligne]. Disponible : <http://macsphere.mcmaster.ca/handle/11375/13516>
- [194] R. Mulligan et A. Bajc, “The pre-late wisconsin stratigraphy of southern Simcoe County, Ontario : implications for ice sheet buildup, decay, and great lakes drainage

- evolution,” *Canadian Journal of Earth Sciences*, p. 1–21, jan 2017. [En ligne]. Disponible : <http://www.nrcresearchpress.com/doi/abs/10.1139/cjes-2016-0160>
- [195] D. Armstrong, “Paleozoic geology of the northern Lake Simcoe area, Southcentral Ontario ; Ontario,” Ontario Geological Survey, Rapport technique 6011, 2000.
- [196] A. Bajc, D. Rainsford et R. Mulligan, “An update on three-dimensional mapping of quaternary deposits in the southern part of the county of Simcoe, Southern Ontario ; in summary of field work and other activities 2012,” Ontario Geological Survey, Open File Report 6280, 2012.
- [197] A. Bajc, R. Mulligan, D. Rainsford et J. Webb, “Results of 2011–13 overburden drilling programs in the southern part of the County of Simcoe, south-central Ontario ; Ontario Geological Survey, Miscellaneous Release—Data 324.” Ontario Geological Survey, Rapport technique, 2015.
- [198] R. Mulligan, “Surficial geology of the Alliston area, Southern Ontario ; in summary of field work and other activities 2011,” Ontario Geological Survey, Rapport technique 6270, 2011.
- [199] D. R. Sharpe, A. Pugin, S. E. Pullan et G. Gorrell, “Application of seismic stratigraphy and sedimentology to regional hydrogeological investigations : an example from Oak Ridges Moraine, Southern Ontario, Canada,” *Canadian Geotechnical Journal*, vol. 40, n°. 4, p. 711–730, aug 2003. [En ligne]. Disponible : <http://www.nrcresearchpress.com/doi/abs/10.1139/t03-020>
- [200] B. Kjarsgaard, D. Sharpe, R. Knight, D. Stepner et H. Russell, “Newmarket till aquitard : Optimum grain packing with a calcite-rich cement,” dans *Regional-scale groundwater geoscience in southern Ontario : an Ontario Geological Survey, Geological Survey of Canada, and Conservation Ontario geoscientists open house*, 2018.
- [201] P. Barnett, D. R. Sharpe, H. Russel, T. A. Brennand, G. Gorrell, F. Kenny et A. Pugin, “On the origin of the Oak Ridges Moraine,” *Can. J. Earth Sci.*, may 2013. [En ligne]. Disponible : <http://geoscan.nrcan.gc.ca/starweb/geoscan/servlet.starweb>
- [202] H. A. J. Russell, D. R. Sharpe, T. A. Brennand, P. J. Barnett et C. Logan, “Tunnel channels of the greater Toronto and Oak Ridges Moraine areas, Southern Ontario,” Natural Resources Canada, Geological Survey of Canada, Rapport technique Open File 4485, 2003. [En ligne]. Disponible : <http://geoscan.ess.nrcan.gc.ca/cgi-bin/starfinder/0?path=geoscan.fl&id=fastlink&pass=&search=R%3D214777&format=FLFULL>
- [203] A. Pugin, S. E. Pullan et D. R. Sharpe, “Seismic facies and regional architecture of the Oak Ridges Moraine area, Southern Ontario,” *Canadian Journal of*

- Earth Sciences*, vol. 36, n°. 3, p. 409–432, mar 1999. [En ligne]. Disponible : <http://www.nrcresearchpress.com/doi/abs/10.1139/e98-104>
- [204] S. Singer, “A hydrogeological study along the north shore of Lake Ontario in the Bowmanville - Newcastle area,” Ontario Ministry of the Environment, Rapport technique, 1974.
- [205] P. Pasquier, N. Benoit, Y. Delahaye, L. Gagnon et R. Mulligan, “Hydraulic Conductivity Database Simcoe County, Southern Ontario,” Geological Survey of Canada, Rapport technique, 2016.
- [206] YPDT-CAMC, “Comprehensive York Peel Durham Toronto - Conservation Authorities Moraine Coalition (YPDT-CAMC) , 2016. Water Management SQL Database. Oak Ridges Moraine Groundwater Program, 2002 - 2018. presently referred to as Oak Ridges Groundwater Program (ORMGP). data provided in 2016,” 2016. [En ligne]. Disponible : <https://oakridgeswater.ca/>
- [207] A. F. Bajc, H. A. J. Russell et D. R. Sharpe, “A three-dimensional hydrostratigraphic model of the Waterloo Moraine area, Southern Ontario, Canada,” *Canadian Water Resources Journal / Revue canadienne des ressources hydriques*, vol. 39, n°. 2, p. 95–119, apr 2014. [En ligne]. Disponible : <http://dx.doi.org/10.1080/07011784.2014.914794>
- [208] R. Chesnaux, C. Baudement et H. Murray, “Assessing and comparing the hydraulic properties of granular aquifers on three different scales,” dans *Geohydro 2011*, 2011.
- [209] R. Mulligan, “Project unit 14-015. subsurface data collection for three-dimensional sediment mapping in the central part of the county of Simcoe, Southern Ontario,” Ontario Geological Survey, Rapport technique, dec 2016.
- [210] ASTM-D5856-15, “Standard test method for measurement of hydraulic conductivity of porous material using a rigid-wall, compaction-mold permeameter,” ASTM International, Rapport technique, 2015. [En ligne]. Disponible : [www.astm.org](http://www.astm.org)
- [211] ASTM-D5084-10, “Standard test methods for measurement of hydraulic conductivity of saturated porous materials using a flexible wall permeameter,” ASTM International, Rapport technique, 2010. [En ligne]. Disponible : [www.astm.org](http://www.astm.org)
- [212] K. R. Bradbury et E. R. Rothschild, “A computerized technique for estimating the hydraulic conductivity of aquifers from specific capacity data,” *Ground Water*, vol. 23, n°. 2, p. 240–246, mar 1985. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1111/j.1745-6584.1985.tb02798.x/abstract>
- [213] S. G. McLin, “Estimating aquifer transmissivity from specific capacity using MATLAB,” *Ground Water*, vol. 43, n°. 4, p. 611–614, jul 2005. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1111/j.1745-6584.2005.0101.x/abstract>

- [214] V. Batu, *Aquifer Hydraulics : A Comprehensive Guide to Hydrogeologic Data Analysis*, 1<sup>er</sup> éd. Wiley-Interscience, feb 1998.
- [215] J. W. Molson et E. O. Frind, “On the use of mean groundwater age, life expectancy and capture probability for defining aquifer vulnerability and time-of-travel zones for source water protection,” *Journal of Contaminant Hydrology*, vol. 127, n<sup>o</sup>. 1, p. 76–87, jan 2012. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0169772211000611>
- [216] J.-P. Chilès et P. Delfiner, *Geostatistics : Modeling Spatial Uncertainty*, ser. Wiley Series in Probability and Statistics. Hoboken, NJ, USA : John Wiley & Sons, Inc., mars 2012. [En ligne]. Disponible : <http://doi.wiley.com/10.1002/9781118136188>
- [217] H. Rezaee, D. Marcotte, P. Tahmasebi et A. Saucier, “Multiple-point geostatistical simulation using enriched pattern databases,” *Stochastic Environmental Research and Risk Assessment*, vol. 29, n<sup>o</sup>. 3, p. 893–913, mar 2015. [En ligne]. Disponible : <https://link.springer.com/article/10.1007/s00477-014-0964-6>
- [218] A. Boucher, J. F. Costa, L. G. Rasera et E. Motta, “Simulation of geological contacts from interpreted geological model using multiple-point statistics,” *Mathematical Geosciences*, vol. 46, n<sup>o</sup>. 5, p. 561–572, jul 2014. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s11004-013-9510-1>
- [219] T. T. Tran, “Improving variogram reproduction on dense simulation grids,” *Computers & Geosciences*, vol. 20, n<sup>o</sup>. 7, p. 1161–1168, aug 1994.
- [220] S. Kullback et R. A. Leibler, “On information and sufficiency,” *The Annals of Mathematical Statistics*, vol. 22, n<sup>o</sup>. 1, p. 79–86, 1951. [En ligne]. Disponible : <http://www.jstor.org/stable/2236703>
- [221] M. L. Ravalec, B. Noetinger et L. Y. Hu, “The FFT moving average (FFT-MA) generator : An efficient numerical method for generating and conditioning Gaussian simulations,” *Mathematical Geology*, vol. 32, n<sup>o</sup>. 6, p. 701–723, aug 2000. [En ligne]. Disponible : <http://link.springer.com/article/10.1023/A%3A1007542406333>
- [222] J. M. Ortiz, “Characterization of high order correlation for enhanced indicator simulation,” PhD thesis, University of Alberta, 2003.
- [223] M. P. Anderson, W. W. Woessner et R. J. Hunt, *Applied Groundwater Modeling*. Elsevier Science, 2015.
- [224] M. Rivest, D. Marcotte et P. Pasquier, “Hydraulic head field estimation using kriging with an external drift : A way to consider conceptual model information,” *Journal of Hydrology*, vol. 361, n<sup>o</sup>. 3, p. 349–361, nov 2008. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0022169408004083>



- [225] L. Y. Hu, “Gradual deformation and iterative calibration of Gaussian-related stochastic models,” *Mathematical Geology*, vol. 32, n<sup>o</sup>. 1, p. 87–108, 2000.
- [226] J. Hadamard, *Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques ; leçons professées à l’Université Yale*. Hermann et cie, 1932.
- [227] M. L. Ravalec, B. Noetinger et L. Y. Hu, “The FFT moving average (FFT-MA) generator : An efficient numerical method for generating and conditioning Gaussian simulations,” *Mathematical Geology*, vol. 32, n<sup>o</sup>. 6, p. 701–723, 2000.
- [228] M. L. Ravalec et E. Mouche, “Calibrating transmissivities from piezometric heads with the gradual deformation method : An application to the culebra dolomite unit at the waste isolation pilot plant (WIPP), new mexico, USA,” *Journal of Hydrology*, vol. 472-473, p. 1–13, nov 2012.
- [229] J. E. Capilla et C. Llopis-Albert, “Gradual conditioning of non-Gaussian transmissivity fields to flow and mass transport data : 1. theory,” *Journal of Hydrology*, vol. 371, n<sup>o</sup>. 1-4, p. 66–74, jun 2009.
- [230] C. Llopis-Albert et J. E. Capilla, “Gradual conditioning of non-Gaussian transmissivity fields to flow and mass transport data : 2. demonstration on a synthetic aquifer,” *Journal of Hydrology*, vol. 371, n<sup>o</sup>. 1-4, p. 53–65, jun 2009.
- [231] B. Marteau, D. Y. Ding et L. Dumas, “A generalization of the local gradual deformation method using domain parameterization,” *Computers & Geosciences*, vol. 72, p. 233–243, nov 2014.
- [232] H. Jeong, S. Ki et J. Choe, “Reservoir characterization from insufficient static data using gradual deformation method with ensemble Kalman filter,” *Energy Sources, Part A : Recovery, Utilization, and Environmental Effects*, vol. 32, n<sup>o</sup>. 10, p. 942–951, mar 2010.
- [233] X. Emery et C. Lantuéjoul, “TBSIM : A computer program for conditional simulation of three-dimensional Gaussian random fields via the turning bands method,” *Computers & Geosciences*, vol. 32, n<sup>o</sup>. 10, p. 1615–1628, 2006. [En ligne]. Disponible : <http://www.sciencedirect.com/science/article/pii/S0098300406000549>
- [234] H. Zhou, L. Li et J. J. Gómez-Hernández, “Three-dimensional hydraulic conductivity upscaling in groundwater modeling,” *Computers & Geosciences*, vol. 36, n<sup>o</sup>. 10, p. 1224–1235, oct 2010.
- [235] J. Molson et E. O. Frind, “SALTFLOW, User Guide, Density-dependent flow and mass transport model in three dimensions, Version 4.0, Université Laval and University of Waterloo, 2017.” University of Waterloo, Rapport technique, déc. 2017.

- [236] R. M. Barnett, J. G. Manchuk et C. V. Deutsch, “Projection pursuit multivariate transform,” *Mathematical Geosciences*, vol. 46, n<sup>o</sup>. 3, p. 337–359, 2014. [En ligne]. Disponible : <https://doi.org/10.1007/s11004-013-9497-7>
- [237] D. G. Luenberger et Y. Ye, *Linear and Nonlinear Programming*. Springer International Publishing, 2016.
- [238] R. P. Brent, *Algorithms for minimization without derivatives*. Englewood Cliffs, N.J : Prentice-Hall, 1972.
- [239] M. L. Ravalec-Dupin et B. Noetinger, “Optimization with the gradual deformation method,” *Mathematical Geology*, vol. 34, n<sup>o</sup>. 2, p. 125–142, 2002.
- [240] M. Vukovic et A. Soro, *Determination of hydraulic conductivity of porous media from grain-size composition*. Littleton, Colo. : Water Resources Publications, 1992.
- [241] J. Doherty, “Calibration and uncertainty analysis for complex environmental models,” Watermark Numerical Computing, Brisbane, Australia, Rapport technique, 2015.
- [242] M. C. M. C. Hill et C. R. Tiedeman, “Effective groundwater model calibration : with analysis of data, sensitivities, predictions, and uncertainty,” dans *Effective groundwater model calibration*. John Wiley & Sons, 2006.
- [243] J. Molson et E. Frind, “On the use of mean groundwater age, life expectancy and capture probability for defining aquifer vulnerability and time-of-travel zones for source water protection,” *Journal of Contaminant Hydrology*, vol. 127, n<sup>o</sup>. 1-4, p. 76–87, jan 2012.
- [244] E. Frind, D. Muhammad et J. Molson, “Delineation of three-dimensional well capture zones for complex multi-aquifer systems,” *Ground Water*, vol. 40, n<sup>o</sup>. 6, p. 586–598, nov 2002.
- [245] R. M. Neupauer et J. L. Wilson, “Adjoint-derived location and travel time probabilities for a multidimensional groundwater system,” *Water Resources Research*, vol. 37, n<sup>o</sup>. 6, p. 1657–1668, jun 2001.
- [246] —, “Numerical implementation of a backward probabilistic model of ground water contamination,” *Groundwater*, vol. 42, n<sup>o</sup>. 2, p. 175–189, mar 2004.
- [247] —, “Backward probability model using multiple observations of contamination to identify groundwater contamination sources at the massachusetts military reservation,” *Water Resources Research*, vol. 41, n<sup>o</sup>. 2, feb 2005.
- [248] L. W. Gelhar et C. L. Axness, “Three-dimensional stochastic analysis of macrodispersion in aquifers,” *Water Resources Research*, vol. 19, n<sup>o</sup>. 1, p. 161–180, 1983. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1029/WR019i001p00161/abstract>

- [249] T. Hermans, F. Nguyen et J. Caers, “Uncertainty in training image-based inversion of hydraulic head data constrained to ERT data : Workflow and case study,” *Water Resources Research*, vol. 51, n°. 7, p. 5332–5352, 2015. [En ligne]. Disponible : <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014WR016460>
- [250] A. A. S. Barfod, T. N. Vilhelmsen, F. Jørgensen, A. V. Christiansen, A.-S. Høyer, J. Straubhaar et I. Møller, “Contributions to uncertainty related to hydrostratigraphic modeling using multiple-point statistics,” *Hydrology and Earth System Sciences*, vol. 22, n°. 10, p. 5485–5508, oct 2018, publisher : Copernicus GmbH. [En ligne]. Disponible : <https://www.hydrol-earth-syst-sci.net/22/5485/2018/>
- [251] K. Mahmud, G. Mariethoz, J. Caers, P. Tahmasebi et A. Baker, “Simulation of earth textures by conditional image quilting,” *Water Resources Research*, vol. 50, n°. 4, p. 3088–3107, apr 2014. [En ligne]. Disponible : <http://onlinelibrary.wiley.com/doi/10.1002/2013WR015069/abstract>
- [252] A. A. Efros et W. T. Freeman, “Image quilting for texture synthesis and transfer,” dans *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '01. ACM, 2001, p. 341–346. [En ligne]. Disponible : <http://doi.acm.org/10.1145/383259.383296>
- [253] D. D’Or, “Spatial prediction of soil properties, the Bayesian maximum entropy approach,” phdthesis, Université catholique de Louvain, 2003.
- [254] T. Hermans, F. Nguyen, M. Klepikova, A. Dassargues et J. Caers, “Uncertainty Quantification of Medium-Term Heat Storage From Short-Term Geophysical Experiments Using Bayesian Evidential Learning,” *Water Resources Research*, vol. 54, n°. 4, p. 2931–2948, 2018. [En ligne]. Disponible : <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017WR022135>
- [255] T. Hermans, N. Lesparre, G. De Schepper et T. Robert, “Bayesian evidential learning : a field validation using push-pull tests,” *Hydrogeology Journal*, vol. 27, n°. 5, p. 1661–1672, aug 2019. [En ligne]. Disponible : <https://doi.org/10.1007/s10040-019-01962-9>
- [256] D. Higdon, J. Swall et J. Kern, *Non-Stationary Spatial Modeling*, ser. Bayesian Statistics 6, J. M. Bernardo, J. O. Berger, A. P. Dawid et A. Smith, édit. Oxford University Press., 1999.
- [257] M. Liang, D. Marcotte et N. Benoit, “A comparison of approaches to include outcrop information in overburden thickness estimation,” *Stochastic Environmental Research and Risk Assessment*, vol. 28, n°. 7, p. 1733–1741, dec 2013. [En ligne]. Disponible : <http://link.springer.com/article/10.1007/s00477-013-0835-6>

- [258] M. Rivest, “Interpolation de la concentration de contaminant par krigeage avec covariance non stationnaire en coordonnées d’écoulement,” PhD thesis, École Polytechnique de Montréal, 2012.
- [259] C. J. Paciorek et M. J. Schervish, “Spatial modelling using a new class of nonstationary covariance functions,” *Environmetrics*, vol. 17, n°. 5, p. 483–506, 2006. [En ligne]. Disponible : <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2157553/>
- [260] P. Shamsipour, D. Marcotte, M. Chouteau, M. Rivest et A. Bouchedda, “3D stochastic gravity inversion using nonstationary covariances,” *GEOPHYSICS*, vol. 78, n°. 2, p. G15–G24, mar 2013. [En ligne]. Disponible : <http://library.seg.org/doi/abs/10.1190/geo2012-0122.1>

## ANNEXE A NOTIONS DE STATIONNARITÉ

La stationnarité est une caractéristique d'une fonction aléatoire  $Z(x)$  qui décrit la variation spatiale ou temporelle des distributions de variables aléatoires  $z(x)$  qui la caractérise. On reconnaît trois types de stationnarité, la stricte, celle du second ordre et l'intrinsèque [143]. La stationnarité stricte est représentée par une fonction aléatoire dont la distribution ne varie pas par translation dans tout l'espace (phénomène homogène). La stationnarité de second ordre est représentée par une fonction aléatoire dont les moments statistiques ne varient pas par translation dans tout l'espace. Dans ce cas, on ne considère que les deux premiers moments représentés par une moyenne constante et une fonction de covariance (définie positive) qui dépend de la distance ( $h$ ) entre deux points. Exceptionnellement, si la fonction aléatoire est gaussienne, la stationnarité devient alors stricte puisqu'une fonction aléatoire gaussienne est complètement définie par les deux premiers moments (moyenne et covariance). La stationnarité intrinsèque quant à elle est représentée par une fonction aléatoire correspondant à un incrément  $Yh(x) = Z(x+h) - Z(x)$  ayant comme propriété une moyenne correspondante à une dérive linéaire et une variance égale à  $2\gamma(h)$ . Le cas d'usage courant est toutefois celui où la moyenne des incréments est supposée égale à 0. La notion d'incrément se généralise à des incréments d'ordre supérieur permettant de filtrer des variations de la moyenne d'ordre plus élevé. La moyenne de ces incréments est 0 et leur variance définit la fonction de covariance généralisée, le variogramme correspondant (à une constante près et à un signe près) à la covariance généralisée d'un incrément d'ordre 0 ( $Z(x+h) - Z(x)$ ).

En géostatistique appliquée, le choix (hypothèse) du type de stationnarité se base sur les données disponibles et les connaissances de la variable étudiée. De manière classique, on pose une hypothèse de stationnarité du second ordre (moyenne, variance et covariance constantes sur le domaine). Les statistiques sont donc indépendantes de l'emplacement sur le domaine. Lorsque l'hypothèse de stationnarité ne s'applique pas, le domaine est dit non-stationnaire. La figure 7 montre un exemple de domaine stationnaire (gauche) et un domaine où la moyenne est non-stationnaire (droite).

La non-stationnarité du modèle peut aussi s'appliquer à la variance et à la fonction de covariance [41, 256–260]. Higdon et al. [256] ont fait des progrès importants dans la définition de structures de covariance non stationnaires. À titre d'exemple, les auteurs ont obtenu une version non-stationnaire de la fonction de covariance stationnaire exponentielle carrée. Ce résultat a été obtenu par la convolution de noyaux de type gaussien dont les paramètres varient spatialement. Le modèle proposé par Higdon et al. [256] a été généralisé par Paciorek

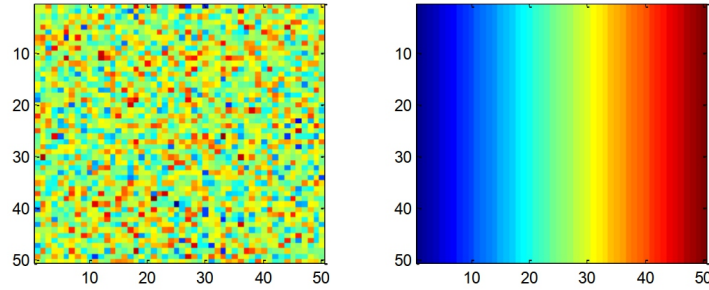


Figure A.1 Domaine stationnaire (gauche) et non-stationnaire (droite).

et Schervish [259] pour l'utilisation de fonctions de covariance de type Matérn et Cauchy non stationnaires. À partir d'un modèle numérique de terrain à fort contraste, Paciorek et Schervish [259] illustrent comment produire le modèle de covariance non stationnaire à partir de deux structures spatiales distinctes.

Soulignons que si la non-stationnarité de la moyenne entraîne la non-stationnarité de la covariance, il ne s'agit que d'un type de non-stationnarité. Il existe plusieurs situations où une non-stationnarité supplémentaire sur la covariance existe indépendamment de la non-stationnarité de la moyenne. Ainsi, on peut avoir deux zones de variance différente et de même moyenne, on peut avoir un champ montrant une portée qui s'accroît selon une certaine direction, ou deux zones de portées différentes. Les directions de l'anisotropie d'une covariance peuvent fluctuer dans l'espace, etc. Également, l'effet proportionnel couramment observé en géostatistique, où la variance fluctue selon la moyenne locale, peut être vu également comme une manifestation de non-stationnarité.

## ANNEXE B    SIMULATIONS GÉOSTATISTIQUES DU TENSEUR COMPLET DE CONDUCTIVITÉ HYDRAULIQUE DE BLOC 3D EN TENANT COMPTE DE LA VARIABILITÉ INTERNE À L'ÉCHELLE LOCALE

Dans les systèmes hydrostratigraphiques régionaux, les hétérogénéités représentées par la conductivité hydraulique ( $K$ ) jouent un rôle central sur les vitesses d'écoulement et la modélisation du transport de contaminants. Les mesures typiques de  $K$  provenant d'essais sur le terrain sont généralement représentatives d'un ou plusieurs hydrofaciès locaux d'une même unité hydrostratigraphique (UHS). À l'échelle régionale, de nombreuses unités UHS existent, chacune montrant une distribution de valeurs de  $K$  s'étendant sur plusieurs ordres de grandeur. Pour une modélisation numérique réaliste, une échelle intermédiaire c'est-à-dire une échelle de bloc, est requise. Chaque bloc doit donc contenir un tenseur 3D de  $K$  équivalent qui prend en compte la connectivité des hydrofaciès et les variations de conductivité à l'échelle locale.

L'approche suivie dans la thèse comporte quatre étapes : 1- simulation à l'échelle locale de  $K$  pour chaque UHS, 2- mise à l'échelle des réalisations locales (quasi-ponctuelles) en bloc 3D représenté par un tenseur complet de  $K$  à l'aide d'un simulateur d'écoulement par éléments finis, 3- définition de la covariance spatiale du tenseur- $K$  de bloc, et 4- simulation directe des tenseurs de bloc  $K$  à l'échelle régionale. Cette approche a été appliquée à la modélisation régionale de l'écoulement des eaux souterraines sur un modèle déterministe 3D complexe du bassin versant Innisfil Creek en Ontario (Canada).

L'analyse des composantes du tenseur  $K$ , mis à l'échelle, a révélé un contrôle clair des composantes principales avec de fortes corrélations entre elles. L'impact associé à l'incertitude du tenseur  $K$  sur une réponse typique de simulation d'écoulement a été évalué et s'est révélé non négligeable.

L'approche proposée a été présentée à la 12e conférence internationale GEOENV tenue à Belfast en 2018. L'article de conférence a été republié (avec permission) dans : Benoit, N., Marcotte, D., Molson, J.W., and Pasquier, P., 2020. Geostatistical simulations of the full 3D block hydraulic conductivity tensor considering inner local-scale variability; Geological Survey of Canada, Open File 8700, 8 p. <https://doi.org/10.4095/321822>.

## ANNEXE C    PSEUDOCODES DES PRINCIPAUX ALGORITHMES

---

### Algorithm 1: Simulation MCP

---

**Input:** image d'entraînement  $TI$ , grille de simulation  $grid$ , données  $data$ , nombre maximal de voisin  $nmax$ , rayon de recherche  $dmax$ , germe de simulation  $seed$ , nombre de réalisation  $nreal$   
**OPTIONS :** interdire le tirage directement de la distribution marginale  $marginal$ , prévenir l'inversion verticale des unités  $invcor$ , imposer une recherche par octant  $oct$ , imposer un nombre maximal de données par octant  $nmaxOct$ , imposer une séquence de simulation multigrille  $multigrid$

**Output:** réalisations des  $UHS$

```

    // Pour simplifier le pseudocode, les options ne sont pas prise en compte à l'exception de invcor.
1  Calcul des transiogrammes des probabilités bivariées  $gh$  à partir de la  $TI$ 
2  Définition d'un vecteur  $gridID$  de 1 de dimension égale à  $grid$  où  $gridID$  prend une valeur de 0 aux
   points simulés
3  Assignment des données sur  $grid$  et une valeur correspondante de 0 à  $gridID$ 
4  Calcul des probabilités marginales  $pmar0$  pour  $TI$ 
5  Définir la séquence aléatoire de la simulation
   /* Tant que tous les points de la grille ne sont pas simulés: */
6  while  $Sum(gridID > 0)$  do
   /* Pour tous les points de la grille: */
7     for  $grid$  do
8         trouver les données voisines  $nvois$  dans  $grid$ 
9         if  $nvois > 0$  then
10            Calcul des probabilités bivariées et compilation de la distribution conditionnelle  $pk0$ 
11        else
12            La distribution conditionnelle  $pk0 = pmar0$ 
13        Effectue le tirage de la loi conditionnelle de  $UHS$  /* Vérifie s'il y a inversion verticale des
           unités si l'option  $invcor = 1$  : */
14        if  $invcor = 1$  then
15            S'il y a inversion verticale d'unité dans la grille de simulation,  $UHS$  est retiré et
            $gridID = 0$  // le point est remplacé à la fin de la séquence de simulation et sera
           resimulé
16        else
17             $gridID = 0$  // le point est simulé et devient une donnée
   /* Fin de la boucle While */

```

---



---

**Algorithm 2:** Méthode de mise à l'échelle de la K

---

**Input:** Conductivité hydraulique estimée à l'échelle locale  $K$ , grille de simulation à l'échelle locale  $gridLoc$ , grille de simulation à l'échelle de l'élément du modèle régional  $gridElem$ , grille de simulation à l'échelle régionale  $gridReg$

*/\* La taille de  $gridElem$  est égale à  $gridLoc$ . La taille maximale dans chacune des directions doit être plus grande que la portée anticipée des différentes composantes du tenseur. \*/*

**Output:** Réalisations des tenseurs de conductivité hydraulique  $tenseursK$

*/\* Pour chacune des unités hydrostratigraphique  $UHS$ : \*/*

- 1 **for**  $UHS$  **do**
  - /\* Simuler la structure spatiale à l'échelle locale: \*/*
  - 2 Définir la distribution statistique de  $statLoc$  de  $K$
  - 3 Modéliser le variogramme 3D  $varioLoc$  de  $K$
  - 4 Simuler sur  $gridLoc$  une réalisation non conditionnelle  $simLoc$  avec  $varioLoc$  contraint par  $statLoc$
  - /\* Mise à l'échelle de la réalisation à l'échelle locale vers l'échelle des éléments du modèle régionale: \*/*
  - 5 **for**  $gridElem$  **do**
    - 6 Extraire  $simLoc$  à l'intérieur d'un bloc  $3 \times 3 \times 3$   $gridElem3 \times 3$  centré sur  $gridElem$
    - /\* Les deux prochaines étapes utilisent le calcul parallèle pour accélérer le temps de calcul \*/*
    - 7 Simuler l'écoulement de  $gridElem3 \times 3$  selon 8 configurations différentes de conditions frontières  $condFront$  à charges constantes
    - 8 Extraire le bloc central  $ElemCentre$  de  $gridElem3 \times 3$  et calculer les flux moyens  $(\bar{q}_x, \bar{q}_y, \bar{q}_z)$  et les gradients moyens  $(\nabla \bar{h}_x, \nabla \bar{h}_y, \nabla \bar{h}_z)$  dans les 3 directions principales pour chacune des conditions frontières.
    - 9 Assembler les gradients moyens dans une matrice [(3 directions x 8  $condFront$ ), 6 composantes] et flux moyens [3 directions x 8  $condFront$ , 1] dans un système d'équation surdéterminé suivant la loi de Darcy
    - 10 Calculer le  $tenseursK$  [6 composantes, 1] par moindre carré de  $gridElem$
    - /\* Simuler les réalisations  $tenseursK$  sur l'ensemble du modèle régionale  $gridReg$ : \*/*
    - 11 Définir les corrélations entre les composantes  $tenseursK$  de  $gridElem$
    - 12 Modéliser le variogramme 3D  $varioElem$  des  $tenseursK$
    - /\* Si une portée de  $varioElem$  est plus grande que la taille de  $gridElem$ , la taille des grilles initiales est sous-estimée. \*/*
    - 13 Définir le modèle de corégionalisation des  $tenseursK$
    - 14 Simuler  $n$  réalisations non conditionnelle  $simReg$  sur la grille du modèle régionale  $gridReg$
    - 15 Effectuer la transformation bivariable de  $simReg$  pour préserver la relation entre les composantes et la contrainte d'inégalité ( $K_v \leq K_h$ )

*/\* Fin de la boucle For \*/*

---

---

**Algorithm 3:** Calage des tenseurs-K par déformation graduelle

---

```

/* L'algorithme permet le calage des tenseurs-K et/ou de la recharge. Pour simplifier le pseudocode la
déformation est effectuée sur les deux paramètres en simultannée. */
Input: Ensemble de  $nr$  réalisations de chacune des composantes des tenseurs de conductivité
hydraulique  $(K_{xx}, K_{yy}, K_{zz})$ , ensemble de  $nr$  réalisations de la recharge  $R$  et une série de
 $nobs$  observations de charges hydrauliques  $Hobs$ .
Output: Composantes calées des tenseurs-K  $(K_{xx}, K_{yy}, K_{zz})$  et recharge calée ( $R$ )
1 Choisir le critère arrêt de l'algorithme d'inversion
/* arrêt est un scalaire qui est itérativement comparée à la valeur de la fonction objectif  $FO$ . */
2 Choisir les paramètres d'optimisation du coefficient de déformation  $t$  : maxiter correspondant au
nombre d'itération maximal de déformation entre deux réalisations, tolx critère d'arrêt de la
déformation entre deux réalisations comparé à  $FO$ , nmax est nombre maximal d'itération avant de
permettre à l'algorithme de réduire l'intervalle du coefficient de déformation  $t$ , valeur minimale
absolue que peut prendre  $t$  tmin. /* L'optimisation du coefficient de déformation  $t$  permet une
sélection objective de ce paramètre ainsi qu'une réduction progressive de l'intervalle possible de
valeurs pour accélérer les calculs. */
3 Sélectionner une réalisation initiale des composantes des tenseurs-K ( $Y_{Kcur}$ ) et de la recharge ( $Y_{Rcur}$ )
4 Initialiser un compteur d'itération à compteur=1
5 Initialiser la fonction objectif où  $FO=1/nobs \times \sum abs(Hsim - Hobs)$ 
/* DÉFORMATION GRADUELLE sur  $nr$  réalisations */
6 while  $FO > \text{arrêt}$  OR compteur  $> nr$  do
7   Sélectionner une nouvelle réalisation des tenseurs-K ( $Y_{Knew}$ ) et de la recharge ( $Y_{Rnew}$ )
8   Initialisation des limites du coefficient de déformation  $t$  entre  $bnd1 = -pi$  et  $bnd2 = pi$ 
   /* Optimisation du coefficient de déformation  $t$  pour une sélection objective et accélérer le temps
   de calcul en réduisant l'intervalle de valeur possible */
9   if compteur  $> nmax$  then
10      $t = \max(abs(t \text{ des } nmax \text{ dernières itérations}))$ 
11     if  $t < 0.1$  then
12       /* valeur minimale que peut prendre  $t$  */
13        $t = tmin$ ;
14      $bnd1 = -t$ 
15      $bnd2 = t$ 
   /* Optimisation de la  $FO$  sur  $t$  entre deux réalisations: */
16   while  $FO > tol_x$  OR niter  $< maxiter$  do
17     Application de la déformation graduelle entre  $Y_{Kcur}$  et  $Y_{Knew}$  ainsi que  $Y_{Rcur}$  et  $Y_{Rnew}$  pour
un  $t$  donné entre  $bnd1$  et  $bnd2$ 
18     applique la transformation bivariable
19     lance le simulateur d'écoulement
20     extraction des charges simulés aux points d'observation
21     Calcul de la  $FO$ 
22   compteur = compteur + 1
/* Fin de la boucle While */

```

---