



**HAL**  
open science

# Analyse de sensibilité globale et estimation Bayésienne de paramètres : application aux transferts en milieu poreux non saturé

Jabran Zaouali

► **To cite this version:**

Jabran Zaouali. Analyse de sensibilité globale et estimation Bayésienne de paramètres : application aux transferts en milieu poreux non saturé. Hydrologie. Université de Strasbourg, 2019. Français. NNT : 2019STRAH019 . tel-03435397

**HAL Id: tel-03435397**

**<https://theses.hal.science/tel-03435397v1>**

Submitted on 18 Nov 2021

**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.

*ÉCOLE DOCTORALE SCIENCES DE LA TERRE ET ENVIRONNEMENT*

**Laboratoire d'hydrologie et de géochimie de Strasbourg**

**THÈSE** présentée par :

**Jabran ZAOUALI**

soutenue le : 26 Septembre 2019

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Science de la Terre/ Hydrogéologie

**Analyse de sensibilité globale et estimation  
Bayésienne de paramètres : application aux  
transferts en milieu poreux non saturé**

**THÈSE dirigée par :**

**M. YOUNES Anis**  
**M. LEHMANN François**

DR CNRS, Université de Strasbourg  
Pr, Université de Strasbourg

# Remerciements

Les travaux de recherche qui ont fait l'objet de ce mémoire ont été réalisés au sein de l'équipe TrHyCo « TRansferts dans les HYdrosystèmes COntinentaux » du Laboratoire d'Hydrologie et de Géochimie de Strasbourg « LHyGÉS », UMR 7515, CNRS. Je profite de ces lignes pour exprimer ma reconnaissance à ceux qui m'ont accompagné durant ces années de thèse.

**M. YOUNES Anis**, mon directeur de thèse, le mot merci ne pourra jamais quantifier les efforts que vous m'avez accordé à l'égard de ma recherche. J'aimerais vous dire à quel point j'ai apprécié vos qualités humaines, vos multiples encouragements, vos conseils ainsi que toutes les heures que vous m'avez largement consacrées tout au long de ce travail. Enfin je resterai très reconnaissant et je voudrais vous dédier cette réussite.

**M. LEHMANN François**, je vous remercie pour votre encadrement, vos directives impressionnantes durant ce travail, et surtout votre encouragement incessant et vos conseils stimulants et avisés à chaque étape de cette thèse.

Je voudrais également remercier **M. ACKERER Philippe**, Directeur de Recherche au CNRS et directeur de LHyGeS, pour m'avoir accueilli au sein du laboratoire et de m'avoir offert de très bonnes conditions de travail. Je tiens également à remercier tous les membres du laboratoire et spécialement **M. FAHS Marwan** pour ses conseils, ses encouragements et pour toute l'aide qu'il m'a apporté.

Je remercie **M. GRUNBERGER Olivier** d'abord pour avoir accepté de participer au jury et à son nom je tiens à remercier toute l'équipe de l'IRD en Tunisie de nous avoir fourni de la quasi-totalité d'équipement nécessaire, je m'adresse en particulier à **M. HAMDI Radouane** pour votre soutien durant la partie expérimentale de la thèse.

**M. MOUSSA Mahmoud** directeur du LMHE à l'ENIT, merci de nous avoir accordé la permission de travail au sein de vos laboratoires, encore un grand merci à mes chères professeurs qui étaient à mes côtés tout au long de mon parcours universitaire **Mme. BOUHLILA Rachida**, **Mme. GALLOUZ Lamia**, et **Mme. SLAMA Fairouz**, merci pour vos conseils et votre soutien.

Je vous remercie encore **Mme. BOUHLILA** ainsi que **M. AMAZIANE Brahim** d'avoir accepté pour être rapporteurs sur cette thèse.

Je pense particulièrement à toutes les personnes qui m'ont soutenu pendant mes séjours à Strasbourg et avec qui j'ai eu d'agréables moments, une pensée spéciale à **BEREZ Amor**, **ELAIB Fethi**, **BEDRI Hamid**, **SAHLI Mehdi**, **ZOUAOUI Wael**, **HAMZAOUI Melek**, **BEN AHMED kmar** ....

Je tiens ensuite à remercier ma femme **Mme. ZAOUALI BOUFARES Zeineb**, Merci d'avoir supporté mes sauts d'humeur ces derniers temps et de m'avoir toujours encouragé et soutenu.

Enfin, parce que la thèse est aussi une aventure psychologique parfois « chaotique », je remercie également mes parents Ali et Amel qui m'ont toujours supporté et encouragé. Ça valait vraiment le coup maintenant que c'est fait ! Aller à l'université et poursuivre le doctorat ont été de très bonnes décisions de la vie.

*Je souhaite à tous une agréable lecture*

.....

# Tables des matières

Remercîment .....	i
Tables des matières .....	iii
Introduction générale.....	1
Chapitre 1. Modèles mathématique.....	4
Introduction .....	5
1.1 Ecoulement d'eau dans un milieu poreux non saturé.....	5
1.1.1 Caractéristiques d'un milieu poreux non saturé.....	5
1.1.2 Hydrodynamique des milieux poreux .....	8
1.2 Potentiel spontané.....	13
1.2.1 Principe de l'électrofiltration .....	14
1.2.2 Modèle de double couche électrique.....	15
1.2.3 Couplage électrocinétique .....	17
1.2.4 Coefficient d'infiltration .....	19
1.2.5 Couplage hydrogéophysique .....	22
1.3 Transport de soluté en milieux poreux non saturés .....	22
1.3.1 L'advection.....	23
1.3.2 Diffusion moléculaire.....	23
1.3.3 Dispersion cinématique .....	23
1.3.4 Dispersion hydrodynamique .....	25
1.3.5 L'équation d'advection-dispersion.....	26
1.3.6 Modèle mobile-immobile.....	26
1.3.7 Couplage écoulement-transport en milieux poreux non-saturé.....	27
Conclusion.....	28
Bibliographie.....	29

Chapitre 2. Méthodes d'analyse de sensibilité globale et d'estimation des paramètres.....	34
2.1 Introduction .....	35
2.2 Analyse de sensibilité globale .....	36
2.2.1 Introduction .....	36
2.2.2 Identification des sources d'incertitude.....	36
2.2.3 Propagation d'incertitude .....	37
2.2.4 Analyse de sensibilité.....	40
2.2.5 Expansion en polynômes de chaos.....	48
2.3 Estimation des paramètres .....	66
2.3.1 Introduction :.....	66
2.3.2 Méthodes de résolution du problème inverse :.....	67
2.3.3 Approximation de premier ordre (FOA) .....	70
2.3.4 Approche Bayésienne.....	75
2.4 Conclusion.....	85
Bibliographie.....	87
Chapitre 3. Analyse de sensibilité globale et estimation bayésienne des paramètres pour le potentiel spontané en milieu poreux non saturé .....	101
3.1 Introduction .....	102
3.2 Description du cas de test et solution numérique .....	104
3.2.1 Description du cas de test .....	104
3.2.2 Modèle mathématique .....	104
3.2.3 Modèle numérique .....	104
3.3 Analyse de sensibilité globale des signaux de SP .....	105
3.3.1 La méthode GSA .....	105
3.3.2 Résultats GSA .....	106
3.4 Estimation des paramètres .....	108

3.4.1 Approches MCMC et FOA .....	108
3.4.2 Résultats d'estimation des paramètres .....	109
3.5 Conclusion .....	112
Chapitre 4. Estimation bayésienne des paramètres d'un sol non remanié à partir d'expériences de drainage et d'infiltration .....	116
4.1 Introduction .....	117
4.2 Modèles expérimentaux et mathématiques .....	118
4.2.1 Expériences de laboratoire .....	118
4.3 Modèle numérique .....	118
4.4 Inférence bayésienne de paramètre .....	119
4.5 Résultats et discussion .....	120
4.6 Conclusion .....	124
Chapitre 5. Estimation bayésienne des paramètres pour le transport non réactif en milieu poreux non saturé .....	127
5.1 Introduction .....	128
5.2 Modèles expérimentaux et mathématiques .....	129
5.2.1 Expérience de laboratoire .....	129
5.2.2 Modèle numérique .....	130
5.2.3 Inférence bayésienne de paramètre .....	132
5.3 Résultats et discussion .....	133
5.4 Conclusion .....	140
Conclusion générale .....	146

# Introduction générale

Divers contaminants générés par les activités humaines et déposés à la surface du sol finissent par s'infiltrer jusqu'à la nappe phréatique. En général, les premiers horizons de sol conditionnent l'essentiel de la migration de ces contaminants vers les eaux souterraines. De ce fait, l'étude des transferts d'eau et de contaminants en milieu poreux non saturé requiert une importance majeure afin de préserver la qualité des eaux et des sols. Dans cet objectif, la modélisation apparaît comme un outil intéressant permettant de comprendre et de reproduire les processus de transfert dans le sol. Ces processus sont décrits par des équations aux dérivées partielles non linéaires incluant un certain nombre de paramètres qui dépendent de la nature du sol et de celle du fluide qui le traverse.

Cependant, la modélisation fournit souvent des résultats entachés d'incertitudes à cause notamment des erreurs de mesures et du manque de données. Dans ces conditions, il est important de comprendre et d'évaluer la contribution de chaque paramètre (ou condition aux limites) à l'incertitude du modèle, c'est l'objet de l'analyse de sensibilité. Ainsi, en appelant la sortie du modèle  $Y = f(X)$  avec  $X$  les conditions aux limites et/ou les paramètres du modèle, la question que nous nous posons est : comment chaque élément de  $X$  contribue à l'incertitude sur  $Y$  ?

Par ailleurs, l'utilisation des modèles sur des cas réels est souvent confrontée à une difficulté majeure qui consiste à attribuer des valeurs aux paramètres que contient le modèle. En effet, certains de ces paramètres ne sont pas directement mesurables et sont estimés par approche inverse. Les problèmes inverses sont généralement difficiles à résoudre à cause de la non-identifiabilité des paramètres. C'est par exemple le cas lorsque la sortie du modèle est insensible aux paramètres ce qui implique la non unicité de la solution. C'est également le cas lors d'un manque d'informations dans les mesures pour assurer une bonne estimation de ces paramètres. Généralement, le problème inverse est résolu en cherchant les paramètres qui minimisent une fonction objective mesurant l'écart entre les observations et les valeurs calculées par le modèle. La minimisation permet de trouver une valeur unique pour chaque paramètre, sans tenir compte des différentes sources d'incertitude affectant le problème inverse. Ces sources d'incertitude sont souvent responsables de la non existence de solutions correspondant aux données observées (existence), ou la présence d'une multiplicité de solutions vérifiant ces données



(unicité). De plus, les petites erreurs de mesure peuvent conduire à d'énormes changements dans le modèle estimé (stabilité). Tout cela rend le problème inverse mal-posé. Pour pallier ce problème, l'approche bayésienne offre une base rigoureuse pour l'estimation des paramètres à partir de données bruitées et pour les modèles incertains, en intégrant les informations diverses disponibles a priori.

L'objectif principal de cette thèse est l'utilisation des méthodes d'analyse de sensibilité globale et d'estimation bayésienne des paramètres pour les transferts en milieux poreux non saturés. Trois applications correspondant à des expériences de drainage et d'infiltration à l'échelle du laboratoire sont considérées. Différentes variables (sorties) seront analysées parmi lesquelles le potentiel spontané (PS) induit par l'écoulement d'eau dans la zone non saturée, le flux cumulé et la concentration d'un contaminant non réactif à la sortie de la colonne.

Ce manuscrit est composé de cinq chapitre :

Dans le premier chapitre, nous rappelons d'une manière succincte, les modèles mathématiques qui régissent le transfert d'eau et de contaminants dans les milieux poreux non saturés ainsi que le modèle électrocinétique qui décrit le potentiel spontané à l'intérieur du milieu poreux. Deux modèles sont utilisés pour le transport de masse. Le premier fait l'hypothèse d'un simple milieu (une seule phase mobile), alors que le deuxième fait l'hypothèse d'un double milieu (deux phases : l'une mobile et l'autre immobile). Dans ce chapitre, nous rappelons également le couplage entre les différents modèles mathématiques à savoir le couplage hydrogéophysique et le couplage écoulement-transport en milieu poreux non saturé.

Dans le deuxième chapitre, nous rappelons les théories de base de l'analyse d'incertitude et de l'analyse de sensibilité globale (GSA) basée sur les indices de Sobol. Plusieurs techniques peuvent être utilisées pour le calcul de ces indices parmi lesquelles le développement en chaos polynomial qui sera détaillé dans ce chapitre. Dans un second temps, quelques approches d'estimation de paramètres et d'incertitude à posteriori seront présentées. Deux types d'approches seront étudiés ; la première utilise l'algorithme de Levenberg-Marquardt pour l'estimation des paramètres et une approximation de premier ordre (FOA) pour l'incertitude à posteriori. La deuxième est une approche bayésienne utilisant un échantillonneur basé sur les chaînes de Markov (Markov Chain Monte Carlo, MCMC). Cette dernière offre une base rigoureuse pour l'estimation des paramètres à partir de données bruitées et pour les modèles incertains, en intégrant toutes les informations disponibles à priori.

Dans le troisième chapitre, on présente la première application de la thèse qui est consacrée au couplage hydrogéophysique en milieu poreux non saturé. Le cas test étudié, correspond à une expérience de percolation drainage pendant laquelle le PS est mesuré à différentes altitudes. . Ce signal PS est sensible aux propriétés hydrauliques et électriques du sol. L'analyse de sensibilité globale du signal PS est basée sur les polynômes de chaos (PC) comme métamodèle. L'estimation des paramètres à partir des mesures PS est effectuée à l'aide de l'échantillonneur MCMC. Les résultats sont ensuite comparés à ceux obtenus avec l'approche FOA. Cette étude a fait l'objet d'un article publié dans le journal *Hydrology and Earth System Sciences* (Younes *et al.*, 2018).

Le quatrième chapitre est dédié à l'identification des propriétés hydrauliques caractérisant l'écoulement non saturé dans un sol non remanié. L'étude comporte deux volets : un volet expérimental et un volet analyse de sensibilité et estimation de paramètres. Deux colonnes de sol non remanié ont été extraites dans la région de Bekalta située dans le centre Est de la Tunisie. Pour la première, on a effectué une expérience de percolation-drainage et pour la seconde une expérience d'infiltration avec un débit d'eau constant à la surface du sol. Les paramètres hydrauliques du sol ont été étudiés à l'aide d'une modélisation inverse utilisant les mesures de la pression à l'intérieur de la colonne ainsi que le débit cumulé à la sortie. Dans ce chapitre, les GSA et les MCMC sont utilisés pour comparer et ensuite comprendre la différence entre les résultats des deux expériences en terme d'incertitude associée aux paramètres hydrauliques du sol. Ce travail est présenté sous la forme d'un article publié dans *Journal of Hydrology* (Younes *et al.*, 2018).

Dans le cinquième et dernier chapitre de cette thèse, une expérience de transfert d'eau et de contaminant a été réalisée au laboratoire en injectant une solution de bromure à la surface d'un sol non remanié. Durant l'infiltration, la pression a été mesurée à différentes altitudes, de même que la concentration à la sortie de la colonne. Ces observations sont ensuite utilisées pour une estimation simultanée de tous les paramètres hydrauliques et de transport. Deux modèles de transport sont étudiés pour la modélisation de la propagation du bromure : le modèle linéaire d'advection-dispersion et le modèle mobile-immobile. Ce dernier fait intervenir deux phases (l'une mobile et l'autre immobile) avec un terme d'échange entre les deux (non-équilibre). L'inversion est effectuée dans un cadre bayésien à l'aide de l'échantillonneur MCMC. Les résultats de l'estimation des paramètres sont analysés et comparés en utilisant les deux modèles de transport. Cette étude a fait l'objet d'une publication dans *Water Journal* (Younes *et al.*, 2019).

# Chapitre 1. Modèles mathématiques

## Plan du chapitre

Chapitre 1. Modèles mathématiques .....	4
Introduction .....	5
1.1 Ecoulement d'eau dans un milieu poreux non saturé.....	5
1.1.1 Caractéristiques d'un milieu poreux non saturé .....	5
1.1.2 Hydrodynamique des milieux poreux .....	8
1.2 Potentiel spontané.....	13
1.2.1 Principe de l'électrofiltration .....	14
1.2.2 Modèle de double couche électrique.....	15
1.2.3 Couplage électrocinétique .....	17
1.2.4 Coefficient d'infiltration .....	19
1.2.5 Couplage hydrogéophysique .....	22
1.3 Transport de soluté en milieux poreux non saturés .....	22
1.3.1 L'advection.....	23
1.3.2 Diffusion moléculaire.....	23
1.3.3 Dispersion cinématique .....	23
1.3.4 Dispersion hydrodynamique .....	25
1.3.5 L'équation d'advection-dispersion.....	26
1.3.6 Modèle mobile-immobile .....	26
1.3.7 Couplage écoulement-transport en milieux poreux non-saturé.....	27
Conclusion.....	28

## Introduction

Ce premier chapitre a comme but de présenter, d'une manière succincte, les principales notions physiques liées aux équations aux dérivées partielles qui décrivent le transfert d'eau et de soluté dans un milieu poreux non saturé ainsi que le modèle couplé « hydrogéophysique ». Dans la première partie de ce chapitre, nous définissons l'aspect et les variables qui caractérisent les milieux poreux non saturés ainsi que l'équation de Richards qui décrit l'écoulement d'eau dans ces milieux. Dans la deuxième partie, nous définissons le potentiel spontané et ses origines ainsi que les équations couplées de l'électrocinétique, basées sur la théorie de double couche électrique. La dernière partie de ce chapitre sera consacrée au transport de soluté dans les milieux poreux non saturés dans l'hypothèse d'un simple milieu (une seule phase mobile) ainsi que dans l'hypothèse d'un double milieu (deux phases : l'une mobile et l'autre immobile).

### 1.1 Ecoulement d'eau dans un milieu poreux non saturé

#### 1.1.1 Caractéristiques d'un milieu poreux non saturé

Généralement, un milieu poreux est défini par un milieu solide qui contient des espaces vides occupés par des fluides (liquide et/ou gaz). Ces espaces peuvent être séparés ou interconnectés et sont appelés les pores. Le sol est donc considéré comme un milieu poreux formé par des particules solides entre lesquels l'eau ou l'air peuvent circuler.

##### 1.1.1.1 Texture granulométrique

La distribution des sols en différentes classes dépend des tailles des particules solides qui les constituent. Cette classification est obtenue par l'analyse granulométrique qui consiste à déterminer le pourcentage des particules dont les diamètres sont inférieurs ou égaux à une valeur fixée. Il existe trois grandes classes de sol (argile, limon, sable) et leurs intersections donnent plusieurs sous classes (Figure 1.1).

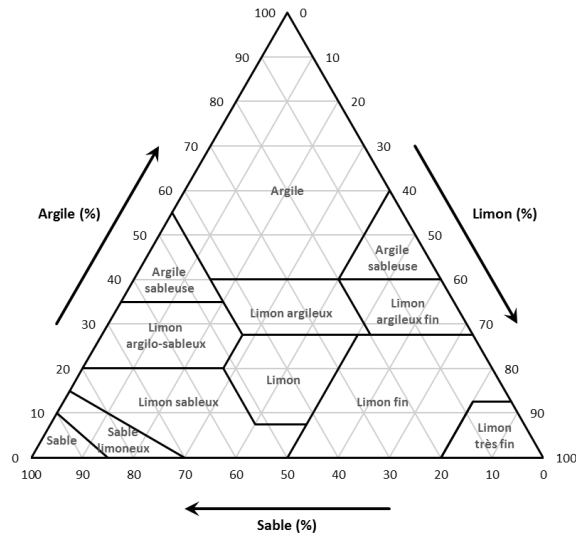


Figure 1.1: Classification des sols suivant les diamètres des particules appelé « Triangle des textures »

### 1.1.1.2 Porosité

La porosité d'un sol est une caractéristique importante. Cette grandeur physique qui représente la densité des pores, conditionne l'écoulement et la rétention de l'eau dans le sol. Elle est définie par le rapport entre le volume de vides  $V_v$  et le volume total  $V_T$  occupé par un échantillon de sol :

$$\phi = \frac{V_v}{V_T} \quad (1.1)$$

Elle est liée à la taille des constituants (texture), leurs arrangements et leurs orientations (structure).

### 1.1.1.3 Teneur en eau

Un milieu poreux est dit saturé, lorsque la totalité du volume de ses pores est occupée par un liquide (l'eau). Dans le cas de présence de phase gazeuse, le milieu est considéré non-saturé. D'où la nécessité de faire intervenir un autre paramètre physique " la teneur en eau " qui désigne la quantité d'eau liquide contenue dans un échantillon.

La teneur en eau volumique  $\theta$  [ $L^3.L^{-3}$ ] est définie par le rapport entre le volume d'eau  $V_w$  [ $L^3$ ] et le volume total  $V_T$  [ $L^3$ ] de l'échantillon :

$$\theta = \frac{V_w}{V_T} \quad (1.2)$$

Cette dernière peut être exprimée en pourcentage en la divisant par la porosité et elle est appelée saturation en eau  $S_w = \theta/\phi$ . Pour un milieu saturé,  $V_w = V_v$ ,  $\theta = \theta_s = \phi$  et  $S_w = 1$  ou 100 %, avec  $\theta_s$  la teneur en eau à saturation [ $L^3.L^{-3}$ ].

#### 1.1.1.4 L'énergie potentielle de l'eau dans le sol

L'état énergétique de l'eau dans le sol se décompose en énergie potentielle et en énergie cinétique (due à la vitesse de déplacement du liquide dans le sol). Cette dernière est généralement très faible. L'énergie potentielle, ou "potentiel", est donc suffisante à représenter l'état énergétique de l'eau dans le sol. Plusieurs facteurs sont à l'origine de cette énergie potentielle, certains sont négligeables telle que la variation des concentrations en sels dissous qui donne le potentiel osmotique  $\psi_o$ , la déformabilité et la compressibilité du sol qui fournit le potentiel d'enveloppe  $\psi_e$  (Campbell, 1985). En un point donné, l'eau du sol présente un potentiel total  $\psi_t$ , qui est la somme des deux principaux potentiels  $\psi_g$  et  $\psi_m$  :

$$\psi_t = \psi_g + \psi_m \quad (1.3)$$

$\psi_g$  : le potentiel gravitaire associé aux forces liées à l'accélération de la pesanteur  $g$ , dont la valeur dépend de la différence de cote  $z$  [ $L$ ] entre un point donné et le niveau référence.

$\psi_m$  : le potentiel matriciel ou de pression correspondant aux forces de liaison liquide/solide.

Physiquement, ce potentiel total est exprimé en énergie par unité de masse [ $L^2.T^{-2}$ ], néanmoins souvent en hydraulique, en hydrologie, ou en hydrogéologie, il est défini comme une énergie par unité de poids [ $L$ ] et il est appelé la charge hydraulique ou charge totale  $H_t = H = h_g + h_m$ .

$h_g$  est l'altitude de l'eau  $[L]$ , si l'axe (Oz) est orienté vers le bas  $h_g = -z$ .

$h_m = h = P/\rho g [L]$  est définie comme la hauteur de la colonne d'eau nécessaire pour provoquer la pression effective de l'eau dans le sol par rapport à la pression atmosphérique  $P = P_e - P_a [M.L^{-1}.T^{-2}]$  (Bernoulli, 1738). Sachant que,  $\rho$  est la masse volumique de l'eau  $[M.L^{-3}]$ ,  $g$  est l'accélération de pesanteur  $[L.T^{-2}]$ ,  $P_e$  et  $P_a$  sont respectivement la pression de l'eau du sol et la pression atmosphérique  $[M.L^{-1}.T^{-2}]$ . En résumant, la charge hydraulique sera définie alors sous cette formule :

$$H = h - z \quad (1.4)$$

La présence de l'air dans un milieu poreux non saturé provoque des forces d'adsorption et de capillarité exercées par la matrice solide sur l'eau afin de la retenir. Ces forces qui s'opposent à l'écoulement de l'eau donnent un potentiel de pression  $\psi_m$  négatif ( $h < 0$ ) car dans ce cas la pression de l'eau du sol est inférieure à la pression atmosphérique ( $P_e < P_a$ ).

## 1.1.2 Hydrodynamique des milieux poreux

### 1.1.2.1 Loi de Darcy et Perméabilité

L'expérience de Darcy (1856) consiste à faire circuler de l'eau dans une colonne de sable de section  $S [L^2]$  et de longueur  $L [L]$ . La mesure du débit  $Q [L^3.T^{-1}]$  et la différence de charge hydraulique entre le bas et le haut de la colonne  $\Delta H$  donnent la fameuse relation suivante :

$$\frac{Q}{S} = K_s \frac{\Delta H}{L} = q \quad (1.5)$$

Où  $q [L.T^{-1}]$  est le flux surfacique appelé également vitesse de Darcy (filtration). La loi de Darcy fait apparaître une constante  $K_s [L.T^{-1}]$  appelée conductivité hydraulique à saturation. Elle caractérise la capacité de circulation de l'eau saturante dans le sable. Cette grandeur est liée à la perméabilité intrinsèque  $k [L^2]$  du milieu poreux par la viscosité dynamique  $\mu$

$[M.L.T^{-1}]$  du fluide et sa masse volumique  $\rho [M.L^{-3}]$  à travers  $K_s = \rho g k / \mu$ , où  $g$  est l'accélération de la pesanteur  $[L.T^{-2}]$ .

Dans le cas d'un sol non saturé, la loi de Darcy est généralisée en utilisant une conductivité hydraulique  $K(\theta) = K_s k_r(\theta)$  qui est fonction de la teneur en eau du sol, où  $k_r(\theta)$  est la perméabilité relative du sol. Pour un milieu poreux homogène et isotrope, la vitesse de filtration d'un fluide incompressible en régime laminaire s'exprime en fonction du gradient de pression et de la gravité sous la forme générale (Buckingham, 1907) :

$$\vec{q} = -K(\theta) \vec{\nabla} \left( \frac{P}{\rho g} - z \right) \quad (1.6)$$

Où  $P = \rho g h$  est la pression d'eau dans le sol  $[M.L^{-1}.T^{-2}]$ ,  $h$  est la pression effective de l'eau qui dépend de la teneur en eau  $[L]$ ,  $z$  est la profondeur orientée vers le bas  $[L]$ . La charge hydraulique  $H [L]$  est définie par  $H = P / \rho g - z = h - z$ .

#### 1.1.2.2 L'équation de continuité

Le principe de conservation de la masse d'un fluide en mouvement est exprimé par l'équation de continuité. Au niveau d'un volume élémentaire, la somme de la variation de la masse du fluide durant un intervalle de temps élémentaire et des flux massiques traversant la surface de ce volume est égale à la quantité de fluide prélevée ou injectée (terme puits/source massique)  $S_m [M.L^{-3}.T^{-1}]$ .

$$\frac{\partial(\rho \phi S_w)}{\partial t} + \nabla \cdot (\rho \vec{q}) = S_m \quad (1.7)$$

#### 1.1.2.3 L'équation de Richards

Pour un fluide incompressible, de masse volumique et de viscosité dynamique constantes, la combinaison de l'équation de continuité et de l'équation de Darcy généralisée donne l'équation de Richards (Richards 1931) :



$$\frac{\partial \theta}{\partial t} - \nabla \cdot [K(\theta) \vec{\nabla}(h(\theta) - z)] = S_v \quad (1.8)$$

Où  $K(\theta) = k_r(\theta) \rho g / \mu$  est la conductivité hydraulique et  $S_v = S_m / \rho$  est le terme puits/source  $[T^{-1}]$ .

Cette équation est basée sur la variable  $\theta$ . Elle peut être écrite avec  $h$  variable principale sous la forme (Hillel 1980) :

$$c_s(h) \frac{\partial h}{\partial t} - \nabla \cdot [K(h) \vec{\nabla}(h - z)] = S_v \quad (1.9)$$

Avec  $c_s(h) = \partial \theta / \partial h$  est la capacité capillaire spécifique  $[L^{-1}]$ .

Elle peut également être écrite sous sa forme mixte en fonction de  $\theta$  et de  $h$  :

$$\frac{\partial \theta}{\partial t} - \nabla \cdot [K(h) \vec{\nabla}(h - z)] = S_v \quad (1.10)$$

Plusieurs études (Milly 1985; Celia et al. 1990; Li 1993) ont montré que les solutions numériques obtenues avec la forme mixte de l'équation de Richards sont plus précises que celles obtenues avec la forme en  $h$  ou la forme en  $\theta$ .

#### 1.1.2.4 Les relations constitutives

Pour résoudre l'équation de Richards, il est essentiel de connaître la courbe de rétention d'eau  $\theta(h)$  qui relie la pression  $h$  à la teneur en eau du sol  $\theta$ , ainsi que l'expression de la conductivité hydraulique  $K$  en fonction de la teneur en eau  $\theta$  ou la pression  $h$ . Ces relations qui dépendent de plusieurs paramètres sont des expressions empiriques. Les modèles les plus célèbres sont ceux de Gardner (1958), Brooks et Corey (1964), Brutsaert (1966), Mualem (1976) et van Genuchten (1980).

- Modèle Brooks et Corey (1964)

Brooks et Corey proposent une relation de rétention d'eau  $\theta(h)$  qui dépend de la teneur en eau résiduelle  $\theta_r$ , la teneur en eau à saturation  $\theta_s$ , un paramètre  $\alpha [L^{-1}]$  qui représente l'inverse de la pression d'entrée d'air ainsi que le paramètre  $\lambda [-]$  qui caractérise la distribution de la taille des pores. Cette relation est écrite sous la forme :

$$S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r} = \begin{cases} \frac{1}{(\alpha|h|)^\lambda} & \text{pour } \alpha|h| < 1 \\ 1 & \text{pour } \alpha|h| \geq 1 \end{cases} \quad (1.11)$$

- Modèle Mualem (1976)

$$K(S_e) = K_s \cdot S_e^{L+2+\frac{2}{\lambda}} \quad (1.12)$$

Ce modèle utilise les paramètres du modèle précédent en ajoutant un nouveau paramètre  $L [-]$  qui tient compte de la tortuosité du milieu.

Le couple de modèle Brooks et Corey – Mualem est souvent utilisé dans les modèles de simulation des écoulements dans les milieux non saturés, ainsi que pour l'estimation des paramètres de ces modèles (Rivière, 1977 ; Rawls et al, 1982).

- Modèle de van Genuchten (1980)

$$S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r} = \begin{cases} \frac{1}{\left(1 + (\alpha|h|)^n\right)^m} & \text{pour } h < 0 \\ 1 & \text{pour } h \geq 0 \end{cases} \quad (1.13)$$

Avec  $n$  et  $m [-]$  deux paramètres du modèle, reliés par l'expression :  $m = 1 - 1/n$ . Ce modèle vise à trouver une relation entre la teneur en eau et la pression d'eau pour tous les types de sols (Figure1.2).

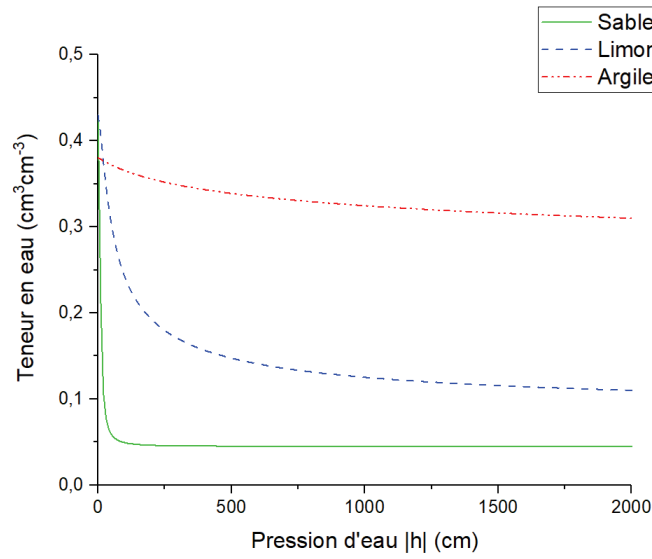


Figure 1.2 : Teneur en eau en fonction de la pression d'eau calculée par modèle van Genuchten – Mualem pour trois classes de sols

Le deuxième couple de modèles souvent utilisé, combine le modèle de rétention de van Genuchten et la relation de Mualem  $K(\theta)$  modifiée (Figure 1.3):

$$K(S_e) = \begin{cases} K_s \cdot S_e^L \left[ 1 - (1 - S_e^{1/m})^m \right]^2 & \text{pour } h < 0 \\ K_s & \text{pour } h \geq 0 \end{cases} \quad (1.14)$$

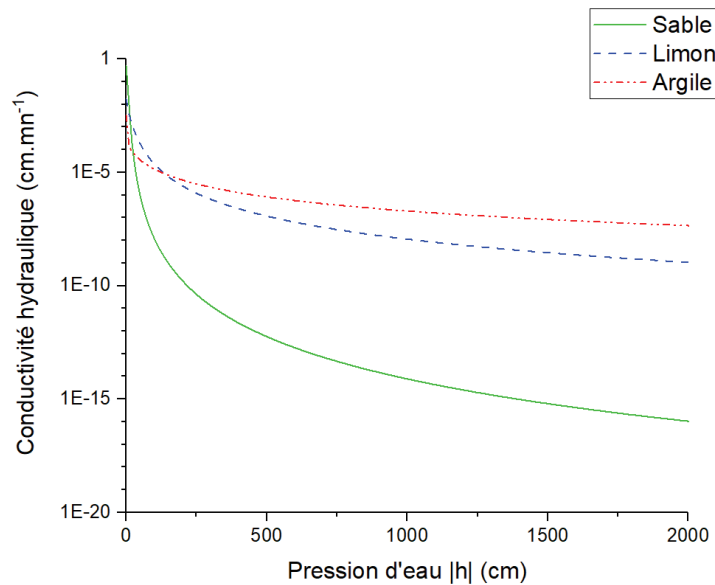


Figure 1.3: Conductivité hydraulique en fonction de la pression d'eau calculée par modèle van Genuchten-Mualem pour trois classes de sols

Carsel et Parrish (1988) ont relié les paramètres hydrauliques du modèle van Genuchten – Mualem à la courbe granulométrique des sols. Les valeurs des paramètres du modèle van Genuchten – Mualem pour différents sols sont résumées dans le Tableau 1.1.

*Tableau 1.1 : Valeurs moyenne des paramètres du modèle van Genuchten – Mualem pour douze classes de sols (Carsel et Parrish, 1988)*

Texture	$K_s [cm.j^{-1}]$	$\theta_s$	$\theta_r$	$\alpha [cm^{-1}]$	$n$
Sand	712.80	0.43	0.045	0.145	2.68
Loamy Sand	350.20	0.41	0.057	0.124	2.28
Sandy Loam	106.10	0.41	0.065	0.075	1.89
Loam	24.96	0.43	0.078	0.036	1.56
Silt	6.00	0.46	0.034	0.016	1.37
Silt loam	10.80	0.45	0.067	0.020	1.41
Sandy clay loam	31.44	0.39	0.100	0.059	1.48
Clay loam	6.24	0.41	0.095	0.019	1.31
Silty clay loam	1.68	0.43	0.089	0.010	1.23
Sandy clay	2.88	0.38	0.100	0.027	1.23
Silty clay	0.48	0.36	0.070	0.005	1.09
Clay	4.80	0.38	0.068	0.008	1.09

Russo (1988), Lenhard *et al.* (1989), Russo *et al.* (1991) et Morel-Seytoux *et al.* (1996) ont étudié l'équivalence entre les modèles de Brooks et Corey – Mualem et van Genuchten – Mualem. Les différents modèles ne caractérisent pas l'écoulement de la même manière. Il faut essayer différentes relations de  $K(\theta)$  et  $\theta(h)$  pour modéliser les écoulements d'eau dans un sol.

## 1.2 Potentiel spontané

Le potentiel spontané est la mesure passive de la distribution électrique dans le sol sans injection de courant. Il existe dans la nature plusieurs mécanismes de polarisation de charges électriques qui sont associés au potentiel spontané. On distingue les phénomènes d'électro-diffusion, électro-rédox, thermoélectrique et le phénomène d'électrocinétique ou électrofiltration qui est la source principale du potentiel spontané qui sera considéré dans la suite de ce travail.

### 1.2.1 Principe de l'électrofiltration

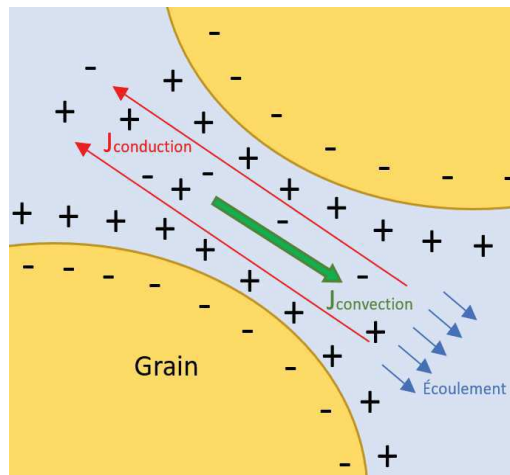


Figure 1.4: Schéma illustratif du principe de l'électrofiltration

L'électrofiltration est le mécanisme physique principal, qui provoque un champ électrique mesurable sous l'effet de la circulation d'un fluide dans un capillaire ou dans un milieu poreux. En effet, lors d'un écoulement dans un milieu poreux engendré par un gradient de pression  $\Delta P$  les particules chargées transportées par l'eau génèrent un courant de convection et afin d'assurer l'électroneutralité, un courant de conduction sera créé (Figure 1.4).

Ce courant de conduction est lié par la loi d'Ohm à une différence de potentiel électrique  $\Delta V$  qui est proportionnelle au gradient de pression  $\Delta P$  :

$$\Delta V = C.\Delta P \quad (1.15)$$

Où,  $C$  est le coefficient d'électrofiltration exprimé en  $V.Pa^{-1}$ .

Ce coefficient est lié aux caractéristiques physico-chimiques de l'électrolyte (l'eau) et du milieu poreux ainsi que les interactions entre eux, ce qui nous conduit à revenir sur l'origine microscopique de l'électrofiltration et à la théorie de double couche électrique (Electrical Double Layer, EDL).

### 1.2.2 Modèle de double couche électrique

La surface des grains qui forment le milieu poreux possède généralement une charge négative provoquée par des interactions chimiques entre les hydroxyles existants à la surface du minéral (silanol par exemple) et l'électrolyte ( $SiOH \Leftrightarrow SiO^- + H^+$  pour la silice, e.g., Revil *et al.*, 1999). Cette surface chargée, génère un champ électrostatique qui influence les ions transportés par les molécules d'eau selon leurs signes de charges : les cations sont attirés au voisinage de la surface des grains alors que les anions sont repoussés. Donc, il existe deux couches entre la surface des minéraux et le centre de l'espace porale.

Le premier modèle proposé par Helmholtz en 1879, simule le milieu poral comme deux couches : la première est chargée négativement et la deuxième positivement. Gouy (1910) et Chapman (1913) ont amélioré ce modèle en ajoutant une couche intermédiaire, entre la surface du minéral et l'électrolyte, dite couche de diffuse, dans laquelle la concentration des espèces chargées diminue en s'éloignant des grains suivant une distribution de Poisson-Boltzman. En 1924, Stern propose un modèle qui prend en compte les dimensions réelles des particules. En effet, Il décrit une couche compacte appelée couche de Stern et une couche diffuse séparées par un plan de cisaillement qui présente la limite entre les particules fixes et les particules mobiles lors d'un écoulement. La couche de Stern se divise elle-même en deux sous-couches : une sous-couche interne de Helmholtz (Inner Helmholtz Plane, IHP) où les ions sont accolés chimiquement à la surface du minéral, et une externe de Helmholtz (Outer Helmholtz Plane, OHP) dans laquelle les ions sont à l'équilibre électrostatique avec les charges adsorbées (physisorption).

Cette distribution de charge fournit un potentiel électrique variable linéairement avec deux pentes dans la couche de Stern, et il décroît en exponentielle dans la couche diffuse jusqu'à la zone de l'électrolyte libre, où le potentiel électrique est nul grâce à l'électroneutralité local (voir la courbe sur la Figure 1.5).

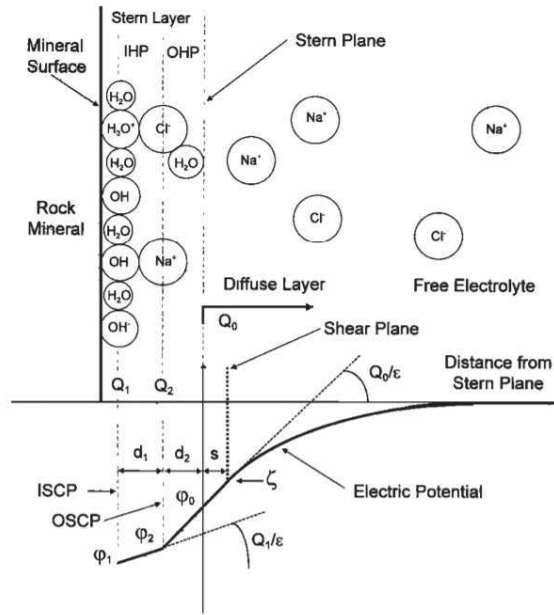


Figure 1.5: Schéma de la double couche électrique (Lorne et al. 1999)

L'expression du potentiel électrique dans la couche diffuse proposée par Debye et Hückel en 1923 suit la distribution de Poisson-Boltzmann :

$$\nabla^2 \varphi = - \sum_{l=1}^L \frac{e z_l N_l}{\varepsilon_0 \varepsilon_f} \exp\left(\frac{e z_l}{kT} \varphi\right) \quad (1.16)$$

Où  $\varphi$  est le potentiel électrique dans la couche diffuse,  $L$  est le nombre d'espèces présentes,  $e$  est la charge élémentaire ( $e = 1.6 \times 10^{-19} C$ ),  $z_l$  est la valence des ions de l'espèce  $l$  présents,  $N_l$  est le nombre d'ions de l'espèce  $l$  présents par unité de volume,  $\varepsilon_0$  et  $\varepsilon_f$  sont respectivement les permittivités électriques du vide et du fluide ( $\varepsilon_0 = 8.85 \times 10^{-12} F.m^{-1}$  et  $\varepsilon_f = 80 F.m^{-1}$ ),  $k$  est la constante de Boltzmann ( $k = 1.28 \times 10^{-23} J.K^{-1}$ ) et  $T$  la température exprimée en  $K$ .

La solution de cette équation est définie par la relation suivante (Pride, 1994) :

$$\varphi(\chi) = \zeta \exp\left(-\frac{\chi}{\chi_D}\right) \quad (1.17)$$

Où  $\chi$  est la distance normale à la surface des grains,  $\zeta$  est le potentiel zêta, qui correspond au potentiel électrique de la surface de cisaillement entre la couche de Stern et la couche diffuse

et  $\chi_D$  présente l'épaisseur de la couche diffuse qui correspond à l'influence électrique de la surface de grain, qui est égale à la longueur de Debye, exprimée par :

$$\frac{1}{\chi_D^2} = \sum_{l=1}^L \frac{(ez_l)^2 N_l}{\varepsilon_0 \varepsilon_f kT} \quad (1.18)$$

Cette longueur qui est de l'ordre de 10 à 300 Å (Revil et Glover, 1997) dépend de la concentration de l'électrolyte en ions. Plus cette concentration est faible, plus l'épaisseur de la couche diffuse est grande, et par conséquent le potentiel mesuré important. Par exemple, la conductivité de l'eau douce est plus faible que la conductivité de l'eau de mer dont la concentration en ions est plus élevée. De ce fait, la longueur de Debye pour l'eau douce est plus importante que pour l'eau de mer.

Plusieurs facteurs peuvent agir sur le potentiel spontané. La vitesse d'écoulement d'électrolyte et le degré de saturation du milieu poreux sont les plus importants paramètres qui caractérisent ce potentiel.

### 1.2.3 Couplage électrocinétique

Dans le cas d'un milieu poreux non saturé, isotrope et homogène dont l'écoulement est laminaire, les flux associés au phénomène physique sont linéairement reliés aux forces appliquées selon la formulation des processus thermodynamiques irréversibles de Onsager (1931). La densité de courant électrique  $J$  [ $I.L^{-2}$ ] et le flux hydraulique  $q$  [ $L.T^{-1}$ ] s'écrivent:

$$\vec{J} = -L_{11} \vec{\nabla} \varphi - L_{12} \vec{\nabla} P \quad (1.19)$$

$$\vec{q} = -L_{21} \vec{\nabla} \varphi - L_{22} \vec{\nabla} P \quad (1.20)$$

Où  $\varphi$  est le potentiel électrique ( $V$ ) [ $M.L^2.T^{-3}.I^{-1}$ ],  $P = \rho g H$  est la pression hydrostatique [ $M.L^{-1}.T^{-2}$ ] qui décrit l'état énergétique de l'eau dans le milieu avec  $\rho$  est la masse volumique de fluide [ $M.L^{-3}$ ],  $g$  est l'accélération de pesanteur [ $L.T^{-2}$ ] et  $H$  est la charge hydraulique [ $L$ ]. Les coefficients propres  $L_{ii}$  sont obtenus à partir des lois physiques telles que la loi d'Ohm et la loi de Darcy et les  $L_{ij}$  sont les paramètres phénoménologiques ou les



coefficients de couplage. Ceci donne  $L_{11} = -\sigma$ , où  $\sigma$  est la conductivité électrique du milieu ( $S.m^{-1}$ )  $[M^{-1}L^{-3}T^3I^2]$  et  $L_{22} = k/\mu$  avec  $k [L^2]$  la perméabilité intrinsèque du milieu poreux et  $\mu [M.L.T^{-1}]$  la viscosité dynamique du fluide.

En utilisant la réciprocité d'Onsager (1931), les coefficients de couplage  $L_{12}$  et  $L_{21}$  sont égaux et sont obtenus à partir de l'expression du courant électrique généré par l'électrofiltration et du flux hydrique généré par l'électro-osmose avec  $L_{12} = L_{21} = \sigma C$  où  $C$  est le coefficient d'électrofiltration  $[L^3T^{-1}I^{-1}]$ .

En générale, le champ électrique a un effet négligeable sur la vitesse de Darcy. Le premier terme de l'électro-osmose peut donc être éliminé et on obtient les deux relations de couplage électrocinétique sous la forme :

$$\vec{J} = -\sigma \vec{\nabla} \varphi - \sigma C \rho g \vec{\nabla} H \quad (1.21)$$

$$\vec{q} = -\frac{k}{\mu} \rho g \vec{\nabla} H \quad (1.22)$$

Le milieu poreux est supposé conservatif, en l'absence de source externe de courant, l'équation de conservation de la charge s'écrit :

$$\nabla \cdot \vec{J} = 0 \quad (1.23)$$

Si le milieu poreux est saturé, homogène et isotrope, le coefficient d'électrofiltration  $C$  et la conductivité électrique  $\sigma$  sont constants, l'équation 1.21 se réduit à l'équation de Poisson suivante :

$$\nabla^2 \varphi = C \rho_w g \nabla^2 H \quad (1.24)$$

En réalité, le coefficient d'électrofiltration  $C$  dépend de la conductivité électrique du milieu  $\sigma$  qui est variable dans le cas des milieux poreux non saturés. L'équation de couplage électrocinétique s'écrit sous la forme :

$$\nabla \cdot (\sigma \vec{\nabla} \varphi) = \nabla \cdot (\sigma C \rho_w g \vec{\nabla} H) \quad (1.25)$$

#### 1.2.4 Coefficient d'infiltration

Pour un écoulement stationnaire d'un électrolyte monovalent dans un capillaire cylindrique saturé en eau dont le rayon est grand devant la longueur de Debye, le Coefficient d'électrofiltration est exprimé suivant la formule de Helmholtz-Smoluchowski (1903) :

$$C_{HS} = \frac{\varepsilon_0 \varepsilon_r \zeta}{\eta \sigma_f} \quad (1.26)$$

Avec  $\varepsilon_0$  et  $\varepsilon_r$  la permittivité électrique du vide et la permittivité électrique relative de fluide,  $\sigma_f$  est la conductivité électrique du fluide ( $S.m^{-1}$ ),  $\mu$  est sa viscosité dynamique et  $\zeta$  est le potentiel zêta ( $V$ ) définie comme le potentiel sur le plan de cisaillement de double couche électrique (Revil *et al.*, 1999a). Cette expression est utilisée pour les milieux poreux saturés. Dans notre cas d'étude il faut rajouter l'effet de la variation de la teneur en eau. Le coefficient d'électrofiltration est relié à la teneur en eau à travers la conductivité électrique du milieu poreux  $\sigma$ . Cette dernière est donnée par la loi d'Archie pour l'écoulement d'eau dans les sols non saturés en fonction du degré de saturation  $S_w$  (Archi, 1942) :

$$\sigma = \frac{1}{F} S_w^{n_a} \sigma_w \quad (1.27)$$

Où  $F = \theta^{-m}$  est le facteur de formation électrique,  $m$  est l'exposant de cimentation du sol,  $n_a$  est l'exposant de saturation d'Archie et  $\sigma_w$  est la conductivité électrique de l'eau. Cette expression est valable en supposant négligeable la conductivité électrique de la surface des grains.

Plusieurs modèles alternatifs ont été présentés pour la relation  $C(\theta)$  telle que Revil *et al.* (1999) b) qui proposent la définition suivante :

$$C = \frac{C_{sat}}{S_e^{n_a} \left[ 1 + 2 \left( \frac{F}{S_e^{n_a}} - 1 \right) \frac{\xi}{S_e} \right]} \quad (1.28)$$

Où  $C_{sat} = C_{HS}$ , le coefficient d'électrofiltration à la saturation définie par l'équation de Helmholtz-Smoluchowski,  $\xi$  est le rapport entre la conductivité électrique de surface et la conductivité électrique du fluide et  $S_e$  est la saturation effective définie par la relation suivante :

$$S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r} = \frac{S_w - S_w^r}{1 - S_w^r} \quad (1.29)$$

avec  $\theta$ ,  $\theta_r$  et  $\theta_s$  les teneurs en eau instantanée, résiduelle et à saturation aussi  $S_w = \theta/\theta_s$  et  $S_w^r = \theta_r/\theta_s$  sont les degrés de saturation en eau instantanée et résiduelle. Ce modèle est utilisé par Darnet *et al.* (2003) et Sailhac *et al.* (2004) pour prédire le coefficient de filtration.

En supposant que la conductivité électrique de la surface est négligeable pour les matériaux poreux dominés par de la silice (c'est-à-dire que  $\xi$  proche de zéro), l'équation (1.28) devient (Darnet et Marquis, 2004) :

$$C = \frac{C_{sat}}{S_e^{n_a}} \quad (1.30)$$

La première expérience qui mesure la teneur en eau et les différences de potentiel spontané a été effectuée par Guichet *et al.* (2003). Ces auteurs ont proposé un coefficient de couplage dépendant linéairement de  $S_e$  :

$$C = C_{sat} S_e \quad (1.31)$$

Par analogie à la relation constructive de la conductivité hydraulique dans les sols non saturés  $K(\theta)$ , des modèles du coefficient d'électrofiltration en fonction de la perméabilité relative  $k_r$  ont été proposés tels que le modèle proposé par Perrier and Morat (2000) :

$$C = C_{sat} \frac{k_r}{\sigma_r} \quad (1.32)$$

Où  $\sigma_r$  est la conductivité électrique relative du milieu poreux définie par  $\sigma_r = S_w^{n_a}$ , la perméabilité relative  $k_r$  est définie par la relation constructive de Van Gunechten-Mualem (équation 1.14).

Cette expression du coefficient d'électrofiltration a été modifiée par Linde et Revil (2007) ainsi que Revil et al. (2007).

$$C = C_{sat} \frac{k_r}{S_w \sigma_r} = \frac{C_{sat} k_r}{S_w^{n_a+1}} \quad (1.33)$$

Ce modèle a été validé en comparaison avec des mesures expérimentales (Revil et Cerepi, 2004) comme présenté dans la Figure 1.6.

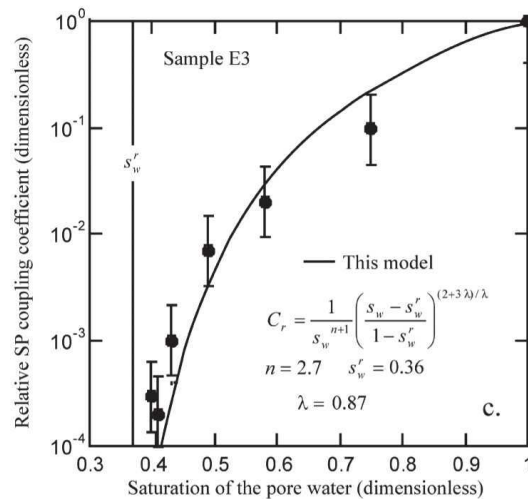


Figure 1.6: Comparaison du modèle de coefficient d'électrofiltration (1.33) avec des données mesurées sur des échantillons de calcaire (Revil et Cerepi, 2004)

### 1.2.5 Couplage hydrogéophysique

Les expressions du coefficient d'infiltration  $C$  (1.33) et de la conductivité électrique du milieu poreux  $\sigma$  (1.27) sont substituées dans l'équation du couplage électrocinétique (1.25). Le gradient de la pression dans cette équation est obtenu à partir de la vitesse de Darcy. Cette dernière est calculée après la résolution de l'équation de Richards à chaque instant  $t$  de la discrétisation temporelle pour obtenir la relation suivante :

$$\nabla \cdot \left( \frac{1}{F} S_w^{n_a} \sigma_w \vec{\nabla} \varphi \right) = \nabla \cdot \left( \frac{1}{F} S_w^{n_a} \sigma_w \frac{C_{sat} k_r(h)}{S_w^{n_a+1}} \frac{\rho_w \mathbf{g}}{-K(h)} \vec{q} \right) \quad (1.34)$$

Sachant que le facteur de formation électrique  $F$  et la conductivité électrique de l'eau  $\sigma_w$  sont supposés des constantes alors on peut les simplifier. En outre, on rappelle l'expression du degré de saturation  $S_w = \theta/\theta_s$  et de la conductivité hydraulique  $K(h) = K_s \cdot k_r(h)$ , c'est-à-dire que l'équation (1.34) peut être formulé sous la forme de l'équation du couplage électrocinétique représentée dans la figure 1.7.

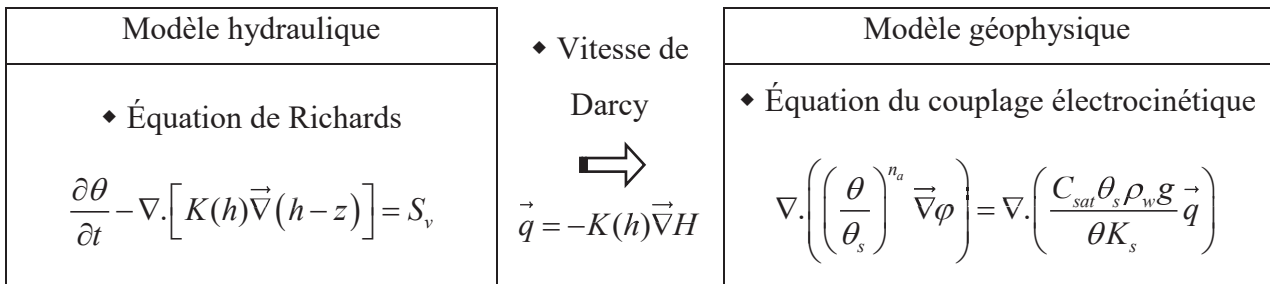


Figure 1.7: Schéma illustratif du couplage hydrogéophysique en milieux poreux non-saturé

### 1.3 Transport de soluté en milieux poreux non saturés

Naturellement, l'eau en mouvement dans les réseaux poreux du sol contient un grand nombre d'espèces chimiques produites par les processus géochimiques, biologiques et par les activités anthropiques. Ces éléments dissouts dans l'eau sont transportés de la surface du sol vers la nappe à travers la zone non saturée. Dans ce travail, nous nous intéressons au transport des espèces n'interagissant pas entre eux ou avec les minéraux des grains solides, appelées traceurs de l'écoulement. Les mécanismes régissant le transfert d'un traceur en milieu poreux, sont l'advection, la diffusion moléculaire et la dispersion cinématique.

### 1.3.1 L'advection

La migration des solutés lors de l'écoulement de l'eau dans les sols est appelée advection ou convection. Elle présente l'entraînement des espèces dissoutes sous la seule action du mouvement du fluide. Le déplacement moyen du soluté est alors directement proportionnel à la vitesse moyenne d'écoulement de l'eau, régie par la loi de Darcy. Le flux convectif du soluté s'exprime par le produit de la concentration  $C$  [ $M.L^{-3}$ ] et de la vitesse de Darcy  $q$  [ $L.T^{-1}$ ].

Pour un fluide incompressible, l'équation de transport par convection est donnée par :

$$\frac{\partial(\theta C)}{\partial t} + \nabla \cdot (\vec{q}C) = 0 \quad (1.35)$$

Où  $\theta$  est la teneur en eau du milieu poreux [-] et  $\theta C$  représente la concentration en soluté de la phase liquide.

### 1.3.2 Diffusion moléculaire

La diffusion moléculaire est un processus indépendant de l'écoulement de l'eau dans le milieu poreux. Elle est liée à la présence de gradient de concentration dans l'eau. Ce mécanisme produit un transfert de masse des zones à forte concentration vers les zones moins concentrées en soluté afin d'homogénéiser la composition du fluide. Ce transfert généré par l'agitation brownienne est modélisé par la loi de Fick (1855) :

$$\frac{\partial(\theta C)}{\partial t} - \nabla \cdot (\underline{\underline{D}}_m \vec{\nabla} C) = 0 \quad (1.36)$$

Où,  $\underline{\underline{D}}_m$  [ $L^2.T^{-1}$ ] est le tenseur diagonal de diffusion moléculaire qui dépend du milieu poreux, du soluté et de la température.

### 1.3.3 Dispersion cinématique

À l'échelle microscopique, le champ de vitesse de l'écoulement dans un milieu poreux est hétérogène, ce qui engendre un phénomène complexe appelé dispersion cinématique. Trois processus sont principalement à l'origine de ce phénomène (Fried et Combarous, 1971) :

- En assimilant l'intérieur d'un pore à un capillaire, le profil de vitesse est parabolique à cause du frottement engendré par la viscosité du fluide. (Figure 1.7 a)
- La variation des vitesses au sein d'un même pore due à la variation des dimensions des pores. (Figure 1.7 b)
- Les lignes de courant fluctuent par rapport à la direction principale de l'écoulement, engendrant alors des variations dans la direction des vitesses. (Figure 1.7 c)

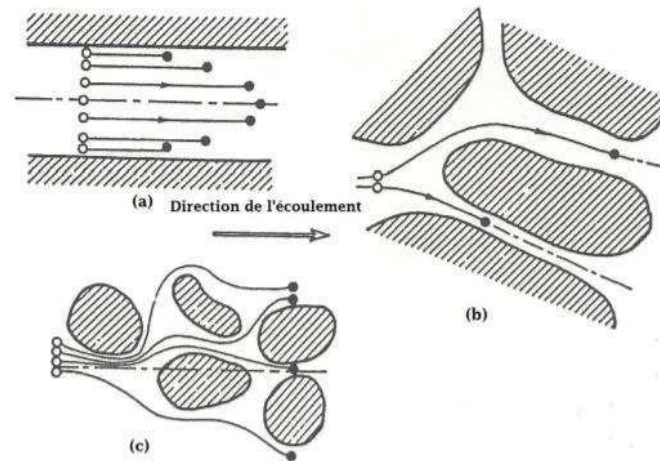


Figure 1.8: Les processus responsables de la dispersion cinématique (Besnard K., 2003)

La dispersion cinématique dépend de la direction moyenne de l'écoulement ; on distingue une dispersion longitudinale et une dispersion transversale qui a lieu dans le plan orthogonal à la direction principale. A cause de l'anisotropie de ce phénomène, la dispersion cinématique  $\underline{\underline{D_c}}$   $[L^2.T^{-1}]$  s'exprime sous une forme tensorielle, symétrique de second ordre :

$$\underline{\underline{D_c}} = \begin{vmatrix} D_{c_{\perp}} & D_{c_{ij}} & D_{c_{ij}} \\ D_{c_{ij}} & D_{c_{\perp}} & D_{c_{ij}} \\ D_{c_{ij}} & D_{c_{ij}} & D_{c_{//}} \end{vmatrix} \quad (1.37)$$

On définit  $D_{c_{//}}$  comme le coefficient de dispersion dans la direction longitudinale et  $D_{c_{\perp}}$  comme le coefficient de dispersion dans la direction transversale à l'écoulement. Ces coefficients sont proportionnels au module de la vitesse de darcy par une forme similaire à la loi de Frick (Scheidegger, 1974) :

$$D_{c_{//}} = \alpha_{//} \|\vec{q}\| \quad (1.38)$$

$$D_{c_{\perp}} = \alpha_{\perp} \|\vec{q}\| \quad (1.39)$$

$$D_{c_{ij}} = (\alpha_{//} - \alpha_{\perp}) \frac{q_i q_j}{\|\vec{q}\|^2} \quad (1.40)$$

Où,  $\alpha_{//}$  et  $\alpha_{\perp}$  sont les coefficients de proportionnalités de la dispersion longitudinale et transversale, en outre,  $q_i$  et  $q_j$  sont les composantes de la vitesse de darcy.

En générale, L'ordre de grandeur de la valeur du coefficient de dispersion transversale est plus faible que celle de la dispersion longitudinale avec un rapport  $\alpha_{//}/\alpha_{\perp}$  variant entre 6 et 20, d'après Anderson (1979) et Koltz et al. (1980).

L'équation de transport par dispersion cinématique est donnée par la loi de Frick :

$$\frac{\partial(\theta C)}{\partial t} - \nabla \cdot (\underline{\underline{D}}_c \vec{\nabla} C) = 0 \quad (1.41)$$

#### 1.3.4 Dispersion hydrodynamique

En pratique, l'effet de la dispersion cinématique et l'effet de la dispersion moléculaire sont indissociables. C'est pourquoi ces deux contributions sont regroupés dans un même tenseur appelé tenseur de dispersion hydrodynamique.

$$\underline{\underline{D}} = \underline{\underline{D}}_m + \underline{\underline{D}}_c \quad (1.42)$$

La variation de la concentration dans le temps et l'espace est exprimée par la loi de Frick :

$$\frac{\partial(\theta C)}{\partial t} - \nabla \cdot (\underline{\underline{D}} \vec{\nabla} C) = 0 \quad (1.43)$$



### 1.3.5 L'équation d'advection-dispersion

En tenant compte à la fois des phénomènes de convection, des phénomènes de dispersion, des termes source  $Q_s$  avec une concentration  $C_s$ , le bilan de matière du transport d'un soluté conservatif amène à considérer une équation de la forme :

$$\frac{\partial(\theta C)}{\partial t} + \nabla \cdot (\vec{q}C) - \nabla \cdot (\underline{\underline{D}} \nabla C) = Q_s C_s \quad (1.44)$$

Ce type de modèles est généralement utilisé pour les transports des solutés dans les milieux où uniquement des phénomènes convectifs et dispersifs sont produits. Dans ce cas, on suppose que la distribution des tailles des pores est monomodale et qu'il n'existe pas de phase immobile qui échange avec le fluide en circulation.

### 1.3.6 Modèle mobile-immobile

Coats et Smith (1964) puis Van Genuchten et Wierenga (1976) ont développé un modèle de transport qui permet de distinguer des régions du volume poral où l'écoulement est nul à cause de la faible connectivité des pores. Par conséquent, le transport de masse s'y fait seulement par l'échange diffusif qui dépend du gradient de concentration entre l'eau mobile et l'eau immobile. Ce modèle est représenté par une équation d'advection-dispersion dans le sous-domaine d'eau mobile et uniquement par une équation de diffusion moléculaire dans le sous-domaine d'eau immobile. Cette représentation est exprimée par (van Genuchten and Wagenet., 1989; Clothier et al., 1998):

$$\frac{\partial(\theta_m C_m)}{\partial t} + \frac{\partial(\theta_{im} C_{im})}{\partial t} + \nabla \cdot (\vec{q}C_m) - \nabla \cdot (\theta_m \underline{\underline{D}} \nabla C_m) = 0 \quad (1.45)$$

$$\frac{\partial(\theta_m C_m)}{\partial t} = \omega(C_m - C_{im}) \quad (1.46)$$

$$\theta = \theta_m + \theta_{im} \quad (1.47)$$

avec ( $m$ ) et ( $im$ ) pour désigner respectivement la phase mobile et la phase immobile. L'échange entre les deux phases est régi par une cinétique de premier ordre avec  $\omega$  le coefficient de transfert de masse entre les deux phases. La teneur en eau de la phase mobile et celle de la phase

immobile peuvent varier dans l'espace et dans le temps, mais nous supposons que le rapport : teneur en eau mobile sur teneur en eau totale ( $f = \theta_m / \theta$ ) reste constant. De ce fait, la fraction  $\theta_m / \theta$  reste également constante. Cette hypothèse est moins restrictive que de supposer  $\theta_{im}$  constant lors de la simulation du transfert de masse en milieu poreux non saturé car une telle hypothèse peut influencer la résolution de l'équation de Richards (Moreira et al., 2016).

### 1.3.7 Couplage écoulement-transport en milieux poreux non-saturé

La résolution numérique du problème de transport, modélisé par l'équation (1.42) dans le cas d'un simple milieu et par le système d'équations (1.44)-(1.45) dans le cas d'un double milieu nécessite la connaissance de la vitesse d'écoulement et de la teneur en eau. Ces deux variables sont obtenues à partir de la résolution de l'équation de Richards à chaque instant  $t$  de la discrétisation temporelle (Figure 1.8).

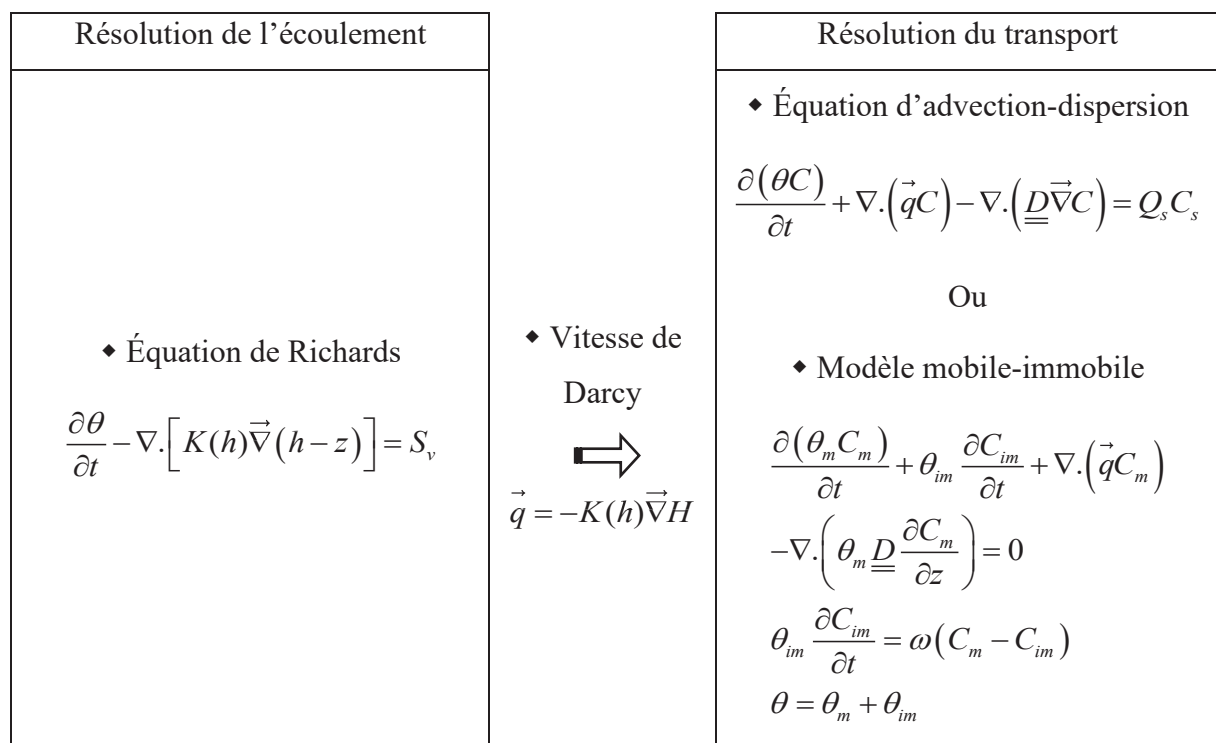


Figure 1.9: Principe de couplage écoulement-transport en milieux poreux non-saturé

## Conclusion

Dans ce chapitre, nous avons présenté les modèles mathématiques permettant de décrire les phénomènes physiques étudiés dans la suite des travaux de la thèse. En effet, l'écoulement d'eau dans un milieu poreux saturé ou non saturé est modélisé par l'équation de Richards accompagnée par le modèle de van Genuchten – Mualem qui donne la conductivité hydraulique en fonction de la pression  $K(h)$  et la relation de rétention d'eau  $\theta(h)$ . De plus, le potentiel spontané est représenté par l'équation du couplage électrocinétique, enfin le phénomène de transport des solutés dans les milieux poreux est simulé par l'équation d'advection-dispersion ou par le modèle mobile-immobile.

Tous ces modèles contiennent un nombre important des paramètres dont la plupart sont non mesurables. Dans le chapitre suivant, nous allons présenter l'analyse de sensibilité globale qui permet l'évaluation de la contribution de chaque paramètre à l'incertitude du modèle, ainsi que le problème inverse qui permet d'estimer les valeurs des paramètres non mesurables.

## Bibliographie

- Anderson, J. E., 1979. "A Theoretical Foundation for the Gravity Equation," *American Economic Review*, 69(1), 106-16.
- Archie, G.E., 1942. The electrical resistivity log as an aid in determining some reservoir characteristics. *Trans. Am. Inst. Min. Metall. Pet. Eng.* 146:54–62.
- Bernoulli D., 1738. *Hydrodynamica*. Published by sumptibus Johannis Reinholdi Dulseckeri.
- Brooks R. H. et A. T. Corey, 1964. Hydraulic properties of porous media. *Hydrol. Pap.* 3. Colorado State Univ., Fort Collins.
- Brutsaert W., 1966. Probability laws for pore-size distributions. *Soil Science* 101(2):85-92.
- Buckingham, E., 1907. Studies on the movement of soil moisture. *Bull.*38. USDA, Bureau of Soils, Washington D.C.
- Campbell G. S., 1974. A simple method for determining unsaturated conductivity from moisture retention data. *Soil Science* 117(6) :311-314.
- Carsel R. F. et R. S. Parrish, 1988. Developing joint probability distributions of soil water retention characteristics. *Water Resources Research* 24:755-769.
- Celia M. A., Bouloutas E. T. and Zarba R. L., 1990. A general mass-conservative numerical solution for the unsaturated flow equation. *Water Resources Research* 26(7):1483-1496.
- Chapman, D.L., 1913. *Philos. Mag.* 25(6).
- COATS, K. H, AND SMITH B. D., 1964. Dead end pore volume and dispersion in porous media. *Soc. Pet. Eng. J.* 4:73-84.
- Darcy, H., 1856. Détermination des lois d'écoulement de l'eau à travers le sable. p. 590–594. In *Les Fontaines Publiques de la Ville de Dijon*. Victor Dalmont, Paris.
- Darnet, M., G. Marquis, and P. Sailhac, 2003. Estimating aquifer hydraulic properties from the inversion of subsurface streaming potential (SP) anomalies. *Geophys. Res. Lett.* 30(13):1679.

- Darnet, M., and G. Marquis., 2004. Modeling streaming potential (SP) signals induced by water movement in the vadose zone. *J. Hydrol.* 285:114–124.
- De Smedt, F., Wierenga, P.J., 1984. Solute transfer through columns of glass beads. *Water Resour. Res.*, 20, 225–232.
- Debye, P. and Hückel, E. 1923. *Phys. Z.*, 24, 185.
- Fick A., « Über Diffusion », *Annalen der Physik und Chemie*, vol. 94, 1855, p. 59–86
- Freeze, R.A., 1971. Three-dimensional, transient, saturated-unsaturated flow in groundwater basin. *Water Resour. Res.* 7 :347-366.
- Gardner W. R., 1958. Some steady-state solutions of the unsaturated moisture flow equation with application to evaporation from a water table. *Soil Science* 85:228-232.
- Gouy, G, 1910. *J. Phys. Radium* 9(4), 457.
- Guichet, X., L. Jouniaux, and J. Pozzi. 2003. Streaming potential of a sand column in partial saturation conditions. *J. Geophys. Res.* 108(B3):2141.
- Helmholtz H., 1879. Studien über elektrische grenzschichten. *Annalen der physik und chemie* 7(7), 337-382.
- Hillel D., 1980. *Fundamentals of Soil Physics*. Academic Press. 6, 314.
- Klotz, D., Seiler, K.-P., Moser, H. and Neumaier, F., 1980. Dispersivity and velocity relationship from laboratory and field experiments. *J. Hydrol.* 45: 169-184.
- Lenhard R. J., J. C. Parker et S. Mishra, 1989. On the correspondence between Brooks-Corey and van Genuchten models. *Journal of Irrigation and Drainage Engineering* 115(4): 744-751.
- Li C. W., 1993. A simplified Newton iteration method with linear finite elements for transient unsaturated flow. *Water Resources Research* 29(4):965-971.
- Linde, N., and A. Revil. 2007. Inverting self-potential data for redox potentials of contaminant plumes. *Geophys. Res. Lett.* 34:L14302.

Lorne B., Perrier F. and Avouav J.,P., 1999. Streaming potential measurements 1. Properties of the electrical double layer from crushed rock samples. *Journal of Geophysical Research*, 104 (B8) ; 17857-17877.

Maraqa, M.A., Wallace, R.B., Voice, T.C., 1997. Effect of degree of saturation on dispersivity and immobile water in sandy columns. *J. Contam. Hydrol.*, 25, 199–218.

Milly P. C. D., 1985. A mass-conservative procedure for time-stepping in models of unsaturated flow. *Advances in Water Resources* 8:32-36.

Morel-Seytoux H. J., P. D. Meyer, M. Nachabe, J. Touma, M. T. van Genuchten et R. Lenhard, 1996. Parameter equivalence for the Brooks-Corey and van Genuchten soil characteristics : preserving the effective capillary drive. *Water Resources Research* 32(5):1251-1258.

Moreira PHS, van Genuchten, MT, Orlande HRB, Cotta RM. 2016. Bayesian estimation of the hydraulic and solute transport properties of a small-scale unsaturated soil column. *Journal of Hydrology and Hydromechanics*, 64(1), pp. 30-44.

Mualem Y., 1976. A New Model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research* 12(3):513-522

Narasimhan, T.N. 2004. Darcy's law and unsaturated. *Vadose zone J.3* :1059.

Narasimhan T.N. 2006. On storage coefficient and vertical Strain. *Ground Water* 44.488-491.

Onsager, L., 1931. Reciprocal relation in irreversible processes : I, *Phys. Rev*, 37(4).

Perrier, F., and P. Morat. 2000. Characterization of electrical daily variations induced by capillary flow in the non-saturated zone. *Pure Appl. Geophys.* 157:785–810.

Pride, S., 1994. Governing equations for the coupled electromagnetics and acoustics of porous media. *Physical Review B*, 50, 15678-15695.

Rawls W. J., D. L. Brakensiek et K. E. Saxton, 1982. Estimation of soil water properties. *Trans. ASAE* 81-2510:1316-1320.

Rivière, A. 1977. *Méthodes granulométriques : Techniques et interprétations*. Paris : Masson.

Revil, A. and Glover, P.W., 1997. Theory of ionic-surface electrical conduction in porous media, *Phys. Rev. B*, 55(3), 1757-1773.

Revil, A., Pezard, P.A. and Glover, P.W.J., 1999. Streaming potential in porous media : 1. Theory of the zeta potential, *J. geophys. Res.*, 104(B9), 20021-20031.

Revil, A., H. Schwaeger, and P.D. Manhardt., 1999. Streaming potential in porous media: 2. Theory and application to geothermal systems. *J. Geophys. Res.* 104(B9):20033–20048.

Revil, A., and A. Cerepi. 2004. Streaming potentials in two-phase flow conditions. *Geophys. Res. Lett.* 31: L11605.

Revil, A., N. Linde, A. Cerepi, D. Jougnot, S. Matthai, and S. Finsterle. 2007. Electrokinetic coupling in unsaturated porous media. *J. Colloid Interface Sci.* 313:315–327.

Russo D., E. Bresler, U; Shani et J. C. Parker, 1991. Analyses of infiltration events in relation to determining soil hydraulic properties by inverse problem methodology. *Water Resources Research* 27(6):1361-1373.

Russo D; 1988. Determining soil hydraulic properties by parameter estimation : on the selection of a model for the hydraulic properties. *Water Resources Research* 24(3): 453- 459.

Smoluchowski, M. 1905. Theory of electrophoretic deposition and surface conduction. (In German.) *Phys. Z.* 6:529–536.

Sailhac, P., M. Darnet, and G. Marquis. 2004. Electrical streaming potential measured at the ground surface : Forward modeling and inversion issues for monitoring infiltration and characterizing the vadose zone. *Vadose Zone J.* 1200–1206.

Stern, O., 1924. Zur theorie der electrolytischen doppelschicht. *Z. Electrochem.*, 30, 508-516.

Van Genuchten M. Th., and Wierenga P.J., 1976. Mass transfer studies in sorbing porous media : 1. Analytical solutions, *Soil Sci. Soc. Amer. J* : 40(4), 473-480.

Van Genuchten M. Th., 1980. A closed-form equation for predicting the hydraulic conductivity of unstrated soils. *Soil Science of America Journal* 44:892-898.

## Tables des figures

Figure 1.1: Classification des sols suivant les diamètres des particules appelé « Triangle des textures » .....	6
Figure 1.2 : Teneur en eau en fonction de la pression d'eau calculée par modèle van Genuchten – Mualem pour trois classes de sols .....	12
Figure 1.3: Conductivité hydraulique en fonction de la pression d'eau calculée par modèle van Genuchten – Mualem pour trois classes de sols.....	12
Figure 1.4: Schéma illustratif du principe de l'électrofiltration .....	14
Figure 1.5: Schéma de la double couche électrique (Lorne et al. 1999) .....	16
Figure 1.6: Comparaison du modèle de coefficient d'électrofiltration (1.33) avec des données mesurées sur des échantillons de calcaire .....	21
Figure 1.7: Schéma illustratif du couplage hydrogéophysique en milieux poreux non-saturé	22
Figure 1.8: Les processus responsable à la dispersion cinématique .....	24
Figure 1.9: Principe de couplage écoulement-transport en milieux poreux non-saturé.....	27

## Tables des tableaux

Tableau 1.1 : Valeurs moyenne des paramètres du modèle van Genuchten – Mualem pour douze classes de sols.....	13
---------------------------------------------------------------------------------------------------------------	----



# Chapitre 2. Méthodes d'analyse de sensibilité globale et d'estimation des paramètres

## Plan du chapitre

Chapitre 2.....	Méthodes d'analyse de sensibilité globale et d'estimation des paramètres	34
2.1	Introduction .....	35
2.2	Analyse de sensibilité globale .....	36
2.2.1	Introduction .....	36
2.2.2	Identification des sources d'incertitude.....	36
2.2.3	Propagation d'incertitude .....	37
2.2.4	Analyse de sensibilité.....	40
2.2.5	Expansion en polynômes de chaos .....	48
2.3	Estimation des paramètres .....	66
2.3.1	Introduction : .....	66
2.3.2	Méthodes de résolution du problème inverse : .....	67
2.3.3	Approximation de premier ordre (FOA) .....	70
2.3.4	Approche Bayésienne.....	75
2.4	Conclusion .....	85

## 2.1 Introduction

La modélisation représente la meilleure solution pour comprendre et reproduire les processus de transfert dans le sol. Cependant, elle fournit souvent des résultats entachés d'incertitudes à cause des erreurs de mesures et du manque de données. Il est donc primordial d'identifier l'origine de ces erreurs et de les quantifier avant d'estimer les paramètres du modèle.

Dans un premier temps, on décrira les théories de base de la propagation d'incertitude et l'analyse de sensibilité globale basée sur les indices de Sobol. Plusieurs techniques peuvent être utilisées pour le calcul de ces indices parmi lesquelles le développement en chaos polynomial qui sera détaillé dans ce chapitre.

Dans un second temps, quelques approches d'estimation de paramètres et d'incertitude à posteriori vont être présentées. On va s'intéresser spécialement à deux types d'approches ; la première utilise l'algorithme de Levenberg-Marquardt pour l'estimation des paramètres et une approximation de premier ordre (FOA) pour l'incertitude à posteriori. La deuxième est une approche bayésienne utilisant un échantillonneur basé sur les chaînes de Markov. Cette dernière offre une base rigoureuse pour l'estimation des paramètres à partir de données bruitées et pour les modèles incertains, en intégrant toutes les informations disponibles à priori.

## 2.2 Analyse de sensibilité globale

### 2.2.1 Introduction

Les modèles mathématiques développés dans le premier chapitre permettent de représenter les processus de transfert dans la zone non saturée à l'aide d'équations aux dérivées partielles non linéaires. Ces modèles sont basés sur une simplification de la réalité et négligent l'influence de plusieurs facteurs (température, hystérésis, géochimie, ...). Ces simplifications impactent la précision et l'applicabilité des modèles dont la fiabilité peut être mise en question. George Box disait « Tous les modèles sont faux, certains sont utiles » (Box et Draper 1987). De plus, les modèles mathématiques développés font intervenir plusieurs paramètres et conditions aux limites dont les valeurs ne sont pas connues avec grande précision. Ceci peut également impacter la précision des modèles. La caractérisation à priori de ces incertitudes, permet de donner des informations sur l'incertitude associée aux prédictions du modèle et d'optimiser les variables décisionnelles.

### 2.2.2 Identification des sources d'incertitude

Kennedy et O'Hagan (2001) ont présenté diverses sources d'incertitude. Ces dernières peuvent être classées en trois familles.

- *Incertaince des paramètres :*

Certains paramètres du modèle mathématique amènent des incertitudes importantes. Ces paramètres sont les conditions aux limites, les valeurs initiales, la géométrie ainsi que des constantes physiques des lois empiriques issue de faits expérimentaux ou validée par l'expérience avec des erreurs d'observation et d'interpolation.

On peut distinguer dans ce cas deux types d'incertitudes :

- L'incertitude aléatoires qui est liée à la qualité de mesurage réalisé par l'opérateur et à la précision des instruments de mesure. Cette incertitude dite erreur de justesse est expliquée

par l'obtention des valeurs mesurées différentes lorsqu'un même opérateur répète plusieurs fois le mesurage d'une même grandeur, dans des conditions identiques.

- L'incertitude systématique ou également connue sous le nom d'incertitude épistémique qui est attachée au mal étalonnage ou à l'utilisation incorrecte de l'appareil de mesure. Elle peut également être due au non-respect du protocole de l'expérience ou au choix d'un échantillon non représentatif. Cette erreur conduit à des mesures proches les unes des autres mais écartées de la valeur réelle d'où l'appellation d'erreur de fidélité.

- *Incertitude structurelle :*

Cette incertitude traduit l'insuffisance et le décalage du modèle. Elle dépend de la précision avec laquelle un modèle mathématique décrit le phénomène réel. En effet, la modélisation de ce dernier est basée sur des différentes hypothèses simplificatrices afin d'approcher la réalité. Ces simplifications fournissent une grande partie de l'incertitude du modèle.

- *Incertitude numérique :*

Cette incertitude est également appelée incertitude algorithmique ou incertitude discrète. Généralement, les modèles mathématiques sont définis par des équations aux dérivées partielles difficiles à résoudre analytiquement. Des méthodes numériques telle que la méthode des éléments finis ou la méthode des différences finies sont souvent utilisées pour approcher le système continu par un autre discret traduisant la variation dans l'espace et dans le temps du phénomène réel. Cette discrétisation provoque des incertitudes non éliminables, mais qui peuvent être contrôlées en réduisant la taille des mailles spatiales et en diminuant les pas du temps.

### 2.2.3 Propagation d'incertitude

Avoir une réponse du modèle représenté par une distribution de probabilité qui dépend de l'incertitude des paramètres d'entrée, est l'objectif de la propagation des incertitudes (Saltelli et al., 2006). Les méthodes de propagation d'incertitude sont regroupées en trois grandes familles :

- *Méthode des moments :*

Soit un modèle mathématique défini par une fonction  $f(x)$  qui représente la sortie scalaire  $Y$  du modèle avec  $x = \{x_1, \dots, x_n\}$ , les entrées scalaires définies sur l'hypercube unité  $\Omega_n = [0, 1]^n$ . Ces entrées sont des variables continues dont la distribution est définie par une fonction de probabilité  $p_x$ . La moyenne et la variance de  $Y = f(x)$  s'écrivent :

$$\mu_Y = E[f(x)] = \int_{\Omega_n} f(x) p_x(x) dx \quad (3.1)$$

$$\sigma_Y^2 = E\left[(f(x) - \mu_Y)^2\right] = \int_{\Omega_n} (f(x) - \mu_Y)^2 p_x(x) dx \quad (3.2)$$

Plusieurs méthodes sont utilisées pour l'évaluation de ces deux moments telle que la méthode de quadrature, basée sur le principe de l'intégration numérique ou la méthode de perturbation, basée sur un développement en série de Taylor.

- *Méthodes d'échantillonnage :*

Ce sont les méthodes les plus utilisées pour quantifier l'incertitude de la réponse à partir de la densité d'incertitude des entrées. Elles consistent à calculer la sortie  $Y = f(x)$  du modèle  $N$  fois en choisissant à chaque fois un vecteur de paramètres différent  $x^i = \{x_1^i, \dots, x_n^i\}$  avec  $i \in [1, N]$ . On calcule par la suite la moyenne et la variance à partir des  $N$  valeurs de la sortie :

$$\mu_Y \approx \frac{1}{N} \sum_{i=1}^N f(x^i) \quad (3.3)$$

$$\sigma_Y^2 \approx \frac{1}{N-1} \sum_{i=1}^N (f(x^i) - \mu_Y)^2 \quad (3.4)$$

Le nombre d'échantillons  $N$  doit être relativement grand afin d'assurer la précision des statistiques de la distribution de la sortie du modèle. Ceci peut nécessiter un temps de calcul important. Malgré cet inconvénient, les méthodes d'échantillonnage sont les plus faciles à utiliser, vu qu'elles utilisent le modèle sans modification et elles le considèrent comme une boîte noire.

La discordance entre les différentes méthodes d'échantillonnage est la manière de choisir l'échantillon  $x^i = \{x_1^i, \dots, x_n^i\}$ . Par exemple, pour la méthode de Monte Carlo, la génération des échantillons est basée sur la distribution de probabilité de chaque paramètre, ce qui provoque une concentration des points autour de la région de forte probabilité et un manque d'information dans les zones à faible probabilité. Afin de balayer tout le domaine d'existence des entrées du modèle, Eglajs et Audze (1977) ont développé la méthode du Latin Hypercube Sampling qui divise le domaine de répartition des paramètres sur le nombre des échantillons  $N$ . Ensuite, l'échantillonnage est réalisé en tirant une valeur dans chaque sous-domaine. McKay et al. (1979) ont montré que méthode du Latin Hypercube Sampling donne une meilleure estimation des statistiques de la sortie que la méthode Monte Carlo. Avec la méthode Latin Hypercube Sampling, l'échantillonnage s'effectue d'une façon indépendante pour chaque paramètre. Iman et Conover (1982) et Iman & Davenport (1982) ont développé une procédure permettant d'imposer des corrélations sur les valeurs de l'échantillon.

- *Méthodes spectrales :*

Les méthodes d'échantillonnage représentent une caractérisation ponctuelle de la réponse, c'est-à-dire la réponse est calculée localement pour un ensemble de paramètres d'entrée. Ces méthodes sont considérées comme une forme discrétisée d'une fonction continue reliant la réponse aux paramètres d'entrée. Ainsi, afin d'identifier l'aspect statistiques de la réponse, il est nécessaire d'avoir une résolution suffisamment fine dans le domaine des paramètres, autrement, un grand nombre de simulation. Au contraire, les méthodes de propagation d'incertitude spectrales (surface de réponse) visent à réduire le temps de calcul en reconstruisant une dépendance fonctionnelle entre la réponse du modèle et les paramètres (Box et Draper 2007 ; Myers et al. 2016). Ces méthodes utilisent des modèles de substitution, appelés des méta-modèles qui remplacent le modèle de calcul d'origine par une fonction facile à évaluer (Hastie

et al. 2001 ; Forrester et al. 2008 ; Storlie et al., 2009). Ces fonctions qui représentent une décomposition de la sortie  $Y = f(x)$ , sont typiquement sous la forme suivante :

$$f(x) = \sum_{i \in \mathbb{N}} f_i \psi_i(x) \quad (3.5)$$

où,  $\psi_i(x)$  constituent une base de second ordre adéquat à l'espace des variables aléatoires  $x$  et  $f_i$  sont les coordonnées de  $f(x)$  sur cette base. Plusieurs techniques ont été développées pour la construction des méta-modèles comme l'expansion en polynômes de chaos (Le Maître et Knio, 2010 ; Xiu, 2010), la méta-modélisation par processus Gaussien (Santner et al., 2003 ; Rasmussen et Williams, 2006) et le réseau de neurones artificiels (Bishop, 1995 ; Ripley, 1996). Toutes ces techniques de substitution sont basées sur trois étapes principales (Simpson et al., 2001 ; Forrester et al., 2008). Premièrement, on construit un plan d'expériences qui représente un ensemble de jeux de paramètres aléatoires obtenu par une méthode d'échantillonnage. Puis, on évalue le modèle direct pour chaque jeu de paramètres du plan d'expérience. Ces données sont ensuite utilisées afin d'identifier les inconnus du méta-modèle choisi.

## 2.2.4 Analyse de sensibilité

### 2.2.4.1 Introduction

On peut définir l'analyse de sensibilité (sensitivity analysis SA) comme l'étude de la répartition de l'incertitude de la sortie  $Y$  d'un modèle mathématique sur différentes sources d'incertitude de ses variables d'entrée  $x = \{x_1, \dots, x_n\}$ . C'est l'étude de la dépendance de la sortie d'un modèle vis à vis de ces facteurs (Saltelli et al., 1999). Elle permet de savoir quel est le paramètre qui a le plus d'impact sur les résultats, celui qui en a le moins et ceux qui interagissent au sein du modèle.

Les objectifs de ce type d'analyse sont multiples (Saltelli et al., 2000b). En effet, elle permet d'analyser le rôle des paramètres dans le modèle, c'est-à-dire la contribution de chaque paramètre seul et les interactions entre eux. En plus, la SA sert à réduire la dimension du problème en identifiant et en négligeant les paramètres les moins influents dans le modèle.

Finalement, elle aide à diminuer l'incertitude et la dispersion de la sortie du modèle en minimisant la variabilité des paramètres les plus influents.

Dans la littérature, il existe un grand nombre de méthodes de SA (Helton et al. 2006 ; Xu et Gertner 2008). On distingue deux grandes classes, à savoir l'analyse de sensibilité locale et globale :

- L'analyse de sensibilité locale (LSA) est basée sur le calcul des dérivées partielles de la sortie du modèle par rapport aux paramètres, *c.à.d.* qu'elle évalue la variation de la réponse en fonction de petites variations des paramètres d'entrée au voisinage de leurs valeurs optimales. Plusieurs méthodes ont été développées afin de calculer ces dérivées d'une manière efficace, tel que la méthode des différences finies, la méthode de différentiation directe ou la méthode de différentiation adjointe (Cacuci 2003). Généralement, ces méthodes sont limitées dans des exemples pratiques car, elles ne tiennent pas compte des interactions entre les variables d'entrée.
- Contrairement à LSA qui examine la variation à petite échelle autour d'une valeur nominale, l'analyse de sensibilité globale (GSA) vise à quantifier l'incertitude de la sortie du modèle dépendant de toute la variation d'un paramètre d'entrée ou une combinaison de plusieurs paramètres. Saltelli et al., (2000a) ont classé les méthodes GSA en des méthodes basées sur la régression et sur la variance. Les méthodes basées sur la régression linéaire de la réponse du modèle par rapport aux paramètres d'entrée sont adaptées aux modèles linéaires ou quasi linéaires.

#### 2.2.4.2 Les méthodes d'A.S.G. basées sur la variance

Ces méthodes, connues sous le nom d'ANOVA (acronyme de ANalysis Of VAriance) (Efron et Stein 1981), décomposent la variance de la réponse du modèle en une somme des contributions de chaque paramètre d'entrée ou de leurs combinaisons. On peut définir l'indice de Sobol' comme la part de la variance de la sortie due à la variation d'un paramètre (Sobol' 1993, 2001). Ces indices peuvent être calculés par plusieurs méthodes telles que la méthode FAST (acronyme de Fourier amplitude sensitivity test) (Cukier et al. 1973 ; Cukier et al. 1978 ; Saltelli et al. 1999 ; Mara 2009), la simulation de Monte Carlo (Saltelli et Sobol' 1995 ; Archer et al. 1997 ; Sobol' et Kucherenko 2005, 2009) ou les méthodes utilisant un méta-modèle telles



que le processus Gaussien (Krige 1951 ; Matheron 1962) et le polynôme de chaos (Wiener 1938 ; Cameron et Martin 1947).

#### 2.2.4.2.1 Décomposition de Sobol'

Soit un modèle mathématique défini par une fonction  $f(x)$  avec la sortie scalaire  $Y$  et les entrées scalaires indépendantes  $x = \{x_1, \dots, x_n\}$  définies sur l'hypercube unité  $\Omega_n = [0, 1]^n$ . Cette fonction qui est de carré sommable ( $f(x) \in L^2$ ), peut se décomposer en une somme de fonctions de dimension croissante (Sobol' 1993):

$$f(x) = f_0 + \sum_{i=1}^n f_i(x_i) + \sum_{1 < i < j < n} f_{ij}(x_i, x_j) + \dots + f_{1..n}(x_1, \dots, x_n) \quad (3.6)$$

Où  $f_0$  est une constante et l'intégrale de chaque terme est nulle ( $\int_0^1 f_{i_1, \dots, i_s}(x_{i_1}, \dots, x_{i_s}) dx_{i_k} = 0$  pour  $k \in [1, s]$ ). D'après cette définition Homma et Saltelli (1996) ont déduit que le nombre des termes de cette décomposition est égale à  $2^n - 1$ . La constante  $f_0$  est la valeur moyenne de la fonction  $f(x)$  ( $f_0 = E(f(x)) = \int_{\Omega^n} f(x) dx$ ). Les termes de la décomposition sont orthogonaux les uns aux autres  $\int_{\Omega_n} f_{i_1, \dots, i_s}(x_{i_1}, \dots, x_{i_s}) f_{j_1, \dots, j_r}(x_{j_1}, \dots, x_{j_r}) dx = 0$  pour  $\{i_1, \dots, i_s\} \neq \{j_1, \dots, j_r\}$ . La variance totale de la réponse du modèle  $V$  s'écrit sous la forme :

$$V = Var(f(x)) = E\left(E(f^2(x)) - E(f(x))^2\right) = \int_{\Omega^n} f^2(x) dx - f_0^2 \quad (3.7)$$

Par conséquent, la variance de la réponse du modèle peut être décomposée sous cette forme :

$$V = \sum_{i=1}^n V_i + \sum_{1 < i < j < n} V_{ij} + \dots + V_{1..n} \quad (3.8)$$

Où  $V_i$  est la variance partielle due à la variation du paramètre  $x_i$  seul et  $V_{i_1, \dots, i_s} = \int_{\Omega^n} f_{i_1, \dots, i_s}^2(x_{i_1}, \dots, x_{i_s}) dx_{i_1}, \dots, dx_{i_s}$  est la variance partielle due aux interactions entre les paramètres  $\{x_{i_1}, \dots, x_{i_s}\}$ . Afin de donner plus de sens à ces variances partielles, on les divise par la variance totale et on obtient alors les variances partielles ou indices de sensibilité, appelés les indices de Sobol' :

$$S_{i_1, \dots, i_s} = \frac{V_{i_1, \dots, i_s}}{V} \quad (3.9)$$

Les valeurs de ces indices appartiennent à l'intervalle  $[0, 1]$  et la somme de tous les indices de sensibilité de la réponse du modèle est égale à 1.

$$1 = \sum_{i=1}^n S_i + \sum_{1 < i < j < n} S_{ij} + \dots + S_{1 \dots n} \quad (3.10)$$

La valeur de l'indice de sensibilité de premier ordre  $S_i = \frac{V_i}{V} = \frac{V(E(f(x)/x_i))}{V(f(x))}$  qui représente l'influence de chaque paramètre  $x_i$  considéré seul, peut être nulle. Cela ne signifie pas forcément que  $x_i$  n'a pas d'influence sur la sortie car il peut participer à la variance en interagissant avec les autres paramètres. Pour évaluer l'effet total d'un paramètre d'entrée  $x_i$ , on définit l'indice de sensibilité total comme la somme de toutes les variances partielles relatives impliquant l'indice  $i$  (Homme et Saltelli, 1996) :

$$S_{T_i} = \sum_{\{i_1, \dots, i_s\} \ni i} \frac{V_{i_1, \dots, i_s}}{V} = \frac{E(E(f(x)/\sim x_i))}{V(f(x))} \quad (3.11)$$

Dans la suite, on va rappeler quelques méthodes d'estimation de ces indices de sensibilité.

## 2.2.4.2.2 Méthodes d'échantillonnage

Parmi les méthodes d'échantillonnage la méthode de Monte Carlo est la plus utilisée, elle consiste à estimer les différentes intégrales dans le calcul de la moyenne et les variances par la méthode de Monte Carlo en utilisant un plan d'expérience de  $N$  échantillons. La valeur moyenne et la variance totale sont estimées par les deux sommes suivantes :

$$\hat{f}_0 = \frac{1}{N} \sum_{k=1}^N f(x^{(k)}) \quad (3.12)$$

$$\hat{V} = \frac{1}{N} \sum_{k=1}^N f^2(x^{(k)}) - \hat{f}_0^2 \quad (3.13)$$

Afin d'estimer les variances partielles, on génère deux plans différents d'expérience en mixant leurs composantes.

$$\hat{V}_i = \frac{1}{N} \sum_{k=1}^N f(x_{\sim i}^{(k),1}, x_i^{(k),1}) f(x_{\sim i}^{(k),2}, x_i^{(k),1}) - \hat{f}_0^2 \quad (3.14)$$

$$\hat{V}_{\sim i} = \frac{1}{N} \sum_{k=1}^N f(x_{\sim i}^{(k),1}, x_i^{(k),1}) f(x_{\sim i}^{(k),1}, x_i^{(k),2}) - \hat{f}_0^2 \quad (3.15)$$

Avec  $x_{\sim i}$  présente le vecteur des paramètres sauf le paramètre  $x_i$ . L'indice de Sobol' de premier ordre  $\hat{S}_i = \hat{V}_i / \hat{V}$  et l'indice de sensibilité total  $\hat{S}_{T_i} = 1 - \hat{V}_{\sim i} / \hat{V}$  calculés par cette méthode contiennent une incertitude importante qui peut être réduite en augmentant le nombre d'échantillons  $N$  et donc le coût en temps de calcul. Notons que pour décrire d'une façon complète la sensibilité du modèle, il est nécessaire d'évaluer  $2^n$  fois le modèle direct, ce qui ne peut être faisable que si le nombre de paramètres est réduit. L'échantillonnage quasi-Monte Carlo (Sobol' 1998 ; Bratley et al. 1992 ; Joe et Kuo 2008) a été proposé pour augmenter la vitesse de convergence et donc réduire le temps de calcul. Notons que souvent, on se limite à estimer les indices de sensibilité de premier ordre et les indices de sensibilité totale, ce qui est équivalent à estimer les indices de sensibilité obtenu par la méthode FAST.

### 2.2.4.3 Méthodes basées sur la méta-modélisation

Les méthodes standards de propagation de l'incertitude et d'analyse de sensibilité reposent sur un grand nombre d'appels au modèle direct. Cela peut conduire à des calculs intraitables si la fonction de modèle  $M$  est coûteuse à évaluer. Pour contourner ce problème, il est possible d'approcher  $M$  par une fonction analytique dont l'évaluation est peu coûteuse et cette approximation est souvent appelée méta-modèle, modèle de substitution ou surface de réponse. Une fois le méta-modèle construit, les simulations intensives peuvent être appliquées à un coût de calcul négligeable.

Diverses techniques de méta-modélisation ont été proposées dans la littérature, telles que les surfaces de réponse polynomiales, la régression multivariée par spline adaptative (MARS) (Vapnik et al., 1997 ; Smola et Schölkopf, 2004), les arbres de régression, le réseau neural artificiel (Chen et al., 2005 ; Chen et al., 2006), la modélisation de processus gaussiens (Sacks et al., 1989 ; Santner et al., 2003) et les polynômes de chaos. Dans la suite, nous nous intéressons à ces deux derniers types d'émulateur qui sont généralement les moins coûteux à construire (Oakley et O'Hagan, 2004 ; sudret, 2008).

Le processus Gaussien est parmi les méthodes méta-modélisation les plus utilisées, il est connu aussi par le nom de Krigeage (Matheron, 1967), est une méthode bien établie en géostatistique afin de prédire la valeur d'un champ aléatoire du modèle physique  $M$  en un point donné  $x$  à partir d'un ensemble de données connus au voisinage de ce point (Sacks et al. 1989 ; Welch et al. 1992 ; Koehler et Owen 1996). L'idée générale du processus Gaussien est proche des méthodes de régression traditionnelles tel que la régression linéaire. En effet, les hypothèses statistiques peuvent être considérées comme un moyen de représenter l'intervalle d'incertitude associé à la prédiction (Santner et al. 2003; Vazquez et Walter 2003). Cette méthode a été largement utilisée dans la littérature pour construire des surfaces de réponse car elle est capable de modéliser précisément des comportements non linéaires (O'Hagan 2006 ; Vazquez 2005).

- *Représentation de la fonction de modèle*

Considérons le modèle  $M$  déterministe d'un système physique dont la réponse  $f(x)$  soit uniquement connue en  $N$  points. La vraie réponse du modèle  $M$  est supposée être un exemple de chemin d'un processus gaussien sous-jacent  $Y(x)$  qui est défini comme :

$$Y(x) = Z(x) + \mu(x) \quad (3.16)$$

Où,  $Z(x)$  est un champ aléatoire centré (moyenne nulle) qui correspond à des fluctuations à petites échelles de la réponse du modèle. Ce terme est considéré comme un champ Gaussien stochastique caractérisé par sa variance  $\sigma^2$  et sa matrice d'autocorrélation  $\rho(x, x')$ . L'autre terme de l'équation précédente  $\mu(x)$  est la moyenne du champ aléatoire  $Y(x)$  qui représente une tendance à grande échelle de la sortie du modèle.  $\mu(x)$  est généralement exprimé comme une combinaison linéaire de fonctions déterministes connues :

$$\mu(x) = \sum_{j=1}^p \beta_j \phi_j(x) \quad (3.17)$$

Avec  $\phi_j(x)$  un vecteur de fonction de régression et  $\beta_j$  les coefficients de régression. En pratique, la séparation des deux termes est une tâche compliquée vu que le choix du terme de régression est souvent arbitraire. Afin de simplifier un peu le problème, Sacks et al., 1989 ont limité le choix de la fonction moyenne en une fonction constante ou linéaire.

- *Distribution conditionnelle de la réponse du modèle*

Considérons un échantillon  $x = (x^{(1)}, \dots, x^{(N)})$  de dimension  $N$  du vecteur aléatoire d'entrée  $x$ , également appelé plan d'expériences. Les évaluations correspondantes du modèle direct, sont stockées dans un échantillon unidimensionnel  $y = (y^{(1)}, \dots, y^{(N)})$ . L'approximation du processus gaussien consiste à construire un métamodèle à partir de la distribution conditionnelle de  $y$ . Par conséquent, il peut être reformulé en tant que problème d'optimisation sous contrainte. En utilisant le théorème fondamental de la prédiction, Santner et al., 2003 ont montré

que la réponse conditionnelle du modèle en un point non évalué  $x$  suit la distribution gaussienne  $N(\mu(x), \sigma^2 R)$ , où  $R_{ij} = \rho(x^{(i)}, x^{(j)})$ .

- *Estimation des paramètres de krigeage*

Afin de trouver le jeu de paramètres optimal de krigeage, on maximise la fonction de vraisemblance qui est définie sous cette forme :

$$p(y|x, \sigma, \beta, R) \propto \frac{1}{\sigma^N \sqrt{\det(R)}} = \exp\left(-\frac{1}{2\sigma^2} (y - \mu(x))^T R^{-1} (y - \mu(x))\right) \quad (3.18)$$

En supposant que  $R$  est donné, on peut déterminer le vecteur des paramètres d'autocorrélation  $\theta$ . Donc, on est capable de résumer les paramètres optimaux du processus Gaussien ci-dessous :

$$\hat{\beta} = (\phi^T(x) R^{-1} \phi(x))^{-1} (\phi^T(x) R^{-1} y) \quad , \quad \hat{\mu}(x) = \sum_{j=1}^p \hat{\beta}_j \phi_j(x) \quad (3.19)$$

$$\hat{\sigma}^2 = (y - \hat{\mu}(x))^T R^{-1} (y - \hat{\mu}(x)) \quad (3.20)$$

$$\hat{\theta} = \arg \min (\hat{\sigma}^2 \det(R)^{1/N}) \quad (3.21)$$

Le choix de la dimension  $N$  du plan d'expériences est un facteur important de l'estimation des paramètres de krigeage, car une grande dimension induit un coût de calcul important. Par conséquent, plusieurs tentatives de réduction de temps de calcul sont apparues tel que O'Hagan, 2006 qui a estimé tous les paramètres en une seule procédure d'optimisation et Welch et al. (1992) et plus récemment de Marrel et al. (2008) qui ont basé leurs travaux sur la sélection des variables afin de choisir des fonctions d'autocorrélation anisotropes pour des plans d'expériences de grandes dimensions.

- *L'incertitude associée au processus Gaussien*

Les indices de sensibilité basés sur la variance pour un nouveau vecteur d'entrée  $x^*$  (Caniou et al. 2011) :

$$s^2(x^*) = \sigma^2 \left( 1 - r^T R^{-1} r + \frac{1 - r^T R^{-1} r}{1^T R^{-1} 1} \right) \quad (3.22)$$

Pour une prédiction du processus Gaussien  $y^* = \mu(x^*) + r^T R^{-1} (y - \mu(x^*))$ , où

$$r_i = \exp \left( - \sum_{j=1}^n (x_j^{(i)} - x_j^*)^{p_j} \right).$$

Dans le cadre stochastique, l'importance de déterminer l'incertitude à l'aide de krigeage est de fournir des régions de l'espace des paramètres, où l'incertitude est faible, appelées des intervalles de confiance.

Pour conclure, le processus gaussien est une des techniques de méta-modélisation les plus utilisées dans la littérature (Kennedy et O'Hagan, 2001 ; Bayarri et al., 2007 ; Higdon et al., 2008). L'idée générale de cette technique est de considérer la réponse du modèle comme un exemple de chemin d'un champ aléatoire gaussien dont les paramètres sont estimés par maximum de vraisemblance. Ce métamodèle est dérivable car il interpole les évaluations de modèle observées. De plus, vu l'absence d'erreur aléatoire en considérant une fonction de modèle déterministe, le krigeage est une méthode attrayante. Notons cependant que le krigeage est laborieux à utiliser dans le cas où le modèle est fortement non linéaire. Une autre limitation est la connaissance des paramètres qui caractérisent le domaine aléatoire sous-jacent. En effet, ces paramètres sont estimés par maximum de vraisemblance, ce qui peut s'avérer coûteux en calcul si le nombre de paramètres inconnus ou le nombre de variables d'entrée est important.

## 2.2.5 Expansion en polynômes de chaos

### 2.2.5.1 Introduction du PCE :

L'expansion en polynôme de chaos (*polynomial chaos expansion* : PCE) est une méthode de méta-modélisation, basée sur la théorie du chaos Homogène introduite par Wiener (1938) et sur les résultats des travaux de Cameron et Martin (1947) qui consistent à développer des fonctions

de carré sommable suivant une base de polynômes deux à deux orthogonaux. Une telle approche a été appliquée au début des années 90 aux problèmes de mécanique des structures comportant des paramètres spatialement aléatoires (Ghanem et Spanos, 1991). Au cours de ces deux dernières décennies, le PCE a reçu beaucoup d'attention tel que Xui et Karniadakis (2002), Fajraoui et al. (2011) et Riva et al. (2015).

Revenons à la définition de la décomposition de Sobol et considérons un modèle mathématique défini par une fonction  $f(x)$  qui représente la sortie scalaire  $Y$  où les entrées scalaires indépendants  $x = \{x_1, \dots, x_n\}$  sont définies sur l'hypercube unité  $\Omega^n$ . Selon la théorie du chaos homogène, une fonction de carré sommable ( $f(x) \in L^2$ ) peut se décomposer sous la forme suivante :

$$f(x) = \sum_{i=0}^{\infty} a_i \psi_i(x) \quad (3.23)$$

Où,  $a_i$  et  $\psi_i(x)$  sont respectivement le coefficient du polynôme de chaos et le polynôme multivarié de degrés  $i$ .

Initialement, Wiener (1938) a conçu le développement en polynômes de chaos qui est également appelé développement de Wiener-Hermite pour modéliser les processus stochastiques à partir de variables aléatoires gaussiennes standards sur une base de polynômes d'Hermite multivariés. En effet, Cameron et Martin (1947) ont constaté que la convergence de ce développement est optimale lorsque les variables suivent une loi normale.

Afin d'élargir le champ d'application de cette méthode et de répondre aux attentes des praticiens utilisant non seulement des variables gaussiennes, Xiu et al., (2003) ont montré que la vitesse de la convergence du développement dépend de la densité de probabilité des variables et selon le schéma de Askey (Askey et Wilson, 1985), ils ont trouvé une correspondance entre les principales lois de probabilité et leurs familles de polynômes orthogonaux. Le Tableau 2-1 donne une partie de cette correspondance. Dans ce cas, la décomposition est appelée expansion en polynômes de chaos généralisée (GPCE). Pour des distributions de probabilité non usuelles, le problème peut être reformulé en utilisant une transformation iso-probabiliste dans laquelle GPCE est appliqué aux variables transformées (Wan and Karniadakis, 2006 ; Witteveen et al., 2007).



Tableau 2-1 Correspondance entre les distributions et les familles de polynômes orthogonaux

Distribution	Support	Polynômes
Gaussienne	$\mathbb{R}$	Hermite
Uniforme	$[a, b]$	Legendre
Gamma	$\mathbb{R}_+$	Laguerre
Beta	$[a, b]$	Jacobi
Poisson	$\mathbb{N}$	Charlier
Binomiale	$\{0, 1, \dots, n\}$	Kravtchouk
Binomiale négative	$\mathbb{N}$	Meixner
Hypergéométrique	$\{0, 1, \dots, n\}$	Hahn

En pratique, le développement en polynômes de chaos est généralement tronqué jusqu'à un degré polynomial maximal  $p$ . Par conséquent, le nombre total de termes  $P$  peut être calculé à partir de ce dernier  $p$  et du nombre  $n$  de variable d'entrée comme suit :

$$P = \binom{n+p}{p} = \frac{(n+p)!}{n!p!} \tag{3.24}$$

Ce développement est similaire à la décomposition de Sobol de point de vu orthogonalité des termes de la décomposition. De plus, les indices de sensibilité sont directement déduits des coefficients du polynôme de chaos  $a_i$ , par conséquent, la décomposition devient sous cette forme :

$$f(x) = \sum_{i=0}^p a_i \psi_i(x) \tag{3.25}$$

### 2.2.5.2 Estimation des coefficients polynomiaux :

Les approches qui estiment les coefficients du PCE sont généralement classées en deux catégories ; *approches intrusives* (Le Maitre et al. 2002, Matthies et Keese, 2005) où des extensions tronquées sont insérées dans le modèle numérique et les coefficients du PCE sont directement déduits en une seule exécution du modèle, ou *non intrusives* (Tatang et al. 1997, Berveiller, et al. 2006b et Blatman & Sudret 2008, 2010b, 2011).

Plusieurs méthodes sont développées pour obtenir ces coefficients, mais, la plus efficace est la méthode des éléments finis stochastiques spectraux (SSFEM) qui peut être considérée comme

une généralisation de la méthode de Galerkin. Cette méthode novatrice a été développée par Xiu & Karniadakis (2002), Sudret et al. (2004), Wan & Karniadakis (2005, 2006).

L'inconvénient des techniques intrusives est la complexité de leur mise en œuvre parce qu'elles nécessitent des développements algorithmiques spécifiques qui dépendent de la nature du problème. En outre, elles sont gourmandes de point de vue du temps de calcul dans les cas des modèles non linéaires (Berveiller, 2005).

Alternativement, les méthodes dites non intrusives considèrent le modèle numérique comme une boîte noire. La détermination des coefficients du PCE est basée sur de simples évaluations du modèle à des points bien choisis de l'espace des paramètres sans aucune modification du code de calcul. Le défi en utilisant ce type des méthodes est la détermination de la technique adéquate qui limite le nombre d'appel du modèle direct sans perdre la précision. La simplicité d'utilisation de ces approches non intrusives nous a encouragé à les adopter au cours de ces travaux de thèse. On va citer deux méthodes parmi les approches non intrusives par la suite.

#### 2.2.5.2.1 La méthode de la projection spectrale :

La méthode de projection spectrale vise à estimer les coefficients  $a_i$  du polynôme de chaos en exploitant l'orthogonalité de la base tronquée  $\{\psi_j, j \leq p\}$ . En effet, la projection de l'équation (2.58) sur cette base donne :

$$E[f\psi_j] = E\left[\sum_{i=0}^p a_i \psi_i \psi_j\right] = \sum_{i=0}^p a_i E[\psi_i \psi_j] \quad (3.26)$$

Avec,  $E[\cdot] = \int_{\Omega^p} (\cdot) d\phi(x)$ .

On peut déduire l'expression du coefficient de polynôme de chaos  $a_i$  :

$$a_i = \frac{E[f(x)\psi_j(x)]}{E[\psi_j^2(x)]} \quad (3.27)$$

Où, le dénominateur dans cette expression est déterminé analytiquement tandis que le numérateur nécessite un calcul d'intégrale :

$$E[f\psi_j] = \int_{\Omega^n} f(x)\psi_j(x)d\phi(x) \quad (3.28)$$

En pratique, ce calcul est estimé à l'aide des techniques d'échantillonnage dont la plus simple méthode est l'estimation de Monte Carlo en considérant  $N$  échantillons de dimension  $n$ . Afin d'améliorer la précision de l'estimation, on augmente le nombre des points  $N$  ou on utilise d'autres méthodes telles que l'hypercube latin (LHS) ou le quasi-Monte Carlo (QMC) qui proposent des stratégies d'échantillonnage plus sélectives. Généralement, la nature multivariée des entrées implique l'emploi d'un schéma en quadrature multidimensionnelle qui sélectionne un ensemble de points spécifiques  $x^{(k)}$  et des poids associés  $\omega^{(k)}$ , de sorte que l'intégrale dans Eq. (2.59) s'écrit de la manière suivante :

$$E[f\psi_j] = \frac{1}{N} \sum_{k=1}^N \omega^{(k)} f(x^{(k)}) \psi_j(x^{(k)}) \quad (3.29)$$

L'inconvénient principal des méthodes de projection est que les coefficients doivent être évalués un à un.

#### 2.2.5.2.2 La méthode de régression :

Contrairement à l'approche précédente, la méthode de régression permet de calculer tous les coefficients du polynôme de chaos simultanément (Choi et al., 2004, Berveiller, 2005 et Berveiller et al., 2006) sous la forme d'un vecteur  $a = \{a_0, \dots, a_p\}^T$  lié au vecteur des polynômes multivariés  $\psi(x) = \{\psi_0(x), \dots, \psi_p(x)\}^T$ . Le problème de régression peut être limité à trouver la valeur  $\hat{a}$  qui minimise la fonction ci-dessous :

$$\eta(a) = E\left[\left(a^T \psi(x) - f(x)\right)^2\right] \quad (3.30)$$

En considérant la condition de minimalité  $d\eta(\hat{a})/da = 0$ , on obtient l'équation suivante :

$$E[\psi(x)\psi^T(x)]\hat{a} = E[\psi(x)f(x)] \quad (3.31)$$

Puisque  $\psi(x)$  est un vecteur aléatoire dont les composantes sont décorrélées par définition donc sa matrice de corrélation  $E[\psi(x)\psi^T(x)]$  tend vers la matrice identité. Enfin, le vecteur des coefficients du PCE s'écrit :

$$\hat{a} = E[\psi(x)f(x)] \quad (3.32)$$

On peut remarquer que ce vecteur est équivalent formellement aux coefficients basés sur la projection, donc on peut dire théoriquement que les coefficients de projection minimisent aussi l'erreur quadratique moyenne.

On considère toujours les échantillons  $y = (y^{(1)}, \dots, y^{(N)})$  l'image de l'ensemble des vecteurs aléatoires d'entrées  $x = (x^{(1)}, \dots, x^{(N)})$  de dimension  $N$  par la fonction  $f(x)$  qui définit le modèle mathématique. En outre, on définit la matrice  $\psi$  par  $\psi = \{\psi_{ij} = \psi_j(x^{(i)}) | i = 1, \dots, N; j = 1, \dots, P\}$ . Par conséquent, on peut représenter le vecteur des coefficients du PCE basé sur la régression vectorielle sous la forme suivante :

$$\hat{a} = (\psi^T \psi)^{-1} \psi^T y \quad (3.33)$$

La matrice  $\psi^T \psi$  est appelée matrice d'information de Fisher et la résolution de l'équation précédente est habituellement effectuée à l'aide des méthodes numériques telle que la décomposition en valeurs singulières (SVD) (Bjorck, 1996).

Il est important de bien choisir le plan d'expérience en utilisant une telle méthode afin que la matrice d'information soit de rang plein ; par exemple, il faut que la taille  $N$  de l'échantillon soit supérieure au nombre de coefficients  $P$ . En ce qui concerne les techniques d'échantillonnage, les techniques quasi-aléatoires telles que MC, LHS, QMC ou les racines de Gauss peuvent être utilisées (Isukapalli, 1999).

### 2.2.5.3 Précision du PCE :

Les expansions en polynôme de chaos représentent un moyen efficace de substitution des modèles numériquement coûteux par des modèles analytiques moins coûteux. Toutefois, il est important de quantifier l'erreur d'approximation afin de construire un intervalle de confiance. Cette erreur est l'erreur quadratique moyenne qui est calculée par la sommation des carrés des écarts entre la réponse du modèle numérique et la valeur donnée par le PCE; on l'appelle généralement l'erreur empirique et elle est définie comme suit :

$$Err_{emp} = \frac{1}{N} \sum_{i=1}^N (f(x^{(i)}) - f^{PC}(x^{(i)}))^2 \quad (3.34)$$

Pour évaluer la précision de l'approximation, on se réfère généralement au coefficient de détermination  $R^2$  :

$$R^2 = 1 - \frac{Err_{emp}}{V(y)} \quad (3.35)$$

Où,  $V(y)$  désigne la variance empirique de l'échantillon de réponse  $y$  :

$$V(y) = \frac{1}{N-1} \sum_{i=1}^N (y^{(i)} - E[y])^2, \quad E[y] = \frac{1}{N} \sum_{i=1}^N y^{(i)} \quad (3.36)$$

Un coefficient de détermination  $R^2$  proche de 1 indique généralement une bonne précision du méta-modèle, alors qu'un  $R^2$  proche de zéro caractérise une mauvaise représentation de la réponse du modèle. Cependant, dans le cas où  $N$  tend vers  $P$ ,  $R^2$  converge systématiquement vers 1. Effectivement,  $R^2 = 1$ , si  $N = P$  vu qu'on interpole des exécutions du modèle, c'est ce qu'on appelle le phénomène de sur-apprentissage et parmi ses cas extrêmes, on considère l'effet Runge (Runge 1901).

Diverses méthodes fondées sur les pénalités ont été proposées dans la littérature afin de réduire la sensibilité des estimations d'erreur au sur-apprentissage et parmi ces alternatives, on utilise le coefficient de détermination ajusté  $R^{2*}$  qui est déduit de l'erreur empirique ajusté :

$$Err_{emp}^* = Err_{emp} \frac{N-1}{N-P-1} \quad (3.37)$$

De cette façon, le coefficient de détermination ajusté  $R^{2*}$  est pénalisé lorsque le nombre des termes inclus dans le métamodèle augmente.

Le problème principal de la surestimation du  $R^2$  est que sa détermination est basée sur le même plan d'expérience que celui utilisé pour construire le PCE. Par conséquent, on pourrait être tenté de calculer l'erreur à partir d'un autre ensemble de points différents, appelé ensemble de test ou ensemble de validation. Donc, on divise l'échantillon de données en deux sous-échantillons : échantillon d'apprentissage et échantillon de validation, cette technique est dite validation croisée (Stone, 1974 ; Geisser, 1975). Il existe un cas particulier de cette dernière appelé « leave-one-out » qui consiste à calculer l'erreur quadratique moyenne de la manière suivante :

$$Err_{LOO} = \frac{1}{N} \sum_{i=1}^N \left( f(x^{(i)}) - f_{x^{(i)}}^{PC}(x^{(i)}) \right)^2 \quad (3.38)$$

Sachant que,  $\hat{f}_{x|x^{(i)}}^{PC}(x^{(i)})$  désigne le métamodèle construit à partir du plan d'expérience  $x|x^{(i)}$  c'est-à-dire en supprimant la nième observation et par analogie au coefficient de détermination  $R^2$ , on peut définir le coefficient  $Q^2$  par :

$$Q^2 = 1 - \frac{Err_{LOO}}{V(y)} \quad (3.39)$$

#### 2.2.5.4 Expansion en polynômes de chaos creux :

L'approximation de la réponse du modèle  $Y$  en fonction de l'expansion en PC s'écrit :

$$Y \simeq f_p(x) = \sum_{\alpha \in \mathcal{A}^{n,p}} a_\alpha \psi_\alpha(x) \quad , \quad \mathcal{A}^{n,p} \equiv \{\alpha \in \mathbb{N}^n : |\alpha| \leq p\} \quad (3.40)$$

Avec,  $\mathcal{A}^{n,p}$  un ensemble fini et non-vide des multi-indices  $\alpha = \{\alpha_1, \dots, \alpha_n\} \in \mathbb{N}^n$ , il définit le schéma de troncature qui considère tous les polynômes jusqu'à un degré total  $|\alpha| \equiv \|\alpha\|_1 = \sum_{i=1}^n \alpha_i$  et il est limité par un degré maximal  $p$ .

Notons que, selon l'équation (2.57) le nombre des coefficients du polynôme qui est égal à  $card(\mathcal{A}^{n,p})$ , croît d'une manière polynomiale en fonction du nombre  $n$  des variables et du degré maximal  $p$  à la fois. Dans ce cas, un grand nombre d'évaluations de modèles est nécessaire pour calculer l'ensemble des coefficients du PCE, c'est ce qu'on dénomme le fléau de la dimension. Plusieurs stratégies ont été proposées afin de réduire le nombre des coefficients dont certaines sont rappelées ici.

##### 2.2.5.4.1 Ensembles des indices aux rangs faibles :

Cette méthode est inspirée par le principe de hiérarchie des effets (the sparsity-of-effect principle Montgomery, 2004, Yuan et al., 2007) qui stipule que la plupart des modèles sont principalement régis par les effets principaux et les interactions d'ordre faible. D'ailleurs, la stratégie des indices aux rangs faibles consiste à retenir seulement les multi-indices  $\alpha$  dont le rang est inférieur à un rang maximal  $j \leq p$  (An and Owen, 2001 ; Todor and Schwab, 2007). Dans ce cas l'ensemble des indices est défini comme :

$$\mathcal{A}_j^{n,p} \equiv \{\alpha \in \mathbb{N}^n : |\alpha| \leq p, \|\alpha\|_0 \leq j\} \quad (3.41)$$

Avec,  $\|\alpha\|_0 = \sum_{i=1}^n 1_{\{\alpha_i > 0\}}$  le rang du multi-indice  $\alpha$  qui désigne les termes non nuls, autrement dit, il indique le nombre des variables actives du  $\psi_\alpha$ . De cette façon, on peut limiter l'ordre d'interaction des modèles dont le nombre des variables  $n$  et du degré maximal  $p$  sont élevés et par conséquent, la cardinalité de l'ensemble  $\mathcal{A}^{n,p,j}$  est moins élevée que celle de  $\mathcal{A}^{n,p}$ .

#### 2.2.5.4.2 Ensembles des indices hyperboliques :

Blatman et Sudret (2009 ; 2010a) ont proposé une stratégie qui permet de sélectionner les coefficients du polynôme ayant une contribution révélatrice sur le modèle, donc, on définit l'ensemble des indices hyperboliques basés sur la norme de  $q$  :

$$\mathcal{A}_q^{n,p} \equiv \{\alpha \in \mathbb{N}^n : \|\alpha\|_q \leq p\} \quad , \quad \|\alpha\|_q = \left( \sum_{i=1}^n \alpha_i^q \right)^{1/q} \quad (3.42)$$

Cette norme pénalise les interactions d'ordre élevé d'avantage lorsque  $q$  est faible. Notons que pour  $q=1$ , l'ensemble  $\mathcal{A}^{n,p,q}$  est équivalent à  $\mathcal{A}^{n,p}$ , et lorsque  $q < 1$  les polynômes de base conservés sont situés sous une hyperbole, d'où l'appellation d'ensembles des indices hyperboliques.

Dans la configuration simple de cette stratégie, les ensembles des indices hyperboliques sont isotropes, c'est-à-dire que toutes les variables d'entrée ont la même importance. Dans le but de pénaliser plus l'interaction d'ordre élevé sur les variables qui ont une contribution faible à la réponse du modèle, Blatman et Sudret 2010a ont proposé un schéma de troncature basé sur un ensembles des indices hyperboliques anisotrope :

$$\mathcal{A}_{q,\omega}^{n,p} \equiv \{\alpha \in \mathbb{N}^n : \|\alpha\|_{q,\omega} \leq p\} \quad , \quad \|\alpha\|_{q,\omega} = \left( \sum_{i=1}^n |\omega_i \alpha_i|^q \right)^{1/q} \quad (3.43)$$

On définit alors l'ensemble des poids  $\omega_i$  qui sont inversement proportionnels aux indices de sensibilités totaux  $S_i^T$  :

$$\omega_i = 1 + K \frac{S_{max}^T - S_i^T}{\sum_{i=1}^n S_i^T} \quad (3.44)$$

Avec,  $S_{max}^T = \max \{S_i^T, i = 1, \dots, n\}$  l'indice de sensibilité total le plus élevé et le coefficient  $K$  désigne le taux d'anisotropie, c'est-à-dire lorsque  $K = 0$ ,  $\mathcal{A}_{q,\omega}^{n,p}$  sera isotrope, alors qu'une valeur élevée de  $K$  donne un ensemble  $\mathcal{A}_{q,\omega}^{n,p}$  plus anisotrope.

Les ensembles hyperboliques et surtout ceux qui sont anisotropes conduisent à une réduction notable du nombre de termes du PCE, néanmoins, il est clair que les indices de sensibilité totaux doivent être connus a priori pour définir l'ensemble de pondérations approprié.

#### 2.2.5.4.3 Algorithmes adaptatifs pour les PCE creux :

- *En utilisant la régression pas à pas (stepwise) :*

En statistique, la régression pas à pas (appelée par étapes) est une méthode d'ajustement des modèles de régression dans laquelle le choix des variables prédictives est effectué automatiquement (Efroymsou, 1960). La construction de la base du PC est effectuée suivant les 5 étapes suivantes (Blatman et Sudret, 2008a, 2010a, 2010b):

##### Étape 1 : Définition

- On choisit un plan d'expérience  $x$ , de taille initiale  $N_{l=0}$ , issu d'un échantillonnage quasi-aléatoire. Les évaluations du modèle  $f(x)$  sont regroupées dans le vecteur  $y$ .
- On définit les paramètres de l'algorithme, c'est-à-dire le degré maximal du PC  $p_{max}$ , les paramètres  $K$  et  $q$  de la norme hyperbolique, le coefficient de précision cible  $Q_{Cible}$  et les valeurs de seuil  $\varepsilon_1$  et  $\varepsilon_2$ .

##### Étape 2 : Initialisation



- Cet algorithme est initialisé par un degré du PC  $p = 0$ , un ensemble des poids  $\omega_i = 1$  ( $i = 1, \dots, n$ ) ensuite, l'ensemble des multi-indices  $\mathcal{A}_{q,1}^{n,0} = \{\mathbf{0}\}$  (avec  $\mathbf{0}$  est l'élément nul de  $\mathbb{N}^n$ ).

### Étape 3 : Enrichissement de la base du PC

On répète pour chaque  $p \in \{1, \dots, p_{max}\}$  :

- *Étape d'addition (Forward step)* : Soit  $C_{q,\omega}^{n,p} \equiv \{\alpha \in \mathbb{N}^n : p-1 \leq \|\alpha\|_{q,\omega} \leq p\}$  l'ensemble des termes candidats, on ajoute chaque terme candidat à l'ensemble  $\mathcal{A}_{q,\omega}^{n,p-1}$  un par un et on calcule les coefficients du PCE par régression (équation 2.66) et le coefficient de détermination  $R^2$  associé. On ne conserve éventuellement que les termes candidats qui entraînent une augmentation significative du coefficient  $R^2$  ( $\Delta R^2 > \varepsilon_1$ ), c'est-à-dire qui diminuent considérablement l'erreur empirique  $Err_{emp}$ .

- *Étape d'élimination (Backward step)* : A tour de rôle, on élimine, provisoirement, de notre ensemble chaque multi-indice dont la norme  $\|\alpha\|_{q,\omega}$  est strictement inférieur à  $p$  et on estime, à chaque fois, le PCE et le coefficient de détermination  $R^2$ . Si on constate une diminution non considérable de  $R^2$  ( $\Delta R^2 < \varepsilon_2$ ), c'est-à-dire une augmentation négligeable de l'erreur empirique  $Err_{emp}$ , on supprime ce terme d'une manière permanente. Finalement, On obtient l'ensemble de troncature  $\mathcal{A}_{q,\omega}^{n,p}$ .

- Après avoir déterminé les coefficients de PCE basés sur  $\mathcal{A}_{q,\omega}^{n,p}$ , on calcule les indices de sensibilité totaux et on actualise les poids  $\omega_{i=1,\dots,n}$  (équation 2.77).

### Étape 4 : Enrichissement du plan d'expérience

- Si le nombre des échantillons  $N_l$  est inférieur au  $2 \times \text{card}(\mathcal{A}_{q,\omega}^{n,p})$ , on enrichit le plan d'expérience. Dans ce cas, on répète les étapes 2 et 3 en utilisant les nouveaux échantillons. Cette étape est effectuée afin d'assurer un bon conditionnement de la matrice de Fisher.

### Étape 5 : Test d'arrêt

- L'algorithme s'arrête dès que  $Q^2 \geq Q_{Cible}$  ou bien évidemment lorsque  $p = p_{max}$ .

- En utilisant la régression par le moindre angle :

La régression par le moindre angle est généralement connue par son appellation en anglais *Least Angle Regression* (LAR). Cette technique statistique développée par Efron et al. (2004), est une sorte d'algorithme permettant d'adapter les modèles de régression linéaire aux données de grandes dimensions. Dans le contexte de PCE, Blatman et Sudret (2011) ont appliqué l'algorithme LAR qui est basé sur le déplacement itératif des régresseurs d'un ensemble candidat à un ensemble actif pour construire des PCE creux qui sont précis même avec des plans d'expérience très petits. Le coefficient associé est directement augmenté jusqu'au point où un autre prédicteur de base a autant de corrélation avec le résidu. Cet algorithme est résumé comme la suite :

1. Au début, les coefficients du PC sont nuls  $a_\alpha = 0$  et le résidu initial est égal au vecteur des observation  $y$ . On commence par les polynômes de la base  $\psi_\alpha$  en tant qu'ensemble candidat et l'ensemble actif est initialement vide  $\phi$ .
2. On cherche le vecteur  $\psi_{\alpha_j}$  (régresseur) qui est le plus corrélé avec le résidu actuel.
3. On décale les coefficients de l'ensemble actif en cours  $a_{\alpha_j}$  vers leurs valeurs de moindres carrés jusqu'à ce qu'un autre régresseur de l'ensemble candidat aura la même corrélation avec le résidu actuel.
4. On calcule l'erreur pour  $Err_{LOO}^j$  l'itération actuel. On stocke l'erreur minimal dans  $Err_{LOO}^{min}$ .
5. On actualise tous les coefficients actifs et on déplace  $\psi_{\alpha_j}$  vers l'ensemble actif.

On répète les étapes précédentes (sauf 1) jusqu'à ce que la taille de l'ensemble actif soit égale à  $\min(P, n-1)$ . L'algorithme peut être aussi arrêté si  $Err_{LOO}^{min} < Err_{Cible}$ .

Une fois que le processus est arrêté, L'ensemble des régresseurs actifs  $\mathcal{A}^{min}$  qui correspond à la plus faible erreur  $Err_{LOO}^{min}$  est sélectionné comme la meilleure base pour développer notre polynôme de chaos creux. Par ailleurs, l'estimation des coefficients de ce dernier est effectuée à l'aide de la régression à moindres carrés ordinaire.

## 2.2.5.5 PCE creux bayésien :

L'idée de générale de ce type d'expansion en polynôme de chaos creux est proposée par Shao et al. (2017), elle est basée sur une approche statistique bayésienne qui effectue des comparaisons quantitatives de modèles concurrents (Neuman , 2003) à l'aide du critère KIC qui sélectionne le meilleur PCE creux pour un échantillon d'entrée/sortie donné.

Considérons  $N_m$  approximations plausibles  $f_{\mathcal{A}_k}$  données par PCE creux :

$$f_{\mathcal{A}_k}(x) \equiv \sum_{\alpha \in \mathcal{A}_k} a_\alpha \psi_\alpha(x) \quad , \quad k=1, \dots, N_m \quad (3.45)$$

Cette équation peut être écrite sous la forme vectorielle :

$$f_{\mathcal{A}_k} = a_k \psi_k \quad (3.46)$$

Avec,  $a_k$  et  $\psi_k$  sont respectivement le vecteur des coefficients et le vecteur des termes polynomiaux. Les échantillons  $y = \{y^{(1)}, \dots, y^{(N)}\}^T$  sont l'image de l'ensemble des vecteurs aléatoires d'entrées  $x = \{x^{(1)}, \dots, x^{(N)}\} : x^{(i)} = \{x_1, \dots, x_n\}$  par la fonction  $f(x)$  qui définit le modèle mathématique. Afin de sélectionner les modèles les plus pertinents parmi l'ensemble  $\{f_{\mathcal{A}_k}, k=1, \dots, N_m\}$ , une approche statistique formelle basée sur le théorème de Bayes est utilisée et appelée *Bayesian model averaging* (BMA), elle estime les probabilités à posteriori pour que chaque modèle soit le meilleur parmi les autres concurrents. Ces probabilités sont données par le théorème de Bayes :

$$\mathcal{P}(f_{\mathcal{A}_k} | y) = \frac{\mathcal{P}(y | f_{\mathcal{A}_k}) \mathcal{P}(f_{\mathcal{A}_k})}{\sum_{i=1}^{N_m} \mathcal{P}(y | f_{\mathcal{A}_i}) \mathcal{P}(f_{\mathcal{A}_i})} \quad (3.47)$$

Où,  $\mathcal{P}(f_{\mathcal{A}_k})$  représente la probabilité à priori que le modèle  $f_{\mathcal{A}_k}$  soit le meilleur de l'ensemble des modèles avant l'échantillonnage, l'expression du dénominateur est une constante de normalisation qu'on peut négliger parfois, et  $\mathcal{P}(y | f_{\mathcal{A}_k})$  est la probabilité que les échantillons expriment une préférence pour le modèle  $f_{\mathcal{A}_k}$ . Le calcul de ce terme, connu sous le nom

*Bayesian model evidence* (BME), est très compliqué. Pour ce faire, Kayshap (1982) a défini un critère, appelé *Kashyap information criterion* (KIC) à partir duquel on peut calculer le BME facilement :

$$KIC_k = -2 \ln \mathcal{P}(y | f_{\mathcal{A}_k}) \simeq -2 \ln \mathcal{P}(y | f_{\mathcal{A}_k}, \tilde{a}_k) - 2 \ln \mathcal{P}(\tilde{a}_k | f_{\mathcal{A}_k}) - P_k \ln(2\pi) - \ln |C_{\tilde{a}\tilde{a}}| \quad (2.48)$$

Avec,  $P_k$  le nombre des termes du modèle  $f_{\mathcal{A}_k}$ ,  $\tilde{a}_k = C_{\tilde{a}\tilde{a}}^{-1} \psi_k^T y / \tilde{\sigma}_k^2$  est l'ensemble des coefficients les plus probable pour le modèle, avec  $\tilde{\sigma}_k^2 = \frac{1}{N} (y - \psi_k \tilde{a}_k)^T (y - \psi_k \tilde{a}_k)$ .  $C_{\tilde{a}\tilde{a}} = (\psi_k^T \psi_k / \tilde{\sigma}_k^2 + C_{\tilde{a}\tilde{a}}^T)^{-1}$  et  $\mathcal{P}(\tilde{a}_k | f_{\mathcal{A}_k}) \sim \mathcal{N}(\tilde{a}_k, C_{\tilde{a}\tilde{a}})$  sont respectivement la matrice de covariance et la distribution à priori associés à l'ensemble  $\tilde{a}_k$ , et finalement,  $\mathcal{P}(y | f_{\mathcal{A}_k}, \tilde{a}_k) \sim \mathcal{N}(f_{\mathcal{A}_k}, \tilde{\sigma}_k^2)$  représente la fonction de vraisemblance qui exprime la probabilité des échantillons pour différentes valeurs de l'ensemble  $\tilde{a}_k$ . D'après son expression, la plus petite valeur de  $KIC_k$  correspond au meilleur modèle  $f_{\mathcal{A}_k}$ .

En se basant sur ce concept, Shao et al. (2017) ont proposé un algorithme qui permet de sélectionner le PCE creux optimal. Les étapes de cet algorithme sont comme suit :

#### Étape 1 : Initialisation

- On choisit un plan d'expérience  $x$ , de taille  $N$ , issu d'un échantillonnage quasi-aléatoire. Le vecteur  $y$  regroupe les évaluations du modèle  $f(x)$ . Ensuite, on transforme les vecteurs  $x$  et  $y$  en vecteurs normalisés  $\bar{x} = (x - E[x]) / V[x]$  et  $\bar{y} = (y - E[y]) / V[y]$ .

- On définit le degré maximal  $p$  et le rang maximal  $j$  initiaux (généralement  $p = 2, j = 1$ ). Puis, le sous-ensemble des multi-indices est créé  $\mathcal{A}_j^{n,p} \equiv \{\alpha \in \mathbb{N}^n : |\alpha| \leq p, \|\alpha\|_0 \leq j\} / \{\mathbf{0}\}$ .

**Étape 2 : Classement par le coefficient de corrélation**

- On calcule le coefficient de corrélation de Pearson  $r_i$  entre chaque polynôme multivarié  $\psi_i(x) \forall i = 1, \dots, P = \text{card}(\mathcal{A}_j^{n,p})$  associé à  $\mathcal{A}_j^{n,p}$  et la réponse du modèle normalisée  $\bar{y}$  avec la formule suivante :

$$r_i = \frac{COV[\bar{y}, \psi_i(x)]}{\sqrt{V[\bar{y}]V[\psi_i(x)]}} \quad (2.49)$$

Avec,  $COV[.,.]$  est l'opérateur covariance entre deux variables aléatoires.

- On trie les valeurs du vecteur  $(r_1^2, r_2^2, \dots, r_p^2)$  par ordre décroissant et selon ce tri, on réorganise les polynômes multivariés de la base dans un nouveau vecteur  $\hat{\psi}$ .

**Étape 3 : Classement par le coefficient de corrélation partiel**

- On calcule le coefficient de corrélation partielle  $\hat{r}_{i|1, \dots, i-1}$  entre chaque polynôme  $\hat{\psi}_i(x)$   $\forall i = 1, \dots, P$  et  $\bar{y}$  selon :

$$\hat{r}_{i|1, \dots, i-1} = \frac{COV[\bar{y}, \hat{\psi}_i(x) | \hat{\psi}_1(x), \dots, \hat{\psi}_{i-1}(x)]}{\sqrt{V[\bar{y} | \hat{\psi}_1(x), \dots, \hat{\psi}_{i-1}(x)]V[\hat{\psi}_i(x) | \hat{\psi}_1(x), \dots, \hat{\psi}_{i-1}(x)]}} \quad (2.50)$$

Où,  $COV[.,. | .]$  est l'opérateur covariance conditionnelle et  $V[. | .]$  est l'opérateur variance conditionnelle.

- On trie le vecteur  $(\hat{r}_1^2, \hat{r}_{2|1}^2, \hat{r}_{3|1,2}^2, \dots, \hat{r}_{P|1, \dots, P-1}^2)$  par ordre décroissant, et comme à l'étape précédente, on classe les polynômes de la base du PC dans le vecteur  $\tilde{\psi}$  selon le vecteur des coefficients de corrélation partiel trié.

- On initialise  $KIC_1 = +\infty$ ,  $\psi_{\mathcal{A}} = \tilde{\psi}_1$  et  $k = 2$ .

**Étape 4 : Identification du PCE creux actuel**

- On définit le méta-modèle  $f_{\mathcal{A}_k}$  donné par PCE creux basé sur  $\psi_k = (\psi_{\mathcal{A}}, \tilde{\psi}_k)$  et on calcule son critère  $KIC_k$ .
- Si  $KIC_k \leq KIC_{k-1}$ , on actualise  $\psi_{\mathcal{A}} = \psi_k$  et  $a_{\mathcal{A}} = \tilde{a}_k$ , sinon  $KIC_k = KIC_{k-1}$ . Puis, on passe à l'itération suivante  $k = k + 1$ .

Cette étape va être répétée jusqu'à  $k = P$ .

**Étape 5 : Enrichissement du  $\mathcal{A}_j^{n,p}$**

- à la fin de l'étape précédente, on a obtenu le modèle  $f_{\mathcal{A}} = a_{\mathcal{A}}\psi_{\mathcal{A}}$  basé sur le sous-ensemble  $\mathcal{A}$ . Si ce dernier contient des éléments de degré  $|\alpha| = p - 1$ , alors  $p = p + 2$ . Ou s'il possède des éléments de rang (ordre d'interaction)  $\|\alpha\|_0 = j$ , alors  $j = j + 1$ . Dans ces deux cas,  $\mathcal{A}_j^{n,p} = \mathcal{A}$  et on reprend l'algorithme à partir de l'étape 2.
- Si le PCE creux actuellement identifié ne contient pas d'éléments de base de degré supérieur à  $p - 1$  ou d'interaction d'ordre  $j$ . Alors, ce PCE représente fidèlement notre modèle mathématique et il est considéré comme le meilleur PCE creux bayésien et on arrête le processus.

2.2.5.6 Post-traitement du PCE :

2.2.5.6.1 Les moments statistiques :

Considérons le méta-modèle  $f_{\mathcal{A}}(x) \equiv \sum_{\alpha \in \mathcal{A}} a_{\alpha} \psi_{\alpha}(x)$  qui est l'approximation de la fonction  $f(x)$  qui représente notre modèle mathématique. Les moments statistiques de la réponse de ce modèle peuvent être estimés par leurs expressions analytiques (eq. 2.6 et 2.7). Cependant, ils peuvent également être obtenus directement à partir des coefficients du PCE dus aux propriétés d'orthogonalités des fonctions de base polynomiales. Les moments du premier et du deuxième ordre de la variable aléatoire c'est-à-dire la valeur moyenne et la variance se présentent respectivement :

$$f_0 = \mu_{f_{\mathcal{A}}} = E[f_{\mathcal{A}}(x)] = a_0 \quad (2.51)$$

$$D = \sigma_{f_{\mathcal{A}}}^2 = V[f_{\mathcal{A}}(x)] = E\left[\left(f_{\mathcal{A}}(x) - \mu_{f_{\mathcal{A}}}\right)^2\right] = \sum_{\alpha \in \mathcal{A}} a_{\alpha}^2 \quad (2.52)$$

### 2.2.5.6.2 Les indices de sensibilité

Une fois que le PCE pour une réponse du modèle est construit, une analyse de sensibilité globale peut être réalisée à un coût de calcul supplémentaire négligeable en calculant analytiquement les indices de Sobol. En effet, en utilisant la propriété d'ortho-normalité de la base polynomiale, Sudret, (2006, 2008) a trouvé une relation entre le développement en PC et la décomposition ANOVA. Pour ce faire, Il a défini l'ensemble des multi-indices  $\mathcal{A}_u$  qui dépend exactement du sous-ensemble de variables  $u = \{i_1, \dots, i_s\} \subset \{1, \dots, n\}$  :

$$\mathcal{A}_u \equiv \{\alpha \in \mathbb{N}^n : k \in u \Leftrightarrow \alpha_k \neq 0\} \quad (2.53)$$

L'unicité de la décomposition de Sobol permet d'obtenir ( $x_u = \{x_{i_1}, \dots, x_{i_s}\}$ ) :

$$f(x) \equiv a_0 + \sum_{u \subset \{1, \dots, n\}} f_u(x_u) \quad (2.54)$$

Avec :

$$f_u(x_u) = \sum_{\alpha \in \mathcal{A}_u} a_{\alpha} \psi_{\alpha}(x) \quad (2.55)$$

Dans ce cas, on peut définir l'indice de sensibilité de Sobol  $S_u$  pour le sous-ensemble des variables  $x_u$  par :

$$S_u = \frac{D_u}{D} \quad , \quad D_u = V[f_u(x_u)] = \sum_{\alpha \in \mathcal{A}_u} a_{\alpha}^2 \quad (2.56)$$

Où,  $D_u$  est la variance partielle. L'indice de sensibilité total de  $x_u$  est donné par :

$$S_u^T = \frac{D_u^T}{D} \quad , \quad D_u^T = \sum_{v \supset u} D_v = \sum_{v \supset u} \sum_{\alpha \in \mathcal{A}_v} a_{\alpha}^2 \quad (2.57)$$

donc, on peut définir comme un cas particulier de l'équation (2.89), l'indice de Sobol de premier ordre de la variable  $x_{i=1,\dots,n}$ , ainsi que son indice de sensibilité total :

$$S_i = \sum_{\alpha \in \mathcal{A}_i} a_\alpha^2 / D \quad , \quad \mathcal{A}_i \equiv \{ \alpha \in \mathcal{A} : \alpha_i > 0, \alpha_{j \neq i} = 0 \} \quad (2.58)$$

$$S_i^T = \sum_{\alpha \in \mathcal{A}_i^T} a_\alpha^2 / D \quad , \quad \mathcal{A}_i^T \equiv \{ \alpha \in \mathcal{A} : \alpha_i > 0 \} \quad (2.59)$$



## 2.3 Estimation des paramètres

### 2.3.1 Introduction :

Le but principal d'un modèle mathématique défini par  $f(x)$  est de comprendre et de reproduire les phénomènes physiques compliqués, c'est ce qu'on appelle le problème direct. Evidemment, il permet de prévoir la réponse du système  $Y$  sous l'effet de conditions imposées, représentées par un ensemble des paramètres  $x = \{x_1, \dots, x_n\}$ . Par ailleurs, l'attribution des valeurs à ces paramètres présente généralement une difficulté majeure lors de l'utilisation de ce type de modèles sur des cas concrets. En effet, certains paramètres ne sont pas mesurables d'une façon directe, on doit alors les estimer à partir de mesures de la variable d'état (la réponse) en utilisant l'approche inverse (Blum 1999; Kern 2008). Cette approche est souvent utilisée dans différentes disciplines scientifiques principalement l'économie l'imagerie et l'intelligence artificielles. En ce qui concerne l'hydrogéologie, les problèmes inverses sont des outils primordiaux d'indentification des caractéristiques hydrauliques des sols (Sun 1995 ; Chavent et al. 1995) ainsi que les paramètres de transport ou de diffusion en milieu poreux (Cartalade et al. 2003 ; Cartalade et al. 2005).

L'estimation des paramètres est une tâche délicate à réaliser à cause de la non-identifiabilité (Vrugt et al. 2001 ; Iskrev 2010). En effet, on risque souvent de tomber sur un problème inverse mal posé au sens mathématique de Hadamard (1952), c'est-à-dire un modèle insensible aux paramètres avec un manque d'information dans les mesures. Ceci peut engendrer la non unicité, l'instabilité ou l'inexistence de la solution. Ces difficultés imposent une étude préliminaire de sensibilité du modèle numérique afin d'éviter l'estimation des paramètres qui n'ont pas d'effet sur la sortie du modèle. Par ailleurs, il est important de quantifier l'incertitude associée aux paramètres puisqu'elle peut influencer la précision de la prédiction (Clément et al. 2004 ; Allaire et al. 2004).

D'une façon générale, la résolution du problème inverse est basée sur la minimisation de la fonction objectif qui mesure l'écart entre les valeurs calculées par le modèle et les observations. Le défi à ce stade, est de définir la fonction objectif adéquate au problème étudié et de bien choisir le schéma numérique qui la minimise ou la maximise quelquefois. Dans la suite, nous allons présenter les différentes méthodes de résolution du problème inverse et notamment l'approche bayésienne qui va être utilisée dans les chapitres suivants.

### 2.3.2 Méthodes de résolution du problème inverse :

Le problème inverse est une approche très utilisée avec un grand nombre de méthodes de résolution de ce problème dans la littérature. Le problème inverse est très dépendant de la fonction objectif à optimiser qui le définit. À son tour, cette fonction varie beaucoup d'un domaine d'application à un autre ; par exemple, en économie on l'appelle *l'utilité* et qui est totalement différente à *la fonction de récompense* utilisée dans le domaine d'intelligence artificielle ou *la fonction fitness* d'un code génétique etc. De plus, plusieurs chercheurs ont développé différentes méthodes qui permettent d'obtenir des solutions aux problèmes inverses et qui peuvent être classées selon les critères suivants :

- *Méthodes directes ou indirectes :*

Les méthodes directes de résolution du problème inverse consistent à reformuler le modèle étudié dès le début en inversant les inconnus et les données (Nelson 1960 ; 1961), c'est-à-dire, un autre modèle représenté par une fonction  $F(Y)$  qui prend comme entrée la réponse  $Y$  du modèle direct et nous fournit les paramètres  $x = \{x_1, \dots, x_n\}$  à estimer. En revanche, les méthodes d'inversion indirectes sont basées sur le calcul itératif de l'écart entre les mesures et les valeurs obtenues à la suite de la résolution du problème direct en modifiant à chaque fois les paramètres jusqu'à l'obtention d'un écart négligeable, c'est le principe de l'optimisation. Généralement, la majorité des techniques d'estimation des paramètres sont reposées sur une approche inverse indirecte.

- *Méthodes linéaires ou non-linéaires*

Un problème inverse linéaire peut être représenté par un système linéaire (Hoeksema et Kitanidis, 1984 ; Rubin et Dangan, 1987). Cela signifie que le vecteur des mesures est égal au vecteur des paramètres optimaux multipliés par un opérateur linéaire appelé matrice d'observation. Dans le cas d'un problème inverse non-linéaire, cet opérateur n'est pas linéaire. Notons que la plupart des problèmes inverses étudiés sont non-linéaires.

- *Méthodes déterministes ou stochastiques*

On peut encore classer les méthodes d'inversion selon la façon d'interprétation de l'espace de recherche des paramètres et on distingue : les méthodes déterministes qui

considèrent cet espace comme étant unique et les méthodes stochastiques qui l'explorent d'une manière aléatoire. Autrement dit, en utilisant l'inversion déterministe on produit ce qu'on considère comme une solution « optimale » unique, c'est-à-dire connue parfaitement. Cependant, les méthodes du problème inverse stochastiques (ou probabilistes) fournissent de nombreuses solutions possibles suivant une distribution aléatoire, donc elles permettent une idée sur la précision de la prédiction des paramètres.

- *Méthodes locales ou globales*

Les méthodes d'estimation des paramètres locales visent à chercher un extremum local de la fonction objectif en utilisant le calcul numérique de sa fonction dérivée. Ce type des méthodes peut affronter quelques problèmes dans le cas d'un problème multimodal. Les méthodes d'inversion globales recherchent l'extremum global de la fonction objectif en parcourant tout l'espace des paramètres. Dans le cas d'indentification d'un grand nombre des paramètres, ce genre des méthodes devient trop coûteux en terme de temps de calcul.

- *Méthodes basées sur le gradient ou sans gradient*

Les méthodes des problèmes inverses basées sur le calcul du gradient, souvent appelées méthodes de descente, font partie des plus anciennes et des plus utilisées. Par exemple, la méthode de l'état adjoint et la méthode de Levenberg Marquard qui sont basées sur le gradient et sont généralement utilisées dans le domaine de l'hydrogéologie (Kool et al. 1987 ; Lehmann et Ackerer, 1997) en raison de leur rapidité. L'inconvénient de ces méthodes est qu'elles sont compatibles qu'avec les fonctions différentiables puisqu'elles sont parmi les méthodes locales donc elles risquent de converger vers un extremum local. Les méthodes d'inversion sans calcul de gradient sont plus robustes et peuvent être utilisées avec des fonctions non différentiables mais elles sont plus coûteuses en temps de calcul.

- *Méthodes combinatoires ou continues*

Les problèmes inverses combinatoires consistent à trouver les paramètres à estimer parmi un ensemble discret de réalisations c'est-à-dire ils ont un nombre fini de valeurs possibles qui sont le plus souvent sous la forme d'entiers ou de binaires, c'est pourquoi on les appelle aussi les méthodes discrètes. En revanche, les variables utilisées dans la fonction objectif

en utilisant les méthodes d'inversion continues doivent être des variables continues. Les problèmes d'inversion continue sont généralement plus simples à résoudre.

- *Méthodes sous contraintes ou sans contraintes*

Les méthodes inverses sous contraintes imposent des conditions sur le choix des paramètres. L'insertion de ces contraintes est effectuée dans la fonction objectif sous forme d'égalités ou d'inégalités entre les paramètres ou simplement sous forme des bornes. L'estimation des paramètres sans contraintes est exécutée d'une manière libre, autrement dit, les paramètres peuvent prendre n'importe quelle valeur de l'espace de recherche.

- *Méthodes mono-objectif ou multi-objectif*

Les méthodes inverses mono-objectif sont définies par une seule fonction objectif, ce qui est souvent le cas pour la modélisation des transferts dans les milieux poreux. Parfois, lors d'un compromis entre plusieurs objectifs contradictoires, on utilise les méthodes d'inversion multi-objectifs qui sont basées sur une liste de fonctions objectif à optimiser simultanément. Ces méthodes peuvent être converties en mono-objectif mais c'est une procédure délicate.

Il existe dans la littérature encore plusieurs critères qui servent à classer les méthodes inverses, tels que les méthodes à contraintes séparables ou non séparables, les méthodes dépendantes du temps ou non et les méthodes monodimensionnelles ou multi-dimensionnelles.

Le choix de la méthode d'estimation dépend principalement de la nature du problème étudié. En effet, les problèmes inverses qui sont appliqués en hydrogéologie sont généralement indirects (vue la complexité de la formulation du problème), continus et mono-objectif. Depuis plus de trois décennies, la méthode de l'état adjoint a été largement utilisée pour estimer les paramètres de transfert dans les sols (Chavent, 1974 ; Sun, 1994). Récemment, une étude comparative de deux types de cette méthode (discrète et continue) a été réalisée par Delay et al. 2017.

Dans le cadre de cette thèse, on a adopté une méthode basée sur l'inférence Bayésienne vu qu'elle offre un cadre rigoureux pour l'estimation des paramètres en partant de données bruitées et pour les modèles incertains. Les résultats de cette estimation pour la première application

(Chapitre 3), sont comparés à ceux obtenus avec l'approximation de premier ordre (FOA) basé sur l'algorithme Levenberg Marquardt.

### 2.3.3 Approximation de premier ordre (FOA)

L'approximation de premier ordre (*First Order Approximation*) est une approche basée sur une méthode locale de recherche de l'optimum comme celle de Levenberg-Marquardt. L'objectif est de déterminer les  $n$  paramètres  $x = \{x_1, \dots, x_n\}$  qui interviennent dans la fonction  $f(x)$  à partir de l'ensemble de  $N_{obs}$  observations  $Y = \{Y_1, \dots, Y_{N_{obs}}\}$ . Pour cela, on commence d'abord par le choix de la fonction objectif qui est une étape décisive lors de ce processus. De nombreuses approches ont été développées pour définir une fonction objectif adéquate à ce type de problème. Le critère quadratique est l'un des plus utilisés grâce à son critère intuitif (Carrera et Neuman 1986). Cette fonction vise à calculer la somme des carrés des écarts pondérés entre les mesures  $Y_i$  et la fonction  $f_i(x)$  :

$$j(x) = (f(x) - Y)^T w (f(x) - Y) \quad (2.60)$$

$$= \sum_{i=1}^{N_{obs}} w_{ii} (f_i(x) - Y_i)^2 \quad (2.61)$$

Sachant que,  $w$  est une matrice diagonale de pondération et ses coefficients  $w_{ii}$ , appelés souvent les poids des mesures sont définis a priori et nous permettent de favoriser une mesure sur une autre selon leurs impacts sur l'optimisation.

Jusqu'à maintenant, on est dans le cadre des méthodes déterministes et afin d'ajouter un aspect stochastique à ce type des méthodes, la matrice de pondération dépend des mesures d'erreur pour chaque observation et des corrélations entre elles. Par exemple, elle peut être ajustée à l'inverse de la matrice de covariance des erreurs de mesure  $w = Cov_Y^{-1}$ , c'est-à-dire que les coefficients de la diagonale prennent l'inverse des variances des mesures  $w_{ii} = 1/\sigma_{Y_i}^2$ . Dans ce cas, la fonction objectif sera nommée *le critère d'erreur du khi-carré* car la somme des carrés des écarts suit une loi du  $\chi^2$  (*khi-carré*).

Puisque le problème inverse considéré est non linéaire, sa résolution est réalisée en minimisant la fonction objectif  $j(x)$  itérativement. En effet, à chaque itération  $k$ , on cherche l'ensemble

des perturbations  $\Delta x^{(k)} = \{\Delta x_1^{(k)}, \dots, \Delta x_n^{(k)}\}$  aux paramètres  $x^{(k)}$  pour obtenir des  $x^{(k+1)} = x^{(k)} + \Delta x^{(k)}$  qui réduisent le critère quadratique  $j(x)$ , cela signifie que,  $j(x)^{(k+1)} < j(x)^{(k)}$  jusqu'à un critère d'arrêt prédéfini. Le défi dans ce genre de méthodes est la manière de trouver la valeur de perturbation. Il existe plusieurs algorithmes dans la littérature, qui ont été développés pour résoudre ce problème.

### 2.3.3.1 L'algorithme de la descente du gradient

Cet algorithme est aussi connu sous le nom de l'algorithme de descente la plus rapide (*steepest descent*) ; il permet de mettre à jour les valeurs des paramètres dans la direction de la descente, c'est-à-dire, la direction opposée au gradient de la fonction objectif par rapport aux paramètres :

$$\nabla j(x) = \frac{\partial}{\partial x} j(x) = 2(f(x) - Y)^T w \frac{\partial}{\partial x} (f(x) - Y) \quad (2.62)$$

$$= 2(f(x) - Y)^T w J \quad (2.63)$$

Où,  $J = \partial f(x) / \partial x$  est la matrice Jacobienne, de dimensions  $N_{obs} \times n$ , représente la sensibilité de la fonction  $f(x)$  à la variation des paramètres  $x$ . Le coefficient de perturbation  $\Delta x$  qui modifie les paramètres dans le sens de la descente la plus rapide est donné par :

$$\Delta x = -\frac{\rho}{2} \nabla j(x) = \rho J^T w (Y - f(x)) \quad (2.64)$$

Avec,  $\rho$  un scalaire positif prédéfini qui désigne la longueur d'un pas dans la direction de la descente. Cette méthode converge bien pour les fonctions objectif simples (Madsen et al., 2004 ; Lourakis, 2005). En outre, ce type de méthodes est le plus performant pour les problèmes ayant un grand nombre de paramètres.

### 2.3.3.2 L'algorithme de Gauss-Newton

L'algorithme de Gauss-Newton (GN) est un processus robuste de résolution des problèmes de moindres carrés non linéaires ; il est spécifique à la minimisation des fonctions objectif de type quadratique. Pour les problèmes de taille moyenne, cet algorithme converge généralement beaucoup plus rapidement que les méthodes du gradient et les autres méthodes qui nécessitent les dérivées secondes (Hessien) tel que l'algorithme de Newton.

L'idée de base de l'algorithme de GN est d'appliquer l'approximation quadratique sur la fonction objectif au voisinage de la solution optimale (Björck, 1996). Le développement en série de Taylor du premier ordre (d'où l'appellation de ce type de méthodes « l'approximation de premier ordre ») permet d'écrire

$$f(x + \Delta x) \approx f(x) + \frac{\partial f(x)}{\partial x} \Delta x = f(x) + J \Delta x \quad (2.65)$$

Cette approximation est substituée dans l'expression de la fonction objectif (2.93) pour avoir une approximation de  $j(x + \Delta x)$  :

$$j(x + \Delta x) \approx j(x) + 2(f(x) - Y)^T w J \Delta x + \Delta x^T J^T w J \Delta x \quad (2.66)$$

En appliquant cette procédure, on obtient une approximation de second ordre de la fonction objectif sachant que  $J^T w J$  n'est qu'une approximation du Hessien  $\nabla^2 j(x)$ . Le gradient de cette approximation est donné par :

$$\frac{\partial}{\partial x} j(x + \Delta x) \approx 2(f(x) - Y)^T w J + 2 \Delta x^T J^T w J \quad (2.67)$$

Enfin, les coefficients de perturbation  $\Delta x$  de GN qui minimisent la fonction objectif sont trouvés à partir de l'équation  $\partial j(x + \Delta x) / \partial h = 0$ . En effet, ils représentent la solution particulière du système suivant :

$$\left[ J^T w J \right] \Delta x = J^T w (Y - f(x)) \quad (2.68)$$

### 2.3.3.3 L'algorithme de Levenberg-Marquardt

L'algorithme de Levenberg-Marquardt (LM) est une interpolation entre l'algorithme de Gauss-Newton et l'algorithme de la descente du gradient. L'ensemble des coefficients de perturbation de LM est la solution particulière du système donné par Levenberg (1944) :

$$\left[ J^T wJ + \lambda I \right] \Delta x = J^T w(Y - f(x)) \quad (2.69)$$

Le concept est d'ajouter un facteur positif d'amortissement  $\lambda$  (Tikhonov, 1943 ;1963), qui est ajusté à chaque itération. Un paramètre d'amortissement nul  $\lambda = 0$  traduit une perturbation de Gauss-Newton et quand  $\lambda$  tend vers l'infinie, la perturbation est celle de la descente de gradient.

Au début, le facteur d'amortissement  $\lambda$  est initialisé à des grandes valeurs de sorte que les premières modifications des paramètres sont effectuées avec des petits pas par la descente de gradient qui est généralement efficace loin d'un minimum. Si tout se passe bien, c'est-à-dire que la fonction objectif diminue, on diminue le facteur d'amortissement  $\lambda$  afin de se rapprocher de l'algorithme GN et d'avoir une convergence accélérée vers un minimum local. Sinon, dans le cas d'une divergence ( $j(x)^{(k+1)} > j(x)^{(k)}$ ), on augmente la valeur de  $\lambda$ .

Marquardt (1963) a proposé une mise à jour du système en normalisant les valeurs de  $\lambda$  par  $J^T wJ$  :

$$\left[ J^T wJ + \lambda \text{diag}(J^T wJ) \right] \Delta x = J^T w(Y - f(x)) \quad (2.70)$$

Ce processus itératif est effectué jusqu'à satisfaire un des critères de convergence suivants :

- *Convergence du gradient* :  $\max |J^T w(Y - f(x))| < \varepsilon_1$
- *Convergence des paramètres* :  $\max |\Delta x/x| < \varepsilon_2$
- *Convergence de la fonction objectif* :  $j(x)/(N_{obs} - n) < \varepsilon_3$

Sinon, l'algorithme s'arrête lorsque le nombre d'itérations dépasse une limite prédéfinie.

#### 2.3.3.4 Analyse de précision

Une étude statistique peut être effectuée une fois que l'ajustement des paramètres est terminé ; autrement dit, une quantification de la précision du problème inverse est réalisée en attribuant un intervalle de confiance à chaque paramètre estimé. Tout d'abord, on commence par l'examen du critère de convergence de la fonction objectif  $j(x)/(N_{obs} - n)$  qui indique la qualité de l'ajustement, en effet, une grande valeur de ce critère indique un mauvais ajustement,



une valeur au voisinage de 1 montre que l'erreur d'ajustement est du même ordre que l'erreur de mesure (comme souhaité), et une petite valeur de ce critère (inférieure à 1) indique que le modèle ajuste trop les données, c'est-à-dire, un problème de Sur-apprentissage (*overfitting*). Ensuite, on calcule la matrice de covariance des paramètres estimés  $Cov_x$  à partir de la covariance des erreurs des mesures  $Cov_Y$  connues à priori qui est l'inverse de la matrice de pondération  $w = Cov_Y^{-1}$ :

$$Cov_x = [J^T w J]^{-1} \quad (2.71)$$

Cette matrice résume les variables statistiques des paramètres obtenus par le problème inverse qui sont les écarts type des paramètres (les variances) et les corrélations entre eux :

$$\sigma_x^2 = V_x = \text{diag}(Cov_x) \quad ; \quad V_{x_i} = Cov_{x_i, x_i} \quad (2.72)$$

$$Cor_{x_i, x_j} = \frac{Cov_{x_i, x_j}}{\sqrt{V_{x_i} V_{x_j}}} \quad (2.73)$$

Lorsque les erreurs de mesures sont inconnues à priori, c'est-à-dire avant l'inversion, l'analyse d'erreur peut être effectuée en considérant une erreur moyenne pour toutes les mesures :

$$\sigma_Y^2 = V_Y = \frac{1}{N_{obs} - n} (f(x) - Y)^T (f(x) - Y) \quad (2.74)$$

Dans ce cas, pour améliorer la précision de l'inversion, on peut réaliser une autre estimation des paramètres en introduisant une nouvelle matrice de pondération  $w$  dont la diagonale est l'inverse des erreurs calculés à posteriori  $w_{ii} = 1/\sigma_Y^2$ . Ensuite, on calcule les erreurs de mesures de nouveau, si les nouvelles valeurs sont voisines des valeurs précédentes, on ne peut pas améliorer d'avantage la précision de notre problème inverse. Sinon, on répète une autre fois l'estimation en changeant de nouveau la matrice de pondération jusqu'à atteindre ce critère :  $\max\left(\left|\sigma_Y^{2(k)} - \sigma_Y^{2(k-1)}\right|/\sigma_Y^{2(k)}\right) < \epsilon_4$ .

Finalement, après avoir vérifié la précision de l'estimation, la largeur de l'intervalle de confiance de 95% des paramètres peut être déterminé à partir de ses variances et la valeur de l'erreur des mesures comme suit :

$$IC(x) = \sqrt{\sigma_x^2} \sqrt{\sigma_Y^2} F^{-1}(0.95 | N_{obs} - n) \quad (2.75)$$

Avec,  $x = F^{-1}(p|\nu) = \{x : F(x|\nu) = p\}$  est une fonction de distribution cumulative inverse où

$$F \text{ est définie par : } F(x|\nu) = \int_{-\infty}^x \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\nu\pi}(1+t^2/\nu)^{(\nu+1)/2}} dt.$$

Au cours des travaux de cette thèse, on a utilisé l'approche FOA pour estimer les paramètres hydrauliques ainsi que les paramètres géophysiques d'un milieu poreux non saturé à partir des mesures de potentiel spontané. Les résultats de cette estimation sont comparés de ceux obtenus avec une méthode probabiliste appelée l'approche Bayésienne.

### 2.3.4 Approche Bayésienne

Dès son apparition, l'inférence bayésienne a envahi tous les domaines scientifiques ; elle représente une base fondamentale pour collecter les données et les analyser même si leur nombre est limité. Le problème inverse bayésien permet d'estimer des paramètres  $x = \{x_1, \dots, x_n\}$  à partir d'un nombre réduit  $N_{obs}$  de données bruitées  $Y = \{Y_1, \dots, Y_{N_{obs}}\}$  et pour un modèle incertain  $\mathcal{M}$ .

Réellement, trouver la meilleure solution n'est pas le but ultime des problèmes inverses bayésiens. En effet, ces méthodes visent à chercher un ensemble de solutions possibles décrites par une densité de probabilité conjointe postérieure, en se servant de toutes les informations disponibles à priori. Dans ce cas, on ne parle pas d'un solveur mais on parle d'un échantillonneur. Pour simplifier, on peut distinguer deux types de distributions de probabilité décrivant  $x$  ; une distribution de probabilité à priori  $\mathcal{P}(x|\mathcal{M}) = \pi$  qui représente « ce qu'on pense des paramètres avant l'utilisation des données  $Y$  » et une distribution de probabilité à posteriori  $\mathcal{P}(x|Y, \mathcal{M}) = \Pi$  qui exprime « ce qu'on pense des paramètres après l'utilisation des données ». Ces deux probabilités sont liées par le fameux théorème de Bayes :

$$\mathcal{P}(x|Y, \mathcal{M}) = \frac{\mathcal{P}(Y|x, \mathcal{M})\mathcal{P}(x|\mathcal{M})}{\mathcal{P}(Y|\mathcal{M})} \quad (2.76)$$

Sachant que,  $\mathcal{P}(Y|x, \mathcal{M})$  est la vraisemblance  $\mathcal{L}$  (*Likelihood*) et  $\mathcal{P}(Y|\mathcal{M})$  est la vraisemblance marginale qui est souvent appelée l'évidence, ce terme qui ne dépend pas des

paramètres joue le rôle d'un facteur de normalisation qui ne fait aucun effet lors de l'estimation de la distribution postérieure. La distribution à posteriori s'écrit donc :

$$\mathcal{P}(x|Y, \mathcal{M}) \propto \mathcal{P}(Y|x, \mathcal{M})\mathcal{P}(x|\mathcal{M}) \quad , \quad \Pi \propto \mathcal{L}\pi \quad (2.77)$$

D'après cette équation, l'approche bayésienne dépend beaucoup de la fonction de vraisemblance (Pawitan, 2001 ; Rohde, 2014) vu que son processus consiste à « deviner » d'une façon intelligente un ensemble de paramètres possédant une distribution de probabilité à priori, ensuite, on calcule la probabilité postérieure de cette conjecture en utilisant la vraisemblance, puis on répète ces étapes afin d'augmenter la probabilité à posteriori.

#### 2.3.4.1 La fonction de vraisemblance

Il est évident que les observations  $Y$  sont incertaines, c'est-à-dire, elles sont perturbées par des bruits stochastiques dont les valeurs réelles sont indéterminables. L'idée de base de la vraisemblance est de trouver une bonne approximation de ces bruits permettant au problème inverse de tenir compte des mesures d'incertitude des mesures  $\sigma^2$ . Par exemple, la fonction de vraisemblance peut suivre une loi normale  $\mathcal{N}(0, \sigma^2)$  comme ci-dessous :

$$\mathcal{L} = \mathcal{P}(Y|x, \mathcal{M}) = \prod_{i=1}^{N_{obs}} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(f_i(x) - Y_i)^2}{2\sigma_i^2}\right) \quad (2.78)$$

En pratique la forme logarithmique de  $\mathcal{L}$  est souvent plus utilisée :

$$\log(\mathcal{L}) = -\frac{1}{2} \sum_{i=1}^{N_{obs}} \log(2\pi) + \log(\sigma_i^2) + (f_i(x) - Y_i)^2 / \sigma_i^2 \quad (2.79)$$

$$= -\frac{1}{2} (f(x) - Y)^T w (f(x) - Y) - \frac{1}{2} \log(\det(w)) - \frac{1}{2} N_{obs} \log(2\pi) \quad (2.80)$$

Où,  $w$  est la matrice de pondération dont la diagonale prend l'inverse des incertitudes des mesures :  $w_{ii} = 1/\sigma_Y^2$ . On remarque que le terme  $\sum_{i=1}^{N_{obs}} (f_i(x) - Y_i)^2 / \sigma_i^2$  représente la fonction objectif quadratique  $j(x)$  plus précisément, *le critère d'erreur du khi-carré*  $\chi^2$ . De plus, pour

un cas particulier ; en supposant que toutes les mesures possèdent une seule erreur représentative  $\sigma^2$ , le log de la vraisemblance devient :  $\log(\mathcal{L}) = c - \frac{1}{2} \chi^2$ .

Une fois que la fonction de vraisemblance est adoptée, le défi dans ce type de méthodes est de choisir le prochain ensemble de paramètres qui fait augmenter la probabilité à posteriori ; c'est le rôle des échantillonneurs.

#### 2.3.4.2 Types des échantillonneurs

Le point commun qui rassemble tous les types d'échantillonneurs utilisés dans la résolution des problèmes inverses bayésiens est le fait qu'ils font tous partie des méthodes d'échantillonnage de Monte-Carlo (MC). Le but de ces méthodes est de générer les jeux de valeurs des paramètres qui satisfassent une loi de probabilité, c'est-à-dire les plus probable en parcourant tous l'espace d'existence de ces paramètres. Il existe de nombreux échantillonneurs de type MC dans la littérature que nous allons citer par la suite ainsi que les techniques basées sur les chaînes de Markov.

- *L'échantillonnage par transformée inverse*

La première méthode d'échantillonnage de MC classique qu'on présente est appelée la méthode de la transformée inverse.

Soit  $F(x)$ ,  $\forall x \in \mathbb{R}$ , une fonction distribution cumulative (*cdf*). Ainsi,  $F: \mathbb{R} \rightarrow [0,1]$  est une fonction positive, strictement croissante et continue sur  $\mathbb{R}$ , en outre, sa limite à gauche  $F(-\infty) = 0$  et à droite  $F(+\infty) = 1$ .  $F$  est inversible et sa fonction inverse  $F^{-1}: [0,1] \rightarrow \mathbb{R}$  est définie comme suit :

$$F^{-1}(y) = \min \{x : F(x) \geq y\}, \quad \forall y \in [0,1] \quad (2.81)$$

Notre objectif est de créer une variable aléatoire  $X$  suivant  $F$  ; c'est-à-dire que  $P(X \leq x) = F(x)$ . Pour cela, on génère la variable aléatoire  $U$  à partir d'une distribution uniforme sur l'intervalle  $[0,1]$ . On peut prouver que si  $X = F^{-1}(U)$ , alors,  $X$  est distribué selon  $F$ .

- *L'échantillonnage de rejet*

Un algorithme d'échantillonnage de rejet (*rejection sampling*) est généralement utilisé pour engendrer indirectement une variable aléatoire suivant une densité de probabilité cible  $f$ , lorsqu'il est impossible de l'obtenir directement de  $f$  à cause de sa complexité par exemple.

On commence d'abord l'échantillonnage à partir d'une autre distribution de probabilité  $g$  qui est appelée distribution d'enveloppe, ensuite, on élimine les éléments indésirables selon une stratégie de rejet afin que notre échantillon suive la distribution cible  $f$ ; pour cela, on choisit un facteur d'échelle  $M$  tel que  $f \leq Mg$ , ce facteur est défini par :

$$M = \sup_x \{f(x)/g(x)\} \in [1, \infty) \quad (2.82)$$

L'algorithme qui permet de gêner un échantillon de densité de probabilité  $f$  est résumé dans les étapes suivantes :

- Générer un échantillon  $x$  de densité de probabilité  $g$ , ainsi qu'un échantillon  $U$  à partir de la loi uniforme  $\cup(0,1)$ .
- Si  $U < f(x)/Mg(x)$ , On accepte  $x$  comme une variable aléatoire de densité  $f$ , sinon on la rejette et on retourne à l'étape précédente.

Cet algorithme doit être répété au moyenne  $M$  fois pour obtenir un échantillon. Notant qu'il est difficile de savoir à l'avance le nombre d'itérations à effectuer pour obtenir un échantillon d'une taille donnée.

- *L'échantillonnage préférentiel*

L'idée générale de l'échantillonnage préférentiel (*importance sampling*) est similaire au principe de l'échantillonnage de rejet. En effet, on choisit encore une densité d'enveloppe  $g$  qui se rapproche de la densité cible  $f$  et on n'a pas besoin de la limiter. C'est pourquoi, l'échantillonnage préférentiel est jugé plus efficace que celui de rejet.

Le problème fondamental de la mise en œuvre de ce type d'échantillonnage est le choix de la distribution d'enveloppe (appelée distribution biaisée). Une fois qu'elle est fixée, on génère à

partir d'elle un échantillon  $x$ , ensuite, on calcule les poids  $\omega_x$  pour mettre l'accent sur les valeurs de  $x$  qui sont « préférés » pour la distribution cible  $f$ . Ces pondérations sont obtenues par le ratio de vraisemblance des deux densité  $\omega_x = f(x)/g(x)$ . Soit la distribution  $g^*$  qui contient l'échantillon du minimum de variance et qui est obtenue par la résolution du problème d'optimisation suivant :

$$g^* = \min_g \left\{ \text{Var}_g \left( x \frac{f(x)}{g(x)} \right) \right\} \quad (2.83)$$

La distribution qui minimise la variance ci-dessus, est :

$$g^*(x) = \frac{|x|f(x)}{\int |x|f(x) dx} \quad (2.84)$$

On remarque que, la distribution  $g^*$  n'est qu'une imitation de la distribution cible  $f$ , mais, il est possible d'engendrer à partir de laquelle des variables aléatoires.

- *L'échantillonnage de Monte-Carlo par chaînes de Markov*

Comme tous les échantillonnages basés sur la méthode de Monte Carlo, l'échantillonneur MCMC (*Markov Chain Monte Carlo*) vise à générer aléatoirement des variables suivant une distribution de probabilité cible en utilisant une autre distribution intermédiaire. L'apport de la méthode MCMC est que le choix des échantillons est effectué selon leurs probabilités, au lieu de parcourir l'espace des paramètres d'une manière équivalente et indépendante. En effet, une telle méthode engendre une séquence de valeurs dépendantes adoptant une procédure itérative stochastique nommée « chaîne de Markov » de manière que l'échantillon final soit en harmonie avec la densité cible.

Cet échantillonneur a été beaucoup adopté dans plusieurs problèmes, principalement, l'estimation bayésienne des paramètres. Ce qui explique le nombre important des algorithmes développés dans la littérature. On va citer les plus célèbres dans la suite :

- *L'algorithme de Metropolis*

L'algorithme de Metropolis (Metropolis et al., 1953), souvent appelé l'algorithme de Metropolis à marche aléatoire (*random walk Metropolis (RWM) algorithm*), est la forme la plus populaire de MCMC pour les débutants, grâce à sa simplicité de compréhension, facilité de sa mise en œuvre et sa compatibilité avec la plupart des problèmes. Cet algorithme est basé sur les étapes suivantes :

- On définit un jeu de paramètres initiale  $x^{(0)} = \{x_1^{(0)}, \dots, x_n^{(0)}\}$ ,  $i = 0$ , à partir de la distribution de probabilité à priori  $\pi$ , et on évalue la distribution de probabilité à posteriori non-normalisée  $\Pi^{(0)}$  (éq. 2.110).
- On essaye un « saut » pour obtenir un jeu de paramètres candidat  $x^* \sim J(x^* | x^{(i-1)})$  à partir de la densité de probabilité de proposition symétrique  $J$  et on calcule  $\Pi^*$ .
- Si  $\Pi^* \geq \Pi^{(i)}$ , on accepte le saut, c'est-à-dire  $x^{(i+1)} = x^*$ , sinon on accepte le saut ayant une probabilité égale à  $\Pi^* / \Pi^{(i)}$ .

Les deux dernières étapes vont être répétées jusqu'à atteindre le nombre de chaînes déjà fixé. Cet algorithme ne s'arrête pas même s'il a trouvé « la bonne » solution, c'est-à-dire un maximum de  $\Pi$ , il continue à effectuer des sauts à l'entour de cette solution afin d'étudier sa précision, c'est le but de l'approche bayésienne.

Cette procédure engendre la distinction entre deux phases : (i) étapes de rodage (*burn-in steps*), lorsqu'on cherche loin de la solution, et (ii) après rodage (*post burn-in*), quand on entre dans une région où  $\Pi$  est élevée.

Cette deuxième phase est ce qui compte pour le problème inverse, c'est pour cela, on a besoin d'un grand nombre de chaînes dans la région de  $\Pi$  élevée (entre  $10^4$  et  $10^5$ ), de plus, on élimine les étapes de rodage lors de post-traitement des données, c'est-à-dire lors de l'étude stochastique.

On remarque qu'on ne s'intéresse qu'aux rapports de distributions de probabilité à posteriori  $\Pi^* / \Pi^{(i)}$ , c'est pourquoi les évidences ne comptent pas dans MCMC.

- *L'algorithme de Metropolis-Hastings*

Comme son nom l'indique, l'algorithme de Metropolis-Hastings (MH) (Hastings, 1970) est seulement une mise à jour de l'algorithme précédent. Il remplace la stratégie de saut définie auparavant par une nouvelle densité de probabilité de proposition  $J$  qui est une distribution asymétrique dépendante des paramètres retenus dans l'itération précédente. L'algorithme s'écrit comme suit :

- On initialise  $x^{(0)}$ , à partir de la distribution de probabilité à priori  $\pi$ .
- On répète pour  $i$  de 1 à  $N_c$  le nombre total des chaînes.
  - On génère  $x^* \sim J(x^* | x^{(i-1)})$
  - On génère  $u \sim U([0,1])$  qui est indépendant de  $x^*$
  - On calcule

$$\alpha = \min \left( 1, \frac{\Pi^* J(x^* | x^{(i-1)})}{\Pi^{(i)} J(x^{(i-1)} | x^*)} \right) \quad (2.85)$$

- Si  $u \leq \alpha$ , alors, on accepte le saut et  $x^{(i)} = x^*$
- Sinon  $x^{(i)} = x^{(i-1)}$

Ainsi, l'algorithme de de Metropolis est un cas particulier du processus ci-dessus où la densité de probabilité de proposition  $J$  est symétrique c'est-à-dire  $J(x^{(i-1)} | x^*) = J(x^* | x^{(i-1)})$ .

- *L'algorithme de Gibbs*

Comme avec d'autres algorithmes MCMC, l'échantillonnage de Gibbs (Geman S. and Geman D., 1984) génère une chaîne d'échantillons de Markov, chacun d'entre eux étant corrélé aux échantillons voisins. En effet, l'échantillonnage est effectué en générant chaque variable à partir de sa distribution postérieure conditionnelle sachant que les autres variables sont fixées à leurs valeurs actuelles. Ce processus sera répété jusqu'à atteindre un nombre choisi d'itérations :

- On initialise  $x^{(0)} = \{x_1^{(0)}, \dots, x_n^{(0)}\}$ ,  $i = 0$ .
- On répète pour  $i$  de 1 à  $N_c$  le nombre des échantillons, et on génère :



$$\begin{aligned}
 x_1^{(i)} &\sim J\left(x_1 \mid x_2^{(i-1)}, x_3^{(i-1)}, \dots, x_n^{(i-1)}\right) \\
 x_2^{(i)} &\sim J\left(x_2 \mid x_1^{(i)}, x_3^{(i-1)}, \dots, x_n^{(i-1)}\right) \\
 &\dots \\
 x_j^{(i)} &\sim J\left(x_j \mid x_1^{(i)}, x_2^{(i)}, \dots, x_{j-1}^{(i)}, x_{j+1}^{(i-1)}, \dots, x_n^{(i-1)}\right) \\
 &\dots \\
 x_n^{(i)} &\sim J\left(x_n \mid x_1^{(i)}, x_2^{(i)}, \dots, x_{n-1}^{(i)}\right)
 \end{aligned}$$

Comme toutes les méthodes MCMC, l'algorithme de Gibbs ne prend pas en compte les premières itérations qui vont être abandonnées (Gilks et al., 1996) afin d'obtenir un ensemble de vecteurs  $x$  qui puissent être considérés comme étant des réalisations de la distribution de probabilité à posteriori  $\mathcal{P}(x|Y, \mathcal{M})$ .

Cet échantillonneur représente le meilleur choix lorsqu'il est possible de déterminer des densités de probabilité conditionnelles, mais ce n'est pas toujours le cas. Afin de contourner ce problème, plusieurs alternatives ont été proposées telles que l'application des méthodes d'échantillonnage de rejet adaptatif (Martino and Míguez, 2010 ; Görür and Teh, 2011) et l'algorithme d'échantillonnage Metropolis avec réjection adaptative (Meyer et al., 2008 ; Martino et al., 2015). De plus, une méthode hybride (*Metropolis within Gibbs*) est souvent utilisée en remplaçant la simulation sous la loi conditionnelle par une étape de Hastings-Metropolis dont la distribution cible est cette loi conditionnelle.

- *Les algorithmes MCMC adaptatifs*

La mise en œuvre des échantillonneurs de Monte-Carlo par chaînes de Markov exige l'intervention de l'utilisateur dans le choix de la loi de probabilité de proposition ou l'espace à balayer des paramètres, par exemple. Les algorithmes MCMC adaptatifs sont conçus afin de limiter ces interventions ; ce type d'algorithmes apprend la manière de modification des paramètres d'implémentation de son propre comportement passé, c'est-à-dire que chaque itération actualise non seulement la chaîne de Markov, mais encore, les paramètres d'implémentation. Comme résultats, on obtient des algorithmes plus performants et plus rapide en les comparant avec les algorithmes originaux RWM et MH.

Au cours des deux dernières décennies, une variété d'approches différentes a été proposée pour accroître l'efficacité de MCMC. Ces approches peuvent être regroupées en deux groupes de méthodes ; un à chaîne unique et un à chaînes multiples.

- Une méthode à chaîne unique suit une seule trajectoire en adaptant d'une façon continue la covariance  $C_x$  de la distribution de probabilité de proposition Gaussienne  $J(x^{(i-1)}|\cdot) = \mathcal{N}_n(x^{(i-1)}, s_n C_x)$ . Cette covariance est calculée en fonction de la trajectoire de la chaîne  $C_x = \text{cov}(x^{(0)}, \dots, x^{(i-1)}) + \varepsilon I_n$ . Sachant que,  $s_n$  est un facteur d'échelle qui dépend uniquement du nombre de paramètres  $n$ ,  $I_n$  est la matrice identité de dimension  $n$  et  $\varepsilon$  représente le facteur de régulation qui empêche la singularité de la matrice, ce facteur est généralement fixé à  $\varepsilon = 10^{-6}$ . Roberts et al., (1997) ont montré que la valeur optimale du facteur d'échelle est  $s_n = 2.4^2/d$  pour les distributions cibles et de proposition Gaussienne.

Parmi les méthodes à chaîne unique, nous pouvons citer : l'algorithme de Metropolis adaptatif (AM) (Gilks et al., 1998 ; Haario et al., 2001 ; Roberts et Rosenthal, 2001 ; Vihola, 2009); l'algorithme de Metropolis adaptatif de rejet retardé (DRAM) (Haario et al., 2006 ; Andrieu et Moulines, 2006 ; Saksman et Vihola, 2010) ; l'algorithme de Wang-Landau (Fort et al. 2014a ; 2014b) ; processus de Markov déterministes par morceaux (PDMP) (Bierkens et al., 2016 ; 2017) et l'algorithme de gradient stochastique avec la dynamique de Langevin (SGLD) (Welling and Teh, 2011 ; Ding et al., 2014 ; Chen et al., 2014).

- Les méthodes à chaînes multiples utilisent différentes trajectoires en parallèle afin d'explorer d'une manière plus efficace la distribution à posteriori. Ce type de méthodes montre plusieurs avantages notamment, lorsqu'il s'agit de distributions postérieures complexes impliquant de longues chaînes, des paramètres corrélés, de nombreux optimums locaux ou lorsqu'il s'agit des problèmes multimodales (Liu et al., 2000 ; ter Braak and Vrugt, 2008 ; Bédard et al., 2012 ; Wang et Dunson, 2013 ; Wang et al. 2015 ; Martino, 2018). Les algorithmes pour les chaînes multiples offrent une large gamme de mesures statistiques afin de vérifier la convergence vers la distribution cible, (Gelman and Rubin, 1992, Angelino et al. 2014).

Divers algorithmes basés sur le principe de chaînes multiples sont proposés dans la littérature, l'algorithme de l'évolution complexe mélangée de Metropolis (SCEM-UA) est parmi les plus utilisés, surtout dans le domaine d'hydrologie et d'hydrogéologie (Vrugt et al. 2003).

En 2006, ter Braak, a développé une autre méthode plus facile à illustrer et implémenter, appelée l'algorithme de l'évolution différentielle par chaînes de Markov (DE-MC), où,  $N_c$  chaînes différentes de Markov sont exécutées simultanément en parallèle pour définir une population  $X^{(t)}$ , à chaque génération  $t$ , qui correspond à une matrice  $N_c \times n$ , sachant que  $n$  est le nombre de paramètres et que chaque ligne représente une chaîne. L'échantillonnage est donc effectué à partir de cette matrice en utilisant l'évaluation différentielle (Storn and Price, 1997 ; Price et al., 2005), ensuite, la condition de rejet est celle de l'algorithme de Metropolis.

DE-MC a été amélioré par Vrugt et al., 2008, 2009, en utilisant un échantillonnage aléatoire auto-adaptatif en sous-espaces, pour donner un algorithme intitulé évolution différentielle adaptative de Metropolis (*DiffeREntial Evolution Adaptive Metropolis* DREAM), cet algorithme a été revisité par Laloy et Vrugt (2012) pour fonctionner non seulement en chaînes multiple mais encore à plusieurs essais, il est mis à disposition grâce au logiciel DREAM<sub>(ZS)</sub>. L'idée est de diviser l'ensemble des paramètres en sous-ensembles de  $n^*$  paramètres afin de permettre une correction des valeurs aberrantes de chaque chaîne et par conséquent, une convergence plus rapide vers la distribution cible.

Au cours de cette thèse, l'estimation bayésienne des paramètres est effectuée à l'aide du logiciel DREAM<sub>(ZS)</sub>, grâce à son efficacité qui a été testée avec succès sur plusieurs problèmes complexes, non linéaires et ayant un nombre énorme des paramètres qui peut aller jusqu'à 241 (Vrugt et al. 2009b ; Scharnagl et al., 2010 ; Laloy et al. 2010 ; Dekker et al., 2012). Cet algorithme génère des candidats simultanément en échantillonnant à partir d'une archive d'états antérieurs collectés dans l'échantillon. On remarque qu'au bout seulement de quelques chaînes parallèles, l'échantillon commence à converger vers la distribution à posteriori cible, c'est-à-dire que la période de rodage est très petite comparant aux autres méthodes (Laloy et Vrugt, 2012).

## 2.4 Conclusion

Dans ce chapitre, nous avons rappelé les outils et méthodes pour l'analyse de sensibilité et l'estimation des paramètres dans le but de pouvoir les utiliser pour le transfert d'eau et de polluants dans les milieux poreux non saturés.

Concernant la première partie de ce chapitre, on a défini l'analyse de sensibilité et son rôle pour un modèle numérique, ensuite, on a dénombré différentes méthodes présentées dans la littérature. La méta-modélisation est souvent utilisée comme une base rigoureuse pour décomposer la variance du modèle, ainsi, elle facilite la détermination, avec un coût de calcul réduit, des indices de sensibilité appelé souvent indices de Sobol. L'expansion en polynômes de chaos est parmi les méthodes les plus efficace de méta-modélisation, c'est pourquoi, nous avons choisi de l'utiliser dans la suite de cette thèse.

La deuxième partie du chapitre a été consacrée pour classifier les différentes méthodes d'estimation des paramètres. Principalement, on a présenté la méthode FOA, basée sur l'algorithme de Levenberg Marquardt, ainsi que la méthode MCMC inspirée d'une approche bayésienne. Cette dernière sera appliquée à l'estimation des paramètres hydraulique des sols ainsi que les paramètres de transport à partir d'expériences que nous avons effectuées au laboratoire. Le prochain chapitre sera consacré à l'analyse de sensibilité globale et à l'estimation bayésienne de paramètres appliquées à un modèle hydrogéophysique en milieu poreux non saturé.



## Bibliographie

- Allaire G., Jouve F., and Toader A. M., 2004. Structural optimization using sensitivity analysis and a level-set method. *J. Comp. Phys.*, 194 :363–393.
- An J., and Owen A., 2001. Quasi-regression". In : *J. Complex.* 17 (4), pp. 588-607.
- Andrieu C., Moulines E., 2006. *Ann. Appl.* 20051462-1505.
- Angelino E., Kohler E., Waterland A., Seltzer, M. and Adams, R. 2014. Accelerating MCMC via parallel predictive prefetching. arXiv preprint arXiv :1403.7265.
- Archer G., Saltelli A., Sobol' I. 1997. Sensitivity measures, ANOVA-like techniques and the use of bootstrap. *J. Stat. Comput. Simul.* 58, pp. 99-120.
- Askey R., Wilson J., 1985. Some basic hypergeometric orthogonal polynomials that generalize Jacobi polynomials. *Mem. Amer. Math. Soc.* 54, no. 319, iv55 pp
- Bayarri M. J., Berger J. O., Cafeo J., Garcia-Donato G., Liu F., Palomo J., Parthasarathy R. J., Paulo R., Sacks J. and Walsh D., 2007. Computer model validation with functional output. *Ann. Statist.* 35, no. 5, 1874--1906.
- Bédard M., Douc R. and Moulines E., 2012. Scaling analysis of multiple-try MCMC methods. *Stochastic Processes and their Applications*, 122 758786.
- Berveiller M., 2005. *Eléments finis stochastiques : approches intrusive et non intrusive pour des analyses d'habilité.* Ph. D. thesis, Université Blaise Pascal, Clermont-Ferrand.
- Berveiller M., Sudret B., and Lemaire M., 2006a. Stochastic finite elements: a non intrusive approach by regression. *Eur. J. Comput. Mech.* 15 (1-3), 81-92.
- Berveiller M., Sudret B., and Lemaire M., 2006b. Stochastic finite elements: a non intrusive approach by regression". In: *Eur. J. Comput. Mech.* 15 (1-3), pp. 81-92.
- Bierkens, J. 2016. Non-reversible Metropolis-Hastings. *Statistics and Computing*, 26 1213-1228.
- Bierkens J., Fearnhead P. and Roberts G. 2016. The zig-zag process and super-efficient sampling for Bayesian analysis of big data. arXiv preprint arXiv :1607.03188.

Bierkens, J., Bouchard-Côté, A., Doucet, A., Duncan, A. B., Fearnhead, P., Roberts, G. and Vollmer S. J., 2017. Piecewise deterministic Markov processes for scalable Monte Carlo on restricted domains. arXiv preprint arXiv :1701.04244.

Bishop C. M. 1995. Neural Networks for Pattern Recognition. Advanced Texts in Econometrics. Oxford, UK: Oxford University Press.

Bjorck A., 1996. Numerical methods for least squares problems. SIAM Press, Philadelphia, PA.

Blatman G. and Sudret B. 2008. Polynomial chaos expansions and adaptive stochastic finite elements using a regression approach". In : Comptes Rendus Mécanique 336 (6), pp. 518-523.

Blatman, G. and B. Sudret 2008b. Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach. Comptes Rendus Mécanique 336 (6), 518-523.

Blatman G. and Sudret B. 2009. Anisotropic parcimonious polynomial chaos expansions based on the sparsity-of-effects principle. In Proc. ICOSSAR'09, Int Conf. on Structural Safety And Reliability, Osaka, Japan.

Blatman G. and Sudret B., 2010a. An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis". In : Prob. Eng. Mech. 25 (2), pp. 183-197.

Blatman G. and Sudret B., 2010b. Efficient computation of global sensitivity indices using sparse polynomial chaos expansions". In : Reliab. Eng. Sys. Safety 95, pp. 1216-1229.

Blatman G. and Sudret B., 2011. Adaptive sparse polynomial chaos expansion based on least angle regression. In : Journal of Computational Physics 230.6, pp. 2345–2367.

Blum J. 1999. Problèmes inverses : de l'expérimentation à la modélisation. OFTA, ARAGO.

Box G. E. P.; Draper N. R. 2007. Response Surfaces, Mixtures, and Ridge Analyses. 2nd ed. Wiley Series in Probability and Statistics. Hoboken, New Jersey, USA : John Wiley & Sons, Inc.

Bradley M.M., Greenwald M.K., Petry M.C., and Lang P. J., 1992. Remembering Pictures. Pleasure and arousal in memory. Journal of Experimental Psychology : Learning Memory and Cognition, 18, 379 – 390.

Cacuci, D. 2003. Sensitivity and uncertainty analysis: theory. Vol. 1. Chapman & Hall/CRC, Boca Raton.

Cameron R., and Martin W., 1947. The Orthogonal Development of Non-Linear Functionals in Series of Fourier-Hermite Functionals. *Annals of Mathematics*, 48(2), second series.

Caniou, Y., Defaux G., Dubourg V., Sudret B., and Petit G. 2011. Caractérisation indirecte de défauts géométriques de forme liés au process de fabrication d'un élément d'essuie-glace. In 20ème Congrès Français de Mécanique.

Carrera J. and Neuman S.P. 1986. Estimation of Aquifer Parameters under Transient and Steady State Conditions I: Maximum Likelihood Method Incorporating Prior Information. *Water Resources Research*, 22, 199-210.

Cartalade A., Montarnal P., Cavanna B., and Blum J., (2003). Estimation des paramètres de transport d'un milieu poreux, approche par état adjoint. Rapport DM2S/SFME/MTMS/RT/03-002/A. CEA/Saclay.

Cartalade A. and al., 2005. Application of inverse modeling methods to thermal and diffusion experiments at mont terri rock laboratory. Proc. of the 2nd international meeting : Clays in Natural and Engineered Barriers for Radioactive waste confinement, Tours.

Chavent G., 1974. Identification of Functional Parameters in Partial Differential Equations. New York : American society of mechanical engineers, 31-48.

Chavent G., Jaffré J., and Roberts J. E., 1995. Mixed-hybrid finite elements and cell-centered finite volumes for two-phase flow in porous media. In A. Bourgeat, C. Carasso, S. Luckhaus, and A. Mikelic, editors, *Mathematical Modelling of Flow Through Porous Media*, pages 100–114, World Scientific, Singapore.

Chen, M., D. Zhang, A. A. Keller, and Lu Z., 2005. A stochastic analysis of steady state two-phase flow in heterogeneous media, *Water Resour. Res.*, 41, W01006.

Chen M., Keller A.A., Zhang D., Lu Z., and Zyvoloski G.A., 2006. A stochastic analysis of transient two-phase flow in heterogeneous porous media, *Water Resour. Res.*, 42, W03425.

Chen T., Fox E. and Guestrin C., 2014. Stochastic gradient hamiltonian monte carlo 1683-1691.



Choi S., Grandhi R., Canfield R., and Pettit C., 2004. Polynomial chaos expansion with Latin Hypercube sampling for estimating response variability. *AIAA Journal* 45, 1191-1198.

Clément F., Khvoenkova N., Cartalade A., and Montarnal P., 2004. Analyse de sensibilité et estimation de paramètres de transport pour une équation de diffusion, approche par état adjoint. INRIA, Rapport n°5132.

Cukier, R. 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I theory. *J. Chem. Phys.* 59 (8), pp. 3873-3878.

Cukier R.I., Fortuin C.M., Shuler K.E., Petschek A.G., and Schaibly J.H., 1973. Study of the Sensitivity of Coupled Reaction Systems to Uncertainties in Rate Coefficients. I. Theory. *The Journal of Chemical Physics.* 59. 3873-3878.

Cukier R., Levine H. ; Shuler K. 1978. Nonlinear sensitivity analysis of multiparameter model systems. *J. Comput. Phys.* 26, pp. 1-42.

De Rocquigny É., 2006. La maîtrise des incertitudes dans un contexte industriel. 1re partie : une approche méthodologique globale basée sur des exemples. *Journal de la société française de statistique*, Tome 147 no. 3, p. 33-71

De Rocquigny É., 2006. La maîtrise des incertitudes dans un contexte industriel. 2nde partie : revue des méthodes de modélisation statistique physique et numérique. *Journal de la société française de statistique*, Tome 147 no. 3, p. 73-106

Delay F., Badri H., Fahs M. et Ackerer Ph., 2017. A comparison of discrete versus continuous adjoint states to invert groundwater flow in heterogeneous dual porosity systems. *Advances in Water Resources*, 110 1-18.

Dempster, A.P., 1967. Upper and lower probabilities induced by a multivalued mapping". *The Annals of Mathematical Statistics.* 38 (2) : 325–339.

Ding N., Fang Y., Babbush R., Chen C., Skeel R. D. and Neven H., 2014. Bayesian sampling using stochastic gradient thermostats 3203-3211.

Efroymson M. A., 1960. Multiple Regression Analysis. In *Mathematical Methods for Digital Computers*, Edited by: Ralston, A. and Wilf, H. S. 191–203. New York: John Wiley.

Efron, B. ; Stein C. 1981. The Jackknife estimate of variance. *Annals Statist.* 9(3), pp. 586-596.

Efron B., Hastie T., Johnstone I. and Tibshirani R., 2004. Least angle regression. In: *Ann. Stat.* 32 (2), pp. 407-499.

Eglajs, V. and Audze P. 1977. "New approach to the design of multifactor experiments". *Problems of Dynamics and Strengths.* 35 (in Russian). Riga : Zinatne Publishing House : 104–107.

Fajraoui N., Ramasomanana F., Younes A., Mara T.A., Ackerer P. and Guadagnini, A. 2011. Use of global sensitivity analysis and polynomial chaos expansion for interpretation of nonreactive transport experiments in laboratory-scale porous media, *Water Resour. Res.*, 47.

Ferson S. and Ginzburg L.R., 1996. Different methods are needed to propagate ignorance and variability, *Reliability Engineering & System Safety*, Volume 54, Issues 2–3, Pages 133-144.

Ferson, S., Kreinovich V., Ginzburg L., Myers D.S., and Sentz K., 2003. Constructing Probability Boxes and Dempster–Shafer Structures Archived 2011-07-22 at the Wayback Machine. Sandia National Laboratories, SAND2002-4015, Albuquerque, NM.

Ferson, S. and Hajagos, J. G., 2004. Arithmetic with uncertain numbers: rigorous and (often) best possible answers. *Reliab. Eng. Sys. Safety*, 85(1-3), 135–152.

Forrester, A. ; Sobester, A. ; Keane, A 2008. *Engineering design via surrogate modelling: a practical guide.* Wiley.

Fort G., Jourdain B., Kuhn E., Lelièvre T. and Stoltz G., 2014. Convergence of the Wang-Landau algorithm. Accepted for publication in *Mathematics of Computation*. arXiv math.PR 1207.6880.

Fort G., Jourdain B., Kuhn E., Lelièvre T. and Stoltz G., 2014. Efficiency of the Wang-Landau algorithm Accepted for publication in *Applied Mathematics Research Express*. arXiv 1310.6550

Geisser S., 1975. The predictive sample reuse method with applications. *J. Amer. Stat. Assoc.* 70, 320-328.

Gelman, A. G. and D. B. Rubin 1992 Inference from iterative simulation using multiple sequences, *Stat. Sci.*, 7, 457–472.

Geman S. and Geman D., 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 6 (6): 721–741.

Ghanem R. and Spanos P., 1991. Stochastic finite éléments. A spectral approach. Springer Verlag.

Gilks W.R., Roberts G.O., and Sahu S.K., 1998. Adaptive Markov chain Monte Carlo through regeneration. *Journal of the American Statistical Association*, 93(443) :1045–1054.

Görür D. and Teh Y. W. 2011. Concave-Convex Adaptive Rejection Sampling". *Journal of Computational and Graphical Statistics*. 20 (3): 670–691.

Haario H., Saksman E. and Tamminen J., 2001. An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242.

Haario H., Saksman E. and Tamminen J., 2005. Componentwise adaptation for high dimensional MCMC, *Stat. Comput.*, 20, 265–274.

Haario, H., Laine M., Mira A. and Saksman E., 2006. DRAM : Efficient adaptive MCMC, *Stat. Comp.*, 16, 339–354.

Hadamard J. 1952. Lectures on Cauchy's problem in linear partial differential equations, New Haven, CT : Yale University Press.

Hastie T., Tibshirani R. and Friedman J., 2001. *The Elements of Statistical Learning*. Springer Series in Statistics Springer New York Inc., New York, NY, USA.

Hastings W.K., 1970. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*. 57 (1) : 97–109.

Helton, J. ; Johnson J. ; Sallaberry C. ; Storlie C. 2006. Survey of sampling-based methods for uncertainty and sensitivity analysis". *Reliab. Eng. Sys. Safety* 91 (10-11), pp. 1175-1209.

Higdon D., Gattiker J., Williams B. and Rightley M., 2008. Computer Model Calibration Using High-Dimensional Output, *Journal of the American Statistical Association*, 103 :482, 570-583,

Hoeksema R. J. and Kitanidis P. K., 1984. An Application of the Geostatistical Approach to the Inverse Problem in Two-Dimensional Groundwater Modeling. *Water Resources Research*, 20(7), 1003-1020.

Homma T. and Saltelli A., 1996. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*. V. 52, Issue 1, Pages 1-17.

Iman, R.L. ; Davenport, J.M. ; Zeigler, D.K. 1980. Latin hypercube sampling (program user's guide). OSTI 5571631.

Iman, R.L. ; Helton, J.C. ; Campbell, J.E. 1981. "An approach to sensitivity analysis of computer models, Part 1. Introduction, input variable selection and preliminary variable assessment". *Journal of Quality Technology*. 13 (3): 174–183.

Iman, R.L. ; Conover, W.J. 1982. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics - Simulation and Computation*, 11(3):311-334

Iman, R.L. ; Davenport, J.M. 1982 Rank correlation plots for use with correlated input variables. *Communications in Statistics - Simulation and Computation*, 11(3):335-360

Iskrev N., 2010. Local identification in dsge models. *Journal of monetary economics* ; 57 :189-202.

Isukapalli S.S., 1999. Uncertainty Analysis of Transport-Transformation Models. Ph. D. thesis, The State University of New Jersey.

Joe, S. and Kuo, F.Y., 2008. Constructing Sobol' sequences with better two-dimensional projections. *SIAM Journal on Scientific Computing*, 30(5), 2635-2654.

Kayshap R.L. 1982. Optimal choice of AR and MA parts in autoregressive moving average models, *IEEE Trans. Pattern Anal. Mach. Intell.* 4 (2) 99–104.

Kennedy M.C., O'Hagan A. 2001, Bayesian calibration of computer models, *Journal of the Royal Statistical Society, Series B* Volume 63, Issue 3, pages 425–464.

Kern M. 2008. Problèmes inverses : Techniques de l'Ingénieur, to appear, TSI.

Koehler J.R. and Owen A.B., 1996. 9 Computer experiments, Handbook of Statistics, Elsevier, Volume 13, Pages 261-308.

Kool, J.B., Parker J.C., and van Genuchten M.Th., 1987. Parameter estimation for unsaturated flow and transport models. A review. Journal of Hydrology. 91 :255–293.

Krige, D.G., 1951. A Statistical Approaches to Some Basic Mine Valuation Problems on the Witwatersrand. Journal of the Chemical, Metallurgical and Mining Society of South Africa, 52, 119-139.

Laloy E. and Vrugt J.A., 2012. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and highperformance computing. Water Resour Res 48 : W01526.

Laloy, E., Fasbender D. and Bielders C.L., 2010. Parameter optimization and uncertainty analysis for plot-scale continuous modeling of runoff using a formal Bayesian approach, J. Hydrol., 380(1–2), 82–93.

Laughlin D. Mc. and Townley L.R., 1996. A reassessment of the groundwater inverse problem. Water Resources Research., 32 (5) :1131–1161.

Le Maître O.P., Reagan M.T., Najm H.N., Ghanem R.G. and Knio O.M., 2002. A Stochastic Projection Method for Fluid Flow : II. Random Process. In : Journal of Computational Physics 181.1, pp. 9–44.

Le Maître O. P. ; Knio O. M. 2010. Spectral Methods for Uncertainty Quantification : With Applications to Computational Fluid Dynamics. Scientific Computation. Dordrecht, Netherlands: Springer.

Lehman F. and Ackerer Ph., 1997. Determining soil hydraulic properties by inverse method in one-dimensional unsaturated flow. Journal of Environmental Quality. 26, p. 76-81.

Levenberg K., 1944. “A Method for the Solution of Certain Non-Linear Problems in Least Squares”. The Quarterly of Applied Mathematics, 2: 164-168.

Liu J. S., Liang F. and Wong W. H., 2000. The multiple-try method and local optimization in Metropolis sampling. J. American Statist. Assoc., 95 121-134.

Lourakis M.I.A., 2005. A brief description of the Levenberg-Marquardt algorithm implemented by levmar, Technical Report, Institute of Computer Science, Foundation for Research and Technology - Hellas.

Madsen K., Nielsen N.B., and Tingleff O., 2004. Methods for nonlinear least squares problems. Technical Report. Informatics and Mathematical Modeling, Technical University of Denmark.

Mangasarian, O.L., 1969. Nonlinear Programming. McGraw-Hill, New York.

Mara, T. A. 2009. Extension of the RBD-FAST method to the computation of global sensitivity indices. Reliab. Eng. Sys. Safety 94 (8), pp. 1274-1281.

Marrel A., Iooss B., Van Dorpe F. and Volkova E., 2008. An efficient methodology for modeling complex computer codes with Gaussian processes. Computational Statistics and Data Analysis, Elsevier, 52, pp.4731-4744.

Martino L. and Míguez J., 2010. A generalization of the adaptive rejection sampling algorithm". Statistics and Computing. 21 (4): 633–647.

Martino L., Read J. and Luengo D., 2015. Independent Doubly Adaptive Rejection Metropolis Sampling Within Gibbs Sampling. IEEE Transactions on Signal Processing. 63 (12): 3123–3138.

Martino L., 2018. A Review of Multiple Try MCMC algorithms for Signal Processing. ArXiv e-prints. 1801.09065.

Marquardt D.W., 1963. "An algorithm for least-squares estimation of nonlinear parameters," Journal of the Society for Industrial and Applied Mathematics, 11(2):431-441.

Matheron, G., 1962. Trait´e de g´eostatistique appliqu´ee, Tome I: M´emoires du Bureau de Recherches G´eologiques et Mini`eres, no. 14, Editions Technip, Paris, 333 p.

Matheron G., 1967. Analyse de la notion de hasard. Journal de la soci´et´e fran¸caise de statistique, Tome 108 pp. 289-290.

Matthies, H. and Keese A., 2005. Galerkin methods for linear and nonlinear elliptic stochastic Partial differential equations. Comput. Methods Appl. Mech. Engrg. 194, 1295-1331.

McKay, M.D. ; Beckman, R.J. ; Conover, W.J. 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code.

McKay M. 1995. Evaluating prediction uncertainty. Tech. rep. NUREG/CR-6311. Los Alamos National Laboratory.

Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H. and Teller E., 1953. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, Volume 21, Issue 6, p.1087-1092.

Meyer R., Cai B. and Perron F., 2008. Adaptive rejection Metropolis sampling using Lagrange interpolation polynomials of degree 2. *Computational Statistics & Data Analysis*. 52 (7): 3408–3423.

Montgomery D., 2004. *Design and analysis of experiments*. John Wiley and Sons, New York.

Myers R. H. ; Montgomery D. C. ; and Anderson-Cook C. M. 2016. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. 4th ed. Wiley Series in Probability and Statistics. Hoboken, New Jersey, USA : John Wiley & Sons, Inc., 2016.

Nelson R. W., 1960. In-place measurement of permeability in heterogeneous media : 1. Theory of a proposed metho. *Journal of Geophysical Research*, 65(6), 1753-1760.

Nelson R. W., 1961. In-place measurement of permeability in heterogeneous media : 2. Experimental and computational considerations. *Journal of Geophysical Research*, 66, 2469-2478.

Neuman S. P., 2003. Maximum likelihood Bayesian averaging of uncertain model predictions, *Stochastic Environmental Research and Risk Assessment* 17 (5) 291–305.

Oakley J.E., and O'Hagan A., 2004. Probabilistic Sensitivity Analysis of Complex Models: A Bayesian Approach. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66(3), 751-769.

O'Hagan A., 2006. Bayesian analysis of computer code outputs : A tutorial, *Reliability Engineering & System Safety*, Volume 91, Issues 10–11, Pages 1290-1300, A.

O'Hagan, C. E. Buck, A. Daneshkhah, J. R. Eiser, Garthwaite, Jenkinson D. J., Oakley J. E., and Rakow T., 2006. *Uncertain Judgements: Eliciting Experts' Probabilities*. Statistics in Practice. Chichester, West Sussex, UK : John Wiley & Sons, Ltd.

Pawitan Y., 2001. In *All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford, UK: Oxford University Press.

Rasmussen C. E. ; Williams C. K. I. 2006. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. Cambridge, Massachusetts, USA: The MIT Press.

Ripley B. D. 1996. *Pattern Recognition and Neural Networks*. Cambridge, UK: Cambridge University Press.

Riva M., Guadagnini A., and Dell'Oca A., 2015. Probabilistic assessment of seawater intrusion under multiple sources of uncertainty, *Adv. Water Res.*, 75, 93–104.

Roberts G.O., Gelman A., and Gilks W.R., 1997. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1) :110–120.

Roberts G.O. and Rosenthal J.S., 2001. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367.

Rohde C. A., 2014. *Introductory Statistical Inference with the Likelihood Function*. Cham, Switzerland: Springer International Publishing.

Rubin Y. and Dangan G., 1987. Stochastic identification of transmissivity and effective recharge in steady groundwater flow : 1. Theory., *Water Resources Research*, 23(7), 1185-1192.

Runge C. 1901. About empirical functions and the interpolation between equidistant ordinates, *Journal of Mathematics and Physics*, 46: 224–243.

Sacks, J., Welch W.J., Mitchell T.J. and Wynn H.P., 1989. Design and Analysis of Computer Experiments. *Statist. Sci.* 4, no. 4, 409--423.

Saksman E. and Vihola M., 2010. On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *The Annals of Applied Probability*, 20(6):2178–2203.

Saltelli A., Sobol' I.M., 1995. Sensitivity analysis for nonlinear mathematical models: numerical experience”, *Matem. Mod.*, 7:11, 16–28.



Saltelli, A. ; Tarantola S. ; Chan K. 1999. A quantitative, model independent method for global sensitivity analysis of model output. *Technometrics* 41 (1), pp. 39-56.

Saltelli A. ; Chan K. ; Scott E. 2000a. *Sensitivity analysis*. J. Wiley & Sons.

Saltelli A. ; Tarantola S. ; Campolongo F. 2000b. Sensitivity analysis as an Ingredient of modeling. *Statis. Sci.* Vol. 15. 4, 377-395.

Saltelli A., Ratto M., Tarantola S. and Campolongo F., 2006. Sensitivity analysis practices : Strategies for model-based inference. *Reliability Engineering and System Safety* 91 (2006) 1109–1125

Santner T. J. ; Williams B. J. ; Notz W. I. 2003. *The Design and Analysis of Computer Experiments*. Springer Series in Statistics. New York: Springer.

Satterthwaite F., 1959. Random balance experimentation. *Technometrics* 1, 11.

Schaibly J.H. and Shuler, K.E., 1973. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. II Applications. *The Journal of Chemical Physics*.

Scharnagl, B., Vrugt J.A., Vereecken H., and Herbst M., 2010. Information content of incubation experiments for inverse modeling of carbon pools in the Rothamsted model : A Bayesian approach, *Biogeosciences*, 7, 763–776.

Shao Q., Younes A., Fahs M., Mara T.A., 2017. Bayesian sparse polynomial chaos expansion for global sensitivity analysis. *Comput. Methods Appl. Mech. Engrg.*

Shafer G., 1976. *A Mathematical Theory of Evidence*, Princeton University Press.

Simpson, T. W. ; Poplinski, J. D. ; Koch, P. N. ; Allen, J. K. 2001. Metamodels for computerbased engineering design : survey and recommendations. *Eng. Comput.*, 17(2):129–150.

Smola A.J. and Schölkopf B., 2004. A tutorial on support vector regression. *Statistics and Computing* 14, 3 ,199-222.

Sobol' I.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals”, *Zh. Vychisl. Mat. Mat. Fiz.*, 7:4, 784–802; *U.S.S.R. Comput. Math. Math. Phys.*, 7:4 , 86–112.

Sobol' I.M., 1993. Sensitivity estimates for nonlinear mathematical models. *Math. Modeling and Comp. Exp.* 1, pp. 407-414.

Sobol' I.M., 1998. On quasi-Monte Carlo integrations. *Mathematics and Computers in Simulation*. Vol. 47, Issues 2–5, Pages 103-112.

Sobol' I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* 55 (1-3), pp. 271-280.

Sobol' I.M., Kucherenko S., 2005. Global sensitivity indices for nonlinear mathematical models. *Review*.

Sobol' I.M., Kucherenko S., 2009. Derivative based global sensitivity measures and their link with global sensitivity indices. *Math. Comput. Simul.* 79 (10), pp. 3009-3017.

Stone M., 1974. Cross-validators choice and assessment of statistical predictions. *J. Royal Stat. Soc., Series B* 36, 111-147.

Storlie C.B., Swiler L.P., Helton J.C., Sallaberry C.J., 2009. Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability Engineering and System Safety*, 94, 1735-1763.

Sudret, B., Delfaux G. and Pendola M., 2004. Finite element reliability analysis of cooling towers submitted to degradation. In : *Proc. 2nd Int. ASRANet Colloquium, Barcelona, Spain*.

Sudret, B., 2006. Global sensitivity analysis using polynomial chaos expansions. In : *Proc. 5th Int. Conf. on Comp. Stoch. Mech (CSM5), Rhodos, Greece*. Ed. by P. Spanos & G. Deodatis.

Sudret, B., 2007. Uncertainty propagation and sensitivity analysis in mechanical models – Contributions to structural reliability and stochastic spectral methods. *Université Blaise Pascal, Clermont-Ferrand, France. Habilitation à diriger des recherches*, 173 pages.

Sudret, B., 2008. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, Vol. 93, Issue 7, Pages 964-979.

Sun N.Z., 1994. *Inverse Problems in Groundwater Modelling*. Kluwer Academic Publishers, Dordrecht, 337.

Sun N. Z., 1995. *Inverse Problems in Groundwater Modeling*. Kluwer Academic Publishers, the Netherland.

Tarantola S., Gatelli D. and Mara T.A., 2006. Random balance designs for the estimation of first order global sensitivity indices. *Reliability Engineering and System Safety*, Elsevier, 91 (6), pp.717-727.

Tatang, M., Pan W., Prinn R. and McRae G., 1997. An efficient method for parametric uncertainty analysis of numerical geophysical models. In: *J. Geophys. Research* 102 (D18), pp. 21925-21932.

ter Braak C.J.F., 2006. A Markov chain Monte Carlo version of the genetic algorithm differential evolution: Easy Bayesian computing for real parameter space, *Stat. Comput.*, 16(3), 239–249.

ter Braak, C.J.F., and J. A. Vrugt 2008. Differential evolution Markov chain with snooker updater and fewer chains, *Stat. Comput.*, 18(4), 435–446.

Tikhonov A.N., 1943. On the stability of inverse problems, *Dokl. Akad. Nauk SSSR*, 39, No. 5, 195-198

Tikhonov A.N., 1963. Solution of incorrectly formulated problems and the regularization method, *Soviet Math Dokl* 4, 1035-1038 English translation of *Dokl Akad Nauk SSSR* 151, 1963, 501-504

Todor R.A. and Schwab C. 2007. Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. *IMA J. Numer. Anal* 27, 232-261.

Vanderbei R.J., 1997. *Linear Programming*. International Series in Operations Research & Management Science.

Vapnik V., Steven E.G. and Smola A., 1997. Support vector method for function approximation, regression estimation and signal processing. *International Conference on Neural Information Processing Systems*. Pages 281-287.

Vazquez E. and Walter E., 2003. Multi-output support vector regression. In: *Proc. 13th IFAC Symposium on System Identification*, Rotterdam, pp. 1820-1825.

Vazquez E., 2005. Modélisation comportementale de systèmes non-linéaires multivariables par méthodes à noyaux et applications. Automatique / Robotique. Université Paris Sud - Paris XI.

Vihola M., 2009. On the stability and ergodicity of an adaptive scaling Metropolis algorithm. Arxiv preprint arXiv : 0903.4061.

Vrugt J. A., Bouten W. and Weerts A. H., 2001. Information Content of Data for Identifying Soil Hydraulic Parameters from Outflow Experiments. Soil Science Society America Journal ; 65 : 19-27

Vrugt J. A., Gupta H. V., Bouten W. and Sorooshian S., 2003. A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, Water Resour. Res., 39(8).

Vrugt, J.A., ter Braak C.J.F., Clark M.P., Hyman J.M., and Robinson B.A., 2008. Treatment of input uncertainty in hydrologic modeling : Doing hydrology backward with Markov chain Monte Carlo simulation, Water Resour. Res., 44, W00B09.

Vrugt J.A., ter Braak C.J.F., Diks C.G.H., Higdon D., Robinson B. A. and Hyman J. M., 2009a. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, Int. J. Nonlinear Sci. Numer. Simul., 10(3), 273–290.

Vrugt J.A., ter Braak C.J.F., Gupta H.V., Robinson B.A. 2009b. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling Stoch Environ Res Risk Assess. 23(7):1011–1026.

Walley P., 1996. Inferences from multinomial data : learning about a bag of marbles. Journal of the Royal Statistical Society, Series B 58: 3–57.

Wan X. and Karniadakis G. E., 2005. An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. J. Comput. Phys. 209, 617-642.

Wan X. and Karniadakis G. E., 2006. Beyond Wiener-Askey expansions : handling arbitrary PDFs. J. Sci. Comput. 27, 455- 464.

Wang X. and Dunson D., 2013. Parallelizing MCMC via weierstrass sampler. arXiv preprint arXiv :1312.4605.

Wang X. Guo F. Heller K. and Dunson D., 2015. Parallelizing MCMC with random partition trees. *Advances in Neural Information Processing Systems* 451-459.

Welch W.J., Robert J.B., Sacks J., Wynn H., Mitchell T., and Morris M., 1992. Screening, Predicting, and Computer Experiments. *Technometrics*.

Welling M. and Te Y., 2011. Bayesian learning via stochastic gradient Langevin dynamics 681-688.

Wiener N., 1938. The Homogeneous Chaos. *American Journal of Mathematics*, Vol. 60, No. 4, pp. 897-936.

Witteveen, J., Sarkar S., and Bijl H., 2007. Modeling physical uncertainties in dynamic stall Induced fluid-structure interaction of turbine blades using arbitrary polynomial chaos. *Computers & Structures* 85, 866-878.

Xiu D. and Karniadakis G. E., 2002. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J. Sci. Comput.* 24 (2), 619-644.

Xiu, D., Lucor D., Su C.H. and Karniadakis G., 2003. Performance evaluation of generalized polynomial chaos. In *Int. Conf. Comput. Science*, pp. 346-354.

Xiu D. 2010. *Numerical Methods for Stochastic Computations : A Spectral Method Approach*. Princeton, New Jersey, USA : Princeton University Press.

Xu C., and Gertner G.Z., 2008. A general first-order global sensitivity analysis method, *Reliability Engineering and System Safety*, 93, issue 7, p. 1060-1071.

Yuan M., Joseph V. R., and Lin Y., 2007. An efficient variable selection approach for analyzing designed experiments. *Technometrics*, 49(4) :430-439.

## Tables des tableaux

Tableau 1.1 : Valeurs moyenne des paramètres du modèle van Genuchten – Mualem pour douze classes de sols.....	13
---------------------------------------------------------------------------------------------------------------	----

# Chapitre 3. Analyse de sensibilité globale et estimation bayésienne des paramètres pour le potentiel spontané en milieu poreux non saturé

## Plan du chapitre

Chapitre 3. Analyse de sensibilité globale et estimation bayésienne des paramètres pour le potentiel spontané en milieu poreux non saturé .....	101
3.1 Introduction .....	102
3.2 Description du cas de test et solution numérique .....	104
3.2.1 Description du cas de test .....	104
3.2.2 Modèle mathématique .....	104
3.2.3 Modèle numérique .....	104
3.3 Analyse de sensibilité globale des signaux de SP .....	105
3.3.1 La méthode GSA .....	105
3.3.2 Résultats GSA .....	106
3.4 Estimation des paramètres .....	108
3.4.1 Approches MCMC et FOA .....	108
3.4.2 Résultats d'estimation des paramètres .....	109
3.5 Conclusion .....	112



## Sensitivity and identifiability of hydraulic and geophysical parameters from streaming potential signals in unsaturated porous media

Anis Younes<sup>1,2,3</sup>, Jabran Zaouali<sup>1</sup>, François Lehmann<sup>1</sup>, and Marwan Fahs<sup>1</sup>

<sup>1</sup>LHyGES, Université de Strasbourg/EOST/ENGEES, CNRS, 1 rue Blessig, 67084 Strasbourg, France

<sup>2</sup>LISAH, Univ. Montpellier, INRA, IRD, SupAgro, Montpellier, France

<sup>3</sup>LMHE, ENIT, Tunis, Tunisia

**Correspondence:** Marwan Fahs (fahs@unistra.fr)

Received: 13 December 2017 – Discussion started: 24 January 2018

Revised: 7 May 2018 – Accepted: 11 June 2018 – Published: 2 July 2018

**Abstract.** Fluid flow in a charged porous medium generates electric potentials called streaming potential (SP). The SP signal is related to both hydraulic and electrical properties of the soil. In this work, global sensitivity analysis (GSA) and parameter estimation procedures are performed to assess the influence of hydraulic and geophysical parameters on the SP signals and to investigate the identifiability of these parameters from SP measurements. Both procedures are applied to a synthetic column experiment involving a falling head infiltration phase followed by a drainage phase.

GSA is used through variance-based sensitivity indices, calculated using sparse polynomial chaos expansion (PCE). To allow high PCE orders, we use an efficient sparse PCE algorithm which selects the best sparse PCE from a given data set using the Kashyap information criterion (KIC). Parameter identifiability is performed using two approaches: the Bayesian approach based on the Markov chain Monte Carlo (MCMC) method and the first-order approximation (FOA) approach based on the Levenberg–Marquardt algorithm. The comparison between both approaches allows us to check whether FOA can provide a reliable estimation of parameters and associated uncertainties for the highly nonlinear hydrogeophysical problem investigated.

GSA results show that in short time periods, the saturated hydraulic conductivity ( $K_s$ ) and the voltage coupling coefficient at saturation ( $C_{sat}$ ) are the most influential parameters, whereas in long time periods, the residual water content ( $\theta_s$ ), the Mualem–van Genuchten parameter ( $n$ ) and the Archie saturation exponent ( $n_a$ ) become influential, with strong interactions between them. The Mualem–van Genuchten pa-

rameter ( $\alpha$ ) has a very weak influence on the SP signals during the whole experiment.

Results of parameter estimation show that although the studied problem is highly nonlinear, when several SP data collected at different altitudes inside the column are used to calibrate the model, all hydraulic ( $K_s, \theta_s, \alpha, n$ ) and geophysical parameters ( $n_a, C_{sat}$ ) can be reasonably estimated from the SP measurements. Further, in this case, the FOA approach provides accurate estimations of both mean parameter values and uncertainty regions. Conversely, when the number of SP measurements used for the calibration is strongly reduced, the FOA approach yields accurate mean parameter values (in agreement with MCMC results) but inaccurate and even unphysical confidence intervals for parameters with large uncertainty regions.

### 1 Introduction

Flow through a charged porous medium can generate an electric potential (Zablocki, 1978; Ishido and Mizutani, 1981; Allègre et al., 2010; Jougnot and Linde, 2013), called streaming potential (SP). SP signals play an important role in several applications related to hydrogeology and geothermal reservoir engineering as they are useful for examining subsurface flow dynamics. During the last decade, surface SP anomalies have been widely used to estimate aquifers' hydraulic properties (Darnet et al., 2003). Interest in SP is motivated by its low cost and its high sensitivity to water flow. Either



coupled or uncoupled approaches can be used for hydraulic parameter estimation from SP signals (Mboh et al., 2012). In the uncoupled approach, Darcy velocities (e.g., Jardani et al., 2007; Bolève et al., 2009) are obtained from the tomographic inversion of SP signals and are then used for the calibration of the hydrologic model. In the coupled approach, anomalies related to the tomographic inversion are avoided by inverting the full coupled hydrogeophysical model (Hinnell et al., 2010).

The SP signals have been widely studied in saturated porous media (Bogoslovsky and Ogilvy, 1973; Patella, 1997; Sailhac and Marquis, 2001; Richards et al., 2010; Bolève et al., 2009, among others). Fewer studies focused on the application of the SP signal in unsaturated flow despite the large interest in such nonlinear problems (Linde et al., 2007; Allègre et al., 2010; Mboh et al., 2012; Jougnot and Linde, 2013). Hence, in this work we are interested in the SP signals in unsaturated porous media. Our main objective is to investigate the usefulness of the SP signals for the characterization of soil parameters. For this purpose, we evaluate the impact of uncertain hydraulic and geophysical parameters on the SP signals and assess the identifiability of these parameters from the SP measurements.

The impact of soil parameters on SP signals is investigated using global sensitivity analysis (GSA). This is a useful tool for characterizing the influential parameters that contribute the most to the variability of model outputs (Saltelli et al., 1999; Sudret, 2008) and for understanding the behavior of the modeled system. GSA has been applied in several areas, for risk assessment for groundwater pollution (e.g., Volkova et al., 2008), nonreactive (Fajraoui et al., 2011) and reactive transport experiments (Fajraoui et al., 2012; Younes et al., 2016), for unsaturated flow experiments (Younes et al., 2013), natural convection in porous media (Fajraoui et al., 2017) and seawater intrusion (Rajabi et al., 2015; Riva et al., 2015). To the best of our knowledge, GSA has never been used for SP signals in unsaturated porous media. Hence, in the first part of this study, GSA is performed on a conceptual model inspired from the laboratory experiment of Mboh et al. (2012) in which SP signals are measured at different altitudes in a sandy soil column during a falling-head infiltration phase followed by a drainage phase. Four uncertain hydraulic parameters, saturated hydraulic conductivity ( $K_s$ ), residual water content ( $\theta_r$ ), fitting Mualem–van Genuchten parameters ( $\alpha, n$ ) and two geophysical parameters (Archie's saturation exponent ( $n_a$ ) and voltage coupling coefficient at saturation ( $C_{sat}$ )), are investigated. GSA of SP signals is performed by computing the variance-based sensitivity indices using polynomial chaos expansion (PCE). To reduce the number of PCE coefficients while maintaining high PCE orders, we use the efficient sparse PCE algorithm developed by Shao et al. (2017) which selects the best sparse PCE from a given data set using the Kashyap information criterion (KIC).

In the second part of this study, we investigate the identifiability of hydrogeophysical parameters from SP measure-

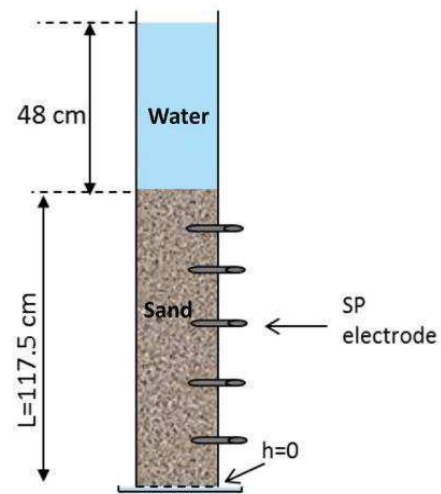


Figure 1. Illustration of the experimental device.

ments. For this purpose, parameter estimation is performed using two different approaches. The first is a Bayesian approach in which model parameters are treated as random variables and characterized by their probability density functions. With this approach, the prior knowledge about the model and the observed data is merged to define the joint posterior probability distribution function of the parameters. In the sequel, Bayesian analysis is conducted using the DREAM<sub>(ZS)</sub> software (Laloy and Vrugt, 2012; Vrugt, 2016) based on the Markov chain Monte Carlo (MCMC) method. MCMC has been successfully used in various inverse problems (e.g., Vrugt et al., 2003, 2008; Arora et al., 2012; Younes et al., 2017). The MCMC method yields an ensemble of possible parameter sets that satisfactorily fit the available data. These sets are then employed to estimate the posterior parameter distributions and hence the optimal parameter values and the associated 95 % confidence intervals (CIs) in order to quantify the parameters' uncertainty. The second inversion approach is the commonly used first-order approximation (FOA) approach based on the standard Levenberg–Marquardt algorithm. Two scenarios are considered to check whether FOA can provide a reliable estimation of parameters and associated uncertainties for the highly nonlinear hydrogeophysical problem investigated in the case of abundant data (small uncertainty regions) and in the case of scarcity of data (large uncertainty regions). In the first scenario, SP data collected from sensors at five different locations are taken into account for the calibration. In the second scenario, only the SP data from one sensor are used for model calibration.

The present study is set out as follows. Section 2 presents the hydrogeophysical model and the reference solution. Section 3 reports on the GSA results of SP signals. Then, Sect. 4 discusses results of parameter estimation with both MCMC and FOA approaches for the two investigated scenarios.

## 2 Test case description and numerical solution

### 2.1 Test case description

The test case considered in this work is similar to the laboratory experiment developed in Mboh et al. (2012), involving a falling-head infiltration phase followed by a drainage phase (Fig. 1). This experiment is representative of several laboratory SP experiments (Linde et al., 2007; Allègre et al., 2010; Jougnot and Linde, 2013, among others). Quartz sand is evenly packed in a plastic tube with an internal diameter of 5 cm to a height of  $L_s = 117.5$  cm. The column is initially saturated with a ponding of  $L_w = 48$  cm above the soil surface. Five sensors allowing SP measurements are installed 5, 29, 53, 77 and 101 cm from the surface, respectively. The column has a zero pressure head maintained at its bottom. At the top of the column, the boundary condition corresponds to a Dirichlet condition with a prescribed pressure head during the falling-head phase, followed by a Neumann condition with zero infiltration flux during the drainage phase. During the falling-head phase, the prescribed pressure head  $h_{\text{top}} = 48$  cm has an exponential behavior driven by the saturated conductivity  $h_{\text{top}} = (L_s + L_w) e^{-\frac{K_s t}{L_s}} - L_s$ . The falling-head phase remains until the ponding vanishes at the critical time  $t_c = -\frac{L_s}{K_s} \ln\left(\frac{L_s}{L_s + L_w}\right)$ .

### 2.2 Mathematical model

The total electrical current density  $j$  ( $\text{A m}^{-2}$ ) is determined from the generalized Ohm's law as follows:

$$j = -\sigma \nabla \varphi + j_s, \quad (1)$$

where  $\varphi$  (V) is the streaming potential,  $j_s$  ( $\text{A m}^{-2}$ ) is the streaming current density and  $\sigma$  ( $\text{S m}^{-1}$ ) is the electrical conductivity distribution that is assumed to be isotropic.

Hence, the conservation equation ( $\nabla \cdot j = 0$ ) is written as

$$\nabla \cdot (\sigma \nabla \varphi) = \nabla \cdot j_s. \quad (2)$$

In addition, the electrical conductivity distribution can be estimated using the saturation  $S_w = \theta/\theta_s$  as follows (Mboh et al., 2012):

$$\sigma = \sigma_{\text{sat}} S_w^{n_a}, \quad (3)$$

where  $\sigma_{\text{sat}}$  ( $\text{S m}^{-1}$ ) is the electric conductivity at saturation and  $n_a$  is the Archie saturation exponent (Archie, 1942).

The streaming current density ( $j_s$ ) can be related to the Darcy velocity ( $q$  ( $\text{cm min}^{-1}$ )) by (Linde et al., 2007; Revil et al., 2007)

$$j_s = \left( -\sigma_{\text{sat}} \frac{\rho g}{K_s} C_{\text{sat}} S_w \right) q, \quad (4)$$

where  $K_s$  ( $\text{cm min}^{-1}$ ) is the saturated hydraulic conductivity,  $\rho$  ( $\text{kg min}^{-3}$ ) is the water density,  $g$  ( $\text{m s}^{-1}$ ) is the gravitational acceleration and  $C_{\text{sat}}$  ( $\text{V Pa}^{-1}$ ) is the voltage coupling coefficient at saturation.

Hence, the combination of the previous Eqs. (1)–(4) leads to the following partial differential equation governing the SP signals:

$$\nabla \cdot (S_w^{n_a} \nabla \varphi) = \nabla \cdot \left( -\frac{\rho g C_{\text{sat}} S_w}{K_s} q \right). \quad (5)$$

However, the flow through an unsaturated soil column can be modeled by the one-dimensional Richard's equation:

$$\frac{\partial \theta}{\partial t} = \left( c(h) + S_s \frac{\theta}{\theta_s} \right) \frac{\partial H}{\partial t} = -\nabla \cdot q, \quad (6)$$

where  $H$  (cm) and  $h$  (cm) are, respectively, the hydraulic and pressure head, such that  $H = h - z$ ;  $z$  (cm) is the depth (downward positive);  $S_s$  (–) is the specific storage;  $\theta_s$  ( $\text{cm}^3 \text{cm}^{-3}$ ) and  $\theta$  ( $\text{cm}^3 \text{cm}^{-3}$ ) are the saturated and actual water contents, respectively;  $c(h)$  ( $\text{cm}^{-1}$ ) is the specific moisture capacity; and  $K(h)$  ( $\text{cm min}^{-1}$ ) is the hydraulic conductivity.

The water velocity ( $q$ ) is given from Darcy's law:

$$q = -K(h) \nabla H. \quad (7)$$

In the current study, the standard models of Mualem (1976) and van Genuchten (1980) are used to relate pressure head, hydraulic conductivity and water content:

$$S_e(h) = \frac{\theta(h) - \theta_r}{\theta_s - \theta_r} = \begin{cases} \frac{1}{(1 + |\alpha h|^n)^m} h < 0 \\ 1 h \geq 0 \end{cases}, \quad (8)$$

$$K(S_e) = K_s S_e^{1/2} \left[ 1 - \left( 1 - S_e^{1/m} \right)^m \right]^2$$

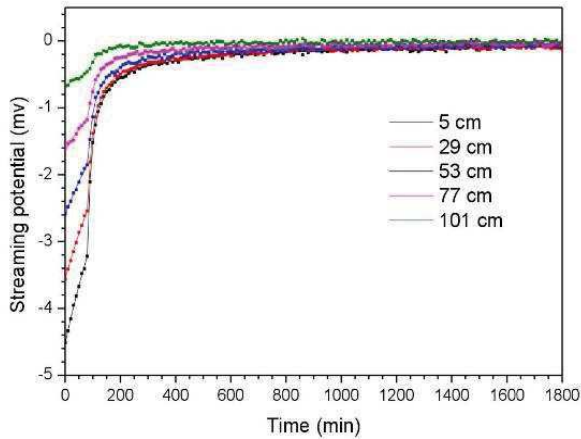
where  $S_e$  (–) is the effective saturation,  $\theta_r$  ( $\text{cm}^3 \text{cm}^{-3}$ ) is the residual water content,  $K_s$  ( $\text{cm min}^{-1}$ ) is the saturated hydraulic conductivity,  $m = 1 - 1/n$ , and  $\alpha$  ( $\text{cm}^{-1}$ ) and  $n$  (–) are the Mualem–van Genuchten-shaped parameters.

### 2.3 Numerical model

Although several numerical techniques have been developed for the solution of the multidimensional Richards equation (e.g., Fahs et al., 2009; Belfort et al., 2009; Younes et al., 2013; Deng and Wang, 2017), the standard finite volume method is used here for the spatial discretization of the one-dimensional Richard's equation (Eq. 6). The integration of this equation over the finite volume ( $i$ ) between  $(i - 1/2)$  and  $(i + 1/2)$  gives

$$\int_{i-1/2}^{i+1/2} \left( c(h) + S_s \frac{\theta}{\theta_s} \right) \frac{\partial H}{\partial t} dz = q_{i-1/2} - q_{i+1/2}. \quad (9)$$

Using expressions of the Darcy velocity at the element interfaces  $q_{i-1/2} = -\frac{K_{i-1/2}}{\Delta z} (H_i - H_{i-1})$  and  $q_{i+1/2} =$



**Figure 2.** Reference SP signals. Solid lines represent the reference SP solution and dots represent the sets of perturbed data serving as conditioning information for model calibration.

$-\frac{K_{i+1/2}}{\Delta z}(H_{i+1}-H_i)$  in the case of a uniform spatial discretization with a spatial step, we obtain

$$\Delta z \left( c_i + S_s \frac{\theta_i}{\theta_s} \right) \frac{\partial H_i}{\partial t} = K_{i+1/2} (H_{i+1} - H_i) \quad (10)$$

$$-K_{i-1/2} (H_i - H_{i-1}). \quad (11)$$

Using  $\tau = S_w^{n_a}$  and  $\delta = \frac{\rho g C_{sat} S_w}{K_s}$ , the integration of Eq. (5) over the finite volume ( $i$ ) yields

$$\frac{\tau_{i+1/2}}{\Delta z} (\varphi_{i+1} - \varphi_i) - \frac{\tau_{i-1/2}}{\Delta z} (\varphi_i - \varphi_{i-1}) \quad (12)$$

$$- \delta_{i+1/2} K_{i+1/2} (H_{i+1} - H_i) \quad (13)$$

$$+ \delta_{i-1/2} K_{i-1/2} (H_i - H_{i-1}) = 0, \quad (14)$$

where the values at the interface  $\tau_{i\pm 1/2}$ ,  $\delta_{i\pm 1/2}$  and  $K_{i\pm 1/2}$  are calculated using the arithmetic mean between adjacent elements (for instance,  $\tau_{i+1/2} = (\tau_i + \tau_{i+1})/2$ ).

Then, the temporal discretization of the obtained nonlinear ODE/DAE system (9–10) is performed with the method of lines (MOL) using the DASKP (Brown et al., 1994) time solver. The MOL is suitable for strongly nonlinear systems since it allows high-order temporal integration methods with formal error estimation and control (Miller et al., 1998; Younes et al., 2009; Fahs et al., 2009, 2011). In the current study, the relative and absolute local error tolerances are fixed to  $10^{-6}$ .

Numerical simulations are performed assuming typical MVG hydraulic parameters for the sandy soil with (according to Carsel and Parrish, 1988)  $K_s = 0.495 \text{ cm min}^{-1}$ ,  $\theta_s = 0.43 \text{ cm}^3 \text{ cm}^{-3}$ ,  $\theta_r = 0.045 \text{ cm}^3 \text{ cm}^{-3}$ ,  $\alpha = 0.145 \text{ cm}^{-1}$  and  $n = 2.68$ . The voltage coupling coefficient at saturation is  $C_{sat} = -2.9 \times 10^{-7} \text{ V Pa}^{-1}$  and the Archie saturation exponent is  $n_a = 1.6$ .

Based on these hydraulic and geophysical parameters, a reference (mesh-independent) solution is obtained using a uniform mesh of 235 cells of 0.5 cm length. Data are generated by sampling the output SP signals every 10 min during 1800 min. Figure 2 shows that the SP signals have an almost linear behavior in the saturated falling-head phase. During the drainage phase, they have a nonlinear behavior and approach zero voltage for the dry conditions occurring toward the end of the experiment. The SP signals are independent Gaussian random noises with a standard deviation of  $2.73 \times 10^{-5} \text{ V}$ . This noise level was obtained by Mboh et al. (2012) from laboratory measurements. The noised data (Fig. 2) are used as “observations” in the calibration exercise.

### 3 Global sensitivity analysis of SP signals

#### 3.1 GSA method

The aim of GSA is to assess the effect of the variation of parameters on the model output (Mara and Tarantola, 2008). Such knowledge is important for determining the most influential parameters as well as their regions and periods of influence (Fajraoui et al., 2011). The sensitivity of a model to its parameters can be assessed using variance-based sensitivity indices. These indices evaluate the contribution of each parameter to the variance of the model (Sobol’, 2001). Polynomial chaos theory (Wiener, 1938) has been largely used to perform variance-based sensitivity analysis of computer models (see for instance, Sudret, 2008; Blatman and Sudret, 2010; Fajraoui et al., 2012; Younes et al., 2016; Shao et al., 2017; Mara et al., 2017). It can be stated that the PCE method is a surrogate-based approach. However, we argue that this method employs ANOVA (analysis Of variance) decomposition and hence can be considered as a spectral method (such as the Fourier amplitude sensitivity test; Cukier et al., 1973; Saltelli et al., 1999). Indeed, with this method, the sensitivity indices are directly obtained from the PCE coefficients without needing to run the surrogate model.

Let us consider a mathematical model with a random response  $f(\xi)$  which depends on ( $d$ ) independent random parameters  $\xi = \{\xi_1, \xi_2, \dots, \xi_d\}$ . With PCE,  $f(\xi)$  is expanded using a set of orthonormal multivariate polynomials (up to a polynomial degree  $p$ ):

$$f(\xi) \approx \sum_{|\alpha| \leq p} s_\alpha \Psi_\alpha(\xi), \quad (15)$$

where  $\alpha = \alpha_1 \dots \alpha_d \in R^d$  is a  $d$ th-dimensional index.  $S_\alpha$  represents the polynomial coefficients and  $\Psi_\alpha$  represents the generalized polynomial chaos of degree  $|\alpha| = \sum_{i=1}^d \alpha_i$ , such as Hermite, Legendre and Jacobi polynomials, for instance. In this work, Legendre polynomials are employed because uniform distributions are considered for the parameters. The

noninformative uniform distributions are used here to express the absence of prior information, which makes all possible values of the parameter equally likely.

Equation (12) is similar to an ANOVA representation of the original model (Sobol' 1993), from which it is straightforward to express  $V[f(\xi)]$ , the variance of  $f(\xi)$  as the sum of the partial contribution of the inputs, as follows:

$$V[f(\xi)] = \sum_{\alpha} v_{\alpha}^2. \quad (16)$$

The first-order sensitivity index  $S_i$  and the total sensitivity index  $ST_i$  are defined by

$$S_i = \frac{V[E[f(\xi)|\xi_i]]}{V[f(\xi)]} \in [0, 1] \quad (17)$$

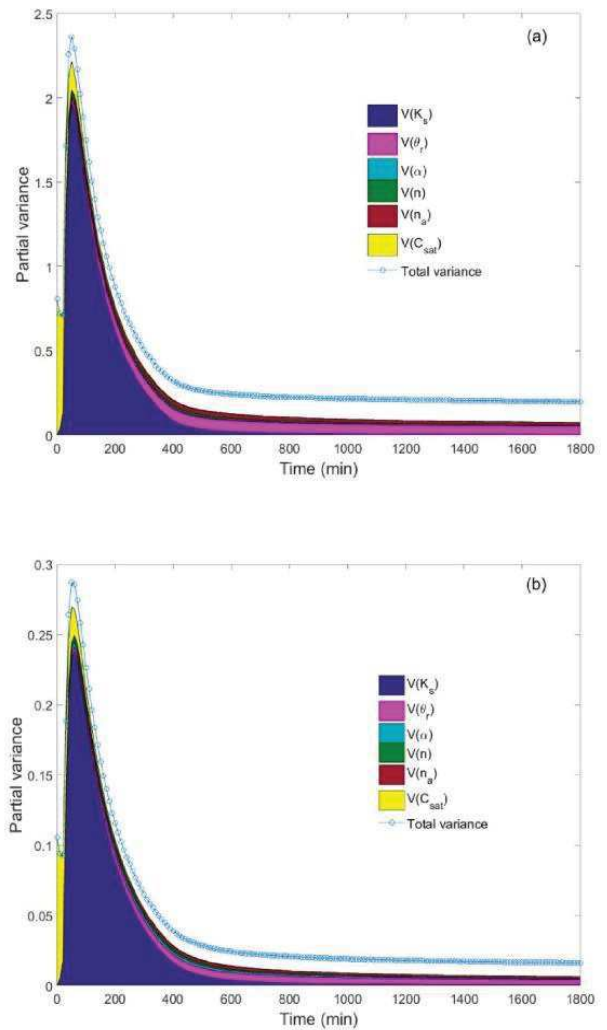
$$ST_i = \frac{E[V[f(\xi)|\xi_{-i}]]}{V[f(\xi)]} \in [0, 1], \quad (18)$$

where  $\xi_{-i} = \xi \setminus \xi_i$ ,  $E(\cdot)$  is the conditional expectation operator and  $V(\cdot)$  the conditional variance.  $S_i$  measures the amount of variance of  $f(\xi)$  due to  $\xi_i$  alone, while  $ST_i \geq S_i$  measures the amount of all contributions of  $\xi_i$  to the variance of  $f(\xi)$ , including its cooperative nonlinear contributions with the other parameters  $\xi_j$ . The input/output relationship is said to be *additive* when  $ST_i = S_i$ ,  $\forall i = 1, \dots, d$ , and in this case  $\sum_{i=1}^d S_i = 1$ .

In the sequel, a PCE is constructed for each SP signal at each observable time. The number of coefficients for a full PCE representation is  $P = (d + p)! / (d! p!)$ . The evaluation of the PCE coefficients requires at least  $P$  simulations of the nonlinear hydrogeophysical model. Note that  $P$  increases quickly with the order of the PCE and the number of parameters. Hence, several sparse PCE representations, in which only the significant coefficients are sought, have been proposed in the literature in order to reduce the computational cost of the estimation of the Sobol indices. For instance, Blatman and Sudret (2010) developed a sparse PCE representation using an iterative forward-backward approach based on nonintrusive regression. Fajraoui et al. (2012) developed a technique whereby only the sensitive coefficients (that affect significantly model variance) are retained in the PCE. Recently, Shao et al. (2017) developed an algorithm based on Bayesian model averaging (BMA) to select the best sparse PCE from a given data set using the KIC (Kayshap, 1982). The main idea of this algorithm is to increase the degree of an initial PCE progressively and compute the KIC until a satisfactory representation of the model responses is obtained. This algorithm is used hereafter to compute the sensitivity indices of the SP signals.

### 3.2 GSA results

The SP responses are considered for uniformly distributed parameters over the large intervals shown in Table 1. These



**Figure 3.** Time distribution of the SP variance at 5 cm (a) and 77 cm (b) depth. The shaded area under the variance curve represents the partial marginal contributions of the random input parameters; the contribution of interactions between parameters is represented by the blank region between the shaded area and the variance curve.

intervals include the reference values reported in Mboh et al. (2012). The sensitivity indices of the six input parameters ( $K_s$ ,  $\theta_r \alpha$ ,  $n$ ,  $n_a$ ,  $C_{sat}$ ) are estimated using an experimental design formed by  $N = 2^{12} = 4096$  parameter sets. The order of the sparse PCE is automatically adapted for each observable time and location. For some observable times, the PCE is highly sparse; it reaches a degree of 31 but only contains 112 nonzero coefficients.

Figure 3 depicts the temporal distribution of the streaming potential variance, represented by the blue curve, and the relative contribution of the parameters, represented by the shaded area. This figure corresponds to the temporal ANOVA decomposition for sensor 1 (5 cm from the soil surface) and

**Table 1.** Reference values, lower and upper bounds for hydraulic and geophysical parameters.

Parameters	Lower bounds	Upper bounds	Reference values
$K_s$ (cm min <sup>-1</sup> )	0.1	2	0.495
$\theta_r$ (cm <sup>3</sup> min <sup>-3</sup> )	0	0.2	0.045
$\alpha$ (cm <sup>-1</sup> )	0.01	0.2	0.145
$n$	1.5	7	2.68
$n_a$ (-)	1	3	1.6
$C_{\text{sat}} \times (-10^{-7})$ (V Pa <sup>-1</sup> )	2	4	2.9

for sensor 4 (77 cm from the soil surface). Interactions between parameters are represented by the blank region between the variance curves and the shaded area. Note that because a Dirichlet boundary condition with zero SP is maintained at the outlet boundary, the variance of the SP signal is zero at the bottom and reaches its maximum value near the soil surface. Hence, the variance is higher for the first sensor, located 5 cm from the soil surface (Fig. 3a) than for sensor 4, located 77 cm from the soil surface (Fig. 3b).

The SP signals at different altitudes exhibit similar behavior (Fig. 3). In the following, we comment on the results of sensor 1 (Fig. 3a). Because  $K_s$  varies between 0.1 and 2 cm min<sup>-1</sup>, the saturated falling-head phase remains until the ponding vanishes at  $t_c = -\frac{L_s}{K_s} \ln\left(\frac{L_s}{L_s+L_w}\right)$ . Depending on the value of  $K_s$  (see Table 1),  $t_c$  varies between  $t_1 = 20$  min and  $t_2 = 403$  min. Thus, in Fig. 3a, we can see that during the first time period ( $t \leq t_1$ ), the SP signal is strongly influenced by the value of the parameter  $C_{\text{sat}}$ . The first-order and total sensitivity indices at  $t = 10$  min (Table 2a) confirm that only the saturated parameters  $K_s$  and  $C_{\text{sat}}$  are influential.  $C_{\text{sat}}$  is about 17 times more influential than  $K_s$ . As expected, the remaining parameters have no influence during the first period. The total variance is 0.72 mv, and there is no interaction between the two parameters  $K_s$  and  $C_{\text{sat}}$  since  $ST_i = S_i$  for both of them and  $\sum_{i=1}^d S_i = 1$ .

During the second period ( $t_1 \leq t \leq t_2$ ), the flow is either saturated or unsaturated depending on the value of  $K_s$ . Figure 3a shows that the variance of the SP signal exhibits its maximum value around 2.4 mv, with strong influences of the parameters  $K_s$  and  $C_{\text{sat}}$  and weak interactions between them (small blank region between the variance curve and the shaded area). These results are confirmed by the sensitivity indices calculated at  $t = 70$  min and reported in Table 2a for sensor 1. Both first-order and total sensitivity indices indicate that  $K_s$  is the most influential parameter. The second influential parameter is  $C_{\text{sat}}$ , which has a total sensitivity index about 12 times less than  $K_s$ . The parameter  $\alpha$  is irrelevant since its total sensitivity index is 109 times less than  $K_s$  and its partial variance is  $V_i = S_i \times V_i = 0.01$  mv, which is less than the 95 % confidence interval associated with the SP measurement ( $\pm 0.055$  mv). The total variance at  $t = 70$  min is calculated to be 2.17 mv, and the output–input relationship is close to being additive since  $\sum_{i=1}^d S_i = 0.94$ , which means

that interactions between parameters exist but are not significant.

During the third period ( $t \geq t_2$ ), the variance of the SP signal reduces to 0.3 mv (Fig. 3a) and significant interactions are observed between parameters (large blank region between the shaded area and the variance curve). Table 2a shows that for  $t = 800$  min, which corresponds to dry conditions, the total variance is 0.22. First-order sensitivity indices are very small, except for  $\theta_r$ . The latter is highly influential since it has a significant first-order sensitivity index ( $S_i = 0.27$ ) and a more significant total-sensitivity index ( $ST_i = 0.74$ ). The parameters  $C_{\text{sat}}$  and  $K_s$  are irrelevant as they have very small first-order and total sensitivity indices. Further, strong interactions are observed between the parameters since the sum of the first-order indices is far from 1 ( $\sum_{i=1}^d S_i = 0.47$ ). The total sensitivity indices are significantly different from first-order sensitivity indices for almost all parameters. For instance, the ratio between these two indices is around 4 for  $\alpha$ , 5 for  $n_a$  and 7 for  $n$ . The total sensitivity index of  $\alpha$  remains small (0.065), whereas significant total sensitivity indices are obtained for  $n$  ( $ST_i = 0.27$ ) and  $n_a$  ( $ST_i = 0.47$ ), which indicates that these two parameters are influential (although their first-order sensitivity indices are small) because of the interaction between the parameters.

Figure 3b shows similar behavior for sensor 4 located 77 cm from the soil surface. The results in Table 2b indicate that the total variance observed at  $t = 10, 70$  and 800 min are around 8 times less than for sensor 1. For the first time period, the first and total sensitivity indices are identical to those observed for sensor 1 since saturated conditions occur inside the whole column and the same effect of  $K_s$  and  $C_{\text{sat}}$  can be observed whatever the location inside the column. For the second time period, the sensitivity indices for sensor 4 (Table 2b) are similar to those observed for sensor 1. However, the results for the third time period show an improvement of the relevance of the parameter  $\alpha$ , with an increase of both first and total sensitivity indices. Indeed, compared to the results of sensor 1, both first-order and total sensitivity indices tripled. Moreover, the total sensitivity index for  $\alpha$  ( $ST_i = 0.22$ ) becomes close to that of  $n$  ( $ST_i = 0.24$ ).

In summary, the GSA applied to SP signals identifies the influential parameters and their periods of influence and shows that

**Table 2.** The first-order sensitivity index  $S_i$  and the total sensitivity index  $ST_i$  for the SP signal 5 (a) and 77 cm (b) below the soil surface at different times.

	$K_s$ ( $\text{cm min}^{-1}$ )	$\theta_r$ ( $\text{cm}^3 \text{min}^{-3}$ )	$\alpha$ ( $\text{cm}^{-1}$ )	$n$	$n_a$	$C_{\text{sat}}$ ( $\text{V Pa}^{-1}$ )
<b>(a) Sensor 1 (5 cm from the soil surface)</b>						
$t = 10 \text{ min}$ (total variance = 0.72)						
$S_i$	0.055	0	0	0	0	0.942
$ST_i$	0.057	0	0	0	0	0.945
$t = 70 \text{ min}$ (total variance = 2.17)						
$S_i$	0.841	0.217	0.005	0.014	0.008	0.045
$ST_i$	0.894	0.043	0.008	0.028	0.021	0.078
$t = 800 \text{ min}$ (total variance = 0.224)						
$S_i$	0.053	0.266	0.015	0.038	0.094	0.008
$ST_i$	0.085	0.738	0.065	0.266	0.472	0.041
<b>(b) Sensor 4 (77 cm from the soil surface)</b>						
$t = 10 \text{ min}$ (total variance = 0.094)						
$S_i$	0.055	0	0	0	0	0.942
$ST_i$	0.057	0	0	0	0	0.945
$t = 70 \text{ min}$ (total variance = 0.2744)						
$S_i$	0.839	0.015	0.014	0.013	0.005	0.053
$ST_i$	0.891	0.028	0.024	0.025	0.011	0.086
$t = 800 \text{ min}$ (total variance = 0.224)						
$S_i$	0.099	0.225	0.054	0.043	0.085	0.01
$ST_i$	0.138	0.621	0.218	0.238	0.379	0.043

- the parameter  $C_{\text{sat}}$  is highly influential during the first time period ( $t \leq t_1$ ) during which no interactions are observed between parameters;
- the parameter  $K_s$  is highly influential during the second time period ( $t_1 \leq t \leq t_2$ ) during which small interactions occur between parameters;
- the parameters  $\theta_r$ ,  $n$  and  $n_a$  are influential during the third time period ( $t \geq t_2$ ) during which dry conditions occur; during this period, strong interactions take place between parameters;
- the parameter  $\alpha$  has no influence on the SP signals during the two first periods and presents a very small influence ( $S_i = 0.015$  and  $ST_i = 0.065$ ) during the third period on sensor 1 (near the surface of the column);
- the relevance of the parameter  $\alpha$  improves with the distance from the soil surface, although the total variance diminishes with respect to this distance. The influence of  $\alpha$  becomes significant ( $ST_i = 0.22$ ) on sensor 4 (located 77 cm from the soil surface) during the third period.

## 4 Parameters' estimation

### 4.1 MCMC and FOA approaches

Calibration of computer models is an essential task since some parameters (like the Mualem–van Genuchten-shaped parameters  $\alpha$  and  $n$ ) cannot be directly measured. In such an exercise, the unknown model parameters are investigated by comparing the model responses to the observations. Recently, Mboh et al. (2012) showed that the inversion of SP signals can yield an accurate estimate of the saturated hydraulic conductivity  $K_s$ , the MVG fitting parameters  $\alpha$  and  $n$  and the Archie saturation exponent ( $n_a$ ). Moreover, they showed that the quality of the estimation was comparable to that obtained from the calibration of pressure heads. In their study, Mboh et al. (2012) used the FOA approach with the shuffled complex evolution optimization algorithm SCE-UA (Duan et al., 1993).

As important as the determination of the optimal parameter sets are the associated 95 % confidence intervals (CIs) to quantify the uncertainty of the estimated values. The determination of CIs is not straightforward if the observed model

responses are highly nonlinear functions of model parameters (Christensen and Cooley, 1999). In the sequel, the parameter estimation is performed using two approaches: the popular FOA approach and the Bayesian approach based on the MCMC sampler. Contrarily to FOA, the MCMC method is robust since no assumptions of model linearity or differentiability are required. Furthermore, prior information available for the parameters can be included. MCMC provides not only an optimal point estimate of the parameters but also a quantification of the entire parameter space. Several MCMC strategies have been developed for Bayesian sampling of the parameter space (Gallagher and Doherty, 2007; Vrugt, 2016). In a groundwater and vadose zone modeling context, the most widely used of these strategies is the Metropolis–Hastings algorithm (Metropolis et al., 1953; Hastings, 1970). It proceeds as follows (Gelman et al., 1996).

- i. Choose an initial candidate  $x^0 = (\xi^0, \sigma^0)$  formed by the initial estimate of the parameter set  $\xi^0$  and the hyperparameter  $\sigma^0$  and a proposal distribution  $q$  that depends on the previous accepted candidate.
- ii. A new candidate  $x^i = (\xi^i, \sigma^i)$  is generated from the current one  $x^{i-1}$  with the generator  $q(x^i | x^{i-1})$  associated with the transition probability  $p(\xi^i | y_{\text{mes}}, \sigma)$ .
- iii. Calculate  $p(\xi^i | y_{\text{mes}}, \sigma)$  and compute the ratio  $\alpha = \frac{p(\xi^i | y_{\text{mes}}, \sigma) q(x^{i-1} | x^i)}{p(\xi^{i-1} | y_{\text{mes}}, \sigma) q(x^i | x^{i-1})}$ . Additionally, draw a random number  $u \in [0, 1]$  from a uniform distribution.
- iv. If  $\alpha \geq u$ , then accept the new candidate, otherwise it is rejected.
- v. Resume from (ii) until the chain  $\{x^0, \dots, x^k\}$  converges or a prescribed number of iterations  $i_{\text{max}}$  is reached.

Many improvements have been proposed in the literature to accelerate the MCMC convergence rate (e.g., Haario et al., 2006; ter Braak and Vrugt, 2008; Dostert et al., 2009, among others). Vrugt et al. (2009a, b) developed the DREAM MCMC sampler based on the differential evolution–Markov chain method of ter Braak (2006) to improve sampling efficiency. DREAM runs multiple Markov chains in parallel and uses subspace sampling and outlier chain correction to speed up MCMC convergence (Vrugt, 2016). Laloy and Vrugt (2012) developed the DREAM<sub>(ZS)</sub> MCMC sampler, in which a candidate for each chain is drawn from an archive of past states denoted  $Z$ , which plays the role of the generator  $q$ . Interested readers are referred to Vrugt (2016) for more details about the properties and implementation of DREAM and DREAM<sub>(ZS)</sub>. In the current study, the DREAM<sub>(ZS)</sub> software is used for the MCMC estimation of the hydrogeophysical parameters. Note that because of the large number of model evaluations required, the MCMC method remains rarely used in practical applications compared to the FOA approach. Indeed, with FOA, the CIs are estimated once by

assuming that the Jacobian remains constant within the CIs. This assumption was found to be reasonably accurate in nonlinear problems by Donaldson and Scnabel (1987). However, recently, several authors stated that parameter interdependencies and model nonlinearities violate this assumption (see, for instance, Vrugt and Bouten, 2002; Vurgin et al. 2007; Gallagher and Doherty, 2007; Mertens et al., 2009; Kahl et al., 2015).

In the following, both MCMC and FOA approaches are employed for the inversion of the highly nonlinear hydrogeophysical problem using SP measurements.

#### 4.2 Parameters estimation results

Hydrogeophysical parameters are estimated using the DREAM<sub>(ZS)</sub> MCMC sampler (Laloy and Vrugt, 2012). Independent uniform distributions are considered for model parameter priors and likelihood hyperparameters (see Table 1). The parameter posterior distribution is written as

$$p(\xi | y_{\text{mes}}, \sigma) \propto \sigma^{-N} \exp\left(-\frac{SS(\xi)}{2\sigma^2}\right), \quad (19)$$

where  $SS(\xi) = \sum_{k=1}^N (y_{\text{mes}}^{(k)} - y_{\text{mod}}^{(k)}(\xi))^2$  is the sum of the squared differences between the observed  $y_{\text{mes}}^{(k)}$  and modeled  $y_{\text{mod}}^{(k)}$  SP signals at time  $t_k$  for  $N$ , the total number of SP observations.

The DREAM<sub>(ZS)</sub> software computes multiple sub-chains in parallel to thoroughly explore the parameter space. Taking the last 25 % of individuals (when the chains have converged) yields multiple sets used to estimate the updated parameter distributions and therefore the optimal parameter values and their CIs. In the sequel, the DREAM<sub>(ZS)</sub> MCMC sampler is used with three parallel chains.

We assume that the saturated water content has been initially measured with a fair degree of accuracy. However, instead of fixing its value (as in Kool et al., 1987, van Dam et al., 1994, and Nützmann et al., 1998, among others), we assign a Gaussian distribution to  $\theta_s$  to take associated uncertainty and its effect on the estimation of the rest of the parameters into account. It is assumed here that the saturated water content was accurately measured to be  $\theta_s = 0.43 \text{ cm}^3 \text{ cm}^{-3}$  by weighing the saturated soil. The corresponding error measurements are independently and normally distributed with a zero mean and a standard deviation  $\sigma_\theta = 0.01 \text{ cm}^3 \text{ cm}^{-3}$ . Hence a Gaussian distribution is assigned to  $\theta_s$ , with a mean value of  $0.43 \text{ cm}^3 \text{ cm}^{-3}$  and a 95 % CI  $[0.41 - 0.45] \text{ cm}^3 \text{ cm}^{-3}$ . The rest of the hydrogeophysical parameters have noninformative uniform distributions over the ranges reported in Table 1. The error (measurement) variance is also considered to be unknown and is simultaneously estimated with the physical parameters. Two scenarios are considered to check whether the FOA approach can provide a reliable estimation of parameters and associated uncertainties for the highly nonlinear hydrogeophysi-

**Table 3.** Estimated mean values (bold), confidence intervals (CIs) and size of the posterior CIs (italic) with MCMC and FOA approaches for scenario 1.

	MCMC	FOA
$K_s$ ( $\text{cm min}^{-1}$ )	<b>0.49</b> (0.487–0.498) <i>0.01</i>	<b>0.49</b> (0.487–0.497) <i>0.01</i>
$\theta_s$ ( $\text{cm}^3 \text{min}^{-3}$ )	<b>0.43</b> (0.41–0.45) <i>0.04</i>	<b>0.43</b> (0.41–0.45) <i>0.04</i>
$\theta_r$ ( $\text{cm}^3 \text{min}^{-3}$ )	<b>0.046</b> (0.025–0.068) <i>0.04</i>	<b>0.046</b> (0.026–0.066) <i>0.04</i>
$\alpha$ ( $\text{cm}^{-1}$ )	<b>0.14</b> (0.12–0.17) <i>0.05</i>	<b>0.14</b> (0.12–0.16) <i>0.04</i>
$n$	<b>2.64</b> (2.54–2.77) <i>0.23</i>	<b>2.64</b> (2.54–2.76) <i>0.22</i>
$n_a$	<b>1.64</b> (1.37–1.98) <i>0.6</i>	<b>1.64</b> (1.38–1.90) <i>0.5</i>
$C_{\text{sat}}$ ( $\text{V Pa}^{-1}$ )	<b>2.90</b> (2.89–2.91) <i>0.02</i>	<b>2.90</b> (2.89–2.91) <i>0.02</i>

**Table 4.** Estimated mean values (bold), confidence intervals (CIs) and size of the posterior CIs (italic) with MCMC and FOA approaches for scenario 2.

	MCMC	FOA
$K_s$ ( $\text{cm min}^{-1}$ )	<b>0.49</b> (0.481–0.495) <i>0.014</i>	<b>0.49</b> (0.474–0.503) <i>0.029</i>
$\theta_s$ ( $\text{cm}^3 \text{min}^{-3}$ )	<b>0.43</b> (0.41–0.45) <i>0.04</i>	<b>0.43</b> (0.41–0.45) <i>0.04</i>
$\theta_r$ ( $\text{cm}^3 \text{min}^{-3}$ )	<b>0.053</b> (0.011–0.093) <i>0.08</i>	<b>0.053</b> (0.002–0.103) <i>0.1</i>
$\alpha$ ( $\text{cm}^{-1}$ )	<b>0.13</b> (0.07–0.20) <i>0.13</i>	<b>0.13</b> (-0.15–0.43) <i>0.58</i>
$n$	<b>2.54</b> (2.44–2.68) <i>0.24</i>	<b>2.56</b> (2.44–2.68) <i>0.24</i>
$n_a$	<b>1.82</b> (1.36–2.41) <i>1.05</i>	<b>1.78</b> (1.29–2.27) <i>0.98</i>
$C_{\text{sat}}$ ( $\text{V Pa}^{-1}$ )	<b>2.89</b> (2.88–2.91) <i>0.03</i>	<b>2.89</b> (2.88–2.91) <i>0.03</i>

cal problem investigated, both in the case of abundant data (small uncertainty regions) and in the case of scarcity of data (large uncertainty regions). In the first scenario, SP data collected from the sensors located at the five locations are taken into account for the calibration. In the second scenario, only the SP data from the first sensor located 5 cm from the soil surface serve as conditioning information for model calibration. Results of the MCMC sampler are compared to those of the FOA approach for both scenarios.

#### 4.2.1 Scenario 1: inversion using all SP measurements

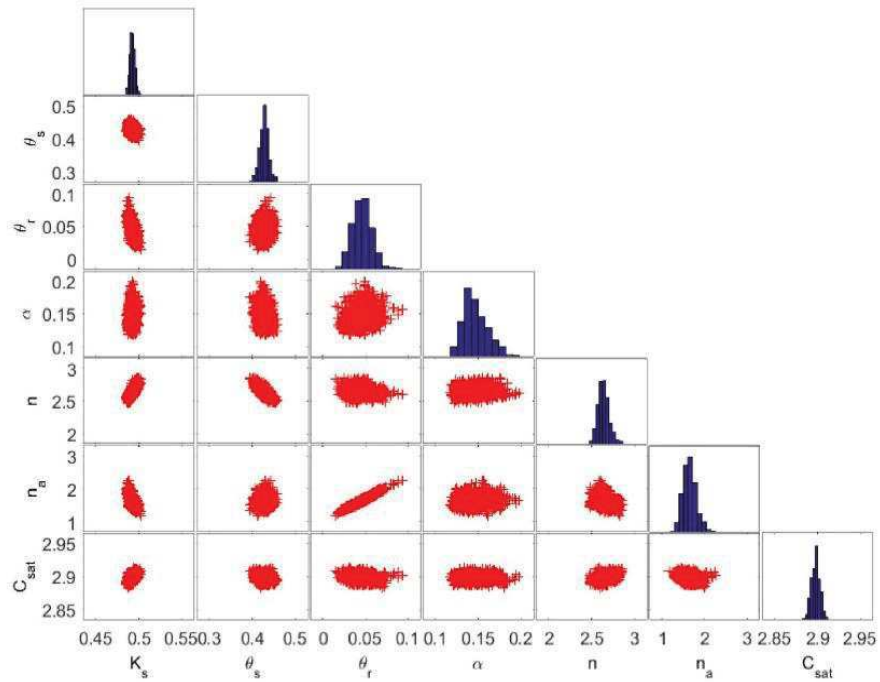
Figure 4 shows the results obtained with MCMC when the SP data of the five sensors are used for the calibration. The “on-diagonal” plots in this figure display the posterior parameter distributions, whereas the “off-diagonal” plots represent the correlations between parameters in the MCMC sample. Figure 4 shows nearly bell-shaped posterior distributions for all parameters. A strong correlation is observed between  $\theta_r$  and  $n_a$  ( $r = 0.98$ ).

From the obtained MCMC sample, it is straightforward to estimate the posterior 95 % confidence interval of each parameter. This as well as the mean estimate value of each parameter obtained with both MCMC and FOA approaches are reported in Table 3.

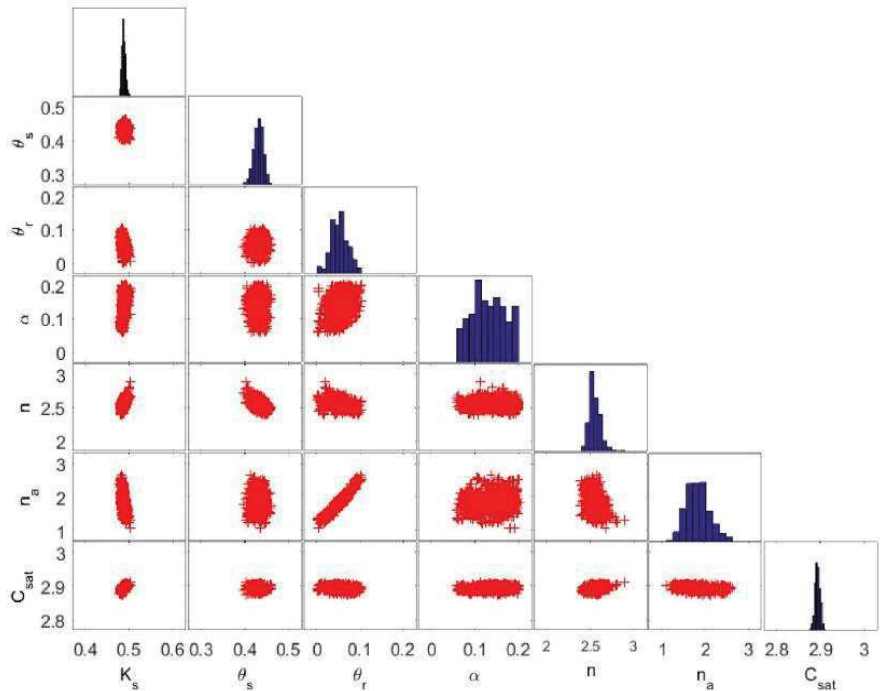
The results of this table show that the parameters are well estimated from the SP measurements since (i) identified mean values are very close to the reference solution, (ii) all confidence intervals include the reference solution and (iii) the confidence intervals are rather narrow. The saturated parameters  $K_s$  and  $C_{\text{sat}}$  are very well estimated (with CIs around 2 %) because of data collected during the falling-head phase during which only these two parameters are influential.

The posterior CI of the parameter  $\theta_s$  is similar to its prior CI. The parameter  $\alpha$  is reasonably well estimated, with a CI around 35 %. Recall that this parameter had very small first-order and total sensitivity indices for sensor 1 but had more significant sensitivity indices for the sensors away from the soil surface (see results for sensor 4 in Table 2b). The parameter  $\theta_r$  is estimated with a CI around 90 % although it was highly influential for all sensors (for instance, a first-order sensitivity index of 0.27 and a total order of 0.74 for sensor 1). The parameters  $n$  and  $n_a$  had similar GSA behavior to small first-order sensitivities (0.038 and 0.094, respectively, for sensor 1) and large total sensitivities (0.266 and 0.4715, respectively, for sensor 1); however, the inversion shows that the parameter  $n$  is well estimated with a CI less than 10 %, whereas the parameter  $n_a$  is less well estimated with a CI around 35 %. These results suggest that GSA out-





**Figure 4.** MCMC solutions in which all SP data are considered for the calibration. The diagonal plots represent the inferred posterior probability distribution of the model parameters. The off-diagonal scatterplots represent the pairwise correlations in the MCMC drawing.



**Figure 5.** MCMC solutions in which calibration is performed using only SP data located 5 cm from the surface. The diagonal plots represent the posterior probability distribution of the parameters. The off-diagonal scatterplots represent the pairwise correlations in the MCMC drawing.

comes should be interpreted with caution in the context of parameter estimation since (i) a parameter which is not relevant for the model output in one sensor can be influential for another sensor and (ii) GSA does not presume the quality of the estimation since two parameters with similar sensitivity indices can have a different quality of estimation with the inversion procedure.

Further, the results of Table 3 show that FOA and MCMC approaches yield similar mean estimated values. Moreover, very good agreement is observed between FOA and MCMC uncertainty bounds. Concerning the efficiency of the two calibration methods for this scenario, the FOA approach is by far the most efficient method since it requires only 95 s of CPU time. The MCMC method was terminated after 16 000 model runs, which required 14 116 s. The convergence was reached at around 12 000 model runs. The last 4000 runs were used to estimate the statistical measures of the posterior distribution. Recall that contrarily to FOA, MCMC can include prior information available for the parameters and allows a quantification of the entire parameter space.

#### 4.2.2 Scenario 2: inversion using only SP measurements near the surface

In this scenario, the number of measurements used for the calibration is strongly reduced. Only SP measurements from sensor 1 (located 5 cm below the soil surface) are considered.

The results of MCMC are plotted in Fig. 5. The correlation observed between  $\theta_r$  and  $n_a$  decreases slightly to  $r = 0.95$ . Almost bell-shaped posterior distributions are observed for all parameters except for the parameters  $\theta_r$  and  $\alpha$ .

The results obtained with MCMC and FOA approaches depicted in Table 4 show the following.

- The FOA approach yields accurate mean estimated values similar to MCMC results for all parameters.
- The MCMC and FOA mean estimated values are close to the reference solution and to the previous scenario. The maximum difference is observed for  $\theta_r$  for which the mean estimated value with scenario 2 is 15 % greater than for scenario 1.
- The MCMC CIs for the parameters  $K_s$ ,  $\theta_s$ ,  $n$  and  $C_{\text{sat}}$  are close to the previous scenario. The parameters  $\theta_s$  and  $n$  are well estimated (CIs < 10 %) and the parameters  $K_s$  and  $C_{\text{sat}}$  are very well estimated (CIs  $\leq 5$  %).
- Due to the reduction of the number of data used for model calibration in scenario 2, the MCMC CIs for the parameters  $n_a$ ,  $\alpha$  and  $\theta_r$  are much larger than in the previous scenario. Indeed, compared to scenario 1, the CI for  $n_a$  and  $\theta_r$  increases by around 60 %, whereas the CI of  $\alpha$  is 3 times larger than for scenario 1.
- The FOA method yields accurate CIs for the parameters  $\theta_s$ ,  $n$ ,  $n_a$  and  $C_{\text{sat}}$ , whereas it overestimates the CIs of

$\theta_r$  (by 24 %),  $K_s$  (by 100 %) and  $\alpha$  (by 427 %). An unphysical uncertainty region (including negative values) is obtained for the parameter  $\alpha$ .

These results show that the FOA can fail to provide realistic parameter uncertainties and can yield larger CIs than their corresponding nonlinear MCMC counterpart. Indeed, the linearization in the FOA method assumes that the Jacobian remains constant across the CIs. This assumption was fulfilled for the first scenario in which a large number of measurements ensured small uncertainty regions. However, the assumption is not fulfilled for some parameters of the current scenario because of the large uncertainty regions induced by the reduction of the number of SP measurements.

Concerning the efficiency of the calibration methods, the FOA required approximately 174 s of CPU time, and the MCMC required many more runs to reach the convergence than in the previous scenario. Indeed, the sampler was used with 50 000 runs (35 000 runs were necessary to reach the convergence).

## 5 Conclusions

In this work, a synthetic test case dealing with SP signals during a drainage experiment has been studied. The test case is similar to the laboratory experiment developed in Mboh et al. (2012), involving a falling-head infiltration phase followed by a drainage phase. GSA and Bayesian parameter inference have been applied to investigate (i) the influence of hydraulic and geophysical parameters on the SP signals and (ii) the identifiability of hydrogeophysical parameters using only SP measurements. The GSA was performed using variance-based sensitivity indices which allow the contribution of each parameter (alone or by interaction with other parameters) to the output variance to be measured. The sensitivity indices have been calculated using a PCE representation of the SP signals. To reduce the number of coefficients and explore PCE with high orders, we used the efficient sparse PCE algorithm developed by Shao et al. (2017), which selects the best sparse PCE from a given data set using the Kashyap information criterion (KIC).

The GSA applied to SP signals showed that the parameters  $C_{\text{sat}}$  and  $K_s$  are highly influential during the first period corresponding to saturated conditions. The parameters  $\theta_r$ ,  $n$  and  $n_a$  are influential when dry conditions occur. In such conditions, strong interactions take place between these parameters. The parameter  $\alpha$  has a very small influence on the SP signals near the soil surface but its sensitivity increases with depth although the total variance decreases with depth.

Parameter estimation has been performed using MCMC and FOA approaches to check whether FOA can provide a reliable estimation of parameters and associated uncertainties for the highly nonlinear hydrogeophysical problem investigated. All hydraulic ( $K_s$ ,  $\theta_r$ ,  $\alpha$  and  $n$ ) and geophysical ( $n_a$  and  $C_{\text{sat}}$ ) parameters can be reasonably estimated in the first sce-

nario for which the whole SP data (measured at five different locations) are used as conditioning information for the model calibration. The confrontation with GSA results shows that the latter should be interpreted with caution when used in the context of parameter estimation since (i) a parameter which is not relevant for the model output in one sensor can be influential for another sensor and (ii) GSA does not presume the quality of the estimation since two parameters with similar sensitivity indices can have a different quality of estimation with the inversion procedure (see, for instance, parameters  $n$  and  $n_a$ ). Furthermore, although the studied problem is highly nonlinear, the FOA approach provides accurate estimations of both mean parameter values and CIs in the first scenario. These results are identical to those obtained with MCMC.

When the number of SP measurements used for the calibration is considerably reduced (i.e., data are scarce), the MCMC inversion provides larger uncertainty regions of the parameters. The FOA approach yields accurate mean parameter values (in agreement with MCMC results) but inaccurate and even unphysical CIs for some parameters with large uncertainty regions.

*Data availability.* No data sets were used in this article.

*Author contributions.* AY framed the research question, worked on sensitivity analysis and finalized the manuscript. JZ and FL worked on parameter estimation and numerical model development. MF performed simulations, analyzed the results and reviewed the manuscript.

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* The authors acknowledge the financial support from the Tunisian–French joint international laboratory NAILA (<http://www.lmi-naila.com/>).

Edited by: Bill X. Hu

Reviewed by: two anonymous referees

## References

- Allègre, V., Jouniaux, L., Lehmann, F., and Sailhac, P.: Streaming potential dependence on water-content in Fontainebleau sand, *Geophys. J. Int.*, 182, 1248–1266, <https://doi.org/10.1111/j.1365-246X.2010.04716.x>, 2010.
- Archie, G. E.: The Electrical Resistivity Log as an Aid in Determining Some Reservoir Characteristics, *Transactions of the AIME*, 146, 54–62, <https://doi.org/10.2118/942054-G>, 1942.
- Arora, B., Mohanty, B. P., and McGuire, J. T.: Uncertainty in dual permeability model parameters for structured soils, *Water Resour. Res.*, 48, <https://doi.org/10.1029/2011WR010500>, 2012.

## A. Younes et al.: Hydrogeophysical parameters estimation

- Belfort, B., Ramasomanan, F., Younes, A., and Lehmann, F.: An efficient Lumped Mixed Hybrid Finite Element Formulation for variably saturated groundwater flow, *Vadoze Zone Journal*, 8, 352–362, <https://doi.org/10.2136/vzj2008.0108>, 2009.
- Blatman, G. and Sudret, B.: Efficient computation of global sensitivity indices using sparse polynomial chaos expansions, *Reliability Engineering & System Safety*, 95, 1216–1229, <https://doi.org/10.1016/j.res.2010.06.015>, 2010.
- Bogoslovsky, V. A. and Ogilvy, A. A.: Deformation of natural electric fields near drainage structures, *Geophys. Prospect.*, 21, 716–723, <https://doi.org/10.1111/j.1365-2478.1973.tb00053>, 1973.
- Bolève, A., Revil, A., Janod, F., Mattiuzzo, J. L., and Fry, J.-J.: Preferential fluid flow pathways in embankment dams imaged by self-potential tomography, *Near Surf. Geophys.*, 7, 447–462, <https://doi.org/10.3997/1873-0604.2009012>, 2009.
- Brown, P. N., Hindmarsh, A. C., and Petzold, L. R.: Using Krylov Methods in the Solution of Large-Scale Differential-Algebraic Systems, *SIAM J. Sci. Comp.*, 15, 1467–1488, <https://doi.org/10.1137/0915088>, 1994.
- Carsel, R. F. and Parrish, R. S.: Developing joint probability distributions of soil water retention characteristics, *Water Resour. Res.*, 24, 755–769, <https://doi.org/10.1029/WR024i005p00755>, 1988.
- Christensen, S. and Cooley, R. L.: Evaluation of confidence intervals for a steady-state leaky aquifer model, *Adv. Water Res.*, 22, 807–817, [https://doi.org/10.1016/S0309-1708\(98\)00055-4](https://doi.org/10.1016/S0309-1708(98)00055-4), 1999.
- Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G., and Schaibly, J. H.: Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I. theory, *J. Chem. Phys.*, 59, 3873–3878, 1973.
- Darnet, M., Marquis, G., and Sailhac, P.: Estimating aquifer hydraulic properties from the inversion of surface Streaming Potential (SP) anomalies, *Geophys. Res. Lett.*, 30, <https://doi.org/10.1029/2003GL017631>, 2003.
- Deng, B. and Wang, J.: Saturated-unsaturated groundwater modeling using 3D Richards equation with a coordinate transform of nonorthogonal grids, *Applied Mathematical Modelling*, 50, 39–52, [doi:10.1016/j.apm.2017.05.021](https://doi.org/10.1016/j.apm.2017.05.021); 2017.
- Donaldson, J. R. and Schnabel, R. B.: Computational Experience with Confidence Regions and Confidence Intervals for Nonlinear Least Squares, *Technometrics*, 29, <https://doi.org/10.2307/1269884>, 1987.
- Dostert, P., Efendiev, Y., and Mohanty, B.: Efficient uncertainty quantification techniques in inverse problems for Richards' equation using coarse-scale simulation models, *Adv. Water Res.*, 32, 329–339, <https://doi.org/10.1016/j.advwatres.2008.11.009>, 2009.
- Duan, Q. Y., Gupta, V. K., and Sorooshian, S.: Shuffled complex evolution approach for effective and efficient global minimization, *J. Optimiz. Theory App.*, 76, 501–521, <https://doi.org/10.1007/BF00939380>, 1993.
- Fahs, M., Younes, A., and Lehmann, F.: An easy and efficient combination of the Mixed Finite Element Method and the Method of Lines for the resolution of Richards' Equation, *Environ. Modell. Soft.*, 24, 1122–1126, <https://doi.org/10.1016/j.envsoft.2009.02.010>, 2009.
- Fahs, M., Younes, A., and Ackerer, P.: An Efficient Implementation of the Method of Lines for Multicomponent Reactive

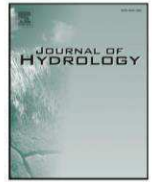
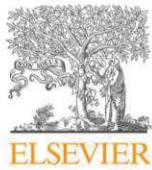
- Transport Equations, *Water Air. Soil Pollut.*, 215, 273–283, <https://doi.org/10.1007/s11270-010-0477-y>, 2011.
- Fajraoui, N., Ramasomanana, F., Younes, A., Mara, T. A., Ackerer, P., and Guadagnini, A.: Use of global sensitivity analysis and polynomial chaos expansion for interpretation of nonreactive transport experiments in laboratory-scale porous media, *Water Resour. Res.*, 47, <https://doi.org/10.1029/2010WR009639>, 2011.
- Fajraoui, N., Mara, T. A., Younes, A., and Bouhlila, R.: Reactive Transport Parameter Estimation and Global Sensitivity Analysis Using Sparse Polynomial Chaos Expansion, *Water Air. Soil Pollut.*, 223, 4183–4197, <https://doi.org/10.1007/s11270-012-1183-8>, 2012.
- Fajraoui, N., Fahs, M., Younes, A. and Sudret, B.: Analyzing natural convection in porous enclosure with polynomial chaos expansions: Effect of thermal dispersion, anisotropic permeability and heterogeneity, *Int. J. Heat Mass Trans.*, 115, 205–224, <https://doi.org/10.1016/j.ijheatmasstransfer.2017.07.003>, 2017.
- Gallagher, M. and Doherty, J.: Parameter estimation and uncertainty analysis for a watershed model, *Environ. Modell. Soft.*, 22, 1000–1020, <https://doi.org/10.1016/j.envsoft.2006.06.007>, 2007.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D.: *Bayesian Data Analysis*, Second Edition, London, Great Britain: Chapman Hall, 696 p., ISBN:0-158-48838-8, 1996.
- Haario, H., Laine, M., Mira, A., and Saksman, E.: DRAM: Efficient adaptive MCMC, *Statist. Comput.*, 16, 339–354, <https://doi.org/10.1007/s11222-006-9438-0>, 2006.
- Hastings, W. K.: *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*, *Biometrika*, 57, <https://doi.org/10.2307/2334940>, 1970.
- Hinnell, A. C., Ferré, T. P. A., Vrugt, J. A., Huisman, J. A., Moysey, S., Rings, J., and Kowalsky, M. B.: Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion, *Water Resour. Res.*, 46, <https://doi.org/10.1029/2008WR007060>, 2010.
- Ishido, T. and Mizutani, H.: Experimental and theoretical basis of electrokinetic phenomena in rock–water systems and its applications to geophysics, *J. Geophys. Res.*, 86, 1763–1775, <https://doi.org/10.1029/JB086iB03p01763>, 1981.
- Jardani, A., Revil, A., Bolève, A., Crespy, A., Dupont, J.-P., Barash, W., and Malama, B.: Tomography of the Darcy velocity from self-potential measurements, *Geophys. Res. Lett.*, 34, <https://doi.org/10.1029/2007GL031907>, 2007.
- Jougnot, D. and Linde, N.: Self-Potentials in Partially Saturated Media: The Importance of Explicit Modeling of Electrode Effects, *Vadose Zone J.*, 12, <https://doi.org/10.2136/vzj2012.0169>, 2013.
- Kahl, G. M., Sidorenko, Y., and Gottesbüren, B.: Local and global inverse modelling strategies to estimate parameters for pesticide leaching from lysimeter studies: Inverse modelling to estimate pesticide leaching parameters from lysimeter studies, *Pest Manage. Sci.*, 71, 616–631, <https://doi.org/10.1002/ps.3914>, 2015.
- Kayshap, R. L.: Optimal choice of AR and MA parts in autoregressive moving average models, *IEEE T. Pattern Anal.*, 4, 99–104, 1982.
- Kool, J. B., Parker, J. C., and van Genuchten, M. T.: Parameter estimation for unsaturated flow and transport models – A review, *J. Hydrol.*, 91, 255–293, [https://doi.org/10.1016/0022-1694\(87\)90207-1](https://doi.org/10.1016/0022-1694(87)90207-1), 1987.
- Laloy, E. and Vrugt, J. A.: High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing, *Water Resour. Res.*, 48, <https://doi.org/10.1029/2011WR010608>, 2012.
- Linde, N., Jougnot, D., Revil, A., Matthäi, S. K., Arora, T., Renard, D., and Doussan, C.: Streaming current generation in two-phase flow conditions, *Geophys. Res. Lett.*, 34, <https://doi.org/10.1029/2006GL028878>, 2007.
- Mara, T. A. and Tarantola, S.: Application of global sensitivity analysis of model output to building thermal simulations, *Build. Sim.*, 1, 290–302, <https://doi.org/10.1007/s12273-008-8129-5>, 2008.
- Mara, T. A., Belfort, B., Fontaine, V., and Younes, A.: Addressing factors fixing setting from given data: A comparison of different methods, *Environ. Modell. Soft.*, 87, 29–38, <https://doi.org/10.1016/j.envsoft.2016.10.004>, 2017.
- Mboh, C. M., Huisman, J. A., Zimmermann, E., and Vereecken, H.: Coupled Hydrogeophysical Inversion of Streaming Potential Signals for Unsaturated Soil Hydraulic Properties, *Vadose Zone J.*, 11, <https://doi.org/10.2136/vzj2011.0115>, 2012.
- Mertens, J., Kahl, G., Gottesbüren, B. and Vanderborcht, J.: Inverse Modeling of Pesticide Leaching in Lysimeters: Local versus Global and Sequential Single-Objective versus Multiobjective Approaches, *Vadose Zone J.*, 8, 793, <https://doi.org/10.2136/vzj2008.0029>, 2009.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of State Calculations by Fast Computing Machines, *J. Chem. Phys.*, 21, 1087–1092, <https://doi.org/10.1063/1.1699114>, 1953.
- Miller, C. T., Williams, G. A., Kelley, C. T., and Tocci, M. D.: Robust solution of Richards’ equation for nonuniform porous media, *Water Resour. Res.*, 34, 2599–2610, <https://doi.org/10.1029/98WR01673>, 1998.
- Mualem, Y.: A new model for predicting the hydraulic conductivity of unsaturated porous media, *Water Resour. Res.*, 12, 513–522, <https://doi.org/10.1029/WR012i003p00513>, 1976.
- Nützmann, G., Thiele, M., Maciejewski, S., and Joswig, K.: Inverse modelling techniques for determining hydraulic properties of coarse-textured porous media by transient outflow methods, *Adv. Water Res.*, 22, 273–284, [https://doi.org/10.1016/S0309-1708\(98\)00009-8](https://doi.org/10.1016/S0309-1708(98)00009-8), 1998.
- Patella, D.: Introduction to ground surface self-potential tomography, *Geophys. Prospect*, 45, 653–681, <https://doi.org/10.1046/j.1365-2478.1997.430277>, 1997.
- Rajabi, M. M., Ataie-Ashtiani, B., and Simmons, C. T.: Polynomial chaos expansions for uncertainty propagation and moment independent sensitivity analysis of seawater intrusion simulations, *J. Hydrol.*, 520, 101–122, <https://doi.org/10.1016/j.jhydrol.2014.11.020>, 2015.
- Revil, A., Linde, N., Cerepi, A., Jougnot, D., Matthäi, S., and Finsterle, S.: Electrokinetic coupling in unsaturated porous media, *J. Colloid Interface Sci.*, 313, 315–327, <https://doi.org/10.1016/j.jcis.2007.03.037>, 2007.
- Richards, K., Revil, A., Jardani, A., Henderson, F., Batzle, M., and Haas, A.: Pattern of shallow ground water flow at Mount Princeton Hot Springs, Colorado, using geoelectric methods, *J. Volcanol. Geotherm. Res.*, 198, 217–232, <https://doi.org/10.1016/j.jvolgeores.2010.09.001>, 2010.
- Riva, M., Guadagnini, A., and Dell’Oca, A.: Probabilistic assessment of seawater intrusion under multiple sources of uncertainty, *Adv. Water Res.*, 75, 93–104, <https://doi.org/10.1016/j.advwatres.2014.11.002>, 2015.

- Sailhac, P. and Marquis, G.: Analytic potentials for the forward and inverse modeling of SP anomalies caused by subsurface fluid flow, *Geophys. Res. Lett.*, 28, 1851–1854, <https://doi.org/10.1029/2000GL012457>, 2001.
- Saltelli, A., Tarantola, S., and Chan, K. P.-S.: A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output, *Technometrics*, 41, 39–56, <https://doi.org/10.1080/00401706.1999.10485594>, 1999.
- Shao, Q., Younes, A., Fahs, M., and Mara, T. A.: Bayesian sparse polynomial chaos expansion for global sensitivity analysis, *Comput. Method Appl. M.*, 318, 474–496, <https://doi.org/10.1016/j.cma.2017.01.033>, 2017.
- Sobol', I. M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Sim.*, 55, 271–280, [https://doi.org/10.1016/S0378-4754\(00\)00270-6](https://doi.org/10.1016/S0378-4754(00)00270-6), 2001.
- Sobol', I. M.: Sensitivity estimates for nonlinear mathematical models, *Math. Model. Comput. Exp.*, 407–414, 1993.
- Sudret, B.: Global sensitivity analysis using polynomial chaos expansions, *Reliab. Eng. Syst. Safe.*, 93, 964–979, <https://doi.org/10.1016/j.res.2007.04.002>, 2008.
- ter Braak, C. J. F.: A Markov Chain Monte Carlo version of the genetic algorithm differential evolution: Easy Bayesian computing for real parameter spaces, *Stat Comput.*, 16, 239–249, <https://doi.org/10.1007/s11222-006-8769-1>, 2006.
- ter Braak, C. J. F. and Vrugt, J. A.: Differential Evolution Markov Chain with snooker updater and fewer chains, *Statist. Comput.*, 18, 435–446, <https://doi.org/10.1007/s11222-008-9104-9>, 2008.
- van Dam, J. C., Stricker, J. N. M., and Droogers, P.: Inverse Method to Determine Soil Hydraulic Functions from Multistep Outflow Experiments, *Soil Sci. Soc. Am. J.*, 58, <https://doi.org/10.2136/sssaj1994.03615995005800030002x>, 1994.
- van Genuchten, M. T.: A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils, *Soil Sci. Soc. Am. J.*, 44, 892, <https://doi.org/10.2136/sssaj1980.03615995004400050002x>, 1980.
- Volkova, E., Iooss, B., and Van Dorpe, F.: Global sensitivity analysis for a numerical model of radionuclide migration from the RRC “Kurchatov Institute” radwaste disposal site, *Stoch. Env. Res. Risk A.*, 22, 17–31, <https://doi.org/10.1007/s00477-006-0093-y>, 2008.
- Vrugt, J. A., Gupta, H. V., Bouten, W. and Sorooshian, S.: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, 39, doi:10.1029/2002WR001642, 2003.
- Vrugt, J. A. and Bouten, W.: Validity of First-Order Approximations to Describe Parameter Uncertainty in Soil Hydrologic Models, *Soil Sci. Soc. Am. J.*, 66, <https://doi.org/10.2136/sssaj2002.1740.2002>.
- Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., and Robinson, B. A.: Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation: FORCING DATA ERROR USING MCMC SAMPLING, *Water Resour. Res.*, 44, <https://doi.org/10.1029/2007WR006720>, 2008.
- Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., and Higdon, D.: Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *Int. J. Nonlinear Sci. Numer. Simul.*, 10, 273–290, <https://doi.org/10.1515/IJNSNS.2009.10.3.273>, 2009a.
- Vrugt, J. A.: Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation, *Environ. Modell. Softw.*, 75, 273–316, <https://doi.org/10.1016/j.envsoft.2015.08.013>, 2016.
- Vrugt, J. A., Robinson, B. A., and Hyman, J. M.: Self-adaptive multimethod search for global optimization in real parameter spaces, *IEEE Trans. Evol. Comput.*, 13, 243–259, <https://doi.org/10.1109/TEVC.2008.924428>, 2009b.
- Vugrin, K. W., Swiler, L. P., Roberts, R. M., Stucky-Mack, N. J., and Sullivan, S. P.: Confidence region estimation techniques for nonlinear regression in groundwater flow: Three case studies, *Water Resour. Res.*, 43, W03423, <https://doi.org/10.1029/2005WR004804>, 2007.
- Wiener, N.: The Homogeneous Chaos, *Am. J. Math.*, 60, <https://doi.org/10.2307/2371268>, 1938.
- Younes, A., Fahs, M., and Ahmed, S.: Solving density driven flow problems with efficient spatial discretizations and higher-order time integration methods, *Adv. Water Res.*, 32, 340–352, <https://doi.org/10.1016/j.advwatres.2008.11.003>, 2009.
- Younes, A., Fahs, M., and Belfort, B.: Monotonicity of the cell-centred triangular MPFA method for saturated and unsaturated flow in heterogeneous porous media, *J. Hydrol.*, 504, 132–141, <https://doi.org/10.1016/j.jhydrol.2013.09.041>, 2013.
- Younes, A., Mara, T. A., Fajraoui, N., Lehmann, F., Belfort, B., and Beydoun, H.: Use of Global Sensitivity Analysis to Help Assess Unsaturated Soil Hydraulic Parameters, *Vadose Zone J.*, 12, <https://doi.org/10.2136/vzj2011.0150>, 2013.
- Younes, A., Delay, F., Fajraoui, N., Fahs, M., and Mara, T. A.: Global sensitivity analysis and Bayesian parameter inference for solute transport in porous media colonized by biofilms, *J. Contam. Hydrol.*, 191, 1–18, <https://doi.org/10.1016/j.jconhyd.2016.04.007>, 2016.
- Younes, A., Mara, T., Fahs, M., Grunberger, O., and Ackerer, P.: Hydraulic and transport parameter assessment using column infiltration experiments, *Hydrol. Earth Syst. Sci.*, 21, 2263–2275, <https://doi.org/10.5194/hess-21-2263-2017>, 2017.
- Zablocki, C. J.: Streaming potentials resulting from the descent of meteoric water: A possible source mechanism for Kilauean self-potential anomalies, *Trans. Geotherm. Resour. Council.*, 2, 747–748, 1978.

# Chapitre 4. Estimation bayésienne des paramètres d'un sol non remanié à partir d'expériences de drainage et d'infiltration

## Plan du chapitre

Chapitre 4. Estimation bayésienne des paramètres d'un sol non remanié à partir d'expériences de drainage et d'infiltration .....	116
4.1 Introduction .....	117
4.2 Modèles expérimentaux et mathématiques .....	118
4.2.1 Expériences de laboratoire .....	118
4.3 Modèle numérique .....	118
4.4 Inférence bayésienne de paramètre .....	119
4.5 Résultats et discussion .....	120
4.6 Conclusion .....	124



Research papers

## Bayesian soil parameter estimation: Results of percolation-drainage vs infiltration laboratory experiments



A Younes<sup>a,b,c</sup>, J. Zaouali<sup>a</sup>, M. Fahs<sup>a,\*</sup>, F. Slama<sup>c</sup>, O. Grunberger<sup>b</sup>, T.A. Mara<sup>d,e</sup>

<sup>a</sup> LHyGES, Univ. de Strasbourg/EOST/ENGEES, CNRS, 1 rue Blessig, 67084 Strasbourg, France

<sup>b</sup> LISAH, Univ Montpellier, INRA, IRD, Montpellier SupAgro, Montpellier, France

<sup>c</sup> LMHE, Ecole Nationale d'Ingénieurs de Tunis, Tunisia

<sup>d</sup> PIMENT, EA 4518, Université de La Réunion, FST, 15 Avenue René Cassin, 97715 Saint-Denis, Réunion, France

<sup>e</sup> Directorate for Modelling, Indicators and Impact Evaluation, Joint Research Centre, European Commission, 21027 Ispra (VA), Italy

### ARTICLE INFO

#### Keywords:

Infiltration experiment  
Drainage experiment  
Richards' equation  
Markov Chain Monte Carlo  
Uncertainty ranges

### ABSTRACT

In this work, we conducted two laboratory column experiments on an undisturbed sandy soil. The first deals with a percolation-drainage experiment whereas the second deals with an infiltration of a constant water flux at the surface of the unsaturated soil. A Bayesian assessment of the soil parameters is performed for both experiments with the Markov Chain Monte Carlo (MCMC) method using measurements of pressure head inside the column and cumulative outflow collected during the experiments. The results show that both experiments can be well reproduced with the mathematical model based on the Richards equation and the van-Genuchten/Mualem models. Furthermore, the inversion of the two laboratory experiments yields similar results in terms of mean estimated parameter values but strong discrepancies occur between confidence intervals used to quantify uncertainty on the estimated parameters. Compared to the percolation-drainage experiment, the infiltration experiment yields more accurate parameters with narrower uncertainty regions.

### 1. Introduction

Accurate knowledge of hydraulic soil parameters is essential before simulation of water flow and/or contaminant transport through the unsaturated zone. Generally, the soil parameters are not known a priori and are estimated using inversion methods by fitting the simulated model responses to the observed data. To this aim, undisturbed soil samples are usually collected from the field and moved to laboratory to perform experiments under controlled initial and boundary conditions.

In the literature, laboratory dynamic flow experiments have been largely used to estimate unsaturated hydraulic soil parameters (e.g., van Dam et al., 1994; Hopmans et al., 2002; Beydoun and Lehmann, 2006; Schelle et al., 2010; Younes et al., 2013; Moreira et al., 2016). Many types of transient flow experiments with different boundary conditions have been investigated in the literature for a reliable estimation of the soil properties. Among the laboratory flow experiments, we can mention the one-step outflow experiment (Kool et al., 1985; Parker et al., 1985), the multistep outflow experiment (van Dam et al., 1994; Durner et al., 1999; Hopmans et al., 2002; Puhmann et al., 2009), controlled infiltration experiments (Mishra and Parker, 1989; Hudson et al., 1996), extended multistep-outflow experiments (Durner and Iden,

2011), evaporation experiments (Wind, 1968; Wendroth et al., 1993; Kumar et al., 2010), a combination of extended multistep-outflow and evaporation experiments (Schelle et al., 2011). Van Dam et al. (1994) showed that multistep-outflow experiments yielded more accurate results than one-step outflow experiments. Durner and Iden (2011) showed that the multistep outflow experiment can yield large uncertainties, in particular with respect to the determination of the saturated hydraulic conductivity. To overcome this drawback, they proposed the extended multistep outflow experiment where a saturated percolation phase is combined to the unsaturated multistep outflow experiment (Durner and Iden, 2011). Schelle et al. (2010) have compared hydraulic properties obtained from multistep-outflow and evaporation methods. They found that the two methods yield similar results in terms of estimated parameter values when the observations are well reproduced by the simulations. If this condition is not fulfilled, a disagreement is observed between the results of the two methods.

In the literature few studies have been devoted to investigate the agreement of the soil properties obtained from different experiment types (e.g. Stolte et al., 1994; Eching et al., 1994; Garnier et al., 1997; Schelle et al., 2010). Furthermore, the comparison is often based on the mean estimated parameter values and omits the associated

\* Corresponding author.

E-mail address: [fahs@unistra.fr](mailto:fahs@unistra.fr) (M. Fahs).

<https://doi.org/10.1016/j.jhydrol.2018.08.082>

Available online 05 September 2018

0022-1694/ © 2018 Elsevier B.V. All rights reserved.

uncertainties. In this work, we are concerned with the comparison of the quality of parameter estimate, both in terms of mean estimated parameter values and uncertainty ranges, between two types of laboratory experiments carried out on the same undisturbed sandy soil column. The first experiment deals with a multi-step outflow experiment where water is extracted from the initially saturated soil column by applying different suctions at the lower soil boundary. The second experiment deals with the infiltration of a constant water flux at the surface of the initially unsaturated soil. For both experiments, the cumulative outflow and the pressure head in three locations inside the column are monitored. The cumulative outflow and pressure head data were shown to be sufficient for estimating the hydraulic soil parameters for a drainage experiment (Younes et al., 2013). The unsaturated flow through the undisturbed soil column for both experiments is first checked to be governed by the Richards equation and the standard Mualem/van Genuchten models (Mualem, 1976; van Genuchten, 1980). Then, the hydraulic soil properties are sought in a Bayesian framework using the Markov Chain Monte Carlo (MCMC) sampler based on the Metropolis-Hastings (Hastings, 1970; Metropolis et al., 1953; Lee, 2004) algorithm. The results of the two experiments are then confronted in terms of mean estimated parameter values and uncertainty ranges.

## 2. Materials and methods

### 2.1. Laboratory experiments

A cylindrical soil column (length 44 cm, diameter 15 cm) was carefully extracted from the BEKALTA province located in the east center of Tunisia (35°35'24.3"N, 11°00'15.8"E) by inserting in the soil a 60 cm rigid PVC tube beveled to form a cutting edge. The soil is composed of 93% sand (2–0.05 mm), 5% silt (0.05–0.002 mm) and 2% clay particles (< 0.002 mm). The soil column is set on a container filled with gravel (Fig. 1). The container is sealed to the PVC column with the exception of an orifice located face to the soil bottom and connected to a pressure reservoir which allows to fix the pressure head  $h$  [L] at the column lower boundary.

The first laboratory experiment (Fig. 1) is similar to the extended multistep outflow experiment proposed in Durner and Iden (2011). Initially, the soil is progressively saturated from the bottom until reaching a ponding of 3 cm. To avoid disturbance of the soil, the saturation was carried out during a long period of 24 h by slowly increasing the pressure head at the bottom of the column using the pressure reservoir. Three pressure sensors monitored the pressure head near the top ( $Z_1 = 3.5$  cm), middle ( $Z_2 = 18.5$  cm) and bottom of the column ( $Z_3 = 33.5$  cm).  $Z$  is the vertical downward axis with the origin located at the soil surface. The accuracies of the tensiometers (METER T5 tensiometers) were approximately  $\pm 2$  cm. The cumulative outflow was measured with an accuracy of  $\pm 2$  g using a digital balance. Error measurements for the two data series are assumed to be independently and normally distributed with zero means and variances  $\sigma_{h_i} = 1$  cm for the pressure head and  $\sigma_Q = 1$  g for the cumulative outflow.

During the first 80 min, the head  $H = h - z$  at the lower boundary of the column was fixed to 3 cm by placing the pressure reservoir 3 cm upper the soil surface. To prevent evaporation, the upper end of the PVC column is sealed with the exception of a small orifice to allow equilibrium with atmospheric pressure. Then, six head steps corresponding to a head  $H$  of 0.5, -5, -10, -19, -30 and -43 cm at respectively,  $t_1=80$ ,  $t_2=752$ ,  $t_3=888$ ,  $t_4=968$ ,  $t_5=1066$  and  $t_6=1164$  min were successively applied at the bottom of the soil column by moving down the pressure reservoir. During percolation, the ponding was filmed with a camera to monitor the water level decrease at the surface of the soil. The column remained saturated during the first 741 min. During this saturated phase, the ponding decreases exponentially (Durner and Iden, 2011) until it vanishes. After that, the drainage process replaces the percolation process.

The second Laboratory experiment (Fig. 2) deals with an infiltration experiment. The lower boundary condition is maintained at a head of -44 cm cm (i.e. a zero pressure head  $h$  is fixed at the bottom of the column during the whole experiment to maintain the pressure reservoir at the same level than the soil bottom). Initial conditions correspond to hydrostatic equilibrium (i.e. the value of the suction given by each tensiometer corresponds to its distance to the bottom of the column). Then, a constant water flow rate of 20 cm<sup>3</sup>/min was applied to the top surface of the soil using a Masterflex® L/S peristaltic pump. Cotton fiber wicks were applied at the surface of the column to allow a uniform distribution of the flow rate over the area of the column. The three pressure sensors were located far from the surface (at  $Z_1 = 18.5$  cm,  $Z_2 = 23.5$  cm and  $Z_3 = 33.5$  cm) to insure a relatively uniform distribution of water across the area of the column. The infiltration occurs during a period of 57.5 min. It starts at  $t = 17.5$  min and was stopped at  $t = 75$  min. The tensiometric data and the cumulative outflow were monitored until  $t = 117.5$  min.

### 3. Numerical model

The saturated-unsaturated water flow through the cylindrical soil column is assumed to be one-dimensional and governed by the following Richards Equation (RE):

$$\begin{cases} \frac{\partial \theta}{\partial t} = \left( c(h) + S_e \frac{\partial}{\partial z} \right) \frac{\partial h}{\partial t} = -\frac{\partial q}{\partial z} \\ q = -K(h) \left( \frac{\partial h}{\partial z} - 1 \right) \end{cases} \quad (1)$$

where  $\theta$  [L<sup>3</sup> L<sup>-3</sup>] is the water content,  $t$  [T<sup>-1</sup>] is time,  $c(h)$  [L<sup>-1</sup>] is the specific moisture capacity,  $S_e$  [-] is the specific storage,  $\theta_s$  [L<sup>3</sup> L<sup>-3</sup>] is the saturated water content,  $q$  [L T<sup>-1</sup>] is the Darcy's velocity,  $z$  [L] is the positive downward vertical coordinate with the origin at the soil surface and  $K(h)$  [L T<sup>-1</sup>] is the hydraulic conductivity. The standard models of Mualem (1976) and van Genuchten (1980) are used to describe the nonlinear relationships between pressure head, conductivity and water content,

$$\begin{aligned} S_e(h) &= \frac{\theta(h) - \theta_r}{\theta_s - \theta_r} = \begin{cases} \frac{1}{(1 + |\alpha h|)^m} & h < 0 \\ 1 & h \geq 0 \end{cases} \\ K(S_e) &= K_s S_e^{1/2} [1 - (1 - S_e^{1/m})^m]^2 \end{aligned} \quad (2)$$

where  $S_e$  (-) is the effective saturation,  $\theta_r$  [L<sup>3</sup> L<sup>-3</sup>] is the residual water content,  $K_s$  [L T<sup>-1</sup>] is the saturated hydraulic conductivity,  $m = 1 - 1/n$ ,  $\alpha$  [L<sup>-1</sup>] and  $n$  [-] are the Mualem van Genuchten shape parameters.

In Eq. (2), the pore-connectivity parameter is fixed to 0.5 since we deal with a sandy soil. For general soils, this parameter could be taken unknown and its value estimated from the Bayesian inversion as in Mara et al. (2016). The system of Eqs. (1) and (2) is solved with the corresponding initial and boundary conditions. For the percolation-drainage experiment, the initial condition is a hydrostatic pressure distribution corresponding to the pressure head of 3 cm at the surface of the column. The boundary conditions are as follows: at the upper boundary, a Dirichlet condition corresponding to the measured ponding for the percolation phase ( $t < 741$ ) followed by a Neumann condition, with a zero flux, for the drainage phase. At the lower boundary, a Dirichlet condition at the bottom column corresponding to the five applied pressure head steps.

For the infiltration experiment, the initial condition is a hydrostatic pressure distribution with a zero pressure at the bottom of the column. The upper boundary condition is a Neumann condition with a zero flux except during the irrigation period (17.5 min <  $t$  < 117.5 min) where the upper flux is  $Q_{inj} = 0.115$  cm/min. The lower boundary condition is of Dirichlet type with a zero pressure head maintained at the bottom of the column during the whole experiment.

The system (1)–(2) with the corresponding initial and boundary conditions is solved using the standard finite volume method for the



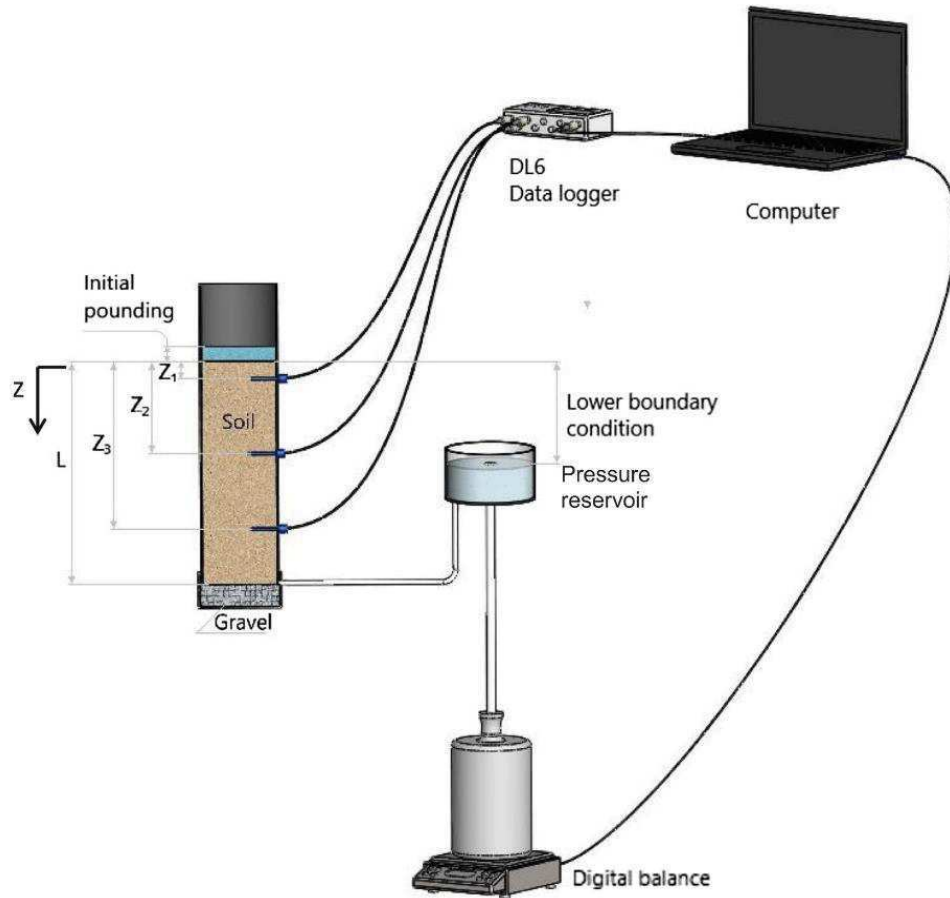


Fig. 1. Schematic representation of the experimental device: percolation-drainage experiment.

spatial discretization. A mesh independent solution is obtained using a uniform mesh of 215 cells of 0.2 cm. The high-order method of lines (MOL) is used for the temporal discretization of the  $h$  form of the RE via the DLSODIS (Double-precision Livermore Solver for ordinary differential equations – implicit form and sparse matrix) time solver (Hindmarsh, 1980). MOL is well suited for solving the nonlinear RE (e.g., Younes et al., 2009; Fahs et al., 2011; Farthing and Ogden, 2017) since it allows variable time-step and variable order integration with time error control (Miller et al., 2006). The algebraic nonlinear system is solved with the modified Newton iteration method (Radhakrishnan and Hindmarsh, 1993).

#### 4. Bayesian parameter inference

Two types of observations were collected during the experiments. The tensiometers provided the pressure head at three different locations inside the column. The digital balance provided the cumulative outflow at the column output. Hence, the vector of observations  $y_{mes}$  is formed by these two types of data. All unsaturated soil parameters are sought by inverse modeling except the saturated water content which was accurately measured by weighing the saturated soil ( $\theta_s = 0.417 \text{ cm}^3 \text{ cm}^{-3}$ ). A very small value is used for the specific storage ( $S_s = 10^{-8} \text{ cm}^{-1}$ ). Hence, the vector of unknowns is  $\xi = (K_s, \theta_r, \alpha, n)$ . These parameters are estimated using a Bayesian approach where the observations are merged with the prior parameter knowledge to define the joint posterior probability distribution function (jpdf) of the parameters. In this work, the jpdfs are evaluated using the DREAM<sub>(ZS)</sub> software (Laloy and Vrugt, 2012) relying on the Markov

Chain Monte Carlo (MCMC) sampler. The MCMC method has been largely used in hydrogeology (e.g. Vrugt et al., 2008; Laloy and Vrugt, 2012; Moreira et al., 2016; Linde et al., 2017; Younes et al., 2017; Hayek et al., 2018). With MCMC, the generated random sequences of parameter sets converge asymptotically toward the target distribution (Gelman et al., 1997). The statistical measures of the parameters such as the mean estimated values and the variances are then computed using the converged sets.

Using the Bayes theorem, the probability density function of the parameters conditioned onto observations write:

$$p(\xi|y_{mes}) \propto p(y_{mes}|\xi)p(\xi), \quad (3)$$

with  $p(\xi|y_{mes})$  the likelihood function measuring how well the model outputs are in agreement with the observations  $y_{mes}$ , and  $p(\xi)$  is the prior information on the parameters.

In this work, independent uniform distributions are assumed for all parameters. The non-informative uniform distributions are used here to express the absence of prior information which makes all possible values of the parameter equally likely. The parameter intervals (Table 1) are based on the literature and are chosen large enough to reflect our poor prior knowledge of the model parameter values.

Under the assumption that the error measurements for the two data series are independently and normally distributed, the joint posterior distribution matches the likelihood function, that is,

$$p(\xi|y_{mes}) \propto \sigma_h^{-N_h} \sigma_Q^{-N_Q} \exp\left(-\frac{SS_h(\xi)}{2\sigma_h^2} - \frac{SS_Q(\xi)}{2\sigma_Q^2}\right) \quad (4)$$

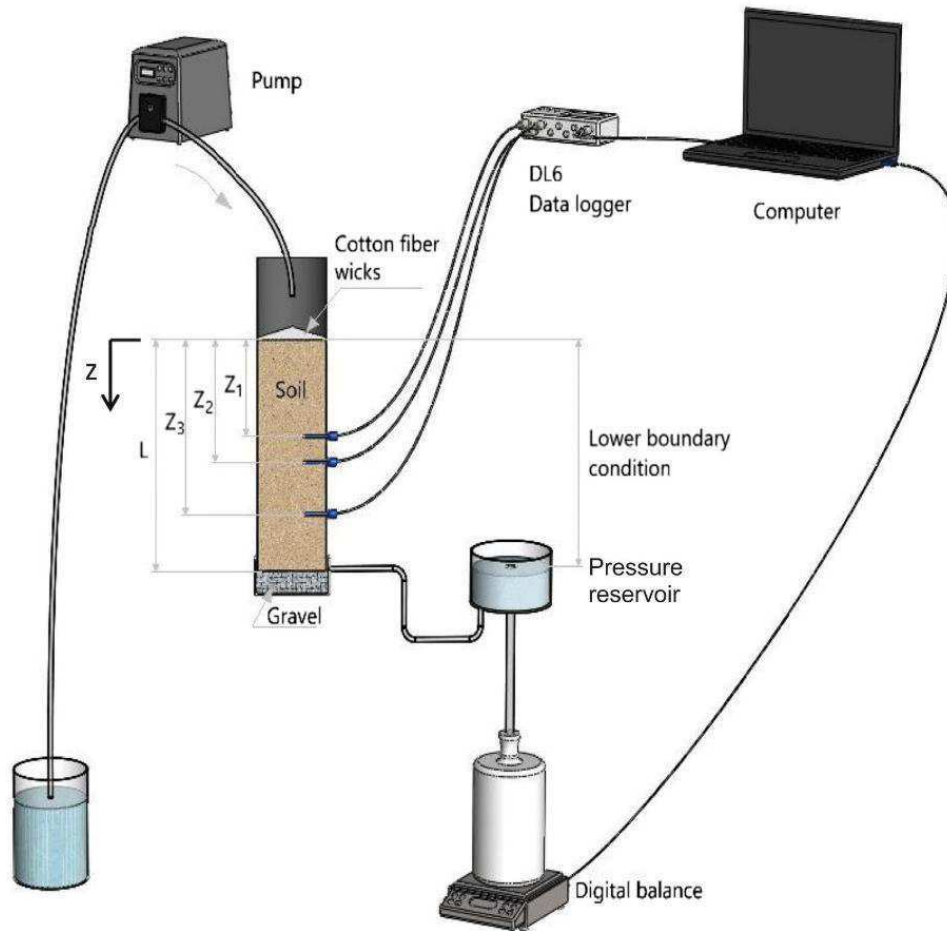


Fig. 2. Schematic representation of the experimental device: infiltration experiment.

**Table 1**  
Lower and upper bounds of the soil parameters.

Parameter	Lower bound	Upper bound
$K_S$ [cm/min]	0.1	2.0
$\theta_r$ [-]	0.05	0.15
$\alpha$ [cm <sup>-1</sup> ]	0.005	1.0
$n$ [-]	1.3	7.0

where  $SS_h(\xi)$  and  $SS_Q(\xi)$  are the sums of the squared differences between the modeled and observed data of the pressure head and cumulative outflow, respectively.

For instance,  $SS_h(\xi) = \sum_{k=1}^{N_h} (h_{mes}^{(k)} - h_{mod}^{(k)}(\xi))^2$  (resp.  $SS_Q$ ), which includes the observed  $h_{mes}^{(k)}$  (resp.  $Q_{mes}^{(k)}$ ) and predicted  $h_{mod}^{(k)}$  (resp.  $Q_{mod}^{(k)}$ ) pressure head (resp. cumulative outflow) at time  $t_k$  for the number of cumulative outflow (resp. concentration) observations  $N_h$  (resp.  $N_Q$ ).

The MCMC sampler is based on the Metropolis-Hasting inference. In the latter, a new candidate  $\xi^i$  is generated at the  $i^{\text{th}}$  iteration from a proposal distribution  $q(\xi^i | \xi^{i-1})$  which only depends on the previous accepted candidate. The new candidate is accepted or rejected depending on the associated Hasting ratio, defined as,

$$\alpha = \min\left(1, \frac{p(\xi^i | y_{mes}) q(\xi^{i-1} | \xi^i)}{p(\xi^{i-1} | y_{mes}) q(\xi^i | \xi^{i-1})}\right) \quad (5)$$

The last 25% sets (when the chains have converged) adequately fit the model onto observations. These sets are then used to assess the

mean estimated parameter values and the associated confidence intervals. In this work, three parallel chains were used with the DREAM<sub>(ZS)</sub> MCMC sampler. The posterior distribution is considered stationary if the Gelman and Rubin (1992) criterion is less than 1.2 and the chains are not autocorrelated.

## 5. Results and discussion

In this section, identifiability of the parameters is investigated for the percolation-drainage and the infiltration experiments. The MCMC results in term of mean estimated parameter values and confidence intervals are depicted in Table 2 for the percolation-drainage experiment and in Table 3 for the infiltration experiment.

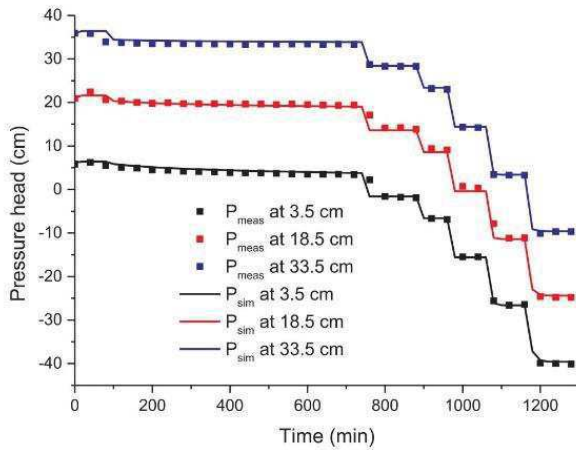
The results of the two experiments show a good agreement between the mean estimated values for the parameters  $K_S$ ,  $\theta_r$  and  $\alpha$ . For the parameter  $n$ , a reasonable agreement with a difference of around 30% is

**Table 2**  
Estimated mean values, confidence intervals (CIs) and size of the posterior CIs for the percolation-drainage experiment.

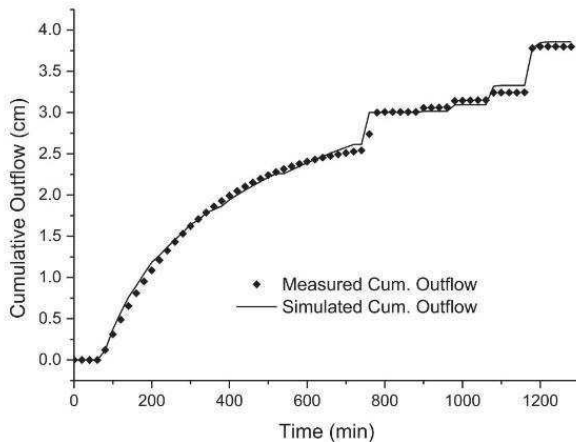
	Unit	Mean estimated value	Confidence Interval	Size of the CI
$K_S$	(cm/min)	0.22	[0.13–1.24]	1.11
$\theta_r$	-	0.1	[0.05–0.15]	0.1
$\alpha$	(cm <sup>-1</sup> )	0.016	[0.013–0.019]	0.006
$n$	-	1.9	[1.34–3.12]	1.78

**Table 3**  
Estimated mean values, confidence intervals (CIs) and size of the posterior CIs for the infiltration experiment.

	Unit	Mean estimated value	Confidence Interval	Size of the CI
$K_S$	(cm/min)	0.21	[0.2–0.22]	0.02
$\theta_r$	-	0.08	[0.05–0.15]	0.1
$\alpha$	( $\text{cm}^{-1}$ )	0.012	[0.011–0.014]	0.003
$n$	-	2.7	[2.4–2.9]	0.5

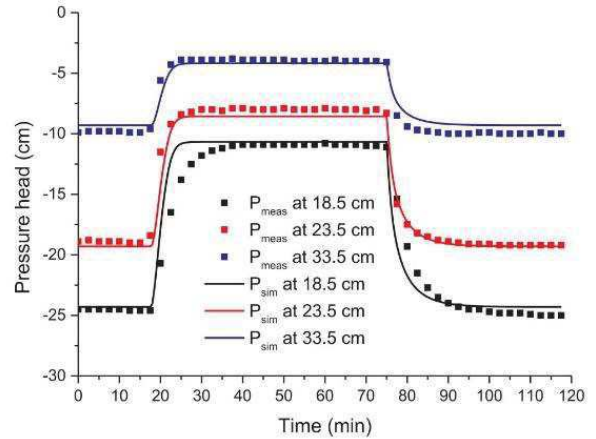


**Fig. 3.** Observed and calibrated pressure head at  $Z_1 = 3.5$  cm,  $Z_2 = 18.5$  cm and  $Z_3 = 33.5$  cm from the soil surface for the percolation-drainage experiment.



**Fig. 4.** Observed and calibrated cumulative outflow for the percolation-drainage experiment.

observed between the two experiments. The obtained mean estimated values are then used to simulate the percolation-drainage and the infiltration experiments. Fig. 3 shows the simulated and measured pressure head by the three tensiometers for the percolation-drainage experiment. In this figure, the pressure head curve of each tensiometer starts at a hydrostatic pressure and then decreases by step due to the different suctions applied at the bottom of the column. A very good agreement can be observed between simulated and measured pressure heads for the three tensiometers. In the Fig. 4, the cumulative outflow displays a parabolic behavior during the percolation phase followed by an increase by step due to the suction caused by the different pressure head levels applied at the lower boundary. The results of Fig. 4 show a good agreement between the simulated and measured cumulative outflow. Therefore, the Figs. 3 and 4 show that using the mean

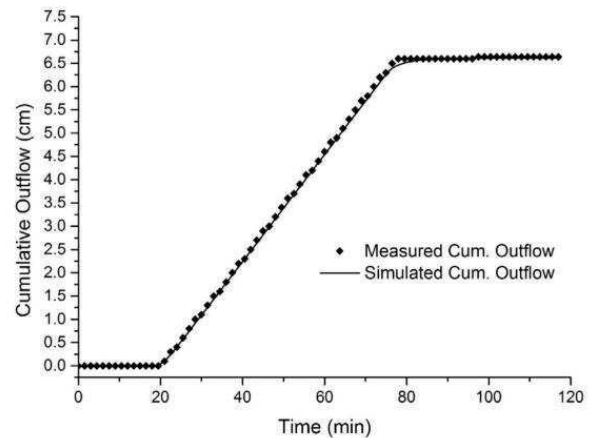


**Fig. 5.** Observed and calibrated pressure head at  $Z_1 = 18.5$  cm,  $Z_2 = 23.5$  cm and  $Z_3 = 33.5$  cm from the soil surface for the infiltration experiment.

estimated parameter values, the mathematical model based on the Richards Equation and the Mualem/van-Genuchten models reproduces relatively well the saturated unsaturated flow processes occurring during the percolation-drainage experiment through the studied undisturbed soil.

In the case of the infiltration experiment, an increase of the pressure head is observed (when the infiltration front reaches the tensiometers) followed by a decrease in the wake of the end of the surface injection (Fig. 5). During the injection, the pressure head measured by the tensiometers is almost constant. A good agreement is observed between the calculated (using mean estimated parameter values) and the measured pressure-heads by the three tensiometers (Fig. 5). Fig. 6 shows that the cumulative outflow starts to increase when the infiltration front reaches the bottom of the column and shows a linear behavior until the end of the period of injection. A very good agreement is observed between the measured and simulated cumulative outflow. Hence, the results of the Figs. 5 and 6 confirm that the used mathematical model is reliable to reproduce the infiltration experiment on the investigated undisturbed soil.

At first sight, and based on the comparison between the mean estimated parameter values, one could think that a similar quality of estimation is obtained from the two experiments. In order to further analyze the quality of the estimation, we compare, in the following, the results of the two experiments in terms of uncertainty associated to



**Fig. 6.** Observed and calibrated cumulative outflow for the infiltration experiment.

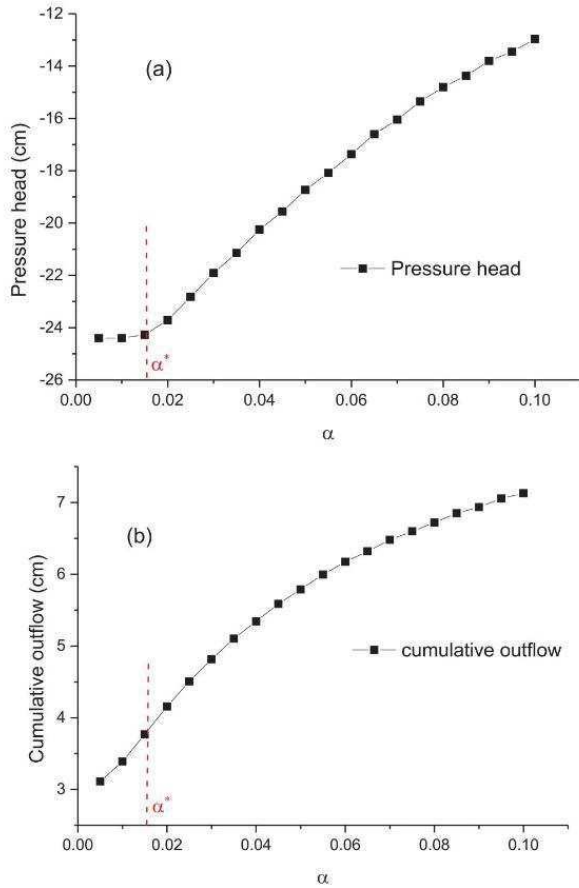


Fig. 7. Marginal effects of the Mualem-van Genuchten shape parameter  $\alpha$  on the pressure head (a) and cumulative outflow (b) at  $t = 1200$  min for the percolation-drainage experiment ( $\alpha^*$  is the mean estimated value).

each estimated parameter.

For the parameter  $\theta_r$ , the results of Tables 2 and 3 show that both the percolation-drainage and the infiltration experiments yield posterior intervals similar to the prior interval in Table 1. This indicates that the residual water content is not identified with neither of the experiments. This non-identifiability is due to the very low sensitivity of the pressure head and the cumulative outflow to  $\theta_r$ . Indeed the results of simulations of the two experiments using different values of  $\theta_r$  yield similar pressure head and cumulative outflow curves. This weak sensitivity is due to the fact that the residual water content was not attained during neither of the laboratory experiments.

The parameter  $\alpha$  is well estimated from the two experiments. Indeed, a narrow uncertainty region is associated to this parameter due to its high influence on the model outputs for both experiments. The influence of  $\alpha$  on the pressure head is especially observable during the transition to dry conditions at  $t = 1200$  min (right after the last pressure head step applied at the lower boundary). Fig. 7 shows the simulated pressure head of the second tensiometer (at 18.5 cm from the soil surface) and the cumulative outflow at  $t = 1200$  min for different values of  $\alpha$  taken from the interval  $[0.005, 0.1] \text{ cm}^{-1}$ . All the other parameters were fixed to their mean estimated values obtained from MCMC results. A similar behavior is observed for the effect of  $\alpha$  on the pressure head of the other tensiometers. The results of Fig. 7 show that the pressure head is weakly sensitive to  $\alpha$  in the region surrounding the mean estimated value  $\alpha^*$  (Fig. 7a) whereas, the cumulative outflow is highly sensitive to  $\alpha$  in that region (Fig. 7b). This suggests that the cumulative outflow measurements contribute the most to the

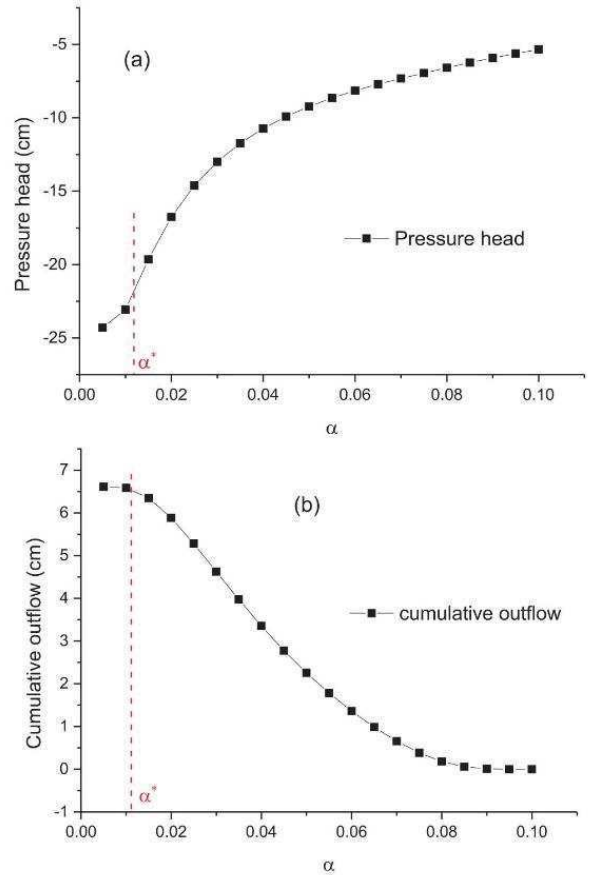


Fig. 8. Marginal effects of the Mualem-van Genuchten shape parameter  $\alpha$  on the pressure head (a) and cumulative outflow (b) at  $t = 80$  min for the infiltration experiment ( $\alpha^*$  is the mean estimated value).

identifiability of  $\alpha^*$  for the percolation-drainage experiment.

The influence of the parameter  $\alpha$  for the infiltration experiment is especially noticeable during the transition to dry conditions at  $t = 80$  min (just after stopping the injection). Fig. 8 shows the marginal effect of the parameter  $\alpha$  on the pressure head (Fig. 8a) and the cumulative outflow at  $t = 80$  min (Fig. 8b). The cumulative outflow is weakly sensitive to  $\alpha$  in the region surrounding  $\alpha^*$ , whereas the pressure head is highly sensitive to  $\alpha$  in that region. This suggests that the pressure head measurements contribute the most to the identifiability of  $\alpha^*$  for the infiltration experiment.

The parameter  $n$  is not well estimated from the percolation-drainage experiment. Indeed, the MCMC sampler yielded a mean estimated value of  $n^* = 1.9$  with a large confidence interval  $[1.34\text{--}3.12]$ . Fig. 9 shows that the pressure head is weakly sensitive to  $n$  (Fig. 9a) whereas the cumulative outflow is almost not sensitive to  $n$  in the region surrounding  $n^*$  (Fig. 9b).

The infiltration experiment yielded a mean estimated value  $n^* = 2.7$  with a reduced confidence interval  $[2.4\text{--}2.9]$ . For this experiment, both pressure head (Fig. 10a) and cumulative outflow (Fig. 10b) are sensitive to  $n$  in the region surrounding  $n^*$ . The pressure head is highly influenced by the value of the parameter  $n$ . For instance, the variation of  $n$  from 2.1 to 3.1 induces a variation of 2.4 cm in the pressure head whereas the same variation of  $n$  induced only a variation of 0.11 cm in the pressure head in the case of the percolation-drainage experiment. It is interesting to note that the large confidence interval  $[1.34\text{--}3.12]$  for the parameter  $n$  from the percolation-drainage experiment involves the reduced confidence interval  $[2.4\text{--}2.9]$  obtained from the infiltration

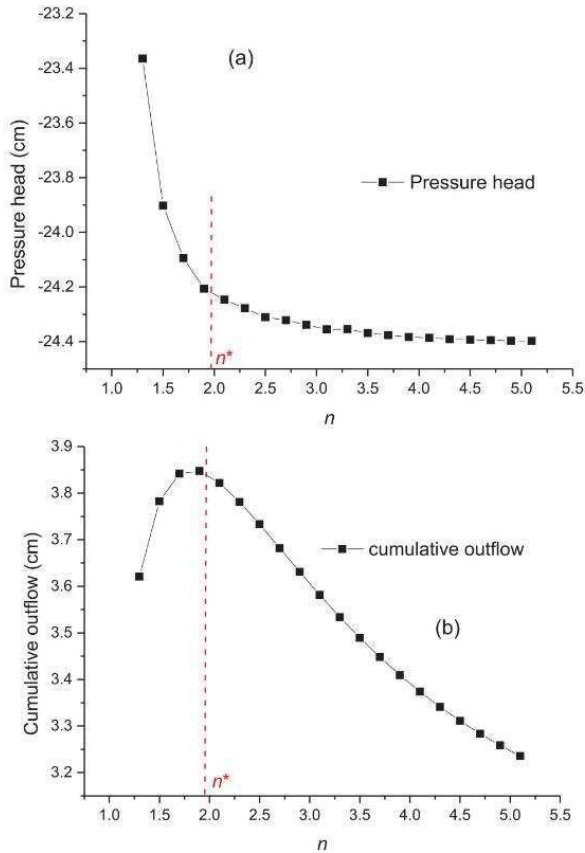


Fig. 9. Marginal effects of the Mualem-van Genuchten shape parameter  $n$  on the pressure head (a) and cumulative outflow (b) at  $t = 1200$  min for the percolation-drainage experiment ( $n^*$  is the mean estimated value).

experiment.

Finally, the inversion of the percolation-drainage experiment yielded a hydraulic conductivity  $K_S^* = 0.22$  cm/min with a large uncertainty region since the confidence interval is [0.13–1.24] cm/min. This large uncertainty region can be explained by the weak sensitivity of the cumulative outflow and the pressure head inside the column to the value of  $K_S$ . For instance, Fig. 11 shows the simulated pressure head at the location of the second tensiometer (at 18.5 cm from the soil surface) at  $t = 1200$  min for different  $K_S$  values taken from its prior interval [0.1, 2] cm/min. This figure shows a weak sensitivity of the pressure head around  $K_S^*$ . Indeed a variation of  $K_S$  from 0.2 to 0.3 cm/min induces only a variation of 0.2 cm of the pressure head. Furthermore, the simulation with the different values of  $K_S$  yielded similar cumulative outflow curves (not shown) which indicates that  $K_S$  is an irrelevant parameter for the cumulative outflow in this case. These results are in agreement with those of Durner (1994) and Schaap and van Genuchten (2005) who noted that the determination of  $K_S$  with multi-step outflow experiments can be highly uncertain in particular for structured soils.

On the other hand, the inversion of the infiltration experiment yielded a hydraulic conductivity of  $K_S^* = 0.21$  cm/min with a very narrow uncertainty region since the confidence interval is [0.2–0.22] cm/min. Therefore,  $K_S^*$  is 50 times more accurate than obtained from the percolation-drainage experiment. This phenomenon is due to the high sensitivity of the pressure head inside the column to the value of  $K_S$  in the case of the infiltration experiment. Note that the cumulative outflow depends only on the inflow rate  $Q_{inj}$  and was not sensitive to  $K_S$ . The sensitivity of the plateau region reached by the pressure head at

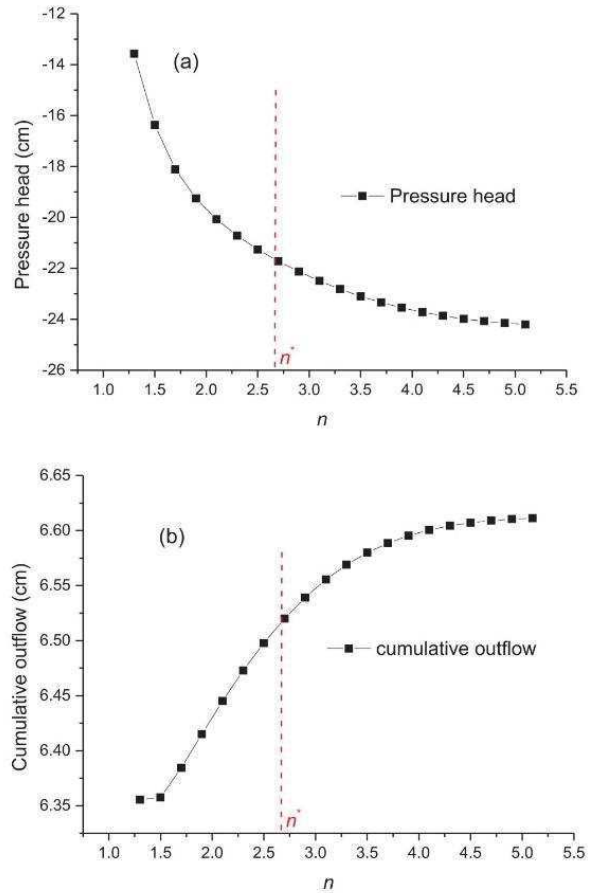


Fig. 10. Marginal effects of the Mualem-van Genuchten shape parameter  $n$  on the pressure head (a) and cumulative outflow (b) at  $t = 80$  min for the infiltration experiment ( $n^*$  is the mean estimated value).

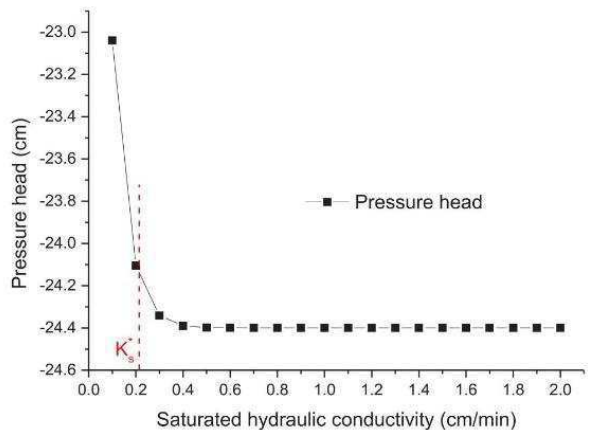


Fig. 11. Marginal effects of the saturated hydraulic conductivity  $K_S$  on the pressure head at 18.5 cm from the soil surface and  $t = 1200$  min for the percolation-drainage experiment ( $K_S^*$  is the mean estimated value).

18.5 cm from the soil surface during the injection is shown in Fig. 12. This figure is obtained by simulating the pressure head at  $t = 50$  min for different  $K_S$  values inside its prior interval [0.1–2] cm/min.

The results of Fig. 12 confirm that the pressure head is highly sensitive to  $K_S$  in the region surrounding  $K_S^*$ . Indeed, the variation of  $K_S$  from 0.2 to 0.3 cm/min induces a variation of about 5 cm of the

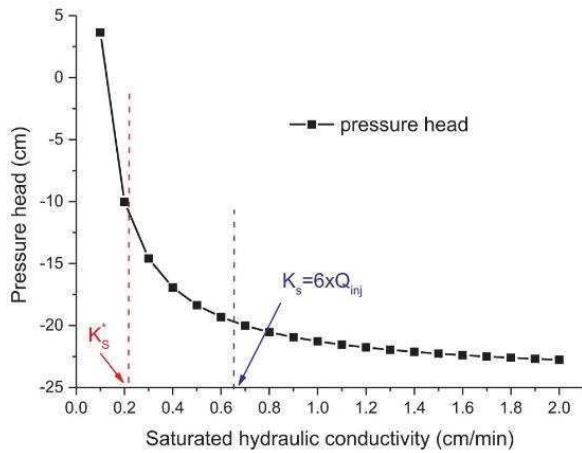


Fig. 12. Marginal effects of the saturated hydraulic conductivity  $K_S$  on the pressure head at 18.5 cm from the soil surface and  $t = 50$  min for the infiltration experiment ( $K_S^*$  is the mean estimated value).

pressure head.

Note that the slope of the curve in Fig. 12 is weak for high values of  $K_S$ . This means that if the sought value  $K_S^*$  is high, the sensitivity of the pressure head to  $K_S$  can be weak. To increase the sensitivity to  $K_S$  in these situations and therefore improve identifiability of  $K_S^*$ , we should use a higher inflow rate. Indeed, the applied inflow rate at the surface of the column should not be very small compared to  $K_S^*$ . Fig. 12 suggests that an inflow rate around 50% of  $K_S^*$  is preferential (very high sensitivity), but the zone of high sensitivity involves inflow rate ranging

until 6 times less than  $K_S^*$ .

Finally, we note that although the results of both experiments concerning the estimation of  $K_S^*$  are different, they remain in agreement since the narrow [0.2–0.22] cm/min confidence interval of  $K_S$  from the infiltration experiment is included in the large [0.13–1.24] cm/min confidence interval of  $K_S$  from the percolation-drainage experiment.

To summarize, the results of the Bayesian inversion show that although the two laboratory experiments yield a good agreement between the mean estimated parameter values, the quality of estimation in terms of uncertainty associated to the estimated parameters shows strong disagreement. Indeed, for the investigated undisturbed soil, the infiltration experiment yielded better quality of hydraulic soil parameter estimation (with narrower uncertainty regions) compared to the percolation-drainage experiment. This can be explained by the higher sensitivity of the model outputs to the hydraulic parameters in the case of the infiltration experiment. The results of the MCMC sampler for the infiltration experiment are shown in Fig. 13. In this figure, the “on-diagonal” plots display the parameter distributions, whereas the “off-diagonal” plots show the correlations between parameters in the MCMC sample. This figure shows that the parameter  $\theta_r$  cannot be identified. The parameters  $K_S$ ,  $\alpha$  and  $n$  are well estimated; their intervals of variation are strongly reduced and they essentially display bell-shaped distributions. Moderate correlations are observed between  $K_S$  and  $n$  ( $r = -0.73$ ),  $\theta_r$  and  $n$  ( $r = 0.76$ ), and  $\alpha$  and  $n$  ( $r = 0.74$ ).

## 6. Conclusions

In this work, a laboratory percolation-drainage and an infiltration experiments were conducted on an undisturbed soil column. Hydraulic soil parameters have then been investigated by inverse modeling using

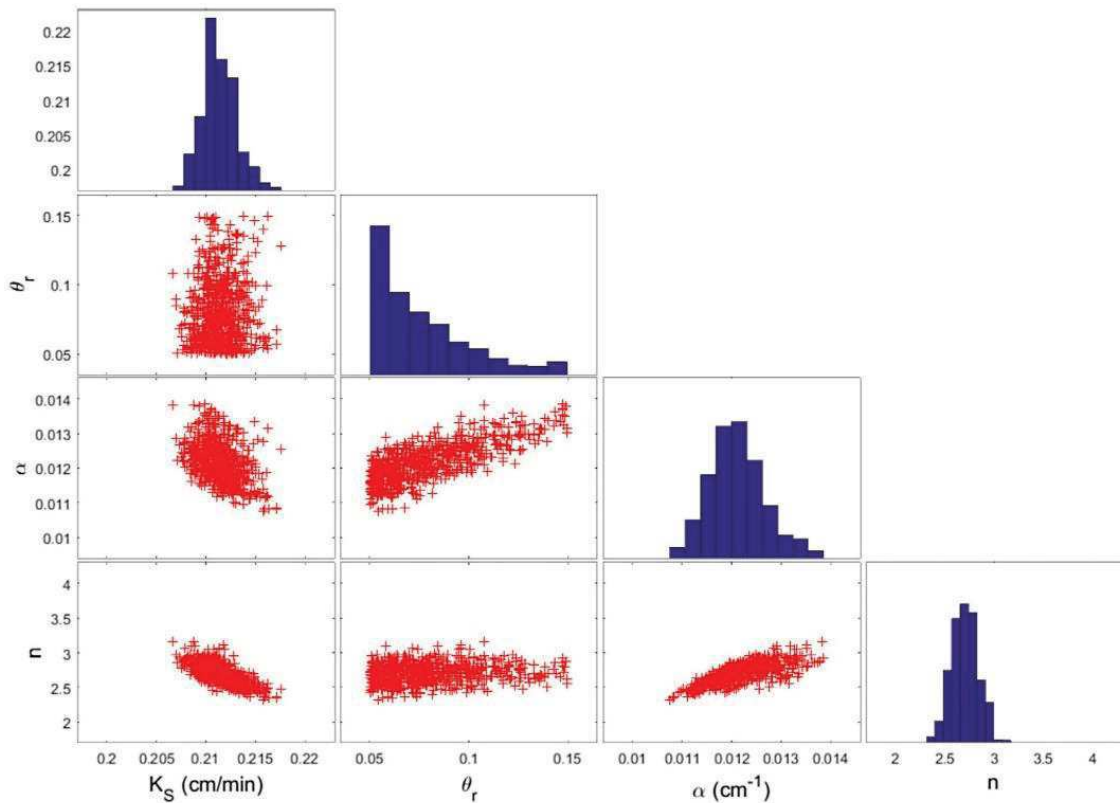


Fig. 13. MCMC solutions for the infiltration experiment. The diagonal plots represent the parameter posterior distributions. The off-diagonal scatterplots show the correlations between the parameters.

pressure head and cumulative outflow measurements. The saturated hydraulic conductivity, the residual water content and the Mualem-van Genuchten shape parameters  $\alpha$  and  $n$  have been estimated in a Bayesian framework using the Markov Chain Monte Carlo (MCMC) sampler.

The results reveal the following conclusions:

1. Saturated-unsaturated flow through the investigated undisturbed soil is well described by the RE combined with the Mualem/van-Genuchten models. The estimated hydraulic soil parameters allow to fit relatively well to the measurements of pressure head and cumulative outflow for both experiments.
2. Similar mean estimated soil parameter values were obtained from the inversion of the percolation-drainage and the infiltration experiments. However, significant differences were observed between uncertainties associated to the estimated parameters.
3. The residual water content was not identified with neither of the experiments since dry conditions were not attained during the two laboratory experiments.
4. Because of its high influence on the model outputs, the parameter  $\alpha$  is well estimated (with a narrow uncertainty region) from both experiments.
5. The parameters  $n$  and  $K_S$  are much better estimated (with much narrower uncertainty regions) using the infiltration experiment than using the percolation experiment. This phenomenon is due to the high sensitivity of the pressure head to these parameters in the case of the infiltration experiment whereas this sensitivity is weak in the case of the percolation-drainage experiment.
6. In the case of infiltration experiment, the injected inflow rate at the surface of the column should not be very small compared to the sought value of saturated conductivity in order to insure a high sensitivity of the pressure head to  $K_S$ .

For practical applications, this last point suggests that before fixing the inflow rate, we should have a rough idea about the value of  $K_S^*$ . This can be easily obtained by varying the inflow rate since a ponding can be observed if  $Q_{inj} > K_S^*$ .

#### Acknowledgments

The author's express their thanks to the Tunisian-French joint international laboratory NAILA (<http://www.lmi-naila.com/>) for financial support and to Radhouane Hamdi for his careful assistance during the experimental work.

#### References

Beydoun, H., Lehmann, F., 2006. Drainage experiment and parameter estimation in unsaturated porous media. *C.R. Géoscience* 338 (3) 180–178.

van Dam, J.C., Stricker, J.N.M., Droogers, P., 1994. Inverse method to determine soil hydraulic functions from multistep outflow experiments. *Soil Sci. Soc. Am. J.* 58, 647–652. <https://doi.org/10.2136/sssaj1994.03615995005800030002x>.

Durner, W., 1994. Hydraulic conductivity estimation for soils with heterogeneous pore structure. *Water Resour. Res.* 30, 211–223.

Durner, W., Iden, S.C., 2011. Extended multistep outflow method for the accurate determination of soil hydraulic properties near water saturation. *Water Resour. Res.* 47, W08526. <https://doi.org/10.1029/2011WR010632>.

Durner, W., Schultze, B., Zurmühl, T., 1999. State-of-the-art in inverse modeling of inflow/outflow experiments. p661–681. In: van Genuchten, M.Th. et al. (Ed.) *Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media*, Proc. Int. Worksh. Riverside, CA. Univ. of California, Riverside.

Eching, S.O., Hopmans, J.W., Wendroth, O., 1994. Unsaturated hydraulic conductivity from transient multistep outflow and soil water pressure data. *Soil Sci. Soc. Am. J.* 58, 687–695.

Fahs, M., Younes, A., Ackerer, P., 2011. An efficient implementation of the method of lines for multicomponent reactive transport equations. *Water Air Soil Pollut.* 215 (1–4), 273–283. <https://doi.org/10.1007/s11270-010-0477-y>.

Farthing, M.W., Ogdén, F.L., 2017. Numerical solution of richards' equation: a review of advances and challenges. *Soil Sci. Soc. Am. J.* 81, 1257–1269. <https://doi.org/10.2136/sssaj2017.02.0058>.

Garnier, P., Rieu, M., Boivin, P., Vauclin, M., Baveye, P., 1997. Determining the hydraulic properties of a swelling soil from a transient evaporation experiment. *Soil Sci. Soc. Am. J.* 61, 1555–1563.

Gelman, A., Carlin, J.B., Stren, H.S., Rubin, D.B., 1997. *Bayesian Data Analysis*.

Chapmann and Hall, London.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7, 457–472.

van Genuchten, M.Th., 1980. A closed form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.* 44 (5), 892–898. <https://doi.org/10.2136/sssaj1980.03615995004400050002x>.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.

Hayek, M., Younes, A., Zouali, J., Fajraoui, N., Fahs, M., 2018. Analytical solution and Bayesian inference for interference pumping tests in fractal dual-porosity media. *Comput. Geosci.*, vol. 22, 1, 413–421. <https://doi.org/10.1007/s10596-017-9701-9> (ISSN 1420-0597).

Hindmarsh, A.C., 1980. LSODE and LSODI, two new initial value ordinary differential equation solvers. *SIGNUM Newslett.* 15 (4), 10–11.

Hopmans, J.W., Simunek, J., Romano, N., Durner, W., 2002. Simultaneous determination of water transmission and retention properties. *Inverse Methods*, p. 963–1008. In: Dane, J.H., Topp, G.C. (Eds.) *Methods of Soil Analysis. Part 4. Physical Methods*. Soil Science Society of America Book Series No. 5.

Hudson, D.B., Wierenga, P.J., Hills, R.G., 1996. Unsaturated hydraulic properties from upward flow into soil cores. *Soil Sci. Soc. Am. J.* 60, 388 ± 96.

Kool, J.B., Parker, J.C., van Genuchten, M.Th., 1985. Determining soil hydraulic properties from one-step outflow experiments by parameter estimation: I. Theory and numerical studies. *Soil Sci. Soc. Am. J.* 49, 1348 ± 54.

Kumar, S., Sekhar, M., Reddy, D.V., Kumar, M.S., 2010. Estimation of soil hydraulic properties and their uncertainty: comparison between laboratory and field experiment. *Hydrol. Process.* 24, 3426–3435. <https://doi.org/10.1002/hyp.7775>.

Mara, T.A., Delay, F., Lehmann, F., Younes, A., 2016. A comparison of two Bayesian approaches for uncertainty quantification. *Environ. Model. Software* 82, 21–30. <https://doi.org/10.1016/j.envsoft.2016.04.010>.

Mishra, S., Parker, J.C., 1989. Parameter estimation for coupled unsaturated flow and transport. *Water Resour. Res.* 25 (3). <https://doi.org/10.1029/WR025i003p00385>.

Moreira, P.H.S., van Genuchten, M.T., Orlande, H.R.B., Cotta, R.M., 2016. Bayesian estimation of the hydraulic and solute transport properties of a small-scale unsaturated soil column. Retrieved 12 Feb. 2018, from. *J. Hydrol. Hydromech.* 64 (1), 30–44. <https://doi.org/10.1515/johh-2016-0002>.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., 1953. Equation of state calculation by fast computing machines. *J. Chem. Phys.* 21, 1087–1092.

Miller, C.T., Abhishek, C., Farthing, M., 2006. A spatially and temporally adaptive solution of Richards' equation. *Adv. Water Resour.* 29, 525–545. <https://doi.org/10.1016/j.advwatres.2005.06.008>.

Mualem, Y., 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resour. Res.* 12, 513–522. <https://doi.org/10.1029/WR012i003p00513>.

Laloy, E., Vrugt, J.A., 2012. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high-performance computing. *Water Resour. Res.* 48, W01526. <https://doi.org/10.1029/2011WR010608>.

Lee, P., 2004. *Bayesian Statistics*. Oxford University Press, London.

Linde, N., Ginsbourger, D., Irving, J., Nobile, F., Doucet, A., 2017. On uncertainty quantification in hydrogeology and hydrogeophysics. *Adv. Water Resour.* 110, 166–181. <https://doi.org/10.1016/j.advwatres.2017.10.014>.

Parker, J.C., Kool, J.B., van Genuchten, M.Th., 1985. Determining soil hydraulic properties from one-step outflow experiments by parameter estimation. II. Experimental studies. *Soil Sci. Soc. Am. J.* 49, 1354–1359.

Puhlmann, H., von Wilpert, K., Lukes, M., Dröge, W., 2009. Multistep outflow experiments to derive a soil hydraulic database for forest soils. *Eur. J. Soil Sci.* 60 (5), 792–806.

Radhakrishnan, K., Hindmarsh, A.C., 1993. Description and use of LSODE, the Livermore solver for ordinary differential equations. LLNL report UCRL-ID-113855.

Schaap, M.G., van Genuchten, M.T., 2005. A modified Mualem-van Genuchten formulation for improved description of the hydraulic conductivity near saturation. *Vadose Zone J.* 5, 27–34.

Schelle, H., Iden, S.C., Peters, A., Durner, W., 2010. Analysis of the agreement of soil hydraulic properties obtained from multistep-outflow and evaporation methods. *Vadose Zone J.* 9, 1080–1091. <https://doi.org/10.2136/vzj2010.0050>.

Schelle, H., Iden, S.C., Durner, W., 2011. Combined transient method for determining soil hydraulic properties in a wide pressure head range. *Soil Sci. Soc. Am. J.* 75 (5), 1681–1693.

Stolte, J., Freijer, J.I., Bouten, W., Dirksen, C., Halbertsma, J.M., van Dam, J.C., van den Berg, J.A., Veerman, G.J., Wösten, J.H.M., 1994. Comparison of six methods to determine unsaturated soil hydraulic conductivity. *Soil Sci. Soc. Am. J.* 58, 1596–1603.

Vrugt, J.A., ter Braak, C.J.F., Clark, M.P., Hyman, J.M., Robinson, B.A., 2008. Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resour. Res.* 44, W00B09. <https://doi.org/10.1029/2007WR006720>.

Wendroth, O., Ehlers, W., Hopmans, J.W., Kage, H., Halbertsma, J., Wösten, J.H.M., 1993. Reevaluation of the evaporation method for determining hydraulic functions in unsaturated soils. *Soil Sci. Soc. Am. J.* 57, 1436–1443. <https://doi.org/10.2136/sssaj1993.03615995005700060007x>.

Wind, G.P., 1968. Capillary conductivity data estimated by a simple method. In: *Water in the unsaturated zone: Proc. Symp., Wageningen, the Netherlands. June 1966. Vol. 1. Int. Assoc. Sci. Hydrol., Gentbrugge, Belgium*, p. 181–191.

Younes, A., Fahs, M., Ahmed, S., 2009. Solving density driven flow problems with efficient spatial discretizations and higher-order time integration methods. *Adv. Water Resour.* 32 (3), 340–352. <https://doi.org/10.1016/j.advwatres.2008.11.003>.

Younes, A., Mara, T., Fahs, M., Grunberger, O., Ackerer, P., 2017. Hydraulic and transport parameter assessment using column infiltration experiments. *Hydrol. Earth Syst. Sci.* 21, 2263–2275. <https://doi.org/10.5194/hess-21-2263-2017>.

Younes, A., Mara, T.A., Fajraoui, N., Lehmann, F., Belfort, B., Beydoun, H., 2013. Use of global sensitivity analysis to help assess unsaturated soil hydraulic parameters. *Vadose Zone J.* 12 (1), 1–12. <https://doi.org/10.2136/vzj2011.0150>.

# Chapitre 5. Estimation bayésienne des paramètres pour le transport non réactif en milieu poreux non saturé

## Plan du chapitre


Chapitre 5. Estimation bayésienne des paramètres pour le transport non réactif en milieu poreux non saturé .....	127
5.1 Introduction .....	130
5.2 Modèles expérimentaux et mathématiques .....	131
5.2.1 Expérience de laboratoire .....	131
5.2.2 Modèle numérique .....	132
5.2.3 Inférence bayésienne de paramètre .....	135
5.3 Résultats et discussion .....	137
5.4 Conclusion .....	140





Article

# Bayesian Simultaneous Estimation of Unsaturated Flow and Solute Transport Parameters from a Laboratory Infiltration Experiment

Anis Younes <sup>1,2,3</sup>, Jabran Zaouali <sup>1</sup>, Sabri Kanzari <sup>4</sup>, Francois Lehmann <sup>1</sup> and Marwan Fahs <sup>1,\*</sup> 

<sup>1</sup> LHyGES, University of Strasbourg/EOST/ENGEES, CNRS, 67084 Strasbourg, France

<sup>2</sup> LISAH, University Montpellier, INRA, IRD, Montpellier SupAgro, 34060 Montpellier, France

<sup>3</sup> LMHE, Ecole Nationale d'Ingénieurs de Tunis, Tunis 1002, Tunisia

<sup>4</sup> INRGRF, Laboratory of Rural Engineering, University of Carthage, Carthage 1054, Tunisia

\* Correspondence: fahs@unistra.fr; Tel.: +33-3-6885-0448

Received: 28 July 2019; Accepted: 7 August 2019; Published: 11 August 2019



**Abstract:** Numerical modeling has become an irreplaceable tool for the investigation of water flow and solute transport in the unsaturated zone. The use of this tool for real situations is often faced with lack of knowledge of hydraulic and soil transport parameters. In this study, advanced experimental and numerical techniques are developed for an accurate estimation of the soil parameters. A laboratory unsaturated flow and solute transport experiment is conducted on a large undisturbed soil column of around 40 cm length. Bromide, used as a nonreactive contaminant, is injected at the surface of the undisturbed soil, followed by a leaching phase. The pressure measurements at different locations along the soil column as well as the outflow bromide concentration are collected during the experiment and used for the statistical calibration of flow and solute transport. The Richards equation, combined with constitutive relations for water content and permeability, is used to describe unsaturated flow. Both linear and non-equilibrium mobile-immobile transport models are investigated for the solute transport. All hydraulic and mass transport parameters are inferred using a one-step Bayesian estimation with the Markov chain Monte Carlo sampler. The results prove that the pressure and concentration measurements are able to identify almost all hydraulic and mass transport parameters. The mobile-immobile transport model better reproduces the infiltration experiment. It produces narrower uncertainty intervals for soil parameters and predictive output concentrations.

**Keywords:** infiltration laboratory experiment; Richards equation; advection-dispersion equation; mobile-immobile model; predictive uncertainty; Markov chain Monte Carlo

## 1. Introduction

Numerical modeling of water flow and solute transport through the vadoze zone is essential to estimate the amount of recharge and/or to prevent soil and subsurface water contamination. It is becoming increasingly important for several applications as water quality management, water resources planning in urban zones, and for mitigating groundwater pollution. The processes of flow and mass transport in the unsaturated zone depend on the driving forces (matric and gravitational potential) and on the soil properties. Thus, to perform real field simulations, the numerical models require an accurate prior knowledge of the soil parameters. These parameters are usually determined using laboratory unsaturated flow and solute transport experiments performed using soil samples collected from the field. Inversion methods can then be applied to fit the model outputs to the laboratory measurements and to assess the soil parameters.

Different laboratory experiments have been conducted for the estimation of the soil parameters e.g., [1–13]. Small scale laboratory column experiments were often used for the characterization

of hydraulic and mass transport properties [1,2,11,14,15]. Mishra and Parker [16] showed that the simultaneous estimation of hydraulic and mass transport parameters yields more accurate results than the two-steps estimation in which hydraulic and transport parameters are determined sequentially. Inoue et al. [17] used electrical conductivity and matric pressure head measurements at different depths for the flow-transport inversion with the local search Levenberg–Marquardt algorithm. Recently, several Bayesian approaches, where measurements are combined with prior parameter information to provide posterior parameter distributions, have been investigated for the estimation of the unsaturated hydraulic soil parameters, among others [10,12,18–21]. The term Bayesian is used to describe statistical inversion by considering [22]: (i) That model variables are random, (ii) that randomness describes the degree of information for their realization, and (iii) that the solution of the estimation problem is the posterior probability distribution from which several statistics can be obtained. Bayesian hydraulic soil parameter estimation was investigated for a laboratory drainage experiment in [10]. In [12], it is shown that the infiltration laboratory experiment yields more accurate hydraulic parameters than the percolation-drainage experiment. Bayesian inference of both hydraulic and transport unsaturated soil parameters has been investigated for hypothetical experiments in the case of linear transport [21] and in the case of pesticide transport [23] in unsaturated soils. Moreira et al. [11] considered the estimation of hydraulic and solute transport parameters based on a Bayesian inversion of a laboratory infiltration experiment performed on a small column of 7.3 cm length. One tensiometric sensor was used to monitor the pressure head in the middle of the column. The pressure head and outflow concentrations measurements were employed for parameters estimation. Because of the insufficient collected data, some hydraulic soil parameters were not included in the inversion procedure to prevent identification difficulties. These parameters were considered as known (measured) and they were assigned informative Gaussian priors. Furthermore, Moreira et al. [11] used a sequential procedure in which the hydraulic parameters were firstly estimated using the pressure head measurements. Then, in a second step, these estimates were employed as prior information with Gaussian distributions to determine the mass transport parameters from the concentration measurements.

In the present study, a laboratory infiltration experiment is conducted in order to estimate soil hydraulic and mass transport parameters. Bromide is used as a nonreactive contaminant and injected at the surface of the soil column. The current research differs from the literature by (i) considering a large column of around 40 cm length of an undisturbed sandy soil, (ii) performing a simultaneous estimation of all hydraulic and mass transport soil parameters using the outflow breakthrough curve of bromide and the pressure head measurements at three different locations, (iii) assuming no prior information for all hydraulic and solute transport parameters for which we attribute uniform prior distributions with large intervals to reflect our poor prior knowledge of their values, and (iv) estimating all parameters using a one-step Bayesian procedure via the Markov chain Monte Carlo (MCMC) method [24].

The direct problem is based on the Richards equation coupled with the Darcy's law and the mass conservation equation [25]. The Mualem and van Genuchten models are used as constitutive relations [26,27]. The linear advection-dispersion and the non-equilibrium mobile-immobile models are investigated for bromide transport. Bayesian inversion results are discussed in terms of fitting between simulations and measurements and in terms of uncertainty ranges associated with the estimated parameter values and to the model outputs.

## 2. Materials and Methods

### 2.1. Laboratory Experiment

A cylindrical PVC tube of 60 cm length, beveled at its end to form a cutting edge, was sunk into the soil in the region of Bekalta in the east center of Tunisia. A homogeneous soil column of 38.4 cm length and 15 cm diameter has been carefully extracted. The soil, essentially formed by sand, was placed over a porous plate of 6 mm high with a saturated water content of  $0.42 \text{ cm}^3 \text{ cm}^{-3}$ . The soil column (sand + porous plate) is set on a container filled with gravel (Figure 1). An orifice,

located face to the soil bottom, allows effluent recovery and fixing the pressure head  $h$  [L] at the lower boundary to zero. The vertical axis  $Z$  is downward oriented with the origin located at the soil surface. Three tensiometric sensors, connected to a data logger and a computer, allow monitoring the pressure head near the top ( $Z_1 = 2.5$  cm), middle ( $Z_2 = 17.5$  cm), and bottom of the column ( $Z_3 = 32.5$  cm). The outflow bromide concentration is measured using an ion selective electrode (HI4102 from HANNA instruments). Initially, hydrostatic equilibrium is obtained by verifying that each tensiometer indicates a suction equal to its distance to the bottom of the column. A constant water flow rate of  $14 \text{ cm}^3/\text{min}$  was applied at the top surface of the soil using a high-accuracy Masterflex® L/S peristaltic pump. Cotton fiber wicks were deposited at the surface of the soil to obtain a uniform distribution of the flow rate over the soil area. The tensiometric data and the cumulative outflow were monitored until  $t = 660$  min. Infiltration occurs during a period of 562 min, it starts at  $t = 24$  min and was stopped at  $t = 586$  min. During a first period ( $24 \text{ min} < t < 220$  min), the pump injects a potassium bromide (KBr) solution from a reservoir with a concentration of  $0.08 \text{ g/L}$ . The period of injection ensures around one pore volume of injected KBr solution. Then, clean water was injected at the surface of the soil with the same flow rate for the period  $220 \text{ min} < t < 586$  min. This ensures around two pore volumes of injected clean water. To limit evaporation, the upper end of the column was sealed with the exception of a small orifice, which ensures equilibrium with atmospheric pressure.

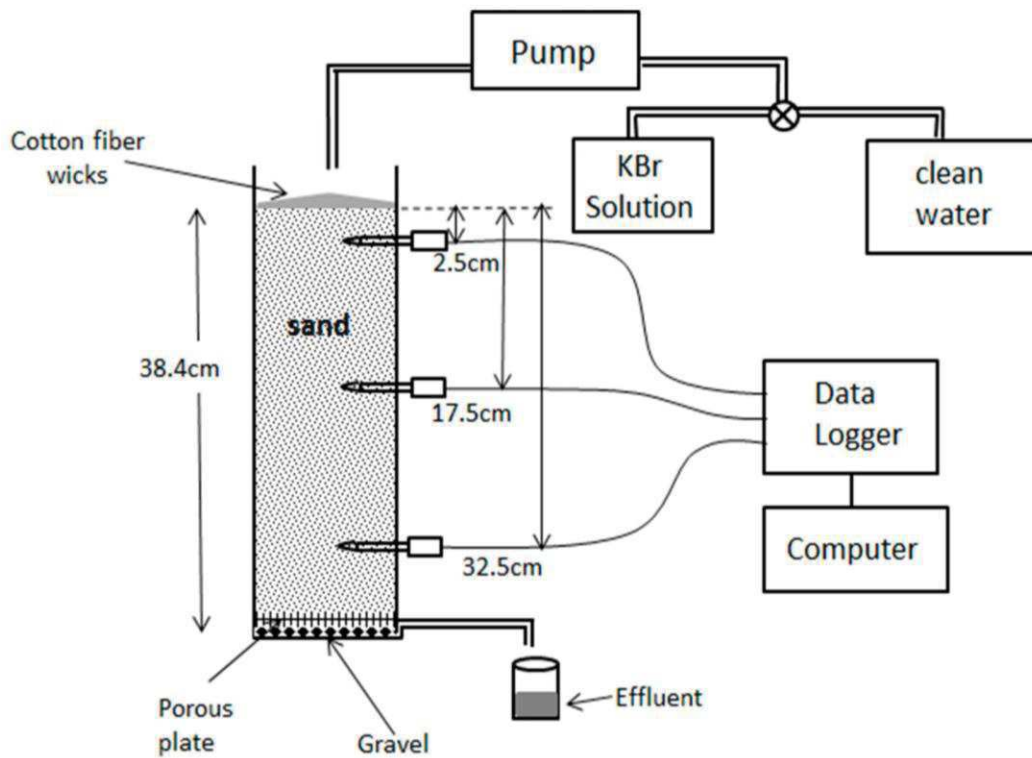


Figure 1. Schematic of the laboratory infiltration experimental setup.

## 2.2. Numerical Model

The 1D unsaturated flow through the column can be described using Richards equation (RE):

$$\begin{cases} \frac{\partial \theta}{\partial t} = (C_s(h) + S_s \frac{\theta}{\theta_s}) \frac{\partial h}{\partial t} = -\frac{\partial q_d}{\partial z} \\ q_d = -K(h) \left( \frac{\partial h}{\partial z} - 1 \right) \end{cases} \quad (1)$$

where  $\theta$  is the water content,  $t$  is time,  $C_s(h)$  represents the specific moisture capacity,  $S_s$  is the specific storage,  $\theta_s$  is the saturated water content,  $q_d$  is the water velocity, and  $K(h)$  is the hydraulic conductivity.

Equation (1) is closed using the constitutive relations of Mualem [26] and van Genuchten [27] expressing the hydraulic conductivity in terms of saturation and the saturation in terms of pressure head, respectively.

$$S_e(h) = \frac{\theta(h) - \theta_r}{\theta_s - \theta_r} = \begin{cases} \frac{1}{(1 + |\alpha h|^n)^m} & h < 0 \\ 1 & h \geq 0 \end{cases} \quad (2)$$

$$K(S_e) = K_s S_e^{1/2} \left[ 1 - (1 - S_e^{1/m})^m \right]^2$$

where  $\theta_r$  is the residual water content,  $S_e$  is the effective saturation,  $K_s$  is the hydraulic conductivity at saturation,  $m = 1 - 1/n$ , and  $\alpha$  and  $n$  are the shape parameters specific to the soil.

The flow system of Equations (1) and (2) is subject to the following boundary conditions. At the upper boundary, we impose a constant flux ( $q_{inj} = 0.08$  cm/min) during the injection period ( $24 \text{ min} < t < 586 \text{ min}$ ); otherwise, the upper flux is zero. A zero-pressure head (Dirichlet boundary condition) is imposed at the column lower boundary. Initial conditions correspond to a hydrostatic pressure distribution ( $h = z - L$ ).

Two transport models are used to investigate bromide transport through the studied soil column. The first model is the well-known linear transport model, based on the advection-dispersion equation:

$$\frac{\partial(\theta C)}{\partial t} + \frac{\partial(q_d C)}{\partial z} - \frac{\partial}{\partial z} \left( \theta D \frac{\partial C}{\partial z} \right) = 0 \quad (3)$$

where  $C$  is the bromide concentration,  $D = a_l q_d + d_m$  is the dispersion coefficient,  $d_m$  is the molecular diffusion coefficient fixed to a small value of  $3.10^{-4}$  cm<sup>2</sup>/min as common in literature, and  $a_l$  is the soil dispersion coefficient. The transport and flow equations are coupled together through the water content  $\theta$  and the Darcy velocity  $q_d$ .

The second transport model is based on the non-equilibrium approach. It is usually called mobile-immobile model. This model proceeds by considering that the liquid phase in the porous medium can exist as mobile and immobile phases. No advective solute transport is assumed in the immobile zone and a first-order kinetic diffusion transport process occurs between the mobile and immobile zones. The mobile-immobile model in variably saturated porous media writes [28,29]:

$$\frac{\partial(\theta_m C_m)}{\partial t} + \theta_{im} \frac{\partial C_{im}}{\partial t} + \frac{\partial(q C_m)}{\partial z} - \frac{\partial}{\partial z} \left( \theta_m D \frac{\partial C_m}{\partial z} \right) = 0 \quad (4)$$

$$\theta_{im} \frac{\partial C_{im}}{\partial t} = \omega (C_m - C_{im}) \quad (5)$$

$$\theta = \theta_m + \theta_{im} \quad (6)$$

where subscripts ( $m$ ) and ( $im$ ) represent the mobile and immobile phases, respectively,  $\omega$  is the first-order exchange coefficient between phases, and  $\theta$  is the total water content. We assume that the mobile and immobile water contents can vary in space and time but the ratio of mobile water content to the total water content ( $f = \theta_m / \theta$ ) is assumed to be constant. As a consequence, the fraction of immobile water  $\theta_{im}$  is also constant during the transient flow (see Equation (6)). This assumption is less restrictive and more representative of reality than assuming a constant immobile water content  $\theta_{im}$  during the unsaturated flow [11].

For both transport models, initial conditions correspond to a zero-solute concentration inside the whole column. Boundary upper condition ( $z = 0$ ) corresponds to a Dirichlet condition with the concentration fixed to that of the bromide solution during pollutant injection and to zero during clean water injection. The lower boundary condition ( $z = L$ ) corresponds to a zero concentration gradient.

Numerical solution of flow and transport equations in unsaturated porous media is a challenging problem [30–32]. The systems of equations in both linear transport and mobile-immobile transport cases are solved with the standard finite volume method. A uniform mesh of 192 cells of 0.2 cm is employed for the discretization of the soil column and porous plate. This spatial discretization

yields mesh independent solutions. Temporal discretization is performed with a variable order ODE (ordinary differential equation) solver via the method of lines (MOL), as in Fahs et al. [33,34]. MOL has been successfully used for solving flow and/or solute transport in unsaturated porous media in many studies e.g., [10,12,21,35,36]. The ODE solver used in this work (DLSODIS: Double-precision Livermore Solver for ordinary differential equations-implicit form and sparse matrix) allows variable time-step and variable order integration with time error control [37,38].

### 2.3. Bayesian Parameter Inference

The vector of observations  $\mathbf{y}_{mes}$  is formed by the measurements of the pressure head at three different locations inside the column and the breakthrough curve of bromide at the outflow. The specific storage is fixed to a very small value ( $S_s = 10^{-8} \text{cm}^{-1}$ ). All hydraulic and solute transport soil parameters are assumed unknowns and sought by inverse modeling. In the case of linear transport model, the estimated parameters are:  $K_S, \theta_S, \theta_r, a_l, \alpha$ , and  $n$  for the soil and the saturated conductivity for the porous plate ( $K_S^p$ ). Hence, the vector of unknown parameters is  $\xi = (K_S, K_S^p, \theta_S, \theta_r, a_l, \alpha, n)$ . In the case of mobile-immobile model, the vector of unknown parameters contains, in addition, the fraction of mobile water content ( $f$ ) and the exchange coefficient ( $\omega$ ). Hence, in this case, it writes  $\xi = (K_S, K_S^p, \theta_S, \theta_r, a_l, \alpha, n, f, \omega)$ .

The soil parameters for both linear and mobile-immobile situations are estimated using a Bayesian approach where the measurements are combined with prior parameter information to determine the posterior probability distribution function (pdf). In the following, the pdfs are inferred using DREAM<sub>(ZS)</sub> software [39] based on the Markov chain Monte Carlo (MCMC) method. The MCMC method has been used by several authors in hydrogeology, e.g., [11,40–44]. With MCMC, random sequences of parameter sets are generated and converge asymptotically toward the target distribution [45]. The latter can then be used to evaluate mean estimated parameter values and confidence intervals to characterize parameter uncertainty. Due to the absence of prior information on the parameter values, independent uniform prior distributions are assumed for all soil parameters. With these non-informative distributions, all possible parameter values are equally likely. Furthermore, the parameter intervals are chosen sufficiently large (Table 1) to reflect our usual poor prior knowledge of the model parameter values.

Error measurements of the tensiometric pressure sensors and of the ion selective electrode are assumed to be normally and independently distributed with unknown standard deviations. The latter are considered as hyperparameters and are simultaneously estimated with the physical parameters. We denote by  $\sigma = (\sigma_h, \sigma_c)$  the vector of hyperparameters.

Setting the calibration problem in a Bayesian framework yields the following posterior pdf,

$$p(\xi, \sigma | \mathbf{y}_{mes}) \propto \sigma_h^{-N_h} \sigma_c^{-N_c} \exp\left(-\frac{SS_h(\xi)}{2\sigma_h^2} - \frac{SS_c(\xi)}{2\sigma_c^2}\right) \quad (7)$$

where  $SS_h(\xi)$  and  $SS_c(\xi)$  are defined as the sums of the square of differences between the observed and predicted data of the pressure head and concentration, respectively.

For instance,  $SS_h(\xi) = \sum_{k=1}^{N_h} (h_{mes}^{(k)} - h_{mod}^{(k)}(\xi))^2$  includes the observed ( $h_{mes}^{(k)}$ ) and predicted ( $h_{mod}^{(k)}$ ) pressure heads at time  $t_k$  for the number of pressure head observations ( $N_h$ ).

With the MCMC sampler, a new candidate  $\mathbf{x}^i = (\xi^i, \sigma^i)$  is generated from a proposal distribution  $q(\mathbf{x}^i | \mathbf{x}^{i-1})$ , which only depends on the previous accepted candidate. The new candidate can be accepted or rejected depending on the Hasting ratio  $\alpha_H$  [46],

$$\alpha_H = \min\left(1, \frac{p(\xi^i, \sigma^i | \mathbf{y}_{mes})q(\mathbf{x}^{i-1} | \mathbf{x}^i)}{p(\xi^{i-1}, \sigma^{i-1} | \mathbf{y}_{mes})q(\mathbf{x}^i | \mathbf{x}^{i-1})}\right) \quad (8)$$

The last parameter sets that adequately fit the model onto observations are used to appraise the quality of parameter estimate (such as the mean and the standard deviation of the estimated parameters). The MCMC sampler is used hereafter with three parallel chains. Results are considered stationary if the criterion, defined by Gelman and Rubin [47], is verified ( $R_{stat} \leq 1.2$ ) and if the chains are not autocorrelated.

**Table 1.** Ranges of the soil parameters.

Parameter	Lower Bound	Upper Bound
$K_S$ (cm/min)	0.01	0.3
$K_S^p$ (cm/min)	$5 \cdot 10^{-4}$	$10^{-2}$
$\theta_S$ (-)	0.2	0.45
$\theta_r$ (-)	0.	0.15
$\alpha$ (cm <sup>-1</sup> )	0.001	0.1
$n$ (-)	1.5	7.0
$a_l$ (cm)	0.1	3.0
$f$ (-)	0.5	1.
$\omega$ (min <sup>-1</sup> )	$10^{-5}$	$10^{-2}$

### 3. Results and Discussion

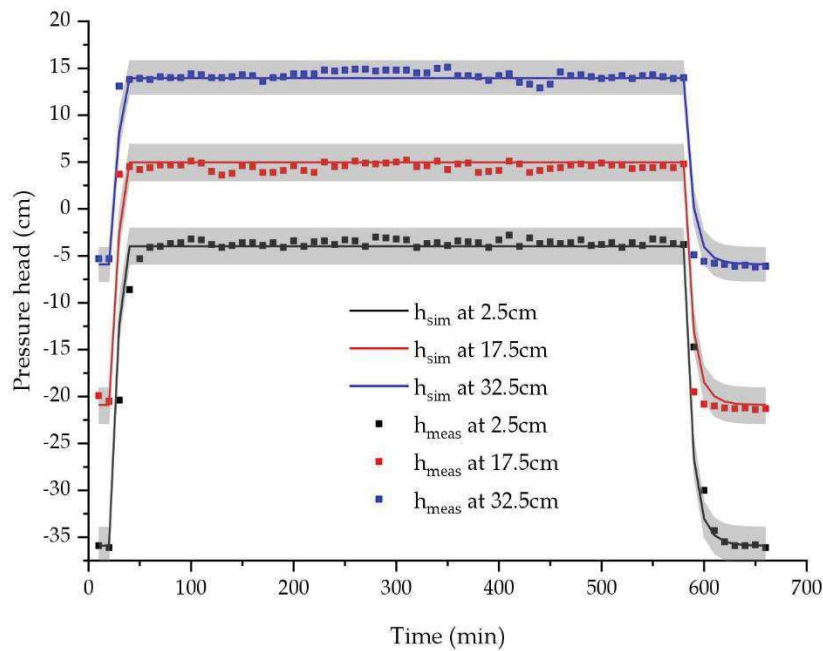
The surface injection induces an increase of the pressure head at the tensiometer locations. When the surface injection is stopped, the measured pressure heads decrease as the water front moves below the tensiometers (Figure 2). For each tensiometer, the pressure head is almost constant during the injection period. Note that since the tensiometers are separated by the same distance (15 cm between each other), the jump in the pressure levels during the injection period between the first and second tensiometers is similar to that between the second and third ones (Figure 2). The results of Figure 2 show that only the tensiometer located at  $Z_1 = 2.5$  cm from the soil surface continuously measures negative pressure heads. Hence, it is the only tensiometer that remains under unsaturated conditions after the arrival of the water front. The other tensiometers (located at  $Z_2 = 17.5$  cm and  $Z_3 = 32.5$  cm from the soil surface) measure positive pressure heads indicating that the latter are under the water table. Note that after the end of the infiltration period (i.e., for  $t > 586$  min), desaturation occurs and the pressure head at each tensiometer decreases until reaching its initial hydrostatic negative value corresponding to  $-35.9$  cm,  $-20.9$  cm, and  $-5.8$  cm at, respectively,  $Z_1 = 2.5$  cm,  $Z_2 = 17.5$  cm, and  $Z_3 = 32.5$  cm from the soil surface. The pressure measurements during this desaturation period (Figure 2) are prone to contain useful information for the characterization of the unsaturated soil parameters. The measured concentration at the column outflow is given in Figure 3. This figure depicts the logarithm of the bromide concentration in the effluent since the potential measured by the ion-selective electrode is directly proportional to the logarithm of the ion concentration.

Parameter identifiability is assessed for the linear transport model using MCMC with 30,000 model runs. Figure 4 shows that the chains have converged after 14,000 model runs. The values of the mean estimated parameters and the corresponding confidence intervals, for the linear transport model, are depicted in Table 2.

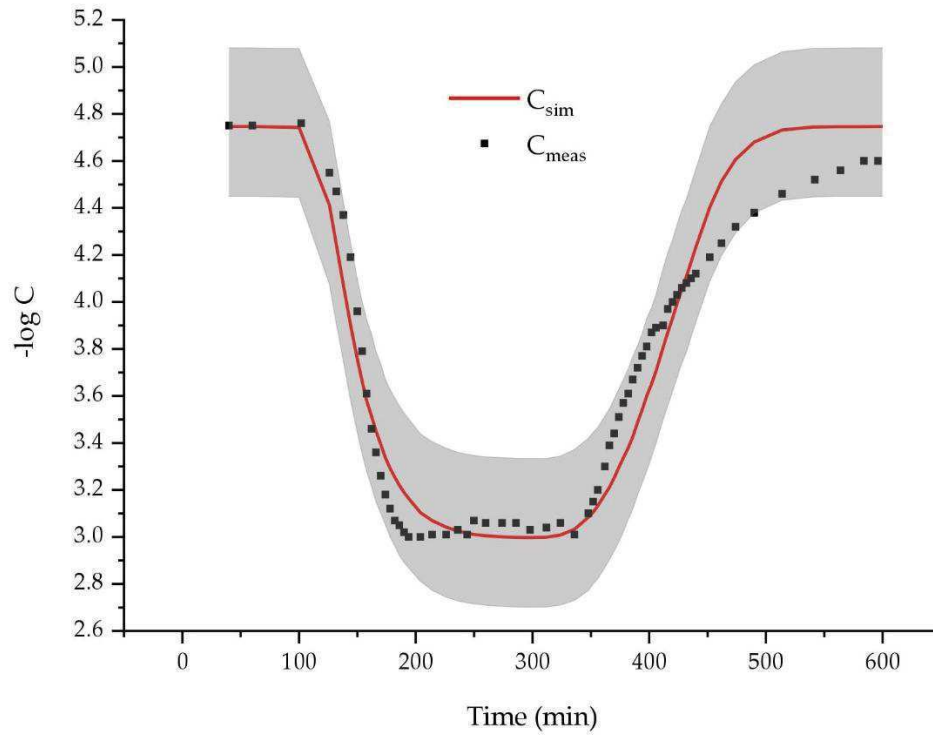
**Table 2.** Estimated mean values, confidence intervals (CIs), and size of the posterior CIs for the linear transport model.

	Unit	Mean Estimated Value	Confidence Interval	Size of the CI
$K_S$	(cm/min)	0.2	[0.19–0.21]	0.013
$K_S^p$	(cm/min)	0.0022	[0.0021–0.0023]	$10^{-4}$
$\theta_S$	(-)	0.33	[0.32–0.34]	0.016
$\theta_r$	(-)	0.1	[0.02–0.15]	0.13
$\alpha$	( $\text{cm}^{-1}$ )	0.015	[0.012–0.018]	0.006
$n$	(-)	2.57	[1.96–3.4]	1.43
$a_l$	(cm)	2.36	[1.98–2.91]	0.93

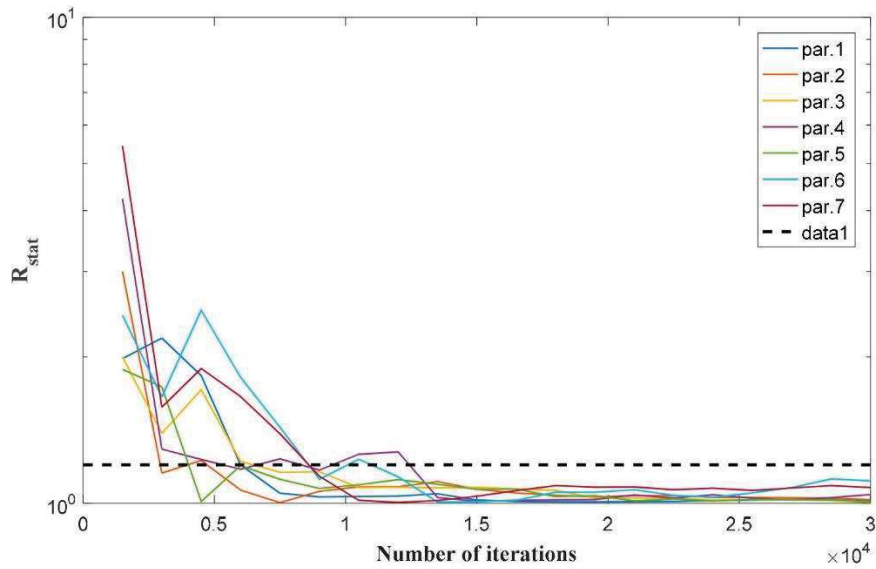
The results of this table show that, except the residual water content, all the parameters are well estimated since the posterior intervals of the estimated parameters (Table 2) are strongly narrowed, as compared to their prior intervals (given in Table 1). The significant difference between the prior and posterior distributions of a parameter testifies the high model sensitivity of this parameter. Furthermore, the marginal distributions of all soil parameters (except  $\theta_r$ ) exhibit almost bell-shaped posterior distributions (Figure 5).



**Figure 2.** Observed (symbols), calibrated (lines), and predictive uncertainty (grey band) for the pressure head at  $Z_1 = 2.5$  cm,  $Z_2 = 17.5$  cm, and  $Z_3 = 32.5$  cm from the soil surface.

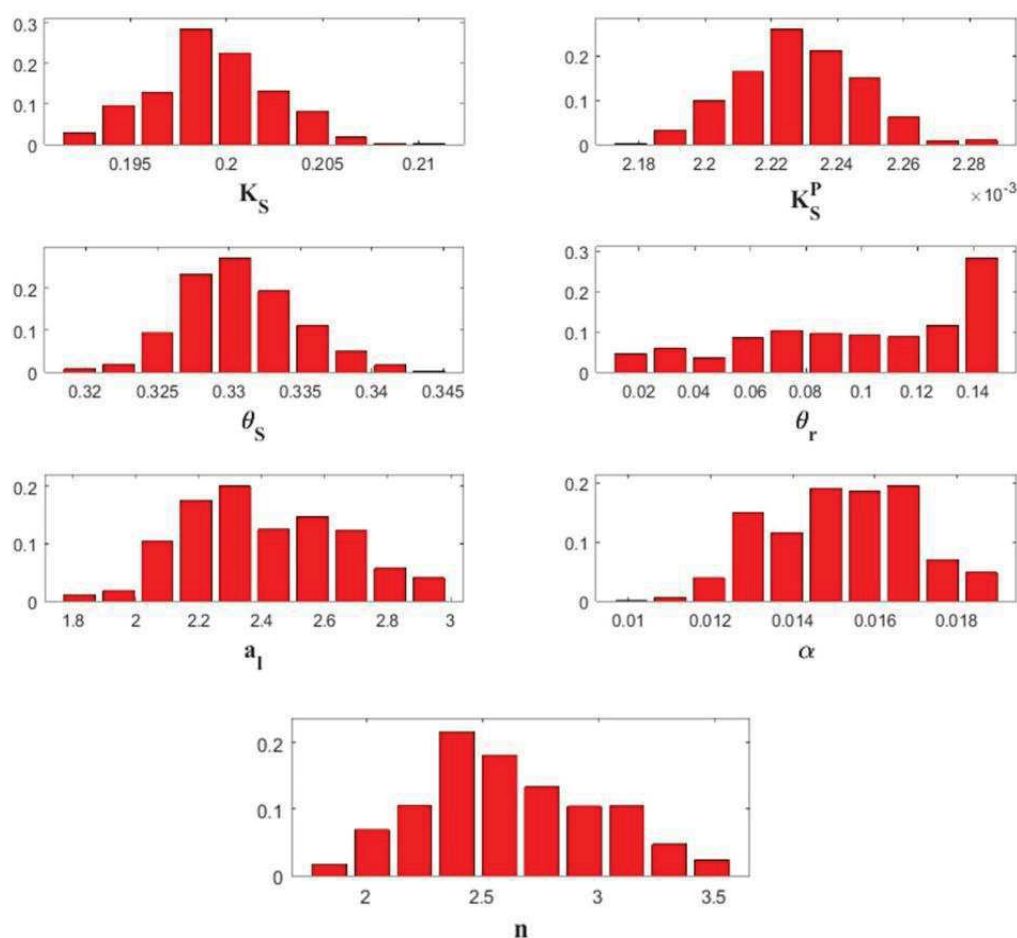


**Figure 3.** Observed (symbols), calibrated (lines), and predictive uncertainty (grey band) for the effluent bromide concentration using the linear transport model.



**Figure 4.** Convergence of the unsaturated soil parameters for linear transport: Variation of the Gelman–Rubin ( $R_{stat}$ ) with the number of iterations. The convergence is achieved if the chains reach the threshold in broken-line ( $R_{stat} \leq 1.2$ ).





**Figure 5.** Histograms of marginal distribution of parameters with the linear transport model.

The saturated conductivities of the soil  $K_S$  and that of the porous plate  $K_S^p$  are accurately estimated and show very narrow uncertainty intervals. The sand soil is around 100 times more permeable than the porous plate. Note that  $K_S$  and  $K_S^p$  are moderately correlated ( $r = -0.85$ ). The saturated water content of the soil is also well estimated with an associated uncertainty less than 5%. The residual water content has no influence on the pressure head nor on the output concentration. Indeed, the posterior interval of  $\theta_r$  is similar to its prior interval, which indicate that this parameter is not influent and, hence, cannot be well estimated. This non identifiability is due to the fact that dry conditions were not attained during the experiment. The results of the Figure 5 show that the parameters  $\alpha$  and  $n$  are quite well estimated, although they are strongly correlated ( $r = 0.99$ ). Finally, the dispersivity transport parameter  $a_l$  is quite well estimated with a high mean estimated value of 2.33 cm and a large confidence interval (Figure 5).

Figures 2 and 3 depict the pressure head and the outflow concentration evaluated at the measurement points using results of the MCMC inversion. A good agreement is observed between the predicted and simulated pressure heads using the mean estimated parameter values. However, a less satisfactory matching is observed between the predicted and measured concentrations. The disagreement between concentration profiles is especially notable at the leaching phase. Indeed, the measured concentrations show an asymmetric and long-tailed breakthrough curve, which cannot be reproduced by the symmetrical simulated concentration profile. The 95% uncertainty bands corresponding to the 97.5 and 2.5 percentiles of the distributions are shown in Figures 2 and 3. In these figures, the green regions represent the total predictive uncertainty, which takes into account both parametric uncertainty and measurement errors. All pressure and concentration observations fit

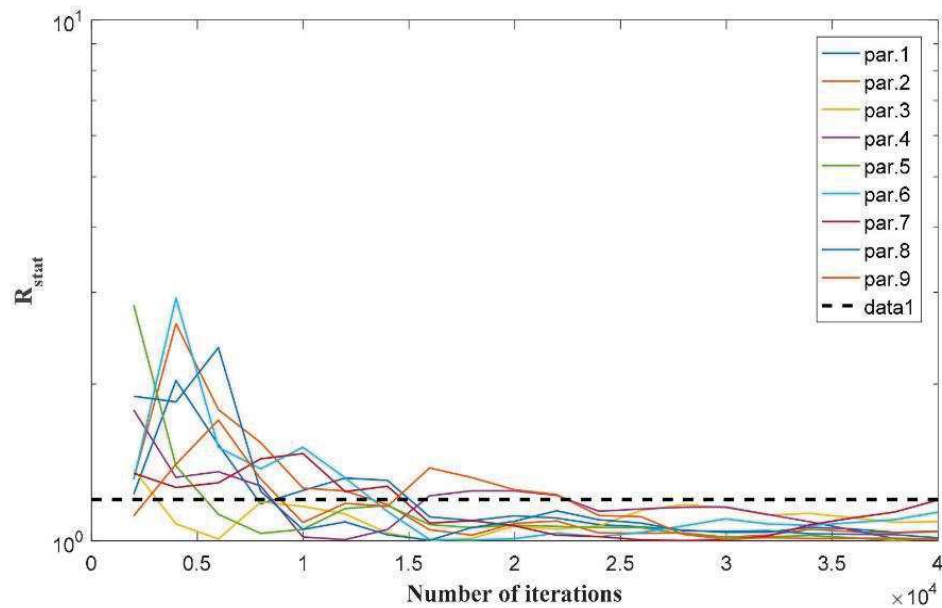
within the 95% total uncertainty range. Small uncertainties were obtained for pressure head responses, while very large uncertainties were assigned to the concentration outputs. This phenomenon is related to the bad matching between measured and simulated concentrations. This bad matching is responsible of the large dispersivity coefficient found by the MCMC sampler in order to reproduce the long-tailed breakthrough curve of the observed concentrations.

The Bayesian inversion of the mobile-immobile model, which includes nine unknown parameters, was performed with the MCMC sampler using 40,000 model runs. The convergence of the MCMC chains requires more simulations to reach the convergence than with the linear transport model. As reported in Figure 6, the overall chains have converged after 24,000 model runs. The mean values of the estimated parameters and the associated confidence intervals are summarized in Table 3. As previously, the posterior confidence intervals of all parameters, except  $\theta_r$ , are much narrower than their prior intervals. The histogram of marginal distribution of  $\theta_r$  shows almost a uniform distribution similar to its prior one. The histograms of marginal distributions of the rest of parameters show almost bell-shaped distributions (Figure 7) significantly different from their prior uniform distributions which indicate the high sensitivity of the parameters.

The results of Table 3 and Figure 7 show that the new two parameters  $f$  and  $\omega$  are well estimated. Note that the fraction of the mobile water content to the total water content  $f$  is very accurately estimated with an associated uncertainty less than 2%. A moderate correlation ( $r = -0.73$ ) is observed between the parameters  $f$  and  $\omega$ . The comparison between the results of the linear transport model (Table 2) and the mobile-immobile one (Table 3) show that the two models yield almost similar mean parameter estimated values and similar confidence intervals for the saturated hydraulic conductivities ( $K_S$  and  $K_S^p$ ) and the van-Genuchten parameters ( $\alpha$  and  $n$ ). As previously, the parameters  $K_S$  and  $K_S^p$  are moderately correlated ( $r = -0.87$ ), whereas the parameters  $\alpha$  and  $n$  are very strongly correlated ( $r = 0.99$ ). The mean estimated values of the saturated water content of the two models are almost identical. However, the associate uncertainty intervals are significantly different. Indeed, the inversion of the mobile-immobile model yields an uncertainty interval three times narrower than with the linear transport model. Hence,  $\theta_S$  is much more accurately estimated with the mobile-immobile transport model, although both models yield the same mean estimated parameter value. As with the linear transport model, the residual water content is irrelevant and, hence, cannot be identified. Among the parameters estimated by the inversion of the two transport models, the dispersivity coefficient shows the most significant differences both in terms of mean estimated value and in term of associate uncertainty range. Indeed, the inversion of the mobile-immobile transport model yields a mean estimated  $a_l$  value that is 3.3 times less than the one obtained with the linear transport model. The confidence interval is also more than 3 times narrower with the mobile-immobile transport model.

**Table 3.** Estimated mean values, confidence intervals (CIs), and size of the posterior CIs for the mobile-immobile model.

	Unit	Mean Estimated Value	Confidence Interval	Size of the CI
$K_S$	[cm/min]	0.2	[0.19–0.21]	0.012
$K_S^p$	[cm/min]	0.0022	[0.0021–0.0023]	$1 \cdot 10^{-4}$
$\theta_S$	[-]	0.32	[0.32–0.33]	0.005
$\theta_r$	[-]	0.11	[0.02–0.15]	0.13
$\alpha$	[cm <sup>-1</sup> ]	0.015	[0.012–0.018]	0.007
$n$	[-]	2.43	[1.95–3.24]	1.28
$a_l$	[cm]	0.7	[0.58–0.86]	0.27
$f$	[-]	0.915	[0.907–0.922]	0.015
$\omega$	[min <sup>-1</sup> ]	$5 \cdot 10^{-4}$	[ $4 \cdot 10^{-4}$ – $5.9 \cdot 10^{-4}$ ]	$1.9 \cdot 10^{-4}$



**Figure 6.** Convergence of the unsaturated soil parameters for the mobile–immobile transport model: Variation of the Gelman–Rubin ( $R_{stat}$ ) with the number of iterations. The convergence is achieved if the chains reach the threshold in broken-line ( $R_{stat} \leq 1.2$ ).

Because the two transport models yield almost the same mean estimated values and uncertainty ranges for all hydraulic parameters ( $K_S, K_S^p, \theta_S, \theta_r, \alpha, n$ ), the matching between measured and calibrated pressure heads by inverting the mobile–immobile transport model (not shown) is similar to the matching depicted in Figure 2 obtained by inverting the linear transport model. Likewise, the total predictive uncertainty associated with the pressure head output with the mobile–immobile model (not shown) is similar to the one observed with the linear transport model (grey bands in Figure 2).

Measured and simulated output concentrations using the mean estimated parameter values obtained from the inversion of the mobile–immobile transport model are plotted in Figure 8. Contrary to the results of the linear transport model (Figure 3), a very good matching is observed between measured and simulated concentrations during the whole experiment (including the leaching phase). Furthermore, the total uncertainty (grey bands) characterizing the concentration output is much narrower with the mobile–immobile transport model (Figure 8) than with the linear one (Figure 3). Hence, for the investigated undisturbed soil, the mobile–immobile transport model better reproduces bromide infiltration and yields much more accurate output concentrations than the linear one.

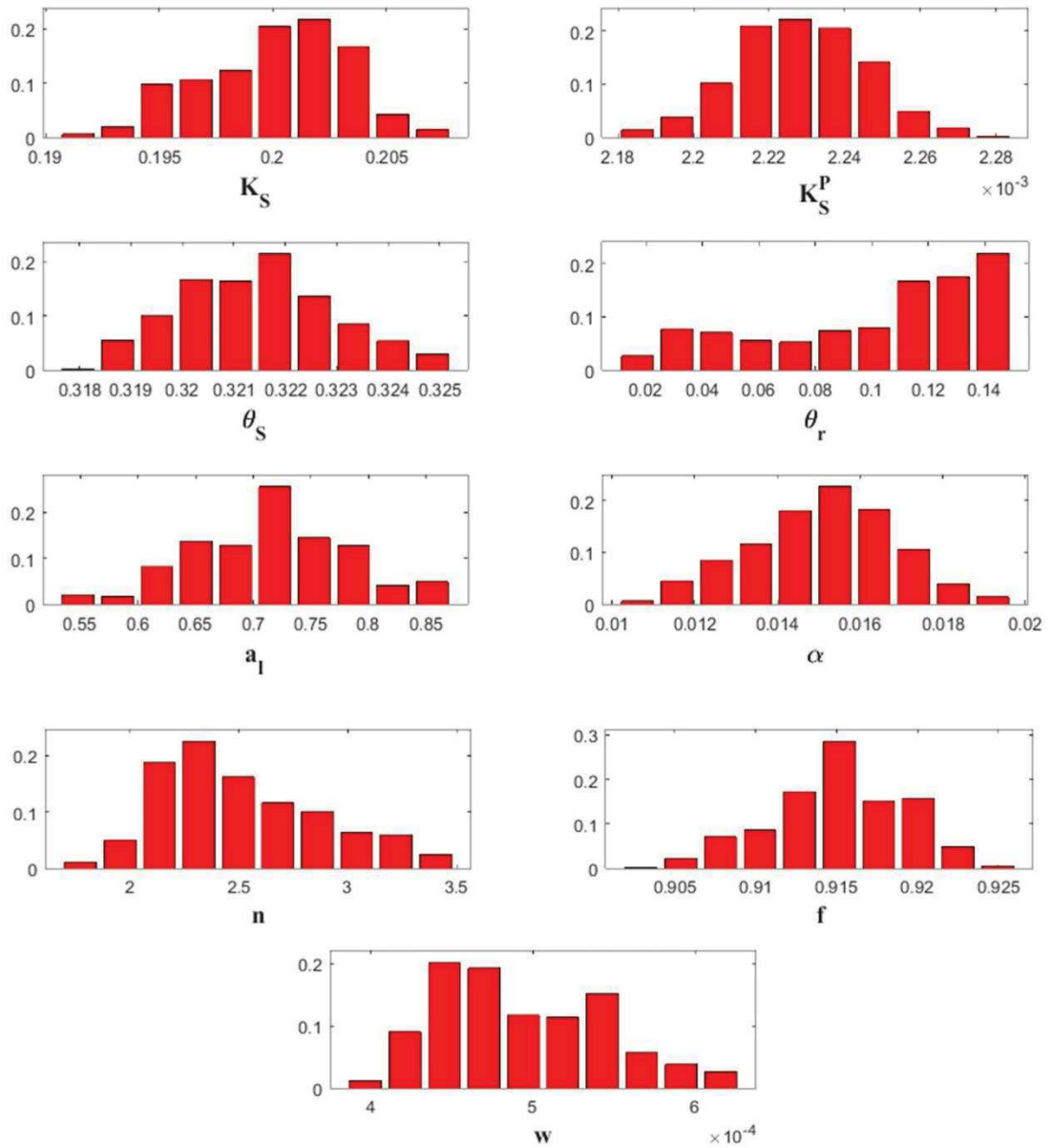
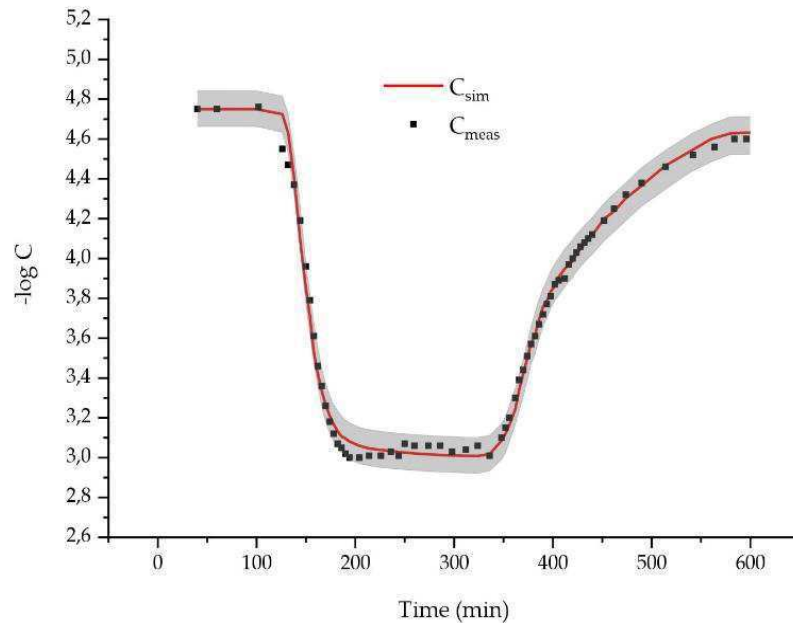


Figure 7. Histograms of marginal distribution of parameters with the mobile-immobile transport model.



**Figure 8.** Observed (symbols), calibrated (lines), and predictive uncertainty (grey band) for the effluent bromide concentration using the mobile-immobile transport model.

#### 4. Conclusions

In this study, a laboratory infiltration experiment was conducted by injecting bromide solution at the surface of an undisturbed soil column. The local pressure head measurements inside the column and the outflow concentration measurements were used to simultaneously estimate all hydraulic and transport soil parameters. Unsaturated water flow is simulated using the Richards equation coupled with the Mualem/van Genuchten models. The linear advection-dispersion and the non-equilibrium mobile-immobile models were investigated for modeling the bromide propagation. A Bayesian inversion was performed using the pressure head and concentration measurements. In the case of linear transport model, seven soil parameters were estimated: The saturated conductivity  $K_S$ , the saturated water content  $\theta_S$ , the residual water content  $\theta_r$ , the dispersivity coefficient  $a_l$ , and the van Genuchten parameters ( $\alpha$  and  $n$ ) of the soil and the saturated conductivity  $K_S^p$  of the porous plate. In the case of mobile-immobile model, the vector of unknown parameters contains, in addition, the fraction of mobile water content  $f$  and the exchange coefficient  $\omega$ . All hydraulic and transport soil parameters were estimated in a one-step procedure using the MCMC sampler starting from non-informative uniform distributions for all parameters. The obtained results are summarized hereafter:

- The flow through the investigated soil can be well reproduced by the Richards equation combined with the Mualem/van-Genuchten models.
- The pressure head and outflow concentration measurements are able to identify all the unsaturated hydraulic and solute transport soil parameters, except the residual water content  $\theta_r$  which was not identified with either of the models since dry conditions were not attained during the laboratory experiment.
- significant difference was observed between the two models for the estimation of the dispersivity coefficient. Indeed, with the mobile-immobile transport model, the estimated value of  $a_l$  was three times lower than the one obtained with the linear transport model. The associated uncertainty was also three times smaller with the mobile-immobile transport model.
- The linear transport model yielded a good agreement between the measured and simulated pressure heads, but a less satisfactory matching was observed between measured and simulated

concentrations. Furthermore, small uncertainties were obtained for pressure head responses, while very large uncertainties were assigned to the output concentrations.

- The mobile–immobile transport model better reproduced the infiltration experiment. Indeed, a very good matching was obtained between measured and simulated concentrations. Furthermore, the concentration uncertainty region was much narrower than with the linear transport model.

Finally, the results of this work point out the necessity of testing different models for interpreting real experiments and to employ accurate Bayesian methods for model calibration. These methods provide accurate information concerning the estimated parameters, their confidence intervals, and model output uncertainties.

**Author Contributions:** J.Z. and S.K. worked on the laboratory experiment; F.L. worked on the numerical simulations; A.Y. and M.F. supervised the research and finalized the manuscript.

**Funding:** This research was partially supported by the French National Research Agency (ANR Hydrocrizsto) and the Tunisian-French joint international laboratory NAILA (<http://www.lmi-naila.com/>).

**Acknowledgments:** The authors express their thanks to Radhouane Hamdi for his careful assistance during the experimental work.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Kool, J.B.; Parker, J.C.; van Genuchten, M.T. Determining Soil Hydraulic Properties from One-step Outflow Experiments by Parameter Estimation: I. Theory and Numerical Studies. *Soil Sci. Soc. Am. J.* **1985**, *49*, 1348–1354. [[CrossRef](#)]
2. Parker, J.C.; Kool, J.B.; van Genuchten, M.T. Determining Soil Hydraulic Properties from One-step Outflow Experiments by Parameter Estimation: II. Experimental Studies. *Soil Sci. Soc. Am. J.* **1985**, *49*, 1354–1359. [[CrossRef](#)]
3. Van Dam, J.C.; Stricker, J.N.M.; Droogers, P. Inverse Method to Determine Soil Hydraulic Functions from Multistep Outflow Experiments. *Soil Sci. Soc. Am. J.* **1994**, *58*, 647–652. [[CrossRef](#)]
4. Durner, W.; Schultze, B.; Zurmühl, T. *State-of-the-Art in Inverse Modeling of Inflow/Outflow Experiments*; University of California: Riverside, CA, USA, 1999.
5. Beydoun, H.; Lehmann, F. Expériences de drainage et estimation de paramètres en milieu poreux non saturé. *C. R. Geosci.* **2006**, *338*, 180–187. [[CrossRef](#)]
6. Puhlmann, H.; Von Wilpert, K.; Lukes, M.; Dröge, W. Multistep outflow experiments to derive a soil hydraulic database for forest soils. *Eur. J. Soil Sci.* **2009**, *60*, 792–806. [[CrossRef](#)]
7. Kumar, S.; Sekhar, M.; Reddy, D.V.; Mohan Kumar, M.S. Estimation of soil hydraulic properties and their uncertainty: Comparison between laboratory and field experiment. *Hydrol. Process.* **2010**, *24*, 3426–3435. [[CrossRef](#)]
8. Schelle, H.; Iden, S.C.; Peters, A.; Durner, W. Analysis of the Agreement of Soil Hydraulic Properties Obtained from Multistep-Outflow and Evaporation Methods. *Vadose Zone J.* **2010**, *9*, 1080–1091. [[CrossRef](#)]
9. Durner, W.; Iden, S.C. Extended multistep outflow method for the accurate determination of soil hydraulic properties near water saturation. *Water Resour. Res.* **2011**, *47*, 13. [[CrossRef](#)]
10. Younes, A.; Mara, T.A.; Fajraoui, N.; Lehmann, F.; Belfort, B.; Beydoun, H. Use of Global Sensitivity Analysis to Help Assess Unsaturated Soil Hydraulic Parameters. *Vadose Zone J.* **2013**, *12*, 1–12. [[CrossRef](#)]
11. Moreira, P.H.S.; van Genuchten, M.T.; Orlande, H.R.B.; Cotta, R.M. Bayesian estimation of the hydraulic and solute transport properties of a small-scale unsaturated soil column. *J. Hydrol. Hydromech.* **2016**, *64*, 30–44. [[CrossRef](#)]
12. Younes, A.; Zaouali, J.; Fahs, M.; Slama, F.; Grunberger, O.; Mara, T.A. Bayesian soil parameter estimation: Results of percolation-drainage vs infiltration laboratory experiments. *J. Hydrol.* **2018**, *565*, 770–778. [[CrossRef](#)]

13. Szabó, N.P. A genetic meta-algorithm-assisted inversion approach: Hydrogeological study for the determination of volumetric rock properties and matrix and fluid parameters in unsaturated formations. *Hydrogeol. J.* **2018**, *26*, 1935–1946. [[CrossRef](#)]
14. Abbaspour, K.C.; Johnson, C.A.; van Genuchten, M.T. Estimating Uncertain Flow and Transport Parameters Using a Sequential Uncertainty Fitting Procedure. *Vadose Zone J.* **2004**, *3*, 1340–1352. [[CrossRef](#)]
15. Abbaspour, K.C.; Schulin, R.; van Genuchten, M.T. Estimating unsaturated soil hydraulic parameters using ant colony optimization. *Adv. Water Resour.* **2001**, *24*, 827–841. [[CrossRef](#)]
16. Mishra, S.; Parker, J.C. Parameter estimation for coupled unsaturated flow and transport. *Water Resour. Res.* **1989**, *25*, 385–396. [[CrossRef](#)]
17. Inoue, M.; Šimůnek, J.; Shiozawa, S.; Hopmans, J.W. Simultaneous estimation of soil hydraulic and solute transport parameters from transient infiltration experiments. *Adv. Water Resour.* **2000**, *23*, 677–688. [[CrossRef](#)]
18. Laloy, E.; Weynants, M.; Bielders, C.L.; Vanclooster, M.; Javaux, M. How efficient are one-dimensional models to reproduce the hydrodynamic behavior of structured soils subjected to multi-step outflow experiments? *J. Hydrol.* **2010**, *393*, 37–52. [[CrossRef](#)]
19. Diamantopoulos, E.; Iden, S.C.; Durner, W. Inverse modeling of dynamic nonequilibrium in water flow with an effective approach. *Water Resour. Res.* **2012**, *48*, 16. [[CrossRef](#)]
20. Mara, T.A.; Delay, F.; Lehmann, F.; Younes, A. A comparison of two Bayesian approaches for uncertainty quantification. *Environ. Model. Softw.* **2016**, *82*, 21–30. [[CrossRef](#)]
21. Younes, A.; Mara, T.; Fahs, M.; Grunberger, O.; Ackerer, P. Hydraulic and transport parameter assessment using column infiltration experiments. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 2263–2275. [[CrossRef](#)]
22. Kaipio, J.; Somersalo, E. *Statistical and computational inverse problems*; Springer: New York, NY, USA, 2005; ISBN 978-0-387-22073-4.
23. Younes, A.; Mara, T.A.; Voltz, M.; Guellouz, L.; Musa Baalousha, H.; Fahs, M. A new efficient Bayesian parameter inference strategy: Application to flow and pesticide transport through unsaturated porous media. *J. Hydrol.* **2018**, *563*, 887–899. [[CrossRef](#)]
24. Lee, P.M. *Bayesian Statistics: An Introduction*, 4th ed.; Wiley: Chichester, West Sussex, UK; Hoboken, NJ, USA, 2012; ISBN 978-1-118-33257-3.
25. Zhu, H.; Liu, T.; Xue, B.; Wang, G. Modified Richards' Equation to Improve Estimates of Soil Moisture in Two-Layered Soils after Infiltration. *Water* **2018**, *10*, 1174. [[CrossRef](#)]
26. Mualem, Y. A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resour. Res.* **1976**, *12*, 513–522. [[CrossRef](#)]
27. Van Genuchten, M.T. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. *Soil Sci. Soc. Am. J.* **1980**, *44*, 892–898. [[CrossRef](#)]
28. Van Genuchten, M.T.; Wagenet, R.J. Two-Site/Two-Region Models for Pesticide Transport and Degradation: Theoretical Development and Analytical Solutions. *Soil Sci. Soc. Am. J.* **1989**, *53*, 1303–1310. [[CrossRef](#)]
29. Clothier, B.E.; Vogeler, I.; Green, S.R.; Scotter, D.R. Transport in unsaturated soil: Aggregates, macropores, and exchange. In *Physical Nonequilibrium in Soils: Modeling and Application*; Ann Arbor Press: Chelsea, MI, USA, 1998; pp. 273–295, ISBN 9781575040493.
30. Farthing, M.W.; Ogden, F.L. Numerical Solution of Richards' Equation: A Review of Advances and Challenges. *Soil Sci. Soc. Am. J.* **2017**, *81*, 1257–1296. [[CrossRef](#)]
31. Ku, C.-Y.; Liu, C.-Y.; Xiao, J.-E.; Yeih, W. Transient Modeling of Flow in Unsaturated Soils Using a Novel Collocation Meshless Method. *Water* **2017**, *9*, 954. [[CrossRef](#)]
32. Zha, Y.; Yang, J.; Zeng, J.; Tso, C.-H.M.; Zeng, W.; Shi, L. Review of numerical solution of Richardson-Richards equation for variably saturated flow in soils. *WIREs Water* **2019**, *6*, e1364. [[CrossRef](#)]
33. Fahs, M.; Younes, A.; Lehmann, F. An easy and efficient combination of the Mixed Finite Element Method and the Method of Lines for the resolution of Richards' Equation. *Environ. Model. Softw.* **2009**, *24*, 1122–1126. [[CrossRef](#)]
34. Fahs, M.; Koohbor, B.; Belfort, B.; Ataie-Ashtiani, B.; Simmons, C.; Younes, A.; Ackerer, P. A Generalized Semi-Analytical Solution for the Dispersive Henry Problem: Effect of Stratification and Anisotropy on Seawater Intrusion. *Water* **2018**, *10*, 230. [[CrossRef](#)]
35. Miller, C.T.; Abhishek, C.; Farthing, M.W. A spatially and temporally adaptive solution of Richards' equation. *Adv. Water Resour.* **2006**, *29*, 525–545. [[CrossRef](#)]

36. Younes, A.; Zaouali, J.; Lehmann, F.; Fahs, M. Sensitivity and identifiability of hydraulic and geophysical parameters from streaming potential signals in unsaturated porous media. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 3561–3574. [[CrossRef](#)]
37. Hindmarsh, A.C. LSODE and LSODI, two new initial value ordinary differential equation solvers. *ACM Signum Newsl.* **1980**, *15*, 10–11. [[CrossRef](#)]
38. Radhakrishnan, K. *Description and Use of LSODE, the Livermore Solver for Ordinary Differential Equations*; NASA report; NASA: Washington, DC, USA, 1993.
39. Laloy, E.; Vrugt, J.A. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing. *Water Resour. Res.* **2012**, *48*, W01526. [[CrossRef](#)]
40. Vrugt, J.A.; ter Braak, C.J.F.; Clark, M.P.; Hyman, J.M.; Robinson, B.A. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation: FORCING DATA ERROR USING MCMC SAMPLING. *Water Resour. Res.* **2008**, *44*, W00B09. [[CrossRef](#)]
41. Linde, N.; Ginsbourger, D.; Irving, J.; Nobile, F.; Doucet, A. On uncertainty quantification in hydrogeology and hydrogeophysics. *Adv. Water Resour.* **2017**, *110*, 166–181. [[CrossRef](#)]
42. Younes, A.; Delay, F.; Fajraoui, N.; Fahs, M.; Mara, T.A. Global sensitivity analysis and Bayesian parameter inference for solute transport in porous media colonized by biofilms. *J. Contam. Hydrol.* **2016**, *191*, 1–18. [[CrossRef](#)]
43. Hayek, M.; Younes, A.; Zouali, J.; Fajraoui, N.; Fahs, M. Analytical solution and Bayesian inference for interference pumping tests in fractal dual-porosity media. *Comput. Geosci.* **2018**, *22*, 413–421. [[CrossRef](#)]
44. Rajabi, M.M.; Ataie-Ashtiani, B.; Simmons, C.T. Model-data interaction in groundwater studies: Review of methods, applications and future directions. *J. Hydrol.* **2018**, *567*, 457–477. [[CrossRef](#)]
45. Gelman, A. *Bayesian Data Analysis*, 3rd ed.; Chapman & Hall/CRC Texts in Statistical Science; CRC Press: Boca Raton, FL, USA, 2014; ISBN 978-1-4398-4095-5.
46. Hastings, W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97–109. [[CrossRef](#)]
47. Gelman, A.; Rubin, D.B. Inference from Iterative Simulation Using Multiple Sequences. *Stat. Sci.* **1992**, *7*, 457–472. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



# Conclusion générale

Cette thèse est consacrée à l'étude des transferts d'eau et de soluté en milieux poreux non saturés à l'aide des méthodes d'analyse de sensibilité globale et d'estimation bayésienne des paramètres. Trois applications de transfert sur colonne de sol à l'échelle du laboratoire sont considérées et différentes variables (sorties) sont analysées. Parmi ces dernières la pression à l'intérieur du sol, le potentiel spontané (PS) induit par l'écoulement d'eau dans la zone non saturée, le flux cumulé et la concentration d'un contaminant non réactif à la sortie de la colonne.

La première application porte sur le couplage hydrogéophysique en milieu poreux non saturé. Le cas test étudié, correspond à une expérience de percolation drainage pendant laquelle le PS est mesuré à différentes altitudes. L'analyse de sensibilité globale du signal PS a été effectuée à l'aide des PCE. L'estimation des paramètres à partir des mesures de PS a été effectuée à l'aide de l'échantillonneur MCMC. Cette étude a permis de montrer que les résultats de l'analyse de sensibilité doivent être interprétés avec précaution dans un contexte d'estimation de paramètres. En effet, un paramètre qui n'est pas influent sur un capteur peut être influent sur un autre. De même, l'analyse de sensibilité globale ne préjuge pas de la qualité de l'estimation puisque deux paramètres avec des sensibilités similaires peuvent avoir une qualité d'estimation très différente. Cette étude a également permis de montrer que le signal PS contient assez d'informations pour une bonne estimation de tous les paramètres hydrauliques et électriques. En effet, quand plusieurs mesures PS (à différentes altitudes) sont utilisées pour la calibration, tous les paramètres sont estimés avec une relative faible incertitude. Dans ce cas, il faut noter que la méthode FOA donne les mêmes résultats que la méthode MCMC mais nécessite beaucoup moins de temps de calcul. Cependant, quand le nombre de mesures PS utilisées pour la calibration est réduit, la méthode FOA donne des valeurs de paramètres en accord avec la méthode MCMC. L'incertitude sur certains paramètres est cependant inexacte (incluant même des valeurs non physiques) avec la méthode FOA contrairement à la méthode MCMC.

La seconde application concerne l'identification des propriétés hydrauliques caractérisant l'écoulement non saturé dans un sol non remanié. L'étude comporte deux volets : un volet expérimental et un volet analyse de sensibilité et estimation de paramètres. Deux expériences sont effectuées avec des mesures de pression à l'intérieur de la colonne ainsi que du débit cumulé à la sortie de celle-ci. Pour la première, on a effectué une expérience de percolation-drainage et pour la seconde une expérience d'infiltration avec une injection d'un débit d'eau

constant à la surface du sol. Les paramètres hydrauliques du sol ont été étudiés à l'aide d'une modélisation inverse utilisant les mesures de la pression à l'intérieur de la colonne ainsi que le débit cumulé à la sortie. Les résultats de l'inversion montrent que les deux expériences permettent une bonne estimation de tous les paramètres hydrauliques du sol. Les valeurs des paramètres trouvées pour les deux expériences sont très proches et permettent de reproduire d'une façon très satisfaisante les données expérimentales. Les méthodes bayésiennes utilisées ont permis également le calcul d'une façon précise des incertitudes associées aux paramètres. Ces incertitudes ont des différences significatives entre les des deux expériences. En effet, les intervalles d'incertitude sont beaucoup plus réduits avec l'expérience d'infiltration qu'avec l'expérience de percolation-drainage. Ce résultat a pu être expliqué par l'analyse de sensibilité qui a montré une plus grande sensibilité des sorties du modèle par rapport aux paramètres dans le cas de l'expérience d'infiltration. Par conséquent, cette dernière est à privilégier pour la caractérisation des propriétés des sols non saturés.

La dernière application est une expérience de transfert d'eau et de contaminant réalisée au laboratoire en injectant une solution de bromure à la surface d'un sol non remanié. Durant l'infiltration, la pression a été mesurée à différentes altitudes, de même que la concentration à la sortie de la colonne. Ces observations sont ensuite utilisées pour une estimation simultanée de tous les paramètres hydrauliques et de transport. Les résultats obtenus montrent que les observations permettent d'identifier tous les paramètres hydrauliques et de transport à l'exception de la teneur en eau résiduelle. En effet, cette dernière n'a pas pu être identifiée parce que le sol n'a pas été suffisamment asséché (l'humidité est restée importante durant toute l'expérience). Le modèle de transport mobile-immobile a permis de mieux reproduire les résultats de l'expérience d'infiltration étudiée. En effet, ce modèle donne une meilleure estimation (avec moins d'incertitude) pour certains paramètres du sol. De plus, il aboutit à un meilleur ajustement du profil de concentration à la sortie de la colonne ainsi qu'à une incertitude associée plus réduite qu'avec le modèle de transport linéaire.

En résumé, dans cette thèse, nous avons combiné les approches expérimentales de laboratoire et les approches de modélisation directe et inverse pour l'étude des transferts en milieux poreux non saturés. Ce travail a pu mettre en valeur l'intérêt de l'analyse de sensibilité globale et de l'estimation bayésienne des paramètres pour l'étude de ces transferts. En effet l'analyse de sensibilité globale permet non seulement de détecter les paramètres influents, mais également les intervalles et les périodes de grande influence. Ceci peut être d'une grande utilité pour l'expérimentateur afin de choisir les positions des capteurs ainsi que le temps où une attention

spéciale doit être accordée aux mesures. L'estimation bayésienne des paramètres permet d'avoir plusieurs informations car elle donne accès non seulement aux valeurs optimales des paramètres, mais également à l'incertitude ainsi qu'à la leur distribution à posteriori.

Dans les perspectives de ce travail, nous envisageons d'explorer les transferts avec sorption et dégradation dans les sols (typiquement contamination par les pesticides). Dans ce cas, le contaminant réagit avec le milieu poreux à cause du phénomène de sorption-désorption et peut se dégrader avec le temps. Nous prévoyons également de combiner les méthodes d'analyse de sensibilité globale et les méthodes bayésiennes d'estimation des paramètres pour l'analyse de la valeur des données (data-worth analysis) où le but est d'étudier à priori l'intérêt des mesures avant de les effectuer. Ceci permet de choisir les mesures les plus intéressantes à effectuer pour la caractérisation des propriétés hydrauliques des sols.

## Analyse de sensibilité globale et estimation Bayésienne de paramètres : application aux transferts en milieu poreux non saturé

### Résumé

L'étude des transferts d'eau et de contaminants en milieu poreux non saturé requiert une importance majeure afin de préserver la qualité des eaux et des sols. Dans cette thèse, nous développons les méthodes d'analyse de sensibilité globale et d'estimation bayésienne de paramètres pour l'étude des transferts en milieux poreux non saturés. Trois applications de transfert sur colonne de sol à l'échelle du laboratoire sont considérées. La première porte sur le couplage hydrogéophysique en milieu poreux non saturé. La seconde concerne l'identification des propriétés hydrauliques caractérisant l'écoulement non saturé dans un sol non remanié à partir d'expériences de drainage et d'infiltration. La dernière application est une expérience de transfert d'eau et de contaminant réalisée en injectant une solution de bromure à la surface d'un sol non remanié.

**Mots clés :** équation de Richards, potentiel spontané, équation d'advection-dispersion, modèle immobile-mobile, analyse de sensibilité globale, polynôme de chaos, chaîne de Markov de Monte-Carlo, expériences de laboratoire.

### Abstract

The study of water and contaminant transfers in unsaturated porous media is of major importance in order to preserve water and soil quality. In this thesis, we develop the methods of global sensitivity analysis and Bayesian parameter estimation for the study of transfers in unsaturated porous media. Three laboratory scale soil transfer applications are considered. The first deals with hydrogeophysical coupling in unsaturated porous media. The second concerns the identification of hydraulic properties characterizing unsaturated flow in undisturbed soil from drainage and infiltration experiments. The last application is a water and contaminant transfer experiment performed by injecting a bromide solution on the surface of undisturbed soil.

**Key words:** Richards' equation, streaming potential, advection-dispersion equation, immobile-mobile model, global sensitivity analysis, chaos polynomial, Monte Carlo Markov chain, laboratory experiments.