



HAL
open science

Caractérisation et modélisation des écoulements dans les milieux fracturés

Romain Le Goc

► **To cite this version:**

Romain Le Goc. Caractérisation et modélisation des écoulements dans les milieux fracturés. Sciences de la Terre. Université Rennes 1, 2009. Français. NNT: . tel-00467417

HAL Id: tel-00467417

<https://theses.hal.science/tel-00467417>

Submitted on 26 Mar 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE / UNIVERSITÉ DE RENNES 1

sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Sciences de la terre

École doctorale sciences de la matière

présentée par

Romain Le Goc

préparée à l'unité de recherche UMR6118, GR

Géosciences Rennes

UFR Sciences et Propriétés de la matière

Caractérisation et
modélisation des
écoulements dans les
milieux fracturés

Thèse soutenue à Rennes

le 22 Décembre 2009

devant le jury composé de :

Kerry GALLAGHER,

professeur à l'université Rennes 1/*président*

Frederick DELAY,

professeur à l'université de Poitiers / *rapporteur*

Philippe RENARD

directeur de recherche à l'université de Neuchâtel (Suisse)

/ *rapporteur*

Jesus CARRERA

Professeur au CSIC (Espagne) / *examineur*

Olaf KOLDITZ

Professeur à l'UFZ (Allemagne) / *examineur*

Philippe DAVY

directeur de recherche CNRS, université de Rennes 1 /

directeur de thèse

Jean-Raynald DE DREUZY

chargé de recherche CNRS, université de Rennes 1 / *co-*

directeur de thèse

Daniel BILLAUX

Itasca Consultant SAS / *invité*

Raymond MUNIER

SKB / *invité*

Résumé

La gestion d'un aquifère repose sur l'étude de ses propriétés hydrauliques. Nous proposons une approche computationnelle afin de caractériser la structure physique d'un aquifère et les propriétés hydrauliques associées, en focalisant sur l'étude des milieux hétérogènes fracturés où les écoulements sont fortement chenalés. Ainsi, une première partie cherche à définir le modèle de fracturation sous-jacent à deux sites naturels suédois à partir d'observations corrigées des biais d'échantillonnage. La modélisation choisie repose sur une distribution des longueurs en loi de puissance et sur un terme de densité dépendant des orientations. Les propriétés hydrauliques des milieux hétérogènes, poreux et fracturés, sont ensuite caractérisées grâce à deux indicateurs statistiques décrivant le degré de chenalisation des écoulements à partir de la distance entre deux chenaux principaux et de la longueur efficace des chenaux. Ces indicateurs permettent de relier les propriétés physiques du milieu (hétérogénéité, organisation) à ses propriétés hydrauliques. Finalement, nous tentons d'identifier les composantes principales de la structure de perméabilité à partir de données de charge hydraulique en résolvant le problème inverse associé. On montre ainsi qu'il est possible d'identifier les principales structures contrôlant les écoulements à condition d'utiliser des données contenant une information sur la perméabilité pertinente à l'échelle du site, une paramétrisation représentative des structures principales et une méthodologie appropriée. Ces différentes approches permettent de définir un modèle pertinent des propriétés hydrauliques d'un milieu complexe.

Managing an aquifer relies on the understanding of its physical properties. We propose in this study a computational approach in order to characterize the physical structure of aquifers and the corresponding hydraulic properties. We focus on heterogeneous fractured media where flow is highly channeled. We first investigate the fracturing properties two Swedish sites by defining the corresponding DFN model. The DFN model framework is a power-law length distribution with a density term that depends on orientation, which parameters are calibrated with observations of fractures (lengths, orientations) by correcting the sampling bias. Hydraulic properties of heterogeneous porous and fractured media are then described by two channeling indicators based on statistical properties of flow and characterizing the distance between channels and their actual effective length. These indicators are used to relate the hydraulic properties to the physical structure of the medium. Finally, we develop a strategy to identify the position and the transmissivity of main flow channels from head data by solving the corresponding inverse problem. We show that this identification is possible provided that 1- data are remotely sensitive to the main structures underlying flow channels, 2- the parameterization focus on these structures and 3- the strategy is carefully set up. All these approaches can be combined to define a relevant model of the hydraulic properties of complex media.

Avant-propos

Je dédie cet ouvrage à tous ceux qui, directement ou indirectement, l'ont rendu possible. Ceci concerne en premier lieu mes parents sans qui, naturellement, je ne serais pas là, ainsi que mon frère qui m'a appris, souvent bien malgré lui, la patience. Même si leur rôle est moins direct, mes oncles, tantes et cousins ont, au moins grâce à leur soutien sans faille, contribué à ce que les travaux présentés ici aillent à leur terme. Un grand merci en particulier à Charlène et Vincent pour les soirées-restaurant qui m'ont permis de me nourrir, si ce n'est de manière équilibré, tout du moins de manière suffisante. Merci également à Matthieu et Mickaël pour les quelques soirées-poker et à tous les autres pour les soirées en général.

A côté de ces personnes je tiens à remercier mes amis, et particulièrement Thomas, Line et Sébastien, qui, en me parlant d'autres choses, ont permis que je garde les pieds sur terre. Je souhaite au passage bon courage à Sébastien qui affronte ces jours-ci ce que j'ai affronté ces derniers mois. J'adresse également tout particulièrement mes félicitations à Line car, si je sais combien il peut être difficile d'« accoucher » d'une thèse, je ne saurais jamais ce que signifie accoucher d'un enfant. Quant à Thomas, il sait qu'il ne se débarrassera pas de moi si facilement.

Que serait un travail de thèse sans l'accompagnement journalier des autres étudiants, du midi au RU ou dans la salle commune au jeudi soir au O' Connells. De peur d'oublier quelqu'un je ne citerai personne mais ceux pour qui les vacances scolaires signifient l'absence de journal et le jeudi le jour "de l'happy hour" se reconnaîtront. Je dois également remercier Bernard et Diane qui, sans le savoir, m'ont fait perdre beaucoup de temps et à qui, en toute honnêteté, j'aurai dû réserver un chapitre de cet ouvrage. Je remercie également mes collègues d'H2OLAB qui comprennent ce que développement collaboratif signifie et qui continueront à subir mes choix de développement encore un certain temps. Un grand merci à Cyril, mon partenaire de kung-fu préféré, qui n'a pas fini de s'entraîner avec moi et à qui je dois encore quelques coups mal placés.

Merci surtout à mes collègues de bureau, tout d'abord Loïc qui a déjà soutenu et est parti vers d'autres cieux, Etienne puis Pietro le florentin et Christoff, pour m'avoir supporté au jour le jour. J'espère qu'ils ne regrettent pas d'avoir partagé le peu d'espace disponible avec moi et je leur souhaite bonne chance pour la suite.

Je termine ces remerciements par les personnes les plus impliquées dans les travaux présentés ici. Tout d'abord je remercie Jean-Raynald, à qui je promets qu'on écrira de nouveaux articles ensemble, de préférence tard le soir. Merci à Philippe, qui en tant que directeur a su diriger ces recherches et leur donner un sens et merci à Caroline qui m'a tout appris sur les réseaux de fractures. Comme je suis plutôt vénal, je remercie également Daniel qui a financé une partie de ces travaux et a même fini par me proposer un travail, comme quoi tout est possible.

Un grand merci à tous le personnel de Géosciences, y compris Jean-Pierre, et au personnel d'Itasca. Merci à tous ceux qui ne se reconnaîtront pas dans ces lignes mais qui ont pourtant joué un rôle.

Sommaire

Résumé	I
Avant-propos	III
Sommaire	V
Liste des illustrations	IX
Introduction	1
Partie 1 État de l'art sur les modèles de perméabilité intégrant la chenalisation des écoulements	5
A. Modèles de perméabilité.....	5
1. Décrire la perméabilité dans les milieux poreux	5
2. Les milieux fracturés	9
B. Calibration des modèles de perméabilité : le problème inverse	14
1. Les principes du problème inverse	14
2. Exemple de résolutions du problème inverse en hydrogéologie	19
3. Adaptation du problème inverse aux milieux fracturés.....	22
4. Conclusion vis-à-vis des milieux très chenalisés	30
C. Positionnement des travaux effectués par rapport aux travaux précédents	32
Partie 2 Modélisation numérique des réseaux de fractures	33
A. Présentation de la plateforme logicielle.....	33
1. Description de PARADIS	35
2. Description de MPFRAC.....	36
3. Les fonctionnalités de la plateforme communes aux différentes applications	37
4. Les partenaires de la plateforme.....	38
B. Principes et fonctionnements de fonctionnalités développées pour le doctorat	38
1. Modèles de réseaux de fractures	39
2. Simulation d'écoulements et de transport inerte dans un DFN 2D	40
3. Analyse d'affleurement	42
4. Problème inverse.....	42
C. Conclusions	44
Partie 3 Définition d'un modèle de fracturation à partir d'observations du milieu naturel : application aux sites de Forsmark et Laxemar (Suède).....	45
A. Introduction	45
B. Description des sites de Forsmark et Laxemar	46
C. Méthode d'analyse de la longueur des traces de fracture appliquée à des affleurements et linéaments Suédois.	48
1. Résumé	48

2.	Fracture trace sizes analysis method applied to Swedish outcrops and lineaments	49
3.	Résultats complémentaires	64
D.	Analyse statistique des domaines de fracturations	65
1.	Définition du modèle de densité des domaines locaux de fracturation	65
2.	Analyse statistiques des domaines de fracturations	68
3.	Résultats complémentaires	84
4.	Conclusions sur l'analyse du terme de densité.....	85
E.	Conclusions sur la définition d'un modèle de site	85
Partie 4	Analyse statistique des écoulements : définitions d'indicateurs de chenalisation.....	89
A.	Indicateurs de chenalisation basés sur les caractéristiques statistiques du flux.....	89
	Résumé	89
B.	Résultats complémentaires	103
1.	Statistiques sur la distribution des valeurs de flux.....	103
2.	L'utilisation du ratio de participation S_2	104
C.	Conclusion.....	105
Partie 5	Identification des chenaux d'écoulement principaux par méthode inverse.....	107
A.	Identification des chenaux d'écoulement dans les milieux fracturés par stratégie inverse. 107	
1.	Résumé	107
2.	Inverse problem strategy to identify flow channels in fractured media.....	108
1.	Introduction.....	109
2.	Parameterization of fracture flow structures in terms of channels.....	110
2.1.	Hierarchical organization of channels	110
2.2.	Quantitative definition of a channel	111
2.3.	Iterative identification of channels	112
3.	Inverse problem methodology and implementation	114
3.1.	Hierarchical identification of channels.....	114
3.2.	Single channel identification	116
3.3.	Definition of the objective function	116
3.4.	Parameter calibration.....	117
3.5.	Numerical method for solving the direct problem.....	118
3.6.	Post-processing analysis of the solution distribution.....	119
3.7.	Synthesis of parameters of the identification strategy.....	120
4.	Results	121
4.1.	Tested configurations.....	121
4.2.	Data	121
4.3.	Results for a simple configuration.....	122

4.4. Results for an intermediary configuration	126
4.5. Results for a complex configuration.....	128
4.6. Analyses of the final models on the set of test-configurations.....	130
5. Discussion	134
6. Conclusion	136
Appendix A. Simulated annealing parameterization	137
B. Influence des conditions aux limites sur la sensibilité des données de charges stationnaires aux structures d'écoulement.	139
1. Résumé	139
2. Influence of boundary conditions on the sensitivity of steady head data to the flow structure	139
C. Conclusions générales	154
Partie 6 Synthèse, discussion et conclusion	157
A. De la caractérisation des milieux complexes.....	157
1. Analyse des données et contrainte sur la paramétrisation	158
2. Ajout d'indicateurs sur les propriétés du milieu	160
3. L'importance de la paramétrisation.	161
4. La calibration des paramètres à partir des données.	162
5. Résumé des différentes approches.....	163
B. Discussion sur les méthodes proposées et les résultats obtenus	164
1. Définition d'un modèle stochastique de fracturation	164
2. Définitions d'indicateurs sur la chenalisation des écoulements.....	165
3. Résolution du problème inverse	165
C. Perspectives	167
D. Conclusion générale	168
Références bibliographiques	157
Partie 7 Annexes	182

Liste des illustrations

- figure 1-1 : Problématique du travail de thèse : comment caractériser les milieux fortement hétérogènes où les écoulements sont très chenalisés ? Par une observation du milieu naturel permettant de modéliser sa structure physique, une analyse des données hydrauliques pour caractériser les propriétés d'écoulements et la définition d'un modèle de site intégrant les propriétés physiques et hydrauliques du milieu par une méthode inverse..... 2
- figure 1-1 : Modèle de champs de perméabilité. En haut se trouve champ de perméabilité et en bas le flux correspondant. Les modèles par zonation et par krigeage sont obtenus à partir de 100 données réparties de manière régulière dans le milieu. 7
- figure 1-2 : Illustration de champs de perméabilité intégrant des propriétés de chenalisation..... 8
- figure 1-3 : trois représentations différentes d'un milieu fracturé. Gauche : approche discrète, le milieu est représenté par un réseau de fracture et les écoulements sont calculés à l'intérieur de chaque fracture. Milieu : approche continue, le milieu est représenté par une grille hétérogène de perméabilité. (Droite) : approche mixte, le milieu est représenté par une grille hétérogène contenant des plans de fractures 10
- figure 1-4 : Calibration du modèle en loi de puissance des longueurs de fracture appliquée au bassin d'Hornelen (Norvège). À gauche, carte de traces de fracture à différentes échelles et à droite, distribution des longueurs normalisée par la surface d'échantillonnage. D'après les figures 1 et 6 de [Bour et al., 2002]..... 11
- figure 1-5 : Réseaux de fractures ayant les mêmes propriétés statistiques. (Haut) : le réseau de fracture. La couleur associée à chaque fracture correspond au logarithme de sa transmissivité. (Bas) : les écoulements correspondant au réseau du haut. Chaque segment est coloré en fonction de la quantité de flux qui le traverse. 13
- figure 1-6 : Illustration de l'algorithme de Newton (modifiée de [Tarantola, 2004]). Pour évaluer la direction de descente au point M on cherche le minimum de la parabole tangente à la courbe au point M et ayant la même courbure que la fonction objectif. Le point M+1 qui a la même abscisse que ce minimum constitue la prochaine étape de l'algorithme. 17
- figure 1-7 : Illustration de la méthode de recuit simulé. L'espace des paramètres est coloré en fonction de la probabilité de transition à partir d'un point initial, les zones en rouge (probabilité forte) seront donc plus échantillonnées lors de la marche aléatoire que les zones en bleus (probabilité faible) (1) état initial : la température est élevée et la majeure partie de l'espace des paramètres est accessible. (2) lorsque la température décroît, on commence à converger et une transition vers les zones de l'espace des paramètres correspondant à une forte valeur de fonction-objectif devient de moins en moins probable. (3) A partir d'une certaine température, seuls les minima locaux sont accessibles. Cependant la probabilité de passer d'une vallée à l'autre reste non négligeable. (4) Vers la fin de l'algorithme, si l'on se trouve dans la vallée du minimum global, la faible température rend tout retour en arrière très peu probable, ce qui permet de converger vers ce minimum. 18
- figure 1-8 : Problème inverse appliqué à un modèle de zones. (1) conductivité hydraulique du milieu de référence (m/d). (2) à (6) : modèle obtenu après résolution du problème inverse contenant de 1 à 5 zones. Les valeurs de RE correspondent à la valeur de la fonction-objectif (eq. (1-4)). D'après [Tsai et al., 2003b]..... 20
- figure 1-9 : Problème inverse avec interpolation. (1) conductivité hydraulique du milieu de référence (m/d). (2) à (5) : modèle obtenu après résolution du problème inverse contenant de 2 à 5

perméabilités de référence. Les valeurs de RE correspondent à la valeur de la fonction-objectif (eq. (1-4)). D'après [Tsai et al., 2005].	20
figure 1-10 : Résultat de l'inversion via la méthode des points pilotes. (1) champs de références et emplacement des points de mesure, (2) tendance générale calculée par krigeage à partir des 13 valeurs de transmissivité accessibles. (3) résultat du problème inverse à partir de 241 points pilotes. La valeur de μ correspond à α^{-1} dans l'équation (1-4). D'après [Alcolea et al., 2006].	21
figure 1-11 : Transmissivités d'un milieu karstique calibrées par un modèle continu. (1) milieu modélisé par une grille de pas discrétisation $p=2.5\text{km}$ et calibré avec les données réelles. (2) milieu modélisé par une grille de $p=1\text{km}$ et calibré avec les données réelles. (3) $p=2.5\text{km}$ et calibré avec les données interpolées. (4) $p=1\text{km}$ et données interpolées. La valeur de HR, pour « head residuals », représente l'écart entre les données réelles et les données simulées. D'après [Larocque et al., 1999].	23
figure 1-12 : modèle poreux-fracturé. (1) classification des cellules de la grille entre fractures (blanc) et matrice (noir). (2). Conductivités initiales calibrées à partir des données de conductivité et de position des fractures. (3) Modèles finales après résolution du problème inverse en utilisant les données de charge.	23
figure 1-13 : Résultats obtenus avec un modèle en treillis régulier avec des transmissivités assignées par cluster de taille L. (1) $L=1$, (2) $L=10$, (3) $L=20$ et (4) $L=40$. La valeur de RE (« residual error ») correspond au minimum de la fonction-objectif.	24
figure 1-14 : Problème inverse appliqué à un milieu poreux fracturé. (a) milieu synthétique de référence avec ouvertures constantes et (b) résultat du problème inverse correspondant au milieu (a). (c) Schéma de la fracturation visible de cas naturel et (d) résultat du problème inverse appliqué au cas naturel.	25
figure 1-15 : Résultats du problème inverse appliqué à l'identification d'une zone de forte transmissivité dans un milieu faiblement perméable. En haut se trouvent les trois organisations de puits testées avec un pompage proche ou éloigné de la zone recherchée et des puits d'observation couvrant une échelle proche de la taille de la structure ou plus petite. La ligne du milieu présentent les résultats lorsque le milieu poreux est homogène de transmissivité $T=10^{-5}$ et la dernière ligne correspond à un milieu poreux hétérogène. La croix signifie qu'aucune solution n'a été obtenue pour cette configuration. La valeur de RSD correspond au carré de l'erreur résiduelle normalisé par le nombre de paramètres. D'après [Tiedeman et al., 1995].	27
figure 1-16 : Identification d'une zone fracturée dans un milieu hétérogène. La zone de référence (rouge, gauche) est bien identifiée après la résolution du problème inverse (bleu, droite). Pour la légende voir figure 1-15. D'après [Renshaw, 1996].	28
Figure 1-17 : Résolution d'un problème inverse pour l'identification de la structure des zones de forte transmissivité. (a) puits d'observations, (b) connectivité entre puits d'après les tests hydrauliques, (c) zones de forte transmissivité 3D définies a priori à partir des tests hydrauliques, (d) et (e) résultats du problème inverse. Modifié de [Day-Lewis et al., 2000].	29
figure 1-18 : Problème inverse utilisant une description fractale du milieu. (1) Cas synthétique de référence, (2) modèle résultat pour le cas synthétique (moyenne de 30 réalisations), (3) modèle résultat du réservoir de Kesterson et (4) modèle résultat de la mine de Stripa. La valeur de RE correspond à la concordance entre les charges simulées et les charges observées. Le point rouge correspond au point de pompage et les points verts aux puits d'observation.	30
figure 2-1 : Présentation des différents ensembles présents dans H2OLAB. La couleur bleu correspond aux fonctionnalités qui ont été entièrement développées dans le cadre de la thèse et la couleur vertes aux fonctionnalités modifiées ou complétées durant la thèse.	34

figure 2-2 : Illustrations des différentes fonctionnalités de PARADIS. (a) et (d) Génération de champs de perméabilité 2D (a) et 3D (d) selon une loi géostatistique. Les champs présentés ici suivent une distribution log-normale ($\mu=0$, $\sigma^2=1$ pour le 2D et $\sigma^2=3$ pour le 3D), de corrélation Gaussienne avec une longueur de corrélation $\lambda=L/10$ où L est la taille du système. (b) et (f) charges correspondantes pour des conditions aux limites de type perméamètre avec $h_{\max}=10$ et $h_{\min}=0$. (c) et (f) simulation du transport de particules (en rouge) en advection-diffusion par une méthode de "particle tracking" (ici avec 10000 particules)..... 36

figure 2-3 : Illustrations des différentes fonctionnalités de MPFRAC en 2D et 3D. (a) et (d) Réseau de fracture et connectivité, les fractures sont colorées par amas connectés et les intersections sont soulignées (points en 2D, segments en 3D). (b) et (e) Flux correspondant. (c) Transport de particules dans le réseau 2D. (f) Réseau de fracture avec une forme non carrée..... 37

figure 2-4 : Schéma d'exécution d'une application de la plateforme H2OLAB. En ovale sont spécifiées les fonctionnalités génériques et en rectangle les fonctionnalités spécifiques à chaque application..... 38

figure 2-5 : Organisation du noyau de MPFRAC. Le réseau de fracture est manipulé à partir des structures globales (réseau, sous-réseau, système géologique) qui mettent à jour les objets génériques (DFN, fractures et intersections). En fonction de la représentation choisie (2D/3D, géométrie des fractures), ces objets génériques sont modélisés par des objets géométriques spécifiques à partir desquels les fonctionnalités de bases sont accessibles (création, intersection,...)39

figure 2-6 : Illustration du calcul de flux sur un réseau de fracture 2D. (a) réseau de fractures, (b) réseau de liens, chaque lien a une valeur de flux. (c) système linéaire obtenu en appliquant la loi de conservation du flux ($q_{\text{entrant}}=q_{\text{sortant}}$) et les conditions aux limites. (d) flux avec des conditions aux limites de type perméamètre..... 40

figure 2-7 : Configurations possibles lors de la simulation du transport de particules. Arrivée à une intersection en Q_1 , une particule se dirigera vers Q_2 , Q_3 ou Q_4 en fonction du sens des différents débits et d'une probabilité dépendant de la méthode choisie entre "complete mixing" et "streamline routing". 41

figure 2-8 : Principe des fonctionnalités de MSFRAC et coopération avec MPFRAC..... 42

figure 2-9 : Organisation d'ISTHMUS. En rose sont représentées les classes génériques et en bleu les classes plus spécifiques dépendant du type de problème inverse et de la méthode de résolution. 43

figure 3-1 : (a) Localisation de Forsmark (A) et de Laxemar, (B) carte Google™ Maps. (b) domaines tectoniques, modifié de [Wahlgren et al., 2008]. 47

figure 3-2 : Illustration des données disponibles sur les longueurs des fractures (a) à l'échelle d'un affleurement, (b) à l'échelle des linéaments et (c) à l'échelle de la Suède..... 47

Figure 3-3 : Fracture trace size pdf from trace field measurements: the power-law behavior of the distribution is clearly visible. We notice censoring and truncating effects at the left- and the right hand of the distribution. 50

Figure 3-4 : Representation of the area normalization accounting for finite-size effect. White indicates the total outcrop area, light grey, the area that might contain the smallest fracture; and dark gray, the area in which the largest fracture could be observed. The difference between these areas causes a bias in the observations. 52

Figure 3-5 : Illustration of the projection of the trace coordinates to a mean plane along the fracture direction. We consider that fractures are modeled by flat ellipse oriented using the measured strike and dip..... 53

Figure 3-6 : non-projected (a) and projected (b) trace of an outcrop in Laxemar (ASM000208). We observe a straightening up of this curved trace of fractures, which having a small dip value.....	54
Figure 3-7 : Fracture trace size distribution for the original fracture network (black, squares) and the projected fracture network (red, square). We do not notice any crucial influence of the projection in the power-law parameters.....	54
Figure 3-8 : Example of non-Poisson structures in the natural fractures: a) “dashed-line, b) “en-echelon and c) “layered” fractures.	55
Figure 3-9 : Contour of the shear stress relaxed normalized with the fracture size.	56
Figure 3-10 : Illustration of non-Poisson structures and description of linkage procedures	56
Figure 3-11 : Illustration of linkage on the fracture traces of a Laxemar outcrop (ASM100235): the whole network (left) and detail (right) where original traces are displayed in black and linked traces in red.	56
Figure 3-12 : Investigated field: the regional model (thick red line), subareas of Simpevarp (on the right, dashed grey line) and Laxemar (on the right, dashed grey line) and the local model Simpevarp V1.2 (dashed black and red line). This picture is extracted from [Darcel et al., 2004].....	58
Figure 3-13 : Variation of orientation between each segment in a fracture trace and the mean fracture trace. The distribution mean α and the square-difference $\sigma\alpha$ are taken into account to compute the orientation criterion ω_{max} used in the linkage procedure. We obtain an orientation criterion between 13 and 25 degrees.	58
Figure 3-14 : Evolution of the number of structure identified as segmented with the distance criterion. The Poisson curves (grey) are calculated on a synthetic random fracture network generated in the corresponding outcrop contour with similar size and orientation distributions. Other curves are obtained using ASM000208 fracture-trace map after a linkage procedure with $\omega_{max} = 20$ degrees and $\tau\Delta\sigma I_{max}$ given in parentheses. We notice that after a linkage with $\tau\Delta\sigma I_{max} = 0.4$ (blue) the corrected outcrop can be seen as Poisson.....	59
Figure 3-15 : Each outcrop is associated with the distance criterion $\tau\Delta\sigma I_{max}$ corresponding to a number of segmented structures similar to a Poisson fracture network with similar orientation and size distributions.....	59
Figure 3-16 : Distribution of the fracture-trace size of a Laxemar outcrop (ASM000208) before (black, squares) and after (blue, circle) the correction procedure. The intermediary distributions with only projection (wine, empty squares) and only area normalization (green, empty circles) are also given for indication. The distribution on corrected data presents less deviation from the expected straight-line and has an exponent lower than the original distribution.	60
Figure 3-17 : Density term (top) and power-law exponent (bottom) calculated on all available corrected outcrops. We observe that the density term is not mainly influenced by the correction procedure while the exponent is significantly decreased. The final density term is found to be in average 2.0 while we have clearly two models for the exponent. ASM000205 and ASM000206 have an exponent around -3.0 while all other outcrops have an exponent around -2.3.....	61
Figure 3-18 : Map of all available lineaments in the Laxemar subarea. There are lineaments smaller than 1km (blue), lineaments larger than 1km (grey) and identified deformation zones (red and green).	62
Figure 3-19 : All normalized density distributions available are plot on this multi-scale diagram. The two average model for outcrops are plot, with exponent equal to 2.3 (wine) and 3.0 (blue). We observe that the second model fits also at the lineament scale and the Swedish lineaments, while the lineaments larger than 1km have a distribution close to the first outcrop model.	63

figure 3-20 : Exposants de la loi de puissance dérivés des traces de fractures des affleurements de Forsmark. D'après [Darcel et al., 2009].	64
figure 3-21 : Distribution des longueurs de fracture à toutes les échelles disponibles pour le site de Forsmark. D'après [Darcel et al., 2009].	65
Figure 3-22 : Discrétisation de l'espace (θ, φ) en 13 zones de surface égale.	66
figure 3-23: Analyse de la variabilité du domaine local de fracturation FFM01 du forage KFM01C (Forsmark). On compare l'évolution de $\sigma_{var}(s)/\alpha_{3D}$ en fonction de s pour les données du site (bleu) à celle pour des données tirées dans une distribution de Poisson (gris). D'après [Darcel et al., 2009].	67
figure 3-24 : Comparaison de σ_{var} et de $\sigma_{\alpha}(s)$ pour tous les domaines locaux de fracturation de Forsmark (étoile) et Laxemar (carré). On calcule $\alpha_{3D}(\theta, \varphi)$ sur la discrétisation en 13 zones (jaune et vert) ou sur l'espace complet pris comme une zone unique (violet et bleu).	68
Figure 3-25 : Sorting process algorithm.	70
Figure 3-26 : Illustration of the resulting classification on single points. With a $d_{max}=4$ the points are distributed into 4 classes represented with different colors. Circles represent class averages and the associated error bar is equal to d_{max} . We observe that the classification is consistent but the limit between two classes might be slightly changed without exceeding d_{max} .	72
Figure 3-27 : Coherence indexes for the classes illustrated on Figure 3-26. All coherence indexes are high, meaning that the resulting classes are consistent. Classes 1 and 2 have smaller coherence indexes because of a possible overlap between them.	72
Figure 3-28 : Discretization of the stereonet into 13 pole zones. The density distribution $\alpha_{3D}\theta, \varphi$ is computed for each zone.	73
Figure 3-29 : Available data on Laxemar (Sweden). Top is the localization of cored borehole and down are the resulting density distributions. Each borehole is split into rock units and we calculate one density distribution per rock unit.	76
Figure 3-30 : Illustration of the stopping criteria d_{max} . With $d_{max}=1.0$ the two distribution on the top will belong to the same class because all $\alpha_{3D}\theta, \varphi$ values are similar. The two distributions on the bottom-left will belong to different classes because $\alpha_{3D}\theta, \varphi$ values are different for the central zone (1) and the middle zones (12-13). On the bottom-right, the two distributions have different shape and the classification procedure separate them into different groups.	76
Figure 3-31 : evolution of the larger heterogeneity in a class as a function of the number of classes. A value $d_{max} = 1.0$ means that in a class we have only identical distributions accounting for uncertainty. A value $d_{max} = 0.8$ leads in our case to seven classes and seems relevant to find similar fracture domains.	77
Figure 3-32 : Resulting classification of the 51 available distributions. Using a stopping criteria $d_{max}=0.8$ leads to seven different classes (1-7). Class means are summarized in (8).	78
Figure 3-33 : I_c and I_{Cw} of all classes resulting from a classification with $d_{max}=0.8$. I_c values indicates that there is overlapping between classes. I_{Cw} indicates that the classes with the most elements (class 4 and class 7) the most consistent while classes with a few numbers of distributions (class 2 and 3) are less significant in terms of fracture domain.	79
Figure 3-34 Contoured plot of the probability for a distribution (y-axis) to belong to each class defined (x-axis). Distributions from 1 to 51 are sorted according to their class of reference, from 1 to 7 (grey dashed-lines indicate the limit between classes — i.e. distribution rows 1 through 7 refer to Class 1 and so on).	80

Figure 3-35 : Index of coherence IC by class (left) and weighted index (right). IC highlights the possible overlapping between classes (e.g. the high IC between classes 2 and 3) while ICw refers to the total number of distributions remaining in each class, accounting for overlapping.	80
Figure 3-36 : Classification based on geological assumptions. Only six classes are defined. In some fracture domains, for example FSM_W, the orientations are clearly heterogeneous.	82
Figure 3-37: Coherence indexes for the classification based on geological settings.....	82
Figure 3-38 : Representation of all density distribution in a 2D space using multidimensional scaling. (left) points are colored according to the statistical classification; (right) points are colored according to the geological classification.....	83
figure 3-39. Index de cohérence pour la classification finale des distributions du site de Forsmark (dmax = 0.9, 11 classes).....	84
figure 3-40 : index de cohérence IC (a) et ICw (b) des domaines de fracturation définis à partir de critères géologiques pour le site de Forsmark.	85
figure 4-1 : Valeurs moyenne du logarithme du flux (a), écarts-type (b) et longueurs de corrélation (c) calculés pour les configurations tests de la figure 4-6. Chaque valeur est prise comme la moyenne de 500 simulations.	103
figure 4-2 : valeurs de $S_2(\varphi)$ pour les différentes configurations présentées à la figure 4-6. Les valeurs présentées correspondent à la moyenne calculée à partir de 500 simulations.....	105
Figure 5-1 : (a) Synthetic fracture network. (b) Flows computed with a constant sub-horizontal hydraulic gradient of direction given by the arrow. All fractures have the same transmissivity. The grey color is proportional to the logarithm of flow normalized by the total flow entering the domain. (c) Simplified fracture structure carrying 70% of the flow. (d) Sketch of the main flow channels.	111
Figure 5-2 : Principle of the identification strategy applied to the flow structure of Figure 5-1d. Channels are identified by decreasing order of importance. At each step, an objective function with a restricted number of parameters is defined and optimized. Steps 1 to 3 illustrate the identification of the first channel with a shape of increasing complexity. Steps 4 to 5 illustrate the addition and refinement of a secondary channel. Step 6 shows how channel connectivity can increase with a channel connecting two previously identified channels. Steps 7 to 9 complete the identification strategy. The background color is correlated to the background permeability. It is expected to decrease with the number of structures identified.	112
Figure 5-3 : Flowchart of the flow channels identification strategy consisting of four nested loops underlined by different grey colors. The four loops are the calibration loop, the single channel identification loop, the hierarchical channels identification loop and the post-processing analysis loop. P is the last parameterization considered as valid, P' is the current parameterization, n_s is the simulation number, n_{max} is the number of simulations to perform and stopping criteria are given by (4) and (5).	113
Figure 5-4 : objective function versus model parameters on a simple case made up of two channels. (a) Configuration used to compute the objective function. Red dots stand for observation locations. Objective function values (6) with $\mu_1=0$ and $\mu_2=1$ as a function of (b) the first channel parameter, (c) the second channel parameters and (d) the matrix permeability. All parameters are varied around their optimal value, all others being left fixed to their optimal value. 1 st and 2 nd connected structures refer to the borders or channels connected by the channel (between 0 and 4). Position of 1 st and 2 nd extremity refers to the coordinate of the intersection between the channel and the border or channel (between 0 and 10). The log-transmissivity of channels varies from -3 to 2 and the log-transmissivity of the background matrix varies from -5 to 0.....	118

Figure 5-5 : Sketch of the post-processing clustering algorithm identifying the prevalent characteristics (right) within a set of individual solutions (left). Colors on the individual solutions refer to the channels identified as prevalent characteristics.	119
Figure 5-6 : Flow rates ϕ of three tested configurations, ordered by increasing complexity. The simplest configuration (a) is composed of two straight-line channels. The intermediary configuration (b) is composed of two channels that are more complex. The complex configuration (c) has several second-order well-connected structures.	121
Figure 5-7 : (a) Objective function against iteration number for the simple configuration of Figure 5-6a. Evolution of the parameter values of (b) the first identified channel, (c) the second identified channel and (d) the background matrix against the number of iterations.....	123
Figure 5-8 : Distribution of the square difference between reference and model results on the whole surface Δh^2 (12) derived from the 50 individual solutions for the simple configuration of Figure 5-6a.....	124
Figure 5-9: illustrations of the results of the post-processing clustering algorithm for the simple configuration of Figure 5-6a with objective functions made up of the two first terms with σ_{ih} equal to 1 (a) and to the inverse of head deviation (b). The last map (c) is obtained with the full objective function made up of the three terms.....	125
Figure 5-10 : Interpolated maps from the 25 data points (left) for the head deviation and (right) for the distance to the nearest channel for the simple configuration of Figure 5-6a.....	125
Figure 5-11: Interpolated maps from the 25 data points for the head deviation (left) and for the distance to the nearest channel (right) for the simple configuration of Figure 5-6b.	126
Figure 5-12: Distribution of the least-square difference between head data and model results on the whole surface Δh^2 (12) derived from the 50 individual solutions for the intermediary configuration (Figure 5-6b).	126
Figure 5-13: Illustrations of the results of the post-processing clustering algorithm with objective functions made up of the two first terms with σ_{ih} equal to 1 (a) and to the inverse of head deviation (b). The last map (c) is obtained for the full objective function composed of the three terms.	127
Figure 5-14: Four individual solutions of the inversion method applied to the intermediary complex configuration of Figure 5-6b. The individual solutions identify the diagonal main flow structure represented in red. Three of them identify the secondary flow structures in green. Embranchments are too small structures to be resolved. The channeled structure is recalled in the background.....	127
Figure 5-15: illustrations of the results of the post-processing synthetic algorithm for the configuration of higher complexity of Figure 5-6c with objective functions composed of the two first terms with σ_{ih} (a) equal to 1 and (b) equal to inverse of the head deviation. The last map (c) is obtained for the full objective function composed of the three terms.....	128
Figure 5-16: Interpolated maps from the 25 data points for the head deviation (left) and for the distance to the nearest channel (right) for the simple configuration of Figure 5-6b.	128
Figure 5-17 : illustrations of the results of the post-processing synthetic algorithm for the most complex configuration (Figure 5-6c) when using only distance information in the identification strategy.....	129
Figure 5-18 : Four individual solutions in red and their associated objective function value obtained for the identification strategy with the full objective function applied to the configuration of Figure 5-6c. The reference configuration is recalled with gray colors in the background.....	129

Figure 5-19: Distribution of the square difference between head data and model results on the whole surface Δh^2 (12) derived from the 50 individual solutions for the most configuration of highest complexity of Figure 5-6c.	130
Figure 5-20 : First 12 configurations (C1 to C12) used to test the identification strategy. Left and right columns show the reference configuration and the modeling result given by the postprocessing analysis. Green dots stand for well locations.	131
Figure 5-21 : Last 8 configurations (C13 to C20) used to test the identification strategy. Left and right columns show the reference configuration and the modeling result given by the postprocessing analysis. Green dots stand for well locations.	132
Figure 5-22 : Quadratic mismatch sum RE between reference data and modeling results as a function of the product of the channeling indicators D_{ic} and D_{cc} normalized by the square of the system size L^2 . Black squares identify configurations for which RE increases with more complex configurations corresponding to smaller $D_{ic} \cdot D_{cc} / L^2$ values. White squares stand for four complex configurations resulting in small errors RE.	132
Figure 5-23: Visual rating of the final post-processing solution versus the visual channel number	134
Figure A-1 : Comparison of flow structures in highly (left) and moderate channeled fracture network (right). TOP: fracture networks were generated with a power-law length-distribution (exponent = 3.0) and uniform random orientations and positions in a system size of 25. Moderate and highly channeled networks were obtained for a density at the percolation threshold (left) and 50% above the percolation threshold (right). Note that the generation seed is the same in the two cases. Boundary conditions are permeameter-like, i.e. they are fixed on two opposite sides of the domain and the perpendicular domain sides are impervious. The middle row represents head from 0 to 10. The bottom row represents the flows. We notice that flow channels are based on existing active transmissivity structures and not strictly depending on the boundary conditions.	141
Figure A-2 : Main synthetic test case. (Right) The conductivity structure mimics the presence of transmissive paths carrying flow. Two main channels are thus created with connected structures having a transmissivity $T=10^{1.4}$ ($n^{\circ}1$ and $n^{\circ}2$), intermediary channels with a transmissivity $T=10^{1.7}$ ($n^{\circ}3$, 4, 7 and $n^{\circ}8$) and minor channels with $T=10^{1.9}$ ($n^{\circ}5$ and $n^{\circ}6$). (Left) corresponding flow in gradient-like boundary conditions. We observe that flow is located in the two main channels.	143
Figure A-3 : Discrepancy maps of the considered test case with gradient-like boundary conditions with different orientations : 0° (left), $+10^{\circ}$ (middle) and -20° (right). The underlying transmissivity structure is superimposed over the discrepancy map to visualize the influence of channels position.	144
Figure A-4 : Discrepancy maps of the considered test case with gradient-like boundary conditions and adding a pumping at different location: (a) in the middle of the field, (b) in the upper part and (c) in the topper part. The underlying transmissivity structure is superimposed over the discrepancy map to visualize the influence of channels position. The red points indicate the position of the pumping well.	145
Figure A-5 : Discrepancy maps of the considered test case with gradient-like boundary applied on a system five times larger than the investigated one : (a) no rotation, no pumping, (b) rotation of $+10^{\circ}$ and (c) pumping at the system center. The underlying transmissivity structure is superimposed over the discrepancy map to visualize the influence of channels position.	146
Figure A-6 : Discrepancy maps with gradient-like boundary rotated at $+10^{\circ}$. It represents the absolute difference between heads calculated with the complete test case and partial solutions	

having from 0 to 8 channels. The partial transmissivity structure is superimposed over the discrepancy map to visualize the influence of channels position. 148

Figure A-7 : models obtained after the resolution of the inverse problem on 25 head data under different hydraulic conditions. Top, channels are colored according to their statistical weight among the solutions; bottom, channels are colored according to their relative transmissivity $\log(T_{\text{channel}}/T_{\text{matrix}})$ 150

Figure A-8 : Final solution resulting from the merge of all single solutions obtained under different boundary conditions. (a) final channeling structure colored by their statistical weight, (b) channels colored by their ratio $\log(T_{\text{channel}}/T_{\text{matrix}})$, (c) distribution of objective function value calculated with all transmissivity structures with gradient-like boundary conditions. 151

Figure A-9 : Mean difference between head data in the channeled case and in the homogeneous case plotted as a function of the distance between the head position and the nearest channel. 153

figure 6-1 : Schématisation des paramètres à prendre en compte lorsque l'on cherche à définir un modèle pertinent. En jaune est représenté l'espace des possibles, c'est-à-dire l'ensemble des possibilités de représentation d'un phénomène. Plus le phénomène est complexe, plus cet espace est vaste. En vert est représentée la sous-partie de l'espace qui peut correspondre aux observations du milieu. Plus il y a d'information disponible, plus cette partie est localisée. La surface hachurée représente l'ensemble des solutions que la paramétrisation choisie permet de représenter. Le modèle défini appartiendra à une intersection entre la zone hachurée et une des zones vertes. Un modèle pertinent doit contenir la réalité (point rouge) et couvrir une surface la plus restreinte possible autour de ce point. 157

figure 6-2 : Schématisation des objectifs recherchés lors de la définition du modèle de fracturation des sites suédois de Forsmark et Laxemar décrite à la partie 3. Par rapport à la figure 6-1, on cherche à contraindre le modèle afin de réduire l'univers des modélisables et à mieux définir, en intégrant la variabilité et l'incertitude sur les données, à mieux définir l'univers des observations. 158

figure 6-3 : Schéma du rôle des indicateurs de chenalisation dans la détermination d'un modèle pertinent. A partir des caractéristiques de la chenalisation du milieu étudié (distance entre chenaux, longueurs caractéristique des chenaux), il est possible d'exclure un grand nombre de modèles dont les caractéristiques sont différentes afin de ne garder que les modèles conformes. 160

figure 6-4 : Schématisation du rôle de la paramétrisation dans la détermination d'un modèle pertinent. Si la paramétrisation intègre efficacement les caractéristiques principales du milieu étudié (zone avec les hachures diagonales), alors la résolution du problème inverse, qui vise à obtenir un modèle proche de la réalité, est facilitée. Par contre, si les caractéristiques utilisées ne sont pas pertinentes (zones avec les hachures carrées), le risque est d'obtenir un modèle très différent de la réalité tout en étant capable de reproduire fidèlement les observations. 161

figure 6-5 : Schématisation du rôle de la méthode d'inversion. Même si la paramétrisation est bien choisie et que l'on a suffisamment de données pour paramétrer le système, il existe toujours une incertitude liée à la méthode elle-même dont il faut tenir compte. 163

figure 6-6 : illustration des méthodes de caractérisation des propriétés hydrauliques des milieux hétérogènes développée pendant le doctorat. 169

tableau 1-1 : Synthèse des différentes méthodes de résolution du problème inverse décrite au paragraphe 2. 31

Table 1 : Parameter list and possible values for the whole identification strategy. 115

Table 2 : Residual errors RE defined by (1), channeling indicators (D_{ic} and D_{cc}) associated to the tested configurations (Figure 20 and Figure 21), visual indicators of the channel number and solution quality. D_{ic} and D_{cc} are in the interval [0;10]. The number of channel and the quality of the solution range between 0 and 5..... 133

Introduction

Une bonne gestion des ressources en eaux fait partie des défis importants des années à venir. L'accès à une eau de bonne qualité tant pour la consommation que pour l'agriculture est une condition essentielle au développement d'un pays. La découverte de nouveaux points de captage, la compréhension des circulations d'eau ainsi que la préservation des ressources est donc au cœur des préoccupations actuelles et l'enjeu majeur de demain. Les ressources en eau potable sont majoritairement composées des eaux de surfaces comme les glaciers, les lacs ou rivières (68,7%) et des eaux souterraines (30,1%) [Gleick, 1996]. Si les premières sont facilement accessibles et ne nécessitent que peu d'infrastructures pour être exploitées, ce sont les plus sensibles aux pressions démographiques et environnementales et leur exploitation est de plus en plus compromise dans de nombreux pays du monde [UNESCO, 2009]. Il est donc naturel de se tourner vers les eaux souterraines, moins exposées et globalement de meilleure qualité. Cependant, si leur position naturelle, de quelques mètres à quelques centaines de mètres sous la surface du sol, les préserve des pollutions chimiques et agricoles, elle rend beaucoup plus compliquée la compréhension de leur fonctionnement. Il n'existe en effet à l'heure actuelle aucun moyen de « voir » l'eau souterraine directement, les techniques d'imageries géophysiques étant toujours trop imprécises pour distinguer plus que quelques événements majeurs ou certaines propriétés moyennes à grande échelle. Il ne reste donc que les observations directes et indirectes du sous-sol et des écoulements, soit en surface, soit au travers de forages.

Les circulations des eaux souterraines résultent des interactions entre le fluide et la roche dont l'aquifère est constitué. Le débit q à l'intérieur de l'aquifère est ainsi lié à la perméabilité k de la roche et à la charge hydraulique h par la loi de Darcy [Darcy, 1856] en tout point (x,y,z) de l'espace:

$$q(x, y, z) = k(x, y, z) \nabla_{x,y,z} h \quad (\text{I-1})$$

Le paramètre k , autrement appelé perméabilité de Darcy, représente la capacité de la roche à laisser s'écouler un fluide. Connu en tout point de l'aquifère, il permet de calculer les flux et donc de modéliser les écoulements à l'intérieur du milieu. En appliquant la loi de conservation de la masse on tombe sur une relation classique en l'hydrogéologie reliant k à h [Bear, 1973]:

$$\nabla \cdot (k \nabla h) = S \frac{\partial h}{\partial t} + Q \quad (\text{I-2})$$

où S est l'emmagasinement spécifique de la roche et Q un terme source. Connaître k , ainsi que S et Q si nécessaire, permet de définir complètement les écoulements d'un aquifère. Une grande partie des études scientifiques porte donc sur la caractérisation de la perméabilité via l'observation du milieu naturel, des expériences en laboratoire et des modèles numériques. Il apparaît que déterminer les valeurs de k est loin d'être une chose facile, la principale difficulté provenant de l'hétérogénéité des propriétés du milieu, cette hétérogénéité pouvant être très forte et ce, à toutes les échelles d'observation. Un modèle numérique est donc généralement utilisé pour lier propriétés du milieu et écoulements. L'objectif est de mieux appréhender la nature du flux, de prédire leur évolution, d'analyser les effets d'un changement dans les conditions hydrauliques comme l'ajout d'un pompage ou d'anticiper les effets d'une pollution de l'aquifère. Ces modèles s'appuient sur une représentation de la perméabilité et des autres

paramètres de l'aquifère (forme, conditions aux limites, etc.) sur laquelle est appliquée une discrétisation des équations d'écoulement, le plus souvent sous forme matricielle. Cette représentation de la perméabilité, ou paramétrisation, doit être compatible avec les propriétés du milieu naturel sous peine d'obtenir des résultats incohérents. En fonction des caractéristiques de la perméabilité, on est ainsi amené à séparer les aquifères en deux catégories, les milieux poreux et les milieux fracturés, dont les propriétés, et donc les modèles associés, peuvent être très différents.

La problématique de ce travail repose sur la caractérisation des écoulements dans les milieux où le flux est fortement chenalisés. Le terme « chenalisé » renvoie ici au phénomène de concentration des écoulements dans une petite partie du milieu et sera défini ultérieurement. La question posée ici est : comment peut-on identifier et décrire les structures responsables des principaux écoulements ? Nous partons du principe que répondre à cette question implique tout d'abord de décrire rigoureusement les propriétés physiques du milieu naturel à partir des observations de ce milieu. Cette description doit ensuite être intégrée à un modèle numérique afin d'analyser les effets de la structure de perméabilité sur les écoulements et plus précisément sur leur chenalisation éventuelle. Finalement, nous proposons une approche de type problème inverse où les propriétés hydrauliques du milieu, et surtout les valeurs de charges, sont utilisées pour caractériser la perméabilité du milieu. Comme les milieux étudiés sont particulièrement complexes, l'idée ici est d'avoir une vision la plus complète possible des propriétés physiques et hydrauliques de la chenalisation. La figure I-1 donne une vision schématique de ces différentes questions.

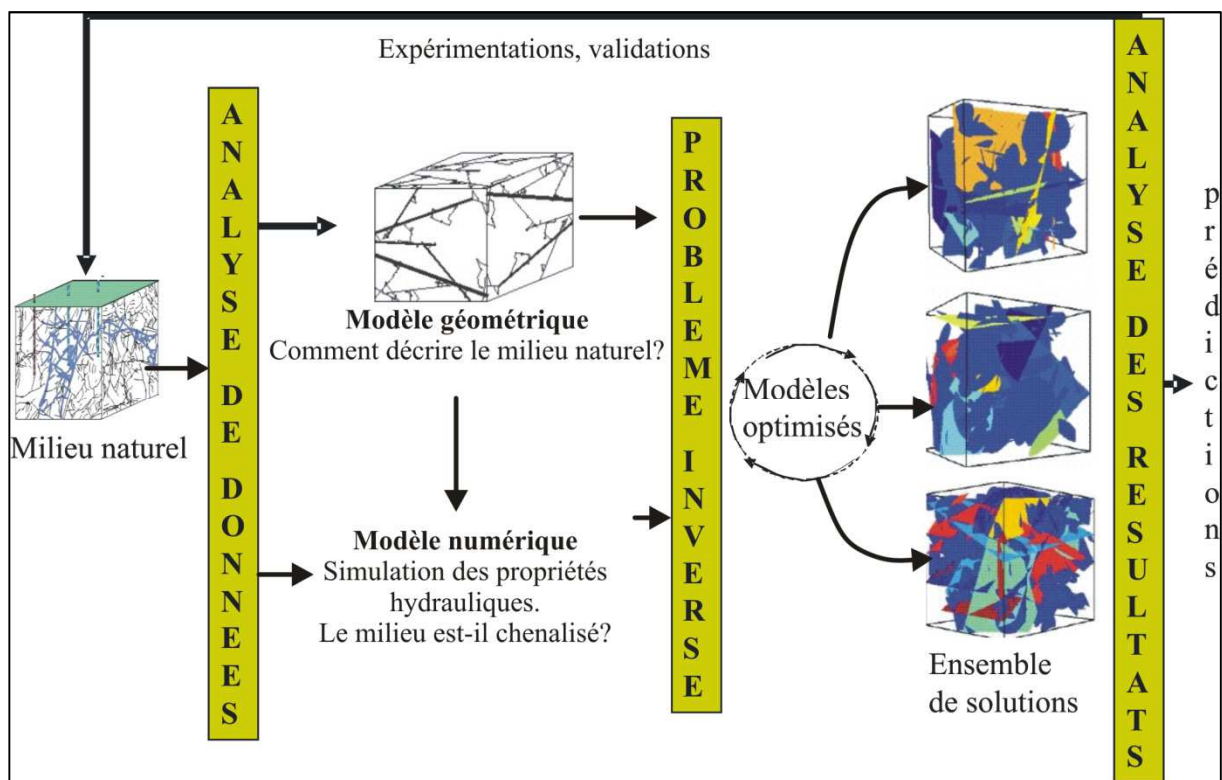


figure I-1 : Problématique du travail de thèse : comment caractériser les milieux fortement hétérogènes où les écoulements sont très chenalisés ? Par une observation du milieu naturel permettant de modéliser sa structure physique, une analyse des données hydrauliques pour caractériser les propriétés d'écoulements et la définition d'un modèle de site intégrant les propriétés physiques et hydrauliques du milieu par une méthode inverse.

Ce travail s'articulera autour de six parties. Pour commencer, une revue des avancées scientifiques sur la caractérisation de la perméabilité dans les milieux poreux et fracturés donnera au lecteur un aperçu des techniques et des modèles généralement utilisés pour décrire les propriétés physiques des aquifères et les lier aux propriétés hydrauliques. Dans une seconde partie, nous décrirons les méthodes numériques développées pour modéliser les propriétés physiques d'un milieu poreux ou fracturé et simuler les écoulements et le transport de particules dans ces milieux. Nous nous pencherons dans une troisième partie sur la caractérisation d'un milieu fracturé en s'appuyant sur les travaux réalisés pour l'agence suédoise des déchets radioactifs *Svensk Kärnbränslehantering AB* (SKB). Cette étude visait à définir un modèle de la fracturation des sites de Forsmark et Oskarshamn en Suède pour l'installation d'un site de stockage des déchets nucléaires. Une quatrième partie détaillera le calcul d'indicateurs mesurant le degré de chenalisation de milieux synthétiques. Ces indicateurs permettent de lier les caractéristiques de la perméabilité à une caractéristique hydraulique, qui est que les écoulements suivent des chemins préférentiels contenus dans une petite partie du milieu considéré. Une cinquième partie sera consacrée à l'identification de ces chemins préférentiels à partir de données hydrauliques via la résolution d'un problème inverse. Il s'agit d'une démarche opposée aux précédentes, puisqu'au lieu de déduire les propriétés des écoulements à partir des propriétés de la perméabilité, on utilise les caractéristiques hydrauliques, liées au flux, pour déterminer les structures physiques sous-jacentes. Finalement, une sixième et dernière partie synthétisera les résultats obtenus et discutera des manières de caractériser les écoulements dans les milieux fortement chenalisés.

Partie 1 État de l'art sur les modèles de perméabilité intégrant la chenalisation des écoulements

La caractérisation des écoulements dans les milieux fortement hétérogènes s'appuie sur une modélisation des propriétés hydrauliques du milieu. Il existe deux approches permettant de définir et de calibrer le modèle. La première consiste à décrire les spécificités de la structure physique du milieu géologique afin d'en déduire le modèle d'écoulement. Il s'agit classiquement de résoudre le problème direct où, dans l'équation (I-1), la perméabilité k est définie à partir des observations du milieu naturel, ce qui permet de déterminer le flux q et la charge hydraulique h . Cependant, cette méthode nécessite une très bonne connaissance de k ce qui, compte tenu du volume et de la qualité de l'information généralement disponible, est une tâche difficile, d'autant plus que les résultats obtenus en terme de propriétés hydrauliques dépendent entièrement du modèle de perméabilité. La seconde approche s'appuie sur la résolution du problème inverse, c'est-à-dire que k est cette fois déterminé à partir de mesures des propriétés hydrauliques, q et h notamment. Comme les milieux auxquels nous nous intéressons ont à la fois une structure et des propriétés hydrauliques complexes, de part l'hétérogénéité de la perméabilité et la chenalisation des écoulements, nous avons abordé les deux types d'approches afin d'extraire le maximum d'information du milieu naturel pour définir des modèles efficaces. Cette partie présente donc les méthodes de modélisation pour les milieux poreux et fracturés en insistant sur les particularités des milieux chenalisés, ainsi que les méthodologies pour définir la perméabilité via le problème direct et surtout le problème inverse.

Définir la chenalisation n'est pas une chose triviale, et la notion de chenal dépend vraisemblablement du phénomène considéré (flux, transport, etc.). Dans les travaux présentés ici, nous définirons un chenal à partir de ses propriétés d'écoulement. On considère un chenal comme étant une sous-partie du milieu considéré, fortement localisée, connectée aux limites du système et porteuse d'une portion significative des écoulements. Cette définition sera affinée par la suite grâce aux différents résultats obtenus, notamment en Partie 4 et Partie 5.

A. Modèles de perméabilité

Lorsque l'on essaie de décrire la perméabilité d'un milieu, la principale difficulté réside dans la prise en compte de son hétérogénéité. Une méthode couramment utilisée est de simplifier cette hétérogénéité en définissant une perméabilité équivalente homogène correspondant à une valeur moyenne du milieu. D'autres méthodes, au contraire, tentent de décrire les variations spatiales des propriétés du milieu à partir d'observations géologiques et de mesures locales. On se référera à [de Marsily et al., 2005] pour une revue des méthodes existantes. La prise en compte des propriétés de la perméabilité dépend du milieu considéré et il existe des méthodes adaptées aux milieux poreux et des méthodes adaptées aux milieux fracturés. Ces méthodes reposent sur une paramétrisation du milieu naturel adaptées à ses caractéristiques.

1. Décrire la perméabilité dans les milieux poreux

Dans les milieux poreux l'eau s'écoule à travers les pores de la roche. Ceci concerne principalement les aquifères sédimentaires, largement utilisés pour l'approvisionnement en eau potable. Dans ce type de milieu, la taille des pores, leur distribution spatiale et leur connectivité définissent la perméabilité de la roche. Dans un premier temps des études ont été menées pour relier la taille des pores et la perméabilité. Il a par exemple été démontré que la

perméabilité intrinsèque est directement reliée à la taille des grains constituant le milieu [Fetter, 2001]. Cependant, la taille des grains comme celle des pores est fortement hétérogène quel que soit le milieu étudié, ce qui induit une forte hétérogénéité de la perméabilité à petite échelle, avec des valeurs pour un même type de roche variant sur plusieurs ordres de grandeurs. Un aquifère pouvant par ailleurs être composé de plusieurs types de roche, l'hétérogénéité est multi-échelle, ce qui a amené *Petitjohn* et ses co-auteurs [Petitjohn et al., 1987] à la classer en plusieurs catégories allant de l'échelle du pore à celle des formations géologiques.

a) *Modèles intégrant l'hétérogénéité*

Cette question est devenue incontournable à partir des années 60 avec le développement des premiers modèles numériques, dont une revue est disponible dans [Prickett, 1975]. La perméabilité est modélisée par une grille en deux ou trois dimensions, le plus souvent régulière, dont chaque cellule contient la perméabilité moyenne locale correspondant à la surface de la maille. L'hétérogénéité se résume alors à l'utilisation de valeurs de perméabilité différentes pour chaque cellule de la grille, ce qui revient à ne prendre en compte que l'échelle d'hétérogénéité la plus large. Bien que ces modèles soient capables de simuler les propriétés hydrauliques du milieu (charge, flux), les résultats restaient peu fiables en termes de prédiction, dans la mesure où ils négligeaient l'hétérogénéité aux échelles inférieures. Pour la prendre en compte, il est nécessaire d'augmenter le nombre de cellules de la grille servant à décrire la perméabilité du milieu. Or, la quantité de données disponibles est très insuffisante pour affecter une perméabilité représentative du milieu à chacune de ces cellules.

C'est pourquoi de nouveaux outils sont apparus, notamment la géostatistique [de Marsily et al., 1984] qui considère les caractéristiques d'un aquifère comme des variables aléatoires et s'appuie sur une description de la distribution spatiale de la perméabilité. *Gelhar* [Gelhar, 1993], entre autre, décrit les principes de la géostatistique appliquée à l'hydrogéologie. Par exemple, le krigeage [Krige, 1951] permet d'interpoler de manière non biaisée les valeurs de perméabilité des cellules pour lesquelles aucune donnée n'est disponible à partir des perméabilités mesurées localement. Le principe consiste à respecter le variogramme des données existantes tout en minimisant l'erreur liée à l'estimation [Cressie, 1993]. Un autre type de méthode repose sur la génération aléatoire de champs de perméabilité respectant les lois géostatistiques décrivant les relations géométriques observées sur le terrain. Les champs ainsi générés sont ensuite utilisés dans des modèles simulant le flux et le transport et ceux respectant le mieux les propriétés hydrauliques du milieu sont conservés [Koltermann and Gorelick, 1996]. Un des modèles de génération les plus utilisés est le modèle multi-Gaussien, qui produit des champs de perméabilité définis par leurs moyennes, variances, covariances et variogrammes. La popularité de ce type de modèle vient de sa simplicité, puisqu'il est complètement caractérisé par sa moyenne et sa covariance, et surtout parce qu'il semble correspondre aux propriétés de certains milieux naturels comme ceux modélisés par [Hoeksema and Kitanidis, 1985].

La figure 1-1 donne un aperçu de ces différents modèles en montrant, pour un champ de perméabilité donné, les modélisations équivalentes par zonation, par krigeage et par un champ multi-Gaussien aux propriétés statistiques équivalentes. Les écoulements correspondant aux modèles sont également illustrés, leurs différences permettant d'illustrer les écarts pouvant survenir entre les écoulements simulés avec de tels modèles et les écoulements réels.

b) Vers la prise en compte de la chenalisation

Les modèles précédant donnent des résultats satisfaisant lorsque le milieu naturel n'est pas trop hétérogène et que les écoulements ne sont pas chenalisés. Cependant, dès que l'on veut rentrer dans les détails, comme modéliser le transport d'un polluant dans le milieu par exemple, ou dès que l'hétérogénéité devient trop forte, il faut prendre en compte de façon plus fine les caractéristiques du milieu. Ces modèles sous-estiment fortement la présence de chemins préférentiels d'écoulement. Dans la mesure où les écoulements ont tendance à éviter les zones de faible perméabilité pour se concentrer dans les zones de fortes perméabilité, la chenalisation est due à la concentration des flux le long d'un petit nombre de chemins d'écoulement préférentiel où la résistance est plus faible [Tsang and Tsang, 1989]. Si l'hétérogénéité favorise ce phénomène, des écoulements préférentiels peuvent se former dans des milieux plus simples. Ainsi dans les expériences de Silliman [Silliman, 1995; Silliman et al., 1998], un milieu simple constitué d'un mélange de deux sables de perméabilité différentes est utilisé en laboratoire afin d'étudier les écoulements et les conséquences sur le transport de particules. Elles mettent en évidence la présence de chemins préférentiels d'écoulement même dans ce type de milieu ainsi que le rôle-clé joué par ce type de structure sur les propriétés hydrauliques. Ce phénomène, loin d'être minoritaire, a été détecté dans un grand nombre de milieux dès que l'hétérogénéité est importante. En se basant sur des modèles numériques en trois dimension, [Moreno and Tsang, 1994] ont montré que les formulations classiques de perméabilité équivalente étaient irrecevables dès que l'écart-type des perméabilités dépassait 2.0 et ce, à cause de la chenalisation des écoulements.

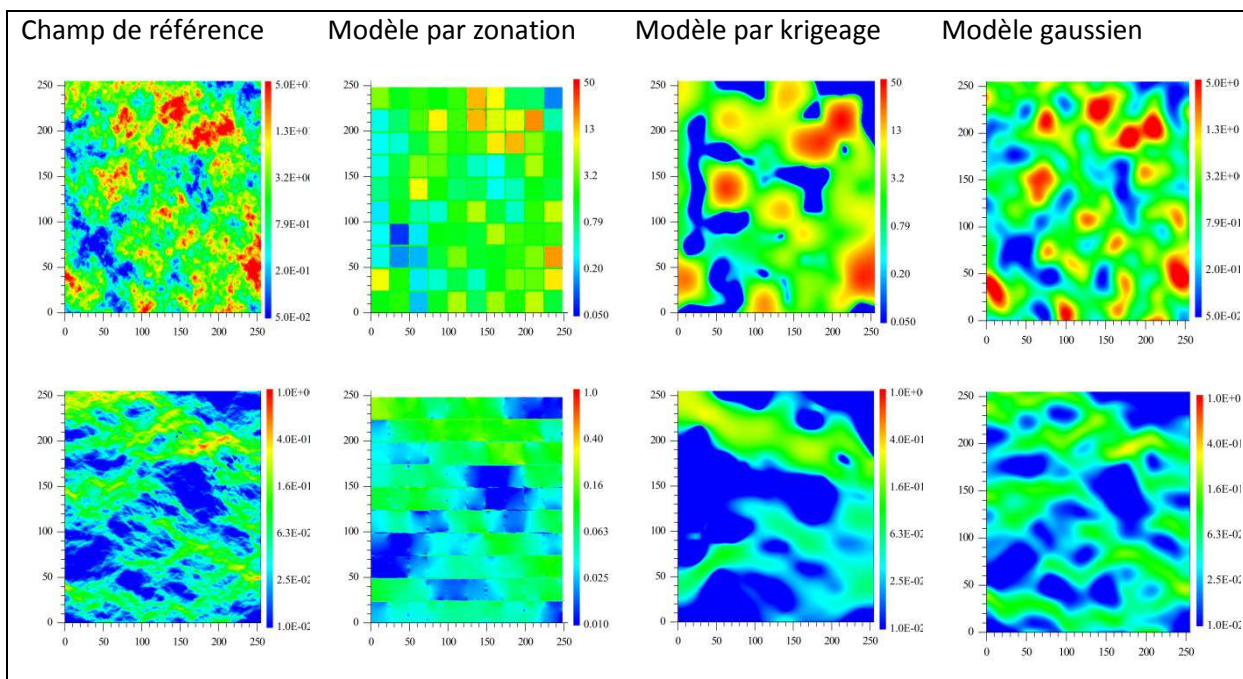


figure 1-1 : Modèle de champs de perméabilité. En haut se trouve champ de perméabilité et en bas le flux correspondant. Les modèles par zonation et par krigeage sont obtenus à partir de 100 données réparties de manière régulière dans le milieu.

Des descriptions plus détaillées de la perméabilité on donc commencé à apparaître pour intégrer spécifiquement la présence de chemins préférentiels d'écoulement. Une première catégorie de modèles cherche, par une réorganisation du champ de perméabilité à recréer les conditions de la chenalisation. C'est par exemple le cas dans l'étude de [Zinn and Harvey, 2003] où un champ de perméabilité multi-Gaussien, donc faiblement corrélé, est transformé en un champ équivalent, i.e. avec les mêmes propriétés statistiques, mais où les valeurs

élevées présentes une forte connectivité. Les propriétés hydrauliques associées sont profondément modifiées par cette réorganisation, notamment en ce qui concerne la perméabilité effective et les temps de transport. Une autre alternative consiste à générer des champs aléatoires de perméabilité présentant des structures très conductrices fortement corrélées en imitant la présence de chenaux dans le milieu et en se basant sur propriétés de ces chenaux, comme la distance inter-chenal [Koltermann and Gorelick, 1996; Ronayne et al., 2008]. Les modèles obtenus permettent alors de reproduire plus efficacement les observations que les modèles ne prenant pas en compte la chenalisation des écoulements. Cette dernière méthode s'appuie sur des indicateurs statistiques décrivant la connectivité des zones de forte perméabilité afin de générer des modèles compatibles avec le milieu naturel. [Western et al., 2001] ont ainsi mis au point un indicateur basé sur la mesure de la connectivité des fortes valeurs de perméabilités basée sur la statistique multipoint permettant de quantifier cette connectivité. [Knudby and Carrera, 2005] ont eux défini deux types d'indicateurs, l'un basé sur la connectivité des fortes valeurs de perméabilité et l'autre basé sur la connectivité des chenaux d'écoulement à partir des propriétés de transport du milieu. Leur conclusion est que la connectivité mesurée pour les flux n'est pas corrélée à la connectivité mesurée pour le transport et que le concept de connectivité est dépendant du processus analysé. L'utilisation de qualificateur des propriétés d'écoulement est cependant amenée à se développer car elle permet d'améliorer les modèles décrits en a) en intégrant la chenalisation des écoulements observée dans les milieux naturels [Renard, 2007]. Des modèles intégrant les effets de la chenalisation sont présentés à la figure 1-2.

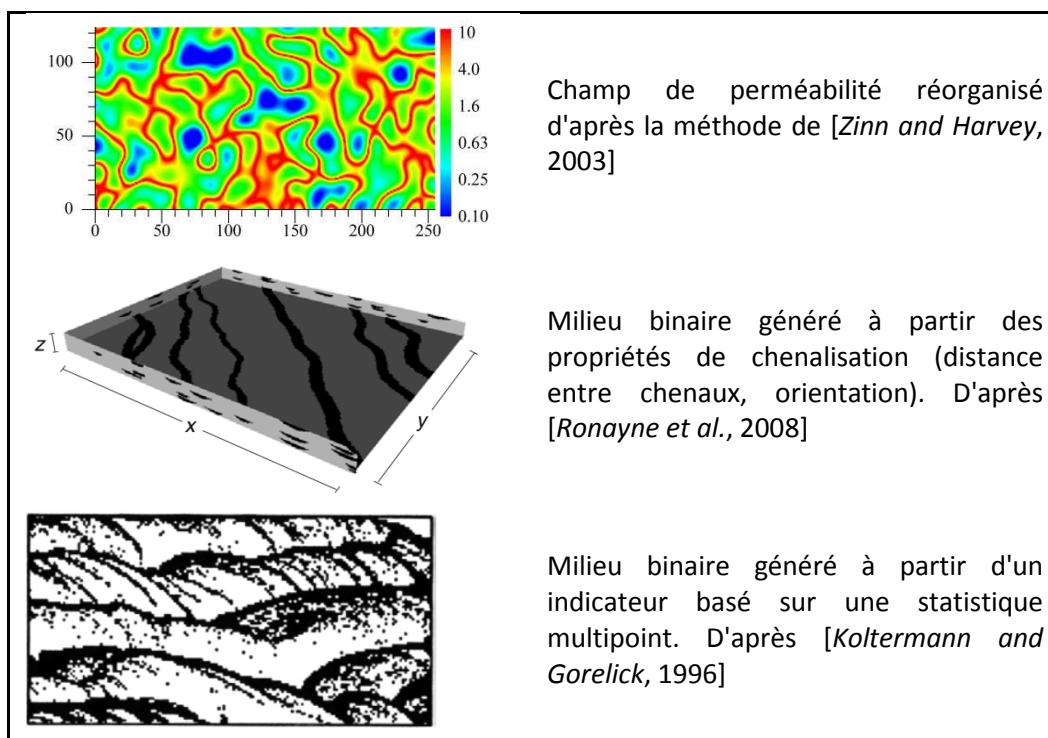


figure 1-2 : Illustration de champs de perméabilité intégrant des propriétés de chenalisation.

c) *Conclusions sur les modèles de perméabilité*

On représente généralement le champ de perméabilité par une grille en 2 ou 3 dimensions dont les valeurs sont générées à partir d'un modèle de référence. Ce modèle peut être une interpolation de données de terrain ou généré statistiquement à partir des propriétés du champ de perméabilité du milieu naturel. Les lois de génération simples, comme le modèle multi-

Gaussien, sont capables de modéliser globalement l'hétérogénéité du milieu en intégrant la variance de la perméabilité et en intégrant des lois de corrélation simples. Cependant, les écoulements calculés à partir de ce type modèle sont souvent dénués de pertinence à cause de la présence, dans le milieu naturel, de chemins préférentiels d'écoulement. Cette chenalisation du flux doit être prise en compte. Les modèles les plus récents intègrent donc les propriétés statistiques de la connectivité entre fortes valeurs de perméabilité.

2. Les milieux fracturés

Dans les milieux fracturés, la roche est quasi-imperméable et l'eau ne peut s'écouler qu'au travers des fractures. Les observations du milieu naturel mettent en évidence une complexité à la fois géométrique, liée au réseau de fracture, et hydraulique, de part l'organisation des écoulements dans ce réseau. L'hétérogénéité est présente à toutes les échelles. A l'échelle de la fracture, sa forme, son ouverture et sa rugosité induisent des propriétés hydrauliques variables spatialement [Tsang and Tsang, 1989]. A l'échelle du réseau, les écoulements dépendent de la distribution des longueurs de fractures, des orientations, des ouvertures ainsi que de la manière dont les fractures sont organisées entre elles [de Dreuzy et al., 2001a; b]. On se réfèrera à [Berkowitz, 2002] pour une étude critique des méthodes de caractérisation des écoulements dans les milieux fracturés.

a) Représentation discrète ou continue ?

Une étude bibliographique permet de dégager deux tendances générales basées sur des hypothèses différentes, voire contradictoires. La première suppose qu'il est absolument nécessaire, au vu de l'importance hiérarchique des structures du réseau de fracture sur les écoulements, de représenter de manière discrète les éléments de ce réseau afin d'en déduire les propriétés hydrauliques. C'est l'hypothèse, par exemple, de Cacas [Cacas et al., 1990a; b], qui a mis au point un modèle stochastique de fractures discrètes pour interpréter les mesures d'écoulements dans le massif cristallin de Fanay-Augères (France). A partir de 20 réseaux de fractures aléatoires en trois dimensions générés à partir des propriétés statistiques des fractures observées, il a en effet réussi à reproduire fidèlement les temps de transport mesurés lors d'expériences hydrauliques. La méthodologie consistant à caractériser au mieux la structure du réseau de fracture pour en déduire les propriétés hydrauliques est ainsi utilisée couramment. La deuxième approche est radicalement différente dans la mesure où elle suppose que connaître les propriétés du milieu fracturé est trop hasardeux et de surcroît inutile pour connaître ses propriétés hydrauliques. Elle préconise de modéliser le milieu naturel par un champ continu de perméabilité dont les propriétés sont adaptées en fonction des expériences hydrauliques. Le même site de Fanay-Augères a été modélisé par un milieu continu par Ando et al. [Ando et al., 2003]. A partir de 100 réalisations respectant les propriétés statistiques déduites du milieu naturel, les tests hydrauliques effectués sur le site ont également été fidèlement reproduits. Une troisième approche tente de réconcilier les approches discrètes et stochastiques en intégrant, dans un champ continu, les fractures porteuses des principaux écoulements du milieu. Ainsi, Rodrigo et al. [Rodrigo et al., 2002] utilisent un modèle stochastique continu incorporant de manière déterministe la position des fractures majeures. Les cellules de la grille correspondant à une fracture ont une transmissivité beaucoup plus forte que la matrice en arrière plan afin d'intégrer la notion de structures préférentielles d'écoulement. Ils montrent que cette représentation permet de simuler des écoulements chenalisés et est donc bien adaptée aux milieux fracturés.

Malgré leurs différences, ces trois approches tentent de modéliser l'hétérogénéité du milieu et d'intégrer la chenalisation des écoulements. La figure 1-3 illustre ces trois modèles et présente

les flux associés. On considère que lorsque le réseau de fracture est dense et contient des structures fortement interconnectées, il peut alors être modélisé par un milieu continu. Par exemple, [Long *et al.*, 1982] ont étudié ce type de modèles et ont défini les conditions à remplir pour que l'approche continue soit efficace. Ces conditions sont que le réseau de fracture sous-jacent est dense et interconnecté, que les conditions aux limites induisent un gradient constant et que les ouvertures sont homogènes. Ces restrictions mettent l'utilisation des modèles continus en question pour des milieux qui sont fortement chenalisés et qui, par conséquent, ne respectent pas ces conditions.

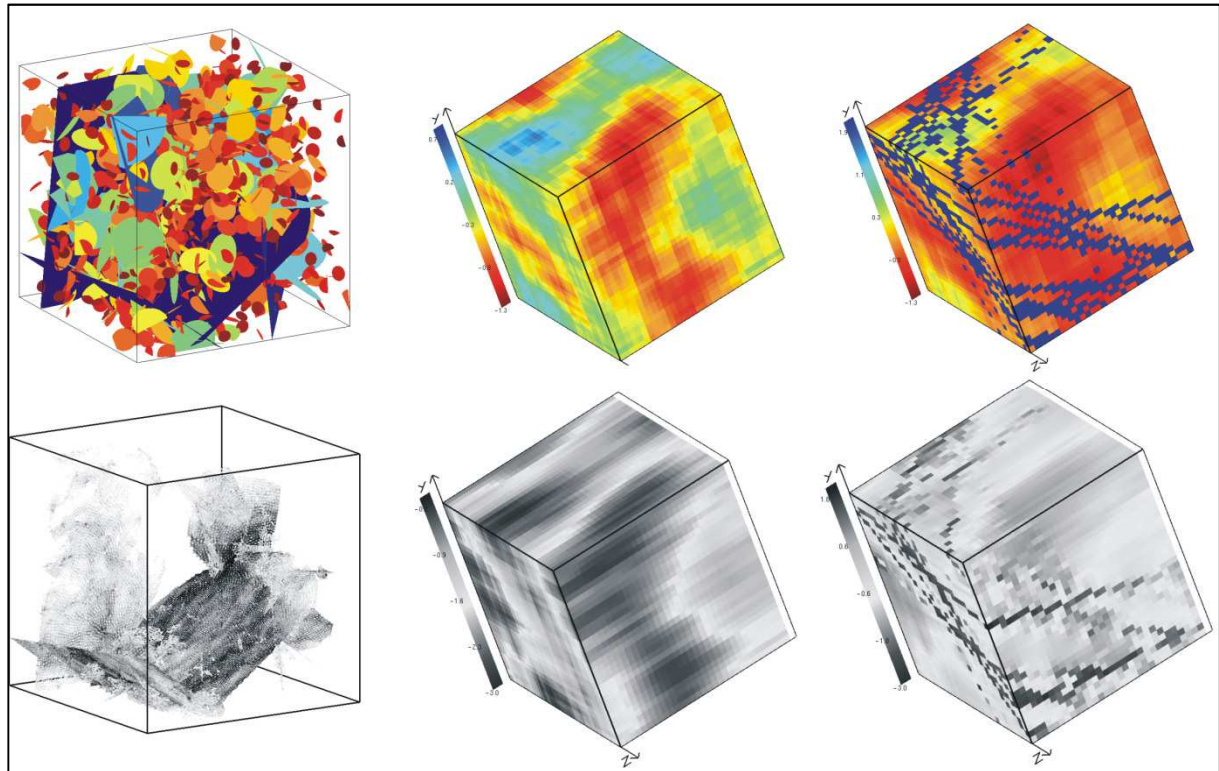


figure 1-3 : trois représentations différentes d'un milieu fracturé. Gauche : approche discrète, le milieu est représenté par un réseau de fracture et les écoulements sont calculés à l'intérieur de chaque fracture. Milieu : approche continue, le milieu est représenté par une grille hétérogène de perméabilité. (Droite) : approche mixte, le milieu est représenté par une grille hétérogène contenant des plans de fractures

b) Modèle statistique de fractures et échantillonnage

Lorsqu'une approche continue n'est pas envisageable, le milieu est généralement modélisé par un réseau de fractures discrètes où chaque fracture est individualisée et représentée par un objet géométrique (segment en 2D, polygone, cercle ou ellipse en 3D) auquel on assigne une valeur de conductivité hydraulique. Décrire la perméabilité du milieu revient ainsi à décrire le réseau de fracture sous-jacent. Connaître l'ensemble du réseau de fracture de manière déterministe est impossible, puisque ceci reviendrait à carter l'ensemble des fractures du milieu, y compris celles invisibles situées profondément sous la surface. Une description statistique des propriétés de fracturation, comme la longueur, l'orientation, la densité et la transmissivité des fractures, est donc requise.

Le modèle décrivant la distribution des longueurs de fracture est crucial dans la mesure où c'est celui qui permet de faire le changement d'échelle, c'est-à-dire relier les observations à l'échelle d'un affleurement (1-10 m) aux observations faites à une échelle régionale (1-100

km). Différents modèles de distributions de longueurs ont été étudiés [Bonnet et al., 2001] et il résulte que les modèles en loi de puissance correspondent le mieux aux observations des milieux naturels et permettent de décrire efficacement les propriétés de densité de fracture aux différentes échelles. Ils sont dérivés du fait que le taux de croissance d'une fracture est apparemment une loi de puissance [Atkinson, 1989]:

$$\frac{dl}{dt} \sim l^a \quad (1-1)$$

Si $a \neq 1$, la distribution de densité de fracture $n(l)$ s'écrit alors [Bour et al., 2002]:

$$n(l, L) = \alpha l^{-a} L^D \quad (1-2)$$

où $n(l, L)$ est le nombre de fracture de taille $[l, l+dl]$ dans un système de taille L et D la dimension massique. Estimer l'exposant de la loi de puissance a ainsi que le terme de densité α à partir des observations du milieu géologique n'est pas triviale. La méthode classique consiste à cartographier les fractures à différentes échelles, de l'échelle du puits via imagerie radar ou analyse de carottages à l'échelle régionale grâce aux images aériennes et satellites, en passant par l'échelle de l'affleurement où les fractures peuvent être cartées précisément. Les distributions ainsi échantillonnées sont analysées pour inférer les valeurs de a et de α . La figure 1-4 donne un exemple de ce type d'analyse dans le bassin d'Hornelen en Norvège [Bour et al., 2002].

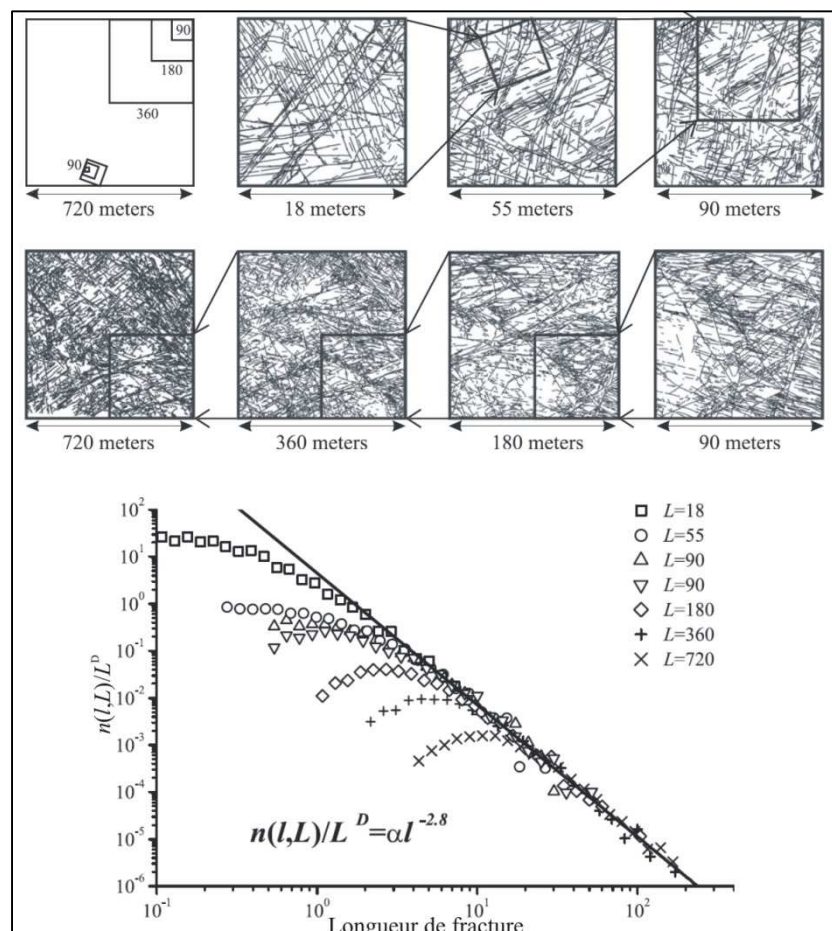


figure 1-4 : Calibration du modèle en loi de puissance des longueurs de fracture appliquée au bassin d'Hornelen (Norvège). À gauche, carte de traces de fracture à différentes échelles et à droite, distribution des longueurs normalisée par la surface d'échantillonnage. D'après les figures 1 et 6 de [Bour et al., 2002].

La principale difficulté rencontrée lors de la calibration du domaine vient de la qualité des données utilisées (nombre, précision) ainsi que biais liés à l'échantillonnage. [Pickering et al., 1995] décrivent ainsi les deux principales causes de ces biais. Tout d'abord, l'échantillonnage est souvent tronqué dans la mesure où il est effectué dans une gamme d'échelle limitée et à un emplacement fixé. Le nombre de fractures observées dont la taille est proche des limites cette échelle est donc biaisée. De plus, les structures de la taille de l'affleurement sont sous-estimées de part la faible probabilité de les observer. Ensuite, un échantillonnage est généralement « censuré », c'est à dire que les fractures dont l'extrémité est située en dehors de la surface d'échantillonnage sont artificiellement coupées et la taille mesurée est inférieure à la taille réelle. A côté de ces biais statistiques, les données utilisées peuvent être naturellement biaisées soit par la topographie du milieu naturel, qui peut masquer, tronquer ou segmenter des fractures, soit par l'érosion subit par la surface d'échantillonnage, qui tend également à déconnecter des ensembles de traces. Des méthodes mathématiques, comme celle de [Riley, 2005], existent pour corriger les biais d'échantillonnage. Elles s'appuient pour cela sur les propriétés supposées de la distribution d'origine et la géométrie des fractures. Cependant, ces méthodes supposent que les fractures ont une forme simple et que leurs propriétés sont indépendantes. C'est pourquoi d'autres méthodes, comme celle de [Dershowitz et al., 1991b], privilégient des simulations stochastiques de réseaux de fractures en intégrant ces biais pour tenter de retrouver la distribution échantillonnée et ainsi déterminer les paramètres de la distribution originale. Ces dernières méthodes sont efficaces lorsqu'il s'agit de retrouver une distribution en loi de puissance à partir d'une distribution tronquée et/ou censurée mais, à notre connaissance, n'intègrent pas les biais liés à la topographie ou à l'altération du sol.

La distribution des orientations est également un paramètre important du réseau de fracture dans la mesure où il influe sur l'orientation des écoulements et la connectivité du milieu. Dans un milieu à trois dimensions, une fracture est définie par l'orientation dans laquelle elle s'est propagée et par son pendage, c'est-à-dire son orientation par rapport à la verticale. Cette distribution est échantillonnée essentiellement à partir des données de puits où le nombre de fractures disponibles est important et éventuellement à partir des cartes d'affleurement lorsque les fractures sont suffisamment ouvertes pour mesurer leur pendage. Comme pour les longueurs, les orientations mesurées sont biaisées par l'orientation du puits, les fractures quasi-parallèles au puits ayant une très faible probabilité d'être échantillonnées. La méthode de correction la plus simple, connue sous le nom de correction de Terzaghi [Terzaghi, 1965], consiste à pondérer le nombre de fracture d'une classe d'orientation par le coefficient $w = \cos \varphi^{-1}$ dépendant de l'angle φ entre le puits et la fracture. Cependant, cette correction tend à surreprésenter les fractures horizontales dans les cas où le puits est vertical. Cette observation a conduit Davy et ses co-auteurs [Davy et al., 2006] à corriger ce coefficient pour les réseaux de fractures dont les longueurs suivent une distribution en loi de puissance, le nouveau terme de correction dépendant de a et de φ .

Les deux paramètres précédents (longueur et orientation) définissent la géométrie du milieu. Cependant les écoulements ne se répartissent pas de manière uniforme dans le réseau de fracture. Comme dans les milieux poreux, où le flux dépend de la perméabilité, chaque fracture ne permet qu'un écoulement limité. Cette limitation est reliée à une valeur de transmissivité quantifiant ses propriétés hydrauliques. Les transmissivités sont naturellement hétérogènes et leur distribution ne peut être évaluée que par des tests hydrauliques, notamment par des tests de pompage. Comme la transmissivité est hétérogène à l'échelle de la fracture, on se contente généralement de définir une valeur moyenne par fracture, l'hétérogénéité globale étant constituée par l'organisation des fractures. A partir de mesure in-situ des transmissivités, [Cacas et al., 1990a] ont proposé et appliqué un modèle en loi log-

normale pour simuler les écoulements. En se basant sur les mesures effectuées sur un puits du laboratoire d'Aspö (Suède), [Gustafson and Fransson, 2006] ont plutôt obtenu une distribution en loi de puissance.

Le modèle de réseau de fracture final intègre l'ensemble des propriétés statistiques calibrées à partir du milieu naturel afin de simuler propriétés hydrauliques. La figure 1-5 donne un aperçu de plusieurs réseaux synthétiques de fractures en deux dimensions aux propriétés statistiques identiques : system carré de taille $L=100$, longueurs distribuées suivant une loi de puissance de coefficient $a=3$, orientations uniformes, transmissivité distribuées suivant une loi log-normale, positions uniformément distribuées et densité commune. On s'aperçoit que, même si les propriétés statistiques sont connues, les propriétés hydrauliques comme le nombre de chenaux d'écoulement, leur connexion et leur position, peuvent varier d'une configuration à l'autre.

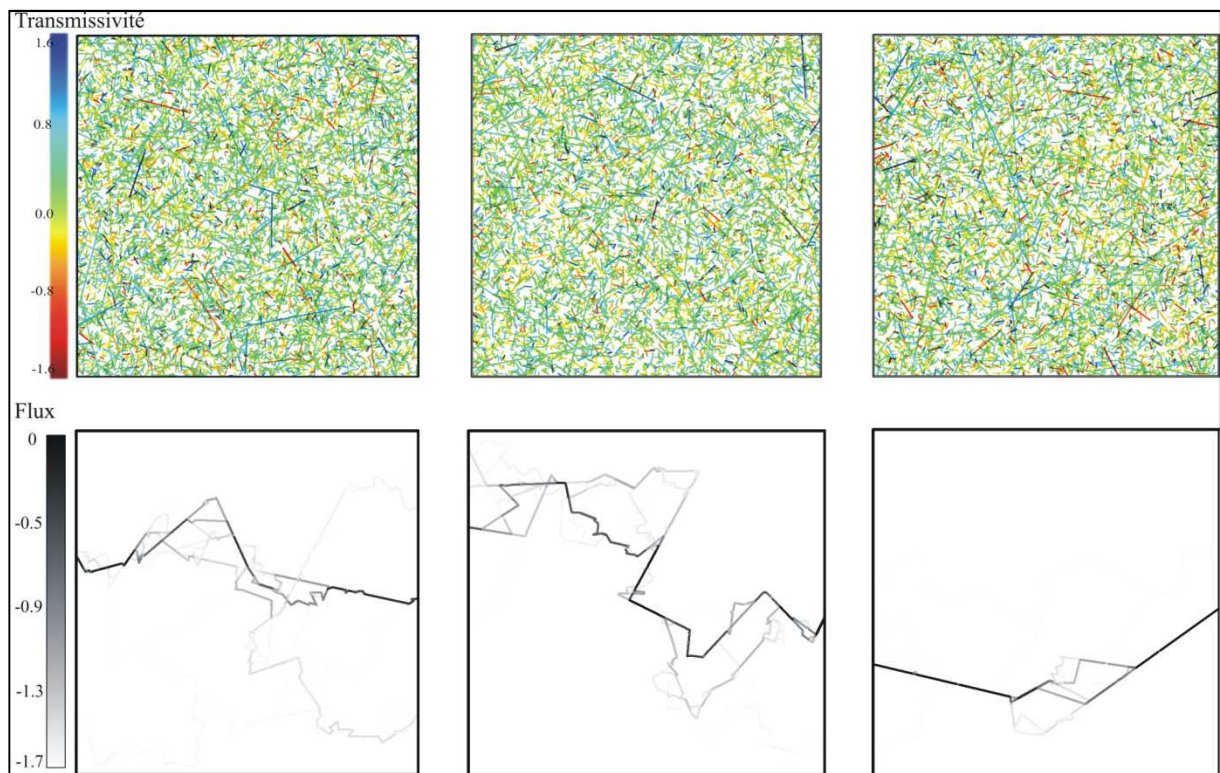


figure 1-5 : Réseaux de fractures ayant les mêmes propriétés statistiques. (Haut) : le réseau de fracture. La couleur associée à chaque fracture correspond au logarithme de sa transmissivité. (Bas) : les écoulements correspondant au réseau du haut. Chaque segment est coloré en fonction de la quantité de flux qui le traverse.

c) *La chenalisation dans les milieux fracturés*

La complexité structurelle des milieux fracturés induisent naturellement une complexité des propriétés hydrauliques. Cette complexité provient de la fracture elle-même, qui provoque une première chenalisation des écoulements [Tsang and Neretnieks, 1998], mais également de l'organisation entre les fractures dont les gammes de longueurs, d'orientations et ouverture sont très larges [de Dreuzy et al., 2001a; b]. Charlaix et al. [Charlaix et al., 1987] ont ainsi montré que les écoulements étaient localisés dans une petite partie du réseau de fracture, et principalement dans les fractures les plus transmissives, les fractures les moins conductrices ayant une influence négligeable sur les propriétés hydrauliques. Appréhender les propriétés hydrauliques requière donc soit un grand nombre de donnée, soit une approche théorique

cohérente avec l'organisation des écoulements, la principale difficulté provenant de la chenalisation des écoulements dans une petite partie de la fracturation. On pourra se référer à [Neuman, 2005] pour avoir une idée générale des méthodes utilisées pour caractériser les propriétés de flux de les milieux fracturés.

d) Conclusions sur les milieux fracturés

La différence entre un milieu poreux et un milieu fracturé vient des structures physiques constituant la perméabilité. Au lieu d'un assemblage de pores, puis de matériaux, la perméabilité provient essentiellement des fractures présentes dans le milieu. L'hétérogénéité, puis la chenalisation des écoulements, est ainsi généralement plus forte dans les milieux fracturés. Si l'approche continue, qui est l'approche de modélisation classique dans les milieux poreux, est souvent appliquée aux milieux fracturés, des approches spécifiques ont été développées, notamment tout ce qui concerne la représentation en réseaux de fractures discrètes. Comme pour les milieux poreux, une description statistique de ces milieux est indispensable dans la mesure où il est impossible d'observer et donc de décrire toutes les fractures du milieu. De plus, la taille des fractures est très variable et couvre une grande gamme d'échelle, provoquant in fine une hétérogénéité à toutes les échelles qu'il est difficile d'homogénéiser. La méthodologie générale pour bâtir un modèle de site nécessite donc un gros effort de récolte d'information sur le milieu naturel et de validation du modèle à partir d'expérimentation sur site.

B. Calibration des modèles de perméabilité : le problème inverse

Lorsqu'on cherche à caractériser les propriétés hydrauliques d'un milieu naturel, on s'appuie sur les modèles conceptuels décrits au paragraphe précédent. Ce sont des représentations simplifiées du milieu s'appuyant sur un certain nombre de paramètres comme la définition d'une loi statistique ou des valeurs de perméabilité par zone. Si le concept sous-jacent au modèle s'inspire des observations du milieu et de ses propriétés supposées, assigner une valeur pertinente à chaque paramètre à partir de ces mêmes observations n'est pas toujours simple et ne garantit pas forcément que les propriétés hydrauliques déduites de ce modèle restent pertinentes. De plus, même dans les milieux les mieux échantillonnés, on ne peut avoir qu'une vision très partielle du milieu et les hypothèses émises lors de la conception du milieu ont alors une influence déterminante sur ses capacités à prédire fidèlement les écoulements. L'étude de *Bredehoeft* [Bredehoeft, 2005] illustre parfaitement ces limitations. Il compare, pour un certain nombre de sites, les prédictions réalisées il y a quelques années et les analyses réalisées a posteriori afin de déterminer si ces prédictions étaient avérées. Il en conclut que si les paramètres sont mal calibrés, par exemple dans le cas du bassin de la Blue River (USA), les niveaux d'eau et de flux prédits peuvent être très sur- ou sous-estimés. Un biais dans la distribution des paramètres dans le cas du site de Phoenix (USA) a eu des conséquences identiques. C'est pourquoi utiliser les données disponibles sur les propriétés hydrauliques du milieu pour calibrer le modèle du site est une nécessité. La perméabilité, responsable des écoulements, est donc déterminée à partir des mesures sur ces écoulements (charges locales, flux entre puits, etc.), ce qui constitue un problème inverse.

1. Les principes du problème inverse

Le problème inverse en hydrogéologie est traité depuis une quarantaine d'année et l'évolution de la manière dont ce problème a été abordé est décrite par de Marsily [de Marsily et al., 2000]. L'aspect mathématique du problème inverse ainsi que certaines méthodes de résolution sont décrites par Tarantola [Tarantola, 2004]. On se contentera, dans cette partie, de décrire

les grands principes du problème inverse pour se concentrer sur la mise en œuvre de sa résolution dans le cadre de l'hydrogéologie, et particulièrement pour les milieux hétérogènes.

a) La fonction-objectif

Le problème inverse sert de lien entre les paramètres du modèle et les données hydrauliques observées sur le milieu naturel. Ce lien est concrètement réalisé par la fonction-objectif. Cette fonction est minimale lorsque le modèle reproduit fidèlement les données d'observation et croît lorsque l'écart entre les données simulées et les données réelles augmente. La formulation la plus utilisée reste la formulation en moindre carré s'écrivant [Chavent, 1991]:

$$F_{\text{obj}}(\mathbf{x}) = \sum_{i=1}^{n_d} \left(\frac{d_i(\mathbf{x}) - d_i^{\text{obs}}}{\sigma_i} \right)^2 \quad (1-3)$$

où \mathbf{x} représente le vecteur des paramètres, n_d le nombre de donnée, $d_i(\mathbf{x})$ la valeur de la $i^{\text{ème}}$ donnée résultant du modèle, d_i^{obs} la même donnée mais issue du milieu naturel et σ_i^d l'incertitude associée à cette donnée. Résoudre le problème inverse revient alors à trouver \mathbf{x} tel que $F_{\text{obj}}(\mathbf{x})$ soit minimale.

Pour parvenir à un résultat, plusieurs difficultés doivent être surmontées [Carrera et al., 2005]. Tout d'abord le problème est mal-posé. Un problème est bien posé si il a une solution, que cette solution est unique et si elle dépend continument des données [Hadamard, 1902], or le problème inverse n'a priori pas de solution unique. Par ailleurs, pour résoudre le problème inverse, il est nécessaire d'estimer un grand nombre de fois la valeur de F_{obj} et donc de résoudre l'équation (I-1), ce qui peut s'avérer coûteux en temps et en puissance de calcul. De part l'hétérogénéité du milieu naturel et sa simplification au travers de la paramétrisation, l'incertitude sur les valeurs des paramètres est également non-négligeable [Moore and Doherty, 2006] et la sensibilité des données aux paramètres peut, selon les conditions, être faible. La solution pour contourner ces difficultés consistent à augmenter la connaissance du milieu, en ajoutant des contraintes sur les valeurs des paramètres en fonction de leurs propriétés supposées, comme des propriétés de connectivité, afin de réduire l'espace des paramètres, ou à leur associer une valeur a priori. Cette valeur a priori est utilisée pour régulariser la fonction objectif en lui ajoutant un critère de plausibilité [Neuman, 1973]. Il s'agit en substance de considérer que le résultat doit satisfaire à la fois le critère sur les données hydrauliques et un critère physique basé sur des mesures locales de perméabilité. La fonction-objectif s'écrit alors :

$$F_{\text{obj}}(\mathbf{x}) = \sum_{i=1}^{n_d} \left(\frac{d_i(\mathbf{x}) - d_i^{\text{obs}}}{\sigma_i^d} \right)^2 + \alpha \cdot \sum_{j=1}^{n_p} \left(\frac{x_j - x_j^{\text{prior}}}{\sigma_j^p} \right)^2 \quad (1-4)$$

où x_j est la valeur du $j^{\text{ème}}$ paramètre, x_j^{prior} sa valeur a priori, σ_j^p l'incertitude associée à cette valeur, n_p le nombre de paramètre et α un coefficient de pondération.

Minimiser la fonction objectif dépend donc d'un subtil équilibre entre les paramètres d'un côté et les données de l'autre. Si le modèle contient un trop grand nombre de paramètre, la quantité de données disponibles sera alors insuffisante pour calibrer efficacement leurs valeurs, aboutissant à un problème non-solvable ou très incertain. C'est pourquoi, en terme de paramétrisation le principe du modèle « le plus simple possible » est souvent retenu [Hill, 2006]. D'un autre côté, un modèle trop simple ne sera pas capable de simuler efficacement les données disponibles et ne rendra pas compte de la complexité du milieu. C'est pourquoi, le véritable principe serait plutôt le modèle « le plus simple possible, mais pas trop simple » [Hunt et al., 2007]. Pour cette raison, Tsai et ses co-auteurs [Tsai et al., 2003a] ont intégré la

paramétrisation dans la formulation du problème inverse. Ils définissent pour cela un problème inverse généralisé dont la fonction-objectif s'écrit :

$$F_{\text{obj}}(S, \mathbf{x}) = \sum_{i=1}^{n_d} \left(\frac{d_i(S, \mathbf{x}) - d_i^{\text{obs}}}{\sigma_i^d} \right)^2 + \alpha \cdot \sum_{j=1}^{n_p} \left(\frac{x_j - x_j^{\text{prior}}}{\sigma_j^p} \right)^2 + G(P(S, \mathbf{x}), P_0(S_0, \mathbf{x}_0)) \quad (1-5)$$

où $P(S, \mathbf{x})$ représente la paramétrisation de structure S et de paramètres \mathbf{x} , P_0 une paramétrisation de référence et $G(P_1, P_2)$ une fonction évaluant les performances de la paramétrisation P_1 par rapport à celle de P_2 . Cette fonction-objectif est alors optimisée en deux temps. Dans un premier temps la paramétrisation est fixée et ses paramètres sont calibrés en minimisant la fonction de l'équation (1-4). Dans un second temps, les performances de cette paramétrisation sont évaluées et comparées à une paramétrisation de référence, généralement celle ayant jusqu'à présent obtenu les meilleurs résultats, et sa structure est optimisée en conséquence par ajout et suppression de paramètres de manière à minimiser l'équation (1-5).

b) Les méthodes d'optimisation de la fonction objectif

Il existe différentes méthodes d'optimisation dont l'efficacité dépend essentiellement des propriétés de la fonction objectif, notamment sa régularité et de la présence ou non de minima locaux.

(1) Les méthodes de gradients

Les méthodes les plus efficaces sont les méthodes de gradient. Elles ne nécessitent que peu d'évaluations de la fonction-objectif et trouvent rapidement le minimum global de la fonction. Par contre elles ne fonctionnent que si le problème est bien posé, et que donc la fonction objectif a été régularisée. Hill en donne une description détaillée [Hill and Tiedeman, 2007]. Ces méthodes s'appuient sur le gradient de la fonction. En effet, un développement en série de la fonction-objectif au voisinage de \mathbf{x}_0 s'écrit :

$$F_{\text{obj}}(\mathbf{x}_0 + \delta \mathbf{x}) = F_{\text{obj}}(\mathbf{x}_0) + \boldsymbol{\gamma}_0 \cdot \delta \mathbf{x} + \frac{1}{2} \cdot \mathbf{H}_0 \cdot \delta^2 \mathbf{x} + O(3) \quad (1-6)$$

où $\boldsymbol{\gamma}_0$ est le gradient de F_{obj} au point \mathbf{x}_0 , \mathbf{H}_0 son Hessian et $O(3)$ un terme qui tend vers 0 plus rapidement qu'un terme du second ordre. F_{obj} est à un optimum pour $\boldsymbol{\gamma}(\mathbf{x})=0$, ce qui revient à poser :

$$\mathbf{x} \sim \mathbf{x}_0 - \mathbf{H}_0^{-1} \cdot \boldsymbol{\gamma}_0 \quad (1-7)$$

Si la fonction est convexe, alors le minimum est obtenu en quelques itérations grâce à l'algorithme de Newton :

$$\mathbf{x}_{n+1} \sim \mathbf{x}_n - \mathbf{H}_n^{-1} \cdot \boldsymbol{\gamma}_n \quad (1-8)$$

Ce principe est illustré à la figure 1-6.

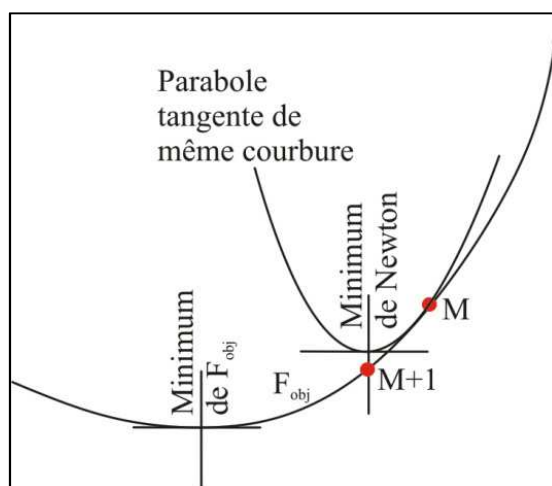


figure 1-6 : Illustration de l'algorithme de Newton (modifiée de [Tarantola, 2004]). Pour évaluer la direction de descente au point M on cherche le minimum de la parabole tangente à la courbe au point M et ayant la même courbure que la fonction objectif. Le point M+1 qui a la même abscisse que ce minimum constitue la prochaine étape de l'algorithme.

Cet algorithme requiert le calcul du Hessien à chaque étape, ce qui est extrêmement coûteux en termes de calcul. C'est pourquoi des méthodes dérivées de la méthode de Newton ont été développées qui évitent son calcul. On citera principalement les méthodes de Gauss-Newton et de Levenberg-Marquart qui s'appuient sur les matrices de sensibilités des paramètres pour trouver une valeur approchée du Hessien.

(2) Les méthodes stochastiques

Si la fonction-objectif ne peut pas être suffisamment régularisée, par exemple en l'absence de valeurs a priori, ou si le gradient de la fonction n'est pas accessible, les méthodes de gradient ne sont pas applicables. Pour minimiser la fonction-objectif, il faut alors utiliser un algorithme non déterministe basé sur un parcours stochastique de l'espace des paramètres. Les méthodes les plus utilisées sont les méthodes de Monte-Carlo comme le recuit simulé [Ingber, 1993b], les algorithmes génétiques [Holland, 1975] et les méthodes évolutionnaires comme le CMA-ES (Covariance Matrix Adaptive Evolution Strategies) [Hansen and Ostermeier, 1996]. Comme il existe une littérature exhaustive pour chacune de ces méthodes, nous nous contenterons de rappeler le principe du recuit-simulé puisque c'est cet algorithme que nous avons utilisé.

Le recuit simulé a été développé par [Kirkpatrick et al., 1983] et s'appuie sur l'algorithme d'échantillonnage de l'espace de paramètre mis au point par Metropolis [Metropolis et al., 1953]. Il s'agit d'une méthode probabiliste et heuristique qui cherche le minimum global en analogie avec la solidification des roches. Elle est efficace sur les problèmes à grande échelle et est applicable même en cas de fonction irrégulière et de minima locaux. A chaque point x de l'espace des paramètres, on associe une énergie correspondant à la valeur de la fonction-objectif. On associe également à chaque étape de l'algorithme une température T , initialement élevée et qui décroît progressivement. A chaque étape, une marche aléatoire permet de se déplacer dans l'espace des paramètres. Passer à un point de plus faible énergie est immédiat alors que se déplacer vers un point de plus haute énergie nécessite d'« utiliser » l'énergie de la température pour y parvenir. Cette utilisation se traduit par une probabilité d'accéder ou non au point suivant s'écrivant :

$$p(\mathbf{x}_1 \rightarrow \mathbf{x}_2) = \min\left(\exp\left(-\frac{F_{\text{obj}}(\mathbf{x}_2) - F_{\text{obj}}(\mathbf{x}_1)}{T}\right), 1\right) \quad (1-9)$$

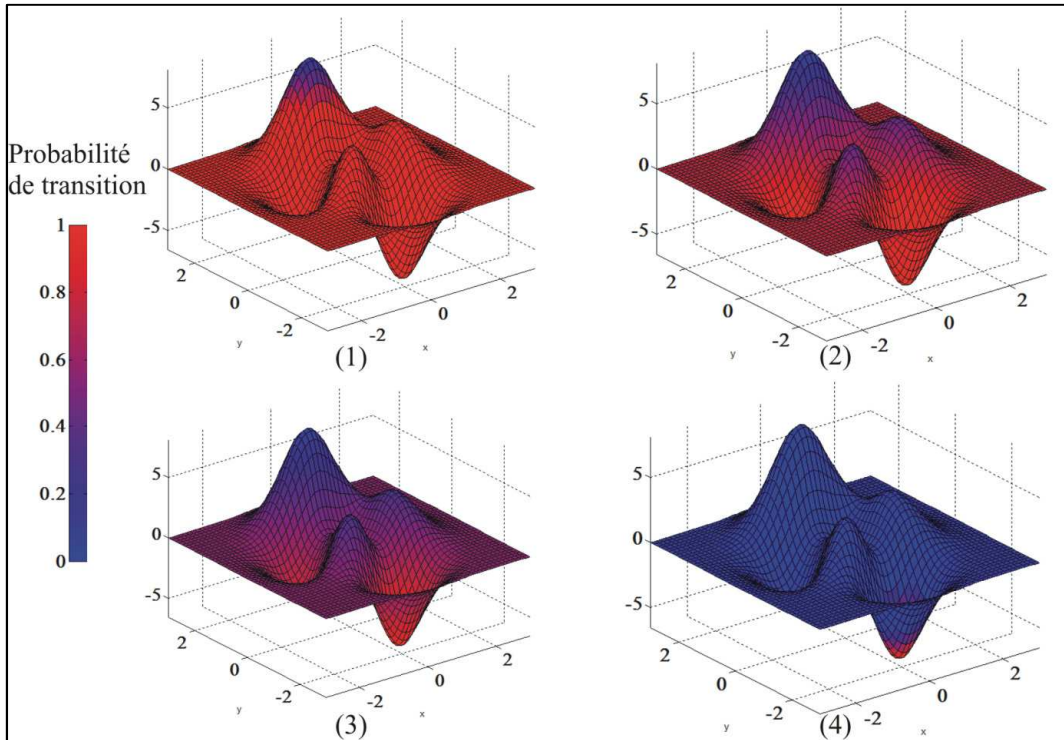
Lorsque la température décroît, l'« espace » accessible se restreint pour finalement se concentrer autour du minimum global (figure 1-7). Les performances du recuit simulé dépendent, en plus de la forme de la fonction-objectif, des paramètres de la marche aléatoire, de la température initiale et du calendrier de décroissance de cette température. La marche aléatoire consiste à choisir \mathbf{x}_2 dans un voisinage de \mathbf{x}_1 dans l'équation (1-9). Ce voisinage peut être local ou global en fonction de la proximité entre \mathbf{x}_1 et \mathbf{x}_2 . Sur les conseils de *Ingber* [Ingber, 1993b], le recuit simulé est plus efficace si on privilégie une recherche globale à forte température et locale lorsque la température est faible. Il conseille donc de choisir \mathbf{x}_2 tel que :

$$\mathbf{x}_2 = \mathbf{x}_1 + \mathbf{y} \cdot (\mathbf{x}_{\text{max}} - \mathbf{x}_{\text{min}}) \quad (1-10)$$

où \mathbf{x}_{max} et \mathbf{x}_{min} sont les valeurs limites des paramètres et \mathbf{y} , qui représente la taille du « bond » dans l'espace, est défini par :

$$y_i = \text{sgn}\left(u - \frac{1}{2}\right) T \left[\left(1 + \frac{1}{T}\right)^{|2u-1|} - 1 \right] \quad (1-11)$$

où u est tiré aléatoirement d'une distribution uniforme $U[0,1]$.



Une fonction de décroissance de la température associée à ce type de parcours aléatoire qui permet d'atteindre le minimum global s'écrit alors, toujours selon *Ingber* [Ingber, 1993b]:

$$T(k) = T_0 \exp\left(-c \cdot k^{\frac{1}{D}}\right) \quad (1-12)$$

où T_0 est la température initiale, D la dimension de l'espace des paramètres, k le nombre de configurations acceptées lors du recuit simulé et c un paramètre « libre » servant à ajuster l'algorithme.

2. Exemple de résolutions du problème inverse en hydrogéologie

Le problème inverse est appliqué couramment en hydrogéologie. Il permet de calibrer les modèles de site à partir des données hydrauliques. Les données les plus utilisées sont les hauteurs de charges, en stationnaire ou en transitoire. On présente ici quelques exemples d'application afin de comprendre les enjeux et les limitations de l'approche inverse.

a) Modèle de zones

Une approche simple et intuitive consiste à découper le milieu naturel en zones de perméabilité constante et à trouver les valeurs de perméabilité associées à chaque zone par un problème inverse. La géométrie des différentes zones peut reposer sur une interprétation géologique du milieu ou sur un découpage autour de points de référence. Ainsi, *Tsai et al.* [Tsai et al., 2003b] résolvent le problème inverse sur un milieu poreux 3D. Les paramètres du milieu sont constitués de 1 à 5 points dont la position et la perméabilité sont inconnues et le milieu est découpé par une zonation de Voronoi basée sur ces points. Ces paramètres sont calibrés à partir de 640 hauteurs piézométriques et 640 concentrations provenant de 16 puits d'observation uniformément répartis sur l'aquifère. La valeur des paramètres est optimisée par un algorithme génétique dont les résultats sont affinés par une méthode de gradient.

La figure 1-8 présente les résultats obtenus pour les différentes paramétrisations (de 1 à 5 zones). L'écart entre les données simulées et les données réelles diminue lorsqu'on augmente le nombre de zones et le modèle à 5 zones, même s'il satisfait les conditions du problème, n'arrive pas à reproduire fidèlement les observations. Ceci tend à indiquer que le modèle en zone n'est pas suffisant si la structure de perméabilité réelle est trop complexe. D'un autre côté, *Sun et Yeh* [Sun and Yeh, 2007a; b] ont appliqué une méthodologie similaire à des milieux poreux synthétiques 2D et ont montré qu'un découpage en 2 ou 3 zones pouvaient modéliser efficacement des structures de perméabilités plus complexes (distribution log-normale, distribution discrètes ou imitant la présence de fractures) si l'on ne cherche qu'à représenter la charge hydraulique. Ce résultat indique qu'en l'absence d'hypothèses sur la structure de la perméabilité, calibrer les paramètres uniquement à partir des valeurs de charge peut s'avérer insuffisant si l'on cherche à caractériser les propriétés à plus petite échelle du milieu considéré.

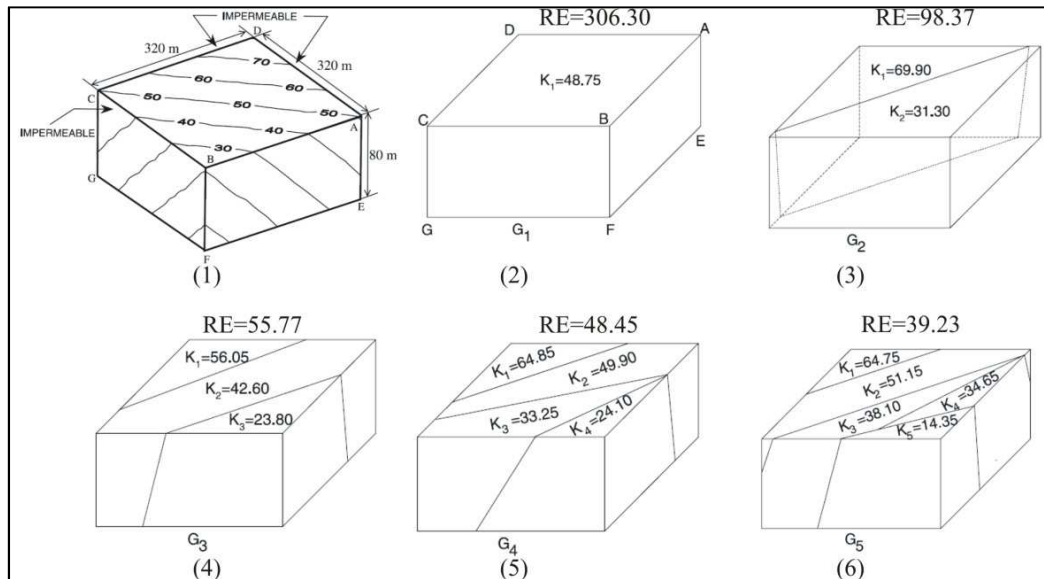


figure 1-8 : Problème inverse appliqué à un modèle de zones. (1) conductivité hydraulique du milieu de référence (m/d). (2) à (6) : modèle obtenu après résolution du problème inverse contenant de 1 à 5 zones. Les valeurs de RE correspondent à la valeur de la fonction-objectif (eq. (1-4)). D'après [Tsai et al., 2003b].

b) Modèles par interpolation

Si le modèle de zone est trop restrictif, de part les discontinuités fortes qu'il induit, la technique qui repose sur l'identification de certaines valeurs de perméabilité pour modéliser l'ensemble du milieu est largement utilisé en hydrogéologie. Au lieu de découper le milieu en zones homogènes, l'autre possibilité est de définir une fonction d'interpolation qui, à partir des valeurs identifiées, reconstruit l'ensemble du champ de perméabilité. Par exemple, Tsai et al [Tsai et al., 2005] ont appliqué la même méthode d'inversion que [Tsai et al., 2003b] sur un exemple similaire. Au lieu d'utiliser une zonation, ils ont modélisé le champ de perméabilité par une interpolation linéaire entre les valeurs des paramètres et ont obtenu des résultats bien meilleurs en termes d'accord entre les données observées et les données simulées (figure 1-9).

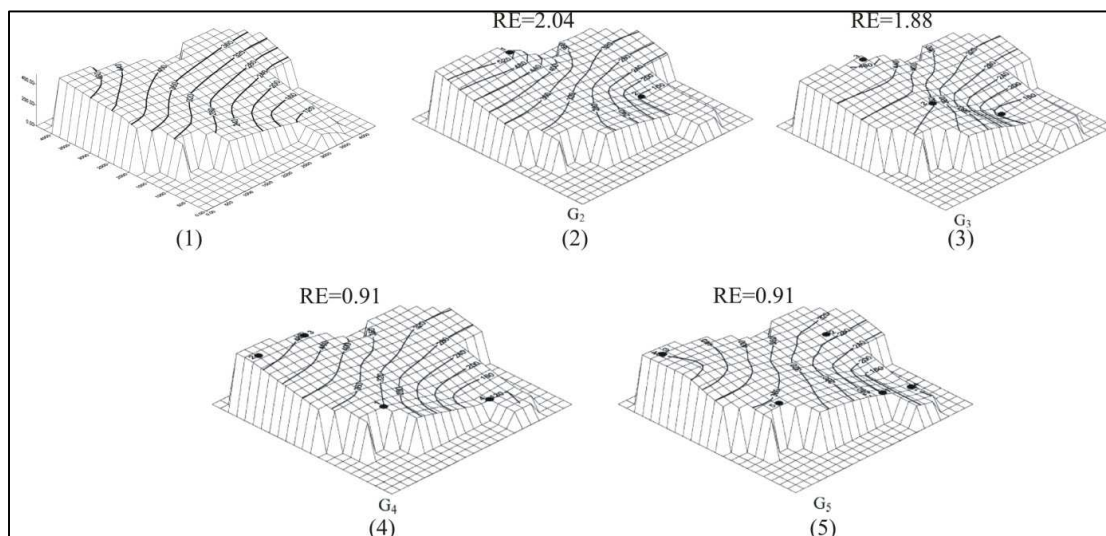


figure 1-9 : Problème inverse avec interpolation. (1) conductivité hydraulique du milieu de référence (m/d). (2) à (5) : modèle obtenu après résolution du problème inverse contenant de 2 à 5 perméabilités de référence. Les valeurs de RE correspondent à la valeur de la fonction-objectif (eq. (1-4)). D'après [Tsai et al., 2005].

La méthode des points-pilotes, initialement conçue par *de Marsily* [*de Marsily et al.*, 1984] et appliquée à l'identification d'un champs de perméabilité par [*LaVenue et al.*, 1995; *RamaRao et al.*, 1995], repose sur un principe similaire. Elle considère que la propriété hydraulique recherchée, comme la perméabilité k , est composée d'une tendance générale k_D et d'une perturbation k_p :

$$k(\mathbf{x}) = k_D(\mathbf{x}) + k_p(\mathbf{x}) \quad (1-13)$$

La tendance générale peut par exemple être déduite par krigeage à partir des valeurs mesurées sur le site afin qu'elle respecte les propriétés géophysiques supposées du milieu. La perturbation quant à elle est déduite par interpolation à partir de points de référence définis par le modélisateur. Le nombre de ces points peut être fixe ou augmenter de manière itérative au cours de la modélisation. La position et la perméabilité de ces points, dits points-pilotes, sont optimisées par la résolution d'un problème inverse, généralement par une méthode de gradient. Cette méthode a par exemple été utilisée par *Alcolea* et ses co-auteurs [*Alcolea et al.*, 2006] sur un champ synthétique 2D de transmissivité. Ils ont utilisés 13 valeurs de transmissivité pour calculer k_D et plusieurs séries de valeurs de charges en transitoire (936 valeurs au total). La fonction-objectif donnée à l'équation (1-4) a été optimisée par une méthode de gradient (Levenberg–Marquardt). Ils ont montré que si le nombre de points-pilotes est suffisamment grand et si les valeurs a priori sont bien pondérées (coefficient α de l'équation (1-4)) alors le champ estimé était très proche du champ réel (figure 1-10).

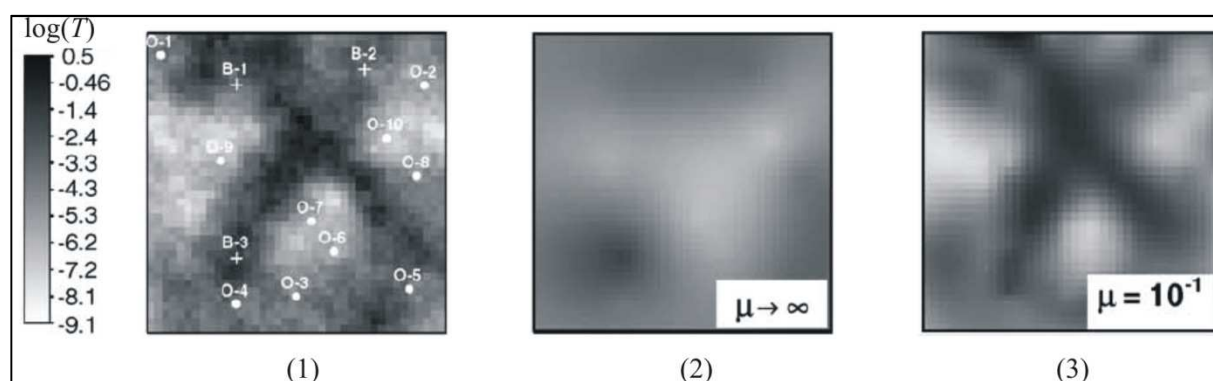


figure 1-10 : Résultat de l'inversion via la méthode des points pilotes. (1) champs de références et emplacement des points de mesure, (2) tendance générale calculée par krigeage à partir des 13 valeurs de transmissivité accessibles. (3) résultat du problème inverse à partir de 241 points pilotes. La valeur de μ correspond à α^1 dans l'équation (1-4). D'après [*Alcolea et al.*, 2006].

c) *Autres méthodes de résolution*

Si les méthodes décrites précédemment font partie des méthodes standards, il existe une multitude de méthode de résolution du problème inverse. Ces méthodes se distinguent par la paramétrisation sous-jacente, discrète ou statistique, la manière d'incorporer les données hydrauliques dans la fonction-objectif et la méthode de résolution. Une description exhaustive de toutes ces méthodes n'est pas l'objectif de cette partie. Il est par contre important de bien comprendre les points clés qui vont permettre à une méthode d'être efficace ou non. Ainsi, plusieurs méthodes géostatistiques ont été utilisées pour modéliser un certains nombres de configurations test lors de la fameuse étude du *WIPP* (Waste Isolation Pilot Plant) [*Zimmerman et al.*, 1998]. Il s'agissait de comparer les capacités de prédiction de sept méthodes géostatistiques, dont les points pilotes, sur des milieux synthétiques présentant divers degrés de complexité. Si toutes les méthodes avaient des performances équivalentes sur

les milieux peu hétérogènes, des différences sont apparues pour les prédictions liées au transport de particules dans les milieux hétérogènes. Les méthodes s'appuyant sur une linéarisation du problème inverse ont en effet montré des difficultés sur les milieux présentant des « anomalies » ou des zones très transmissives. Dans les méthodes non linéaires, la capacité à prendre en compte les particularités du milieu test, comme la connectivité des fortes perméabilités, s'est révélé déterminante dans les configurations présentant des discontinuités.

Il semble donc que le plus important, dans la résolution du problème inverse, est la capacité du modèle sous-jacent à gérer les hétérogénéités du milieu et particulièrement ses discontinuités. Ainsi, concernant les milieux où les écoulements sont très chenalisés, l'effort de modélisation et notamment l'accent mis sur les paramètres les plus sensibles semblent être plus important que la méthode d'optimisation utilisée.

3. Adaptation du problème inverse aux milieux fracturés

Les méthodes traditionnelles montrent leurs limites lorsqu'il s'agit de modéliser les milieux très hétérogènes, particulièrement en présence de chemins préférentiels d'écoulement. C'est pourquoi une attention particulière a été portée sur le problème inverse dans les milieux fracturés. La résolution du problème inverse dans les milieux fracturés dépend de la paramétrisation choisie, soit grâce à un milieu continu, soit grâce à un modèle discret (voir paragraphe A.2).

a) Modèles continus

L'aquifère karstique de Laroche-foucault (France) a été caractérisé via la résolution d'un problème inverse par *Larocque et al.* [*Larocque et al.*, 1999]. Le milieu est modélisé par une grille régulière et deux pas de discrétisation sont utilisés. Les transmissivités sont déduites de 70 charges en transitoires pour la discrétisation la plus grossière et 117 charges pour la discrétisation la plus fine. Afin d'évaluer l'importance de la répartition des observations, ils ont également utilisé 54 données de charges régulièrement réparties sur l'aquifère et calculées par interpolation (krigeage) des données existantes. Pour diminuer le nombre de paramètres du modèle, *Larocque* a utilisé une méthode de réduction de l'espace des paramètres [*Siegel et al.*, 1996]. Les transmissivités sont tout d'abord évaluées en 4 points et les valeurs de la grille de discrétisation sont calculées par krigeage à partir de ces valeurs. Le nombre de points de référence est ensuite augmenté progressivement et les valeurs déterminées précédemment sont utilisées comme valeurs initiales. Les valeurs des paramètres sont optimisées à chaque étape par une méthode de gradient (Quasi-Newton). La figure 1-11 montre les transmissivités obtenues avec les deux discrétisations et les deux types de données, réelles et interpolées.

Les modèles obtenus présentent un bon accord avec les données observées et sont cohérents avec les valeurs de transmissivité obtenues par tests de pompage et de traçage. L'utilisation de charges interpolées s'est montrée aussi efficace que l'utilisation de valeur réelle, ce qui peut permettre de combler un manque de donnée ou une mauvaise répartition des valeurs à condition que le champ de charge soit suffisamment régulier. L'intérêt des modèles continus est que le nombre de paramètres à inverser reste limité et peut être ajusté en fonction de la quantité de d'information disponible. Cependant, dans l'étude précédente, l'utilisation d'une grille plus fine n'a pas permis d'améliorer les résultats, ce qui signifie que raffiner un modèle continu ne permet pas d'améliorer ses capacités de modélisation, notamment en présence de fortes hétérogénéités. Si les résultats sont satisfaisant pour le milieu étudié, considéré comme modérément karstique, la pertinence des modèles continus pour des milieux plus hétérogènes, avec des écoulements plus chenalisés, peut être par contre mise en question.

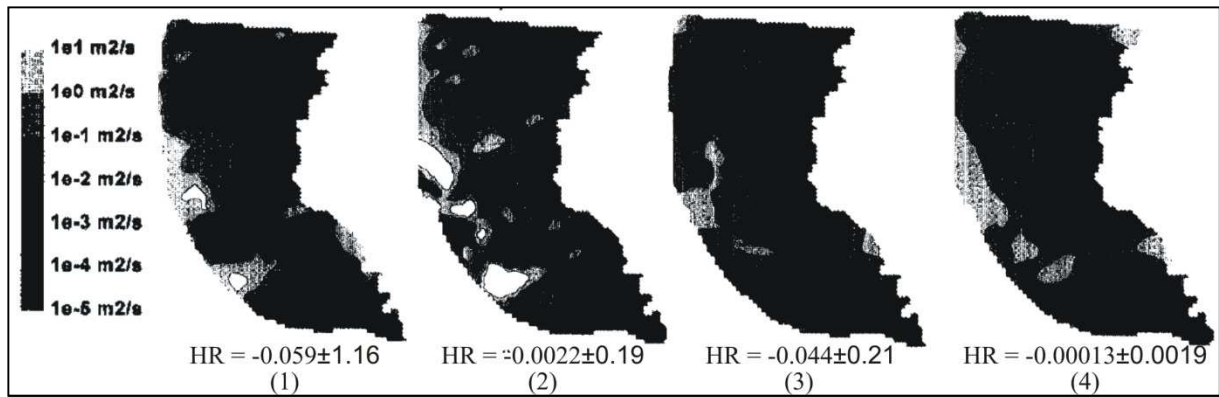


figure 1-11 : Transmissivités d'un milieu karstique calibrées par un modèle continu. (1) milieu modélisé par une grille de pas discrétisation $p=2.5\text{km}$ et calibré avec les données réelles. (2) milieu modélisé par une grille de $p=1\text{km}$ et calibré avec les données réelles. (3) $p=2.5\text{km}$ et calibré avec les données interpolées. (4) $p=1\text{km}$ et données interpolées. La valeur de HR , pour « head residuals », représente l'écart entre les données réelles et les données simulées. D'après [Larocque et al., 1999].

En effet, si les écoulements sont très chenalisés, certaines structures, comme des fractures majeures dans le cas d'un milieu fracturé, devient prédominantes dans les propriétés hydrauliques. La solution adoptée par *Franssen and Gómez-Hernández* [Franssen and Gómez-Hernández, 2002] a été d'ajouter de manière déterministe la position des 21 fractures majeures dans la grille de perméabilité. Ils ont appliqué cette paramétrisation au site fracturé d'Aspö (Suède) et résolu un problème inverse pour déterminer la perméabilité moyenne de chacune de ces fractures et celle de la matrice poreuse. Les valeurs de perméabilité sont ensuite réparties dans la grille 3D en fonction du type de cellule (appartenant à la matrice ou une fracture) suivant une loi de distribution centrée sur la valeur moyenne correspondante. Ces valeurs moyennes sont obtenues en résolvant le problème par méthode d'auto calibration séquentielle (sequential self-calibrated approach, *Gómez-Hernández et al.* [Gomez-Hernandez et al., 1997]). Les données utilisées étaient constituées de mesures de conductivités hydrauliques déterminées par essais de puits et de hauteurs piézométriques. 270 conductivités étaient disponibles dans la matrice et 23 dans les plans de fracture, combinées à 51 charges hydrauliques. Les résultats ont montré un bon accord entre les charges simulées et les charges observées et les conductivités des fractures étaient significativement supérieures aux conductivités de la matrice poreuse (figure 1-12), ce qui valide ce type d'approche pour les milieux fracturés. Cependant, la position des fractures majeures est supposée connue ce qui est rarement le cas dans les milieux fortement chenalisés.

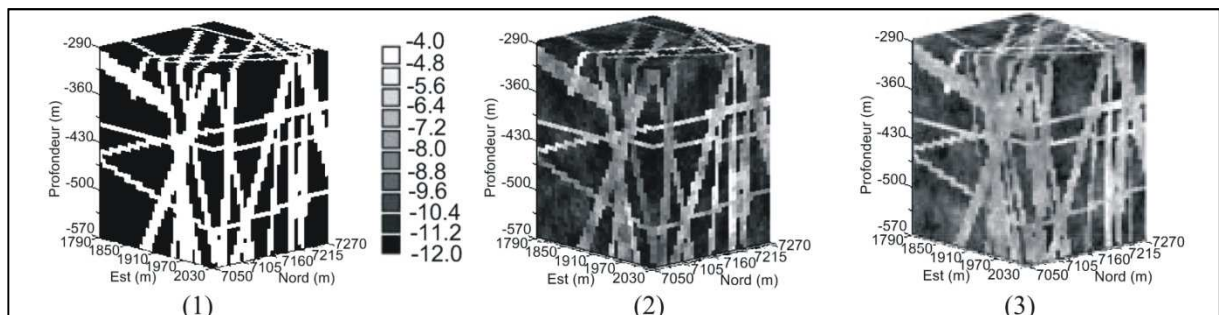


figure 1-12 : modèle poreux-fracturé. (1) classification des cellules de la grille entre fractures (blanc) et matrice (noir). (2). Conductivités initiales calibrées à partir des données de conductivité et de position des fractures. (3) Modèles finaux après résolution du problème inverse en utilisant les données de charge.

Une représentation intermédiaire entre un milieu continu et un milieu discret est utilisée par Nakao [Nakao et al., 2000]. Le milieu est modélisé par un treillis régulier où chaque élément représente une fracture. Ainsi, la position et l'orientation des fractures ne sont pas représentées, ce qui rapproche ce modèle d'un modèle continu. Les fractures sont regroupées en amas anisotropes de taille fixée et une valeur de transmissivité est associée à chaque amas. Les inconnues du problème sont le centre et la transmissivité de chaque amas. Elles sont identifiées par une méthode de recuit simulé à l'aide de charges hydrauliques en transitoires. Ce problème est résolu en appliquant différentes tailles d'amas afin de déterminer la corrélation spatiale de la distribution de transmissivités. Les résultats présentent une bonne concordance entre les données simulées et les observations, validés par des tests d'injection non utilisés dans la résolution du problème inverse, et permettent d'estimer la corrélation spatiale du champ de transmissivité (figure 1-13). Par contre, compte-tenu de la paramétrisation utilisée, le modèle ne permet pas d'identifier les structures les plus transmissives, ce qui implique qu'elle ne peut être appliquée qu'à des milieux sans discontinuités majeures. Pour cette raison, ce type de modèle se rapproche des représentations par un milieu continu.

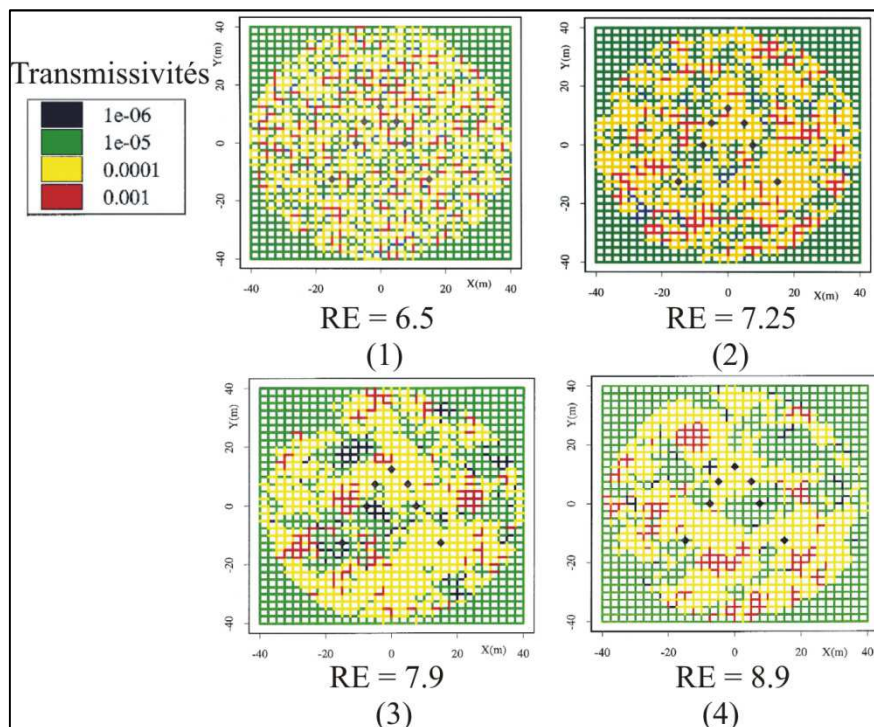


figure 1-13 : Résultats obtenus avec un modèle en treillis régulier avec des transmissivités assignées par cluster de taille L . (1) $L=1$, (2) $L=10$, (3) $L=20$ et (4) $L=40$. La valeur de RE (« residual error ») correspond au minimum de la fonction-objectif.

La représentation en treillis a également été utilisée par Gwo [Gwo, 2001] pour modéliser et identifier les chemins préférentiels d'écoulement d'un milieu poreux fracturé. Dans ce milieu, la porosité, la perméabilité et la dispersivité de la matrice sont fixées, mais la position des segments correspondant au réseau de fracture ainsi que leur ouverture, choisie dans une gamme limitée, sont inconnues. Ces inconnues sont déterminées par la résolution d'un problème inverse utilisant d'une à trois courbes de restitution d'un traceur inerte comme données de référence. Le problème est ensuite résolu par un algorithme génétique qui identifie, pour chaque montant du treillis, s'il appartient au réseau de fractures et le cas échéant la classe de transmissivité à laquelle il appartient. Si les résultats sont significatifs sur un cas très simple, dès que le réseau de fracture est un peu plus complexe, comme avec de la tortuosité par exemple, la méthode est moins efficace. Gwo l'a néanmoins appliqué à une

expérience en colonne mesurant, en laboratoire, les écoulements sur un milieu naturel. Le modèle ainsi généré, même si ses propriétés physiques semblent assez éloignées du milieu de référence, arrive à reproduire les courbes de restitutions (figure 1-14). Le fait d'intégrer dans le modèle l'existence de chemins préférentiels d'écoulement permet de bien simuler les observations, même si des travaux supplémentaires sont nécessaires pour arriver à faire correspondre la structure physique du milieu avec le modèle.

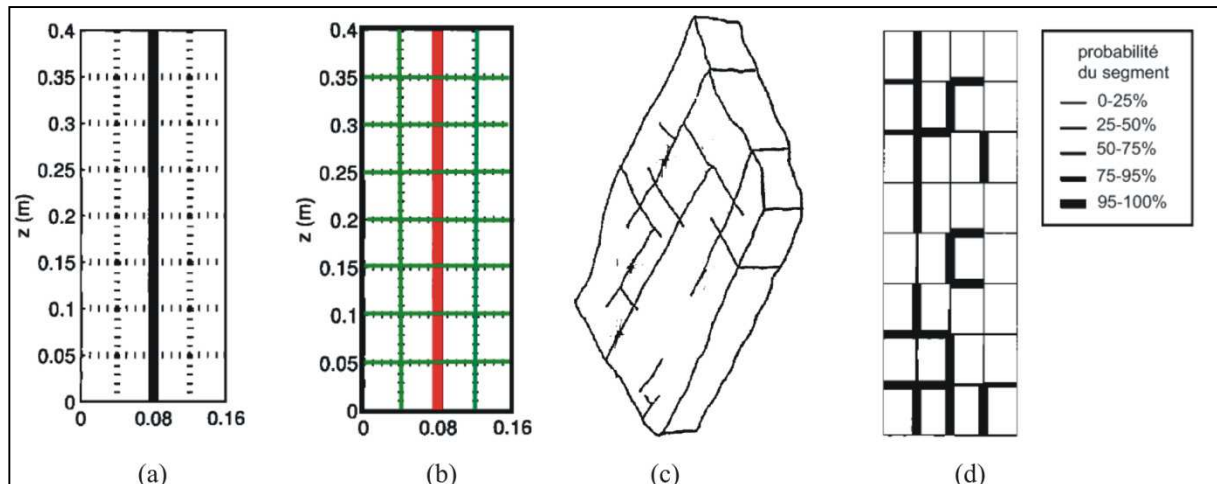


figure 1-14 : Problème inverse appliqué à un milieu poreux fracturé. (a) milieu synthétique de référence avec ouvertures constantes et (b) résultat du problème inverse correspondant au milieu (a). (c) Schéma de la fracturation visible de cas naturel et (d) résultat du problème inverse appliqué au cas naturel.

b) Modèles en réseau de fractures discrètes (DFN)

Les approches continues ne sont pas forcément pertinentes pour les milieux fracturés où les écoulements sont très chenalisés dans la mesure où seulement quelques structures majeures dominent les écoulements. Les modèles en réseau de fracture présentent l'avantage de représenter ces structures en même temps que l'ensemble des fractures du système. Cependant, calibrer l'ensemble du réseau de fracture, avec la position et la transmissivité de chaque élément, est impossible au vu du nombre de paramètres nécessaires. C'est pourquoi une simplification est nécessaire, afin de réduire le nombre de paramètres.

Un premier type de simplification est de considérer que le réseau est ordonné en famille de fractures ayant des caractéristiques communes. Ainsi, dans le cadre de la modélisation du site de El Berrocal (Espagne), *Donado et al* [*Donado et al.*, 2005] résolvent le problème inverse appliqué à un milieu fracturé. Cependant la description du réseau fracture (taille des fractures, positions, orientations) est déterminé par approche directe à partir de critères tectoniques. Le réseau est ensuite découpé en 5 grandes familles de fractures et seuls la transmissivité et l'emmagasinement de chaque famille est déterminé par approche inverse. Ces paramètres sont calibrés à partir de charges hydrauliques et de concentrations par une méthode de maximum de vraisemblance [*Carrera and Neuman*, 1986] et les résultats sont comparés avec ceux obtenus en utilisant un modèle de milieu poreux équivalent. La conclusion est que, en réduisant le nombre de paramètres, les modèles en DFN permettent de caractériser les écoulements et les modèles résultant de la résolution du problème inverse sont au moins aussi efficaces que ceux obtenus par une approche continue. Une approche similaire est développée à l'IFP par *Bruyelle et Lange* [*Bruyelle and Lange*, 2009]. La structure du réseau de fracture est encore une fois fixée à l'aide de mesures tectoniques, les fractures sont réparties en deux

sous-ensembles et la transmissivité de chaque ensemble est déterminée via une approche inverse s'appuyant sur un algorithme évolutionnaire (CMA-ES, [Hansen and Ostermeier, 1996]) et en utilisant des données de tests hydrauliques. Les résultats montrent une rapide convergence de la méthode et l'efficacité de ce type d'approche dans les milieux fracturés.

Les méthodes précédentes considèrent que la structure du réseau de fracture est connue et résolvent le problème inverse uniquement sur les transmissivités. Ce problème est beaucoup plus simple que le problème inverse complet qui doit être résolu à la fois sur la structure et sur les charges. Il suffit de quelques hypothèses sur les transmissivités pour que le problème soit bien posé et il peut être résolu à partir de données sur le flux. Il est équivalent au problème inverse dans les milieux poreux utilisant un terme de régularisation important, basé sur la structure de la perméabilité. Ce type d'approche ne peut donc concerner que les réseaux de fracture dont les observations géologiques permettent une connaissance précise de la fracturation. Dans un milieu où les écoulements sont fortement chenalisés, les structures principales qui dominent les écoulements ne sont par contre pas connues. *Tiedeman et al.* [Tiedeman et al., 1995] sont, à notre connaissance, les premiers à avoir tenté de localiser et de caractériser ce type de structure. Le modèle utilisé était volontairement simple puisqu'il ne comportait qu'une seule structure de transmissivité infinie, assimilée à une zone fracturée, dans un milieu homogène ou faiblement hétérogène. Les paramètres du modèle étaient constitués de la transmissivité du milieu homogène, de la géométrie de la zone de forte transmissivité, modélisée par une ellipse (position, longueur, orientation) et d'une constante C représentant la hauteur de charge aux limites du système. Ces paramètres ont été calibrés par recuit simulé en utilisant les charges en stationnaire mesurées en 8 puits d'observation lors d'un pompage dans un neuvième puits. Le problème est ensuite résolu plusieurs fois en utilisant différentes conditions initiales. Comme le problème est appliqué à un milieu synthétique, les valeurs optimales des paramètres sont connus ce qui permet de vérifier les résultats obtenus en terme de charge et de structure. Trois configurations de position des puits sont analysées et le milieu entourant la zone fracturée est soit homogène soit aléatoirement hétérogènes, ce qui donne six configurations tests. Les résultats sont illustrés à la figure 1-15.

Lorsque le milieu poreux est homogène, la géométrie de la zone de forte transmissivité et la transmissivité du milieu poreux sont bien estimées quelque soit la configuration de puits utilisée, et les charges résultantes sont en parfait accord avec les charges de référence, même si la position des solutions secondaires indique que la méthode peut converger dans un minimum local en fonction des conditions initiales. Lorsque le milieu poreux est hétérogène, l'identification est plus difficile et la configuration des puits joue un rôle important. Ainsi lorsque le pompage est proche de la structure à identifier, la méthode est plus efficace même si elle surestime la longueur de zone de forte transmissivité, particulièrement lorsque les puits couvrent une échelle inférieure à la taille de la zone. Pour un pompage plus éloigné, l'identification échoue et aucune solution n'est trouvée.

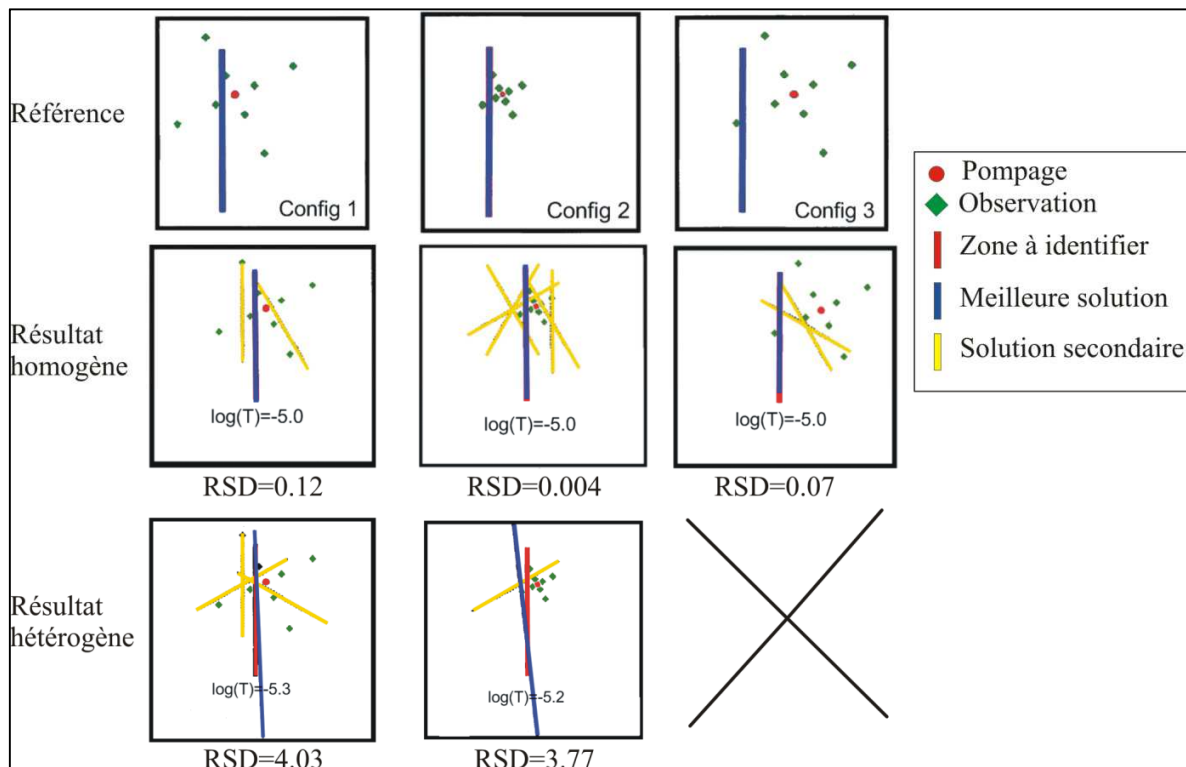


figure 1-15 : Résultats du problème inverse appliqué à l'identification d'une zone de forte transmissivité dans un milieu faiblement perméable. En haut se trouvent les trois organisations de puits testées avec un pompage proche ou éloigné de la zone recherchée et des puits d'observation couvrant une échelle proche de la taille de la structure ou plus petite. La ligne du milieu présentent les résultats lorsque le milieu poreux est homogène de transmissivité $T=10^{-5}$ et la dernière ligne correspond à un milieu poreux hétérogène. La croix signifie qu'aucune solution n'a été obtenue pour cette configuration. La valeur de RSD correspond au carré de l'erreur résiduelle normalisé par le nombre de paramètres. D'après [Tiedeman et al., 1995].

L'étude de *Tiedeman et al.* montre qu'il est possible, sous certaines conditions, de caractériser une structure de forte transmissivité à partir de donnée de charge en résolvant le problème inverse. *Renshaw* [Renshaw, 1996] est allé un peu plus loin en résolvant un problème similaire mais où la zone fracturée avait une géométrie plus complexe, correspondant à l'intersection entre deux fractures, le milieu poreux étant hétérogène. Le problème inverse a été résolu par une méthode de cokrigage itérative [Yeh et al., 1996] dans une configuration de puits favorable : pompage proche de la structure et puits couvrant une échelle équivalente à la taille de la zone fracturée. Ce test correspond au cas 4 de la figure 1-15 mais avec une zone fracturée correspondant à deux structures principales connectées. Ses résultats (figure 1-16) montrent qu'il est possible d'identifier une géométrie plus complexe, même dans milieu poreux hétérogène. La question se pose par contre de savoir s'il est possible d'identifier les structures les plus transmissives d'un réseau plus complexe de fractures.

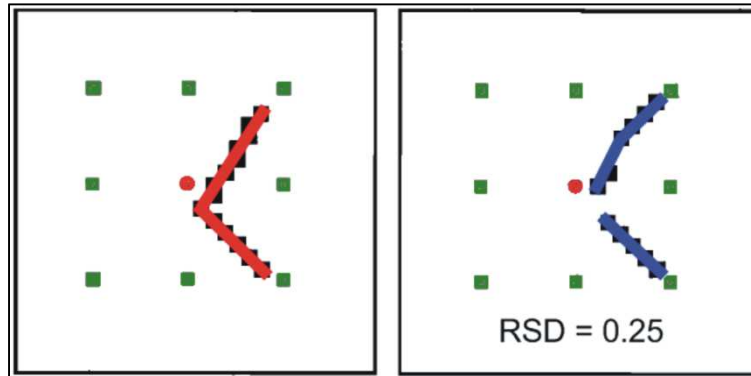


figure 1-16 : Identification d'une zone fracturée dans un milieu hétérogène. La zone de référence (rouge, gauche) est bien identifiée après la résolution du problème inverse (bleu, droite). Pour la légende voir figure 1-15. D'après [Renshaw, 1996].

Day-Lewis et al. [Day-Lewis et al., 2000] se sont penchés sur cette question. Ils ont tout d'abord tenté de déterminer la géométrie des zones de forte transmissivité d'un aquifère fracturé près de Mirror Lake (USA). Le milieu était discrétisé sous forme d'une grille régulière et chaque cellule de la grille pouvant appartenir soit à une zone de forte transmissivité, soit à la matrice poreuse de plus faible perméabilité. Le type de chaque cellule a été défini par la résolution d'un problème inverse grâce à des données indirectes sur la structure du réseau de fracture. Le type des cellules placées le long des puits de pompage ainsi que la connectivité entre les différentes zones du milieu était défini en amont par des tests hydrauliques. La fonction-objectif était définie par l'écart entre le variogramme supposé du champ de perméabilité, le nombre estimé de zones de forte transmissivité et la connectivité. Cette fonction-objectif a ensuite été optimisée par une méthode de recuit simulé, ce qui a permis de définir quatre zones. La transmissivité de ces zones, considérée comme isotrope, et la transmissivité de la matrice poreuse, considérée comme anisotrope, ont été ensuite calibrées par une méthode inverse s'appuyant sur des données hydrauliques recueillies en 13 puits via des tests hydrauliques et résolue par une méthode de gradient (Gauss-Newton). Les modèles résultant concordent mieux avec les données hydrauliques que les modèles négligeant la présence de zones de forte transmissivité interconnectées. Les résultats sont illustrés à la figure 1-18. Dans cette approche, si la position et la forme des zones les plus transmissives n'étaient pas connues a priori, elles ont néanmoins été calibrées par des données indirectes sur la structure du réseau de fracture. Le problème était donc différent de celui de [Tiedeman et al., 1995] et de [Renshaw, 1996], où la structure était calibrée avec des données hydrauliques. Il peut donc être considéré comme intermédiaire entre les problèmes où la structure est parfaitement connue et ceux où elle est complètement inconnue.

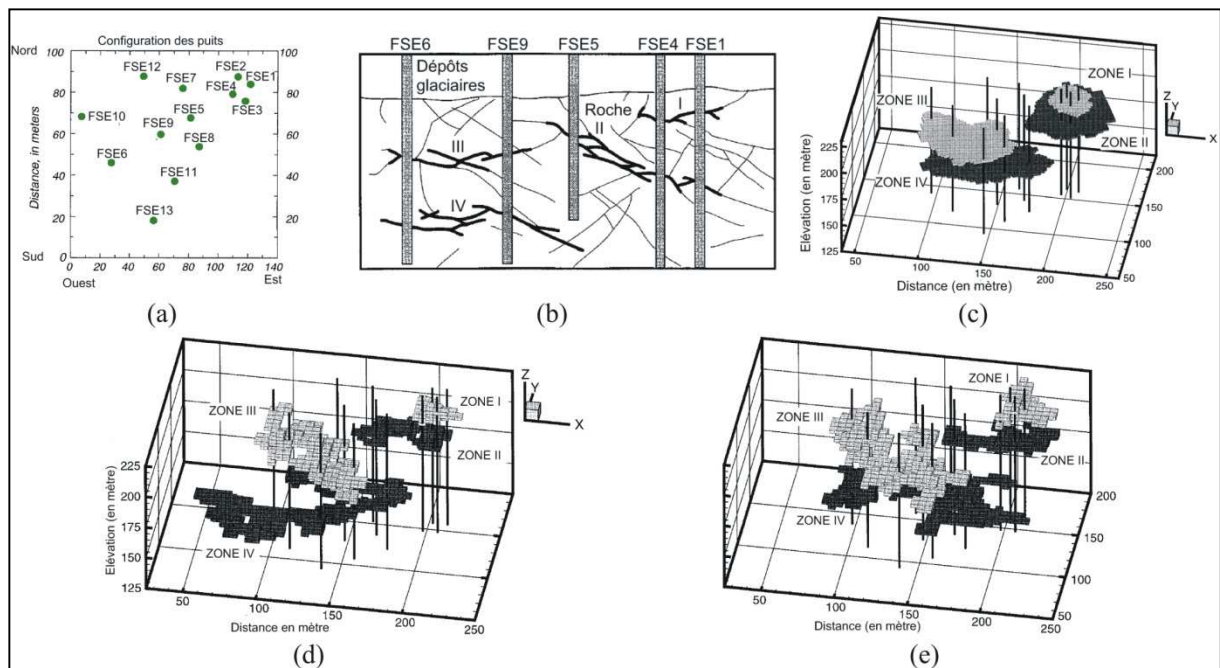


Figure 1-17 : Résolution d'un problème inverse pour l'identification de la structure des zones de forte transmissivité. (a) puits d'observations, (b) connectivité entre puits d'après les tests hydrauliques, (c) zones de forte transmissivité 3D définies a priori à partir des tests hydrauliques, (d) et (e) résultats du problème inverse. Modifié de [Day-Lewis et al., 2000].

c) *Autres paramétrisations du milieu fracturé*

Une approche différente a été utilisée par *Doughty et al.* [Doughty et al., 1994]. Comme les structures d'hétérogénéité sont multi-échelles, ils ont utilisé un modèle s'appuyant sur un système de fonctions itérées (IFS) capable de décrire une géométrie fractale. Ce système s'appuie sur plusieurs opérations transformant un ensemble de points en un autre que l'on applique itérativement. Pour un milieu en 2 dimensions, une fonction est définie par 4 paramètres. Cette modélisation est appliquée dans un premier temps à un réseau synthétique (4 paramètres) et dans un second temps à deux milieux naturels, le réservoir de Kesterson (Californie), modélisé par 16 paramètres, et la mine de Stripa (Suède), modélisée par 24 paramètres. Les paramètres sont calibrés à l'aide de données de charges observées au cours d'un pompage et optimisés par une méthode de recuit simulé. Les résultats sont illustrés à la figure 1-18. Les résultats du cas synthétique sont encourageants, même si l'inversion ne renvoie pas de résultat unique. Il faut alors combiner plusieurs réalisations et identifier leurs caractéristiques communes. Les modèles sur les cas réels montrent que des perfectionnements doivent être apportés, surtout concernant la mine de Stripa. *Doughty* [Doughty et al., 1994] suggère de combiner des données hydrauliques issues de plusieurs pompages voire d'intégrer des données de flux. Elle suggère également d'améliorer les modèles résultant en améliorant les performances de l'algorithme d'inversion.

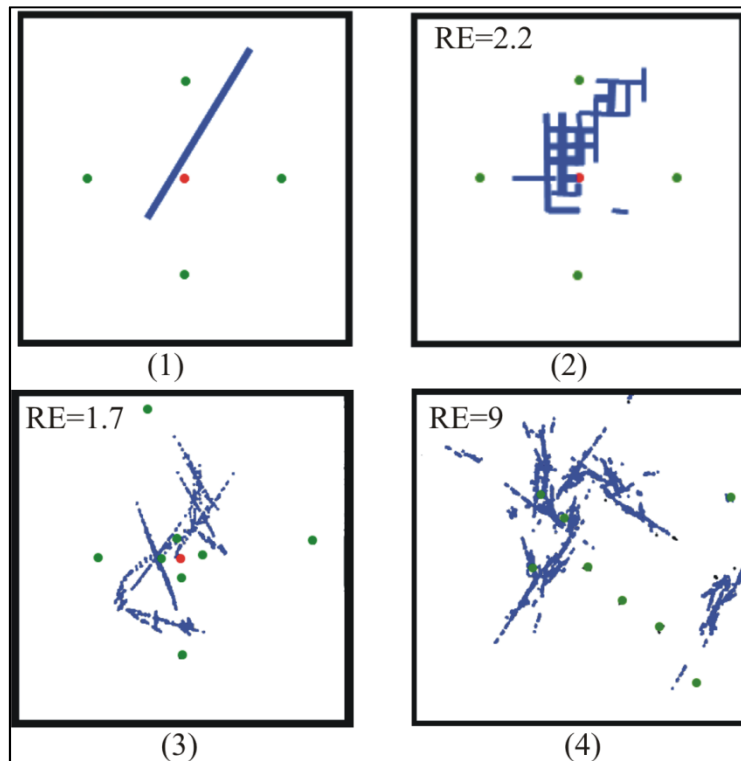


figure 1-18 : Problème inverse utilisant une description fractale du milieu. (1) Cas synthétique de référence, (2) modèle résultat pour le cas synthétique (moyenne de 30 réalisations), (3) modèle résultat du réservoir de Kesterson et (4) modèle résultat de la mine de Stripa. La valeur de RE correspond à la concordance entre les charges simulées et les charges observées. Le point rouge correspond au point de pompage et les points verts aux puits d'observation.

4. Conclusion vis-à-vis des milieux très chenalisés

Lorsqu'on veut résoudre le problème inverse dans les milieux hétérogènes, poreux et fracturés, l'approche la plus courante est de s'affranchir de la géométrie des structures majeures soit par l'utilisation d'un modèle continu, soit en fixant la géométrie à partir d'observations géologiques. Dans ce cas, le problème inverse est très proche du problème inverse classique pour les milieux poreux, où l'on cherche à calibrer les transmissivités à partir de données hydrauliques. Cette approche montre de très bons résultats à conditions que les écoulements soient suffisamment répartis pour que leur interpolation soit efficace ou si la structure d'une éventuelle chenalisation est connue en amont. Dans le cas contraire, le problème inverse consiste à identifier à la fois la transmissivité et la géométrie des structures dominant les écoulements. Les travaux de *Tiedeman* [*Tiedeman et al.*, 1995], de *Renshaw* [*Renshaw*, 1996] et de *Doughty* [*Doughty et al.*, 1994] montrent qu'il est possible de résoudre un tel problème, même si à l'heure actuelle le nombre de paramètres utilisés pour décrire le milieu est resté très limité. Le tableau 1-1 résume les différents travaux présentés au paragraphe précédent.

Auteurs	Type de modèle	Données	Paramètres	Algorithme
<i>Tsai et al.</i> [Tsai et al., 2003b] <i>Sun and Yeh</i> [Sun and Yeh, 2007a; b]	Zones homogènes.	Charges hydrauliques et concentrations (transitoire – 640 valeurs).	Centre et perméabilité de chaque zone (max. 15 paramètres).	Algorithme génétique et méthode de gradient
<i>Tsai et al.</i> [Tsai et al., 2005]	Milieu continu défini par interpolation (voisins naturels).	Charges hydrauliques (transitoire – 70 valeurs) et transmissivités locales (9 valeurs).	Position et perméabilité de chaque point (max. 15 paramètres).	Algorithme génétique et méthode de gradient
<i>Alcolea et al.</i> [Alcolea et al., 2006] <i>de Marsily et al.</i> [de Marsily et al., 1984; LaVenue et al., 1995]	Champ géostatistique de perméabilité continu (deux krigeages)	Charges hydrauliques (transitoire – 936 valeurs) et transmissivités locales (13 valeurs)	Position et perméabilité des points pilotes (jusqu'à 241 points).	Méthode de gradient
<i>Zimmerman et al.</i> [Zimmerman et al., 1998]	Champ géostatistique de transmissivité	Charges hydrauliques (transitoires – 32 valeurs) et transmissivités locales (41 valeurs)	Divers paramètres de la loi géostatistiques	Méthodes géostatistiques (cokrigeage, self-calibration, points pilotes, etc.)
<i>Larocque et al.</i> [Larocque et al., 1999]	Milieu continu défini par interpolation (krigeage)	Charges hydrauliques (transitoires – 70 à 117 valeurs)	Transmissivité de points fixés (64 à 256)	Méthode de gradient
<i>Franssen et Gómez-Hernández</i> [Franssen and Gómez-Hernández, 2002] <i>Rodrigo et al.</i> [Rodrigo et al., 2002]	Plans de fracture dans un milieu continu	Charges hydrauliques (transitoires – 51 valeurs) et valeurs de transmissivité (~300 valeurs)	Propriétés statistiques de ln(K) par fracture et pour la matrice (22 paramètres)	Méthodes géostatistiques (self calibration, conditionnal probabilities)
<i>Nakao et al.</i> [Nakao et al., 2000]	Treillis régulier dont la transmissivité est définie par cluster	Charges hydrauliques (transitoire – huit puits d'observation)	Centre et transmissivité de chaque cluster (~100 paramètres)	Recuit simulé
<i>Gwo</i> [Gwo, 2001]	Treillis régulier où chaque segment est défini comme appartenant à la matrice	Concentration (transitoire – 1 à 3 courbes de restitution)	Type de chaque segment (2 possibilités) et classe de transmissivité (4 possibilités) – ~70 paramètres	Algorithme génétique
<i>Donado</i> [Donado et al., 2005] <i>Bruyelle and Lange</i> [Bruyelle and Lange, 2009]	Réseau de fracture	Charges hydrauliques et concentrations (transitoire)	Perméabilité et emmagasinement de 2 à 5 familles de fracture (max. 10 paramètres)	Maximum Likelihood + méthode de gradient ou algorithme évolutionnaire (CMA-ES)
<i>Tiedeman et al.</i> [Tiedeman et al., 1995] <i>Renshaw</i> [Renshaw, 1996]	Une ou deux zones fracturées dans un milieu faiblement hétérogène	Charges (stationnaires – 8 valeurs)	Position, longueur et orientation des zones fracturées (4 à 8 paramètres)	Recuit simulé ou cokrigeage
<i>Day-Lewis et al.</i> [Day-Lewis et al., 2000]	Grille régulière avec des cellules correspondant à la matrice et d'autres au milieu fracturé	Propriété du champ de perméabilité (variogramme, connectivité) et charges hydrauliques (transitoires – issues de 13 puits)	Type de chaque cellule puis transmissivités des zones identifiées	Recuit simulé (structure) et méthode de gradient (transmissivités)
<i>Doughty et al.</i> [Doughty et al., 1994]	Fonctions itératives définissant une géométrie fractale	Charges hydrauliques transitoires – issues de 5 à 9 puits)	Paramètres des fonctions IFS (4 à 24 paramètres)	Recuit simulé

tableau 1-1 : Synthèse des différentes méthodes de résolution du problème inverse décrite au paragraphe 2.

C. Positionnement des travaux effectués par rapport aux travaux précédents

La problématique de la thèse porte sur la caractérisation des structures d'écoulement dans les milieux où le flux est fortement chenalisé. L'étude des travaux menés au cours de ces dernières années montre qu'un gros effort a été porté sur la paramétrisation du milieu naturel. Cette paramétrisation doit en effet tenir compte de l'hétérogénéité spatiale de la perméabilité et de ses caractéristiques, notamment de connectivité, ou tout du moins être capable de bien la modéliser. Lorsque les flux sont fortement chenalisés, la compréhension des propriétés d'écoulement et de leur lien avec la perméabilité sous-jacente devient en effet critique. On s'oriente donc de plus en plus vers des modèles intégrant les structures majeures d'écoulement soit de manière déterministe, soit par l'utilisation d'indicateurs géostatistiques sur ces structures [Renard, 2007]. Nous proposons ainsi deux nouveaux indicateurs mesurant le degré de chenalisation des milieux hétérogènes. Les travaux présentés ici concernent essentiellement les milieux fracturés. Les écoulements dans les milieux fracturés sont plus enclins à être chenalisés de par la présence même des fractures et leur organisation. C'est pourquoi ces indicateurs doivent être applicables à la fois aux milieux poreux et aux milieux fracturés. Évaluer le degré de chenalisation permet également, le cas échéant, d'adapter la paramétrisation du milieu afin de mieux intégrer les propriétés d'écoulement.

Nous avons, au cours de ces travaux, privilégié les modèles discrets (DFN) dans la mesure où, selon nous, ils sont mieux à même de décrire la structure géologique d'un milieu fracturé tout en tenant compte des effets d'échelle. Pour être efficaces, ces modèles doivent intégrer le maximum d'information sur la géologie des fractures. Les propriétés de la fracturation sont déduites à partir de données de terrain et les trois principales difficultés viennent des biais dans les échantillonnages, de l'incertitude associée à chaque donnée et surtout de la variabilité naturelle de la fracturation, variabilité spatiale et variabilité d'échelle. Nous avons porté une attention toute particulière à ces difficultés lors de la caractérisation de la fracturation d'un site naturel afin que les propriétés hydrauliques déduites de cette caractérisation soit les plus proches possible des propriétés naturelles.

Finalement, définir correctement la structure du milieu est une condition nécessaire mais pas suffisante pour obtenir des modèles prédictifs fiables. La meilleure solution, à l'heure actuelle, restent la résolution d'un problème inverse où le modèle de perméabilité est adapté en fonction des propriétés hydrauliques mesurées sur site. Le problème inverse en hydrogéologie est largement appliqué aux milieux poreux en supposant une certaine continuité de la perméabilité, mais certains travaux ont montré qu'il était possible d'identifier les principales structures porteuses des écoulements. Nous avons donc continué dans cette voie afin d'améliorer les capacités d'identification de ces structures, toujours en utilisant une approche discrète. Nous aborderons le problème inverse complet, où la structure et les propriétés sont inconnues et proposerons une méthode de résolution basée sur l'identification des structures les plus importantes en termes de propriétés hydrauliques.

Les modèles numériques sont de plus en plus utilisés pour comprendre et définir les propriétés physiques et hydrauliques des milieux naturels. Ils permettent en effet, à partir d'une paramétrisation du milieu, de simuler un grand nombre de phénomènes à moindre coût, de l'écoulement le plus simple au transport réactif de polluants au cours d'un pompage. L'un des objectifs de la thèse a donc été de mettre au point des méthodes numériques qui, à partir d'informations sur le milieu, permettent de générer un modèle dont la structure de perméabilité correspond aux observations. Ces méthodes numériques doivent également permettre de calculer les propriétés hydrauliques des modèles générés et les résultats obtenus doivent pouvoir être utilisés par une méthode chargée de résoudre le problème inverse.

La qualité d'un outil de modélisation repose sur trois critères :

- Les méthodes développées doivent avoir de bonnes performances en termes de temps de calcul et d'espace mémoire afin de permettre la modélisation de milieux complexes ou sur plusieurs échelles.
- Ses fonctionnalités doivent être suffisamment souples pour intégrer un grand nombre de paramètres extérieurs comme les propriétés du milieu, le type de conditions aux limites, la précision demandée, etc., afin de pouvoir explorer un maximum de configurations.
- Le code numérique doit rester clair, adaptable et extensible, tout d'abord parce que sa conception et sa mise en œuvre fait intervenir un grand nombre d'acteurs aux compétences différentes et ensuite parce que l'investissement réalisé par le développement de méthodes efficaces doit être capitalisé et pérennisé.

Les différents travaux présentés dans cet ouvrage utilisent des méthodes numériques pour modéliser le milieu naturel (poreux ou fracturé), dériver ses propriétés hydrauliques et calculer les résultats présentés. C'est pourquoi cette partie est consacrée à la présentation de la plateforme logicielle H2OLAB qui constitue le cadre des fonctionnalités élaborées durant la thèse et a permis d'obtenir les résultats présentés dans les autres parties de ce document. Nous n'exposerons en détail que les fonctionnalités réalisées pour ou en compléments du doctorat après avoir décrit de manière générale les applications de la plateforme.

A. Présentation de la plateforme logicielle

La plateforme logicielle H2OLAB est le fruit d'une collaboration entre le laboratoire de Géosciences Rennes, l'IRISA de Rennes, le CDCSP de Lyon et le laboratoire LMPG du Havre. Elle est consacrée à l'étude du flux et du transport dans les milieux poreux hétérogènes et les réseaux de fractures via la réalisation de simulations stochastiques. Elle est composée d'un ensemble de classes et de fonctionnalités développées en C++ respectant les pré-requis décrits au paragraphe précédent. Ces classes sont regroupées selon leur degré de généralité et leur champ d'application en plusieurs projets qui sont développés par les différents acteurs. Ces projets peuvent être regroupés thématiquement de la manière suivante :

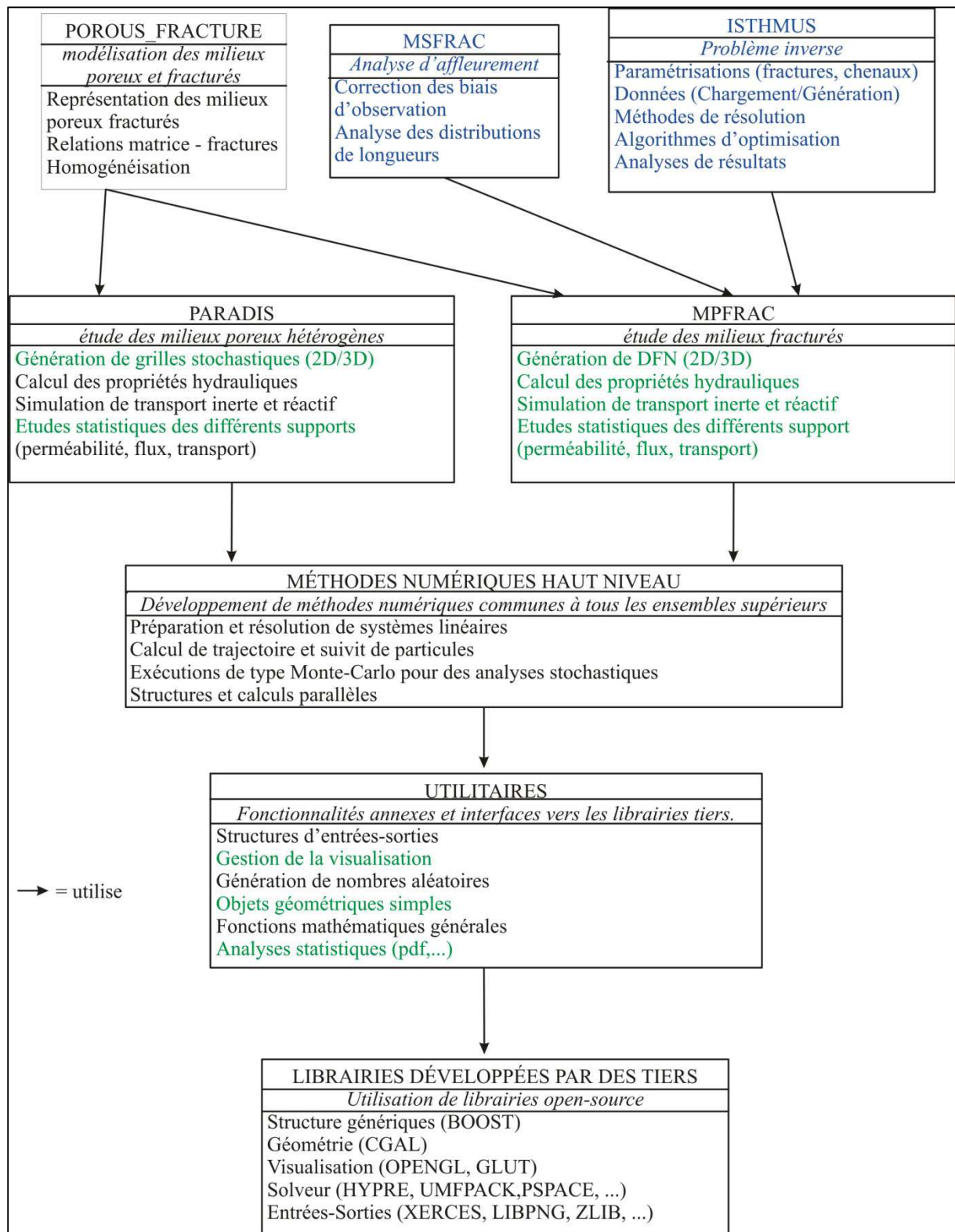


figure 2-1 : Présentation des différents ensembles présents dans H2OLAB. La couleur bleu correspond aux fonctionnalités qui ont été entièrement développées dans le cadre de la thèse et la couleur vertes aux fonctionnalités modifiées ou complétées durant la thèse.

- Un premier ensemble concerne l'étude des milieux poreux hétérogènes, modélisés par une grille hétérogène. Il contient un module de génération, un module de calcul du flux et un modèle de simulation de transport.
- Un second ensemble concerne l'étude des milieux fracturés, modélisés par un réseau de fracture. Il contient les mêmes modules que pour l'ensemble précédent.
- Un troisième ensemble est consacré aux développements de méthodes numériques communes aux deux ensembles précédant comme le suivi de trajectoires de particules, la résolution de systèmes linéaires, la parallélisation du code ou les méthodes de Monte-Carlo.
- Un quatrième ensemble contient de nombreux utilitaires communs aux différents ensembles. Il contient par exemple les outils de visualisations, d'entrées / sorties, de génération de nombres aléatoires, de mathématiques générales, etc.
- Un dernier ensemble contient l'ensemble des bibliothèques développées par des tiers et utilisées dans la plateforme. On y trouve par exemple une bibliothèque permettant la gestion d'objets géométriques, les solveurs hautes performances, les bibliothèques pour afficher à l'écran ou celles contenant des structures génériques comme les tableaux à plusieurs dimensions, les matrices, etc.

A côté des ces applications qui permettent l'analyse des milieux poreux et fracturés on trouve des applications plus spécifiques traitant un problème particulier. C'est notamment le cas du problème inverse, présenté en partie 5, de l'analyse d'affleurements, présentée en partie 3, ou de l'étude des milieux poreux fracturés. L'organisation de ces différents ensembles est présentée à la figure 2-1. En bleu sont spécifiées les parties entièrement développées dans le cadre des travaux de doctorat et en vert les fonctionnalités modifiées ou complétées durant la thèse.

1. Description de PARADIS

PARADIS (pour PARallel DISpersion) contient les outils d'analyse des milieux poreux hétérogènes 2D et 3D. Il est constitué d'un module de génération géostatistique permettant de générer aléatoirement des champs de perméabilité corrélés ou non (corrélation gaussienne, log-normale ou en loi de puissance) et des champs aux propriétés fractales. Il permet également de simuler le flux en stationnaire ou en transitoire et de simuler le transport inerte ou réactif de particules sur des domaines réguliers en 2D ou 3D. La méthode utilise les volumes finis pour discrétiser les équations de flux et se sert du module de résolution de systèmes linéaires pour calculer q et h en fonction de k et des conditions aux limites. Il permet notamment d'analyser les corrélations entre le champ de vitesse et le champ de perméabilité [de Dreuzy and Davy, 2007; Le Borgne et al., 2007b] et possède également un module de calcul statistique. Il utilise le module de parallélisation pour calculer la dispersion macroscopique et analyser le comportement asymptotique du transport de particule sur de grands domaines. Le transport y est simulé par advection-diffusion grâce à une méthode de "particle tracking" [Beaudoin et al., 2007; de Dreuzy et al., 2007]. La figure 2-2 présente un exemple de ces différentes fonctionnalités. Aux fonctionnalités déjà existantes nous avons rajouté la génération de champs de perméabilités dont les fortes ou les faibles valeurs de k sont corrélées selon la méthode de Zinn et Harvey [Zinn and Harvey, 2003] ou l'ajout de structures de fortes perméabilités dans le champ géostatistique selon la méthode Knudby et Carrera [Knudby and Carrera, 2005]. Nous avons également ajouté le calcul d'un certain nombre d'indicateurs statistiques concernant la chenalisation comme ceux décrits à la partie 4.

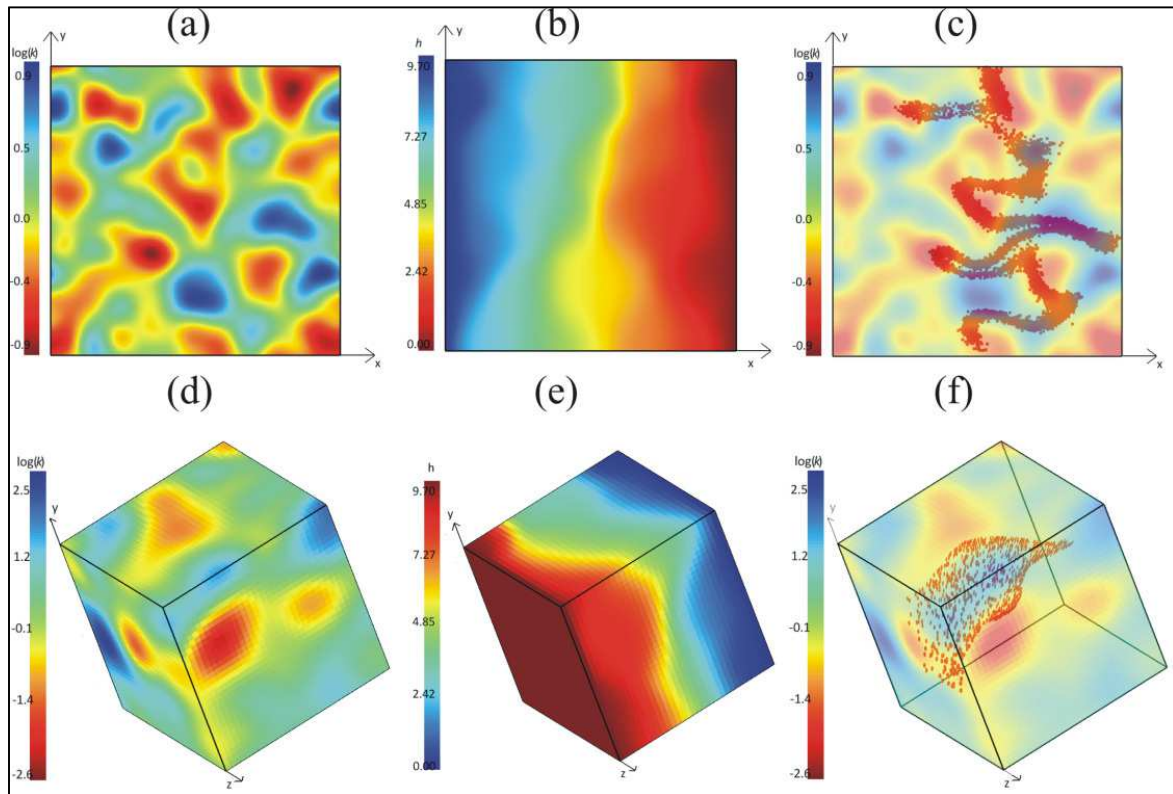


figure 2-2 : Illustrations des différentes fonctionnalités de PARADIS. (a) et (d) Génération de champs de perméabilité 2D (a) et 3D (d) selon une loi géostatistique. Les champs présentés ici suivent une distribution log-normale ($\mu=0$, $\sigma^2=1$ pour le 2D et $\sigma^2=3$ pour le 3D), de corrélation Gaussienne avec une longueur de corrélation $\lambda=L/10$ où L est la taille du système. (b) et (f) charges correspondantes pour des conditions aux limites de type perméamètre avec $h_{\max}=10$ et $h_{\min}=0$. (c) et (f) simulation du transport de particules (en rouge) en advection-diffusion par une méthode de "particle tracking" (ici avec 10000 particules).

2. Description de MPFRAC

Le logiciel MPFRAC (pour Massive Parallel FRACTure networks) est consacré à la modélisation des milieux fracturés sous forme de réseaux fractures. Bien qu'une version préliminaire existait en amont des travaux présentés dans ce mémoire, cette partie a été entièrement repensée et recodée afin de remplir les objectifs de performances et de généricité nécessaires à son intégration dans la plateforme H2OLAB. Pour cela, une nouvelle structure modélisant le milieu géologique et le réseau de fractures a été conçue de manière à factoriser les algorithmes de base indépendamment de la géométrie (2D/3D, forme des fractures, etc.). MPFRAC contient :

- un noyau générique qui gère la structure physique du réseau de fracture. Il calcule si besoin les intersections entre fractures et la connectivité du réseau. L'ensemble des fonctionnalités est générique et fait référence à des fonctions spécifiques uniquement pour les opérations géométriques. Il s'appuie pour cela sur la bibliothèque CGAL [Lutz Kettner and Näher, 2004].
- un module de génération. Ce module permet de générer de manière stochastique les paramètres des fractures comme leurs positions, leurs orientations, leurs longueurs ou leurs transmissivité.
- un module d'analyse statistique sur la structure du réseau de fracture (distributions, connectivité, etc.).

- un module de calcul d'écoulement contenant une interface générique, le calcul du flux en 2D par une méthode de discrétisation aux intersections et le calcul du flux en 3D par maillage du réseau de fracture et discrétisation par la méthode des éléments finis mixtes hybrides [Erhel *et al.*, 2009; Pichot *et al.*, 2009].
- un module de calcul et suivi du parcours de particules dans un milieu fracturé 2D.

Les fonctionnalités développées dans le cadre du doctorat sont présentées en détail au paragraphe B. La figure 2-3 illustre les principales fonctionnalités de MPFRAC.

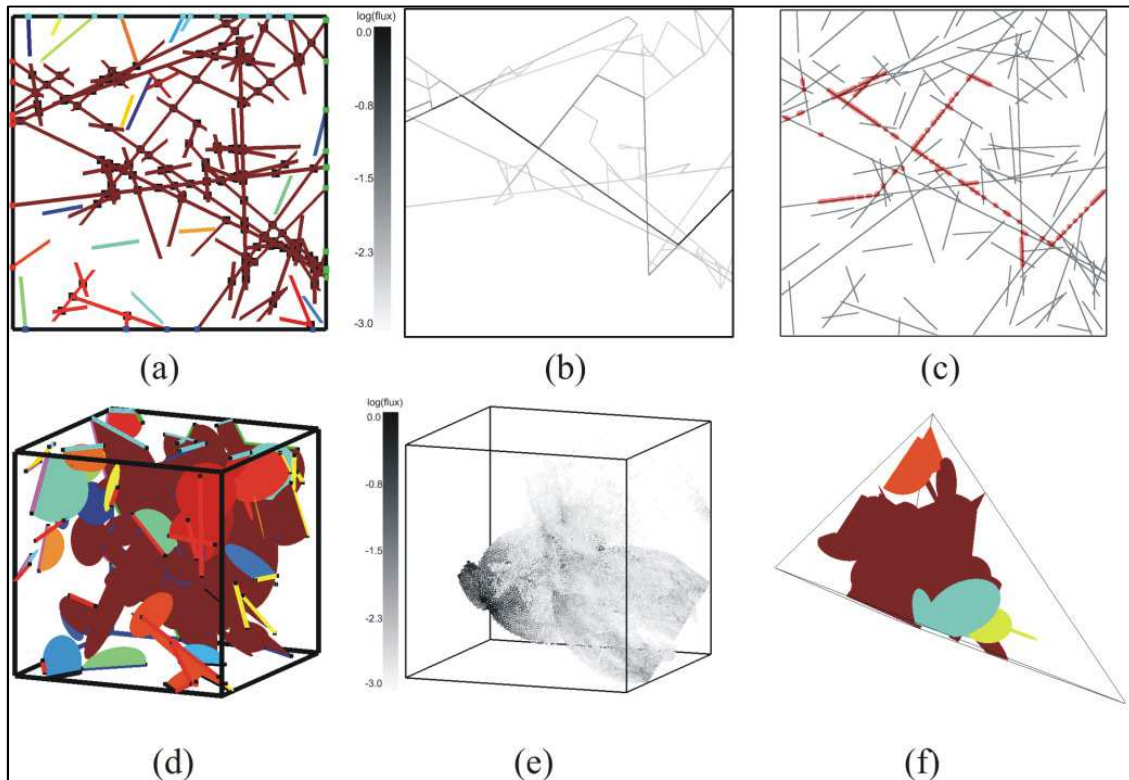


figure 2-3 : Illustrations des différentes fonctionnalités de MPFRAC en 2D et 3D. (a) et (d) Réseau de fracture et connectivité, les fractures sont colorées par amas connectés et les intersections sont soulignées (points en 2D, segments en 3D). (b) et (e) Flux correspondant. (c) Transport de particules dans le réseau 2D. (f) Réseau de fracture avec une forme non carrée.

3. Les fonctionnalités de la plateforme communes aux différentes applications

L'intérêt de développer une plateforme logicielle permet de factoriser les fonctionnalités utiles aux différentes applications. Par exemples, les formats d'entrées-sorties sont communs à toutes les applications de la plateforme. Ils sont stockés sous format XML précisant, pour chaque paramètre d'entrée ou de sortie, son type (entier, booléen, flottant), une valeur par défaut et sa description. Une autre fonctionnalité pouvant être utilisées pour toutes les applications est l'exécution par méthode de Monte-Carlo d'un grand nombre de simulations afin d'analyser de manière stochastique les relations entre paramètres. Ces exécutions peuvent être réalisées en série ou en parallèle grâce à l'implémentation d'un module générique basé sur la librairie MPI [Gropp *et al.*, 1999]. De plus, la plupart des applications génèrent un système linéaire. La plateforme fournit une représentation générique de ce système et contient une interface permettant sa résolution avec de nombreux solveurs (principalement UMFPACK [Davis, 2004] et HYPRE [Falgout and Yang, 2002]). Ainsi, chaque application suit une procédure similaire exploitant au mieux les fonctionnalités communes de la plateforme. La figure 2-4 donne une illustration d'une exécution d'une des applications de la plateforme. Par

ailleurs, la plateforme contient de nombreux outils permettant de valider les codes développés et de s'assurer de leur pérennité par des tests de non-régression.

4. Les partenaires de la plateforme

Un autre intérêt d'utiliser une plateforme numérique de développement est de pouvoir faire participer simultanément plusieurs partenaires qui améliorent les parties génériques ou spécifiques en fonction de leurs compétences. Ainsi, l'IRISA est chargée de la gestion de la plateforme et du calcul scientifique haute performance (parallélisation, solveurs linéaires), le CDCSP des méthodes mathématiques de discrétisation des équations physiques, Géosciences Rennes de la physique des écoulements et des liens entre les propriétés physiques et les propriétés hydrauliques pendant que le LMPG se concentre sur le transport réactif. Ce développement collaboratif implique la mise en place d'outils de travail commun, comme un serveur subversion (SVN) permettant d'éviter les conflits lors du développement et un espace collaboratif contenant notamment un forum de discussion et un planificateur de tâche.

B. Principes et fonctionnements de fonctionnalités développées pour le doctorat

Cette partie a pour objectif de préciser les fonctionnalités développées dans le cadre du travail de thèse.

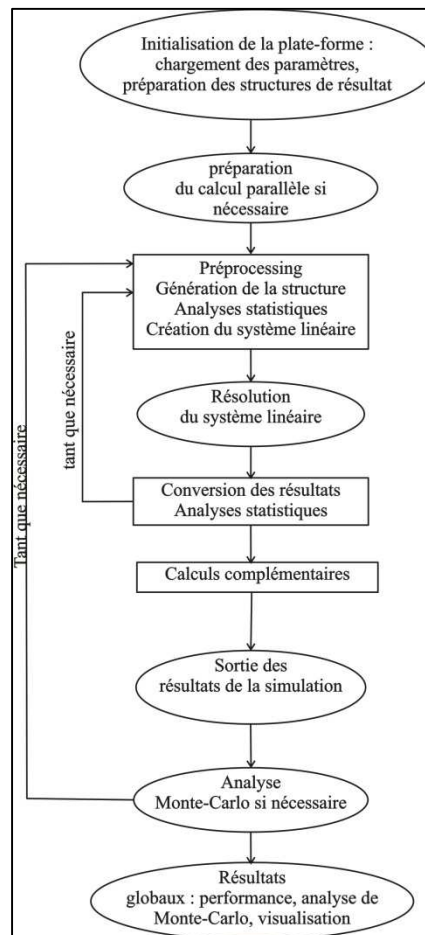


figure 2-4 : Schéma d'exécution d'une application de la plateforme H2OLAB. En ovale sont spécifiées les fonctionnalités génériques et en rectangle les fonctionnalités spécifiques à chaque application.

1. Modèles de réseaux de fractures

La modélisation des milieux fracturés par DFN consiste à représenter l'ensemble des fractures du milieu. Dans la mesure où connaître l'intégralité de la fracturation est impossible, le modèle repose sur une description statistique des propriétés des fractures, c'est-à-dire les longueurs, orientations, positions et ouvertures. MPFRAC est capable de générer un ensemble de fractures respectant ces propriétés et de les ajouter dans un réseau de fracture. Il est également capable de séparer le milieu en plusieurs sous-parties afin de permettre un traitement en parallèle ou d'analyser séparément différentes parties du réseau de fracture. Il possède des fonctionnalités génériques, c'est-à-dire indépendantes de la géométrie des fractures et de la dimension euclidienne, permettant de calculer la connectivité du réseau. Il est composé d'un ensemble de classes C++ dont l'organisation est illustrée à la figure 2-5.

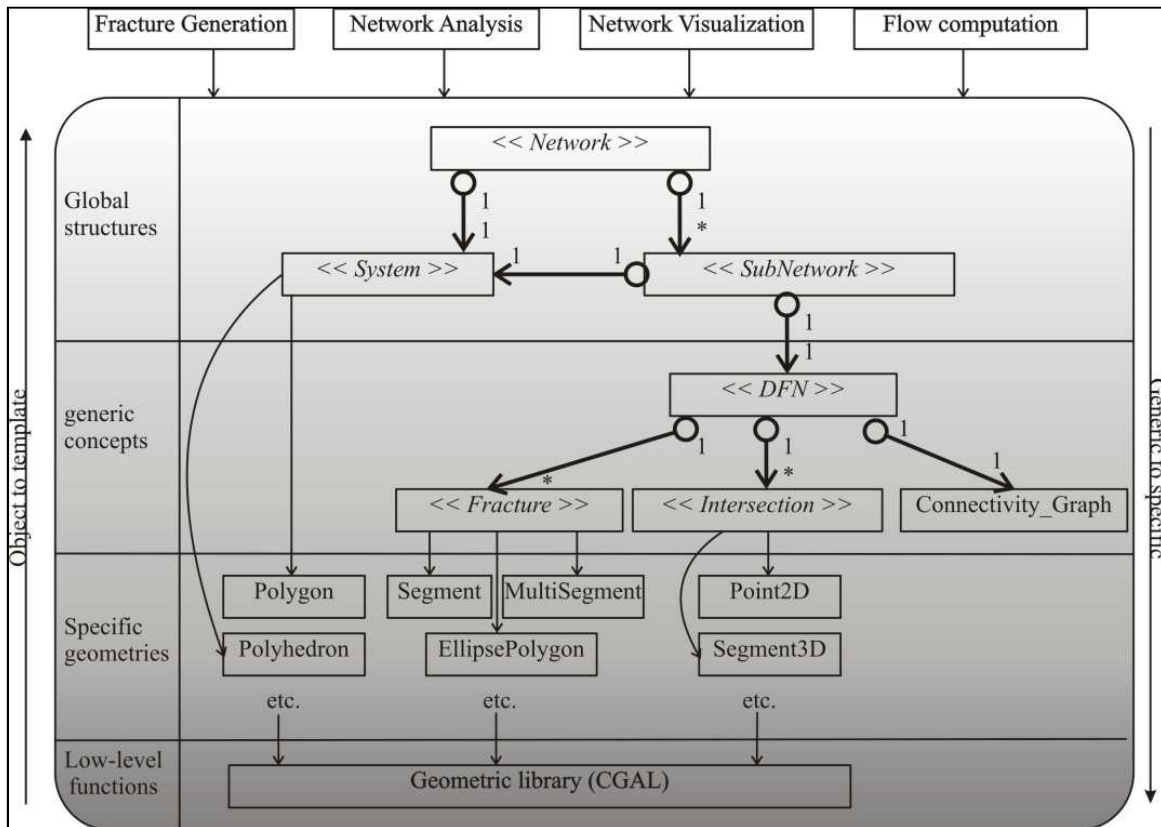


figure 2-5 : Organisation du noyau de MPFRAC. Le réseau de fracture est manipulé à partir des structures globales (réseau, sous-réseau, système géologique) qui mettent à jour les objets génériques (DFN, fractures et intersections). En fonction de la représentation choisie (2D/3D, géométrie des fractures), ces objets génériques sont modélisés par des objets géométriques spécifiques à partir desquels les fonctionnalités de bases sont accessibles (création, intersection,...)

Le réseau de fracture est représenté par une structure globale comprenant un ensemble de sous-réseaux et un système géologique. Un système géologique est composé d'une géométrie et de propriétés spécifiques pendant que le réseau permet d'accéder aux éléments inférieurs et de gérer les interactions entre sous-réseaux. Chaque sous-réseau contient un DFN et peut interagir avec ses voisins. Le DFN contient les fractures et la structure de connectivité, c'est-à-dire les intersections entre fractures et le graphe de connectivité permettant de calculer les amas connectés. Chaque élément générique (système, fracture, intersection) est modélisé par une classe-patron héritant de l'objet géométrique le représentant (polygone, polyèdre, segment, etc.).

Ainsi, lors du calcul des amas connectés par exemple, le réseau global demande à chaque sous-réseau de calculer ses amas propres à partir du DFN puis réunit ces amas en fonction de la connectivité entre sous-réseau. Cette méthode est complètement générique et seules les fonctions calculant réellement l'intersection entre objets (intersection fracture-fracture et intersection fracture-système) sont spécifiques au choix de la représentation. Le réseau globale est également utilisé par un générateur stochastique, qui ajoute au réseau les fractures générées, un analyseur, qui récupère et analyse les propriétés des fractures et des intersections, et par le module de calcul de flux.

2. Simulation d'écoulements et de transport inerte dans un DFN 2D

Cette partie décrit le calcul du flux dans un milieu fracturé 2D. Le flux dans un réseau en trois dimensions est calculé différemment puisqu'il nécessite de mailler le réseau de fracture et de discrétiser les équations de flux selon un schéma numérique particulier, ce qui constitue un sujet de recherche en soi. La première étape consiste à discrétiser le réseau de fracture en un réseau de lien. Pour cela, seul le ou les amas connectés aux conditions aux limites sont conservés. Chaque point du réseau (extrémité, intersection) est ensuite numéroté et relié à ses voisins grâce au graphe de connectivité. L'équation de conservation du flux est alors appliquée à chaque point en intégrant les conditions aux limites selon le principe des volumes finis. Cette équation assure que le flux entrant de chaque nœud est égal au flux sortant. Cette équation, appliquée à un réseau de N_n nœuds, aboutit à la création d'un système linéaire de la forme $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$ où \mathbf{A} est une matrice carré de dimension $N_n \times N_n$, \mathbf{x} le vecteur des variables de dimension N_n et \mathbf{b} le terme source dépendant des conditions aux limites de dimension N_n . Ce système linéaire est finalement résolu par les outils de la plateforme H2OLAB, ce qui permet de déterminer q puis h . La figure 2-6 illustre ce processus.

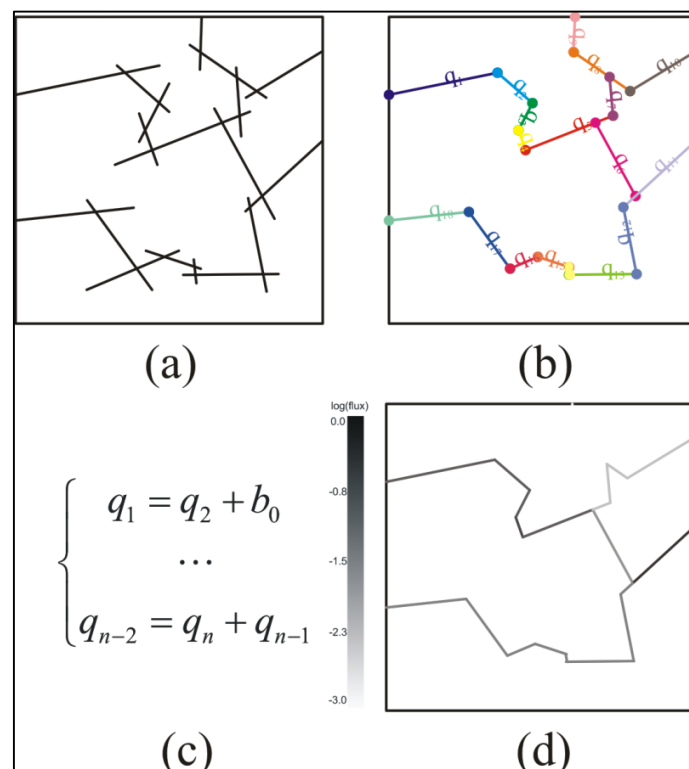


figure 2-6 : Illustration du calcul de flux sur un réseau de fracture 2D. (a) réseau de fractures, (b) réseau de liens, chaque lien a une valeur de flux. (c) système linéaire obtenu en appliquant la loi de conservation du flux ($q_{\text{entrant}}=q_{\text{sortant}}$) et les conditions aux limites. (d) flux avec des conditions aux limites de type perméamètre.

A partir du flux il est possible de simuler le transport de particule par une méthode de "particle tracking". Comme la matrice est imperméable, la particule ne peut se déplacer que dans le réseau de fracture. Le mouvement d'une particule dépend de l'advection, liée à la vitesse d'écoulement du fluide, de la diffusion, liée au gradient de concentration, et de la dispersion, liée aux mélanges locaux des fluides. On ne considère, dans le cas des réseaux de fracture, que l'advection. Le déplacement d'une particule ne dépend alors que de la vitesse du fluide, proportionnelle au débit. La difficulté se situe au niveau des intersections entre fracture. La répartition des particules dépend de l'organisation des débits (figure 2-7) et de la méthode de mélange choisie, "complete mixing" ou "streamline routing" [Hull and Koslow, 1986; Berkowitz et al., 1994].

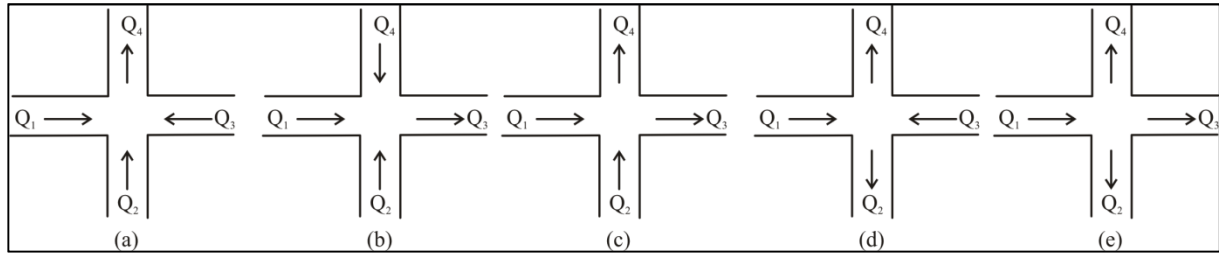


figure 2-7 : Configurations possibles lors de la simulation du transport de particules. Arrivée à une intersection en Q_1 , une particule se dirigera vers Q_2 , Q_3 ou Q_4 en fonction du sens des différents débits et d'une probabilité dépendant de la méthode choisie entre "complete mixing" et "streamline routing".

Lorsqu'une seule voie est possible (figure 2-7-a ou figure 2-7-b), alors la particule continuera naturellement dans cette voie avec une probabilité $p=1$. Lorsque les trois voies sont possibles (figure 2-7-e), alors on applique la méthode de "complete mixing" et la probabilité de chaque voie dépend de son flux. On obtient alors :

$$\begin{cases} p_{12} = \frac{Q_2}{(Q_2 + Q_3 + Q_4)} \\ p_{13} = \frac{Q_3}{(Q_2 + Q_3 + Q_4)} \\ p_{14} = \frac{Q_4}{(Q_2 + Q_3 + Q_4)} \end{cases} \quad (2-1)$$

où p_{ij} est la probabilité que la particule, venant de Q_i , choisissent la voie Q_j . Les choses sont plus compliquées en ce qui concerne les cas illustrés à la figure 2-7-c et à la figure 2-7-d. Dans le cas de la figure 2-7-d, les deux voies possibles obligent la particule à bifurquer et ont donc une importance similaire. C'est pourquoi on applique également la méthode de "complete mixing" :

$$\begin{cases} p_{12} = \frac{Q_2}{(Q_2 + Q_4)} \\ p_{14} = \frac{Q_4}{(Q_2 + Q_4)} \end{cases} \quad (2-2)$$

Par contre, à la figure 2-7-c, la particule a le choix entre la voie qui lui fait face et une bifurcation. Dans ce cas, la méthode de "streamline routing" spécifie que la voie la plus proche, donc la bifurcation, est prioritaire si son débit est supérieur. En effet, dans ce cas, la particule est "happée" par le courant se dirigeant vers cette voie sans avoir le temps de passer dans la voie de face. Si par contre son débit est inférieur, alors la particule peut avoir le temps d'atteindre la voie qui lui fait face et les deux voies ont une probabilité non nulle. En

appliquant la méthode de "streamline routing", on obtient les probabilités de transition suivantes :

$$\left\{ \begin{array}{l} \left\{ \begin{array}{l} p_{13} = 0 \\ p_{14} = 1 \end{array} \right. \quad \text{si } Q_4 > Q_3 \\ \left\{ \begin{array}{l} p_{13} = \frac{Q_3 - Q_4}{Q_3} \\ p_{14} = \frac{Q_4}{Q_3} \end{array} \right. \quad \text{si } Q_3 > Q_4 \end{array} \right. \quad (2-3)$$

Ainsi, à partir des valeurs de flux de chaque segment, on est capable de simuler le transport de particules à l'intérieur d'un réseau de fracture en deux dimensions.

3. Analyse d'affleurement

L'application servant à l'analyse d'affleurements, MSFRAC (Multi-Segments FRACture networks), dont le principe est donné en partie 3, utilise également la plateforme H2OLAB. C'est une application différente de MPFRAC puisqu'il s'agit, dans ce cas, d'intégrer dans le DFN des observations de terrain, et notamment les cartes de traces de fracture. Pour cela, MSFRAC utilise le noyau de MPFRAC pour gérer les réseaux de fractures mais possède son propre module permettant de convertir les cartes d'affleurement en ensemble de fractures. Il possède également un module d'analyse permettant de corriger les données brutes des biais d'observation et un module de statistiques spécifique permettant d'analyser les distributions de longueurs sur les données corrigées. La figure 2-8 résume ces fonctionnalités.

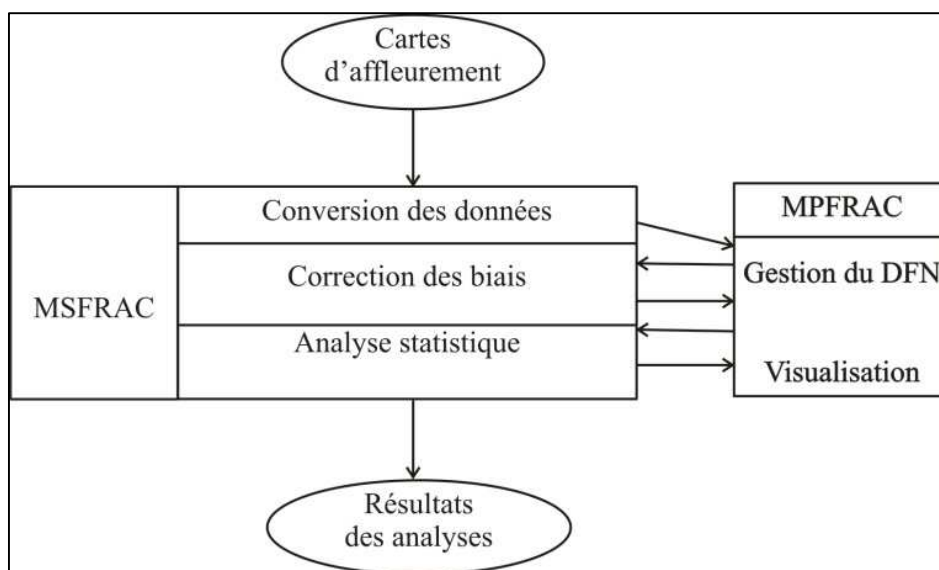


figure 2-8 : Principe des fonctionnalités de MSFRAC et coopération avec MPFRAC.

4. Problème inverse

MPFRAC permet la résolution du problème direct, c'est-à-dire la détermination des propriétés hydrauliques à partir de la structure physique du milieu. L'un des objectifs du doctorat a été le développement d'une application permettant de résoudre le problème inverse, c'est-à-dire la détermination de la structure physique à partir des propriétés hydrauliques. Cette application, baptisée ISTHMUS (Inverse Solver Technique for Heterogeneous Media Using Simulations), a été intégrée à la plateforme H2OLAB à la fois pour bénéficier des outils génériques contenus dans la plateforme et également parce que la résolution du problème inverse passe par un grand nombre de résolutions du problème direct. Cette application a été développée en

respectant les principes de la plateforme, notamment en ce qui concerne la généricité et les possibilités d'extension. La figure 2-9 montre le schéma structural d'ISTHMUS.

L'application contient un ensemble de classes génériques permettant de résoudre le problème inverse en hydrogéologie. Elle contient une interface représentant les données utilisées et une interface pour la paramétrisation. La valeur des paramètres est optimisée via la classe *Inverse_Solution* à partir d'une ou plusieurs méthodes d'inversion. Ces méthodes sont gérées par la classe *Inverse_Method* et le problème direct est résolu par la classe *Direct_Solver*. Les fonctionnalités des classes génériques sont spécifiées pour chaque problème en fonction de la méthode de résolution choisie. Les données peuvent être de différents types (charges en stationnaire, en transitoire, temps de transport, etc.). De plus, les paramètres dépendent du type de paramétrisation (réseau de fracture, réseau de chenaux voire grille régulière). Le solveur du problème est implémenté de manière à calculer les mêmes données que les données utilisées, ce qui dans notre cas se traduit par une interface permettant d'utiliser MPFRAC ou PARADIS. La méthode d'optimisation peut également être choisie parmi un algorithme de recuit simulé, un algorithme basé sur le gradient de la fonction voire d'une méthode évolutive faisant varier la paramétrisation et utilisant une ou plusieurs méthodes pour optimiser la fonction objectif. Concernant les méthodes de gradient, dans le cas des milieux poreux celui-ci est calculé par la méthode de l'état-adjoint [Le Goc and de Dreuzy, 2007]. Le calcul du gradient a été mis au point dans une étude préalable aux travaux présentés ici et est fourni en annexe. Cette organisation permet une grande souplesse d'utilisation tout en multipliant les possibilités, puisqu'un problème inverse pourra être résolu par plusieurs algorithmes d'inversion sans modifier les autres parties du code. De plus, un analyseur générique permet de stocker et d'analyser différents résultats communs à tous les problèmes inverses comme les valeurs de fonction-objectif ou l'évolution des paramètres au cours de l'inversion.

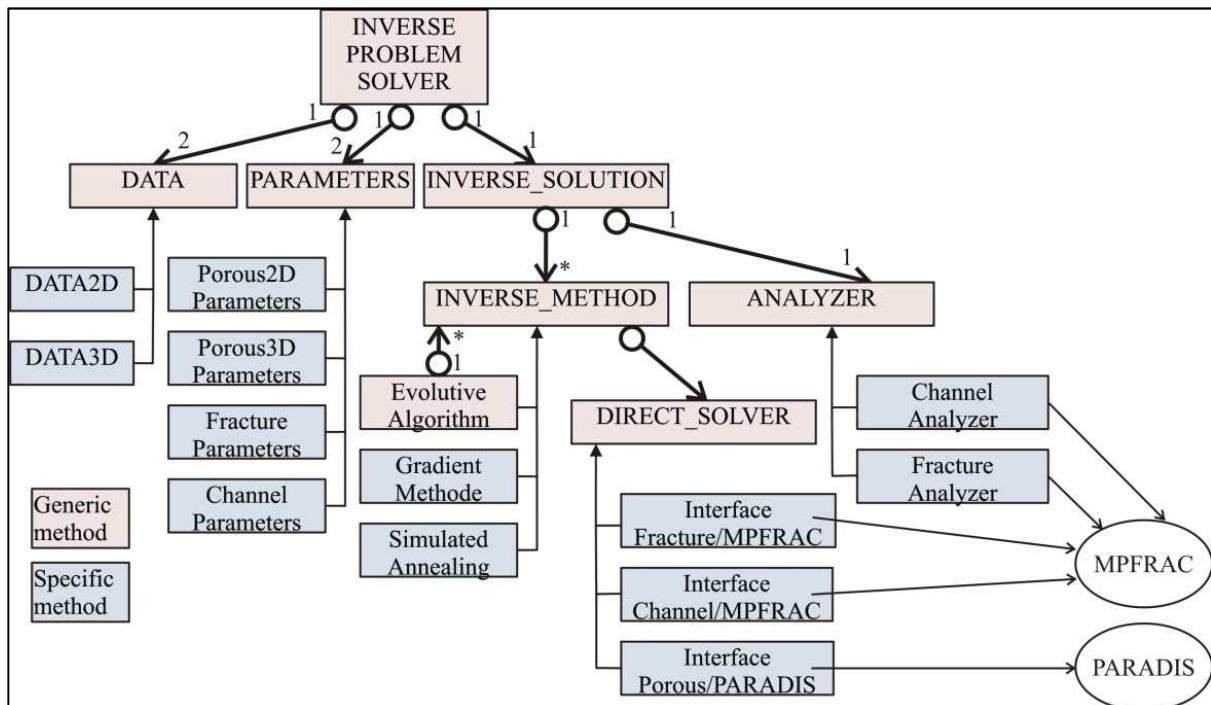


figure 2-9 : Organisation d'ISTHMUS. En rose sont représentées les classes génériques et en bleu les classes plus spécifiques dépendant du type de problème inverse et de la méthode de résolution.

C. Conclusions

Les travaux réalisés pendant le doctorat nécessitent l'utilisation d'outils numériques performants. Le développement de ces outils a donc constitué une partie importante des travaux effectués. L'utilisation d'une plateforme logicielle permet de bénéficier de nombreuses fonctionnalités développées par des intervenants différents et dont les performances, ainsi que les résultats, sont vérifiés. De plus, une telle plateforme permet de capitaliser les travaux des uns et des autres en conservant les méthodes mises au point au sein d'une structure dynamique et évolutive. Il s'agit donc d'un outil de valorisation essentiel qui, au-delà des résultats scientifiques, permet de conserver et de partager les méthodes ayant permis d'obtenir ces résultats.

Partie 3 Définition d'un modèle de fracturation à partir d'observations du milieu naturel : application aux sites de Forsmark et Laxemar (Suède)

Svensk Kärnbränslehantering AB (SKB) est l'agence suédoise chargée de gérer le combustible nucléaire et le stockage des déchets radioactifs. La sélection d'un site pour accueillir un futur centre de stockage constitue un de ses objectifs majeurs. Deux sites ont été pressentis, Forsmark près de Östhammar et Laxemar près d'Oskarshamn, au sud-ouest de la Suède. Le futur site devra accueillir tous les déchets radioactifs des centrales nucléaires suédoises, qui seront déposés dans une installation spécifique située à une profondeur de 500 mètres dans une roche cristalline. Afin de choisir le site présentant le plus de garanties en termes de sécurité, d'intenses investigations ont été menées depuis plusieurs années. Elles visent à recueillir le maximum d'information sur la géologie des deux sites candidats afin de comprendre leurs caractéristiques géologiques et de déterminer un modèle statistique de leur fracturation. Malgré la grande quantité d'observations disponibles, comprenant notamment la cartographie de plusieurs affleurements à différentes échelles et l'analyse de la fracturation le long de dizaines de puits, caractériser les propriétés du milieu demande beaucoup d'efforts à cause de la forte hétérogénéité du milieu naturel, tant au niveau de la structure du réseau de fracture que de la variabilité des propriétés géologiques. La nature multi-échelle de la fracturation constitue une difficulté supplémentaire lorsque l'on cherche à intégrer des informations de natures et d'échelles différentes dans le modèle de site. Cela nécessite donc un cadre théorique et des méthodes de modélisation compatibles avec les particularités des milieux étudiés, notamment en ce qui concerne les changements d'échelle entre l'échelle locale des observations et l'échelle, plus large, du modèle.

A. Introduction

Les travaux présentés ici s'appuient sur la modélisation de la fracturation en réseau de fractures discrètes (DFN). Il s'agit de définir les paramètres du DFN, qui permettent de décrire quantitativement le réseau de fracture, à partir des observations géologiques. Ces paramètres peuvent être séparés en deux ensembles : les paramètres d'échelle, constitués de la distribution des longueurs et de la dimension fractale, et les termes de densité, comprenant la distribution des orientations et la densité de la fracturation. Lorsque l'on cherche à définir ces paramètres, plusieurs questions doivent être abordées :

1. Tout d'abord la définition d'un modèle de site doit permettre de caractériser la fracturation, ce qui implique que le modèle doit être compatible avec les propriétés physiques sous-jacentes et les observations. Elle doit également permettre d'apporter des contraintes pertinentes lors de l'évaluation des risques présentés par l'enfouissement de déchets radioactifs, ce qui signifie que le modèle doit avoir de bonnes capacités de prédiction en ce qui concerne les propriétés hydrauliques et le transport réactif.
2. Un modèle repose sur l'utilisation de données issues du milieu naturel. Cependant ces données peuvent être biaisées par les conditions d'échantillonnage ou l'altération de la roche. En ce qui concerne la taille des fractures, la segmentation des fractures observées sur un affleurement peut modifier profondément la distribution des longueurs. De plus, même des fractures en-échelon ou se présentant en surface sous la forme de lignes brisées peuvent être connectées en profondeur, ce qui modifie la connectivité du milieu et donc ses propriétés hydrauliques. Finalement, la segmentation crée des configurations

fortement corrélées difficilement reproductibles lors de la génération de réseaux de fractures synthétiques. Il est donc nécessaire de la corriger pour obtenir des paramètres valides.

3. Les données utilisées proviennent principalement de deux supports d'observations, les puits et les affleurements. Si les données de puits permettent de définir les paramètres de densité du réseau de fracture, seules les données recueillies sur les affleurements permettent de définir les paramètres d'échelle. Cette distinction crée une difficulté supplémentaire dans la mesure où les deux types de paramètres sont nécessaires pour définir le modèle.
4. Définir le modèle décrivant les longueurs des fractures est une tâche critique dans la mesure où c'est ce modèle qui va permettre de relier les différentes échelles. Les modèles en loi de puissance sont largement utilisés pour décrire la distribution des longueurs et sont de plus cohérents avec les analyses préalables des affleurements effectuées par SKB. Néanmoins, la cohérence entre l'échelle des affleurements (1-10 m) et l'échelle des linéaments (1-10 km) pose toujours problème lorsque les exposants mesurés diffèrent entre ces deux échelles.
5. Les modèles utilisés contiennent la définition de domaine de fracturation, où les propriétés statistiques des fractures sont supposées homogènes à la variabilité près. Les paramètres de densité ne peuvent en effet pas être considérés comme constants puisque les propriétés géologiques varient naturellement en fonction de l'échelle considérée et que le réseau de fracture est intrinsèquement hétérogène. Caractériser cette variabilité et l'intégrer dans le modèle est donc une tâche nécessaire.
6. L'analyse de la fracturation le long des puits doit permettre de définir les zones où les propriétés de fracturation sont similaires afin de décrire les domaines de fracturation. La notion de similarité dépend fortement de la variabilité du milieu naturel et de l'incertitude sur les paramètres du modèle. On estimera en effet que deux modèles sont équivalents si leur différence est inférieure à la variabilité intrinsèque des zones considérées et à l'incertitude. C'est pourquoi il est essentiel de prendre en compte la variabilité et de l'incertitude lors de la définition des paramètres du modèle de fracturation.

L'intégralité de l'analyse des deux sites candidats a donné lieu à la publication d'un rapport pour SKB [Darcel et al., 2009]. Les travaux effectués dans le cadre de la présente thèse ont porté essentiellement sur la caractérisation de la distribution des longueurs de fracture (questions 2 et 4) et la définition de zones de fracturation à partir de l'analyse des données de forage (question 6). Ces travaux ont débouché sur la rédaction de deux articles scientifiques dont le texte est donné dans les paragraphes suivants. Ces articles se focalisent sur le site de Laxemar. Ils sont précédés d'un résumé en français et suivi par des résultats complémentaires concernant le site de Forsmark. Une première partie s'attache à décrire les deux sites étudiés ainsi que les données utilisées pour effectuer les différentes analyses et une dernière partie synthétise et conclut cette partie.

B. Description des sites de Forsmark et Laxemar

Les sites candidats à l'enfouissement des déchets nucléaires se trouvent au sud-ouest de la Suède. Ils sont localisés sur le bouclier scandinave (figure 3-1). Le sous-sol est principalement composé de granites très peu perméables. L'essentiel de la transmissivité provient donc du réseau de fracture, ce qui justifie l'étude et la modélisation des propriétés de fracturation du milieu. Afin d'analyser les propriétés de la fracturation, on dispose d'une grande quantité de données provenant de relevés le long d'une vingtaine de puits pour chacun des sites ainsi que de cartes des traces de fractures sur 6 et 8 affleurements pour Forsmark et Laxemar. Chaque puits a une longueur variant d'une centaine de mètres à un kilomètre. A

chaque intersection entre le puits et une fracture sont relevées la position de l'intersection et les orientations de la fracture (inclinaison, pendage) ainsi que les incertitudes liées aux conditions de mesure [Stigsson and SKB, 2007; 2008]. Les affleurements couvrent une échelle d'une dizaine de mètre et, pour chacun d'eux, toutes les traces de fractures dont la taille est supérieure à 0.5m sont cartées afin de répertorier leurs longueurs et leurs orientations. À côté de ces affleurements sont également fournies des cartes de linéaments couvrant une échelle de 100 mètres à quelques kilomètres, ce qui permet d'échantillonner les longueurs à une échelle supérieure. Finalement, la carte des failles majeures à l'échelle de la Suède est également disponible. La figure 3-2 présente un exemple de cartes d'affleurement, de linéaments et des failles majeures.

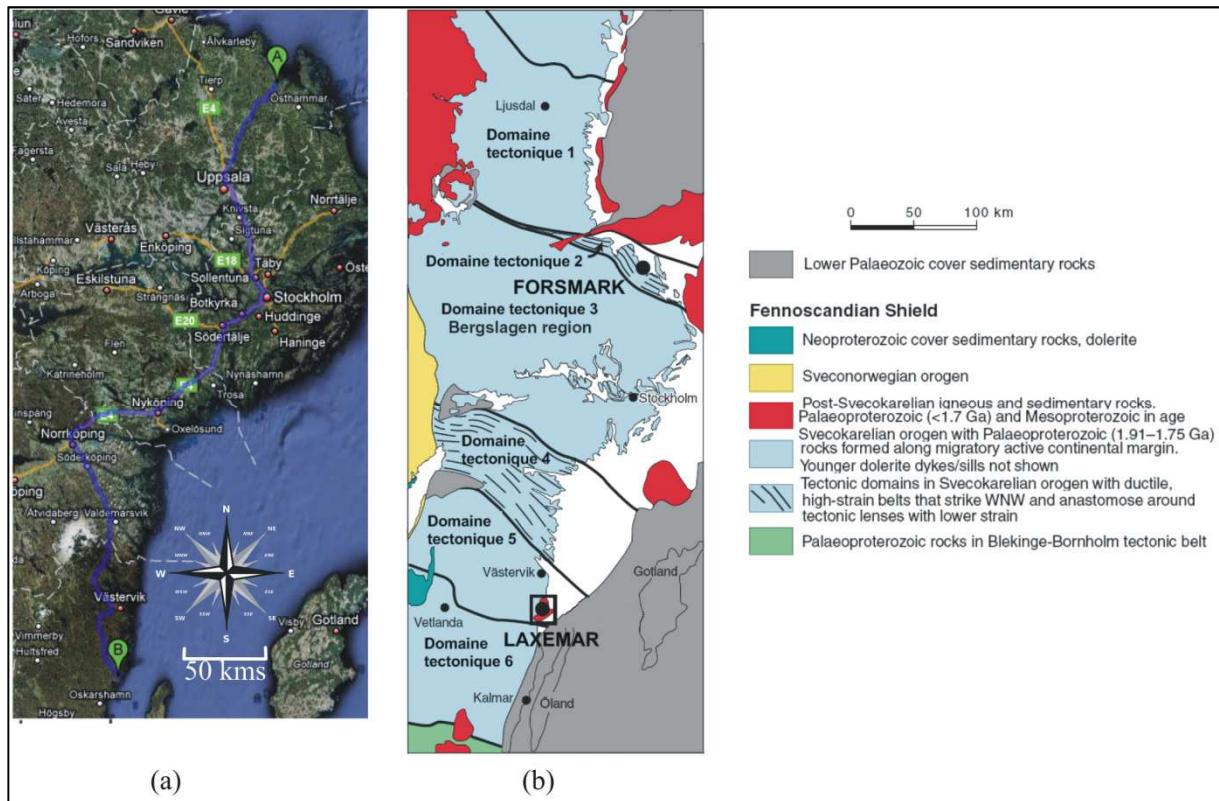


figure 3-1 : (a) Localisation de Forsmark (A) et de Laxemar, (B) carte Google™ Maps. (b) domaines tectoniques, modifié de [Wahlgren et al., 2008].

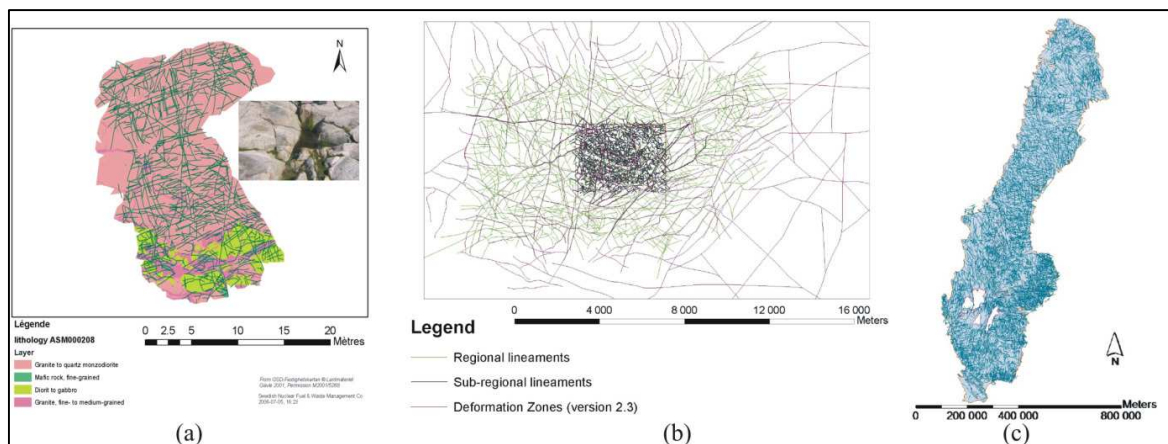


figure 3-2 : Illustration des données disponibles sur les longueurs des fractures (a) à l'échelle d'un affleurement, (b) à l'échelle des linéaments et (c) à l'échelle de la Suède.

C. Méthode d'analyse de la longueur des traces de fracture appliquée à des affleurements et linéaments Suédois.

La distribution des longueurs de fractures peut être modélisée par une loi de puissance sous la forme [Bour et al., 2002]:

$$n(l, L) = \alpha \cdot L^D \cdot l^{-a} \quad (3-1)$$

où n représente le nombre de fracture dans un système de taille L dont la longueur appartient à l'intervalle $[l, l+dl]$, D est la dimension fractale du milieu, inférieure ou égale à la dimension topologique, α est le terme de densité et a l'exposant de la loi de puissance. Si D peut être mesuré indépendamment (voir par exemple [Bonnet et al., 2001]), α et a doivent être déterminés à partir des observations du milieu naturel. Comme il s'agit d'une loi de puissance, ces données doivent être disponibles sur plusieurs ordres de grandeur afin d'en déduire un modèle pertinent. Les travaux présentés ici décrivent une méthode d'analyse des traces à l'échelle d'un affleurement. De plus les données disponibles aux échelles plus larges (linéaments) sont également utilisées pour dériver les paramètres a et α pour le site de Laxemar (Suède).

1. Résumé

L'objectif du papier est de définir les paramètres de la fracturation du site de Laxemar (Suède). Les paramètres de l'équation (3-1) sont dérivés des traces de fractures relevées sur 8 affleurements du site couvrant une échelle de 1 à 10 mètres. Les données sont tout d'abord corrigées des différents biais d'observation, comme l'effet topographique, les effets de troncation et les déconnexions de fractures. Les exposants calculés pour chaque affleurement sont comparés entre eux et comparés aux observations à grande échelle (échelle des linéaments) afin d'évaluer la robustesse de la méthode et de dériver les paramètres du modèle régional. La correction des biais tend à réduire la valeur de a et n'a pas d'incidence significative sur la valeur de α . Le biais dans les observations réduit ainsi artificiellement l'importance des grandes failles, ce qui peut avoir de lourdes conséquences sur le modèle de fracturation et les propriétés hydrauliques déduites de ce modèle. Parmi les huit affleurements étudiés, six ont des caractéristiques communes, les longueurs suivant une loi de puissance $n(l, L) = 2.0 \cdot L^2 \cdot l^{-2.3}$ (modèle I). Deux affleurements présentent une distribution différente et suivent une loi de puissance $n(l, L) = 2.0 \cdot L^2 \cdot l^{-3.0}$ (modèle II). Ce dernier modèle est cohérent avec les observations réalisées sur les cartes de linéaments couvrant une échelle allant de 10^2 à 10^5 mètres. Le changement de distribution entre l'échelle des affleurements et l'échelle des linéaments s'explique par le changement de régime dans la densité de fracturation. Les affleurements du modèle peuvent être considérés comme des réseaux "clairsemés" alors qu'à l'échelle des linéaments, comme pour les deux affleurements du modèle II, ils peuvent être considérés comme denses. La description de ces modèles, ainsi que leurs implications en termes de propriétés hydrauliques, a fait l'objet d'une publication spécifique fournie en annexe [Davy et al., 2009].

2. Fracture trace sizes analysis method applied to Swedish outcrops and lineaments

R. Le Goc ^{a,b}, C. Darcel ^a, P. Davy ^a and J.-R. de Dreuzy ^a

^a Géosciences Rennes, UMR 6118 CNRS, Université de Rennes 1, CS 74205, F-35042 Rennes Cedex, FRANCE

^b Itasca Consultants SAS, 64 chemin des mouilles, F-69134, Écully Cedex, France

Abstract:

Some natural fracturing properties, like fracture size, have a power-law behavior. Finding the parameters of the distribution from data measurements is a critical but non-trivial task. The power-law exponent of the fracture size distribution is indeed used as a scaling exponent but has to be derived from biased data. Finding the properties of the natural rock from biased data requires techniques to correct the data. We present here a method that works directly on the fractures traces measured in outcrops to determine the power-law exponent of the fracture size distribution. This method is based on the projection of the fracture traces on a flat topography, a linkage procedure that reconnect together traces disconnected during the sampling process and a correction of finite-size effects in the computation of the distribution. We finally apply this method to Swedish outcrops (1-10m) and use additional observations on lineaments (10²-10⁵m) to find the fracturing properties at the regional scale. The results indicate that there are two scaling models, one at the outcrop scale and another for the lineament scale.

a) Introduction

Mechanical and flow behaviour of rocks are strongly dependant to fractures and fracture network. Creating an equivalent discrete fracture network (DFN) and modelling physical properties on fractured rocks requires an appropriate statistical description of the fracture characteristics.

Different parameters can describe the rock properties: the fracture sizes, orientations (strike and dip) and the fracture density can define the spatial pattern of fractures. From all these parameters, we can compute distributions using field data. These data are available at different scales and dimensions and the main issue in the fracturing pattern analysis is to determine the manner in which fracture system properties at different scales relate to each other, i.e. their scaling properties and, in a statistical framework, their scaling laws.

Studying the fracture sizes belongs to these main issues. Fracture size plays indeed an important role in the overall fracture network connectivity. That is why this paper will only focus on the analysis of fracture size distributions and particularly on the way of finding the appropriate parameters of the distribution from field data.

We will first describe the different way of sampling fracture size distributions and the inherent difficulties, then we will present some known method to overcome these problems, next we will explain our methodology to correct fracture traces distributions to finally outline the results of the correction procedure.

Svensk Kärnbränslehantering AB (SKB) supplied all the data used in this paper. SKB is responsible for the nuclear wastes repository in Sweden. We focus here on data from the Simpevarp site. See SKB public report for more details on the site and/or on the data [Cronquist *et al.*, 2004; Forssberg *et al.*, 2005; Cronquist *et al.*, 2006].

b) Fracture size distribution model

The power-law scaling of fracture traces is supported both by the literature [Bonnet *et al.*, 2001] and by previous analyses performed at the outcrop scale towards SKB sites.

(1) Correcting bias in the sampled data

The power-law parameters estimation is very sensitive to noise [Oltmans and Verheijen, 1997]. That is why working on good quality measurements is essential. The only relevant data of fracture size is trace size measurements, sampled on outcrops or lineaments.

Several corrections are necessary to go from initial raw data to final size distributions, which DFN model parameters are derived. Essentially this concerns two aspects: sampling biases and adequacy between fracture model definition and apparent data. There are three common biases in rock fracture studies, truncation and censoring [Pickering *et al.*, 1995] and the orientation bias [Terzaghi, 1965]. Truncation occurs since the fracture size range is limited - bounded below by the resolution of the outcrop mapping and above by the scale of the outcrop - and is, in practice, significantly smaller than the range of sizes covered by the fracture size distribution. The main cause of censoring is that fracture traces may be larger than the outcrop mapped. Their true size is also larger than the observed one. Orientation bias is mainly caused by small-dip fractures; the probability for these fractures to be visible on surface is indeed very low.

In addition, significant biases can arise from apparent segmentation and deformation of fracture traces, due to exposure conditions, weathering and map departure from planarity (local topography). All these factors directly influence the distribution shape (Figure 3-3). Besides, the level of interpretation required to interpret data according to the fracture model meaning - for instance when several en-echelon segmented traces are merged into one feature- will also affect the distribution shape. This disconnection leads to an overestimate of small fractures at the expense of large fractures.

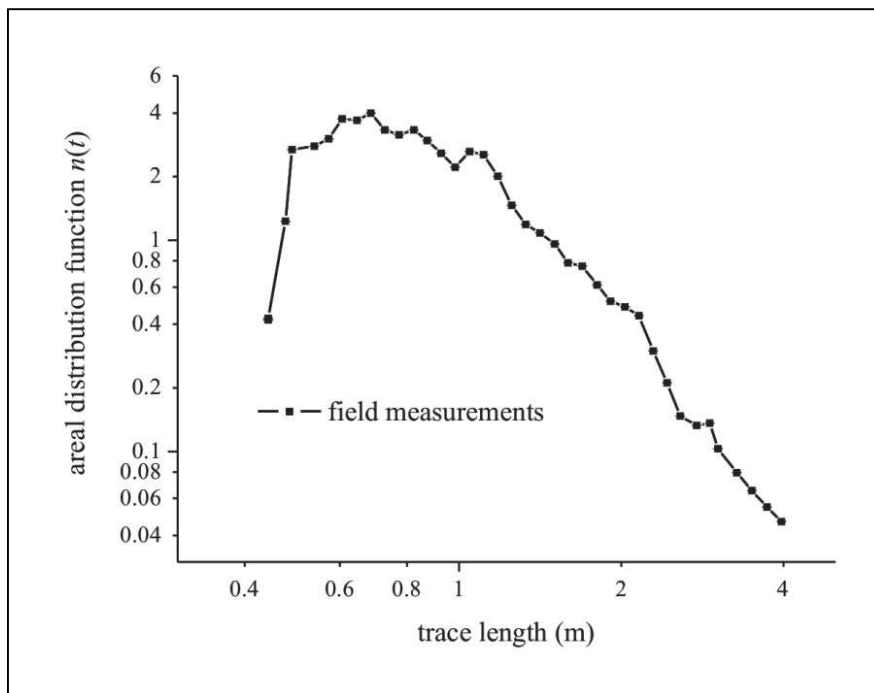


Figure 3-3 : Fracture trace size pdf from trace field measurements: the power-law behavior of the distribution is clearly visible. We notice censoring and truncating effects at the left- and the right hand of the distribution.

Techniques for correcting sampling bias already exist. [Priest and Hudson, 1981] devised formulae to identify the underlying distribution from a biased sample, assuming an exponential parent size distribution. Later [Munier, 1995; Riley, 2005] applied this method on power-law distributions. Because, in some circumstances, these methods can assume nonstandard forms for which there are no standard analytical tools, there might be conceived as counterintuitive. Another approach is based on the assumption that circular, perfectly flat, infinitely thin discs can model fractures, and that all the properties of the fracture system are independent. For example [Marrett and Allmendiger, 1991] and others like [Piggot, 1997; Berkowitz and Adler, 1998; La Pointe, 2002] have demonstrated that trace sizes on outcrops are also power-law distributed with a shape parameters $k_{2D} = k_{3D} - 1$. Nevertheless, these results should be undertaken with caution, in particular because of clustering and anti-clustering effects near large fractures [Ackermann and Schlische, 1997]. Due to these uncertainties another approaches based on stochastic simulations were developed. The basic idea is to mimic the real sampling conditions with simulated fracture networks [Dershowitz et al., 1991a; Dershowitz et al., 1991b]. An objective function is defined to compare artificial fracture traces (and eventually boreholes) and field observations. These methods are based on stochastic simulations associated to simulated annealing or conjugated gradient algorithms.

(2) Fitting the power-law model

A complete description of the power law model can be found in [Newman, 2005]. The standard strategy to determine a power-law behaviour is to create a histogram of the distribution, which appears as a straight line when plotted on log-log scales. The probability density function of a power-law distribution can be expressed as:

$$f(r) = \frac{k \cdot r_{\min}^k}{r^{k+1}} \quad (3-2)$$

where r_{\min} is the location parameter (smallest value of r) and k the shape parameter [Munier, 2004]. To estimate the parameters of the distribution it is possible to plot the histogram on a log-log scale and manually fit the model parameters using, for example, least-square regression. However, this method is prone to errors, particularly in the right-end of the distribution: the power-law distribution dwindles in this region because each bin has only a few samples in it. We strongly recommend a more robust method -i.e. less sensitive to outliers, such like maximum likelihood estimators [Press et al., 2007]. This method leads to [Newman, 2005]:

$$k = N / \sum_{i=1}^N \ln \frac{r_i}{r_{\min}} \quad (3-3)$$

where N is the total number of elements in the distribution, k is the shape parameter and $r_{\min} = \min(r_i)$. An estimate of the expected statistical error on k , σ_k , is given by:

$$\sigma_k = N^{0.5} / \sum_{i=1}^N \ln \frac{r_i}{r_{\min}} = \frac{k}{N^{0.5}} \quad (3-4)$$

Following equation (3-2), the density term is:

$$\alpha = N \cdot k \cdot r_{\min}^k \quad (3-5)$$

and its associated uncertainty is:

$$\sigma_{\alpha} = N \cdot r_{\min}^k \cdot (\sigma_k + k + \ln r_{\min}) \quad (3-6)$$

We propose here a new method to find the fracture-size-distribution parameters from biased trace-size data. Unlike other methods that correct distributions in a statistical framework, using assumptions on the parent distribution, we attempt to correct data in a deterministic manner - i.e. we apply a correction directly on fracture traces and next apply finite size corrections to fracture size distribution to remove, or at least reduce, biases.

c) Procedure for fracture-traces correction

As described in the previous paragraph, the data collected from the field are damaged by the sampling condition. We describe here a method that attempt to improve the data quality to restore the original distribution.

In this section, we show how fracture traces are initially corrected directly from observation biases including the topographical effect and apparent disconnection. From the outcrop, it is only possible to measure biased traces. Topography or ground weathering can disconnect traces and the surface where fractures are sampled can affect the distribution. A linkage procedure is also required and the proposed method focuses on finite-size effects and disconnected traces.

(a) Finite size correction

In an outcrop, due to the fracture density, the probability of seeing a large fracture is lower than the probability of seeing a small fracture. Therefore, to perform the area normalization we do not use the total area of the outcrop for every class of fracture length but only the area in which fractures can be observed, -i.e., the area where the fracture centre can be located.

Thus, we compute the corrected area using a grid interpolation of the outcrop surface. We count only cells at a distance greater or equal than $t/2$ from the boundary, where t is the mean fracture-trace size of the current class, and multiply this number by the area of a cell (Figure 3-4).

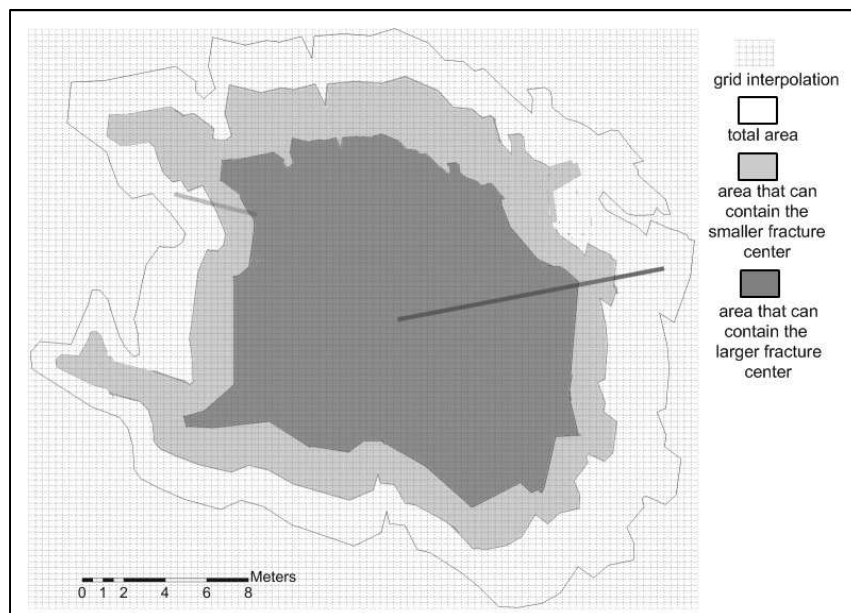


Figure 3-4 : Representation of the area normalization accounting for finite-size effect. White indicates the total outcrop area, light grey, the area that might contain the smallest fracture; and dark grey, the area in which the largest fracture could be observed. The difference between these areas causes a bias in the observations.

Then we compute the fracture-size distribution only on traces entirely included in the sampling surface. Next, we normalize the distribution using the corrected area previously computed. Finally, we estimate the distribution parameters using the maximum likelihood estimator.

(b) *Projection on a flat topography*

Maps defined for fracture trace sampling are not perfectly planar but rather naturally irregular. This potentially entails a bias into the recording, such as apparent fracture trace segmentation due to sharp accident or trace curvature increase due to round topographical variation. These effects are naturally more pronounced for subhorizontal fractures. They can affect the following fracture-trace linkage process.

The basic idea is to project all fracture traces on a flat surface, in order to reduce the perturbation due to topography. We choose as reference plane $\Pi: z = z_{\text{mean}}$ where z_{mean} is the mean height of the fracture traces coordinates. Fractures are modeled by flat planes, whose orientations are the strike and dip. Each trace segment is projected along the fracture plane in the fracture dip direction (Figure 3-5). The directing vector of the projection can be written:

$$\vec{v} = (-\sin \theta \cdot \cos \varphi, \cos \theta \cdot \cos \varphi, -\sin \varphi) \quad (3-7)$$

where θ is the fracture strike and φ the fracture dip.

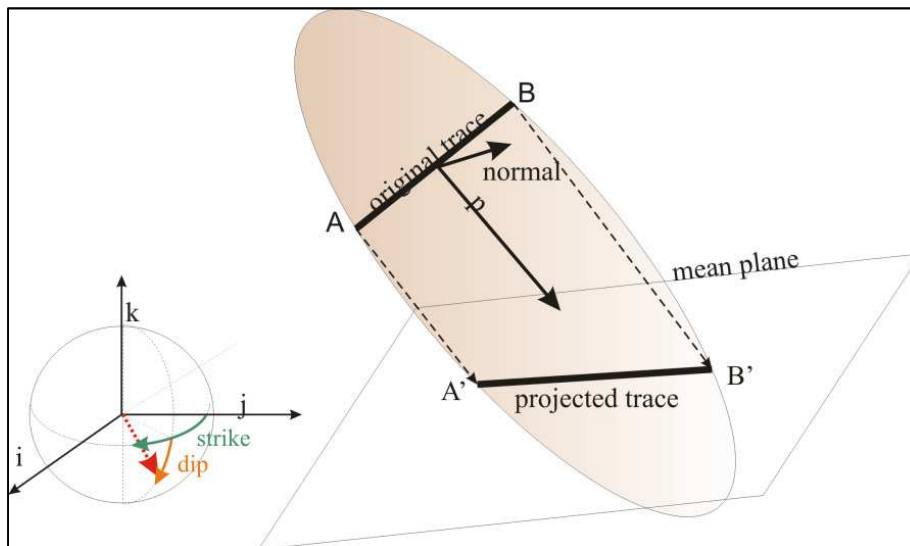


Figure 3-5 : Illustration of the projection of the trace coordinates to a mean plane along the fracture direction. We consider that fractures are modeled by flat ellipse oriented using the measured strike and dip.

In Figure 3-6, an example of non-projected and projected traces is provided. The projection tends to straighten curved traces for fractures of small dip. Figure 3-7 shows the lack of influence of the projection on the final fracture-trace size distribution. Fractures shortened through the projection do not significantly influence the estimate of the fracture length distribution. They essentially fall within the range of truncation effects.

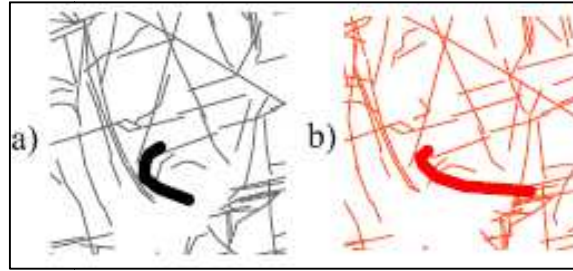


Figure 3-6 : non-projected (a) and projected (b) trace of an outcrop in Laxemar (ASM000208). We observe a straightening up of this curved trace of fractures, which having a small dip value.

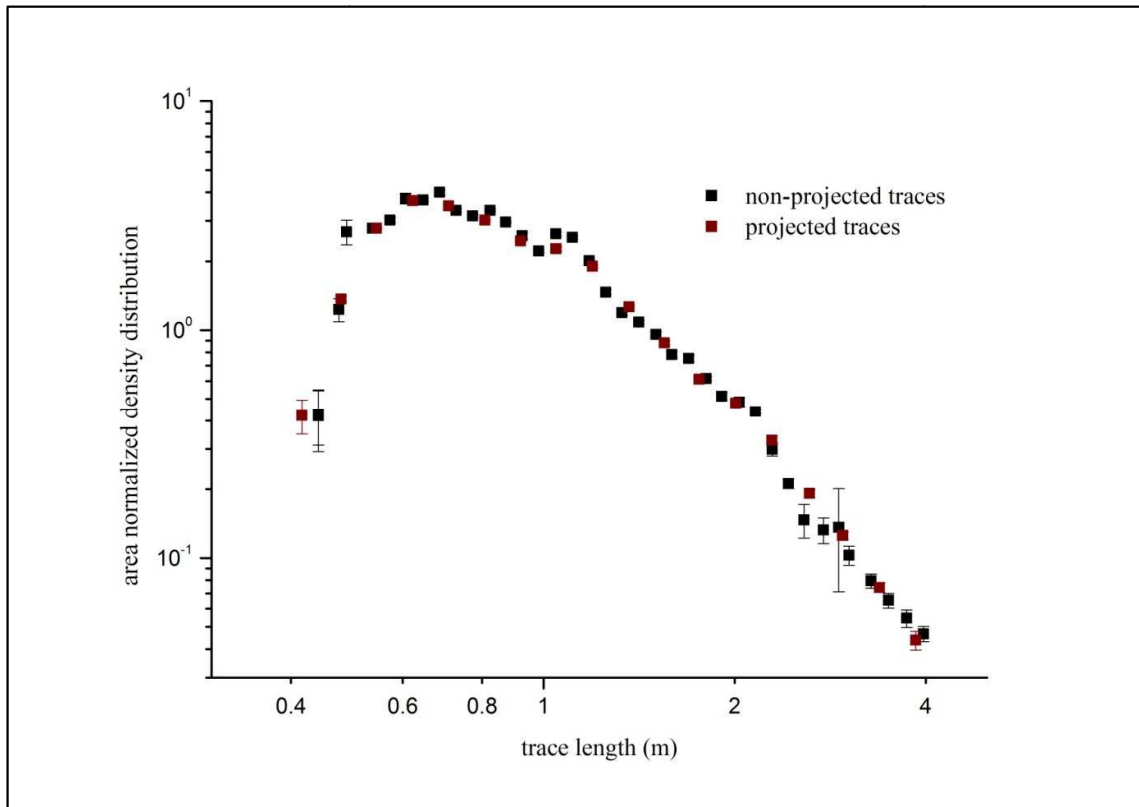


Figure 3-7 : Fracture trace size distribution for the original fracture network (black, squares) and the projected fracture network (red, square). We do not notice any crucial influence of the projection in the power-law parameters.

(c) Linkage of segmented traces

The second step concerns the identification, and consecutive linkage, of apparently segmented traces. It is first motivated by observation difficulties and secondly by the existence of non-Poisson structures (Figure 3-8). Our hypothesis is that en-echelon fractures are elements of an underlying structure and that they are connected at depth. Moreover, non-Poisson structures are difficult to generate in synthetic, statistically based, networks. We set up methods to identify these configurations and link structures of the type a) and b). Dashed lines and en-echelon fractures are close configurations, which we distinguish by use of the following procedure. We consider the projection of a segment A on the line supporting the segment B. If the projection of B and A intersect, the configuration is en-echelon, otherwise the configuration is dashed-line. Layered fractures are identified when the projection of A is included in B or conversely.

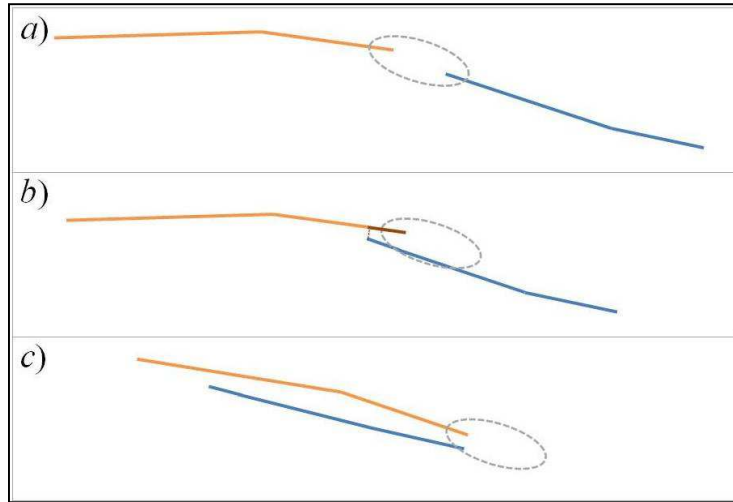


Figure 3-8 : Example of non-Poisson structures in the natural fractures: a) “dashed-line, b) “en-echelon and c) “layered” fractures.

Fracture traces to be linked should be recognized according to two intuitive criteria: similarity of their respective orientations (solid angle between fracture poles) and vicinity of their tips [Fox *et al.*, 2007]. The orientation criterion is simply based on a threshold, i.e. the solid angle between fractures must not exceed a threshold (noted ω_{\max}). In complement, the maximal distance between fracture tips derives from mechanical considerations. Around each fracture tip is defined a zone of influence, into which fractures should be linked and beyond which no linkage occurs. For each fracture tip the zone of influence is an ellipse, defined in polar coordinates, whose size is related to stress conditions and to the fracture length (t_{frac}) itself, such as [Atkinson, 1989]:

$$\frac{\Delta\sigma_I}{\tau} = \left(\frac{t_{\text{frac}}^2}{4r^2} \cdot f(\theta) \right)^{0.5} \quad (3-8)$$

where

- t_{frac} is the fracture size
- (r, θ) are the polar coordinates of the observation point.
- $\Delta\sigma_I = (\sigma_{11}^r - \sigma_{11}^c)$ is the difference between remote normal stress σ_{11}^r and the stress on the crack σ_{11}^c ,
- τ is the shear stress induced by the fracture at the fracture tip is according to

$$\tau = 1/2 \cdot [(\sigma_{22} - \sigma_{11})^2 + 4 \cdot \sigma_{12}^2]^{0.5} \quad (3-9)$$

- σ_{11} , σ_{22} and σ_{12} are the components of the stress tensor.

One fracture trace is selected for linkage if it intersects the curve given by equation (3-8). The curve is an ellipse as shown by Figure 3-9. In the linkage process, this comes to define an equivalent second criterion: a distance threshold $(\tau/\Delta\sigma_I)_{\max}$, which reflects the stress conditions. The zone of influence increases while increasing $(\tau/\Delta\sigma_I)_{\max}$.

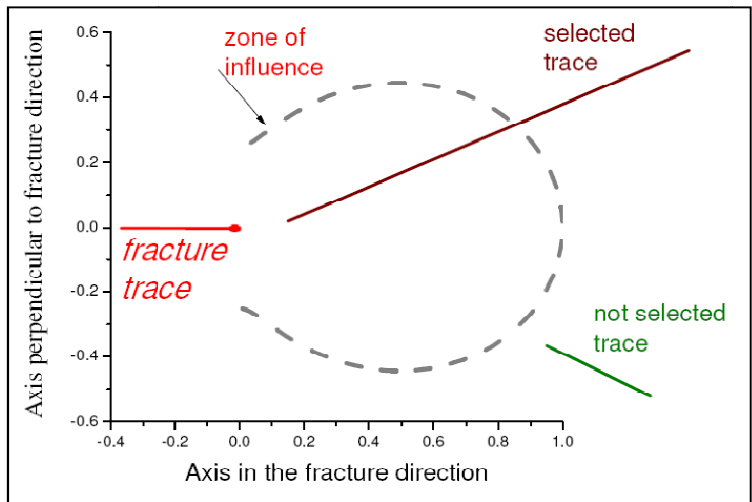


Figure 3-9 : Contour of the shear stress relaxed normalized with the fracture size.

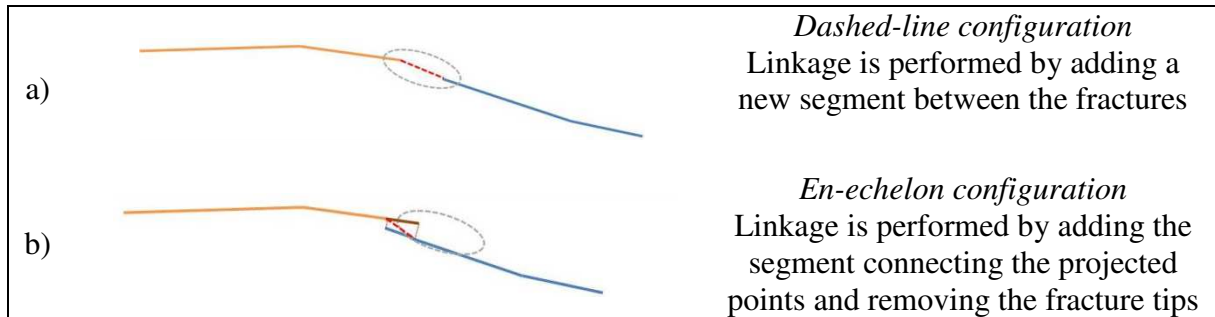


Figure 3-10 : Illustration of non-Poisson structures and description of linkage procedures

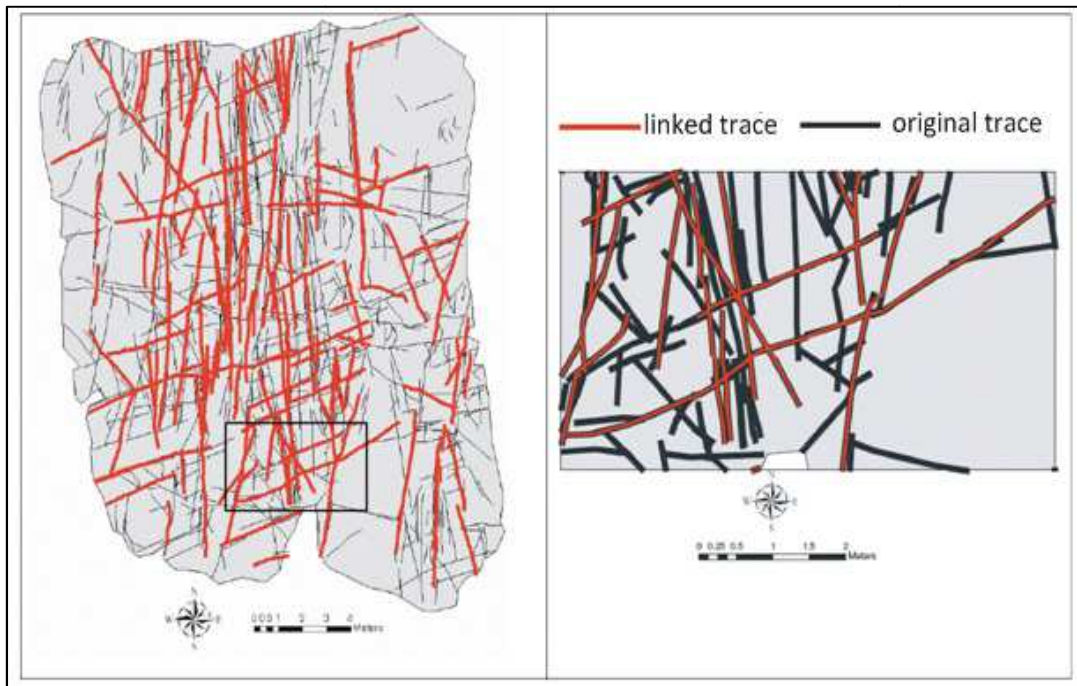


Figure 3-11 : Illustration of linkage on the fracture traces of a Laxemar outcrop (ASM100235): the whole network (left) and detail (right) where original traces are displayed in black and linked traces in red.

Using relevant values of these criteria, concurrent traces may be linked. From the possible couples, we select systematically the closest fractures. Figure 3-10 finally illustrates how the reconnection is performed. For a dashed-line configuration, a new segment is added between the fractures (Figure 2-7a) and, for an en-echelon configuration, the segment connecting the projected points is added while the fracture tips are accordingly adjusted. In addition, we verify visually, that the reconnected trace shapes remain acceptable. A typical outcome of the linkage process is presented in Figure 3-11.

d) Application of the correction procedure to Simpevarp data

The Swedish Nuclear Fuel and Waste Management Company (SKB) is undertaking site characterization at two different locations, the Forsmark and Simpevarp areas, with the objective of sitting a geological repository for spent nuclear fuel. The Simpevarp candidate area consists of two subareas, named the Laxemar subarea and the Simpevarp subarea.

(1) Site and data description

The investigation area is located in the province of Småland, within the municipality of Oskarshamn, adjacent to the Oskarshamn nuclear power station on the southeast coast of Sweden. All the data used in this paper are located within an area with a size of 13 by 21 km (Figure 3-12). Fracture traces from four outcrops in the Laxemar subarea (named ASM000208, ASM000209, ASM100234 and ASM100235) and four other outcrops in the Simpevarp subarea (named ASM000025, ASM000026, ASM000205 and ASM000206) are mapped, providing us trace size at the meter scale, while lineament maps at the 100m-scale and the 1km-scale are also available. We applied the complete correction procedure on outcrop data while lineament data were only affected by the correction of finite-size effects.

(2) Analyses of fracture traces at the outcrop scale

As described previously, two parameters are involved in the linkage procedure, the distance criteria $(\tau/\Delta\sigma_I)_{\max}$ and the orientation criteria ω_{\max} . Selecting optimal values for these criteria relies on the following properties:

- The linked fracture network has a proportion of segmented structures ("en-echelon" and "dashed-line" fractures) consistent with Poisson random fracture network.
- The orientation variability in a fracture trace, i.e. the difference between the orientation of each segment in the trace and the mean trace orientation, remains stable after linkage.
- They form no (or almost no) aberration in the linked fractures.

Thus, we first calculated for each outcrop the orientation variability distribution. Assuming that this distribution is Gaussian with a mean value α and a square-difference σ_α , we set $\omega_{\max} = \alpha + \sigma_\alpha$. Figure 3-13 sums up the resulting values. Then, for each outcrop we calculate the number of structures identified as segmented as a function of the size criterion. This number is calculated at different step of the linkage and compared with the number of segmented structures in a random outcrop with similar size and orientation distributions (computed on the non-corrected outcrop map). The results for ASM000208 are given on Figure 3-14. We notice that using a distance criterion $(\tau/\Delta\sigma_I)_{\max} = 0.4$ leads to a linked network with a similar quantity of segmented structures as in Poisson fracture networks. The resulting $(\tau/\Delta\sigma_I)_{\max}$ values are summed up in Figure 3-15.

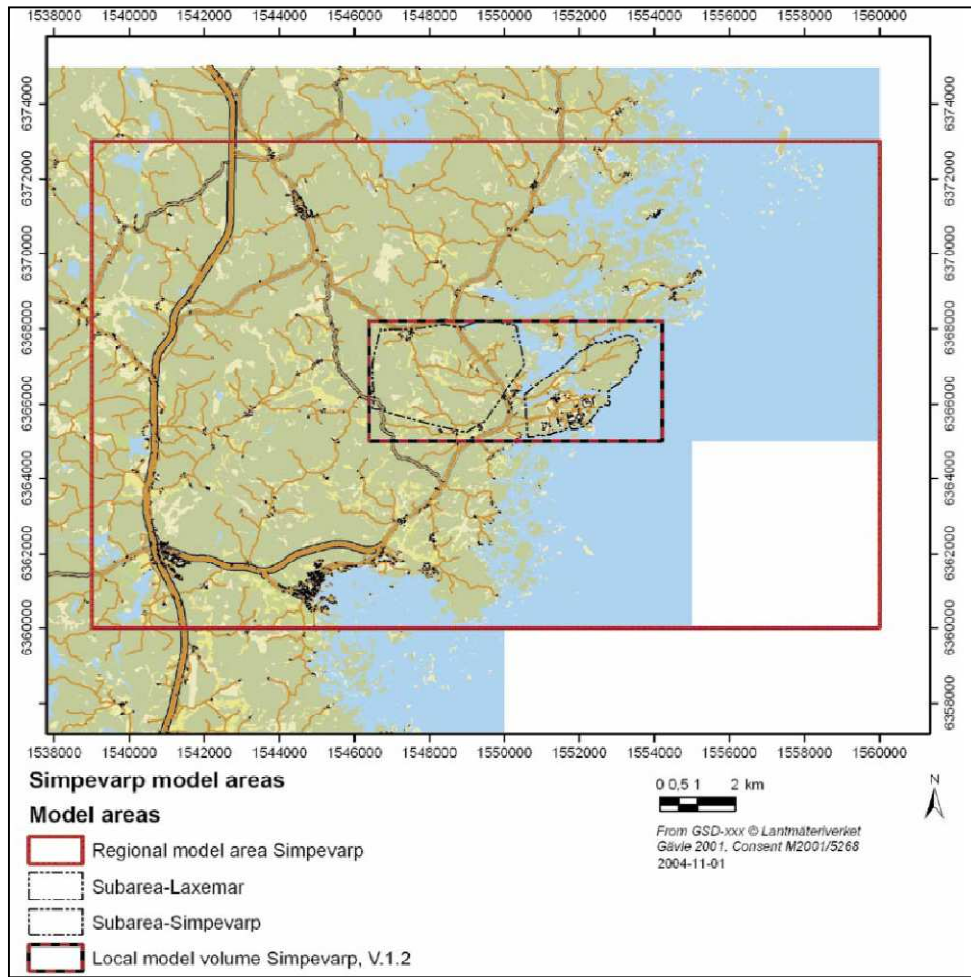


Figure 3-12 : Investigated field: the regional model (thick red line), subareas of Simpevarp (on the right, dashed grey line) and Laxemar (on the right, dashed grey line) and the local model volume Simpevarp V1.2 (dashed black and red line). This picture is extracted from [Darcel et al., 2004].

Outcrop name	α (deg)	σ_α (deg)	ω_{\max} (deg)
ASM100234	7.51	15.91	23.42
ASM100235	8.46	17.19	25.65
ASM000208	7.10	13.40	20.50
ASM000209	8.93	15.11	24.04
ASM000025	3.85	9.16	13.01
ASM000026	4.40	8.58	12.981
ASM000205	4.59	9.59	14.181
ASM000206	6.40	13.16	19.56

Figure 3-13 : Variation of orientation between each segment in a fracture trace and the mean fracture trace. The distribution mean α and the square-difference σ_α are taken into account to compute the orientation criterion ω_{\max} used in the linkage procedure. We obtain an orientation criterion between 13 and 25 degrees.

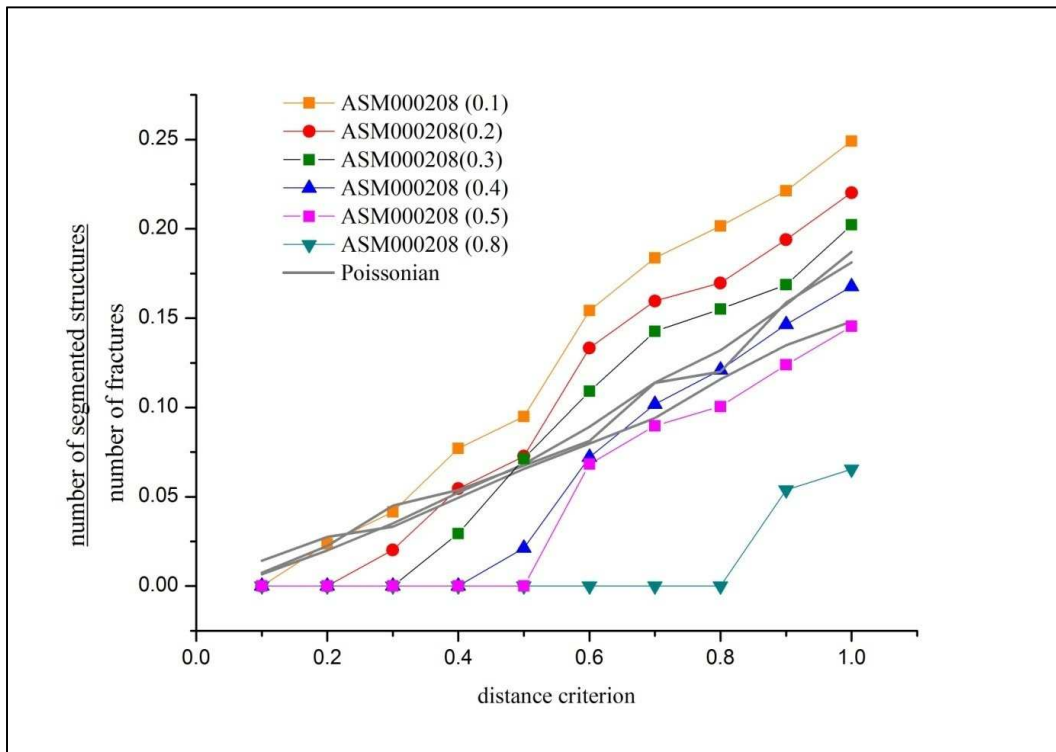


Figure 3-14 : Evolution of the number of structure identified as segmented with the distance criterion. The Poisson curves (grey) are calculated on a synthetic random fracture network generated in the corresponding outcrop contour with similar size and orientation distributions. Other curves are obtained using ASM000208 fracture-trace map after a linkage procedure with $\omega_{max} = 20$ degrees and $(\tau/\Delta\sigma_I)_{max}$ given in parentheses. We notice that after a linkage with $(\tau/\Delta\sigma_I)_{max} = 0.4$ (blue) the corrected outcrop can be seen as Poisson.

Outcrop name	$(\tau/\Delta\sigma_I)_{max}$
ASM100234	0.30
ASM100235	0.30
ASM000208	0.40
ASM000209	0.35
ASM000025	0.35
ASM000026	0.35
ASM000205	0.25
ASM000206	0.35

Figure 3-15 : Each outcrop is associated with the distance criterion $(\tau/\Delta\sigma_I)_{max}$ corresponding to a number of segmented structures similar to a Poisson fracture network with similar orientation and size distributions.

Once all the parameters are defined, we applied the complete correction procedure on all outcrops and calculated the size-distribution parameters, i.e. the shape parameter and the density term. The size distribution is illustrated on Figure 3-16 for the non-corrected traces, projected traces, after correcting only finite-size effects and after the complete correction procedure, including linkage. We notice that the two main corrections (finite-size and linkage) decrease the power-law exponent, meaning that the exponents estimated on non-correction distributions can be overestimated.

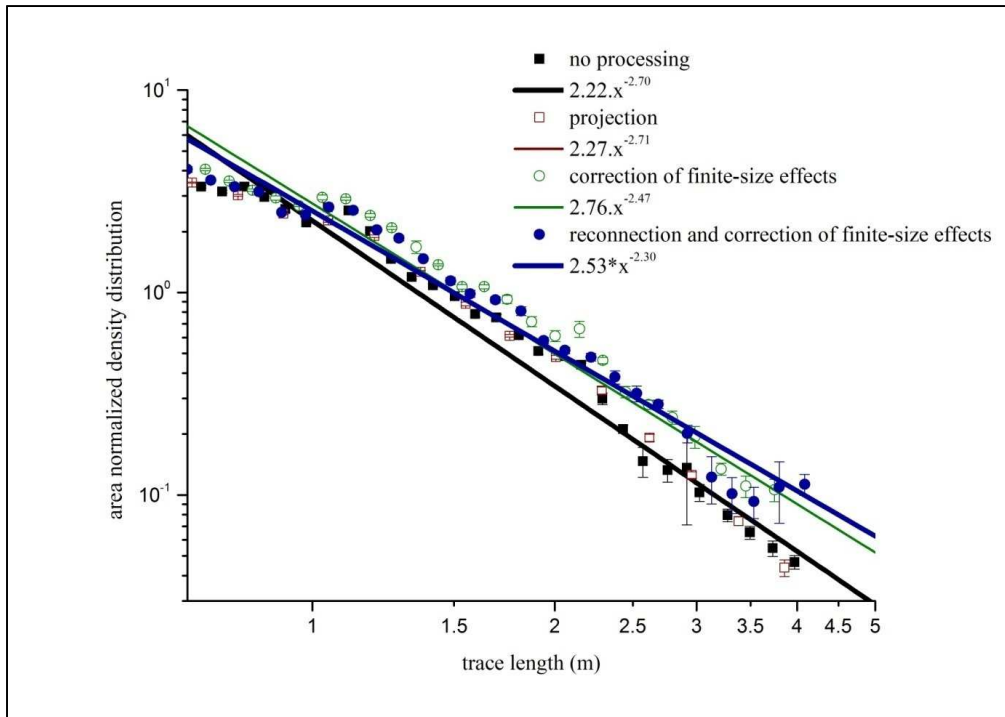


Figure 3-16 : Distribution of the fracture-trace size of a Laxemar outcrop (ASM000208) before (black, squares) and after (blue, circle) the correction procedure. The intermediary distributions with only projection (wine, empty squares) and only area normalization (green, empty circles) are also given for indication. The distribution on corrected data presents less deviation from the expected straight-line and has an exponent lower than the original distribution.

A compilation of all measured fracture-size distribution parameters is provided in Figure 3-17. In general, global correction entails a significant decrease of the power-law exponent while the density term remains similar. For instance, the initial mean without correction is 2.58 whereas it decreases to 2.32 after correction. The scaling-exponent decrease is related logically to the size increase of some fractures during the reconnection process. Therefore, due to apparent disconnection, the original data appear to under-rate large fractures and over-rate small fractures. Thus, the type of correction applied entails significant variations of the resulting parameters of the fracture-size distribution model. In complement, further observations reinforce the confidence in the corrections. The range of exponent values is tightened after correction. In the DFN modeling perspective, we define the mean power-law exponent and dispersion by averaging the values derived from each dataset of a site (disks and error bars in Figure 3-17). It appears that dispersion around the mean decreases when the whole correction process is completed. The square difference between the different exponents has a value of 0.13 without correction and a value of 0.08 after correction. Interesting is that we observe two different classes of outcrops, the first one containing almost all outcrops with a shape parameter equals to 2.3 (model I) and a second class containing two outcrops of the Simpevarp area (ASM000205 and ASM000206) with a parameter equals to 3.0 (model II).

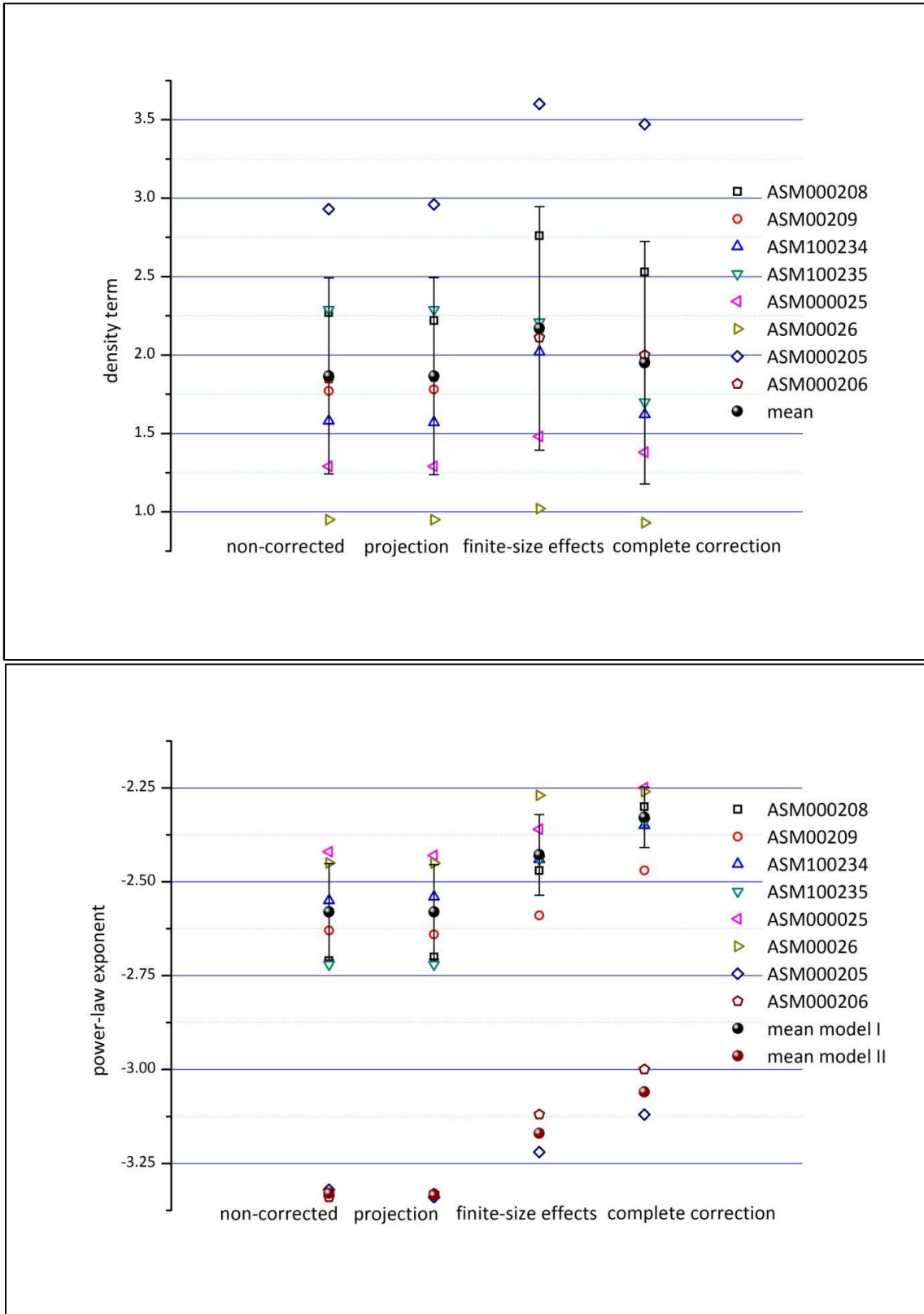


Figure 3-17 : Density term (top) and power-law exponent (bottom) calculated on all available corrected outcrops. We observe that the density term is not mainly influenced by the correction procedure while the exponent is significantly decreased. The final density term is found to be in average 2.0 while we have clearly two models for the exponent. ASM000205 and ASM000206 have an exponent around -3.0 while all other outcrops have an exponent around -2.3.

(3) Analyses of traces at the lineament scale

Deriving the power-law exponent is a critical task because this exponent is used as a scale exponent to link observations made at the outcrop scale with observations made at the kilometer scale (Figure 3-18).

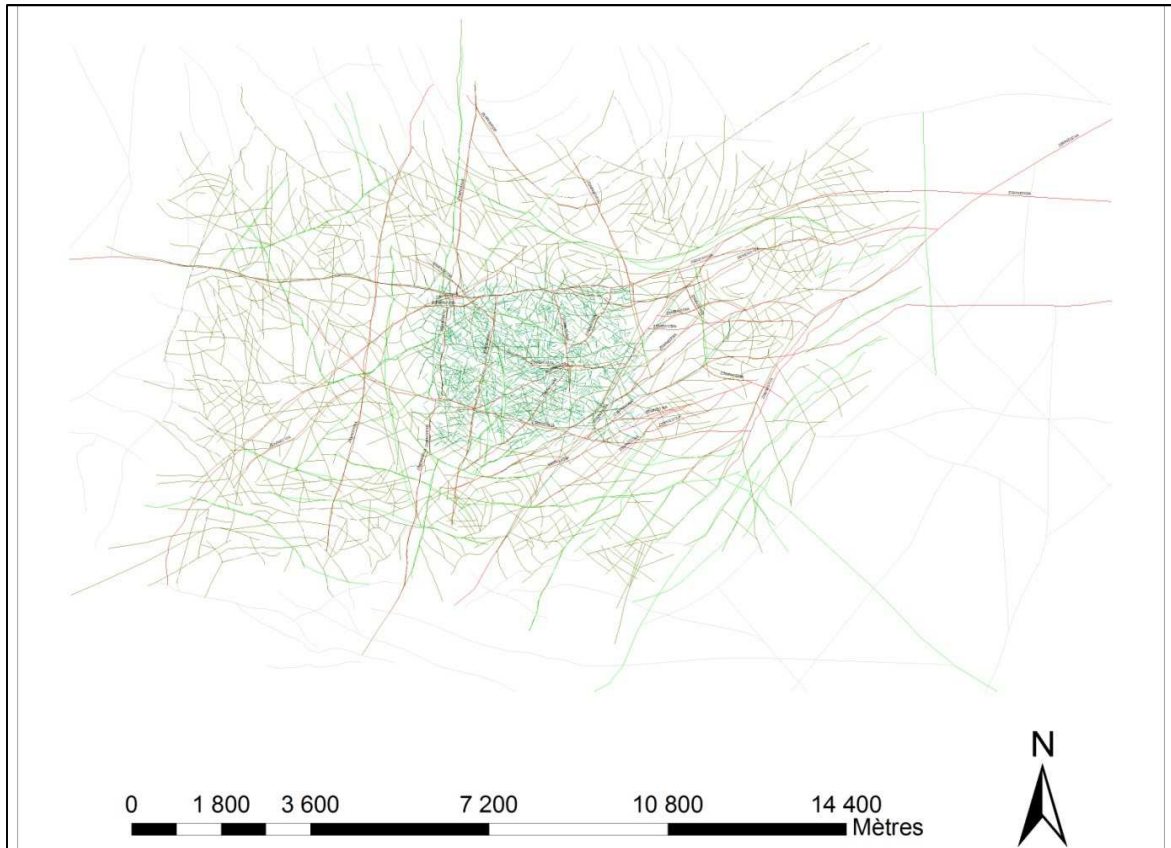


Figure 3-18 : Map of all available lineaments in the Laxemar subarea. There are lineaments smaller than 1km (blue), lineaments larger than 1km (grey) and identified deformation zones (red and green).

The lineament data analyzed here differ straightly from the outcrop trace datasets since they arise from various methods of sampling and especially since they are already linked or coordinated. That is why only a correction of the finite-size effects is applied. All normalized density distributions are combined on multiscale log-log diagrams in Figure 3-19 below. All available scales of observation of DFNs are included, from the outcrop meter up to the lineament kilometer scales. We add intentionally the Swedish faults, representing all faults at the Swede scale, to test the power-law model on a wide range of scale (5 degrees of magnitude).

As a rough conclusion, the scaling behavior of the fracture size distribution follows two major trends: on small scales, an exponent around 2.3 (outcrop model I); and, on large scales, an exponent of 2.0 (outcrop model II). The physical modeling framework proposed by [Davy *et al.*, 2009] explains perfectly the two regimes. This physical model assumes that small fractures cannot cross-larger fault and demonstrate that, in this case, the trace-size distribution has an exponent equals to 3.0 after a certain critical scale, whatever the original exponent was. We notice however that the large lineament model has an exponent equals to 2.2, very close to the outcrop model I, while it should follow the outcrop model II. We suppose that the linkage of these data, performed by SKB specialists, was overestimated.

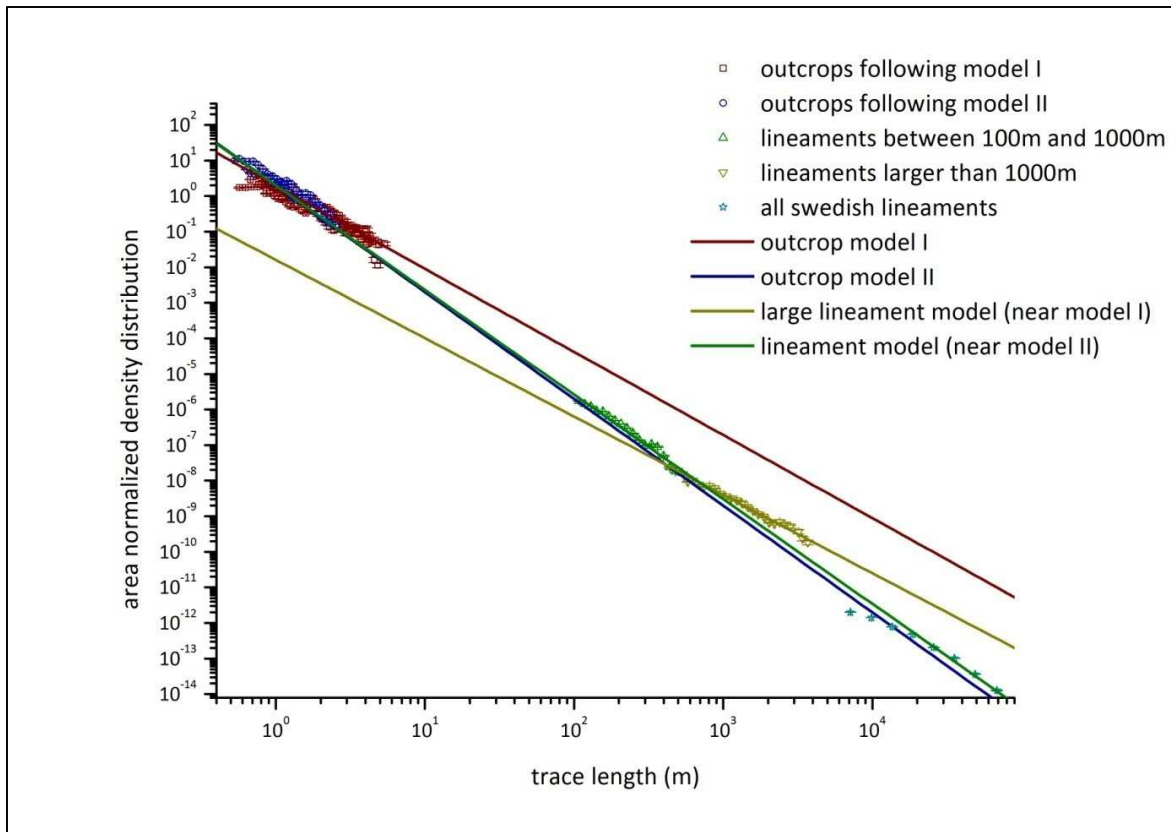


Figure 3-19 : All normalized density distributions available are plot on this multi-scale diagram. The two average model for outcrops are plot, with exponent equal to 2.3 (wine) and 3.0 (blue). We observe that the second model fits also at the lineament scale and the Swedish lineaments, while the lineaments larger than 1km have a distribution close to the first outcrop model.

e) Conclusion

We propose a new method to correct bias in measurement to determine the real fracture-size distribution from biased fracture-trace size distribution. The correction procedure entails a significant decrease of the shape parameter. It can have a significant impact on the fracturing properties, mainly in term of connectivity. This method is here applied to parameters following a power-law distribution but can naturally be extended to other distribution, as assumptions on the distribution shape are not made in the correction procedure.

Acknowledgements: The authors are thankful to Svensk Kärnbränslehantering AB for providing data and funding this study and would like specifically thank Dr. Raymond Munier for its appropriate remarks and suggestions.

3. Résultats complémentaires

La fracturation du site de Forsmark est étudiée de la même manière que celle de Laxemar, présentée au paragraphe précédent. On se concentre ici sur la description des paramètres de la distribution des longueurs calculés aux différentes échelles (affleurement, linéament). Les paramètres des affleurements sont donnés à la figure 3-20. Après correction des biais d'observation, on obtient un modèle en loi de puissance s'écrivant $n(l,L) = 1.42 \cdot L^2 \cdot l^{-2.23}$.

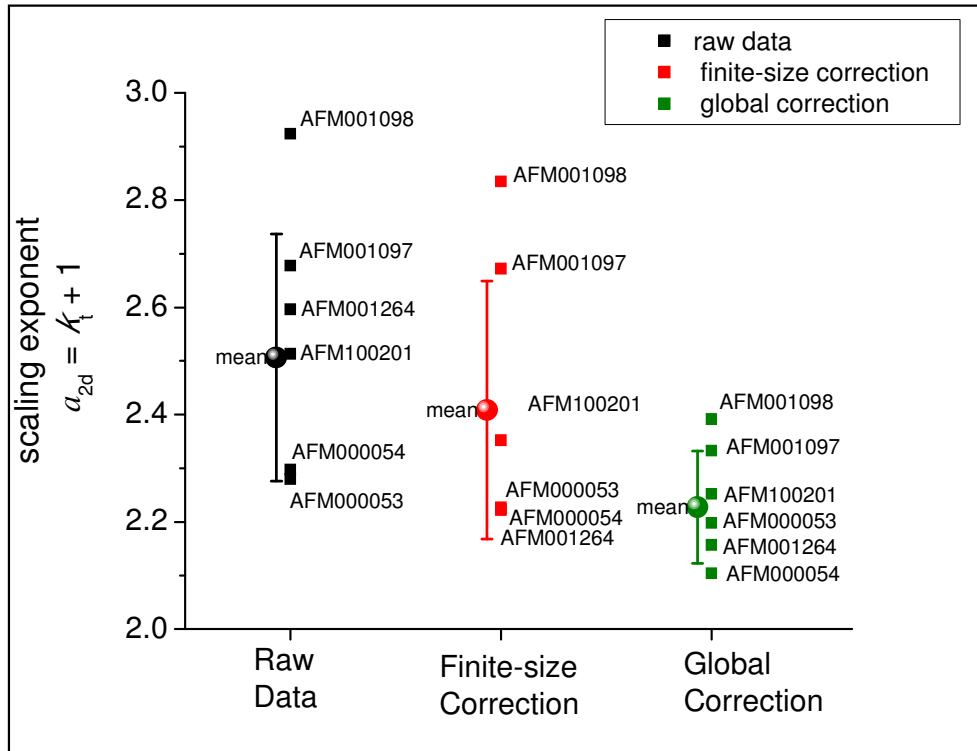


figure 3-20 : Exposants de la loi de puissance dérivés des traces de fractures des affleurements de Forsmark. D'après [Darcel et al., 2009].

On utilise également les données tirées des cartes de linéaments pour dériver les paramètres du modèle régional. Les courbes des distributions aux différentes échelles sont données à la figure 3-21. On retrouve des distributions proches de modèles I (clairsemé) et II (dense) défini à Laxemar, avec des affleurements tous proches du modèle I et des linéaments respectant le modèle II. L'hétérogénéité est cependant plus forte, avec un modèle intermédiaire, celui des linéaments subrégionaux (en noir sur la figure 3-21) présentant une distribution tout à fait différente. On explique cette forte divergence avec les modèles de référence par une forme très particulière de ces linéaments, vraisemblablement due à la méthode d'échantillonnage.

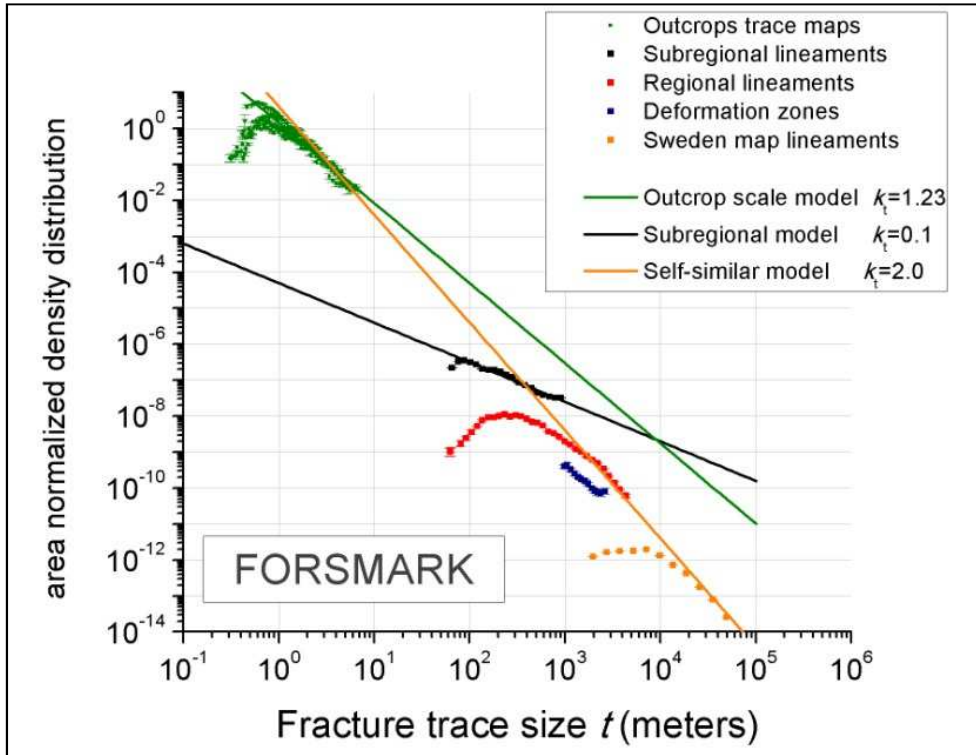


figure 3-21 : Distribution des longueurs de fracture à toutes les échelles disponibles pour le site de Forsmark. D'après [Darcel et al., 2009].

D. Analyse statistique des domaines de fracturations

L'exposant de la loi de puissance est un paramètre essentiel puisqu'il est nécessaire pour comparer des domaines de fracturation de tailles différentes. Cependant, la définition des domaines de fracturation repose également sur la cohérence de la densité de fracturation à l'échelle du domaine, en relation avec les distributions d'orientation des fractures. Afin d'analyser ces distributions, on dispose des données obtenues par imagerie d'un grand nombre de forages répartis sur les sites de Forsmark et Laxemar. Chaque forage est découpé en zones de fracturation définies après observation de la géologie locale de chaque puits par les géologues de SKB. Nous nous reposerons sur ces domaines locaux afin de déterminer les domaines de fracturation à l'échelle globale. Pour chacun de ces domaines, on dispose de l'orientation du puits d'observation, de l'orientation de la fracture par rapport au puits ainsi que des incertitudes associées à ces valeurs.

1. Définition du modèle de densité des domaines locaux de fracturation

La première des tâches est de définir, pour chacun des domaines locaux, la densité de fracturation. Celle-ci s'écrit :

$$N(l, L, \theta, \varphi) = \alpha_{3D}(\theta, \varphi) \cdot L^{D_{3D}} \cdot l^{-a_{3D}} \quad (3-10)$$

où θ et φ correspondent respectivement au pendage et à l'inclinaison des fractures. En normalisant par la surface du domaine, on obtient le modèle de distribution normalisé s'écrivant :

$$n(l, L, \theta, \varphi) = \frac{N(l, L, \theta, \varphi)}{L^{D_{3D}}} = \alpha_{3D}(\theta, \varphi) \cdot l^{-a_{3D}} \quad (3-11)$$

L'exposant de la loi de puissance ne peut être déterminé à partir des observations puisque les longueurs de fracture ne sont pas disponibles. C'est pourquoi α_{3D} est calculé en amont à partir des affleurements et des linéaments. Le terme de densité $\alpha_{3D}(\theta, \varphi)$ représente la densité de fracturation pour les fractures de taille 1. Il peut être calculé à partir des mesures de la densité 1D le long d'un puits de forage en prenant en compte la stéréologie des fractures [Davy *et al.*, 2006].

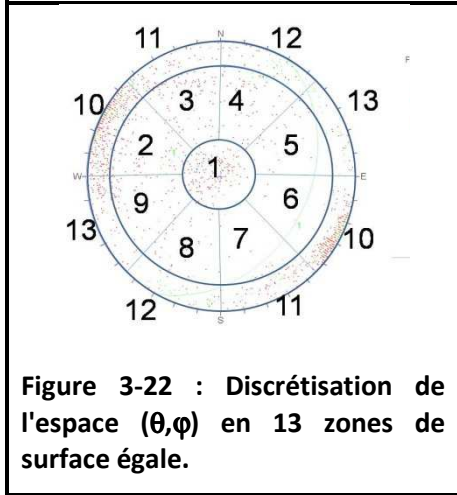


Figure 3-22 : Discretisation de l'espace (θ, φ) en 13 zones de surface égale.

En pratique, $\alpha_{3D}(\theta, \varphi)$ est calculé à partir d'une discrétisation de l'espace des orientations en zones polaires $z_i = \{\theta \in [\theta_1, \theta_2]; \varphi \in [\varphi_1, \varphi_2]\}$. Le nombre de zones définit la précision avec laquelle les termes de densité des différentes zones de fracturation seront ensuite comparés et doit être compatible avec la quantité d'information disponible. Nous avons choisi de discrétiser l'espace en 13 zones en découpant les pendages en trois zones $[0^\circ, 22.62^\circ]$, $[22.62^\circ, 72.08^\circ]$ et $[72.08^\circ, 90^\circ]$. Les inclinaisons de la zone centrale sont quant à eux découpés en 8 intervalles de 45° . Ces valeurs assurent que chaque zone couvre une surface égale. La répartition est illustrée à la Figure 3-22.

Lorsque l'on définit le terme $\alpha_{3D}(\theta, \varphi)$ à partir des valeurs discrètes relevées sur un puits d'observation, il est nécessaire d'intégrer l'incertitude sur les valeurs de θ et de φ , notées σ_θ et σ_φ . En considérant que l'incertitude est gaussienne, on définit la probabilité p_i pour une mesure d'orientation donnée (θ, φ) d'appartenir à l'intervalle $[\theta_1, \theta_2] \cap [\varphi_1, \varphi_2]$ par :

$$p_i = \frac{1}{2} \left[\operatorname{erf} \left(\frac{\theta_2 - \theta_i}{\sigma_{\theta_i} \sqrt{2}} \right) - \operatorname{erf} \left(\frac{\theta_1 - \theta_i}{\sigma_{\theta_i} \sqrt{2}} \right) \right] \cdot \frac{1}{2} \left[\operatorname{erf} \left(\frac{\varphi_2 - \varphi_i}{\sigma_{\varphi_i} \sqrt{2}} \right) - \operatorname{erf} \left(\frac{\varphi_1 - \varphi_i}{\sigma_{\varphi_i} \sqrt{2}} \right) \right] \quad (3-12)$$

où erf est la fonction d'erreur :

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-z^2} dz \quad (3-13)$$

L'incertitude associée à cette probabilité est prise comme son écart type σ_{p_i} :

$$\sigma_{p_i}^2 = p_i \cdot (1 - p_i) \quad (3-14)$$

Par ailleurs, chaque observation doit être corrigé du biais d'observation $\psi(\omega)$ lié à l'orientation ω du puits, ce qui donne comme distribution pour un ensemble de N observations :

$$\alpha_{3D}([\theta_1, \theta_2], [\varphi_1, \varphi_2]) = \sum_{i=1}^N [p_i \cdot \psi(\omega_i)] \quad (3-15)$$

associée à un écart type de :

$$\sigma_\alpha([\theta_1, \theta_2], [\varphi_1, \varphi_2])^2 = \sum_{i=1}^N \left[p_i \cdot (1 - p_i) \cdot \psi(\omega_i)^2 + \left(p_i \cdot \frac{d\psi}{d\omega}(\omega_i) \cdot \sigma_{\omega_i} \right)^2 \right] \quad (3-16)$$

où ψ est le terme de correction des biais d'observation.

Le terme de densité doit également intégrer la variabilité du milieu naturel pour représenter fidèlement les propriétés stochastique du réseau de fracture. En effet, l'hétérogénéité de la fracturation se traduit par une dispersion spatiale de $\alpha_{3D}(\theta, \varphi)$ dont l'importance dépend de l'échelle d'observation. Cette variabilité peut être mesurée en fonction de l'échelle s d'observation par la formulation de l'écart type :

$$\sigma_{\text{var}}(\theta, \varphi, s)^2 = \frac{1}{N_s} \left[\sum_{i=1}^{N_s} \alpha_{3D}(\theta, \varphi, s) \right]^2 - \frac{1}{N_s} \sum_{i=1}^{N_s} \alpha_{3D}(\theta, \varphi, s)^2 \quad (3-17)$$

où $\alpha_{3D}(\theta, \varphi, s)$ est le terme de densité calculé pour un échantillon de taille s et N_s le nombre d'échantillons considérés à l'échelle du domaine. En faisant varier s , on observe que la variabilité décroît avec l'échelle en suivant une loi de puissance de coefficient -0.3 alors que la variabilité d'un modèle poissonien décroît selon une loi de puissance de coefficient -0.5 (figure 3-23). A partir de cette distribution, on est capable d'évaluer la variabilité à l'échelle d'un domaine de fracturation (~100m).

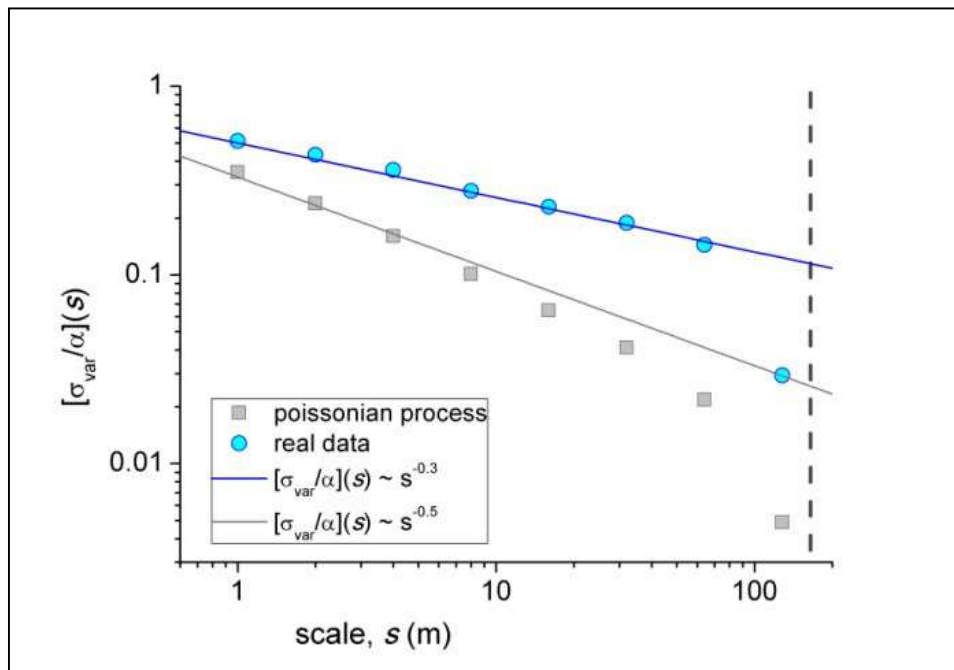


figure 3-23: Analyse de la variabilité du domaine local de fracturation FFM01 du forage KFM01C (Forsmark). On compare l'évolution de $\sigma_{\text{var}}(s)/\alpha_{3D}$ en fonction de s pour les données du site (bleu) à celle pour des données tirées dans une distribution de Poisson (gris). D'après [Darcel et al., 2009].

A partir des équations (3-15), (3-16) et (3-17) on est capable de définir la distribution de $\alpha_{3D}(\theta, \varphi)$ à partir des données pour un domaine local de fracturation. L'incertitude globale associée à chaque valeur de $\alpha_{3D}(\theta, \varphi)$ correspond à la plus grande cause d'imprécision, c'est-à-dire soit l'incertitude liées aux mesures, soit la variabilité du milieu naturel. Ces deux valeurs sont comparées, pour tous les domaines, à la figure 3-24. L'incertitude des domaines où la densité de fracturation est faible est dominée par l'incertitude sur les données alors que la variabilité est dominante pour les domaines où la densité est plus importante. On conserve, comme imprécision finale et pour chaque domaine local de fracturation, la plus grande des valeurs entre l'erreur de mesure et la variabilité.

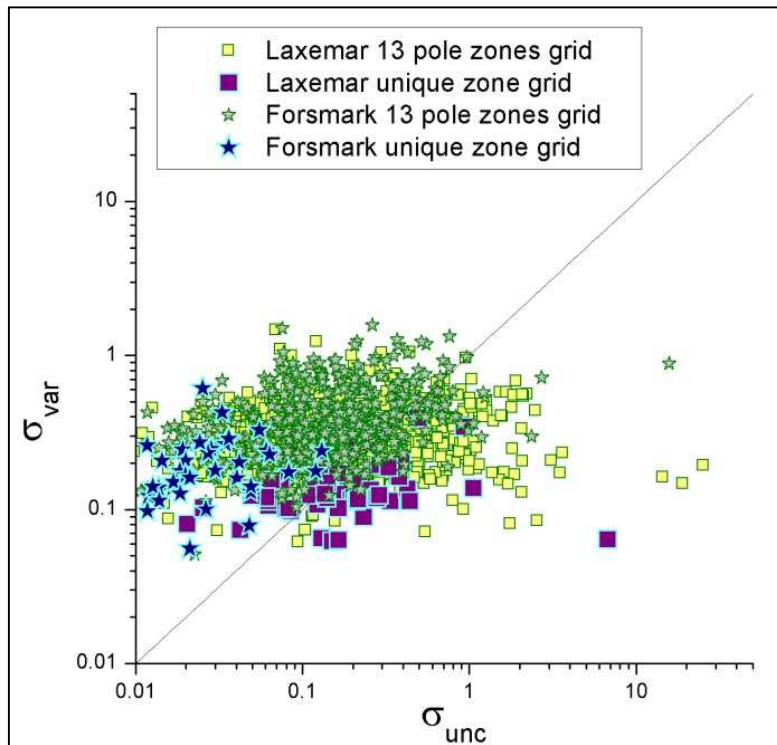


figure 3-24 : Comparaison de σ_{var} et de $\sigma_{\alpha}(s)$ pour tous les domaines locaux de fracturation de Forsmark (étoile) et Laxemar (carré). On calcule $\alpha_{3D}(\theta, \varphi)$ sur la discrétisation en 13 zones (jaune et vert) ou sur l'espace complet pris comme une zone unique (violet et bleu).

On calcule pour chacun des domaines locaux des deux sites, soit 51 pour Laxemar et 60 pour Forsmark, la distribution $\alpha_{3D}(\theta, \varphi)$ associée à son incertitude. Ces domaines sont définis localement à l'échelle d'un puits. Afin de définir des domaines globaux de fracturation, à l'échelle du site, il est nécessaire de comparer ces distributions entre elles afin de déterminer lesquelles sont compatibles et doivent être regroupées.

2. Analyse statistiques des domaines de fracturations

La comparaison des domaines locaux de fracturation ainsi que la méthode permettant de regrouper les domaines présentant des fracturations similaires est fourni sous forme d'un article scientifique.

a) Résumé

Le terme de densité $\alpha_{3D}(\theta, \varphi)$ (équation (3-15)) est défini de manière stochastique par une fonction de densité de probabilité (PDF). Nous avons mis au point une méthode permettant de comparer un ensemble de PDFs et de les organiser en classes. Chaque classe ainsi définie peut être considérée comme une propriété stochastique moyenne représentant l'ensemble des distributions de la classe. Chaque classe contient des distributions similaires, c'est-à-dire que la distance entre les distributions et la moyenne de la classe est inférieure à une valeur donnée d_c . Le calcul des distances, ainsi que le critère de classement, intègrent l'incertitude associée à chaque valeur d'une PDF afin de définir des classes statistiquement cohérentes. De manière générale, fixer $d_c=1$ signifie qu'on regroupe dans une même classe toutes les distributions statistiquement égales à l'incertitude près. Le classement repose sur une méthode d'agglomération hiérarchique itérative qui consiste à regrouper, à chaque étape, les deux distributions les plus proches. Ce procédé est répété itérativement jusqu'à atteindre la distance maximale autorisée.

On applique cette méthode aux distributions de densité des domaines locaux définis sur le site de Laxemar. En fixant $d_c=0.8$, on regroupe les 51 domaines locaux en 7 classes définissant les domaines globaux de fracturation. À chacune des classes on associe un indice de cohérence permettant de quantifier les chevauchements entre classes liés à la variabilité et à l'incertitude. Les 7 domaines de fracturation ainsi définis sont comparés au 6 domaines définis par SKB sur la base de critères géologiques (type de roche, fréquences des fractures, homogénéité). Les deux ensembles de domaines sont clairement différents. Les domaines définis par SKB sont plus cohérents spatialement, c'est-à-dire que deux domaines locaux proches géographiquement ont plus de chance d'appartenir au même domaine global. Cependant leur cohérence statistique en termes d'orientation est loin d'être satisfaisante. D'un autre côté, les domaines de fracturation définis uniquement à partir du terme de densité regroupent des domaines locaux géographiquement dispersés. L'introduction, dans la méthode de classification, d'un critère spatial, devrait permettre de pallier ce problème.

b) Automatic comparison method applied to fracture density distributions.

R. Le Goc ^{a,b}, C. Darcel ^a, P. Davy ^a and J.-R. de Dreuzy ^a

^a Géosciences Rennes, UMR 6118 CNRS, Université de Rennes 1, CS 74205, F-35042 Rennes Cedex, FRANCE

^b Itasca Consultants SAS, 64 chemin des mouilles, F-69134, Écully Cedex, France

Abstract:

Probability Density Functions (PDFs) are widely used to describe the statistical properties of natural processes. Discrete Fracture Network (DFN) models can be described with stochastic laws defining the fracturing densities and scaling laws. If the scaling law is generally defined globally, i.e. at the complete domain scale, the density distribution may be defined at smaller scales by defining sub-domains, each sub-domain having a specific density distribution. These sub-domains are defined locally, i.e. at the borehole scale, using geological observations and fracturing frequency patterns. Relating the locally defined domain to large-scale global domain require to compare and group the local domains into classes of similarity. We propose a method that finds these classes based on the statistical properties (the PDF) of each local domain, which integrate the natural variability that occurs in natural geologic media. The classification process uses a hierarchical clustering algorithm to group together local domains with similar PDFs. This method is then applied to the Laxemar area (Sweden), where a facility to dispose nuclear wastes should be constructed. The 51 locally defined fracture domains are thus grouped into seven global fracture domains presenting compatible fracturing properties. These domains statistically defined are finally compared to domains defined geologically. The results indicate that the proposed methodology leads to global domains with fracturing properties that are more consistent while geological domains are spatially better organized. We thus suggest that the classification process should be improved by including an additional criterion based on the physical distance between local sub-domains.

(1) Introduction

Observation of natural processes can be described probabilistically using probability density functions (PDF). A PDF represents the density distribution of a physical quantity over the sample space. A comparison between several of these quantities based on their pdf is often required but remains a non-trivial task. Thus, an objective method is required that is able to

find similarities and differences between objects. Such methods are used in a wide category of scientific fields like biology and biochemistry [Cheng et al., 2006; Datta and Datta, 2006], mathematics [Garai and Chaudhuri, 2004], communication [Kim and Yamashita, 2007], climate research [Vrac et al., 2007], soil analysis [Sarbu et al., 2007] and rock mechanics [Hammah and Curran, 1998; Hofrichter and Winkler, 2006]. The method we propose in this paper is relevant for all these fields but is applied here only fracture properties description in natural media.

We will first describe the mathematical framework of the comparison method, and then we specify the method for fracture density distributions and finally apply the method to field data. The paper ends with a discussion on validity and limitation of the method.

(2) Mathematical Framework

The proposed method is a classification method: it generates classes with specific characteristics in partitioning a space into different subspaces. It is inspired by the agglomerative hierarchical clustering algorithm [Chelcea et al., 2006] and generates disjoint classes to emphasize object groups with same properties.

(a) Sorting process algorithm

Let E be an object set containing n objects. We consider collection of non-void subsets of E called classes (for further details see [Gordon, 1999a]). If C is a class collection and X is an element of C , C is a valid division of E if C contains all the subset of E and for every X and Y of C , $X \cap Y = \emptyset$. The principle of the agglomerative hierarchical clustering is, from a class collection containing all singletons, to aggregate iteratively two classes until the largest degree of heterogeneity of element X in C reaches a certain value. Moreover, two classes X and Y are aggregated only if they are the nearest in the set. When two classes are aggregated, the distance between them, written as $d(X, Y)$, represents the degree of heterogeneity of $Z = X \cup Y$. The algorithm steps are described on Figure 3-25.

- (0) Set $i=0$ and define M_0 containing all the singletons of E .
- (1) Set $i=i+1$ and merge the two classes X_i and Y_i such as $\forall \{X, Y\} \subseteq M_{i-1}, d(X_i, Y_i) \leq d(X, Y)$.
- (2) Set $M_i = M_{i-1} \cup \{X_i \cup Y_i\} - \{X_i\} - \{Y_i\}$
- (3) Optimization of classes:
 $\forall \{X, Y\} \subseteq M_i, \forall x \in X, \bar{x} = \text{mean}(X), \bar{y} = \text{mean}(Y+x), \text{if } \max(d(x, \bar{x})) > \max(d(x, \bar{y}))$
 set $X = X \setminus x$ and $Y = Y + x$
- (4) Repeat (3) while it is effective.
- (5) Repeat (1) to (4) until $\forall \{X_i, Y_i\} \subseteq M_i, \max(d(X_i, Y_i)) \geq d_{\max}$, with d_{\max} a fixed criterion.

Figure 3-25 : Sorting process algorithm.

At the end of the classification procedure, we obtain a collection of classes C where all classes can be considered as different (with respect to d_{\max} , where d_{\max} is the stopping criterion) from the others and where all elements of a class can be considered as equivalent (with the same characteristic).

(b) *Coherence index*

Ideally, the resulting classification contains classes that do not overlap. In reality, however, some distributions might be associated with two or more classes. An efficient validity indicator of the classification is required to compute, for each distribution, its probability of belonging to all classes. In other words, distributions and sorting process are used initially to define classes with statistical properties expected to represent general trends over the total space. Next, the probability for each distribution to belong to each of the defined classes is considered. Accordingly, the coherence index is based on the probability, for a distribution, of belonging to a class. It expresses as follow:

$$p(x \in X, Y) = \frac{d^*(x, \bar{y})^{-1}}{\sum_i^{N_c} d^*(x, \bar{y}_i)^{-1}} \text{ where } d^*(x, \bar{y}) = \begin{cases} d(x, \bar{y}) & \text{if } d(x, \bar{y}) < d_{max} \\ 0 & \text{else} \end{cases} \quad (3-18)$$

where $d^*(x, \bar{y})$ is the distance between a distribution x and a Class Y denoted by its mean distribution \bar{y} , N_c is the number of class and \bar{y}_i refers to the mean distribution of class Y_i . These probabilities are meant to evaluate the resulting classification. We compute a coherence index for each class X with the formula:

$$IC(X) = \frac{1}{N_x} \sum_{i=1}^{N_x} p(x \in X, X) \quad (3-19)$$

where N_x is the number of distributions in class X . This coherence index corresponds to the average of the membership probability of distributions in class X and denotes the robustness of the class. Similarly, this coherence index can be used to calculate the overlapping between two different classes X and Y :

$$IC(X, Y) = \frac{1}{N_x} \sum_{i=1}^{N_x} p(x \in X, Y) \quad (3-20)$$

This second index corresponds to the probability of the distributions of class X to belong to class Y . We use also a complementary index, IC_w , as indicator of robustness and overlapping:

$$IC_w(X, Y) = \sum_{i=1}^{N_x} p(x \in X, Y) \quad (3-21)$$

This second parameter provides a complementary piece of information that accounts for the statistical weight (or number of distributions) of each class.

Once a classification is performed, the class robustness is assessed further through computation of these indexes. The real robustness of a class X arises from the comparison between $IC(X)$ and $IC(X, Y)$ for all available classes. If $IC(X)$ is significantly larger than any $IC(X, Y)$, the class is clearly differentiated and its intern consistency is given by the value of $IC(X)$.

Figure 3-26 illustrates the procedure on a very simple example. Sixty values are selected and the distance is taken as the difference between values. Coherent classes are obtained, which components depend on d_{max} (here equal to 4.0 to have an illustrative number of classes). The coherence indexes IC and IC_w for these classes are given in Figure 3-27. We observe high values for all classes, meaning that the classification is correct. Classes 1 and 2 have smaller indexes because of a possible overlapping between them. The boundary between the two classes could indeed be moved without questioning $d_{max} = 4.0$. In this case, the optimization procedure plays a major role (step 3 in Figure 3-25). IC_w provides an indication of the

statistical weight of each class. Thus, classes 1 and 2 have very larger index as they are the most significant classes according to their numbers of elements while class 4 is insignificant as it contains only one element, although its IC is very high.

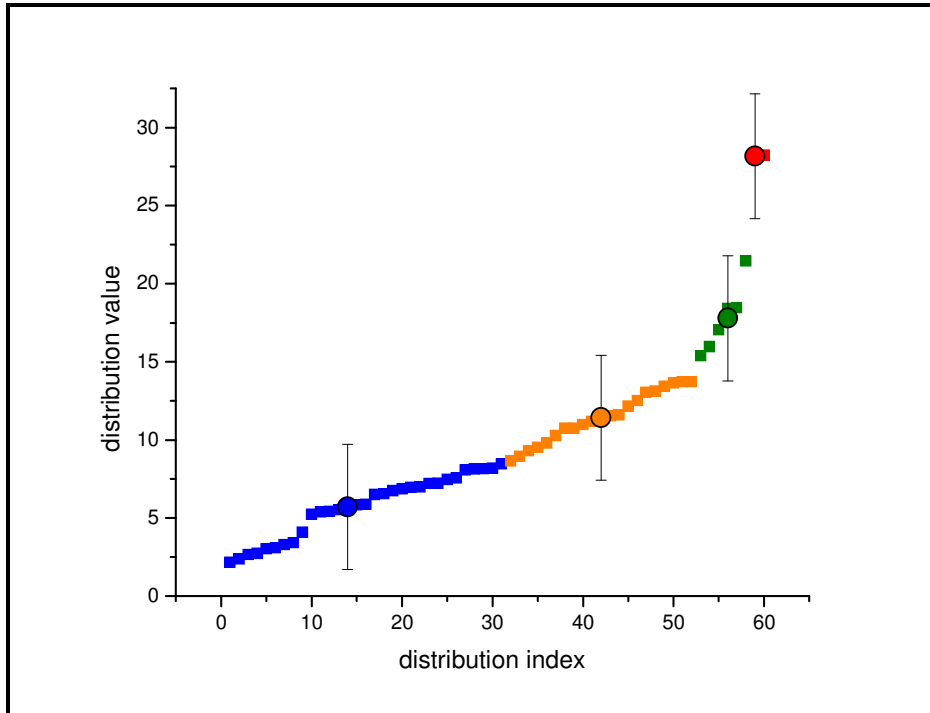


Figure 3-26 : Illustration of the resulting classification on single points. With a $d_{\max}=4$ the points are distributed into 4 classes represented with different colors. Circles represent class averages and the associated error bar is equal to d_{\max} . We observe that the classification is consistent but the limit between two classes might be slightly changed without exceeding d_{\max} .

Class X	$IC(X)$	$IC_w(X)$
Class 1 (blue)	0.91	28.10
Class 2 (orange)	0.82	17.22
Class 3 (green)	0.94	5.61
Class 4 (red)	1	1

Figure 3-27 : Coherence indexes for the classes illustrated on Figure 3-26. All coherence indexes are high, meaning that the resulting classes are consistent. Classes 1 and 2 have smaller coherence indexes because of a possible overlap between them.

(3) Application to fracture density function

Fracture density functions are a relevant way to describe fracturing properties of natural media. Comparing these functions is required to build a 3D-discrete fracture network (DFN) model.

(a) Description of the application framework

Natural fractured media can be described using the Discrete Fracture Network models. We have chosen the power scaling models. Such models are formalized in [Davy *et al.*, 1990], and later further elaborated in [Bour *et al.*, 2002]. In this framework, fracturing properties are described with a fracture density distribution $N_{3D}(L, l, \theta, \varphi) = \alpha_{3D}(\theta, \varphi) \cdot l^{-(k_{3D}+1)} \cdot L^{D_{3D}}$

where L is the system size, l the fracture length, θ the fracture strike and φ the fracture dip. The length term $l^{-(k_{3d}+1)}$ is the scaling term whereas $\alpha_{3D}(\theta, \varphi)$ is the density term with respect to orientation. This term is relevant to compare different domains with different properties.

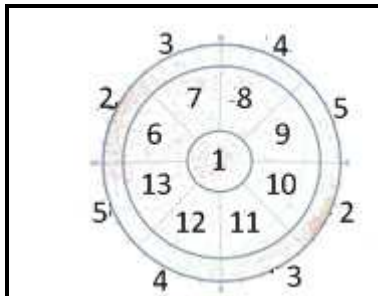


Figure 3-28 : Discretization of the stereonet into 13 pole zones. The density distribution $\alpha_{3D}(\theta, \varphi)$ is computed for each zone.

The density term $\alpha_{3D}(\theta, \varphi)$ represents the fracture density for a fracture of size 1 and can be derived from borehole and outcrops data. Stereological formula can be used to extend 1D and 2D data to the 3D model. First, $\alpha_{1D}(\theta, \varphi)$ depends on the fracture orientation (strike and dip). The fracture pole orientations, and their uncertainties, are calculated from orientation of intersecting fractures and borehole parameters. The density term mentioned above is function of orientation: in practice, the (θ, φ) in $\alpha_{3d}(\theta, \varphi)$ refers to the subdivision of a dataset according to orientation.

The binning over orientations reflects the degree of detail given to the relation between fracture density and orientations. It can be used to discriminate between fracture sets. However, its level of detail relies greatly on the necessity to have a statistically relevant number of fracture in the bins, this to be able to derive a sound calculation of $\alpha_{3d}(\theta, \varphi)$. The selected division encompasses 13 zones, distributed as illustrated on Figure 3-28. This is a regular binning where all bins have the same area and allowing a simultaneous consideration of dip and strike. In this case, we obtain α_{3d} such :

$$\alpha_{3D} = \iint_{\theta, \varphi} \alpha_{3D}(\theta, \varphi) \cdot d\theta \cdot d\varphi \quad (3-22)$$

Here, uncertainty arises from measure precision. Accordingly, the density term $\alpha_{3d}(\theta, \varphi)$ is associated with its uncertainty, noted $\sigma_{unc}(\theta, \varphi)$, such that the effective density associated to a dataset is simply $\alpha_{3d}(\theta, \varphi) \pm \sigma_{unc}(\theta, \varphi)$. The uncertainty term is computed by assuming a normalized, centered Gaussian law for each fracture intercept. This leads to:

$$\alpha_{3D}([\theta_1, \theta_2], [\varphi_1, \varphi_2]) = \sum_{i=1}^N [p_i \cdot \psi(\omega_i)]$$

$$\sigma_{\alpha}([\theta_1, \theta_2], [\varphi_1, \varphi_2])^2 = \sum_{i=1}^N \left[p_i \cdot (1 - p_i) \cdot \psi(\omega_i)^2 + \left(p_i \cdot \frac{d\psi}{d\omega}(\omega_i) \cdot \sigma_{\omega_i} \right)^2 \right] \quad (3-23)$$

with $p_i = \frac{1}{2} \left[\operatorname{erf} \left(\frac{\theta_2 - \theta_i}{\sigma_{\theta_i} \sqrt{2}} \right) - \operatorname{erf} \left(\frac{\theta_1 - \theta_i}{\sigma_{\theta_i} \sqrt{2}} \right) \right] \cdot \frac{1}{2} \left[\operatorname{erf} \left(\frac{\varphi_2 - \varphi_i}{\sigma_{\varphi_i} \sqrt{2}} \right) - \operatorname{erf} \left(\frac{\varphi_1 - \varphi_i}{\sigma_{\varphi_i} \sqrt{2}} \right) \right]$

where $\psi(\omega)$ represents a correction term related to orientation bias [Davy *et al.*, 2006].

(b) Basic class definition for fracture fields

During the classification process, sets of distribution are grouped together and replaced with the class average. We have to ensure that comparing class averages and individual elements is similar. So the following definitions have been selected. A class of distributions X , containing N_x elements, is defined as a set of distributions $\{\alpha_{3D}^i, i=1 \dots N_x\}$. Its mean value, α_{mean} , is the average value of α_{3d}^i weighted by $(1/\sigma_{\alpha_{3D}^i})$ the inverse of the uncertainty on α_{3D}^i :

$$\alpha_{mean} = \frac{\sum_{i=1}^{N_x} \frac{\alpha_{3D}^i}{\sigma_{\alpha_{3D}^i}}}{\sum_{i=1}^{N_x} \frac{1}{\sigma_{\alpha_{3D}^i}}} \quad (3-24)$$

The standard deviation associated to this distribution, σ_{mean} , is given by:

$$\frac{1}{\sigma_{mean}} = \frac{1}{N_x} \sum_{i=1}^{N_x} \frac{1}{\sigma_{\alpha_{3D}^i}} \quad (3-25)$$

The above definitions guarantee the conservation of the final properties of a class that would encompass all the distributions: from any stage of the sorting process, the mean distribution resulting from the merging of all distributions ends up to a unique global distribution, $(\alpha_{mean})_{TOT}$, where:

$$(\alpha_{mean})_{TOT} = \frac{\sum_{i=1}^{N_x} \frac{\alpha_{3D}^i}{\sigma_{\alpha_{3D}^i}} + \sum_{i=N_x+1}^{N_{TOT}} \frac{\alpha_{3D}^i}{\sigma_{\alpha_{3D}^i}}}{\sum_{i=1}^{N_x} \frac{1}{\sigma_{\alpha_{3D}^i}} + \sum_{i=N_x+1}^{N_x} \frac{1}{\sigma_{\alpha_{3D}^i}}} = \frac{N_x \frac{\alpha_{mean}}{\sigma_{mean}} + \sum_{i=N_x+1}^{N_{TOT}} \frac{\alpha_{3D}^i}{\sigma_{\alpha_{3D}^i}}}{N_x \frac{1}{\sigma_{mean}} + \sum_{i=N_x+1}^{N_x} \frac{1}{\sigma_{\alpha_{3D}^i}}} \quad (3-26)$$

In the above equation, the distributions contribute equivalently either individually with their α_{3D}^i and $\sigma_{\alpha_{3D}^i}$, or through the class to which they belong with α_{mean} and σ_{mean} . Accordingly, the sorting process defines classes whose equivalent parameters (mean and standard deviation) have the following properties.

- Within a class, the weight attributed to each distribution, in order to define the mean density, is inversely proportional to the standard deviation. In other words, distributions with relatively small standard deviations are more significant; conversely, the contribution of distributions with very large standard deviations is insignificant in the definition of the mean trend of a class.
- The standard deviation associated with a class arises from the harmonic mean of the standard deviations of the density distribution. Accordingly, it tends to reflect the smallest standard deviations.

Therefore, the more distributions are included in a class, the more precise is the resulting definitions of the class parameters.

(c) *Definition of a distance*

The distance between two distributions is defined to reflect their probability of belonging to the same parent distribution. It requires that both mean and standard deviations of a distribution be defined. It is quantified with the difference between density values, normalized by their related standard deviations. First, for two distributions α_{3D}^m and α_{3D}^n , the distance is defined by:

$$d(m, n) = \frac{1}{N_z} \sum_{i=1}^{N_z} \frac{|(\alpha_{3D}^m)_i - (\alpha_{3D}^n)_i|}{(\sigma_{\alpha_{3D}^m})_i + (\sigma_{\alpha_{3D}^n})_i} \quad (3-27)$$

where N_z refers to the number of grid-pole zones used to discretize the stereonet, in our case 13 (Figure 3-28).

Thus defined, distance can be computed between two basic distributions, or a distribution and the equivalent distribution associated to a class. In addition, an internal distance d_c is associated to each class as the largest distance between a class element and the class average. It reflects the internal class cohesion.

(d) *Sorting procedure and critical threshold*

During the sorting process (Figure 3-25), the class building procedure is iterated as long as the most-dispersed class displays an internal distance, d_c , smaller than a threshold value d_{\max} . Distributions are aggregated as long as the criterion is fulfilled.

The critical threshold reflects the distance below (or above) which two distributions (or two classes) are set similar (or not):

- If d_c is very small (if the gap between values is much smaller than the associated uncertainties), then both distributions are statistically undistinguishable.
- Conversely, if the distance is very large, both distributions are statistically distinct.

In practice, having d_{\max} smaller or equal to 1 reflects that standard deviations do overlap. Thus, fixing a threshold below 1 is more restrictive, and, in contrast, accepting a larger d_{\max} tends to aggregate distributions according to a relative closeness (within the total set of data).

(4) *Practical use of the method*

The class-identification process is applied to field data provided by the Swedish national agency for nuclear wastes (Svensk Kärnbränslehantering AB, SKB).

(a) *Description of the available data*

SKB investigation area is located in Sweden and covers an area of around 11 square kilometers. In this area 28 cored borehole were drilled (Figure 3-29-a). The borehole lengths range from a hundred of meters up to a thousand of meters. The mean borehole orientations, taken as the direction defined by the first and last fracture intercept, are mostly close to vertical. Core observations and borehole logging provide information about rock matrix properties and fractures occurrences and complexity. When fracture intercept identification is possible, main information as relative and absolute position, strike and dip of the fracture plane and the confidence level are provided. Rock units are delimited from single borehole interpretation. From these data, we calculate one fracture density function per rock unit per borehole, which yields 51 functions (Figure 3-29-b). Each distribution is named using its borehole (e.g. KFM01A) and its local fracture domain (e.g. FFM01). The density-function classification is a way to find similar distributions corresponding to geographical area with similar fracturing properties in order to integrate them into a 3D DFN model.

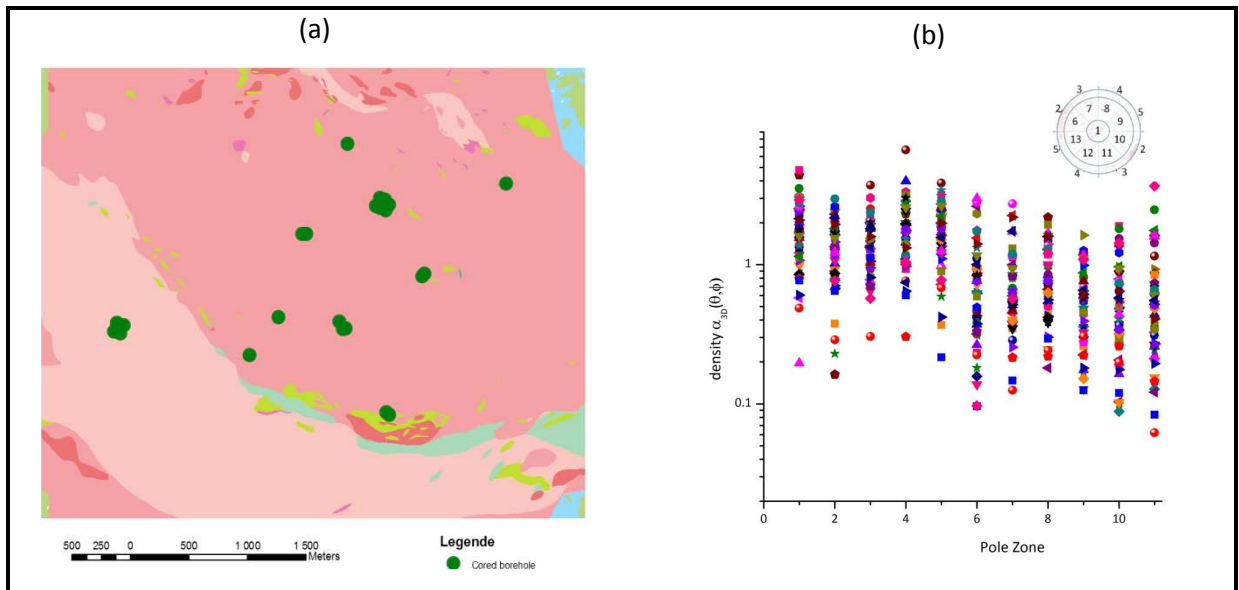


Figure 3-29 : Available data on Laxemar (Sweden). Top is the localization of cored borehole and down are the resulting density distributions. Each borehole is split into rock units and we calculate one density distribution per rock unit.

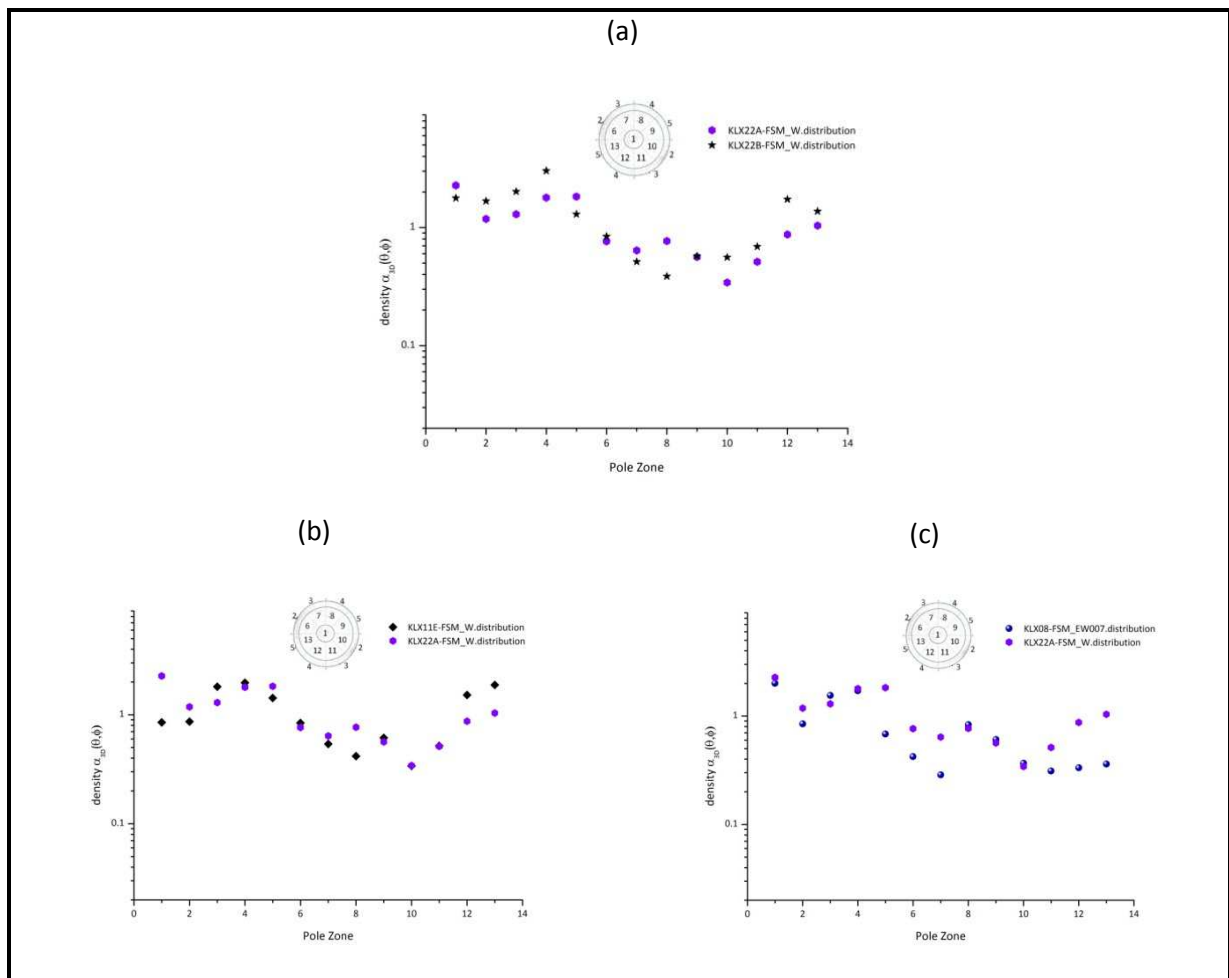


Figure 3-30 : Illustration of the stopping criteria d_{max} . With $d_{max}=1.0$ the two distribution on the top will belong to the same class because all $\alpha_{3d}(\theta, \phi)$ values are similar. The two distributions on the bottom-left will belong to different classes because $\alpha_{3d}(\theta, \phi)$ values are different for the central zone (1) and the middle zones (12-13). On the bottom-right, the two

distributions have different shape and the classification procedure separate them into different groups.

(b) Classification parameters

We apply the algorithm of Figure 3-25 to the data. Selecting an appropriate value for d_{max} is not trivial. As explained in the previous section, the ideal value for d_{max} is 1.0. Using this value means that the resulting classes will contain only identical distributions, according to the associated uncertainties (Figure 3-30). For example, on Figure 3-30-a, the two distributions are much closed to each other's and will be grouped together by the classification algorithm. However, the two distributions on Figure 3-30-b will be considered as different because of the density in zones 1, 12 and 13, although the global trend is similar. The two distributions on Figure 3-30-c will be grouped into two different classes, as their shapes are different, leading to high differences in most of the zones. Besides this criterion, the evolution of the number of classes while increasing d_{max} can be used to define a relevant threshold. Figure 3-31 displays the evolution of d_{max} during the classification process from the beginning (1 class per distribution) to the larger class (all the distribution into the same class). Variations of d_{max} with the class number are rather smooth. This reflects the continuous nature of the fracturing throughout the site, rather than a clear discrimination into domains with very different statistical properties. For two adjacent values ($n; n + 1$), a significant shift of d_{max} values reinforces the relevancy of classification " $n + 1$ classes" with regard to the " n classes". This is illustrated in Figure 3-31 between 5-6 classes on one side and 7 on the other. So the optimized classification is chosen according to this occurrence of a drop ($d_{max}=0.8$).

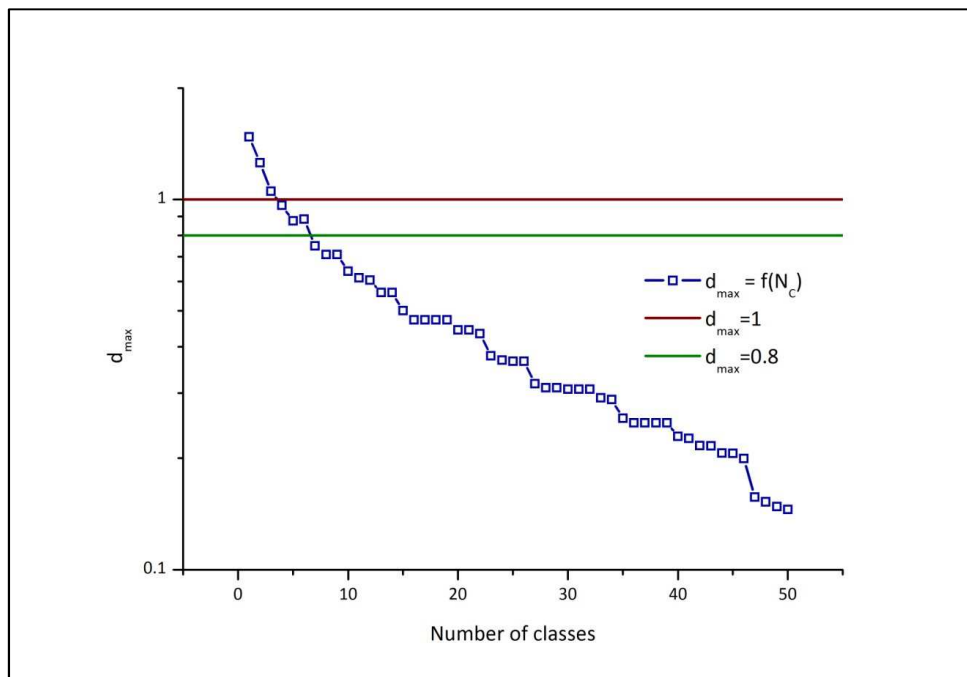


Figure 3-31 : evolution of the larger heterogeneity in a class as a function of the number of classes. A value $d_{max} = 1.0$ means that in a class we have only identical distributions accounting for uncertainty. A value $d_{max} = 0.8$ leads in our case to seven classes and seems relevant to find similar fracture domains.

(c) Resulting classification

Figure 3-32 displays the resulting classes after classification. In general, the mean class distributions display higher densities in the central zone (zone 1, gathering subhorizontal

fractures) and in the subvertical zones (zones 2, 3, 4 and 5). In other words, fracturing is dominated by subhorizontal or subvertical direction for the fracture pole orientations.

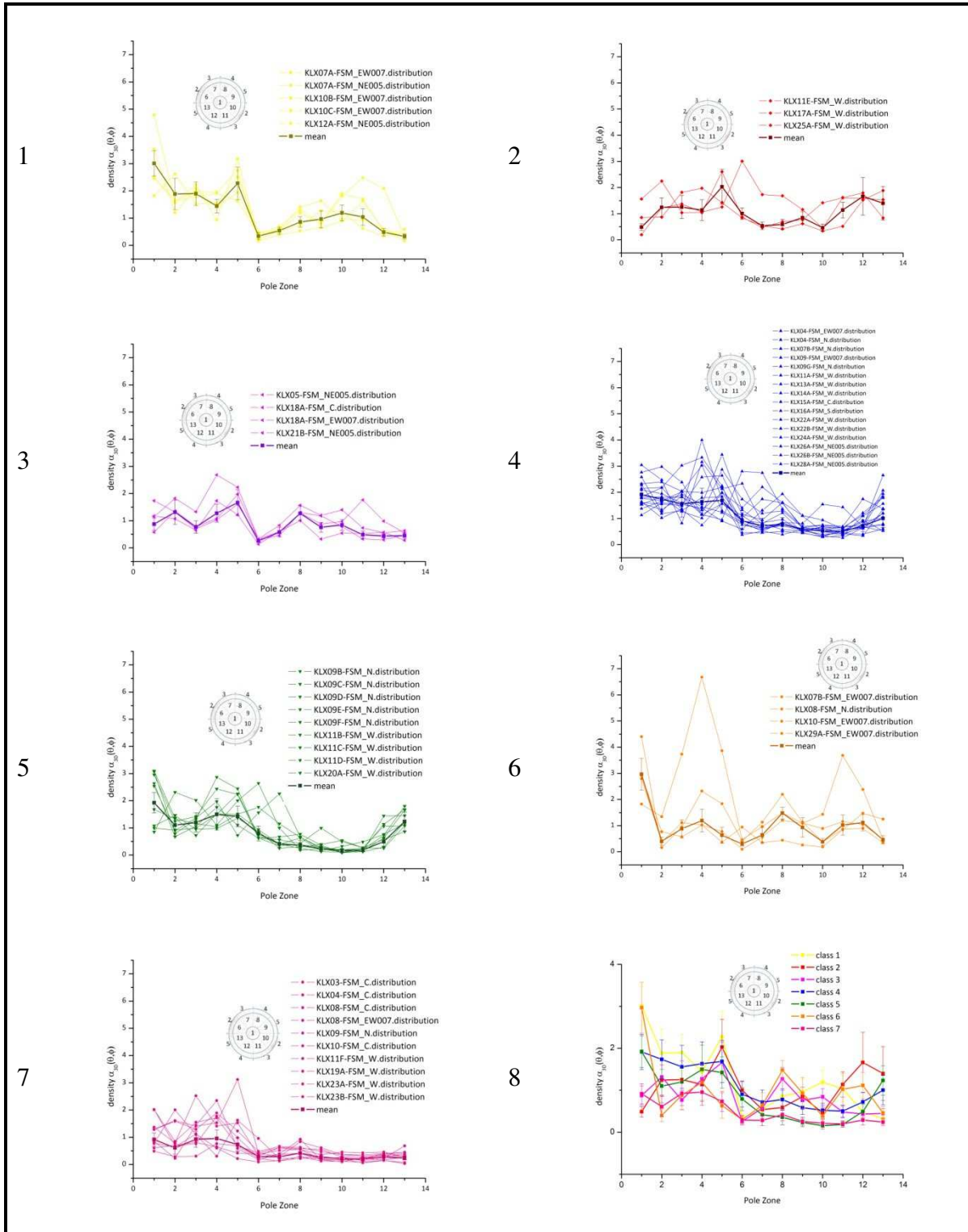


Figure 3-32 : Resulting classification of the 51 available distributions. Using a stopping criteria $d_{\text{max}}=0.8$ leads to seven different classes (1-7). Class means are summarized in (8).

In addition, most classes thus formed display local agreement and discrepancies from one another. Two sets of classes can be defined. The first set includes classes 7 and 3, which exhibits mostly lower densities than other classes, especially in the central and middle zones than the second set. The second set encompasses the remaining classes, where the discrimination is based on differences in orientations rather than in the global density. Note that classes 5 and 7 display very similar profiles according to orientation values and differ essentially through a density shift (with a factor roughly equal to 2.5) that prevents them from being merged into one single class. The coherence index IC (Figure 3-33) indicates indeed that classes overlaps, which again highlights the difficulty of defining clear boundaries between classes or domains. The weighted coherence index (Figure 3-33) indicates however that the largest classes (4 and 7) are more consistent than the smallest classes (2 and 3), pointing out that the classes with few elements might belong to transition areas between fracture domains.

<i>Class name</i>	<i>Number of elements</i>	<i>IC(class c)</i>	<i>ICw(class c)</i>
Class 1	5	0.46686	2.33431
Class 2	3	0.449	1.34701
Class 3	4	0.3136	1.25439
Class 4	16	0.39179	6.26862
Class 5	9	0.38596	3.47365
Class 6	4	0.53659	2.14638
Class 7	10	0.46346	4.63456

Figure 3-33 : Ic and ICw of all classes resulting from a classification with $d_{max}=0.8$. IC values indicates that there is overlapping between classes. ICw indicates that the classes with the most elements (class 4 and class 7) the most consistent while classes with a few numbers of distributions (class 2 and 3) are less significant in terms of fracture domain.

(d) *Class robustness*

As defined in equation (3-23), the probability of each distribution to belong to a given class is computed. Even if each distribution does belong to a particular class, it could have a non-null probability of belonging to another class. Computations are based on distances from one class to another. In order to get a first overview of the coherence analysis, the probability of each distribution (y-axis) belonging to one of the classes (x-axis) is plotted in the form of a contour map (Figure 3-34). In the plot, high probabilities of occurrence are represented in red and lower probabilities in blue. In addition, the representation is such that distributions from one class are grouped (see the figure legend for details). If most of the distribution have a clearly higher probability of belonging to the class identified by the classification process, belonging some distributions have a significant probability of belonging to another class (e.g. distributions n°25-28), which explains the overlap between classes. Finally, we quantify the classification strength by considering the coherence parameters defined in equation (3-20) and (3-21). Both parameters arise from a combination of the above-mentioned probabilities; they are plotted in Figure 3-35.

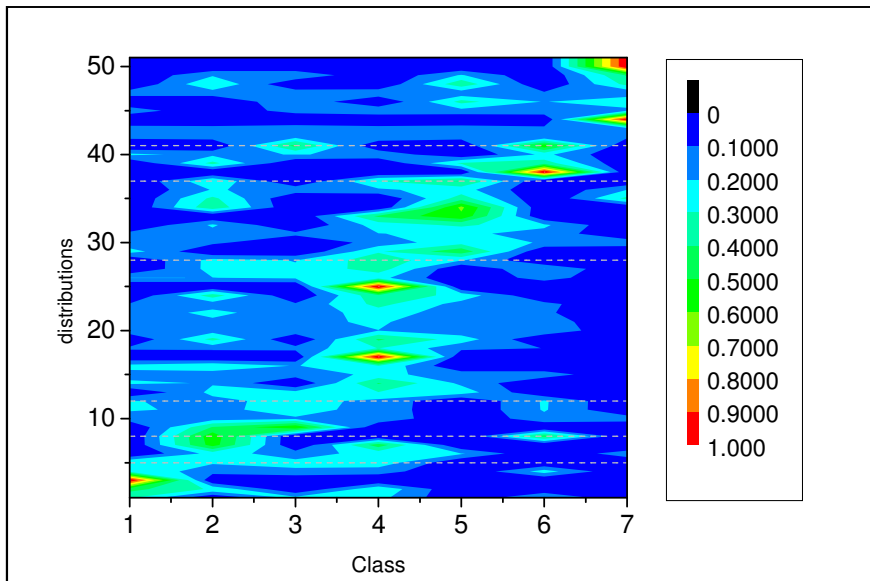


Figure 3-34 Contoured plot of the probability for a distribution (y-axis) to belong to each class defined (x-axis). Distributions from 1 to 51 are sorted according to their class of reference, from 1 to 7 (grey dashed-lines indicate the limit between classes — i.e. distribution rows 1 through 7 refer to Class 1 and so on).

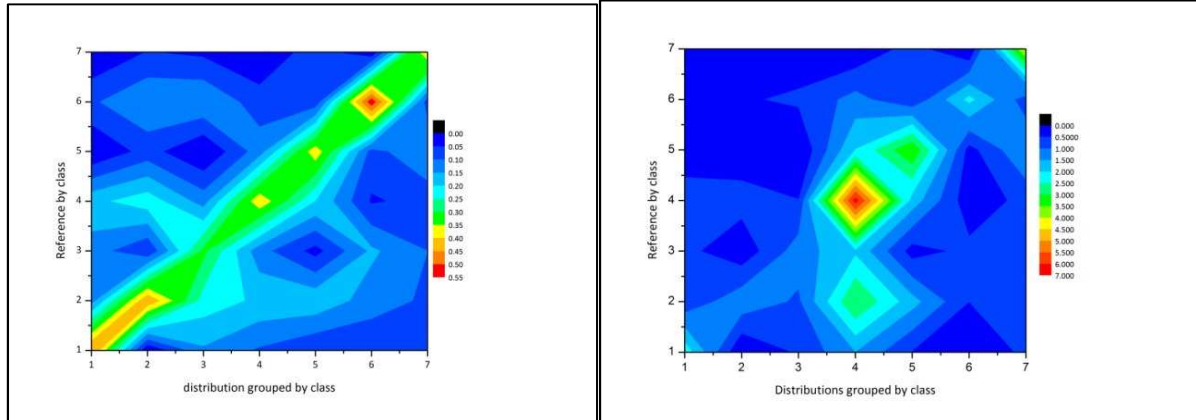


Figure 3-35 : Index of coherence IC by class (left) and weighted index (right). IC highlights the possible overlapping between classes (e.g. the high IC between classes 2 and 3) while ICw refers to the total number of distributions remaining in each class, accounting for overlapping.

The *IC* criterion expresses a degree of confidence in the class definition without encompassing the basic number of distributions involved in each class. Thus, *IC* neglects the initial total number of distributions encompassed into a class. Indeed, if a class is associated with a single distribution, its mean distribution will be exactly equal to the distribution, leading to a probability of 1 to belong to the class. In contrast, a class containing several distributions will inevitably encompass more variability, which presumably will decrease its internal coherence. Therefore, the second consistency index, IC_w , is weighted by the respective number of distributions into a class. This gives a direct level of representativeness for each class from the initial set of distributions and allows comparison of various classifications from an initial set. It refers to the total number of distributions remaining in each class when overlapping is taken into account. Again, classes 4 and 7, which contain the most of distributions, are the best-defined classes.

(5) Comparison with geological fracture domain

A classification based on geological assumptions can be considered and compared to the previous classification.

(a) Geological fracture domain

Geological fracture domains have been defined by SKB geologists [Olofsson *et al.*, 2007]. They have taken into account different deterministic geological aspects: homogeneity in rock domains, fracture frequency distribution patterns and geology. Using these criteria, six separate domains are recognized. Figure 3-36 represents the density distribution of the geological domains. We can notice that the resulting classification is different from the classification based on purely statistical assumptions.

The two classifications are different. The geological fracture domains are based on larger amount of criteria than the statistical classification based only on the density distributions. However, some points are noticeable. A first remark is that, in the statistical classification, different borehole parts may belong to the same class while they do not belong to the same local fracture domain. For example, class n°3 contains KLX18A-FFM_C and KLX18A-FFM_EW007. The separation into different local domains could be explained by geological criteria (like rock properties) that are not taken into account in the statistical classification while they are similar regarding their orientation distribution. Accounting for this effect, the fracture domain FFM_N and FFM_W are combined into classes n°4 and n°5 in the statistical classification, which question their definition. Smaller fracture domains like FFM_NE005 are completely included in other classes, meaning that their relevancy might be discussed. The coherence indexes of this classification (eq. (3-20) and (3-21)) are given in Figure 3-37. We notice that the orientations of the domains defined with geological settings are poorly consistent.

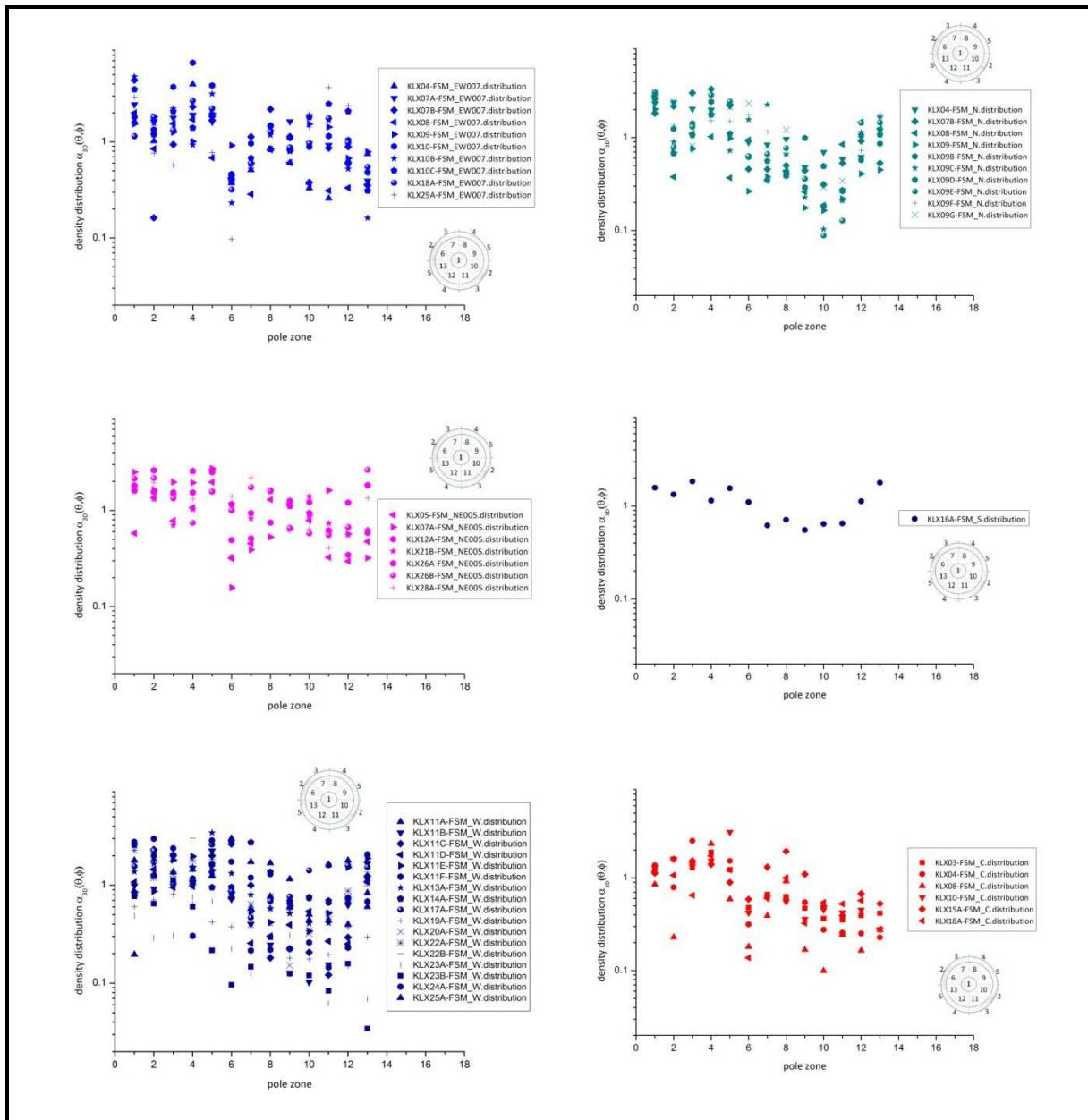


Figure 3-36 : Classification based on geological assumptions. Only six classes are defined. In some fracture domains, for example FSM_W, the orientations are clearly heterogeneous.

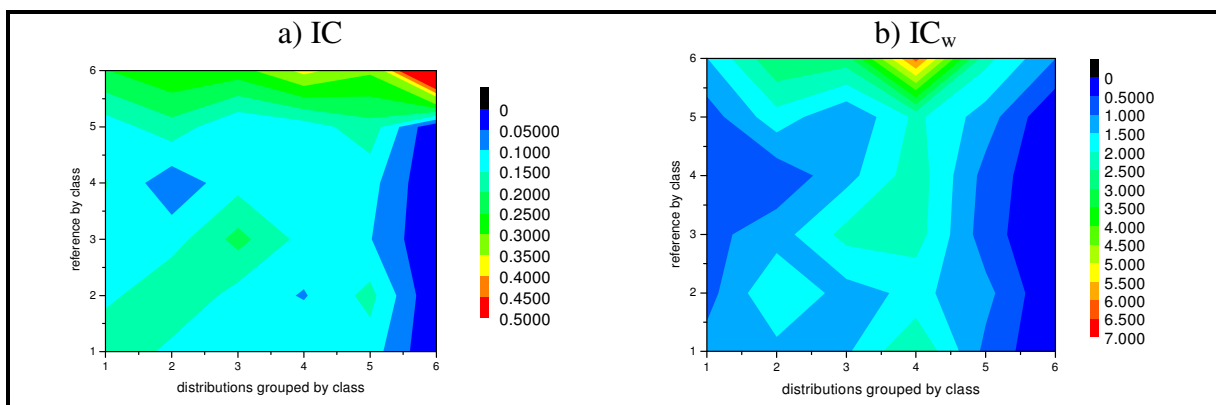


Figure 3-37: Coherence indexes for the classification based on geological settings.

(6) Discussion

The proposed classification method is a useful tool to compare physical properties at different space location in order to find a consistency on larger scale. This classification, based only on statistical assumptions, can be compared to classical interpretation based on other physical criteria.

Figure 3-38 displays a projection of all the studied distributions in the 2D space using multidimensional scaling (MDS) [Cox and Cox, 1994]. MDS is a widely used for the visualization of data and consists in mapping high-dimensional data to a low dimensional representation, mainly 2D or 3D. The Matlab Toolbox for Dimensionality Reduction has been used to compute the 2D coordinates [van der Maaten, 2007].

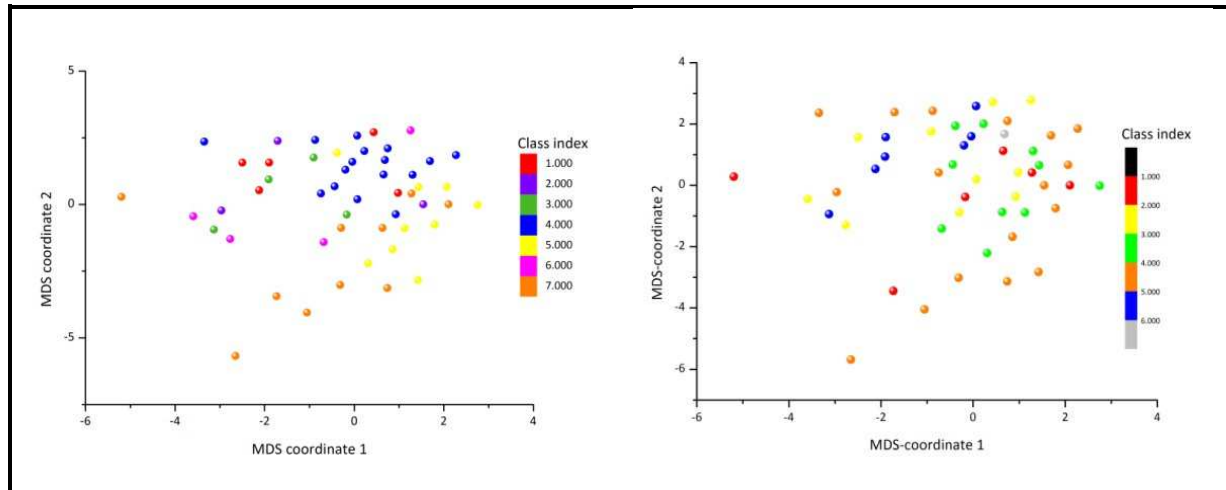


Figure 3-38 : Representation of all density distribution in a 2D space using multidimensional scaling. (left) points are colored according to the statistical classification; (right) points are colored according to the geological classification.

The proposed classification method is efficient on non-overlapping clusters. In our case the resulting classification must be discussed using the given coherence indexes. Too small indexes indicate that the founded classes may overlap and that a refinement of the classification may be required. As displayed on Figure 3-38 the founded classes are consistent for distribution far from each other (classes n°4 and n°7) and contain "transition" classes (e.g. class n°5) between major classes. We observe that the geological fracture domains are not consistent with the statistical distribution, including for extreme values, while they are spatially consistent –i.e. distributions located closed to each other's are more likely to belong to the same class. The definition of a fracture domain might be also adjusted to account for the statistical classification and the statistical classification should be improved by including, in the definition of the distance between distributions, an additional term accounting for the physical distance between data sets.

The main difference between the method proposed in this article and classical clustering method is that the uncertainty in values is taken into account at each step to include in the analysis data uncertainty and natural variability. That is why overlapping effects are more important in our case, but including these factors, particularly natural variability, is a critical task in natural media analysis, to set up consistent 3D DFN models.

(7) Conclusion

We propose in this paper a method to compare automatically a large number of distributions. These distributions are grouped together into similarity classes, where the similarity index is

defined by the larger distance between a distribution and the class mean. This method is inspired by hierarchical clustering methods but includes uncertainty. Including uncertainty (due to data sampling uncertainties or natural variability) is indeed a critical task. Side effects are that the resulting classification contains classes that may overlap.

A comparison between the proposed classification and a classification based on geological assumption is also described. A conclusion is that the geological fracture domains encompass more criteria but may underestimate the density distribution properties. As a result, we suggest a method that is based on statistical properties, with a refinement accounting for geological properties.

3. Résultats complémentaires

L'analyse du terme de densité est appliquée de la même manière au site de Forsmark. Sur ce site, 60 domaines locaux de fracturation ont été définis par SKB. L'évolution du nombre de classes en fonction de d_{max} montre différence significative entre $d_{max}=0.9$ et $d_{max}=1.0$. La méthode de classification est donc appliquée avec une valeur de $d_{max}=0.9$. Le résultat donne 11 classes différentes, dont les index de cohérence IC et IC_w sont illustrés à la figure 3-39. Si le chevauchement entre classe est moins prononcé que pour le site de Laxemar, en partie de part le plus grand nombre de classe, la cohérence de chaque classe est moins importante. Cette remarque, comme le fait d'obtenir plus de classes, dénotent de la plus grande hétérogénéité du site de Forsmark. On retrouve donc les difficultés observées précédemment concernant la définition des limites entre classes de fracturation.

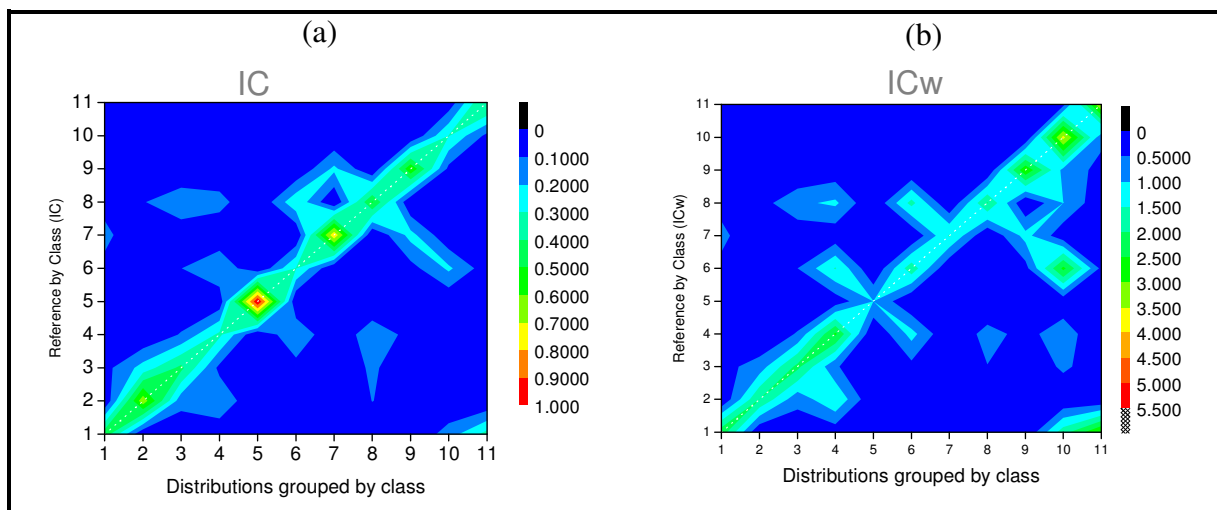


figure 3-39. Index de cohérence pour la classification finale des distributions du site de Forsmark ($d_{max} = 0.9$, 11 classes).

La comparaison avec les domaines de fracturation définis à partir de critères géologiques montre une fois de plus que les domaines définis statistiquement, même s'ils manquent de cohérence spatiale, présentent des distributions d'orientation plus homogènes. En effet, la cohérence statistique des domaines définis géologiquement (voir figure 3-40), est très faible, en particulier concernant IC . De plus les valeurs de IC_w montrent que les distributions regroupées dans les domaines 1 et 2 sont très hétérogènes et pourraient, en ce qui concerne le terme de densité, appartenir à n'importe quelle classe.

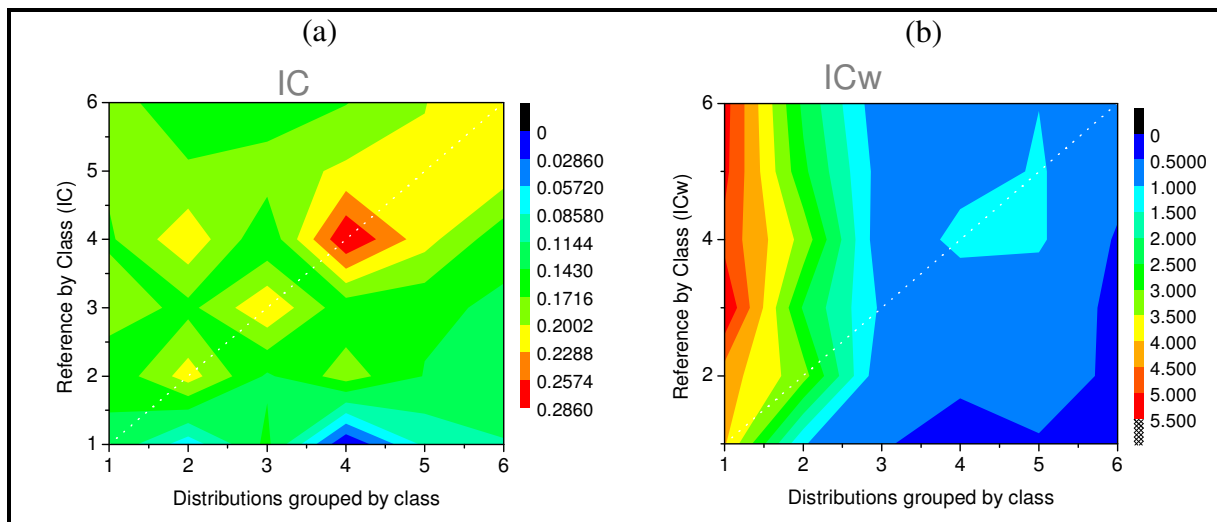


figure 3-40 : index de cohérence IC (a) et ICw (b) des domaines de fracturation définis à partir de critères géologiques pour le site de Forsmark.

4. Conclusions sur l'analyse du terme de densité

Le grand nombre de données disponibles sur le site de Forsmark et Laxemar ont permis de mener une étude approfondie du terme de densité en fonction de l'orientation des fractures. A partir des observations le long de forage et de la définition de domaines locaux de fracturation, on a été amené à définir, pour chaque domaine, le terme de densité $\alpha_{3d}(\theta, \varphi)$ du modèle statistique en incluant à la fois l'incertitude liée aux imprécisions inhérentes à l'échantillonnage des données et l'incertitude liée à la variabilité du milieu naturel. Afin de définir un modèle à l'échelle du site, les domaines définis à l'échelle du puits ont été regroupés par classes en fonction de la similarité de leur $\alpha_{3d}(\theta, \varphi)$. Il s'avère que la principale difficulté lors de ce regroupement provient de la variabilité de la fracturation des sites étudiés et de l'absence de rupture nette entre les différents domaines à l'échelle du site. Néanmoins, il reste possible de distinguer les différentes zones dont les propriétés de fracturation sont très différentes, même si les frontières entre zones restent floues. Comparée aux méthodes classiques de définition des domaines de fracturation, basées sur les observations géologiques et les principaux traits de la fracturation (densité globale, orientation dominante), la méthode stochastique donne des classes où la fracturation est plus homogène à l'intérieur d'un domaine et mieux différenciée par rapport aux autres domaines. Ces résultats permettent donc de mieux contraindre les paramètres du modèle de réseau de fracture. Le principal désavantage de la méthode est de ne considérer que les propriétés statistiques des domaines et pas leur position géographique. Inclure des critères sur la distance entre domaines par exemple devrait permettre de mieux contraindre les résultats et de mieux définir les frontières entre les différentes zones.

E. Conclusions sur la définition d'un modèle de site

La modélisation de la fracturation d'un site naturel par un réseau stochastique de fractures discrètes requiert la détermination de deux paramètres : le paramètre d'échelle, constitué de l'exposant de la loi de puissance modélisant la distribution des longueurs de fractures, et le paramètre de densité, qui définit le nombre de fracture à l'échelle du mètre pour chaque classe d'orientation. Ces deux paramètres sont calibrés à l'aide des données les plus pertinentes parmi les données disponibles. Le coefficient de la loi de puissance doit être déterminé à l'aide de données couvrant des échelles sur plusieurs ordres de grandeur afin de déterminer les modèles locaux et régionaux. Les données les plus précises concernant les longueurs des

fractures proviennent des affleurements, où les traces de fractures sont cartées précisément avec leurs longueurs et leurs orientations. Elles sont complétées par les données provenant des linéaments, c'est-à-dire des grandes failles et zones de déformation localisées et mesurées par photos aériennes ou imagerie radar. Concernant le terme de densité, les données les plus précises sont les données recueillies le long des forages. Comme ils couvrent plusieurs centaines de mètres, la quantité de données accessibles est très importante et permet de définir précisément la densité sur un grand nombre de classes d'orientation. Les données disponibles sont généralement biaisées à cause des conditions d'observation. La taille de l'affleurement pour les longueurs de fracture ou l'orientation du puits pour les orientations influent directement sur la capacité à observer certaines fractures. De plus, sur un affleurement, les fractures peuvent être artificiellement déconnectées à cause de la topographie ou de l'érosion. Il est donc nécessaire de compenser ces biais soit par des méthodes statistiques, soit par des méthodes déterministes. Sous-estimer ce biais peut conduire à sur ou sous-estimer les paramètres du DFN jusqu'à 12%.

Les données sont observées à une échelle locale, particulièrement en ce qui concerne la densité de fracturation. Il est donc nécessaire de faire un changement d'échelle entre les paramètres définis localement et le modèle régional. Ce changement d'échelle repose sur le paramètre d'échelle, c'est-à-dire l'exposant de la loi de puissance, qui va permettre de déduire les intensités de fracturation à la plus grande échelle à partir des données des échelles inférieures. De plus, il est nécessaire de regrouper les modèles locaux en fonction de leurs similarités et de déterminer lesquels appartiennent au même modèle globale. Si l'hétérogénéité du site conduit à calculer des paramètres avec une forte variabilité, elle rend également compliquée la définition des domaines de fracturation à l'échelle du site. S'il est en effet possible de caractériser les domaines où la fracturation est très différente, les frontières entre ces domaines sont incertaines à cause, principalement, de la continuité des changements de propriétés et de l'incertitude sur celles-ci. L'hétérogénéité reste pour ces raisons, la difficulté majeure à laquelle le modélisateur est confronté. Il est à noter également que les données fournies par SKB sont d'une qualité remarquable et que la quantité d'information recueillie sur le site est très importante. C'est pourquoi nous avons pu mener une étude aussi précise des propriétés de la fracturation et obtenu des résultats statistiquement pertinents. Si la quantité et la qualité des données étaient moins importantes, il aurait alors fallu utiliser des hypothèses simplificatrices afin de mieux contraindre les paramètres du modèle de DFN.

Pour terminer, il est nécessaire de rappeler que la définition d'un modèle de site est réalisée dans une optique particulière. Dans le cas de l'étude présentée ici, il s'agissait d'évaluer les risques de contamination de l'environnement en cas de stockage en profondeur de déchets radioactifs. Le modèle doit donc permettre de déterminer la probabilité que le site de stockage soit en contact avec une faille majeure et de quantifier la connectivité du réseau de fracture. Cette connectivité a été étudiée par *Andresen et al.* en complément de l'analyse de site [*Andresen et al.*, 2008]. Ils ont utilisé les cartes de traces de fractures corrigées des effets de déconnexion provenant des affleurements de Laxemar. Le texte de cette étude est fourni en annexe. À partir d'une représentation topographique du réseau de fracture basé sur la connectivité entre fractures, ils ont montré que les réseaux de Laxemar étaient mieux connectés que des réseaux de fractures aléatoires présentant les mêmes propriétés stochastiques. Les corrélations entre fracture introduisent une connectivité à petite échelle, c'est-à-dire que fractures proches les unes des autres sont bien connectées, et à grande échelle, c'est-à-dire que les clusters locaux sont bien connectés les uns aux autres. Ces résultats sont cohérents avec le modèle de fracturation universelle de [*Davy et al.*, 2009]. Ce modèle prend pour hypothèse que les fractures voient leur croissance stoppée lorsqu'elles rencontrent une

fracture plus grande. Les simulations numériques qui ont servi à valider ce modèle et l'étude de la connectivité des réseaux de fractures produits ont été effectuées dans le cadre de ce doctorat. Le modèle prédit un changement de pente dans la distribution des longueurs de fracture aux grandes échelles, à partir du moment où les fractures passent d'un régime dilué, c'est-à-dire avec de faibles interactions fracture-fracture, à un régime denses où les interactions fracture-fracture deviennent très importantes. Concernant la connectivité, les réseaux ainsi produits restent proches du seuil de percolation même en augmentant la densité de fracture à l'échelle du modèle. Une description détaillée du modèle ainsi que son application à plusieurs cas naturels sont présentées en annexe.

La conclusion des études de connectivité indique que le modèle de site doit non seulement respecter les paramètres de la fracturation mesurés sur le milieu naturel, mais également ses caractéristiques annexes comme la connectivité. Pour cela il est possible de générer un grand nombre de réseaux respectant les paramètres de la fracturation et de ne sélectionner que les modèles ayant les propriétés recherchées. Une autre solution est de combiner l'analyse directe de la fracturation à un problème inverse utilisant des données induites par les propriétés physiques comme les propriétés hydrauliques.

Partie 4 Analyse statistique des écoulements : définitions d'indicateurs de chenalisation

Modéliser un milieu géologique répond à une demande en termes de connaissance du milieu et de prédictions. Le modèle doit donc être compatible avec les propriétés étudiées afin de remplir ses objectifs concernant la prédiction de scénarios géologiques ou hydrauliques. Dans les milieux fortement hétérogènes les écoulements peuvent être concentrés dans un petit nombre de structures principales [Watanabe *et al.*, 1997]. Ces structures créent des chemins d'écoulement préférentiels qui sont les éléments de premier ordre contrôlant les propriétés hydrauliques [Moreno and Neretnieks, 1993]. Si le flux est fortement chenalisé, il est donc nécessaire que le modèle du milieu intègre ces structures ou du moins soit capable de simuler leurs effets. C'est pourquoi quantifier le degré de chenalisation est une étape indispensable lors de la modélisation d'un milieu fortement hétérogène. Cette partie traite de la définition de deux indicateurs de chenalisation et de leur application à un grand nombre de configurations tests constituées de milieux poreux hétérogènes et de milieux fracturés. Les résultats sont donnés sous la forme d'un article scientifique publié dans la revue *Advances in Water Resources* [Le Goc *et al.*, 2010b]. Un premier paragraphe introduit cet article et un second paragraphe ajoute quelques résultats supplémentaires, notamment sur l'utilisation d'indicateurs plus intuitifs de la chenalisation. Un troisième paragraphe conclut cette partie et discute de ces résultats en termes de modélisation.

A. Indicateurs de chenalisation basés sur les caractéristiques statistiques du flux

Ce paragraphe commence par le résumé de l'article scientifique décrivant la formulation et l'application de deux nouveaux indicateurs de chenalisation.

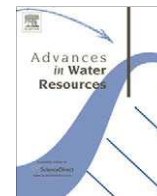
Résumé

Afin de quantifier l'importance de la chenalisation des écoulements dans un milieu fortement hétérogène, on introduit deux nouveaux indicateurs, D_{ic} et D_{cc} , basés sur la distribution lagrangienne des flux. Ces indicateurs reposent sur le ratio de participation, S_2 , défini par [Davy *et al.*, 1995]. Ils caractérisent les grandeurs extrêmes de la distribution des largeurs des tubes de courant et de la distribution des variations du flux le long des lignes de courant. S_2 est égal au rapport normalisé entre le carré du moment d'ordre 1 et le moment d'ordre 2 et est sensible aux grandes valeurs d'une distribution. Comparés aux indicateurs existants, D_{ic} et D_{cc} apportent une information supplémentaire sur la géométrie des chenaux d'écoulement. Ils sont de plus applicables aux milieux poreux et aux milieux fracturés et sont généralement moins variables pour les milieux générés avec les mêmes paramètres. A partir de valeurs calculées sur un large panel de champs de perméabilité poreux et de réseaux de fractures, on montre qu'ils décrivent de manière pertinente deux propriétés géométriques des écoulements. D_{ic} fournit une échelle caractéristique des zones de faible flux pour les milieux poreux et une distance caractéristique entre les structures effectivement porteuses de flux dans les milieux fracturés. De son côté, D_{cc} fournit une échelle caractéristique de la continuité des zones de fort flux dans les milieux poreux et une longueur caractéristique des chenaux dans les milieux fracturés. D_{ic} est principalement déterminé par la densité de chenaux et la variabilité de la perméabilité alors que D_{cc} est plus influencé par la nature de la structure de corrélation comme la connectivité des zones de forte perméabilité dans les milieux poreux et la distribution des longueurs pour les réseaux de fracture. Ces deux indicateurs sont donc complémentaires et permettent de caractériser deux propriétés différentes de la chenalisation.



Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Statistical characteristics of flow as indicators of channeling in heterogeneous porous and fractured media

R. Le Goc^{a,b,*}, J.-R. de Dreuzy^a, P. Davy^a

^a Géosciences Rennes, UMR 6118 CNRS, Université de Rennes 1, CS 74205, F-35042 Rennes Cedex, France

^b Itasca Consultants SAS, 64 chemin des mouilles, F-69134 Écully Cedex, France

ARTICLE INFO

Article history:

Received 1 July 2009

Received in revised form 4 December 2009

Accepted 6 December 2009

Available online xxxx

Keywords:

Flow channels

Heterogeneous media

Connectivity

Fracture network

Channeling indicators

ABSTRACT

We introduce two new channeling indicators D_{ic} and D_{cc} based on the Lagrangian distribution of flow rates. On the basis of the participation ratio, these indicators characterize the extremes of both the flow-tube width distribution and the flow rate variation along flow lines. The participation ratio is an indicator biased toward the larger values of a distribution and is equal to the normalized ratio of the square of the first-order moment to the second-order moment. Compared with other existing indicators, they advantageously provide additional information on the flow channel geometry, are consistently applicable to both porous and fractured media, and are generally less variable for media generated using the same parameters than other indicators. Based on their computation for a broad range of porous and fracture permeability fields, we show that they consistently characterize two different geometric properties of channels. D_{ic} gives a characteristic scale of low-flow zones in porous media and a characteristic distance between effectively flowing structures in fractured cases. D_{cc} gives a characteristic scale of the extension of high-flow zones in porous media and a characteristic channel length in fractured media. D_{ic} is mostly determined by channel density and permeability variability. D_{cc} is, however, more affected by the nature of the correlation structure like the presence of permeability channels or fractures in porous media and the length distribution in fracture networks.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Spatial heterogeneity in hydraulic conductivity affects fluid flow and solute transport in complex natural media like fractured media [38], alluvial systems [11] and strongly heterogeneous porous media [28] and has been a subject of research for decades ([8] and references therein). It is a function of contrasts between high permeability and low permeability values. As flow tends to avoid low- k zones for high- k zones, heterogeneity induces the development of preferential flow paths [20,23] also called “paths of least resistance” [39], along which flow is focused. Their effects on up-scaled/effective hydrologic properties have been observed in laboratory and numerical studies. Fogg [10] performed a numerical study on the hydraulic conductivity distribution in the Wilcox aquifer and suggests that flow is mainly controlled by the continuity and connectivity of sand deposits rather than by local hydraulic conductivity values. Hanor [16] drew similar conclusions for the Livingston site. Silliman [34] illustrated the formation of preferential flow paths with laboratory experiments. Knudby et al. [22] and,

Ronayne and Gorelick [30] showed how the estimate of aquifer properties, like the effective permeability of a system, should take channeling into account. Ronayne et al. [31] used statistical channeling properties to estimate aquifer parameters in a system affected by channeling. Similarly, Kerrou et al. [19] showed that not accounting explicitly for channeling in a sequential self-calibration approach resulted in flow underestimation and strong deviations in capture zone estimates. Trincherio et al. [37] showed that for moderate heterogeneities, both the connectivity of high- k values and apparent porosity are key in predicting transport times efficiently. Although channeling is important for flow and transport properties, its quantification remains a matter of debate. Two types of indicators have been proposed: indicators derived from the comparison of up-scaled hydraulic properties with their small-scale counterparts, and statistical indicators calculated from the permeability and flow fields. The first category of estimators is based on hydraulic properties that are sensitive to channeling. The simplest estimator is the effective permeability, K_{eff} , known to be sensitive to flow organization [14]. In 2D multi-log-Gaussian isotropic weakly-correlated fields, the equivalent permeability is equal to the geometric mean $K_{eff} = K_g$ [27]. If the connectivity of the higher- K zones is greater than that of the lower- K zones, K_{eff} is larger than K_g [32] within the limit that $K_{eff} \leq K_a$ (where K_a is the arithmetic mean) [40]. The type of average measured by the

* Corresponding author. Address: Géosciences Rennes, UMR 6118 CNRS, Université de Rennes 1, CS 74205, F-35042 Rennes Cedex, France. Tel.: +33 223236624; fax: +33 223236090.

E-mail address: r.legoc@itasca.fr (R. Le Goc).

power averaging exponent CF_1 [9,18] has thus been considered as a measure of channeling [20]:

$$CF_1 : K_{\text{eff}} = \left(\frac{1}{V} \int_V K(x)^{CF_1} dV \right)^{\frac{1}{CF_1}} \quad (1)$$

CF_1 varies between -1 and 1 for the harmonic and arithmetic means, respectively, and is equal to zero for the geometric mean corresponding to isotropic weakly-correlated multi-Gaussian fields. As transport is also strongly affected by channeling, breakthrough curve properties have been proposed as estimators of the channeling degree [41]. Knudby and Carrera [20] used the ratio CT_1 of the average arrival time \bar{t} to the time at which 5% of the solute have broken through the domain boundary t_5 :

$$CT_1 = \bar{t}/t_5 \quad (2)$$

When preferential flow paths exist, t_5 becomes much smaller than \bar{t} , CT_1 increases and the field should be considered as increasingly connected. The apparent hydraulic diffusivity has been proposed as an intermediary characteristics between flow and transport connectivities [21]. Park et al. [29] suggested that the normalized travel time and distance be used to investigate preferential flow.

The second category of estimators uses statistical characteristics of the permeability field or of the flow field. N -point spatial connectivity statistics are dedicated to the measurement of connectivity and were applied to permeability fields to estimate the presence of high- k connected patterns [17,24]. Western et al. [42] used a directional multi-point geostatistical indicator and showed that it could capture the difference between random and channeled fields with similar k -distributions, unlike non-directional indicators. Fripiat et al. [13] suggested that the presence of preferential flow paths or flow barriers could be identified using head and flow variances, since head variance is negatively correlated to connectivity while flow variance is positively correlated to the effective permeability increase. Bruderer-Weng et al. [3] used the multifractal spectrum of the flow field to quantify channeling in heterogeneous pipe networks. The distribution of flow has also been used for quantifying channeling in fractured networks [6].

The multiplicity of the proposed indicators shows that channeling cannot be restricted to a single simple characteristic. The concept of channeling also strongly depends on the application targeted. The relevant use of channeling indicators probably differ between flow and transport applications [33]. In this study, we focused first on the geometrical characterization of channels, i.e. on the channels themselves rather than on their consequences in terms of flow or transport. In this respect, the first category of indicators based on equivalent medium properties are limited by the fact that they measure the consequences of channeling rather than channeling itself. The limitation of the indicators based on permeability statistics arises from the measurement of a single cause of channeling cause (the connectivity of high- k zones) where channeling is also induced by the variability of permeability [26]. The advantage of those indicators based on the statistical properties of the flow field is the measurement of channeling itself. As opposed to the multifractal dimensions and the variance of head or flow, we look for indicators based on the geometrical properties of the channels that additionally identify channeling consistently in both porous and fractured media.

Even though channeling occurs under many different circumstances, it has two recurrent characteristics. First, flow is localized within a few structures. Second, channeling locally maintains high flow rates over long distances. On the basis of these two characteristics, we aimed at defining quantitative channeling indicators that meet the three following constraints. First, they must be globally consistent with the visually intuitive classification of channeling.

Second, they must provide a quantification of channeling. Third, they must be applicable simultaneously to porous and fractured media.

We define two new indicators in Section 2. We compute their value for the broad range of synthetic fields introduced in Section 3. In Section 4, we analyze first their consistency with the expected ranking of channeling and then their dependency on the permeability correlation structures. Finally, we compare them to other existing indicators in Section 5.

2. Flow-based indicators

A channeled medium is defined as a medium where flow is localized within a few structures and where preferential flow locally maintains high flow rates over long distances. To this end, we defined two channeling indicators, one quantifying the localization of flow within the system and the other quantifying the continuity of flow paths. Since the proposed indicators were not straightforward, we introduce them using preliminary attempts based on simpler quantities. The objective was to show the relevance of the more complex indicators finally adopted. The first indicator should characterize the relative volume occupied by the high-flow zones. The simplest indicator could be the relative volume occupied by flows larger than a given threshold value. Although simple, this indicator depends on the arbitrary choice of the threshold value. Rather than a deterministic indicator, we looked for a statistical characteristic biased toward the higher values of the flow distribution. Since the moments of the flow distribution $M_k(\Phi)$ are increasingly sensitive to the highest values with increasing orders k , the idea was to compare moments of increasing orders like in the participation ratio S_2 [5,35] equal to:

$$S_2(\Phi) = M_1(\Phi)^2 / (M_0(\Phi) \cdot M_2(\Phi)) \quad (3)$$

where Φ stands for the spatial distribution of flow rates. For Φ discretized on a domain of n cells of volumes V_i , $M_k(\Phi)$ writes:

$$M_k(\phi) = \sum_{i=1}^n \phi_i^k \times V_i \quad (4)$$

where ϕ_i is the mean value of Φ over the grid cell i . Table 1 shows S_2 values for usual distributions. When the distribution variability vanishes, S_2 tends to 1. By contrast, S_2 systematically decreases with higher variability whatever the distribution type. For the lognormal distribution, S_2 is solely function of the lognormal variance.

Whereas $S_2(\Phi)$ gives indications about the surface occupied by the largest flow rates, it does not account for the distribution of this surface within the domain. Consequently, we did not use $S_2(\Phi)$ but $S_2(W_n)$, where W_n is the distribution of flow-tube widths carrying all the same fraction $1/n$ of the total flow. Since $S_2(W_n)$ is biased toward the larger W_n values, it characterizes the extension of the low-flow zones and hence the distance between main flow channels. We defined and computed W_n in the specific context of permeameter-like boundary conditions defined for a square domain by fixed heads on two opposite sides and no flow on the other sides. The definition of W_n may also be adapted for different boundary conditions. In convergent flow conditions, W_n would

Table 1
Participation ratio S_2 for classical distributions.

Distribution type	Parameters	S_2
Gaussian	Mean μ and variance σ_G	$S_2 = (1 + (\sigma_G/\mu)^2)^{-1}$
Log-normal	Log-normal variance σ_{LN}^2	$S_2 = \exp(-\sigma_{\text{LN}}^2)$
Binary distribution	v_1 with a probability p and 0 with a probability $1 - p$	$S_2 = p$

be defined by the distance between flow lines normalized by the distance to the well. With permeameter-like boundary conditions, we first determined n equivalent flow tubes defined as the tubes carrying all the same fraction $1/n$ of the total flow (Fig. 1, middle column). Then, we computed the participation ratio $S_2(W_n)$ on the flow-tube width distribution.

We introduce the meaning of $S_2(W_n)$ with the case of p regularly spaced flow tubes of width L/p , where L is the system size, and $n - p$ flow tubes of negligible width within the channels. The distribution of flow-tube widths is thus a binary distribution of values L/p with a probability p/n , and 0 with a probability $(1 - p/n)$. From Table 1, it leads to $S_2(W_n) = p/n$. In this case, $S_2(W_n)$ is directly proportional to the number of channels p . When the number of channels p is equal to n , $S_2(W_n)$ reaches a value of 1, like in homogeneous flow fields. In fact, for a homogeneous case, all flow tubes have the same width and $S_2(W_n) = 1$. Using this same example, we derived the characteristic distance between channels D_{ic} from $S_2(W_n)$. Since the distance between two channels is equal to $D_{ic} = L/p$ and $S_2(W_n) = p/n$, then:

$$\frac{D_{ic}}{L} = \frac{1}{n \cdot S_2(W_n)} \quad (5)$$

D_{ic} ranges from L/n in homogeneous fields to L in a unique channel conveying all the flow, for which $S_2(W_n) = 1/n$. L/n can be a priori interpreted as a channel resolution. The selection of an appropriate value for n will be investigated at the beginning of Section 4. D_{ic} is a characteristic distance between channels. However, it does not provide any information on the channel persistence throughout the system. For example, in Fig. 1 the field at the top and the field in the middle have two different flow fields with about the same D_{ic}/L ratio equal to 0.09, but high flow rates are visually maintained over a longer distance in the middle field than in the top field.

We looked for a second indicator designed to differentiate these two fields by characterizing the distance over which flow rates are continuously high. We first tried the Lagrangian correlation length of flow rates. It was however not consistent with our intuition of channel persistency. For example, the correlation length of the

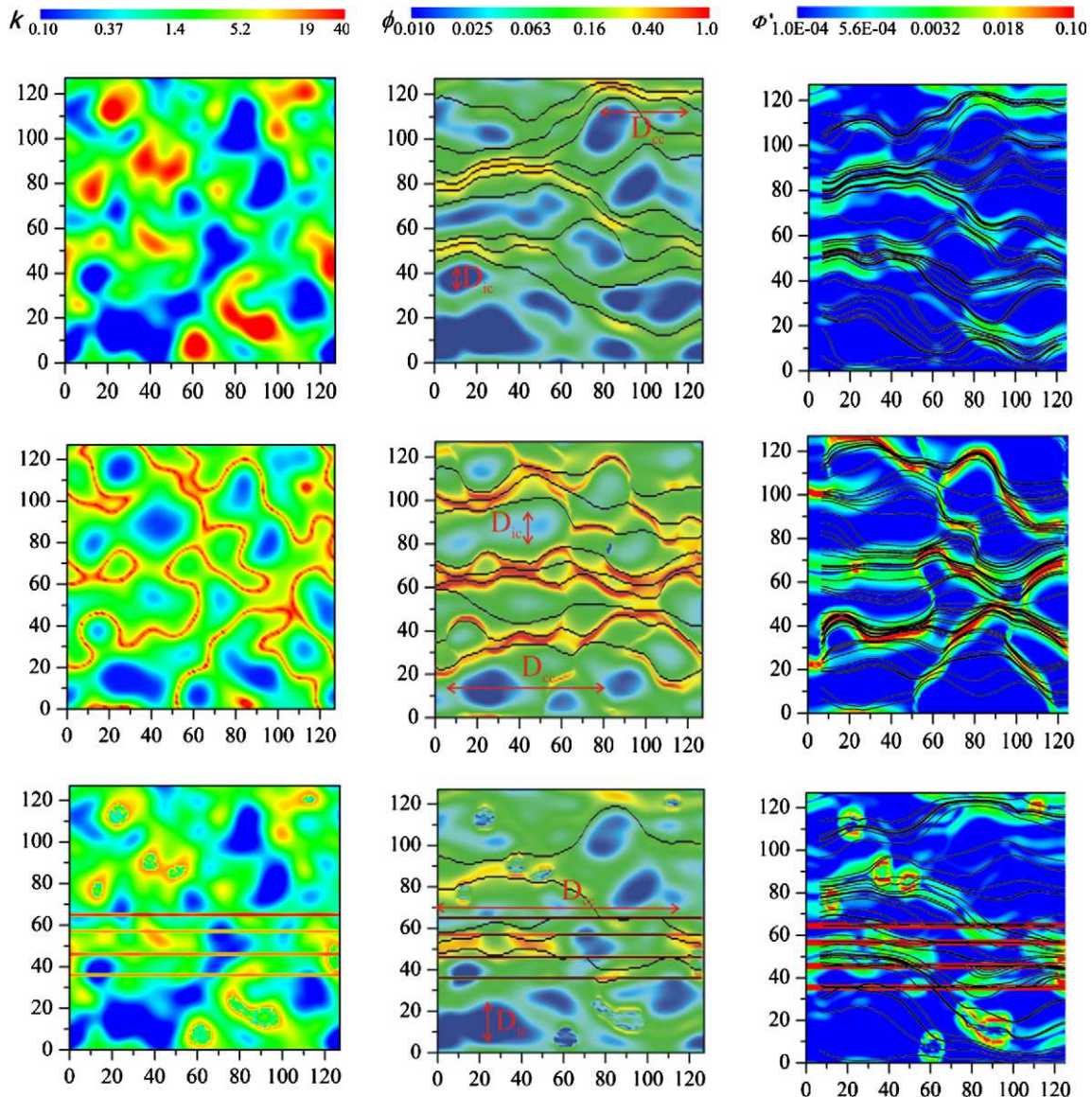


Fig. 1. Permeability fields K (left), flow rates ϕ (middle) and Lagrangian derivatives of flow rates ϕ' (right) in three permeability fields, all based on the same multi-Gaussian distribution with a Gaussian correlation. The correlation length is fixed to $1/16$ th of the system size. Correlations are either not modified (top) or increased either by connecting the larger permeability values (middle) [43] or by introducing fracture-like structures (bottom) [20]. Ten flow tubes as well as the D_{ic} and D_{cc} values defined in Section 2 are superimposed on the fields of the middle column. Fifty flow lines are superimposed on the distribution values of Lagrangian flow derivatives of the right column.

rearranged field in Fig. 1 (middle) is smaller than the correlation length of the non-rearranged field (Fig. 1, top). The correlation length is not only sensitive to the large flow rates but also to all other values. It thus fails to characterize high-flow zone connectivity. Like for the previous indicator, computing the Lagrangian correlation length from the sole velocities larger than a given threshold faces the same problem of the arbitrary choice of the threshold. Moreover, the channels may display some discontinuities that hinder the relevance of a threshold (Fig. 1, right bottom). We found that characterizing flow channel discontinuities is easier than flow channel persistence since discontinuities are more localized. Large values of the spatial derivatives of flow rates are more localized at the entrance and exit of channels than in the remaining of the field. On the contrary, the variations of flow rates are smaller and distributed evenly in non-channeled media. To characterize the distribution of the flow transitions taken as the Lagrangian derivatives of flow rates Φ' , we used again the participation ratio S_2 on Φ' (right column in Fig. 1). Numerically, $S_2(\Phi')$ was calculated according to (3) from the moments of the distribution of Φ' discretized along the flow lines:

$$M_k(\Phi') = \sum_{j=1}^p \sum_{i=1}^m \Delta s_i^j \left| \frac{\Delta \phi_i^j}{\Delta s_i^j} \right|^k \quad (6)$$

with j the flow line index, p the number of flow lines, s_i^j the i th position along the flow line j , m the number of positions along the flow line, $\Delta \phi_i^j$ and Δs_i^j the flow rate variation and distance between two consecutive points. Flow lines were computed by using a particle-tracking algorithm. We chose p equal to 10^4 and m so that Δs_i^j was of the order of the grid cell size after ensuring that larger p and m values did not modify the results. $S_2(\Phi')$ defines a characteristic distance l_{eff} over which flow rates are actually variable divided by the average flow line length (L') [35]. By contrast, a characteristic scale D_{cc} over which flow rates are only slowly varying is function of $L' - l_{\text{eff}}$:

$$D_{\text{cc}}/L' = (L' - l_{\text{eff}})/L' = (1 - S_2(\Phi')) \quad (7)$$

D_{cc} will be taken as a characteristic channel length. In media with discontinuous flow paths, Φ' has a narrow and spatially uniform distribution, leading to $S_2(\Phi') = 1$ and $D_{\text{cc}}/L' = 0$. In highly-channeled media, the distribution Φ' contains values close to zero except at the channel extremities, leading to small $S_2(\Phi')$ values and D_{cc}/L' close to 1. For example, the D_{cc}/L' value for the middle field in Fig. 1 is equal to 0.72. It is larger than the value of 0.43 for the field at the top of Fig. 1 following the intuition that persistence is larger in the middle field.

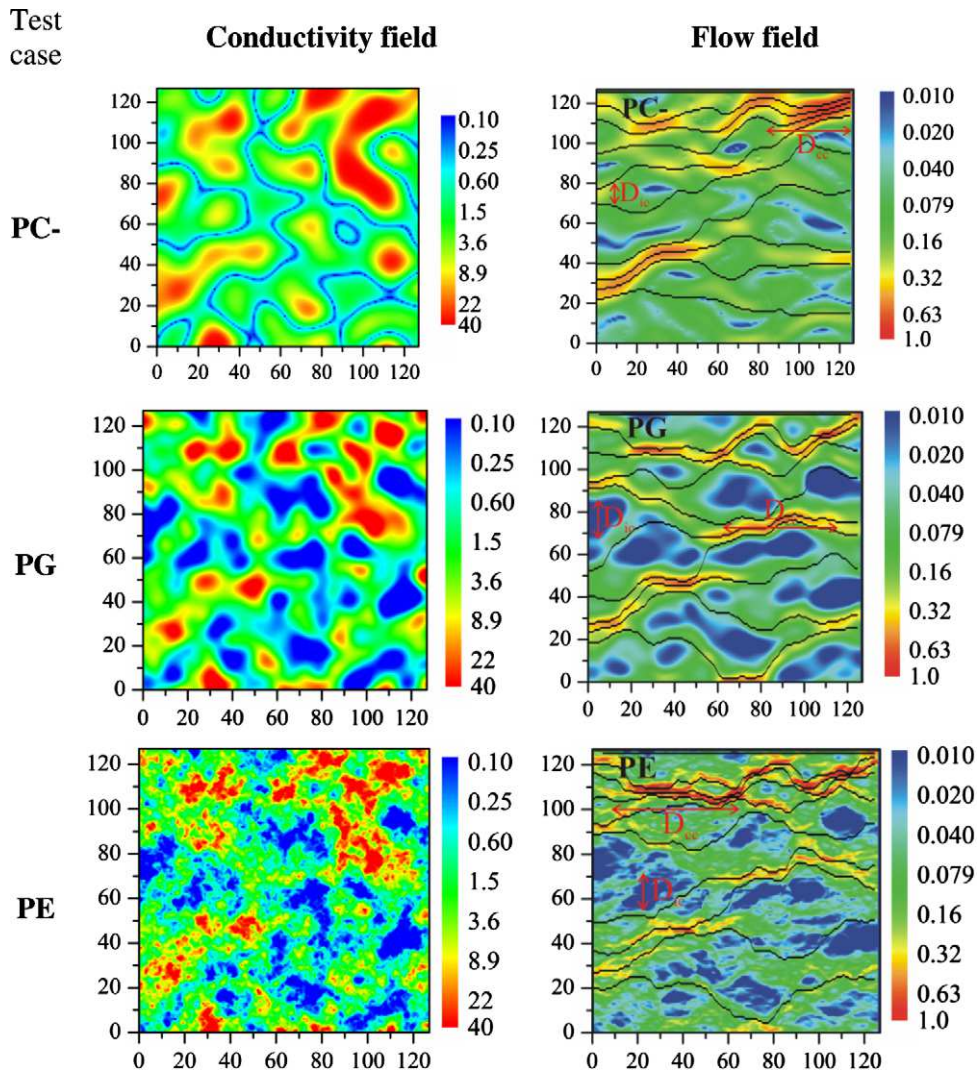


Fig. 2. Permeability fields (left) associated with their corresponding flow fields (right). Permeabilities are normalized by their geometrical mean value. Flows are normalized by their maximal value. Test case names in the first column refer to Table 2. D_{ic} and D_{cc} values defined in Section 2 are superimposed on the flow fields of the right column.

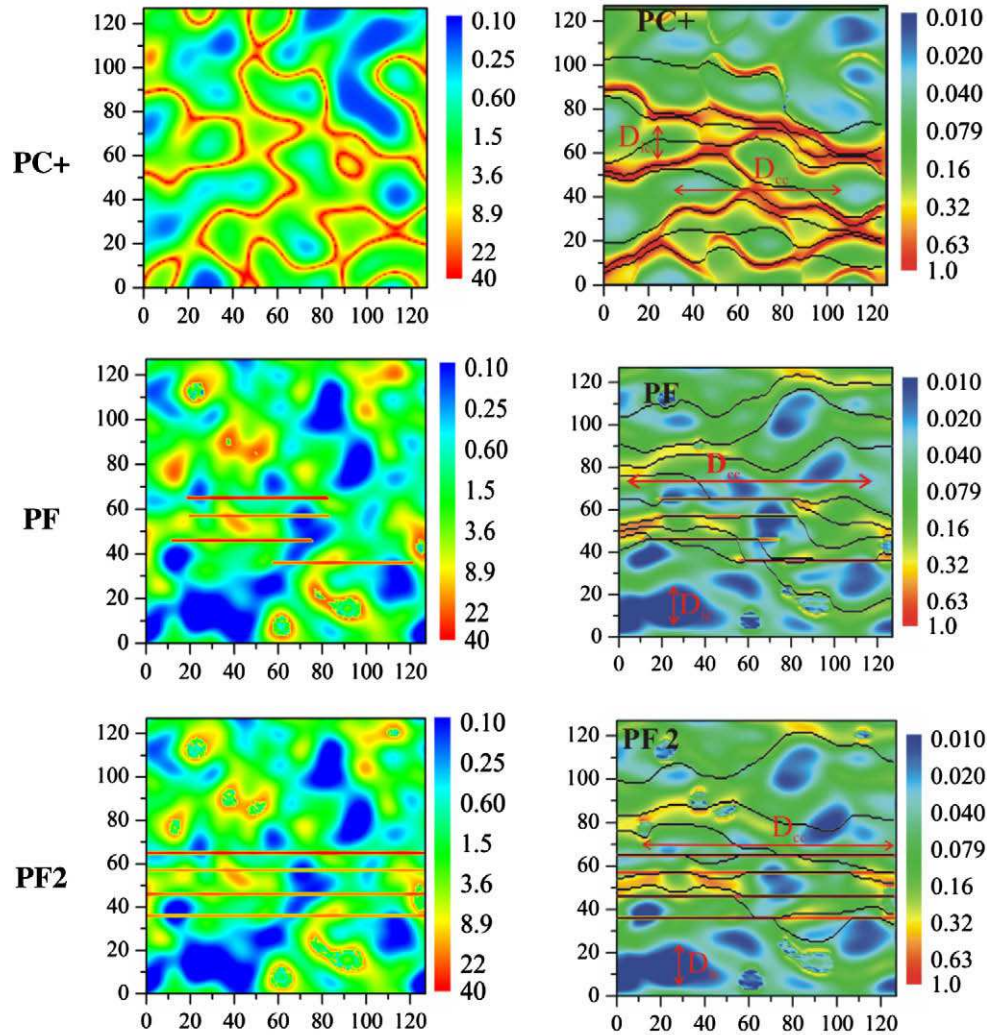


Fig. 2 (continued)

D_{ic}/L and D_{cc}/L' are statistically-derived indicators designed to characterize flow localization and flow continuity. They are dimensionless quantities ranging between 0 and 1 that can be used to compare channeling in different systems. In Section 3, we define a broad range of synthetic porous fields and fracture networks for which D_{ic} and D_{cc} will be computed in Section 4.

3. Tested media and computational methods

Indicators D_{ic} and D_{cc} will be compared in the different synthetic fields displaying various connectivity degrees presented in this section.

3.1. Field generation and flow computational method

Simulations were performed in four steps, consisting in the generation of the tested fields, simulation of flow, derivation of the flow lines and computation of the different indicators. The generation of the multi-Gaussian porous fields was performed via a Fourier transform [15] using the software FFTW [7,12]. Some fields are then rearranged according to the rearrangement methods described in [43,20] if required. In order to avoid side effects, particularly with large correlation lengths, all fields were generated within a 1280×1280 grid of which the central part (128×128)

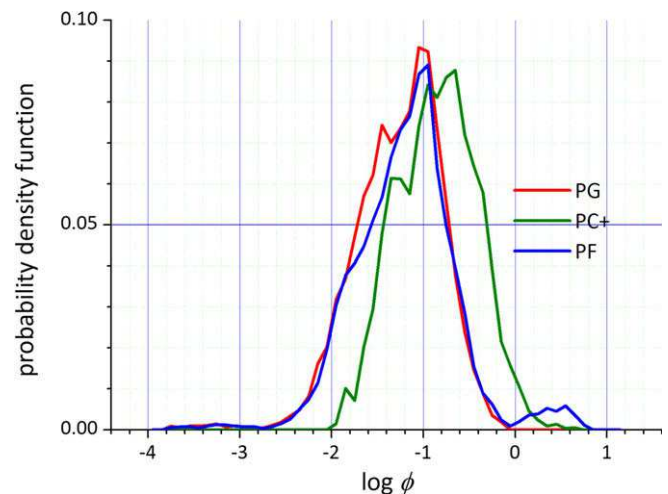


Fig. 3. Probability density function of the logarithm of flow rates in the classical multi-Gaussian fields with a Gaussian correlation (PG) and in the fields with the two rearrangement methods PC+ and PF. The C+ rearrangement method globally shifts the flow rate distribution to higher flow values while the F rearrangement introduces a second peak of higher flow values.

was kept for the analysis, so $L = 128$ was the characteristic system size. The original field was taken with a $\log-k$ mean equal to zero, a

variance (σ_y^2) equal to 1 and 3 and a correlation length (λ) equal to 8 and 64. Fracture networks were generated within a system size equal to $10 * l_{\min}$, where l_{\min} is the size of the smallest fracture.

In porous media, the flow equation was discretized on the structure of the medium according to a finite volume framework with harmonic inter-cell permeabilities [25]. As previously mentioned, permeameter-like boundary conditions were imposed on the sides of the domain, i.e. fixed heads on two opposite borders and no flow on the others. The discretized flow equations ended up to a linear system $\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$ solved by the multifrontal method implemented in the software UMFPACK [4]. Flow lines were constructed using a particle-tracking algorithm. Particles were injected through a vertical segment positioned in a central part at a distance of one correlation length from the inlet and proportionally to flow in order to

avoid boundary effects [1,36]. Indicators were computed from 500 Monte-Carlo realizations for each tested case.

3.2. Description of the test cases

We used a broad range of 2D synthetic porous and fractured media characterized by the histogram of their permeability distribution and their connectivity structures. The synthetic porous media have all a lognormal permeability distribution of variance σ_y^2 , where y stands for $\log(k)$ and k is the permeability, but differ by their correlation structure (see Fig. 2). For the same correlation length λ , we used six correlation patterns. The tested structures are identified by P as in porous and one or two other letters specifying the correlation pattern. The first two fields have Gaussian and

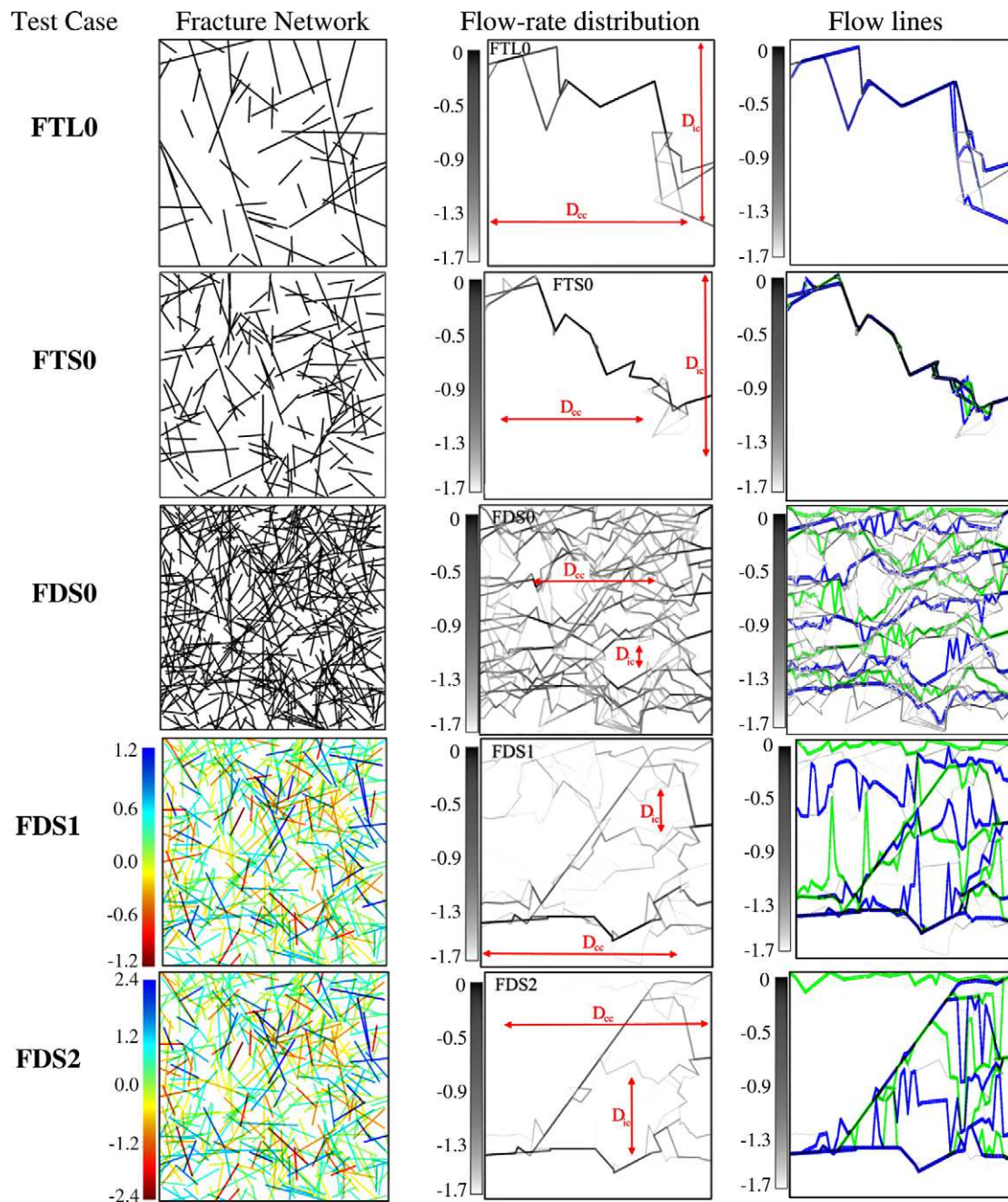


Fig. 4. (Left) Fracture test cases, colored according to their log-transmissivity (middle) corresponding flow fields with D_{ic} and D_{cc} and (right) flow tubes. Flows are normalized by their maximal value and colored with a logarithmic scale. Test case names in the first column refer to Table 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2

Porous and fractured test cases ranked visually by increasing order of channeling.

Field type	Properties	Short name
Porous	Rearranged with the C-method, $\sigma_y^2 = 1$ or 3 , $\lambda = L/16$ or $L/2$	PC–
Porous	Gaussian correlation, $\sigma_y^2 = 1$ or 3 , $\lambda = L/16$ or $L/2$	PG
Porous	Exponential correlation, $\sigma_y^2 = 1$ or 3 , $\lambda = L/16$ or $L/2$	PE
Porous	Rearranged with the C+ method, $\sigma_y^2 = 1$ or 3 , $\lambda = L/16$ or $L/2$	PC+
Porous	Rearranged with the F method, $\sigma_y^2 = 1$ or 3 , $\lambda = L/16$ or $L/2$	PF
Porous	Rearranged with the F2 method, $\sigma_y^2 = 1$ or 3 , $\lambda = L/16$ or $L/2$	PF2
Fractured	Dense ($d = 3p_c$), dominated by short fractures ($a = 3.5$), uniform fracture transmissivity ($\sigma_{y \log T}^2 = 0$)	FDS0
Fractured	Dense ($d = 3p_c$), dominated by short fractures ($a = 3.5$), distributed fracture transmissivity ($\sigma_{y \log T}^2 = 1$)	FDS1
Fractured	Dense ($d = 3p_c$), dominated by short fractures ($a = 3.5$), distributed fracture transmissivity ($\sigma_{y \log T}^2 = 2$)	FDS2
Fractured	Sparse ($d = p_c$), dominated by short fractures ($a = 3.5$), constant fracture transmissivity ($\sigma_{y \log T}^2 = 0$)	FTS0
Fractured	Sparse ($d = p_c$), dominated by long fractures ($a = 2.0$), constant fracture transmissivity ($\sigma_{y \log T}^2 = 0$)	FTL0

Table 3

Mean and variance on 500 realizations for the different indicators and test cases. N/A stands for indicators that cannot be computed in the corresponding cases.

CASE	σ_y^2	λ	D_{ic}/L	$\sigma^2(D_{ic}/L)$	D_{cc}/L'	$\sigma^2(D_{cc}/L')$	CF_1	$\sigma^2(CF_1)$	CT_1	$\sigma^2(CT_1)$
PC–	1	8	0.063	0.0040	0.4	0.0820	–0.20	0.17	1.4	0.22
			0.073	0.011	0.38	0.066	–0.21	0.19	1.6	0.38
	3	64	0.061	0.0080	0.40	0.18	–0.10	0.50	1.4	0.35
			0.070	0.022	0.44	0.20	–0.10	0.51	1.4	0.34
PG	1	8	0.065	0.0027	0.41	0.032	–0.0042	0.076	1.6	0.25
			0.092	0.0082	0.44	0.037	0.0014	0.071	2.3	0.64
	3	64	0.064	0.014	0.29	0.088	–0.00023	0.43	1.3	0.28
			0.096	0.049	0.31	0.13	0.020	0.41	1.6	0.55
PE	1	8	0.063	0.0031	0.40	0.016	–0.035	0.075	1.5	0.25
			0.088	0.010	0.45	0.022	–0.026	0.073	2.2	0.60
	3	64	0.062	0.0087	0.37	0.030	–0.015	0.27	1.4	0.26
			0.089	0.032	0.39	0.065	0.00018	0.26	1.7	0.55
PC+	1	8	0.071	0.0064	0.64	0.066	0.17	0.15	1.7	0.42
			0.094	0.020	0.60	0.075	0.23	0.16	2.1	0.67
	3	64	0.064	0.016	0.57	0.21	0.039	0.53	1.5	0.50
			0.077	0.035	0.51	0.22	0.065	0.53	1.7	0.73
PF	1	8	0.066	0.0030	0.74	0.082	0.23	0.12	1.9	0.54
			0.10	0.011	0.86	0.067	0.29	0.09	3.4	1.8
	3	64	0.057	0.0053	0.70	0.12	0.31	0.31	1.4	0.25
			0.071	0.017	0.70	0.15	0.31	0.31	1.8	0.68
PF2	1	8	0.067	0.0037	0.78	0.076	0.42	0.12	3.5	2.4
			0.12	0.0030	0.92	0.052	0.64	0.078	15	9.6
	3	64	0.056	0.060	0.71	0.11	0.33	0.32	1.5	0.55
			0.069	0.019	0.72	0.14	0.35	0.33	2.6	2.6
FDS0	N/A		0.11	0.10	0.59	0.036	N/A		1.7	0.30
FDS1			0.24	0.15	0.86	0.050			7.2	8.4
FDS2			0.29	0.14	0.91	0.048			26	26
FTS0			0.64	0.18	0.65	0.089			2.2	5.3
FTL0			0.82	0.21	0.90	0.045			1.2	6.9

exponential correlation structures (PG and PE in Fig. 2). The two next ones result from the rearrangement methods by Zinn and Harvey [43], yielding fields of highly-connected high or low permeabilities (PC+ and PC– on Fig. 2). The two last ones are Gaussian-correlated fields to which are added highly permeable fracture-like structures oriented parallel to the average head gradient and spanning either half of the system (PF on Fig. 2) or the whole system (PF2 on Fig. 2) [20]. These rearrangement methods provide different flow distributions (Fig. 3). The C+ method increases the mean flow rate compared to the Gaussian correlated field, while the F method adds a second peak of larger flow rates to the histogram. Note, however, that the rearrangement methods do not modify the permeability histogram.

Flow channeling was also observed in fractured media because of both the fractures and the network-scale heterogeneities [38]. In this paper, we concentrated on the network-scale complexity stemming at first from the power-law fracture length distribution:

$$n(l) \sim l^{-a} \quad (8)$$

where a is a characteristic exponent between 1.5 and 4 [2]. We chose five types of fracture networks differing by their fracture length and transmissivity distributions and by their density. They are identified by the letter F followed by three additional letters. The first one, FTL0, corresponds to fracture networks at percolation threshold (structures just connected) with a power-law length distribution yielding to large fractures corresponding to a equal to 2.0 (Fig. 4). The four other networks show a smaller probability of occurrence of large fractures ($a = 3.5$) and are respectively at threshold (FTS0, Fig. 4) and dense with a density three times as large as that of threshold. The dense fracture networks differ by their fracture transmissivity distributions of lognormal standard deviation σ_y equal to 0 in FDS0, 1 in FDS1 and 2 in FDS2. Flow fields displayed in Fig. 4 (second column) show different flow structures from highly-channelled (FTL0, FTS0 and FDS2) to well-distributed (FDS0).

4. Results

After a visual inspection of the different test cases (Figs. 2 and 4), we ranked them by their apparent channeling degree (Table 2).

The order was derived separately in the porous and fracture cases according to the flow-tube widths and regularity. In porous media, this order is consistent with CF_1 values (Table 3). All results discussed in the following paragraph are given in Table 3.

4.1. Relation between D_{ic} and the number of considered flow tubes

The interchannel distance D_{ic} (5) depends on the proportion of flow used to define a channel. If n is the number of flow tubes, each flow tube carries $1/n$ part of the total flow. Fig. 5 displays the rela-

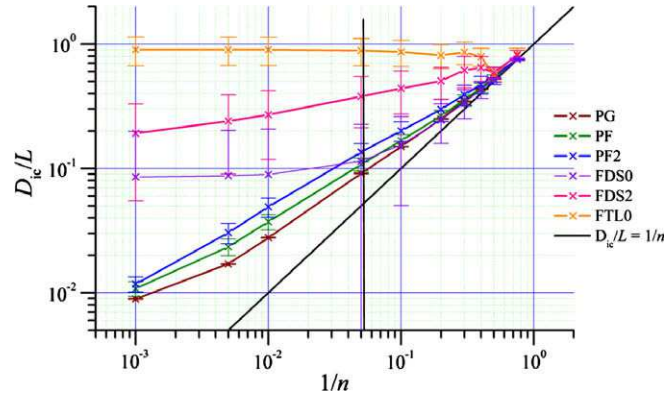


Fig. 5. D_{ic}/L versus $1/n$ in various porous and fractured test cases. The vertical black line represents the values used in the current study. $D_{ic}/L = 1/n$ is the lower limit representing a homogeneous field while $D_{ic}/L = 1$ is the upper limit representing a field with all flows concentrated in a unique channel.

tion between D_{ic} and $1/n$. When $1/n$ tends to 1, D_{ic} tends to L , meaning that no channel contains all the flow by itself. For the smallest values of $1/n$, the fracture cases reach a plateau characteristic of the smallest distance between flowing fractures. $1/n$ can be interpreted as characteristic of the smallest flow channel that can be identified. In Fig. 5, we chose a value of $n = 20$, for which all test cases have a characteristic interchannel distance larger than the interfracture distance. The value of D_{ic} remains dependant of n but the relative order for the different test cases remains the same whenever $n \leq 20$. We will thus compare D_{ic} values between test cases rather than their absolute values in single test cases. The chosen value of n is the flow-tube resolution. If $D_{ic} < L/n$, the medium will be considered as homogeneous in the sense that interchannel distances are smaller than L/n . The value of n should be increased for distinguishing closer channels.

4.2. Channeling characteristics D_{ic} and D_{cc}

Fig. 6a and b display the two new indicators D_{ic}/L and D_{cc}/L' in highly-heterogeneous porous cases ($\sigma_y^2 = 3$) and fracture cases. PG and PE configurations have very close values of D_{ic} and D_{cc} despite the visual ranking, meaning that the short-range correlations do not affect the channeling degree. Fig. 6c and d display the two other indicators CF_1 and CT_1 . CT_1 performs poorly while CF_1 captures the channeling increase of porous configurations (PC+, PF and PF2). However, CF_1 can be used in porous cases but is not available in fractured media. Moreover, CT_1 is not discriminating in porous cases and does not account for the apparent ranking of channeling in the fracture test cases. D_{ic}/L and D_{cc}/L' are thus more adapted to characterize channeling consistently in porous and fractured cases.

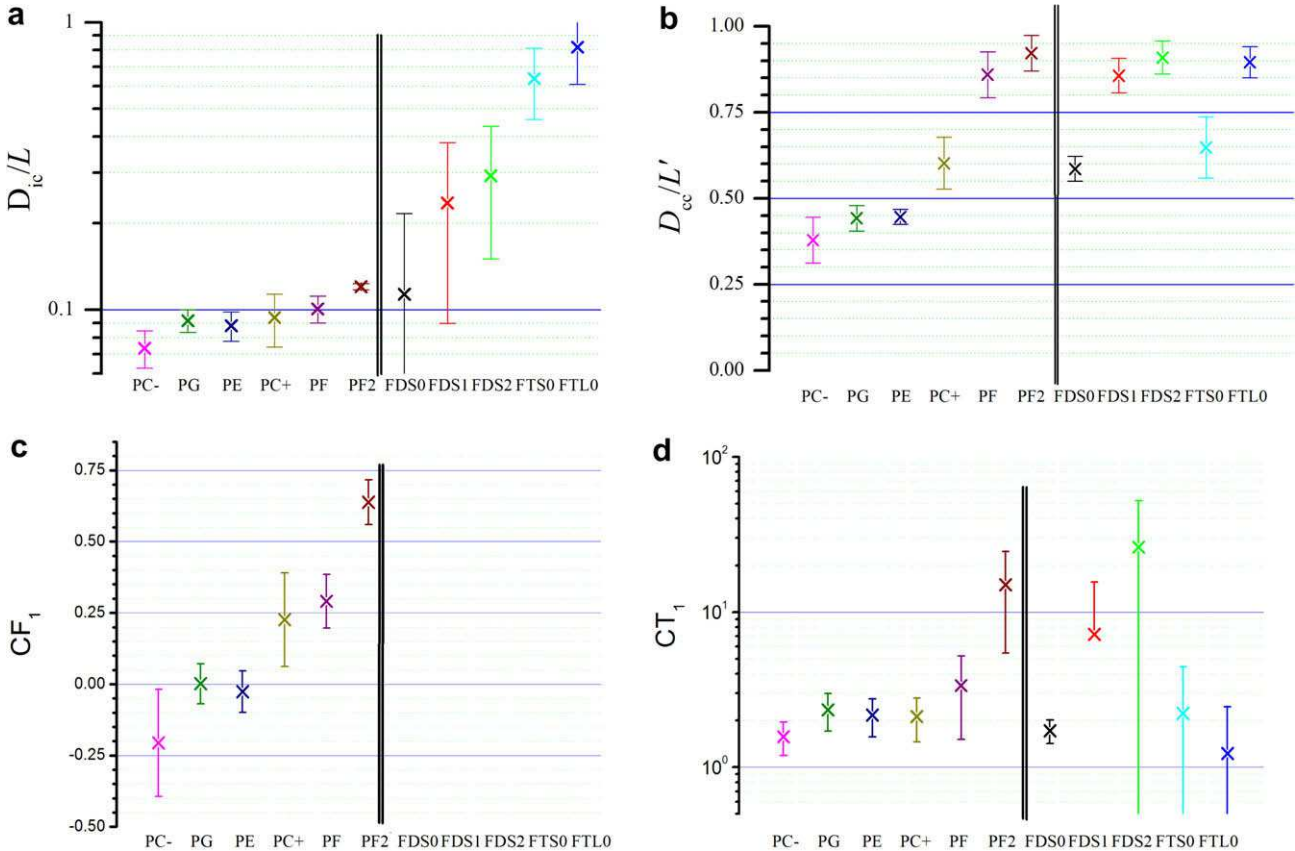


Fig. 6. Values of (a) D_{ic}/L , (b) D_{cc}/L' , (c) CF_1 and (d) CT_1 . For the different test cases ranked by their increasing intuitive rating of channeling (Table 2), Error bars are the standard deviations of the underlying distributions. Parameters of porous cases are $\sigma_y^2 = 3$ and $\lambda = 8$. The vertical double bar separates porous and fracture test cases.

The characteristic interchannel distance D_{ic}/L consistently increases with more visual channeling. This increase is much smaller in the porous cases than in the fracture cases. D_{cc}/L' also increases in porous media and has significantly larger values in all channeled fracture and porous fracture cases (FDS1, FDS2, FTL0, PF and PF2). In fact, in fractured media, flow is focused within the fractures and the variations of flow rates are more restricted than in porous media. D_{cc}/L' reaches values close to one equivalent to little variation of flow rates within flow lines in FTL0, FDS2, PF and PF2. The comparison of the variations in D_{ic}/L and D_{cc}/L' shows that in the porous cases, D_{cc}/L' increases over a range twice as large as that of D_{ic}/L from PC- to PC+. In the fracture case, however, D_{ic}/L is more consistent with the visual ranking of channeling than D_{cc}/L' . These results indicate that a flow organization indicator (D_{cc}) better characterizes porous flow channeling while a flow localization indicator (D_{ic}) better characterizes fracture flow channeling.

The variability of D_{ic}/L is much larger in fractured media than in porous media (Error bars in Fig. 6a), which means that D_{ic} does not vary much in porous configurations where channels are distributed over the field. However, D_{ic} is highly variable in fracture configurations where channels can be either very clustered or spread.

The absence of systematic correlation between D_{ic}/L and D_{cc}/L' shown in Fig. 7 confirms that D_{ic}/L and D_{cc}/L' characterize two different channeling properties. Fig. 7 also shows that D_{ic}/L and D_{cc}/L' consistently characterize channeling in respectively the fracture and the porous cases. First, the visually ranked non-channeled fracture case FDS0 is in fact close to a highly-correlated porous case (PC+, $\sigma_y = 3$, $\lambda = 8$). Second, the porous fracture cases PF and PF2 located at the top left corner of Fig. 7 have larger D_{cc}/L' values than the porous cases and smaller D_{cc}/L' values than the fracture channeled cases.

Based on Fig. 7, we distinguish three types of flow structures. First, weakly-channeled flow structures are characteristic of multi-Gaussian fields (PG, PE and PC-) and lead to small D_{ic}/L and D_{cc}/L' values. Second, the mildly-channeled flow structures were obtained for high- k connected patterns (PC+, PF and PF2 with $\sigma_y^2 = 1$) and have small D_{ic}/L values and large D_{cc}/L' values. Third, the highly-channeled media have large D_{ic}/L and D_{cc}/L' values, like FTL0 or PF2 with $\sigma_y^2 = 3$. The latter case corresponds to extreme channeling for which flow is both highly localized and highly continuous in a very small number of channels.

4.3. Relation between channeling characteristics and k-field parameters

In this section, we look for a finer understanding of indicators D_{ic}/L and D_{cc}/L' by analyzing their dependence on the structures of the porous (Figs. 8 and 9) and fractured test cases (Figs. 10–12). We then comment on the variation trends and amplitudes. D_{ic}/L systematically increases with more heterogeneity. In fact, D_{ic}/L increases with σ_y^2 in porous media (Fig. 8) and with σ_y^2 in fracture networks (Fig. 12). Larger σ_y^2 values imply that flows focus in sparser transmissivity zones. D_{ic}/L also systematically decreases in denser fracture networks, i.e. when increasing the number of connected parallel fracture paths (Fig. 11). Increasing the probability of occurrence of long fractures with smaller a values yields similar causes and effects (Fig. 10). Similarly, increasing the correlation length λ in porous media from small to median values induces more channeling. The sole non-obvious D_{ic}/L variation is its decrease from intermediary to large correlation lengths, approxi-

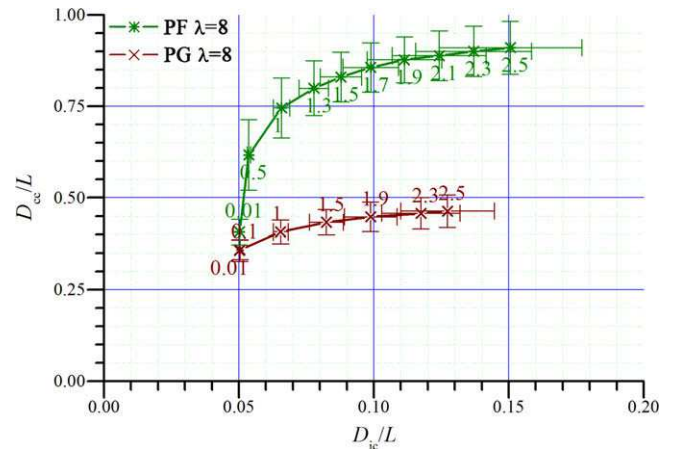


Fig. 8. Variation of D_{cc}/L' versus D_{ic}/L in porous test cases PG and PF for varying permeability standard deviations σ_y . σ_y values are given next to the corresponding symbols. When σ_y tends to zero, D_{ic} tends to $1/n$ and D_{cc} tends to zero as the permeability values tend to be homogeneous.

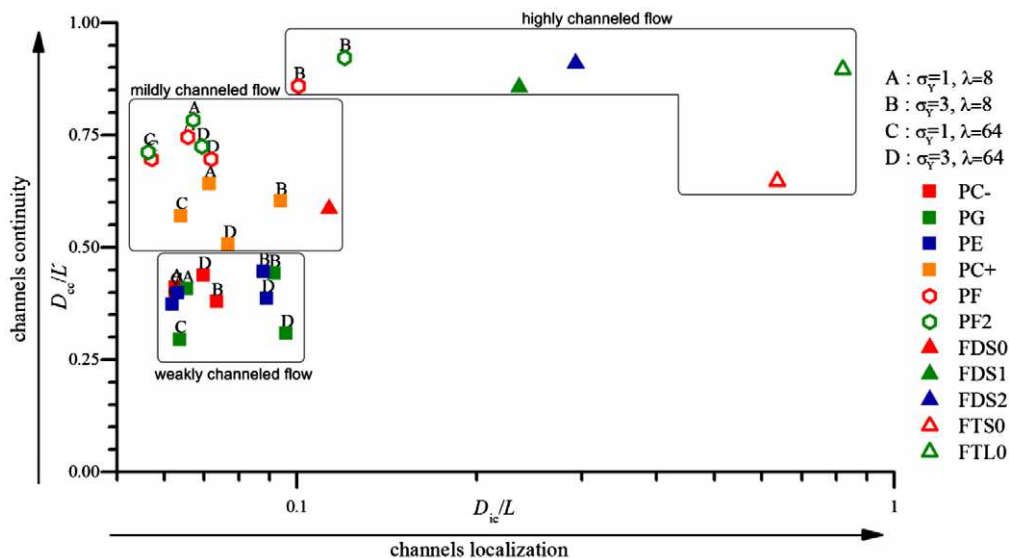


Fig. 7. D_{ic}/L versus D_{cc}/L' in porous and fractured fields. Porous cases are represented by squares, fractured cases by triangles and porous fractured cases by hexagons. The combination of the two indicators give an estimation of the channeling degree with weakly-channeled configurations (small D_{ic} , small D_{cc}), mildly-channeled configurations (high D_{cc} , small D_{ic}) and highly-channeled configurations (high D_{ic} , high D_{cc}).

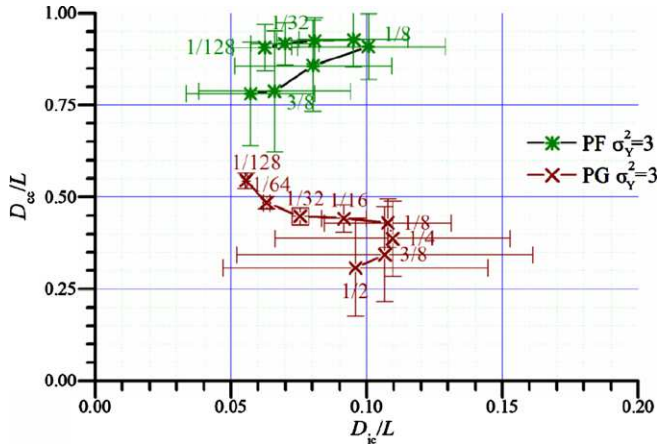


Fig. 9. Variation of D_{cc}/L' versus D_{ic}/L in porous test cases PG and PF for varying permeability correlation lengths λ . λ values are given next to the corresponding symbols.

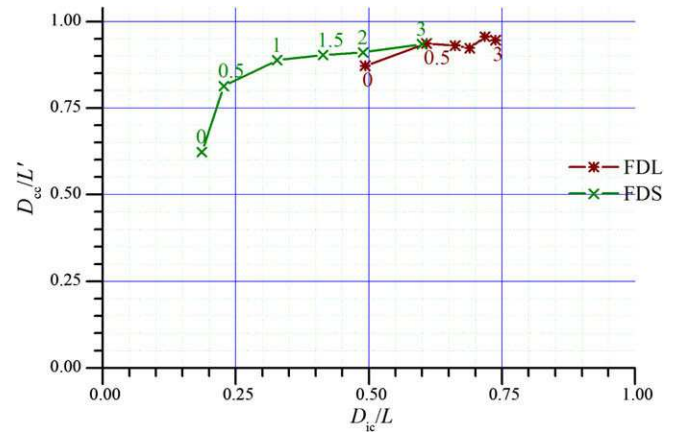


Fig. 12. D_{cc}/L' versus D_{ic}/L in fractured test cases with varying variances of the transmissivity distribution σ_y^2 . Values of σ_y^2 are given next to the corresponding symbols.

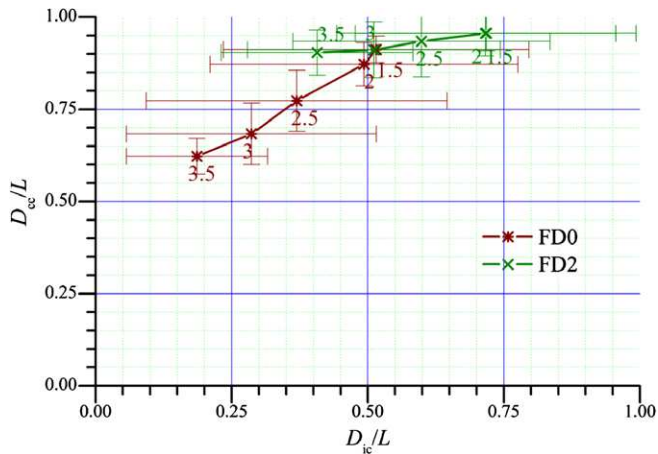


Fig. 10. D_{cc}/L' versus D_{ic}/L in fractured test cases with varying power-law length exponents a . Values of a are given next to the corresponding symbols.

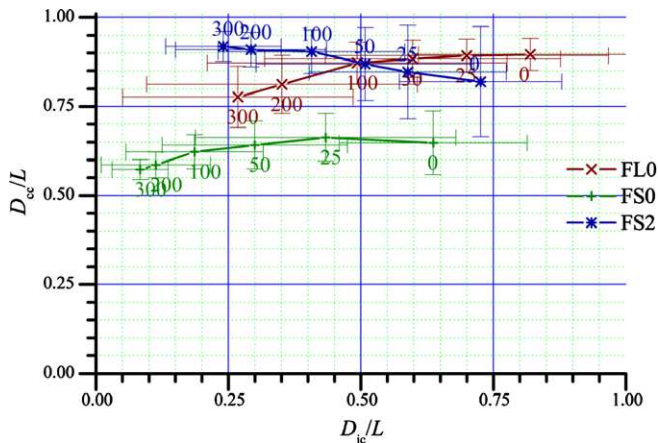


Fig. 11. D_{cc}/L' versus D_{ic}/L in fractured test cases with varying fracture densities d . Values of d are given next to the corresponding symbols. Density is measured as the percentage of fractures above percolation threshold. It is 0 at percolation threshold and 100 in networks having twice as much fractures as in networks at percolation threshold.

mately from $\lambda \sim L/8$ to $\lambda = L/2$ (Fig. 9). This may be due to two reasons. First, the standard deviation of D_{ic}/L steeply increases with λ

Table 4

Variation range (maximal minus minimal values) of D_{ic}/L and D_{cc}/L' according to porous and fracture parameters. High values in bold are due to transitions from porous-like to fracture-like structures rather than to variations with σ_y^2 .

	Porous cases		Fractured cases	
	D_{ic}/L	D_{cc}/L'	D_{ic}/L	D_{cc}/L'
Full range of variation	0.12	0.66	0.73	0.40
σ_y^2	0.08 (PG)	0.1 (PG)	0.42 ($a = 3.5$)	0.31 ($a = 3.5$)
	0.10 (PF)	0.55 (PF)	0.24 ($a = 2.0$)	0.07 ($a = 2.0$)
λ	0.06 (PG)	0.09 (PG)		
	0.04 (PF)	0.14 (PF)		
a			0.35 ($\sigma_y^2 = 0$)	0.30 ($\sigma_y^2 = 0$)
			0.22 ($\sigma_y^2 = 3$)	0.05 ($\sigma_y^2 = 3$)
d			0.55 ($a = 3.5$)	0.1 ($a = 3.5$)
			0.55 ($a = 2.0$)	0.12 ($a = 2.0$)

to the point where its variations become smaller than its variability. Second, when the large correlation length λ is comparable to the system size L , for example for $\lambda = L/2$, the system becomes more homogeneous and channels are more regularly distributed, explaining the smaller value of D_{ic}/L . Apart from these side effects, D_{ic}/L is first determined by the density of potential channels given by the correlation length in porous media and by fracture density and length distributions. Among the potential channels, only those made up of the higher permeabilities lead to effective channels. This is confirmed by the variation ranges of D_{ic}/L presented in Table 4. The largest variation ranges are due to the fracture density first and to the difference of D_{ic}/L values between porous and fracture cases (Fig. 6). They cover at least two thirds of the full variation of D_{ic}/L . The variation ranges according to σ_y^2 are smaller but not negligible and account on average for less than half of the full variation range of porous and fracture cases.

D_{cc}/L' is less variable than D_{ic}/L . It varies significantly only in PF as a function of σ_y^2 (Fig. 8), in FDS as a function of σ_y^2 (Fig. 12), in PG as a function of λ (Fig. 9) and in FD0 as a function of a (Fig. 10). We argue in the following that the sole genuine variation is the last one. The two first variations are due to the transition from porous to fractured cases. Beyond the transition ($\sigma_y^2 > 1$), D_{cc}/L' is almost constant. In porous-fracture fields, small σ_y^2 values correspond to almost pure porous cases without fractures whereas high σ_y^2 values lead to fracture-like cases. In dense fracture networks with $\sigma_y^2 = 0$, the fracture network looks like a porous medium. Increasing σ_y^2 triggers channeling while keeping D_{cc}/L' almost constant. D_{cc}/L' decreases with the correlation length. This counter-intuitive result is

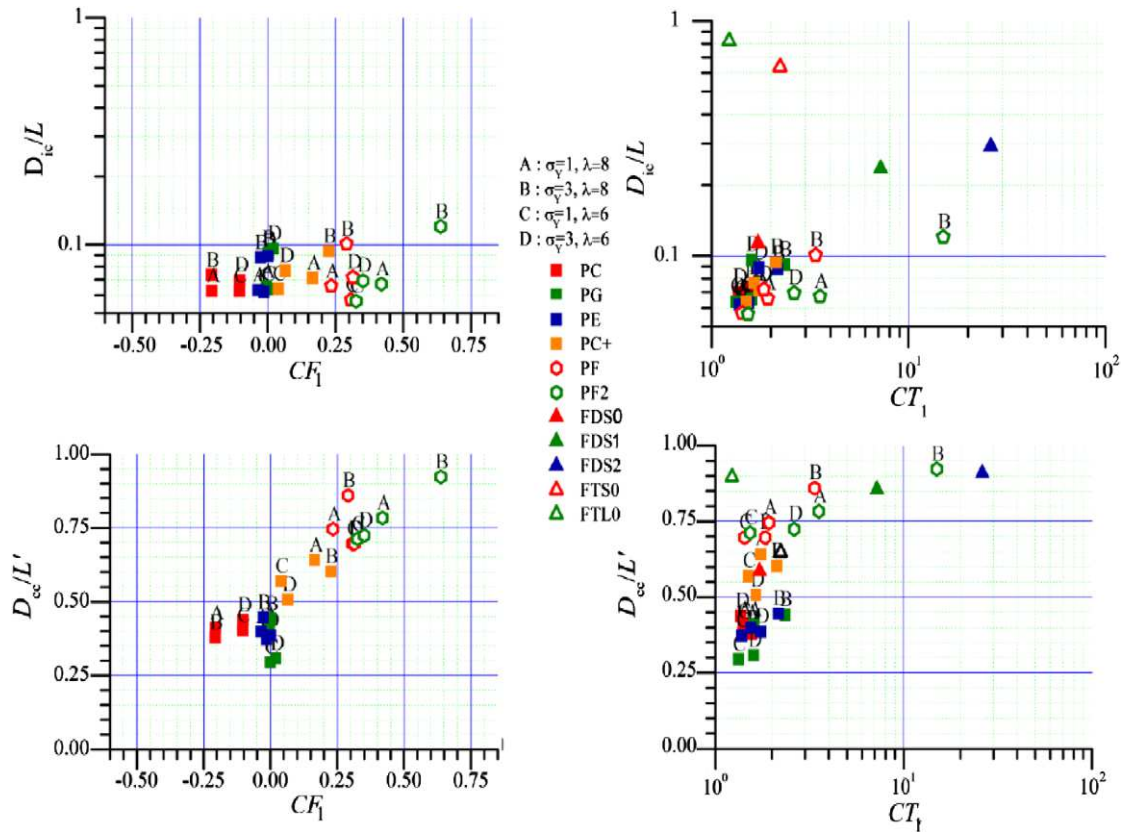


Fig. 13. Channeling characteristics (D_{ic}/L and D_{cc}/L') versus indicators CF_1 and CT_1 defined in [20]. Note that D_{ic} is available only in porous cases.

only apparent and mostly due to the simultaneous variations of L and λ in our simulation settings. In fact, we have found that D_{cc}/L' depends more on L/λ than on λ . The range of variations of D_{cc}/L' is reduced from 0.25 to 0.09 when decreasing the range of variations of L/λ from 64 to 8. Disregarding these dummy variations, the sole genuine variation that is not a transition is the increase in D_{cc}/L' with smaller a values for dense networks (FDO). What fundamentally changes in the latter case is the nature of the correlations. D_{cc}/L' seems to be more affected by the nature of the correlation than by the variability of permeability. Aside from the large variation ranges due to a transition from porous-like to fracture-like media marked in bold in Table 4, it is the sole case where the variation range of D_{cc}/L' is significant. More precisely, it is of the order of three quarters of the full range of variations in all fracture networks, where all other cases are restricted to one quarter. The channel continuity measured by D_{cc}/L' is thus much more influenced by the nature of the correlation structure than by the other parameters including the permeability variability, the fracture density and the correlation length. D_{cc}/L' can be considered as an indicator of the nature of correlation. Finally, the absence of correlation between variability and D_{cc}/L' expresses that there is a fundamental limit in channeling related to the local permeability structure rather than to the permeability variability.

5. Discussion

As concluded in the previous section, the D_{cc}/L' ratio depends much more on the nature of the correlation than on the other parameters. The continuity of channels is at first a function of the occurrence of underlying-connected permeability structures. D_{ic}/L is more intuitively a function of the density of paths and of the selection of the highest permeability paths.

Although D_{ic}/L and D_{cc}/L' have been defined as statistical characteristics of the fields as shown in Fig. 2. We recall that D_{ic} has been computed with a separation of flow tubes into 20 parts and thus measures statistical properties of structures carrying at least 1/20 of the total flow. D_{ic} is only slightly variable for porous media and gives a characteristic scale of the low-flow zones (blue patches). In simple multi-Gaussian cases (PG), D_{ic} is close to the correlation length of the velocity field. D_{ic} is much more variable in the fracture networks where it is not too far from the distance between effectively flowing structures. On the other hand, D_{cc} is more variable in the porous media than in the fracture networks. In the porous media, it gives a characteristic scale of the extension of the high-flow zones (red¹ flow tubes in Fig. 2) and in fractured media, it gives a characteristic length of flow channels.

As displayed in Fig. 7, D_{cc}/L' and D_{ic}/L are not strongly correlated. Their dependences on the different model parameters (Figs. 8–12) explain the lack of strong correlation. Consequently, D_{ic} and D_{cc} actually do measure two different channeling characteristics that are only weakly interdependent. In other words, they complementarily characterize channeling.

Fig. 13 displays the relation between the new indicators D_{ic}/L and D_{cc}/L' and the existing indicators CF_1 and CT_1 [20]. CT_1 does not systematically identify the sparse fracture cases (FTS0 and FTL0). Furthermore, CT_1 values within a single case are highly variable in channeled media (Fig. 14). CT_1 is thus not a good channeling indicator. The correlation of CF_1 with D_{cc}/L' is apparently better than with D_{ic}/L . $D_{cc}/L' < 0.5$ corresponds to weakly negative CF_1 values, indicating flow fields slightly more influenced by low-per-

¹ For interpretation of the references to color in this figure, the reader is referred to the web version of this article.

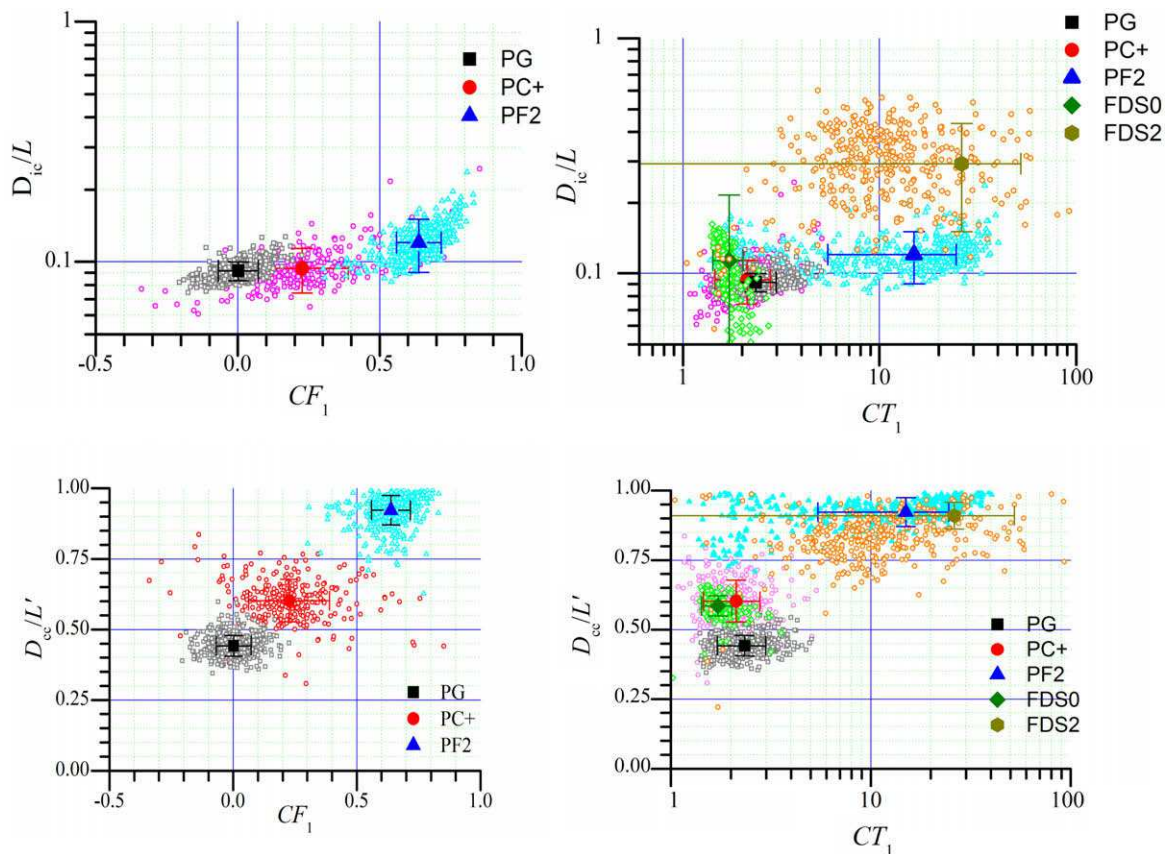


Fig. 14. D_{ic}/L and D_{cc}/L' versus CF_1 and CT_1 . This figure shows the variability of the indicators by depicting the 500 values for PG, PC+ and PF2 ($\sigma_v^2 = 3, \lambda = 8$) and, if available, FDS0 and FDS2. Small points stand for the 500 results and large points for the associated mean. Error bars stand for the standard deviation. D_{ic} and D_{cc} are generally less variable in the same configurations than CF_1 , and particularly than CT_1 .

meability zones. $D_{cc}/L' > 0.5$ corresponds to positive CF_1 values indicating flow fields more influenced by the high-permeability zones making up the channels. The advantage of D_{cc}/L' over CF_1 is that the correlation structures are better distinguished. For example, PC+ and PF have very close CF_1 values (Fig. 6) but differ by their D_{cc}/L' values. This difference stems from the more contrasted channels in the PF configuration relative to the PC+ configuration, despite the similar channeling intensity. However, it must be noted that CF_1 better distinguishes PF from PF2 than D_{ic}/L and D_{cc}/L . The second advantage of D_{cc}/L' over CF_1 is that it provides information on the flow structure in both porous and fracture cases.

D_{ic}/L and D_{cc}/L' could also be readily computed in anisotropic and 3D fields. We expect both anisotropy and 3D to increase D_{cc}/L' without significantly modifying D_{ic}/L . D_{ic}/L will remain linked to the volume of the low-flow zones, the characteristic size of which will not be strongly modified. However, we expect anisotropy to increase D_{cc}/L' just by the effect of the higher velocity correlations in the flow direction. 3D could also potentially provide longer and more tortuous channels around the low-flow zones, and hence increase D_{cc}/L' .

The derivation of D_{ic}/L and D_{cc}/L' in natural cases is more difficult because of the lack of data that would lead to their direct estimate. They could however be inferred from the geometrical and hydraulic characteristics of the permeability field either with the results of this study or with closer simulations. It would be interesting, in field cases, to condition the estimation of these indexes on permeability and flow values and thus to lower the non-negligible variability displayed in Fig. 14.

6. Summary and conclusion

Channeling has been observed both in field and in synthetic contexts. However, its characterization has been essentially qualitative. We introduce two statistical indicators based on the distribution of flow rates and compute them on a wide variety of porous and fracture permeability fields. The tested fields range, in porous media, from multi-Gaussian fields with classical correlation laws (Gaussian and exponential) to permeability fields rearranged to enhance channeling using the method by Zinn and Harvey and to permeability fields rearranged to mimic the presence of fractures within the field. The tested fractured media cover a broad range of fracture lengths, densities and transmissivity distributions.

The first indicator, D_{ic} , is related to the characteristic inter-channel distance. It is based on the participation ratio S_2 applied to the distribution of flow-tube widths. Although statistically derived, this ratio can be interpreted as a characteristic scale of the low-flow zones perpendicular to the flow. It is moreover highly sensitive to the variability of the permeability, as well as to the permeability correlation pattern. The second indicator, D_{cc} , is related to a characteristic extension of channels. It is too computed with the participation ratio S_2 applied to the Lagrangian derivative distribution of flow rates. It is highly sensitive to the nature of the permeability correlation structure, which is also an important channeling cause. Both indicators consistently characterize flow channeling in porous and fractured media, with D_{ic} being more sensitive in fractured media and D_{cc} in porous media. As they are weakly correlated, they measure different channeling characteristics that are weakly dependent and are complementary to charac-

terize channeling in porous and in fractured media. As a result, they are complementary to identify and quantify channeling in various media, from non-channeled fields like multi-Gaussian permeability fields with common correlation laws to highly-channeled media like porous-fractured fields with a large variability and fracture network with large fractures and broadly distributed transmissivity values. We will use the proposed indicators in further studies to distinguish weakly-, mildly- and highly-channeled media in order to choose the most relevant modeling framework and identification strategies.

Acknowledgements

The French National Research Agency ANR is acknowledged for its financial founding through the MOHINI project (ANR-07-VULN-008) and for providing for numerical methods and computational means through the MICAS project (ANR-07-CIS7-004). The authors thank two anonymous reviewers for their highly detailed and constructive reviews.

References

- [1] Beaudoin A, de Dreuzy J-R, Erhel J. An efficient parallel particle tracker for advection–diffusion simulations in heterogeneous porous media. In: Euro-Par 2007 parallel processing. Berlin/Heidelberg: Springer; 2007.
- [2] Bonnet E, Bour O, Odling NE, Davy P, Main I, Cowie P, et al. Scaling of fracture systems in geological media. *Rev Geophys* 2001;39:347–83.
- [3] Bruderer-Weng C, Cowie P, Bernabé Y, Main I. Relating flow channelling to tracer dispersion in heterogeneous networks. *Adv Water Resour* 2004;27:843–55.
- [4] Davis TA. Algorithm 832: UMFPACK, an unsymmetric-pattern multifrontal method. *ACM Trans Math Softw* 2004;30:196–9.
- [5] Davy P, Hansen A, Bonnet E, Zhang S-Z. Localization and fault growth in layered brittle–ductile systems: implications for deformations of the continental lithosphere. *J Geophys Res* 1995;100:6281–94.
- [6] de Dreuzy J-R, Davy P, Bour O. Hydraulic properties of two-dimensional random fracture networks following a power law length distribution 1. Effective connectivity. *Water Resour Res* 2001;37:2065–78.
- [7] de Dreuzy JR, Beaudoin A, Erhel J. Asymptotic dispersion in 2D heterogeneous porous media determined by parallel numerical simulations. *Water Resour Res* 2007;43:13.
- [8] de Marsily G, Delay F, Gonçalves J, Renard P, Teles V, Violette S. Dealing with spatial heterogeneity. *Hydrogeol J* 2005;12:161–83.
- [9] Desbarats AJ. Spatial averaging of hydraulic conductivity in three-dimensional heterogeneous porous media. *Math Geol* 1992;24:249–67.
- [10] Fogg GE. Groundwater flow and sand body interconnectedness in a thick, multiple aquifer system. *Water Resour Res* 1986;22:679–94.
- [11] Fogg GE, Carle SF, Green C. Connected-network paradigm for the alluvial aquifer system. In: Zhang D, Winter CL, editors. Theory, modeling and field investigation in hydrogeology: a special volume in honor of Shlomo P Neuman's 60th birthday. Geological Society of America; 2000. p. 25–42.
- [12] Frigo M, Johnson SG. The design and implementation of FFTW3. *Proc IEEE* 2005;93:216–31.
- [13] Frippiat CC, Illangasekare TH, Zyvoloski GA. Anisotropic effective medium solutions of head and velocity variance to quantify flow connectivity. *Adv Water Resour* 2009;32:239–49.
- [14] Guswa AJ, Freyberg DL. On using the equivalent conductivity to characterize solute spreading in environments with low-permeability lenses. *Water Resour Res* 2002;20.
- [15] A.L. Gutjahr, Fast Fourier transforms for random field generation, New Mexico Tech project report 4-R58-2690R, 1989.
- [16] Hanor JS. Effective hydraulic conductivity of fractured clay beds at a hazardous waste landfill; Louisiana Gulf Coast. *Water Resour Res* 1993;29:3691–8.
- [17] Journel A, Alabert FG. Focusing on spatial connectivity of extreme-valued attributes: stochastic indicator models of reservoir heterogeneities. *AAPG Bull* 1989;73.
- [18] Journel AG, Deutsch CV, Desbarats AJ, Stanford A. Power averaging for block effective permeability, SPE California regional meeting, 2–4 April 1986, Soc. Petro. Eng., Oakland, California, 1986.
- [19] Kerrou J, Renard P, Hendricks Franssen H-J, Lunati I. Issues in characterizing heterogeneity and connectivity in non-multiGaussian media. *Adv Water Resour* 2008;31:147–59.
- [20] Knudby C, Carrera J. On the relationship between indicators of geostatistical, flow and transport connectivity. *Adv Water Resour* 2005;28:405–21.
- [21] Knudby C, Carrera J. On the use of apparent hydraulic diffusivity as an indicator of connectivity. *J Hydrol* 2006;329:377–89.
- [22] Knudby C, Carrera J, Bumgardner JD, Fogg GE. Binary upscaling – the role of connectivity and a new formula. *Adv Water Resour* 2006;29.
- [23] Koltermann CE, Gorelick SM. Heterogeneity in sedimentary deposits: a review of structure imitating, process-imitating and descriptives approaches. *Water Resour Res* 1996;32:2617–58.
- [24] Krishnan S, Journel A. Spatial connectivity: from variograms to multiple-point measures. *Math Geol* 2003;35:915–25.
- [25] Krueel-Romeu R, Noetinger B. Calculation of internodal transmissivities in finite difference models of flow in heterogeneous porous media. *Water Resour Res* 1995;31:943–59.
- [26] Le Borgne T, de Dreuzy JR, Davy P, Bour O. Characterization of the velocity field organization in heterogeneous media by conditional correlation. *Water Resour Res* 2007;43:10.
- [27] Matheron G. Elements pour une théorie des milieux poreux, Masson et Cie, 1967.
- [28] Moreno L, Tsang CF. Flow channeling in strongly heterogeneous porous media. *Water Resour Res* 1994;30:1421–30.
- [29] Park C-H, Beyer C, Bauer S, Kolditz O. A study of preferential flow in heterogeneous media using random-walk particle tracking. *Geosci J* 2008;12.
- [30] Ronayne MJ, Gorelick SM. Effective permeability of porous media containing branching channel networks. *Phys Rev E* 2006;73.
- [31] Ronayne MJ, Gorelick SM, Caers J. Identifying discrete geologic structures that produce anomalous hydraulic response: an inverse modeling approach. *Water Resour Res* 2008;44.
- [32] Sánchez-Vila X, Carrera J, Girardi JP. Scale effects in transmissivity. *J Hydrol* 1996;183:1–22.
- [33] Scheibe T, Yabusaki S. Scaling of flow and transport behavior in heterogeneous groundwater systems. *Adv Water Resour* 1998;22:223–38.
- [34] Silliman SE. An interpretation of the difference between aperture estimates derived from hydraulic and tracer tests in a single fracture. *Water Resour Res* 1989;25:2275–83.
- [35] Sornette A, Davy P, Sornette D. Fault growth in brittle–ductile experiments and the mechanics of continental collisions. *J Geophys Res* 1993;98(12):11.
- [36] Tompson AFB, Gelhar LW. Numerical-simulation of solute transport in 3-dimensional, randomly heterogeneous porous-media. *Water Resour Res* 1990;26:2541–62.
- [37] Trinchero P, Sánchez-Vila X, Fernández-García D. Point-to-point connectivity, an abstract concept or a key issue for risk assessment studies? *Adv Water Resour* 2008;31:1742–53.
- [38] Tsang C-F, Neretnieks I. Flow channeling in heterogeneous fractured rocks. *Rev Geophys* 1998;36.
- [39] Tsang YW, Tsang CF. Flow channeling in a single fracture as a two dimensional strongly heterogeneous permeable medium. *Water Resour Res* 1989;25:2076–80.
- [40] Warren JE, Price HH. Flow in heterogeneous porous media. *Soc Petr Eng* 1961;1:153–69.
- [41] Wen X-H, Gomez-Hernandez JJ. Numerical modeling of macrodispersion in heterogeneous media – a comparison of multi-Gaussian and non-multi-Gaussian models. *J Cont Hydrol* 1998;30:129–56.
- [42] Western AW, Blöschl G, Grayson RB. Toward capturing hydrologically significant connectivity in spatial patterns. *Water Resour Res* 2001;37:83–97.
- [43] Zinn B, Harvey CF. When good statistical models of aquifer heterogeneity go bad: a comparison of flow, dispersion and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields. *Water Resour Res* 2003;39.

B. Résultats complémentaires

Les indicateurs de chenalisation présentés dans cette partie reposent sur les propriétés lagrangiennes du flux (tube de courant, ligne de flux) plutôt que sur la distribution des valeurs de flux. Ce paragraphe présente donc les résultats obtenus à partir d'indicateurs plus classiques, comme l'écart type ou la longueur de corrélation.

1. Statistiques sur la distribution des valeurs de flux

La chenalisation des écoulements est provoquée par l'hétérogénéité du milieu et par la structure de corrélation entre les zones de forte transmissivité. Intuitivement, on peut donc supposer que la variabilité des valeurs de flux, mesurée par l'écart-type, ainsi que leur corrélation, mesurée par la longueur de corrélation, sont des outils pertinents pour quantifier le chenalisation des écoulements. De plus, la moyenne des écoulements devrait permettre de mesurer l'importance des chemins d'écoulements préférentiels dans la mesure où si le flux est fortement chenalisé, on peut s'attendre à obtenir une vitesse moyenne plus importante. Nous avons donc calculé ces différentes grandeurs pour certaines des configurations présentées au paragraphe précédent. La figure 4-1 montre les résultats obtenus.

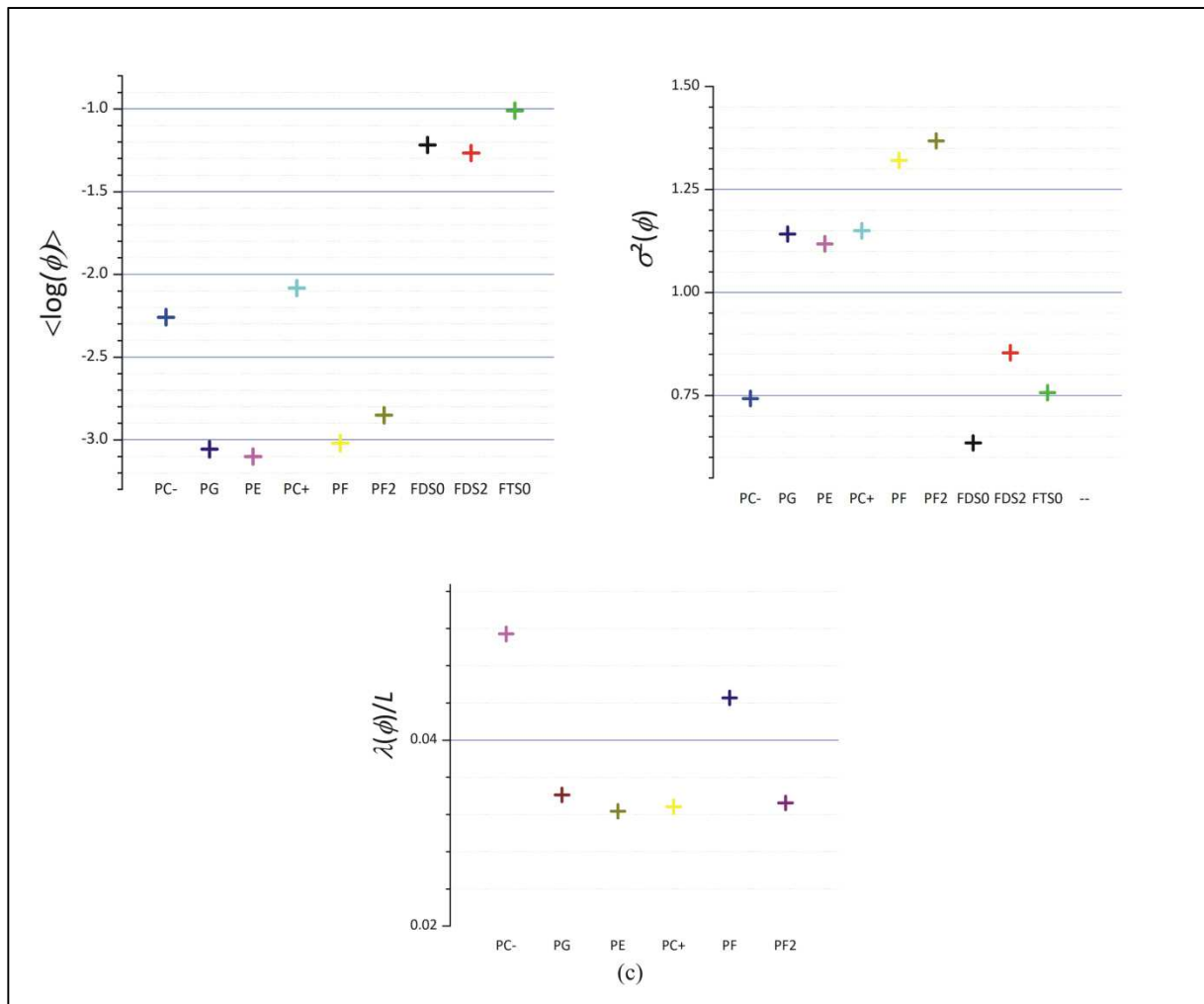


figure 4-1 : Valeurs moyenne du logarithme du flux (a), écarts-type (b) et longueurs de corrélation (c) calculés pour les configurations tests de la figure 4-6. Chaque valeur est prise comme la moyenne de 500 simulations.

Premièrement, on constate sur la figure 4-1-a, qu'une forte connectivité des zones les plus transmissives n'a qu'une incidence marginale sur la valeur moyenne des écoulements. C'est

particulièrement visible pour les configurations PF, réorganisées pour créer des structures connectées de forte transmissivité similaires à des fractures, qui ont un flux moyen proche de celui des configurations PG qui ne bénéficient pas de cette réorganisation. De la même manière, FDS0 et FTS0, qui diffèrent par leur densité de fracturation, FTS0 étant au seuil de percolation et FDS0 étant 3 fois plus dense, ont également des flux moyens très proches. On ne peut donc pas utiliser la valeur moyenne du flux pour distinguer un milieu chenalisé d'un milieu non chenalisé.

Si la moyenne n'est pas pertinente, on s'attend néanmoins à ce que la variabilité des valeurs de flux soit plus importante dans un milieu chenalisé, avec des écoulements importants dans les chemins préférentiels et des écoulements très faibles en dehors de ces chemins. A partir des résultats de la figure 4-1-b, on observe que si l'écart-type est effectivement légèrement supérieur pour les configurations PF par rapport aux configurations PG, les configurations PC+, réorganisées statistiquement pour favoriser la connectivité des fortes valeurs de perméabilités, ont un écart-type proche de celui des milieux non-réorganisés. Ceci signifie que la localisation des fortes valeurs de perméabilité ne se traduit pas nécessairement par une amplification des écarts entre flux faibles et forts mais plutôt par une plus grande organisation entre ces valeurs. De plus, la variabilité dans les milieux fracturés est nécessairement plus faible que dans les milieux poreux, dans la mesure où les écoulements sont de toute façon localisés à l'intérieur du réseau de fracture. Par conséquent, la variabilité des valeurs de flux ne permet pas non plus de quantifier la chenalisation des écoulements.

Comme illustré à la figure 4-1-c, la longueur de corrélation ne permet pas de distinguer les milieux fortement chenalisés. En effet, cette grandeur est sensible à la fois aux corrélations entre les fortes valeurs de flux, mais également entre les faibles valeurs, ce qui explique la valeur importante, en comparaison, des configurations PC-. De plus les indicateurs basés sur le variogramme de la distribution ne sont pas pertinents pour mesurer la connectivité du milieu, comme l'ont montré [Western et al., 1998] et [Knudby and Carrera, 2005].

2. L'utilisation du ratio de participation S_2

Puisque les indicateurs les plus intuitifs ne permettent pas de mesurer le degré de chenalisation d'un milieu, il est nécessaire d'utiliser des outils plus élaborés comme le ratio de participation S_2 [Davy et al., 1995]. Cependant, l'application directe de S_2 au champ de flux n'est pas satisfaisante. La figure 4-2 montre les valeurs de $S_2(\varphi)$, où φ représente le flux.

S_2 quantifie la surface occupée par les fortes valeurs d'une distribution par rapport à la surface totale. Une valeur proche de 1 signifie que toutes les valeurs de la distribution sont proches les unes des autres alors qu'une valeur proche de 0 signifie que quelques valeurs sont significativement plus fortes que le reste de la distribution. En l'appliquant aux valeurs de flux (figure 4-2-a), on espère donc que les milieux chenalisés, où les fortes valeurs sont localisées dans les chemins préférentiels d'écoulement, présenteront une valeur de S_2 significativement inférieure aux milieux non chenalisés. Si les résultats obtenus conviennent pour la plupart des configurations test, on observe cependant que la valeur de S_2 pour les configurations PC+ est très proche des configurations PG et PE, non organisées. Ceci s'explique par le fait que S_2 n'est pas sensible à l'organisation entre fortes et faibles valeurs, mais uniquement à l'importance relative des unes par rapport aux autres. Par ailleurs, la comparaison entre les valeurs de S_2 des milieux poreux et des milieux fracturés est impossible dans la mesure où le support est différent. Dans les milieux poreux, les écoulements se produisent sur toute la surface du milieu alors que dans les milieux fracturés, ces écoulements sont localisés dans les fractures, ce qui donne des valeurs artificiellement basses. Il est par conséquent très difficile de classer les milieux fracturés par rapport aux milieux poreux.

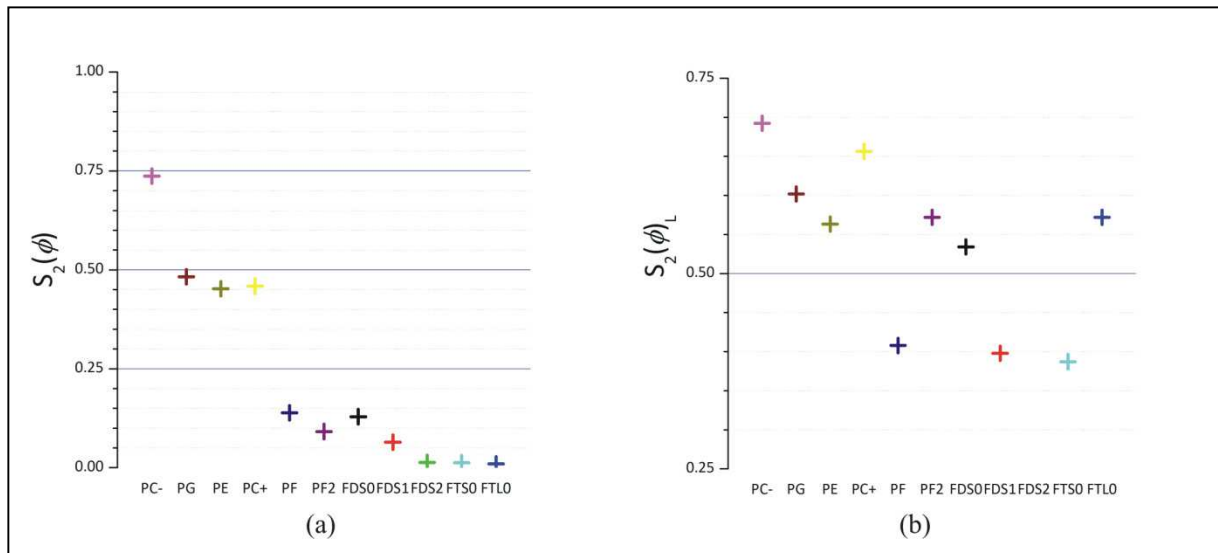


figure 4-2 : valeurs de $S_2(\phi)$ pour les différentes configurations présentées à la figure 4-6. Les valeurs présentées correspondent à la moyenne calculée à partir de 500 simulations.

Une solution aurait pu être de calculer S_2 sur la distribution lagrangienne des flux (figure 4-2-b). En effet, le champ lagrangien suit les lignes de flux, ce qui résout la question du support tout en favorisant les chemins de fort écoulement. Cependant cette solution présente le même inconvénient que l'utilisation de la longueur de corrélation et on obtient un S_2 fort pour certains milieux très chenalisés (PF2 et FTL0 par exemple) dans la mesure où ceux-ci ont des lignes de flux concentrées dans les chemins préférentiels et on donc une distribution plutôt homogène dominée par les fortes valeurs. La valeur de S_2 est dans ce cas proche des valeurs obtenues pour les milieux non chenalisés, où les distributions sont également plutôt homogènes, mais dominées par les faibles valeurs. Le problème vient ici du fait que S_2 n'est pas sensible aux valeurs absolues mais à l'écart relatif entre les faibles et les fortes valeurs, ce qui ne permet pas de distinguer, par exemple une distribution D_1 d'une distribution D_2 dont les valeurs seraient toutes égales à n fois les valeurs de D_1 .

En conclusion, quantifier le degré de chenalisation des écoulements dans les milieux poreux et hétérogène nécessite d'utiliser des indicateurs élaborés, appliqués à des propriétés statistiques plus caractéristiques de la chenalisation que la simple distribution des valeurs de flux. C'est pourquoi les indicateurs proposés se basent sur la distribution des largeurs des tubes de courant et sur la dérivée spatiale du champ lagrangien.

C. Conclusion

Les milieux fortement hétérogènes, où les écoulements sont très chenalisés, ont des propriétés hydrauliques particulières, dominées par les structures très transmissives qui forment des chemins préférentiels d'écoulement. Il est donc difficile de remplacer ces milieux par des modèles homogénéisés n'intégrant pas ces caractéristiques. C'est pourquoi il est important d'évaluer le degré de chenalisation des écoulements afin de mettre au point des modèles pertinents. Comme la chenalisation est provoquée par l'hétérogénéité du milieu et par la présence de motifs de corrélation entre zones de forte transmissivité, des indicateurs statistiques simples comme la variabilité des valeurs de flux ne permettent pas de distinguer un milieu chenalisé d'un milieu non chenalisé. C'est pourquoi nous avons mis au point des indicateurs basés sur les propriétés physique des chenaux, c'est-à-dire sur le fait qu'ils localisent les forts écoulements et qu'ils sont continus tout au long du milieu. Les indicateurs présentés dans cette partie évaluent la distance caractéristique entre deux chenaux principaux (D_{ic}) et la persistance du chenal dans le milieu (D_{cc}).

À partir de ces indicateurs, on a mis en évidence que les formes des corrélations entre fortes valeurs amplifiaient la chenalisation en augmentant la persistance des chenaux, ce qui se traduit par la création de chemins préférentiels d'écoulement. L'hétérogénéité du milieu se traduit par une localisation des fortes valeurs de flux, ce qui augmente la distance entre les différents chenaux. Dans les milieux poreux, l'hétérogénéité est produite par la variabilité des valeurs de perméabilité alors que les corrélations sont formées par l'organisation de ces fortes valeurs. Dans les milieux fracturés, l'hétérogénéité est produite par la distribution des ouvertures des fractures et par leur densité spatiale alors que les corrélations dépendent essentiellement de la présence de grandes fractures et éventuellement de l'organisation de ces grandes fractures. Il apparaît également que les milieux fracturés sont particulièrement sujets à la chenalisation, dans la mesure où les écoulements sont naturellement localisés dans le réseau de fracture. Seuls les milieux fracturés très denses, avec des ouvertures et des longueurs homogènes peuvent être considérés comme non chenalisés, ce qui rejoint les conclusions de Long et al. [Long et al., 1982].

Ces indicateurs ont permis de distinguer trois types de milieux en fonction de leur degré de chenalisation. Les milieux non chenalisés comme ceux obtenus par des champs multi-Gaussiens sont caractérisés par des valeurs de D_{ic} et de D_{cc} proches de 0. Pour ces milieux, les modèles homogénéisés sont particulièrement bien adaptés. Les milieux moyennement chenalisés, comme ceux obtenus avec des champs de perméabilités réorganisés ou pour des réseaux de fractures denses, sont caractérisés par une faible valeur de D_{ic} mais une valeur de D_{cc} plus importante. Dans ces milieux, on note la présence de chemins préférentiels d'écoulement sans pour autant que cette organisation domine largement les propriétés hydrauliques, ce qui signifie que le flux en dehors des chemins préférentiels ne peut être négligé. Finalement, les milieux fortement chenalisés ont des valeurs de D_{ic} et de D_{cc} proches de la taille du système. Dans ces milieux, les écoulements sont concentrés dans un ou deux chenaux principaux et les structures formant ces chenaux dominant largement les propriétés d'écoulement. Ces milieux sont les plus compliqués à modéliser dans la mesure où il est nécessaire d'intégrer le rôle des chenaux principaux d'écoulement dans le modèle. Cependant, comme ces structures sont très localisées, la probabilité de les observer directement sur le terrain est quasi-nulle. C'est pourquoi on doit utiliser soit l'approche stochastique décrite à la partie 3, qui présente l'inconvénient de nécessiter un très grand nombre de données pour être pertinente, soit l'approche inverse, décrite à la partie 5.

Partie 5 Identification des chenaux d'écoulement principaux par méthode inverse

L'organisation des écoulements en chemins préférentiels influe de manière significative sur les propriétés hydrauliques des milieux hétérogènes. Il est donc nécessaire de prendre en compte cette chenalisation afin d'obtenir des modèles pertinents. La difficulté majeure réside dans le fait que la chenalisation se traduit par une localisation des écoulements dans un petit nombre de structures. Ces structures, bien qu'ayant une influence majeure sur les propriétés hydrauliques, représentent une partie infinitésimale de la structure de perméabilité. S'il est possible de définir un modèle général de perméabilité à partir des observations du sous-sol, par interpolation ou en définissant un modèle statistique (voir partie 3), ou d'évaluer le degré de chenalisation des écoulements en mesurant certaines propriétés hydrauliques (voir partie 4), localiser et caractériser les structures porteuses des principaux écoulements n'est pas possible à partir d'observations directes du milieu. Néanmoins, comme les propriétés hydrauliques sont fortement dépendantes de la structure de la chenalisation, il est possible de les utiliser pour conditionner la structure des perméabilités et si possible d'identifier les éléments les plus transmissifs. C'est toute la problématique du problème inverse en hydrogéologie.

Cette partie est consacrée au problème inverse appliqué aux milieux fracturés présentant une forte chenalisation des écoulements. Un premier paragraphe introduit la méthode de résolution mise au point dans le cadre de ce doctorat sous la forme d'un article scientifique soumis à la revue *Advances in Water Resources* [Le Goc et al., 2010a]. Un second paragraphe fait le lien entre la structure de chenalisation, l'information contenue dans les charges hydrauliques et les conditions aux limites. Ce second paragraphe est également présenté sous la forme d'un article scientifique et fera l'objet d'une soumission à la revue *Ground Water*. Finalement, un dernier paragraphe synthétise les résultats obtenus et conclut cette partie.

A. Identification des chenaux d'écoulement dans les milieux fracturés par stratégie inverse.

Ce paragraphe introduit l'article scientifique consacré au problème inverse dans les milieux fortement chenalisés.

1. Résumé

L'identification de la structure physique des milieux hétérogènes passe par la résolution d'un problème inverse. Pour résoudre ce problème, il est nécessaire de prendre en compte ses caractéristiques. On doit en effet faire face à la non-linéarité du problème et au fait qu'il est mal posé et sous-déterminé. De plus, l'hétérogénéité de la fracturation des milieux étudiés complique la résolution. Puisque les écoulements dans les milieux fracturés sont généralement très chenalisés, on propose une méthode de résolution basée sur la localisation et la caractérisation des structures sous-jacentes aux chenaux d'écoulement afin de réduire le nombre de paramètres du problème inverse. Ce problème est plus complexe que le problème inverse classique puisqu'il s'agit d'identifier à la fois la structure du milieu et ses transmissivités. On représente la structure de chenalisation par la géométrie, la connectivité et la transmissivité des chenaux qui la composent. On augmente ensuite progressivement le

nombre de paramètre du modèle afin d'éviter les problèmes de sur et sous-paramétrisation au travers d'un procédé d'identification en plusieurs étapes. A chaque étape, la valeur des paramètres est optimisée grâce à un algorithme de recuit-simulé. Comme le problème est mal posé, on applique le procédé d'identification avec différentes conditions initiales et les solutions obtenues sont analysées par un algorithme d'agglomération itératif afin de déterminer les caractéristiques les plus pertinentes. La méthodologie est appliquée sur un ensemble de configurations ayant des degrés et des formes de chenalisation différents. A partir de mesures de charges hydrauliques en stationnaire, on montre que la méthode proposée permet de localiser efficacement les structures contenant les principaux chenaux d'écoulement dans la plupart des cas. Pour les configurations les plus simples, tous les chenaux sont identifiés en appliquant le procédé une seule fois. Lorsque la structure des écoulements est plus complexe, les solutions individuelles sont moins précises mais la méthode d'analyse permet d'identifier malgré tout ces structures. Les modèles synthétiques basés sur la combinaison des structures les plus pertinentes intègrent en effet les caractéristiques les plus importantes de la chenalisation et les données simulées sont alors fidèles aux données observées. En conclusion, l'approche discrète se révèle efficace pour modéliser les écoulements des milieux fortement chenalisés. Les données de charges contiennent en effet suffisamment d'information pour localiser les chenaux principaux y compris pour des écoulements complexes.

2. Inverse problem strategy to identify flow channels in fractured media

R. Le Goc ^{a,b}, J.-R. de Dreuzy ^a and P. Davy ^a

^a Géosciences Rennes, UMR 6118 CNRS, Université de Rennes 1, CS 74205, F-35042 Rennes Cedex, FRANCE

^b Itasca Consultants SAS, 64 chemin des mouilles, F-69134, Écully Cedex, FRANCE

Abstract

We set up a methodology for identifying highly localized flow channels embedded in a significantly less permeable media from steady-state head and geometrical data. This situation is typical of fractured media where flows are often strongly channeled at the scales of interest (10 m-1 km). The objective is to identify both geometrical and hydraulic characteristics of the conducting structures. Channels are identified by decreasing order of importance by using successive optimizations of an objective function. The identification strategy takes advantage of the hierarchical flow organization to restrict the dimension of the solution space of each individual optimization. Because of the successive optimizations, main flow channels determine strongly the characteristics of secondary channels. Additionally, main flow channels can be slightly modified by secondary channels through the introduction of a regularization term on the main channel characteristics in the objective function. As the objective function is strongly non-convex with numerous local minima, inversion is performed by a stochastic algorithm (simulated annealing). We assess the whole identification strategy on a set of twenty elementary synthetic cases with varying channeling degrees. Flow configurations are determined by a uniform head gradient in steady state. Boundary conditions are fixed. Data are taken on a regular grid of 25 piezometers. Flow structures dominated by at most two simple channels can be identified solely with head data. More complex flow structures with multiple interconnections cannot however be identified with head data alone. Identification becomes successful when including geometrical information on the distances of the piezometers to the closest channel. It succeeds in locating as much as

three complex and intricately connected channels. Tested configurations consisting in more than four complex channels were not identified because of the absence of strong enough hierarchy in the channel structure.

Keywords: inverse problem; simulated annealing; fractured media; flow channels

1. Introduction

Flow channeling is widespread in fractured media, mainly in rock of small permeability where flows concentrate in highly transmissive flow paths within a restricted number of fractures [Tsang and Neretnieks, 1998; Martinez-Landa and Carrera, 2005]. Highly channeled media refer here to media where the main channels, i.e. channels carrying the largest part of the flow, are sparse at the scale of interest. The hydrological properties of highly channeled media are mainly controlled by the continuity and connectivity of the channels [Fogg, 1986; Silliman, 1989; Hanor, 1993]. In such media, characterizing the hydraulic parameters comes down to identify these main channels and their hydraulic properties. Existing approaches based on extensive resistivity [Francese et al., 2009] or hydraulic data [Hao et al., 2008] yield zones of higher permeability rather than highly localized channels. The objective of this article is to explore the possibilities of identifying highly localized flow channels from head data. This is a rather new challenge of inverse problems in hydrogeology, where most of the developments have focused on continuous porous media with permeability fields parameterized by zones or geostatistical functions [McLaughlin and Townley, 1996; de Marsily et al., 2000; Carrera et al., 2005]. We thus look for solving the inverse problem with a parameterization relying on discrete flow structures.

Parameterization is known to be a key issue for handling inverse problems, since optimization algorithms can effectively identify only a limited number of parameters [Zimmerman et al., 1998]. Applied to fractured media, continuous-like parameterizations have led to the definition of wide fracture zones within a stochastic less permeable continuous medium [Lavenue and de Marsily, 2001; Vesselinov et al., 2001]. On the opposite, 3D Discrete Fracture Network parameterizations have led to more localized flow structures but failed in identifying the major flow channels [Cacas et al., 1990a; b]. Because of the complexity of the problem, either simplifications or introduction of additional constraints have been considered. Gwo [2001] simplified it by discretizing the channels on a regular grid. Grid elements were either highly transmissive fractures or almost impervious matrix. Fracture geometries and transmissivities were identified by a genetic algorithm. Flow path geometries could be correctly estimated only for single straight-line flow paths. Lighter parameterizations have also been proposed first by replacing the network structure by a scale invariant Iterated Function System (IFS) conditioned to transient flow tests [Doughty et al., 1994]. Optimization performed by a stochastic method (simulated annealing) yielded a broad range of highly different structures. Another simplification consisted in identifying in 2D only one or two single structures [Tiedeman et al., 1995; Renshaw, 1996] with simulated annealing or co-kriging. Identification has been successful only in the cases where the observation network surrounds the transmissive structure and covers an area close to the size of the fracture zone. The identification has been improved by introducing a decreasing precision with the distance to the well [Basquet et al., 2005]. Considering the lack of constraints, Day-Lewis [2000] used inter-well connectivity data and hydraulic data to estimate successively the geometry and the transmissivity of four fractured zones. He showed that accounting for discrete highly transmissive structures improves significantly the model accuracy. Following the same spirit, standard parameterizations used in the petroleum industry and in the storage of high-level

nuclear wastes now rely first on the definition of a reference geometrical structure of the highly transmissive zones and secondly on the identification of hydraulic properties of fixed fracture families [Donado et al., 2005; Bruyelle and Lange, 2009; Grenier et al., 2009].

Compared to these approaches, we propose a parameterization designed to identify both the channels geometry and their permeability. The parameterization is based on the main flow channels rather than on the fracture zones. It leads to much more localized flows than those obtained in heterogeneous porous parameterizations, while it requires fewer parameters and details than DFN-like parameterizations. The identification strategy is based on a hierarchical analysis of the flow paths. Structures of decreasing importance are identified by successive objective function optimizations performed by a simulated annealing algorithm. A step-by-step procedure is developed where results from first-order channels are slightly modified by adding second-order links. The objective of this paper is to define the parameterization in terms of flow channels and to setup the identification methodology on 2D synthetic flow channel structures from simple hydraulic and geometrical data. Interests and limitations of these data as well as possible extensions of the inversion methodology with other data will be discussed in section 5. Beforehand, we present the parameterization adapted to highly channeled media in section 2 and the strategy to solve the inverse problem in section 3. We test the methodology in section 4 on increasingly complex fracture flow structures with either head or distance data or a combination of both of them.

2. Parameterization of fracture flow structures in terms of channels

We consider that flows are highly channeled when they are focused in a very few structures at the scale of interest. This paper pertains only to these highly channeled flow cases. When flows are more evenly distributed in the medium, the more classical geostatistical based parameterizations should be used [McLaughlin and Townley, 1996; de Marsily et al., 2000; Carrera et al., 2005]. The discussion (section 5) will assess quantitatively the zones of relevance of these two different approaches by using the channeling indicators previously developed in Le Goc et al. [Le Goc et al., 2010b].

2.1. Hierarchical organization of channels

In fractured media, flows are highly channeled in a small number of fractures below the homogenization scale (REV). The homogenization scale is roughly the distance between two main channels [Long et al., 1982]. Below this scale, flows are organized hierarchically in channels of decreasing importance [de Dreuzy et al., 2001a; b]. Flows are focused in preferential flow channels, or "paths of least resistance" [Tsang and Tsang, 1989]. Figure 5-1 gives an example of channeled flows obtained on a synthetic 2D fracture network. Among all generated fractures (Figure 5-1a), only a few of them carry a non-negligible amount of flow (Figure 5-1b) and an even smaller number carry 70% of the total flow (Figure 5-1c). Channels are visually identified and organized by decreasing order of importance on Figure 5-1d with a dominant channel in purple carrying 25% of flow, a secondary channel in red carrying 15% of the flow and higher-order structures carrying from 5% to 15% of the flow. The highly simplified network of Figure 5-1d leads to a structure of flow very similar to that of the simplified network of Figure 5-1c by increasing judiciously the segment transmissivity. All remaining flowing structures of Figure 5-1a can be integrated in the background matrix.

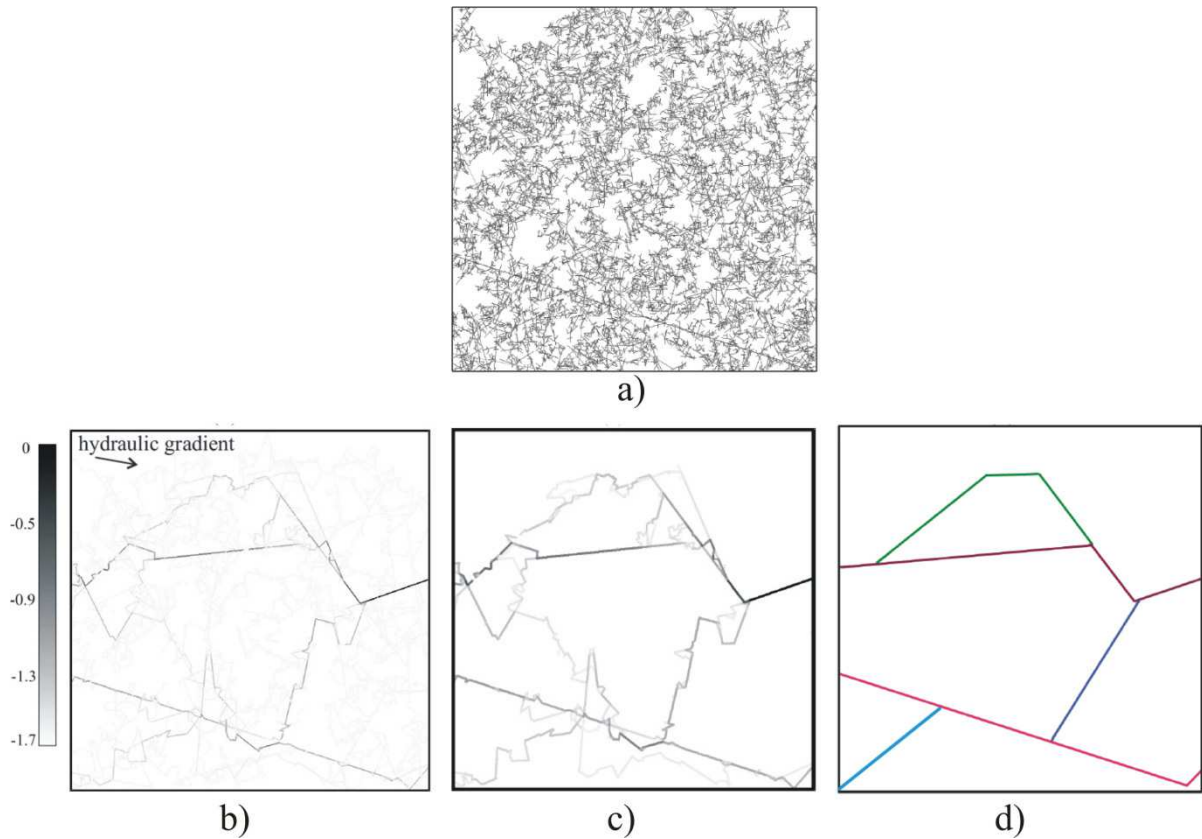


Figure 5-1 : (a) Synthetic fracture network. (b) Flows computed with a constant sub-horizontal hydraulic gradient of direction given by the arrow. All fractures have the same transmissivity. The grey color is proportional to the logarithm of flow normalized by the total flow entering the domain. (c) Simplified fracture structure carrying 70% of the flow. (d) Sketch of the main flow channels.

2.2. Quantitative definition of a channel

We define a channel as a succession of highly transmissive flow paths maintaining high flow rates over long distances. Figure 5-1d displays five channels of different complexity along which flows are at least one tenth of the total flow entering the domain. We base the channel parameterization on this definition. A channel is represented by a broken line connected either to the boundary conditions or to other channels. Each channel has a single uniform transmissivity. The full fracture network is parameterized by a restricted set of channels superimposed to a homogeneous background matrix of smaller permeability. This parameterization can be organized in three classes of parameters made up respectively of the channels geometry and topology (connectivity between channels), of the channel transmissivities and of the matrix permeability.

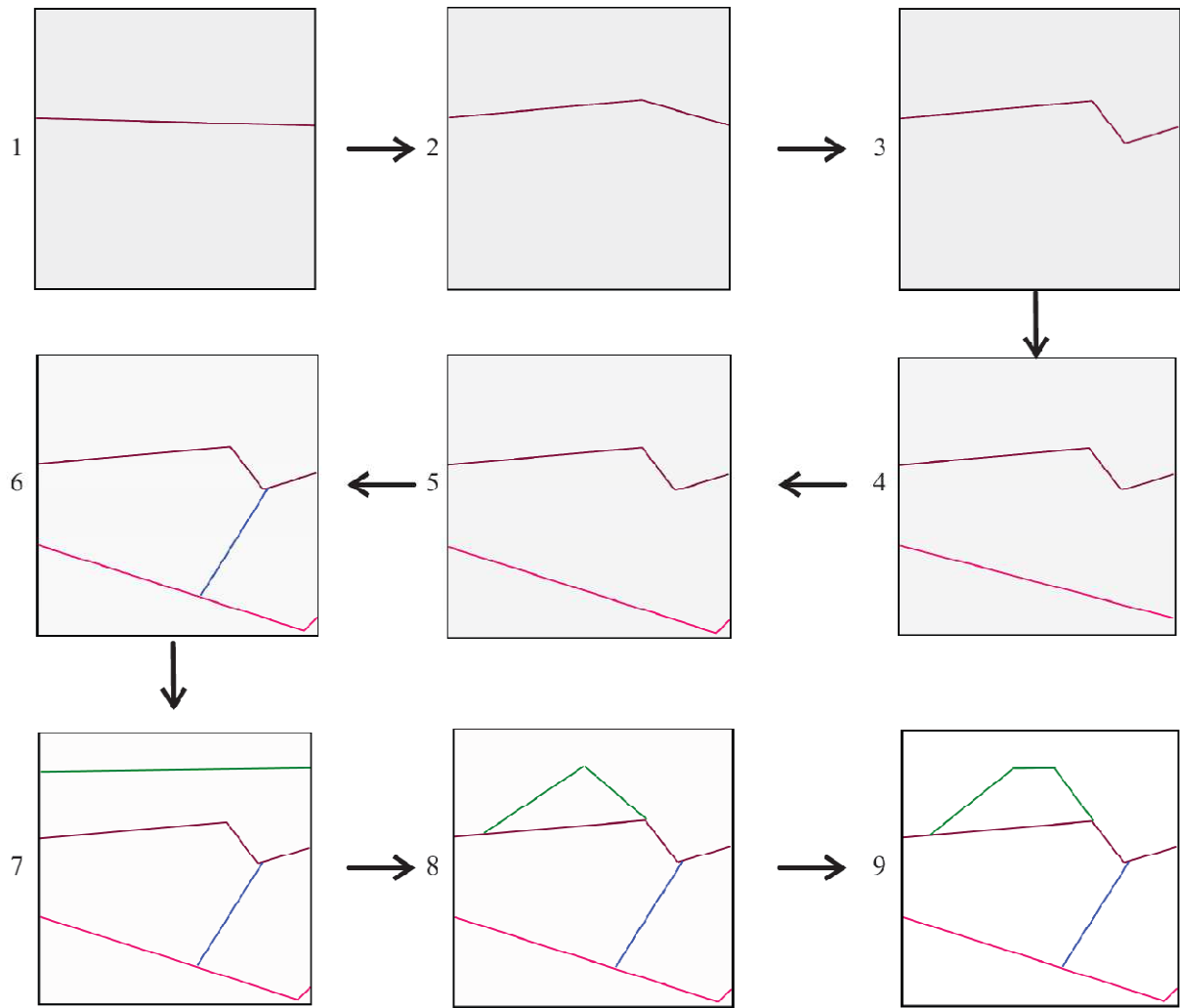


Figure 5-2 : Principle of the identification strategy applied to the flow structure of Figure 5-1d. Channels are identified by decreasing order of importance. At each step, an objective function with a restricted number of parameters is defined and optimized. Steps 1 to 3 illustrate the identification of the first channel with a shape of increasing complexity. Steps 4 to 5 illustrate the addition and refinement of a secondary channel. Step 6 shows how channel connectivity can increase with a channel connecting two previously identified channels. Steps 7 to 9 complete the identification strategy. The background color is correlated to the background permeability. It is expected to decrease with the number of structures identified.

2.3. Iterative identification of channels

The identification strategy relies on the hierarchical organization of the flow structure. It consists in an iterative identification of the channels by decreasing order of importance. Figure 5-2 illustrates this strategy on the example of Figure 5-1. At each step, we solve an inverse problem on a single channel while modifying only marginally the already identified ones (e.g. Figure 5-2, steps 1 to 3). We thus restrict the complexity of each inverse problem by reducing strictly the parameter space. Variations of the previously identified parameters are highly restricted by the addition of a regularization term in the objective function. Only the newly added parameters can freely vary in their full possible range. Because of the hierarchical organization of the channels, our strategy is to refine the shape of the most important channels before introducing less important additional channels. This is illustrated on the simplified flow structure of Figure 5-1d by an idealized succession of identification

steps (Figure 5-2). The channel refinement is performed by increasing the number of its composing segments (Figure 5-2, steps 2, 3, 5, 8 and 9). Once a channel has been identified, we add a new channel in the identification process (Figure 5-2, steps 1, 4, 6, 7). Each added channel increases the connectivity between the system limits (Figure 5-2, steps 1, 4, 6) or between the channels (Figure 5-2, step 7). The strategy is to identify first the main channel (Figure 5-2, steps 1 to 3) and then the other channels by decreasing order of importance (Figure 5-2, steps 4 to 9). One of the critical point of the algorithm is the criterion used to shift from the refinement of an existing channel to the addition of a new channel. This will be detailed in the following section.

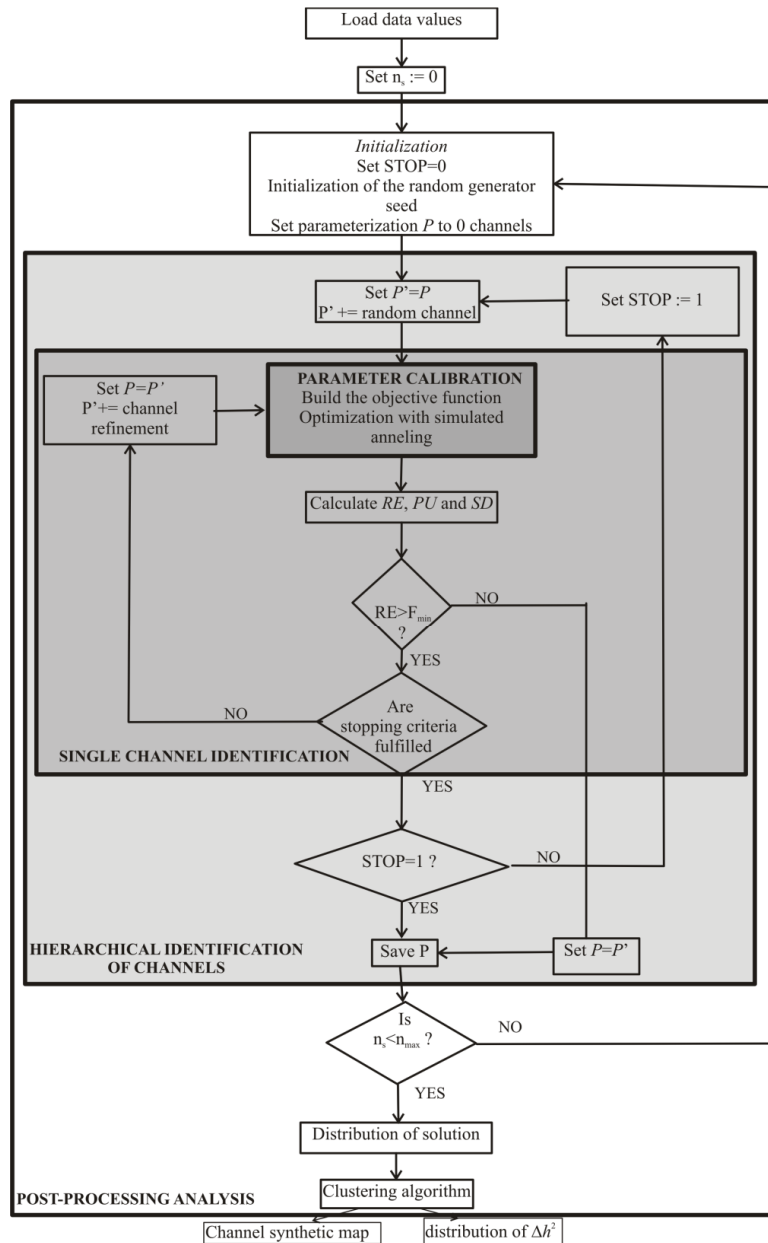


Figure 5-3 : Flowchart of the flow channels identification strategy consisting of four nested loops underlined by different grey colors. The four loops are the calibration loop, the single channel identification loop, the hierarchical channels identification loop and the post-processing analysis loop. P is the last parameterization considered as valid, P' is the current parameterization, n_s is the simulation number, n_{max} is the number of simulations to perform and stopping criteria are given by (4) and (5).

3. Inverse problem methodology and implementation

The algorithm is composed of four nested loops (Figure 5-3). As will be shown in section 3.3, the objective function is highly non-convex and possesses numerous local minima. We thus use a stochastic optimization yielding a distribution of sub-optimal solutions rather than a unique optimal solution. The most external loop drives a complete identification process that ends up to the final solution. This solution is derived from all sub-optimal solutions obtained by the identification strategy. This strategy, directly derived from section 2.3, is made up of three loops. Its first loop is the hierarchical identification of the channels. It adds and identifies a new channel at each iteration. It calls the single channel identification loop that refines the structure of the channel by adding new segments to the broken line. The most internal parameter calibration loop identifies the characteristics of the segment. It is in this loop that occurs the classical optimization algorithm with the definition of the objective function and its minimization according to the segment parameters. Once the distribution of solutions has been built up, we look for the similarities and differences of the solutions with a clustering algorithm. In this section, we present successively these four loops with two additional sections dedicated to the definition of the objective function and to the method used for solving the direct problem.

3.1. Hierarchical identification of channels

We first describe the “hierarchical identification of channels” loop in Figure 5-3. The principle of this loop is to add new channels as long as they yield a better agreement between model and data. Each step of the loop consists in initializing a new channel, optimizing its structure parameters from the single channel identification loop and, after a stopping criterion, analyzing the benefits of its addition. The new channel is initialized by drawing its characteristics from uniform distributions of its parameters taken between the extreme values given in Table 1. The positions of its extremities are drawn uniformly on the already existing channels and boundaries. Its \log_{10} -transmissivity is chosen at least two orders of magnitude greater than the background permeability in a uniform distribution in the interval $[\min(\log_{10}T_{\text{mat}})+2, \max(\log_{10}T_{\text{mat}})+2]$, where $\min(\log_{10}T_{\text{mat}})$ and $\max(\log_{10}T_{\text{mat}})$ are the limit values allowed for decimal logarithm of the background matrix transmissivity. The improvement of the model by the tentative addition of the channel is assessed by two criteria. The first criterion is given by a measure of the agreement between model and data. The measure is the classical quadratic mismatch sum RE between reference data d and modeling results $d'(n_c)$ obtained with the n_c number of channels of the current model:

$$RE(n_c) = \sum_{i=1}^{N_w} \left(\frac{d_i - d'_i(n_c)}{\sigma_i^d} \right)^2 \quad (1)$$

where N_w is the number of data observations, d_i and d'_i are the i^{th} observed and simulated data and σ_i^d is the error associated to the i^{th} observed data. The second criterion is given by the comparison of the parameter uncertainty PU to the structure discrimination SD . PU is based on the covariance matrix of the estimated parameters defined by Yeh and Yoon [Yeh and Yoon, 1981] and is defined for a parameterization with n_c channels and $NP(n_c)$ parameters as [Yeh and Yoon, 1981; Tsai et al., 2003a]:

Parameter	Definition	Range of possible values	Value chosen for the method
T_{mat}	Transmissivity of the background matrix	10^{-5} to 10^0	Initialized at 10^{-3}
T_{ch}	Transmissivity of channels	10^{-3} to 10^2	Drawn in a uniform distribution
$B1_{ch}, B2_{ch}$	Borders or channels connected by the current channel	0 to n_b+n_c where n_c is the number of channels and n_b the number of system borders	Drawn in a uniform distribution
$C1_{ch}, C2_{ch}$	Position of the channel's extremity	0 to the border/channel length	Drawn in a uniform distribution
$(x,y)_{ch}$	Position of points defining the channel shape	Inside the system	Drawn in a uniform distribution
σ_i^h	Weight associated to the i^{th} head value (6)	10^{-3} to ∞	1 or equal to the difference between the channeled and the homogeneous configuration
σ_i^d	Weight associated to the i^{th} distance value (6)	>0	1
σ_i^k	Weight associated to the i^{th} parameter (6)	>0	1
α_1	Stopping criterion associated to RE (4)	$]0;1[$	1
α_2	Stopping criterion associated to PU and SD (5)	≥ 1	2
$\lambda_{j,k}$	Weighting term associated to previously identified parameters	>0	$n_{cur}(k) - n_k$ where n_{cur} is the current step and n_k the step where the k_{th} parameter were introduced.
ξ_i	Weighting term associated to channels during the post-processing	>0	F_{obj}/S_i with F_{obj} the objective function value corresponding to the channel and S_i its sensitivity.
It_{max}	Allowed number of iteration of the simulated annealing	≥ 1	$5 \cdot 10^4$
F_{obj}^{min}	Aimed value of the objective function	≥ 0	10^{-4}
T_{init}	Initial temperature of the simulated annealing	>0	$10^0, 10^1$ or 10^2
n_s	Number of single solutions used in the post-treatment	≥ 1	50
d_c	Threshold of the clustering algorithm	>0	Between 0.5 and 3.5 as a function of the configuration

Table 1 : Parameter list and possible values for the whole identification strategy.

$$PU(n_c) = \left\| \left(\frac{N_w - NP(n_c)}{RE(n_c)^2} (J_d^T \cdot J_d) \right)^{-1} \right\| \quad (2)$$

where J_d is the Jacobian matrix of RE .

SD is used as a criterion to discriminate two parameterizations and is based on the distance between the two corresponding solutions. It is given by *Tsai et al.* [*Tsai et al.*, 2003a]:

$$SD(n_c + 1) = \sqrt{\sum_{i=1}^{N_w} \left(\frac{d'_i(n_c + 1) - d'_i(n_c)}{N_w} \right)^2} \quad (3)$$

The addition of the channel " n_c+1 " significantly contributes to the improvement of the model provided that:

$$\frac{RE(n_c + 1)}{RE(n_c)} < \alpha_1 \quad (4)$$

with $\alpha_1 \leq 1$ and that:

$$\frac{PU(n_c + 1)}{SD(n_c + 1)} < \alpha_2 \quad (5)$$

The first criterion means that the agreement between model and data is improved. The second criterion ensures that the two models with n_c and n_c+1 channels can be discriminated according to the uncertainty of the parameters. It avoids over-parameterization, which leads to a higher parameter uncertainty [*Tsai et al.*, 2003a; *Hill*, 2006]. These criteria have been used for the identification of zones in heterogeneous continuous media with $\alpha_1=1$ and $\alpha_2=2$ [*Tsai et al.*, 2003a; *Chang et al.*, 2005]. We take these same simple values in this study.

3.2. Single channel identification

All channels are identified iteratively. At first, each channel is a single segment initialized in the hierarchical identification of channels loop. Its parameters are the position of its extremities and its transmissivity. They are optimized by the parameter calibration loop. Once optimized, the segment is the starting element of a broken line. At each further step, a new node is added to the broken line. Its position as well as the transmissivity of the broken line is optimized. The stopping criteria of the refinement are identical to those of the hierarchical identification of channels loop ((4) and (5)). The full identification procedure stops when both the improvement of the last channel and the addition of a new channel successively fail.

3.3. Definition of the objective function

The determination of the broken line parameters is performed by the parameter calibration loop based on the optimization of an objective function made up of two terms. The first term is the mismatch between modeled and reference head data RE (1). The second term is a regularization term for the current and previously identified channels [*Carrera et al.*, 2005]. The objective function writes:

$$F_{\text{obj}}(n_c, NP(n_c)) = \sum_{i=1}^{N_w} \left(\frac{d_i - d'_i(n_c)}{\sigma_i^d} \right)^2 + \sum_{j=1}^{n_c-1} \sum_{k=1}^{NP(j)} \lambda_{j,k} \cdot \left(\frac{p_k(j) - p'_k(j)}{\sigma_k^p} \right)^2 \quad (6)$$

where p_k is the k^{th} parameter introduced in the model with n_c channels, p'_k is its a priori value derived from previous steps, σ_k^p is the associated error, $NP(j)$ is the number of parameters describing the j^{th} channel, $NP(n_c)$ is the sum of $NP(j)$ for the n_c first channels and $\lambda_{j,k}$ is the strength of the regularization of the k^{th} parameter of the j^{th} channel.

The regularization term (second term) restricts the possible variations of the parameters of the n_c-1 previously identified channels. This strategy is not classical but critical in our identification strategy. The objective is to allow some degree of freedom to the previously identified parameters. In fact, previously identified channels may be impacted by ulterior channel additions and modifications. The regularization term enforces the former parameters to remain close to their previously identified values. Let note n_{cur} the current step index and n_k the step index where the parameter k was introduced. The weighting coefficient $\lambda_{j,k}$ is a function linearly increasing with the number of steps between n_{cur} and n_k :

$$\lambda_{j,k} = n_{\text{cur}}(k) - n_k \quad (7)$$

Increasing the value of $\lambda_{j,k}$ restricts the possible modifications of the already identified corresponding parameter. It means that the lately identified structures of less importance have a decreasing impact on the previously identified structures.

3.4. Parameter calibration

The single channel identification loop issues at each iteration the optimization of the objective function (6) with a restricted number of parameters. In the case of the addition of a new channel, the number of parameters to identify is 6: the position of the extremities of the channel (2 parameters by extremity), the log-transmissivity of the channel and the transmissivity of the background matrix. In the case of the modification of a channel, the number of parameter to identify is 4: the position of the new broken line node (2 parameters), the log-transmissivity of the channel and the log-transmissivity of the background matrix. The background “matrix” transmissivity is systematically calibrated as it is expected to decrease through the identification process. We add that we have found very early in this study that the choice of log-transmissivity leads to much better results than transmissivity. This is mainly due to the variation range of transmissivities over several orders of magnitudes while geometrical parameters have a narrower variation range.

The calibration algorithm has been chosen according to the properties of the objective function. Figure 5-4 shows that the objective function is not convex and displays numerous local minima. We thus choose a stochastic inversion algorithm [Tarantola, 2004] among which we selected simulated annealing as the simplest and most commonly used in groundwater calibration problems [Marryote et al., 1993; Zheng and Wang, 1996; Day-Lewis et al., 2000; Nakao et al., 2000]. We focus in this paper on the parameterization rather than on the optimization algorithm. That is why we took the most classical stochastic algorithm. We give in Appendix A the chosen parameterization of simulated annealing. The algorithm is

stopped either when it reaches a maximum number of iterations it_{\max} or when the objective function falls below a minimum value F_{obj}^{\min} .

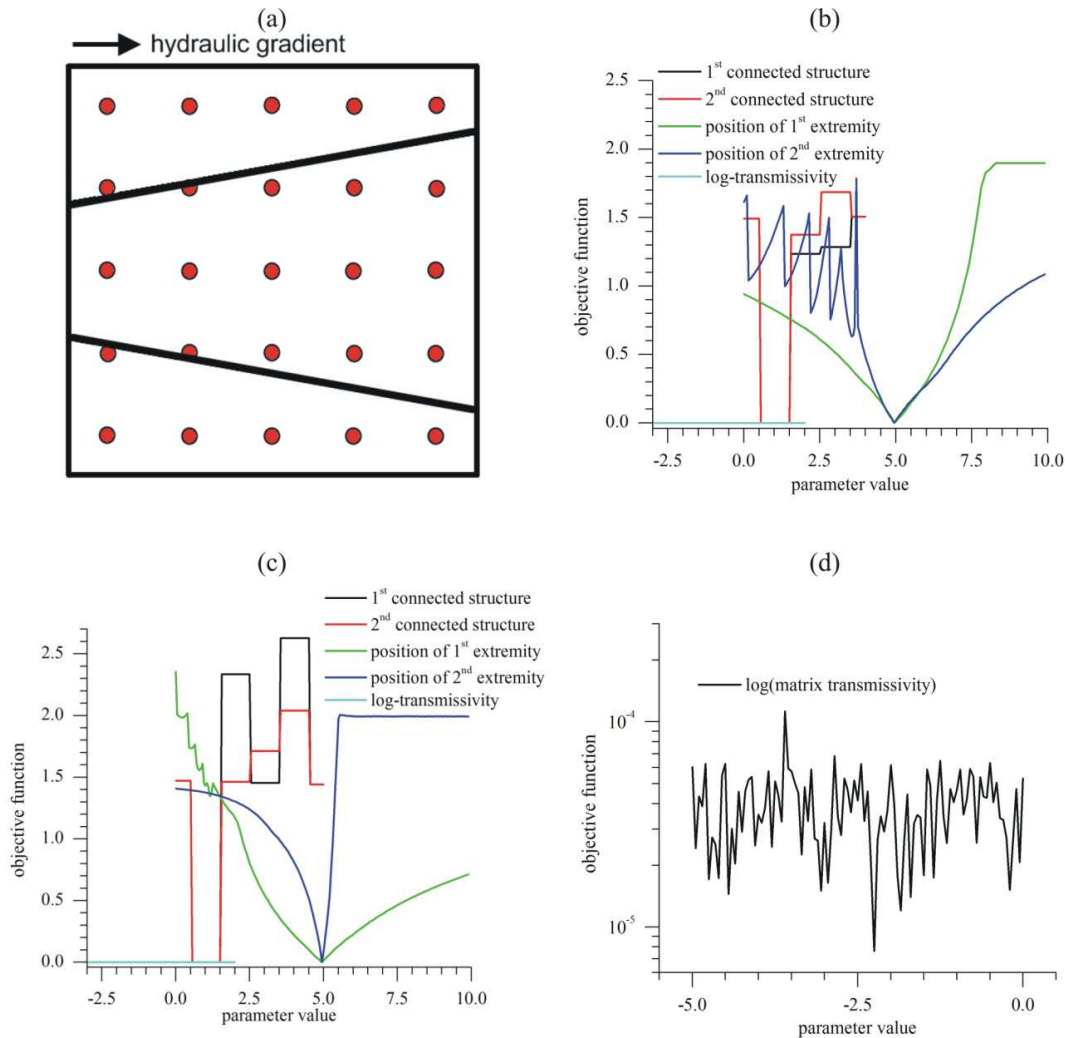


Figure 5-4 : objective function versus model parameters on a simple case made up of two channels. (a) Configuration used to compute the objective function. Red dots stand for observation locations. Objective function values (6) with $\mu_1=0$ and $\mu_2=1$ as a function of (b) the first channel parameter, (c) the second channel parameters and (d) the matrix permeability. All parameters are varied around their optimal value, all others being left fixed to their optimal value. 1st and 2nd connected structures refer to the borders or channels connected by the channel (between 0 and 4). Position of 1st and 2nd extremity refers to the coordinate of the intersection between the channel and the border or channel (between 0 and 10). The log-transmissivity of channels varies from -3 to 2 and the log-transmissivity of the background matrix varies from -5 to 0.

3.5. Numerical method for solving the direct problem

The direct problem consists in solving the flow equation in well-connected channels embedded in a uniform homogeneous medium. We set up a fast flow simulation method by discretizing the matrix so that the discretization nodes of the matrix include the wells where reference data are made available. Practically, the discretization grid is regular with a characteristic scale equal to the inter-well distance. The discretized matrix superimposed on the channels issues a network of 1D segments in which we solve the classical flow equation:

$$\nabla \cdot (T\nabla h) = 0 \quad (8)$$

where T is the segment transmissivity. Boundary conditions are described in section 4.1. The discretized equation leads to a linear system solved by the multifrontal method implemented in the software UMFPAK [Davis, 2004].

3.6. Post-processing analysis of the solution distribution

Both because of the existence of local minima and because of the use of a stochastic inversion method, the solution does not necessarily match an absolute minimum of the objective function. When running the identification strategy several times with different initial conditions, each solution depends on the initial conditions of the algorithm represented here by the position of the newly added channels and points within the channels and on the modifications of configurations made at each step of the simulated annealing algorithm. For a given problem, we obtain a distribution of sub-optimal solutions. We build up the distribution of solutions in the most external loop of the inversion scheme (Figure 5-3). We set up a methodology to analyze the similarities and differences of the simulations based on hierarchical clustering algorithms [Gordon, 1999b; Chelsea et al., 2005]. It consists in finding the most common channels and in quantifying their probability of occurrence (Figure 5-5).

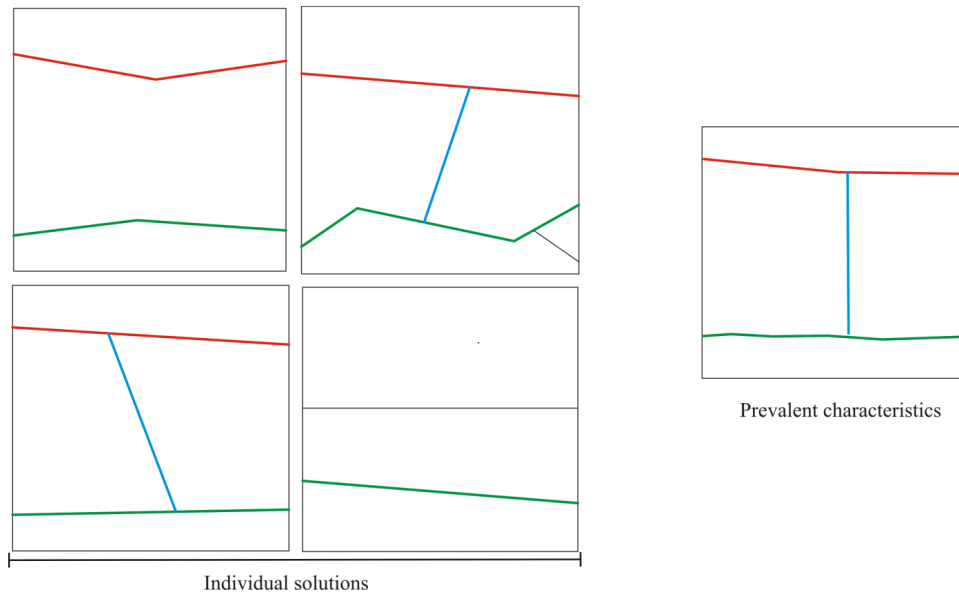


Figure 5-5 : Sketch of the post-processing clustering algorithm identifying the prevalent characteristics (right) within a set of individual solutions (left). Colors on the individual solutions refer to the channels identified as prevalent characteristics.

The degree of similarity between two channels i and j is given by the distance between them:

$$d_{ij} = \left[\int_{\Delta_{ij}} \frac{(y^i(x) - y^j(x))^2 dx}{\Delta_{ij}^2} \right]^{\frac{1}{2}} \quad (9)$$

where i and j are the indices of the two channels, $y^i(x)$ and $y^j(x)$ characterize the locations of the channels and Δ_{ij} is the overlapping x -range between the channels. The clustering algorithm gathers the channels in meta-channels by minimizing the distances between their elements. It

is controlled by the maximum distance d_c between two channels in a meta-channel. Large d_c values reduce the number of meta-channels but may lower their internal consistency. On the opposite, small d_c values yield larger numbers of meta-channels with higher internal consistency. After having investigated the clustering results for several different configurations, we fix d_c between 5% and 35% of the characteristic domain scale L .

Each meta-channel is then characterized by its mean position and transmissivity and the uncertainty of its parameters. The mean location of a meta-channel is derived from the position of its components weighted by the coefficients ξ_i given by:

$$\xi_i = \frac{S_i}{F_{\text{obj}}^i} \quad (10)$$

where F_{obj}^i is the value of the objective function for the solution containing the channel i and S_i is the weight of the channel within its solution. The transmissivity of the meta-channel is chosen as the geometric mean weighted by the coefficients ξ_i of its components. Larger weights ξ_i are obtained for smaller objective function values and for channels that are more sensitive. S_i is more difficult to obtain than the final value of the objective function. We have tried several possibilities and end up with a not so simple but intuitive definition. First, we rank the channels within a solution by their relative weight. The weight of the channel i (S_i) within its solution is given by the difference of RE values (1) between the whole solution and the whole solution to which has been removed the channel i . Secondly, from a system void of channels, we add progressively the channels within the solution by their decreasing weight as defined just before. We define S_i as the difference of RE values before and after the addition of the channel i . With this method, the channel with the largest S_i is likely to be the main channel. The uncertainty on the meta-channel position is obtained as the standard deviation of its components weighted by the coefficients ξ_i .

3.7. Synthesis of parameters of the identification strategy

The identification strategy defined in the previous paragraph is controlled by several parameters (Table 1). First, the initial conditions, i.e. the initial parameterization and the initialization of the pseudo-random number generator, condition the results of the simulated annealing algorithm. Second, the stopping criteria of the single channel identification loop, i.e. the maximal number of iteration it_{max} and the expected objective function value $F_{\text{obj}}^{\text{min}}$, control the balance between the optimization duration and accuracy. The refinement level of the final parameterization depends on the stopping criteria of the hierarchical identification loop α_1 (4) and α_2 (5). The optimization algorithm itself is controlled by the initial temperature, the temperature schedule and the size of the searching neighborhood (Appendix A, (17)). The shape of the objective function is a critical factor that depends on the weighting coefficients $\lambda_{j,k}$ of the regularization term in (6). The final clustering algorithm based on the analysis of single solutions is controlled by the number of solutions n_s used in the analysis, by the weighting coefficient ξ_i and by the stopping threshold d_c . Parameter values have been assigned either as their most common values used in other studies or with tests in simple configurations containing one or two straight channels.

4. Results

We apply the identification strategy to a set of synthetic flow structures of increasing complexity. On a bi-processor quad core Intel® Xeon at 2.33GHz, the time required to get a single solution varies from 20 minutes to 1 hour depending on the number of parameters of the final model. Thus, 50 solutions for 20 configurations required around 28 days of cumulated computation. This time can be reduced by using parallel computations. It would be lowered to 3.5 days by distributing computations over 8 CPUs.

4.1. Tested configurations

We select 20 synthetic channeled configurations by increasing order of complexity. Other parameters than those of the channel structures are kept simple. We derive the flow configurations from fracture networks embedded in a square domain. The fracture flow patterns are dominated by large fractures and are slightly above percolation threshold. The boundary conditions are derived from a fixed regional gradient not aligned with the edge of the system. The head gradient is fixed at 1. We retain only the largest channels identified as the shortest paths between the limits of the domain. Three typical configurations of increasing complexity are illustrated on Figure 5-6.

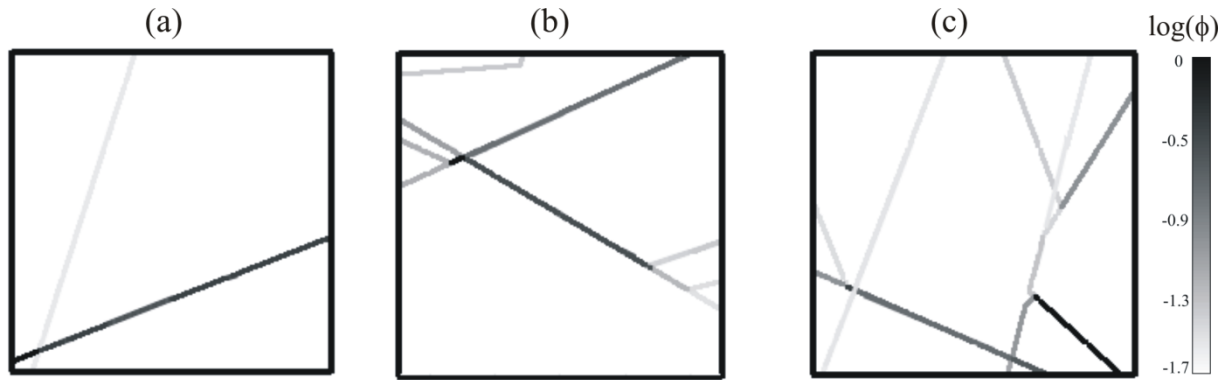


Figure 5-6 : Flow rates ϕ of three tested configurations, ordered by increasing complexity. The simplest configuration (a) is composed of two straight-line channels. The intermediary configuration (b) is composed of two channels that are more complex. The complex configuration (c) has several second-order well-connected structures.

4.2. Data

In this study, we use elementary data consisting in steady-state well heads h and in the distances r of the nearest channel from the wells. We assume that these distances can be provided either by well tests [Chang *et al.*, 2008] or by geophysical measurements like GPR [Grasmueck, 1996; Tsoflias *et al.*, 2004]. Introducing explicitly these two types of data in the objective function (6) leads to:

$$\begin{aligned}
 F_{\text{obj}}(n_c, NP(n_c)) &= \sum_{i=1}^{N_w} \left[\left(\frac{h_i - h'_i(n_c)}{\sigma_i^h} \right)^2 + \left(\frac{r_i - r'_i(n_c)}{\sigma_i^r} \right)^2 \right] \\
 &+ \sum_{j=1}^{n_c-1} \sum_{k=1}^{NP(j)} \lambda_{j,k} \cdot \left(\frac{p_k(j) - p'_k(j)}{\sigma_k^p} \right)^2
 \end{aligned} \tag{11}$$

where h stands for reference heads, h' for modeled heads, r for reference distances and r' for modeled distances. σ_i^h and σ_i^r are the uncertainties linked to head and distance data. The choice of σ_i^h and σ_i^r balance the relative importance of head and distance data. More generally, our interest is to determine the relative importance of hydraulic and geometrical data. We have used four different sets of values for σ_i^h and σ_i^r . The two first ones consist in using either only head data ($\sigma_i^h=1$ and $\sigma_i^r = \infty$) or only distance data ($\sigma_i^h = \infty$ and $\sigma_i^r = 1$). The third one uses both head and “head deviation” data and no distance data ($\sigma_i^r = \infty$). The principle is to use the head uncertainty σ_i^h to increase the weight of the heads highly sensitive to the channel structure. We rate the sensitivity to the channel structure by the head deviation equal to the difference between the reference head and the head computed for a homogeneous case having the same boundary conditions. σ_i^h should thus be inversely proportional to this difference. As, for the cases tested hereafter, the inverse of the head deviation varies around 1, the coefficient of proportionality is kept at one. In other words, the objective with the knowledge of the head deviations is to filter out some of the influence of the boundary conditions. The fourth choice consists in using conjointly heads, head deviations and distance data. It is built up on the previous choice by adding the distance data ($\sigma_i^r = 2$). σ_i^r is thus of the same order as σ_i^h . We expect that these four choices conceal an increasing amount of information.

For the tested configurations, N_w wells are regularly distributed over the square domain according to a grid pattern. We take the inter-well distance d_w as the fundamental length unit. The domain size L is expressed as a function of N_w and d_w as $L=N_w^{1/2} \cdot d_w$. With $N_w=25$ and a typical inter-well distance of the order of 100 m the domain size would be of the order of 1 km. This is typical of a dense observation network [Audouin *et al.*, 2008]. We sample reference head data as well as distances to the nearest channel from the 25 wells. The 20 configurations will be taken as references for assessing the inversion methodology run solely with the knowledge of the 25 head and distance data.

4.3. Results for a simple configuration

We present first the results of the previous inversion strategy on the simple configuration of Figure 5-6a consisting of two straight channels. At first, we use only the hydraulic head data ($\sigma_i^h=1$, $\sigma_i^r=\infty$). Figure 5-7a to Figure 5-7d display the evolution of the objective function and of the parameter values during the optimization process represented by the number of iterations. The parameter defining the connected structures is between 0 and the number of system borders plus the number of existing channels. The location of channel extremities and nodal points ranges between 0 and the system size. The transmissivities are positive. The evolution of the objective function (Figure 5-7a) follows several staircase-like decreases typical of simulated annealing algorithms [Ingber, 1993b; Sambridge and Mosegaard, 2002]. The objective function alternates between sharp decreases and long plateaus linked to the existence of local minima. A recurrent feature is also the increase of the plateau duration in terms of iteration. This is especially apparent in the third calibration loop of Figure 5-7a. The lengthening of the plateaus comes from the decrease of the simulated annealing temperature. Lower temperatures induce stricter acceptance tests trapping the solution in sub-optimal minima, which reduces the probability of getting out of the local minima. The addition of new parameters by the single and hierarchical channel identification loops triggers successive sharp increase and decrease. As a whole, the channel addition or modification can reduce either significantly or marginally the objective function as shown respectively by the addition of the second channel and its modification (Figure 5-7a). The identification strategy stops on

a configuration of two channels, showing that the identification correctly avoids over-parameterization.

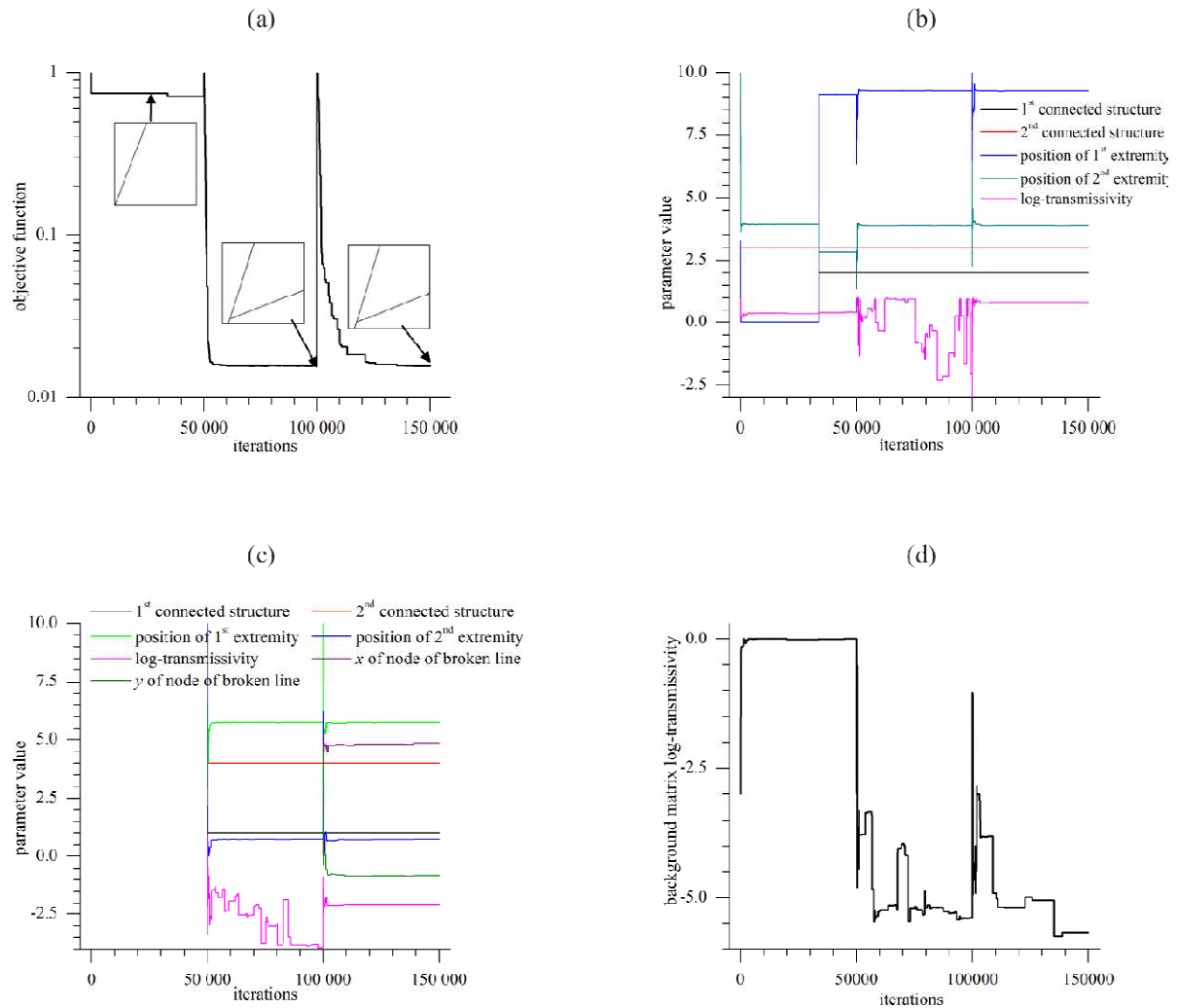


Figure 5-7 : (a) Objective function against iteration number for the simple configuration of Figure 5-6a. Evolution of the parameter values of (b) the first identified channel, (c) the second identified channel and (d) the background matrix against the number of iterations.

The channel and background log-transmissivities values fluctuate much more than the geometrical parameters, as expected, because head data are less sensitive to permeability than flow data. Moreover, transmissivities are only relative to each other, since heads are only sensitive to the difference between channels transmissivity, i.e. the absolute flow is not known. The hierarchy in channel importance is respected as the first identified channel has a transmissivity around three orders of magnitude larger than the second one. The background matrix plays efficiently its role of substitution of less important channels. It has indeed a high relative transmissivity value ($T_{mat}=10^0$) at first with only one channel and a small value ($T_{mat}<10^{-5}$) when the optimal number of channels is reached. The evolution of the parameter values shows that channel parameters are identified successively with, first, the index of the system edge on which are located the extremities, second, the location of the extremity within this edge and, third, the channel log-transmissivity (Figure 5-7b-c).

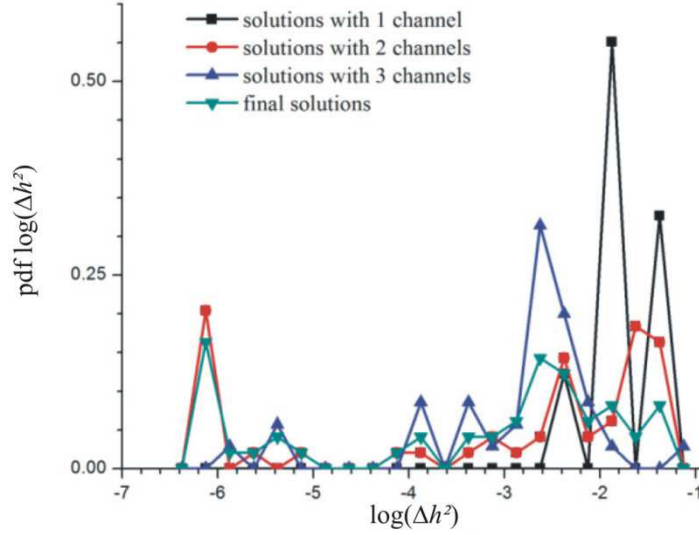


Figure 5-8 : Distribution of the square difference between reference and model results on the whole surface Δh^2 (12) derived from the 50 individual solutions for the simple configuration of Figure 5-6a.

Figure 5-8 gives the distribution of the squared difference between head data and results Δh^2 over the whole domain derived from the 50 individual solutions obtained with:

$$\Delta h^2 = \frac{1}{S} \int_S (h - h')^2 dS \quad (12)$$

where S is the system surface. Practically this integral is computed by using a 100x100-discretization. The interest of Δh^2 over RE is to give a more precise quantitative measure of the agreement between the initial structure and the model results. The distribution is given at the intermediary steps containing 1, 2 and 3 channels and at the final stage of the identification (Figure 5-8). For the intermediary step containing a single channel, the distribution of Δh^2 is very broad containing head maps both close to the objective ($\Delta h^2 < 10^{-6}$) and quite far from it ($\Delta h^2 \sim 10^{-1}$) where head values vary between 0 and 10. The addition of one and two channels significantly restricts the occurrence of large Δh^2 values. For the final solutions, most of the solution has indeed Δh^2 values smaller than 10^{-2} and an important part even have Δh^2 values smaller than 10^{-6} . The addition of a third channel does not systematically improve the solution especially for Δh^2 values smaller than 10^{-6} . Thus, most final solutions (38/50) are composed of two channels and only a few (12/50) lead to better agreements with three channels. The good agreement of the 50 solutions to the initial structure is further illustrated by the results of the clustering algorithm. Figure 5-9 shows the mean and standard deviation of the position of the identified channels. The percentage next to the channel indicates the proportion of channels within the 50 individual solutions represented by this mean channel. Channels are colored according to their relative transmissivity. We present the same synthetic map resulting from three of the four possible choices of objective functions obtained by using either only hydraulic heads ($\sigma_i^h = 1$ and $\sigma_i^r = \infty$) (Figure 5-9a), or hydraulic heads and their deviations ($\sigma_i^h = \text{“head deviation”}$ and $\sigma_i^r = \infty$) (Figure 5-9b) or hydraulic heads, their deviations and distances ($\sigma_i^h = \text{“head deviation”}$ and $\sigma_i^r = 1$) (Figure 5-9c). The influence of the modifications of the objective function is very small. For this simple case, the three tested objective functions end up to a globally good agreement of the 50

solutions to the initial structure. Both channels are well located and the ratio between their transmissivities (1.8), (Figure 5-9c) is close to the real one (1.0).

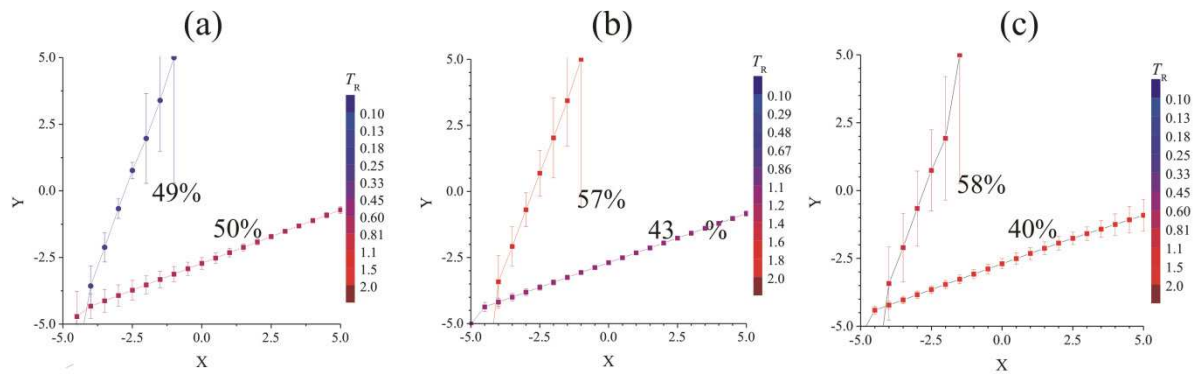


Figure 5-9: illustrations of the results of the post-processing clustering algorithm for the simple configuration of Figure 5-6a with objective functions made up of the two first terms with σ_i^h equal to 1 (a) and to the inverse of head deviation (b). The last map (c) is obtained with the full objective function made up of the three terms.

Figure 5-10 shows the head deviation σ_h and distance data interpolated from the 25 data points. We recall that the head deviation at a given point is equal to the difference between heads obtained on the reference structure and on the homogeneous case with the same boundary conditions. Fundamentally, the two maps give graphic illustrations of the information contained in the head and distance data. For such a simple configuration, both head deviation and distance data provide visually rough indications on the channel positions, explaining the relative absence of sensitivity of the results to the type of objective function.

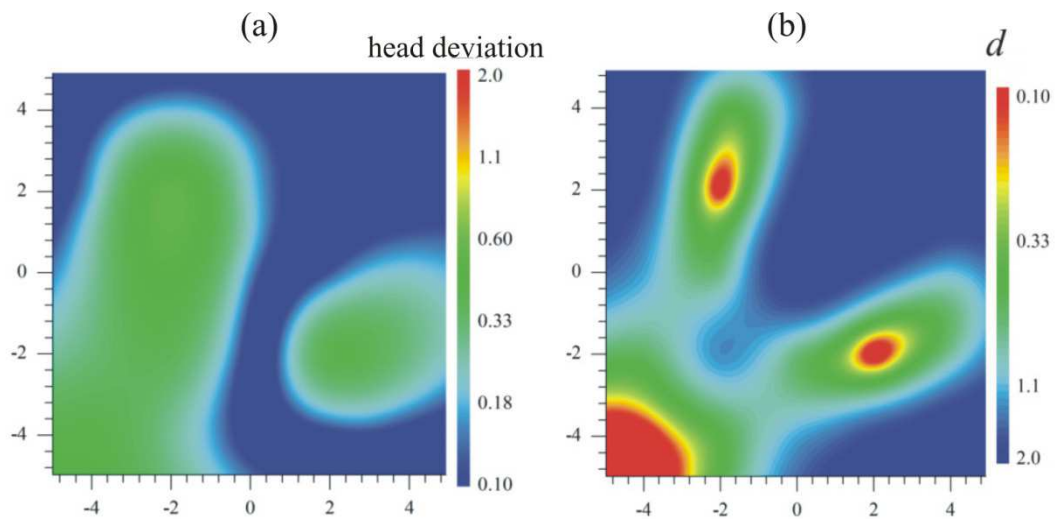


Figure 5-10 : Interpolated maps from the 25 data points (left) for the head deviation and (right) for the distance to the nearest channel for the simple configuration of Figure 5-6a.

Through this simple configuration, we have defined an interpretation framework of the results. We will use the same analysis framework for more complex configurations. We have also shown that the algorithm finds solutions close to the initial configuration with a very good agreement for the channel position but with less accuracy on the channel transmissivity. In such a simple case, the choice of the objective function is not critical. In the next sections, we test the identification strategy on more complex configurations.

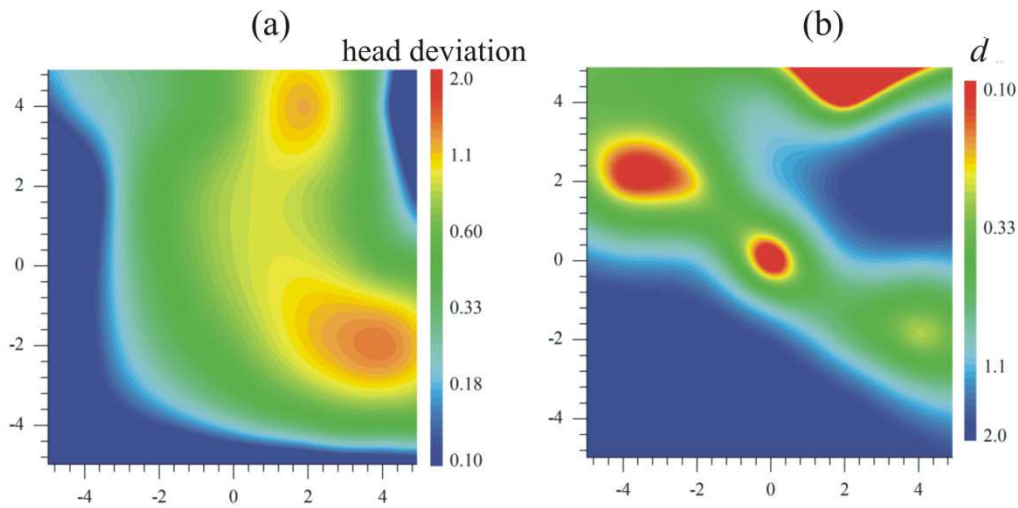


Figure 5-11: Interpolated maps from the 25 data points for the head deviation (left) and for the distance to the nearest channel (right) for the simple configuration of Figure 5-6b.

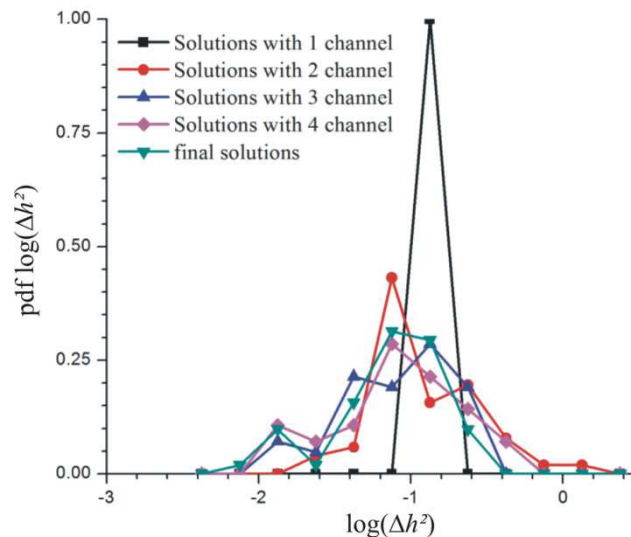


Figure 5-12: Distribution of the least-square difference between head data and model results on the whole surface Δh^2 (12) derived from the 50 individual solutions for the intermediary configuration (Figure 5-6b).

4.4. Results for an intermediary configuration

We apply the identification strategy on the more complex configuration of Figure 5-6b. Flows are organized around a main channel with two embranchments at its limits and a second interconnected channel. The interpolations of head and distance data are displayed in Figure 5-11, the distribution of single solutions, obtained with heads and distances, in Figure 5-12 and the results of the post-processing analysis are displayed in Figure 5-13. The interpolated distances to the nearest channel (Figure 5-11b) clearly reveal the main channel while the secondary channel is well marked neither on the distance nor on the head deviation interpolated maps (Figure 5-11). As for the previous simpler configuration, adding a second channel to the first one improves significantly the fit between the real and modeled heads (Figure 5-12). With three channels, most of the models yield head maps close to the reference

heads ($\Delta h^2 < 10^{-1}$). The distribution of Δh^2 is only slightly shifted to lower values by the introduction of additional channels. The map drawn from the results of the post-processing cluster algorithm shows the presence of the two main flow channels representing a bit less than 45% of all the determined channels while the remaining channels (~55%) are not in the reference configuration (Figure 5-13a). Introducing the head deviation in the objective function displaces the fictitious channel to the corner of the domain crossing a zone of low sensitivity (Figure 5-13b). It does not however lower its probability of occurrence. Only the introduction of the distance information through the third term of the objective function (6) significantly reduces the probability of occurrence of the not existing channel (Figure 5-13c). It also improves the probability of occurrence of the two main channels and gives a map much closer than for the two other cases. The improvement is linked to the large distances to the nearest channel in the bottom zone (Figure 5-11). Channels are likely to be absent from this zone. We also note that the standard deviation of the channel location is inversely correlated to the head deviation (Figure 5-13b). The two channel transmissivities are almost equal with a ratio equal to 0.9 (Figure 5-13c). They correctly match the initial value of 1.0. Four of the fifty solutions obtained with heads and distances are displayed on Figure 5-14. They all contain the most important flow channel and three of the four contain the secondary channel. The embranchments are not resolved.

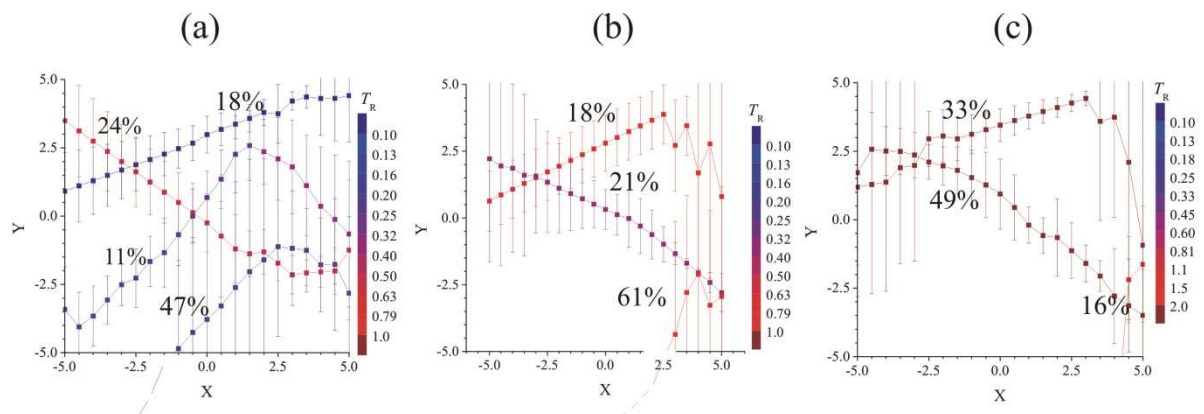


Figure 5-13: Illustrations of the results of the post-processing clustering algorithm with objective functions made up of the two first terms with σ_i^h equal to 1 (a) and to the inverse of head deviation (b). The last map (c) is obtained for the full objective function composed of the three terms.

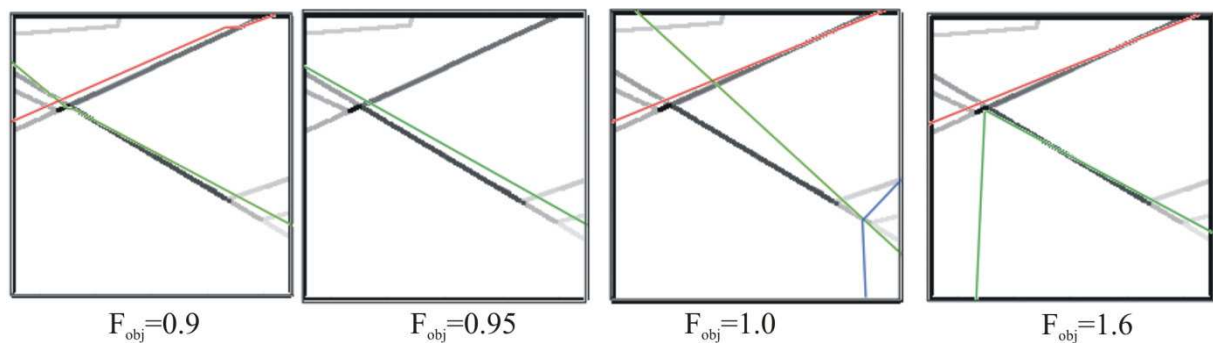


Figure 5-14: Four individual solutions of the inversion method applied to the intermediary complex configuration of Figure 5-6b. The individual solutions identify the diagonal main flow structure represented in red. Three of them identify the secondary flow structures in green. Embranchments are too small structures to be resolved. The channeled structure is recalled in the background.

On this intermediary complex configuration, the identification strategy based only on head data without head deviations identifies the main flow channels. It also yields other not-existing channels in zones poorly sensitive in terms of head. The introduction of additional information as head deviation or distances does not modify the two main channels but significantly lowers the probability of occurrence of the not-existing channels.

4.5. Results for a complex configuration

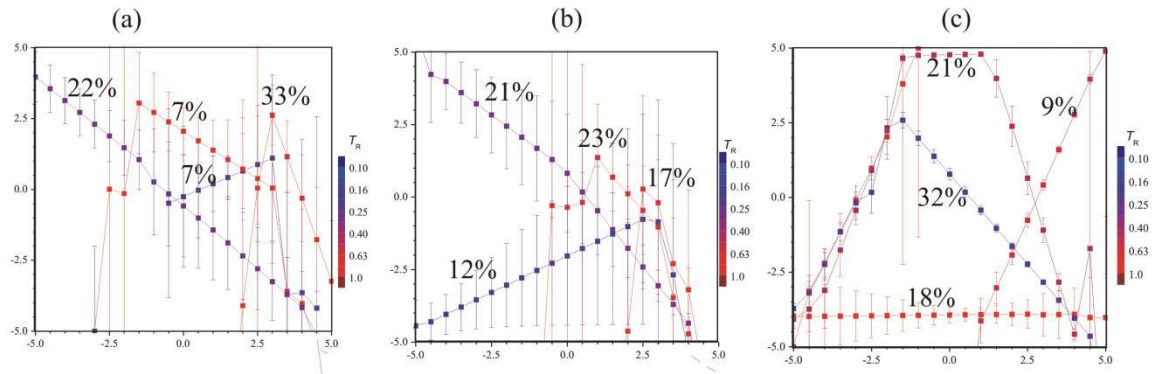


Figure 5-15: illustrations of the results of the post-processing synthetic algorithm for the configuration of higher complexity of Figure 5-6c with objective functions composed of the two first terms with σ_i^h (a) equal to 1 and (b) equal to inverse of the head deviation. The last map (c) is obtained for the full objective function composed of the three terms.

The more complex configuration displayed in Figure 5-6c is made up of several interconnected conducting segments. First-order segments are located in the bottom part of the domain. Two other second-order structures cross the whole domain from top to bottom. The optimal solutions yielded by the full inversion procedure are displayed on Figure 5-15 with increasingly complex objective functions. With head data alone, the most transmissive structures are sub-vertical and located in the right part of the domain (Figure 5-15a-b). They identify partly the right vertical crossing channel. The two diagonal structures crossing the domain from the left to the right do not exist in the reference configuration. Their importance is limited because of their relatively small transmissivity.

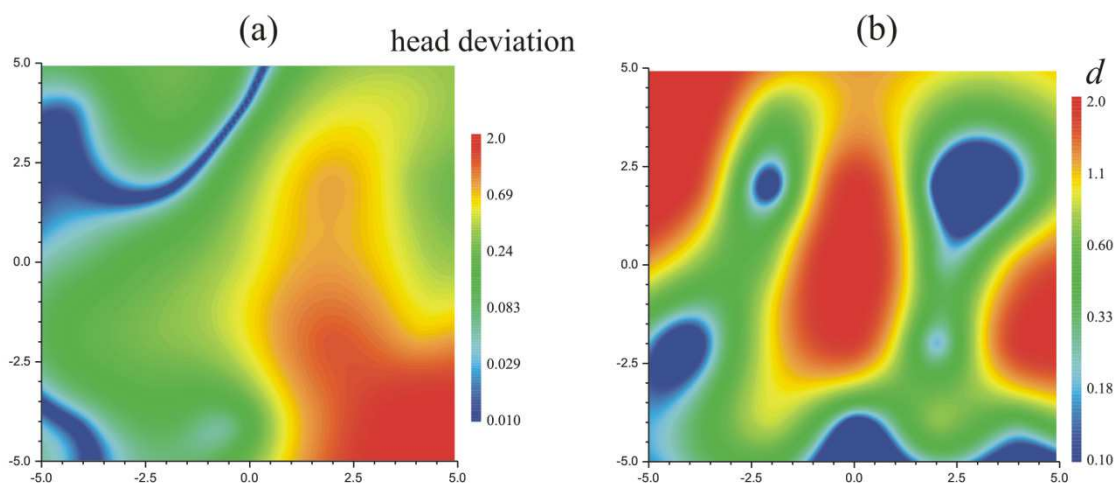


Figure 5-16: Interpolated maps from the 25 data points for the head deviation (left) and for the distance to the nearest channel (right) for the simple configuration of Figure 5-6b.

The identification is consistent with the head deviation map (Figure 5-16a) showing that head deviation is focused in the right part of the domain. With distance data alone ($\sigma_1^h=\infty$, $\sigma_1^r=1$, Figure 5-17), the optimal results of the inversion are made up of a correctly identified vertical structure on the left.

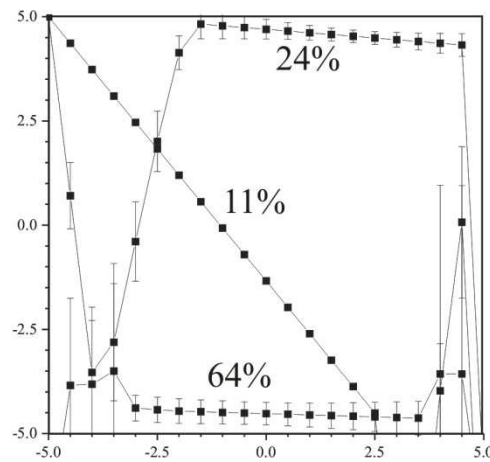


Figure 5-17 : illustrations of the results of the post-processing synthetic algorithm for the most complex configuration (Figure 5-6c) when using only distance information in the identification strategy.

The other channels connect zones of small distances but are not located along existing features. It seems to be the limit of the use of the sole distance data. The small distances indicate the presence of channels but the extraction of channels from these low distance zones is not successful with the presented hierarchical identification strategy. Combining head and distance data results in a much better identification. Most of the principal channels are resolved. Single solutions contain at most three channels including, for more than half of them, the main channels (Figure 5-18).

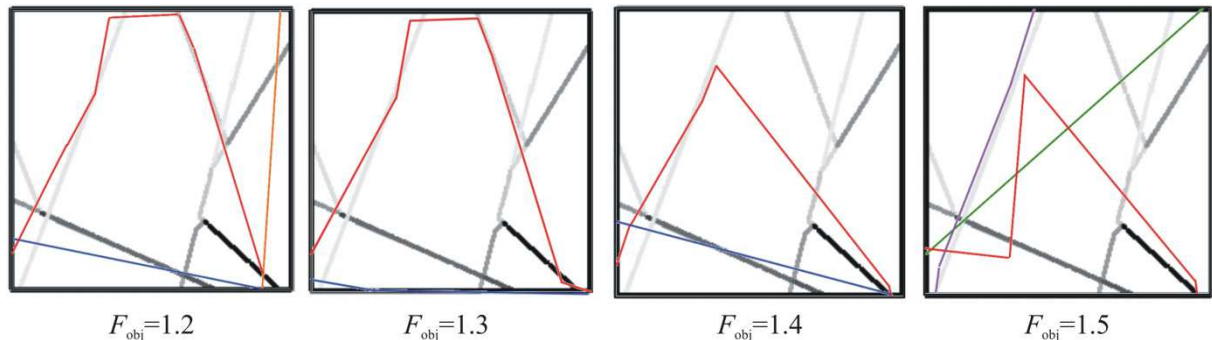


Figure 5-18 : Four individual solutions in red and their associated objective function value obtained for the identification strategy with the full objective function applied to the configuration of Figure 5-6c. The reference configuration is recalled with gray colors in the background.

From two channels, the addition of an extra channel does not improve significantly the solution (Figure 5-19). The optimal solution is much more accurate (Figure 5-15c) and only the location of the bottom channel is inaccurate. It may be due to the small distance to flow structures to the wells at the bottom of the field. The diagonal structure already observed with only distance or head data is persistent when using both head and distance data. Its importance is however limited by its transmissivity twice to four times smaller than that of the other channels. Finally, the use of additional information lowers the dispersion of channels. It

is first observed in the error bars of the identified channels like in the previous intermediary case (Figure 5-13). It is also apparent in the increase of the cumulative percentage of the identified structures. It increases from 69% in Figure 5-15a to 82% in Figure 5-15c.

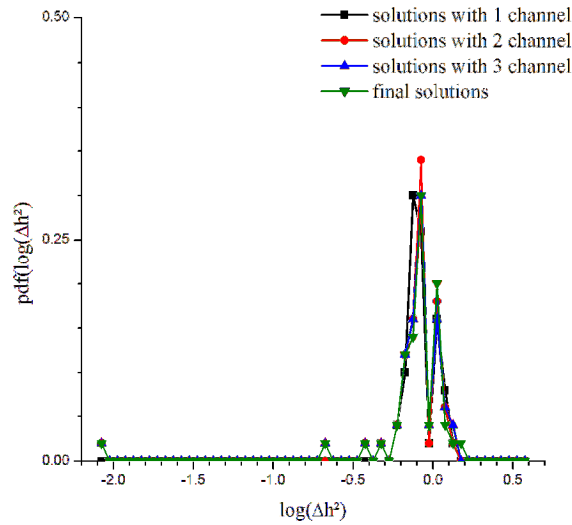


Figure 5-19: Distribution of the square difference between head data and model results on the whole surface Δh^2 (12) derived from the 50 individual solutions for the most configuration of highest complexity of Figure 5-6c.

With this complex configuration, identification can be performed only by using simultaneously head and distance data. Transmissivity appears to be a regulating factor of the importance of the different structures. Resulting structures not present in the reference configuration have a significantly lower transmissivity.

4.6. Analyses of the final models on the set of test-configurations

In this paragraph, we investigate the performances of the identification strategy on the full set of tested configurations introduced in section 4.1. We use the most advanced objective function where σ_i^h are fixed at the head deviations and $\sigma_i^r = 2$. The 20 configurations are numbered by apparent increasing degree of complexity. They are displayed on the right columns of Figure 5-20 and Figure 5-21. From C1 to C20, networks contain more intricately connected channels. Even if this visual rating is subjective, it is closely consistent with the product of the channeling indicators D_{cc} and D_{ic} introduced in *Le Goc et al.* [*Le Goc et al.*, 2010b] (Table 2). D_{cc} is a characteristic channel continuity scale taken as the distance over which flow rates remain consistently high. D_{ic} is a characteristic inter-channel distance. The degree of channeling increases with $D_{cc} \cdot D_{ic}$.

The results of the identification strategy are compared to their reference counterpart on Figure 5-20 and Figure 5-21. The configurations analyzed in the previous sections are C2, C10 and C12. The agreement of the results to the reference globally decreases with the apparent complexity measured by the number of the configuration. This is confirmed for most of the tested configurations by the increase of RE , calculated with the optimal configuration, as a function of the configuration complexity measured by $D_{cc} \cdot D_{ic} / L^2$ (black squares in Figure 5-22). Three more complex configurations (C14, C15, C20) are departing from this tendency with still low values of RE (white squares in Figure 5-22). These values of RE smaller than for those of the other configurations having similar $D_{ic} \cdot D_{cc} / L^2$ values come from the absence of the largest head discrepancies amplified by the square of the difference in RE (1).

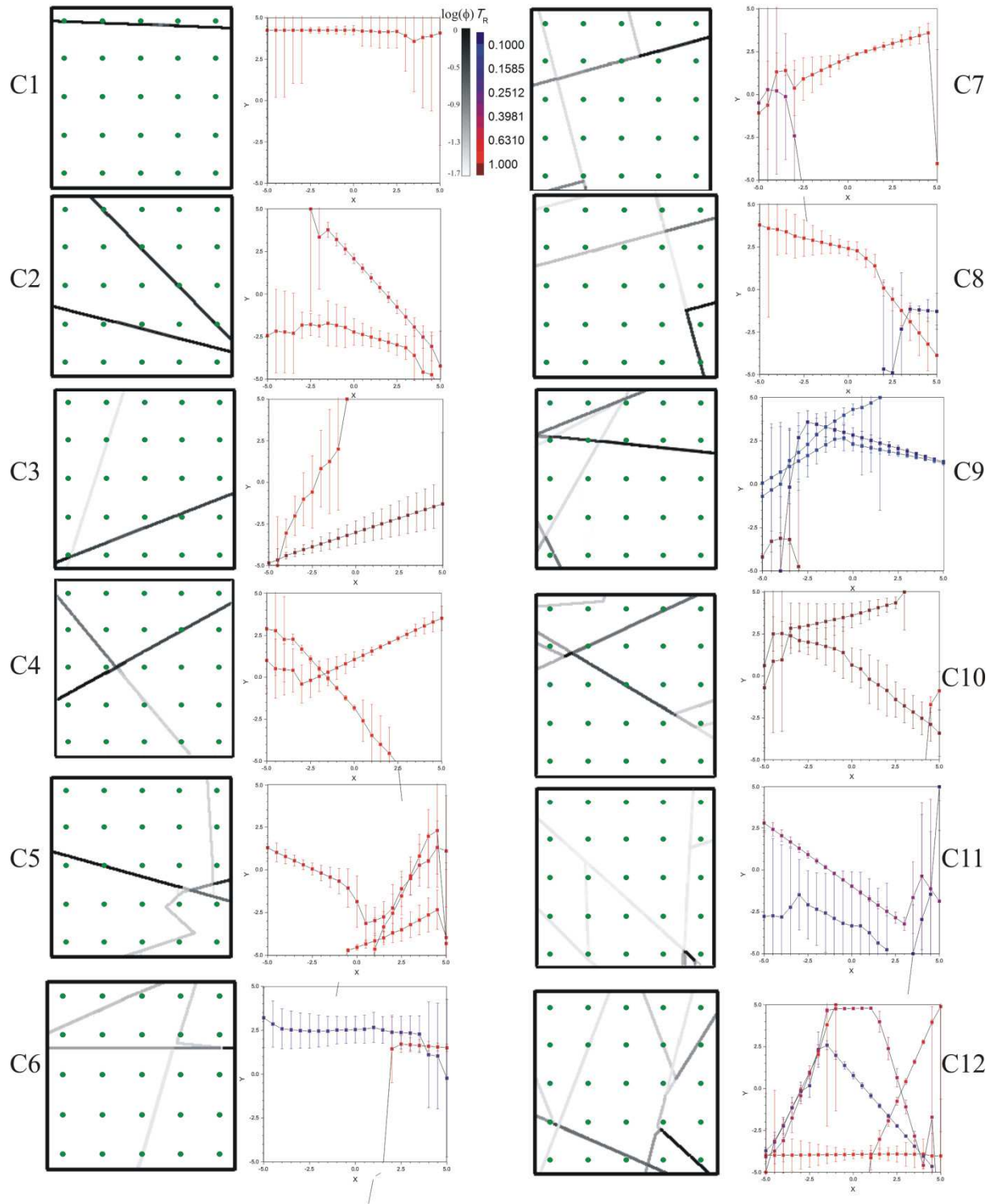


Figure 5-20 : First 12 configurations (C1 to C12) used to test the identification strategy. Left and right columns show the reference configuration and the modeling result given by the postprocessing analysis. Green dots stand for well locations.

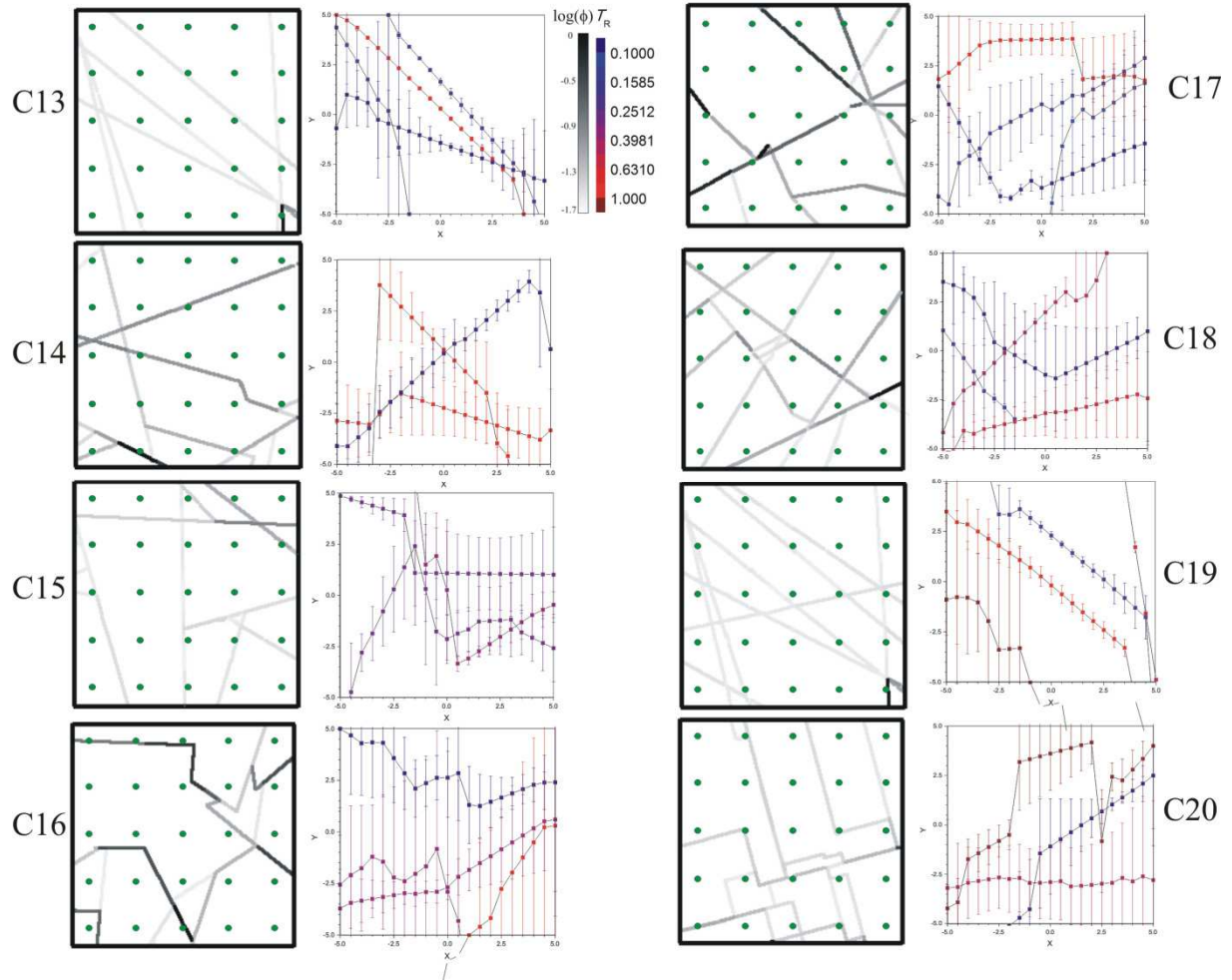


Figure 5-21 : Last 8 configurations (C13 to C20) used to test the identification strategy. Left and right columns show the reference configuration and the modeling result given by the postprocessing analysis. Green dots stand for well locations.

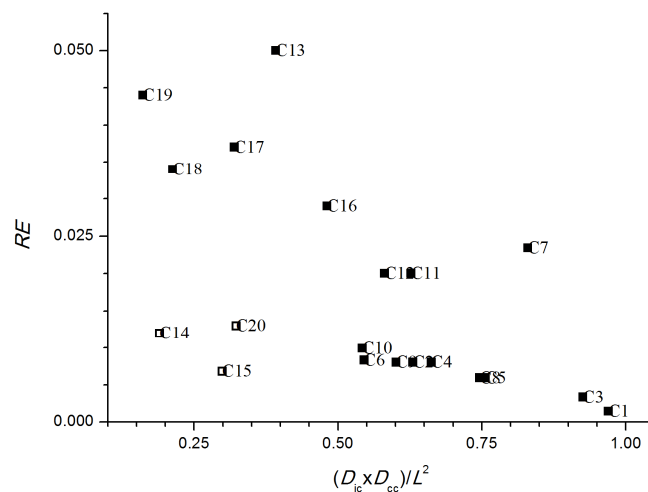


Figure 5-22 : Quadratic mismatch sum RE between reference data and modeling results as a function of the product of the channeling indicators D_{ic} and D_{cc} normalized by the square of the system size L^2 . Black squares identify configurations for which RE increases with more complex configurations corresponding to smaller $D_{ic} \cdot D_{cc} / L^2$ values. White squares stand for four complex configurations resulting in small errors RE .

Configuration	RE	D_{ic}/L	D_{cc}/L	d_c	Visual Number of channels	Visual quality of the solution
C1	0.0014	0.97	1.0	3.5	1	5
C2	0.0080	0.65	0.97	1.0	2	5
C3	0.0033	0.94	0.98	1.0	2	5
C4	0.0080	0.72	0.92	1.5	2	5
C5	0.0059	0.85	0.89	1.0	2	4.5
C6	0.0083	0.62	0.88	2.0	2.5	4
C7	0.023	0.86	0.96	1.5	2	4
C8	0.0059	0.79	0.94	1.5	2	4
C9	0.0080	0.65	0.92	0.5	2.5	3.5
C10	0.0099	0.65	0.97	0.7	2	4
C11	0.020	0.64	0.98	1.0	2.5	3.5
C12	0.020	0.75	0.78	0.5	3.5	4
C13	0.050	0.40	0.98	1.0	3.5	3
C14	0.012	0.34	0.56	0.5	4	1
C15	0.0068	0.46	0.65	1.3	4	1
C16	0.029	0.63	0.77	1.0	3	4
C17	0.037	0.42	0.76	0.4	3.5	2
C18	0.034	0.30	0.71	1.4	5	3
C19	0.044	0.26	0.62	1.0	6	1
C20	0.013	0.43	0.76	1.5	4	1.5

Table 2 : Residual errors RE defined by (1), channeling indicators (D_{ic} and D_{cc}) associated to the tested configurations (Figure 5-20 and Figure 5-21), visual indicators of the channel number and solution quality. D_{ic} and D_{cc} are in the interval [0;10]. The number of channel and the quality of the solution range between 0 and 5.

D_{ic} and D_{cc} give a global rating of the flow structure complexity. The final value of RE gives a global rate of agreement between the final solution and the reference structure. As we are also interested in the quality of the identification of the structural characteristics, we use two additional visual indicators for the number of independent channels and for the agreement of the solution to the reference structure. These indicators are clearly subjective. However, for the simple structures handled here, we assume that the variability of their estimation remains limited. The number of independent channels ranges from 1 for C1 to 6 for C19. The visual quality of the agreement between reference and resulting structures increases from 1 to 5 (Table 2). As previously noticed, the increase in channel number and thus in the flow structure complexity induces on average a decrease of the solution quality (Figure 5-23). We identify two different groups. The first group is made up of the well-identified structures characterized by a visual rating greater or equal to 3. Apart from C17 and C18, it also corresponds to the networks made up of less than 4 independent channels. The other configurations including C14, C15, C17, C20 have more than 4 independent channels and lead to visual rating smaller or equal to 2. We finally note that the three complex configurations C14, C15 and C20 having small RE value are among the worse ones in terms of visual rating. The mismatch between resulting and reference heads thus gives only a very crude estimate of the inversion quality.

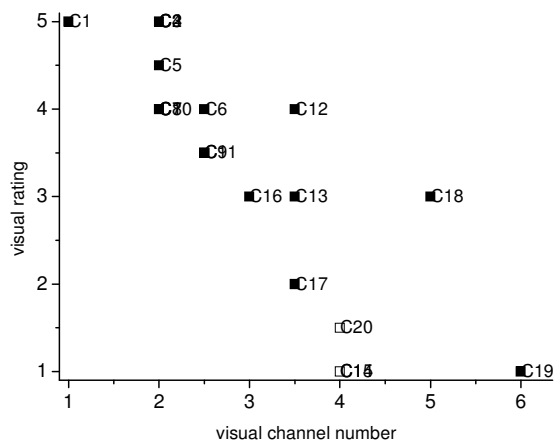


Figure 5-23: Visual rating of the final post-processing solution versus the visual channel number

5. Discussion

In this work, we handle complex structures but simple data. This work shares similar channeled structure complexity with natural media but is not close in terms of available data. The density of data is indeed greater than that for most of hydrogeological sites. Our first objective is to define an identification strategy adapted to highly channeled flow structures with simple data. The complexity comes from the discrete nature of the channels embedded in a significantly less conductive equivalent matrix. Apart from the first four configurations (C1 to C4), all the other configurations (C5 to C20) have some channel and connectivity complexities. The random generation of the configurations ensures that the configurations are not too simple. On this hand, we get closer to the field complexity. We address the parameterization issue rather than the numerical performances issue. We use simulated annealing as one of the simplest optimization algorithm adapted to highly non-convex objective functions. Simulated annealing however requires a high number of direct flow simulations explaining why we kept a steady-state flow configuration. In this other hand, we are not close to the commonly available transient-state drawdown field data. In the following of this section, we first discuss the knowledge that can be gained from the simple steady-state flow configuration and secondly the possible extensions of the presented identification strategy to transient-state data and to unknown boundary conditions.

Based on the steady state flow configuration, we discuss the content of the data in terms of information on the flow channel structure. The 20 tested configurations can be organized in four levels of complexity. The first level corresponds to configurations with one or two straight channels (C1 to C4). Their location can be identified only with steady-state head data. In this case, the wells surround the channels and the characteristic inter-well distance is smaller than the channel scale, a case for which identification has already been proved possible with only head data and a restricted number of nine wells [Tiedeman *et al.*, 1995]. Without flow or permeability data, fracture transmissivities are very poorly resolved. The second level corresponds to configurations with around two tortuous channels (C5 to C8). In these cases, the identification strategy cannot be successful with the sole head data but also requires the deviation of heads induced by the channels. As we have seen in section 4.4, it means that some of the effects of the boundary conditions have to be filtered out. Tortuosity is a second order structural characteristic requiring finer data at least partly provided by head

deviation. The third level corresponds to configurations made up of three complex and interconnected channels (C9-C13, C16). Head data and deviations cannot by themselves identify the channel structures. However, the addition of the distance of the wells to the nearest channel leads to determine the main flow structures. The fourth level of complexity corresponds to structures made up of more than four channels. Identification does not yield very relevant channel structures. These results can be synthetically explained in two ways. The first explanation comes from the potential limitations of the identification strategy. Because of the absence of fracture aperture distribution, the assumption of hierarchical organization of the flow channels is likely to break down when the number of channels increases. The second explanation relies on the comparison of the characteristic data and structure scales. For the first and second level of complexity, the inter-well distance is around twice as large as the inter-channel distance (Figure 5-4a). For the third level of complexity, the characteristic scale of the important channel interconnections becomes of the order of the inter-well distance. For the fourth level of complexity, two important channels can be found between two wells. The identification strategy as it has been performed in this study leads to reasonable channel structures only when the inter-well distance is of the order or smaller than the characteristic scale of the channel structure. The second explanation is that the identification strategy can work only when the distance between important structures is larger than the inter-well distance. To discriminate between these two possible interpretations, we have performed the identification strategy with 100 rather than 25 wells. We do not find significant improvements of the channel structure. Limitations thus come from the absence of strong flow hierarchy in the most complex structures tested in this study. The second possible interpretation will be investigated in further studies on fracture structures with broad transmissivity distributions.

We have worked with elementary data, boundary conditions and optimization algorithms towards the improvement of the parameterization of complex discrete channel structures. To get closer to field settings, the challenge is to extend the identification strategy to different data types with possibly partially known or fully unknown boundary conditions. So far, most inverse problem methods seeking for discrete flow structures have been performed on more complex data but with simpler parameterization. *Doughty et al.* [1994] sought for IFS parameters from transient-state drawdown data. *Day-Lewis et al.* [2000] determined inter-well connectivity structures also from transient-state drawdown data. *Bruyelle et al.* [*Bruyelle and Lange, 2009*] worked on a known fracture network structure and restrict the parameterization to the identification of the fracture transmissivities. *Gwo* [*Gwo, 2001*] determined the occupation probability of links on a discrete structure from transport solute breakthrough data. The advantage of the transient-state flow data over their steady-state counterpart is to be more informative of the flow structure and less dependent on boundary conditions. At not too late times, the hydraulic response is more influenced by the pumping location than by the far-away boundary conditions. Pumping systematically in different wells provides a possible evolving coverage of the sensitive zones while modifying potentially the main flow structures (hydraulic tomography) [*Hao et al., 2008*]. As our identification strategy relies on the hierarchical organization of the flow channels, it may lead to different and only partially consistent flow structures for evolving pumping locations. Combining them in a consistent discrete transmissivity structure is the first critical issue of shifting from steady-state to transient-state flow data. The second issue is the computational load issued by the transient-state simulations. Choosing 10 discrete times on 10 pumping tests performed in 10 of the 25 previous wells leads to the necessity to solve 250 flow problems at fixed times. The most appropriate method is the Laplace method with the Stehfest algorithm for the inverse Laplace transform [*Stehfest, 1970a; b*]. To achieve a reasonable precision, it requires at each time the

solution of eight full flow problems. On this example, the computational load increases by a factor of 2000 compared to a single steady-state flow simulation. The optimization algorithms should thus be improved to lower the number of flow simulations by one to three orders of magnitude. We argue that this may be possible first because we took a very simple simulated annealing algorithm and did not optimize its performances and secondly because transient-state flow data contain more information than steady-state flow data. The simulated annealing algorithm may be replaced by more advanced optimization methods like Covariance Matrix Adaptation or Monte-Carlo Markov Chain methods [Hansen and Ostermeier, 1996; Fu and Gómez-Hernández, 2009]. The combination of the different terms of the objective functions may also be improved to minimize the required number of simulations [Carrera and Neuman, 1986; McLaughlin and Townley, 1996; Bruckner et al., 1998]. The other interest in modifying the objective function is to increase its regularity and use gradient-like optimization methods. The optimization method may combine an initial stochastic search and a subsequent gradient-like scheme like what is done for the identification of permeability zones [Tsai et al., 2003a].

6. Conclusion

We propose an identification strategy for discrete channel flow structures typical of fractured media. We define a parameterization based on the hierarchical organization of the flow channels. Channels are determined iteratively by decreasing order of importance. We set up the methodology on a set of 20 configurations ranging from a couple of straight channels to more intricate interconnected configurations having up to six independent channels. All other flow settings and optimization algorithms have been taken as simple as possible. Boundary conditions are derived from a uniform head gradient. Data are made up of steady-state hydraulic heads and distances from the wells to the nearest channel.

The objective is to identify both geometrical and hydraulic characteristics of the conducting structures. Channels are identified by decreasing order of importance by using successive optimizations of an objective function. The identification strategy takes advantage of the hierarchical flow organization to restrict the dimension of the solution space of each individual optimization. Because of the successive optimizations, main flow channels determine strongly the characteristics of secondary channels. Additionally, main flow channels can be slightly modified by secondary channels through the introduction of a regularization term on the main channel characteristics in the objective function. The regularization term is weighting the rate of variations of the formerly identified structures compared to the newly added structures. Refinement of main channels is performed first before introducing additional less essential channels. The identification is stopped when both the improvements of the channel structures and of the channel number become marginal. A matrix has been introduced in order to replace the channels that should be identified later. The matrix permeability sharply decreases when the identification proceeds. Modifications of the objective function have been introduced to allow marginal modifications of the main flow structures by secondary flow structures. The classical simulated annealing method has been chosen as the optimization algorithm because of the strong non-convex nature of the objective function. For each configuration, the identification strategy has been run 50 times yielding 50 solutions. A post-processing algorithm extracts the prevalent channels from the 50 solutions.

The simplest configurations made up of a couple of straight channels are identified with only the steady-state head data. Identification of similar configurations having tortuous rather than straight channels requires additionally the knowledge of the head deviations. For intricate interconnected configurations made up of up to three tortuous channels, the identification

strategy needs additional geometrical data taking the form of the distances from the wells to the closest channel. The tested configurations consisting in more than four complex channels were not identified because of the absence of strong enough hierarchy in the channel structure.

The hierarchical channel identification strategy has only been tested on steady-state data. Its extension to transient-state data requires critical improvements of the optimization algorithms. We argue that this may be possible because transient-state data are more informative than their steady-state counterpart. The other advantage of well test or flowmeter data is to be less dependent of boundary conditions.

Appendix A. Simulated annealing parameterization

Simulated annealing performs a random walk in the parameter space directed towards the minimization of the objective function [Kirkpatrick *et al.*, 1983]. The acceptance of parameter sets leading to an increase of the objective function is necessary to get out of local minima. Its probability however decreases slowly in order to force the algorithm to converge to the global minimum. The simulated annealing algorithm is parameterized by the acceptance mechanisms of worse parameter sets, by the definition of the random walk in the parameter space and by its condition of termination. The probability of acceptance is managed by an energy criterion characterized by a temperature. The temperature scheduled has been chosen according to Ingber [1993a]:

$$T_k = T_{init} \cdot \exp\left(-c_i \cdot k^{\frac{1}{D}}\right) \quad (13)$$

where T_{init} is the initial temperature, D the dimension of the parameter space, k the number of accepted states and c_i a user-defined value that can be adapted to improve the algorithm performances and which default value is set to:

$$c_i = -\log 10^{-5} \cdot \exp(-\log 100 * 1/D) \quad (14)$$

The acceptance test defines if a new parameter state is accepted or rejected. It writes:

$$u < \exp\left(-\frac{F_i - F_{i-1}}{T_i}\right) \quad (15)$$

where u is a random value drawn from a uniform distribution on $[0;1]$, F_i is the value of the objective function with the new parameters, F_{i-1} is the value of the objective function at the last accepted state and T_i is the current temperature. Thus, if the new parameters induce a decrease of the objective function, they are accepted for sure. Otherwise, they are accepted with a probability that depends on how they penalize the solution. This probability decreases when the temperature becomes smaller.

The random walk within the parameter space is characterized by the parameter modifications from steps i to $i+1$. For the parameter p^j :

$$p_i^j = p_{i-1}^j + y_i \cdot (p_{max}^j - p_{min}^j) \quad (16)$$

where p_{max}^j and p_{min}^j are the minimal and maximal possible value for p^j and y_i is the following scaling factor:

$$y_i = \text{sign}(u - 0.5) \cdot T_i \cdot \left[\left(1 + \frac{1}{T_i} \right)^{|2 \cdot u - 1|} - 1 \right] \quad (17)$$

where u is a random value drawn from a uniform distribution on $[0;1]$. With this formulation, the random walk shifts progressively from global to local thanks to the decrease of temperature. The optimization stops when the objective function becomes smaller than a fixed minimal value $F_i \leq F_{min}$ or when the maximal number of allowed iteration is reached.

Acknowledgements

The French National Research Agency ANR is acknowledged for its financial founding through the MOHINI project (ANR-07-VULN-008) and for its contribution to the development of numerical methods through the MICAS project (ANR-07-CIS7-004). Additional funding was provided by the French Association for Research and Technology ANRT (CIFRE-747/2006).

B. Influence des conditions aux limites sur la sensibilité des données de charges stationnaires aux structures d'écoulement.

La méthode d'identification décrite au paragraphe précédent dépend des données utilisées pour calibrer les paramètres. Dans cette partie, on étudie le lien entre la structure des écoulements, les conditions aux limites et l'information contenue dans les charges hydrauliques stationnaires. Elle est proposée sous forme d'un article scientifique et est introduite par un résumé en français.

1. Résumé

Dans les milieux hétérogènes, et plus particulièrement dans les milieux fracturés, un petit nombre de structures très transmissives contrôlent les propriétés hydrauliques. L'identification de ces structures est cruciale afin de comprendre et de simuler la dynamique des écoulements. Dans la mesure où une identification de ces structures à partir d'observations directes est impossible, il est nécessaire de résoudre un problème inverse s'appuyant sur une observation des propriétés hydrauliques comme les données de charge. On s'intéresse ici à la capacité de ces données de charges à apporter une information pertinente sur la caractéristique des structures de chenalisation. A la différence du problème classique où on identifie un champ de perméabilité, les chenaux d'écoulement sont très sensibles aux conditions hydrauliques et aux conditions aux limites. On montre ici que les charges hydrauliques peuvent réciproquement être très sensibles aux chenaux et apporter des informations importantes sur leur position et leur transmissivité. Néanmoins, certains chenaux se trouvent dans des "zones d'ombre", c'est-à-dire dans des parties du milieu où les charges sont très peu sensibles aux écoulements. En changeant les conditions hydrauliques, par exemple en sollicitant le milieu, il est possible de déplacer voire de réduire ces zones d'ombre afin de renforcer la sensibilité des charges aux chenaux. Par conséquent, en appliquant plusieurs conditions hydrauliques, il est possible d'identifier une plus grande partie de la structure de transmissivité. On illustre ce résultat en résolvant le problème inverse à partir de données de charge obtenus avec et sans sollicitation et en changeant la position du puits de pompage. Les résultats obtenus sont ensuite analysés afin d'extraire les structures les plus pertinentes. On montre ainsi qu'il est possible, à partir de données de charge, d'identifier la plupart des chenaux principaux. En conclusion, les charges hydrauliques sont donc des données bien adaptées à l'identification des écoulements et qu'en sollicitant le milieu améliore significativement la quantité et la qualité de l'information disponible et permet d'obtenir de meilleurs résultats lors de la résolution du problème inverse.

2. Influence of boundary conditions on the sensitivity of steady head data to the flow structure

R. Le Goc ^{a,b}, J.-R. de Dreuzy ^a and P. Davy ^a

^a Géosciences Rennes, UMR 6118 CNRS, Université de Rennes 1, CS 74205, F-35042 Rennes Cedex, FRANCE

^b Itasca Consultants SAS, 64 chemin des mouilles, F-69134, Écully Cedex, FRANCE

Abstract

In heterogeneous media like fractured crystalline rock, fluid flow is often controlled by a limited number of highly conductive features. Identifying the few structures that dominate the flow behavior is of great interest to better comprehend the fluid dynamics. As a direct identification of these structures

is very uncertain, an inverse approach is required. We investigate the ability of head data to provide relevant information about the characteristics of the channeling structure. Unlike the classical inverse problem, flow channels are highly sensitive to the hydrological settings and the sensitivity of head values varies with the boundary conditions. We show that head data are very sensitive to main flow channels and provides relevant information on their localization. Nevertheless, some flow channels fall into "shadow zones", i.e. zones that do not influence the head values. Changing the hydraulic conditions, by adding a pumping well in the system, can efficiently reduce the extent of these zones and can enhance the influence of main channels in their Head sensitivity to channels

a) Introduction

Highly channeled flows are widespread in fractured media. For almost impervious rock types like crystalline rocks, flows concentrate in fractures and even more in some highly transmissive flow paths within fractures [Koltermann and Gorelick, 1996; Tsang and Neretnieks, 1998]. What we call "extremely channeled flows" occur when the channels are very sparse at the scale of interest. For example, the water inflow within 50 meters-long gallery of Stripa was found to be all concentrated in a single zone and at 80% in a single fracture, all other mapped fractures being dry [Rasmuson and Neretnieks, 1986; Olsson, 1992]. Similarly, at the Fanay-Augères uranium mine (France), although the mapped fracture pattern suggested a highly connected fracture network, hydraulic tests indicated that the flow system was dominantly controlled by a single major fault [Billaux *et al.*, 1989]. In media affected by this phenomenon, the challenge is to locate main channels. Direct prospection by well drilling has a very low probability of hitting them, meaning that all kind of measurements performed directly on cores and very close to the well require an important sampling and interpretation effort to be relevant [Le Borgne *et al.*, 2006] and classical hydraulic tests cannot generally capture the characteristics of such media [Sanchez-Vila *et al.*, 2006]. Similarly, geophysical signals are generally not sensitive to the infinitesimal fraction of volume occupied by the channels. Head data, conversely, can be very useful provided that they remain remotely sensitive enough to the channel structure. Renshaw [Renshaw, 1996] shows indeed that it was possible to estimate the geometry of an isolated fractured zone using head data with a co-kriging method. Detecting the largest permeability channels from heads is typically an inverse problem as causes are determined from their consequences.

(1) Specificities of channeled media

Channels, as opposed to permeability fields, heavily depends on boundary conditions and source terms [Neuman, 2005]. Their use in other flow conditions than the one used for their identification is a fundamental issue like the use of data obtained with different flow conditions for identifying the same flow channels. Thus, the inverse problem for extremely channeled flows yields weaker information consisting in flow channels than the classical inverse problem resulting in intrinsic permeability fields. It is not linked to the inverse problem per se or the way it is formulated but more fundamentally to the medium heterogeneity. Stochastic hydrogeology gives an illustration of this dependence [Zhang, 2002]. So, the identification of main flow channels does not consist in drawing flow tubes from the interpolation of existing head data as in the Cauchy problem that is known to be instable [de Marsily *et al.*, 2000]. It rather consists in looking for a transmissivity structure underlying the flow channel. Channeling is indeed either caused by a large contrast between high- k and low- k values [Koltermann and Gorelick, 1996] or by the presence of high- k connected structures within the system [Journal and Alabert, 1989; Western *et al.*, 2001]. When both effects are combined, channeling is extremely important [Knudby and Carrera, 2005]. In fractured media channeling is mainly driven by the presence of large fractures and possibly by broadly distributed apertures [Le Goc *et al.*, 2010b]. This is the main large

structures that have to be identified to capture correctly flow properties. Figure A-1 shows two moderately and strongly channeled synthetic fracture networks (top line) in horizontal and vertical flows. More precisely, boundary conditions are horizontal and vertical permeameter-like. The dependence of flow channels to boundary conditions is obvious, but for the extremely channeled flow case, the fractures carrying the flow are mainly the same. Note also that head sensitivity concentrates in the channeled zone, leaving large "shadow zones" where the channeling structures have a smaller influence on heads. It corresponds typically to fractures perpendicular to the head gradient or dead ends that do not show up in the head signal.

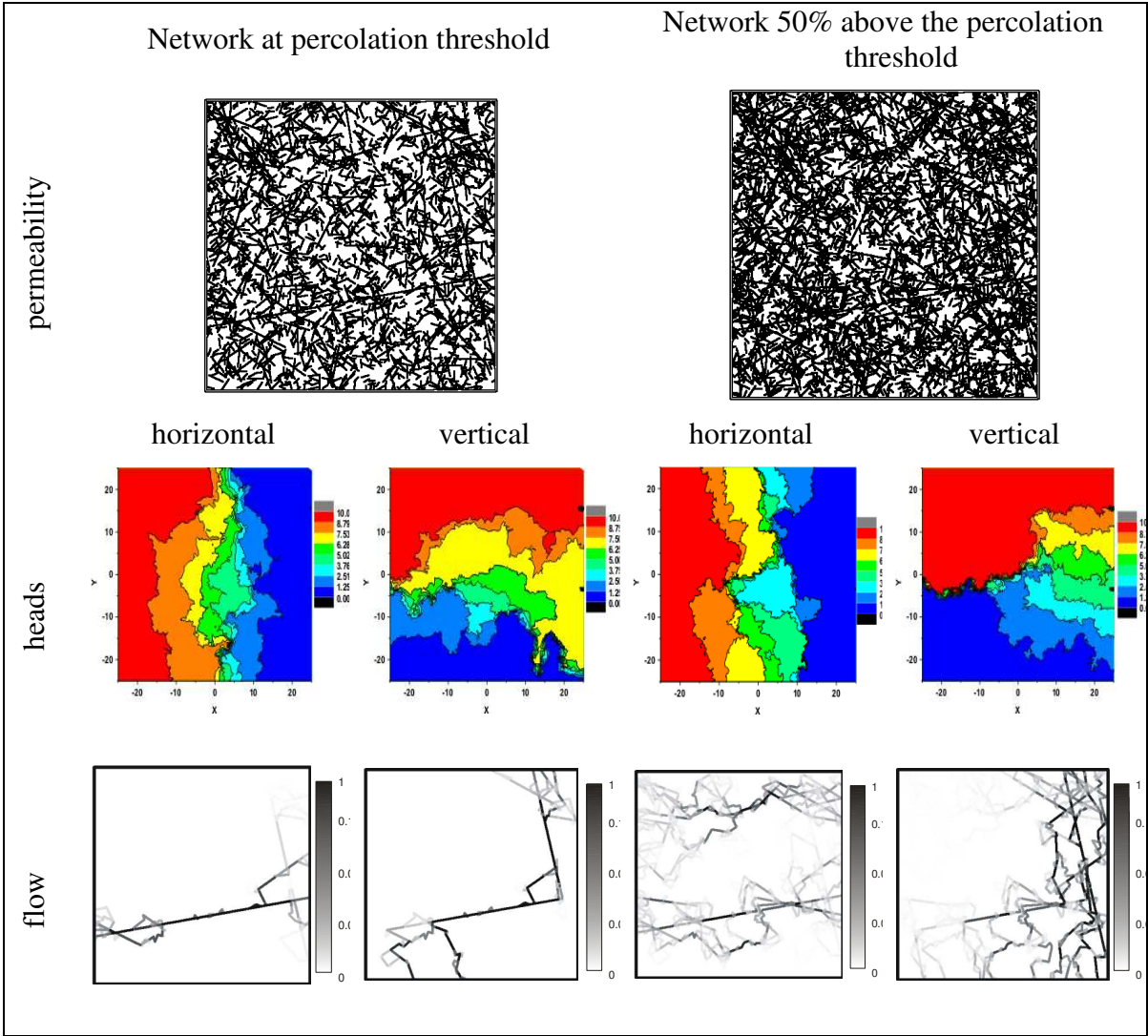


Figure A-1 : Comparison of flow structures in highly (left) and moderate channeled fracture network (right). TOP: fracture networks were generated with a power-law length-distribution (exponent = 3.0) and uniform random orientations and positions in a system size of 25. Moderate and highly channeled networks were obtained for a density at the percolation threshold (left) and 50% above the percolation threshold (right). Note that the generation seed is the same in the two cases. Boundary conditions are permeameter-like, i.e. they are fixed on two opposite sides of the domain and the perpendicular domain sides are impervious. The middle row represents head from 0 to 10. The bottom row represents the flows. We notice that flow channels are based on existing active transmissivity structures and not strictly depending on the boundary conditions.

As a conclusion, for extremely channeled media, the dependence of the identified structure to the boundary conditions remains a key question for assessing the relevance of the inverse problem. It does however not discard it for two reasons. First, without additional information, the dependence to the boundary conditions cannot be filtered out from the inverse problem whatever the methodology. Identification of channels can be considered as a necessary intermediary between the head data and the permeability field. Secondly, variability of boundary conditions is limited. Boundary conditions change mainly in the two circumstances of hydrologic variations and hydraulic tests. Hydraulic tests are highly interesting as they change the sensitivity map in highlighting channels that do not show up in natural conditions. In a first a version, the inverse problem could thus be performed independently on different boundary conditions and yield different flow channel structures, which analysis would give the permeability field.

We analyze in section 0 the information on the flow structure concealed in head data with varied boundary conditions. We introduce in section c) the method used to solve the inverse problem and discuss the results. Section d) is a discussion about the importance of locating main channels and the possible improvement required to characterize main flow structures.

b) Head sensitivity to channels

This paragraph investigates the feasibility of solving the inverse problem on channeled flow media. The first issue is to evaluate the information on flow channels concealed by head data in relation with the boundary conditions.

(1) Test cases

We first define a synthetic test case used to challenge head data sensitivity. We generate a determinist fracture network presenting a high degree of channeling using H2OLAB, an object-oriented software with a modular approach. Fractures are modeled with 2D segments and Darcy's law is discretized at the segment scale with the stream-line routine at each intersection [Hull and Koslow, 1986] to get the flow. It ends up to a linear system solved with the software UMFPACK [Davis, 2004].

The main synthetic test case consists in a 10x10 field containing a fracture network leading to a hierarchical organization of flow channels. It is composed by organized fractures with different transmissivity values (from $10^{1.4}$ m²/s to $10^{1.9}$ m²/s), which leads to two main flow channels and secondary channels in gradient-like boundary conditions, i.e. heads are fixed at $h=10$ and $h=0$ on side $x+$ and $x-$ and fixed with a regular gradient on side $y+$ and $y-$ (Figure A-2). This test case is a schematic simplification of what we expect from highly channeled media, ignoring the non-significant structures like fractures carrying a negligible quantity of flow.

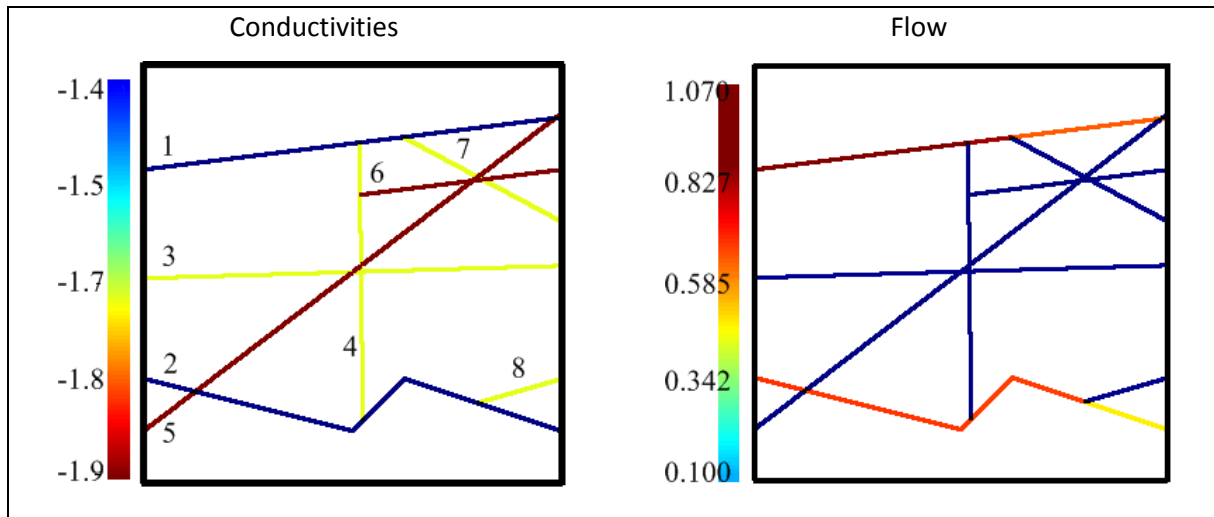


Figure A-2 : Main synthetic test case. (Right) The conductivity structure mimics the presence of transmissive paths carrying flow. Two main channels are thus created with connected structures having a transmissivity $T=10^{1.4}$ ($n^{\circ}1$ and $n^{\circ}2$), intermediary channels with a transmissivity $T=10^{1.7}$ ($n^{\circ}3, 4, 7$ and $n^{\circ}8$) and minor channels with $T=10^{1.9}$ ($n^{\circ}5$ and $n^{\circ}6$). (Left) corresponding flow in gradient-like boundary conditions. We observe that flow is located in the two main channels.

(2) Discrepancy maps and boundary conditions

Natural boundary conditions are generally fixed by environmental settings. The flow organization within the transmissivity structure depends in this case of various parameters including the orientation of flow channels in relation to boundary conditions [Fan et al., 2007] and the relative transmissivity of each channel. Kahn and co-workers [Khan et al., 2007] show that the spatial structure of sensitivity denotes then that sensitive areas are located along main flow channels that are connected to source or sink, i.e. boreholes or boundaries that transmit fluids into or out the model. Sensitivities are generally used to estimate the piece of information contained in a data value, but a sensitivity analysis requires a model of the transmissivity structure. We propose a tool to estimate this information prior to the modeling stage to comprehend globally the interactions between flow structure, boundary conditions and head sensitivity. In this context, we calculate the absolute differences between heads in the homogeneous case and heads in the studied case, with identical boundary conditions. We refer to this tool as a "discrepancy map" as it depicts the dissimilarities in heads due to the transmissivity structure. Heads are calculated by applying a synthetic homogeneous matrix composed by infinite vertical and horizontal fractures with a low-transmissivity ($T_{mat}=10^{-6}$) that cross at head positions. The advantage of this method is that it avoids a double permeability model and flow is calculated with the same algorithm with and without the matrix. In the present cases, a 100x100 grid of head position is applied to get sufficiently accurate maps.

(3) Information provided by heads in natural boundary conditions

We first investigate the influence of the orientation of the hydraulic gradient on the information contained by head data. We apply gradient-like boundary conditions to the test case of Figure A-2 and apply a rotation of 0° , 10° and -20° . Figure A-3 displays the resulting discrepancy maps. We observe first that heads are generally the most discrepant near the channels carrying the most significant part of flow ($n^{\circ}1$ and $n^{\circ}2$ on Figure A-2). It means that heads contain a significant piece of information on the main channels position. Furthermore, this information is generally decorrelated, i.e. each piece of data is mainly influenced by the

nearest channel, which means that data near a channel are required to locate it. Moreover, the piece of information used to identify a channel is different from the piece of information used to locate another distant channel, which is important in the prospect of solving the inverse problem. Some channels, meanwhile, are completely transparent in the discrepancy maps. This is what we called "shadow zones" in introduction. As illustrated on Figure A-3, channel n°3 does not influence the head values for the selected boundary conditions, although it carries a non-negligible part of the flow. The reason is that it is almost parallel to the two main channels, which lead to a competition between them to influence the heads. As it is less transmissive, only channels n°1 and 2 influence heads. The channel n°4, which is perpendicular to the flow direction, is also hardly visible on the discrepancy maps. In this case, its orientation plays most probably a significant role in its ability to control the head variations.

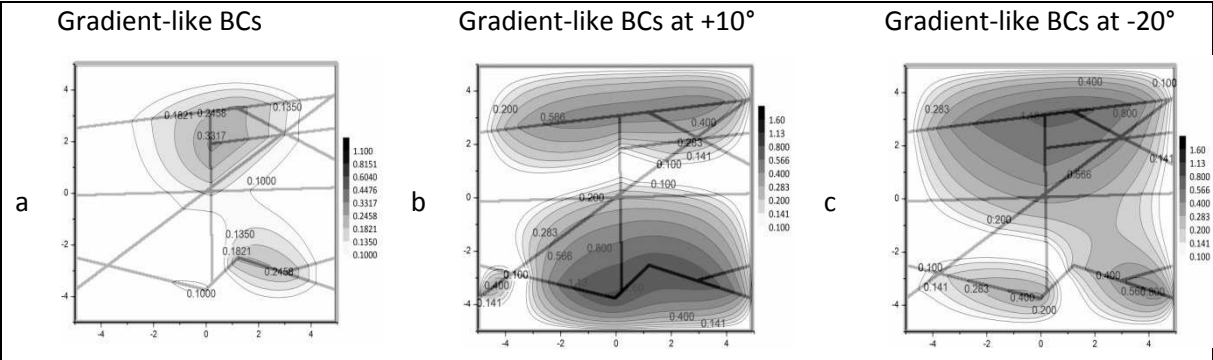


Figure A-3 : Discrepancy maps of the considered test case with gradient-like boundary conditions with different orientations : 0° (left), +10° (middle) and -20° (right). The underlying transmissivity structure is superimposed over the discrepancy map to visualize the influence of channels position.

The influence of the orientation of the boundary conditions is clearly visible in the respective influence of each channel. By comparing Figure A-2b and Figure A-2c, we observe that a rotation of a few degree changes deeply the quantity and the position of the information concealed by heads. The hierarchy between channels is reversed as for gradient-like boundary conditions at 10°, the most significant channel is at the bottom of the field (n°2 on Figure A-2) while for boundary conditions at -20° the top channel (n°1 on Figure A-2) dominates the head variations. In the same time, channel n°5, which has a favorable orientation in Figure A-2b and is so visible in the discrepancy maps, becomes undetectable when the orientation of the boundary conditions is changed (Figure A-2c). It illustrates the strong dependence between the main orientation of the hydraulic gradient and the ability of a channel to influence head values, and therefore the capacity to solve the inverse problem. Consequently, head data sampled during periods of low water-levels are not necessarily influenced by the same flow channels than head data sampled during periods of high water-levels.

(4) Adding information in heads in changing the boundary conditions

The "shadow zones" represent a problem when solving the inverse problem, as potentially important structures are invisible in the data. One can hardly modify the global orientation of flow. It is however quite simple to make demands on the field in pumping and/or injecting a liquid within an existing well. Depending on the location of the well and its connection to fracture system, one can hope a displacement of the "shadow zones" to induce variations in the sensitivity of head values. Thus, we simulate a pumping at three different locations to investigate the changing in head variations. A pumping well is only represented by its

position and we add two artificial fractures in the system to connect the well to the fracture system. These fractures are composed by a vertical and a horizontal segment, which origins are at the well location and targets on the nearest fracture. We assign the transmissivity of the connected fractures to the artificial segments. We consider then a pumping as an imposed head at the well (here equal to $h=2$). Boundary conditions are taken as gradient-like without rotation. Figure 4 shows the resulting discrepancy maps.

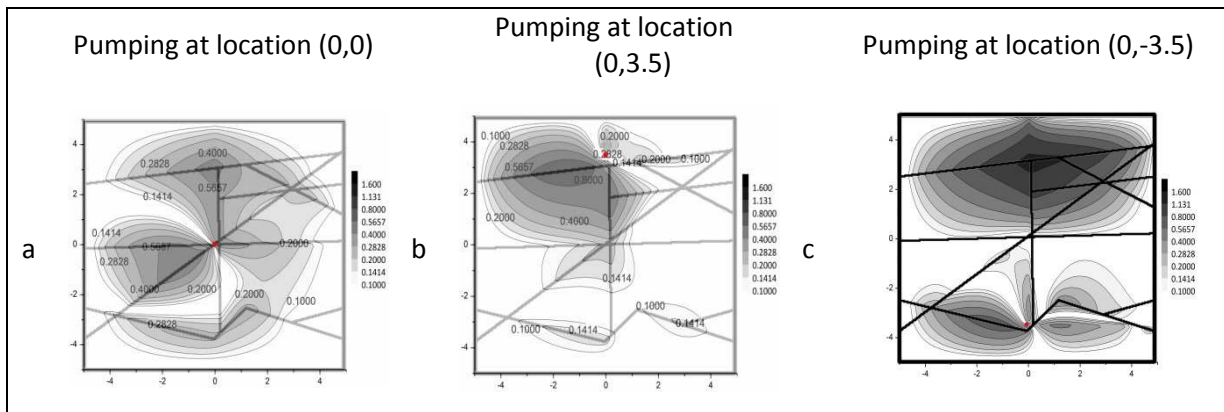


Figure A-4 : Discrepancy maps of the considered test case with gradient-like boundary conditions and adding a pumping at different location: (a) in the middle of the field, (b) in the upper part and (c) in the topper part. The underlying transmissivity structure is superimposed over the discrepancy map to visualize the influence of channels position. The red points indicate the position of the pumping well.

Without pumping, head variations were small excepting in the central part of the field (Figure A-3a). With a pumping, the discrepancy in heads depends of the distance to the pumping well and is disturbed by the channeling structure. The area of influence is strongly dependent on the channel directly connected to the pumping well since its control on head data is significantly increased. This is particularly interesting in cases where the channel was in a "shadow zone" in natural boundary conditions. For example, channel n°3 has a strong influence on head values when the pumping well is located close to it, at the center of the field (Figure A-4a). We notice also that channels not directly connected to the well can see their importance grow, if they are linked to the channel connected to the well. For example, channel n°1 has a strong influence on Figure A-4c even if the pumping is located at the bottom of the field. The dissymmetry between Figure A-4b and Figure A-4c can be explained by the position of the pumping relative the channel. On Figure A-4b, the pumping is located near the top channel but outside the complete channeling structure. In this case, the nearest channel is the most affected by the pumping. On Figure A-4c, against, the pumping is located inside the structure, allowing the top channel to be significantly sensitive to the pumping location. The increase in influence is thus transmitted from the pumping well to the connected structure, and to the possibly neighboring channels. Consequently, mixing head data under different boundary conditions, with pumping and/or injection, is valuable as it provides additional information on the transmissivity structure inexpensively by enhancing the data sensitivity and by removing most of the "shadow zones". Some channels (e.g. n°8), however, are unlikely to be detected with head data, unless a pumping well hits them directly. It reflects the important role of the global orientation of the hydraulic gradient, even with pumping.

(5) Reducing the influence of boundary conditions

The control of boundary conditions on head values lead to the presence of "shadow zones". One should expect that fixing them away from the system limits would reduce their influence and possibly remove these zones. Moreover, if the natural conditions are usually unknown on

the system boundaries, they can generally be inferred at a larger scale by the hydrological settings. We investigate the sensitivity of head data with gradient-like boundary conditions fixed at the limits of a system 5 times larger than the studied one, i.e. in our case a 50x50 meters square. Existing fractures that intersect the real system limits are connected to the boundaries of the larger system in increasing their size in the appropriate direction. To keep a similar hydraulic gradient on the area of interest, we multiply by 5 the imposed head values, i.e. heads are fixed at $h=50$ and $h=0$ on side $x-$ and $x+$. We then calculate head values on the channeled case and use them as boundary conditions for the homogeneous case. The resulting discrepancy map is given in Figure A-5a. On Figure A-5b, the hydraulic gradient is oriented at $+10^\circ$ and on Figure A-5c, a pumping is applied in the center of the field.

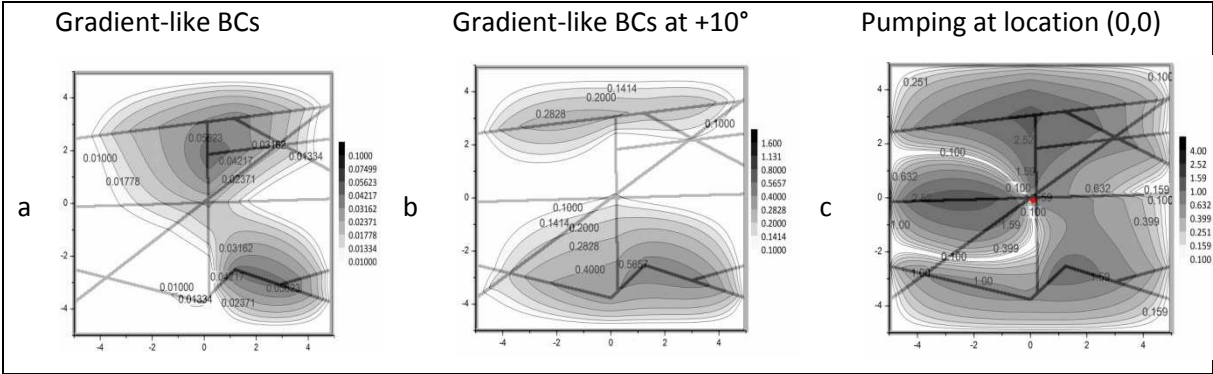


Figure A-5 : Discrepancy maps of the considered test case with gradient-like boundary applied on a system five times larger than the investigated one : (a) no rotation, no pumping, (b) rotation of $+10^\circ$ and (c) pumping at the system center. The underlying transmissivity structure is superimposed over the discrepancy map to visualize the influence of channels position.

Under these conditions, the sensitivity of heads is reduced in natural boundary conditions but the area of influence of each channeling structure is not modified (Figure A-5a,b versus Figure A-3a,b). We deduce that heads are largely impacted by the heads imposed on the limits, which reduces the effect of the transmissivity structure. Fixing the boundary conditions away from the system is not, in this case, an efficient way to enhance the data sensitivity. With a pumping well (Figure A-5c), for against, the discrepancy map reveals that the transmissivity structure influence more head values and the shadow zones have almost completely vanished. The reason is that the influence of pumping is limited when the boundary conditions are fixed at the system limits but is freely propagated if they are applied away from the system. Therefore, when the hydraulic system is under stress, it is more dependent on the boundary conditions than when flow is under natural conditions. Consequently, moving boundary values away from the system enhances the influence of the transmissivity structure only when there is pumping or injection and has an opposite effect without solicitation.

(6) Conclusions on the data sensitivity

The relation between the transmissivity structure, the boundary conditions and the information concealed by head data is, if not obvious, clearly quantifiable with discrepancy maps. Flow is indeed governed by the hydrological settings and the transmissivities. Boundary conditions are the first-order term, insofar as they directly control the significance of a particular channel on heads. The transmissivity structure is the second-order term and plays a role important enough to be detected in heads. The information concealed by heads contains indications on the channeling structure (localization, transmissivity) once the influence of boundary conditions is filtered out. It is furthermore possible to change the

hydraulic conditions by pumping or injecting water in wells in order to highlight structures that fall into "shadow zones".

Heads influenced by a channeling structure are mainly located near this structure and to the possibly connected channels. They are remotely sensitive enough to detect channels, the influenced volume (or surface) being larger than the volume (or surface) occupied by the underlying transmissivity structure. Head values are thus relevant data that can be used to solve the inverse problem. As the sensitivity presents a hierarchical organization related to that of channels, it permits the iterative identification of channels, which accuracy will depend on the quantity of available data and their localization relative to the underlying structures and the boundary conditions. Adding a pumping well enhances the sensitivity of data and increase the influence of channels in its vicinity. Using heads in varied boundary conditions should thus improve the inverse model.

c) Solving the inverse problem

Observation made on discrepancy maps are validated with a method that solves the inverse problem on highly channeled media. In this section, we use heads obtained with various hydraulic conditions as inputs to the solving method and we analyze the models resulting from the resolution.

(1) Characteristics of the method

Solving the inverse problem for highly channeled media requires accounting for the particularities of the transmissivity structure. The parameterization must focus on the structures carrying the main flow channels and a care use of the head data must be done. A complete procedure dedicated to highly channeled media has thus be developed, which details are given in. The main optimization step is based on a simulated annealing algorithm and the procedure relies on three pillars:

- It models only main flow channels, other structures, assumed almost insignificant, are modeled with a homogeneous matrix.
- The number of parameter is iteratively increased. The inverse problem is first solved on a simple model containing only one channel. Then, new channels are added in the model if the inverse solution is satisfactory in terms of objective function and uncertainty.
- Instead of using a large amount of computer resources to calibrate a complete model, several simpler solutions are calculated and finally analyzed to infer the final model.

The concepts developed in this paper are consistent with the observation made on discrepancy maps. Head data are indeed mainly influenced by main channels, meaning that a relevant model should contain only these channels. Moreover, in the iterative parameterization, one should expect that the channel that has the main influence on the head data be identified in the simplest model while other channels are thereafter detected in models with a more complex parameterization, provided that head data are sensitive enough. With such a scenario, the homogeneous matrix modeling all second-order structures should be less and less significant.

So, the methodology described in [Le Goc *et al.*, 2010a] is applied on the test case presented on Figure A-2 under different conditions, with "natural" gradient-like boundary conditions and adding a pumping at different location of the aquifer. [Le Goc *et al.*, 2010a] developed also a method that is able to merge different solutions, coming from different execution of the solving procedure, to create a global solution. This method is based on a hierarchical clustering algorithm that merges together solutions that are just a few different from each

other. It could be used to merge the solutions obtained with different hydraulic conditions in order to identify as most transmissivity structures as possible by a comparison of all solutions, regardless of their boundary conditions.

(2) Relation between the iterative procedure with discrepancy maps

The iterative refinement of the parameterization used in [Le Goc et al., 2010a] would be particularly efficient if the flow structure is hierarchical and if head data integrate this hierarchy. This assumption can be verified using discrepancy maps. A discrepancy map can indeed be defined as the absolute difference between heads in the complete case and heads obtained with the best simpler model.

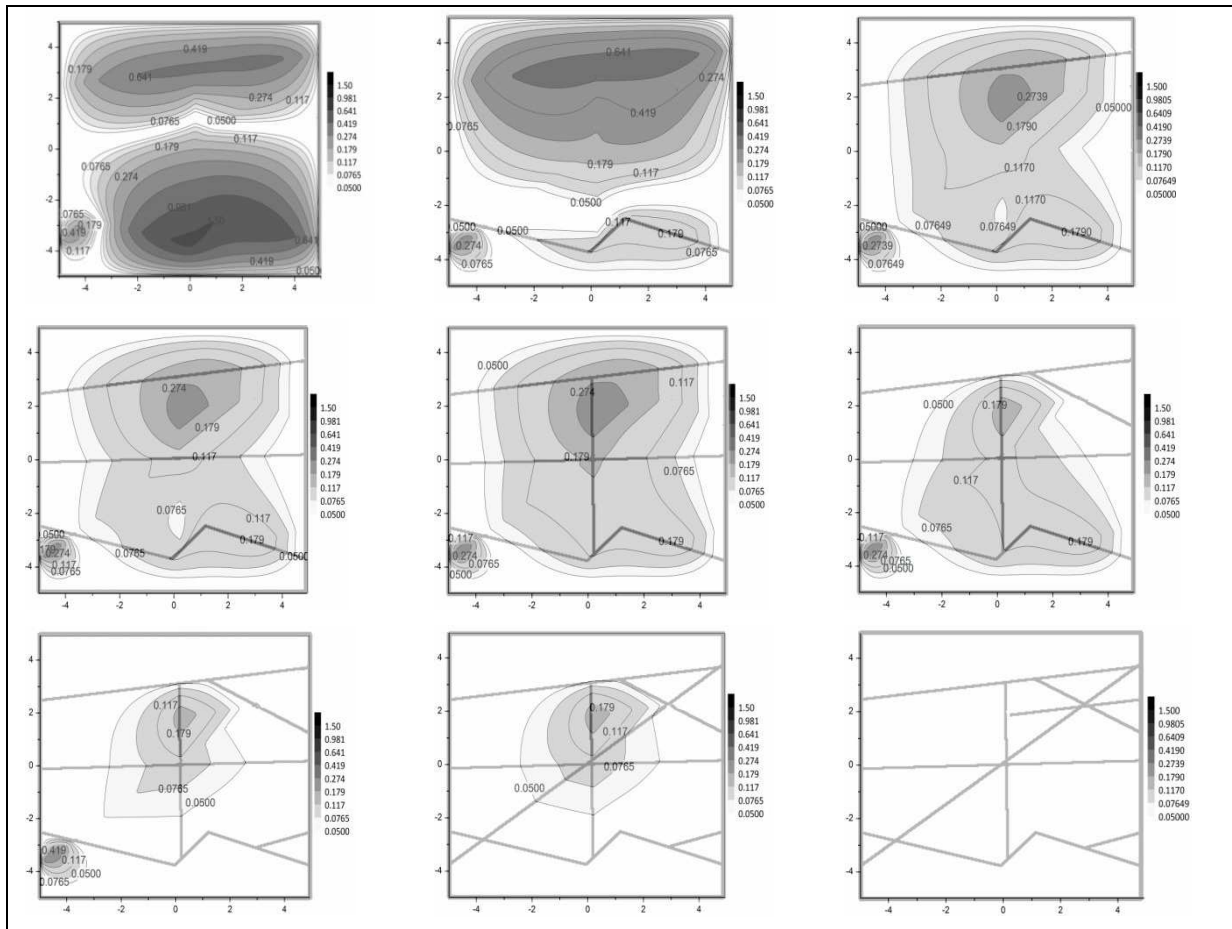


Figure A-6 : Discrepancy maps with gradient-like boundary rotated at $+10^\circ$. It represents the absolute difference between heads calculated with the complete test case and partial solutions having from 0 to 8 channels. The partial transmissivity structure is superimposed over the discrepancy map to visualize the influence of channels position.

On Figure A-6, we provide such maps, beginning with a homogeneous model (classical discrepancy map) and increasing regularly the number of channels in the model, mimicking the expected evolution of the iterative parameterization. On each map is depicted the remaining information available in head data to calibrate new parameters. We notice that the two main channels (n°1 and 2) contain the major part of information while head variations due to other channels are more diffused. Illustrative examples are visible between pictures 5 and 6 and between pictures 6 and 7 on Figure A-6. The information induced by channels n°7 and 8 is located close to the nearest main channel (respectively n°1 and 2) rather than near this one, which means that their localization will be more complex. We notice also that the

intersection between the channels n°4 and 5 contains a persistent piece of information correlated to several channels, including channels not directly connected this intersection like channels n°7 and n°8. The question if such persistent information is useful to solve the inverse problem remains open. As a conclusion, using progressively information concealed in head is feasible, particularly for main channels inferring the major part of information. Second order channels can also be identified using heads near intersections although this knowledge is not directly related to these flow paths. Based on this information, one should suppose that the inverse procedure will correctly identify the two main channels but will only partially match the other channels.

(3) Results

We solve the direct problem on the transmissivity structure described on Figure A-2. A homogeneous matrix with a transmissivity $T_{\text{mat}}=10^{-6}$ is applied to get 25 head values located on a 5x5 regular grid. We apply first gradient-like boundary conditions rotated at 10° and then add a fixed-head pumping to the coordinates (0,0), next to the coordinates (0,3.5) and finally to the coordinates (0,-3.5), as in the discrepancy maps illustrated on Figure A-3 and Figure A-4. These head data are then used as inputs for the procedure that solves the direct problem. Similarly to [Le Goc et al., 2010a], the settings of the inverse problem solver are a maximal number of iteration per optimization fixed at 50000, an initial temperature equal to 1, 10 and 100. Fifty single solutions are calculated and merged with a distance criteria equal to $d_{\text{max}}=5.0$. The resulting models, depicted on Figure A-7, are thus obtained only with the information provided by head data.

Without pumping, only the two main channels are identified, including the middle part of channel n°5. This is consistent with the observation of the corresponding discrepancy map (Figure A-3b). Head data are indeed mainly influenced by the two main channels, other structures playing a minor role or falling into shadow zones. Channel n°2 is less accurately located because its shape is more complex. This is a known limitation of the method used to solve the inverse problem. The top channel transmissivity is reduced compared to its real value ($\log(T_{\text{channel}}/T_{\text{matrix}})=4.6$) while the bottom channel has a relevant transmissivity value, although head data are generally insufficient to calibrate correctly the real value under natural boundary conditions [Le Goc et al., 2010a]. Results obtained with pumping are better, insofar as they locate more efficiently the main structures. It reinforces the assumption that the physical structure carrying the channels, which does not depend on the boundary conditions, is more important than the flow channels, which are dependent of the hydraulic conditions. Furthermore, adding pumping wells permits the detection of channels that are invisible in the solution without pumping, mainly channel n°3 and channel n°5, although the localization is less accurate. Changing the hydraulic conditions has so successfully improved the identification of the transmissivity structure. Moreover, it reduces the uncertainty on the channels position. The statistical weight of each channel is highly dependent of the pumping location, where it reflects only the hierarchy in channeling with natural boundary conditions. A pump favors logically the structures near the pumping well, which was the objective of their addition in the system. The transmissivity values are calibrated in a similar manner, i.e. channels near the borehole have an accurate value while channels that are more distant have a smaller transmissivity value than the real one.

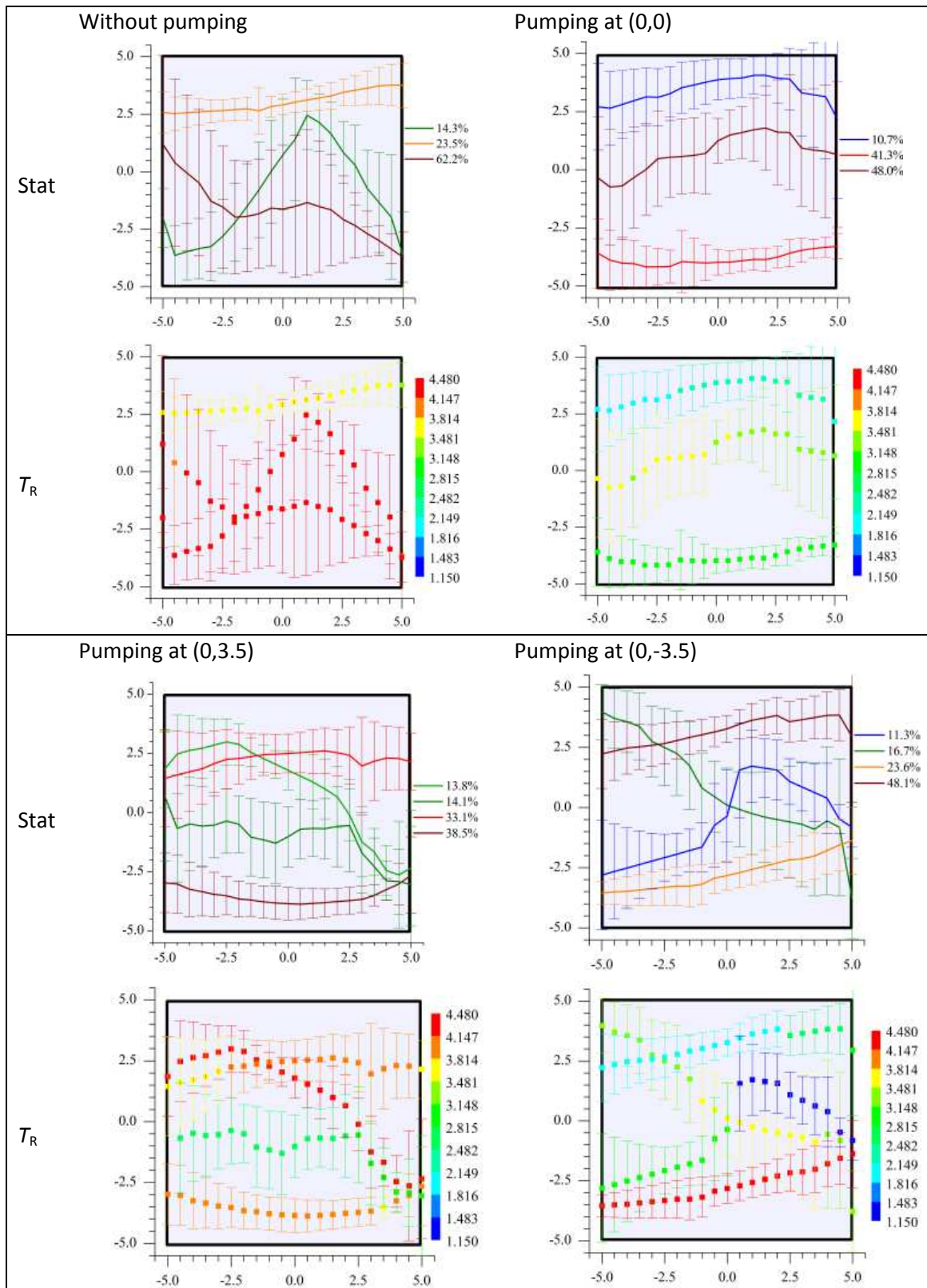


Figure A-7 : models obtained after the resolution of the inverse problem on 25 head data under different hydraulic conditions. Top, channels are colored according to their statistical weight among the solutions; bottom, channels are colored according to their relative transmissivity $\log(T_{\text{channel}}/T_{\text{matrix}})$.

The resulting models obtained under different hydraulic conditions are strongly correlated to the discrepancy maps. It validates first the use of such maps to estimate a priori the quantity of information available in head data. It suggests also that the method used solve the inverse problem is efficient, as most of the information provided by head data is used to characterize the transmissivity structure. We show that adding a pumping in the system is a relevant manner to improve the identification of the flow channels. Nevertheless, it seems that the identification abilities of the method depends in all cases of the global orientation of the hydraulic gradient, inasmuch as channel n°4, which is perpendicular to the main flow direction, was never identified, even with a pumping. Applying the merging procedure to the solutions obtained with different hydraulic conditions is logically the next step, in order to combine the identification abilities of each case. The localization of main structures should be improved but the effect on transmissivity values is more hazardous. As a pump tends to favor the transmissivity of the nearest channel, one should check that the pumping wells applied in solutions used to create the final model are distributed fairly over the field, which is the case in our example. Figure A-8 shows the model resulting from the combination of all solutions.

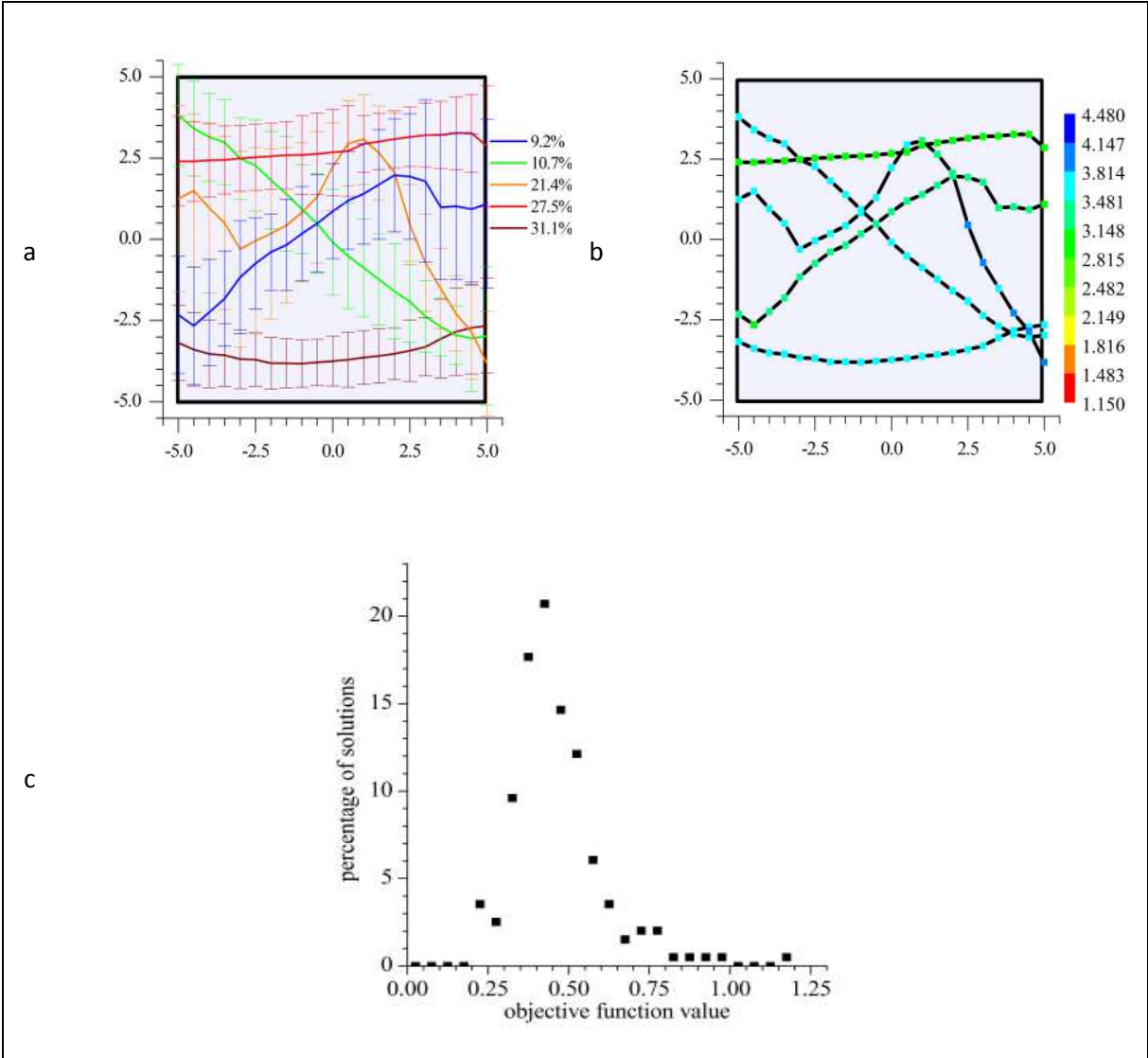


Figure A-8 : Final solution resulting from the merge of all single solutions obtained under different boundary conditions. (a) final channeling structure colored by their statistical weight, (b) channels colored by their ratio $\log(T_{\text{channel}}/T_{\text{matrix}})$, (c) distribution of objective function value calculated with all transmissivity structures with gradient-like boundary conditions.

We first notice that the two main channels (n°1 and n°2), which were identified in almost all solutions of Figure A-7, are accurately localized (Figure A-8a). The position and shape of the middle channel (n°3), which were only roughly calibrated on single solutions, are not improved and is even disturbed by secondary channels of solutions in Figure A-7. It means that additional hydraulic conditions have to be used to find this channel. For against, channel n°5 is well located, although it was only partially visible under gradient-like boundary conditions without pumping and invisible with a pumping. Interesting is also that channels n°6 and n°8, which do not appear on Figure A-7, are visible on the right side of Figure A-8a. They are nevertheless not individually identified, but are parts of other structures, which shape has been adapted to integrate these channels. On Figure A-8b are depicted the ratio $\log(T_{\text{channel}}/T_{\text{matrix}})$. Unlike the resulting models of Figure A-7, where the transmissivity of some channels was favored according to the position of the well, the transmissivities are more consistent with the real solution and lead to a equivalent permeability $K_{\text{eq}}=3.07 \times 10^{-5}$ while, for the real solution, $K_{\text{eq}}=6.8 \times 10^{-6}$. It is a good estimate accounting that no flow value are used in the inverse procedure. The distribution of the objective function values, corresponding to all single solutions used in the final model and calculated with gradient-like boundary conditions, shows that most of them have a value smaller than 0.5, even solutions calibrated with different hydraulic conditions. It reinforces the assumption that a channeling structure obtained with a pumping can efficiently match head values under other hydraulic conditions.

In conclusion, using different hydraulic conditions has significantly improved the final model, since each condition provides slightly different information on the channeling structure. Thus, if some channel (e.g. n°4) are still not localized, one should expect that using the other available well as pumping well will provide the required information to localize efficiently all the significant channels.

d) Discussion

The inverse problem applied on highly channeled media is different from the classical inverse problem as the objects that have to be identified are the main discrete structures carrying the flow. Unlike a continuous permeability field, these structures have a particular signature on the head values that depends on their characteristics, i.e. their position and their transmissivity. If only a few structures are relevant in terms of flow, the inverse problem might be easier than for continuous fields, as the difference between heads in the channeled medium and heads in the equivalent homogeneous medium gives a good overview on the underlying transmissivity structure. The fact that flow channels are highly dependent on boundary conditions is a strong handicap in porous media, as their identification requires solving a more than ill-posed problem [*de Marsily et al.*, 1999]. In highly channeled media, against, as flow is focused within a few physical structures that do not depend on boundary conditions, this drawback becomes an advantage, since varying the hydrologic conditions changes the prominence of individual channels in terms of flow and so, changes the information concealed by head data, as shown in the discrepancy maps. When observation wells are available on the field, it is furthermore quite easy to transform one of the well into a pumping well to get valuable data on the transmissivity structure.

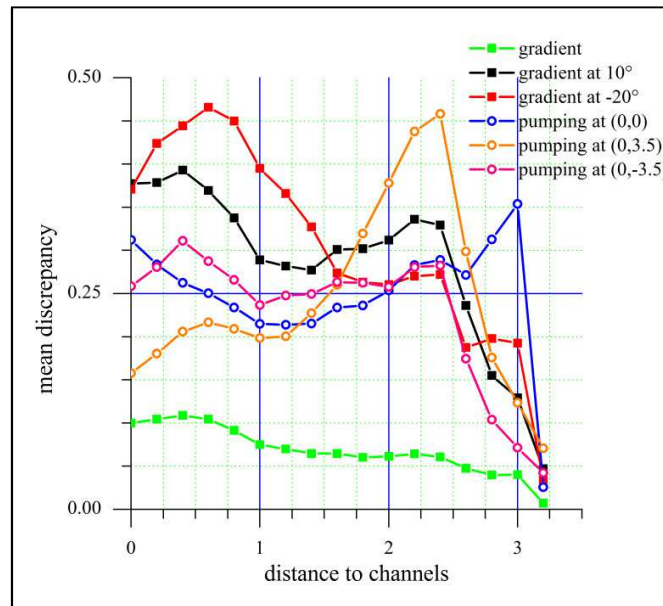


Figure A-9 : Mean difference between head data in the channeled case and in the homogeneous case plotted as a function of the distance between the head position and the nearest channel.

The question of the scale of relevance of head data remains crucial and is related to the characteristic scale of the transmissivity structure. Data too close to each other may contain redundant information on the same channel while too spaced data might miss the zone of influence of another channel. Tiedeman [Tiedeman *et al.*, 1995] studied the influence of data location on the identification of a single fractured zone and concluded that the pumping well should be close to the main channel and that observation wells should cover an area similar to the dimension of the fracture zone. This conclusion is consistent with the observation of the discrepancy maps, insofar as a pumped well favors the localization of channels in the vicinity and observation wells must be located in the zone of influence of main channels. This area of influence can be estimated with variograms derived from discrepancy maps. Figure A-9 depicts the mean discrepancy between the homogeneous and the channeled case as a function of the distance between the head position and the nearest channel. We observe a sudden drop in the channel influence after a distance of 2.5 meters, or 25% of the system size. Even if a deeper analysis is required, it tends to indicate that each piece of data is influenced by only one channel if the distance between two channels is larger than one quarter of the system size. If channels intersect, Figure A-6 shows that the vicinity of the intersection contains information correlated to the position of various channels and then could be used to better characterize the channeling structure connectivity.

The use of information concealed by head data to solve the inverse problem is also a challenging task, since additional information on the physical structures is often required in the classical inverse problem [Zimmerman *et al.*, 1998]. In highly channeled media, the method developed in [Le Goc *et al.*, 2010a] is able to deal with head values to locate channels and possibly characterize their transmissivity. Combining head data obtained with various hydrologic conditions improve significantly the characterization of the transmissivity structure as it takes advantage of the flow sensitivity to the boundary conditions to optimize the amount of information provided by head data. This approach is similar to the method used in the characterization of the flow system with hydraulic data, where various pumping conditions are used to better characterize the transmissivity structure [Le Borgne *et al.*, 2007a]. Further investigation should thus be conducted to estimate the interest of other hydrologic conditions, such as a double pumping test or a coupling injection and pumping.

e) Conclusion

The amount of information concealed by head data is estimated using discrepancy maps, i.e. maps of the difference between heads in a homogenous case and heads in a channeled case, applying the same hydraulic conditions. In highly channeled media, the signature of each main channel is clearly visible, while second-order structures might fall into "shadow zones", i.e. zones that have no influence on head data. Applying a pumping on the field changes the position of these shadow zones and changes the way heads are influenced by the different structures involved in the channeling. Thus, we suggest that varying the hydraulic conditions is an effective way to increase the amount of information available to characterize the transmissivity structure.

Therefore, we calculated head values in steady state on a synthetic test case. These heads were obtained in different hydraulic conditions, i.e. with gradient-like boundary conditions and then with a pumping test made at different location of the field. They are then used as inputs of an inverse problem, that is solved with the method described in [Le Goc et al., 2010a]. The resulting models illustrated a strong correlation between the identified transmissivity structures and the discrepancy maps. Combining the results coming from the different conditions to the limits has successfully improved the characterization of the transmissivity structure, as expected after the observation of the discrepancy maps, showing that head values are relevant data to characterize the transmissivities, especially when various conditions are used.

C. Conclusions générales

Pour définir la structure de la perméabilité dans les milieux où les écoulements sont fortement chenalés, on doit utiliser les données hydrauliques dans la mesure où les observations directes de la perméabilité ne permettent pas de caractériser les quelques structures principales dont dépendent les propriétés hydrauliques du milieu. Pour résoudre ce problème inverse il est nécessaire de gérer à la fois la paramétrisation, c'est-à-dire la manière dont on représente le milieu, et les données disponibles.

Tout d'abord, la paramétrisation se doit d'être cohérente avec les propriétés physiques du milieu. Si celui-ci contient des éléments très transmissifs qui créent de forts écoulements à travers tout le milieu, ce qui arrive fréquemment dans les milieux fracturés, alors la paramétrisation doit permettre de représenter ce type d'élément pour que le modèle résultant soit pertinent. C'est pourquoi nous proposons de modéliser les chenaux d'écoulement par des structures très transmissives et connectées par construction soit aux limites du système, soit entre elles. De plus, la quantité d'éléments pouvant être représentés dépend de la quantité d'information disponible. C'est pourquoi nous avons privilégiée une procédure itérative qui augmente, tant que les données le permettent, le nombre de paramètres du modèle. Finalement, même si la paramétrisation est adaptée, la nature mal posée du problème inverse rend difficile la détermination d'un modèle unique permettant de simuler de manière cohérente les propriétés hydrauliques. Au lieu d'augmenter arbitrairement le temps de calcul pour tenter d'améliorer ce modèle, nous proposons de résoudre le problème inverse plusieurs fois à partir de plusieurs conditions initiales et d'analyser les résultats obtenus pour identifier les caractéristiques les plus pertinentes et de les utiliser pour définir le modèle final.

Les données utilisées pour calibrer le modèle doivent contenir une information pertinente concernant les caractéristiques principales de la chenalisation. Ces données doivent de plus être sensibles à une échelle suffisamment large pour permettre la détection de structures non directement connectées au puits d'observation. On a vu que les données de charge

remplissaient ces conditions et qu'il était de surcroit possible d'améliorer la quantité d'information disponible en variant les conditions d'échantillonnage. Mesurées dans les conditions naturelles, elles induisent la présence de zones d'ombre dans lesquelles les chenaux d'écoulement sont invisibles. Mais si on combine ces données avec les charges mesurées lors d'une sollicitation du milieu, il est possible de réduire voire d'éliminer ces zones d'ombre afin de caractériser la majorité des structures principales d'écoulement.

Enfin, la méthode reliant données et paramètres repose sur une fonction objectif la plus sensible possible aux éléments recherchés. Si l'ajout d'un terme de régularisation est souvent utilisé et a prouvé son efficacité, on a montré qu'une pondération des données en fonction de leur sensibilité est beaucoup plus efficace et permettait d'éviter de sur-paramétrer les zones les moins sensibles. La fonction-objectif doit pour finir être optimisée par une méthode capable de gérer les propriétés liées à la forte hétérogénéité du milieu, comme la non-convexité et la non-régularité. Nous avons proposé d'utiliser un algorithme de recuit simulé dans la mesure où celui-ci est capable de trouver le minimum global même dans ces conditions défavorables. L'utilisation d'une autre méthode, comme un algorithme évolutionnaire, n'est pas à exclure puisque si le recuit simulé est efficace, il demande malgré tout d'évaluer un grand nombre de fois la fonction-objectif. Il présente néanmoins l'avantage d'être facilement adaptable au contexte des milieux fortement hétérogènes.

Les travaux réalisés pendant le doctorat se placent dans le contexte plus général de l'étude des systèmes complexes. Afin de comprendre les caractéristiques de ces systèmes, trois modes d'investigation sont possibles : l'approche théorique, l'approche expérimentale et l'approche computationnelle. Cette dernière approche, que nous avons utilisée, consiste en la conception de modèles numériques. Elle repose sur le fait que représenter un phénomène à grande échelle est indépendant de sa description à petite échelle. Par contre, il est nécessaire d'utiliser le niveau adéquat de représentation pour caractériser le phénomène qui nous intéresse. La question centrale est donc « quelles sont les caractéristiques pertinentes qui doivent être modélisée ? » [Goldenfeld and Kadanoff, 1999].

Nous proposons dans cette partie de discuter les différents travaux présentés précédemment en prenant pour cadre général la définition d'un modèle représentatif d'un système complexe, ici constitué des écoulements dans les milieux hétérogènes. Dans un deuxième temps, nous reprendrons les résultats obtenus afin de discuter leur pertinence et leurs éventuelles limitations. Ces deux discussions permettront d'envisager les perspectives possibles du travail accompli. Finalement, une conclusion générale mettra un point final à ce mémoire.

A. De la caractérisation des milieux complexes

Un modèle repose sur trois caractéristiques [Kühne, 2006] : un phénomène naturel, une sélection de ses caractéristiques pertinentes et la capacité du modèle à se substituer au phénomène dans le cadre d'une étude plus générale. Ces trois caractéristiques reposent sur les trois composantes que sont le milieu naturel, la paramétrisation et les observations. La figure 6-1 est une schématisation des interactions entre ces trois composantes.

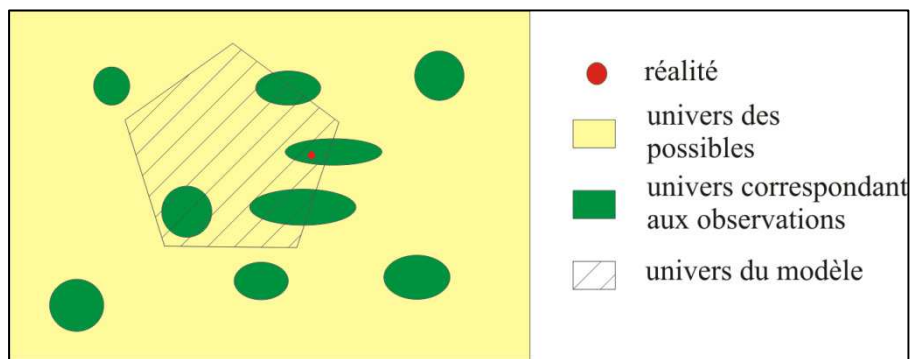


figure 6-1 : Schématisation des paramètres à prendre en compte lorsque l'on cherche à définir un modèle pertinent. En jaune est représenté l'espace des possibles, c'est-à-dire l'ensemble des possibilités de représentation d'un phénomène. Plus le phénomène est complexe, plus cet espace est vaste. En vert est représentée la sous-partie de l'espace qui peut correspondre aux observations du milieu. Plus il y a d'information disponible, plus cette partie est localisée. La surface hachurée représente l'ensemble des solutions que la paramétrisation choisie permet de représenter. Le modèle défini appartiendra à une intersection entre la zone hachurée et une des zones vertes. Un modèle pertinent doit contenir la réalité (point rouge) et couvrir une surface la plus restreinte possible autour de ce point.

La première chose à prendre en compte est la nature du phénomène que l'on cherche à modéliser. Plus le phénomène est complexe, plus l'« univers des possibles » est vaste. Si la figure 6-1 est une représentation en 2 dimensions, les univers décrits sont hautement multidimensionnels. Par exemple, si un milieu homogène peut être caractérisé par une seule

valeur, un milieu hétérogène devra être défini par un très grand nombre de caractéristiques. Parmi cet « univers des possibles », la réalité, c'est-à-dire le phénomène que l'on cherche à comprendre, représente un volume infinitésimal. Cependant c'est ce volume, et ce volume seulement, qui nous intéresse. Pour le trouver, on ne peut s'appuyer que sur la connaissance que l'on a du milieu et un choix de paramétrisation. Premièrement, les données disponibles définissent un « univers des observations » correspondant aux sous-parties de l'espace des possibles qui permettent de reproduire les données observées. A l'intérieur de ces sous-parties se trouve la réalité, mais la quantité de données disponibles est généralement très insuffisante pour la déterminer avec précision et la quantité de représentations acceptables en ce qui concerne la concordance avec les observations reste très importantes, surtout si le phénomène est complexe. Deuxièmement, en définissant une paramétrisation, on définit un « univers des modélisables » qui, en fonction du nombre de paramètres et des contraintes additionnelles sur ces paramètres, peut avoir une forme plus ou moins régulière et être plus ou moins étendu. L'intersection entre « l'univers des modélisables » et « l'univers des observations » doit contenir la réalité pour que le problème soit résoluble. Mais même dans ce cas, les possibilités peuvent être encore trop nombreuses et certaines d'entre elles peuvent même très éloignées de la réalité. Le risque reste donc important de définir un modèle qui, bien que cohérent avec les données disponibles, ne correspond pas du tout au milieu naturel et donnera donc des prédictions complètement fantaisistes. Il est en effet impossible de déterminer a priori si le modèle obtenu est proche du milieu naturel ou au contraire très éloigné. Cependant, la question de savoir si la véracité du modèle est quelque chose d'essentiel ou si la reproductibilité des données observées suffit reste ouverte et dépend grandement des exigences du modélisateur.

Si l'on se ramène à la caractérisation des écoulements dans les milieux hétérogènes, la propriété recherchée, i.e. la perméabilité, présente une complexité structurelle importante liée à son hétérogénéité à toutes les échelles, ce qui signifie que « l'univers des possibles » est vaste. Les travaux présentés aux différentes parties de cet ouvrage cherchent donc à trouver le moyen de déterminer le mieux possible la position de la « réalité » dans cet univers.

1. Analyse des données et contrainte sur la paramétrisation

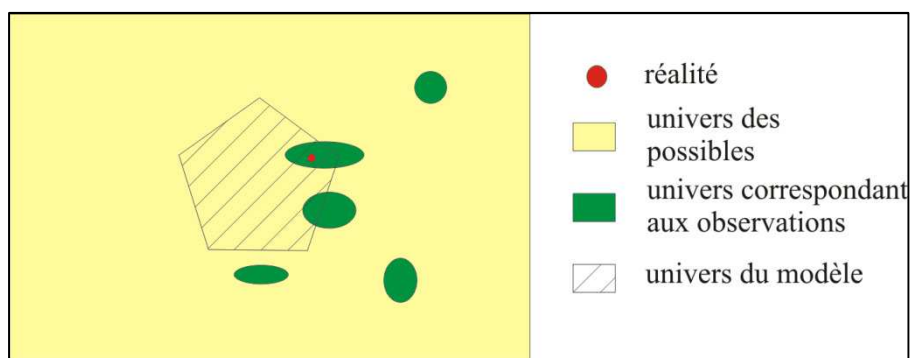


figure 6-2 : Schématisation des objectifs recherchés lors de la définition du modèle de fracturation des sites suédois de Forsmark et Laxemar décrite à la partie 3. Par rapport à la figure 6-1, on cherche à contraindre le modèle afin de réduire l'univers des modélisables et à mieux définir, en intégrant la variabilité et l'incertitude sur les données, à mieux définir l'univers des observations.

La première solution et l'une des plus intuitives consiste à se concentrer sur ce que l'on connaît du milieu. Les travaux présentés en partie 3 présente l'analyse des données

disponibles pour définir un modèle de fracturation, comme les longueurs et les orientations mesurées sur des affleurements et le long de puits d'observation. Cette analyse cherche à remplir deux objectifs. Le premier est de mieux définir l'« univers des observations » en intégrant l'incertitude et la variabilité des propriétés observées. Le second est de contraindre le modèle afin de réduire l'« univers des modélisables » (figure 6-2).

L'avantage, dans l'étude proposée, est que les données disponibles sont nombreuses et de bonne qualité. Mais même dans ce cas, la vision que l'on peut avoir du milieu naturel reste partielle et peut même être parfois erronée. Ce problème provient principalement de trois facteurs. Tout d'abord, chaque donnée est sujette aux incertitudes liées à l'échantillonnage. À l'incertitude liée aux appareils de mesure et aux éventuelles erreurs lors du relevé s'ajoutent en effet les incertitudes inhérentes aux données elles-mêmes. Celles-ci peuvent avoir été altérées par le temps ou modifiées par la topographie. C'est notamment le cas des traces de fractures qui peuvent avoir été déconnectées, ce qui se traduit par le relevé de deux traces disjointes là où il n'existe qu'une trace unique. La conséquence sur la distribution des traces de fracture est que l'exposant estimé de la loi de puissance sous-jacente est plus grand que l'exposant réel, ce qui revient à surestimer le rôle des petites fractures par rapport aux grandes fractures. La fenêtre d'observation étant limitée, la probabilité d'observations des plus grandes fractures est réduite et le fait que certaines traces de fracture peuvent être coupées produit les mêmes effets. C'est pourquoi nous avons développé un ensemble de méthodes destinées à réduire ces biais d'observations afin de travailler sur des données qui reflètent le mieux possible la structure réelle du réseau de fracture.

Le deuxième facteur expliquant que les données observées peuvent donner une vision floue de la réalité vient de la variabilité du milieu naturel. En effet, des observations faites à la même échelle peuvent donner l'impression que l'on mesure des milieux aux propriétés différentes alors qu'il s'agit en fait du même milieu dont les caractéristiques varient spatialement. Plutôt que de définir deux types de modèles pour ces données, ce qui revient à multiplier le nombre de zones vertes à la figure 6-2, il est nécessaire d'évaluer cette variabilité afin de l'intégrer dans l'analyse. Une fois la variabilité (et l'incertitude) prise en compte, il est possible de comparer deux jeux de données et de déterminer si ces données sont identiques ou non, c'est-à-dire si elles appartiennent au même modèle sous-jacent.

Finalement, le troisième facteur pouvant altérer la vision du milieu naturel au travers des observations vient de la nature de l'hétérogénéité de la perméabilité. En plus d'être variable spatialement, elle est également multi-échelle, c'est-à-dire que les observations faites à une échelle donnée, par exemple à l'échelle d'un affleurement, peuvent être différentes des observations faites à plus grande échelle. Là encore, cela ne signifie pas forcément que l'on mesure deux choses différentes, ces données pouvant en effet être expliquées par le même modèle sous-jacent. Dans le cas des milieux fracturés dont les longueurs suivent une loi de puissance, la loi d'échelle est représentée par l'exposant de la loi de puissance. Comme cet exposant est particulièrement sensible aux biais d'échantillonnage, les méthodes de correction mises en place sont indispensables pour effectuer le changement d'échelle. Si la loi d'échelle sous-jacente au milieu naturel est correctement décrite, il est alors possible de lier entre eux les modèles locaux, déduits des observations à l'échelle d'un puits. Cette liaison est naturellement dépendante de la variabilité du milieu, d'où l'importance de sa prise en compte. Dans le cas des propriétés de fracturation, la comparaison s'effectue au niveau du terme de densité en intégrant la distribution des orientations. La méthodologie que nous avons mise au point repose sur un algorithme d'agglomération itératif permettant de définir des classes de fracturation regroupant plusieurs modèles locaux. Cette classification ne correspond pas, dans notre cas, à un modèle approché plus simple mais au contraire à un modèle exact, en termes probabilistes, de la nature du milieu étudié.

La correction de ces trois facteurs comporte deux avantages. Tout d'abord, il est plus simple de reproduire les données observées lorsque celles-ci sont épurées du maximum d'erreurs. De plus, il est possible de définir le modèle sous-jacent à une échelle supérieure en intégrant les conséquences de l'hétérogénéité. Ceci se traduit concrètement, si l'on se réfère à la figure 6-2, par une suppression des zones vertes provenant uniquement des biais d'observation et par une réduction de la taille de ces zones dont une partie provient de la mauvaise intégration des effets d'échelle et de la variabilité. Deuxièmement, la paramétrisation pourra être définie de manière à produire des modèles respectant les propriétés observées, notamment en ce qui concerne la distribution des tailles de fracture ou la distribution des orientations. Ces nouvelles contraintes permettent de réduire la taille de la zone hachurée ce qui se traduit par une diminution du nombre de modèles possibles et permet de déterminer plus facilement le modèle pertinent correspondant à la réalité.

2. Ajout d'indicateurs sur les propriétés du milieu

Après l'analyse des données disponible, le moyen de réduire le nombre de modèles possibles est de définir des indicateurs caractérisant les propriétés les plus pertinentes du milieu naturel (figure 6-3). Ces indicateurs permettent, en plus de la comparaison avec les données directes sur le milieu, d'évaluer la pertinence des modèles à partir d'une information indirecte basée sur les effets de l'hétérogénéité du milieu. C'est dans ce cadre qu'on a défini les indicateurs de chenalisation de la partie 4.

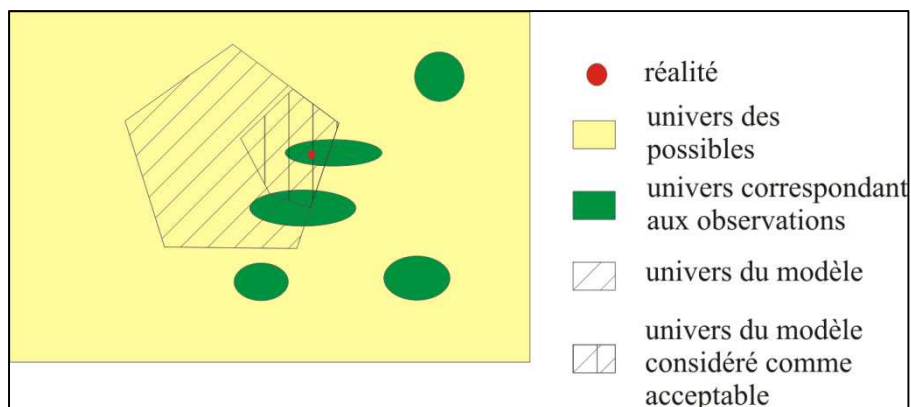


figure 6-3 : Schéma du rôle des indicateurs de chenalisation dans la détermination d'un modèle pertinent. A partir des caractéristiques de la chenalisation du milieu étudié (distance entre chenaux, longueurs caractéristique des chenaux), il est possible d'exclure un grand nombre de modèles dont les caractéristiques sont différentes afin de ne garder que les modèles conformes.

Les modèles que l'on cherche à définir dans le cadre des milieux fortement hétérogènes servent principalement à caractériser les écoulements afin d'en déduire les propriétés hydrauliques. La principale source de la complexité est constituée par les discontinuités provoquées par la chenalisation des écoulements. Les indicateurs mis au point à la partie 4 visent à caractériser cette chenalisation. On considère alors la distance caractéristique entre chenaux, reliée à l'hétérogénéité du milieu, et la longueur caractéristique des chenaux, reliée à la nature de la corrélation entre les fortes valeurs de perméabilité. En effet, les milieux étudiés ont deux propriétés importantes. La première est l'hétérogénéité entre les valeurs de perméabilité. Dans la mesure où les écoulements privilégient les chemins reliant les fortes valeurs, plus le milieu est hétérogène plus le nombre de chenaux est faible. Ceci se traduit par une augmentation de la distance entre deux chenaux mesurée par le premier indicateur D_{ic} . La deuxième propriété dépend de la structure même du milieu. La perméabilité n'est en effet pas

aléatoire mais organisée et la connectivité entre ses fortes valeurs peut profondément modifier les propriétés hydrauliques. Si les fortes perméabilités sont bien connectées, les chemins d'écoulements seront mieux définis et produiront un flux important sur une grande distance, caractérisée par D_{cc} . A partir de ces indicateurs, il est donc possible d'estimer la nature de la chenalisation du milieu considéré et d'éliminer les modèles dont les caractéristiques sont différentes dans la mesure où, même s'ils permettent de reproduire les perméabilités ou les charges hydrauliques observées, leurs propriétés hydrauliques seront vraisemblablement trop différentes pour fournir des prédictions fiables.

Ces indicateurs peuvent également être utilisés pour contraindre le modèle. En effet, il existe un lien entre les propriétés en matière de perméabilité et les caractéristiques de la chenalisation, tant pour des milieux poreux que pour des milieux fracturés. Si ce lien est connu, il est alors facile de ne paramétrer que des modèles dont les propriétés sont par construction proches des propriétés du milieu naturel. Cette méthodologie n'est bien sûr valable que si les indicateurs utilisés sont pertinents vis-à-vis des objectifs du modèle. Lorsque l'on cherche à déterminer les propriétés hydrauliques, qui dépendent des écoulements, utiliser des indicateurs comme D_{ic} et D_{cc} revient à ajouter de l'information indirecte sur les spécificités du phénomène qui nous intéresse.

3. L'importance de la paramétrisation.

Au-delà des données directes et indirectes sur la structure du milieu étudié, un modèle pertinent s'appuie sur une paramétrisation capable de représenter les particularités du milieu, comme les structures sous-jacentes aux chenaux d'écoulement. C'est ce type de paramétrisation qui est utilisée à la partie 5 et calibrée en résolvant le problème inverse. L'objectif est de réduire l'« univers des modélisables » en ajoutant des contraintes sur les paramètres (figure 6-4). Dans notre cas, ces contraintes concernent la connectivité des structures recherchées ; chaque chenal doit être connecté soit aux bords du système soit à un autre chenal. De plus, on impose que les structures recherchées soient localisées et permettent de maintenir de forts écoulements à travers le milieu. Ces contraintes correspondent à la manière dont on a défini un chenal d'écoulement à partir des indicateurs de chenalisation de la partie 4. Le choix d'une paramétrisation repose donc essentiellement sur la connaissance que l'on a du milieu et de ses propriétés supposées.

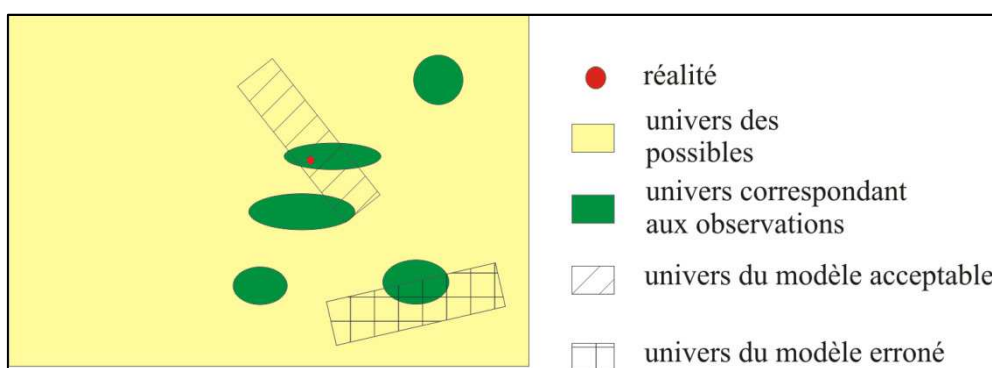


figure 6-4 : Schématisation du rôle de la paramétrisation dans la détermination d'un modèle pertinent. Si la paramétrisation intègre efficacement les caractéristiques principales du milieu étudié (zone avec les hachures diagonales), alors la résolution du problème inverse, qui vise à obtenir un modèle proche de la réalité, est facilitée. Par contre, si les caractéristiques utilisées ne sont pas pertinentes (zones avec les hachures carrées), le risque est d'obtenir un modèle très différent de la réalité tout en étant capable de reproduire fidèlement les observations.

L'idée est ensuite de calibrer les paramètres à partir des données disponibles afin de retrouver la réalité. Pour cela, on peut utiliser l'intégralité des informations extraites du milieu naturel,

du modèle de fracturation défini à la partie 3 aux données hydrauliques en passant par les indicateurs de chenalisation de la partie 4. In fine, dans la mesure où toute l'information disponible restera insuffisante par rapport à la complexité du milieu, c'est la paramétrisation qui permettra de faire concorder toutes ces observations. Son choix est donc crucial, dans la mesure où une mauvaise paramétrisation peut malgré tout permettre de reproduire les observations sans pour autant que le modèle corresponde au milieu étudié. Cette difficulté est illustrée par la zone avec les hachures carrées de la figure 6-4. S'il existe un jeu de paramètres permettant de reproduire les observations, alors il n'y a aucun moyen de distinguer la mauvaise paramétrisation de la bonne. La seule différence se fera sur les prédictions issues de ces modèles ou par l'ajout de nouvelles données qui viendront contredire l'un des modèles. Cette difficulté, illustrée par l'étude de [Bredehoeft, 2005], est un des inconvénients de l'utilisation de modèles et il n'existe pas de moyen sûr de résoudre cette difficulté. C'est pourquoi à la partie 5, les modèles obtenus ont été validés non seulement sur la valeur de la fonction-objectif, qui définit l'écart entre les valeurs simulées et les valeurs observées, mais également en comparant le modèle obtenu avec la solution de référence. Cette solution est connue dans la mesure où la méthode est testée sur des cas synthétiques mais n'est pas disponible pour les cas réels. Dans ce cas, une solution consiste à exclure certaines données lors de la calibration et de les utiliser pour vérifier le modèle a posteriori. Cependant le fait d'exclure des données de la calibration est un luxe, au vu de la quantité d'information généralement disponible.

4. La calibration des paramètres à partir des données.

Les paramètres du modèle sont calibrés à partir des données disponibles, des propriétés indirectes incluses dans la paramétrisation et des indicateurs éventuels. Il s'agit ici d'identifier l'intersection entre l'« univers des modélisables » et l'« univers des observations ». Cette dernière étape est tout aussi importante que les autres dans la mesure où elle peut ajouter de l'incertitude et est le plus souvent difficile à mettre en œuvre. Si le nombre de paramètres est peu important, il est toujours possible de générer un grand nombre de modèles afin de ne conserver que celui (ou ceux) présentant une bonne concordance avec les observations. Cette méthode est néanmoins hasardeuse dans la mesure où il est très difficile de générer suffisamment de modèles de manière à échantillonner efficacement l'espace des paramètres. C'est pourquoi les méthodes basées sur la résolution du problème inverse ont été développées. En hydrogéologie, le problème inverse est mal posé, non-linéaire et sous-déterminé. Il existe donc beaucoup de méthodes pour le résoudre, chacune possédant ses avantages et ses inconvénients. Si l'on étudie ces méthodes de plus près, la spécificité de chacune réside dans les propriétés indirectes qu'elles cherchent à respecter. La plupart des méthodes géostatistiques par exemple cherchent à trouver le modèle le moins hétérogène possible, considérant que l'hétérogénéité crée des fluctuations locales qui peuvent être traitées de manière globale. Certaines méthodes, pour être plus efficaces, traitent l'hétérogénéité comme une composante à part, comme dans le cas des points-pilotes.

L'approche que nous avons proposée à la partie 5 prend le contre-pied de ces méthodes dans la mesure où on cherche à calibrer essentiellement les discontinuités en simplifiant la composante continue. Cette approche est efficace à condition de bien identifier ces structures. En effet, si une approximation dans les propriétés générales peut n'avoir qu'une conséquence mineure sur le résultat, se tromper sur une des structures les plus transmissives provoque immédiatement une erreur importante sur le modèle. Afin de limiter ce risque d'erreur, deux processus ont été mis en place. Tout d'abord la paramétrisation est itérative. L'objectif est

alors de détecter une éventuelle mauvaise calibration le plus tôt possible afin d'éviter de rajouter des paramètres qui n'amélioreront pas le modèle. Cette approche itérative permet d'éviter l'ajout de chenaux non pertinents dans la paramétrisation mais limite également le nombre de paramètres que l'on peut identifier puisque dès qu'un chenal est mal calibré, la méthode s'arrête. C'est pourquoi le deuxième processus consiste à résoudre plusieurs fois le problème inverse et à analyser les modèles obtenus avec l'idée que, si pour certains d'entre eux l'identification s'est arrêtée prématurément, d'autres contiendront les chenaux absents.

Le deuxième intérêt de calculer plusieurs modèles et de les analyser pour identifier les caractéristiques pertinentes est qu'il est alors possible d'évaluer l'incertitude associée au modèle final. Cette incertitude est liée à l'incertitude sur les données, à l'approximation liée à la paramétrisation choisie et à la méthode d'inversion elle-même. Ainsi, au lieu d'obtenir un modèle unique, supposé correspondre à la réalité, on obtient un modèle moyen dont les paramètres comportent une incertitude. C'est ce qui est illustré à la figure 6-5 où le résultat obtenu au final, n'est pas représenté par un point mais par une zone représentant l'ensemble des solutions possibles.

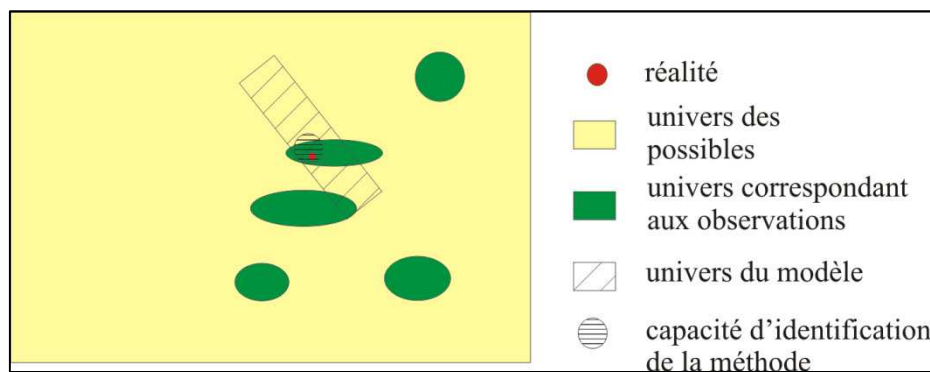


figure 6-5 : Schématisation du rôle de la méthode d'inversion. Même si la paramétrisation est bien choisie et que l'on a suffisamment de données pour paramétrer le système, il existe toujours une incertitude liée à la méthode elle-même dont il faut tenir compte.

Évaluer l'incertitude associée au modèle final est généralement quelque chose de difficile. S'il est parfois possible de calculer l'incertitude sur les paramètres à partir de l'incertitude sur les données et du schéma numérique utilisé, les incertitudes liées à la modélisation elle-même sont difficiles à quantifier [Moore and Doherty, 2006]. L'analyse proposée, si elle est basée sur un nombre significatif de solutions, permet d'estimer cette incertitude.

5. Résumé des différentes approches.

En conclusion, tous les travaux présentés ici cherchent à prendre en compte l'hétérogénéité des milieux naturels. Cette prise en compte peut tout d'abord intervenir au niveau des observations. Il convient alors d'analyser ces données en corrigeant les erreurs contenues dans les mesures et d'estimer la variabilité du milieu. Il est ensuite possible d'interpréter à l'échelle du site les données mesurées localement via un changement d'échelle. Deuxièmement, l'hétérogénéité de la structure du milieu provoque une hétérogénéité des propriétés hydrauliques. Il est donc important de caractériser cette hétérogénéité grâce à des indicateurs. Ces indicateurs peuvent ensuite être intégrés au modèle final soit en conditionnant ses propriétés soit en évaluant a posteriori sa pertinence. Troisièmement, l'hétérogénéité doit être prise en compte dans la structure même du modèle grâce à une paramétrisation pertinente. Il est en effet plus facile de calibrer un modèle dont la paramétrisation est capable de représenter les caractéristiques principales du milieu. Finalement, la méthode de calibration elle-même doit faire face à l'hétérogénéité et la définition de familles de modèle permet d'intégrer cette hétérogénéité dans les résultats.

B. Discussion sur les méthodes proposées et les résultats obtenus

Les thématiques abordées dans ce mémoire ont pour but une meilleure prise en compte de la complexité du milieu naturel. L'objectif de ce paragraphe est de discuter les résultats obtenus.

1. Définition d'un modèle stochastique de fracturation

La partie 3 est consacrée à la définition d'un modèle 3D de la fracturation de sites naturels. Le seul moyen de définir ce modèle est de traiter les données disponibles. On a vu néanmoins que ces données pouvait contenir des biais qui, non corrigés, altèrent la vision que l'on peut avoir des propriétés du milieu naturel. Si la méthode proposée permet de corriger les biais liés à la déconnexion des traces de fracture et à la sous-représentation des plus grandes fracture, on peut légitimement se poser la question de savoir si d'autres sources de biais sont possibles. Cette question est importante dans la mesure où, en corrigeant les biais identifiés, l'exposant de la loi de puissance a été modifié de l'ordre de 12%. La première cause de biais (biais « géométrique ») provient d'une incompatibilité entre la géométrie de la surface d'échantillonnage et la géométrie des objets mesurés. S'il existe de nombreuses études portant sur l'influence de ces biais et proposant des méthodes pour les corriger, toutes ses méthodes reposent sur une géométrie supposée des fractures ou sur une propriété statistique assumée. S'il n'existe aucune méthode universelle, l'objet fracture est en lui-même quelque chose de complexe et il est vraisemblable que les biais d'observation restent mal connus, malgré tous les efforts accomplis. De surcroît, à ces biais géométriques se superposent des « biais environnementaux » plus difficiles à quantifier. Ces biais sont donc plus rarement pris en compte lors de l'analyse des milieux naturels. Dans l'étude présentée à la partie 3, les biais géométriques sont corrigés par des méthodes classiques prenant en compte la distribution et la géométrie supposée des fractures. Cette correction des effets de taille finie compte pour moitié dans la modification de la valeur de l'exposant de la loi de puissance. L'autre moitié provient des biais environnementaux. Ces biais sont, dans notre cas, corrigés par une méthode qui ne fait aucune supposition sur la distribution sous-jacente, ce qui permet de penser que c'est une méthode de correction efficace. Rien n'empêche cependant de penser que d'autres biais restent présents et que leur prise en compte pourrait substantiellement modifier les résultats obtenus. On ne peut malheureusement pas forcément compter sur la quantité de donnée pour compenser les biais inconnus. En effet, si ces biais sont systématiques, en faveur des grandes fractures ou des fractures verticales par exemple, augmenter le nombre de données ne donne pas une vision plus réaliste du phénomène étudié. Les biais non systématiques augmentent la variabilité visible du phénomène qui, par contre, peut être prise en compte lors de la définition du modèle.

Les modèles utilisés pour décrire les données sont des modèles stochastiques, basés sur une loi de probabilité pour les longueurs de fractures et pour les orientations. Les modèles stochastiques permettent généralement de bien reproduire les données observées mais leur pertinence est limitée par le fait qu'ils n'intègrent généralement pas les propriétés organisationnelles du milieu. Par exemple, un modèle stochastique de réseau de fracture peut avoir, en moyenne, une connectivité très différente des réseaux naturels [Andresen *et al.*, 2008] parce qu'il ne prend pas en compte les corrélations entre paramètres. Comme ces corrélations peuvent être indirectes, c'est-à-dire non mesurable par une loi de régression entre deux paramètres, elles sont difficiles à détecter et donc à intégrer au milieu. La loi d'échelle dérivée de [Davy *et al.*, 2009] en est un exemple flagrant. Ce modèle introduit une corrélation entre la position d'une fracture par rapport à ses voisines et sa longueur, corrélation décrite par une loi simple mais difficile à caractériser statistiquement. Elle pourrait en effet se traduire

par plusieurs types de corrélations statistiques comme une corrélation entre la longueur d'une fracture et la longueur de ses voisines ou une corrélation entre la densité locale autour d'une fracture et sa longueur, etc. C'est pourquoi un modèle stochastique purement descriptif n'est pas forcément pertinent pour modéliser un phénomène, même s'il parvient à reproduire les données observées, dans la mesure où certaines de ses propriétés peuvent ne pas être respectées. Ces propriétés peuvent être liées soit à la structure du milieu étudié, comme sa connectivité, où au rôle joué par certains éléments statistiquement rares mais très influents. On peut donc supposer que le modèle purement descriptif présenté à la partie 3, basé sur une distribution des longueurs et des orientations, est moins efficace pour caractériser les propriétés hydrauliques que le modèle « amélioré » décrit par [Davy *et al.*, 2009] et qui décrit l'organisation des fractures entre elles.

2. Définitions d'indicateurs sur la chenalisation des écoulements

Utiliser des indicateurs supplémentaires permet d'intégrer, dans un modèle stochastique, les propriétés du milieu qui ne sont pas décrites par les paramètres du modèle. La question est donc de savoir quelles propriétés doivent avoir les indicateurs pour améliorer les capacités de représentation du modèle. Les indicateurs D_{ic} et D_{cc} , présentés à la partie 4 ont été définis de manière à caractériser les propriétés géométriques des chenaux d'écoulement. Ils sont donc basés sur l'organisation des écoulements bien que calculés à partir de leur distribution. Il existe en effet des indicateurs uniquement basés sur la distribution de la perméabilité, comme la moyenne, la variance et même le ratio de participation, et des indicateurs basés sur la structure, l'organisation des valeurs de perméabilité. Le premier type d'indicateur a été largement utilisé pour décrire les propriétés physiques des milieux hétérogènes afin d'en déduire les propriétés hydrauliques. L'inconvénient de ces indicateurs est de ne fournir qu'une information à très grande échelle et à négliger toutes les propriétés plus locales qui ont pourtant un rôle déterminant dans les propriétés hydrauliques. C'est pourquoi l'utilisation d'indicateurs du second type est de plus en plus fréquente afin de pallier les limitations des premiers indicateurs. Le principal support théorique pour le développement d'indicateurs sur la structure de la perméabilité est la statistique multipoint, qui cherche à quantifier ses corrélations locales. Cependant, comme l'ont montré différentes études [Western *et al.*, 1998; Knudby and Carrera, 2005], même la statistique multipoint a ses limitations. En fait, dans l'analyse des propriétés hydrauliques, la direction dans laquelle la statistique est faite joue un rôle important. C'est pourquoi D_{ic} et D_{cc} présentent un avantage déterminant. Comme ils sont basés sur les propriétés lagrangienne du flux, ils quantifient des propriétés dans la direction du flux, sans pour autant recourir à des artifices de constructions. Un autre avantage est qu'au contraire de certains indicateurs qui fournissent une grandeur statistique difficilement interprétable, ils peuvent être considérés comme la distance caractéristique entre deux chenaux principaux d'écoulement et la longueur caractéristique des chenaux où les écoulements sont significatifs. Ces valeurs ont une réalité physique et sont donc intimement liées à la structure de la perméabilité. Cette réalité est de plus valable tant pour les milieux poreux et les milieux fracturés. D_{ic} et D_{cc} peuvent donc servir pour caractériser un milieu, mais également permettre de remplacer un milieu complexe par un modèle simple tout en respectant les propriétés hydrauliques. En effet, si les milieux poreux et les milieux fracturés ont des structures différentes et que leurs propriétés hydrauliques sont gouvernés par des paramètres différents, ils pourraient servir de critères communs permettant de comparer, voir de substituer, une représentation par une autre.

3. Résolution du problème inverse

La nature même des systèmes complexes rend la résolution du problème inverse difficile. Compte tenu de la quantité d'information disponible, il est nécessaire de bien poser le

problème et de déterminer quelles sont les caractéristiques du milieu que l'on cherche à identifier et quelles sont les propriétés que l'on va utiliser pour réussir cette identification.

Pour l'identification des structures responsables de la chenalisation des écoulements fracturés (Partie 5), nous avons émis l'hypothèse que les écoulements étaient dominés par certaines structures principales dont l'organisation hiérarchique permettait de négliger les éléments les moins significatifs. C'est sur cette hypothèse qu'ont été conçue à la fois la paramétrisation basée sur les chenaux d'écoulement et la méthode d'identification itérative. Cette hypothèse est appuyée par les simulations d'écoulement dans les milieux chenalisés et par la littérature. Néanmoins, si la méthode est particulièrement efficace sur les milieux synthétiques qui remplissent les conditions de l'hypothèse, on peut se poser la question de sa justesse dans les milieux où l'hypothèse n'est pas valable. Tout d'abord, si plusieurs structures ont exactement la même importance en termes d'écoulement, la méthode commencera par identifier l'une ou l'autre de ces structures. Si ces structures sont déconnectées, alors l'identification d'une d'entre elle correspond à un minimum local de la fonction-objectif. Dans ce cas, il n'y aura pas de problème et le choix du minimum par la méthode dépendra des conditions initiales. Par contre, si les structures sont connectées ou s'il existe un minimum local intermédiaire correspondant à l'identification d'une structure différente permettant de reproduire les effets des deux structures, il est possible que la méthode commence par ajouter cette structure à la paramétrisation. Dans ce cas, soit le modèle final contiendra cette structure superflue, soit, à l'étape suivante, cette structure sera recalibrée vers une des structures existantes grâce à l'ajout d'un nouveau chenal. Dans ce cas, la quantité respective de chacun des modèles dans les solutions individuelles ainsi que les valeurs de fonction-objectif obtenues détermineront quelles structures seront conservées après la phase de post-traitement. Le même genre de difficulté apparaît si le nombre de structures non négligeables est important. En effet, la méthode proposée peut difficilement identifier simultanément plus de trois ou quatre chenaux d'écoulement. Là encore, si plusieurs chenaux ont une importance similaire, chaque résolution pourra identifier des chenaux différents en fonction des conditions initiales et c'est la méthode d'analyse qui triera les résultats les plus pertinents. Finalement, le modèle proposé remplace toutes les structures considérées comme moins importantes par une matrice homogène. Néanmoins, il est possible que les structures secondaires, bien que négligeables indépendamment, aient ensemble une influence non-négligeable sur les flux. Dans la méthode proposée, cette influence ajoutera dans le meilleur des cas de l'incertitude aux résultats obtenus et dans le pire des cas provoquera la présence de chenaux superflus destinés à compenser ces effets. Une fois de plus, tout dépendra du nombre de solution individuelle et de la capacité de la méthode d'analyse à ne conserver que les chenaux pertinents. Ces constatations montrent que plus la structure recherchée s'éloigne des hypothèses posées ou devient complexe, plus la méthode d'analyse décrite à la partie 5.3.6 devient critique. La méthode proposée est basée sur une technique d'agglomération des solutions dépendant de la distance physique entre les chenaux identifiés et leur poids respectif. Le poids associé à chaque chenal est ici défini à partir de la valeur de la fonction-objectif obtenue par la solution contenant le chenal et à sa sensibilité. L'utilisation de critères plus proches des propriétés recherchées ou basée sur une étude statistique plus poussée devrait permettre d'améliorer les capacités d'identification des méthodes dans les cas les plus difficiles.

Si, comme le suppose [Knudby and Carrera, 2005], les propriétés de chenalisation sont différentes pour les écoulements et pour le transport, alors toute la pertinence de la méthode repose sur les données utilisées. On a en effet montré que les valeurs de charge hydraulique étaient sensibles aux chemins d'écoulements préférentiels, ce qui permet à la méthode

d'inversion d'identifier ces chenaux. Si l'objectif du modèle est de simuler le transport de particules et si les structures prépondérantes sont différentes entre les écoulements et le transport, alors il est nécessaire d'utiliser des données supplémentaires sensibles aux propriétés de transports, comme les données de concentration ou de tests de traceurs. Dans ce cas, la méthode proposée devrait permettre d'identifier ces structures à condition qu'elles aient les mêmes propriétés, c'est-à-dire si elles sont connectées aux limites du système et sont importantes sur toute leur longueur. Dans le cas contraire, il serait nécessaire d'adapter la paramétrisation aux nouvelles structures, en utilisant une matrice hétérogène par exemple, mais la méthodologie générale ne serait pas remise en cause.

La question des données est donc essentielle. Les modèles obtenus à la Partie 5 montrent que les données de charge, qui sont également les plus utilisées, sont intégratrices et restent sensibles aux chenaux qui ne sont pas dans leur voisinage immédiat. Ceci indique que l'information qu'elles contiennent peut être considérée comme globale, c'est-à-dire pertinente à l'échelle du site. Cependant, cette information a également une partie locale non négligeable. En effet, plus le chenal est proche d'un puits, plus les données issues de ce puits sont affectées par ce chenal. De plus, la capacité à identifier certains détails de la structure de chenalisation (voir Figure 5-15 et Figure 5-17) confirme le caractère local d'une partie de l'information. Par comparaison, des mesures de flux apportent une information extrêmement localisée, pertinente uniquement pour le volume immédiatement autour du point de mesure. L'échelle de pertinence des données n'est pas toujours évidente, ce qui est notamment visible sur les cartes des Figure A-3 à Figure A-6. Mais la question de savoir quelles sont les quantités d'information globale et d'information locale contenues dans une donnée et quelle est l'information requise pour résoudre le problème inverse reste un sujet d'étude. En effet, répondre à cette question permettrait d'estimer a priori le type, la quantité et éventuellement la position des données requises pour définir un modèle pertinent. La partie 5 montre que ces critères sont liées à la fois à la structure du milieu sous-jacent et aux conditions aux limites. Hors les conditions aux limites d'un système sont souvent mal connues. Si la technique consistant à appliquer ces conditions sur des limites éloignées du système étudié est apparemment efficace (voir page 145), l'extrapolation des conditions aux limites à partir des données observées ou l'ajout de ces conditions dans les inconnues du problème pourrait être une solution plus générale et plus robuste.

C. Perspectives

Nous avons abordé l'étude et la modélisation du système complexe régissant les propriétés hydrauliques dans les milieux hétérogènes à partir des observations, de l'analyse statistique des propriétés d'écoulement de la résolution du problème inverse. La première des choses à effectuer pour compléter les travaux présentés dans ce mémoire serait de mieux analyser le lien entre ces trois aspects.

Pour cela, il faudrait tout d'abord étudier les propriétés hydrauliques de modèle de fracturation défini à la partie 3 avec et sans les corrélations induites par le modèle défini par [Davy *et al.*, 2009]. Comparer ces propriétés avec les propriétés mesurées sur le terrain permettrait de qualifier la pertinence du modèle obtenu et l'amélioration apportée par la prise en compte des interactions entre fractures. Il serait également possible de comparer les résultats obtenus avec un modèle stochastique différent, comme un modèle continu, afin d'évaluer si, pour le cas considéré, la représentation en DFN apporte une vraie différence en termes de performance du modèle. Cette distinction pourrait se baser sur la présence, dans le milieu, de chemins préférentiels d'écoulement, évaluée en utilisant les indicateurs de la partie 4. Si le milieu est suffisamment chenalisé, il serait également intéressant d'appliquer la méthode d'identification de la partie 5 afin de déterminer la position des structures les plus importantes par rapport aux

écoulements. Leur intégration aux différents modèles devrait alors permettre d'améliorer la fiabilité des prédictions réalisées, ce qui validerait l'ensemble des approches utilisées pendant le doctorat.

Une deuxième possibilité serait d'étudier les propriétés hydrauliques de plusieurs milieux différents mais dont les indicateurs de chenalisation D_{ic} et D_{cc} auraient les mêmes valeurs. En effet, si ces milieux ont des propriétés similaires, il serait alors possible de substituer un modèle comprenant de nombreux paramètres par un modèle plus simple dont le degré de chenalisation serait identique. Même si les propriétés hydrauliques sont différentes, il serait intéressant de comprendre quel(s) paramètre(s) supplémentaire(s) influent sur ces propriétés afin de définir les caractéristiques les plus essentielles déterminant les propriétés hydrauliques d'un milieu. Pour être parfaitement juste, la première des choses seraient naturellement de tester ces indicateurs sur des milieux 3D afin de mieux appréhender leur comportement. De plus, il reste toujours à pouvoir évaluer précisément les valeurs de D_{ic} et de D_{cc} sur des milieux réels, où la distribution des écoulements est inconnue.

Finalement, le problème inverse reste un sujet d'étude avec un fort potentiel. Une première étape pourrait être de compléter la méthode d'identification de la partie 5 afin d'étendre son domaine d'application. Changer, dans la paramétrisation, la matrice homogène par un milieu hétérogène permettrait tout d'abord son application à des milieux poreux dont le degré de chenalisation est généralement inférieur aux milieux fracturés. Les paramètres décrivant ce milieu pourraient alors être calibrés soit en même temps que les chenaux principaux, soit en amont, afin de modéliser les propriétés hydrauliques moyennes, soit en aval, afin de compléter le modèle composé uniquement des chenaux. Définir à quel moment ces paramètres devront être calibrés n'a pour le moment pas de réponse et dépend certainement des propriétés physique du milieu à modéliser. Deuxièmement, il serait également nécessaire d'étudier quelle amélioration est apportée par l'ajout de nouvelles données dans les modèles obtenus, tant sur le plan quantitatif (nombre, position des données) que sur le plan qualitatif (données avec de l'erreur ou non, utilisation d'un autre type de donnée comme des courbes de rabattement). Cette étude permettrait de mieux comprendre le lien entre données et modèles et permettrait également d'intégrer dans la calibration des données indirectes comme les propriétés de la fracturation étudiée en partie 3 ou les indicateurs de chenalisation étudiés en partie 4. Certaines pistes de recherche existent dans la littérature [Kitanidis, 1998; Tiedeman et al., 2004; Chang et al., 2005] et il serait intéressant de continuer dans cette voie. De plus, les capacités de prédiction des modèles étant liées aux données utilisées, il est indispensable d'évaluer ces capacités pour l'ensemble des modèles présentés ici.

D. Conclusion générale

Pour caractériser les écoulements dans les milieux hétérogènes il est nécessaire de mettre au point des modèles pertinents. L'ensemble des travaux présentés visent à déterminer les paramètres déterminants de la modélisation en se basant sur les caractéristiques principales des milieux étudiés. Après avoir fait un tour des méthodes existantes en partie 1 et décrit les méthodes numériques utilisées en partie 2, nous avons abordé ce problème sous trois angles différents (figure 6-6).

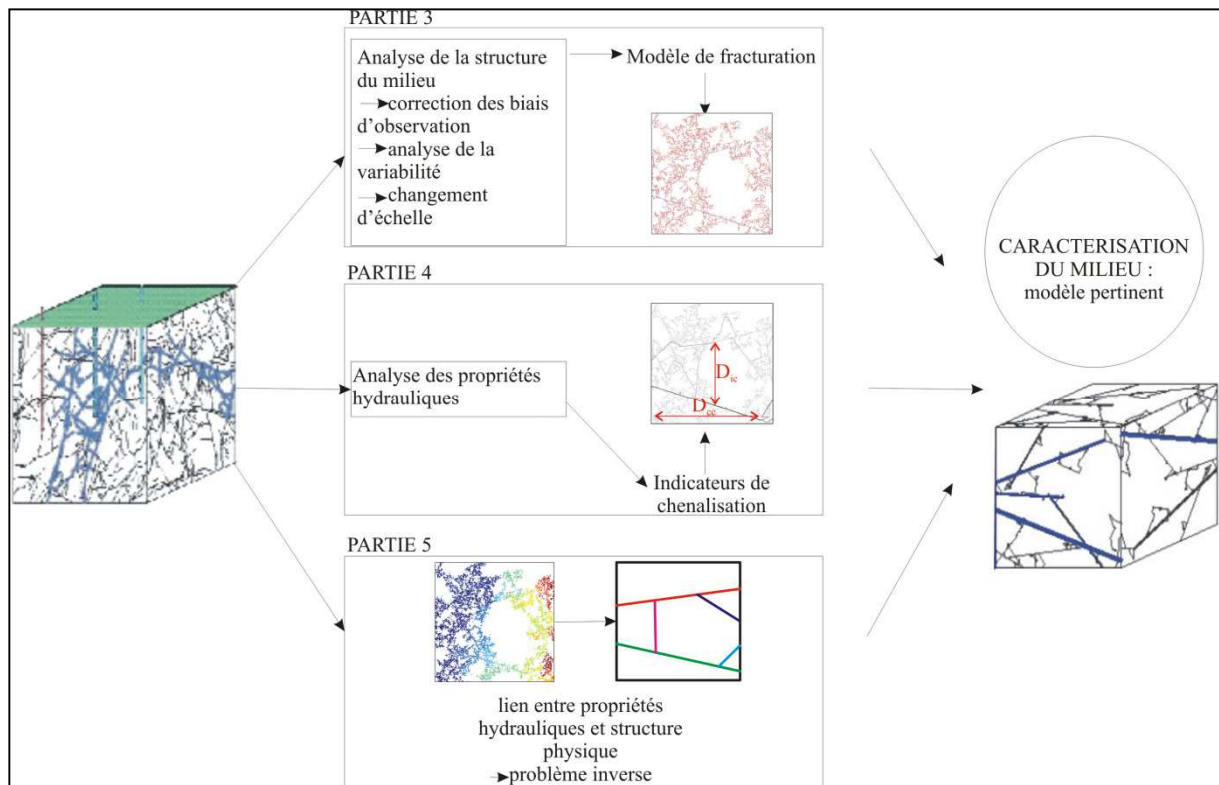


figure 6-6 : illustration des méthodes de caractérisation des propriétés hydrauliques des milieux hétérogènes développée pendant le doctorat.

La partie 3 se concentre sur le modèle décrivant la structure de transmissivité d'un milieu fracturé. A partir des données provenant de deux sites naturels suédois, nous avons défini un modèle de la fracturation. Afin de définir un modèle approprié, nous avons tout d'abord corrigé les biais d'observation. A partir des données disponibles sur la longueur des fractures nous avons calculé le terme d'échelle en utilisant un modèle en loi de puissance pour la distribution des longueurs. L'exposant de la loi de puissance a été calculé à partir des données provenant de plusieurs affleurements et la constance de l'exposant aux échelles supérieures a été étudiée à partir de données provenant de cartes de linéaments. Ceci nous a permis d'observer un changement de régime dans la distribution des longueurs pour les échelles supérieures à une dizaine de mètres. Ce changement, qui n'est pas du à une mauvaise calibration de l'exposant aux échelles inférieures, a été expliquée par le fait que, lors de la croissance des fractures, une fracture plus petite ne peut croiser une fracture plus grande. Ce modèle, décrit par [Davy *et al.*, 2009], explique le changement de pente de la distribution et valide a posteriori les paramètres du modèle. Le terme de densité de la loi de puissance a été calculé à l'échelle locale à partir des données provenant de carottages et d'imageries de tube de puits, ce qui nous a permis de déduire le modèle 3D de la fracturation à l'échelle du site par une méthode de changement d'échelle. Cette méthode s'appuie sur l'exposant défini précédemment ainsi que sur une méthode de comparaison entre domaines locaux intégrant la variabilité du milieu naturel.

En partie 4, nous avons défini des indicateurs statistiques permettant de mesurer le degré de chenalisation des écoulements dans les milieux hétérogènes. Ces deux indicateurs, D_{ic} et D_{cc} , peuvent être rapprochés des propriétés géométriques des chenaux d'écoulement. D_{ic} caractérise en effet la distance entre deux chenaux majeurs et D_{cc} caractérise la longueur effective des chenaux, c'est-à-dire la longueur où les écoulements sont continuent forts. Nous avons également relié la valeur des indicateurs aux caractéristiques de la structure de perméabilité sous-jacente à partir de simulations sur plusieurs cas test poreux et fracturés.

Nous avons montré que l'hétérogénéité, caractérisée par la variance des perméabilités dans les milieux poreux et la densité de fracturation ainsi que la distribution des transmissivités dans les milieux fracturés, tendait à augmenter D_{ic} . De son côté, la nature de la corrélation, défini par la connectivité des zones de fortes perméabilités dans les milieux poreux et la distribution des longueurs dans les milieux fracturés, tendait à augmenter D_{cc} . Compte-tenu de la différence de nature de la structure physique entre les milieux poreux et fracturés, il apparaît que D_{ic} caractérise mieux la chenalisation dans les milieux fracturés pendant que D_{cc} est plus efficace pour les milieux poreux. Les indicateurs mis au point mesurent donc deux caractéristiques différentes de la chenalisation et semble plus robuste et plus discriminant que les indicateurs existants.

Enfin, en partie 5, nous avons tenté de définir la structure de transmissivité des milieux fracturés à partir de données hydrauliques, et notamment de données de charge en stationnaire, en résolvant le problème inverse. De par les difficultés à résoudre ce type de problème, nous nous sommes concentrés sur l'identification des quelques structures prédominantes vis-à-vis des propriétés hydrauliques. Pour cela, nous avons paramétré le milieu par le réseau des chenaux principaux et considéré l'ensemble des structures moins importantes comme une matrice homogène. Nous avons mis au point une méthode d'identification itérative commençant par identifier le chenal le plus important puis complexifiant la paramétrisation tant que l'information contenue dans les données le permettait. Si cette méthode est efficace pour les milieux simples et modérément complexes, les solutions obtenues peuvent être très différentes d'une résolution sur l'autre pour les milieux les plus complexes. Ceci s'explique par la nature mal-posé du problème inverse et par le fait que la méthode proposée ne peut identifier plus de trois ou quatre chenaux simultanément. C'est pourquoi nous avons ajouté une étape d'analyse qui, à partir des résultats de plusieurs résolutions, est capable d'extraire les structures les plus pertinentes afin de définir le modèle global. Dans un deuxième temps, nous avons étudié l'influence des données utilisées sur les résultats de la méthode. Pour cela nous avons étudié l'information contenue dans les données de charge et montré que même si elles sont sensibles aux chenaux d'écoulement, cette sensibilité n'était pas répartie de manière homogène dans l'espace. En effet, en fonction des conditions aux limites, on peut observer la présence de zones d'ombre dans le milieu où les structures n'ont aucune influence sur les charges. En ajoutant une sollicitation dans le milieu, par un pompage par exemple, il est possible de déplacer voire de supprimer ces zones d'ombre. On montre ainsi qu'en combinant les observations obtenues soit sans sollicitation soit avec des pompages en différents endroits, il est possible d'augmenter la sensibilité des données, y compris dans les cas où les conditions aux limites ne permettaient pas d'identifier la structure de chenalisation.

Ces trois approches permettent de mieux caractériser le milieu naturel. En combinant tous les résultats obtenus, il devrait donc être possible de définir un modèle pertinent, c'est-à-dire un modèle capable de reproduire les observations et intégrant les caractéristiques structurels du milieu naturel. Les prédictions en termes de vulnérabilité et de prévention des risques tiré de ce genre de modèle devrait en toute logique être exactes et suffisamment précises pour permettre une meilleure gestion des ressources en eaux.

Références bibliographiques

- Ackermann, R. V., and R. W. Schlische (1997), Anticlustering of small normal faults around larger faults, *Geology*, 25(12), 1127-1130.
- Alcolea, A., et al. (2006), Pilot points method incorporating prior information for solving the groundwater flow inverse problem, *Adv. Water Resources*, 29, 1678-1689.
- Ando, K., et al. (2003), Stochastic continuum modeling of flow and transport in a crystalline rock mass: Fanay-Augeres, France, revisited, *Hydrogeology Journal*, 11(5), 521-535.
- Andresen, C. A., et al. (2008), Complex network properties of outcrop fractures, paper presented at The 27th IUGG Conference on Mathematical Geophysics.
- Atkinson, B. K. (1989), Chapter 8, in *Fracture Mechanics of Rocks*, edited, Academic Press, Incorporated.
- Audouin, O., et al. (2008), Flowpath structure in a limestone aquifer: multi-borehole logging investigations at the hydrogeological experimental site of Poitiers, France, *Hydrogeology Journal*, 16(5), 939-950.
- Basquet, R., et al. (2005), Fracture Flow Property Identification: An Optimized Implementation of Discrete Fracture Network Models, paper presented at 14th SPE Middle East Oil&Gas Show and Conference, Bahrain, 12-15 March 2005.
- Bear, J. (1973), *Dynamics of Fluids in Porous Media*, Dover Publications.
- Beaudoin, A., et al. (2007), An Efficient Parallel Particle Tracker for Advection-Diffusion Simulations in Heterogeneous Porous Media in *Euro-Par 2007 Parallel Processing*, edited by Springer, Berlin / Heidelberg.
- Berkowitz, B. (2002), Characterizing flow and transport in fractured geological media : a review, *Advances Water Resources*, 25, 861-884.
- Berkowitz, B., and P. Adler (1998), Stereological analysis of fracture network structure in geological formations, *J. Geophys. Res.*, 103(B7), 15339-15360.
- Berkowitz, B., et al. (1994), Mass transfer at fracture intersections: An evaluation of mixing models, *Water Resour. Res.*, 30.
- Billiaux, D., et al. (1989), Three-dimensional statistical modelling of a fractured rock mass--an example from the Fanay-Augères mine, *International Journal of Rock Mechanics and Mining Science & Geomechanics Abstracts*, 26(3-4), 281-299.
- Bonnet, E., et al. (2001), Scaling of fracture systems in geological media, *Reviews of Geophysics*, 39(3), 347-383.
- Bour, O., et al. (2002), A statistical scaling model for fracture network geometry, with validation on a multi-scale mapping of a joint network (Hornelen Basin Norway), *J. Geophys. Res.*, 107, 2113.
- Bredehoeft, J. (2005), The conceptualization model problem-surprise, *Hydrogeology Journal*, 13, 37-46.
- Bruckner, G., et al. (1998), An inverse problem from the 2D groundwater modelling., *Inverse Problems*, 14, 835-851.

- Bruyelle, J., and A. Lange (2009), An extended evolution strategy for the characterization of fracture conductivities from well tests, in *Proceedings of the 11th Annual conference on Genetic and evolutionary computation*, edited, ACM, Montreal, Québec, Canada.
- Cacas, M. C., et al. (1990a), Modeling fracture flow with a stochastic discrete fracture network: Calibration and validation, 1, the flow model, *Water Resour. Res.*, 26, 479-489.
- Cacas, M. C., et al. (1990b), Modeling fracture flow with a stochastic discrete fracture network: Calibration and validation, 2, The transport model, *Water Resour. Res.*, 26, 491-500.
- Carrera, J., et al. (2005), Inverse problem in hydrogeology., *Hydrogeology Journal*, 13, 206-222.
- Carrera, J., and S. P. Neuman (1986), Estimation of Aquifer Parameters Under Transient and Steady State Conditions: 1. Maximum Likelihood Method Incorporating Prior Information, *Water Resour. Res.*, 22.
- Chang, L.-F., et al. (2005), Optimal observation network design for parameter structure identification in groundwater modeling., *Water Resources Research*, 41(W03002).
- Chang, Y. C., et al. (2008), Determination of the parameter pattern and values for a one-dimensional multi-zone unconfined aquifer, *Hydrogeology Journal*, 16(2), 205-214.
- Charlaix, E., et al. (1987), Permeability of a random array of fractures of widely varying apertures, *Transport in Porous Media*, 2(1), 31-43.
- Chavent, G. (1991), On the theory and practice of nonlinear least-squares., *Advances in Water Resources*, 14(2), 55-63.
- Chelcea, S., et al. (2006), A New Agglomerative 2–3 Hierarchical Clustering Algorithm in *Innovations in Classification, Data Science, and Information Systems*, edited, Springer, Berlin Heidelberg.
- Chelcea, S., et al. (2005), A New Agglomerative 2–3 Hierarchical Clustering Algorithm, in *Innovations in Classification, Data Science, and Information Systems*, edited, pp. 3-10.
- Cheng, D., et al. (2006), A divide-and-merge methodology for clustering *ACM transactions on database systems*, 31(4), 1499-1525.
- Cox, T., and M. Cox (1994), *Multidimensional scaling*, Chapman & Hall, London, UK.
- Cressie, N. A. C. (1993), *Statistics for Spatial Data*, Revised ed., 928 pp., Wiley-Interscience, New York.
- Cronquist, L., et al. (2004), Oskarshamn site investigation. Detailed fracture mapping of two outcrops at Laxemar. SKB P-04-274, Golder Associates AB.
- Cronquist, T., et al. (2006), Oskarshamn site investigation. Detailed outcrop mapping on drillsite KLX11, Golder Associates AB.
- Darcel, C., et al. (2004), Alternative DFN model based on initial site investigations at Simpevarp., Svensk Kärnbränslehantering AB.

- Darcel, C., et al. (2009), Statistical methodology for discrete fracture model – including fracture size, orientation uncertainty together with intensity uncertainty and variability, Svensk Kärnbränslehantering AB (SKB).
- Darcy, H. (1856), Détermination des Lois D'Ecoulement De L'eau à Travers Le Sable, Appendix, Note D, in *Fontaines Publiques de la Ville de Dijon*, edited by P. edited by V. Dalmont, pp. 590-594.
- Datta, S., and S. Datta (2006), Methods for evaluating clustering algorithms for gene expression data using a reference set of functional classes, *BMC bioinformatics*, 7(397).
- Davis, T. A. (2004), Algorithm 832: UMFPACK, an unsymmetric-pattern multifrontal method, *ACM transactions on Mathematical Software*, 30(2), 196-199.
- Davy, P., et al. (2006), A note on the angular correction applied to fracture intensity profiles along drill core, *J. Geophys. Res.-Solid Earth*, 111(B11), 7.
- Davy, P., et al. (1995), Localization and fault growth in layered brittle-ductile systems: Implications for deformations of the continental lithosphere, *J. Geophys. Res.*, 100(B4), 6281-6294.
- Davy, P., et al. (2009), An Universal Model of Fracture Scaling and its consequence for crustal hydro-mechanics, *Journal of Geophysical Research - Solid Earth*, submitted.
- Davy, P., et al. (1990), Some consequences of a proposed fractal nature of continental faulting., *Nature*, 348, 56-58.
- Day-Lewis, F. D., et al. (2000), Identifying fracture-zone geometry using simulated annealing and hydraulic-connection data, *Water Resources Research*, 36(7), 1707-1721.
- de Dreuzy, J.-R., et al. (2001a), Hydraulic properties of two-dimensional random fracture networks following a power law length distribution 1. Effective connectivity, *Water Resour. Res.*, 37(8), 2065–2078.
- de Dreuzy, J.-R., et al. (2001b), Hydraulic Properties of Two-Dimensional Random Fracture Networks following a Power Law Length Distribution 2. Permeability of Networks Based on Lognormal Distribution of Apertures, *Water Resour. Res.*, 37(8).
- de Dreuzy, J. R., et al. (2007), Asymptotic dispersion in 2D heterogeneous porous media determined by parallel numerical simulations, *Water Resources Research*, 43(10), 13.
- de Dreuzy, J. R., and P. Davy (2007), Relation between fractional flow models and fractal or long-range 2-D permeability fields, *Water Resources Research*, 43(4), 13.
- de Marsily, G., et al. (2005), Dealing with spatial heterogeneity, *Hydrogeology Journal*, 12, 161-183.
- de Marsily, G., et al. (1999), 40 years of inverse problems in hydrogeology, *Earth & Planetary Sciences*, 329, 73-87.
- de Marsily, G., et al. (2000), Four decades of inverse problems in hydrogeology, in *Theory, modeling and field investigation in hydrogeology : a special volume in honor of Shlomo P. Neuman's 60th birthday*, edited by D. Zhang and C. L. Winter, pp. 1-17, Geological society of America, Boulder (Colorado, USA).
- de Marsily, G., et al. (1984), Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model, in

- Geostatistics for natural resources characterization. Part 2.*, edited by G. Verly, et al., pp. 831-849, Kluwer Academic Publishers.
- Dershowitz, W., et al. (1991a), FracMan Version 2.3. Interactive discrete feature data analysis, geometric modeling, and exploration simulation. User documentation, edited, Golder Associates Inc. , Redmond Washington USA.
- Dershowitz, W., et al. (1991b), Discrete fracture modelling for the Stripa site characterization and validation drift inflow predictions. Stripa Project Technical Report 91-16, Svensk Kärnbränslehantering AB.
- Donado, L. D., et al. (2005), Calibration of hydraulic and tracer tests in fractured media represented by a DFN model, paper presented at International Conference on Calibration and Reliability in Groundwater Modelling : From Uncertainty to Decision Making, IAHS, Wallingford, UK, Hague , The Netherlands.
- Doughty, C., et al. (1994), Hydrologic characterization of heterogeneous geologic media with an inverse method based on iterated function systems, *Water Resources Research*, 30(6), 1721-1745.
- Erhel, J., et al. (2009), Flow Simulation in three-dimensional discrete fracture networks, *SIAM J. Sci. Comput.*, 31(4), 2688-2705.
- Falgout, R. D., and U. M. Yang (2002), hypre: a Library of High Performance Preconditioners, in *Computational Science - ICCS 2002 Part III, Lecture Notes in Computer Science*, edited by P. M. A. Sloot, et al., pp. 632-641, Springer-Verlag.
- Fan, Y., et al. (2007), Groundwater flow and groundwater-stream interaction in fractured and dipping sedimentary rocks: Insights from numerical models, *Water Resources Research*, 43(1), 13.
- Fetter, C. W. (2001), Properties of Aquifers, in *Applied Hydrogeology*, edited, pp. 81-89, Prentice-Hall, Inc., Upper Saddle River (New Jersey, USA).
- Fogg, G. E. (1986), Groundwater flow and sand body interconnectedness in a thick, multiple aquifer system, *Water Resources Research*, 22(5), 679-694.
- Forsberg, O., et al. (2005), Oskarshamn site investigation. Detailed outcrop mapping at the drill site of KLX09 in Laxemar. SKB P-05-260, Golder Associates AB.
- Fox, A., et al. (2007), Statistical geological discrete fracture network model, Forsmark modelling stage 2.2, Svensk Kärnbränslehantering AB.
- Francese, R., et al. (2009), A structural and geophysical approach to the study of fractured aquifers in the Scansano-Magliano in Toscana Ridge, southern Tuscany, Italy, *Hydrogeology Journal*, 17(5), 1233-1246.
- Franssen, H. J. W. M. H., and J. J. Gómez-Hernández (2002), 3D inverse modelling of groundwater flow at a fractured site using a stochastic continuum model with multiple statistical populations, *Stochastic Environmental Research and Risk Assessment*, 16(2), 155-174.
- Fu, J., and J. Gómez-Hernández (2009), A Blocking Markov Chain Monte Carlo Method for Inverse Stochastic Hydrogeological Modeling, *Mathematical Geosciences*, 41(2), 105-128.

- Garai, G., and B. Chaudhuri (2004), A novel genetic algorithm for automatic clustering, *Pattern Recognition Letters*, 25(2), 173-187.
- Gelhar, L. W. (1993), *Stochastic subsurface hydrology*, Prentice Hall, Old Trappan, N. J.
- Gleick, P. H. (1996), Water resources, in *Encyclopedia of Climate and Weather*, edited by S. H. Schneider, pp. 817-823, Oxford University Press, New York.
- Goldenfeld, N., and L. P. Kadanoff (1999), Simple lessons from complexity, *Science*, 284(5411), 87-89.
- Gomez-Hernandez, J., et al. (1997), Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data. 1. Theory, *Journal of Hydrology*, 1-4(203), 162-174.
- Gordon, A. D. (1999a), *Classification*, 2nd ed., Chapman and Hall.
- Gordon, A. D. (1999b), Hierarchical classifications, in *Classification, 2nd edition (Monographs on Statistics and Applied Probability, 82)*, edited, pp. 69-111, Chapman & Hall / CRC.
- Grasmueck, M. (1996), 3-D ground-penetrating radar applied to fracture imaging in gneiss, *Geophysics*, 61(4), 1050-1064.
- Grenier, C., et al. (2009), Evaluation of retention properties of a semi-synthetic fractured block from modelling at performance assessment time scales (Aspo Hard Rock Laboratory, Sweden), *Hydrogeology Journal*, 17(5), 1051-1066.
- Gropp, W., et al. (1999), *Using MPI: Portable Parallel Programming with the Message-Passing Interface*, MIT Press.
- Gustafson, G., and Å. Fransson (2006), The use of the Pareto distribution for fracture transmissivity assessment, *Hydrogeology Journal*, 14(1), 15-20.
- Gwo, J.-P. (2001), In Search of Preferential Flow Paths in Structured Porous Media Using a Simple Genetic Algorithm, *Water Resour. Res.*, 37(6).
- Hadamard, J. (1902), Sur les problèmes aux dérivées partielles et leur signification physique, *Princeton University Bulletin*, 49-52.
- Hammah, R., and J. Curran (1998), Fuzzy Cluster algorithm for the automatic identification of joint sets, *Int J Rock Mech Min Sci* 35, 889-905.
- Hanor, J. S. (1993), Effective hydraulic conductivity of fractured clay beds at a hazardous waste landfill; Louisiana Gulf Coast, *Water Resources Research*, 29(11), 3691-3698.
- Hansen, N., and A. Ostermeier (1996), Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation., paper presented at 1996 IEEE International Conference on Evolutionary Computation.
- Hao, Y. H., et al. (2008), Hydraulic tomography for detecting fracture zone connectivity, *Ground Water*, 46(2), 183-192.
- Hill, M. C. (2006), The practical use of simplicity in developing groundwater models, *Ground Water*, 44(6), 775-781.
- Hill, M. C., and C. R. Tiedeman (2007), Estimating parameter values, in *Effective groundwater model calibration*, edited, pp. 67-92, John Wiley and Sons, Inc., Hoboken, New Jersey (USA).

- Hoeksema, R. J., and P. K. Kitanidis (1985), Analysis of the spatial structure of properties of selected aquifers, *Water Resources Research*, 21(4), 563-572.
- Hofrichter, J., and G. Winkler (2006), Statistical analysis for the hydrogeological evaluation of the fracture networks in hard rocks *Environmental Geology*, 9(6), 821-827.
- Holland, J. H. (1975), *Adaptation in natural and artificial systems: an introductory analysis with application to biology, control and artificial intelligence*, University of Michigan Press.
- Hull, L. C., and K. Koslow (1986), Streamline routing through fracture junctions, *Water Resources Research*, 22(12), 1731-1734.
- Hunt, R. J., et al. (2007), Are models too simple? Arguments for increased parameterization, *Ground Water*, 45(3).
- Ingber, A. L. (1993a), Adaptive Simulated Annealing (ASA), Lester Ingber Research, McLean, VA.
- Ingber, L. (1993b), Simulated Annealing - Practice versus Theory, *Math. Comput. Model.*, 18(11), 29-57.
- Journel, A., and F. G. Alabert (1989), Focusing on spatial connectivity of extreme-valued attributes : stochastic indicator models of reservoir heterogeneities, *AAPG Bulletin*, 73.
- Khan, D. K., et al. (2007), Approximate sensitivity coefficients for integrating hydraulic head data into geological models, *Journal of Hydrology*, 347(3-4), 460-473.
- Kim, K., and E. Y. Yamashita (2007), Using a K-means clustering algorithm to examine patterns of pedestrian involved crashes in Honolulu, Hawaii, *Journal of Advanced Transportation*, 41(1), 69-89.
- Kirkpatrick, S., et al. (1983), Optimization by simulated annealing, *Science*, 220(4598).
- Kitanidis, P. K. (1998), How Observations and Structure Affect the geostatistical Solution to the Steady-State Inverse Problem, *Ground Water*, 36(5).
- Knudby, C., and J. Carrera (2005), On the relationship between indicators of geostatistical, flow and transport connectivity, *Adv. Water Resources*, 28, 405-421.
- Koltermann, C. E., and S. M. Gorelick (1996), Heterogeneity in sedimentary deposits: a review of structure imitating, process-imitating and descriptives approaches, *Water Resources Research*, 32(9), 2617-2658.
- Krige, D. G. (1951), A statistical approach to some mine valuations and allied problems at the Witwatersrand, Master's thesis of the University of Witwatersrand.
- Kühne, T. (2006), Matters of (Meta-)Modeling, *Softw. Syst. Model.*, 5(4), 369-385.
- La Pointe, P. R. (2002), Derivation of parent fracture population statistics from trace length measurements of fractal fracture populations, *Int. J. of Rock Mechanics and Mining Sciences*, 39, 381-388.
- Larocque, M., et al. (1999), Determining karst transmissivities with inverse modeling and an equivalent porous media, *Ground Water*, 37(6), 897-903.

- LaVenue, A. M., et al. (1995), Pilot Point Methodology for Automated Calibration of an Ensemble of Conditionally Simulated Transmissivity Fields 2. Application, *Water Resour. Res.*, 31(3), 495-516.
- Lavenue, M., and G. de Marsily (2001), Three-Dimensional Interference Test Interpretation in a Fractured Aquifer Using the Pilot Point Inverse Method, *Water Resour. Res.*, 37.
- Le Borgne, T., et al. (2006), Assessment of preferential flow path connectivity and hydraulic properties at single-borehole and cross-borehole scales in a fractured aquifer, *Journal of Hydrology*.
- Le Borgne, T., et al. (2007a), Comparison of alternative methodologies for identifying and characterizing preferential flow paths in heterogeneous aquifers, *Journal of Hydrology*, 345(3-4), 134-148.
- Le Borgne, T., et al. (2007b), Characterization of the velocity field organization in heterogeneous media by conditional correlation, *Water Resources Research*, 43(2), 10.
- Le Goc, R., and J.-R. de Dreuzy (2007), Adjoint-state method for diffusion equation in the finite-difference and finite-volume frameworks, *Computers & Geosciences*, 33(2), 291-293.
- Le Goc, R., et al. (2010a), Inverse problem strategy to identify flow channels in fractured media, *Adv. Water Resources*, Submitted.
- Le Goc, R., et al. (2010b), Statistical characteristics of flow as indicators of channeling in heterogeneous porous and fractured media, *Advances in Water Resources*, In Press, Corrected Proof.
- Long, J. C. S., et al. (1982), Porous Media Equivalents for Networks of Discontinuous Fractures, *Water Resour. Res.*, 18(3).
- Lutz Kettner, and S. Näher (2004), Two Computational Geometry Libraries: LEDA and CGAL, in *Handbook of Discrete and Computational Geometry*, edited by J. E. Goodman and J. O'Rourke, pp. 1435-1463, CRC Press LLC, Boca Raton, FL.
- Marrett, R., and R. W. Allmendinger (1991), Estimates of strain due to brittle faulting: sampling of fault populations, *Journal of Structural Geology*, 13(6), 735-738.
- Marryote, R. A., et al. (1993), Optimal Groundwater Management; 2. application of simulated annealing to a field-scale contamination site., *Water Resources Research*, 29(4), 847-860.
- Martinez-Landa, L., and J. Carrera (2005), An analysis of hydraulic conductivity scale effects in granite (Full-scale Engineered Barrier Experiment (FEBEX), Grimsel, Switzerland), *Water Resources Research*, 41(3), 13.
- McLaughlin, D., and L. R. Townley (1996), A reassessment of the groundwater inverse problem., *Water Resources Research*, 32(5).
- Metropolis, N., et al. (1953), Equation of state calculation by fast computing machines, *J. Chem. Phys.*, 21(6), 1087-1092.
- Moore, C., and J. Doherty (2006), The cost of uniqueness in groundwater model calibration, *Adv. Water Resources*, 29, 605-623.
- Moreno, L., and I. Neretnieks (1993), Fluid flow and solute transport in a network of channel, *J. Contam. Hydrol.*, 14, 163-192.

- Moreno, L., and C. F. Tsang (1994), Flow channeling in strongly heterogeneous porous media, *Water Resources Research*, 30, 1421-1430.
- Munier, R. (1995), Studies of geological structures at Äspö. Comprehensive summary of results. SKB HRL Progress Report 25-95-21., Svensk Kärnbränslehantering AB.
- Munier, R. (2004), Statistical analysis of fracture data, adapted for modelling Discrete Fracture Networks-Version 2. SKB R-04-66, Svensk Kärnbränslehantering AB.
- Nakao, S., et al. (2000), Hydraulic well testing inversion for modeling fluid flow in fractured rocks using simulated annealing: a case study at Raymond field site, California, *Journal of Applied Geophysics*, 45, 203-223.
- Neuman, S. P. (1973), Calibration of distributed parameter groundwater flow models viewed as multiple objective decision process under uncertainty, *Water Resources Research*, 9(4), 1006-1021.
- Neuman, S. P. (2005), Trends, prospects and challenges in quantifying flow and transport through fractured rocks, *Hydrogeology Journal*, 13, 124-147.
- Newman, M. (2005), Power laws, Pareto distributions and Zipf's law, *Contemporary Physics*, 46, 323-351.
- Olofsson, I., et al. (2007), Site descriptive modelling Forsmark, stage 2.2. A fracture domain concept as a basis for the statistical modelling of fractures and minor deformation zones, and interdisciplinary coordination, Svensk Kärnbränslehantering AB.
- Olsson, O. (1992), Site characterization and validation. , *Stripa Project Technical Report, TR 92-22*.
- Oltmans, H., and P. J. T. Verheijen (1997), Influence of noise on power-law scaling functions and an algorithm for dimension estimations, *Physical Review E*, 56(1).
- Petitjohn, F. J., et al. (1987), *Sand and sandstone*, 2nd ed., 553 pp., Springer-Verlag.
- Pichot, G., et al. (2009), A Mixed-Hybrid Mortar Method for solving flow in Discrete Fracture Networks, *Applicable Analysis*, 2009.
- Pickering, G., et al. (1995), Sampling power-law distributions, *Tectonophysics*, 248, 1-20.
- Piggot, A. (1997), Fractal relations for the diameter and trace lengths of disc-shaped fractures, *J. Geophys. Res.*, 102, 18121-18125.
- Press, W. H., et al. (2007), Modeling of data: Robust Estimation, in *Numerical Recipes: The Art of Scientific Computing*, edited, Cambridge University Press.
- Prickett, T. (1975), Modeling techniques for groundwater evaluation, in *Advances in Hydroscience*, edited by C. Ven Te, pp. 1-143, Academic Press, New York.
- Priest, S. D., and J. A. Hudson (1981), Estimation of discontinuity spacings and trace length using scanline surveys, *Int. J. of Rock Mechanics and Mining Sciences*, 18, 183-197.
- RamaRao, B. S., et al. (1995), Pilote point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields. 1. Theory and computational experiments, *Water Resources Research*, 31(3), 475-493.
- Rasmuson, A., and I. Neretnieks (1986), Radionuclide transport in fast channels in crystalline rock, *Water Resources Research*, 22, 1247-1256.

- Renard, P. (2007), Stochastic Hydrogeology: What professionals really need?, *Ground Water*, 45(5), 531-541.
- Renshaw, C. E. (1996), Estimation of Fracture Zone Geometry from Steady-State Hydraulic Head Data Using Iterative Sequential Cokriging, *Geophys. Res. Lett.*, 23(19).
- Riley, M. S. (2005), Fracture trace length and number distributions from fracture mapping, *J. Geophys. Res.*, 110(B08414).
- Rodrigo, J., et al. (2002), Gaussian vs non-Gaussian approaches in the identification of hydraulic conductivity fields in fractured media, paper presented at Conference on Calibration and Reliability in Groundwater Modelling (ModelCARE 2002), Int Assoc Hydrological Sciences, Prague, Czech Republic, Jun 17-20.
- Ronayne, M. J., et al. (2008), Identifying discrete geologic structures that produce anomalous hydraulic response: an inverse modeling approach, *Water Resources Research*, 44(W08426).
- Sambridge, M., and K. Mosegaard (2002), Monte Carlo methods in geophysical inverse problems., *Reviews of Geophysics*, 40(3).
- Sanchez-Vila, X., et al. (2006), Representative hydraulic conductivities in saturated groundwater flow, *Rev. Geophys.*, 44.
- Sarbu, C., et al. (2007), Fuzzy divisive hierarchical clustering of soil data using Gustafson-Kessel algorithm, *Chemometrics and Intelligent Laboratory Systems*, 86(1), 121-129.
- Siegel, P., et al. (1996), Automatic calibration of groundwater flow parameters for an unconfined aquifer northeast of Vienna, in *Calibration and Reliability Modelling, Proceedings of the ModelCARE96 Conference, Golden, Colorado (USA), IAHS publ. n°237*, edited, pp. 53-60, Wallingford, Oxfordshire (UK).
- Silliman, S. E. (1989), An interpretation of the difference between aperture estimates derived from hydraulic and tracer tests in a single fracture, *Water Resources Research*, 25(10), 2275-2283.
- Silliman, S. E. (1995), Particle transport through two-dimensional, saturated porous media: influence of physical structure of the medium, *Journal of Hydrology*, 167(1-4), 79-98.
- Silliman, S. E., et al. (1998), The Use of Laboratory Experiments for the Study of Conservative Solute Transport in Heterogeneous Porous Media, *Hydrogeology Journal*, 6.
- Stehfest, H. (1970a), Algorithm 368. Numerical Inversion of Laplace Transforms, *Communications of the ACM*, 13(1), 47-49.
- Stehfest, H. (1970b), Remark on Algorithm 368. Numerical Inversion of Laplace Transforms, *Communications of the ACM*, 13(10), 624.
- Stigsson, M., and SKB (2007), P-07-178 Forsmark site investigation - Analysis of uncertainty and changes in orientation of fractures coupled to PFL anomalies, Svensk Kärnbränslehantering AB (SKB).
- Stigsson, M., and SKB (2008), P-08-104 Analysis of uncertainty in orientation of fractures coupled to PFL anomalies - Site descriptive modelling SDM-Site Laxemar, Svensk Kärnbränslehantering AB.
- Sun, N. Z., and W. W.-G. Yeh (2007a), Development of objective-oriented groundwater models: 1. Robust parameter identification, *Water Resources Research*, 43.

- Sun, N. Z., and W. W.-G. Yeh (2007b), Development of objective-oriented groundwater models: 2. robust experimental design, *Water Resources Research*, 43.
- Tarantola, A. (2004), *Inverse Problem Theory and model parameter estimation.*, SIAM.
- Terzaghi, R. D. (1965), Sources of errors in joint surveys, *Geotechnique*, 15, 287-304.
- Tiedeman, C. R., et al. (2004), A method for evaluating the importance of system state observations to model predictions, with application to the Death Valley regional groundwater flow system, *Water Resources Research*, 40(12).
- Tiedeman, C. R., et al. (1995), Characterization of a high-transmissivity zone by well-test analysis: Steady state case., *Water Resources Research*, 31(1), 27-37.
- Tsai, F. T.-C., et al. (2003a), A combinatorial optimization scheme for parameter structure identification in Ground-Water modeling, *Ground Water*, 41(2), 156-169.
- Tsai, F. T.-C., et al. (2005), Geophysical parameterization and parameter structure identification using natural neighbors in groundwater inverse problems, *Journal of Hydrology*, 308, 269-283.
- Tsai, F. T.-C., et al. (2003b), Global-local optimization for parameter structure identification in three-dimensional groundwater modeling, *Water Resources Research*, 39(2).
- Tsang, C.-F., and I. Neretnieks (1998), Flow Channeling in heterogeneous fractured rocks, *Reviews of Geophysics*, 36(2).
- Tsang, Y. W., and C. F. Tsang (1989), Flow channeling in a single fracture as a two dimensional strongly heterogeneous permeable medium, *Water Resources Research*, 25(9), 2076-2080.
- Tsoflias, G. P., et al. (2004), Vertical fracture detection by exploiting the polarization properties of ground-penetrating radar signals, *Geophysics*, 69(3), 803-810.
- UNESCO (2009), *Water in a changing world*, 3rd edition, UNESCO.
- van der Maaten, L. J. P. (2007), *An Introduction to Dimensionality Reduction Using Matlab*, Maastricht University, Maastricht, The Netherlands.
- Vesselinov, V. V., et al. (2001), Three-Dimensional Numerical Inversion of Pneumatic Cross-Hole Tests in Unsaturated Fractured Tuff 2. Equivalent Parameters, High-Resolution Stochastic Imaging and Scale Effects, *Water Resour. Res.*, 37(12), 3019–3041.
- Vrac, M., et al. (2007), Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing, *Climate Research*, 34(3), 169-184.
- Wahlgren, C.-H., et al. (2008), R-08-54 - Geology Laxemar. Site descriptive modelling SDM - Site Laxemar, Svensk Kärnbränslehantering AB (SKB).
- Watanabe, K., et al. (1997), Channel network modeling of the granite in the Himachi area, Japan. , in *Earthquake proof design and active fault*, edited, pp. 247-262, Elsevier Sci.
- Western, A. W., et al. (1998), How well do indicator variograms capture connectivity of soil moisture, *Hydrol. Proces.*, 12, 1851-1868.
- Western, A. W., et al. (2001), Toward capturing hydrologically significant connectivity in spatial patterns, *Water Resources Research*, 37(1), 83-97.

- Yeh, T.-C. J., et al. (1996), Estimation of fracture zone geometry from steady-state hydraulic head data using iterative sequential cokriging, *Geophysical Research Letters*, 23(19), 2685-2688.
- Yeh, W. W. G., and Y. S. Yoon (1981), Aquifer parameter identification with optimum dimension in parameterization, *Water Resources Research*, 17(3), 664-672.
- Zhang, D. (2002), *Stochastic method for flow in porous media: coping with uncertainty*, Academic press.
- Zheng, C., and P. Wang (1996), Parameter structure identification using tabu search and simulated annealing, *Adv. Water Resources*, 19(4), 215-224.
- Zimmerman, D. A., et al. (1998), A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resources Research*, 34(6), 1373-1413.
- Zinn, B., and C. F. Harvey (2003), When good statistical models of aquifer heterogeneity go bad: a comparison of flow, dispersion and mass transfer in connected and multivariate Gaussian hydraulic conductivity fields, *Water Resources Research*, 39(3).

Annexes

- R. Le Goc, and J.-R. de Dreuzy, Adjoint-state method for diffusion equation in the finite-difference and finite-volume frameworks. *Computers & Geosciences* 33 (2007) 291-293
- P. Davy, R. Le Goc, C. Darcel, O. Bour, J.-R. de Dreuzy, and R. Munier, An Ubiquitous Model of Fracture Scaling and its consequence for crustal hydro-mechanics. *Journal of Geophysical Research - Solid Earth* submitted (2009).

Short note

Adjoint-state method for diffusion equation in the finite-difference and finite-volume frameworks

R. Le Goc*, J.-R. de Dreuzy

Géosciences Rennes, UMR CNRS 6118, Université de Rennes 1, 35042 Rennes Cedex, France

Received 19 December 2005; received in revised form 3 May 2006; accepted 13 June 2006

Keywords: Adjoint state; Finite difference; Finite volume; Gradient; Inverse problem

The adjoint state is extensively used for the optimization of the objective function in inverse problems. We derive its expression for diffusion in finite-difference and finite-volume frameworks and show why the classical formula used in finite elements is not directly applicable.

The inverse problem is generally expressed in terms of the optimization of an objective function F proportional to the squared difference between the observations h_0 and the results h of the direct model:

$$F = \frac{1}{2} \sum_{i=1}^n (h_i - h_{0i})^2, \quad (1)$$

where n is the number of data (Chavent, 1991; Jarny et al., 1991). For the diffusion equation in steady state, the direct problem writes

$$\nabla \cdot (K \nabla h) = 0, \quad (2)$$

where K is the conductivity. Appropriate boundary conditions are added to this equation. The following methodology is an application of the adjoint-state method to finite differences including for example groundwater flow. We deliberately keep the development simple for the sake of clarity. It can be

generalized to more complex transient state problems, including additional terms depending on K in the objective function and source terms in the diffusion equation (Medina and Carrera, 2003; Townley and Wilson, 1985).

Whatever the numerical method used, Eq. (2) is discretized on a grid with a conductivity, K_{mesh} in each mesh. In finite-difference and finite-volume methods, the discretization uses an inter-mesh conductivity, K_{inter} that depends on the neighbouring mesh conductivities (Fig. 1). This dependence can be non-linear. Once discretized, Eq. (2) leads to the following system of equation:

$$A[K_{inter}(K_{mesh})] \cdot h = b[K_{inter}(K_{mesh})], \quad (3)$$

where the dimension of matrix A is equal to the number of meshes. In finite-element framework, we note that the discretization of (2) also leads to a linear system of equation whose matrix depends directly and linearly on K_{mesh} .

In gradient methods, the optimization of the objective function F requires the computation of its gradient:

$$\delta F = \frac{\partial F}{\partial h} \delta h, \quad (4)$$

where the small variation δh around the solution h is induced by a small variation of the parameters

*Corresponding author. Tel.: +33 223236624; fax: +33 223236090.

E-mail address: romain.le-goc@univ-rennes1.fr (R. Le Goc).

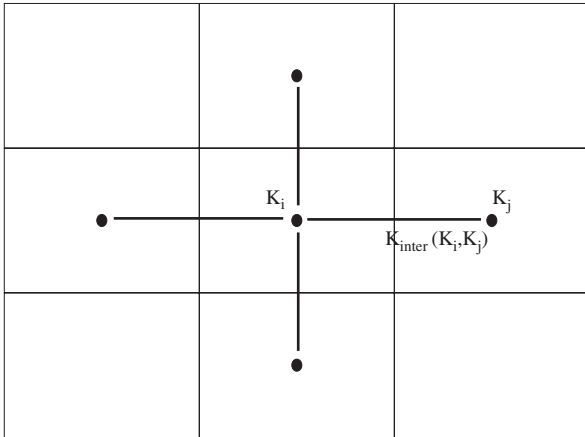


Fig. 1. Finite-difference and finite-volume methods use an inter-mesh conductivity K_{inter} depending on neighbouring mesh conductivities K_i and K_j . Gradient of objective function must include gradient of K_{inter} against K_i and K_j , which is a constant for arithmetic mean but conductivity dependent for geometric and harmonic means.

δK_{mesh} around K_{mesh} (Tarantola, 1987). The differentiation of (3) provides for the relation between δh and δK_{mesh} (Jarny et al., 1991):

$$A[K_{inter}(K_{mesh} + \delta K_{mesh})] \cdot (h + \delta h) = b[K_{inter}(K_{mesh} + \delta K_{mesh})]. \quad (5)$$

To simplify notations, we set

$$A_K = A[K_{inter}(K_{mesh})], \\ b_K = b[K_{inter}(K_{mesh})].$$

To first order, the development of Eq. (5) leads to

$$\left(A_K + \frac{\partial A_K}{\partial K_{inter}} \cdot \frac{\partial K_{inter}}{\partial K_{mesh}} \delta K_{mesh} \right) \cdot (h + \delta h) - \left(b_K + \frac{\partial b_K}{\partial K_{inter}} \cdot \frac{\partial K_{inter}}{\partial K_{mesh}} \delta K_{mesh} \right) = 0 \quad (6)$$

and by neglecting the second-order term and by using Eq. (3) we obtain

$$\left(\frac{\partial A_K}{\partial K_{inter}} \cdot h - \frac{\partial b_K}{\partial K_{inter}} \right) \cdot \frac{\partial K_{inter}}{\partial K_{mesh}} \delta K_{mesh} + A_K \cdot \delta h = 0. \quad (7)$$

By inverting A_K , δh can be expressed as a function of δK_{mesh} , which is equivalent in complexity to solving a number of linear systems equal to the number of meshes. It can also be computed from a much more cost-effective adjoint-state method that requires solving only one linear system. The adjoint state is obtained by first multiplying Eq. (7) by a

test-vector ψ :

$${}^T \psi \cdot \left(\frac{\partial A_K}{\partial K_{inter}} \cdot h - \frac{\partial b_K}{\partial K_{inter}} \right) \cdot \frac{\partial K_{inter}}{\partial K_{mesh}} \delta K_{mesh} + {}^T \psi \cdot A_K \cdot \delta h = 0 \quad (8)$$

and summing Eqs. (8) and (4):

$$\delta F = {}^T \psi \cdot \left(\frac{\partial A_K}{\partial K_{inter}} \cdot h - \frac{\partial b_K}{\partial K_{inter}} \right) \cdot \frac{\partial K_{inter}}{\partial K_{mesh}} \delta K_{mesh} + \left({}^T \psi \cdot A_K + \frac{\partial F}{\partial h} \right) \cdot \delta h. \quad (9)$$

Secondly, we choose ψ so the second-term function of δh vanishes:

$${}^T A_K \cdot \psi + \frac{\partial F}{\partial h} = 0 \quad (10)$$

leading to the expression of δF as a function of δK_{mesh} :

$$\delta F = {}^T \psi \cdot \left(\frac{\partial A_K}{\partial K_{inter}} \cdot h - \frac{\partial b_K}{\partial K_{inter}} \right) \cdot \frac{\partial K_{inter}}{\partial K_{mesh}} \delta K_{mesh}. \quad (11)$$

The computation of the gradient of the objective function requires solving the two linear systems (Eqs. (3) and (10)) to determine h and ψ , respectively.

For finite elements, A_K and b_K depend directly on the mesh conductivity K_{mesh} . It formally corresponds to $K_{inter} = K_{mesh}$ and Eq. (11) simplifies to the classical formula:

$$\frac{\delta F}{\delta K_{mesh}} = \left(\frac{\partial(A_K \cdot h - b_K)}{\partial K_{mesh}} \right) \cdot \psi. \quad (12)$$

In most references, the expression of the adjoint state is given either by Eq. (12) (Carrera et al., 2005; Chavent, 1991; Sun, 1995) or by the variational analysis of Eq. (2) (Sun, 1995). Both of these expressions are specific to the finite-element framework. They cannot be used for finite volumes or finite differences. As shown by Eq. (11), finite-difference and finite-volume adjoint-state formulation must include the correcting term $\partial K_{inter} / \partial K_{mesh}$ that expresses the influence of the inter-mesh conductivity average. Numerically, the components of the gradient are computed independently:

$$\frac{\delta F}{\delta K_{mesh}^i} = {}^T \psi \cdot \left[\sum_{K_{inter}(K_{mesh}^i)} \left(\frac{\partial A_K}{\partial K_{inter}} \cdot h - \frac{\partial b_K}{\partial K_{inter}} \right) \cdot \frac{\partial K_{inter}}{\partial K_{mesh}^i} \right], \quad (13)$$

Table 1
Comparison of results of different methods to compute a gradient

Inter-mesh permeabilities	Finite difference (perturbation of 10-1) (%)	Finite difference (perturbation of 10-5) (%)	Finite difference (perturbation of 10-10) (%)	Direct derivation (%)	Adjoint-state method (%)
Harmonic mean	11.975	0.001	0.105	0.000	0.000
Arithmetic mean	5.282	0.001	0.055	0.000	0.000
Geometric mean	13.941	0.005	0.200	0.000	0.000

This example is based on a 2D aquifer modelled by a 2D-grid with 10×10 meshes. Results are mean differences with true gradient value expressed in percentage.

where the summation is performed only on the inter-mesh conductivities K_{inter} that depend on the i th mesh conductivity, K_{mesh}^i .

We have numerically validated this methodology in the groundwater flow framework, where K is permeability. K_{inter} is the inter-mesh permeability that is either the arithmetic, geometric or harmonic mean of the two neighbouring mesh permeabilities (Krueel-Romeu and Noetinger, 1995). If K_{inter} is taken as the harmonic average of the meshes' permeabilities, K_1 and K_2 , such as $K_{inter} = 2 \cdot K_1 \cdot K_2 \cdot (K_1 + K_2)^{-1}$, K_{inter} is strongly non-linear and leads to $\partial K_{inter} / \partial K_1 = 2(K_2 / (K_1 + K_2)^2)$. We found for 1D and 2D regular grids the same gradient of the objective function (1) using adjoint-state method (Eq. (13)), small perturbation analysis (Ho and Cao, 1991) or direct derivation (Table 1).

References

- Carrera, J., Alcolea, A., Medina, A., Hidalgo, J., Slooten, L.J., 2005. Inverse problem in hydrogeology. *Hydrogeology Journal* 13, 206–222.
- Chavent, G., 1991. On the theory and practice of nonlinear least-squares. *Advances in Water Resources* 14 (2), 55–63.
- Ho, Y.C., Cao, X.R., 1991. *Perturbation Analysis of Discrete Event Dynamic Systems*. Kluwer Academic, Dordrecht, The Netherlands.
- Jarny, Y., Özisik, M.N., Bardon, J.P., 1991. General optimization method using adjoint equation for solving multidimensional inverse heat conduction. *International Journal of Heat Mass Transfer* 34, 2911–2919.
- Krueel-Romeu, R., Noetinger, B., 1995. Calculation of internodal transmissivities in finite difference models of flow in heterogeneous porous media. *Water Resources Research* 31 (4), 943–959.
- Medina, A., Carrera, J., 2003. Geostatistical inversion of coupled problems: dealing with computational burden and different types of data. *Journal of Hydrology* 281 (4), 251–264.
- Sun, N.Z., 1995. *Inverse Problems in Groundwater Modeling*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Tarantola, A., 1987. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier, New York, NY.
- Townley, L.R., Wilson, J.L., 1985. Computationally efficient algorithms for parameter estimation and uncertainty propagation in numerical models of groundwater flow. *Water Resources Research* 21 (12), 1851–1860.

1 **A Universal Model of Fracture Scaling and its** 2 **consequence for crustal hydro-mechanics**

3 P. Davy¹, R. Le Goc^{1,2}, C. Darcel², O. Bour¹, J.R. de Dreuzy¹, and R. Munier³

4 ¹ *Géosciences Rennes, UMR 6118 CNRS and University of Rennes I, Campus de Beaulieu,*
5 *35042 Rennes cedex, France*

6 ² *Itasca Consultants S. A., groupe HClasca, 64 chemin des mouilles, 69130 Ecully, France*

7 ³ *Svensk Kärnbränslehantering AB, Box 5864, SE-102 40 Stockholm, Sweden*

8 **Abstract**

9 We argue that most of fracture systems are spatially organized according to two main regimes: a
10 “dilute” regime for the smallest fractures, where they can grow independently of each other, and
11 a “dense” regime, for which the density distribution is controlled by the mechanical interactions
12 between fractures. We derive a density distribution for the dense regime by acknowledging that,
13 statistically, fractures do not cross a larger one. This very crude rule, which expresses the
14 inhibiting role of large fractures against smaller ones but not the reverse, actually appears to be a
15 very strong control on the eventual fracture density distribution since it results in a self-similar
16 distribution whose exponents and density term are fully determined by the fractal dimension D
17 and a dimensionless parameter γ that encompasses the details of fracture correlations and
18 orientations. The range of values for D and γ appears to be extremely limited, which makes this
19 model quite universal. This theory is supported by quantitative data on either fault or joint
20 networks. The transition between the dilute and dense regimes occurs at about a few tenths of
21 kilometers for faults systems, and a few meters for joints. This remarkable difference between
22 both processes is likely due to a large-scale control (localization) of the fracture growth for
23 faulting that does not exist for jointing. Finally, we discuss the consequences of this model on
24 the flow properties, and show that these networks are in a critical state, with a large number of
25 nodes carrying a large amount of flow.

26 **1. Introduction**

27 Measuring the complexity of fracture networks has been an issue for the last twenty years with
28 consequences on brittle strength, rock permeability or earthquake dynamics [*Allegre et al.*, 1982;
29 *Crampin*, 1999; *Davy et al.*, 2006; *King*, 1983; *Renshaw*, 1999; *Turcotte*, 1986]. A special focus
30 has been given to the scaling properties of fracture networks that link rock samples to crustal
31 scales (see [*Bonnet et al.*, 2001] and references therein). Power laws have been found to likely
32 describe ubiquitously the scaling properties of fracture density, with two consequences: fracture
33 networks are scale-free objects – power law is the only mathematical function that does not
34 require a scale parameter –, and long fractures are much more numerous than for any else
35 distribution type (lognormal in particular, which was often taken as the fitting function).
36 Although ubiquitous in natural systems, the reason why this power-law scaling emerges is still
37 an issue. It is supposed to derive from a complex, a few say critical, self-organized dynamics,
38 one of whose characteristics is to produce long-range correlated patterns that make large-scale
39 structure probable and “genetically” (thus statistically) linked to smaller ones [*Bak et al.*, 1988;
40 *Bak and Tang*, 1989; *Sornette et al.*, 1990; *Sornette*, 2006].

41 In this paper, we aim at giving a theoretical framework for fracture scaling models based on
42 very basic properties. It is beyond the scope of this paper to discuss how complex forms derive
43 from complex processes; we just acknowledge a few basic properties of fracturing, and we
44 elaborate on the consequences. The points are:

45 – A fracture induces a perturbation of the stress (and strain) field that modify – both enhance
46 and decrease – the occurrence probability for another fracture to initiate and/or grow at its
47 vicinity. The extent of this perturbation is of the order of the fracture length (see amongst
48 others [*Atkinson*, 1987; *Chinnery and Petrak*, 1968; *Segall and Pollard*, 1980]).

49 – For quasi-static fracturing, most of the fracture growth – with the exception of boundary-
50 related fractures – is controlled by the perturbation of stress field induced by the fracture
51 itself and by neighbor fractures [*Atkinson*, 1987]. If the fracture does not produce by itself the
52 reasons for stopping its own growth, it is reasonable to think that the eventual fracture pattern

53 and its scaling law result from a basic property of the fracture-to-fracture mechanical
54 interactions [Renshaw and Pollard, 1994; Segall and Pollard, 1983; Segall, 1984].

55 We first discuss both end-members of the fracturing process: the dilute case where fractures are
56 likely growing independently from each other, and the dense case, where the fracture growth is
57 fully controlled by the presence of surrounding fractures. We then combine both into a
58 geometrical model of fracture organization that likely applies over the whole range of fracture
59 scales. A key result is that the large-scale part of this model appears to be almost universal, i.e.
60 that both the scaling exponents and the density term are, although not strictly fixed, highly
61 constrained and almost independent of the details of the fracturing process, of the boundary
62 conditions, of the intensity of fracturing, etc. The relevance of this model can thus be easily
63 checked with actual fracture distribution; we show that it successfully fits the density
64 distribution of fracture networks from meter to tenths of kilometer scales. Eventually we discuss
65 the consequences of the derived organization in terms of fracture connectivity and permeability.

66 **2. The fracture organization model from fracture growth** 67 **consideration**

68 1.1. The “dilute” case

69 When the density of fracture is low, fractures can freely grow and the eventual size distribution
70 depends on the length dependency of the growth-rate function. If the growth rate is a power law:

$$71 \frac{dl}{dt} \sim l^a, \quad (1)$$

72 as it is likely (see [1987; Lyakhovsky, 2001] and references therein for a discussion about rock
73 mechanics), then the fracture density distribution $n(l)$ eventually scales as $n(l) \sim l^{-a}$ if $a \neq 1$, and
74 $n(l) \sim \exp(-l/l_0)$ if $a = 1$ [Sornette and Davy, 1991]. Estimating a is an open issue, especially for
75 geological systems. Lyakhovsky [Lyakhovsky, 2001] reported exponent a (quoted $m/2$ in the
76 article) in the range 1-2.5 from laboratory tests on rock samples, and numerical simulations for
77 quasi-static crack growth (actually 1.3-1.8 for the latter), but he mentioned smaller exponents

78 when the growth is controlled by the transport of reactive species to the crack tip, and larger
 79 ones for large crack lengths. *Sornette and Davy* [*Sornette and Davy*, 1991] argue, from
 80 thermodynamic arguments, that the exponent a is likely 2 if the damping force that
 81 counterbalances the applied stress does not depend on l and $\frac{dl}{dt}$. A few experiments on rock-like
 82 materials (dry sand in [*Davy et al.*, 1990; *Davy et al.*, 1995; *Sornette et al.*, 1993], plaster in
 83 [*Mansfield and Cartwright*, 2001]) and numerical simulations [*Cowie et al.*, 1995; *Hardacre*
 84 *and Cowie*, 2003a; b] also address this issue by analyzing the evolution of quite complex fault
 85 network. The power-law model was found to correctly fit fracture length distributions even for
 86 low-density networks, except for the largest fracture; most of the 2-D exponents (i.e. measured
 87 on fault traces) are close to 2.

88 Note that the growth rate equation (1) includes fracture linkage mechanisms, ubiquitously
 89 observed in fracture networks, and partly responsible of the displacement-length relationship
 90 ([*Cladouhos and Marrett*, 1996; *Cowie and Scholz*, 1992; *Schlagenhauf et al.*, 2008; *Schultz*,
 91 2000; *Schultz and Fossen*, 2002; *Xu et al.*, 2006] amongst others).

92 Thus a general expression of the density length distribution of “dilute” fracture networks can be:

$$93 \quad n_{\text{dilute}}(l, L) = \alpha l^{-a} L^D, \quad (2)$$

94 where $n_{\text{dilute}}(l, L)$ is the number of fractures in a system of size L (either an area in 2-D or a
 95 volume in 3-D) whose length l' is in the range $[l, l+dl]$. D is formally the mass (or correlation)
 96 dimension of the fracture-center network that is smaller or equal to the topological dimension
 97 (see [*Bonnet et al.*, 2001; *Bour et al.*, 2002] for more explanations). As the exponent a , the D is
 98 supposed to be related to some fundamental properties of the fracturing process. The parameter
 99 α is the density term that is likely increasing during fracture growth.

100 Note that the power-law distribution is the eventual result of the growth equation (1), which
 101 means that it is reached only after a certain time. Before, the fracture length distribution is likely
 102 depending on the initial conditions or on the fracture nucleation process.

103 1.1. **The dense case**

104 For dense networks (we will define later how “dense” they must be), fracture lengths are likely
105 controlled by the fracture-to-fracture mechanical interactions. The cause of limiting fracture
106 growth has been widely discussed in the past [*Crampin*, 1994; 1999; *Pollard and Aydin*, 1988;
107 *Renshaw and Pollard*, 1994; *Renshaw*, 1997; *Renshaw et al.*, 2003]. A detailed knowledge of
108 these interactions is hardly measurable, and beyond the capacity of numerical models, especially
109 when it exists a large numbers of fractures. Even if the calculation was feasible, it would be
110 difficult to relate the characteristics of the stress field to the eventual fracture distribution. To
111 overcome this difficulty, we develop a very simple model that is based on basic geometrical
112 rules of the fracture-to-fracture interactions for dense networks. By doing this, we aim at
113 capturing the fundamental characteristics of network geometry that are consistent with the
114 mechanical interactions.

115 We first point out that the very basic reason for a fracture to stop growing is to meet another
116 fracture. This assumption is quantitatively supported by the large numbers of T-shape
117 intersections (when a fracture stops on another) compared to X-shape (both fractures cross each
118 other), which makes fracture pattern really different from Poissonian or random equivalents. As
119 an example of this assessment, the number of T-tips per fracture reported on a series of Swedish
120 outcrops (see a detailed description below) ranges from 0.6-1.6 with an average of about 1.2,
121 which is by far much larger than any Poissonian realization (for which a T-tip configurations is
122 unlikely).

123 Secondly we acknowledge the inhibiting role of large fractures against smaller ones, but we
124 assume that the reverse proposition is not likely to occur.

125 Both these assumptions amount to saying that the length of a fracture is about the distance to the
126 closest larger fracture. In the following paragraphs, we demonstrate that this rule leads to a self-
127 similar distribution of the fracture density distribution.

128 We consider the general case of a fracture network distribution characterized by a density
129 distribution $n(l, L)$, which is the number of fractures of length in the range $[l, l+dl]$ within a
130 volume (or surface in 2D) of typical size L . The cumulative distribution,

131 $C(l, L) = \int_l^\infty n(l', L) dl',$

132 is the number of fractures larger than l . If D is the network dimension (which can be a non-
133 integer value for fractal networks), the average distance between the centroids of a fracture of
134 length l and a larger one is [*Bour and Davy, 1999*]:

152 The second interesting property of this distribution is that its density depends only on two
153 parameters of the initial growing pattern: the correlation dimension D and the geometric
154 parameter γ .

155 D is upper bounded by the dimension of the growth space, but can be smaller because of
156 fracture-to-fracture correlations.. A collection of fractal dimensions have been measured mainly
157 from 2d outcrop or lineament maps (see the review in [Bonnet *et al.*, 2001]) with values in 2d
158 ranging from 1 to 2, and reasonably from 1.7 to 2 (the small values are associated with networks
159 where only a few large fractures are mapped, making the measure somehow meaningless). Little
160 is known about the factors that control D (see however [De Arcangelis *et al.*, 1989; Hansen *et*
161 *al.*, 1991; Herrmann and Roux, 1990]). As it measures the fracture-center scaling organization,
162 we expect it to express more the physics of the nucleus phase than of the growth phase. Large-
163 scale localization such as encountered in faulting is the type of correlation that may lower D .
164 Note that D is generally measured from the pair correlation function [Hentschel and Proccacia,
165 1983], which is very well defined for large enough network [Vicsek, 1992]. D was found to be
166 smaller for fault network [Bour and Davy, 1999] than for joints [Bour *et al.*, 2002], consistent
167 with the existence of large-scale correlation for the former.

168 γ is expected to vary with fracture orientation distribution. For instance subparallel fracture sets
169 are supposed to encounter another large fracture at a larger distance than perpendicular fracture
170 sets. We present in a next paragraph 2-D simulations that give insight on the range of γ values.

171 Above all, γ encompasses the complexity of fracturing that departs from the simplistic
172 formulated rule (“*a fracture stops growing when intersecting a larger one*”), which is basically
173 more probabilistic than absolutely true.

174 As already mentioned, the distribution is rather universal in the sense that neither D nor γ are
175 supposed to vary a lot. We thus call this model UFM for Universal Fracture Model, the basic
176 equation (4) the UFM equation, and (5) the UFM distribution.

177 2.1. **The transition from “dilute” to “dense”**

178 We now analyze the transition from “dilute” to “dense” network, assuming that they follow
179 equations (2) and (5) respectively. We study the case $a < D+1$, which is likely in geological
180 systems (see [Bonnet *et al.*, 2001] and the discussion above). Note that the reverse case ($a > D+1$)
181 would end up to a transition from “dense” to “dilute”.

182 We first analyze the distance between a fracture and its larger neighbor in the “dilute” case. This
183 can be done by using equation (3) with $C(l, L)$ derived from equation (2):

$$184 \quad d_{\text{dilute}}(l) \sim \left(\frac{a-1}{\alpha} \right)^{\frac{1}{D}} l^{\frac{a-1}{D}}$$

185 Since $a < D+1$, the distance increases with l at a rate slower than l (see Figure 1).

186 It exists a finite length l_c for which $d_{\text{dilute}}(l_c) = \gamma l_c$. For fractures with a length $l < l_c$, the distance
187 between fractures (d_{dilute}) is larger than the fracture length; thus statistically fractures are not
188 intersecting their larger neighbor and thus can grow freely according to a growth equation like
189 (2). For $l > l_c$, fractures stop growing freely since they are likely crossing a larger one. Their
190 length cannot exceed the distance $d(l)$, and thus they obey the UFM length distribution. The
191 transition length is given by the expression:

$$192 \quad l_c = \left(\frac{a-1}{\alpha} \gamma^D \right)^{\frac{1}{D+1-a}} \quad (7)$$

193 Increasing the density term α leads to a decrease of the fracture-to-fracture distance, and thus to
194 a decrease of l_c , and to a larger scale range for the “dense” regime.

195 In short, we propose that there is a break in the length distribution scaling due to the transition
196 between a freely growing (“dilute”) regime, and a regime where the fracture-to-fracture
197 interactions are prevailing.

198 **3.1. Numerical simulations**

199 We perform basic numerical simulations with 2d networks both to check mathematics and give
200 a few insights on the γ parameter. We assume that the UFM rule (“*a fracture stop growing when*
201 *intersecting a larger one*”) is strictly ensured. Fractures are first stochastically drawn according
202 to a “dilute” length distribution with $a=2.3$. Then when a fracture intersects a larger one, we
203 remove the shortest tip. The simulations were made with 2D networks with different
204 orientations and initial distributions. An example of the eventual length distribution is given in
205 Figure 2.

206 The general principles developed above are validated by numerical simulations and in particular
207 the UFM distribution. For uniformly distributed orientations, or two-orthogonal sets, γ is equal
208 to about 1, giving a density of the UFM distribution $D\gamma^D=2$. For two fracture sets with an angle
209 of 20° , γ is larger, as expected, in the range 1.5-1.7, giving a density term $D\gamma^D$ of 5-6.

210 **3. Does UFM applies to geological fracture distribution**

211 We argue that the UFM theory is ubiquitous in natural fracture networks, keeping in mind that
212 the testing of any scaling distribution requires an exhaustive sampling of fracture networks that
213 is barely achieved in geological studies [Bonnet *et al.*, 2001]. We analyze three fracture sets for
214 which we are confident about the data completeness and the mapping method.

215 Note that the UFM theory has been developed for 3D networks; but the only data that allows us
216 to test the theory are from 2D fracture traces. We demonstrate in the appendix that equation (5)
217 remains valid for fracture traces if D is replaced by $D-1$ and γ slightly modified.

218 The first data set is the joint network of the Hornelen basin (Norway) that was mapped carefully
219 at different scales [Odling, 1997], and whose scaling properties were thoroughly analyzed in
220 [Bour *et al.*, 2002]. Seven fracture networks were mapped with outcrop scales ranging from
221 18 m to 720 m. Although the mapping technique was different for small and large outcrops, all
222 the density distributions were found to be consistent with the density equation:

223
$$n(l, L) = 4.5 l^{-2.8} L^{1.8} \quad (8).$$

224 The density term, $\alpha=4.5\pm 1$, corresponds to $\gamma=1.7\pm 0.2$ according to eq.(4). *Odling* [1997] argue
225 that 1 m is a possible physical limit to this power-law behavior for this fracture system. No
226 upper limit to this power-law scaling was detectable.

227 The second site series was developed by the Swedish Nuclear Fuel and Waste management
228 company, SKB, as part of an ongoing investigation conducted to locate a repository for spent
229 nuclear fuel [*Fox et al.*, 2007; *La Pointe et al.*, 2008; *SKB*, 2004a; b; *Stephens et al.*, 2008;
230 *Wahlgren et al.*, 2008]. The area was deeply investigated with both fracture maps at outcrop
231 scales (0.5-10 m) and regional scales (100 m-10 km) and a host of deep (~1 km) cored
232 boreholes. At the outcrop scale, fractures are mostly joints and barely faults with a detectable
233 displacement. The large-scale (100 m-10 km) lineaments that are widespread in the
234 Scandinavian shield are quite heavily fractured shear zones or thick faults with a gouge. The
235 caution with which outcrops have been mapped, with an exhaustive sampling of fractures whose
236 lengths is larger than 50 cm, enables to give an estimate of the scaling exponent despite the
237 limited range of scales. This was not possible for lineament maps, where an exhaustive
238 detection of fractures, and of fracture limits, is much less reliable. The fracture density
239 distribution is shown in Figure 3. A fit is calculated for each outcrop map (dashed lines in
240 Figure 3), and by comparing fracture networks at different scales (bold lines in Figure 3). The
241 exponents measured on outcrops are $a=2.2$, for all of them except one, where a higher density of
242 small fractures is visually detectable (Figure 4). For this latter outcrop, the “through-scale”
243 model is consistent with the fit with the following scaling law:

$$244 \quad n(l, L) = 4(\pm 1) l^{-3} L^2 \quad (9).$$

245 The last example is the San Andreas fault map published in [*Davy*, 1993] from the compilation
246 of [*Jennings*, 1988]. In terms of fracturing mechanisms, we may not expect faulting to be
247 similar to jointing, and we thus expect highly different density distributions. But, except the
248 differences in orientations, the argument developed in the UFM theory should remain valid and
249 leads to similar conclusions. [*Davy*, 1993] and [*Bour and Davy*, 1999] noticed that the length
250 distribution has a power-law scaling only for faults smaller than 10 km (Figure 5). The

251 departure from power law for large faults have been interpreted as a segmentation of the largest
252 faults [Davy, 1993], or as a truncating effect [Bour and Davy, 1999]. The exponent a of the
253 small-scale power law ranges between 1.5 and 2, depending on the correction applied for large
254 fractures. A correlation dimension $D=1.7$, also smaller than the dimension of joints, have been
255 derived from a correlation analysis. We apply the UFM equation to the density distribution
256 (solid line in Figure 5). The fit is very good for faults larger than 20 km up to the largest
257 recorded fault of 100 km.

258 The UFM equation obtained for the San Andreas, with $L=450$ km, writes as:

$$259 \quad n(l, L) = 4 l^{-2.7} L^{1.7} \quad (10),$$

260 All the equations (8), (9), and (10) are consistent with equation (5), with the same density term
261 and $a=D+1$. They just differ in the fractal dimension D .

262 At last, we mention that the break in the scaling distribution, which is constitutive of the
263 presented theory, has also been observed on the displacement-length relationship of fractures.
264 Based on the analysis of the Krafla fracture swarm (Iceland), which is assumed to be primarily
265 jointing (opening mode) by [Hatton *et al.*, 1994], [Main *et al.*, 1994] and [Renshaw and Park,
266 1997] found a more-than-linear scaling for fractures smaller than 2.5-70 m, and a linear or less-
267 than-linear scaling for larger fractures. The interpretation of the scaling break is diverse:
268 enhancement of small-fracture aperture by largest one [Renshaw and Park, 1997], thermal
269 effects at large-fracture tips [Hatton *et al.*, 1994], or co-operative behavior of large-fracture
270 growth [Main *et al.*, 1994]. The self-similar model, which entails displacement proportional to
271 length, is consistent with the linear relationship observed for large fractures.

272 **4. More about faults and faulting**

273 The large value of the crossover scale l_c for fault networks (~ 20 km for the San Andreas system
274 shown in Figure 5) entails that most of faults are in the “dilute” regime. The dense regime may
275 appear rather as an abnormal distribution queue of the “dilute” regime than as a broad scaling
276 relationship. This effect was already emphasized on geological networks and experiments. It has

277 been modeled by a gamma or exponential function [Ackermann *et al.*, 2001; Cowie *et al.*, 1994;
278 Davy, 1993; Hardacre and Cowie, 2003a; b; Spyropoulos *et al.*, 1999], or by a censoring
279 function [Bour and Davy, 1999; Pickering *et al.*, 1995]. Apart from the latter that is a pure
280 statistical effect, the main mechanical causes invoked for this cut-off are the segmentation of
281 large fractures [Davy, 1993], or some finite-size effects induced by the thickness of the
282 mechanical layer [Ackermann *et al.*, 2001; Spyropoulos *et al.*, 1999]. In the next paragraph of
283 this section, we discuss the pertinence of the UFM theory and fit relative to processes and data.
284 We illustrate the discussion with the experiments from [Spyropoulos *et al.*, 1999], [Ackermann
285 *et al.*, 2001], and [Davy *et al.*, 1995], and with the numerical simulations from [Hardacre and
286 Cowie, 2003a].

287 In the experiments from [Spyropoulos *et al.*, 1999] and [Ackermann *et al.*, 2001], the fracture
288 length distribution evolves toward an exponential function. Cracks are propagating within a
289 brittle layer that is stretched from below by a thick rubber pad. This strong coupling between the
290 brittle layer and the rubber pad controls the eventual spacing between cracks [Spyropoulos *et al.*
291 *et al.*, 1999]; the brittle thickness is the mechanical length that eventually controls the length
292 distribution. The UFM theory, which postulates that fracture growth is limited by the interaction
293 between fractures, is not appropriate to describe this mechanical system, and thus it cannot
294 predict the dependency on the layer thickness. This is valid for all subparallel joints whose
295 spacing appears to be controlled by a mechanical layer [Wu and Pollard, 1995].

296 The numerical simulations from [Hardacre and Cowie, 2003a; b] seem to end up to similar
297 conclusions: the exponential distribution successfully describes the length distribution of active
298 faults at the later stages. But the boundary conditions are different from those of the
299 aforementioned experiments since simulations are in 2D and faults propagate in mode II (i.e.
300 growth is parallel to length). The only characteristic length scale is the system size, which is a
301 limit for fault length rather than a mechanical control. Even if the UFM distribution is *a priori*
302 not suitable for fitting data, we check how it may explain part of the data shown in [Hardacre
303 and Cowie, 2003a]. We plot the cumulative UFM equation (11):

$$C(l, L) = \int_l^{l_{\max}} n(l', L) dl' = \int_l^{l_{\max}} D \gamma^D l'^{-D+1} L^D dl' = \gamma^D \left(\frac{l}{L} \right)^{-D} \quad (11)$$

305 on the graph of the fig.7 of [Hardacre and Cowie, 2003a], with $\gamma^D=2$ and $D=1.8$, as found in the
 306 previous paragraph for fault networks, and a length $L=7$ km, that is the largest fault length
 307 (Figure 6). Although any power-law exponent cannot be derived from the graph, we find that
 308 the UFM equation fits well the distribution extreme at large length. Note however that these
 309 simulations must be used with caution because 1) faults are not explicitly modeled and the
 310 procedure to identify them is quite complex, and 2) the number of faults is small (~ 100), which
 311 casts some doubt about the relevance of the large fault distribution.

312 We do a similar exercise on the experiments from [Davy et al., 1990; Davy et al., 1995;
 313 Schueller et al., 2005; Sornette et al., 1993], whose boundary conditions are of the same nature
 314 as those used in [Hardacre and Cowie, 2003a]. Again, we find that the UFM distribution fits
 315 well the large end-member of the length distribution (Figure 7).

316 Even if these two examples are successful in demonstrating the validity of the UFM fit at large
 317 lengths, we acknowledge that a fit is far to make a sound scientific proof. We also point out that
 318 the scaling information contained in UFM (l^{-D+1}) is not actually testable on fault systems
 319 because of the very small range of length concerned by the fit. The UFM equation is neither
 320 better nor worse than an exponential function considering the large scattering that is intrinsic to
 321 the density of scarce data (Figure 7).

322 Nevertheless, the UFM equation provides a good fit without any tuning parameter (considering
 323 that $D\gamma^D=4$ has been obtained for all networks, that the fractal dimension can be calculated
 324 independently, and that L is the system size). In contrast, the exponential fit requires two tuning
 325 parameters, one of which is a characteristic length and the other a density. In Figure 7, the
 326 exponential fit yields a characteristic length of 3 cm, much smaller than L (35-75 cm) and larger
 327 than the brittle layer thickness (5 mm).

328 These considerations give credit to the ability of the UFM theory to describe fault distribution.
 329 This is not a definite conclusion but definitely an orientation for further work.

330 **5. Differences between faulting and jointing**

331 Even if a successful analysis of a few networks is not a definite proof, the UFM seems to be a
332 promising theory for both fault and joint networks although both fracture systems present large
333 differences in terms of processes, density, orientation and spatial organization. In this paragraph,
334 we propose to discuss further the differences between faulting and jointing.

335 First of all, the reason why the UFM theory could be universal partly stays in the simplicity of
336 its basic rule (“*a fracture stop growing when intersecting a larger one*”). Nevertheless the
337 distribution of joint (Scandinavian outcrops from Norway (Hornelen basin) and Sweden
338 (Simpevarp, Laxemar)) and fault networks (San Andreas, USA) that we analyzed present a few
339 significant differences (see also [Bonnet *et al.*, 2001]):

- 340 – l_c is much larger for faults (~20 km) than for joints (1-10 m). A small l_c indicates a higher
341 density in the “dilute” regime (or a larger density parameter α in eq.2).
- 342 – The orientation distribution of faults and joints is quite different. Faults tend to be organized
343 around two orientation poles while the orientation distribution of joints is broader.
- 344 – The fractal dimension of faults is slightly smaller (1.5-1.8 on 2D outcrops) than of joints (1.8-
345 2). Even if this difference is not so large, it is significant and emphasizes a higher clustering
346 of faults than of joints. These results highlight the long-range correlations of the faulting
347 process that emerge from the spatial organization of deformation and/or stress [Davy *et al.*,
348 1990; Sornette *et al.*, 1993]. Note that the density distribution, even if highly quantitative, is
349 quite weak in revealing the details of the fracture organization such as correlation patterns.

350 How can we explain these factual differences? To our knowledge, this issue is still totally open
351 in the literature. We cannot pretend to have a definite answer; we just point out a few reasons
352 that deserve being tested in further work.

353 First of all, the UFM predicts that only a few faults propagate throughout the entire system. This
354 is true both for faults and joints. The fact that l_c for fault systems is close to the system size
355 means that only the largest faults are close enough to interact mechanically. A second
356 observation is that fault networks stop growing when large-scale deformation localization
357 appears [Hardacre and Cowie, 2003a; Sornette *et al.*, 1993]. This last stage is concomitant to

358 the development of the largest faults. The fact that l_c remains large even at the eventual stages of
359 the fault development emphasizes a quasi inhibition of the growth of new small faults after the
360 eventual localization. This is well demonstrated by the numerical simulations of [Hardacre and
361 Cowie, 2003a; b], where both inactive and active faults are analyzed. All this points are
362 relatively well described and justified by experiments, numerical simulations and theory on
363 faulting.

364 The case of jointing is more difficult since little is known about it (see however [Pollard and
365 Aydin, 1988]). In particular there is no indication, to our knowledge, about the general evolution
366 of the network growth, whether it eventually stops or not. The fact that the UFM distribution
367 applies on a very large range of scales indicates 1) that fracture density is limited by the
368 fracture-to-fracture mechanical interactions in a way quantitatively similar to the process that
369 we describe above, and 2) that, in contrast to faulting, the “dense” regime applies to scales much
370 smaller than the system size, meaning that the growth of network density is not limited the
371 large-scale localization, as it is for faulting. We suspect that this difference between faulting and
372 jointing is related to the existence of internal stresses in the latter, such as bulk volume changes
373 or gravity forces. [Schmittbuhl and Roux, 1994] demonstrate that internal stresses have
374 important consequences on fracture processes and scaling, but they do not calculate explicitly
375 fracture lengths. In particular, they conclude that, “when the internal stresses become larger, the
376 behavior displays a non-vanishing plasticity, as well as a diffuse damage” (sic). This damage,
377 and the role played by micro-cracking, is a large difference with faulting and consistent with a
378 high density of micro-cracks and thus a small value of l_c . Apart from this study, there are a few
379 studies that identify an increase of the density of small faults, generally because of the coupling
380 with another process such as ductile materials [Davy and Cobbold, 1991; Davy et al., 1995], or
381 dynamic waves [Poliakov and Herrmann, 1994].

382 To conclude, we conjecture that the difference between faulting and jointing is the capacity of
383 small fractures to grow even after the largest fractures appear, and we suspect that this was
384 made possible in joints due to internal stresses.

385 **6. Highlighting the consequences of the UFM organization**

386 Because the UFM is based on a local condition of fracture connectivity, it is likely having
387 important consequences on network connectivity, and thus on transport properties
388 (permeability, dispersivity, etc.). The connectivity of fracture networks with a power-law length
389 distribution has been studied by [*Bour and Davy*, 1997; 1998; *Darcel et al.*, 2003; *Renshaw*,
390 1999]. Large-scale connectivity is ensured both by the presence of large fractures and by the
391 clustering of smaller ones. In d dimensions, connectivity was found to be controlled by the
392 percolation parameter:

$$393 \quad p = \int_{l_{\min}}^{l_{\max}} \frac{n(l, L) \cdot l_{\text{inc}}^d}{L^d} dl \quad (12),$$

394 where l_{inc} is the part of fracture length that is included in the system of size L [*Bour and Davy*,
395 1997; 1998]; l_{\min} and l_{\max} are the smallest and largest fracture length, supposed to be much
396 smaller and larger (respectively) than any characteristic length of the problem. Equation (12) is
397 based on concepts of excluded volume (l_{inc}^d) [*Balberg et al.*, 1984]; it has been demonstrated to
398 correctly quantify connectivity threshold for 2D and 3D networks ($d=2$, and $d=3$, respectively).
399 The extension to fractal networks is still an issue: [*Berkowitz et al.*, 2000] postulated that d can
400 be replaced by the fractal dimension D of network in equation (12); but [*Darcel et al.*, 2003]
401 demonstrated from numerical simulations that the resulting expression is not appropriate for 2D
402 fractal networks. They actually found that equation (12) correctly describe percolation only if
403 $a \leq D+1$, and if d is the dimension of the embedding space (that is always larger than the fractal
404 dimension D). If $a > D+1$, the connectivity of networks is dominated by the lacunarity of fractal
405 networks, which make infinitely large network always disconnected. Here we consider the
406 former case ($a \leq D+1$), and assume that the expression (12) is valid for 2D and 3D fractal
407 networks, with d the dimension of space (either 2 or 3).

408 Since l_{inc} is l if $l < L$, and $l_{\text{inc}} \propto L$ if $l > L$, the integral can be split into both contributions of small
409 and large fractures.

410
$$p \approx \int_{l_{\min}}^L \frac{n(l,L) \cdot l^d}{L^d} dl + \int_L^{l_{\max}} n(l) dl ,$$

411 In the case $a < D+1$, which is the likely conditions of the “dilute” regime, both terms, as well as
 412 p , are varying as L^{D+1-a} with a few important consequences:

413 – Connectivity increases with the system size L . Thus, statistically, small systems are
 414 unconnected while large ones are fully connected [Bour, 1997; Bour and Davy, 1998; Darcel
 415 *et al.*, 2003].

416 – The ratio between both contributions (large and small fractures) is scale independent.

417 – Because the contribution of large fractures is not negligible, the unconnected-to-connected
 418 transition is no more equivalent to a second-order phase transition [de Dreuzy *et al.*, 2001].

419 Its width in terms of range of p values does not vanish for very large systems in contrast with
 420 the classical percolation theory, where the percolation threshold is equivalent to a 2nd-order
 421 phase transition [Stauffer, 1979].

422 The self-similar model with $a=D+1$, as is the UFM, displays similar property to the classical
 423 percolation theory. The percolation parameter is constant (i.e. independent of L) [Darcel *et al.*,
 424 2003], as well as the number of fractures larger than the system-size L :

425
$$N(l > L) = \int_L^{l_{\max}} n(l,L) dl = \int_L^{l_{\max}} D\gamma^D l^{-(D+1)} L^D dl = \gamma^D ,$$

426 The likely model for real fracture networks contains both previous cases: $a < D+1$ for small
 427 fractures in the dilute regime, and a self-similar distribution for large fractures in the dense
 428 regime. We calculate the percolation parameter by taking equation (2) as representative of the
 429 dilute regime, and equation (5) of the UFM distribution; the transition length between both
 430 regimes is l_c defined in equation (7). The percolation parameter writes as the sum of three
 431 integrals whose bounds and values depend on the relative position of the crossover scale l_c to
 432 system size L :

433
$$p|_{L < l_c} = \int_{l_{\min}}^L \frac{n_{\text{dilute}}(l,L) \cdot l^d}{L^d} dl + \int_L^{l_c} n_{\text{dilute}}(l,L) dl + \int_{l_c}^{l_{\max}} n_{\text{dense}}(l,L) dl$$

$$434 \quad p|_{L>l_c} = \int_{l_{\min}}^{l_c} \frac{n_{\text{dilute}}(l, L) \cdot l^d}{L^d} dl + \int_{l_c}^L \frac{n_{\text{dense}}(l, L) \cdot l^d}{L^d} dl + \int_L^{l_{\max}} n_{\text{dense}}(l, L) dl$$

435 By using equations 2, 5, and 7, and assuming $l_{\min} \ll L$ and $l_{\max} \gg L$, we obtain:

$$436 \quad p|_{L<l_c} = \frac{a-1}{d+1-a} \gamma^D \left(\frac{L}{l_c}\right)^{D+1-a} + \left[\gamma^D \left(\frac{L}{l_c}\right)^{D+1-a} - \gamma^D \left(\frac{L}{l_c}\right)^D \right] + \gamma^D \left(\frac{L}{l_c}\right)^D = \frac{d\gamma^D}{d+1-a} \left(\frac{L}{l_c}\right)^{D+1-a}$$

$$437 \quad \text{If } D \neq d, p|_{L>l_c} = \frac{a-1}{d+1-a} \gamma^D \left(\frac{L}{l_c}\right)^{D-d} + \frac{D}{d-D} \gamma^D \left(1 - \left(\frac{L}{l_c}\right)^{D-d}\right) + \gamma^D = \frac{d}{d-D} \gamma^D \left(1 - \frac{D+1-a}{d+1-a} \left(\frac{L}{l_c}\right)^{D-d}\right)$$

$$438 \quad \text{If } D = d, p|_{L>l_c} = \frac{a-1}{d+1-a} \gamma^d + d\gamma^d \ln\left(\frac{L}{l_c}\right) + \gamma^d = d\gamma^d \left(\frac{1}{d+1-a} + \ln\left(\frac{L}{l_c}\right)\right)$$

439 For systems smaller than l_c , the percolation parameter increases with the system size almost as it
440 does for the simple power-law model. At $L=l_c$, the percolation parameter is equal to:

$$441 \quad p|_{L=l_c} = \frac{d}{d+1-a} \gamma^D$$

442 For larger systems ($L>l_c$), the percolation parameter slightly increases with L as a constant plus
443 an additional term that can be either a slow power law (for $D<d$), or a logarithmic function (for
444 $D=d$). This varying term comes from the contribution of fractures smaller than L . The graph in
445 Figure 8 shows the percolation parameter as a function of the ratio L/l_c for 2D fracture traces
446 with the parameters derived from the analysis of fracture maps presented in the previous
447 paragraph ($\gamma^D \approx 2$, and $d+1-a \approx 0.8-1$). At $L=l_c$, the percolation parameter is slightly below
448 the percolation threshold in 2D [Balberg *et al.*, 1984; Bour and Davy, 1997; Robinson, 1984].
449 The percolation threshold is reached for L between 1.3-2 times l_c . Note that α , the density term
450 in the dilute regime, is taken into account by l_c .

451 The above analysis gives a rationale for the intuitive statement that fracture networks are likely
452 below percolation threshold in the “dilute” regime, and above but close to the percolation
453 threshold in the dense regime.

454 However the UFM is not only a self-similar model; it also contains spatial correlations between
455 fractures that are likely playing a role on large-scale connectivity, and on connectivity changes.

456 Since the UFM is by construction very close to percolation threshold, we expect small change
457 of the fracture position or permeability to have very large consequences on flow.

458 To illustrate this statement, we calculate flow on the UFM simulations presented above. We
459 compare it to 2 models: small-crack networks, whose behavior is described by the classical
460 percolation theory, and a power-law length distribution model with an exponent $a=2.2$. As a
461 measure of the connectivity organization, we calculate the distribution of incoming flow at the
462 network nodes at percolation threshold and slightly above (Figure 9). The small-fracture case is
463 indicative of what this distribution may look like. At the percolation threshold, there are a large
464 number of nodes carrying a large amount of total flow. Removing these nodes will thus entail
465 large consequences on the flow organization, up to the total disconnection of the system. Above
466 percolation threshold, the flow is much more distributed over nodes; at $p=1.25p_c$ for instance
467 (Figure 9, bottom graph), no node carries more than 10% of total flow.

468 At percolation threshold, the flow frequency distribution is about similar for the three systems,
469 with a long tail that emphasizes a large number of “critical” nodes (Figure 9, upper graph).
470 Above threshold, the UFM simulations remain long-tailed, quite similar to the percolation
471 threshold case, while the two other models are much more short-tailed (Figure 9, bottom graph).

472 This result highlights how critical the UFM connectivity must be. We may expect small
473 variations to induce dramatic changes of the hydraulic network properties. Let us imagine, for
474 instance, that fractures are ending up very close to their inhibiting neighbor but without
475 intersecting it; the geometric distribution remains about similar than eq.(5), but the connectivity
476 is going to be dramatically reduced. This property puts the emphasis on the detail of fracture
477 intersection (in a mechanistic sense) as a critical control on network permeability. In that
478 respect, the network-connectivity issue is more a concern for mechanical investigations than for
479 classical “percolation-like” network studies.

480 **7. Conclusion**

481 We argue that most of fracture systems are spatially organized according to two main regimes: a
482 “dilute” regime for the smallest, where fractures grow independently of each other, and a

483 “dense” regime, for which the density distribution is controlled by the mechanical interactions
484 between fractures.

485 In the dilute regime, the density of fractures is increasing during the fracturing process, and we
486 do not expect any universality in the density parameter. Based on considerations about fracture
487 growth rate and on fault trace length measurements, we conjecture that the fracture length
488 distribution could tend to a power-law scaling with a 2D exponent (i.e. measured on fault traces)
489 close to -2.

490 In the dense regime, we derive the density distribution by acknowledging that, statistically,
491 fractures do not cross a larger one. This simple rule expresses the inhibiting role of large
492 fractures against smaller ones but not the reverse. This very crude description of mechanical
493 interactions actually appears to be a very strong control on the eventual fracture property
494 distribution. Assuming that fracture centers are fractal with a dimension D , the only length
495 distribution that satisfies the UFM rule is a power law with an exponent $-(D+1)$. The fracture
496 density is fully determined by the fractal dimension D and a dimensionless parameter γ that
497 encompasses the details of fracture correlations and orientations. The range of values for γ
498 appears to be extremely limited, which makes the fracture density distribution quite well
499 constrained. For this reason, we call this model UFM for Universal Fracture Model.

500 The smallest fractures are in the dilute regime and the largest in the dense one. The transition is
501 a length scale l_c , which decreases while increasing the density of the dilute regime.

502 This theory is supported by quantitative data on fracture networks. We report three exhaustive
503 studies on fracture distributions for both joint and fault networks, for which the UFM
504 successfully provides an excellent fit for the density distributions at large fracture lengths. For
505 smaller fractures, the distribution scaling appears different with smaller power-law length
506 exponents, as it is expected for the dilute regime. The transition between this regime and the
507 UFM distribution is about a few tenths of kilometers for fault systems, and a few meters for
508 joints. This is a remarkable difference between both fracture processes whose origin is an open
509 issue. For faulting, the transition length l_c appears to be close to the dimension of the actual
510 mechanical system, entailing that only a few very large faults are in the dense/UFM regime.

511 This is consistent with the almost ending of fault growth – in the dilute regime – once the
512 eventual large-scale localization is achieved. For jointing, we suspect that large-scale
513 localization is no more an inhibiting factor of fracture growth; only the fracture-to-fracture
514 interactions, which are basic to UFM, make fracture growth stop. We argue that this difference
515 may be due to the prevailing role of internal stresses in the jointing process. In addition, we
516 point out that the UFM does not apply to fracture systems whose growth is primarily controlled
517 by an external force or constraint.

518 Since the UFM distribution is well connected by definition, we expect important consequences
519 in the flow properties of fracture network. We calculate the percolation parameter for such a
520 system and demonstrate that it exceeds the percolation threshold once the UFM regime
521 develops. We also give a few insights about the flow organization of fracture networks; in
522 particular, we show that, even above the percolation threshold, the network remains critical with
523 a large number of nodes carrying a large amount of flow.

524 **8. References**

- 525 Ackermann, R. V., et al. (2001), The geometric and statistical evolution of normal fault
526 systems: an experimental study of the effects of mechanical layer thickness on scaling laws,
527 *Journal of Structural Geology*, 23(11), 1803-1819.
- 528 Allegre, C. J., et al. (1982), Scaling rules in rock fracture and possible implications for
529 earthquake prediction, *Nature*, 297(5861), 47-49.
- 530 Atkinson, B. K. (1987), *Fracture mechanics of rock*, xi, 534 p. pp., Academic Press, London ;
531 Orlando.
- 532 Bak, P., et al. (1988), Self-organized criticality, *Phys. Rev. A*, 38(1), 364-374.
- 533 Bak, P., and C. Tang (1989), Earthquakes as a self-organized critical phenomenon, *Journal of*
534 *Geophysical Research-Solid Earth and Planets*, 94(B11), 15635-15637.
- 535 Balberg, I. C., et al. (1984), Excluded volume and its relation to the onset of percolation, *Phys.*
536 *Rev. B*, 30(7), 3933-3943.
- 537 Berkowitz, B., et al. (2000), Scaling of fracture connectivity in geological formations, *Geophys.*
538 *Res. Lett.*, 27(14), 10.1029/1999GL011241, 2061-2064.
- 539 Bonnet, E., et al. (2001), Scaling of fracture systems in geological media, *Reviews of*
540 *Geophysics*, 39(3), 10.1029/1999RG000074, 347-384.
- 541 Bour, O. (1997), Transfert de fluides dans les milieux fracturés, *Thèse de l'Université de*
542 *Rennes, Mémoires de Géosciences-Rennes*.
- 543 Bour, O., and P. Davy (1997), Connectivity of random fault networks following a power law
544 fault length distribution, *Water Resour. Res.*, 33(7), 10.1029/97WR00433, 1567-1584.
- 545 Bour, O., and P. Davy (1998), On the connectivity of three-dimensional fault networks, *Water*
546 *Resour. Res.*, 34(10), 10.1029/98WR01861, 2611-2622.

- 547 Bour, O., and P. Davy (1999), Clustering and size distributions of fault patterns: theory and
548 measurements, *Geophys. Res. Lett.*, 26(13), 10.1029/1999GL900419, 2001 - 2004.
- 549 Bour, O., et al. (2002), A statistical scaling model for fracture network geometry, with
550 validation on a multiscale mapping of a joint network (Hornelen Basin, Norway), *J. Geophys.
551 Res.*, 107(B6), 10.1029/2001JB000176.
- 552 Chinnery, M. A., and J. A. Petrak (1968), The dislocation fault model with a variable
553 discontinuity, *Tectonophysics*, 5(6), 513-529.
- 554 Cladouhos, T. T., and R. Marrett (1996), Are fault growth and linkage models consistent with
555 power-law distributions of fault lengths?, *Journal of Structural Geology*, 18(2-3), 281-293.
- 556 Cowie, P. A., and C. H. Scholz (1992), Growth of faults by accumulation of seismic slip, *J.
557 Geophys. Res.*, 97(B7), 11085-11095.
- 558 Cowie, P. A., et al. (1994), Quantitative fault studies on the East Pacific Rise: A comparison of
559 sonar imaging techniques, *J. Geophys. Res.*, 99, 10.1029/94jb00041.
- 560 Cowie, P. A., et al. (1995), Multifractal scaling properties of a growing fault population,
561 *Geophysical International Journal*, 122, 457-469.
- 562 Crampin, S. (1994), The fracture criticality of crustal rocks, *Geophysical Journal International*,
563 118(2), 428-438.
- 564 Crampin, S. (1999), Implications of rock criticality for reservoir characterization, *Journal of
565 Petroleum Science and Engineering*, 24(1), 29-48.
- 566 Darcel, C., et al. (2003), Connectivity properties of two-dimensional fracture networks with
567 stochastic fractal correlation, *Water Resour. Res.*, 39(10), 10.1029/2002WR001628, SBH 1-1 -
568 SBH 1-13.
- 569 Davy, P., et al. (1990), Some consequences of a proposed fractal nature of continental faulting,
570 *Nature*, 348, 10.1038/348056a0, 56-58.
- 571 Davy, P., and P. R. Cobbold (1991), Experiments on shortening of a 4-layer model of the
572 continental lithosphere, *Tectonophysics*, 188(1-2), 1-25.
- 573 Davy, P. (1993), On the frequency - length distribution of the San Andreas fault system, *J.
574 Geophys. Res.*, 98(B7), 10.1029/93JB00372, 12141 - 12152.
- 575 Davy, P., et al. (1995), Localization and fault growth in brittle-ductile systems. Implications to
576 deformations of the continental lithosphere, *J. Geophys. Res.*, 100, 6281-6294.
- 577 Davy, P., et al. (2006), Flow in multiscale fractal fracture networks, in *Fractal Analysis for
578 Natural Hazards*, edited by G. Cello and B. D. Malamud, pp. 31-45.
- 579 De Arcangelis, L., et al. (1989), Scaling in fracture, *Physical Review B*, 40(1), 877-880.
- 580 de Dreuzy, J.-R., et al. (2001), Hydraulic properties of two-dimensional random fracture
581 networks following a power law length distribution 1. Effective connectivity, *Water Resour.
582 Res.*, 37(8), 10.1029/2001WR900011.
- 583 Fox, A., et al. (2007), Statistical geological discrete fracture network model. Forsmark
584 modelling stage 2.2, 1-271 pp, Svensk Kärnbränslehantering AB, Stockholm.
- 585 Hansen, A., et al. (1991), Scale-invariant disorder in fracture and related breakdown
586 phenomena, *Physical Review B*, 43(1), 665-678.
- 587 Hardacre, K. M., and P. A. Cowie (2003a), Controls on strain localization in a two-dimensional
588 elastoplastic layer: Insights into size-frequency scaling of extensional fault populations, *J.
589 Geophys. Res.*, 108(B11), 10.1029/2001jb001712, 15.
- 590 Hardacre, K. M., and P. A. Cowie (2003b), Variability in fault size scaling due to rock strength
591 heterogeneity: a finite element investigation, *Journal of Structural Geology*, 25(10), 1735-1750.
- 592 Hatton, C. G., et al. (1994), Non-universal scaling of fracture length and opening displacement,
593 *Nature*, 367, 160-162.

- 594 Hentschel, H. G. E., and I. Proccacia (1983), The infinite number of generalised dimensions of
595 fractals and strange attractors, *Physica D: Nonlinear Phenomena*, 435-444.
- 596 Herrmann, H. J., and S. Roux (Eds.) (1990), *Statistical models for the fracture of disordered*
597 *media*, 351 pp., North-Holland, Amsterdam.
- 598 Jennings, C. W. (1988), Fault map of California with locations of volcanoes, thermal springs
599 and thermal wells, 4rth ed., Calif. Div. of Mines and Geol., Sacramento.
- 600 King, G. (1983), The accommodation of large strains in the upper lithosphere of the earth and
601 other solids by self-similar fault systems: the geometrical origin of b-Value, *Pure Appl.*
602 *Geophys.*, 121(5), 761-815.
- 603 La Pointe, P., et al. (2008), Geological discrete fracture network model for the Laxemar site.
604 Site Descriptive Modelling -Site Laxemar, 1-260 pp, Svensk Kärnbränslehantering AB,
605 Stockholm.
- 606 Lyakhovskiy, V. (2001), Scaling of fracture length and distributed damage, *Geophysical Journal*
607 *International*, 144(1), 114-122.
- 608 Main, I. G., et al. (1994), One Slope or Two? Detecting Statistically Significant Breaks of Slope
609 in Geophysical Data, with Application to Fracture Scaling Relationships, *Geophys. Res. Lett.*,
610 26, 10.1029/1999gl005372.
- 611 Mansfield, C., and J. Cartwright (2001), Fault growth by linkage: observations and implications
612 from analogue models, *Journal of Structural Geology*, 23(5), 745-763.
- 613 Odling, N. E. (1997), Scaling and Connectivity of joint systems in sandstones from western
614 Norway, *Journal of Structural Geology*, 19(10), 10.1016/S0191-8141(97)00041-2, 1257-1271.
- 615 Pickering, G., et al. (1995), Sampling Power-Law Distributions, *Tectonophysics*, 248(1-2), 1-20.
- 616 Poliakov, A. N. B., and H. J. Herrmann (1994), Self-organized criticality of plastic shear bands
617 in rocks, *Geophys. Res. Lett.*, 21(19), 2143-2146.
- 618 Pollard, D. D., and A. Aydin (1988), Progress in understanding jointing over the past century,
619 *Geological Society of America Bulletin*, 100(8), 10.1130/0016-
620 7606(1988)100<1181:piujot>2.3.co;2, 1181-1204.
- 621 Renshaw, C. E., and D. D. Pollard (1994), Numerical simulation of fracture set formation: A
622 fracture mechanics model consistent with experimental observations, *J. Geophys. Res.*, 99(B5),
623 9359-9372.
- 624 Renshaw, C. E. (1997), Mechanical controls on the spatial density of opening-mode fracture
625 networks, *Geology*, 25(10), 923-926.
- 626 Renshaw, C. E., and J. C. Park (1997), Effect of mechanical interactions on the scaling of
627 fracture length and aperture, *Nature*, 386, 482-484.
- 628 Renshaw, C. E. (1999), Connectivity of joint networks with power law length distributions,
629 *Water Resour. Res.*, 35(9), 10.1029/1999WR900170, 2661 - 2670.
- 630 Renshaw, C. E., et al. (2003), Role of heterogeneity in elastic properties and layer thickness in
631 the jointing of layered sedimentary rocks, *Geophys. Res. Lett.*, 30(24), 10.1029/2003gl018489,
632 4.
- 633 Robinson, P. C. (1984), Numerical calculations of critical densities for lines and planes, *J. Phys.*
634 *A: Math. Gen.*, 17, 2823-2830.
- 635 Schlagenhauf, A., et al. (2008), Incremental growth of normal faults: Insights from a laser-
636 equipped analog experiment, *Earth Planet. Sci. Lett.*, 273(3-4), 299-311.
- 637 Schmittbuhl, J., and S. Roux (1994), The influence of internal stresses on the fracture of
638 heterogeneous media, *Modelling and Simulation in Materials Science and Engineering*, 2(1),
639 21-52.
- 640 Schueller, S., et al. (2005), Brittle-ductile coupling: Role of ductile viscosity on brittle
641 fracturing, *Geophys. Res. Lett.*, 32(10), 10.1029/2004GL022272, 1-4.

- 642 Schultz, R. A. (2000), Growth of geologic fractures into large-strain populations: review of
643 nomenclature, subcritical crack growth, and some implications for rock engineering,
644 *International Journal of Rock Mechanics and Mining Sciences*, 37(1-2), 403-411.
- 645 Schultz, R. A., and H. Fossen (2002), Displacement-length scaling in three dimensions: the
646 importance of aspect ratio and application to deformation bands, *Journal of Structural Geology*,
647 24(9), 1389-1411.
- 648 Segall, P., and D. D. Pollard (1980), Mechanics of Discontinuous Faults, *J. Geophys. Res.*, 85,
649 10.1029/JB085iB08p04337.
- 650 Segall, P., and D. D. Pollard (1983), Joint formation in granitic rock of the Sierra Nevada,
651 *Geological Society of America Bulletin*, 94(5), 10.1130/0016-7606, 563-575.
- 652 Segall, P. (1984), Formation and growth of extensional fracture sets, *Geological Society of*
653 *America Bulletin*, 95(4), 10.1130/0016-7606(1984)95<454:fagoef>2.0.co;2, 454-462.
- 654 SKB (2004a), Preliminary site description Forsmark area - version 1.1, Svensk
655 Kärnbränslehantering, A. B., Stockholm.
- 656 SKB (2004b), Preliminary site description. Simpevarp area - version 1.1, Svensk
657 Kärnbränslehantering, A. B., Stockholm.
- 658 Sornette, A., et al. (1993), Fault growth in brittle - ductile experiments and the mechanics of
659 continental collisions, *J. Geophys. Res.*, 98(B7), 10.1029/92JB01740, 12111 - 12140.
- 660 Sornette, D., et al. (1990), Structuration of the lithosphere in plate tectonics as a self-organized
661 critical phenomenon, *J. Geophys. Res.*, 95(B11), 10.1029/90JB00824, 17353 - 17361.
- 662 Sornette, D., and P. Davy (1991), Fault growth model and the universal fault length distribution,
663 *Geophysical Research Letters*, 18(6), 10.1029/91GL01054, 1079 - 1081.
- 664 Sornette, D. (2006), *Critical phenomena in natural sciences chaos, fractals, selforganization*
665 *and disorder : concepts and tools*, 2nd ed., 528. pp., Springer, Berlin, Heidelberg.
- 666 Spyropoulos, C., et al. (1999), Experimental Evidence for Different Strain Regimes of Crack
667 Populations in a Clay Model, *Geophys. Res. Lett.*, 26, 10.1029/1999gl900175.
- 668 Stauffer, D. (1979), Scaling theory of percolation clusters, *Phys. Rep.*, 54(1), 1-74.
- 669 Stephens, M. B., et al. (2008), Geology Forsmark. Site descriptive modelling Forsmark stage
670 2.2, 1-224 pp, Svensk Kärnbränslehantering AB, Stockholm.
- 671 Turcotte (1986), A fractal model for crustal deformation, *Tectonophysics*, 132, 261-269.
- 672 Vicsek, T. (1992), *Fractal Growth Phenomena*, 488 pp., World Scientific, London.
- 673 Wahlgren, C.-H., et al. (2008), Geology Laxemar. Site descriptive modelling SDM-Site
674 Laxemar, Svensk Kärnbränslehantering AB, Stockholm.
- 675 Wu, H., and D. D. Pollard (1995), An experimental study of the relationship between joint
676 spacing and layer thickness, *Journal of Structural Geology*, 17(6), 897-905.
- 677 Xu, S. S., et al. (2006), Effect of sampling and linkage on fault length and length–displacement
678 relationship, *International Journal of Earth Sciences*, 95(5), 841-853.

679

680

681 **9. Appendix. The stereology issue**

682 Let assume a 3d density distribution of the form:

683 $n_{3d}(l, L, \theta, \varphi) = \alpha_{3d}(\theta, \varphi) L^{D_{3d}} l^{-a_{3d}},$

684 where $n_{3d}(l, L, \theta, \varphi) dl d\theta d\varphi$ is the number of fractures of length in the range $[l, l+dl]$, strike in
 685 $[\theta, \theta+d\theta]$, dip in $[\varphi, \varphi+d\varphi]$, within a typical volume of size L . If fractures are likely modeled by
 686 2D disks in a 3D space, the distribution of 2d fracture traces is:

687
$$n_{2d}(t, L) = \frac{\sqrt{\pi}}{2} \frac{\Gamma(\frac{a_{3d}}{2})}{\Gamma(\frac{a_{3d}+1}{2})} \left(\int_0^{\frac{\pi}{2}} \int_0^{\pi} \alpha_{3d}(\theta, \varphi) \sin \varphi d\theta d\varphi \right) L^{D_{3d}-1} t^{-a_{3d}+1},$$

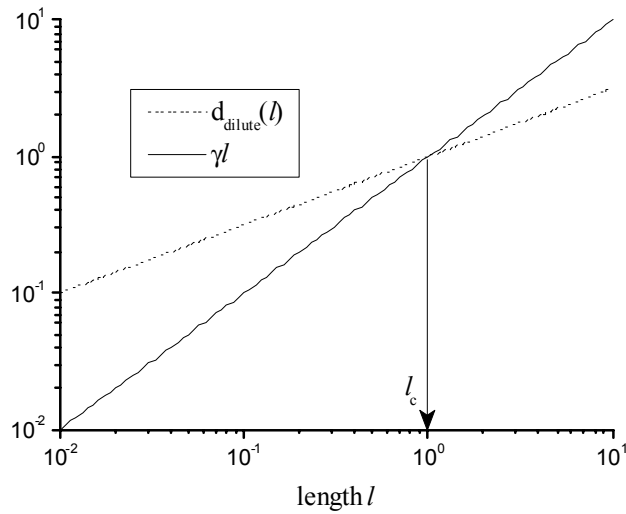
688 where Γ is the gamma function. The density term is thus a complex function of the angular
 689 distributions. For uniform orientations, the 2d density distribution writes as:

690
$$n_{2d}(t, L) = \frac{\alpha}{\sqrt{\pi}} \frac{\Gamma(\frac{a_{3d}}{2})}{\Gamma(\frac{a_{3d}+1}{2})} L^{D_{3d}-1} t^{-a_{3d}+1},$$

691 where α is the integral of α_{3d} over strike and dip. This stereological rule, plus the trivial
 692 relationship $D_{3d} = D_{2d} + 1$ transforms equation (5) into:

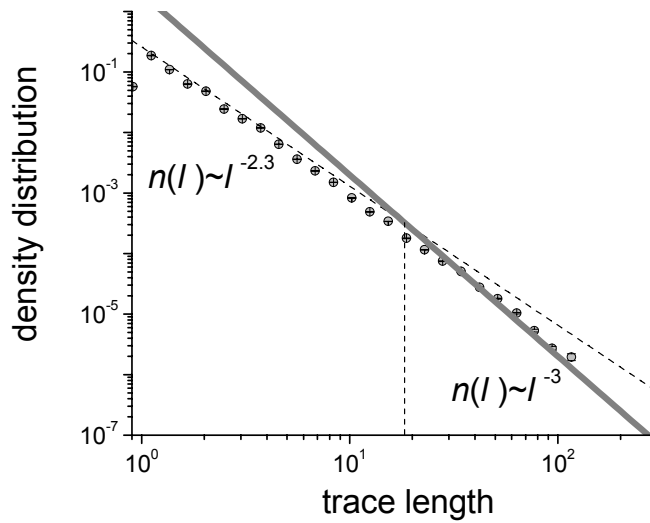
693
$$n_{2d}(t, L) = D_{3d} \gamma^{D_{3d}} \frac{\Gamma(\frac{D_{3d}+1}{2})}{\sqrt{\pi} \Gamma(\frac{D_{3d}+2}{2})} L^{D_{3d}-1} t^{-D_{3d}} \approx 0.4(D_{2d}+1) \gamma^{D_{2d}+1} L^{D_{2d}} t^{-D_{2d}+1}.$$

694 This equation is not very different from equation (5) written for 2D fractures. The scaling is
 695 similar and the density term is of the same order of magnitude. This emphasizes that the basic
 696 geometrical argument that lead to equation (5), i.e. the fact that a fracture stops on its largest
 697 neighbor, is still valid on 2d traces.



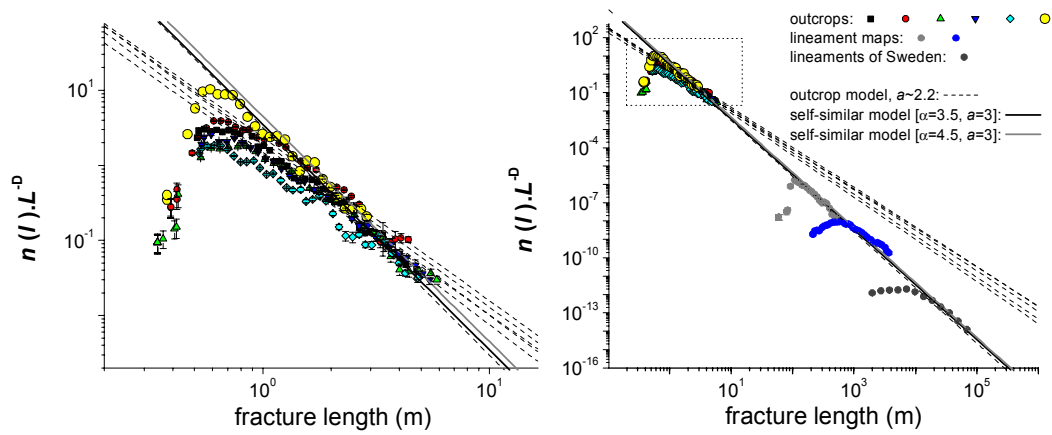
699

700 Figure 1. Sketch of the evolution of the distance between fractures for each class of fracture length l .



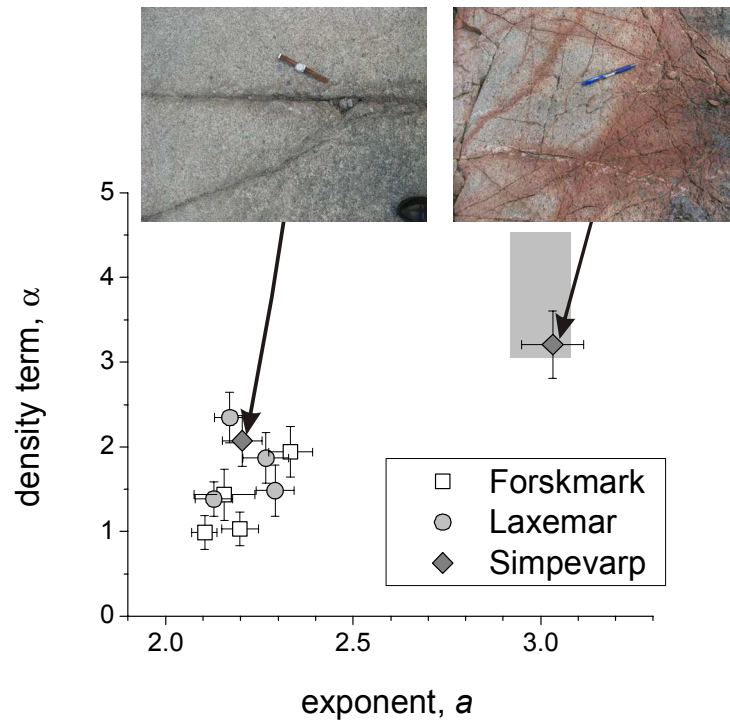
701

702 Figure 2. Length distribution of a fracture network generated according to the following rules: fractures
 703 are generated using a power-law length distribution with an exponent $a=2.3$ (dashed line), and the
 704 eventual distribution results from removing the shortest tip of fractures that intersect a larger one. The
 705 bold grey curve is the fit for the large-length distribution.



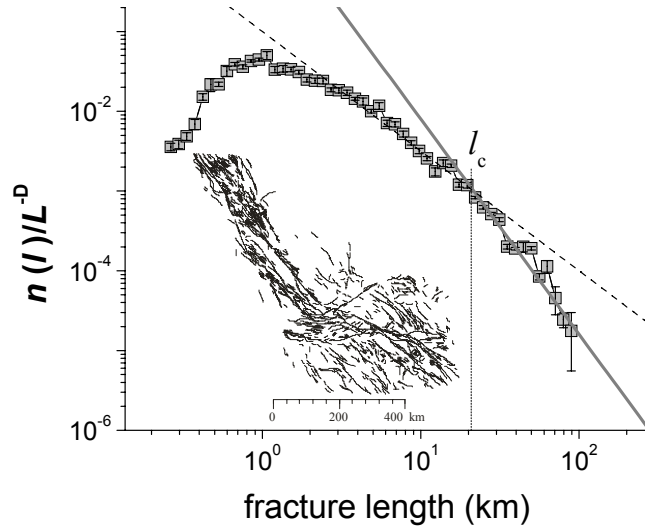
706

707 Figure 3. Density of fracture length per unit area for both outcrop and lineament maps in Sweden.
 708 Outcrops and lineaments are from the Simpevarp and Laxemar area. The last curve is derived from a
 709 compilation of lineaments over Sweden. The dashed lines are power-law fits for outcrops. The solid lines
 710 are self-similar equations with a density term of 3.5 (black) and 4.5 (grey). The upper graph is the portion
 711 of the full lower graph that contains outcrop data.



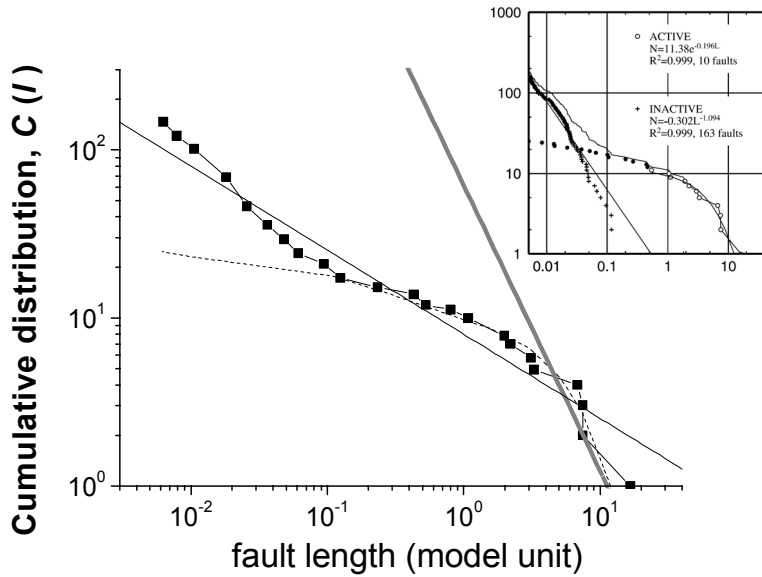
712

713 Figure 4. Plot of the density term vs the power-law length exponent a for all the Swedish outcrops
 714 mapped in the SKB project. The two photographs refer to fracture patterns whose characteristics are
 715 indicated in the figure. The grey square represents the “through-scale” model (see text).



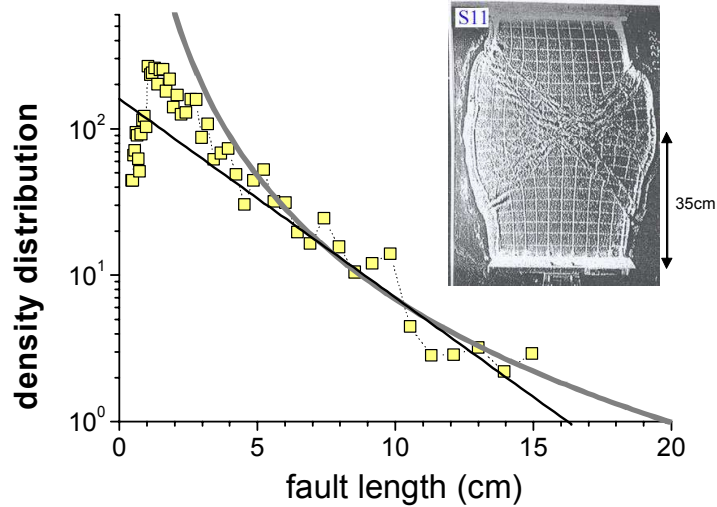
716

717 Figure 5. The length distribution of the San Andreas fault system from [Davy, 1993], with the fault trace
 718 map shown inside the plot. The small-scale “dilute” distribution (dashed line) has a power-law length
 719 exponent $a=1.5$. The large-scale “dense” distribution is given in the text.



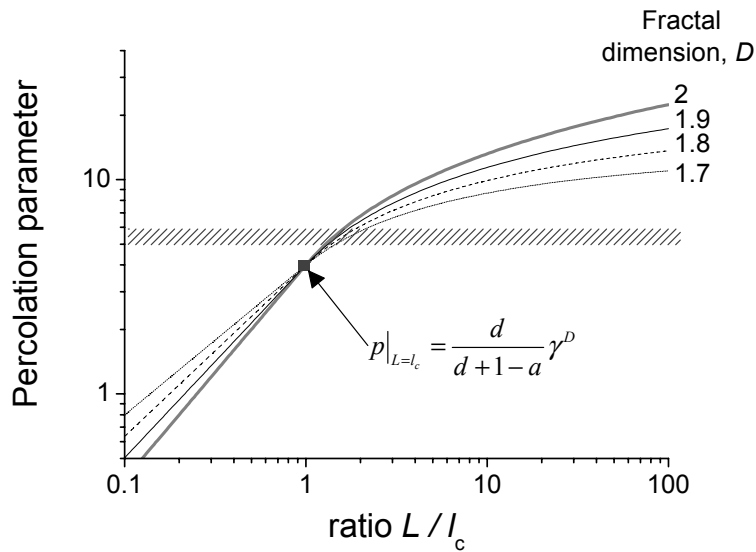
720

721 Figure 6. Cumulative distribution of faults recorded in the simulation presented in [Hardacre and Cowie,
 722 2003a] (original figure is shown on the upper right corner of the figure). The simulation dimensions are
 723 5×10 km. Both active (dashed lines) and total (square points and thin black solid line) faults are shown.
 724 The bold grey solid line is the UFM distribution with $L=7$ km and $\gamma^D=2$.



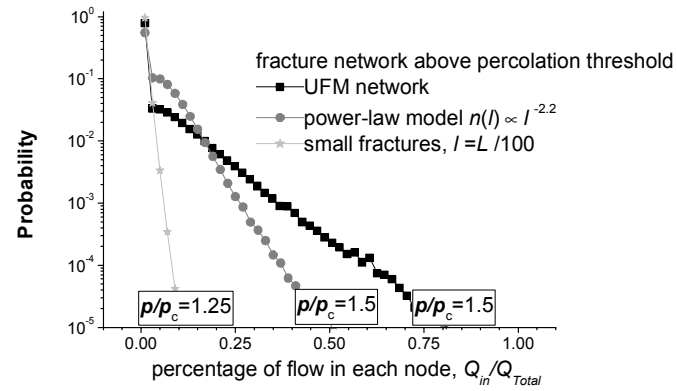
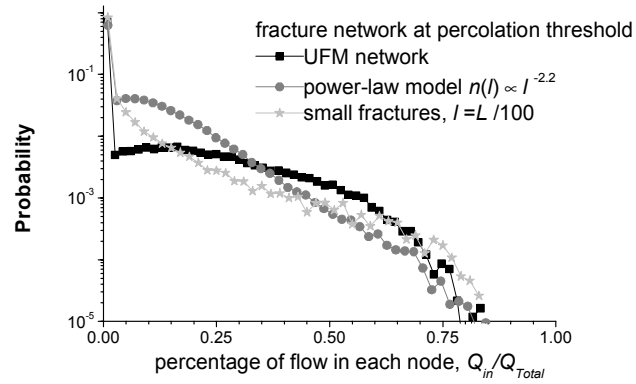
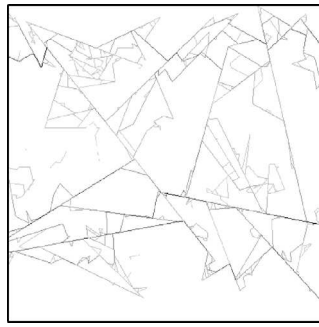
725

726 Figure 7. The graph represents the density distribution of fault lengths for the last recorded stage (15% of
 727 shortening) of the experiment R15 from [Davy *et al.*, 1995] (yellow squares). The experiment is shown in
 728 the top right corner. Its dimensions are 75x35 cm. The UFM equation (grey bold solid line) is calculated
 729 with $D\gamma^D=4$, $D=1.8$ and $L=50$ mm (the length of the largest fault with an angle of 30° to compression). The
 730 exponential fit (black thin solid line) is $n(l) = 160 \cdot \exp(-l/3.2)$.



731

732 Figure 8. Evolution of the percolation parameter with the ratio of L to l_c for different values of the fractal
 733 dimension D . The exponent a and d are taken equal to 2, as well as γ^D . The shaded rectangle represents
 734 the expected range of percolation threshold.



735

736 Figure 9. A flow pattern of a UFM realization (upper graph); the frequency distribution of flow at network

737 nodes calculated at percolation threshold (middle graph), and slightly above percolation for $p/p_c=1.25$ or

738 1.5 (bottom graph).

