

Quantification des incertitudes liées aux simulations d'écoulement dans un réservoir pétrolier à l'aide de surfaces de réponse non linéaires

Emmanuel Fetel

▶ **To cite this version:**

Emmanuel Fetel. Quantification des incertitudes liées aux simulations d'écoulement dans un réservoir pétrolier à l'aide de surfaces de réponse non linéaires. Sciences de la Terre. Institut National Polytechnique de Lorraine, 2007. Français. NNT : 2007INPL005N . tel-01752837

HAL Id: tel-01752837

https://hal.univ-lorraine.fr/tel-01752837

Submitted on 29 Mar 2018

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.

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Institut National
Polytechnique de Lorraine

École Nationale Supérieure de Géologie

École doctorale RP2E

Quantification des incertitudes liées aux simulations d'écoulement dans un réservoir pétrolier à l'aide de surfaces de réponse non linéaires

THÈSE

présentée et soutenue publiquement le 12 Janvier 2007

pour l'obtention du

Doctorat de l'Institut National Polytechnique de Lorraine

Spécialité Géosciences

par

Emmanuel FETEL

Composition du jury

Rapporteurs :	Jean-Paul CHILÈS Richard SINDING-LARSEN
Examinateurs :	Jean-Laurent MALLET William J. MILLIKEN
Directeur :	Jean-Jacques ROYER

Centre de Recherches Pétrographiques et Géochimiques
Laboratoire d'Infographie et d'Analyse de Données
Rue du Doyen Marcel Roubault - 54500 Vandœuvre

Mis en page avec la classe thloria.

Remerciements

Tout d'abord, je tiens à remercier Jean-Jacques Royer, mon directeur de thèse, qui a cru en mes capacités et m'a accueilli durant ces années de thèse. Je tiens aussi à remercier l'ensemble de la famille Gocad et le premier d'entre eux, Jean Laurent Mallet, pour avoir su créer un cadre de travail passionnant et sans lequel je n'aurais probablement pas suivi la voie dans laquelle je me suis engagé. Je tiens également à le remercier pour les idées qu'il a bien voulu sortir de son chapeau pour moi, en particulier en ce qui concerne la méthode de lissage d'histogramme (cf. chapitre 5).

Mes remerciement vont ensuite à l'Energy Technology Company (ETC) de Chevron qui a financé ces travaux de thèse. De ce fait, j'ai eu le plaisir de discuter avec William J. Milliken, Alexandre Castellini, Yeten Burak et Jorge Landa dont les conseils et commentaires pertinents m'ont guidés et éclairés durant ces travaux, en particulier, dans cette dernière phase qu'est la rédaction du manuscrit.

Je suis également reconnaissant à Jean-Paul Chilès, Professeur à l'École des Mines de Paris et à Richard Sinding-Larsen, Professeur à l'Université de Science et Technologie (NTNU) de Trondheim en Norvège pour m'avoir fait l'honneur d'être les rapporteurs de ce mémoire.

Ma deuxième salve de remerciements s'adresse plus particulièrement à mes coreligionnaires. Aux premiers desquels mes infatigables relecteurs qui ont grandement contribué à améliorer la lisibilité de ce texte (Je sais néanmoins qu'ils trouveront un moyen de se venger, si ce n'est déjà fait !) : Laeti (ma première collègue de bureau qui a eu le plaisir de me supporter durant 4 ans), Sarah (collègue de bureau bis qui n'a eu, elle, le droit de m'avoir que pendant 8 mois) et puisqu'il faut sacrifier à la tradition [Ledez, 2003], Christian Lecarlier-de-Veslud (a.k.a Cri-Cri [Voillemont, 2003]) à l'esprit consciencieux et cartésien [Viseur, 2001] mais non dénué d'humour [Macé, 2006]. Je remercie également les thésards du labo avec lesquels j'ai eu le plaisir de discuter passionnément, d'avoir quelques prises de bec et surtout qui ont eu à subir les surfaces de réponse pendant quatre ans : Pierre K. (que je ne désespère pas de battre un jour au squash), Marco (mais si l'école doctorale sert à quelque chose...), Luc (qui au moment même où j'écris ces lignes est contraint de suivre une conférence à Lake Tahoe, .. trop dure la vie !), Anne-Laure (la charmante blondinette d'outre-manche !) et Laurent C. (qui m'a battu d'un fil) ainsi que les

Anciens : Bruno, Pimu, Rémi, Souchy, Tobias, plus tant d'autres que je dois oublier et le fils prodigue qui est revenu (Guillaume).

Bien évidemment, je remercie chaleureusement Mme Cugurno ainsi que les secrétaires de l'ASGA pour leurs disponibilité, sympathie et dont l'aide m'a été précieuse pour survivre dans les rouages de l'administration.

Je souhaite également bon vent à tous les futurs thésard, masters, 2A, etc. qui auraient l'envie, pour ne pas dire le courage, de s'engager en géologie numérique ... sincèrement il n'y a aucune raison d'être *poor* ou *desesperate*.

Cali.
a.k.a Dooschy, Bernardo, Manu, etc.



À mes parents, et à leur soutient sans faille durant mes longues années d'étude

Il y a deux manières d'ignorer les choses : la première, c'est de les ignorer ; la seconde, c'est de les ignorer et de croire qu'on les sait. La seconde est pire que la première. [Victor Hugo]

Table des matières

Chapitre 1 Introduction **1**

1.1 Caractérisation et modélisation du sous-sol 1

1.1.1 Les différentes étapes de la modélisation du sous-sol 1

1.1.2 Rôle de la simulation d'écoulement dans la modélisation du sous-sol . . . 4

1.2 Modèles et techniques pour la simulation d'écoulement en milieux poreux 4

1.2.1 Modèles physiques pour la description des fluides 4

1.2.2 Résolution numérique 5

1.2.3 Problématiques spécifiques aux simulations d'écoulement 7

1.3 Les incertitudes dans la modélisation du sous-sol 7

1.3.1 Classification des incertitudes 7

1.3.2 Objectifs de la caractérisation et de la quantification des incertitudes . . . 8

1.3.3 Méthodes pour la caractérisation des incertitudes liées aux simulations d'écoulement 9

1.4 Objectif et organisation de la thèse 13

Chapitre 2 Construction et analyse de surface de réponse non linéaire **15**

2.1 Nature des données disponibles et notations 16

2.1.1 Données disponibles 16

2.1.2 Notations 17

2.2 État de l'art : méthodes pour la construction d'une surface de réponse 18

2.3 Krigeage dual avec terme de régularisation 20

2.3.1 Principe 20
2.3.2 Formalisme dual 22
2.3.3 Terme de régularisation 23
2.3.4 Remarques finales 24
2.4 Interpolation lisse discrète 25
2.4.1 Méthodologie 25
2.4.2 Structures de données 28
2.4.3 Contraintes DSI pour la construction de surfaces de réponse 29
2.4.4 Remarques finales 32
2.5 Méthodes d'analyse d'une surface de réponse 33
2.5.1 Quantification des incertitudes sur les prédictions de production 33
2.5.2 Analyse de sensibilité aux paramètres incertains 34
2.5.3 Inversion bayésienne de données de production 41
2.6 Cas d'études 45
2.6.1 Modèle Anticlinal avec 4 paramètres incertains 46
2.6.2 Mesures de prédictivité des méthodes de construction sur le modèle "*IC Fault*" 55
2.6.3 Évaluation de la performance des méthodes de construction 58
2.7 Conclusions 58

Chapitre 3 Prise en compte de données secondaires **61**

3.1 Approche proposée 62
3.2 Planification d'expérience pour l'intégration d'une variable secondaire 63
3.3 Méthodes pour la construction de surfaces de réponse en présence de donnée secondaire 65
3.3.1 Contraintes pour l'interpolation lisse discrète 65
3.3.2 Krigeage avec dérive externe 68
3.4 Quantification de l'apport d'information 71
3.4.1 Mesure de "*Root Mean Square Error*" 72

3.4.2 Validation . . . 72
3.5 Cas d'étude basé sur le modèle de réservoir SPE10 . . . 74
3.5.1 Présentation du modèle . . . 74
3.5.2 Construction de surface de réponse du débit en huile après 500 jours . . . 76
3.5.3 Analyse et discussion des résultats . . . 78
3.6 Conclusions . . . 80

Chapitre 4 Prise en compte de paramètres incertains stochastiques 83

4.1 Construction de surfaces de réponse en présence de paramètres stochastiques . . . 85
4.1.1 Principe . . . 85
4.1.2 Estimation par modélisation séparée . . . 86
4.2 Extension des méthodes d'analyse à la prise en compte de paramètres stochastiques 87
4.2.1 Analyse de sensibilité de la production vis-à-vis de paramètres incertains 87
4.2.2 Inversion bayésienne d'un historique de production . . . 91
4.3 Application sur un modèle de réservoir fluviatile . . . 92
4.3.1 Description du modèle de réservoir . . . 92
4.3.2 Construction et analyse des surfaces de réponse . . . 95
4.3.3 Remarques finales . . . 99
4.4 Prise en compte de paramètres continus peu influents . . . 100
4.5 Conclusions . . . 102

Chapitre 5 Méthode pour le lissage d'histogramme 105

5.1 Lissage d'histogramme : définitions, objectifs et méthodes . . . 106
5.1.1 Histogramme et diagramme de fréquence . . . 106
5.1.2 État de l'art des méthodes de lissage d'histogramme . . . 107
5.2 Approche proposée . . . 110
5.2.1 Principes . . . 110
5.2.2 Contraintes pour l'interpolation lisse discrète . . . 110
5.2.3 Résultats préliminaires . . . 114

5.3 Analyse de la sensibilité de la méthode à ses différents paramètres 114

5.4 Prise en compte de données lissée pour la simulation de propriété par "*cloud tranform*" 119

5.5 Conclusions 121

Conclusions et perspectives **123**

Bibliographie **127**

Annexe A Simulation d'écoulement sur lignes de courant dans le géomodeleur Gocad **137**

A.1 Principe d'une simulation d'écoulement biphasique sur lignes de courant 138

A.1.1 Équations fondamentales 138

A.1.2 Méthodologie 139

A.1.3 Traitement de la gravité 140

A.2 Résolution numérique 141

A.2.1 Modèle discret 142

A.2.2 Résolution de l'équation de pression 142

A.2.3 Méthode de tracé des lignes de courant 147

A.2.4 Résolution de l'équation de saturation 151

A.2.5 Transfert des résultats des lignes de courant vers la grille 156

A.2.6 Modélisation de l'effet de la gravité 156

A.3 Apport de StreamLab-2.0.0 157

A.4 Structure de données pour la représentation et la manipulation de lignes de courant 158

A.4.1 Structures de données pré-existantes 159

A.4.2 Implantation proposée 161

A.5 Validation du calcul 163

A.5.1 Description du modèle 163

A.5.2 Résultats 164

A.6 Conclusions 167

Annexe B Description des structures de données NdPointSet et NdCube 169

B.1 Implantation d'objets dans le géomodeleur Gocad 169

B.2 La structure de donnée NdPointSet 170

B.3 La structure de donnée NdCube 172

B.4 Remarques finales 174

Chapitre 1

Introduction

Sommaire

1.1 Caractérisation et modélisation du sous-sol 1
1.1.1 Les différentes étapes de la modélisation du sous-sol 1
1.1.2 Rôle de la simulation d'écoulement dans la modélisation du sous-sol . . 4
1.2 Modèles et techniques pour la simulation d'écoulement en milieux poreux 4
1.2.1 Modèles physiques pour la description des fluides 4
1.2.2 Résolution numérique 5
1.2.3 Problématiques spécifiques aux simulations d'écoulement 7
1.3 Les incertitudes dans la modélisation du sous-sol 7
1.3.1 Classification des incertitudes 7
1.3.2 Objectifs de la caractérisation et de la quantification des incertitudes . . 8
1.3.3 Méthodes pour la caractérisation des incertitudes liées aux simulations d'écoulement 9
1.4 Objectif et organisation de la thèse 13

1.1 Caractérisation et modélisation du sous-sol

1.1.1 Les différentes étapes de la modélisation du sous-sol

La caractérisation et la modélisation du sous-sol dans le cadre d'un gisement pétrolier est un processus long et complexe. Les principaux objectifs sont d'évaluer la position, la quantité et l'organisation spatiale des accumulations (modélisation statique), mais aussi de proposer un éventuel schéma de production (nombre et position des puits, sélection de méthodes d'injection d'eau, de CO_2 ou de polymères, etc.) afin d'optimiser la récupération des hydrocarbures (modélisation dynamique).

Cette caractérisation met en œuvre de nombreux outils et méthodologies liés à des disciplines extrêmement variées (géologie structurale, sédimentologie, géostatistique, ingénierie de réservoir,

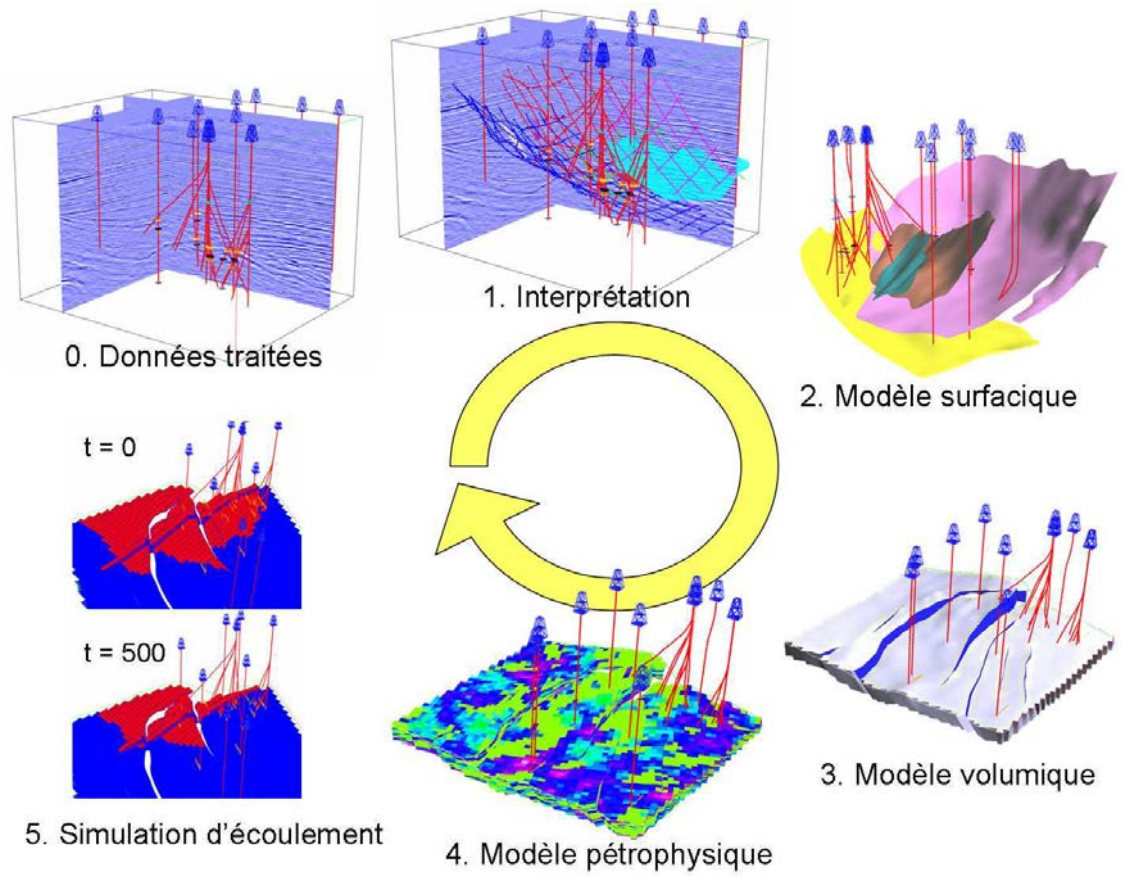


FIG. 1.1 – Étapes de l'exploration et de la modélisation d'un gisement pétrolier. Le travail présenté dans ce mémoire concerne la cinquième simulation d'écoulement.

etc.). Ces outils visent à produire un modèle représentatif de la réalité du sous-sol à partir des données disponibles qui sont, le plus souvent, éparses, hétérogènes et imprécises. On distingue en particulier :

- les données sismiques tri-dimensionnelles qui couvrent l'ensemble du volume étudié mais ont une résolution assez faible (de 5 à 10m lors des campagnes sismiques les plus précises) ;
- les données de puits (carottes, logs, cuttings, etc.) précises (la résolution est centimétrique pour les carottes à décimétrique pour les logs) mais éparses et ponctuelles à l'échelle du réservoir ;
- les informations qualitatives que sont l'analyse du contexte régional d'un réservoir, l'étude d'analogues et l'utilisation de concepts géologiques tels que la stratigraphie séquentielle ou bien, en quelque sorte, les images d'entraînement dans les statistiques multi-points.

À partir de ces données, l'étude et la modélisation d'un gisement pétrolier sont divisées en plusieurs grandes étapes (figure (1.1)) :

1. traitement et filtrage des données brutes tels que le géo-référencement, la conversion temps/profondeur du signal sismique, etc. L'objectif est d'obtenir un ensemble de données cohérentes sur lesquelles les géologues et les ingénieurs peuvent bâtir un modèle du réservoir ;

2. construction d'un modèle structural qui représente les failles et les horizons majeurs présents dans le réservoir et délimitant ses différents compartiments ;

3. construction d'un modèle volumique discret. Celui-ci représente le volume à l'intérieur du réservoir et sert essentiellement de support de calculs ;

4. estimation d'un ou plusieurs modèles de propriétés dans le volume étudié (faciès, porosité, net-to-gross, perméabilité, etc.). Ce modèle est établi, le plus souvent, à l'aide de méthodes géostatistiques qui intègrent les données de puits, les données sismiques et les connaissances géologiques sur le réservoir ;

5. création d'un modèle de simulation d'écoulement en milieux poreux. Ce modèle est dynamique et a pour objectif de permettre la prédiction de la production en hydrocarbures et l'évolution de la distribution des fluides. Il s'appuie sur (1) le modèle volumique et le modèle pétrophysique qui décrivent la structure et les propriétés du réservoir, (2) la connaissance de la nature des fluides à prendre en compte et de leurs propriétés (viscosité, température, relations PVT, etc.), et (3) l'état initial du réservoir et les conditions aux limites (puits, aquifère, etc.). La simulation d'écoulement étant un processus coûteux en calculs et en temps, il est courant de construire un modèle volumique grossier et de calculer les propriétés équivalentes correspondant à celles simulées à l'étape 4 à l'aide de méthodes de mise à l'échelle (ou d'"*upscaling*").

Ces étapes forment un processus de modélisation intégré : toutes les méthodes envisagées partagent les même données, mais interagissent aussi entre elles. En effet, au cours de la modélisation, certains résultats peuvent naturellement contraindre le choix et l'utilisation d'algorithmes spécifiques pour les étapes suivantes, mais surtout, ils peuvent amener à remettre en cause des hypothèses et choix faits précédemment. Cependant, on peut distinguer les étapes (1) à (4) de l'étape (5). Les premières permettent d'obtenir un modèle statique du réservoir qui inclut la description de sa géométrie et de ses propriétés (porosité, perméabilité, etc.). Ce modèle statique suffit à l'estimation du volume poreux du gisement et des réserves en place. Ensuite, la simulation d'écoulement (étape 5) utilise ce modèle pour étudier la dynamique des fluides dans le réservoir[1]. Cette étape se distingue par l'utilisation de nouvelles sources de données spécifiques (historique de production, test de pression aux puits, analyse PVT, etc.) et par un certain nombre de problématiques qui lui sont propres.

[1] Actuellement cette distinction s'estompe légèrement dû à l'utilisation de données dynamiques pour vérifier les hypothèses faites lors de la construction du modèle statique.

1.1.2 Rôle de la simulation d'écoulement dans la modélisation du sous-sol

Comme décrit précédemment, la simulation d'écoulement en milieux poreux est l'un des aboutissements majeurs de l'ensemble du processus de modélisation et de caractérisation du sous-sol. C'est une vaste discipline en ingénierie de réservoir qui inclut des connaissances en mécanique des fluides, géologie et modélisation numérique. Son but principal est de prédire le transport dans le milieu poreux d'une ou plusieurs espèces chimiques (ou composantes) au sein d'une ou plusieurs phases fluides (huile, eau, gaz, etc.). Les écoulements sont le plus souvent simulés en utilisant un ou plusieurs programmes informatiques qui calculent une solution numérique approchée des équations aux dérivées partielles gouvernant les phénomènes physiques complexes ayant lieu dans le réservoir. Les principales applications de ces outils sont :

1. la prédiction du potentiel de production d'un réservoir et ce sous différents scénarios possibles, par exemple, différents plans de développement de puits ou différents processus de production (injection d'eau, de gaz, ou polymères, etc.) ;

2. l'identification du ou des mécanismes de production (déplétion pure, expansion/libération du gaz dissout dans l'huile, présence d'un aquifère, injections, etc.) et la compréhension des processus physiques ayant lieu à l'intérieur du réservoir ;

3. la gestion et l'optimisation de la production d'un réservoir ;

4. la prise en compte d'un historique de production pour améliorer la description et la modélisation de la géométrie et des propriétés du réservoir par le biais de méthodes d'inversion (*history matching*) ;

5. la quantification des incertitudes sur les résultats de simulation d'écoulement. En effet, la connaissance d'un réservoir est partielle et basée sur une quantité limitée de données. Or il faut pouvoir mesurer et apprécier le risque qui se trouve associé à la plupart des décisions de gestion d'un réservoir.

La section suivante introduit et présente les grandes composantes d'une simulation d'écoulement. L'objectif n'est pas de détailler le fonctionnement d'un simulateur d'écoulement, pour cela le lecteur est invité à consulter l'abondante littérature sur le sujet, par exemple Aziz et Settari [1979]; Dake [2001] ou Ertekin et al. [2001] ainsi que l'annexe A qui décrit un simulateur sur lignes de courant implanté durant ces travaux de thèse.

1.2 Modèles et techniques pour la simulation d'écoulement en milieux poreux

1.2.1 Modèles physiques pour la description des fluides

Afin de pouvoir rendre compte de la complexité du milieu poreux, de la diversité des réservoirs pétroliers, du type d'écoulement (monophasique, biphasique etc.) et des procédés d'exploitation

(déplétion, injection d'eau ou de gas etc.) différents modèles de comportement physique des fluides ont été mis en place et décrits dans la littérature. L'objectif est de disposer d'outils capables de décrire la réalité du sous-sol mais aussi suffisamment simples pour être le moins coûteux possible en temps et en calculs. On distingue notamment :

- le modèle monophasique : il permet de simuler un réservoir dans lequel une seule phase est présente ;
- le modèle huile morte ("Dead-Oil") : il considère un écoulement diphasique huile-eau et suppose que l'huile est à une pression suffisamment basse pour qu'elle ait perdu ses composantes volatiles et ne contienne plus de gaz dissout ;
- le modèle huile noire ("Black-Oil") encore appelé modèle "Live Oil" : à l'opposé du précédent, celui-ci permet de modéliser un gisement où une fraction de gaz est dissoute dans l'huile et vice-versa. C'est actuellement le modèle le plus utilisé dans l'industrie pétrolière. En effet, il est suffisamment flexible pour rendre compte de la plupart des comportements fluides, mais aussi suffisamment simple en terme de nombre d'inconnues pour limiter le volume de calculs nécessaires ;
- le modèle compositionel : les fluides sont divisés en n_p phases (avec le plus souvent $n_p \leq 3$) et n_c composantes ou espèces chimiques, chaque espèce pouvant être présente dans la plupart des phases. Ce modèle permet de rendre compte de la diversité des composants chimiques dans les hydrocarbures ainsi que leurs comportements au sein de chaque phase. À cause du grand nombre d'inconnues et de la complexité des phénomènes à modéliser, ce modèle requiert un volume important de calculs, ce qui limite son utilisation. Cependant, avec l'augmentation de la puissance des ordinateurs il devient de plus en plus répandu.

Ces modèles physiques se traduisent par une ou plusieurs équations aux dérivées partielles (EDP) fortement couplées. Celles-ci sont obtenues en combinant les lois de conservation d'une ou plusieurs grandeurs physiques (masse, énergie, etc.) et la loi de Darcy qui relie le flux avec la différence de pression entre deux points du réservoir.

1.2.2 Résolution numérique

Une fois le modèle physique mis en place, il est illusoire de chercher une solution analytique au problème sauf au prix d'hypothèses simplificatrices. Il faut donc chercher une solution approchée en se basant sur une représentation discrète du sous-sol et des méthodes de résolution numérique des EDP. Le calcul de cette solution approchée s'appuie sur une discrétisation spatiale et temporelle des différentes équations puis nécessite un programme informatique spécifique. Les différentes composantes d'un simulateur numérique d'écoulement sont :

1. **La discrétisation spatiale** des équations dirigeant l'écoulement. Dans le cadre d'un simulateur d'écoulement isotherme en milieu poreux, cette discrétisation s'effectue généralement suivant l'approche des volumes finis. Cette technique basée sur l'expression des flux massiques à la frontière de volumes de contrôle a pour avantage d'assurer le respect strict des lois de conservation. L'estimation des flux au niveaux de chaque face du volume de

contrôle est réalisée à l'aide de schéma de type différences finies. On distingue les approximations à deux point (méthodes "TPFA") simples à mettre en œuvre et celles à plusieurs points (méthodes "MPFA") qui ont été développées plus récemment afin de mieux rendre compte de tenseurs de perméabilité complets ainsi que de grilles non K-orthogonales.

2. **La discrétisation temporelle** des équations. Dans un premier temps, il est nécessaire de spécifier la position dans le temps de la discrétisation. Considérons que le but est d'estimer la solution au temps $n+1$. Si un traitement explicite est choisi, alors les termes utilisés pour calculer le flux proviennent du pas de temps précédent, n, et leurs valeurs sont connues. À l'opposé, un traitement implicite requiert que la discrétisation s'effectue en $n+1$ et la valeur exacte des termes nécessaires est inconnue. Deux approches sont principalement utilisées, (1) l'approche entièrement implicite (ou FIM pour "*Fully Implicit Method*") et (2) les approches mixtes, en particulier l'approche IMPES (IMplicite en Pression Explicite en Saturation). Les approches mixtes utilisent le fait que les variables et les équations ont des comportements mathématiques et numériques différents ce qui permet de les résoudre séparément. La sélection du temps de discrétisation à un impact quant à la rapidité, la précision et la convergence des calculs. Il s'agit donc d'un choix important de modélisation qui doit être justifié au regard des phénomènes à modéliser et des objectifs de l'étude.

3. **La résolution du système linéaire** : toutes les méthodes de discrétisation ont pour but de transformer en chaque nœud du maillage une EDP et une équation algébrique faisant intervenir la valeur de l'inconnue au nœud et celles de certains de ses voisins. Exprimé sur l'ensemble du maillage, le processus de discrétisation conduit à un ou plusieurs systèmes linéaires qu'il faut résoudre en fonction de l'état initial et de conditions aux limites. Il existe de très nombreuses méthodes pour résoudre un tel système. Le choix, le détail et l'implantation de la méthode de résolution choisie est au-delà de la portée de ce travail. Pour plus de détails, le lecteur est invité à consulter la littérature sur le sujet, en particulier Barret et al. [1994] et Press et al. [2002].

4. **Le calcul d'un critère stabilité** : le choix d'un processus de discrétisation a des implications quant à la stabilité, la convergence, la rapidité et la précision de la simulation. Pour assurer la convergence et la stabilité de la simulation, il est souvent nécessaire de limiter le pas de temps. Théoriquement, les méthodes de type FIM sont extrêmement stables ce qui permet de considérer un pas de temps relativement large, mais au prix d'un système matriciel complexe à inverser. À l'opposé, les méthodes mixtes impliquant moins de variables au niveau implicite sont plus rapides. Cependant, elles sont aussi moins stables et peuvent, si le comportement des fluides est fortement non-linéaire, nécessiter un pas de temps prohibitivement petit. Ainsi, il est nécessaire de trouver un équilibre entre la taille du système et le nombre de fois qu'il sera nécessaire de l'inverser.

Pour plus de détails sur les différentes étapes d'une simulation, le lecteur est invité à consulter la littérature sur le sujet, mais aussi l'annexe A. Celle-ci détaille un simulateur d'écoulement biphasique en milieux poreux sur lignes de courant développé durant ces travaux. Son implantation a permis de disposer dans le cadre de cette thèse d'un outil fiable et modulable au sein du géomodeleur Gocad pour servir de base à certains des travaux présentés dans ce manuscrit.

1.2.3 Problématiques spécifiques aux simulations d'écoulement

La simulation d'écoulement a pour objectif de calculer de manière fiable et précise le potentiel de production d'un modèle de réservoir donné ainsi que son évolution dans le temps. Cependant les problèmes spécifiques suivants se posent :

- la simulation est un processus long, complexe et coûteux en volume de calculs mais surtout en temps : une simulation d'écoulement sur un modèle de taille raisonnable (10^5~10^6 cellules) peut ainsi durer plusieurs heures ;
- la quantité de résultats à exploiter est importante : les mesures peuvent être locales et associées aux différents puits (débits, saturations, rapport gaz sur huile, etc.) ou globales sur l'ensemble du réservoir (production cumulée, pression, etc.). De plus, leur évolution dans le temps doit être considérée ;
- l'exploitation qualitative de ces résultats est difficile à cause de la nature complexe de la dépendance des résultats de simulation vis-à-vis des paramètres du modèle.

Ces problèmes déjà significatifs pour l'étude d'un modèle de réservoir donné deviennent bien plus importants dans le cadre de l'analyse des incertitudes liées aux simulations d'écoulement à cause du grand nombre de modèles possibles qu'il est nécessaire de gérer.

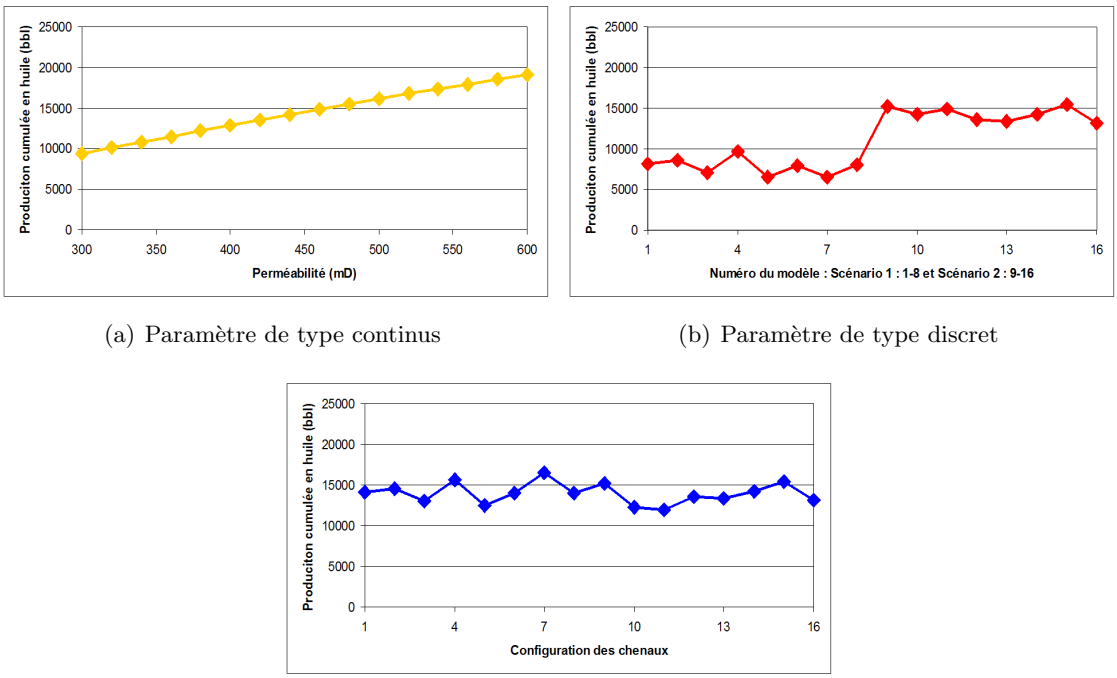
1.3 Les incertitudes dans la modélisation du sous-sol

La modélisation et la caractérisation du sous-sol est par essence incertaine : d'une part, les données disponibles sont limitées, éparses, hétérogènes et, hormis les données de carottes, ce sont des mesures indirectes des propriétés du sous-sol ; d'autre part, les algorithmes employés pour modéliser le sous-sol sont eux-mêmes incertains : ils ne sont qu'une approximation de la réalité et comportent de nombreuses inconnues (vitesse de propagation des ondes en profondeur ou taille des objets géologiques à simuler, etc.) ; pour finir, la sélection des différentes méthodes employées repose sur des choix subjectifs de modélisation qui apportent leurs parts d'incertitude sur le résultat obtenu. L'ensemble de ces incertitudes influence de manière considérable les résultats de la modélisation et doit être pris en compte. En effet, d'une manière générale, ce n'est pas tant le véritable potentiel de production d'un réservoir que l'incertitude que l'on a dessus qui décide de la vie ou de la mort d'un projet.

1.3.1 Classification des incertitudes

Bien que les sources d'incertitudes soient nombreuses et variées, on distingue différentes catégories de paramètres incertains [Zabalza-Mezghani et al., 2004] (figure (1.2)) :

- *Continus* : un paramètre incertain est dit continu quand il prend ses valeurs de façon continue entre deux bornes extrêmes et que les variations de la variable de production sont corrélées avec celles du paramètre en question [Zabalza-Mezghani et al., 2001]. Cette catégorie est généralement la plus étudiée et inclut des paramètres très divers comme la



(a) Paramètre de type continus

(b) Paramètre de type discret

(c) Paramètre de type stochastique

FIG. 1.2 – Classification des paramètres incertains.

moyenne d'une propriété pétrophysique, la position des contacts fluides ou les propriétés des fluides du réservoir (viscosité, densité etc.) ;

- *Discrets* : les paramètres discrets sont ceux qui ne peuvent prendre qu'un nombre fini de valeurs. Des exemples de paramètres incertains discrets sont : l'utilisation de différents algorithmes pour modéliser un phénomène donné, différents scénarios de dépôts des sédiments (environnement deltaïque, fluviatile, etc.), différents modèles structuraux ou bien le comportement binaire d'une faille (ouverte/fermée), etc. ;
- *Stochastiques* : un paramètre appartient à cette catégorie (1) quand il n'y a pas de relation directe entre les variations du paramètre et celles du résultat de simulation et (2) que le paramètre en question peut prendre un nombre infini de valeurs possibles. Le germe géostatistique considéré lors de la simulation d'un faciès ou d'une propriété pétrophysique est un exemple de paramètre stochastique. Un léger changement dans ce germe peut conduire à des modèles pourtant équiprobables mais dont les comportements lors des simulations d'écoulement sont significativement différents.

1.3.2 Objectifs de la caractérisation et de la quantification des incertitudes

La caractérisation et la quantification des incertitudes est un exercice difficile qui demande de faire face à sa propre ignorance [Massonnat, 2000; Caumon et al., 2004]. En particulier, il est nécessaire de bien identifier et exprimer de façon claire les hypothèses et choix subjectifs sous-

jacents à la modélisation [Dubrule, 1998]. En outre, la caractérisation d'un gisement pétrolier répond à un problème économique : elle doit pouvoir servir de base à la prise de décisions telles que la mise en production ou non d'un gisement, le forage de nouveaux puits ou bien le changement de mode de production, etc. Il faut donc disposer d'outils et de modèles spécifiques pour : (1) quantifier l'incertitude sur la production, c'est-à-dire identifier pour une variable cible (huile en place, production cumulée, etc.) sa valeur la plus probable mais aussi son intervalle de variation, (2) évaluer l'influence des paramètres incertains ainsi que leurs interactions afin de pourvoir définir de nouvelles analyses et acquisitions de données pour ceux jugés importants et fixer ceux dont l'effet est négligeable, et (3) mesurer comment l'apport de nouvelles informations permet de réduire globalement les incertitudes, par exemple grâce au formalisme bayésien.

1.3.3 Méthodes pour la caractérisation des incertitudes liées aux simulations d'écoulement

La première étape de toute méthode de quantification des incertitudes consiste tout d'abord à sélectionner les principaux paramètres du réservoir (perméabilité de la roche, transmissibilité des failles, viscosité des fluides, etc.) ayant un impact significatif sur la production. Cette sélection s'effectue à l'aide d'études d'analogues, du jugement d'experts de la zone étudiée ou de résultats d'études précédentes. Puis, pour chacun des paramètres, les différentes valeurs qu'il peut prendre ainsi que leurs probabilités sont définies. Le plus souvent, cela se traduit par une loi de probabilité qui, théoriquement, devrait être multivariée pour tenir compte de corrélations entre les paramètres incertains. Cependant, par souci de simplicité et parce qu'il est difficile d'appréhender ces corrélations, les différents paramètres sont souvent supposés indépendants. Ensuite, idéalement, toutes les combinaisons possibles devraient être considérées et simulées. Cela permettrait d'évaluer précisément la relation existant entre les paramètres incertains du modèle et les capacités de stockage ou de production du réservoir. C'est envisageable si le nombre de paramètres incertains reste modeste et surtout si le volume de calculs et le temps nécessaire à la construction et l'évaluation d'un modèle sont faibles. Ainsi, par exemple, cette approche est au cœur de l'a méthodologie JACTA [Corre et al., 2000; Charles et al., 2001] qui concerne l'étude des incertitudes statiques sur la structure du réservoir, mais aussi la répartition des faciès et son remplissage pétrophysique.

Cependant, dans le cadre des simulations d'écoulement, dû au temps de calcul et à la puissance limitée des ordinateurs, seul un nombre limité de modèles peut être simulé. Cela conduit alors à un sous-échantillonnage de l'espace des paramètres incertains et donc à une mauvaise évaluation des incertitudes sur le réservoir étudié. Pour surmonter ces limites, différentes approches sont possibles. La première est basée sur des simulations conventionnelles d'écoulement, et consiste à effectuer un plan d'expérience puis à utiliser une ou plusieurs fonctions ou surfaces de réponse. Alternativement, une seconde approche consiste à négliger une partie du comportement physique des fluides dans le réservoir en utilisant un certain nombre d'hypothèses simplificatrices, l'objectif global étant de réduire de façon significative le temps de calcul nécessaire pour simuler un modèle donné.

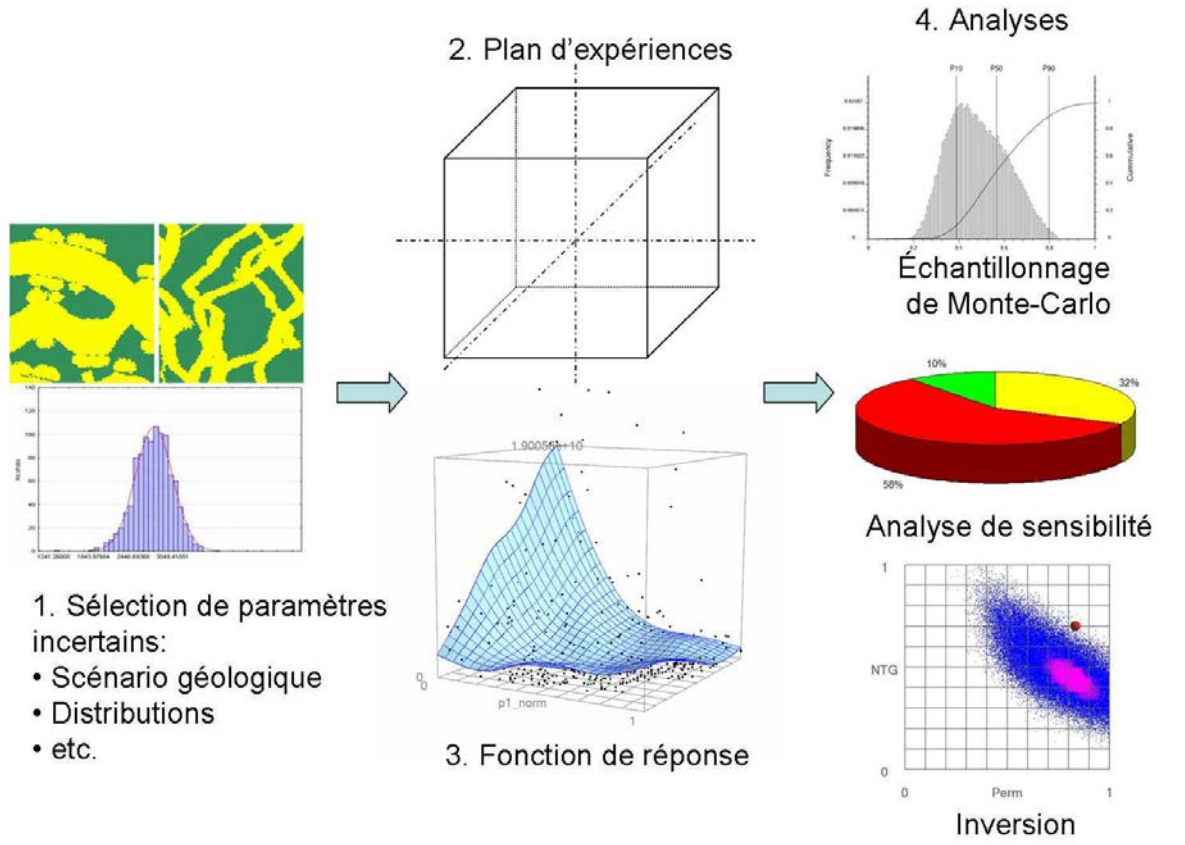


FIG. 1.3 – Approche combinant la planification d'expériences et les surfaces de réponse pour la quantification et l'analyse des incertitudes sur les résultats de simulation d'écoulement. Cette approche forme un cadre robuste statistiquement, cependant (1) le nombre de simulation est limité et (2) les plans d'expérience prennent difficilement en compte des paramètres corrélés et ceux de type stochastique et discret.

Approche par plans d'expérience et fonctions de réponse

La planification d'expérience (ou "*experimental design*") est une méthodologie qui cherche à sélectionner au mieux les combinaisons (ou niveaux) de paramètres incertains à simuler. Cette sélection est construite de telle manière qu'un maximum d'information peut être extrait d'un nombre minimal de simulations [Box et Draper, 1987; Damsleth et al., 1992]. Les premières applications de cette approche dans l'industrie pétrolière remontent à Chu [1990] et Damsleth et al. [1992]. Depuis, différents types de plans d'expérience sont disponibles et ont été publiés dans la littérature. On distingue notamment :

– les plans à deux niveaux pour lesquels chaque paramètre ne peut prendre que deux valeurs extrêmes. Traditionnellement, ce type de plan est exécuté afin de passer en revue les différents paramètres incertains. En effet, il permet de quantifier l'effet linéaire des

différents paramètres : il est alors possible d'éliminer ceux jugés négligeables et de réduire la dimension du problème ;

- les plans à trois niveaux pour lesquels les différents paramètres sont représentés par trois valeurs possibles. Ce type de plan est plus coûteux en nombre de simulation mais il permet d'identifier d'éventuels effets quadratiques sur la production [Box et Draper, 1987] ;
- les plans plus complexes tels que les hypercubes latins [Morris et Mitchell, 1995] ou les plans uniformes [Fang et al., 2000] qui permettent de mieux explorer l'espace des incertitudes en considérant un plus grand nombre de niveaux par paramètre au prix, cependant, de plus de simulations. Ces plans sont souvent construits à l'aide d'algorithmes d'optimisation (on parle alors de plan aidé par ordinateur) : les simulations à réaliser sont déterminées en fonction de critères de remplissage de l'espace des paramètres incertains [Liefvendahl et Stocki, 2004], des distributions a priori des paramètres [McKay et al., 1979] ou de données secondaires [Minasny et McBratney, 2006].

Le choix d'une planification s'effectue suivant le nombre de simulations qu'il est possible d'effectuer, mais aussi suivant les objectifs de l'étude. A ce sujet, une discussion et une comparaison des différents plans considérés dans l'industrie pétrolière peuvent être trouvées dans [Yeten et al., 2005].

D'autre part, les fonctions de réponses (encore appelées surfaces de réponses) sont définies comme étant des fonctions mathématiques multivariées approchant la relation, le plus souvent inconnue et non linéaire, qui relie les paramètres incertains et les résultats de simulation d'écoulement. Cela fournit un outil d'analyse rapide et simple à manipuler pour explorer l'espace des incertitudes. En particulier, au lieu d'effectuer une simulation coûteuse en temps il est préférable d'échantillonner la ou les surfaces correspondantes.

La combinaison de ces deux approches permet d'obtenir un cadre de travail statistiquement robuste pour quantifier les incertitudes liées aux écoulements tout en gardant les phénomènes physiques à modéliser aussi complexes qu'ils doivent l'être. Cette approche est désormais répandue dans l'industrie pétrolière et a déjà été employée pour l'analyse de la capacité de production d'un réservoir [Dejean et Blanc, 1999; Corre et al., 2000; Friedmann et al., 2001; Manceau et al., 2001; Castellini et al., 2004; Zabalza-Mezghani et al., 2004; **?**], le calage d'un historique de production [Landa et Guyaguler, 2003; Feraille et Roggero, 2004] ou l'optimisation du placement de puits [Guyaguler et Horne, 2001].

Approche par classification de modèles

L'approche par classification de modèles a été la première fois suggérée comme une alternative efficace aux simulations d'écoulement dans Ballin et al. [1993]. Elle consiste à utiliser un simulateur d'écoulement rapide comme substitut aux simulateurs conventionnels, précis mais coûteux en temps. Elle suppose de définir un ou plusieurs critères de classification dont le comportement est similaire aux vraies simulations d'écoulement et qui identifient les mêmes modèles médians et extrêmes (P_{10}, P_{50} et P_{90}) [Deutsch, 1996]. De nombreux attributs de classification ont été proposés dans la littérature : Alabert et Modot [1992]; Hird et Dubrule [1995] ainsi que

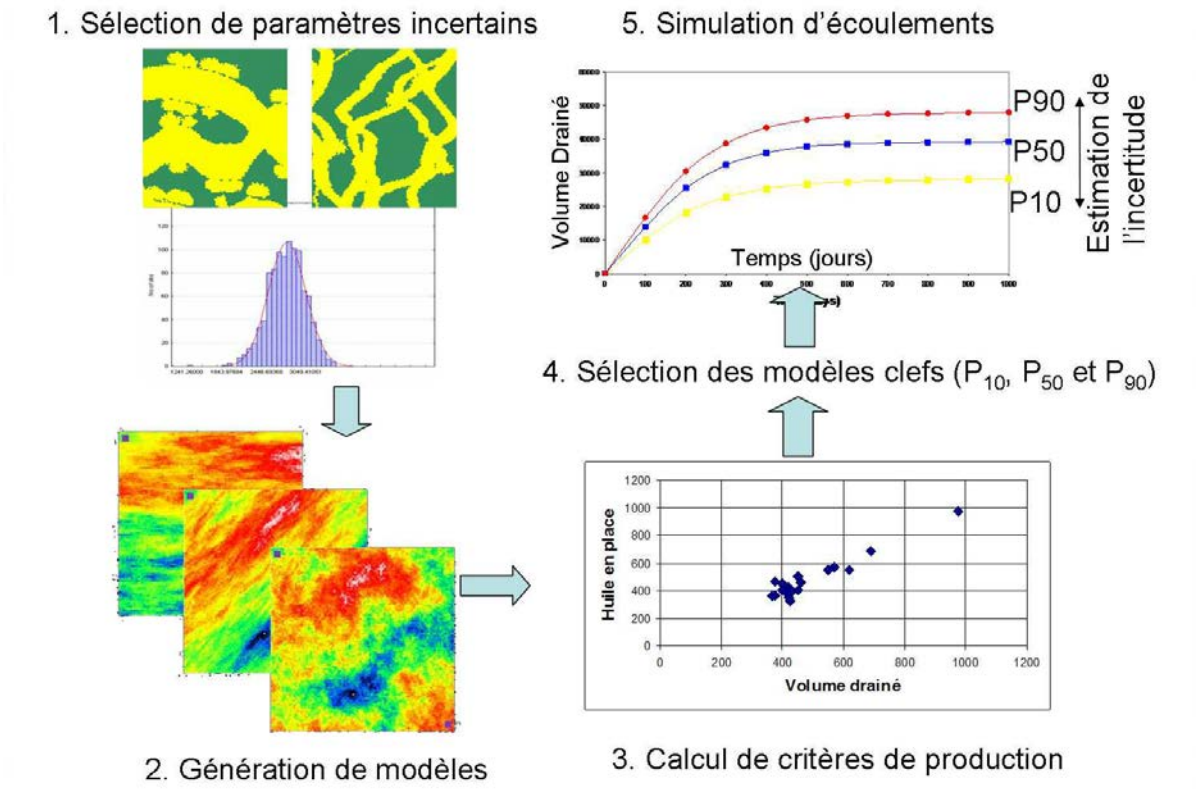


FIG. 1.4 – Approche par classification de modèles pour la quantification et l'analyse des incertitudes sur les résultats de simulation d'écoulement. L'avantage de ce type d'approche est l'étape de génération de modèles qui permet de tenir compte de corrélations entre les paramètres mais aussi d'inclure sans difficulté les paramètres de type discrets et stochastiques. Cependant, le résultat final consiste uniquement en trois modèles empêchant toute analyse statistique.

Deutsch [1996] ont utilisé différentes mesures basées sur la perméabilité du milieu poreux en supposant que la connectivité des corps les plus perméables devaient diriger l'écoulement des fluides dans le réservoir ; Ballin et al. [1993], puis Saad et al. [1996] ont préféré effectuer des simulations de traceurs chimiques en milieu monophasique, Idrobo et al. [2000] ont proposé le concept de volume drainé basé sur le temps de vol tandis que Peddibhotla et al. [1996]; Brandsaeter et al. [2001]; Gilman et al. [2002] et Ates [2005] ont utilisé les résultats de simulations sur lignes de courant.

Cette approche basée sur la classification et la sélection de modèles géologiques à l'aide de simulations alternatives a déjà été appliquée avec succès à la prédiction des incertitudes liées à la production d'hydrocarbures [Ates et al., 2003] et au problème de calage d'historique de production [Wang et Kovseck, 2003]. Cependant, une telle approche suppose une relation forte entre le ou les critères sélectionnés et la production, mais aussi consiste à contourner et négliger une partie du comportement physique des fluides. L'idée générale est de réduire le temps nécessaire pour simuler un modèle donné et ainsi pouvoir en traiter un plus grand nombre. Mais

cela ne permet que de diriger l'interprétation générale de l'écoulement dans le réservoir et non de le quantifier précisément.

1.4 Objectif et organisation de la thèse

Cette thèse se place dans le cadre de la quantification des incertitudes liées aux simulations d'écoulement à l'aide de la planification d'expérience et de surfaces ou fonctions de réponse. Plus particulièrement, le premier objectif de cette thèse est l'étude et l'évaluation de la construction de surfaces de réponse non linéaires à l'aide de l'interpolateur lisse discret et du krigeage dual. Un second objectif est la sélection et le développement d'outils d'analyse de telles surfaces. Ce manuscrit est organisé de la façon suivante :

- Après une revue bibliographique des principales méthodes utilisées pour la construction de fonctions de réponses dans l'industrie pétrolière, le **chapitre 2** décrit, dans un premier temps, deux approches développés dans ces travaux : l'interpolation lisse discrète et le krigeage dual. Dans un second temps les méthodes implantées pour pouvoir analyser ce type de fonction de réponse sont présentées : il s'agit de l'analyse de sensibilité basée sur la variance et de la technique d'inversion bayésienne. Pour finir, l'applicabilité de ces méthodes est étudiée sur plusieurs exemples synthétiques.
- Le **chapitre 3** est consacré à la prise en compte de données secondaires, telles que les résultats de simulations d'écoulement alternatives. La technique proposée combine les résultats de simulations d'écoulement conventionnelles, généralement coûteuses en temps, avec une surface de réponse secondaire issues de simulations rapides alternatives. Cette surface de réponse alternative permet de contraindre l'interpolation de celle des données primaire là où il n'y a pas de données. Il est ainsi possible de rendre en compte de courbures et non linéarités dans la réponse qui sont capturées par les simulations alternatives mais ne le sont pas par une planification d'expérience conventionnelle à deux ou trois niveaux.
- Puis, le **chapitre 4** a pour objet d'étendre les techniques présentées au chapitre (2) à la prise en compte de paramètres incertains de type stochastiques qui se caractérisent par un effet purement aléatoire sur la production. Cette approche s'appuie en particulier sur une modélisation séparée de la moyenne et de la dispersion des valeurs de production à l'aide de surfaces de réponse. Plusieurs exemples sont utilisés pour exposer différentes applications possibles de cette approche.
- Enfin, le **chapitre 5** présente une méthode de lissage d'histogramme dans un espace mutlidimensionnel. Cette méthode est basée sur l'interpolation lisse discrète et les structures de données développées durant ces travaux. Elle illustre l'application des techniques décrites dans les chapitres précédents en dehors du domaine des surfaces de réponse.

L'ensemble de ces chapitres décrivent les travaux effectués durant cette thèse. D'autres travaux sont exposés ou détaillés en annexe : l'**annexe A** présente le simulateur d'écoulement biphasique sur lignes de courant implanté durant cette thèse ; tandis que l'**annexe B** détaille

les structures informatiques implantées pour représenter et manipuler un ensemble de points de données et une surface de réponse en dimension N.

Chapitre 2

Construction et analyse de surface de réponse non linéaire

Sommaire

2.1 Nature des données disponibles et notations 16
2.1.1 Données disponibles 16
2.1.2 Notations 17
2.2 État de l'art : méthodes pour la construction d'une surface de réponse 18
2.3 Krigeage dual avec terme de régularisation 20
2.3.1 Principe 20
2.3.2 Formalisme dual 22
2.3.3 Terme de régularisation 23
2.3.4 Remarques finales 24
2.4 Interpolation lisse discrète 25
2.4.1 Méthodologie 25
2.4.2 Structures de données 28
2.4.3 Contraintes DSI pour la construction de surfaces de réponse 29
2.4.4 Remarques finales 32
2.5 Méthodes d'analyse d'une surface de réponse 33
2.5.1 Quantification des incertitudes sur les prédictions de production 33
2.5.2 Analyse de sensibilité aux paramètres incertains 34
2.5.3 Inversion bayésienne de données de production 41
2.6 Cas d'études 45
2.6.1 Modèle Anticlinal avec 4 paramètres incertains 46
2.6.2 Mesures de prédictivité des méthodes de construction sur le modèle "*IC Fault*" 55
2.6.3 Évaluation de la performance des méthodes de construction 58
2.7 Conclusions 58

Une *surface de réponse*, encore appelée *fonction de réponse* ou *proxy* est une fonction mathématique représentant le comportement d'un système physique réel ou simulé. L'objectif est

d'approcher la relation, souvent inconnue et non-linéaire, qui existe entre un ensemble de paramètres du système avec la ou les variables de sortie. Une telle fonction est particulièrement intéressante dans le cas où l'évaluation directe du système est impossible ou trop coûteuse à simuler.

Ce chapitre se focalise sur la construction et l'analyse de surfaces de réponse de forme complexe, voire non-linéaire, avec de multiples paramètres d'entrée dans le cadre de la simulation d'écoulement dans un réservoir à hydrocarbures. Dans ce cas, le système physique est le simulateur d'écoulement, les variables d'entrée sont les paramètres incertains du modèle géologique (par exemple la position des failles et des horizons, les contacts fluides, les propriétés physiques des fluides, etc.) tandis que les variables de sortie sont les résultats de simulation d'écoulement (par exemple les débits de production et d'injection aux différents puits, la production cumulée, la pression dans le réservoir, etc.). À cette fin, nous avons développé et implanté un ensemble d'outils sous forme de plusieurs librairies informatiques dont l'objectif est de permettre la construction et l'analyse de surfaces de réponse. En outre, ces outils serviront de base aux méthodologies décrites dans les chapitres suivants. Les travaux présentés dans ce chapitre consistent principalement en l'implantation et l'évaluation de deux approches de construction de surface de réponse : la première est basée sur la technique du krigeage tandis que la seconde est une extension de l'interpolateur lisse discret (ou DSI pour *"Discrete Smooth Interpolation"*, [Mallet, 1992]). Ces méthodes sont accompagnées d'outils d'analyse adaptés au caractère non-linéaire du problème.

La première section de ce chapitre présente les données disponibles et les problématiques qui y sont associées ainsi que les différentes notations utilisées. Ensuite, après une revue bibliographique des méthodes les plus répandues en ingénierie de réservoir, nous exposerons les deux approches proposées pour la construction de surfaces de réponse puis les méthodes sélectionnées pour leur analyse. Enfin, l'efficacité de ces techniques sera évaluée sur plusieurs exemples synthétiques de réservoir à hydrocarbures.

2.1 Nature des données disponibles et notations

2.1.1 Données disponibles

Les données disponibles pour la construction de surfaces ou fonctions de réponse sont généralement issues de la théorie des plans d'expérience. Dans le détail, il s'agit :

1. des paramètres incertains jugés avoir un impact significatif sur les résultats de simulation d'écoulement. Chaque paramètre est, de manière générale, associé à une loi de probabilité qui résume l'information disponible avant toute simulation et analyse. Le plus souvent, cette loi est uniforme entre les valeurs minimales et maximales du paramètre mais elle peut prendre une forme plus complexe (triangulaire, normale, etc.) voire non-paramétrique, comme, par exemple, celle d'un histogramme ;

2. des modèles géologiques pour lesquels une simulation d'écoulement est effectuée. Ces modèles se caractérisent par différentes valeurs (encore appelées niveaux) de paramètres incer-

tains. Le choix de ces modèles se base à la fois sur l'expertise des ingénieurs réservoir et sur une planification d'expériences ayant pour objectif d'extraire le maximum d'information à partir du nombre nécessairement limité de simulations.

3. des résultats de simulation d'écoulement. Ceux-ci sont résumés sous forme de plusieurs variables suivant leur nature (débit, pression, etc.) et le temps auquel ils sont mesurés ou simulés. Idéalement, du fait de corrélations existant entre les différentes variables, les surfaces de réponse correspondantes devraient être construites conjointement. Cependant, en l'absence d'informations spécifiques et par soucis de simplicité elles sont traités de manière indépendante.

4. d'autres données plus atypiques peuvent être disponibles, par exemple, le gradient des résultats de simulation par rapport aux paramètres incertains ou l'apport de simulations alternatives.

D'un point de vue pratique, les paramètres d'entrée définissent un espace multi-dimensionnel, appelé espace ou domaine d'incertitude, dans lequel un modèle géologique peut être représenté sous la forme d'un unique point. Ses coordonnées sont les valeurs des paramètres incertains correspondants tandis que les résultats de simulation sont stockés sous la forme d'une ou plusieurs propriétés. Dans ce cadre, les surfaces de réponse sont des outils pour interpoler, et éventuellement extrapoler, les résultats de simulation dans l'espace multi-dimensionnel des incertitudes à partir des points de données. Pour cela, de nombreuses méthodes ont été décrites dans la littérature. Ces dernières doivent néanmoins affronter les contraintes suivantes :

1. le nombre de paramètres incertains reste, même pour les projets les plus matures, relativement élevé (n > 10). Ceci est dû à la connaissance limitée de la réalité du sous-sol, mais aussi à la complexité des phénomènes physiques à simuler.

2. en raison du coût en temps et en volume de calculs d'une simulation d'écoulement le nombre de points de données (c'est-à-dire le nombre de modèles géologiques effectivement simulés) est nécessairement limité. Combiné avec le grand nombre de paramètres incertains possibles, cela conduit à un espace d'incertitude extrêmement vide ;

3. au cours du processus il est possible d'éliminer un ou plusieurs paramètres jugés peu significatifs. L'idée générale est de projeter les points dans un sous-espace pour augmenter leur densité. Cependant, cela amène un bruit dans les résultats de simulation qui gène la construction de la surface de réponse ;

4. dû à la complexité des phénomènes physiques ayant lieu dans un réservoir pétrolier, la relation à modéliser est le plus souvent non-linéaire.

2.1.2 Notations

Les notations ci-dessous sont utilisées dans la suite de ce manuscrit :
- n : nombre de paramètres incertains ;

- $X_i,\ i = 1, \cdots, n$: différents paramètres incertains ;
- $\mathcal{X} \in \mathbb{R}^n$: espace d'incertitude défini par les paramètres $X_i,\ i = 1, \cdots, n$;
- $\mathbf{x} = [x_1, \cdots x_i, \cdots x_n]^t$: point quelconque de l'espace $\mathcal{X}$;
- $z(\mathbf{x})$: surface de réponse à interpoler ou extrapoler dans $\mathcal{X}$. La plupart du temps cette fonction sera inconnue et $\hat{z}(\mathbf{x})$ indiquera une estimation de cette fonction ;
- n_d : nombre de points de données disponibles ;
- $\mathbf{p}_j = [\ p_{j,1}, \ldots, p_{j,n}\ ]$ avec $j = 1, \cdots, n_d$: position des points dans l'espace $\mathcal{X}$ où la valeur exacte de la réponse $z(\mathbf{p}_j)$ est connue.

En outre, dans le cas où plusieurs variables de production doivent être prise en compte simultanément, on notera :

- n_z, leur nombre ;
- $\mathbf{z}(\mathbf{x}) = [z^{(1)}(\mathbf{x}), \ldots, z^{(n_z)}(\mathbf{x})]^t$, le vecteur colonne rassemblant leurs valeurs.

2.2 État de l'art : méthodes pour la construction d'une surface de réponse

Le problème de l'interpolation ou l'extrapolation d'une fonction dans un espace multidimensionnel à partir d'un ensemble de points est un problème récurrent de modélisation auquel différentes solutions ont déjà été apportées. Cette section introduit les principales méthodes employées en ingénierie de réservoir pour la construction de surfaces de réponse. Cette revue s'appuie en particulier sur les études comparatives publiés dans Yeten et al. [2005] et NIST/SEMATECH [2006]. Après une courte description des approches considérées, nous nous sommes intéressés à dégager dans quelle mesure elles permettent de répondre aux quatre problématiques citées ci-dessus (section 2.1.1).

La **régression linéaire multiple** est la méthode la plus fréquemment envisagée dans la littérature pour construire une surface de réponse. Le principe est d'ajuster une fonction composée de sous-fonctions simples, telles que des polynômes, à un ensemble d'observations. Le calcul des coefficients de régression, encore appelé ajustement du modèle de régression, est effectué au sens des moindres carrés, c'est-à-dire qu'il minimise la somme des résidus au carré entre les valeurs observées, $z(\mathbf{p}_i)$, et les valeurs estimées, $\hat{z}(\mathbf{p}_i)$ avec $i = 1, \cdots, n_d$. D'un point de vue pratique, le coût en calcul d'une régression linéaire ainsi que l'estimation de la fonction en un point donné est négligeable, ce qui permet d'appliquer cette technique à des problèmes de grande dimension. D'autre part, cet estimateur peut prendre en compte au sens des moindres carrés différentes données redondantes et/ou contradictoires. Cependant, cette méthode est aussi connue pour être sensible aux points absurdes (encore appelés "*outliers*") et au bruit lorsque les données sont éparses et irrégulières. En outre, elle ne permet pas de modéliser des comportements complexes et non-linéaires. L'autre défaut majeur est qu'elle repose sur la connaissance a priori du modèle de régression ce qui est rarement le cas. Ainsi, le plus souvent un polynôme de degré deux est choisi [Yeten et al., 2005; Zabalza-Mezghani et al., 2004].

Le **krigeage** [Matheron, 1965; Goovaerts, 1997] est une technique d'interpolation qui a pour objectifs de respecter strictement les points de données et de tenir compte de corrélations spatiales par le biais d'un modèle de variogramme. Cette technique a été initialement développée dans le cadre des Géosciences pour l'estimation de propriétés du sous-sol telles que la porosité, la perméabilité ou la teneur d'un minerais, etc. Depuis peu, elle est aussi envisagée en ingénierie de réservoir pour la construction de surface de réponse [Pan et Horne, 1998; Guyaguler, 2002; Landa et Guyaguler, 2003; Castellini et al., 2004; Scheidt et Zabalza-Mezghani, 2004]. En effet, cette technique est connue pour sa robustesse et sa capacité à modéliser avec précision des comportements non-linéaires. Différentes variantes du krigeage existent suivant les hypothèses faites sur l'espérance de la fonction à estimer ou sur la présence de données secondaires. En outre, différents modèles de variogramme peuvent être considérés suivant leur comportement à l'origine (linéaire, quadratique, etc.) et à l'infini (c'est-à-dire avec ou sans palier). Les avantages du krigeage pour la construction de surfaces de réponse sont :

- d'être facilement extensible à des problèmes de grande dimension ($n > 3$) ;
- de ne pas dépendre d'un modèle de régression pré-défini et ainsi de pouvoir modéliser des phénomènes complexes ;
- d'être peu coûteux en volume de calculs, en effet, si l'on suppose un voisinage unique, la matrice de covariance entre les points de données ne doit être calculée et inversée qu'une seule fois. De plus, si l'on considère le formalisme du krigeage dual [Galli et Murillo, 1984; Royer et Vierra, 1984; Goovaerts, 1997] il n'est pas nécessaire d'obtenir l'inverse de la matrice et l'estimation de la surface de réponse en un point de l'espace se réduit à une multiplication vecteur-vecteur, dont la taille dépend du nombre de points de données ;
- d'être un estimateur exact, c'est-à-dire que les valeurs connues aux points de données sont strictement respectées ;
- et enfin, de fournir comme résultat une fonction analytique.

La principale limitation à l'utilisation du krigeage est le choix du variogramme représentant la covaraince spatiale entre les points de données. Ce choix doit se baser sur les données disponibles et constitue un choix de modélisation. Or, dû à la grande dimensionalité du problème et au faible nombre de données, ce choix est difficile et sous-contraint. Une solution intéressante a été proposée dans Feraille et Roggero [2004] : elle consiste, à partir d'un modèle donné (gaussien, exponentiel, puissance, etc.), à optimiser ses paramètres de manière à maximiser la prédictivité de la surface de réponse. Cependant cette approche ne résous pas le choix du type de modèle et reste coûteuse en temps.

La technique des **splines de type plaque mince** provient de l'analogie physique entre les fonctions splines et la flexion d'une plaque mince de métal forcée à passer par un ensemble de points de contrôle. Essenfeld [1969] souligne les propriétés et caractéristiques de cette technique favorable à son utilisation en ingénierie de réservoir. Cette approche initialement développée en deux dimensions (fonctions splines cubiques) peut être facilement extensible à des problèmes de dimension quelconque. Grâce à la possibilité de calculer les dérivées de la fonction interpolée, cette technique est principalement employée pour résoudre des problèmes d'optimisation et de calage d'historique de production [Lee et al., 1986; Archer et al., 2005]. Plus récemment, **?**

proposent une amélioration de cette approche en utilisant un ensemble de splines emboîtées combiné avec une planification d'expérience évolutive pour la construction de surfaces de réponse. Il existe une claire analogie entre cette approche et le krigeage avec tendance linéaire (*"Kriging with Trend"*). La différence tient aux fonctions basales utilisées : un variogramme dans le cas du krigeage, et une fonction radiale imitant la flexion d'une plaque mince de métal dans le cas des fonctions splines. L'un des intérêts des splines est qu'il est possible d'introduire un terme de rigidité dans la plaque mince. De ce fait, la fonction interpolée ne passe plus nécessairement par les points de données et peut donc tenir compte de la présence de bruit dans les données.

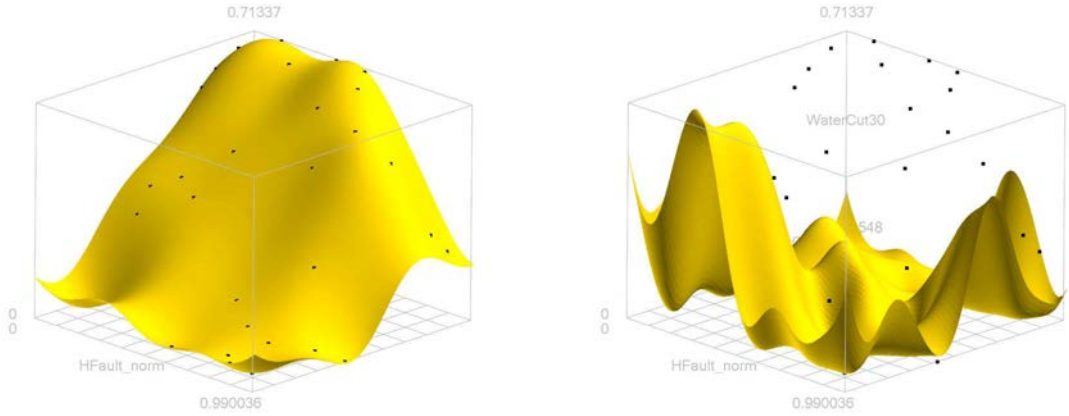
Enfin, une approche de plus en plus populaire consiste à utiliser des **réseaux de neurones artificiels** (ou ANN pour *"Artificial Neural Network"*), comme par exemple Guyaguler [2002] dans le cadre de l'optimisation du placement de puits. Les réseaux de neurones sont une méthode d'interpolation non-linéaire qui imite le comportement d'un système nerveux biologique. Les ANNs sont composées de couches interconnectées et, elles-mêmes, divisées en sous-unités (les neurones). Chaque neurone effectue tout d'abord une somme pondérée de ses entrées, applique une fonction de transfert qui lui est propre et transmet le résultat à la couche suivante. L'intérêt des réseaux de neurones vient de leur capacité à apprendre sur un jeu de donnée test en optimisant leurs pondérateurs internes en fonction de l'erreur de prédiction. Ces réseaux ont la capacité de filtrer les données et donc de tenir compte de la présence de bruit dans les données. Enfin, en variant le nombre de neurones, le nombre de couches et le type des fonctions de transfert, il est théoriquement possible de modéliser tout type de surface de réponse. Cependant, non seulement il est difficile de déterminer la taille et la nature du réseau à utiliser, mais surtout de par leur faible nombre, les données disponibles ne sont pas suffisantes pour la phase d'apprentissage ce qui conduit à des erreurs de modélisation le plus souvent indétectables.

Dans le cadre de ces travaux de thèse, nous avons décidé de développer et d'implanter deux nouvelles approches. La première est issue des méthodes utilisées dans la littérature et combine la variante dual du krigeage avec le terme de régularisation associé aux splines de type plaque mince. L'objectif est de combiner les avantages des deux approches pour disposer d'un estimateur efficace et répondant aux différentes contraintes énoncées ci-dessus. La seconde approche est basée sur l'interpolateur lisse discret qui est l'un des piliers du géomodeleur Gocad. Cette technique est connue, en particulier, pour ses capacités d'intégration de données à la fois éparses, hétérogènes et bruitées. Ces deux approches sont décrites ci-dessous (sections 2.3 et 2.4).

2.3 Krigeage dual avec terme de régularisation

2.3.1 Principe

Forme générale pour une variable scalaire Comme décrit précédemment, le krigeage est une technique d'interpolation dont l'objectif est de tenir compte de corrélations spatiales entre données. Celles-ci sont modélisées par un corrélogramme (ou son pendant, le variogramme) qui mesure la covariance, notée $C(h)$, entre deux points en fonction de la distance h les séparant. Ici, par soucis de simplicité, le variogramme est supposé isotrope. D'un point de vue pratique,



(a) fonction estimée $\hat{z}(\mathbf{x})$ (b) variance d'estimation $\sigma_z^2(\mathbf{x})$

FIG. 2.1 – Exemple de construction par krigeage simple d'une fonction de réponse. Le krigeage est effectué en considérant un variogramme de type puissance avec $\omega = 1.75$. (a) fonction interpolée et (b) variance d'estimation. A noter, en particulier, la forme non-linéaire de la surface et le caractère exact du krigeage : la fonction interpolée passe par les points de données (en noir) où la variance d'estimation est alors nulle. Les données proviennent du modèle IC-Fault présenté dans la section (2.6.2).

le krigeage est un estimateur linéaire : l'estimation de la fonction de réponse en un point $\mathbf{x}$ est une combinaison linéaire des valeurs connues. La forme générale de l'estimateur est :

$$\hat{z}(\mathbf{x}) - m_z(\mathbf{x}) = \sum_{j=1}^{n_d} \lambda_j(\mathbf{x}) \ \big(z(\mathbf{p}_j) - m_z(\mathbf{p}_j)\big) \tag{2.1}$$

où les $\lambda_j(\mathbf{x})$ sont les poids du krigeage associés à la donnée $z(\mathbf{p}_j)$, tandis que $m_z(\mathbf{x})$ et $m_z(\mathbf{p}_j)$ sont l'espérance de z en, respectivement, $\mathbf{x}$ et $\mathbf{p}_j$. Différentes variantes du krigeage existent suivant le modèle choisi pour $m_z(\mathbf{x})$. La plus couramment utilisée pour la construction de surfaces de réponse est celle du krigeage simple ("*Simple Kriging*"). Dans ce cas, l'espérance m($\mathbf{x}$) est supposée connue, et le plus souvent, mais pas nécessairement, constante sur l'ensemble du domaine d'étude. L'estimateur s'écrit alors sous la forme matricielle suivante :

$$\hat{z}(\mathbf{x}) = \boldsymbol{\lambda}^t(\mathbf{x}) \ \mathbf{z}_p \tag{2.2}$$

où :
- $\boldsymbol{\lambda}(\mathbf{x}) = [\lambda_1(\mathbf{x}), \cdots, \lambda_{n_d}(\mathbf{x})]^t$ représente le vecteur colonne des poids du krigeage ;
- $\mathbf{z}_p = [z(\mathbf{p}_1), \cdots, z(\mathbf{p}_{n_d})]^t$ rassemble les valeurs connues de la fonction à interpoler.

Système du krigeage Le calcul des $\lambda_j(\mathbf{x})$ s'effectue en tenant compte de la corrélation entre les données, tout en s'assurant que l'estimateur est sans biais et que la variance des erreurs d'estimation est minimale. Dans le cas du krigeage simple, on montre que l'estimation de ces

poids s'obtient en résolvant le système linéaire suivant [Matheron, 1970] :

$$\mathbf{C}\ \boldsymbol{\lambda}(\mathbf{x})\ =\ \mathbf{c}(\mathbf{x}) \tag{2.3}$$

où $\mathbf{C}$ est la matrice de taille $n_d \times n_d$ des covariances spatiales entre les points de données, tandis que $\mathbf{c}(\mathbf{x})$ est le vecteur de taille n_d des covariances spatiales entre les points de données et le point où la fonction doit être interpolée :

$$\mathbf{C} = \begin{bmatrix} C(\|\mathbf{p}_1 - \mathbf{p}_1\|) & \cdots & C(\|\mathbf{p}_1 - \mathbf{p}_{n_d}\|) \\ \vdots & \vdots & \vdots \\ C(\|\mathbf{p}_{n_d} - \mathbf{p}_1\|) & \cdots & C(\|\mathbf{p}_{n_d} - \mathbf{p}_{n_d}\|) \end{bmatrix} \quad \text{et} \quad \mathbf{c}(\mathbf{x}) = \begin{bmatrix} C(\|\mathbf{p}_1 - \mathbf{x}\|) \\ \vdots \\ C(\|\mathbf{p}_{n_d} - \mathbf{x}\|) \end{bmatrix} \tag{2.4}$$

Sachant que la matrice $\mathbf{C}$ est symétrique, l'estimation de la fonction en un point quelconque de l'espace est alors égale à :

$$\hat{z}(\mathbf{x}) = \mathbf{c}^t(\mathbf{x})\ \mathbf{C}^{-1}\ \mathbf{z}_u \tag{2.5}$$

En complément, il est possible de calculer la variance $\sigma^2(\mathbf{x})$ et tout point telle que :

$$\sigma^2(\mathbf{x})\ =\ C(0)\ -\ \boldsymbol{\lambda}^t(\mathbf{x})\mathbf{c}(\mathbf{x})\ =\ C(0)\ -\ \mathbf{c}^t(\mathbf{x})\ \mathbf{C}^{-1}\ \mathbf{c}(\mathbf{x}) \tag{2.6}$$

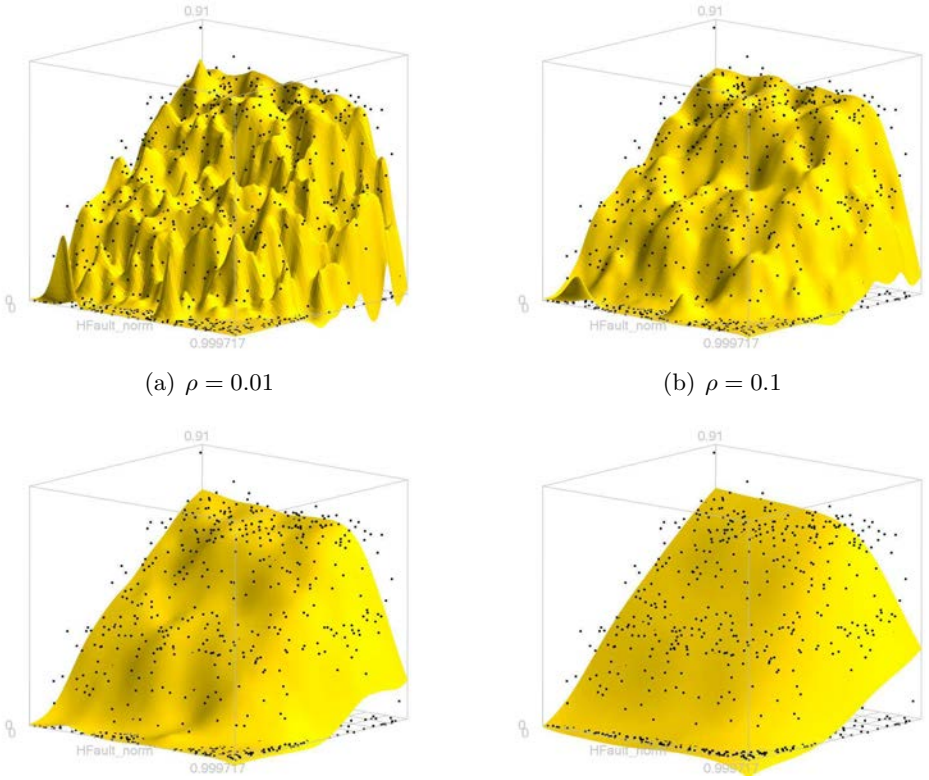
Cette valeur permet de mesurer la fiabilité de l'estimation et a été utilisée par exemple dans le cadre de plans d'expérience évolutifs [Scheidt et Zabalza-Mezghani, 2004]. La figure (2.1) montre un exemple simple d'interpolation par krigeage d'une fonction de réponse ainsi que la variance d'estimation correspondante.

Voisinage global Le calcul par la méthode du krigeage de $\hat{z}(\mathbf{x})$ et $\sigma^2(\mathbf{x})$ en un point quelconque $\mathbf{x}$ nécessite l'inversion de la matrice de covariance $\mathbf{C}$ dont la taille $n_d \times n_d$ dépend du nombre de points de données. Dans le cas général, il est nécessaire de limiter les calculs à un voisinage local autour du point où s'effectue l'interpolation. Les différents coefficients changent donc à chaque estimation ce qui conduit à une très grande répétition de calculs. Dans le cadre de la construction de surface de réponse, n_d reste relativement petit, il est alors possible de considérer un voisinage global, encore appelé voisinage unique, pour lequel la matrice $\mathbf{C}$ reste fixe. Il n'est donc nécessaire de l'inverser qu'une seule fois et l'estimation en tout point se composera alors d'une simple multiplication matrice-vecteur.

2.3.2 Formalisme dual

Le formalisme dual est une ré-écriture du système du krigeage particulièrement utile dans le cadre d'un voisinage global. L'estimateur est alors exprimé comme une combinaison des valeurs de covariances au lieu des valeurs connues de la fonction à estimer. L'idée principale est de constater que, dans le cas d'un voisinage global, le terme $\mathbf{C}^{-1}\ \mathbf{z}_u$ est un vecteur, fixe et indépendant de la position à estimer. On peut donc ré-écrire l'estimateur sous la forme suivante :

$$\hat{z}(\mathbf{x}) = \mathbf{c}^t(\mathbf{x})\ \mathbf{k} \quad \text{avec} \quad \mathbf{k} = \mathbf{C}^{-1}\ \mathbf{z}_u \tag{2.7}$$



(a) $\rho = 0.01$ (b) $\rho = 0.1$

(c) $\rho = 1.0$ (d) $\rho = 10.0$

FIG. 2.2 – Effet du terme de régularisation, ρ, sur la construction d'une surface de réponse à l'aide de la méthode du krigeage. Les données proviennent du modèle IC-Fault présenté dans la section (2.6.2).

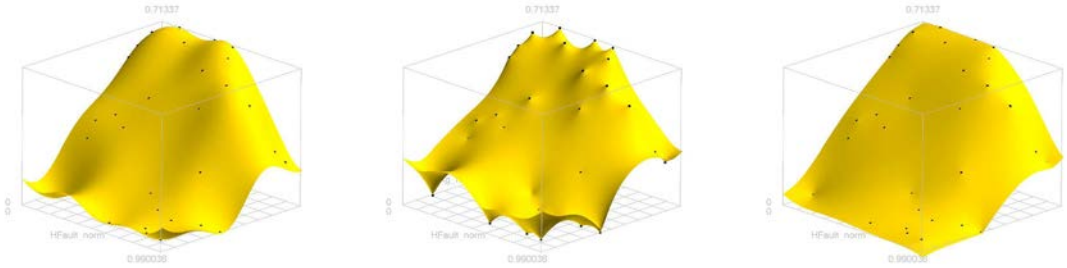
Ce formalisme, appelé **krigeage dual** permet donc de réduire le problème de l'estimation de la fonction par krigeage à une multiplication vecteur-vecteur de taille n_d. En outre, pour calculer $\mathbf{k}$, il n'est pas nécessaire de calculer explicitement $\mathbf{C}^{-1}$, il suffit de résoudre le système linéaire suivant :

$$\mathbf{C}\,\mathbf{k} = \mathbf{z}_u \tag{2.8}$$

Ceci peut être réalisé à l'aide d'un solveur numérique de système linéaire classique comme par exemple la bibliothèque de calcul LASPACK-1.12.2 [Skalicky, 1996] utilisée dans cette étude. Cela permet donc une accélération notable des calculs. Néanmoins, cette technique ne peut s'appliquer pour le calcul de la variance d'estimation. En effet, pour cela il est nécessaire soit de calculer explicitement $\mathbf{C}^{-1}$ soit de résoudre un système linéaire de la forme $\mathbf{C}\,\mathbf{b} = \mathbf{c}(\mathbf{x})$ à chaque estimation (équation (2.6)). Cependant, cette variance n'est pas nécessaire dans notre cas.

2.3.3 Terme de régularisation

Le krigeage est un estimateur exact : au niveaux des points de données, la valeur estimée est égale à la valeur réelle. Dans le cas de données bruitées ou redondantes, il est intéressant



(a) Variogramme gaussien (b) Variogramme exponentiel (c) Variogramme sphérique

FIG. 2.3 – Comparaison de la construction d'une surface de réponse par krigeage avec un variogramme de type gaussien (a) exponentiel (b) ou sphérique (c) mais de même portée. Les points noirs indiquent les points de données. Noter, en particulier, la forme non-linéaire de la surface, le respect strict des données et les changements dans la forme de la surface de réponse liés aux variogrammes utilisés. Les données proviennent du modèle IC-Fault présenté dans la section (2.6.2).

de relâcher l'interpolation en introduisant un terme de régularisation qui consiste à ajouter aux éléments diagonaux de la matrice de covariance un poids constant positif ρ. Les poids du krigeage sont alors la solution du système suivant :

$$\mathbf{K}\ \boldsymbol{\lambda}(\mathbf{x}) = \mathbf{c}(\mathbf{x}) \quad \text{avec} \quad \mathbf{K} = \mathbf{C} + \rho \cdot \mathbf{I}_{n_d} \tag{2.9}$$

où I_{n_d} est une matrice identité de dimension $(n_d \times n_d)$. Il est évident que si $\rho = 0$, alors $\mathbf{K} = \mathbf{C}$, et on retrouve le système du krigeage précédent, tandis que si $\rho \to +\infty$, on obtient :

$$\mathbf{K} \simeq \rho \cdot \mathbf{I}_r \quad \text{d'où} \quad \boldsymbol{\lambda}(\mathbf{x}) = \frac{1}{\rho}\ \mathbf{c}(\mathbf{x}) \quad \text{et} \quad \hat{z}(\mathbf{x}) - m_z\mathbf{x} \simeq 0 \simeq \mathbf{0} \tag{2.10}$$

L'utilisation d'un terme de régularisation, ρ, provient de l'approche "splines de type plaque mince" et représente une rigidité dans la plaque. Le résultat obtenu capture alors les tendances générales sans nécessairement passer par les points de données (figure 2.2) : il est donc possible par ce moyen de tenir compte de données bruitées. Cette approche est semblable sur certains points à l'introduction d'un effet de pépite dans le variogramme ou à l'utilisation du krigeage filtrant [Ma et Royer, 1988]. Une comparaison de ces approches permettrait de juger leurs avantages et inconvénients respectifs.

2.3.4 Remarques finales

En résumé, l'approche basée sur la forme duale du krigeage et l'introduction d'un terme de régularisation est adaptée à la construction de surfaces de réponse non-linéaires dans un espace de grande dimension et en présence de données bruitées. De plus, le résultat obtenu est une fonction analytique dont l'estimation en un point quelconque du domaine d'étude ne nécessite que le calcul d'un produit scalaire. Cependant, comme souligné précédemment, cette approche

reste limitée par le choix du variogamme expérimental qui a un impact majeur sur l'estimation de la surface de réponse (figure 2.3) et qui est largement sous-contraint au vu des données disponibles.

2.4 Interpolation lisse discrète

L'interpolation lisse discrète (ou DSI pour "*Discrete Smooth Interpolation*") est une technique d'interpolation proposée par Mallet [1992]. Cette technique, basée sur les moindres carrés, est adaptée à la modélisation de phénomènes non-linéaires et permet d'intégrer suivant un formalisme cohérent diverses données à la fois éparses, hétérogènes et bruitées. Il nous a donc paru judicieux d'évaluer son utilisation comme technique de construction de surfaces de réponse en reprenant et étendant les travaux d' Amiotte et al. [2002]. Cette section présente tout d'abord les principes de l'interpolation lisse discrète, puis les structures de données et contraintes spécifiquement mises en place et implantées dans le cadre de ces travaux de thèse.

2.4.1 Méthodologie

Cette section présente le principe général de l'interpolateur lisse discret. Cet exposé se limite aux informations nécessaires à la compréhension de ces travaux. Pour plus de détails, le lecteur est invité à consulter Mallet [1992, 1997] ou encore Mallet [2002] où sont développés les fondements théoriques de la méthode ainsi que Cognot [1996] et Muron et al. [2005] pour des indications sur les implantations, respectivement, locale et matricielle de cette méthode.

Principe général

L'interpolation lisse discrète a pour objectif d'interpoler une fonction scalaire ou vectorielle φ au niveau des nœuds α d'un modèle ou maillage discret Ω [1]. Ici, la fonction à interpoler est la surface de réponse, notée précédemment $z(\mathbf{x})$. Cette interpolation s'effectue en respectant au sens des moindres carrés un ensemble $\mathcal{C}$ de contraintes c. Le terme "*contrainte*" désigne toute combinaison ou relation linéaire imposée aux valeurs de la fonction à interpoler au niveau des nœuds α. Les relations possibles sont extrêmement variées, par exemple, il peut s'agir d'imposer le respect de certaines données, de borner la fonction à interpoler ou bien d'assurer la continuité du résultat, etc.

Suivant le formalisme employé dans le cadre de l'interpolation lisse discrète, l'ensemble des contraintes s'écrivent sous la forme algébrique suivante :

$$\{c \in \mathcal{C} \text{ est respectée}\} \Leftrightarrow \sum_{\alpha \in \Omega} A_c(\alpha).\varphi(\alpha) \bowtie b_c \tag{2.11}$$

ou bien sous la forme matricielle :

$$\{c \in \mathcal{C} \text{ est respectée}\} \Leftrightarrow \mathbf{A}_c^t.\boldsymbol{\varphi} \bowtie b_c \tag{2.12}$$

[1] Dans le suite, seules les contraintes pour l'interpolation des fonctions scalaires sont décrites. L'extension aux problèmes vectoriels est décrite en détail dans Mallet [2002]

avec $\mathbf{A}_c = [A_c(\alpha), \forall\alpha \in \Omega]^t$ et $\boldsymbol{\varphi} = [\varphi(\alpha), \forall\alpha \in \Omega]^t$. Cette formulation est appelée forme linéaire canonique des contraintes DSI. Dans cette équation, $A_c(\alpha), \forall\alpha \in \Omega$ et b_c sont des coefficients propres à chaque contrainte tandis que $\bowtie$ représente l'un des trois opérateurs suivants :

$$\bowtie \in \ \{\simeq, =, >\}. \tag{2.13}$$

Comme décrit dans Mallet [2002], cette approche permet de formaliser de façon unique une très grande variété de contraintes. En règle générale, on distingue trois grands types de contraintes :

- les contraintes dures $\{=, >\}$ ou "*hard constraints*" doivent être respectées impérativement. Celles-ci nécessitent le plus souvent un traitement post-processus spécifique.
- les contraintes souples $\{\simeq\}$ ou "*soft constraints*" sont respectées au sens des moindres carrés. Pour cela, dans un premier temps, le degré de violation individuel de chaque contrainte, noté $\rho(\varphi|c)$, est défini égal à :

  $$\rho(\varphi|c) = \left|\mathbf{A}_c^t.\boldsymbol{\varphi} - b_c\right|^2 \tag{2.14}$$

  puis, dans un second temps, ces termes sont sommés pour obtenir le degré global de violation des contraintes, $\rho(\varphi)$. Cette somme est pondérée par un ensemble de facteurs ω_c qui traduisent le niveau de confiance dans les différentes contraintes :

  $$\rho(\varphi) = \sum_{c\in\mathcal{C}} \omega_c.\rho(\varphi|c) \tag{2.15}$$

- la contrainte de rugosité sert à assurer la continuité de la fonction interpolée. L'objectif est double : cela permet de lisser le résultat et d'assurer l'existence d'une solution dans les parties du maillage discret ou aucune autre contrainte n'a été mise en place. En pratique, la rugosité est divisée en termes élémentaires, notés $R(\varphi|\alpha)$. Ceux-ci s'appliquent en chaque nœud α du domaine et à son voisinage, $N(\alpha)$, qui le plus souvent correspond aux nœuds connectés à α. Suivant le formalisme employé dans Mallet [2002] les rugosités élémentaires s'écrivent sous la forme :

  $$R(\varphi|\alpha) = \left[\sum_{\beta\in N(\alpha)} v(\alpha,\beta).\varphi(\beta)\right]^2 = |\mathbf{V}_\alpha \cdot \boldsymbol{\varphi}|^2 \tag{2.16}$$

  où, $v(\alpha, \beta)$ est une fonction de pondération des différents voisins de α. Pour finir, d'une façon similaire aux contraintes souples, les différentes contributions locales sont pondérées et sommées dans le terme de rugosité globale, noté $R(\varphi)$:

  $$R(\varphi) = \sum_{\alpha\in\Omega} \omega_\alpha.R(\varphi|\alpha) \tag{2.17}$$

  Bien que dans un souci de clarté, les notations soient différentes, il est important de noter que la contrainte de rugosité est semblable à un ensemble de contraintes souples avec $v(\alpha, \beta) \Leftrightarrow A_c(\alpha)$ et $b_c = 0$.

Dans la plupart des méthodologies et cas d'études présentés dans ce manuscrit, les contraintes considérées seront généralement de type souple. Cependant, dans la suite quelques exemples de contraintes d'inégalité seront présentés et détaillés.

Construction du système DSI

L'interpolation d'une fonction par l'approche DSI repose sur la minimisation des termes de violation des contraintes, $\rho(\varphi)$, et de rugosité globale, $R(\varphi)$. Afin de les minimiser simultanément, une rugosité générale $R^*(\varphi)$ est définie telle que :

$$R^*(\varphi) \;=\; R(\varphi) \;+\; (\phi \cdot \omega)\; \rho(\varphi) \tag{2.18}$$

où ϕ et ω sont deux facteurs d'ajustement strictement positifs. Le premier, ϕ, appelé "*fitting factor*", sert à spécifier l'importance des deux termes. Ainsi lorsque $\phi > 1$, les contraintes seront d'avantage respectées que la rugosité et inversement lorsque $\phi < 1$. Le second facteur, ω, appelé "*balancing ratio*", permet d'équilibrer les contributions respectives de $R(\varphi)$ et $\rho(\varphi)$ et ainsi d'adimensionner le facteur ϕ. L'algorithme DSI vise alors à minimiser $R^*(\varphi)$ c'est-à-dire à résoudre le système :

$$\frac{\partial R^*(\varphi)}{\partial \varphi} \;=\; 0 \tag{2.19}$$

Pour cela, on montre que cette rugosité générale peut s'écrire sous la forme d'une fonction objective quadratique [Mallet, 2002] dont le minimum est la solution recherchée :

$$\left| \begin{array}{l} R^*(\varphi) = \boldsymbol{\varphi}^t \cdot \mathbf{W}^* \cdot \boldsymbol{\varphi} \;-\; 2\; \boldsymbol{\varphi}^t \cdot \mathbf{Q}^* + (b^*)^2 \\ \\ \text{avec} \left\{ \begin{array}{l} \mathbf{W}^* = (\phi \cdot \omega) \cdot \displaystyle\sum_{c \in \mathcal{C}} \omega_c \cdot \mathbf{A}_c \cdot \mathbf{A}_c^t \;+\; \sum_{\alpha \in \Omega} \mathbf{V}_\alpha \cdot \mathbf{V}_\alpha^t \\ \mathbf{Q}^* = (\phi \cdot \omega) \cdot \displaystyle\sum_{c \in \mathcal{C}} \omega_c \cdot \mathbf{A}_c \cdot b_c \\ (b^*)^2 = (\phi \cdot \omega) \cdot \displaystyle\sum_{c \in \mathcal{C}} \omega_c \cdot b_c^2 \end{array} \right. \end{array} \right. \tag{2.20}$$

où $[W]$ est la matrice de rugosité globale de taille $(m \times m)$, m étant le nombre de nœuds du modèle discret. Si l'on considère les équations (2.19) et (2.20), il vient :

$$\frac{\partial R^*(\varphi)}{\partial \varphi} = 2 \cdot \mathbf{W}^* \cdot \varphi \;-\; 2 \cdot \mathbf{Q}^* \tag{2.21}$$

La recherche de la solution de l'interpolation lisse discrète consiste alors à résoudre le système linéaire de taille $(m \times m)$ suivant :

$$\mathbf{W}^* \cdot \boldsymbol{\varphi} = \mathbf{Q}^* \tag{2.22}$$

Méthodes de résolution

Plusieurs méthodes ont été envisagées et proposées pour la résolution de ce problème d'optimisation linéaire (équation 2.22). On distingue notamment :

– la *formulation locale* : cette approche consiste à parcourir itérativement l'ensemble des nœuds et à réactualiser les valeurs de φ en fonction des contraintes locales et de la rugosité. Pour cela, il est nécessaire de reformuler les contraintes telles que décrites dans Mallet [2002] ou Cognot [1996]. Cette procédure de type Gauss-Seidel permet d'obtenir une solution approchée du problème et ce pour un faible coût en terme de mémoire et puissance

de calcul. Cela explique le choix initial de cette formulation et son implantation au sein du géomodeleur Gocad. Cette approche souffre néanmoins d'un taux de convergence lent malgré différentes optimisations comme les étapes d'initialisation [Cognot, 1996] ou les algorithmes multi-grilles [Labat, 2004].

– la *formulation matricielle* : cette approche consiste en une résolution directe du système linéaire correspondant à l'équation (2.22). Cette solution a cependant été délaissée en raison de la puissance limitée des ordinateurs. Récemment, grâce à l'augmentation significative de cette puissance, une nouvelle implantation générique de cet algorithme a été proposée dans Muron et al. [2005]. La seule limitation est l'utilisation de contraintes de type dynamique dont les coefficients évoluent en fonction de la converge de l'algorithme impliquant alors une modification de la matrice $\mathbf{W}^*$ et donc un ralentissement des calculs. Des travaux récents s'attachent néanmoins à résoudre ce problème [Muron et al., 2005].

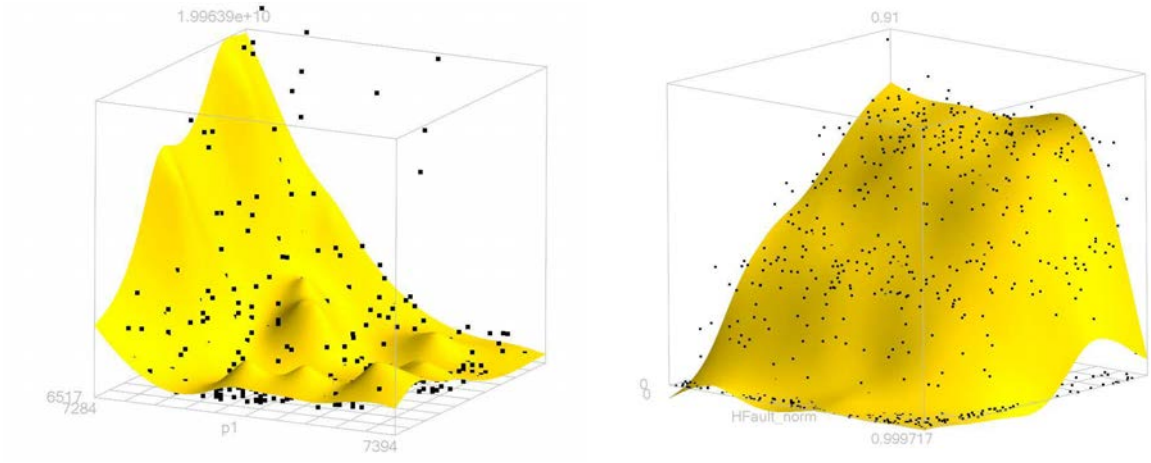
Dans le cadre de ce travail, la formulation matricielle globale a été choisie car elle converge plus rapidement et utilise directement la forme canonique des contraintes. Le développement de nouvelles applications et de nouvelles contraintes s'en trouve accéléré et plus robuste.

2.4.2 Structures de données

L'interpolation lisse discrète a initialement été développée en 3D pour la modélisation d'objets géologiques tels que des failles ou des horizons par le biais de surfaces triangulées. Son application dans le cadre de la construction de surface de réponse a nécessité le développement et l'implantation de structures de données spécifiques pouvant être plongées dans des espaces de dimension quelconque :

– la structure **NdPointSet** représente un ensemble de points dans un espace de dimension quelconque. Dans cette structure, chaque point est représenté par deux tableaux, le premier rassemble les coordonnées du point dans l'espace tandis que le second a une taille variable et sert à stocker les valeurs de propriété qui lui sont associées. Cela permet d'avoir dans un élément unique et compact en mémoire toutes les informations nécessaires. En outre, de façon globale, tous les points sont indépendants, ce qui facilite l'ajout, le retrait ou la modification de données.

– la structure **HyperCube** est une grille structurée régulière de dimension quelconque et est employée comme support de calculs lors de l'interpolation. Cette structure possède une géométrie entièrement implicite définie à partir (1) des points de la grille de coordonnées minimales et maximales et (2) du nombre de nœuds dans chaque direction de l'espace. En parallèle, le stockage des propriétés s'effectue sous la forme d'un tableau contigu en mémoire et dont la taille est égale à celle du nombre de nœuds de la grille. Les avantages de cette structure sont (1) son faible coût mémoire et (2) la forme compacte de stockage des propriétés qui permet un accès aisé aux différentes informations et des calculs rapides. Enfin, différentes fonctionnalités permettant de passer de la position dans le tableau de propriétés (index ijk) à la position dans la grille (index (i, j, k)) et aux coordonnées dans l'espace des paramètres incertains (coordonnées (x, y, z)) ont été implantées.



(a) (b)

FIG. 2.4 – Exemples de construction de surfaces de réponse à l'aide de l'interpolation lisse discrète (DSI). À noter dans les deux cas la forme complexe et non-linéaire de la surface interpolée ainsi que le filtrage du bruit dans les données grâce à l'approche basée sur les moindres carrés. Les données proviennent du modèle IC-Fault présenté dans la section (2.6.2).

Une description plus approfondie de ces structures et de leur implantation au sein du géomodeleur Gocad se trouve en annexe B. En résumé, les caractéristiques et contraintes principales de ces structures sont :

- leur implantation dans un espace de dimension quelconque ;
- le découplage entre la géométrie supposée fixe et les propriétés dont la quantité est variable ;
- leur l'utilisation comme supports de calculs : elles ont donc été implantées de façon à minimiser leur coût mémoire et à maximiser la rapidité d'accès aux diverses informations (coordonnées et propriétés) de chaque point ;
- leur implantation dans le géomodeleur Gocad : celle-ci est basée sur la structure **GObj** (pour "*geological object*"). Cette intégration a permis d'utiliser des fonctionnalités déjà mises en place au sein du logiciel : les outils graphiques, la gestion d'une librairie d'objets, la possibilité d'effectuer des sauvegardes sous forme de fichiers ascii ou binaire, la gestion générique de propriétés, etc.

2.4.3 Contraintes DSI pour la construction de surfaces de réponse

Pour l'interpolation d'une surface de réponse deux contraintes sont nécessaires : la contrainte floue de point de contrôle (ou "*fuzzy control point*") permettant d'imposer le respect lâche de points de données et la contrainte de rugosité assurant la continuité du résultat. Ces contraintes sont décrites dans la section suivante. Par la suite, d'autres contraintes comme celles de gradient ou de dérivée seconde qui ont été développées durant ces travaux seront détaillées le cas échéant.

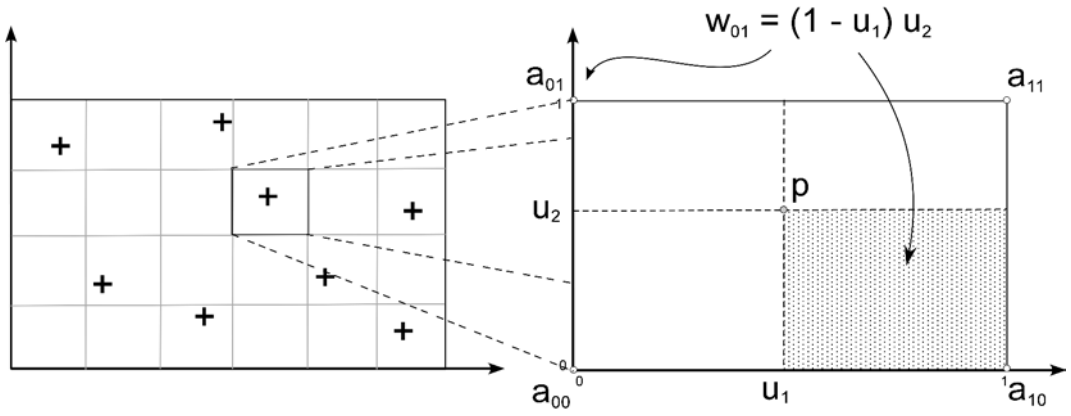


FIG. 2.5 – Interpolation barycentrique dans une cellule régulière en deux dimensions

Contrainte de "*fuzzy control point*"

Interpolation barycentrique Le modèle considéré comme support de la l'interpolation est une grille discrète structurée régulière. La fonction interpolée φ n'est donc connue qu'aux nœuds de la grille. Pour évaluer la valeur de la fonction en un point quelconque différents des nœuds de la grille, une interpolation barycentrique est envisagée. Tout d'abord, considérons une cellule de dimension quelconque (ou n-cellule) et notons les coordonnées de ses sommets $\mathbf{a}_{i_1 \ldots i_n}$ avec $i_1 \ldots i_n$ valant 0 ou 1 (figure (2.5)). L'interpolation barycentrique de φ en un point quelconque $\mathbf{x}$ situé dans la n-cellule s'écrit alors :

$$\varphi(\mathbf{x}) = \sum_{i_1=0}^{1} \ldots \sum_{i_n=0}^{1} w_{\mathbf{a}_{i_1 \ldots i_n}}(\mathbf{x}) \cdot \varphi(\mathbf{a}_{i_1 \ldots i_n}) \tag{2.23}$$

Les termes $w_{\mathbf{a}_{i_1 \ldots i_n}}(\mathbf{x})$ sont les coefficients d'interpolation liés au sommet $\mathbf{a}_{i_1 \ldots i_n}$ définis égaux à :

$$w_{\mathbf{a}_{i_1 \ldots i_n}}(\mathbf{x}) = b_{i_1}(u(x_1)) \times \ldots \times b_{i_n}(u(x_n)) \tag{2.24}$$

où les $u(x_1), \ldots, u(x_n)$ représentent la transformation de la cellule considérée de l'espace "réel" vers un hypercube unité et où $b_i(u)$ est un coefficient égal à :

$$b_i(u) = \left| \begin{array}{rcl} 1-u & \text{si} & i = 0 \\ u & \text{si} & i = 1 \end{array} \right. \tag{2.25}$$

Cette interpolation barycentrique permet donc de relier la valeur d'une fonction en un point quelconque du domaine étudié aux valeurs connues exclusivement aux nœuds. Cette procédure peut être envisagée dans deux cas : tout d'abord lors de l'évaluation des valeurs de réponse durant des analyses mais aussi lors de l'interpolation des surfaces comme base aux contraintes DSI pour imposer le respect de points de données.

Expression des coefficients La contrainte de " *fuzzy control point*" cherche à imposer le respect des valeurs connues de la fonction au niveau des points de données lors de l'interpolation.

C'est une contrainte de type flou qui s'écrit suivant le formalisme DSI sous la forme :

$$\forall j \in [\![1, n_d]\!], \quad \hat{z}(\mathbf{p}_j) \simeq z(\mathbf{p}_j) \quad \Leftrightarrow \quad \sum_{\beta \in \Omega} A_c(\beta)\varphi(\beta) = b_c \tag{2.26}$$

avec comme coefficients de contrainte :

$$\left| \begin{array}{rcll} A_c(\beta) & = & w_\beta(\mathbf{p}_j) & \text{si } \beta \text{ est un sommet de la n-cellule contenant } \mathbf{p}_j \\ A_c(\beta) & = & 0 & \text{sinon} \\ b_c & = & z(\mathbf{p}_j) & \end{array} \right. \tag{2.27}$$

et où le terme $w_\beta(\mathbf{p}_j)$ correspond au coefficient d'interpolation barycentrique à la position $\mathbf{p}_j$ et relatif au nœud $\beta \in \Omega$ comme défini précédemment (équation (2.24)).

Contrainte de rugosité

L'objectif du terme de rugosité est d'obtenir une certaine régularité dans le résultat interpolé mais aussi d'assurer l'existence d'une solution aux nœuds où aucune contrainte n'a été mise en place. En pratique, la contrainte de rugosité la plus couramment répandue est la contrainte à *pondération harmonique* [Mallet, 2002]. L'idée générale consiste, pour un nœud, donné, à attribuer un poids identique à tous ses voisins. Cependant, cette contrainte est biaisée si les nœuds voisins ne sont pas équidistants et souffre d'effets de bord qui réduisent les capacités d'extrapolation de la méthode.

Afin de palier ces défauts, nous avons implanté sur la structure de donnée **HyperCube** une contrainte dite de *gradient constant*. Le but est d'imposer la continuée du gradient de la fonction interpolée d'un nœud à l'autre du maillage discret. Dans le cas d'une grille structurée régulière de dimension quelconque, il est tout d'abord nécessaire de traiter toutes les directions de façon indépendante en imposant la relation :

$$\left(\frac{\partial \varphi(\alpha)}{\partial x_i}\right)^- = \left(\frac{\partial \varphi(\alpha)}{\partial x_i}\right)^+, \ \forall i \in [\![1, n]\!] \tag{2.28}$$

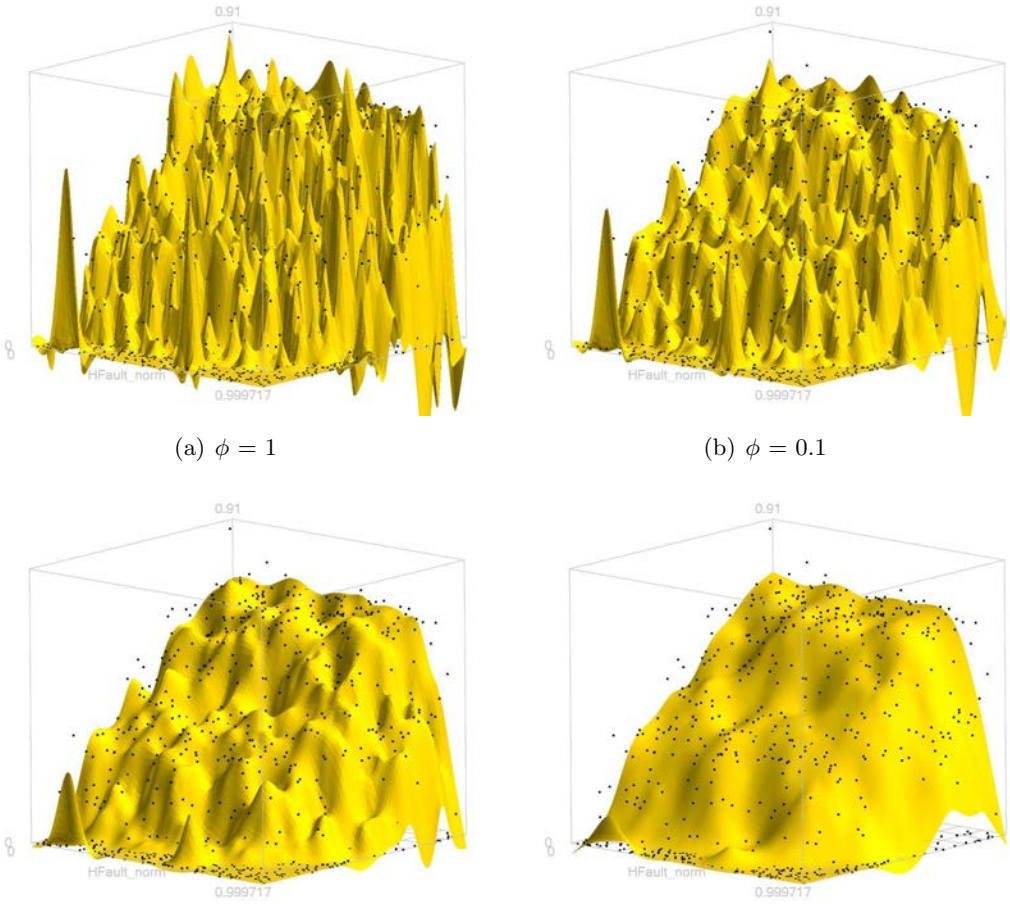
où les signes "$-$" et "$+$" indiquent respectivement les dérivées à gauche et à droite du nœud α dans la direction i considérée. Ces dérivées sont ensuite discrétisées à l'aide de schémas décentrés. Il vient alors :

$$\frac{1}{|\Delta_i^-|}\ \big(\varphi(\alpha) - \varphi(\alpha^-)\big) \ = \ \frac{1}{|\Delta_i^+|}\ \big(\varphi(\alpha^+) - \varphi(\alpha)\big)\,, \ \forall i \in [\![1, n]\!] \tag{2.29}$$

où, dans la direction i considérée, α^- et α^+ sont les nœuds voisins de α tandis que $|\Delta_i^-|$ et $|\Delta_i^+|$ représentent la taille des cellules du maillage discret. Ici ces deux valeurs sont identiques. Ainsi, suivant le formalisme de l'équation 2.16, les poids $v(\alpha, \beta)$ valent :

$$v(\alpha, \beta) \ = \ \left\{ \begin{array}{rll} -2 & \text{si} & \beta = \alpha \\ 1 & \text{si} & \beta = \alpha^+ \text{ ou } \alpha^- \end{array} \right. \tag{2.30}$$

Cette formulation est très voisine de la contrainte à pondération harmonique. La différence majeure concerne les bords du domaine. En effet, suivant l'approche proposée aucune contrainte



(a) $\phi = 1$ (b) $\phi = 0.1$

(c) $\phi = 0.01$ (d) $\phi = 0.001$

FIG. 2.6 – Effet du facteur d'ajustement ϕ sur la forme des surface de réponse : à $\phi = 1$, les données ont autant de poids que la rugosité et inversement à $\phi = 0.001$, la rugosité est le facteur prépondérant. Les données proviennent du modèle IC-Fault présenté dans la section (2.6.2).

de gradient constant n'est imposée dans la direction orthogonale au bord. A l'opposé, dans le cas d'une pondération harmonique, un gradient nul est implicitement imposé. En outre, la formulation décrite ci dessus est plus générale car elle permet de prendre en compte des mailles de taille différente.

2.4.4 Remarques finales

En conclusion, l'interpolation d'une fonction quelconque par la méthode DSI est semblable à la résolution d'un système de moindres carrés et se résume à un problème d'optimisation linéaire. L'objectif est de minimiser une fonction quadratique à deux termes : l'un représente un ensemble de contraintes que la fonction doit respecter (par exemple des données) et l'autre s'assure d'une

certaine régularité dans la solution. Les avantages attendus de cette approche dans le cadre de la construction de surfaces de réponse sont :

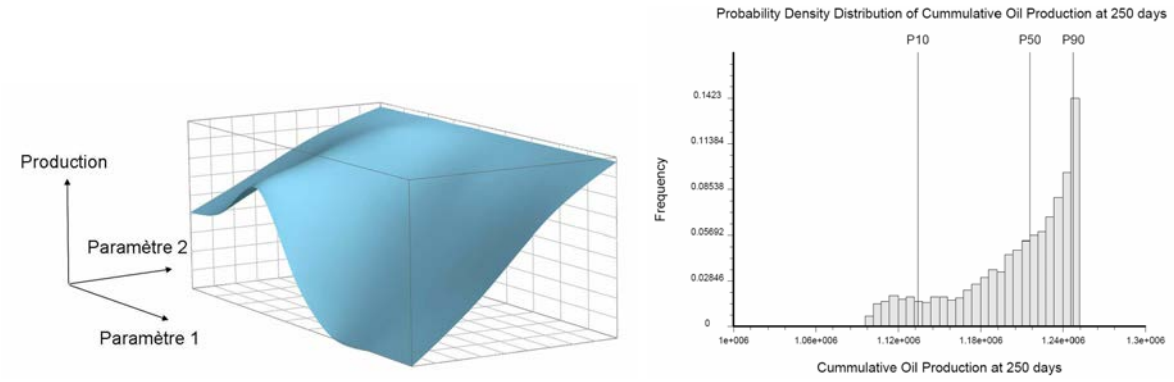
- l'extension facile de la méthode à des problèmes de dimension quelconque, en effet, à aucun moment il n'est nécessaire de le spécifier ;
- l'absence de tout modèle de régression pré-défini combinée avec l'utilisation d'un modèle sous-jacent discret permettant d'approcher des phénomènes non-linéaires ;
- l'utilisation de la formulation matricielle permettant d'accélérer la résolution et la convergence du système linéaire (équation 2.21) ;
- le formalisme mathématique unique combiné à un code informatique générique pour l'écriture des contraintes. Il permet de prendre en compte toute donnée ou information qui peut s'écrire sous forme d'une contrainte, c'est-à-dire sous la forme d'une relation linéaire entre les valeurs de la fonction à interpoler aux nœuds du domaine. Le développement de nouvelles contraintes et applications s'en trouve naturellement facilité.

2.5 Méthodes d'analyse d'une surface de réponse

Dans le cadre de l'analyse de simulations d'écoulement, les surfaces de réponse sont un outil mathématique approchant la relation entre un ou plusieurs paramètres incertains du réservoir et les variables de production. Leur utilisation à la place d'un simulateur d'écoulement permet d'analyser pour un coût de calcul négligeable le comportement du réservoir. Les approches développées durant ces travaux de thèse et décrites ci-dessous ont été sélectionnées pour leur capacité à prendre en compte des surfaces de réponse de forme complexe (en particulier non linéaire) et plongées dans un espace incertain de dimension quelconque. Plus particulièrement, nous nous sommes intéressés à (1) la *quantification des incertitudes* sur les prédictions de production, (2) *l'analyse de sensibilité* des variables de production vis-à-vis des paramètres incertains et enfin (3) *l'inversion bayésienne* de données historiques de production.

2.5.1 Quantification des incertitudes sur les prédictions de production

L'un des principaux intérêts des ingénieurs réservoir dans l'utilisation des surfaces de réponse est la possibilité de quantifier les incertitudes (et par là même le risque) pesant sur les prévisions de production. Pour cela, le domaine des paramètres incertains est échantillonné à l'aide de méthodes aléatoires de type Monte-Carlo tandis que la production est estimée à l'aide des surfaces de réponse. Les résultats permettent ensuite de construire un histogramme de chacune des variables de production (figure 2.7). Ces histogrammes rendent compte de l'incertitude pesant sur les résultats de simulation d'écoulement : en particulier, les percentiles de la réponse peuvent être aisément déterminés pour être considérés par la suite dans des modèles économiques. Ils servent alors à la prise de décision lors des opérations de gestion d'un réservoir pétrolier (mise en production, forage de nouveaux puits, etc.).



(a) (b)

FIG. 2.7 – Calcul de l'histogramme d'une variable de production par une méthode d'échantillonnage de type Monte-Carlo. (a) Exemple de surface de réponse correspondant à la production cumulée en huile sur 250 jours de la variante du modèle SPE 10 considéré au chapitre (3) et (b) histogramme correspondant à 10.000 échantillons.

Les paramètres incertains sont supposés le plus souvent indépendants et suivant une loi uniforme entre leurs valeurs minimales et maximales. L'échantillonnage revient alors à effectuer des tirages indépendants de nombres aléatoires pour chacune des dimensions de l'espace. Dans le cas où une plus grande précision dans l'estimation de l'histogramme est nécessaire, des méthodes de tirage pseudo-aléatoires peuvent être considérées. Ces méthodes ont pour objectif de répartir les points dans l'espace de telle façon que, pour un même nombre de points, l'erreur d'estimation diminue par rapport aux méthodes de calcul classique [Press et al., 2002]. Si les paramètres ne sont plus indépendants et suivent une loi de densité multivariée complexe, il est nécessaire d'utiliser des approches plus avancées telles que l'échantillonnage de Gibbs ou l'algorithme de Métropolis [Metropolis et Ulman, 1949] décrit en détail dans la section (2.5.3).

2.5.2 Analyse de sensibilité aux paramètres incertains

L'analyse de sensibilité est un élément clef de la quantification des incertitudes liées aux simulations d'écoulement. Cela consiste à mesurer l'influence relative des paramètres incertains (et de leur interaction) sur les variables de production [Cukier et al., 1973; Sobol', 1990; Saltelli et al., 2000]. Les objectifs de ce type d'analyse sont :

1. de définir de nouveaux axes de travail en vue de réduire l'incertitude sur les paramètres jugés influents ;
2. de mettre de côté les paramètres dont l'effet est peu important en fixant leur valeur et ainsi de réduire la dimension de l'espace d'incertitude ;
3. d'éliminer les interactions négligeables des plans d'expériences (procédure encore appelée factorisation).

Dans notre cas, l'objectif est d'évaluer la sensibilité (traduite sous la forme d'un indice) des variables de production par rapport aux paramètres incertains du modèle. Là encore, pour des raisons d'efficacité, au lieu d'utiliser directement le simulateur d'écoulement, l'analyse est basée sur les surfaces de réponse.

D'après Ravalico et al. [2005], les critères permettant de juger de l'efficacité des différentes méthodes de calcul d'indice de sensibilité sont : (1) le nombre de paramètres qu'il est pratiquement possible de prendre en compte, (2) la possibilité de capturer les interactions entre paramètres ; en effet, ceux-ci étant le plus souvent corrélés, leur impact sur la production peut être plus important que s'il sont considérés seuls et, enfin, (3) la possibilité de traiter des phénomènes complexes non linéaires et non monotones. En outre, Ravalico et al. [2005] précisent aussi qu'il serait désirable que les méthodes d'analyse de sensibilité ne nécessitent pas la connaissance de la distribution de chaque paramètre incertain.

Méthodes locales et globales

Deux types d'approches peuvent être distingués : les analyses de sensibilité *locales* (ou LSA pour "*Local Sensitivity Analysis*") et *globales* (ou GSA pour "*Global Sensitivity Analysis*"). Les méthodes de type LSA consistent à estimer localement les dérivées partielles des variables de production vis-à-vis des paramètres incertains et à les normaliser par l'espérance ou l'écart type des différentes variables [Homma et Saltelli, 1996]. Cette approche est clairement appropriée si seule une région spécifique de l'espace doit être étudiée, par exemple lors d'un processus d'optimisation [Cooke et Noortwijk, 1999]. Cependant, dans le contexte de la quantification des incertitudes, l'intervalle de variation des paramètres incertains est le plus souvent large et ne peut être couvert par une approche locale. À l'opposé, les approches globales, dont une bonne revue peut être trouvée dans Helton et al. [2005], sont plus ambitieuses et cherchent à estimer les variations de la production induites par un paramètre seul, ou un ensemble de paramètres, de façon "globale", c'est-à-dire sur l'ensemble du domaine d'incertitude. Parmi les approches possibles, la plus utilisée est celle basée sur les coefficients de régression standardisés ou de mesure de corrélations, combinée avec un test statistique de Student [Saltelli et al., 1999]. Cependant, cette approche repose sur une méthode de régression pour la construction de la surface de réponse et la connaissance a priori du modèle de régression. Elle est en outre limitée à des réponses quasi-linéaires et monotones et ne peut généralement pas mesurer des interactions au-delà de l'ordre deux. Les problèmes non linéaires peuvent être envisagés si les valeurs réelles des différentes entrées/sorties sont remplacées par leur rang [Saltelli et Sobol', 1995]. Néanmoins, cette transformation empêche toute analyse quantitative. Pour finir, les approches les plus prometteuses sont les analyses basées sur la variance (ou VBSA pour "*Variance Based Sensitivity Analysis*") qui estiment l'indice de sensibilité comme la contribution de chaque paramètre (et leurs interactions) à la variance de la variable de sortie.

Analyse de sensibilité basée sur la variance

Principe Le principe des méthodes de type VBSA consiste à décomposer la variance de la variable de réponse Z, en sous-termes reliés aux paramètres incertains. Considérons que ces paramètres $(X_1, ..., X_n)$ soient indépendants et continus, la variance $\mathrm{Var}(Z)$ se décompose alors de façon unique sous la forme :

$$\mathrm{Var}(Z) = \sum_{i=0}^{n} D_i + \sum_{i=0}^{n} \sum_{j=0}^{j<i} D_{ij} + \sum_{i=0}^{n} \sum_{j=0}^{j<i} \sum_{k=0}^{k<j} D_{ijk} + ... + D_{1...n} \tag{2.31}$$

où

$$D_i = \mathrm{Var}(E(Z|X_i)), \quad D_{ij} = \mathrm{Var}(E(Z|X_i, X_j)) - D_i - D_j, \quad \ldots \tag{2.32}$$

Cette décomposition est semblable à celle utilisée pour l'analyse de variance (ou ANOVA) et a déjà été largement discutée dans le cadre de calcul d'indices de sensibilité dans Archer et al. [1997]; Rabitz et al. [1999]; Saltelli et al. [2000]. En outre, dans le cas de modèle purement linéaire, cette approche est équivalente aux coefficients de régression standardisés [Saltelli et Bolado, 1998]. Dans l'équation (4.5), le terme D_i exprime l'influence du paramètre X_i seul (encore appelé effet de premier ordre), tandis que D_{ij} capture l'influence (ou effet de second ordre) de l'interaction entre les paramètres X_i et X_j, et ainsi de suite. L'indice de sensibilité d'un paramètre ou d'un ensemble de paramètres est alors mesuré comme le rapport entre sa contribution et la variance globale de la variable de sortie :

$$S_{i_1,i_2,..,i_s} = \frac{D_{i_1,i_2,..,i_s}}{D} \tag{2.33}$$

L'interprétation de ces termes est assez intuitive. En effet, $E(Z|X_i = x_i^*)$ exprime l'espérance de la réponse lorsque une valeur particulière x_i^* du paramètre X_i a été choisie. Si cette espérance varie beaucoup en fonction de x_i^*, alors le paramètre X_i est jugé influent. Or ses variations se mesurent naturellement par le biais de $\mathrm{Var}(E(Y|X_i = x_i^*))$. De la même façon, $\mathrm{Var}(E(Y|X_i, X_j))$ mesure les variations de $E(Y|X_i, X_j)$ en fonction deux valeurs de X_i et X_j, $D_{ij} = \mathrm{Var}(E(Z|X_i, X_j)) - D_i - D_j$ est donc une mesure de l'influence de l'interaction pure entre X_i et X_j sur la réponse Z.

Méthodes d'estimation des indices de sensibilité Les procédures disponibles pour évaluer efficacement les différents termes de l'équation (4.5) sont essentiellement le test de sensibilité basé sur l'amplitude de Fourier (ou FAST pour "*Fourier Amplitude Sensitivity Test*") [Cukier et al., 1973] et son successeur Extended FAST [Saltelli et al., 1999] ou bien la méthode de Sobol' [Sobol', 1990] et ses dérivées [**?**]. Enfin, Iman et Hora [1990] ainsi que McKay [1995] ont proposé différents schémas alternatifs mais peu utilisés. Comparées aux autres approches, celles de type FAST sont les plus précises, stables et rapides en terme de calculs [Homma et Saltelli, 1996; Saltelli et Bolado, 1998; Ravalico et al., 2005]. Cependant, d'une part elles ne peuvent pas estimer séparément les effets de premier ordre et d'ordre plus élevé, et d'autre part elles supposent que les paramètres suivent une loi uniforme entre leurs valeurs minimales et maximales, ce qui est problématique pour l'introduction d'un terme de dispersion tel que proposé au chapitre (4). C'est pourquoi l'approche dite de Sobol' a été sélectionnée et implantée durant ces travaux de thèse et est présentée en détail dans les sections suivantes.

Méthode d'estimation de Sobol'

Formulation analytique La méthode d'estimation des indices de sensibilité de Sobol', initialement publiée dans Sobol' [1990], se base sur l'expression des termes d'espérance et de variance sous la forme d'intégrales multiples. En effet, si l'on considère que les paramètres d'entrée incertains $(X_1, \ldots, X_n)$ sont indépendants et suivent une loi de densité continue, on obtient les formulations traditionnelles [Soong, 1981] suivantes pour le calcul de l'espérance :

$$E(Z) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} z(\mathbf{x})\ f_{X_1}(x_1) \ldots f_{X_n}(x_n)\ d\mathbf{x} \tag{2.34}$$

la variance totale :

$$Var(Z) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} (z(\mathbf{x}) - E(Z))^2\ f_{X_1}(x_1) \ldots f_{X_n}(x_n)\ d\mathbf{x} \tag{2.35}$$

et les termes de variance partielle :

$$Var(E(Z|X_{i_1}, \ldots, X_{i_s})) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} z^2_{i_1 \ldots i_s}(x_{i_1}, \ldots, x_{i_s}) \\ f_{X_{i_1}}(x_{i_1}) \ldots f_{X_{i_s}}(x_{i_s})\ dx_{i_1} \ldots dx_{i_s}\ -\ E(Z)^2$$

avec

$$z_{i_1 \ldots i_s}(x_{i_1}, \ldots, x_{i_s}) = \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} z(\mathbf{x}) \prod_{j \neq i_1 \ldots i_s} f_{X_j}(x_j) \quad d\mathbf{x}/(dx_{i_1} \ldots dx_{i_s}) \tag{2.36}$$

où $f_{X_i}(x_i)\ \forall\ i \in \{1, \ldots, n\}$ représente la loi de densité associée au paramètre incertain X_i, $d\mathbf{x}$ indique une intégration sur l'ensemble des variables $(X_1, \ldots, X_n)$ tandis que $d\mathbf{x}/(dx_{i_1} \ldots dx_{i_s})$ indique une intégration sur l'ensemble des variables exceptées $X_{i_1}, \ldots, X_{i_s}$. De plus, si l'on suppose que ces variables sont normalisées et suivent une loi uniforme entre 0 et 1, c'est-à-dire :

$$f_{X_i}(x_i) = \begin{cases} 1 & \text{si } x_i \in [0,1] \\ 0 & \text{sinon} \end{cases}\ ,\ \forall\ i \in \{1, \ldots, n\} \tag{2.37}$$

il vient naturellement

$$E(Z) = \int_0^1 \ldots \int_0^1 z(\mathbf{x})\ d\mathbf{x} \tag{2.38}$$

$$Var(Z) = \int_0^1 \ldots \int_0^1 (z(\mathbf{x}) - E(Z))^2\ d\mathbf{x} \tag{2.39}$$

et

$$Var(E(Z|X_{i_1}, \ldots, X_{i_s})) = \int_0^1 \ldots \int_0^1 z^2_{i_1 \ldots i_s}(x_{i_1}, \ldots, x_{i_s})\ dx_{i_1} \ldots dx_{i_s} - E(Z)^2 \tag{2.40}$$

avec

$$z_{i_1 \ldots i_s}(x_{i_1}, \ldots, x_{i_s}) = \int_0^1 \ldots \int_0^1 z(\mathbf{x})\ d\mathbf{x}/(dx_{i_1} \ldots dx_{i_s}) \tag{2.41}$$

Intégration numérique de Monte-Carlo Le calcul analytique des termes (2.38) à (2.41) est complexe à réaliser dans le cas d'une fonction $z(\mathbf{x})$ non linéaire et/ou un espace incertain de grande dimension. Pour cela, il est préférable de réaliser une intégration numérique de Monte-Carlo : considérons une fonction f définie dans un volume multi-dimensionnel V. L'idée de base de cette technique d'intégration est de sélectionner aléatoirement, mais distribués de manière uniforme, N points $(p_1, \ldots, p_N)$ dans V et d'estimer l'intégrale de f sur le volume (ainsi que l'erreur d'estimation) [Press et al., 2002] :

$$\int_{\mathcal{V}} f dV \simeq V \times \frac{1}{N} \sum_{i=1}^{N} f(x_i) \pm \frac{V}{\sqrt{N}} \times \sqrt{\frac{1}{N} \sum_{i=1}^{N} f^2(x_i) - (\frac{1}{N} \sum_{i=1}^{N} f(x_i))^2} \tag{2.42}$$

L'intérêt de cette approche est qu'elle est souple, indépendante de la forme de f et facile à implanter dans ses versions les plus simples. Cependant, comme souligné dans Press et al. [2002], l'erreur d'estimation décroît en $o(N^{-1/2})$, avec N nombre d'échantillons. Pour cela, une alternative efficace consiste à utiliser une technique de génération de nombres pseudo-aléatoires qui permet d'obtenir une meilleure précision pour un même nombre de points. Comme suggéré dans Saltelli et al. [2000], ce travail utilise le générateur de nombre initialement proposé dans Sobol' [1967] et dont une variante efficace, décrite par Antonov et Saleev [1979], peut être trouvée dans Press et al. [2002]. Cette séquence est construite de telle façon que l'erreur d'estimation décroît en $o(N^{-2/3})$.

Formules d'estimation L'application de la technique d'intégration numérique de Monte-Carlo, décrite ci-dessus, permet d'obtenir les estimation suivantes pour les termes d'espérance $\hat{E}(Z)$ et de variance totale $\hat{\mathrm{Var}}(Z)$:

$$\hat{E}(Z) = \frac{1}{N} \sum_{m=1}^{N} z(\mathbf{x}_m) \tag{2.43}$$

et

$$\hat{\mathrm{Var}}(Z) = \frac{1}{N} \sum_{m=1}^{N} z^2(\mathbf{x}_m) \; - \; \hat{E}(Z)^2 \tag{2.44}$$

où N est le nombre d'échantillons, les $\mathbf{x}_m, \; \forall m \in 1, \ldots, N$ forment un ensemble de points sélectionnés aléatoirement dans l'espace incertain pour lesquels la réponse $z(\mathbf{x}_m)$ est évaluée tandis que le chapeau indique une estimation. Le calcul des termes conditionnels est plus complexe. En effet, considérons les termes de premier ordre, c'est-à-dire de la forme $\mathrm{Var}(E(Z|X_i)), \forall \; i \in 1, \ldots, n$, il vient des équations (2.40) et (2.41) que :

$$\mathrm{Var}(E(Z|X_i)) = \int_0^1 \left[\int_0^1 \ldots \int_0^1 z(\mathbf{x}) d\mathbf{x}/dx_i \times \int_0^1 \ldots \int_0^1 z(\mathbf{x}) d\mathbf{x}/dx_i \right] dx_i - E(Z)^2 \tag{2.45}$$

suivant la technique d'intégration numérique de Monte-Carlo, on a :

$$\int_0^1 \ldots \int_0^1 z(\mathbf{x}) d\mathbf{x}/dx_i \simeq \sum_{m=1}^{N} z(\mathbf{v}_m, x_i) \tag{2.46}$$

où $\mathbf{v}_m$ est un point aléatoire pris dans l'espace $[0,1]^{n-1}$ défini à partir de toutes les variables exceptée X_i. L'équation (2.45) vaut alors :

$$\hat{\mathrm{Var}}(E(Z|X_i)) \simeq \frac{1}{N}\sum_{m=1}^{N}\left[\frac{1}{N_1}\sum_{m_1=1}^{N_1} z(\mathbf{v}_{m_1}, x_{i,m}) \times \frac{1}{N_2}\sum_{m_2=1}^{N_2} z(\mathbf{w}_{m_2}, x_{i,m})\right] - \hat{E}(Z)^2 \tag{2.47}$$

Les termes $\mathbf{v}$ et $\mathbf{w}$ représentent deux points aléatoires distincts pris dans $[0,1]^{n-1}$, tandis que $x_{i,m}$ est une valeur tirée de la loi de densité associée à X_i. L'équation (2.47) peut ensuite être simplifiée sous la forme :

$$\hat{Var}(E(Z|X_i)) = \frac{1}{N}\sum_{m=1}^{N} z(\mathbf{v}_m, x_{i,m}) \times z(\mathbf{w}_m, x_{i,m}) - \hat{E}(Z)^2 \tag{2.48}$$

D'une manière analogue, les termes conditionnels d'ordre deux peuvent être estimés à partir de :

$$\hat{Var}(E(Z|X_i, X_j)) = \frac{1}{N}\sum_{m=1}^{N} z(\mathbf{v}_m, x_{i,m}, x_{j,m}) \times z(\mathbf{w}_m, x_{i,m}, x_{j,m}) - \hat{E}(Z)^2 \tag{2.49}$$

où $\mathbf{v}_m$ et $\mathbf{w}_m$ sont alors tirés aléatoirement $[0,1]^{n-2}$. Pour les termes d'ordre supérieur la formule générale d'estimation s'écrit sous la forme :

$$\hat{Var}(E(Z|X_{i_1}, \ldots, X_{i_s})) = \frac{1}{N}\sum_{m=1}^{N} z(\mathbf{v}_m, \mathbf{s}_m) \times z(\mathbf{w}_m, \mathbf{s}_m) - \hat{E}(Z)^2 \tag{2.50}$$

où $\mathbf{v}_m$ et $\mathbf{w}_m$ sont générés dans $[0,1]^{n-s}$ et $\mathbf{s}_m$ est le vecteur représentant les variables $X_{i_1}, \ldots, X_{i_s}$.

Erreur d'estimation L'intégration numérique de Monte-Carlo revient à calculer la moyenne d'une fonction donnée sur un échantillon. Par exemple, $\hat{E}(Z)$ correspond à la moyenne de $z(\mathbf{x}_m)$. Comme souligné par Homma et Saltelli [1996] et Press et al. [2002] l'erreur d'estimation faite lors des calculs peut être appréciée par le biais de l'écart type de cette fonction sur le même échantillon. Ainsi, il vient :

$$\text{Erreur}(\ \hat{E}(Z)\ ) \simeq\ \pm\sqrt{\frac{\langle z^2(\mathbf{x})\rangle - \langle z(\mathbf{x})\rangle^2}{N}} \tag{2.51}$$

avec

$$\langle z(\mathbf{x})\rangle = \frac{1}{N}\sum_{m=1}^{N} z(\mathbf{x}_m) \quad \text{et} \quad \langle z^2(\mathbf{x})\rangle = \frac{1}{N}\sum_{m=1}^{N} z^2(\mathbf{x}_m) \tag{2.52}$$

où $\mathbf{x}_m$ représente un échantillon aléatoire de N points dans $[0,1]^n$. De la même façon, $\hat{\mathrm{Var}}(Z) - \hat{E}^2(Z)$ correspond à la moyenne de $z^2(\mathbf{x}_m)$. L'erreur d'estimation de la variance totale de la production peut être approchée par :

$$\text{Erreur}(\ \hat{\mathrm{Var}}(Z) - \hat{E}^2(Z)\ ) \simeq\ \pm\sqrt{\frac{\langle z^4(\mathbf{x})\rangle - \langle z^2(\mathbf{x})\rangle^2}{N}} \tag{2.53}$$

avec

$$\langle z^2(\mathbf{x})\rangle = \frac{1}{N}\sum_{m=1}^{N} z^2(\mathbf{x}_m) \quad \text{et} \quad \langle z^4(\mathbf{x})\rangle = \frac{1}{N}\sum_{m=1}^{N} z^4(\mathbf{x}_m) \tag{2.54}$$

Enfin comme suggéré dans Homma et Saltelli [1996], l'erreur d'estimation des termes de variance partielle se généralise sous la forme (les notations sont les mêmes que celles utilisés dans la section précédente) :

$$\text{Erreur}(D_{i_1,i_2,..,i_s}) = \frac{1}{\sqrt{N}} \sqrt{F - I^2} \tag{2.55}$$

avec

$$F = \frac{1}{N} \sum_{m=1}^{N} z(\mathbf{v}_m, \mathbf{s}_m) \times z(\mathbf{w}_m, \mathbf{s}_m) \quad \text{et} \quad I = \frac{1}{N} \sum_{m=1}^{N} [z(\mathbf{v}_m, \mathbf{s}_m) \times z(\mathbf{w}_m, \mathbf{s}_m)]^2 \tag{2.56}$$

Implantation Le principal avantage de cette approche est que la difficile estimation des intégrales multiples est réduite à un simple problème d'échantillonnage de la fonction $z(\mathbf{x})$. En outre, l'estimation de l'ensemble des termes de variance totale et partielle s'effectue suivant un schéma semblable : il suffit de tirer aléatoirement dans $[0, 1]^n$ N paires de points partageant une ou plusieurs valeurs de coordonnées. En pratique, l'approche décrite dans Homma et Saltelli [1996] est envisagée ici, celle-ci se divise en trois étapes :

1. génération suivant un algorithme de tirages aléatoires (ou pseudo aléatoires) une matrice de N paires de points et de taille $(N \times 2n)$;
2. évaluation lors du calcul des estimations (2.43), (2.44), et (2.50) les vecteurs $\mathbf{x}_m$, $\mathbf{v}_m$ et $\mathbf{s}_m$ dans les n premières colonnes de la matrice ;
3. sélection, à partir des n dernières colonnes, les valeurs du vecteur $\mathbf{w}_m$, représentant les variables qui doivent être ré-échantillonnées pour une paire de points donnée.

Validation de l'approche

Afin de démontrer l'efficacité de l'analyse de sensibilité basée sur l'approche de Sobol' et de valider son implantation, le modèle analytique suivant à trois variables est considéré :

$$z(x_1, x_2, x_3) = \sin(x_1) \; + \; 7 \; sin^2(x_2) \; + \; 0.2 \; x_3^4 \; sin(x_1) \tag{2.57}$$

Les variables $\{x_1, x_2, x_3\}$ sont supposées indépendantes et suivant toutes une loi uniforme dans $[-\pi, +\pi]$. Cette fonction est similaire à celle proposée dans Ishigami et Homma [1990] et est d'intérêt pour son comportement non linéaire et parce qu'il est possible de calculer les différents indices analytiquement.

Ce tableau compare les valeurs analytiques aux résultats obtenus suivant l'approche proposée à partir de différentes tailles d'échantillons. Ces résultats montrent la capacité de l'approche à estimer en un temps raisonnable (variant entre 0.001 secondes pour 100 points à 73.0 secondes pour 10^6 points) la sensibilité de la fonction aux trois variables d'entrée. En outre, ils montrent l'augmentation de la précision de l'estimation en fonction du nombre de points dans l'échantillon. Par exemple, sur ce modèle analytique, dès l'utilisation de 10 000 points, on dispose d'une précision suffisante pour discriminer les variables ainsi que leurs interactions et les classer suivant leur importance. La possibilité de mesurer l'impact des interactions est capital. En effet, dans le cas présent, il existe une forte interaction entre x_1 et x_3 alors que cette dernière variable prise seule n'a aucun effet. Or, elle aurait été négligée s'il n'avait été possible de calculer D_{13}.

Variables	Analytique	N=100	N=10^3	N=10^4	N=10^5
S_1	0.379	0.378	0.379	0.381	0.380
S_2	0.194	0.182	0.176	0.193	0.192
S_3	0.0	0.248	0.003	3E-4	-1E-4
S_{12}	0.0	0.064	0.012	-0.001	-2.1E-4
S_{13}	0.426	0.129	0.440	0.426	0.428
S_{23}	0.0	0.064	0.012	-0.001	-2.1E-4
Résidu	-	-0.065	-0.013	0.0025	5.2E-4

TAB. 2.1 – Valeurs estimées des indices de sensibilité correspondant à la fonction $z(x_1, x_2, x_3)$ (eqn. (2.57)) en fonction de la taille des échantillons et comparées aux résultats analytiques tels que publiés dans Homma et Saltelli [1996].

Remarques finales

En conclusion, les remarques suivantes sur l'approche proposée par Sobol' pour l'estimation de la sensibilité d'une surface de réponse vis-à-vis de paramètres incertains peuvent être faites :

- les indices calculés sont des indices globaux permettant de capturer la sensibilité de la réponse sur l'ensemble du domaine d'incertitude ;
- la méthode permet de prendre en compte à la fois les indices de premier ordre mais aussi les interactions entre paramètres ;
- le problème de l'estimation des indices est réduit à un simple problème d'échantillonnage et ainsi est indépendant de la forme de la fonction dont il faut calculer la sensibilité ;
- le calcul des différents termes se fait suivant une procédure unique et simple, ce qui favorise son implantation et sa robustesse ;
- l'hypothèse d'une distribution uniforme des variables X_i a de nombreux avantages en terme de calculs, cependant elle n'est pas obligatoire, ainsi d'autres distributions paramétriques (gaussienne, triangulaire, etc.) ou non paramétriques peuvent être considérées ;
- enfin, la seule hypothèse réellement contraignante est celle de l'indépendance entre les paramètres incertains qui, bien que nécessaire, est rarement observée dans le cadre d'un modèle géologique.

2.5.3 Inversion bayésienne de données de production

Objectifs

L'inversion bayésienne est une technique d'analyse de données qui cherche à mesurer l'impact de l'introduction de nouvelles informations sur celles déjà disponibles. Dans le cadre des simulations d'écoulement, cela concerne la prise en compte de données réelles de production lors des procédures de calage d'historique (ou "*history matching*").Considérons qu'un certain nombre

de modèles correspondants à différentes combinaisons de paramètres incertains soient simulés. la procédure de calage consiste à sélectionner le ou les modèles pour lesquels la différence entre les résultats de simulation et les données est la plus faible. Dans ce cadre, l'inversion bayésienne a pour objectif de capturer l'impact de cette sélection sur la distribution des paramètres incertains.

Le principal résultat de cette analyse est une distribution (dite *a posteriori*) jointe multivariée des paramètres incertains qui tient compte des données de production. Dans la suite, cette loi de densité sera notée $p(\mathbf{x}|\mathbf{z} = \mathbf{z}_{obs})$ [1]. Elle exprimera la probabilité qu'un modèle ou qu'une combinaison de paramètres incertains, $\mathbf{x}$, soit vraisemblable au regard de l'information connue *a priori* (c'est-à-dire concernant les paramètres du modèle mais aussi les données historiques de production) et du modèle physique reliant les paramètres incertains aux résultats de simulation, ici les surfaces de réponse. À partir de cette loi, il est possible de calculer les différents histogrammes marginaux correspondants aux paramètres incertains mais aussi utiliser un ensemble de points ayant la même densité que cette loi pour échantillonner d'autres surfaces de réponse après la dernière mesure de production : cela permet alors de procéder à une mise à jour des prévisions de production.

Principe

Comme décrit ci-dessus, l'inversion bayésienne a pour but de calculer la loi de densité jointe *a posteriori*, notée ici $p(\mathbf{x}|\mathbf{z} = \mathbf{z}_{obs})$ des paramètres incertains [Tarantola, 2005]. Ce calcul s'appuie sur la forme classique de la formule de Bayes :

$$p(\mathbf{x}|\mathbf{z} = \mathbf{z}_{obs}) = k\ p_0(\mathbf{x})\ \mathcal{L}(\mathbf{x}) \tag{2.58}$$

où

- k est une constante de normalisation ;
- $p_0(\mathbf{x})$ représente la loi de densité, dite *a priori*, des paramètres incertains. Cette loi résume l'état des connaissances avant la prise en compte des nouvelles données de production. En l'absence de toute information, les paramètres sont supposés indépendants et suivant une loi uniforme entre leurs valeurs minimales et maximales, respectivement x_i^- et x_i^+. On a alors :

$$p_0(\mathbf{x}) = \prod_{i=1}^{n} p_0(x_i), \quad \text{avec,} \quad \forall i \in [1, n], \quad p_0(x_i) = \begin{cases} \frac{1}{x_i^+ - x_i^-} & \text{if } x_i \in [x_i^-, x_i^+] \\ 0 & \text{sinon} \end{cases} \tag{2.59}$$

- $\mathcal{L}(\mathbf{x})$ est une fonction de vraisemblance (encore appelée "*likelihood*"). Elle mesure la probabilité qu'une combinaison de paramètres incertains donnée explique les données historiques de production.

[1] Par rapport aux notations utilisées jusqu'à présent où les données de production z étaient traitées de façon indépendantes, ici elles sont toutes rassemblées sous forme d'un vecteur de taille n_z noté $\mathbf{z} = [z^1, \ldots, z^{n_z}]$. En effet, l'un des intérêts de l'inversion bayésienne est de combiner les sources d'informations comme des données de nature différente (production cumulée d'huile/eau, débits au puits, pressions aux puits, ..) ou leur évolution dans le temps.

Le calcul de la fonction de vraisemblance $\mathcal{L}(\mathbf{x})$ s'appuie sur les données historiques de production traduites sous la forme d'une loi de densité $p_0(\mathbf{z})$ et sur les surfaces de réponse des variables de production à prendre en compte $\mathbf{z}(\mathbf{x})$. Ces surfaces approchent le comportement du simulateur d'écoulement et permettent de relier les paramètres incertains aux données de production. Dans le cas général, elles sont supposées univalentes[1] : pour un point $\mathbf{x}$ donné, il n'existe qu'une seule valeur de réponse possible. La fonction de vraisemblance devient alors :

$$\mathcal{L}(\mathbf{x}) = p_0(\mathbf{z}(\mathbf{x})) \tag{2.60}$$

Deux cas sont alors généralement admis pour le calcul de $p_0(\mathbf{z}(\mathbf{x}))$:

1. l'absence de toute erreur de mesure dans les données observées de production, la loi $p_0(\mathbf{z}(\mathbf{x}))$ est alors un Dirac :
$$p_0(\mathbf{z}(\mathbf{x})) = \delta(\mathbf{z}(\mathbf{x}) - \mathbf{z}_{obs}) \tag{2.61}$$
2. l'existence d'une erreur de distribution normale centrée sur zéro et de covariance $\mathbf{C}_d$ [2] :
$$p_0(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^{n_d}\ \det(\mathbf{C}_d)^{-1}}}\ e^{-\frac{1}{2}(\mathbf{z}-\mathbf{z}_{obs})^t \mathbf{C}_d (\mathbf{z}-\mathbf{z}_{obs})} \tag{2.62}$$

Le cas (1) est le plus simple : il considère que seuls les points de l'espace d'incertitude où la production prédite par les surfaces de réponse est égales aux données historiques sont vraisemblables. Cependant, l'estimation de la fonction de vraisemblance est alors sensible à la présence de bruit dans les données et aux erreurs de modélisation. Cela conduit le plus souvent à trouver un ensemble vide de modèles acceptables. Aussi, en pratique, il est préférable de se placer dans le cas (2).

Calcul de la loi de densité a posteriori

Comme souligné dans Tarantola [2005] et au vu de la complexité des surfaces de réponse envisagées, il est impossible d'obtenir une formulation analytique pour calculer la loi de densité a posteriori $p(\mathbf{x}|\mathbf{z} = \mathbf{z}_{obs})$. Son estimation doit donc reposer sur une méthode numérique. Une approche efficace est l'utilisation de l'algorithme d'échantillonnage de Métropolis [Metropolis et Ulman, 1949]. Celui-ci a pour objectif de parcourir aléatoirement une loi de densité initiale et, en utilisant une règle probabiliste, de modifier le parcours (en acceptant certains points et en rejetant d'autres) afin d'échantillonner une loi cible [Mosegaard et Tarantola, 1995] (figure (2.8)). Dans notre cas :

- la loi de densité initiale est $p_0(\mathbf{x})$: elle a l'avantage d'être le plus souvent paramétrique (ici uniforme, voir équation (2.59)) et est donc aisée à parcourir à l'aide d'algorithmes classiques d'échantillonnage de type Monte-Carlo ;
- la cible est la loi de densité a posteriori, $p(\mathbf{x}|\mathbf{z} = \mathbf{z}_{obs})$;

[1] Voir le chapitre 4 pour un exemple de traitement de surfaces non univalentes

[2] Les surfaces de réponse correspondant à ces différentes données étant construites de façon indépendantes, on supposera ici que c'est réellement le cas et que cette matrice de covariance est diagonale.

– tandis que la règle (appelée règle de Métropolis) d'acceptation ou de rejet d'un point est basée sur la fonction de vraisemblance $\mathcal{L}(\mathbf{x})$.

La règle de Métropolis est la suivante : considérons un parcours aléatoire du domaine d'incertitude, celui-ci est une succession de sauts d'un point $\mathbf{x}_\alpha$ à un second point $\mathbf{x}_\beta$. L'acceptation ou le rejet de ce saut s'effectue suivant la règle suivante (d'après Tarantola [2005]) :

– si $\mathcal{L}(\mathbf{x}_\beta) > \mathcal{L}(\mathbf{x}_\alpha)$, la transition est immédiatement acceptée,

– sinon, la décision d'accepter la transition s'effectue avec la probabilité :

$$P_{\alpha\to\beta} = \frac{\mathcal{L}(\mathbf{x}_\beta)}{\mathcal{L}(\mathbf{x}_\alpha)} \tag{2.63}$$

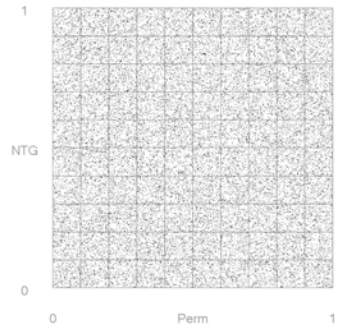
Si le nouveau point $\mathbf{x}_\beta$ n'est pas accepté, un nouveau saut aléatoire à partir de $\mathbf{x}_\alpha$ est proposé et ainsi de suite jusqu'à obtenir le nombre voulu de points pour l'échantillonnage de $p(\mathbf{x}|\mathbf{z} = \mathbf{z}_{obs})$.

Dans ce travail, deux algorithmes de parcours aléatoire ont été testés. Le premier consiste à tirer aléatoirement et de manière indépendante des points dans tout le domaine. Or, dans le cas où les zones vraisemblables sont très localisées et/ou l'espace est de grande dimension, cette approche est lente à converger car un grand nombre de points est rejeté au cour du processus. Une alternative consiste à utiliser un algorithme de type Markov Chain Monte-Carlo (MCMC) où la position du nouveau point se détermine toujours aléatoirement mais à une distance donnée h du point précédent dans la chaîne. Les tirages aléatoires sont alors localisés uniquement autour des zones vraisemblables. La distance h peut être soit fixée à l'avance soit déterminée aléatoirement à chaque saut suivant une loi de distribution donnée. On montre que cet algorithme obtient des résultats identiques au premier mais limite largement le nombre de points rejetés et ainsi accélère les calculs. Un point capital est la détermination de la distance h. Comme discuté dans Tarantola [2005], ce facteur contrôle d'une part la précision des estimations (si h est trop grande) mais aussi la possibilité d'échantillonner différentes zones vraisemblables et la rapidité des calculs (si h est trop petite). Il est donc indispensable de trouver un équilibre. D'après Tarantola [2005], une règle empirique efficace consiste à sélectionner h de telle façon que le taux d'acceptation soit environ de 30-50%. Les avantages de cette approche d'estimation sont :

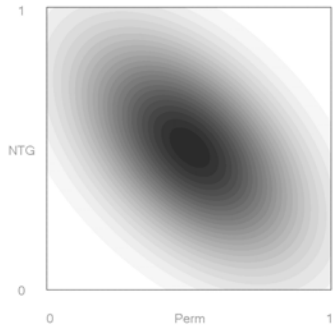
1. comme montré dans Mosegaard et Tarantola [1995, 2002] cet algorithme échantillonne effectivement la probabilité a posteriori sans idées préconçues sur sa forme exacte. Le seul élément nécessaire est une méthode pour évaluer la fonction de vraisemblance en tous points $\mathbf{x}$ du domaine d'incertitude ;

2. l'ensemble du travail repose encore une fois sur un simple échantillonnage de type Monte-Carlo qui permet de prendre en compte tout type de fonctions de réponse continues, discrètes ou non linéaires.

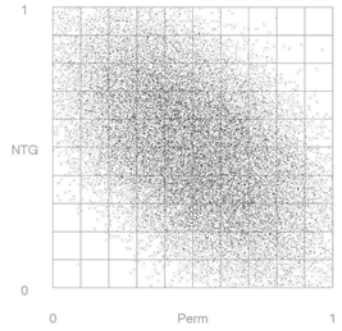
Ces avantages font donc de cette approche une méthode de choix dans notre bibliothèque d'outils pour analyser une surface de réponse et tenir compte d'éventuelles données historiques de production.



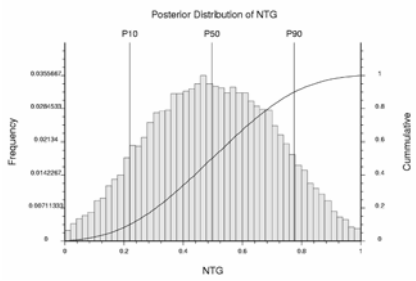
(a) échantillonnage de la loi de densité a priori



(b) fonction de vraisemblance



(c) échantillonnage de la loi de densité a posteriori



(d) histogramme marginal a posteriori du Net-To-Gross

Fig. 2.8 – Exemple de mise en œuvre de l'algorithme de Métropolis [Metropolis et Ulman, 1949] dans un espace à deux dimensions. La distribution *a priori* est uniforme (a) tandis que la fonction de vraisemblance suit une loi gaussienne (b). Chaque échantillon est composé d'environ 30000 points. Enfin, (d) représente l'histogramme marginal a posteriori d'un des deux paramètres incertains calculé à partir de l'échantillon (c).

2.6 Cas d'études

Cette section s'appuie sur différents modèles synthétiques pour illustrer et discuter les méthodologies décrites précédemment. Le premier est un simple modèle d'anticlinal dont différents paramètres (position du toit, position du contact eau-huile, propriétés des fluides, etc.) sont jugés incertains. Ce modèle a pour but d'illustrer l'ensemble de la chaîne d'analyse des résultats de simulation vis-à-vis de paramètres incertains. Dans un second temps, les capacités et les performances des techniques de construction de surfaces de réponse proposées dans ces travaux sont évaluées plus en avant : les modèles considérés sont le même que précédemment est considéré mais avec un nombre plus grand de paramètres ainsi que le "IC Fault Model" Carter et al. [2006] connu pour son comportement complexe.

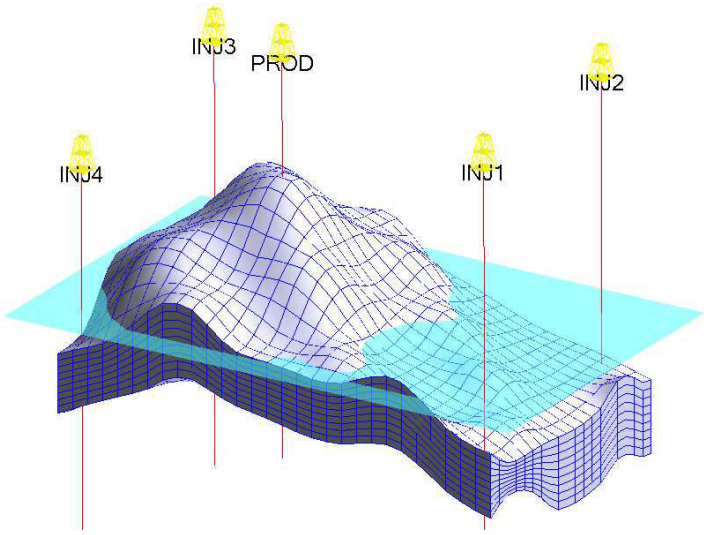


FIG. 2.9 – Vue d'ensemble du modèle synthétique de réservoir utilisé pour la validation des méthodes d'analyse proposées. En outre sont représentés les différents puits ainsi que la position du contact eau-huile de l'une des réalisations.

2.6.1 Modèle Anticlinal avec 4 paramètres incertains

Présentation du modèle

Ce modèle synthétique de réservoir représente un anticlinal s'étendant sur 10 000 pieds dans le sens de la longueur, 6 000 pieds dans celui de la largeur et d'épaisseur moyenne égale à 230 pieds. Le maillage utilisé pour représenter le réservoir est une grille structurée irrégulière de dimensions 29×19×9 cellules (figure (2.9)).

Le modèle pétro-physique (porosité et perméabilité) suppose que les propriété sont uniformes dans l'ensemble du réservoir. La porosité est une propriété scalaire tandis que la perméabilité est un tenseur diagonal avec $k_x = k_y$ et $k_z = (k_v/k_h) \times k_x$. La roche est supposée incompressible.

Ce réservoir ne contient que deux fluides, l'huile et l'eau, supposés immiscibles et incompressibles. Il sont décrits ar leur masse volumique et leur viscosité dynamique ainsi qu'une courbe de perméabilité relative qui exprime l'évolution de la perméabilité de la roche aux fluides en fonction de leur saturation. Le modèle choisi ici est de type "*XCurves*" où $kr_w(S_w) = S_e$ et $kr_o(S_w) = 1 - S_e$ avec S_e, saturation en eau effective du milieu. Ce modèle pourra être modifié en variant les saturations résiduelles en eau et/ou en huile. À l'état initial, les fluides sont supposés en équilibre : le contact eau-huile marque la séparation des deux phases et sert directement de base au calcul des saturations initiales.

Le modèle contient cinq puits : quatre injecteurs d'eau situés aux extrémités du modèle et un producteur localisé au sommet de l'anticlinal. Ces puits traversent et sont ouverts sur l'ensemble de l'épaisseur du réservoir. Ils sont représentés à l'aide d'un modèle de puits de type Peaceman avec un facteur de "skin" nul, un rayon de 0.3 pieds et un facteur de complétion de 1. Ces puits

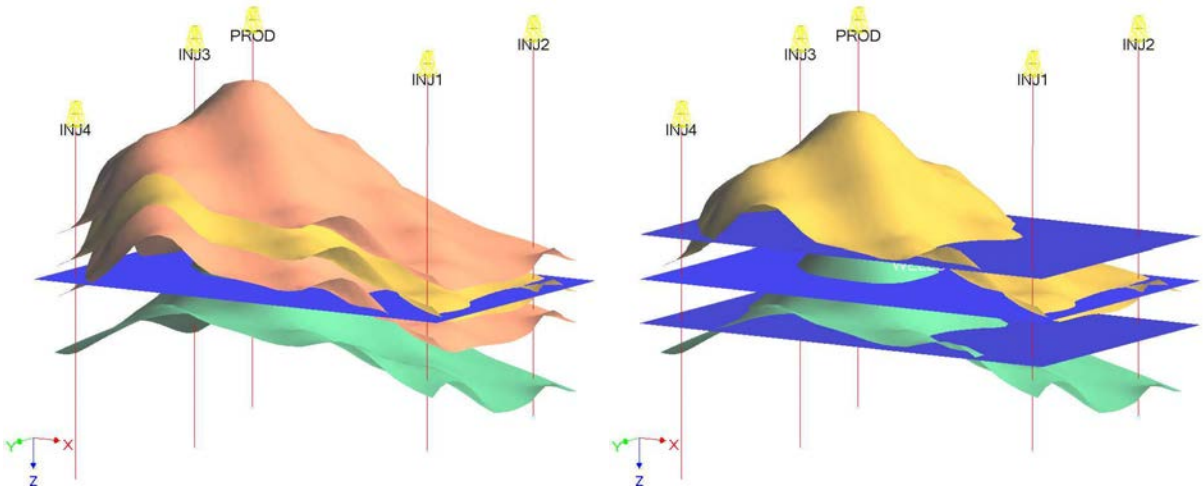


FIG. 2.10 – Gauche : Position du toit du réservoir (paramètre TOP) dans les cas haut (+70 ft), moyen, et bas (-70 ft). Droite : Position du contact eau-huile (paramètre OWC), également dans les cas haut (+100 ft), moyen, et bas (-100 ft).

Variables	Value minimale	Valeur moyenne	Valeur maximale
TOP	-70 ft	0 ft	+70 ft
OWC	-3900 ft	-3800 ft	-3700 ft
PERM	600 mD	800 mD	1000 mD
k_v/k_h	0.1	0.3	0.5

TAB. 2.2 – Valeurs minimales, moyennes et maximales des quatre paramètres incertains utilisés dans le cas d'étude (1). Pour la planification des expériences, chaque paramètre est représenté par cinq niveaux possibles ce qui conduit à un total de 625 simulations d'écoulement.

sont tous contrôlés en pression sans conditions d'arrêt à 5200 psi pour les injecteurs et 1000 psi pour le puits producteur.

Paramètres incertains de l'étude et plan d'expérience

Ce réservoir comprend quatre paramètres incertains (voir figure (2.10) et table (2.2)) :
- la position du toit du réservoir (TOP),
- la position du contact eau-huile (OWC),
- la perméabilité horizontale (PERM) du milieu,
- le rapport de la perméabilité verticale sur la perméabilité horizontale (k_v/k_h).

Le plan d'expérience employé ici considère (1) que chaque paramètre prend 5 valeurs possibles et (2) que toutes les combinaisons (soit un total de 625 modèles) doivent être simulées. Les paramètres restant du modèles sont estimés certains et fixes. En particulier, la porosité est constante à 20%, la viscosité dynamique et la masse volumique de l'huile sont, respectivement,

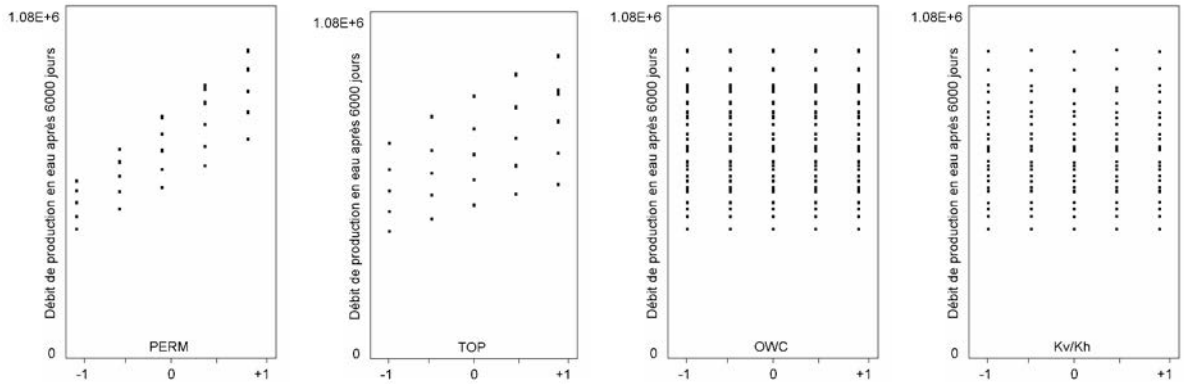
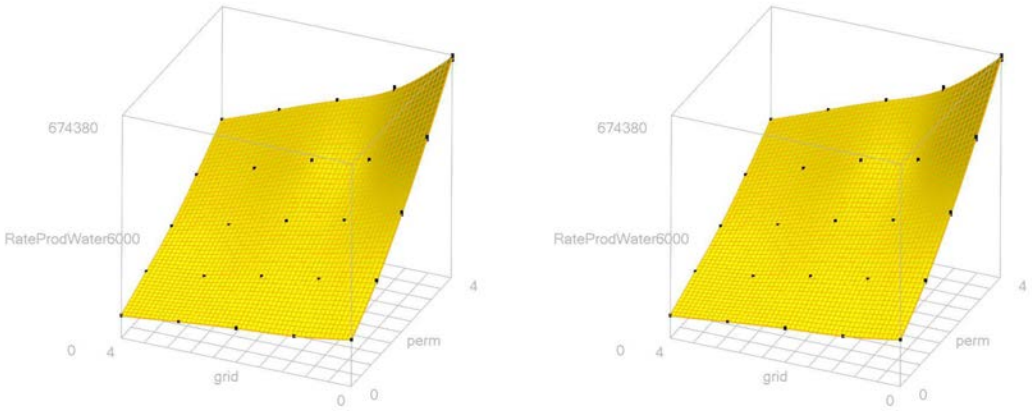


FIG. 2.11 – Effet des facteurs incertains (PERM, TOP, OWC, et k_v/k_h) sur le débit d'eau au puits producteur à 6000 jours.



(a) Interpolation lisse discrète

(b) Krigeage dual

FIG. 2.12 – Surfaces de réponse du débit en eau au puits producteur à 6000 jours par rapport aux paramètres PERM et TOP. Les méthodes de construction employées ici sont (a) l'interpolateur lisse discret avec contrainte de gradient constant et (b) le krigeage dual.

de 2.0 cp et 53 lf.ft^{-3} tandis que celles de l'eau sont de 0.3 cp et 64 lf.ft^{-3}. Enfin, les saturations résiduelles en huile et en eau sont nulles ($S_{wc} = S_{or} = 0$).

Les 625 modèles de réservoir ont été simulés sur 6000 jours à l'aide du simulateur d'écoulement sur lignes de courant commercial 3DSL [Streamsim, 2004] avec une sortie tous les 1000 jours. Les variables cibles sont la production cumulée ainsi que le débit au puits producteur en huile et en eau.. Le modèle étant de taille très réduite chaque simulation dure environ 1.5 min soit un total de 15 heures 37 minutes.

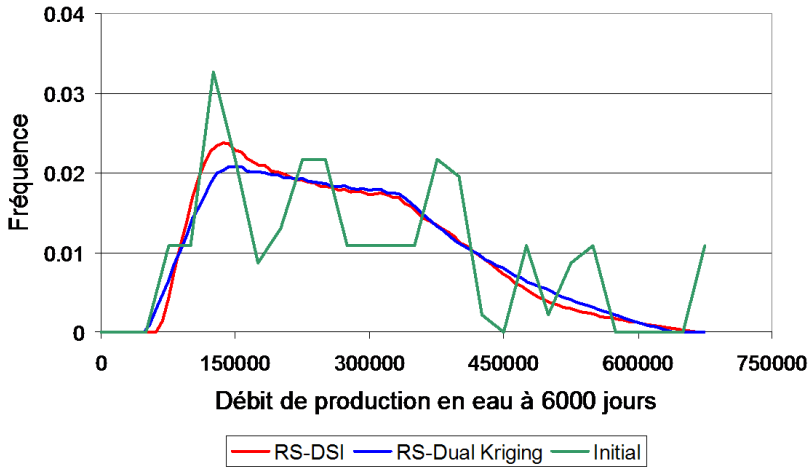


FIG. 2.13 – Fréquence des valeurs de débit en eau au puits producteur à 6000 jours calculé à partir des 625 données initiales (vert), et de 10^6 échantillons des surfaces de réponse par l'approche DSI (en rouge) et par krigeage dual (en bleu).

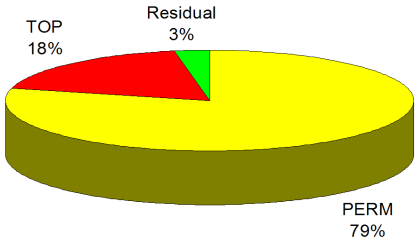
Construction de surfaces de réponse

La première étape de l'analyse des résultats de simulation consiste a représenter l'évolution des variables de production en fonction des valeurs prises par les quatre paramètres incertains. La figure (2.11) s'intéresse, par exemple, au débit d'eau au puits producteur après 6000 jours. Cela permet de constater l'effet important de la position du toit du réservoir puis de la perméabilité du milieu alors que la position du contact eau-huile ainsi que le rapport k_v/k_h ne semblent avoir aucun effet. L'évolution du débit en eau à 6000 jours par rapport aux quatre paramètres incertains est ensuite modélisée à l'aide de surfaces de réponse. Les méthodes employées sont :

- l'interpolation lisse discrète sur une grille régulière structurée de 16 000 nœuds (soit 20 dans chaque directions de l'espace) avec un facteur d'ajustement de 0.01, favorisant ainsi la continuité du résultat par rapport au respect des données ;
- le krigeage dual avec un variogramme théorique de type puissance ($\omega = 1.75$) et sans terme de régularisation.

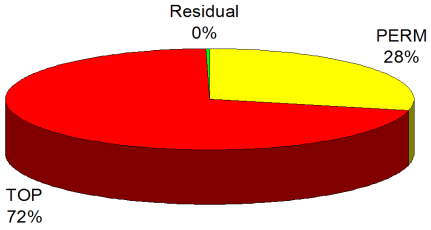
Dans les deux cas, le temps de construction est négligeable. La figure (2.12) montre les surfaces de réponse obtenues dans le plan (TOP-PERM) car seuls ces paramètres ont un effet sensible sur la production. En parallèle, la figure (2.13) compare les fréquences des différentes valeurs de production calculées à partir (1) des 625 points de données initiaux et (2) d'échantillons de points aléatoires pour lesquels la production a été estimée à l'aide des surfaces de réponse. Ces figures montrent que les deux approches proposées produisent des résultats semblables et cohérents avec les données, on ne peut cependant pas distinguer qu'elle serait la meilleure approche.

Variable	Variance	Indices de sensibilité
PERM	1.13e+010	78.82 %
TOP	2.65e+009	18.39 %
OWC	-5.7e+006	0.00 %
K_v/K_h	-1.9e+006	0.00 %
Total	1.44e+010	97.21 %



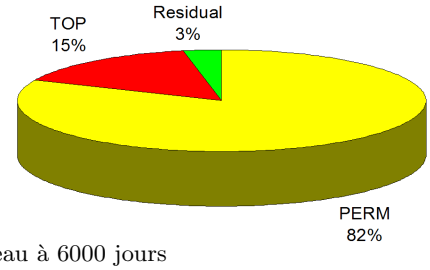
(a) Débit en eau à 6000 jours

Variable	Variance	Indices de sensibilité
PERM	6.13e+009	28.48 %
TOP	1.53e+010	71.21 %
OWC	6.28e+007	0.02 %
K_v/K_h	5.91e+007	0.02 %
Total	2.15e+010	99.73 %



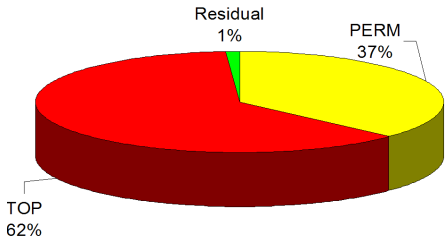
(b) Débit en huile à 6000 jours

Variable	Variance	Indices de sensibilité
PERM	5.42e+016	81.54 %
TOP	1.01e+016	15.33 %
OWC	-2.2e+013	-0.00 %
K_v/K_h	-2.6e+013	-0.00 %
Total	6.63e+016	96.87 %



(c) Production cumulée en eau à 6000 jours

Variable	Variance	Indices de sensibilité
PERM	3.55e+017	36.89 %
TOP	5.96e+017	61.88 %
OWC	-8.5e+014	-0.00 %
K_v/K_h	-8.2e+014	-0.00 %
Total	9.63e+017	96.87 %



(d) Production cumulée en huile à 6000 jours

FIG. 2.14 – Indices de sensibilité des productions cumulées et des débit au puits producteur en huile et en eau à 6000 jours par rapport aux paramètres incertains

Paramètre incertain	Valeur de référence
Position du toit	+59 ft
Position du contact eau-huile	-3843 ft
Perméabilité du milieu	718 mD
rapport K_v/K_h	0.284

TAB. 2.3 – Paramètres du modèle de référence utilisé pour générer un historique de production synthétique lors de l'inversion bayésienne.

Analyse de sensibilité

La seconde étape consiste à procéder à l'analyse de sensibilité des résultats de simulation par rapport aux paramètres incertains. Le schéma d'intégration employé ici est basé sur un échantillonnage de Sobol' de 10 000 points, ce qui d'après Press et al. [2002] permet d'obtenir une précision de calcul d'environ 1 %. La figure (2.14) montre les indices de sensibilité de premier ordre des différents débits et les cumuls de production en huile comme en eau à 6000 jours. Ces résultats indiquent que :

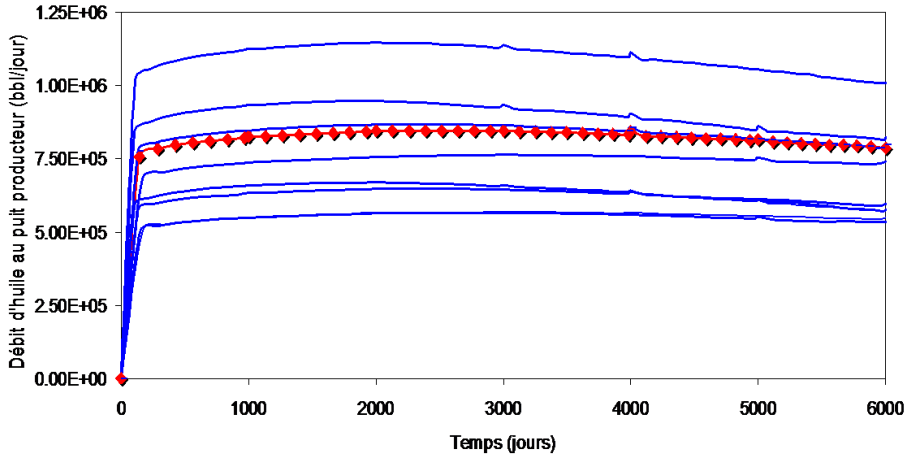
- la position du toit et la perméabilité du milieu (paramètres TOP et PERM) sont les plus influents ;
- pour la production (cumulée et débit) en huile, la position du toit est plus influente (approx. 62 - 72 %) que la perméabilité du milieu (approx. 28% - 37%) et inversement pour les variables liées à l'eau.
- les paramètres OWC et Kv/Kh ont un effet négligeable ;
- les interactions semblent être aussi négligeables, en effet les quatre variables seules expliquent déjà au minimum 96.87 % de la variance totale.

Ces résultats confirment les observations déjà effectuées sur les surfaces de réponse (figure (2.12)). Ils montrent la capacité du schéma de calcul à mesurer la sensibilité des surfaces de réponse par rapport aux paramètres incertains et à quantifier des constatations qualitatives, qui n'auraient pu être réalisées dans le cas d'un ensemble de paramètres incertains de plus grande dimension.

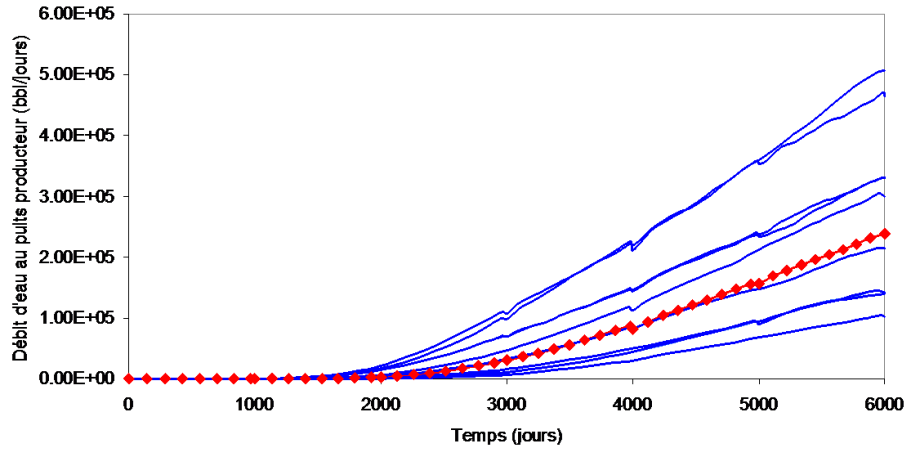
Inversion bayésienne de données de production

Pour illustrer l'inversion bayésienne, le modèle dont les paramètres sont indiqués table (2.3) est choisi comme référence. Ce modèle est simulé sur 6000 jours (figure (2.15)) pour générer un historique de production sur 3000 jours, le restant étant mis de côté pour tester les résultat de l'inversion. La procédure adoptée est la suivante :

1. calcul de la fonction de vraisemblance en supposant une erreur de mesure dans les données historique de production de 10% environ ;
2. échantillonnage suivant l'algorithme de Métropolis de la loi de densité multivariée a posteriori des paramètres incertains ;
3. calcul des histogrammes marginaux a posteriori de chaque paramètre ;



(a) Débit d'huile au puits producteur



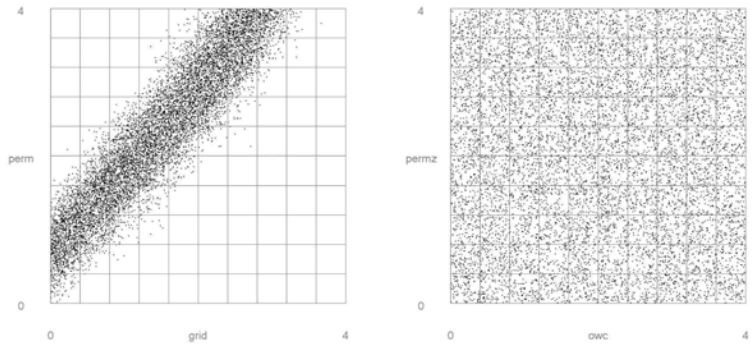
(b) Débit d'eau au puits producteur

FIG. 2.15 – Débits en huile et en eau au puits producteur du modèle de référence comparés à 10 autres résultats de simulation choisis aléatoirement.

4. calcul de l'histogramme a posteriori des productions cumulées en huile et en eau à 6000 jours.

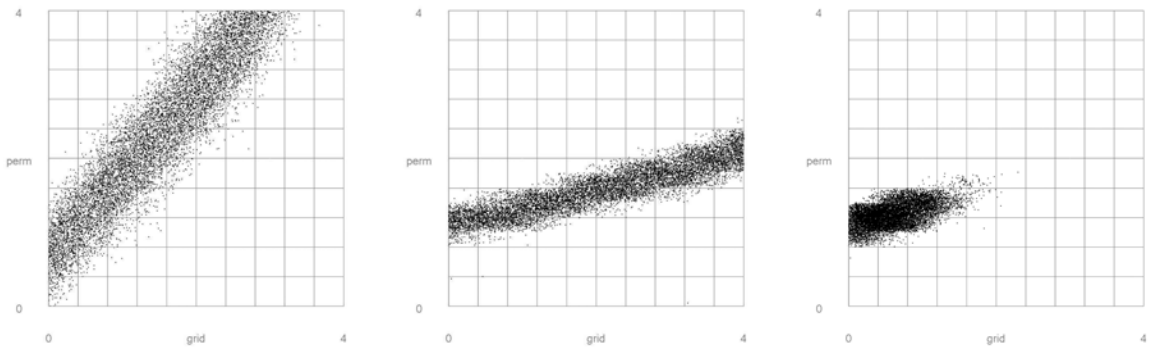
Les résultats sont exposés sur les figures (2.16) à (2.19) :

– La figure (2.16) montre les résultats de l'échantillonnage de la loi de densité multivariée des paramètres incertains obtenue après inversion bayésienne dans le cas ou seule la production d'huile cumulée sur 3000 jours est prise en compte. Les paramètres PERM et TOP ayant effet sur la production, leur intervalle de distribution peut être réduit grâce à l'inversion bayésienne au contraire des paramètres OWC et Kv/Kh.



(a) (b)

FIG. 2.16 – Résultat de l'échantillonnage de la distribution a posteriori si seule la production d'huile cumulée sur 3000 jours est prise en compte : représentation dans le sous-espace PERM-TOP (a) et dans le sous-espace OWC-Kv/Kh (b).



(a) (b) (c)

FIG. 2.17 – Résultat de l'échantillonnage de la distribution a posteriori dans la cas ou seules les production cumulées en huile (a) et en eau (b) sur 3000 jours sont prises en compte et (c) dans le cas où toutes les données sont prises en compte.

- La figure (2.17) illustre comment la combinaison de plusieurs informations de natures différentes (ici les débits en huile et en eau) peut améliorer la réduction des incertitudes sur les paramètres incertains.
- Ensuite, la figure (2.18) présente les histogrammes marginaux a posteriori des différents paramètres. A l'opposé de la figure (2.16), ici toutes des données sont prises en compte. Les résultats confirment ceux obtenus précédemment. En outre, il est important de noter que le modèle de référence (indiqué par la flèche rouge) se retrouve toujours parmi les valeurs vraisemblables et dans leur l'intervalle P_{10} - P_{90}.
- Enfin, la figure (2.19) illustre la possibilité d'utiliser les points représentant les modèles a posteriori vraisemblables pour échantillonner les surfaces de réponse après la dernière donnée de production. Cela montre la réduction de l'incertitude sur les prédictions de

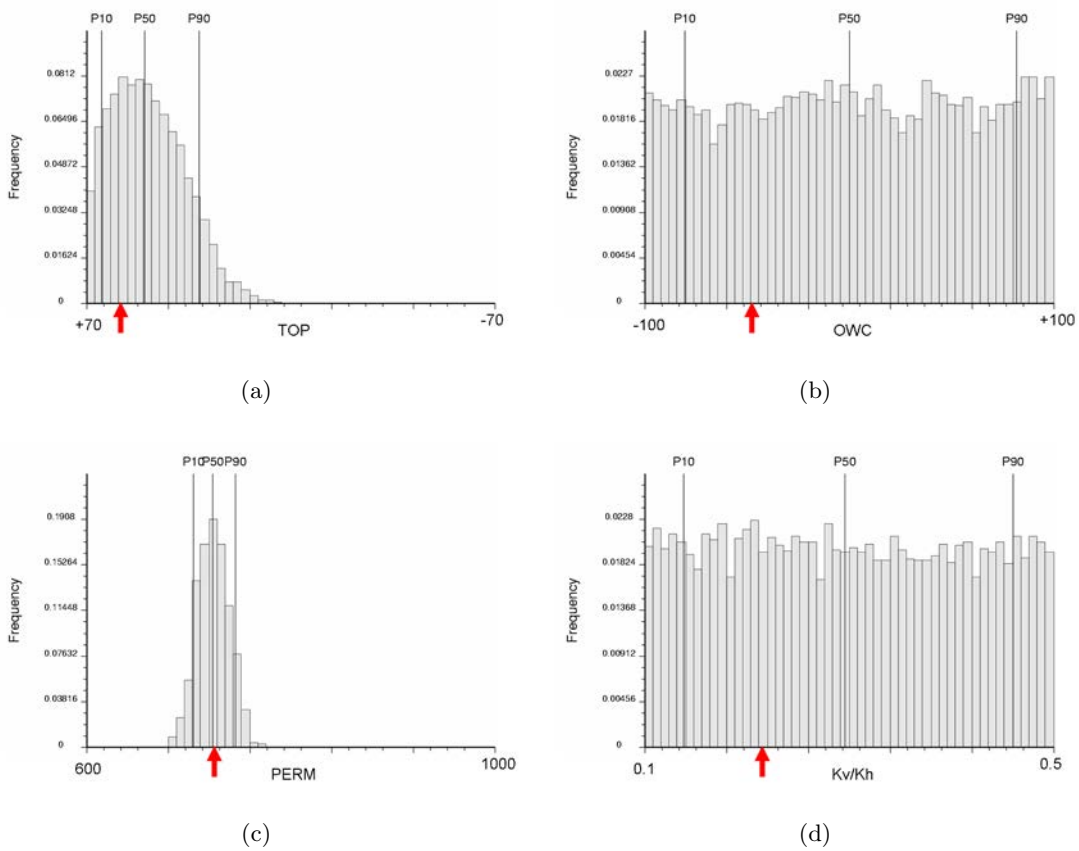


FIG. 2.18 – Histogrammes marginaux a posteriori des paramètres incertains du modèle.

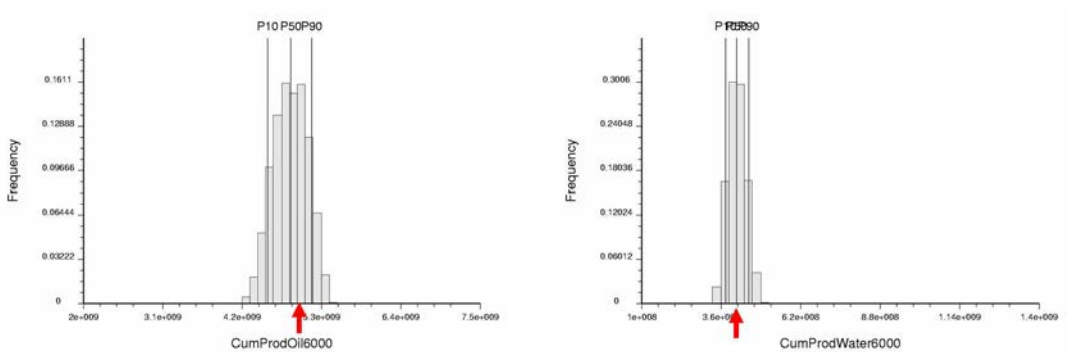


FIG. 2.19 – Prédictions de production à 6000 jours à partir des échantillons de modèles vraisemblables obtenus lors de l'inversion bayésienne.

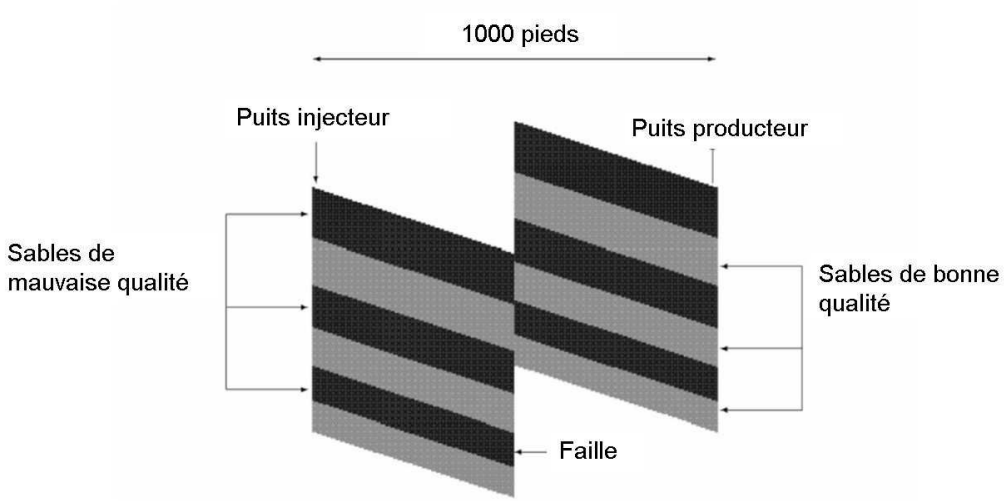


FIG. 2.20 – Modèle de réservoir utilisé dans le cas d'étude "*IC Fault*", d'après Carter et al. [2006]. Le modèle représente un réservoir faillé composé d'une alternance de sables de bonne et mauvaise qualité. Noter, en particulier, la diminution de l'épaisseur des couches depuis le toit vers le mur.

production (à comparer aux résultats montrés figure (2.13)). En outre, les valeurs de production (en huile et en eau) correspondant au modèle de référence sont bien dans les intervalles P_{10} - P_{90} prédits après l'inversion.

En conclusion, l'inversion bayésienne est une technique d'analyse statistique qui permet bien de réduire l'incertitude sur les paramètres incertains mais aussi sur les prédictions de production en tenant compte d'un historique de production. Les résultats obtenus ici sont en accord avec ceux précédemment décrits. Là encore, l'estimation des distributions a posteriori repose sur l'utilisation de technique d'échantillonnage de type Monte-Carlo. Cette technique est donc parfaitement adaptée à la gestion de surfaces de réponse de forme quelconque.

2.6.2 Mesures de prédictivité des méthodes de construction sur le modèle "*IC Fault*"

Présentation du modèle

Le modèle "*IC Fault*" représente la section d'un réservoir sableux traversée par une faille verticale (figure (2.20)). Il est composé de six couches de sable de bonne et mauvaise qualité alternant verticalement. On suppose que les couches sont homogènes et que chaque type de sable peut être décrit par une unique valeur de porosité et de perméabilité. L'épaisseur des couches diminue avec la profondeur depuis 12.5 pieds (toit du réservoir) à 7.5 pieds (au mur) pour une épaisseur totale de 60 pieds. Enfin, la longueur totale du réservoir est de 1000 pieds. Ce modèle simplifié représente un environnement de type deltaïque marin peu profond. L'intérêt de la faille est qu'elle change la continuité des différents corps sableux et donc la manière dont les fluides se

déplacent dans le réservoir. Le modèle utilisé pour la simulation des écoulements est composé de 100×12 cellules de telle façon que chaque unité de sable soit divisée en deux couches d'épaisseur égale et que la longueur de chaque bloc vaille 10 pieds.

Le modèle est initialement rempli d'huile avec cependant une saturation résiduelle en eau de 30%. Deux puits sont situés aux deux extrémités du réservoir : l'injection d'eau s'effectue à l'extrémité gauche du modèle tandis que le puits producteur est situé à droite. Les deux puits traversent l'ensemble du réservoir et sont simplement contrôlés en pression.

Paramètres incertains et plan d'expérience

Trois paramètres du modèle sont jugés incertains :

- le rejet de faille compris entre 0 et 60 pieds ;
- la perméabilité des couches de bonne qualité comprise entre 100 et 200 mD ;
- la perméabilité des couches de mauvaise qualité entre 0 et 50 mD.

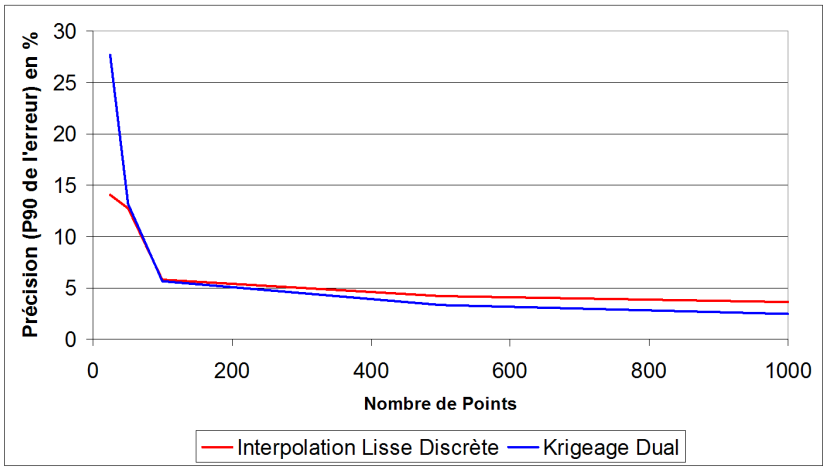
Les variables cibles des simulations d'écoulement sont les débits de production en huile et en eau au puits producteur ainsi que le débit d'injection d'eau. Les simulations ont été conduites sur 36 mois avec un enregistrement des résultats tous les mois. Ce modèle est initialement dédié à l'étude de procédures de calage d'historique [Bush et Carter, 1996; Carter, 2004; Carter et al., 2006]. On dispose donc d'un plan d'expérience intensif du domaine d'incertitude composé de 159 645 simulations d'écoulement. Comme montré lors des études précédentes, ce modèle se caractérise par un comportement fortement non linéaire et cyclique de la production vis-à-vis des paramètres incertains et notamment du rejet de la faille. Par exemple, Bush et Carter [1996] puis Carter et al. [2006] ont démontré la difficulté d'obtenir un modèle calé à l'historique en se basant sur des algorithmes classiques d'optimisation sur ce modèle.

Évaluation des méthodes de construction de surfaces de réponse

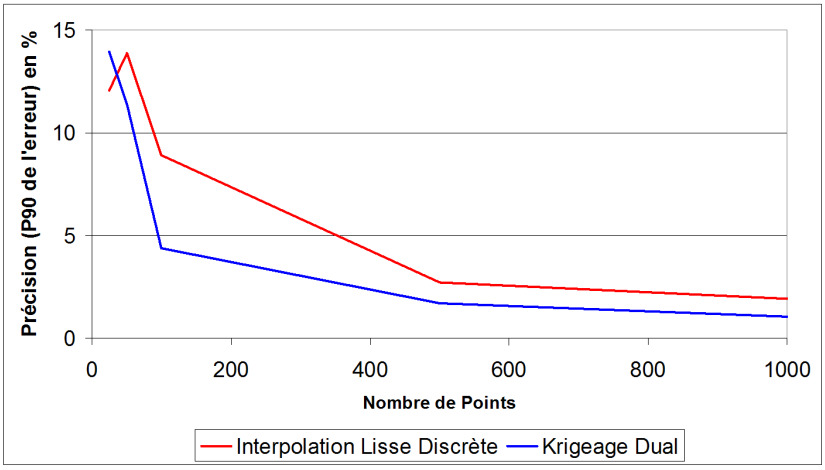
Le modèle "*IC Fault*" a pour objectif de permettre l'évaluation de la qualité des surfaces de réponse construites à l'aide des méthodes décrites précédemment : le krigeage dual (avec, ici, un variogramme de type puissance ($\omega = 1.75$)) et l'interpolation lisse discrète. sur une grille de 64 000 nœuds. L'approche choisie pour mesurer la qualité de l'interpolation des surfaces de réponse comporte deux étapes :

- un premier ensemble de modèles est tiré aléatoirement parmi les 159 645 simulations disponibles et est utilisé pour la construction des surfaces de réponse. Afin d'étudier la qualité des méthodes d'interpolation lorsque le nombre de points de donnée diminue, la taille de cet ensemble varie de 1000 points à 25 points seulement ;
- un second ensemble de 500 points est sélectionné pour estimer la prédictivité des surfaces. La mesure choisie est le P_{90} de l'erreur égale à :

$$Q_i = \left(\frac{z(\mathbf{p}_i) - \hat{z}(\mathbf{p}_i)}{z(\mathbf{p}_i)} \right)^2 \quad \forall i \in \{1 \dots 500\}, \tag{2.64}$$



(a) Fraction d'eau dans les fluides en production



(b) Débit de production en huile

FIG. 2.21 – Évolution de la prédictivité des surfaces de réponse construites par la méthode de l'interpolation lisse discrète et du krigeage dual en fonction du nombre de points pour le cas d'étude "IC Fault". La mesure de prédictivité utilisé est le P_{90} de l'erreur d'estimation au carré et calculée sur 500 points de test.

où $z(\mathbf{p}_i)$ représente la valeur réelle de production au point $\mathbf{p}_i$ tandis que $\hat{z}(\mathbf{p}_i)$ indique une estimation par le biais de la surface de réponse. Ce choix permet de rendre compte des points où l'erreur de prédiction est grande tout en évitant les points extrêmes qui, s'ils ne sont pas inclus dans les données lors la construction de la surface de réponse, ne peuvent être reproduits.

La figure (2.21) montre l'évolution de cette mesure de prédictivité en fonction du nombre de points disponibles pour les deux méthodes proposées. Les deux variables étudiées sont la fraction d'eau dans les fluides en production et le débit de production en huile. Ces résultats démontrent

les bonnes capacités des deux méthodes à interpoler une variable de production au comportement non linéaire dans un espace de dimension quelconque (ici n=3). En outre, ils démontrent aussi que les deux approches développées ici produisent des surfaces de réponse de qualité semblable.

2.6.3 Évaluation de la performance des méthodes de construction

Ce cas d'étude reprend le modèle d'anticlinal utilisé dans la section (2.6.1) avec quelques différences. L'objectif est de tester le temps de calcul requis pour l'interpolation d'une fonction de réponse par les approches de l'interpolation lisse discrète et du krigeage dual. L'évolution de ce temps est mesurée en fonction de la dimension de l'espace d'interpolation mais aussi du nombre de points de donnée disponibles. Le modèle initial est alors étendu à neuf paramètres incertains introduits un à un suivant l'ordre du tableau (2.4). Le plan d'expérience consiste en 2500 modèles ou combinaisons de paramètres incertains sélectionnés suivant une approche uniforme [Fang et al., 2000]. De même que précédemment, l'écoulement des fluides dans ces modèles est simulé sur 6000 jours. La variable cible est ici uniquement le débit en eau au puits producteur à la fin de la simulation.

Deux cas sont considérés : tout d'abord, le nombre de points de donnée disponibles est considéré fixe à 500 tandis que le nombre de paramètres incertains varie de 3 à 9. Dans un second temps, l'impact du changement du nombre de points de donnée (de 25 à 2500) est évalué tandis que le nombre de paramètres est fixé à 4. Dans les deux cas, le temps mesuré est celui nécessaire pour (1) construire l'estimateur de la fonction de réponse et (2) évaluer la réponse sur un échantillon aléatoire de 10 000 points. Pour la méthode de l'interpolation lisse discrète, le modèle discret sous-jacent contient 10 nœuds dans toutes les directions de l'espace tandis que dans le cas du krigeage dual, le variogramme choisi est de type puissance avec $\omega = 1.75$. Les résultats rassemblés tableaux (2.5) et (2.6) indiquent que :

- la performance de l'interpolation lisse discrète dépend directement de la taille du modèle discret et donc de la taille de l'espace sous-jacent, mais reste relativement insensible au nombre de points disponibles ;
- inversement, la performance du krigeage dual est invariante vis-à-vis du nombre de paramètres mais diminue avec l'augmentation du nombre de points de donnée disponibles.

Ces résultats sont cohérents avec la taille des systèmes linéaires (2.5) et (2.22) mis en jeu dans les deux approches. Aussi, l'utilisation de l'interpolation lisse discrète se justifie pour un système pour lequel le modèle discret est de taille limitée et/ou le nombre de points de donnée est grand, et inversement pour le krigeage.

2.7 Conclusions

Dans ce chapitre, nous avons présenté une chaîne d'opérations basée sur l'utilisation de surfaces de réponse pour analyser la relation existant entre la production d'un réservoir à hydrocarbure et un ensemble de paramètres incertains. A cette fin, durant ces travaux de thèse, nous avons spécifiquement :

Paramètre incertain	Valeur minimale	Valeur maximale	Nombre de niveaux
Géométrie			
TOP	-25 m	+25 m	5
BOTTOM	-25 m	+25 m	5
OWC	-3850 pieds	-3750 pieds	5
Propriété			
PERM	100	1300	5
KV/Kh	0.001	0.1	5
PORO	0.15	0.25	5
Fluide			
SWC	0	0.2	5
ViscOil	0.9	1.7	5
DensOil	47	57	5

Tab. 2.4 – Paramètres incertains du modèle anticlinal à 9 paramètres

Nombre de paramètres	3	4	5	6	7	8	9
Krigeage dual	2.985	2.078	2.469	2.235	2.703	2.312	3.188
Interpolation lisse discrète	0.39	4.656	39.453	1078.45	-	-	-

Tab. 2.5 – Temps de calculs (en secondes) nécessaire à la construction de la surface de réponse du débit de production en huile à 6000 jours en fonction du nombre de paramètres incertains. Dans ce cas, on dispose de 500 points de données et le temps indiqué correspond au temps requis pour estimer la réponse sur un échantillon de 10000 points.

Nombre de points	25	50	100	500	1000	2000	2500
Krigeage dual	0.141	0.281	0.547	2.640	6.032	16.187	23.078
Interpolation lisse discrète	4.094	4.318	4.234	4.406	4.765	4.421	4.875

Tab. 2.6 – Temps de calculs (en secondes) nécessaire à la construction de la surface de réponse du débit de production en huile à 6000 jours en fonction du nombre de points de données disponibles. Le nombre de paramètres incertains est de 4 et le temps indiqué correspond au temps requis pour estimer la réponse sur un échantillon de 10000 points.

- développé deux approches alternatives de construction d'une surface de réponse, l'une est basée sur la technique de l'interpolation lisse discrète et l'autre sur le krigeage dual ;
- sélectionné différentes approches d'échantillonnage de type Monte-Carlo pour quantifier l'incertitude sur les prévisions de production ;
- sélectionné un schéma d'analyse de la sensibilité de la production vis-à-vis des paramètres incertains ;

– et sélectionné la technique de l'inversion bayésienne pour pouvoir tenir compte de données historiques de production et réduire l'incertitude existant sur les paramètres incertains ainsi que les prévisions de production.

Le choix de l'ensemble des techniques s'est concentré sur les trois aspects suivants : (1) le nombre important de paramètres incertains (au moins n $>$ 3) qui doivent être pris en compte, (2) l'existence de bruit dans les données provenant le plus souvent de paramètres jugés peu influents et négligés et (3) la forme complexe de la relation entre les paramètres incertains et la production du réservoir. Pour cela, la plupart des méthode se basent sur des techniques d'échantillonnage de type Monte-Carlo. L'ensemble de ces méthodes et les structures de données nécessaires ont été implantées sous la forme de plusieurs librairies au sein du géomodeleur Gocad. Ce développement s'est effectué dans le langage informatique C++ et l'accent a été mis sur (1) l'efficacité, à la fois en terme de temps mais aussi de coût mémoire, des algorithmes et des structures de données, et sur (2) la réutilisation du code de calcul pour le développement de nouvelles méthodes et applications telles que celles décrites dans les chapitres suivants.

L'évaluation de la construction de surface de réponse à l'aide de l'interpolateur lisse discret constitue une part essentielle de ce travail. Cette méthode repose sur l'interpolation au sens des moindre carrés d'une fonction définie aux nœuds d'un modèle discret tout en respectant un ensemble de contraintes. Le terme contrainte signifiant tout type de relation linéaire entre les valeurs de la fonction. Cette approche a été sélectionnée d'une part pour sa capacité à intégrer suivant un formalisme mathématique simple un grand nombre de données de nature et de qualité hétérogènes que ce soit des données de type flou, dur ou des relations d'inégalité, et d'autre part à représenter des comportements ou phénomènes non linéaires. L'implantation de cette méthode a nécessité le développement au sein du géomodeleur Gocad d'un ensemble de structures de données (représentant soit les données elles-mêmes ou bien le modèle discret support de l'information) et des contraintes spécifiques au problème posé. Différents exemples ont montré les possibilités de cette approche. Néanmoins comme souligné sur le dernier exemple, la caractéristique de cette méthode est son coût en calcul qui dépend largement de la taille du modèle discret sous-jacent mais est peu sensible à la quantité de données à intégrer. Or, d'une part, la taille du modèle discret augmente de façon exponentielle avec la dimension de l'espace des paramètres incertains et, d'autre part, le nombre de données dans le cadre des simulations d'écoulement reste toujours limité. De ce fait, si la technique de l'interpolation lisse discrète permet, comme nous l'avons montré, de construire effectivement des fonctions de réponse elle n'est pas applicable pratiquement dans le cas général.

C'est pourquoi nous proposons et développons une méthode alternative d'interpolation. Celle-ci est basée sur le formalisme dual du krigeage auquel un terme de régularisation est ajouté. L'intérêt de cette technique est qu'elle présente des résultats similaires à l'algorithme l'interpolation lisse discrète mais surtout que son coût en calculs dépend plus du nombre de points de données que du nombre de paramètre incertains. Cette technique est donc parfaitement adaptée aux cas où le nombre de paramètres incertains est important. Dans la suite (chapitres 3 et 4), nous nous intéresserons donc à développer des méthodes de construction pouvant utiliser indifféremment les deux techniques. L'exception sera la présentation d'une méthode de lissage d'histogramme (chapitre 5) qui nécessite un certain nombre de contraintes spécifiques à DSI.

Chapitre 3

Prise en compte de données secondaires

Sommaire

3.1 Approche proposée 62
3.2 Planification d'expérience pour l'intégration d'une variable secondaire 63
3.3 Méthodes pour la construction de surfaces de réponse en présence de donnée secondaire 65
3.3.1 Contraintes pour l'interpolation lisse discrète 65
3.3.2 Krigeage avec dérive externe 68
3.4 Quantification de l'apport d'information 71
3.4.1 Mesure de "*Root Mean Square Error*" 72
3.4.2 Validation 72
3.5 Cas d'étude basé sur le modèle de réservoir SPE10 74
3.5.1 Présentation du modèle 74
3.5.2 Construction de surface de réponse du débit en huile après 500 jours . . 76
3.5.3 Analyse et discussion des résultats 78
3.6 Conclusions 80

Un frein majeur à l'analyse du comportement de la production d'un réservoir pétrolier vis-à-vis de paramètres incertains reste le coût en temps d'une simulation d'écoulement. En effet, il n'est possible de prendre en compte qu'un nombre limité de modèles (un modèle correspondant à une combinaison donnée de paramètres incertains) ce qui laisse l'espace d'incertitude vide de données. Ce chapitre présente une nouvelle approche de construction de surface de réponse dont le but est de compléter les résultats de simulations conventionnelles (encore appelées variables ou données primaires) par ceux de simulations rapides alternatives (variables ou données secondaires). Ces simulations alternatives sont par exemple la mesure de la connectivité du réservoir, la simulation sur ligne de courant ou bien la simulation d'un traceur chimique, etc. L'objectif est de réduire considérablement le temps nécessaire à la réalisation d'une simulation et ainsi d'être en mesure de prendre en compte plus de modèles. Cela permet d'améliorer la prédictivité des

surfaces de réponse en utilisant la variable secondaire pour définir la forme de la réponse entre les points correspondant à la variable primaire.

Après une description plus détaillée de l'approche proposée, nous présenterons les outils développés et implantés durant ces travaux de thèse pour traiter ce problème spécifique :

1. un plan d'expérience basé sur les hypercubes latins afin de planifier de façon optimale à la fois les simulations conventionnelles et alternatives ;
2. l'extension des méthodes de construction de surfaces de réponse décrites dans le chapitre précédent de telle façon qu'elles puissent prendre en compte une éventuelle donnée secondaire ;
3. l'utilisation de la validation croisée pour juger et discuter de l'utilité de l'intégration des données secondaires.

3.1 Approche proposée

L'analyse des incertitudes liées aux écoulements fluides dans un réservoir pétrolier souffre d'un problème de sous-échantillonnage de l'espace des paramètres incertains. Cela conduit, le plus souvent, à ne prendre en compte que les tendances linéaires de la réponse. Or, négliger les courbures et non-linéarités peut avoir un impact majeur sur l'analyse des variables de production.

Comme décrit dans le chapitre introductif, deux approches peuvent être considérées pour surmonter ce problème de sous-échantillonnage :

- la première consiste à utiliser un plan d'expérience combiné à une surface de réponse [Box et Draper, 1987]. La planification d'expérience sert à optimiser la sélection des modèles qui doivent être simulés de manière à en extraire le maximum d'information. Ses principaux avantages sont qu'elle est basée sur plusieurs techniques d'analyse robustes et qu'elle permet de conserver une description du comportement physique des fluides aussi complexe que nécessaire.
- l'alternative consiste à négliger une partie du comportement physique des fluides lors des simulations d'écoulement voire même à ne considérer que de simples mesures statiques liées à la production. Cette approche réduit considérablement le temps nécessaire pour réaliser une simulation et permet de prendre en compte un plus grand nombre de modèles. Les simulations simplifiées sont, par exemple, le calcul de mesures de connectivité du réservoir [Deutsch, 1996; Hird et Dubrule, 1995], l'utilisation de la théorie de la percolation [King et al., 2001], la simulation sur lignes de courant [Idrobo et al., 2000; Ates, 2005], la simulation de traceur chimique [Saad et al., 1996], etc. Cette seconde approche est utile pour diriger l'interprétation des écoulements dans un réservoir, identifier les mécanismes de production ou bien classer les modèles suivant leur potentiel de production (procédure appelée "*model ranking*"), etc. Cependant, cette approche est purement qualitative et ne permet pas une quantification précise des incertitudes sur la performance d'un réservoir.

Nous proposons de combiner les deux approches en utilisant les résultats de simulations alternatives comme une donnée secondaire lors de la construction d'une surface de réponse. En quelque

sorte, cette donnée secondaire définit la forme globale de la surface qui est alors mise à l'échelle par les résultats de simulations conventionnelles.

L'intégration de variables secondaires est un problème récurrent en géologie pétrolière et ingénierie de réservoir. Un exemple classique est l'utilisation de données sismiques lors de la simulation des propriétés pétrophysiques du sous-sol (porosité, perméabilité) à partir des puits. La sismique est une source de faible résolution mais couvre l'ensemble du volume du réservoir, par opposition aux données de puits qui sont beaucoup plus précises mais ponctuelles. De la même manière, dans ce travail, l'idée est d'utiliser les simulations alternatives comme une source exhaustive d'information secondaire. Exhaustive signifie que cette donnée est connue au niveau des mêmes points que la variable primaire mais aussi de tous ceux où une interpolation doit être réalisée [Goovaerts, 1997]. Dans le cas où cette information n'est pas exhaustive, une approximation raisonnable consiste à utiliser une seconde surface de réponse. Pour cela nous proposons la procédure suivante :

1. réalisation de simulations d'écoulement conventionnelles ;
2. réalisation de simulations alternatives suivant un plan d'expérience qui, pour des raisons de calibration inclus, les modèles déjà considérés à l'étape 1 ;
3. construction de la surface de réponse correspondant à la variable secondaire ;
4. construction de la surface de réponse de la variable primaire en tenant compte de la précédente.

3.2 Planification d'expérience pour l'intégration d'une variable secondaire

Le plan d'expérience choisi dans ces travaux pour sélectionner les modèles à simuler est basé sur l'approche des hypercubes latins. La construction de ce type de plan s'effectue par ordinateur telle que la position de ces modèles dans l'espace soit optimisée par itérations successives [McKay et al., 1979; Iman et Conover, 1980]. Différents critères d'optimisation sont décrits dans Box et Draper [1987]; Morris et Mitchell [1995] ou Bursztyn et Steinberg [2004], ils correspondent le plus souvent à une mesure de la distance moyenne entre les modèles. En pratique, un hypercube latin est une matrice de taille $m \times n$, où n est le nombre de paramètres incertains et m le nombre de simulations à réaliser, librement fixé à la création du plan tel que : $m > n$. Enfin, dans cette cette matrice, les colonnes sont toutes définies comme une permutation de l'ensemble $\{1, \ldots, m\}$ (figure (3.1)). Les principales caractéristiques de cette planification sont que :

- le nombre de niveaux par paramètre incertain est identique au nombre de simulations à réaliser, c'est à dire à m ;
- quelles que soient les simulations, elles ne partagent jamais les mêmes valeurs (ou niveaux) de paramètre incertains.

Ce type de plan peut être généré suivant un processus entièrement aléatoire, cependant, il est souvent nécessaire de l'optimiser. Cela repose sur l'utilisation d'algorithme de recuit simulé tel que l'algorithme "*columnwise-pairwise*" proposé dans Liefvendahl et Stocki [2004]. Ce type de

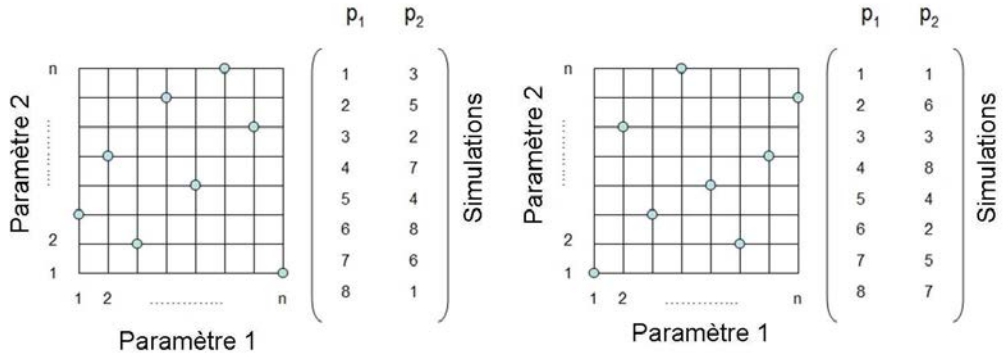
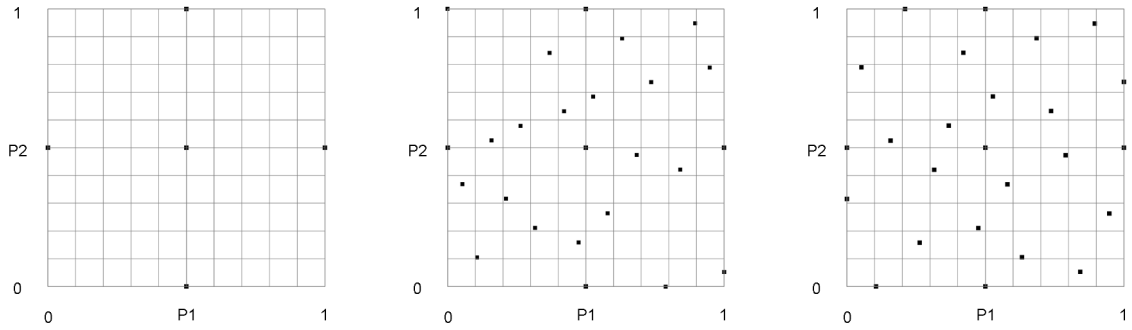


FIG. 3.1 – Exemples de plans de type hypercube latin et les matrices correspondantes avec 2 paramètres incertains et 8 simulations.



(a) Planification des simulation conventionnelles

(b) Planification des simulation alternatives avant optimisation

(c) Planification optimisée des simulations alternatives

FIG. 3.2 – Exemple de construction de plan d'expérience pour deux paramètres incertains avec l'approche proposée : (a) plan pour les simulations primaires (plan de type Box-Benkhen [Box et Draper, 1987]), (b) plan pour les simulations alternatives, de type hypercube latin avec 20 nouvelles simulations avant optimisation et (c) après optimisation suivant le critère de "force répulsive" proposé dans Bursztyn et Steinberg [2004]. A noter l'inclusion des simulations conventionnelles dans la planification des simulations alternatives.

plan assure ainsi la couverture la plus uniforme possible du domaine d'incertitude tout en laissant la liberté de fixer le nombre de simulations.

Ici, la planification d'expérience par hypercube latin a été modifiée pour inclure de façon optimale à la fois les simulations primaires et secondaires. En pratique, deux plans sont successivement construits (figure 3.2) :

– le premier décrit les modèles à simuler de manière conventionnelle et est le plus souvent de type factoriel à deux niveaux, de type Placket-Burman ou composite central ;

– le second concerne les simulations alternatives et est un hypercube latin. En pratique, les modèles sélectionnés précédemment sont introduits dans la matrice mais considérés fixes durant la procédure d'optimisation.

Cette procédure assure que la planification finale soit optimale en terme de distribution des modèles à simuler, et que les deux plans aient un certain nombre de points en commun pour la calibration des résultats de simulation. De manière générale, ce plan a les caractéristiques favorables suivantes : (1) le nombre de modèles à simuler est contrôlable, (2) un grand nombre de niveaux sont considérés par paramètre et (3) l'ensemble de l'intervalle d'un paramètre donné est échantillonné de manière uniforme.

3.3 Méthodes pour la construction de surfaces de réponse en présence de donnée secondaire

Cette section détaille les techniques de construction envisagées pour prendre en compte une source d'information secondaire. Celles-ci sont des extensions de l'interpolation lisse discrète et du krigeage dual introduits dans le chapitre précédent. Ici, les résultats des simulations conventionnelles se présentent toujours sous la forme d'un ensemble de points plongés dans l'espace des paramètres incertains et auxquels est attaché un ensemble de propriétés correspondant aux résultats de simulation. À l'opposé, les simulations alternatives doivent être exhaustives. Pour cela leurs résultats sont modélisés sous la forme d'une surface de réponse obtenue suivant une approche classique.

3.3.1 Contraintes pour l'interpolation lisse discrète

Dans le cadre de l'interpolation lisse discrète, l'introduction d'une donnée secondaire passe par la définition de contraintes sur les dérivées premières (le gradient) et secondes de la fonction à interpoler. Ces contraintes sont détaillées dans les sections suivantes, où la fonction à interpoler sera notée classiquement φ, tandis que celle des données secondaires sera notée ψ.

Contraintes sur le gradient

D'une façon générale, deux gradients sont considérés similaires s'ils sont colinéaires, orientés dans la même direction et de normes identiques. Les deux premières caractéristiques reviennent à spécifier que les deux gradients en un nœud α donné, notés $\mathbf{grad}\varphi(\alpha)$ et $\mathbf{grad}\psi(\alpha)$ ont, respectivement :

– un produit vectoriel nul :

$$\mathbf{grad}\varphi(\alpha) \times \mathbf{grad}\psi(\alpha) = \mathbf{0} \tag{3.1}$$

– et un produit scalaire positif ou nul :

$$\mathbf{grad}\varphi(\alpha) \cdot \mathbf{grad}\psi(\alpha) \geq 0 \tag{3.2}$$

Ces équations peuvent être linéarisées et traduites sous forme d'un ensemble de n contraintes d'égalité et d'une contrainte d'inégalité. Ces contraintes sont appelées respectivement contrainte de gradient parallèle et contrainte de direction du gradient et sont décrites en détails dans Labat [2004]; Fetel [2005] et Kedzierski et al. [2005]. À l'opposé, l'égalité des normes ne peut pas être linéarisée, il n'est donc pas possible de l'écrire sous la forme d'une contrainte DSI.

Ainsi s'il est toujours possible d'utiliser le gradient pour contraindre la surface de réponse correspondant à la variable primaire par celle de la variable secondaire, deux problèmes se posent néanmoins : (1) l'utilisation de contrainte d'inégalité qui réduit l'efficacité des calculs et surtout (2) l'impossibilité de spécifier la norme du gradient qui empêche tout transfert d'information quantitative.

Contraintes sur la dérivée seconde

Le calcul et l'analyse de la dérivée seconde d'une fonction permet de décrire globalement sa forme, c'est-à-dire les zones concaves et convexes. Dans le cadre présent, cette dérivée peut être considérée suivant deux approches possibles : il s'agit soit de spécifier que les deux surfaces ont des dérivées secondes de même signe ce qui permet d'indiquer les mêmes zones concaves et convexes, soit qu'elles sont égales, ce qui permet de contraindre l'amplitude de la dérivée en plus de son signe.

Notations Afin de simplifier l'estimation de la dérivée seconde d'une surface de réponse donnée par rapport aux paramètre incertains, nous avons choisi de considérer les n différentes directions de l'espace indépendamment. En outre, considérant que le modèle discret support de l'interpolation est une grille structurée régulière, il est possible d'approcher chacune des n dérivées par un schéma de discrétisation de type centré. Ainsi, soit un nœud donné α, alors la dérivée seconde, notée $s_{\varphi,i}(\alpha)$, de la fonction φ (respectivement ψ) dans la direction $x_i, \forall i \in [1, \ldots, n]$, vaut :

$$\frac{\partial^2 \varphi}{\partial x_i^2} = s_{\varphi,i}(\alpha) \simeq \frac{1}{(\Delta u_i)^2} \left[\; 2\varphi(\alpha) \; - \; \left(\; \varphi(\alpha_{x_i}^+) + \varphi(\alpha_{x_i}^-) \; \right) \; \right] \tag{3.3}$$

où, pour la direction x_i considérée, Δu_i représente la taille du maillage tandis que $\alpha_{x_i}^+$ et $\alpha_{x_i}^-$ sont respectivement les voisins à droite et à gauche du nœud α. Cette approximation peut être réalisée dans l'ensemble du modèle sauf au niveau des frontières où aucune contrainte n'est alors considérée. Idéalement, il serait aussi nécessaire d'utiliser les termes croisés de la dérivée seconde, c'est-à-dire de la forme :

$$\frac{\partial^2 \varphi}{\partial x_i \partial x_j} \tag{3.4}$$

où x_i et x_j sont deux directions distinctes de l'espace. Cependant, le nombre de termes de ce type croît de façon exponentielle avec la dimension de l'espace ce qui nuit à l'efficacité des calculs.

Contrainte sur le signe de la dérivée seconde Comme décrit précédemment, le signe de la dérivée seconde permet de renseigner sur les zones concaves et convexes de la fonction à prendre

en compte. Pour cela, il suffit de spécifier une contrainte d'inégalité indiquant que le produit des deux dérivées dans la direction $x_i, \forall i \in [1, \ldots, n]$ choisie, est positif ou nul. Il vient alors :

$$s_{\varphi,i}(\alpha) \cdot s_{\psi,i}(\alpha) \geq 0 \quad \Leftrightarrow \quad \sum_{\beta \in \Omega} A_c(\beta) \cdot \varphi(\beta) \geq b_c, \quad \forall i \in [1, \ldots, n] \tag{3.5}$$

avec

$$\left| \begin{array}{rcccl} A_c(\beta) & = & s_{\psi,i}(\alpha) & \text{si} & \beta = \alpha \\ A_c(\beta) & = & -\frac{1}{2} s_{\psi,i}(\alpha) & \text{si} & \beta = \alpha_{x_i}^{+} \quad or \quad \alpha_{x_i}^{-} \\ A_c(\beta) & = & 0 & \text{sinon} & \\ b_c & = & 0 & & \end{array} \right. \tag{3.6}$$

Contrainte de dérivée seconde Cette contrainte spécifie l'égalité des dérivées des surfaces de réponse primaire (que l'on cherche à interpoler) et secondaire. Elle est plus générale que la précédente : celle-ci cherche à imposer non seulement le signe de la dérivée mais aussi son amplitude. Considérons une direction de l'espace donnée $x_i, \forall i \in [1, \ldots, n]$, cette relation peut s'écrire suivant le formalisme des contraintes DSI comme :

$$s_{\varphi,i}(\alpha) \simeq s_{\psi,i}(\alpha) \quad \Leftrightarrow \quad \sum_{\beta \in \Omega} A_c(\beta) \varphi(\beta) \simeq b_c, \quad \forall i \in [1, \ldots, n] \tag{3.7}$$

avec

$$\left| \begin{array}{rcccl} A_c(\beta) & = & 1 & \text{si} & \beta = \alpha \\ A_c(\beta) & = & -\frac{1}{2} & \text{si} & \beta = \alpha_{x_i}^{+} \quad or \quad \alpha_{x_i}^{-} \\ A_c(\beta) & = & 0 & \text{sinon} & \\ b_c & = & s_{\psi,i}(\alpha) & & \end{array} \right. \tag{3.8}$$

Également, il est important de noter que les résultats de simulations conventionnelles et alternatives sont rarement exprimés dans les mêmes unités. Cela peut biaiser l'estimation des dérivées secondes, ainsi il est préférable de procéder à une étape de normalisation avant l'imposition des contraintes soit en centrant et réduisant les différentes variables, ou bien en les mettant à l'échelle entre 0 et 1.

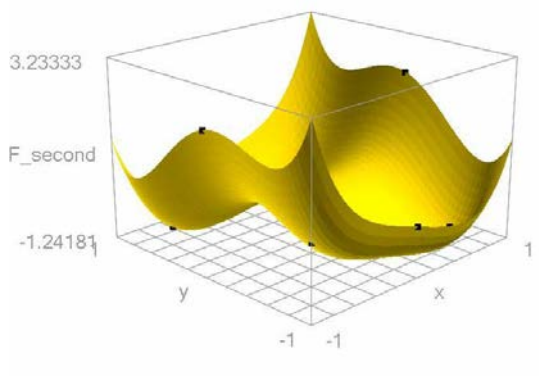
Exemple analytique

Afin d'illustrer l'approche proposée, le modèle analytique à deux paramètres (x et y) suivant est considéré :

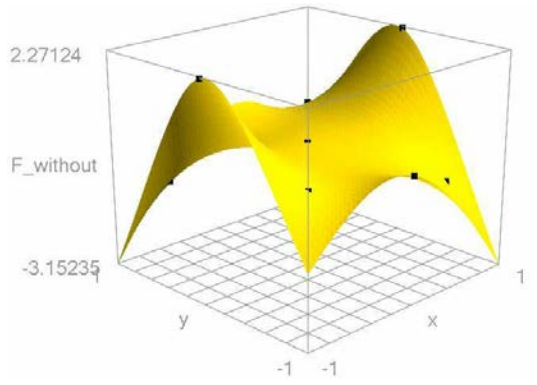
$$F(x, y) = 4x^4 - \frac{21}{10}x^4 + \frac{1}{3}x^6 + xy + 4y^2 + 4y^4 \tag{3.9}$$

Ce modèle proposé dans Scheidt et Zabalza-Mezghani [2004] est d'intérêt ici pour sa forme hautement non linéaire qui est difficilement capturée par une planification d'expérience classique à deux ou trois niveaux. Dans la suite :

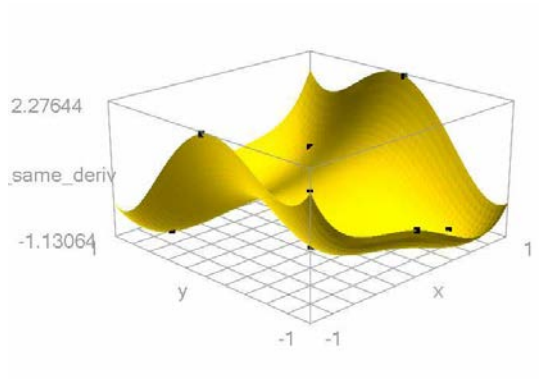
- les données primaires sont issues d'une planification d'expérience de type composite central, soit un total de 9 points ;
- la données secondaire est égale à la fonction elle-même.



(a) Fonction F(x, y) (équation (3.9))



(b) Interpolation lisse discrète en ne prenant en compte que les points de donnée primaire



(c) Interpolation lisse discrète avec contrainte de dérivée seconde

FIG. 3.3 – Exemple analytique d'application de l'interpolation lisse discrète en présence de données secondaires.

Ainsi, le résultat de l'interpolation devrait être lui aussi, in fine, égal à cette fonction.

La figure (3.3) montre les résultats obtenus par l'interpolation lisse discrète en tenant compte d'une contrainte sur la dérivée seconde, l'expérience ayant montré que les autres types de contraintes ne sont pas appropriés. Ces résultats illustrent l'intérêt de l'apport d'une donnée secondaire dans la construction. Les surfaces obtenues respectent les données primaires, mais en plus possèdent la forme de la surface de réponse correspondant à la variable secondaire. On peut cependant noter que par endroit les résultats de l'interpolation ne sont pas entièrement satisfaisants, en particulier loin des données primaires. Cela semblerait être dû à la contrainte de gradient constant qui est imposée comme contrainte de rugosité.

3.3.2 Krigeage avec dérive externe

De nombreuses variantes du krigeage prenant en compte une ou plusieurs données secondaires ont déjà été développées et proposées dans la littérature, voir par exemple Goovaerts [1997]. Parmi elles, la variante du krigeage avec dérive externe (ou KED, "*Kriging with External Drift*")

[Maréchal, 1984] a été choisie pour cette étude. En effet, cette approche considère que la moyenne de la fonction à interpoler suit une tendance locale calculée à partir de la donnée secondaire. En pratique, c'est une extension du krigeage avec tendance (ou KT, "*Kriging with Trend*") [Huijbregts et Matheron, 1971] pour lequel la tendance s'exprime en fonction de la position où s'effectue l'interpolation.

Comparé au krigeage simple avec moyenne locale changeante (ou SKlm, "*Simple Kriging with varying local means*"), l'avantage du KED est que la relation avec la donnée secondaire est une fonction linéaire dont les coefficients sont calculés lors de l'interpolation et non spécifiés à l'avance. D'autre part, le KED est plus satisfaisant que le co-krigeage [Myers, 1984] et ses variantes, car ces derniers nécessitent l'évaluation du coefficient de corrélation, voire du co-variogramme, entre les deux sources de donnée ce qui n'est pas toujours possible au vu du faible nombre de résultats de simulations conventionnelles.

Principes

Comme énoncé ci-dessus, le krigeage avec dérive externe suppose que la moyenne de la fonction à interpoler est une fonction de la donnée secondaire. On a ainsi, d'après Goovaerts [1997] :

$$m_Z(\mathbf{x}) \;=\; a_0(\mathbf{x}) \;+\; a_1(\mathbf{x}) y(\mathbf{x}) \tag{3.10}$$

où $m_Z(\mathbf{x})$ est la moyenne de la fonction à interpoler, $y(\mathbf{x})$ représente la donnée secondaire tandis que $a_0(\mathbf{x})$ et $a_1(\mathbf{x})$ sont deux coefficients à déterminer. Lors de l'interpolation, ces coefficients sont constants pour un voisinage donné mais peuvent changer dans le domaine. Ici, ils sont constants dans l'ensemble du domaine d'interpolation car le voisinage est considéré global. Comme montré dans Maréchal [1984]; Deutsch et Journel [1992] et Goovaerts [1997], le krigeage avec dérive externe sous la forme :

$$\hat{z}(\mathbf{x}) = \left[\boldsymbol{\lambda}^{KED}(\mathbf{x})\right]^t \; \mathbf{z}_p \tag{3.11}$$

avec les poids du krigeage $\boldsymbol{\lambda}^{KED}(\mathbf{x})$ solution du système linéaire à $n+2$ inconnues suivant :

$$\begin{bmatrix} C(\|\mathbf{p}_1 - \mathbf{p}_1\|) & \cdots & C(\|\mathbf{p}_1 - \mathbf{p}_{n_d}\|) & 1 & y(\mathbf{p}_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C(\|\mathbf{p}_{n_d} - \mathbf{p}_1\|) & \cdots & C(\|\mathbf{p}_{n_d} - \mathbf{p}_{n_d}\|) & 1 & y(\mathbf{p}_{n_d}) \\ 1 & \vdots & 1 & 0 & 0 \\ y(\mathbf{p}_1) & \vdots & y(\mathbf{p}_{n_d}) & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1^{KED}(\mathbf{x}) \\ \vdots \\ \lambda_{n_d}^{KED}(\mathbf{x}) \\ \mu_0^{KED} \\ \mu_1^{KED} \end{bmatrix} = \begin{bmatrix} C(\|\mathbf{p}_1 - \mathbf{x}\|) \\ \vdots \\ C(\|\mathbf{p}_{n_d} - \mathbf{x}\|) \\ 1 \\ y(\mathbf{x}) \end{bmatrix} \tag{3.12}$$

et se simplifie sous la forme :

$$\begin{bmatrix} \mathbf{C} & \mathbf{F} \\ \mathbf{F}^t & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\lambda}^{KED}(\mathbf{x}) \\ \boldsymbol{\mu}^{KED}(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \mathbf{c}(\mathbf{x}) \\ \mathbf{f}(\mathbf{x}) \end{bmatrix} \tag{3.13}$$

où $\mathbf{C}$ et $\mathbf{c}(\mathbf{x})$ sont les matrices de covariance déjà vues équation (2.6), $\boldsymbol{\mu}^{KED}(\mathbf{x})$ un ensemble de coefficients de Lagrange et $\mathbf{F}$ et $\mathbf{f}(\mathbf{x})$ sont deux matrices correspondant aux contraintes dues à la donnée secondaire.

Formalisme dual

Sachant que la matrice $\mathbf{C}$ est définie et positive, d'après Galli et Murillo [1984], il vient :

$$\begin{bmatrix} \mathbf{C} & \mathbf{F} \\ \mathbf{F}^t & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{U} & \mathbf{V} \\ \mathbf{V}^t & \mathbf{W} \end{bmatrix} \tag{3.14}$$

avec

$$\begin{cases} \mathbf{W} = -(\mathbf{F}^t\ \mathbf{K}^{-1}\ \mathbf{F})^{-1} \\ \mathbf{U} = \mathbf{K}^{-1}\ (\mathbf{I} - \mathbf{F}\ \mathbf{V}^t) \\ \mathbf{V}^t = -\mathbf{W}\ \mathbf{F}^t\ \mathbf{K}^{-1} \end{cases} \tag{3.15}$$

Ainsi, l'estimateur du krigeage devient :

$$\hat{z}(\mathbf{x}) = \mathbf{c}^t(\mathbf{x})\ \mathbf{U}\ \mathbf{z}_p + \mathbf{f}^{\,t}(\mathbf{x})\ \mathbf{V}^t\ \mathbf{z}_p \tag{3.16}$$

De la même façon qu'au chapitre précédent, $\mathbf{b} = \mathbf{U}\ \mathbf{z}_p$ et $\mathbf{d} = \mathbf{V}^t\ \mathbf{z}_p$ sont deux vecteurs indépendants des coordonnées du point $\mathbf{x}$ où se réalise l'interpolation. Le calcul de ces vecteurs peut s'effectuer à l'aide du système suivant, où $\mathbf{A}$ et $\mathbf{B}$ sont deux matrices quelconques [Galli et Murillo, 1984] :

$$\begin{bmatrix} \mathbf{b} \\ \mathbf{d} \end{bmatrix} = \begin{bmatrix} \mathbf{U} & \mathbf{A} \\ \mathbf{V}^t & \mathbf{B} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{z}_p \\ \mathbf{0} \end{bmatrix} \tag{3.17}$$

Sachant que $\mathbf{A}$ et $\mathbf{B}$ peuvent être choisies arbitrairement, prenons par exemple $\mathbf{A} = \mathbf{V}$ et $\mathbf{B} = \mathbf{W}$, on a ainsi :

$$\begin{bmatrix} \mathbf{b} \\ \mathbf{d} \end{bmatrix} = \begin{bmatrix} \mathbf{U} & \mathbf{V} \\ \mathbf{V}^t & \mathbf{W} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{z}_p \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \mathbf{F} \\ \mathbf{F}^t & \mathbf{0} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \mathbf{z}_p \\ \mathbf{0} \end{bmatrix} \tag{3.18}$$

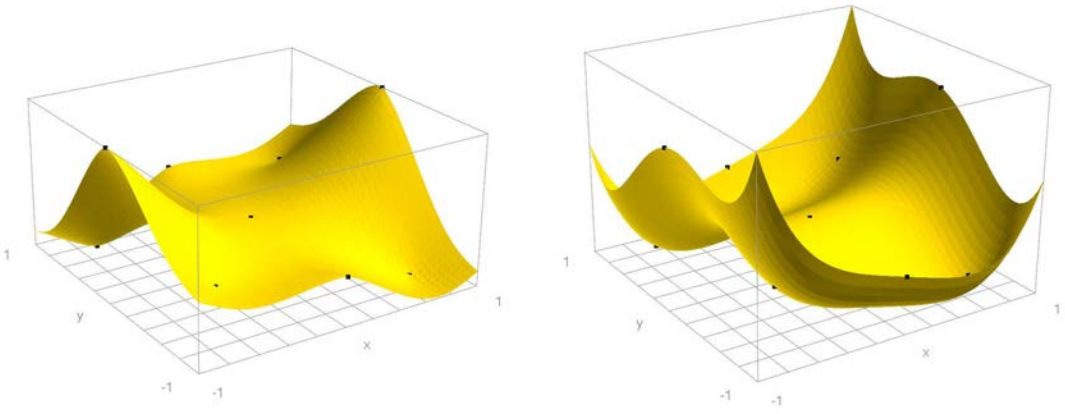
Par conséquent $\mathbf{b}$ et $\mathbf{c}$ sont les solutions du système linéaire suivant :

$$\begin{bmatrix} \mathbf{C} & \mathbf{F} \\ \mathbf{F}^t & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{b} \\ \mathbf{d} \end{bmatrix} = \begin{bmatrix} \mathbf{z}_p \\ \mathbf{0} \end{bmatrix} \tag{3.19}$$

Ce système ne dépend pas de la position où doit se dérouler l'interpolation, il n'est donc nécessaire de le résoudre qu'une seule fois pour un ensemble de points de donnée. Le second avantage de cette approche est qu'il n'est pas nécessaire de calculer explicitement l'inverse de la matrice du système de krigeage (équation (3.19)).

Exemple analytique

Le modèle analytique utilisé dans la section précédente (équation (3.9)) est aussi envisagée pour illustrer la technique du krigeage dual avec dérive externe. Les données primaires et secondaires sont, là encore, les mêmes. Les résultats obtenus après interpolation sont exposés figure (3.4). Ils montrent que, comme attendu, la solution finale est identique à la fonction d'entrée F(x,y).



(a) Interpolation avec un Krigeage simple

(b) Interpolation avec un Krigeage avec dérive externe

FIG. 3.4 – Comparaison des résultats d'interpolation sans tenir compte des données secondaires (krigeage simple) et avec données secondaire (krigeage avec dérive externe). Le variogramme choisi est de type puissance avec $\omega = 1.75$.

Ces résultats sont clairement satisfaisants et bien meilleurs que ceux obtenus suivant l'interpolation lisse discrète. La différence majeure vient certainement du fait que le krigeage dual s'appuie directement sur les données tandis que l'interpolation lisse discrète s'appuie sur la dérivée seconde. Or il existe un nombre infini de fonctions qui partagent la même dérivée seconde. En outre, l'interpolation lisse discrète considère une contrainte supplémentaire de rugosité. A l'opposé, le krigeage dual n'a pas nécessairement les capacités d'intégration et de compromis qu'a l'interpolation lisse discrète, en particulier, dans les cas où les données primaires et secondaires seraient contradictoires.

3.4 Quantification de l'apport d'information

Un problème essentiel de l'approche reste la fiabilité de la donnée secondaire choisie. En effet, dans le cadre d'une étude de cas réel, il n'est pas possible de disposer d'une surface de réponse de référence pour juger de la qualité de l'interpolation. Or, comme montré sur les exemples précédents et sachant que le nombre de points correspondant à la donnée primaire est nécessairement petit, l'impact de la variable secondaire est important. Ainsi, une méthodologie convaincante et applicable doit permettre de répondre aux deux questions suivantes :

- Parmi toutes les simulations alternatives ou données secondaires possibles, quelle est la plus appropriée pour le problème étudié ?
- Comment déterminer si l'information secondaire considérée améliore effectivement la qualité des prédictions ?

3.4.1 Mesure de "*Root Mean Square Error*"

Pour répondre à ces questions, nous avons développé une approche basée sur la technique de la validation croisée (ou "*cross-validation*"). Cette méthode de validation statistique qui consiste à partitionner les données en deux sous-ensembles : le premier sert dans la construction de la surface de réponse, tandis que le second est préservé afin de confirmer et valider la qualité de l'interpolation. Les deux ensembles sont, de façon générale, appelés respectivement ensemble d'entraînement (ou "*training set*") et ensemble de test (ou "*testing set*"). Différentes variantes de la validation croisée existent suivant la taille de l'ensemble de test, que la validation soit répétée avec différentes partitions et qu'il y ait remise ou non des points lors du calcul d'une nouvelle partition. La validation permet alors d'obtenir une mesure de l'erreur de prédiction pour chacun des points de donnée. Ces erreurs sont exprimées ensuite sous la forme d'une mesure globale comme, par exemple, la racine carré de la moyenne des erreurs de prédiction au carré (ou RMSE, "*Root Mean Square Error*") afin de juger de la qualité de la surface de réponse.

L'analyse proposée consiste à effectuer une validation croisée des surfaces de réponse construites avec et sans la ou les donnée(s) secondaire(s) et de comparer les valeurs de prédictivitée. Ainsi, il est attendu que :

- si une donnée secondaire n'apporte pas d'information, la RMSE ne devrait pas changer ;
- si l'information apportée est contradictoire, cette valeur devrait augmenter ;
- si la donnée est informative, cette valeur devrait diminuer.

Dans notre cas, la méthode de validation croisée est de type "jackkniffe", dans laquelle à chaque itération seul un point du jeu de donnée est retiré. Tandis que la mesure de prédictivité du modèle est basée sur le RMSE.

3.4.2 Validation

Afin de tester cette technique d'analyse, la fonction F(x,y), (équation (3.9)) est de nouveau choisie. Ici :

- les données primaires sont reparties suivant un hypercube latin comprenant 20 points où la fonction F(x,y) est évaluée ;
- la donnée secondaire est toujours une surface de réponse égale à la fonction F(x,y) elle-même.

Ensuite, deux cas d'étude sont successivement considérés :

- le premier suit une approche classique et ne prend pas en compte la donnée secondaire ;
- l'autre suit la méthode proposée. En outre, afin de simuler différentes sources de données alternatives de qualité diverses, un bruit aléatoire est ajouté à la surface de réponse correspondant aux données secondaires. Ce bruit est généré suivant une loi uniforme dont l'amplitude varie de $\pm$ 0. (i.e. aucun bruit) à $\pm$ 1.5 (figure (3.6)).

La méthode de construction est l'interpolation lisse discrète avec contrainte de gradient constant comme rugosité et une contrainte de dérivée seconde. L'évolution de la mesure RMSE de la validation croisée en fonction de l'amplitude du bruit est rapportée table (3.7). Ces résultats montrent que cette technique permet de quantifier l'apport des données secondaires dans la

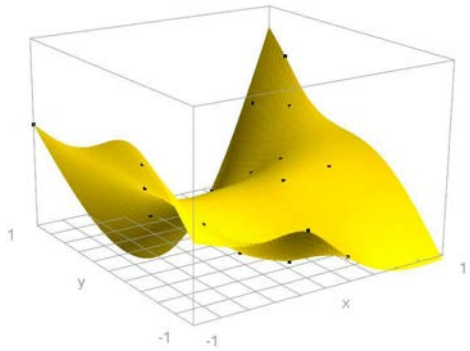
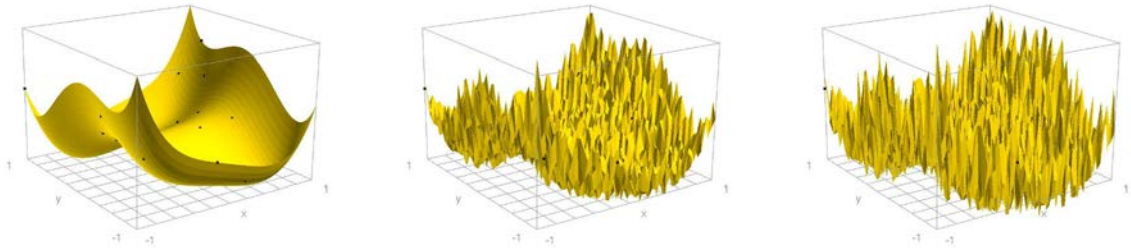


FIG. 3.5 – Résultat de l'interpolation lisse discrète dans le cas où aucune donnée secondaire n'est considérée, la mesure de RMSE vaut alors 0,625.



(a) Amplitude du bruit : 0 (b) Amplitude du bruit : 0.5 (c) Amplitude du bruit : 1

FIG. 3.6 – Exemples de données secondaires plus ou moins bruitées. Les surfaces de réponse construites avec ces données ont pour valeur de RMSE : 0.492, 0.582 et 0.692, respectivement.

Coeff. Bruit	0.	0.25	0.5	0.75	1.0	1.25	1.5
RMSE	0.492	0.527	0.582	0.639	0.692	0.741	0.785

FIG. 3.7 – Évolution de la mesure RMSE en fonction de l'amplitude du bruit ajouté à la variable secondaire.

construction des surfaces de réponse : la prédictivité du modèle (c'est-à-dire l'inverse de la mesure RSME) augmente avec leur prise en compte (la RMSE passe de 0.625 à 0.492). Cette technique de mesure de prédictivité semble en outre être capable de distinguer si les données secondaires sont bruitées ou non : la RMSE augmente de 0.492, s'il n'y a pas de bruit, à 0.785 dans le cas contraire.

Cette technique semble donc bien efficace pour mesurer la prédictivité d'une surface de réponse ainsi que l'apport des données secondaires. Cependant, deux problèmes se posent : (1) le nombre de points où est connue la donnée primaire est limité, ce qui nuit à la robustesse de la mesure RSME et, (2) cette approche n'inclut, pour le moment, aucun test statistique pour comparer les mesures de RMSE obtenues et permettre de déterminer les valeurs significativement différentes.

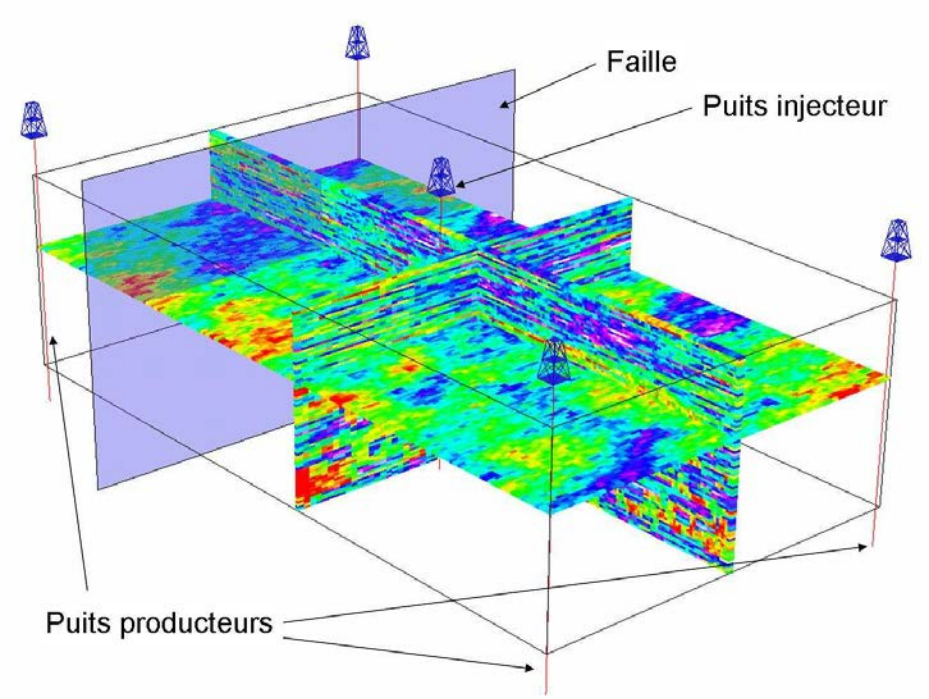


FIG. 3.8 – Version modifiée du modèle de réservoir SPE 10.

3.5 Cas d'étude basé sur le modèle de réservoir SPE10

Le modèle SPE10 a été choisi pour illustrer et discuter l'approche proposée. Ce modèle synthétique de réservoir a été construit lors du "SPE 10^{th} Comparison Solution Project" pour la comparaison de techniques de mise à l'échelle grossière (ou "*upscaling*") [Christie et Blunt, 2001].

3.5.1 Présentation du modèle

Description du réservoir Le modèle est une grille cartésienne de taille $1200 \times 2200 \times 170$ ft et divisée en $60 \times 220 \times 85$ cellules (soit un total de 1,122,000 cellules). Le réservoir ne présente aucune structure et ne possède pas d'aquifère. Le modèle pétrophysique s'appuie sur la séquence stratigraphique "BRENT" définie en Mer du Nord. La partie supérieure du réservoir (70 ft) imite les faciès de la formation Tarbert qui représentent la progradation d'un environnement côtier tandis que la partie inférieure (100 ft) imite ceux de la formation fluviatile Ness supérieur.

Le réservoir est entièrement traversé par cinq puits : quatre producteurs situés aux extrémités de la grille et un injecteur au centre. L'écoulement entre la roche et ces puits est modélisé à l'aide de la relation de Peaceman avec un rayon de puits de 0.15 ft, un facteur de skin de 0. et un facteur de complétion de 1. Tous les puits sont contrôlés en pression : l'injecteur est maintenu à une pression (BHP) de 4500 psi tandis que les producteurs sont fixés à 4000 psi. Par rapport au modèle initial, une faille partiellement étanche a été ajoutée à la position $y = 1650$ ft isolant

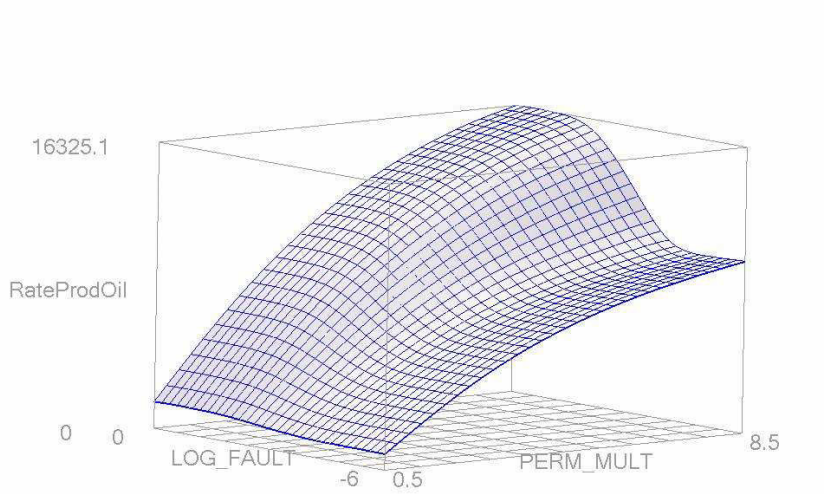


FIG. 3.9 – Surface de référence du débit global de production en huile après 500 jours en fonction des paramètres incertains : multiplicateur global de perméabilité et transmissibilité de la faille.

Paramètre incertain	Valeur basse	Valeur moyenne	Valeur haute
PERM	0.5	4.0	8.5
FAULT	10^{-6}	10^{-3}	1.0

TAB. 3.1 – Intervalles de variation des paramètres jugés incertains dans le modèle de réservoir SPE 10.

ainsi deux des puits producteurs de la zone d'injection (figure (3.8)). Cette faille est modélisée à l'aide d'un multiplicateur de transmissibilité. Enfin, le modèle est initialement rempli d'huile et subit une injection d'eau. Les deux fluides sont supposés incompressibles et immiscibles. Leurs viscosités sont de 2 cp pour l'huile et de 0.6 cp pour l'eau, tandis que leurs masses volumiques sont, respectivement, de 53 et 64 lb.ft^{-3}.

Paramètres incertains et profil de production La variable de production cible est le débit de production global en huile du réservoir après 500 jours. Les paramètres jugés incertains sont un multiplicateur global de perméabilité (noté PERM) et la transmissibilité de la faille (notée FAULT). Les intervalles de variation des différents paramètres sont résumés table (3.1). Afin d'étudier de manière exhaustive le comportement du réservoir, 56 simulations d'écoulement sur 500 jours ont été effectuées à l'aide du simulateur commercial sur ligne de courant 3DSL [Streamsim, 2004]. Ces simulations ont permis de construire le profil de référence pour le débit de production global en huile après 500 jours (figure (3.9)). Celui-ci montre l'effet quadratique de la perméabilité et celui non linéaire de la faille sur la production. Sur cette figure et dans le

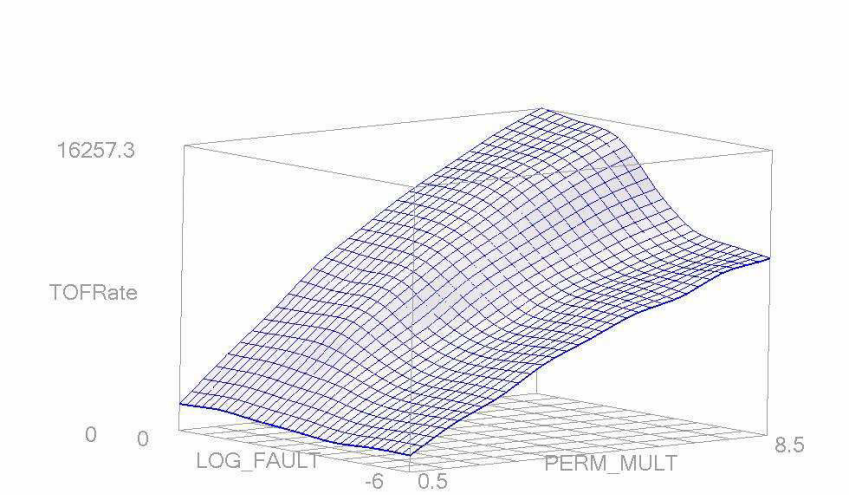
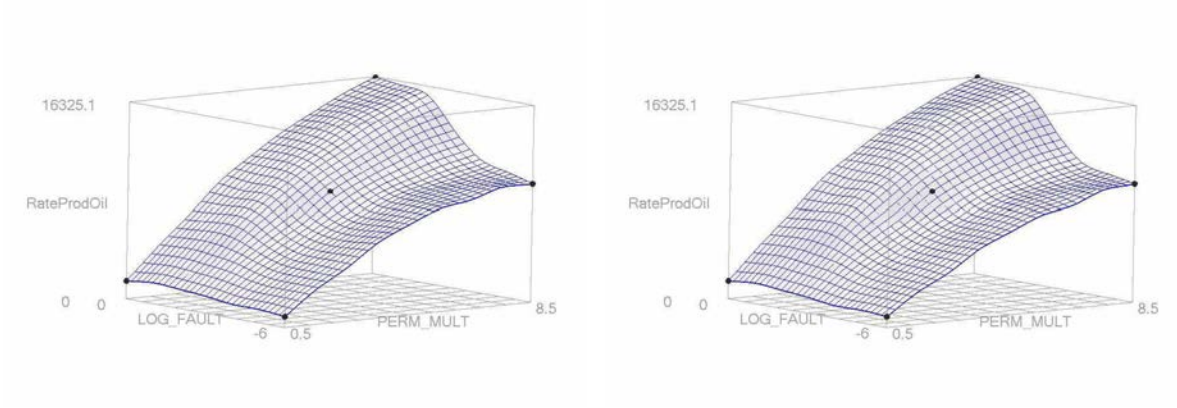


FIG. 3.10 – Surface de réponse du débit de volume drainé basée sur le temps de vol après 500 jours de production vis-à-vis des paramètres incertains du modèle.

reste de l'étude, afin d'éviter des effets d'échelle, le multiplicateur de transmissibilité de la faille est toujours transformé en son logarithme.

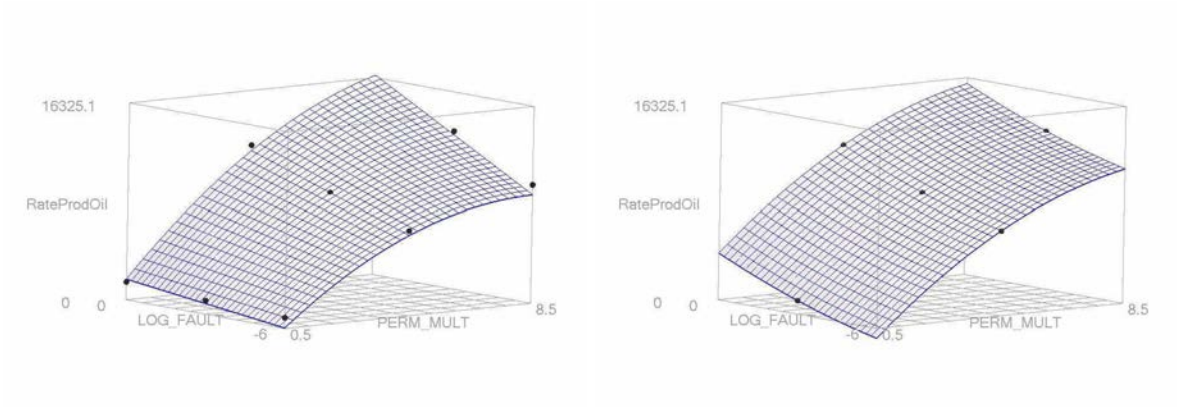
3.5.2 Construction de surface de réponse du débit en huile après 500 jours

Simulation d'écoulement alternative La simulation secondaire considérée dans cette étude est l'estimation du volume drainé basée sur le temps de vol proposée dans Idrobo et al. [2000]. Le temps de vol (voir annexe A) est une quantité fondamentale des simulations sur lignes de courant. En pratique, c'est le temps mis par une particule neutre pour parcourir une distance donnée le long d'une ligne de courant. À partir de cette variable, l'efficacité du drainage du réservoir à un temps t est évaluée comme le volume rocheux dont le temps de vol vers le(s) puits producteur(s) est inférieur ou égal à t [Idrobo et al., 2000]. Cette mesure est l'une des plus simples rendant compte des interactions entre le modèle statique de réservoir (porosité et perméabilité) et les conditions physiques de l'écoulement (propriétés des fluides, pression et débits aux puits, etc.). Elle a déjà été utilisée avec succès dans différentes études pour la quantification des incertitudes sur la prédiction de production et le calage d'historique de production [Gilman et al., 2002; Wang et Kovseck, 2003; Ates, 2005]. Cette estimation de volume drainé est connue pour approcher de manière plutôt convaincante la production cumulée en huile d'un réservoir. Ici, pour approcher le débit de production global en huile, nous utiliserons la dérivée de ce volume par rapport au temps. Cette nouvelle quantité est appelée dans la suite "*Débit de volume drainé basé sur le temps de vol*" ou "*TOF Swept Volume Rate*". La surface de réponse de cette variable est exposée figure (3.10). Lorsque l'on compare cette surface avec la surface de référence (figure (3.9)), il



(a) Interpolation lisse discrète (b) Krigeage avec dérive externe

FIG. 3.11 – Surfaces de réponse du débit global de production en huile après 500 jours tenant compte de la donnée secondaire : (a) interpolation lisse discrète avec contrainte de dérivée seconde et (b) krigeage avec dérive externe.



(a) (b)

FIG. 3.12 – Surfaces de réponse correspondant au débit de production globale en huile après 500 jours : (a) approche basée sur un plan d'expérience de type composite central et (b) sur un plan de type Box-Behnken.

apparaît que les paramètres incertains influent d'une façon similaire sur les deux variables de production.

Surface de réponse avec donnée secondaire La surface de réponse correspondant à l'estimation du débit de volume drainé basé sur le temps de vol après 500 jours est ensuite utilisée comme donnée secondaire. Celle-ci est combinée avec les résultats de simulations d'écoulement conventionnelles issues d'une planification d'expérience à deux niveaux de type factoriel complet et un point central (soit un total de cinq simulations). Les surfaces de réponse construites à

Erreur d'estimation	Moyenne	Maximale
(1) Plan de type Box Behnken (BBD)	8.7 %	189 %
(2) Plan de type composite central (CCD)	7.9 %	95 %
(3) Interpolation Lisse Discrète (DSI)	4.9 %	20 %
(4) Krigeage avec dérive externe (KED)	3.2 %	12 %

TAB. 3.2 – Valeurs moyennes et maximales de l'erreur de prédiction du débit de production d'huile à 500 jours par rapport au modèle de référence des surfaces de réponse construites suivant une approche classique basée sur : (1) une planification de type Box-Behnken et (2) composite centrale comparée avec la méthodologie proposée : (3) interpolation lisse discrète avec contraintes de dérivée seconde et (3) krigeage avec dérive externe.

Débit en huile (en 10^3 stb/jours)	P_{10}	P_{50}	P_{90}	σ
Surface de référence	2.912	8.560	14.428	4.061
Plan Box Behnken (BBD)	2.851	9.045	13.214	3.874
Plan central composite (CCD)	2.483	8.606	13.263	3.893
Interpolation lisse discrète (DSI)	2.834	8.716	13.860	3.926
Krigeage avec dérive externe (KED)	2.810	8.530	14.074	3.997

TAB. 3.3 – Paramètres clefs des histogrammes du débit global de production en huile après 500 jours représentés figure (3.13).

partir de l'interpolation lisse discrète avec contrainte de dérivée seconde et du krigeage avec dérive externe sont exposées figure (3.11).

Approches classiques de modélisation Afin de pouvoir comparer les résultats obtenus, une approche classique basée sur une planification d'expérience à trois niveaux et un modèle polynômial comme surface de réponse est réalisée. Deux cas sont envisagés : le premier considère un plan à trois niveaux de type composite central avec une régression linéaire d'un modèle quadratique ; le second considère un plan de type Box-Behken combiné avec une régression linéaire d'un modèle toujours quadratique mais sans les termes croisés. La figure (3.12) montre les surfaces de réponse correspondantes. Ces deux approches échouent à capturer effectivement la non linéarité et les effet de courbures présents dans le modèle de référence (figure (3.9)).

3.5.3 Analyse et discussion des résultats

Les résultats obtenus montrent que les surfaces de réponse tenant compte de la donnée secondaire approchent mieux la réalité (ici, la surface de référence) que celles construites suivant une méthodologie conventionnelle. Ces résultats sont, en outre, confirmés par l'analyse de l'erreur de prédiction résumée table (3.2).

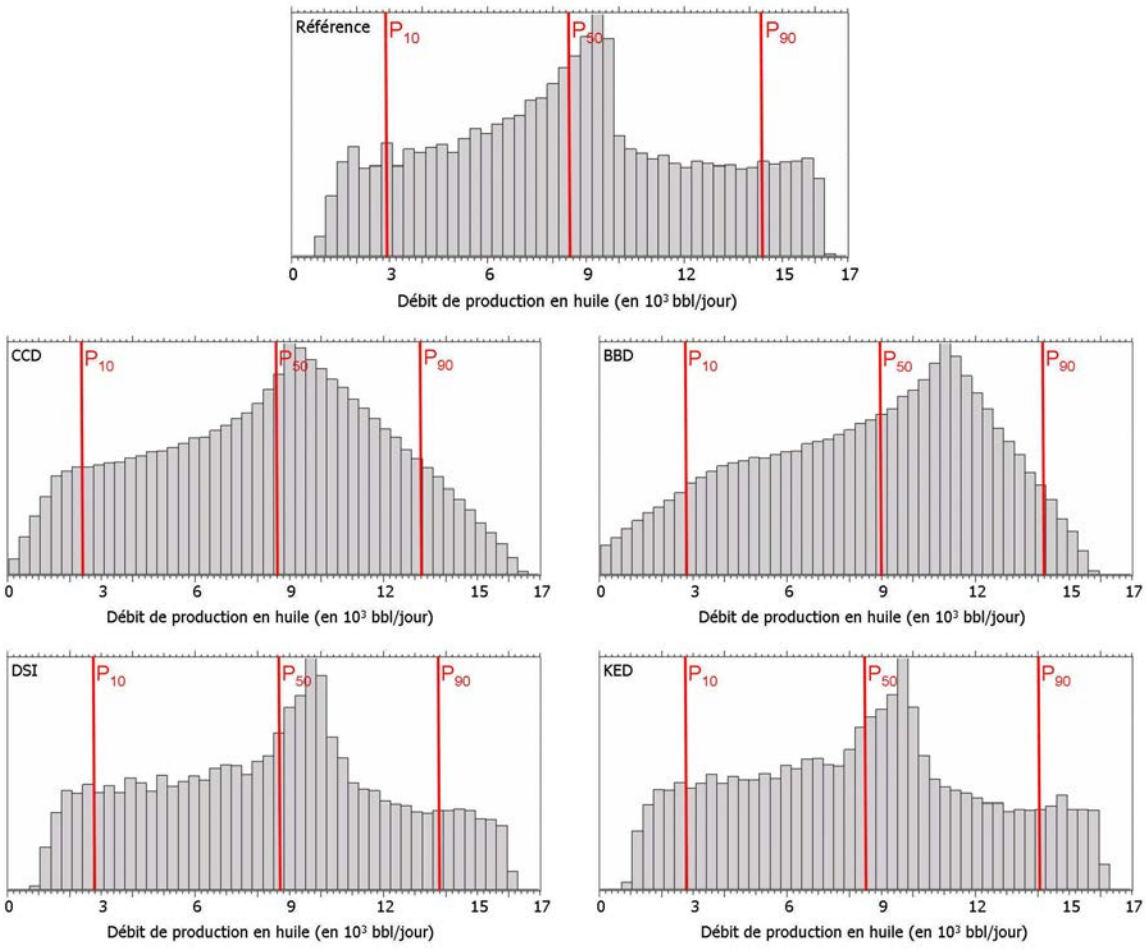


FIG. 3.13 – Histogrammes du débit global de production en huile après 500 jours, obtenus à partir d'un échantillonnage intensif des différentes surfaces de réponse utilisées dans cette étude.

Quantification des incertitudes Afin d'évaluer l'impact de l'approche proposée sur la quantification des incertitudes liées aux prédictions de production, les différentes surfaces de réponse considérées dans cette étude ont été échantillonnées de manière intensive (10 000 échantillons). Les deux paramètres incertains (multiplicateur global de perméabilité et transmissibilité de la faille) sont supposés indépendants et suivent une loi uniforme entre leurs valeurs minimales et maximales respectives. Les histogrammes correspondants sont exposés figure (3.13) tandis que la table (3.3) résume quelques valeurs clefs des distributions. Ces résultats illustrent l'amélioration de la quantification des incertitudes dues à l'introduction des données secondaires. En particulier sur ce modèle de réservoir, l'utilisation d'une simple approche classique conduirait à une sous-estimation de l'incertitude pesant sur le débit de production en huile.

Coût en temps de calcul Une dernière question reste : quel est le coût en temps du calcul de la variable secondaire et de son intégration dans la construction de la surface de réponse ? Dans cette étude basée sur le modèle de réservoir SPE 10, le temps nécessaire pour réaliser une

Méthodologie	Nb. de simulations conventionnelles	Nb. de simulations alternatives	Temps de calcul total (heures)
Modèle de référence	56	0	224
Plan de type Box-Behnken	5	0	20
Plan de type central composite	9	0	36
Approche proposée	5	56	34

TAB. 3.4 – Temps de calcul requis pour la construction des surfaces de réponse suivant les méthodologies envisagées durant cette étude.

simulation d'écoulement conventionnelle sur lignes de courant est en moyenne de 4 heures sur un Intel Pentium 4 - 3.20 GHz avec 1. GB de RAM. À l'opposé, le temps requis pour estimer à partir du temps de vol le débit de volume drainé est de 15 minutes seulement. Enfin, le temps de construction des surfaces de réponse est considéré négligeable. À partir de ces données, le nombre de chaque type de simulation et le temps nécessaire pour chacune des méthodologies présentées dans ce cas d'étude sont résumés table 3.4. Cela montre que le sur-coût en temps de calcul n'est que de 33 % environ comparé à une approche classique, ce qui est parfaitement acceptable au vu du gain de prédictivité obtenu.

3.6 Conclusions

Dans ce chapitre, nous avons proposé et développé une approche innovante pour l'intégration d'une donnée secondaire dans la construction de surface de réponse. La technique combine les résultats de simulations d'écoulement conventionnelles, généralement coûteuses en temps, avec une surface de réponse secondaire issue de simulations rapides alternatives telles que la mesure de la connectivité du réservoir, l'utilisation de la théorie de la percolation, ou l'attribut de volume drainé basé sur le temps de vol, etc. Cette surface de réponse alternative permet de contraindre l'interpolation de celle des données primaire là où il n'y a pas de données. En particulier, grâce à cette approche il est possible de prendre en compte les courbures et non linéarités dans la réponse qui sont capturées par les simulations alternatives mais ne le sont pas par une planification d'expérience conventionnelle à deux ou trois niveaux.

Pour répondre à ce problème, l'approche proposée comprend : (1) une planification d'expérience basée sur les hypercubes latins qui intègre à la fois les simulations primaires et secondaires, (2) une extension des deux techniques de construction de surfaces de réponse proposées au chapitre précédent pour l'intégration de donnée secondaire et (3) une analyse de prédictivité des surfaces basée sur la technique de la validation croisée pour juger de l'apport d'information et de l'utilité de prendre en compte les résultats de simulations alternatives. Cette approche a été illustrée sur une version modifiée du modèle SPE 10. L'analyse des résultats a montré que les surfaces tenant compte de simulations alternatives requièrent un sur-coût en temps parfaitement acceptable par rapport à une approche conventionnelle au vu de l'amélioration de la prédictivité des surfaces.

Deux points majeurs sont toujours problématiques et constituent une perspective à ces travaux. Le premier reste la validité de la donnée secondaire considérée. En effet, l'hypothèse majeure qui sous-tend ces travaux est qu'il doit exister une corrélation entre les résultats des simulations classiques et ceux des simulations alternatives. Il serait intéressant de disposer d'outils d'analyse statistique pour le vérifier. Une première approche a été proposée cependant celle-ci ne permet pas de distinguer quelles sont les modifications de la prédictivité des surfaces de réponse effectivement significatives. Le second point est que les données secondaires susceptibles d'être utilisées reposent le plus souvent sur une simplification du problème physique à résoudre. Or, hormis le simulateur d'écoulement lui-même, il n'existe pas de méthode capable de prédire le comportement d'un réservoir face à tous ses paramètres jugés incertains. Ainsi les problématiques suivantes restent à étudier : (1) la détermination des paramètres incertains ayant le même effet sur les simulations alternatives et conventionnelles, et (2) la combinaison de plusieurs sources de données secondaires.

Chapitre 4

Prise en compte de paramètres incertains stochastiques

Sommaire

4.1 Construction de surfaces de réponse en présence de paramètres stochastiques 85
4.1.1 Principe 85
4.1.2 Estimation par modélisation séparée 86
4.2 Extension des méthodes d'analyse à la prise en compte de paramètres stochastiques 87
4.2.1 Analyse de sensibilité de la production vis-à-vis de paramètres incertains 87
4.2.2 Inversion bayésienne d'un historique de production 91
4.3 Application sur un modèle de réservoir fluviatile 92
4.3.1 Description du modèle de réservoir 92
4.3.2 Construction et analyse des surfaces de réponse 95
4.3.3 Remarques finales 99
4.4 Prise en compte de paramètres continus peu influents 100
4.5 Conclusions 102

Les paramètres incertains d'un réservoir se distinguent suivant la relation entre leurs variations et celles qu'ils induisent dans la production. Dans le cas de paramètres stochastiques, cette relation est purement aléatoire [Zabalza et al., 2000]. Un exemple typique en caractérisation de réservoir est le germe géostatistique nécessaire à la simulation de faciès ou de propriétés pétrophysiques. Un léger changement dans ce germe conduit à des modèles équiprobables présentant des comportements significativement différents vis-à-vis des écoulements.

L'intégration de tels paramètres en ingénierie de réservoir est peu aisée car il est souvent nécessaire de générer un grand nombre de réalisations pour couvrir toutes les configurations possibles, ce qui est particulièrement coûteux en temps et en volume de calculs. C'est pourquoi, dans le cadre de la quantification des incertitudes, l'analyse se limite le plus souvent aux caractéristiques statiques du réservoir, telles que la connectivité des corps les plus perméables, le volume

d'huile en place, etc. (voir [Alabert et Modot, 1992; Deutsch, 1996; Massonnat, 1997, 2000]). De manière plus récente, ce type de paramètre a été aussi envisagé lors du calage d'historique de production dans les méthodes de déformation graduelle [Hu, 2000] ou de perturbation de probabilité (encore appelée "*Probability Pertubation Method*") [Hoffman et Caers, 2003]. Néanmoins, celles-ci ne cherchent pas à couvrir l'ensemble des modèles possibles mais plutôt à en trouver un seul satisfaisant et respectant les données.

En simulation de réservoir, les surfaces de réponse ont pour objectif de couvrir l'ensemble des incertitudes relatives à un modèle de réservoir. Néanmoins cette approche

- reste déterministe : à chaque combinaison de paramètres incertains ne correspond qu'une valeur de production possible ;
- repose sur l'existence d'une relation continue et lisse entre les variations des variables d'entrée et de sortie [Box et Draper, 1987].

Cette méthodologie n'est donc pas adaptée à la prise en compte de paramètres de type stochastique. La modélisation jointe par modèles linéaires généralisés, proposée par Zabalza et al. [1998], est la principale technique décrite dans la littérature pour résoudre ce problème. Ces travaux consistent à prendre en compte l'impact de paramètres stochastiques sur une variable de production par le biais de deux surfaces de réponse distinctes :

- l'une correspond au comportement de la production moyenne vis-à-vis des paramètres continus ;
- l'autre représente la dispersion dans la production due aux paramètres stochastiques.

Cette approche permet de rendre compte des deux types de paramètres tout en gardant leurs intervalles de variation aussi large que nécessaire. La construction de ces deux surfaces peut être réalisée de manière disjointe. Cependant, pour tenir compte de l'hétéroscédasticité éventuelle de la production [1] Zabalza et al. [1998] proposent d'utiliser la technique de régression linéaire généralisée [Nelder et Wedderburn, 1983]. Cette méthode cherche à estimer et à optimiser, suivant un processus itératif, les deux surfaces afin de permettre une meilleure modélisation de la production. Une description plus détaillée de cette méthode se trouve dans Zabalza-Mezghani [2000] et Zabalza-Mezghani et al. [2004], ainsi que des exemples d'applications dans Manceau et al. [2001] pour la quantification des incertitudes sur la production du modèle de réservoir PUNQ et dans Zabalza-Mezghani et al. [2001] au sujet du calage d'historique de production.

Après une revue plus détaillée de la méthode proposée par Zabalza-Mezghani [2000], nous proposons l'adaptation et l'implantation des méthodes d'analyse développées au chapitre (2) pour la gestion de paramètres stochastiques. En particulier, nous nous intéresserons à l'analyse de sensibilité basée sur la variance et à l'inversion bayésienne de données historiques de production. Pour finir, deux exemples sont envisagés : le premier s'appuie sur un modèle synthétique de réservoir fluviatile pour illustrer notre méthodologie, mais aussi mettre en évidence la nécessité d'intégrer ce type de paramètre lors de la caractérisation du comportement dynamique d'un réservoir. Le second reprend le modèle Anticlinal décrit précédemment et explore la possibilité de traiter un certain nombre de paramètres continu jugés peu influents de manière stochastique.

[1] En géostatistique une variable aléatoire est hétéroscédastique si sa variance change en fonction de la position où elle est échantillonnée, dans le cas contraire on parle de variable homoscédastique.

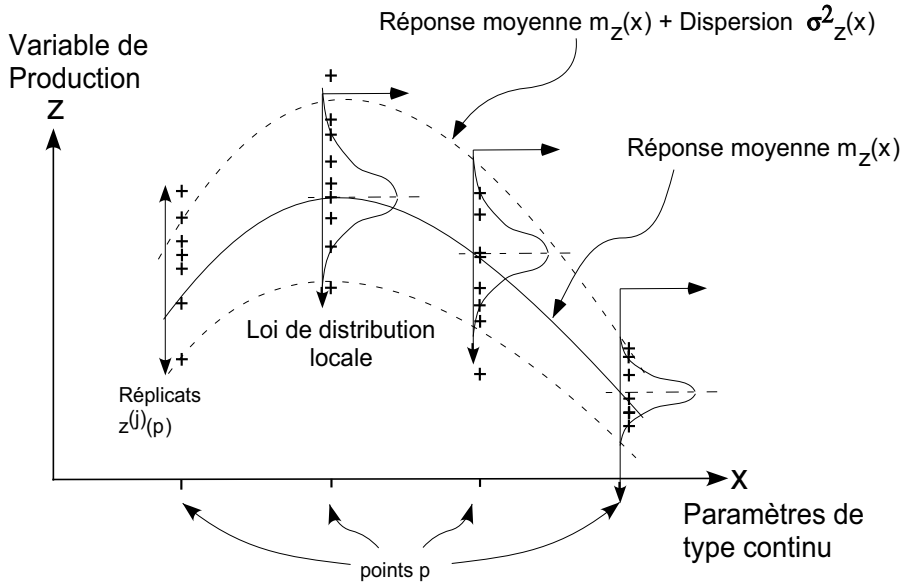


FIG. 4.1 – Approche suivie pour la construction de surface de réponse en présence de paramètres incertains de type stochastique.

4.1 Construction de surfaces de réponse en présence de paramètres stochastiques

Cette section décrit l'approche sélectionnée pour construire une surface de réponse en présence de paramètres incertains de type stochastique. Ces travaux sont basés sur la méthodologie originale proposée par Zabalza et al. [1998]. L'objectif n'est pas de proposer une nouvelle méthode de construction, mais de développer et spécifier le cadre et le formalisme sur lesquels s'appuient les méthodes d'analyse présentées dans la section (4.2).

4.1.1 Principe

La prise en compte de paramètres stochastiques nécessite de sortir du cadre classique des surfaces de réponse déterministe tel que présenté dans les chapitres précédents. Aussi, pour une combinaison de paramètres continus, la production suit une loi de densité définie localement (figure (4.1)), au lieu de ne prendre qu'une unique valeur. Cette loi est supposée normale, dans la suite, on notera $m_z(\mathbf{x})$ sa moyenne et $\sigma_z^2(\mathbf{x})$ sa variance. Considérons un point $\mathbf{x}$ dans l'espace des paramètres continus, il vient alors :

$$z(\mathbf{x}) \ \sim \ \mathcal{N}(m_z(\mathbf{x}), \sigma_z^2(\mathbf{x})) \tag{4.1}$$

la variable de réponse $z(\mathbf{x})$ suit ainsi la loi de densité suivante :

$$f_z(z, \mathbf{x}) \ = \ \frac{1}{\sqrt{(2\pi)\ \sigma_z^2(\mathbf{x})}}\ e^{-\frac{1}{2}\ \frac{(z-m_z(\mathbf{x}))^2}{\sigma_z^2(\mathbf{x})}} \tag{4.2}$$

Considérons maintenant $u(\mathbf{x})$, définie comme la différence entre les valeurs effectives de production et la valeur de production moyenne. Cette nouvelle variable suit logiquement une loi de densité normale centrée sur zéro et de même variance que $z(\mathbf{x})$. La production en un point $\mathbf{x}$ de l'espace des paramètres de type continu s'écrit alors sous la forme :

$$z(\mathbf{x}) \;=\; m_d(\mathbf{x}) + u(\mathbf{x}) \quad \text{avec} \quad u(\mathbf{x}) \sim \mathcal{N}(0, \sigma_z^2(\mathbf{x})) \tag{4.3}$$

Par ce biais, l'effet des paramètres stochastiques peut être séparé de celui de ceux de type continu. En outre, si l'on considère l'équation (4.3), l'effet de $u(\mathbf{x})$ sur la production est continu. Par conséquent, les méthodes d'analyse décrites précédemment peuvent aussi s'appliquer dans le cas présent : il suffit de rassembler l'ensemble des phénomènes stochastiques dans la variable $u(\mathbf{x})$. En particulier, il est donc possible d'analyser et de quantifier la sensibilité de la production (1) vis-à-vis des paramètres stochastiques, mais également (2) vis-à-vis de leurs interactions avec les autres paramètres du réservoir. Cependant, il est important de noter que ce traitement est global et ne permet pas une analyse individuelle de chaque paramètre stochastique.

4.1.2 Estimation par modélisation séparée

Le problème envisagé se réduit alors à celui de la construction dans l'espace des paramètres continus de deux surfaces de réponse distinctes. La première est la moyenne de la variable de production $m_z(\mathbf{x})$ tandis que la seconde correspond à sa dispersion $\sigma_z^2(\mathbf{x})$.

Dans un premier temps, la planification d'expérience réalisée pour rassembler les données nécessaires consiste à (1) choisir un plan d'expérience généralement à deux ou trois niveaux pour l'ensemble des paramètres continus, et (2) considérer, pour chaque point $\mathbf{p}$ du plan précédent, plusieurs répétitions de traitement de telle sorte que les paramètres stochastiques changent tandis que les autres restent fixes. En pratique, tous les points correspondants à une répétition ont les mêmes coordonnées dans l'espace des paramètres continus, mais se distinguent par différentes valeurs de production. Dans la suite, le nombre de répétitions est noté r et les différentes valeurs de production $z^{(j)}(\mathbf{p}), j \in [\![1, r]\!]$. Cette planification a pour objectifs (1) de capturer la dispersion due aux paramètres stochastiques et (2) de permettre le calcul, lors d'une étape de pré-processus, d'une estimation de la production moyenne $\hat{m}_z(\mathbf{p})$ et de sa variance $\hat{\sigma}_z^2(\mathbf{p})$ en chaque point $\mathbf{p}$, telles que :

$$\hat{m}_z(\mathbf{p}) = \frac{1}{r} \sum_{j=0}^{r} z^{(j)}(\mathbf{p}) \quad \text{et} \quad \hat{\sigma}_z(\mathbf{p}) = \frac{1}{r} \sum_{j=0}^{r} (z^{(j)}(\mathbf{p}) - \hat{m}_z(\mathbf{p}))^2 \tag{4.4}$$

La dernière étape est de construire, à partir de ces estimations locales, les surfaces de réponse correspondantes. Pour cela, les méthodes décrites au chapitre (2) peuvent être directement appliquées car on retrouve le cadre déterministe classique.

Durant ces travaux de thèse, nous avons préféré cette approche dite séparée avec pré-processus où les deux surfaces sont construites indépendamment l'une de l'autre à la modélisation jointe proposée dans Zabalza et al. [1998]. Bien que potentiellement moins précise, elle nous parait plus flexible : (1) il est possible de considérer des méthodes d'estimation de surface de réponse autres que la régression linéaire, par exemple l'interpolation lisse discrète ou le krigeage dual décrits précédemment ; (2) la dispersion dans la production peut être modélisée autrement

qu'avec une loi normale. Il suffit de s'intéresser à d'autres valeurs clés de la dispersion : par exemple le minimum et le maximum de la production si une distribution uniforme est envisagée.

4.2 Extension des méthodes d'analyse à la prise en compte de paramètres stochastiques

Cette section s'appuie sur le formalisme présenté précédemment pour étendre les méthodes d'analyse envisagées au chapitre (2) dans le cas où un ou plusieurs paramètres stochastiques doivent être pris en compte.

4.2.1 Analyse de sensibilité de la production vis-à-vis de paramètres incertains

L'analyse de sensibilité cherche à quantifier l'influence des paramètres incertains du réservoir, ainsi que celle de leurs interactions, sur la production. La méthode d'analyse choisie dans ces travaux est une extension de la méthode VBSA [Sobol', 1990; Saltelli et al., 1999] qui s'appuie sur la décomposition de la variance des variables de production Var(Z) sous la forme (voir section (2.5.5)) :

$$\mathrm{Var}(Z) = \sum_{i=0}^{n} D_i + \sum_{i=0}^{n}\sum_{j=0}^{j<i} D_{ij} + \sum_{i=0}^{n}\sum_{j=0}^{j<i}\sum_{k=0}^{k<j} D_{ijk} + ... + D_{1...n} \tag{4.5}$$

où les différents termes mesurent l'influence des paramètres incertains (et de leurs interactions) par le biais de leurs contribution à $\mathrm{Var}(Z)$:

$$D_i = \mathrm{Var}(E(Z|X_i)), \quad D_{ij} = \mathrm{Var}(E(Z|X_i, X_j)) - D_i - D_j, \quad \ldots \tag{4.6}$$

Les avantages de cette approche sont sa flexibilité et la possibilité de prendre en compte des surfaces de réponse de forme complexe. En outre, bien que la description originale considère que les paramètres incertains suivent une distribution uniforme entre 0 et 1, il est possible de l'étendre de manière à prendre en compte tout type de distribution.

Principes Suivant le formalisme choisi, l'ensemble des paramètres stochastiques sont remplacés par une unique variable u suivant une loi de densité normale (équation (4.3)). Les termes d'espérance et de variance totale, équations (2.38) et (2.39), utilisés dans le calcul des indices de sensibilité deviennent alors :

$$E(Z) = \int_{-\infty}^{+\infty} \int_0^1 \ldots \int_0^1 (m_z(\mathbf{x}) + u(\mathbf{x})) \frac{e^{-\frac{u(\mathbf{x})^2}{2\sigma_z^2(\mathbf{x})}}}{\sqrt{2\pi\sigma_z^2(\mathbf{x})}}\, d\mathbf{x}du \tag{4.7}$$

et

$$\mathrm{Var}(Z) = \int_{-\infty}^{+\infty} \int_0^1 \ldots \int_0^1 (m_z(\mathbf{x}) + u(\mathbf{x}) - E(Z))^2 \frac{e^{-\frac{u(\mathbf{x})^2}{2\sigma_z^2(\mathbf{x})}}}{\sqrt{2\pi\sigma_z^2(\mathbf{x})}}\, d\mathbf{x}du \tag{4.8}$$

Le calcul des termes de variance partielle, de la forme $\mathrm{Var}(E(Z|X_{i_1},\ldots,X_{i_s},U))$, est plus complexe. En effet, la décomposition utilisée (équation 4.5) suppose que les différents paramètres soient indépendants, en particulier que u est indépendant des $x_i, i \in 1,\ldots,n$. Cela implique que u soit constant dans tout le domaine d'incertitude et que la variance de la production soit de type homoscédastique, c'est-à-dire :

$$u(\mathbf{x}) = u \quad \text{et} \quad \sigma_z^2(\mathbf{x}) = \sigma_z^2 = \text{cst.}\ \forall\ \mathbf{x} \in \mathcal{X} \tag{4.9}$$

Il vient alors que les termes non conditionnés par u valent :

$$\mathrm{Var}(E(Z|X_{i_1},\ldots,X_{i_s})) = \int_0^1 \ldots \int_0^1 g_{i_1\ldots i_s}^2(x_{i_1},\ldots,x_{i_s})\ dx_{i_1}\ldots dx_{i_s} - E(Z)^2 \tag{4.10}$$

avec

$$g_{i_1\ldots i_s}(x_{i_1},\ldots,x_{i_s}) = \int_{-\infty}^{+\infty} \int_0^1 \ldots \int_0^1 (m_z(\mathbf{x}) + u)\frac{e^{-\frac{u^2}{2\sigma_z^2}}}{\sqrt{2\pi\sigma_z^2}}\ d\mathbf{x}du/(dx_{i_1}\ldots dx_{i_s}) \tag{4.11}$$

Tandis qu'à l'inverse, les termes conditionnés par u sont égaux à :

$$\mathrm{Var}(E(Z|X_{i_1},\ldots,X_{i_s},U)) = \int_{-\infty}^{+\infty} \int_0^1 \ldots \int_0^1 g_{i_1\ldots i_s,u}^2(x_{i_1},\ldots,x_{i_s},u)\frac{e^{-\frac{u^2}{2\sigma_z^2}}}{\sqrt{2\pi\sigma_z^2}}\ dx_{i_1}\ldots dx_{i_s}du - E(Z)^2 \tag{4.12}$$

avec

$$g_{i_1\ldots i_s,u}(x_{i_1},\ldots,x_{i_s},u) = \int_0^1 \ldots \int_0^1 (m_z(\mathbf{x}+u)\ d\mathbf{x}/(dx_{i_1}\ldots dx_{i_s}) \tag{4.13}$$

Dans le cas contraire, la variance de la production est hétéroscédastique et dépend de la position $\mathbf{x}$ où elle est estimée. Les différentes variables ne sont alors plus indépendantes et cette approche n'est plus rigoureusement exacte (en particulier lors de l'estimation des termes de la forme $\mathrm{Var}(E(Z|X_{i_1},\ldots,X_{i_s},U))$). Elle reste cependant applicable car, comme montré dans **?**, les indices de sensibilité sont toujours une mesure valable pour classer les paramètres incertains suivant leur importance. Par la suite, cette approche sera toujours envisagée par la suite et les différents paramètres incertains seront traités de manière indépendante bien que cela soit approximatif.

Formules d'estimation L'estimation des indices de sensibilité s'appuie sur la technique d'intégration numérique de type Monte-Carlo proposée par Sobol' [1990] et Saltelli et al. [1999] (voir section (2.5.2)). Ici, l'espace d'échantillonnage correspond aux n paramètres continus et à la variable u : les premiers sont supposés suivre une loi uniforme entre 0 et 1 tandis que u a une distribution normale. Celle-ci est centrée sur 0 et a pour variance $\sigma_z^2(\mathbf{x})$ qui est estimée localement à partir de la surface de réponse correspondante. Les formules d'estimation de l'espérance et de la variance totale de la production deviennent :

$$\hat{E}(z) \simeq \frac{2u_0}{N}\sum_{m=1}^{N}\ (m_z(\mathbf{x}_m) + u_m)\ \frac{e^{-\frac{u_m^2}{2\ \sigma_z^2(\mathbf{x}_m)}}}{\sqrt{2\pi\sigma_z^2(\mathbf{x}_m)}} \tag{4.14}$$

et

$$\hat{\mathrm{Var}}(z) \simeq \frac{2u_0}{N} \sum_{m=1}^{N} (m_z(\mathbf{x}_m) + u_m)^2 \ \frac{e^{-\frac{u_m^2}{2\ \sigma_z^2(\mathbf{x}_m)}}}{\sqrt{2\pi\sigma_z^2(\mathbf{x}_m)}} \ - \ \hat{E}^2(d_i) \tag{4.15}$$

où N représente le nombre de points d'estimation, $\mathbf{x}_m$ est un point aléatoire dans l'espace des paramètres stochastiques $[0,1]^n$ et u_m est une valeur aléatoire dans l'intervalle $[-u_0, +u_0]$ [1] et représente la dispersion due aux paramètres stochastiques. De la même façon, les termes de variance partielle indépendants de u sont estimés égaux à :

$$\begin{aligned}\hat{\mathrm{Var}}(E(Z|X_{i_1}, \ldots, X_{i_s})) \simeq \tfrac{(2u_0)^2}{N} \sum_{m=1}^{N} g(\mathbf{v}_m, x_{i_1,m}, \ldots, x_{i_s,m}, u_{1,m}) \ \times \\ g(\mathbf{w}_m, x_{i_1,m}, \ldots, x_{i_s,m}, u_{2,m}) \ - \ \hat{E}^2(z)\end{aligned} \tag{4.16}$$

avec

$$g(\mathbf{v}, x_{i_1}, \ldots, x_{i_s}, u) \ = \ (m_z(\mathbf{v}, x_{i_1}, \ldots, x_{i_s}) + u) \times \frac{e^{-\frac{u^2}{2\sigma_z^2(\mathbf{v}, x_{i_1}, \ldots, x_{i_s})}}}{\sqrt{2\pi\sigma_z^2(\mathbf{v}, x_{i_1}, \ldots, x_{i_s})}} \tag{4.17}$$

où $\mathbf{v}_m$ et $\mathbf{u}_m$ sont là encore deux points aléatoires distincts dans l'espace des paramètres continus exceptés $x_{i_1}, \ldots, x_{i_s}$, c'est-à-dire $[0,1]^{n-s}$ tandis que $u_{1,m}$ et $u_{2,m}$ sont deux valeurs choisies dans $[-u_0, +u_0]$. Enfin, les $x_{i_1,m}, \ldots, x_{i_s,m}$ prennent leurs valeurs dans $[0,1]^s$. À l'opposé, les termes dépendants de la variable u sont estimés égaux à :

$$\begin{aligned}\hat{\mathrm{Var}}(E(Z|X_{i_1}, \ldots, X_{i_s}, U)) \simeq \tfrac{2u_0}{N} \sum_{m=1}^{N} g(\mathbf{v}_m, x_{i_1,m}, \ldots, x_{i_s,m}, u_m) \ \times \\ g(\mathbf{w}_m, x_{i_1,m}, \ldots, x_{i_s,m}, u_m) \times \frac{e^{-\frac{u^2}{2\sigma_z^2(\mathbf{v}_m, x_{i_1,m}, \ldots, x_{i_s,m})}}}{\sqrt{2\pi\sigma_z^2(\mathbf{v}_m, x_{i_1,m}, \ldots, x_{i_s,m})}} \ - \ \hat{E}^2(z)\end{aligned} \tag{4.18}$$

avec

$$g(\mathbf{v}, x_{i_1}, \ldots, x_{i_s}, u) = m_z(\mathbf{v}, x_{i_1}, \ldots, x_{i_s}) + u \tag{4.19}$$

Validation Le même modèle que celui décrit section (2.5.2) est utilisé pour valider la méthode d'analyse de sensibilité proposée ci-dessus. Ce modèle se compose de trois paramètres incertains x_1, x_2, et x_3 suivant tous une loi uniforme entre $-\pi$ et $+\pi$. Il a la forme suivante :

$$z(x_1, x_2, x_3) \sim \mathcal{N}(m_z(x_1, x_2, x_3), \ \sigma_z^2(x_1, x_2, x_3)) \tag{4.20}$$

où

$$m_z(x_1, x_2, x_3) = \sin(x_1) + 7 \ \sin^2(x_2) + 0.2 \ x_3^4 \ \sin(x_1). \tag{4.21}$$

Deux cas sont envisagés pour la variance $\sigma_z^2(x_1, x_2, x_3)$:

[1] Idéalement, sachant que $u(\mathbf{x})$ suit une distribution normale, cet intervalle devrait être $[-\infty, +\infty]$. Cependant, pour éviter tout problème numérique et d'éventuelles valeurs aberrantes, il a été choisi de se limiter à l'intervalle $[-u_0, +u_0]$ où u_0 est un facteur égal au maximum de $3 \times \sigma_z(\mathbf{x}), \forall \mathbf{x} \in \mathcal{X}$. En effet, la probabilité associée aux points en dehors de cet intervalle est négligeable.

Variables	Valeurs analytiques	Cas d'étude (1)	Cas d'étude (2)
D	31.572	35.539	45.590
D_1	11.97	12.009	11.931
D_2	6.125	6.161	6.153
D_3	0.0	0.030	-0.023
D_{12}	0.0	-0.036	0.015
D_{13}	13.474	13.396	13.459
D_{23}	0.0	-0.031	0.022
D_u	-	4.044	12.467
D_{u1}	-	-0.062	0.322
D_{u2}	-	-0.080	0.201
D_{u3}	-	0.0	0.563

TAB. 4.1 – Variances totale et partielles du modèle analytique considéré pour valider l'approche proposée (équation (4.20). Dans cette table, la première colonne contient les résultats analytiques lorsqu'aucune dispersion n'est prise en compte, les deux autres contiennent les valeurs correspondant aux cas d'une dispersion homoscédastique (cas (1)) et hétéroscédastique (cas (2)).

– soit elle est constante dans tout le domaine des paramètres continus (cas d'une réponse homoscédastique) telle que :

$$\sigma_z(x_1, x_2, x_3) = 2, \quad \forall\ (x_1, x_2, x_3) \in [-\pi, +\pi]^3 \tag{4.22}$$

– soit elle est supposée liée à la production moyenne et varie dans l'espace des paramètres incertains (cas de réponse hétéroscédastique) telle que :

$$\sigma_z(x_1, x_2, x_3) = 0.2 \times |m_z(x_1, x_2, x_3)|, \quad \forall\ (x_1, x_2, x_3) \in [-\pi, +\pi]^3 \tag{4.23}$$

L'intérêt de cette fonction est tout d'abord sa forme relativement complexe mais surtout la possibilité de calculer de manière analytique les différents termes de variances totale et partielles dans les cas où aucune dispersion n'est prise en compte et où la variance est homoscédastique [Ishigami et Homma, 1990; **?**]. Dans le cas hétéroscédastique, ces termes ne peuvent être que grossièrement estimés. En pratique, les calculs sont réalisés suivant un échantillonnage de Sobol' de 100 000 points dans l'espace des trois paramètres incertains (x_1, x_2, et x_3). Les résultats obtenus sont exposés table (4.1).

Ces résultats montrent que dans le cas d'une variance homoscédastique il y a un très bon accord entre les résultats analytiques et les estimations. En effet, les variances liées aux paramètres continus sont égales, tandis que la variance partielle due au terme de dispersion vaut 4.044 et, comme attendu, est égale à l'augmentation de la variance totale. D'autre part, dans le cas (2), les variances partielles des paramètres continus sont bien retrouvées et l'augmentation de variance totale est bien répartie sur le terme lié au paramètre stochastique (noté D_u) et sur ceux correspondants à son interaction avec les autres paramètres incertains (notés D_{u1}, D_{u2} et D_{u3}). Cette exemple illustre la capacité de l'analyse de sensibilité basée sur la variance à prendre

en compte un paramètre incertain de type stochastique. Cependant, il est important de noter que, dans le cas d'une dispersion hétéroscédastique, ces résultats ne sont qu'approximatifs et ne permettent qu'un classement des différents paramètres incertains. En outre, si le modèle comprend plusieurs paramètres stochastiques, ceux-ci étant combinés au sein d'une unique variable $u(\mathbf{x})$, il n'est pas possible de mesurer l'influence individuelle de chacun d'entre eux.

4.2.2 Inversion bayésienne d'un historique de production

Dans le cadre de l'analyse des incertitudes sur les simulations d'écoulement, l'inversion bayésienne est une technique d'analyse qui quantifie l'impact de l'apport de données historiques de production sur la distribution des paramètres incertains. Comme décrit ci-dessus, cette technique cherche à estimer la loi de densité multivariée des différents paramètres en tenant compte (1) de l'information connue a priori sur les paramètres, (2) des données de production et (3) du modèle physique reliant les paramètres incertains aux données de production.

En pratique, cette technique est basée sur la formule de Bayes [1] :

$$p(\mathbf{x}|\mathbf{z} = \mathbf{z}_{obs}) = k\ p_0(\mathbf{x})\ \mathcal{L}(\mathbf{x}) \tag{4.24}$$

où

- k est une constante de normalisation ;
- $p_0(\mathbf{x})$ est la loi de densité a priori des paramètres incertains ;
- $\mathcal{L}(\mathbf{x})$ est la fonction de vraisemblance et mesure la probabilité qu'une combinaison de paramètres incertains explique les données de production.

Par rapport au cas déterministe classique, la prise en compte d'un ou plusieurs paramètres incertains stochastiques nécessite de changer la façon dont est estimée la fonction de vraisemblance. Comme montré dans Tarantola [2005], celle-ci est égale à :

$$\mathcal{L}(\mathbf{x}) = \int_{\mathcal{X}} k\ p_0(\mathbf{z})\ \theta(\mathbf{z}|\mathbf{x})\ d\mathbf{x} \tag{4.25}$$

où $\mathcal{X}$ représente l'espace des paramètres incertains de type continu, $p_0(\mathbf{z})$ est la loi de densité a priori des données de production et est estimée à partir des données historiques, tandis que $\theta(\mathbf{z}|\mathbf{x})$ représente le modèle physique reliant la production aux paramètres incertains.

Dans le cas classique déterministe $\theta(\mathbf{z}|\mathbf{x})$ est un Dirac : on retrouve alors l'équation (2.60). En présence de paramètres de type stochastique, $\theta(\mathbf{z}|\mathbf{x})$ est égale à la loi considérée pour modéliser la dispersion dans les données de production :

$$\theta(\mathbf{z}|\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^{n_z}\ \det(\mathbf{C}_T(\mathbf{x}))^{-1}}}\ e^{-\frac{1}{2}(\mathbf{z}-\mathbf{m}_z(\mathbf{x}))^t \mathbf{C}_T(\mathbf{x})(\mathbf{z}-\mathbf{m}_z(\mathbf{x}))} \tag{4.26}$$

[1] De même qu'au chapitre (2) toutes les données de production sont analysées simultanément et représentées sous la forme d'un vecteur de taille n_z, lui-même noté $\mathbf{z}(\mathbf{x}) = [z^1(\mathbf{x}), \ldots, z^{n_z}(\mathbf{x})]$.

où $\mathbf{C}_T(\mathbf{x})$ est un opérateur de covariance construit à partir des n_z modèles de dispersion des différentes variables de production [1]. Parallèlement, deux cas sont considérés pour la loi de densité a priori des données de production $p_0(\mathbf{z})$:

– soit il n'y a pas d'erreurs de mesure sur les données historiques, ou celles-ci sont négligeables, on a alors :

$$p_0(\mathbf{z}) = \delta(\mathbf{z} - \mathbf{z}_{obs}) \tag{4.27}$$

d'où

$$\mathcal{L}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^{n_z}\ \det(\mathbf{C}_T(\mathbf{x}))^{-1}}}\ e^{-\frac{1}{2}(\mathbf{z}_{obs}-\mathbf{m}_z(\mathbf{x}))^t\mathbf{C}_T(\mathbf{x})(\mathbf{z}_{obs}-\mathbf{m}_z(\mathbf{x}))} \tag{4.28}$$

– soit ces erreurs suivent une loi normale centrée sur $\mathbf{z}_{obs}$ et de covariance $\mathbf{C}_d$:

$$p_0(\mathbf{z}) = \frac{1}{\sqrt{(2\pi)^{n_d}\ \det(\mathbf{C}_d)^{-1}}}\ e^{-\frac{1}{2}(\mathbf{z}-\mathbf{z}_{obs})^t\mathbf{C}_d(\mathbf{z}-\mathbf{z}_{obs})} \tag{4.29}$$

On montre alors que [Tarantola, 2005] :

$$\mathcal{L}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^{n_z}\ \det(\mathbf{C}_D(\mathbf{x}))^{-1}}}\ e^{-\frac{1}{2}(\mathbf{z}_{obs}-\mathbf{m}_z(\mathbf{x}))^t\mathbf{C}_D(\mathbf{x})(\mathbf{z}_{obs}-\mathbf{m}_z(\mathbf{x}))} \tag{4.30}$$

avec

$$\mathbf{C}_D(\mathbf{x}) = \mathbf{C}_d + \mathbf{C}_T(\mathbf{x}) \tag{4.31}$$

Pour finir, l'estimation de la loi de densité a posteriori des paramètres incertains s'effectue à l'aide de l'algorithme de Métropolis [Metropolis et Ulman, 1949]. Les procédures d'analyse décrites précédemment se basant sur cette technique d'échantillonnage sont alors rigoureusement les mêmes.

4.3 Application sur un modèle de réservoir fluviatile

Cette section s'appuie sur un modèle synthétique de réservoir fluviatile pour illustrer et discuter la méthodologie proposée. Dans cet exemple, le paramètre stochastique est le germe géostatistique utilisé lors de la simulation de différentes réalisations de faciès dans la réservoir. Nous nous sommes intéressé à évaluer son impact sur la production et surtout à montrer la nécessité d'intégrer ce type de paramètre lors de la caractérisation du comportement dynamique du réservoir.

4.3.1 Description du modèle de réservoir

Modèle

Le modèle de réservoir consiste en une grille cartésienne de dimension : $210 \times 500 \times 45$ pieds et divisée en $21 \times 50 \times 45$ cellules (figure (4.2)). Seuls deux faciès sont envisagés : un faciès chenal

[1] Ici ces variables sont supposées indépendantes, ainsi $\mathbf{C}_T$ est une matrice diagonale dont les coefficients sont calculés à partir des surfaces de réponse représentant la variance des différentes variables ${\sigma^2}_{z^j}(\mathbf{x})\ \forall\ j \in [1, n_z]$.

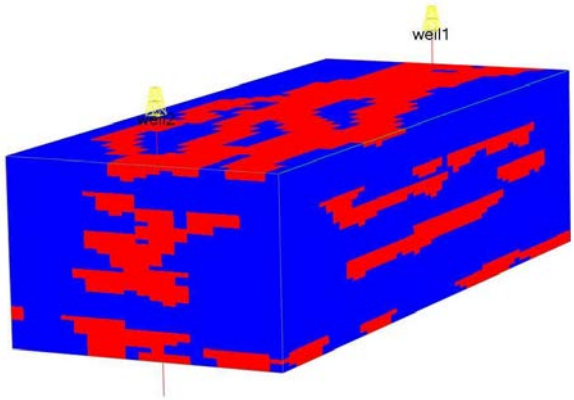


FIG. 4.2 – Modèle synthétique de réservoir fluviatile. Les deux faciès considérés dans ce modèle sont le faciès chenal (en rouge) et le faciès de plaine d'inondation (en bleu).

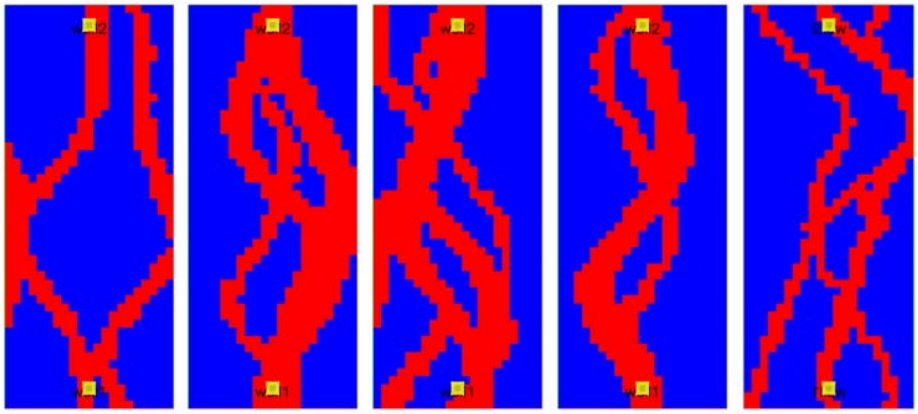


FIG. 4.3 – Exemples de simulations de faciès par l'algorithme FuvSim avec différents germes géostatistiques ; la proportion de chenal est fixée à 30%.

et un faciès de plaine d'inondation. Tous deux sont simulés à l'aide de l'algorithme de simulation basée objet FluvSim [Deutsch et Journel, 1992]. Du point de vue des propriétés pétrophysiques, la plaine d'inondation est considérée imperméable tandis que le chenal est supposé homogène avec une valeur de porosité égale à 20% et une perméabilité incertaine entre 300 et 600 mD.

Ce réservoir ne contient que deux fluides, l'huile et l'eau, considérés à la fois immiscibles et incompressibles. Leurs masses volumiques sont, respectivement, de 53 et 64 lb.ft^3, tandis que leurs densités sont de 2.0 et 0.6 cP. Les courbes de perméabilité relative considérées dans cette étude sont de type Corey sans aucune saturation résiduelle que ce soit en huile ou en eau : $kr_w(S_w) = S_w^2$ et $kr_o(S_w) = (1 - S_w)^2$. Enfin, le réservoir est supposé entièrement rempli d'huile à l'état initial et subit une injection d'eau.

Deux puits verticaux sont localisés aux deux extrémités du réservoir et le traversent entièrement (figure (4.2)). Ils sont modélisés à l'aide de la formule de Peaceman avec un facteur de

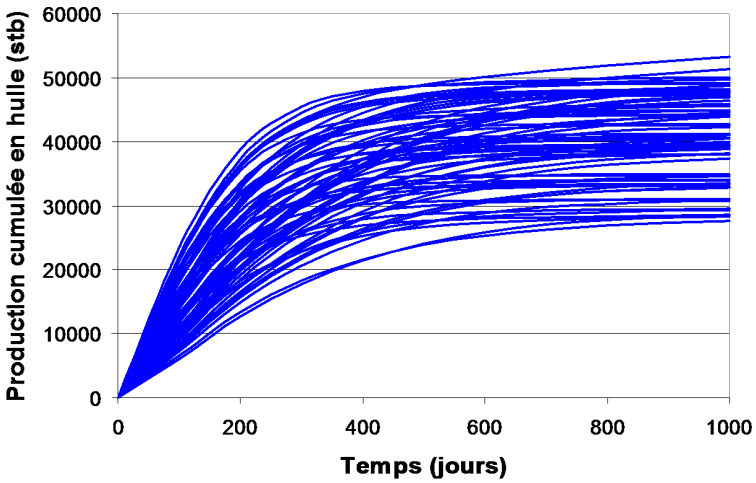


FIG. 4.4 – Évolution de la production cumulée en huile sur 1000 jours pour l'ensemble des 72 modèles considérés durant l'étude.

"skin" nul, un facteur de complétion de 1 et un rayon de 0.15 pieds. Lors des simulations, ces puits sont contrôlés, sans condition d'arrêt, en pression à 4150 psi pour le puits injecteur et à 4000 psi pour le puits producteur.

Paramètres incertains et planification d'expérience

Les paramètres incertains du modèle sont :

- la perméabilité globale des chenaux (PERM), qui suit une loi de densité uniforme entre 300 et 600 mD ;
- la proportion de faciès chenal dans le réservoir ou "*net-to-gross*" (NTG), qui est uniformément distribuée entre 20 et 30 % ;
- la configuration des chenaux. Les différentes réalisations de faciès sont obtenues à l'aide de l'algorithme de simulation Fluvsim à partir d'un simple changement de la graine géostatistique (figure (4.3)).

Dans la suite, la proportion et la perméabilité du faciès chenal sont traitées comme des paramètres incertains de type continu. Alternativement, les changements de configuration des chenaux sont traités de manière stochastique.

La planification d'expérience réalisée ici comprend trois niveaux pour chaque paramètre continu et huit répétitions de traitement pour prendre en compte l'effet du paramètre stochastique, soit un total de $3 \times 3 \times 8 = 72$ simulations. Celles-ci ont été effectuées à l'aide du simulateur d'écoulement sur lignes de courant commercial 3DSL [Streamsim, 2004]. La production est simulée sur 1000 jours avec une sortie tous les 100 jours. La variable cible est la production cumulée en huile au cours du temps (figure (4.4)).

4.3.2 Construction et analyse des surfaces de réponse

Surfaces de réponse

La figure (4.5) montre différentes surfaces de réponse à 100, 400 et 1000 jours construites suivant l'approche proposée. L'algorithme de construction est l'interpolation lisse discrète sur une grille régulière en deux dimensions de 25×25 nœuds. La rugosité est contrainte en imposant un gradient constant avec un fitting factor de 1. Ces résultats illustrent la capacité de la méthode proposée à (1) prendre en compte les courbures présentes dans le modèle de la production moyenne, mais aussi à (2) capturer la dispersion dans la production due aux paramètres de type stochastique. En effet, les points représentant les simulations sont bien compris dans l'intervalle défini par la production moyenne plus ou moins trois fois son écart type.

Parallèlement, il est d'étudier de manière qualitative l'influence des différents paramètres incertains. Les plus influents sur la production sont la proportion et la perméabilité des chenaux. Cependant, ils évoluent au cours du temps : à 100 jours la perméabilité est plus influente que la proportion des chenaux et inversement à 1000 jours. De plus, l'impact du paramètre stochastique semble constant tandis que les interactions entre paramètres semblent négligeables.

Analyse de sensibilité

Une analyse de sensibilité de la production vis-à-vis des paramètres incertains a été effectuée aux différents pas de temps entre 100 et 1000 jours. Les résultats sont exposés figure (4.6). La procédure d'échantillonnage est basée sur le générateur de nombres pseudo aléatoires de Sobol' et comprend 10^5 points d'évaluation.

Ces résultats sont cohérents avec les constatations effectuées d'après la forme globale des surfaces de réponse. En particulier, les paramètres les plus influents sont effectivement la proportion et la perméabilité des chenaux. On vérifie aussi que l'influence de la proportion de chenaux augmente au cours du temps, tandis que celle de la perméabilité diminue et s'annule à partir de 600 jours. Cela peut être logiquement lié avec le pourcentage de volume drainé dans le réservoir : à 100 jours, le volume drainé est faible c'est donc la perméabilité qui influe, à l'opposé, à 1000 jours, tout le réservoir est drainé, c'est donc la proportion (et le volume) des chenaux qui importe. Pour finir, le paramètre stochastique a bien un effet constant d'environ 15-18% tandis que les interactions avec les paramètres continus sont négligeables.

Inversion bayésienne d'historique de production

Afin d'illustrer la technique de l'inversion bayésienne, un modèle correspondant à une proportion de faciès chenal de 27% et à une perméabilité de chenal de 550 mD est sélectionné comme référence. Ce modèle est ensuite simulé sur 500 jours pour générer un historique de production. La réalisation de faciès a été sélectionnée de telle sorte que sa production soit éloignée de la production moyenne telle que prédite par les surfaces de réponse. L'objectif est d'obtenir un modèle au comportement légèrement atypique bien qu'aussi probable que n'importe quel autre.

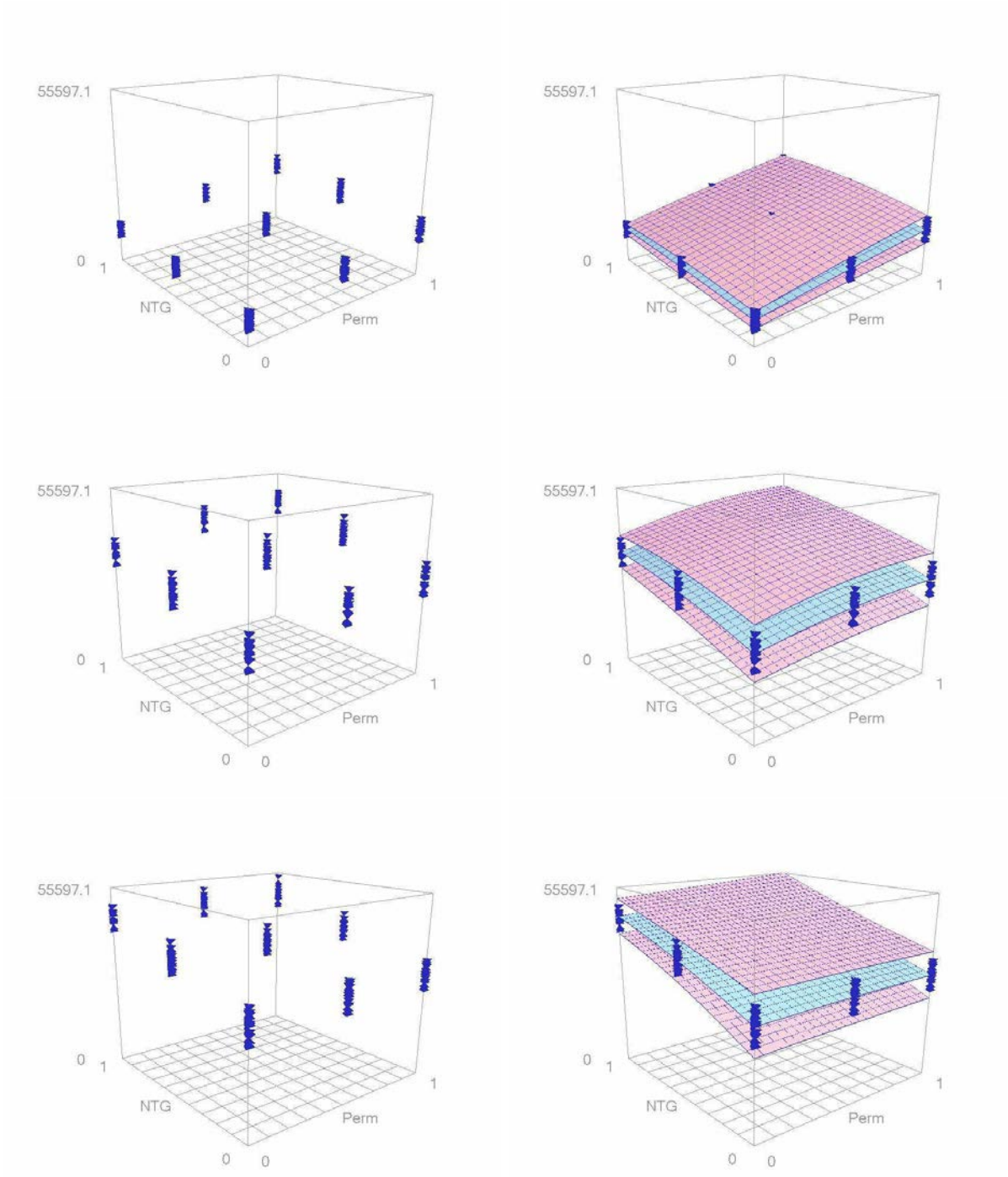


FIG. 4.5 – Points de données (à gauche) et surfaces de réponse interpolées à l'aide de l'algorithme DSI (à droite) correspondant à la production cumulée en huile du réservoir après 100, 400 et 1000 jours. Sur les différentes figures, les deux paramètres incertains ont été mis à l'échelle entre 0 et 1 et la surface bleue représente la production moyenne tandis que les surfaces roses représentent la valeur moyenne plus ou moins trois fois l'écart type de la production.

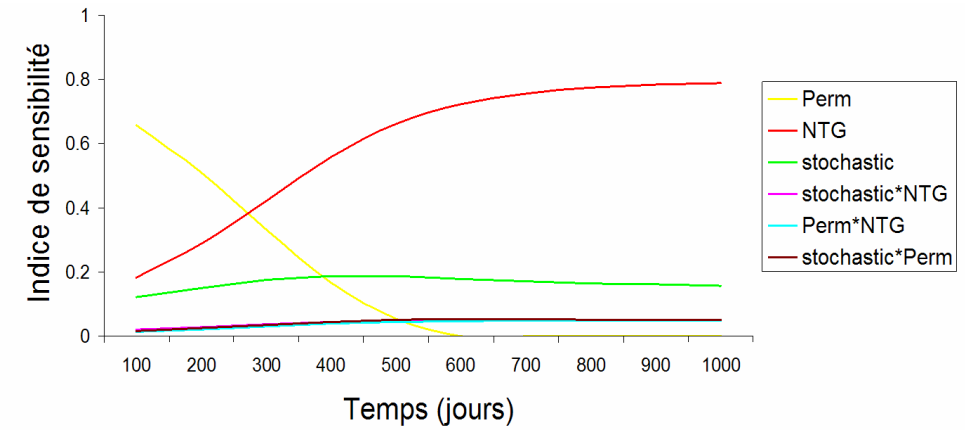


FIG. 4.6 – Évolution dans le temps de la sensibilité de la production cumulée en huile vis-à-vis des paramètres incertains du modèle : perméabilité des chenaux (Perm), proportion du faciès chenal (NTG) et configuration des chenaux (stochastic).

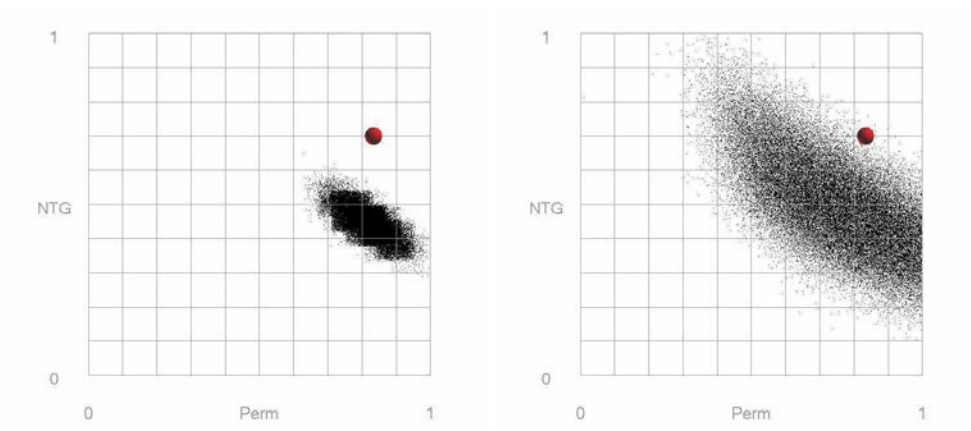
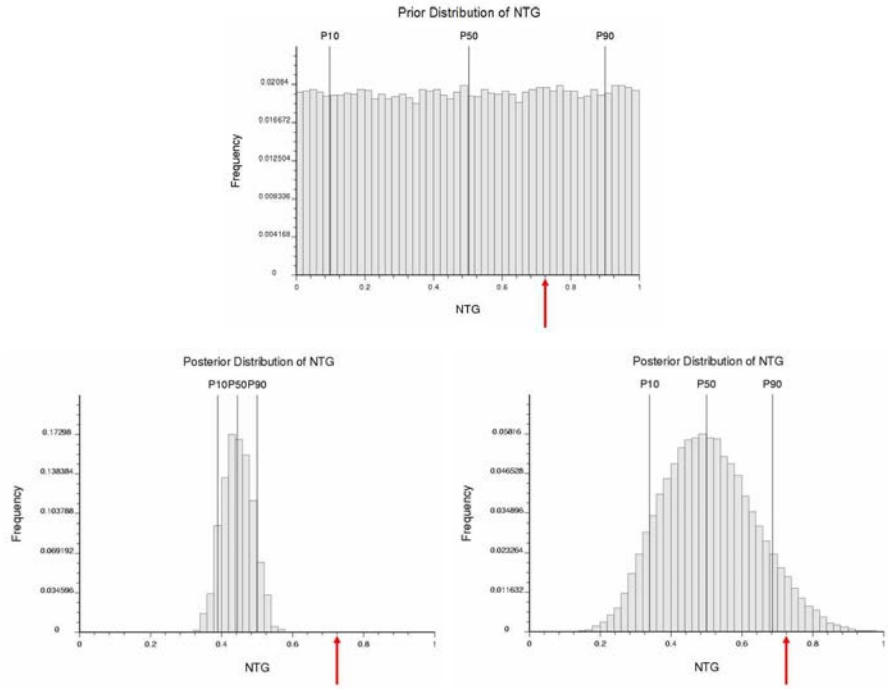


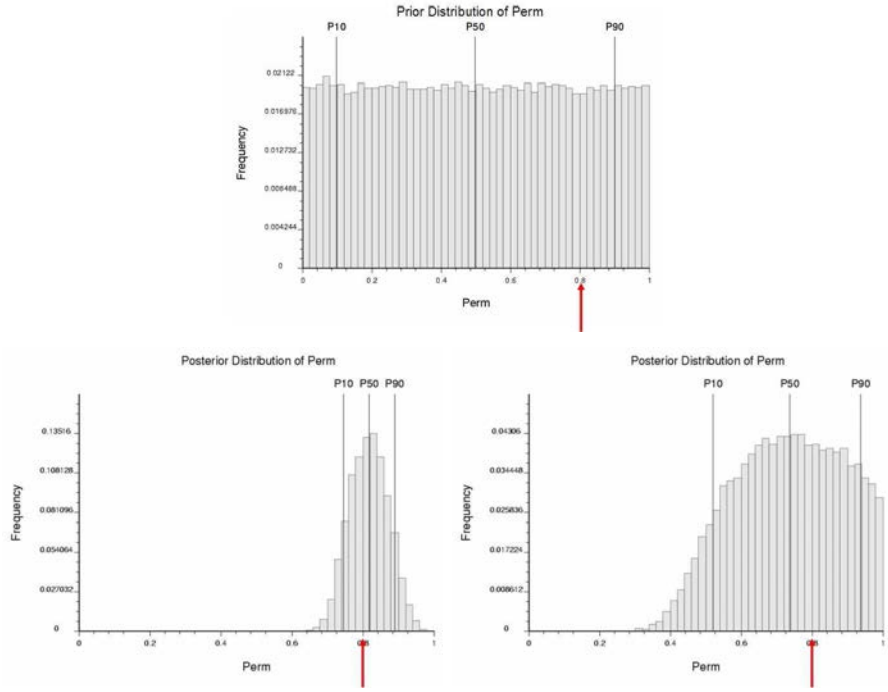
FIG. 4.7 – Échantillonnage des modèles jugés vraisemblables après inversion bayésienne dans le cas (a) où aucun terme de dispersion n'est pris en compte et (b) suivant l'approche proposée. La boule rouge indique le modèle de référence.

Dans la suite, deux études d'inversion bayésienne ont été menées en parallèle. La première considère, suivant une approche classique déterministe, qu'il n'y a pas de dispersion dans la production et ne prend en compte que les surfaces de réponse correspondant à la production moyenne. La seconde est l'approche proposée pour tenir compte de l'effet dû aux changements de réalisation de faciès. Pour chaque cas, les analyses suivantes ont été réalisées :

- l'estimation de la fonction de vraisemblance à partir des surfaces de réponse et de l'historique de production ;
- l'échantillonnage de cette fonction à l'aide de l'algorithme de Métropolis ;
- l'estimation des histogrammes marginaux a posteriori des paramètres incertains de type continu : proportion et perméabilité du faciès chenal ;



(a) Proportion de faciès chenal



(b) Perméabilité des chenaux

FIG. 4.8 – Histogrammes correspondant à la perméabilité et à la proportion du faciès chenal dans le réservoir. Pour ces deux paramètres incertains on trouve, en haut, l'histogramme a priori ; en bas à gauche, l'histogramme a posteriori obtenu suivant l'approche classique déterministe ; en bas à droite, l'histogramme a posteriori obtenu suivant l'approche proposée. La flèche rouge indique le modèle de référence.

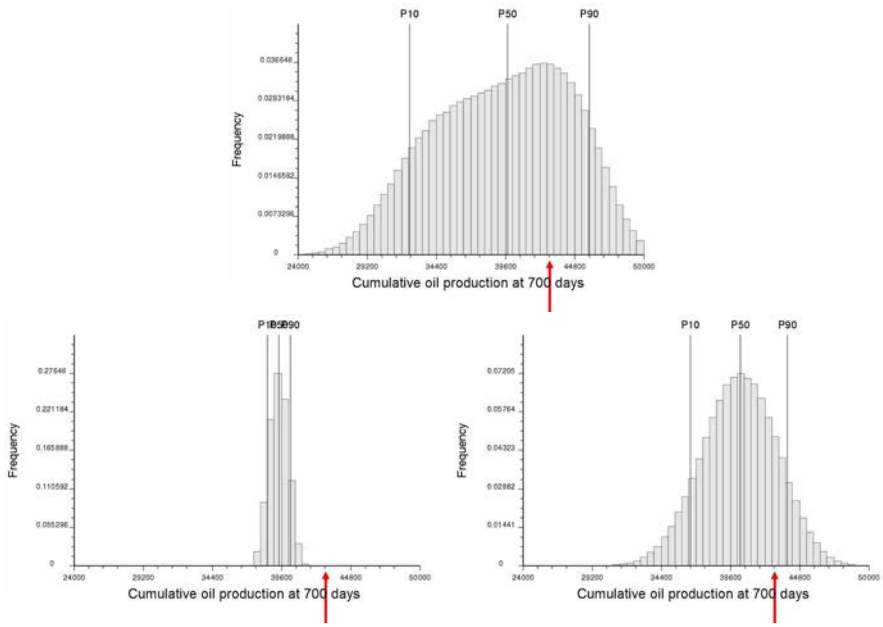


FIG. 4.9 – Histogrammes de la production cumulée en huile après 700 jours. En haut, histogramme de la production a priori avant inversion bayésienne ; en bas à gauche, histogramme a posteriori obtenu suivant une approche déterministe ; en bas à droite, histogramme a posteriori obtenu suivant l'approche proposée. Sur l'ensemble des figures, la flèche rouge indique la valeur de production à 700 jours obtenues à partir du modèle de référence.

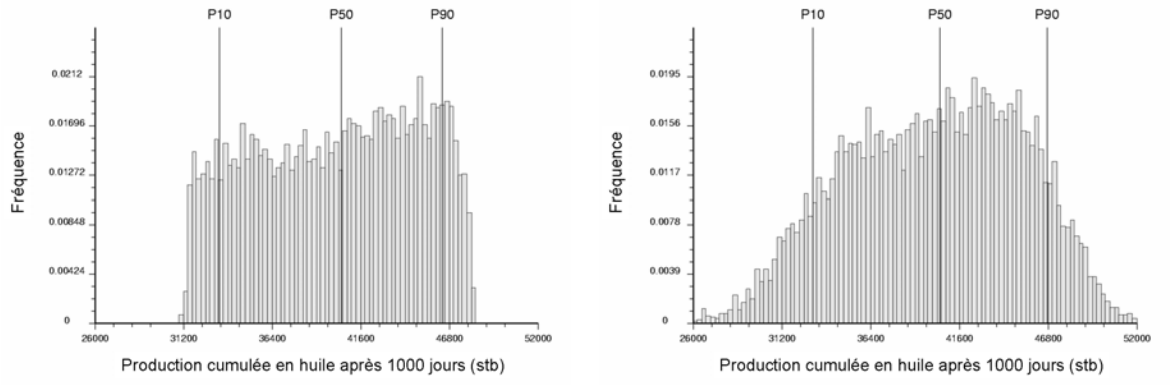
– l'estimation de l'histogramme a posteriori de la production à 700 jours.

Les résultats correspondant aux deux derniers points sont montrés figures (4.7) à (4.9). Celles-ci illustrent, en particulier :

– comment, si l'on compare les histogrammes a priori et a posteriori, l'incertitude sur les paramètres incertains et sur la production est réduite par l'utilisation de l'inversion bayésienne pour la prise en compte de l'historique de production ;
– mais surtout, l'impact majeur du paramètre incertain de type stochastique. En effet, dans le cas où celui-ci n'est pas pris en compte, le modèle de référence ne fait pas partie des modèles jugés vraisemblables, et, en outre, la valeur effective de la production à 700 jours n'est pas non plus dans l'intervalle P_{10}-P_{90}, tel que prédit après l'inversion bayésienne.

4.3.3 Remarques finales

D'une manière plus générale, cet exemple montre l'impact important sur la production que peuvent avoir un ou plusieurs paramètres de type stochastique. Il montre aussi la nécessité de tenir compte de ce type de paramètre pour caractériser précisément le comportement dynamique d'un réservoir. En effet, si l'analyse ne s'était concentrée que sur les histogrammes de la production cumulée en huile à 1000 jours, par exemple, comme ceux exposés figure (4.10), le paramètre stochastique aurait été jugé négligeable. En effet, les histogrammes obtenus suivant



(a) Approche déterministe

(b) Approche proposée

FIG. 4.10 – Histogrammes de la production cumulée en huile à 1000 jours obtenus suivant (a) une approche déterministe classique et (b) suivant l'approche proposée. A noter, que bien que le paramètre stochastique soit pris en compte, la forme générale de l'histogramme et l'intervalle $P_{10} - P_{90}$ sont similaires.

l'approche déterministe classique et suivant l'approche proposée sont semblables. Or l'indice de sensibilité de ce paramètre est sensiblement constant à environ 15-18 %, de plus, s'il n'est pas pris en compte, l'inversion bayésienne est incapable d'identifier comme vraisemblable le modèle de référence.

4.4 Prise en compte de paramètres continus peu influents

Le modèle Anticlinal, décrit dans la section (2.6.3), comprend neuf paramètres incertains, or parmi ceux-ci seul un nombre limité est réellement significatif et a un impact sur la production. Classiquement, les paramètres jugés peu influents sont éliminés de l'analyse et fixés à leur valeur moyenne. L'intérêt est de réduire la dimension de l'espace incertain, d'augmenter la densité des données et d'améliorer la précision des analyses. Cependant, la somme de l'effet de ces paramètres, jugés individuellement négligeables, peut être, elle, non-négligeable. Est-il possible d'en tenir compte et de les intégrer dans la chaîne de modélisation ? Nous proposons, dans cette section, de les traiter sous la forme d'un paramètre stochastique et d'employer la technique d'analyse décrite ci-dessus.

Analyse de sensibilité La première étape consiste à analyser la sensibilité des variables de production (les débits de production en huile et en eau après 6000 jours) vis-à-vis des paramètres incertains. Pour cela, les surfaces de réponse sont interpolées par la technique du krigeage dual avec sur un variogramme de type puissance ($\omega = 1.75$) et sans terme de régularisation. Puis, l'estimation des indices de sensibilité est réalisée grâce un échantillonnage de Sobol' de 10 000 points. Les résultats obtenus (figure (4.11)) permettent d'identifier les paramètres suivants comme les plus influents sur la production. Concernant le débit de production en l'huile, on a dans l'ordre :

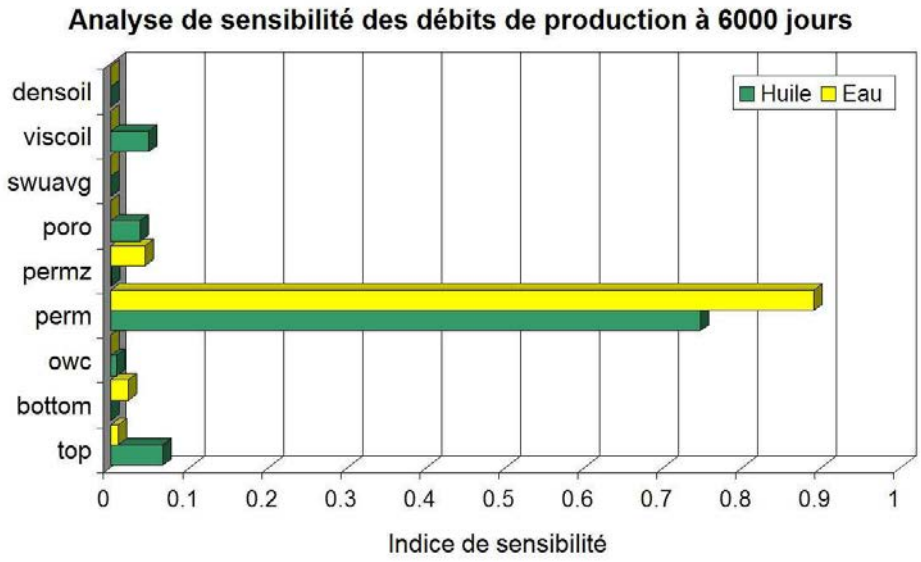


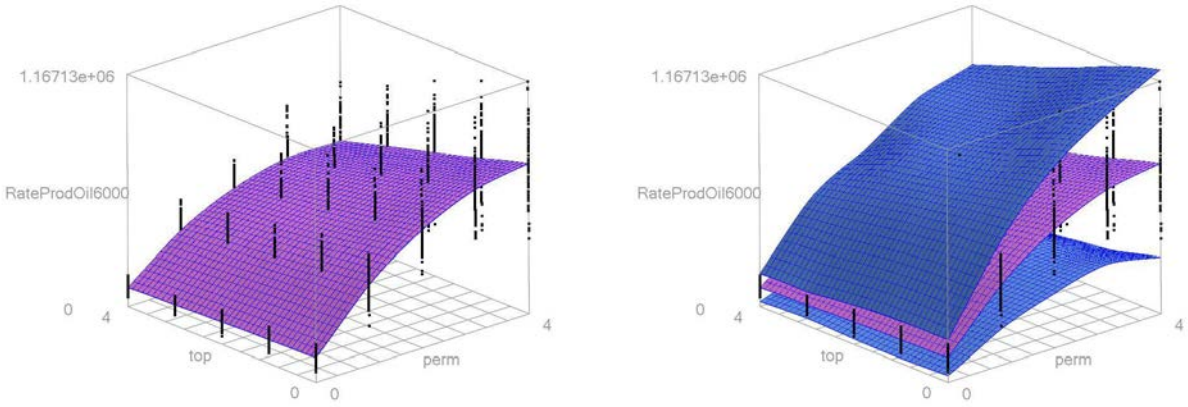
FIG. 4.11 – Indices de sensibilité de premier ordre des débits de production en huile et en eau après 6000 jours sur le modèle Anticlinal présenté au chapitre (2).

- la perméabilité du réservoir (perm) à 75 % ;
- la position du toit du réservoir (top) à 7 % ;
- la viscosité de l'huile (viscoil) à 5 % ;
- la porosité du réservoir (poro) à 4 % ;

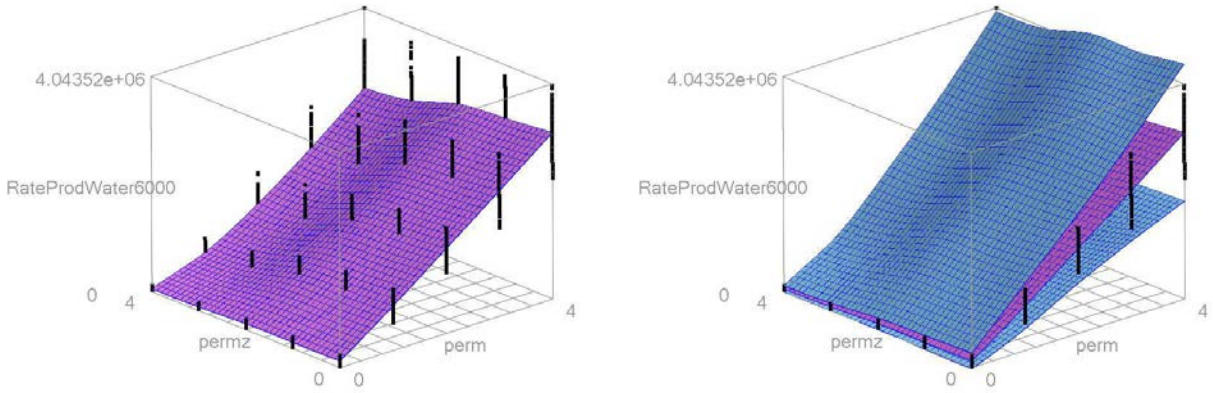
et concernant le débit de production en eau :

- la perméabilité du réservoir (perm) à 89 % ;
- le ratio k_v/k_h (permz) à 4 % ;
- la position du mur du réservoir (bottom) à 2 %.

Construction et analyse des surfaces de réponse Par la suite, pour chacune des variables de production, seuls les deux paramètres incertains les plus influents sont conservés. Les autres sont remplacés par une unique variable dont l'effet sur la production est supposée stochastique. Les surfaces de réponse obtenues par interpolation lisse discrète sont exposées figure (4.12). Elles semblent rendre compte de manière satisfaisante à la fois de l'effet des paramètres continus mais aussi de l'effet de dispersion dû à l'ensemble des paramètres jugés peu influents et négligés. Pour vérifier cela, une nouvelle analyse de sensibilité est réalisée et les résultats sont comparés à ceux obtenus précédemment (figure (4.13)). Dans les deux cas l'influence attribuée aux paramètres les plus influents est la même. D'autre part, les résidus correspondant, dans le cas (1) à la somme des indices de sensibilité des paramètres peu influents et, dans le cas (2) à la variable stochastique, sont aussi semblables. Ces résultats indiquent que le traitement stochastique d'un certain nombre de variables ne semble pas introduire de biais dans les analyses.



(a) Débit de production en huile à 6000 jours

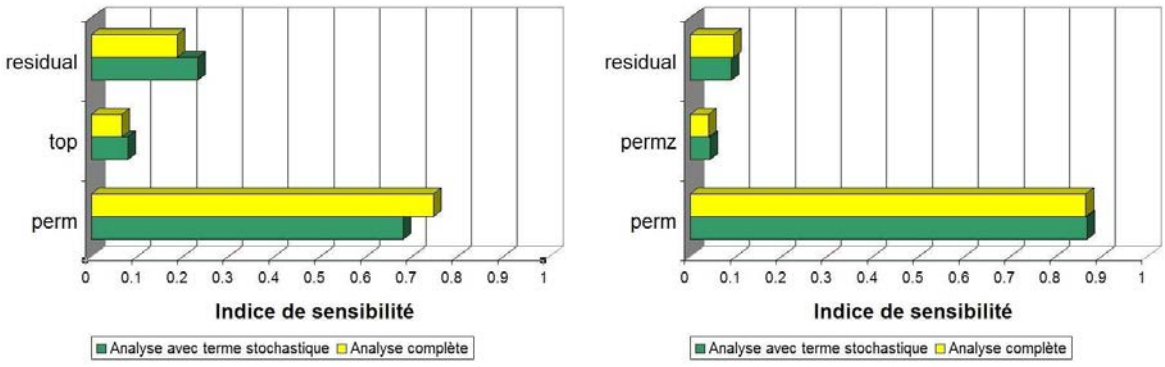


(b) Débit de production en eau à 6000 jours

FIG. 4.12 – Surfaces de réponse des débits de production en huile et en eau après 6000 jours vis-à-vis de la perméabilité et de la position du toit du réservoir, les autres paramètres incertains étant traités de manière stochastique. Dans les différentes figures, la surface en rouge correspond à la production moyenne tandis que les surfaces bleues correspondent à la production moyenne plus ou moins trois fois son écart-type.

4.5 Conclusions

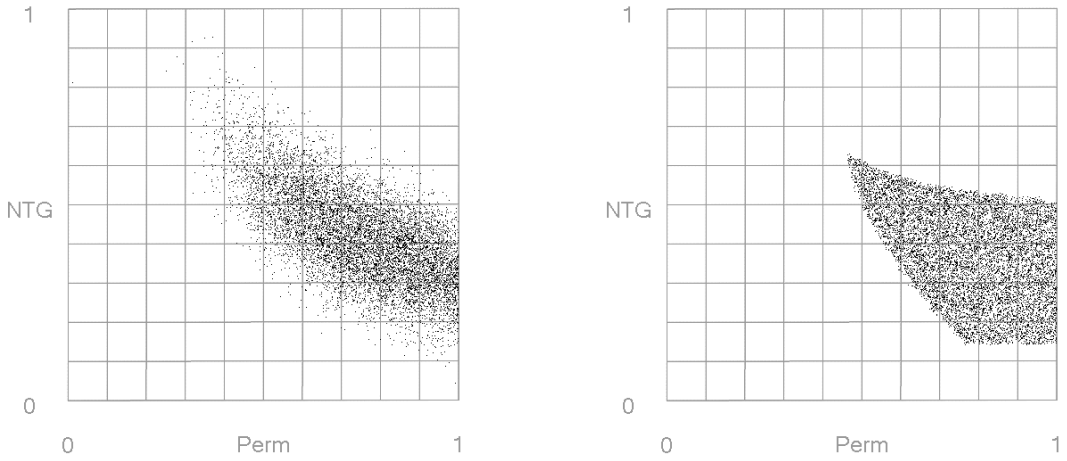
La prise en compte des incertitudes liées aux paramètres incertains de type stochastique nécessite un traitement particulier. En effet, ceux-ci se caractérisent par un effet purement aléatoire sur la production et ne peuvent être inclus dans une approche déterministe classique. À partir de la méthode proposée par Zabalza-Mezghani [2000], nous avons développé et implanté un ensemble d'outils et de techniques d'analyse pour résoudre ce problème. En particulier, nous proposons (1) d'utiliser l'analyse de sensibilité basée sur la variance pour mesurer l'influence des paramètres incertains continus comme stochastiques et (2) d'utiliser l'inversion bayésienne pour tenir compte d'un éventuel historique de production. Ces techniques ont été validées sur un exemple synthétique de réservoir fluviatile. Cet exemple insiste, en outre, sur la nécessité de



(a) Débit de production en huile

(b) Débit de production en eau

FIG. 4.13 – Comparaison des indices de sensibilité des débits de production en huile en en eau calculés sur le modèle Anticlinal (1) en tenant compte de l'ensemble des paramètres et (2) suivant l'approche envisagée dans laquelle certains paramètres sont traités de manière stochastique.



(a) Distribution normale

(b) Distribution uniforme

FIG. 4.14 – Exemples d'échantillons a posteriori de modèles vraisemblables obtenus après inversion bayésienne dans le cas où la dispersion dans la production est supposée suivre (a) une loi normale ou (b) une loi uniforme.

prendre en compte ce type de paramètre : le changement de configuration des chenaux influe fortement la production et s'il n'est pas pris en compte, lors du calage d'historique, le modèle de référence n'est pas estimé vraisemblable. Nous proposons aussi d'étendre l'utilisation de cette méthodologie au traitement de paramètres continus mais peu influents sur la production. Classiquement, ceux-ci sont éliminés en les fixant à leurs valeurs moyennes. Cela augmente la densité des données dans l'espace incertain et améliore la robustesse des résultats d'analyse. Nous proposons de les traiter de manière stochastique afin de pouvoir apprécier leur effet durant toute

la chaîne de modélisation des incertitudes. Cette approche a été illustrée sur un exemple pour lequel le traitement stochastique de paramètre continu ne semble pas introduire de biais.

Ce travail peut être poursuivi dans différentes directions. Tout d'abord, au lieu de se limiter à modéliser l'effet des paramètres stochastiques sur la production à l'aide d'une loi normale, il serait intéressant d'envisager d'autres types de distribution. En particulier, celles-ci pourraient être soit paramétriques (uniforme, triangulaire, etc.) ou bien non paramétriques en se basant, par exemple, sur l'estimation des déciles de la distribution. La figure (4.14) illustre ainsi l'effet du changement de ce type de distribution sur les résultats de l'inversion bayésienne. Comme le montre cette figure, la sélection d'une distribution a un impact important sur les résultats d'analyse et doit constituer un choix de modélisation. D'autre part, le formalisme proposé devrait pouvoir être étendu à la prise en compte de paramètres de type discret tels que le statut binaire (ouvert/fermé) d'une faille, différents scénarios géologiques ou différentes cartes structurales d'un réservoir, etc. Or, de tels paramètres ont aussi un impact majeur sur la réponse en production d'un réservoir et doivent pouvoir être pris en compte. Un point intéressant serait en particulier la détermination de leurs interactions avec les autres paramètres du modèle qui ne sont pas nécessairement évidentes.

Chapitre 5

Méthode pour le lissage d'histogramme

Sommaire

5.1 Lissage d'histogramme : définitions, objectifs et méthodes 106
5.1.1 Histogramme et diagramme de fréquence 106
5.1.2 État de l'art des méthodes de lissage d'histogramme 107
5.2 Approche proposée 110
5.2.1 Principes 110
5.2.2 Contraintes pour l'interpolation lisse discrète 110
5.2.3 Résultats préliminaires 114
5.3 Analyse de la sensibilité de la méthode à ses différents paramètres 114
5.4 Prise en compte de données lissée pour la simulation de propriété par "*cloud tranform*" 119
5.5 Conclusions 121

Je tiens particulièrement à remercier mon co-directeur de thèse Jean-Laurent Mallet pour m'avoir suggéré les idées sur le lissage d'histogramme développées dans ce chapitre.

L'estimation précise de la loi de densité de probabilité ou de la distribution cumulée d'un ensemble de variables comme l'amplitude sismique, la porosité et la perméabilité est un problème récurrent dans l'exploration, la calibration et l'analyse des données du sous-sol. En effet, ce type de loi permet de caractériser une variable aléatoire et d'appréhender ses relations avec les autres variables du système étudié [Soong, 1981]. Par exemple, dans le cadre d'une inversion bayésienne, l'étude de la loi de densité révèle les régions de l'espace des paramètres incertains où le calage avec l'historique de production est le meilleur [Choudhuri, 1998; Tarantola, 2005]. En modélisation de réservoir, la loi de distribution cumulée permet de contraindre la simulation des propriétés pétrophysiques du sous-sol (faciès, porosité, perméabilité, etc.) [Deutsch, 1996]. Dans le détail, deux lois sont considérées dans le cas de variables continues :

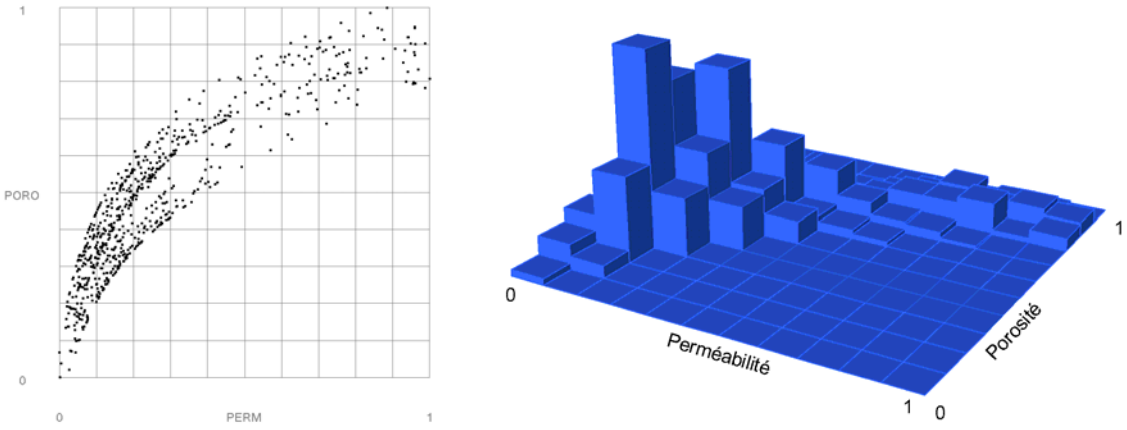


FIG. 5.1 – Exemple de calcul d'histogramme bivarié. À gauche, ensemble de points initial ; droite : histogramme expérimental avec 10 classes par variable.

- la loi de probabilité cumulée multivariée (ou "*mcdf*" pour "*multivariate cumulative distribution function*") qui correspond à la probabilité que les différentes variables étudiées soient inférieures ou égales à un ensemble de valeurs donné ;
- la loi de densité multivariée (ou "*mpdf*" pour "*multivariate probability density function*") qui correspond à la probabilité d'un événement donné, c'est-à-dire que les variables prennent un ensemble de valeurs données.

Néanmoins, ces lois sont généralement inconnues et hors de portée du modélisateur : la combinaison de processus complexes et mal connus ainsi que des problèmes d'échantillonnage rendent la plupart des lois statistiques (gaussienne, de Student, de Poisson, etc.) difficile à justifier et/ou inapplicables. C'est pourquoi, suivant l'approche fréquentiste, elles sont souvent approchées à partir d'histogrammes ou de diagrammes de fréquence expérimentaux déduits des données elles-mêmes. Cependant, dans ce cas, les résultats sont, le plus souvent biaisés, peu représentatifs de la population étudiée. Les solutions à ce problème sont soit d'adapter un modèle paramétrique aux fréquences observées, soit de lisser l'histogramme ou le diagramme de fréquence étudié.

Ce chapitre décrit une méthode non-paramétrique de lissage d'histogramme basée sur l'interpolation lisse discrète et développée durant ces travaux de thèse. Après une revue des principales approches proposées dans la littérature pour résoudre ce problème, nous préciserons l'approche choisie et l'illustrerons sur un exemple synthétique.

5.1 Lissage d'histogramme : définitions, objectifs et méthodes

5.1.1 Histogramme et diagramme de fréquence

Un histogramme est un outil visuel pour appréhender la distribution d'une population donnée. Considérons une variable aléatoire X et divisons son intervalle de variation en un nombre

fini de classes. L'histogramme correspond au nombre de fois qu'une classe est représentée dans un échantillon donné de X. Le diagramme de fréquence est simplement un histogramme normalisé par le nombre de mesures et prenant ses valeurs dans l'intervalle [0, 1]. Il est naturellement possible d'étendre ces concepts à un ensemble de variables pour obtenir des lois multivariées (voir, par exemple, figure (5.1)). Ces outils permettent d'apprécier la répartition des données, leurs intervalles de variation ainsi que les fréquences relatives des différentes classes [Soong, 1981; Isaaks et Srivastava, 1989; Deutsch, 1996].

Suivant l'approche fréquentiste, un diagramme de fréquence calculé à partir des données observées est une approximation de la loi de densité. Cela est d'autant plus vrai que la taille de l'échantillon augmente et que la taille des différentes classes diminue. Cependant, cette approche, particulièrement en Géosciences, a de nombreuses limitations : (1) les données disponibles sont peu abondantes, (2) elle couvrent rarement l'ensemble de l'intervalle de variation de la population étudiée et (3) elles sont souvent peu représentatives. D'autre part, la largeur et le nombre des classes influent de façon notable sur l'aspect global des diagrammes expérimentaux : cela conduit soit à perdre les variations dans la distribution de faible résolution ou bien à des fluctuation erratiques.

Les solutions sont soit d'adapter un modèle analytique aux fréquences observées (approche paramétrique) ou bien de lisser le diagramme de fréquence étudié (approche non-paramétrique). L'approche paramétrique [Johnson et Kotz, 1970; Scott, 1992] repose sur la connaissance a priori du type de la distribution (par exemple uniforme, triangulaire, gaussienne, etc.) et consiste à optimiser ses paramètres vis-à-vis des données observées. Le modèle obtenu, défini de manière continue, ne présente plus de problèmes liés à la résolution des différentes classes. Cependant, le résultat dépend du choix du modèle de distribution et il est difficile de vérifier son adéquation avec les données. À l'opposé, une approche non-paramétrique [Silverman, 1986; Scott, 1992; Deutsch, 1996] estime la loi de densité finale à l'aide de techniques d'interpolation et de lissage de fonctions discrètes. Ce procédé retire les fluctuations dans le diagramme, augmente la résolution des différentes classes et permet d'étendre la distribution au-delà des valeurs minimales et maximales de l'échantillon [Isaaks et Srivastava, 1989; Deutsch et Journel, 1992].

5.1.2 État de l'art des méthodes de lissage d'histogramme

Les principales méthodes de lissage publiées dans la littérature sont le plus souvent basées sur l'utilisation de noyaux ou de fonctions splines :

– L'utilisation de fonctions noyaux (ou *kernel function*) initialement proposée pour le lissage de loi univariée [Parzen, 1962] a été généralisé et étendue au traitement de données multivariées grâce aux travaux de Terrell et Scott [1992]. La valeur de la loi de densité en un point donné est alors estimée comme une combinaison linéaire des observations voisines. Les différents poids sont calculés à partir d'une fonction noyau et de la distance entre les observations et la position étudiée. Différentes fonctions peuvent être employées ; leur choix s'appuie sur des considérations telles que la différenciabilité du résultat ou le coût en calcul.

– A l'opposé, les splines sont des fonctions polynômiales définies par morceaux dont la forme est contrôlée par un ensemble de nœuds dont la position est optimisée au sens des moindres carrés par rapport aux données observées [Eubank, 1988]. Cette approche est similaire aux méthodes paramétriques, mais elle est aussi beaucoup plus flexible car aucune hypothèse n'est faite a priori sur le type de la distribution.

Ces deux approches sont de manière générale adaptées à l'analyse de données statistiques dans un cadre classique. Cependant, comme souligné par Jones [1991] et Deutsch [1996], (1) elles échouent si les données sont trop peu nombreuses, (2) elles peuvent conduire à des valeurs absurdes (c'est-à-dire en dehors des limites physiques d'une variable, par exemple [0,1] pour la porosité), et (3) elles ne peuvent pas honorer simultanément différentes statistiques (moyenne, variance, quantiles, etc.) de la distribution. En outre, elles n'autorisent pas un contrôle des valeurs extrêmes. Or celui est primordial, en effet ces valeurs sont à la fois les pus intéressantes et peu échantillonnées.

En ingénierie de réservoir, les données effective du champ étudié sont peu nombreuses et suivent un échantillonnage biaisé : il est fréquent que les parties du réservoir intéressantes (c'est-à-dire à forte porosité et perméabilité) soient plus représentées dans l'échantillon. Il est donc courant de considérer un certain nombre de mesures statistiques (type de distribution, moyenne, variance, etc.) provenant d'analogues ou de l'avis d'experts. Les géostatisticiens ont donc développés leurs propres routines de lissage. La plus répandues est basée l'utilisation du recuit simulé [Deutsch, 1996] et a été implantée dans la GSLIB [Deutsch et Journel, 1992]. Cette approche consiste à perturber de façon itérative un diagramme de fréquence expérimental tout en minimisant une fonction objective donnée. Cette fonction comprend différents termes contrôlant :

– la différence entre le diagramme initial et le résultat lissé ;
– la rugosité du résultat final ;
– le respect d'un certain nombre de mesures statistiques : dans un cas univarié, il s'agit de la moyenne, de la variance ou de quantiles de la distribution ; dans le cas d'une loi multivariée, s'ajoutent des mesures de covariance entre variables ou leurs distributions marginales.

Cette approche, relativement efficace, s'applique pour des des jeux de données et une et deux dimensions [1]. Une alternative proposée par Mallet et Sthuka [2000] est l'utilisation de l'interpolation lisse discrète pour estimer la loi de densité à partir d'un histogramme grossier tout en respectant au sens des moindres carrés un certain nombre de contraintes (moyennes, variances, covariances, etc.). Ces deux méthodes nécessitent cependant la construction d'une solution initiale cohérente de la loi de densité, ce qui n'est pas toujours possible si le nombre de variables à prendre en compte est trop élevé et que la quantité de données disponibles est trop faible.

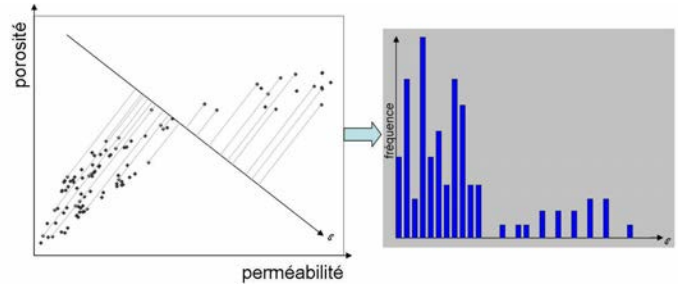


FIG. 5.2 – Exemple de construction d'un histogramme marginal suivant une direction quelconque dans l'espace des données par projection orthogonale.

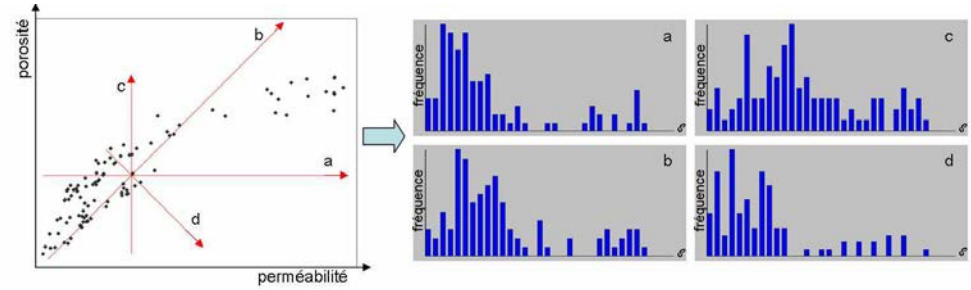


FIG. 5.3 – Ensemble d'histogrammes marginaux correspondants à la projection orthogonale d'un ensemble de données 2D suivant différentes directions distribuées régulièrement dans l'espace. Chaque histogramme permet d'illustrer différentes caractéristiques des données.

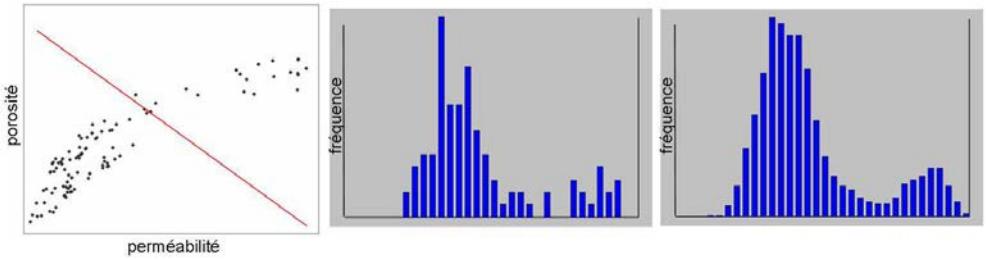


FIG. 5.4 – Exemple de lissage d'histogramme unidimensionel. On trouve de gauche à droite : les points de données initiaux, un diagramme de fréquence unidimensionnel brut et le même diagramme lissé à l'aide d'une fonction noyau. Ceci permet de le rendre plus représentatif de la population étudiée, d'accroître sa résolution et de retirer d'éventuels artefacts.

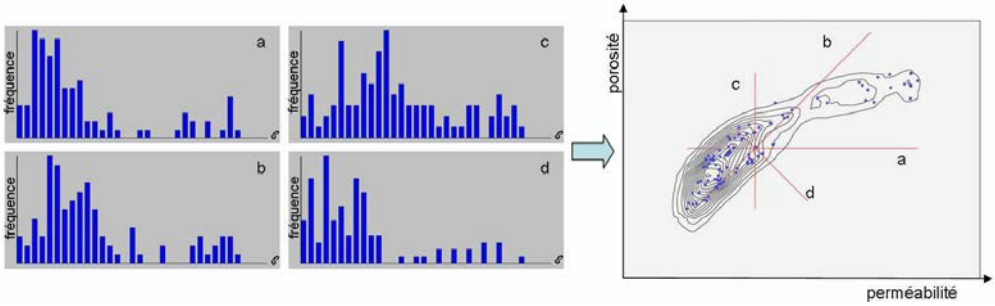


FIG. 5.5 – Interpolation lisse discrète d'un diagramme de fréquence multivarié à partir d'histogrammes marginaux.

5.2 Approche proposée

5.2.1 Principes

La méthodologie développée durant ces travaux s'appuie sur l'exploration et la caractérisation des données par le biais d'un grand nombre d'histogrammes (ou de diagrammes de fréquence) marginaux calculés suivant différentes directions de l'espace. L'idée principale est alors de les introduire sous la forme de contraintes lors de l'interpolation lisse discrète de la loi de densité. L'intérêt de cette approche est de limiter la quantité de données nécessaires. En effet, là où les autres méthodes nécessitent l'estimation d'un histogramme en dimension n, ici, il s'agit de construire plusieurs histogrammes, certes en grand nombre, mais unidimensionels. En pratique, la procédure est divisée en quatre étapes :

1. Calcul des diagrammes de fréquence marginaux (figures (5.2) et (5.3)) : ceux-ci sont simplement calculés par projection orthogonale des données le long de lignes reparties dans l'espace. Ces lignes sont construites de telle manière (1) qu'elles se croisent au centre de gravité des données et (2) qu'elles soient distribuées de manière uniforme dans les différentes directions de l'espace. Ces diagrammes illustrent alors différentes caractéristiques (valeur moyenne, dispersion, asymétrie, etc.) de la population étudiée ;
2. Lissage, si nécessaire, de ces diagrammes à l'aide de techniques classiques telles que l'utilisation de noyaux ou de fonction splines (figure (5.4)). L'objectif est d'accroître leur résolution et d'éliminer d'éventuels artefacts.
3. Interpolation du diagramme de fréquence multivarié à l'aide de la méthode DSI (figure (5.5)) en tenant compte (1) des différents diagrammes marginaux et (2) d'éventuelles informations sur les mesures statistiques de la distribution finale (moyennes, variances, covariances, etc.) ;
4. Normalisation des résultats qui doivent satisfaire la condition suivante :

$$P(-\infty < x_1 < +\infty \cap \cdots \cap -\infty < x_n < +\infty) = 1 \tag{5.1}$$

   Idéalement il est possible d'implanter la relation (5.1) sous forme d'une contrainte pour l'interpolation lisse discrète. Cependant, au vu de son coût prohibitif en termes de calcul, il est préférable de procéder à un post-processus.

5.2.2 Contraintes pour l'interpolation lisse discrète

Cette section décrit les différentes contraintes nécessaires à l'interpolation lisse discrète d'une loi de densité multivariée suivant l'approche proposée. Ces contraintes ont été développées et implantées dans le même cadre que celui utilisé pour l'interpolation de surfaces de réponse. En particulier, ce travail s'appuie sur les structures de données représentant un ensemble de points et une grille structurée régulière dans un espace de dimension quelconque. Là encore, le résultat final se présentera sous la forme d'une propriété définie aux nœuds de la grille.

[1] Les fonctionnalitées correspondantes implantées dans la GSLIB sont respectivement, **histsmth** et **scatsmth** [Deutsch et Journel, 1992].

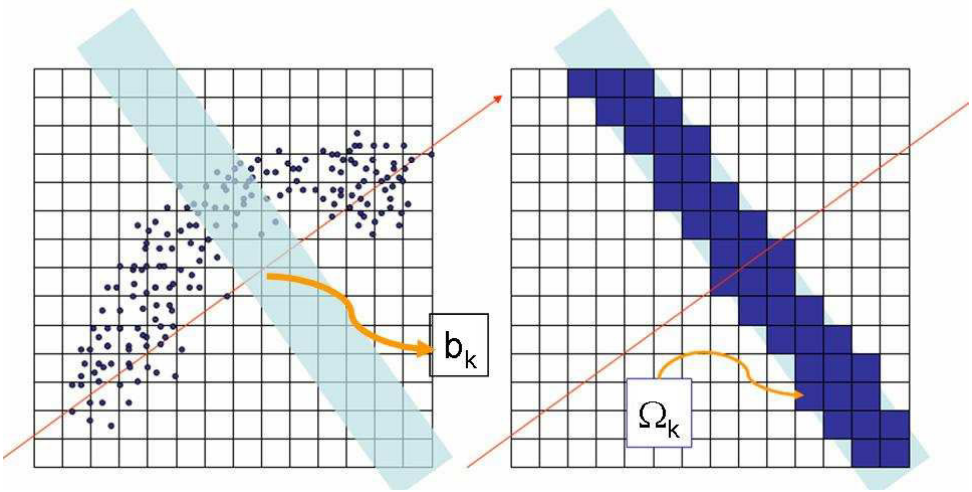


FIG. 5.6 – Contrainte DSI pour imposer le respect d'un diagramme de fréquence

Notations Dans la description des différentes contraintes, les notations suivantes, cohérentes avec celles employées par Mallet [2002], ont été utilisées :

- $X_1, \cdots, X_n$ représente les n variables aléatoires dont la loi de densité multivariée doit être estimée ;
- β représente un nœud donné de la grille régulière structurée support de l'interpolation et $\mathbf{x}(\beta) = [x_1(\beta), \cdots, x_n(\beta)]^t$ sa position ;
- Ω est l'ensemble des nœuds β du modèle discret ;
- et $f_{X_1 \ldots X_n}(\beta)$ représente la valeur de la loi de densité au nœud β.

Contrainte de diagramme de fréquence marginal

La méthode proposée repose sur l'utilisation d'un certain nombre de diagrammes de fréquence marginaux lors de l'interpolation de la loi de densité. En pratique, le respect de ces diagrammes est imposé par le biais d'un ensemble de contraintes DSI mises en place pour chacune de de ses classes. Soit une classe donnée, la contrainte DSI correspondante (1) considère tous les nœuds du modèle discret qui, par projection orthogonale, appartiennent à cette classe et (2) spécifie que l'intégration de la loi de densité sur ces nœuds est égale à la fréquence associée à la classe.

Considérons un diagramme de fréquence donné et notons $c_k, k \in \{1, \ldots, n_k\}$ l'une de ses classes ainsi que b_k la fréquence qui lui est associée. Notons, en outre, Ω_k l'ensemble des nœuds β de Ω qui, par projection orthogonale, appartiennent à c_k (figure (5.3)). La contrainte à imposer est :

$$\sum_{\beta \in \Omega_k} f_{X_1 \ldots X_n}(\beta) = b_k \tag{5.2}$$

Il vient alors que les coefficients de la contrainte DSI correspondante sont :

$$\left| \begin{array}{rcl} A_c(\beta) & = & 1, \ \forall \beta \in \Omega_k \\ A_c(\beta) & = & 0, \ \text{sinon} \\ b_c & = & b_k \end{array} \right. \tag{5.3}$$

Un problème lors de la construction des diagrammes de fréquence marginaux est le choix de la largeur des différentes classes. En effet, comme souligné par Soong [1981], celle-ci influence directement leurs fréquences relatives et l'information qui peut être extraite des données. Cela peut conduire à la perte d'informations de haute résolution si la largeur choisie est trop grande ou bien à des fluctuations erratiques dans le cas contraire. Une seconde contrainte est la taille des cellules de la grille utilisée lors de l'interpolation. En effet, lors de la mise en place des contraintes, il est nécessaire qu'à chaque classe corresponde un certain nombre de nœuds du maillage. Différents tests ont montré que choisir une largeur de classe supérieure ou égale à deux fois la taille des cellules de la grille donne des résultats satisfaisants. En outre, signalons que la résolution de la grille de calcul détermine largement les performances de la méthode (voir l'étude présentée section (2.6.3)). Il est ainsi nécessaire de trouver un équilibre entre la quantité de données disponibles et la taille de la grille de calcul.

Dans la suite, nous utiliserons l'heuristique proposée par Sturges [1926] repris dans Soong [1981] qui suggère que le nombre de classes k d'un diagramme de fréquence 1D, soit déterminé à partir du nombre de données disponibles n, tel que :

$$k = 1 + 3.3 \log_{10} n \tag{5.4}$$

et nous adapterons en conséquence la résolution de la grille sur laquelle est réalisée l'interpolation.

Contraintes sur les mesures caractéristiques d'une distribution

Dans la cadre des Géosciences, les données disponibles sont limitées en quantité mais aussi non pleinement représentatives de la population étudiée. Il est donc nécessaire, en imposant un certain nombre de caractéristiques issues de données analogues ou établie par un expert, par exemple :

- la moyenne d'une variable aléatoire ;
- la covariance entre deux variables aléatoires ;
- ou le quantile d'une des distributions marginales.

Ces contraintes peuvent être cumulées et pondérées, l'algorithme DSI ayant pour objectif de les honorer toutes aux sens des moindres carrés.

Moyenne d'une variable aléatoire La moyenne m_i d'une variable aléatoire quelconque X_i, peut facilement être estimée à partir de la loi de densité multivariée suivant la relation suivante :

$$m_i = \int_{-\infty}^{+\infty} x_i \; f_{X_i}(x_i) \; dx_i \tag{5.5}$$

avec

$$f_{X_i}(x_i) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_{X_1,\dots,X_n}(\mathbf{x}) \; d\mathbf{x}/dx_i \tag{5.6}$$

où $d\mathbf{x}/dx_i$ indique une intégration sur l'ensemble des variables sauf x_i. Afin de pouvoir imposer cette moyenne lors de l'interpolation, il est tout d'abord nécessaire de linéariser l'équation (5.5)

sur le modèle discret Ω, tel que :

$$m_i \simeq \sum_{\beta \in \Omega} x_i(\beta)\ f_{X_1,\dots,X_n}(\beta) \tag{5.7}$$

Les coefficients de la contrainte DSI correspondante sont alors :

$$\left| \begin{array}{rcl} A_c(\beta) & = & x_i(\beta) \quad \forall \beta \in \Omega \\ b_c & = & m_i \end{array} \right. \tag{5.8}$$

Covariance de deux variables aléatoires Considérons deux variables aléatoires X_i et X_j dont les moyennes, respectivement m_i et m_j, sont connues. Leur covariance, notée C_{ij}, est calculée classiquement à partir de la loi de densité multivariée telle que :

$$C_{ij} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_i\ x_j\ f_{X_i,X_j}(x_i,x_j)\ dx_i dx_j - m_i\ m_j \tag{5.9}$$

avec

$$f_{X_i,X_j}(x_i,x_j) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_{X_1,\dots,X_n}(\mathbf{x})\ d\mathbf{x}/(dx_i dx_j) \tag{5.10}$$

De même que précédemment, cette relation est ensuite linéarisée sous la forme :

$$C_{ij} \simeq \sum_{\beta \in \Omega} \{x_i(\beta) \cdot x_j(\beta)\}\ f_{X_1,\dots,X_n}(\beta) - m_i \cdot m_j \tag{5.11}$$

Les coefficients de la contrainte DSI sont alors :

$$\left| \begin{array}{rcl} A_c(\beta) & = & x_i(\beta) \cdot x_j(\beta) \quad \forall \beta \in \Omega \\ b_c & = & C_{ij} + m_i \cdot m_j \end{array} \right. \tag{5.12}$$

Il est intéressant de noter que si $i = j$, alors C_{ii} correspond à la variance marginale (σ_i^2) de la variable aléatoire X_i. Il est alors possible de l'imposer comme contrainte mais avec les coefficients suivants :

$$\left| \begin{array}{rcl} A_c(\beta) & = & x_i^2(\beta) \quad \forall \beta \in \Omega \\ b_c & = & \sigma_i^2 + m_i^2 \end{array} \right. \tag{5.13}$$

Pour finir, deux remarques sur cette contrainte peuvent être faites : tout d'abord, il s'agit d'imposer une mesure de covariance *linéaire*, il est donc conseillé, si l'on souhaite utiliser ce type de contrainte de procéder à une transformation des données ; d'autre part, cette contrainte nécessite de connaître préalablement les moyennes m_i et m_j des variables étudiées, or il est important de souligner que ces mesures sont rarement robustes en particulier dans le cas de jeu de données de taille limitée.

Quantile d'une distribution Notons q_i le quantile d'une variable aléatoire X_i et correspondant à une valeur de fréquence cumulée Q_i. Par définition, la relation entre ces deux variables et la loi de densité multivariée est :

$$Q_i = P(X_1 < +\infty, \cdots, X_i < q_i, \cdots, X_n < +\infty) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{q_i} \dots \int_{-\infty}^{+\infty} f_{X_1,\cdots,X_n}(\mathbf{x}) \cdot d\mathbf{x} \tag{5.14}$$

Considérons maintenant Ω_{Q_i} l'ensemble des nœuds β du maillage discret tel que $x_i(\beta) < q_i$. Il est alors possible de linéariser l'équation (5.15) sous la forme :

$$Q_i \simeq \sum_{\beta \in \Omega_{Q_i}} f_{X_1,\cdots,X_n}(\beta) \tag{5.15}$$

ce qui conduit naturellement aux coefficients de contrainte DSI suivants :

$$\left| \begin{array}{rcll} A_c(\beta) & = & 1 & \forall \beta \in \Omega_{Q_i} \\ A_c(\beta) & = & 0 & \text{sinon} \\ b_c & = & Q_i & \end{array} \right. \tag{5.16}$$

L'intérêt à pouvoir imposer le quantile d'une distribution est double : tout d'abord les quatiles sont souvent des mesures relativement robuste, ainsi la médiane d'une distribution (i.e. le P_{50}) est moins sensibles aux données extrêmes que sa moyenne ; et d'autre part les quantiles permettent de contraindre les valeurs extrêmes d'une distribution par exemple le P_{10} ou le P_{90}. En effet, ces valeurs sont, d'un point de vue de la modélisation, les plus intéressantes mais aussi relativement mal échantillonnées.

Remarques Les contraintes décrites ci-dessus permettent d'imposer des mesures statistiques globales de la distribution. Un problème reste cependant ouvert : doit-on et peut-on traduire ces informations sur les histogrammes marginaux utilisés lors de l'interpolation ? Ou bien doit on laisser la possiblité à la technique d'interpolation de choisir de respecter ou non ces informations vis-à-vis des données elle-mêmes ?

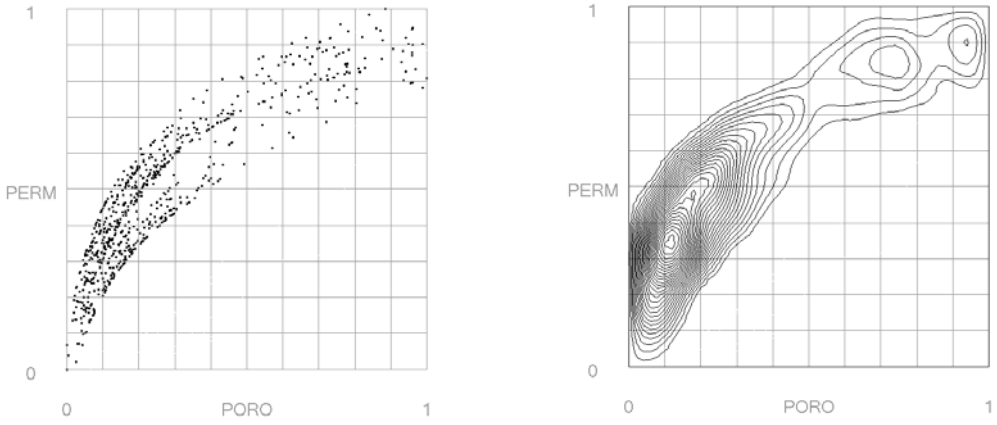
5.2.3 Résultats préliminaires

Cette section présente un exemples de lissage d'histogramme pour illustrer et discuter la méthodologie proposée. Le jeu de données consiste en 720 points de calibration entre des mesures de perméabilité et de porosité dans un réservoir pétrolier.

La figure (5.7) présente les résultat obtenus après lissage dans le cas où aucune mesure statistique externe n'est prise en compte. L'interpolation s'appuie sur le calcul en pré-processus de 24 diagrammes marginaux unidimensionels de 30 classes chacun et distribués de manière uniforme dans l'espace. La loi de densité obtenue est relativement satisfaisante : elle respecte les données initiales et ne présente pas de fluctuations erratiques. Cet exemple illustre la capacité de l'approche proposée a combiner différents histogrammes marginaux pour reconstruire un diagramme global bi-dimensionnel.

5.3 Analyse de la sensibilité de la méthode à ses différents paramètres

Il est particulièrement difficile de juger de valider les lois de densité de probabilité obtenues et ce quelque soit l'approche utilisée. Il n'existe pas d'indicateur simple et quantifiable qui pourrait



(a) Données initiales

(b) Diagramme de fréquence lissé

FIG. 5.7 – Exemple de lissage de loi de densité bivariée de mesures de porosité et de perméabilité : (a) données initiales composée de 720 points de calibration (b) résultat du lissage suivant la méthodologie proposée sans contrainte statistique globale spécifique.

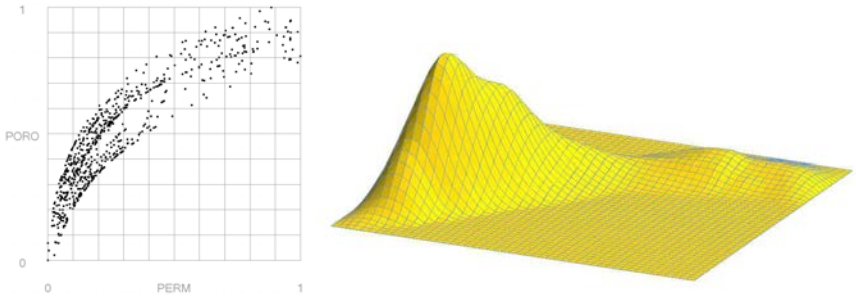
servir de base à une comparaison. Dans ces travaux, les critères envisagés sont essentiellement qualitatifs et visuels. Dans le détail, nous nous sommes intéressés à vérifier que le résultat (1) font apparaître les structures déjà présentes dans les donnés (en particulier l'existence de deux populations et la forme incurvée de la loi de densité), (2) est globalement lisse et (3) ne contient pas d'artefact ou de fluctuation erratique. En outre nous avons évaluer la robustesse du calcul ainsi que le temps d'exécution.

Différents facteurs interviennent dans la procédure de lissage et peuvent influer sur les résultats. On distingue, notamment :

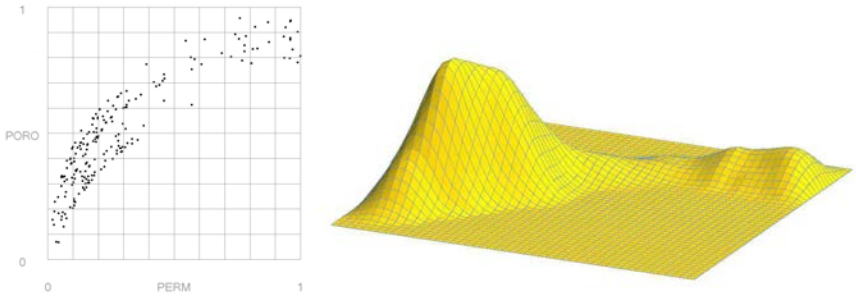
- le nombre de points de données disponibles ;
- le nombre de diagrammes de fréquence marginaux envisagés comme contrainte lors de l'interpolation ;
- le nombre de classes ou de bins pour chacun de ces diagrammes ;
- la procédure de lissage des diagrammes univariés intermédiaires employés pour contraindre l'interpolation finale.

Afin d'étudier leurs impacts respectifs sur la méthode de lissage, la procédure décrite précédemment est réemployée avec différentes valeurs pour le paramètre étudié :

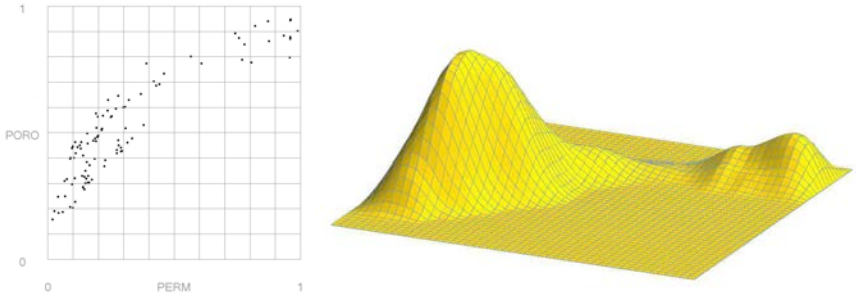
- La figure (5.8) illustre les résultats obtenus à partir de jeux de données de taille de plus en plus réduite (720, 200, 100 et 50 points de calibration) et générés par échantillonnage aléatoire des données initiales. Les différents diagrammes de fréquence lissés font apparaître globalement la même structure et les mêmes deux sous populations pour d'une part des valeurs faibles de porosité/perméabilité et d'autre part des valeurs élevées. Ces résultats sont pleinement satisfaisants et semblent indiquer que l'approche proposée est suffisamment robuste pour prendre en compte des jeux de données de taille réduite.



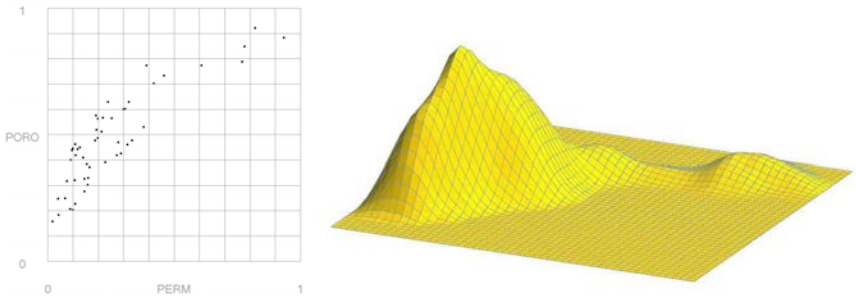
(a) Données initiales (720 points)



(b) 200 points

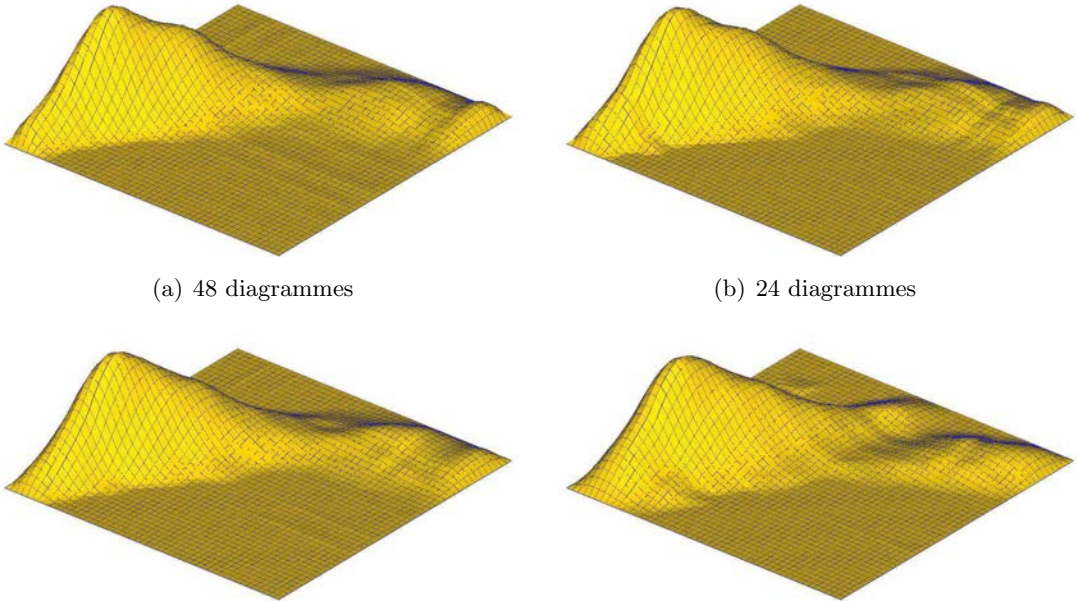


(c) 100 points



(d) 50 points

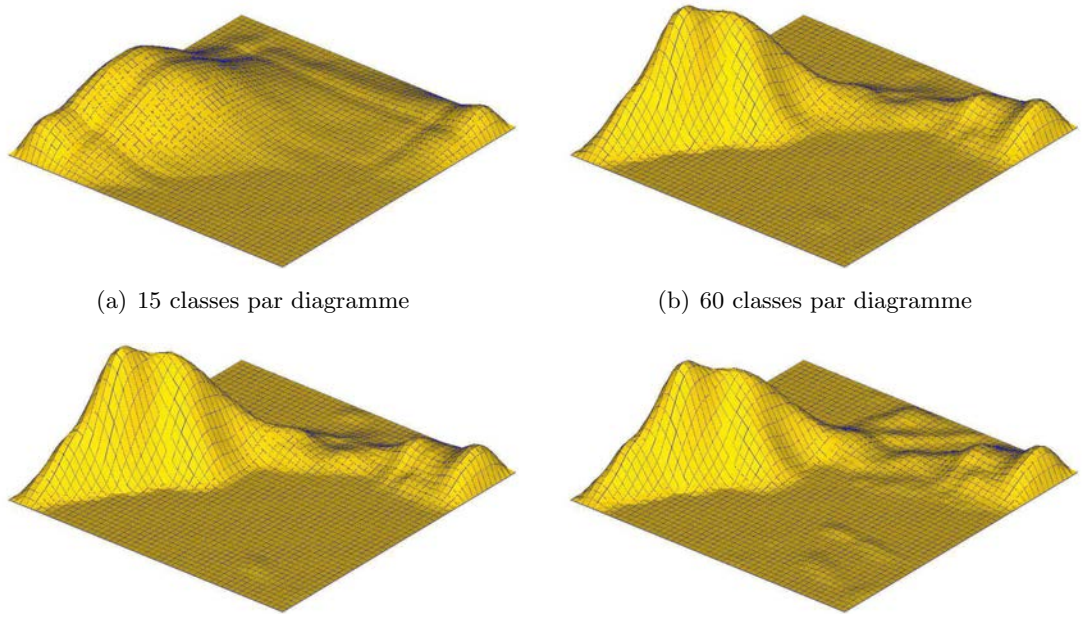
FIG. 5.8 – Influence du nombre de points de données disponibles sur la procédure de lissage. Les ensembles de 200, 100 et 50 points ont été obtenus par échantillonnage aléatoire des données initiales.



(a) 48 diagrammes (b) 24 diagrammes

(c) 12 diagrammes (d) 6 diagrammes

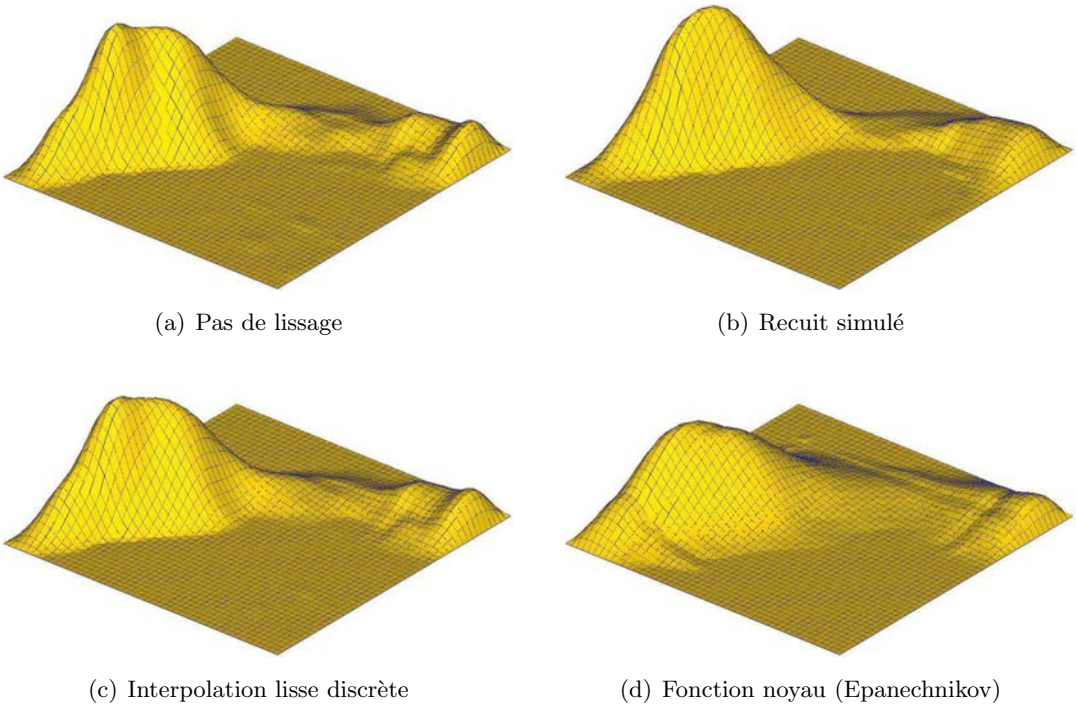
FIG. 5.9 – Influence du nombre de diagrammes marginaux considérés comme contraintes lors de la procédure de lissage.



(a) 15 classes par diagramme (b) 60 classes par diagramme

(c) 100 classes par diagramme (d) 200 classes par diagramme

FIG. 5.10 – Influence du nombre de classes des différents diagrammes marginaux considérés comme contraintes lors de la procédure de lissage.



(a) Pas de lissage

(b) Recuit simulé

(c) Interpolation lisse discrète

(d) Fonction noyau (Epanechnikov)

FIG. 5.11 – Influence d'une procédure de lissage intermédiaire des diagrammes de fréquence marginaux. Les méthode envisagées sont : le recuit simulé [Deutsch, 1996], l'interpolation lisse discrète suivant la procédure présentée précédemment mais sur un diagramme univarié et l'utilisation de la fonction noyau d'Epanechnikov telle que proposée dans Mallet et Sthuka [2000].

– La figure (5.9) montre l'impact du changement du nombre de diagrammes marginaux pris en compte comme contrainte lors de l'interpolation. Là encore, les résultats sont cohérents et montrent les principales structures présentes dans les données. On note cependant une perte d'informations sur les variations de fréquence de faible amplitude ainsi que l'apparition d'artefacts dus à l'interpolation.

– La figure (5.10) s'intéresse à l'influence du nombre de classes envisagées lors de la construction des diagrammes marginaux. Il apparaît, comme attendu, que plus ce nombre est important plus le résultat final est précis et représentatif de la population étudiée. Cependant à partir d'un certain niveau, ici 200 classes, celui-ci présente un certains nombre d'artefacts. Deux explications sont possibles : (1) les données ne sont pas suffisantes pour construire des diagrammes marginaux représentatifs de la population étudiée (ils présentent alors un certain nombre de fluctuations erratiques) et, (2) lors de la mise en place des contraintes, certaines classes ne sont plus associées à des nœuds de la grille et ne sont simplement pas prises en compte lors le l'interpolation.

– La figure (5.11) expose les loi de densité obtenues avec différentes procédures de lissage intermédiaires de diagrammes marginaux. Ce choix de modélisation a un impact fort sur le résultat et doit être évalué avec soin. Dans le détail, on constate que (1) l'utilisation d'une étape de lissage permet d'éliminer un certain nombre de fluctuations dans la solution

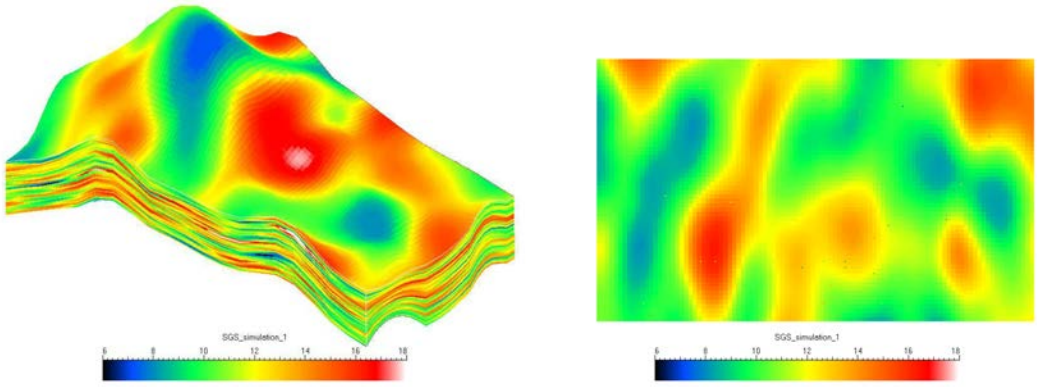


FIG. 5.12 – Champ de porosité obtenu à partir d'une simulation gaussienne séquentielle basée sur un variogramme de type gaussien : (a) vue d'ensemble et (b) section k=24.

finale et (2) que parmi les méthodes employées, la fonction noyau semble présenter les plus mauvais résultats (en particulier on note la présence d'artefacts non représentatifs de la population étudiée dans le résultat final).

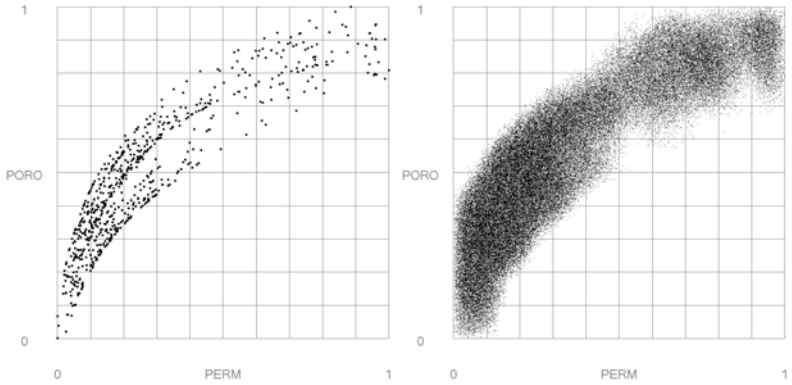
5.4 Prise en compte de données lissée pour la simulation de propriété par "*cloud tranform*"

La méthode du "*cloud transform*" est une technique de simulation géostatistique qui a pour objectif de calculer différentes réalisations d'une propriété dans un réservoir à partir (1) d'une donnée primaire connue de manière exhaustive et (2) de la loi de densité bivariée entre les deux variables. Cette approche est particulièrement intéressante dans les cas où certains types de données sont difficiles à obtenir, par exemple la perméabilité qui ne peut pas être mesurée in-situ dans le puits. Il est donc nécessaire de les relier à d'autres propriétés plus accessibles et facilement mesurables comme la porosité.

Si la loi de densité n'est pas disponible, elle peut être estimée à partir d'un ensemble de points de calibration. La procédure suivie consiste alors à découper la variable primaire en classes puis, pour chacune de ces classes, à construire la distribution cumulée expérimentale de la variable à estimer.

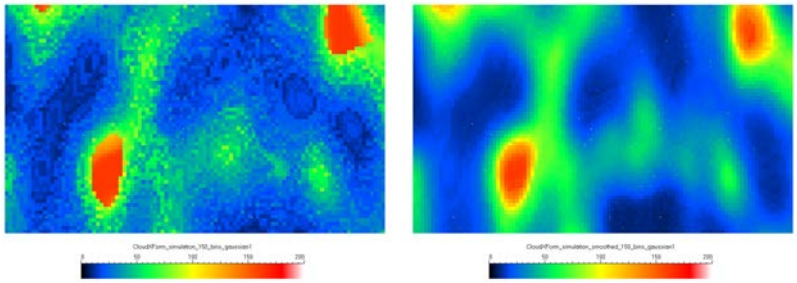
L'exemple proposé s'appuie sur un modèle synthétique de réservoir illustré figure (5.12). Dans un premier temps, la porosité a été simulée à partir des données de puits. La méthode employée est une simulation séquentielle gaussienne [Goovaerts, 1997] basée sur un variogramme de type gaussien. La perméabilité est ensuite simulée suivant la méthode de "*cloud transform*" à partir de cette carte et des données de calibration présentées section (5.3). Deux cas sont successivement envisagés :

1. aucun lissage n'est effectué : la loi bivariée est estimée directement à partir des 720 points de données disponibles ;



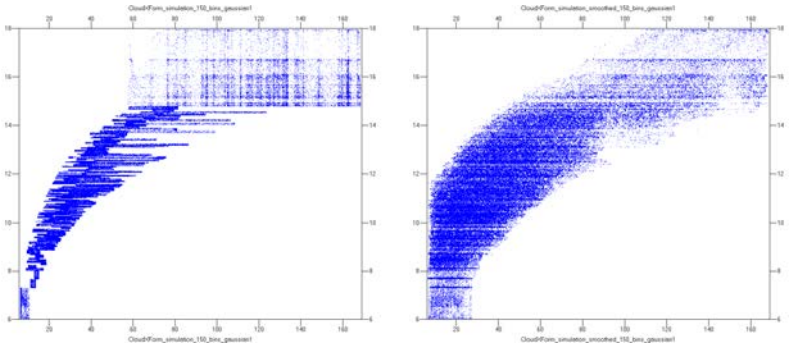
(a) sans lissage (b) après lissage

FIG. 5.13 – Ensembles de points de calibration pris en compte dans la méthode du "*cloud transform*" dans les deux cas : à gauche, données initiales sans lissage ; à droite, données obtenues par échantillonnage de la loi de densité bivariée après lissage suivant l'approche proposée.



(a) sans lissage (b) après lissage

FIG. 5.14 – Réalisations de perméabilité obtenues suivant la méthode "*cloud transform*" à partir du champ de porosité (figure (5.12)) et des ensembles de points de calibration présentés ci-dessus. Ici, de même que la simulation de la porosité, la fonction de covariance de la perméabilité est de type gaussien.



(a) sans lissage (b) après lissage

FIG. 5.15 – Graphiques 2D représentant la relation entre les valeurs de porosité et de perméabilité dans le réservoir dans les cas sans lissage (à gauche) et prenant en compte les données lissées (à droite).

2. les données initiales sont lissées suivant l'approche proposée. La loi de densité bivariée est alors échantillonnée suivant la règle de Métropolis : cela permet d'étoffer l'échantillon de points de calibration et de mieux caractériser la relation entre la porosité et la perméabilité.

Les résultats obtenus sont exposés et comparés figures (5.13) à (5.15) :

- la figure (5.13) montre les deux ensembles de points pris en compte dans la méthode du "*cloud transform*" pour calibrer la relation entre la porosité et la perméabilité ;
- la figure (5.14) compare les réalisations de perméabilité sur une section du modèle du réservoir ;
- la figure (5.15) expose pour les deux cas la relation existant dans le réservoir entre les valeurs simulées de porosité et de perméabilité.

Ces résultats illustrent l'intérêt du lissage de la loi bivariée et de son échantillonnage. Lors de la simulation, à partir des données lissées, il est possible de considérer un plus grand nombre de classes de porosité et pour chacune d'entre elles la distribution cumulée de perméabilité locale est beaucoup plus lisse et semble plus représentative de la population étudiée. Il n'y a plus les fluctuations erratiques dans la réalisation de perméabilité qui ne sont dues qu'à un manque de donnée (figure (5.14)). Cependant les donnés initiales semble être hétéroscédastique, c'est-à-dire que la variance de la porosité change en fonction de la perméabilité et vice-versa. Or cela n'apparaît plus de façon aussi claire dans les données. Il serait intéressant de déterminer si cela est inhérent à toute méthode de lissage ou spécifique à celle proposée dans ces travaux.

5.5 Conclusions

Pour résoudre le problème de lissage d'histogramme ou de diagramme de fréquence multivarié nous avons développé une approche basée sur l'interpolation lisse discrète. L'idée fondamentale est de procéder, dans un premier temps, à l'exploration des données à l'aide de diagrammes de fréquence marginaux puis, dans un second temps, à les utiliser comme contrainte lors de l'interpolation. En outre nous avons proposé un ensemble de contraintes pour la prises en compte de statistiques externes (moyenne, variance, etc.) issues de l'analyse de données analogues ou établie par un expert. L'intérêt de cette méthode est grand en Géosciences, où il est fréquent de devoir traiter des donnés limitées en quantité mais aussi pas toujours pleinement représentatives de la population étudiée.

Les résultat obtenus montrent que cette technique permet d'augmenter la résolution des différentes classes, d'étendre la distribution au-delà des valeurs minimales et maximales de l'échantillon observé et d'éliminer les fluctuations erratiques dues au manque de données. Ainsi, dans le cadre de la simulation des propriétés du sous-sol, l'utilisation du lissage permet d'obtenir des résultats plus robustes et représentatifs de la population et du phénomène étudié.

Conclusions et perspectives

Conclusions

La quantification des incertitudes constitue une étape clé de la modélisation du sous-sol : d'une part, les données disponibles sont limitées, éparses et hétérogènes en qualité et en résolution, d'autre part, le modèle obtenu n'est qu'une approximation de la réalité et comporte de nombreuses inconnues. L'ensemble de ces incertitudes influence de manière considérable les résultats de la modélisation et doit être pris en compte car celle-ci répond à un problème économique : elle doit pouvoir servir de base à la prise de décisions dans l'exploration et la gestion d'un gisement. Dans ce cadre, les simulations d'écoulement prennent une place singulière car elles permettent d'évaluer la production d'un réservoir. Cependant, elles sont particulièrement coûteuses en temps et en calculs et l'exploitation qualitative des résultats est difficile à cause de la nature complexe de la dépendance des résultats de simulation vis-à-vis des paramètres du modèle.

L'objectif de ces travaux de thèse était double, il s'agissait de développer, d'une part, un simulateur sur lignes de courant à l'échelle d'un réservoir pétrolier ; d'autre part, un ensemble de méthodes basées sur l'utilisation de surface de réponse pour caractériser et analyser la relation, souvent inconnue et non-linéaire, entre un ensemble de paramètres incertains et la production d'un réservoir à hydrocarbures.

Dans un premier temps, à partir des travaux de Audigane [2000] et Voillemont [2003], nous avons développé le plugin Gocad StreamLab-2.0.0 dédié à la simulation d'écoulement biphasique sur lignes de courant à l'échelle d'un réservoir pétrolier présentant de fortes hétérogénéités. Par rapport aux travaux précédents, le code de calcul a été spécialisé sur les grilles structurées hexaédriques, et des algorithmes plus rapides, bien que moins précis, ont été sélectionnés et mis en œuvre. Une seconde amélioration concerne l'implantation d'une structure de donnée, plus compacte et efficace, pour représenter et manipuler les lignes de courant. L'objectif de ce travail n'est pas de fournir un simulateur d'écoulement de type commercial, mais un ensemble d'outils pour effectuer une évaluation rapide du potentiel de production d'un réservoir pétrolier.

La seconde partie constitue le cœur des travaux présentés dans ce mémoire :

– Le premier point a été le développement et l'évaluation de l'interpolation lisse discrète et du formalisme dual du krigeage pour la construction d'une surface de réponse. La validation de ces méthodes sur différents exemples laisse apparaître que le krigeage dual est la méthode

la plus adaptée au problème posé au vu du nombre de paramètres incertains et de la quantité limitée de données de simulation à prendre en compte. En parallèle, nous avons sélectionné et implanté une méthode d'analyse de sensibilité basée sur la variance et une technique d'inversion bayésienne pour tenir compte de données historiques de production.

– Une contribution innovante de ces travaux est l'intégration d'une donnée secondaire dans la construction de surface de réponse. La technique originale proposée combine les résultats de simulations d'écoulement conventionnelles avec une surface de réponse secondaire issue de simulations rapides alternatives. Là encore, les techniques d'interpolation utilisées sont l'interpolation lisse discrète et le krigeage dual avec dérive externe. L'application de cette technique sur un exemple de réservoir pétrolier a montré que la prise en compte de simulations alternatives requiert un surcoût en temps parfaitement acceptable par rapport à une approche conventionnelle au vu de l'amélioration de la prédictivité des surfaces de réponse.

– Puis, nous avons mis l'accent sur la prise en compte de paramètres incertains de type stochastique caractérisés par un effet purement aléatoire sur la production et ne pouvant être inclus dans une approche déterministe classique. À partir de la méthode proposée par Zabalza-Mezghani [2000], nous avons développé et adapté les techniques d'analyses de sensibilité et d'inversion bayésienne. La méthodologie a été validée sur un exemple de réservoir fluviative. Cet exemple insiste en outre sur la nécessité de prendre en compte ce type de paramètre : il influe fortement la production et s'il n'est pas pris en compte lors du calage d'historique, le modèle de référence n'est pas estimé vraisemblable. Nous proposons aussi d'étendre l'utilisation de cette méthodologie au traitement de paramètres continus mais peu influents sur la production.

– Enfin, le dernier chapitre présente une méthode de lissage d'histogramme dans un espace mutli-dimensionnel. Cette méthode est basée sur l'interpolation lisse discrète et les structures de données développées durant ces travaux. Elle illustre l'application des techniques décrites dans les chapitres précédents en dehors du domaine des surfaces de réponse.

Le choix des techniques implantées s'est concentré sur les trois aspects suivants : (1) le nombre important de paramètres incertains qui doivent être pris en compte, (2) l'existence de bruit dans les données provenant le plus souvent de paramètres jugés peu influents et négligés et (3) la forme complexe de la relation entre les paramètres incertains et la production du réservoir. Ces outils ont été implantés sous la forme d'un plugin Gocad en langage C++. L'accent a été mis sur (1) l'efficacité, à la fois en terme de temps mais aussi de coût mémoire, des algorithmes et structures de données, et (2) la réutilisation du code de calcul pour le développement de nouvelles méthodes et applications

Perspectives

D'un point de vue technique, les méthodologies proposées durant ces travaux peuvent être améliorées sur plusieurs points :

1. Il n'y a pas dans la bibliothèque d'outils proposée de méthode d'analyse robuste de la qualité des surfaces de réponse, or les méthodes employées comportent certaines inconnues qui influent sur les résultats et doivent être contraintes.

2. Au sujet de l'intégration de données secondaires, un problème majeur reste la validité de la donnée secondaire elle-même. Il serait intéressant de disposer de méthodes répondant aux trois questions suivantes : Parmi toutes les simulations alternatives ou données secondaires possibles, quelle est la plus appropriée pour le problème étudié ? Comment déterminer si l'information secondaire considérée améliore effectivement la qualité des prédictions ? Enfin, si plusieurs données sont disponibles, est-il possible de les combiner ?

3. Dans le traitement de paramètres stochastiques que nous proposons, la dispersion dans la production est modélisée sous la forme dune loi normale. Une perspective serait d'évaluer d'autres types de lois et de définir des outils déterminant l'adéquation entre cette loi et la dispersion.

4. Il serait intéressant de mieux intégrer les outils proposés et développés au sein du simulateur d'écoulement lui-même afin de permettre une meilleur automatisation des procédures d'analyse.

De manière plus générale, la quantification des incertitudes liées aux résultats de simulation d'écoulement devrait être poursuivie dans deux directions principales :

- Jusqu'à présent les techniques basées sur l'utilisation de surfaces de réponse se sont surtout focalisées sur la prise en compte de paramètres incertains de type continu et de résultats de simulations conventionnelles. Il faut poursuivre l'intégration d'autres type de paramètres (par exemple stochastiques et discrets) qui sont parfois simplement ignorés, voire totalement occultés et d'autre type de simulations. Enfin il serait aussi intéressant de séparer les paramètres dits incontrôlables des paramètres contrôlables. Les premiers sont essentiellement les différents paramètres géologiques du réservoir tandis que les seconds sont des paramètres de production tels que le nombre de puits, les pressions de production etc. Ces derniers peuvent être manipulés afin d'optimiser la production et nécessitent, en effet, un traitement spécifique.
- Un frein majeur à l'analyse du comportement de la production d'un réservoir pétrolier vis-à-vis de paramètres incertains reste le coût en temps d'une simulation d'écoulement ce qui conduit à un sous-échantillonnage de l'espace des paramètres incertains. Il est donc nécessaire de développer des techniques de planification d'expérience intelligente permettant de mettre à jour les structures clé dans la réponse d'un réservoir mais aussi adaptée à la prise en compte de sources d'information et de type de paramètres plus divers.

Pour conclure, les méthodologies développées durant ces travaux nous semblent adaptées dans une certaine mesure à l'analyse de tout problème ou la relation entre un ensemble de paramètres d'entrée et de variables de sortie doit être caractérisée. Dans le cadre des Géosciences, différentes applications sont possibles comme par exemple, la caractérisation statique des réseaux de fractures par combinaison de plusieurs indicateurs de facturation [Mace et Fetel, 2006] ou la caractérisation dynamique de ces réseaux par inversion bayésienne.

Bibliographie

Alabert, F. G. et Modot, V. (1992). Stochastic Models of Reservoir Heterogeneity : Impact on Connectivity and Average Permeabilities. *Proceedings of the* 67^{th} *SPE Annual Technical Conference and Exhibition, Washingtown, D.C., U.S.A., October 4-7.* SPE 24893.

Amiotte, J., Voillemont, J., et Mallet, J. (2002). Hyper dsi : N-dimensional space interpolation. *Proceedings of the 22nd Gocad Meeting, Nancy, France, June 24-27.*

Antonov, I. A. et Saleev, V. M. (1979). An Economic Method of Computing LP_τ - sequences,. *USSR Computational Mathematics and Mathematical Physics*, 19(1) :252–256.

Archer, G., Saltelli, A., et Sobol', I. (1997). Sensitivity measures, ANOVA like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58 :99–120.

Archer, R., Zakeri, G., et Vaudrey, T. (2005). Splines as an Optimization Tool in Petroleum Engineering. *Proceedings of the SPE Annual Technical conference and Exhibition, Dallas, Texas, U.S.A., October 9-12.* SPE 95601.

Ates, H. (2005). *Use of Streamline Simulations for Integrated Reservoir Modeling.* PhD thesis, The University of Tulsa, Oklahoma, U.S.A.

Ates, H., Bahar, A., El-Abd, S., Charfeddine, M., Kelhar, M., et Datta-Gupta, A. (2003). Ranking and Upscaling of Geostatistical Reservoir Models Using Streamline Simulation : A Field Case Study. *Proceedings of the SPE 13th Middle East Oil Show and Conference, Bahrain, June 9-12*, SPE 81497.

Audigane, P. (2000). *Caractérisation microsismique des masifs rocheux fracturés. Modélisation thermo-hydraulique. Application au concept géothermique de soultz.* PhD thesis, Institut Polytechnique de Lorraine, CRPG-CNRS, Spécialité Géosciences. (In French).

Aziz, K. et Settari, A. (1979). *Petroleum Reservoir Simulation.* Applied Science Publishers, Essex, England.

Ballin, P., Aziz, K., Journel, A., et Zuccolo, L. (1993). Quantifying the Impact of Geological Uncertainty on Reservoir Performing Forecasts. *Proccedings of the* 21^{st} *SPE Symposium on Reservoir Simulation, New Orleans, Lousiana, U.S.A., February 20 - March 3*, SPE 25238.

Barret, R., Berry, M., Chan, T., Demmel, J., Donato, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C., et der Vorst, H. V. (1994). *Templates for the Solution of Linear Systems : Building Blocks for Iterative Methods.* SIAM, Philadelphia, PA. http ://www.netbib.org/templates/Templates.html.

Batycky, R. P. (1997). *A Three-Dimensional Two-Phase Field Scale Streamline Simulator*. PhD thesis, Stanford University, Palo Alto, California.

Batycky, R. P., Blunt, M. J., et Thiele, M. R. (1996). A 3D Field Scale Streamline Based Reservoir Simulator. *Proceedings of the SPE Annual Technical Conference and Exhibition, Denver, Colorado, 6-9 October*. SPE 36726.

Batycky, R. P., Thiele, M. R., et Blunt, M. J. (1997). A Streamline-Based Reservoir Simulation of the House Moutain Waterflood. *Proceedings of the SCRF Meeting*, page 8.

Bear, J. (1972). *Dynamics of Fluids in Porous Media*. American Elsevier, New York.

Berenblyum, R., Shapiro, A., et Stenby, E. (2004). Reservoir streamline simulation accounting for effects of capillirity and wettability. *9th European Conference on the Mathematics of Oil Recovery, Cannes, France, August 30 - September 1*.

Blunt, M., Liu, K., et Thiele, M. (1996). A Generalized Streamline Method to Predict Reservoir Flow. *Petroleum Geoscience*, 2 :256–269.

Box, G. et Draper, N. (1987). *Empirical Model-Building and Response Surfaces*. J. Wiley & Sons, New York, New York, U.S.A.

Brandsaeter, I., Wist, H. T., Naess, A., Lia, O., Arntzen, O. J., Ringrose, P. S., Martinius, A. W., et Lerdahl, T. R. (2001). Ranking of Stochastic Realisation of Complex Tidal Reservoirs Using Streamline Simulation Criteria. *Petroleum Geoscience*, 7 :53–63.

Bursztyn, D. et Steinberg, D. (2004). Comparison of Designs for Computer Experiments. *Journal of Statistical Planning and Inference*. Article In Press.

Bush, M. D. et Carter, J. N. (1996). *Applications of a modified genetic algorithm to parameter estimation in the petroleum industry : Intelligent engineering systems through artificial neural networks*, volume 6. ASME Press, New York. p. 397.

Carter, J. (2004). Using bayesian statistics to capture the effects of modelling errors in inverse problems. *Mathematical Geology*, 6(2) :187–216.

Carter, J., Ballester, P., Tavassoli, Z., et King, P. (2006). Our calibrated model has poor predictive value : An example from the petroleum industry. *Reliability Engineering and System Safety*, 91 :1373–1381.

Castellini, A., Landa, J. L., et Kikani, J. (2004). Practical Methods For Uncertainty Assesment of Flow Predictions For Reservoir with Significant History. *Proceedings of the* 9^{th} *European Conference on the Mathematics of Oil Recovery, Cannes, France, August 30 - September 2*, A033.

Caumon, G., Strebelle, S., Caers, J. K., et Journel, A. G. (2004). Assesment of global uncertainty for early appraisal of hydrocarbon fields. *Proceedings of the SPE Annual Technical Conference and Exhibition, Houston, Texas, September 26-29*, SPE 89943.

Charles, T., Guemene, J. M., Vincent, B. C. G., et Dubrule, O. (2001). Experience with the quantification of subsurface uncertainties. *Proceedings of the SPE Asia Pacific Oil and Gas Conference and Exhibition, Jakarta, Indonesia, April 17-19*, SPE 68703.

Chiang, W. H. et Kinzelbach, W. (2001). *3D Groundwater Modeling with PMWIN*. Springer, Berlin, Germany.

Choudhuri, N. (1998). Bayesian bootstrap credible sets for multidimensional mean functional. *Ann. Statist.*, 26(6) :2104–2127.

Christie, M. et Blunt, M. (2001). Tenth SPE Comparative Solution Project : A Comparison of Upscaling Techniques. *SPE Reservoir Evaluation & Engineering*, 4(4) :308–317.

Chu, C. (1990). Prediction of Steamflood Performance in Heavy Oil Reservoirs Using Correlations Developed by Factorial Design Method. *Proceedings of the 6Oth SPE California Regional Meeting, Ventura, California, U.S.A., April 4-6*, SPE 20020.

Cognot, R. (1996). *La méthode D.S.I. : optimisation, implémentation et applications*. PhD thesis, Institut National Polytechnique de Lorraine, Nancy, France.

Cooke, R. M. et Noortwijk, J. M. (1999). Local Probabilistic Sensitivity Measures for Comparing FORM and Monte Carlo Calculations Illustrated with Dike Ring Reliability Calculations. *Computer Physics Communications*, 117 :86–98.

Corre, B., Thore, P., de Feraudy, V., et Vincent, G. (2000). Integrated Uncertainty Assessment For Project Evaluation and Risk Analysis. *Proceedings of the SPE European Petroleum Conference, Paris, France, October 24-25*, SPE 65205.

Cukier, R. I., Fortuin, C. M., Schuler, K. E., Petschek, A. G., et Schaibly, J. H. (1973). Study of the Sensitivity of Coupled Reaction Systems to Uncertainties in Rate Coefficients. *Journal of Chemical Physics*, 59 :3873–3878.

Dake, L. P. (2001). *Fundamentals of Reservoir Engineering*. Developements in petroleum science. Elsevier.

Damsleth, E., Hage, A., et Volden, R. (1992). Maximum Information at Minimum Cost : A North Sea Field Development Study with an Experimental Design. *Journal of Petroleum Technology*, pages 1350–1356. SPE 23139.

Datta-Gupta, A. et King, M. J. (1995). A Semianalytic Approach to Tracer Flow Modeling in Heterogeneous Permeable Media. *Advances in Water Ressources*, 18 :9–24.

Dejean, J.-P. et Blanc, G. (1999). Managing Uncertainties on Production Predictions Using Integrated Statistical Methods. *Proceedings of the 1999 SPE Annual Technical Conference and Exhibition, Houston, Texas, U.S.A., October 3-6*. SPE 56696.

Deutsch, C. (1996). Constrained Smoothing of Histograms and Scatterplots With Simulated Annealing. *Technometrics*, 38(3) :266–274.

Deutsch, C. V. et Journel, A. G. (1992). *GSLIB Geostatistical Software Library and User's Guide*. Oxford University Press, New York, NY, U.S.A. 334 p.

Dubrule, O. (1998). *Geostatistics in Petroleum Geology*. The American Association of Petroleum Geologist, Tulsa, Oklahoma.

Ertekin, T., Abou-Kassem, J. H., et King, G. R. (2001). *Basic Applied Reservoir SSimulation*, volume 7. SPE Textbook.

Essenfeld, M. (1969). The Use of Cubic Splines : A Method of Interpolation with Immediate Applications in Petroleum Engineering. SPE 2493.

Eubank, R. L. (1988). *Spline Smoothing and Nonparametric Regression.* Marcel Dekker, New York, N.Y.

Fang, K.-T., Lin, D., Winker, P., et Zhang, Y. (2000). Uniform Design : Theory and Application. *Technometrics*, 42(3) :237–248.

Feraille, M. et Roggero, F. (2004). Uncertainty Quantification for Mature Field Combining the Bayesian Inversion Formalism and Experimental Design Approach. *Proceedings of the 9^{th} European Conference on the Mathematics of Oil Recovery, Cannes, France, August 30 - September 2.* A030.

Fetel, E. (2005). Constraining response surface with secondary information for reservoir flow performance assessment. *Proceedings of the 25^{th} Gocad Meeting held in Nancy, France, June 6-9.*

Friedmann, F., Chawathé, A., et Larue, D. K. (2001). Assessing Uncertainty in Channelized Reservoirs Using Experimental Designs. *Proceedings of the 2001 SPE Annual Technical Conference and Exhibition, New Orleans, Louisianna, September 30 - October 3*, SPE 71622.

Galli, A. et Murillo, E. (1984). Dual Kriging - Its Propeties and it Uses in Direct Contouring. In Verly, G., David, M., Journel, A., et Marechal, A., editors, *Geostatistics for Natural Ressources Characterization - Part 2*, NATO-ASI - Advanced Geostatistics, pages 621–634.

Gilman, J. R., Meng, H.-Z., Uland, M. J., Dzurman, P. J., et Cosic, S. (2002). Statistical Ranking of Stochastic Geomodels Using Streamline Simulation : A Field Application. *Proceedings of the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, September 29 - October 2*, SPE 77374.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evalutation.* Applied Geostatistics Series. Oxford University Press, New York, New York, U.S.A. 483 p.

Guyaguler, B. (2002). *Optimization of Well Placement and Assessment of Uncertainty.* PhD thesis, Stanford University, Stanford, California, U.S.A.

Guyaguler, B. et Horne, R. N. (2001). Uncertainty Assesment of Well Placement Optimization. *Proceedings of the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, September 30 - October 3*, SPE 71625.

Helton, J., Davis, F., et Johnson, J. (2005). A Comparison of Uncertainty and Sensitivity Analysis Results Obtained with Random and Latin Hypercube Sampling. *Reliability Engineering and System Safety*, 89 :305–330.

Hird, K. B. et Dubrule, O. (1995). Quantification of Reservoir Connectivity for Reservoir Description Applications. *Proceedings of the SPE Annual Technical Conference and Exhibition, Dallas, Texas, U.S.A., October 22-25*, SPE 30571.

Hoffman, B. et Caers, J. (2003). Geostatistical history matching using the regional probability perturbation method. *SPE Annual Conference and Technical Exhibition, Denver, Oct. 5-8.*

SPE 84409, page 16pp.

Homma, T. et Saltelli, A. (1996). Importance Measures in Global Sensitivity Analysis of Nonlinear Models. *Reliability Engineering and System Safety*, 52 :1–17.

Hu, L. (2000). Gradual deformation and iterative calibration of gaussian-related stochastic models. *Mathematical Geology*, 32(1) :87–108.

Huijbregts, C. et Matheron, G. (1971). Universal kriging - An optimal approach to trend surface analysis. *Decision Making in the Mineral Industry*, 12 :159–169.

Idrobo, E. A., Choudhary, M. K., et Datta-Gupta, A. (2000). Swept Volume Calculations and Ranking of Geostatistical Reservoir Models Using Streamline Simulation. *Proccedings of the SPE/AAPG Western Regional Meeting, Long Beach, California, U.S.A., June 19-23*, SPE 62557.

Iman, R. et Conover, W. (1980). Small sample sensitivity analysis techniques for computer models with an application to risk assessment. *communications in Statistics Theory and Methods A9*, pages 1749–1874.

Iman, R. et Hora, S. (1990). A robust measure of uncertainty importance for use in fault tree system ananlysis. *Risk Analysis*, 8 :71–90.

Isaaks, E. H. et Srivastava, R. M. (1989). *A Introduction to Applied Geostatistics*. Oxford University Press, New York, NY, U.S.A.

Ishigami, T. et Homma, T. (1990). An Importance Quantification Technique in Uncertainty Analysis for Computer Models. *In Proceedings of the ISUMA90, First International Symposium on Uncertainty Modelling and Analysis, University of Maryland, USA, 3-5 December*, pages 398–403.

Johnson, N. et Kotz, S. (1970). *Continuous Univariate Distributions*. John Wiley, New York, N.Y.

Jones, M. (1991). On correcting for variance inflation in kernel density estimation. *Computational Statistics and Data Analysis*, 11 :3–15.

Kedzierski, P., Solleuz, A. L., Mallet, J., et Royer, J. (2005). Sedimentological and stratigraphic modeling combining membership functions and sequence stratigraphy principles. *Proceedings of the 25th Gocad Meeting held in Nancy, France, June 6-9*.

King, P. R., Buldyrev, S. V., Dokholyan, N. V., Havlin, S., Lee, Y., Paul, G., Stanley, H., et Vandesteeg, N. (2001). Predicting Oil Recovery Using Percolation Theory. *Petroleum Geosciences*, 7 :105–107.

Labat, L. (2004). *Simulation stochastiques de faciès pas la méthode des membership functions*. PhD thesis, Institut National Polytechnique de Lorraine, Nancy, France.

Landa, J. L. et Guyaguler, B. (2003). A Methodology for History Matching and the Assessment of Uncertainties Associated with Flow Prediction. *Proceedings of the SPE Annual Technical Conference and Exhibition, Denver, Colorado, U.S.A., October 5-8*, SPE 84465.

Lee, T., Kravaris, C., et Seinfeld, J. (1986). History Matching by Spline Approximation and Regularization in Single Phase Areal Reservoirs. *SPE Reservoir Engineering, September 1986*, pages 521–534. SPE 13931.

Liefvendahl, M. et Stocki, R. (2004). A Study on Algorithms For Optimization of Latin Hypercubes. *Journal of Statistical Planning and Inference*.

Ma, Y. et Royer, J. (1988). Local geostatistical filtering - application to remote sensing. In Namyslowska-Wilczynska, B. et Royer, J., editors, *Geomathematics and geostatistics analysis applied to space and time dependent data - Part 1*.

Mace, L. et Fetel, E. (2006). Fracture Density Prediction Using Response Surface in Naturally Fractured Reservoirs. *Proceedings of the 26th Gocad Meeting held in Nancy, France, June 6-9*.

Mallet, J. L. (1992). Discrete Smooth Interpolation in Geometric Modeling. *Computer Aided Design*, 24(4) :177–191.

Mallet, J. L. (1997). Discrete Modeling for Natural Objects. *Mathematical Geology*, 29 :199–219.

Mallet, J. L. (2002). *Geomodeling*. Oxford University Press, New York, NY, U.S.A. 678 p.

Mallet, J. L. et Sthuka, A. (2000). Modeling multivariate density and application. *Proceedings of the 20st Gocad Meeting, Nancy, France*.

Manceau, E., Mezghani, M., Zabalza-Mezghani, I., et Roggero, F. (2001). Combination of Experimental Design and Joint Modeling Methods for Quantifying the Risk Associated With Deterministic and Stochastic Uncertainties - An Integrated Test Study. *Proceedings of the 2001 SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, U.S.A., September 30 - October 3*. SPE 71620.

Maréchal, A. (1984). Kriging seismic data in présence of faults. In Verly, G., David, M., Journel, A., et Maréchal, A., editors, *Geostatistics for Natural Resources Characterization*, pages 271–294. Reidel, Dordrecht.

Massonnat, G. (1997). Sampling space of uncertainty through stochastic modelling of geological facies. *Proceedings of the 1997 SPE Annual Technical Conference and Exhibition, San Antonio, Texas, October 5-8*, SPE 38746.

Massonnat, G. (2000). Can we sample the complete geological uncertainty space in reservoir-modeling uncertainty estimates ? *SPE Journal*, vol. 5(1) :pg. 46–59.

Matheron, G. (1965). *Les Variables Régionalisées et leur Estimation, une Application de la Téorie des Fonctions Aléatoires aux Sciences de la Nature*. Masson et Cie, Paris.

Matheron, G. (1970). La théroie des variables régionalisées et ses applications. *Fasicule 5, Les cahiers du centre de morphologie mathématique, Ecole des Mines de Paris, Fontainebleau*.

McKay, M. (1995). Evaluating prediction uncertainty. *Los Alamos National Laboratories Report NUREG/CR-6311, LA-UR-12915-MS*.

McKay, M., Beckman, R., et Conover, W. (1979). A comparison of three methods for selecting values of input varibles in the analysis of output from a computer code. *Technometrics*, 21 :239–245.

Metropolis, N. et Ulman, S. (1949). The monte-carlo method. *Journal of the American Statistical Association*, 44 :335–341.

Minasny, B. et McBratney, A. (2006). A conditioned latin hypercube method for sampling in the presence of ancillary information. *Computer & Geosciences*, 32 :1378–1388.

Morris, M. et Mitchell, T. (1995). Exploratory Designs For Computational Experiments. *Journal of Statistical Planning and Inference*, 43 :381–402.

Mosegaard, K. et Tarantola, A. (1995). Monte carlo sampling of solutions to inverse problems. *Journal of Geophysical Research*, 100(B7) :12,431–12,447.

Mosegaard, K. et Tarantola, A. (2002). Probabilistic approach to inverse problems. *International Handbook of Earthquake & egineering Seismology, Part A, Academic Press, New York*, pages 237–265.

Muron, P., Tertois, A.-L., Mallet, J.-L., et Hovadik, J. (2005). An Efficient and Extensible Interpolation Framework Based on the Matrix Formulation of the Discrete Smooth Interpolation. *Proceedings of the 25th Gocad Meeting held in Nancy, France, June 6-9*.

Myers, D. (1984). Cokriging - new developments. In Verly, G., David, M., Journel, A., et Maréchal, A., editors, *Geostatistics for natural ressources characterization*, pages 295–305. Reidel, Dordrecht, Holland.

Nelder, J. et Wedderburn, R. (1983). Generalized liner model. *The Journal of the Royal Statistical Society*, 135 :370–384.

NIST/SEMATECH (2006). *e-Handbook of Statistical Methods*. http ://www.itl.nist.gov/div898/handbook/.

Pan, Y. et Horne, R. N. (1998). Improved Methods for Multivariate Optimization of Field Development Scheduling and Well Placement Design. *Proceedings of the 1998 SPE Annual Technical Conference and Exhibition held in New Olrleans, Louisiana, U.S.A., September 27-30*. SPE 49055.

Parzen, E. (1962). On the estimation of a probability density function and the mode. *Ann. Math. Stat.*, 33 :1065–1076.

Peaceman, D. W. (1983). Interpretation of Well-Block Pressures in Numerical Reservoir Simulation with Nonsquare Grid Blocks and Anisotropic Permeability. *SPE Journal*. SPE 10528.

Peddibhotla, S., Cubillos, H., Datta-Gupta, A., et Wu, C. H. (1996). Rapid Simulation of Multiphase Flow Through Fine-Scale Geostatistical Realizations Using a New, 3-D Streamline Model : A Field Example. *Proceedings of the Petroleum Computer Conference, Dallas, Texas, U.S.A., June 2-5*. SPE 36008.

Pollock, D. W. (1988). Semianalytical Computation of Path Line for Finite Difference Models. *Ground Water*, 26(6) :743–750.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., et Flannery, B. P. (2002). *Numerical recipes in C++ : the art of scientific computing*. Cambridge University Press, Cambridge, UK.

Prévost, M., Edwards, M., et Blunt, M. (2001). Streamline Tracking on Curvilinear Structured and Unstructured Grids. *Proceedings of the SPE Reservoir Simulation Symposium, Houston, Texas, February 11-14.* SPE 66347.

Rabitz, H., Alì, F., Shorter, J., et Shim, K. (1999). Efficient input-output model representations. *Computer Physics Communications*, 117 :11–20.

Ravalico, J. K., Maier, H. R., Dandy, G. C., Norton, J. P., et Croke, B. F. W. (2005). A Comparison of Sensitivity Analysis Techniques for Complex Models for Environmental Management. *Proceedings of the International Congress on Modelling and Simulation, Melbourbe, Autralia, October 12-15*, pages 2533–2539.

Royer, J.-J. et Vierra, P. (1984). Dual Formalism of Kriging. In Verly, G., David, M., Journel, A., et Marechal, A., editors, *Geostatistics for Natural Ressources Characterization - Part 2*, NATO-ASI - Advanced Geostatistics, pages 691–702.

Saad, N., Maroongroge, V., et Kalkomey, C. T. (1996). Ranking Geostatistical Models Using Tracer Production Data. *Proccedings of the European 3D Reservoir Modelling Conference, Stavanger, Norway, April 16-17*, SPE 35494.

Saltelli, A. et Bolado, R. (1998). An Alternative Way to Compute Fourier Amplitude Sensitivity Test (FAST). *Computational Statistics & Data Analysis*, 26 :445–460.

Saltelli, A. et Sobol', I. M. (1995). About the Use of Rank Transformation in Sensitivity Analysis of Model Output. *Reliability Engineering and System Safety*, 50 :225–239.

Saltelli, A., Tarantola, S., et Campolongo, F. (2000). Sensitivity Analysis as an Ingredient of Modeling. *Statistical Science*, 15(4) :377–395.

Saltelli, A., Tarantola, S., et Chan, K. P. S. (1999). A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. *Technometrics*, 41(1) :39–56.

Sato, K. (2001). Comparison of the Boundary Element Methods for Streamline Tracking. *Journal of Petroleum Science and Engineering*, 30 :29–42.

Scheidt, C. et Zabalza-Mezghani, I. (2004). Assessing Uncertainty and Optimizing Production Schemes - Experimental Designs for Nonlinear Prodcution Responses Modelling An Application to Early Water Breakthrough. *Proceedings of the 9th European Conference on the Mathematics of Oil Recovery, Cannes, France, August 30 - September 2*, A029.

Scott, D. (1992). *Multivariate Density Estimation : Theory, Practice and Visualization*. John Wiley, New York, NY.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York, NY, U.S.A.

Skalicky, T. (1996). *Laspack reference manual*. Dresden Univ. of Fechnology - Institue for Fluid Mechanics. http ://www.tudresden.de/mwism/skalicky/laspack/laspack.html.

Sobol', I. M. (1967). On the Distribution of Points in a Cube and the Approximate Evaluation of Integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4) :86–112.

Sobol', I. M. (1990). Sensitivity Estimates for Nonlinear Mathematical Models. *Matematicheskoe Modelirovanie*, 2 :112–118. [Translated as Sensitivity Analysis for Nonlinear Mathematical Models. *Math. Modeling comput. Experiment* 1 (1993) 407-414].

Soong, T. T. (1981). *Probabilistic Modeling and Analysis in Science and Engineering.* John Wiley & Sons, New York, NY, U.S.A.

Streamsim (2004). *3DSL, User Manual - Version 2.20.* Streamsim Technologies, Inc. http ://www.streamsim.com.

Sturges, H. (1926). The choice of a class-interval. *J. Amer. Statist. Assoc.*, 21 :65–66.

Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation.* SIAM, Philadelphia, PA. 342 p.

Tchelepi, H. et Orr, F. (1994). Interaction of Viscous Fingering Permeability Heterogeneity, and Gravity Segregation in Three Dimensions. *SPE Reservoir Engineering*, 9 :266–271.

Terrell, G. et Scott, D. (1992). Variable kernel density estimation. *Ann. Stat.*, 20(3) :1236–1265.

Thiele, M., Batycky, R., Blunt, M., et Jr., F. O. (1994). Simulating Flow in Heterogeneous Systems Using Streamtubes and Streamlines. *Proceedings of the SPE/DOE Improved Oil Recovery Symposium, Tulsa, Oklahoma*, SPE 27834.

Voillemont, J. (2003). *Caractérisation par micro-sismicité induite des milieux poreux fracturés. Modélisations par la méthode des lignes de courant d'un site géothermique HDR (Soultz-sous-Forêts, France).* PhD thesis, Institut Polytechnique de Lorraine, CRPG-CNRS, LIAD. (In French).

Voillemont, J. et Royer, J. (2001). Building 3d streamlines in gocad. *Proceedings of the 21st Gocad Meeting, Nancy, France.*

Voillemont, J. et Royer, J. (2003). Streamline-based reservoir characterization : Latest advances. *Proceedings of the 23rd Gocad Meeting, Nancy, France.*

Wang, Y. et Kovseck, A. R. (2003). Integrating Production History into Reservoir Models Using Streamline-Based Time-Of-Flight Ranking. *Petroleum Geoscience*, 9 :163–174.

Yeten, B., Castellini, A., Guyaguler, B., et Chen, W. (2005). A Comparison Study on Experimental Design and Response Surface Methodology. *Proceedings of the 2005 SPE Reservoir Simulation Symposium, Houston, Texas, U.S.A., January 31-February 2*, SPE 93347.

Zabalza, I., Blanc, G., Collombier, D., et Mezghani, M. (2000). Use of experimental design in resolving inverse problems : Application to history matching. *Proceedings of the 7th European Conference on the Mathematics of Oil Recovery, Baveno, September 5-8.*

Zabalza, I., Dejean, J., et Collombier, D. (1998). Prediction and density estimation of a horizontal well productivity index using generalized linear model. *Proceedings of the 6th European Conference on the Mathematics of Oil Recovery, Peebles, Netherland, September 8-11.*

Zabalza-Mezghani, I. (2000). *Analyse statistique et plannification d'expérience en ingéniérie de réservoir.* PhD thesis, Univ. de Pau et des pays de l'Adour.

Zabalza-Mezghani, I., Manceau, E., Feraille, M., et Jourdan, A. (2004). Uncertainty Management : From Geological Scenarios to Production Scheme Optimization. *Journal of Petroleum Science & Engineering*, 44 :11–25.

Zabalza-Mezghani, I., Manceau, E., et Roggero, F. (2001). A New Approach For Quantifying the Impact of Geostatistical Uncertainty on Production Forecasts : The Joint Modelling Method. *Proceedings of the IAMG, Cancun, Mexique, September 6-12*.

Annexe A

Simulation d'écoulement sur lignes de courant dans le géomodeleur Gocad

Sommaire

A.1 Principe d'une simulation d'écoulement biphasique sur lignes de courant . . . 138
A.1.1 Équations fondamentales . . . 138
A.1.2 Méthodologie . . . 139
A.1.3 Traitement de la gravité . . . 140
A.2 Résolution numérique . . . 141
A.2.1 Modèle discret . . . 142
A.2.2 Résolution de l'équation de pression . . . 142
A.2.3 Méthode de tracé des lignes de courant . . . 147
A.2.4 Résolution de l'équation de saturation . . . 151
A.2.5 Transfert des résultats des lignes de courant vers la grille . . . 156
A.2.6 Modélisation de l'effet de la gravité . . . 156
A.3 Apport de StreamLab-2.0.0 . . . 157
A.4 Structure de données pour la représentation et la manipulation de lignes de courant . . . 158
A.4.1 Structures de données pré-existantes . . . 159
A.4.2 Implantation proposée . . . 161
A.5 Validation du calcul . . . 163
A.5.1 Description du modèle . . . 163
A.5.2 Résultats . . . 164
A.6 Conclusions . . . 167

Les premières utilisations de la technique des lignes de courant dans le géomodeleur Gocad remontent aux travaux de Audigane [2000] puis Voillemont [2003] pour la simulation de transfert de chaleur et l'injection ponctuelle de traceurs chimiques au sein d'un massif granitique fracturé.

Ici, nous nous sommes intéressés à la simulation d'écoulement biphasique (mélange huile-eau) en milieu poreux à l'échelle d'un réservoir pétrolier. C'est-à-dire : (1) sur des modèles discrets de plusieurs millions de cellules, (2) dans un milieux poreux présentant de fortes hétérogénéités et (3) avec des conditions de puits dynamiques. Après un rappel du principe d'une simulation d'écoulement sur lignes de courant ainsi que les hypothèses encadrant ce travail, nous détaillerons les méthodes numériques et les structures de données spécifiquement mises en place pour résoudre ce problème. Le code de calcul sera validé sur un exemple de réservoir pétrolier synthétique.

A.1 Principe d'une simulation d'écoulement biphasique sur lignes de courant

A.1.1 Équations fondamentales

Comme la plupart des processus physiques, le comportement d'un fluide peut être décrit par une ou plusieurs équations de conservation. Dans le cas d'écoulements isothermaux en milieu poreux, il s'agit de l'équation de conservation de la masse. Celle-ci spécifie que, pour un volume de contrôle donné, la masse de matière échangée avec l'extérieur doit être égale à celle s'accumulant à l'intérieur. Si l'on considère un système biphasique (huile-eau), des fluides non miscibles et incompressibles ainsi que l'absence de pression capillaire, la conservation de la masse se traduit par les trois équations aux dérivées partielles suivantes [Aziz et Settari, 1979; Ertekin et al., 2001; Dake, 2001] :

1. *L'équation de pression* :

$$\nabla \cdot \mathbf{u}_t = -\nabla \cdot \mathbf{K} \cdot (\lambda_t \ \nabla P + \lambda_g \ \nabla D) = q_s \tag{A.1}$$

où $\mathbf{u}_t$ (en $m.s^{-1}$) est la vitesse totale de Darcy (c'est-à-dire, la vitesse moyenne du mélange huile-eau dans le milieu poreux), $\mathbf{K}$ (en m^2) le tenseur de perméabilité intrinsèque de la roche, P (en Pa), la pression totale des fluides, D (en m) l'altitude ou la profondeur selon l'orientation du modèle de réservoir, q_s (en $m^3.s^{-1}$) une éventuelle injection ou production de fluide, et enfin, λ_t (en $Pa^{-1}.s^{-1}$) et λ_g (en $m^{-1}.s^{-1}$) sont, respectivement, la mobilité totale et la mobilité gravitaire calculées à partir des propriétés physiques des fluides, telles que :

$$\lambda_t = \frac{k_{rw}}{\mu_w} + \frac{k_{ro}}{\mu_o} \quad \text{et} \quad \lambda_g = \frac{k_{rw}\rho_w g}{\mu_w} + \frac{k_{ro}\rho_o g}{\mu_o} \tag{A.2}$$

où, g (en $m.s^{-2}$) est l'accélération due à la gravité, et pour une phase donnée, $i \in [o, w]$ [1], k_{ri} représente la perméabilité relative [2], μ_i (en $Pa.s$) la viscosité dynamique et ρ_i ($kg.m^{-3}$) la masse volumique.

2. *L'équation de saturation en eau* (aussi appelée équation de Buckley-Leverett) :

$$\phi \frac{\partial S_w}{\partial t} + \mathbf{u}_t \cdot \nabla f_w + \nabla \cdot \mathbf{g}_w = f_{w,s} \ q_s \tag{A.3}$$

[1] les indices o et w correspondent respectivement à l'huile (oil) et à l'eau (water)

[2] la perméabilité relative représente l'évolution de la perméabilité du milieu poreux à un fluide donné en fonction de sa saturation

où S_w est la saturation en eau, c'est-à-dire la proportion du volume poreux remplie d'eau, ϕ la porosité effective du milieu, f_w est la fraction du flux liée à l'eau, égale à :

$$f_w = \frac{kr_w/\mu_w}{kr_w/\mu_w + kr_o/\mu_o} \tag{A.4}$$

et $\mathbf{g}_w$ est le flux d'eau dû à la gravité :

$$\mathbf{g}_w = \mathbf{K} \cdot \left(g f_w \frac{kr_o\ (\rho_o - \rho_w)}{\mu_o} \nabla D \right) \tag{A.5}$$

3. Et, enfin, *la définition de la saturation* :

$$S_o + S_w = 1 \tag{A.6}$$

Les équations (A.1) à (A.6) forment un système où les inconnues sont la pression, P, et les saturations en fluide, S_w et S_o. Les autres variables sont considérées connues et doivent être définies dans la description du problème. Bien que ce système soit composé d'équations fortement couplées, il est important de noter qu'elles ont des comportements mathématiques et numériques différents ce qui permet de les résoudre séparément. Cette considération est la base de la formulation IMPES (pour IMplicite en Pression Explicite en Saturation) et donc de la méthode de résolution sur lignes de courant. Le champ de pression est d'abord estimé en premier implicitement afin de pouvoir calculer les lignes de courant, puis dans un second temps, l'évolution de la distribution de saturation est estimée explicitement le long des lignes de courant. Pour plus de détails sur la formulation de ces équations, le lecteur est invité à se reporter à l'abondante littérature, par exemple Aziz et Settari [1979] pour des simulations conventionnelles ou Thiele et al. [1994] et Batycky et al. [1997] pour des simulations sur lignes de courant.

A.1.2 Méthodologie

Le principe de la simulation d'écoulement sur lignes de courant est de découpler un problème complexe tridimensionnel sous forme d'un ensemble de sous-problèmes unidimensionnels. Pratiquement, cela consiste à simuler le déplacement de particules fluides le long de lignes de courant définies comme étant, en tout point du domaine d'étude, tangentes au champ de vitesses. Dans le détail, la chaîne d'opérations nécessaires pour réaliser une simulation d'écoulement sur lignes de courant peut être résumée de la façon suivante :

1. Résolution de l'équation de pression et calcul du champ de vitesse dans la grille tridimensionnelle.

2. Calcul de l'intervalle de temps, Δt_p, durant lequel le champ de pression est considéré valide. En effet, le champ de pression peut être estimé non valide (1) si la distribution de saturation dans le réservoir évolue de façon significative de telle façon que cela affecte la géométrie des lignes de courant, (2) si les conditions aux limites changent ou alors (3) à cause de limitations dues aux méthodes numériques employées pour calculer le champ de pression.

3. Tracé des lignes de courant à partir du champ de vitesse.

4. Transfert de l'état du front de saturation et des vitesses de la grille vers les lignes de courant.

5. Calcul du temps de vol et résolution de l'équation de saturation le long des lignes de courant pendant toute la durée de validité du champ de pression. Lors de cette étape, plusieurs pas de temps numériques spécifiques (noté Δt_s) sont nécessaires, avec généralement $\Delta t_p >> \Delta t_s$.

6. Transfert de l'état du nouveau front de saturation des lignes de courant vers la grille.

7. Post processus, le cas échéant, pour simuler des phénomènes purement diffusifs ou n'agissant pas le long des lignes de courant, comme par exemple la gravité [Blunt et al., 1996] ou la pression capillaire [Berenblyum et al., 2004].

8. Enfin retour à l'étape 1 si les conditions d'arrêt de la simulation ne sont pas atteintes. Ces conditions peuvent être soit une durée globale de simulation soit le non respect de contraintes numériques (par exemple, non respect de la loi de conservation de la masse, non convergence du solveur numérique lors de la résolution de l'équation de pression, fermeture de l'ensemble des puits, etc.).

Ce type de simulation a les avantages suivants : (1) les particules fluides suivent les lignes de courant. Or celles-ci forment un chemin plus naturel qu'entre les différentes cellules d'un modèle discret tel que ceux considérés dans les méthodes conventionnelles ; (2) le long des lignes de courant, le problème est unidimensionnel et peut être résolu à l'aide de schémas numériques connus pour leur robustesse, ou, dans certaines situations, à l'aide de solutions analytiques ; et, (3) le temps de validité du champ de pression et donc de la géométrie des lignes de courant est beaucoup plus grand que celui de la distribution de saturation. Par conséquent, il est possible de résoudre plusieurs fois l'équation de la saturation en gardant le champ de pression fixé, contrairement aux méthodes conventionnelles qui nécessitent le calcul de la pression et de la saturation à chaque pas de temps, qui par conséquent est le plus petit (Δt_s).

Cependant elle reste limitée sur quelques points : il est nécessaire que les phénomènes à modéliser s'effectuent principalement le long des lignes de courant et, pour des raisons d'efficacité de calcul, que le champ de pression puisse être considéré constant pendant un grand pas de temps. Par conséquent, cette technique de simulation est difficilement applicable aux problèmes incluant des phénomènes physiques complexes et des sources d'instabilité comme la diffusion (par ex., lors de transfert de chaleur) ou la compressibilité des fluides (par ex., en présence de gaz).

A.1.3 Traitement de la gravité

Dans le cadre d'un écoulement tridimensionnel biphasique, la gravité et la différence de densité entre les phases jouent un rôle important et peuvent avoir un impact négatif sur les taux de récupération en huile [Tchelepi et Orr, 1994]. Ainsi, leurs effets doivent être intégrés dans tout simulateur d'écoulement, en particulier, tridimensionnel. Or dans ce cas, le vecteur

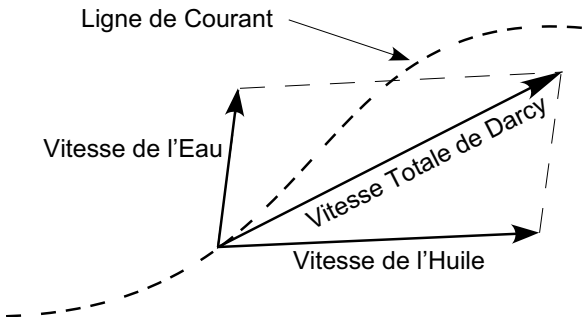


FIG. A.1 – Sous l'action de la gravité, les vitesses de chacune des phases ne sont pas colinéaires à la vitesse totale de Darcy. Par conséquent, la trajectoire réelle des particules de fluide n'est plus confondue avec celle des lignes de courant.

correspondant à la vitesse totale de Darcy (qui définit la trajectoire des lignes de courant) et les vecteurs vitesses de chacune des phases ne sont pas colinéaires (figure (A.1)). Par conséquent, la trajectoire réelle des particules de fluide n'est plus confondue avec celle des lignes de courant.

La principale solution pour résoudre ce problème est de découpler les écoulements convectifs le long des lignes de courant de l'effet de ségrégation des fluides dû à la gravité. Cette approche, initialement décrite dans Blunt et al. [1996] puis Batycky [1997], consiste en deux étapes :

1. L'équation de saturation (A.3) est résolue sans tenir compte de la gravité, c'est-à-dire sous la forme :
$$\phi \frac{\partial S_w}{\partial t} + |\mathbf{u}_t| \frac{\partial f_w}{\partial s} = 0 \tag{A.7}$$
où, pour une particule donné le long d'une ligne de courant, s représente son abscisse curvilinéaire et $|\mathbf{u}_t|$ sa vitesse scalaire.
2. L'effet de gravité est ensuite résolu par post-processus le long de pseudo "lignes" de gravité alignées avec l'accélération de la pesanteur $\mathbf{g}$. A cette étape, les fluides sont simplement déplacés verticalement suivant leurs densités :
$$\phi \frac{\partial S_w}{\partial t} + \frac{1}{\phi} \frac{\partial G(S_w)}{\partial z} = 0 \tag{A.8}$$

A.2 Résolution numérique

Bien que les méthodes analytiques de résolution des Équations aux Dérivées Partielles (EDP) soient nombreuses et performantes, elles reposent sur des hypothèses simplificatrices qui rendent leur utilisation dans le cadre d'un réservoir pétrolier illusoire. Il est donc nécessaire de chercher une solution approchée à ce problème en se basant sur une représentation discrète du sous-sol et des méthodes numériques de résolution des EDP. Dans cette section, nous détaillerons les différentes étapes d'une simulation sur lignes de courant en précisant les méthodes numériques choisies.

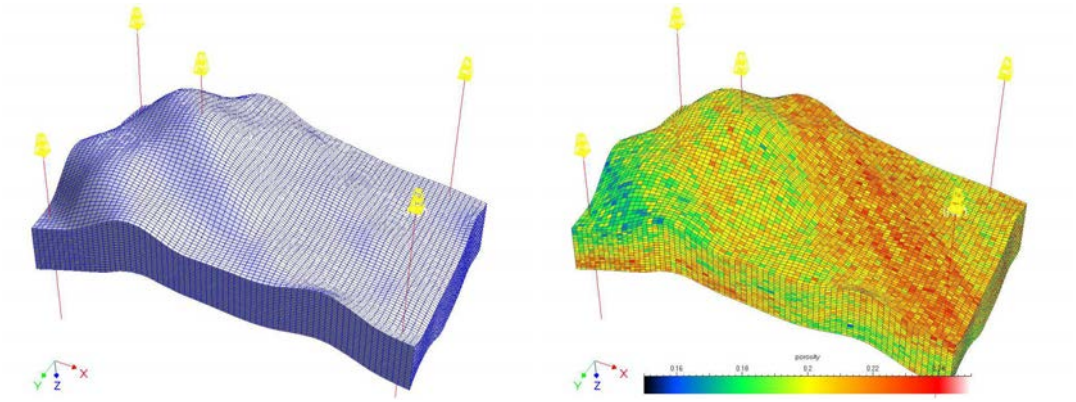


FIG. A.2 – Grille stratigraphique représentant un anticlinal : À gauche, géométrie globale du réservoir et position des puits, à droite, champ de porosité calculé par simulation gaussienne séquentielle.

A.2.1 Modèle discret

Le modèle discret choisi pour représenter la géométrie et les hétérogénéités pétrophysiques du sous-sol est une grille hexaédrique de type stratigraphique supposée non faillée. Ce type de grille est *structurée* : chaque cellule possède toujours le même nombre de voisins, *irrégulière* : les cellules peuvent être de taille et de forme différentes. Enfin le stockage des propriétés est de type *centré* : pour une propriété donnée, chaque cellule ne contient qu'une valeur considérée homogène dans tout le volume. Dans le cadre des simulations d'écoulement, trois propriétés sont nécessaires : (1) la porosité du milieu, ϕ, (2) le tenseur de la perméabilité intrinsèque de la roche, $\mathbf{K}$, qui est supposé diagonal, c'est-à-dire que ses axes principaux sont alignés avec les axes de chaque cellule de la grille et (3) la distribution initiale de la saturation. Cette dernière provient, d'une manière générale, soit d'une simulation d'écoulement précédente soit d'un calcul de distribution supposant que les fluides sont à l'état d'équilibre dans le réservoir.

Le choix de ce type de grille est motivé par plusieurs raisons :
- leur flexibilité leur permet de s'adapter à la plupart des géométries de réservoir ;
- leur compacité en terme de mémoire autorise la représentation et la manipulation de modèles comprenant plusieurs millions de cellules ;
- de nombreux algorithmes robuste de construction ont été décrits et implantés, en particulier, dans le géomodeleur Gocad ;
- et, pour le moment, c'est le plus répandu dans l'industrie pétrolière.

A.2.2 Résolution de l'équation de pression

Discrétisation de l'équation de pression

La résolution de l'équation aux dérivés partielles correspondant à la pression est réalisée à l'aide de la méthode des volumes finis. En effet, contrairement aux autres méthodes que sont les

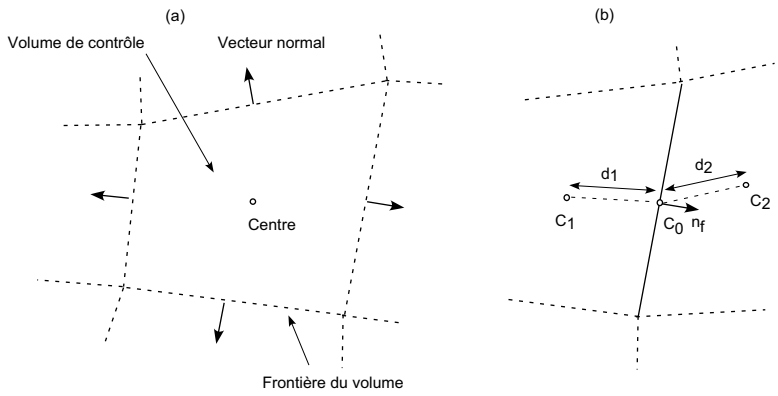


FIG. A.3 – (a) Représentation géométrique (2D) d'un volume de contrôle et (b) définition des principaux termes intervenant dans le calcul de la transmissibilité entre deux volumes adjacents.

différences finies et les éléments finis, cette approche repose sur le strict respect de l'équation de conservation de la masse sur un volume de contrôle. Or ce respect est l'un des éléments clefs lors d'une simulation d'écoulement. De plus, dans notre cas, les volumes de contrôle sont les cellules mêmes de la grille, il n'est donc pas nécessaire de les calculer.

Principe Considérons un volume de contrôle donné Ω de frontière $\partial\Omega$. Tout d'abord, il est nécessaire de réécrire l'équation de pression sur la frontière du volume en appliquant le théorème de Gauss-Ostrogradsky (encore appelé théorème de la divergence). L'équation (A.1) devient alors :

$$\int_{\partial\Omega} \mathbf{u}_t(\zeta).\mathbf{n}(\zeta)\ d\zeta = -q_s \tag{A.9}$$

où, $\mathbf{u}_t(\zeta)$ représente la vitesse de Darcy évaluée sur $\partial\Omega$ tandis que $\mathbf{n}$ est la normale sortante à la frontière, $\partial\Omega$.

Discrétisation Considérons maintenant un volume de contrôle discret, sa frontière est divisée en n_f facettes dont l'aire et la normale sortante sont notées, respectivement, a_f et $\mathbf{n}_f$ (figure (A.3)). L'équation (A.9) peut alors être approchée par la forme discrète suivante :

$$\sum_{i=1}^{n_f} (a_f\ \mathbf{u}_{t,f}.\mathbf{n}_f) = -q_s \tag{A.10}$$

L'estimation du terme $\mathbf{u}_{t,f}$ représentant la vitesse de Darcy entre deux cellules s'effectue alors à l'aide d'un schéma de type différences finies sur deux points. Ainsi, soit deux cellules adjacentes, notées C_1 et C_2, il vient :

$$\vec{u}_{t,C_1\to C_2} = T_{C_1\to C_2}\ \lambda_{t,C_1\to C_2}\ (P_2 - P_1) + \lambda_{g,C_1\to C_2}\ (D_2 - D_1) \tag{A.11}$$

où, $T_{C_1\to C_2}$ est la transmissibilité de C_1 vers C_2. Ce terme dépend des perméabilités de chaque cellule (k_1 et k_2) dans la direction considérée et des distances (d_1 et d_2) entre le centre de chaque

cellule et le centre de l'interface. La transmissibilité $T_{C_1 \to C_2}$ est alors calculée telle que :

$$T_{C_1 \to C_2} = \frac{k_1/d_1 \times k_2/d_2}{k_1/d_1 + k_2/d_2} \tag{A.12}$$

Pour finir, $\lambda_{t,C_1 \to C_2}$ et $\lambda_{g,C_1 \to C_2}$ représentent les mobilités, respectivement, totales et gravitaires. Dans le cas général, pour des raisons de stabilité numérique, ces termes sont évalués dans la cellule en amont de l'écoulement (ou cellule "*up-wind*"). Cela nécessite de connaître le sens de l'écoulement antérieurement au calcul or dans notre cas ce n'est pas possible. Aussi un schéma centré a été préféré :

$$\lambda_{t,C_1 \to C_2} = \frac{d_1 \lambda_{t,C_1} + d_2 \lambda_{t,C_2}}{d_1 + d_2} \quad \text{et} \quad \lambda_{g,C_1 \to C_2} = \frac{d_1 \lambda_{g,C_1} + d_2 \lambda_{g,C_2}}{d_1 + d_2} \tag{A.13}$$

Conditions aux limites

La résolution d'un problème comprenant des équations aux dérivées partielles tel que le calcul du champ de pression, nécessite de connaître l'équation à résoudre elle-même, mais aussi les conditions appliquées aux limites du domaine d'étude. Dans notre cas, celles-ci sont définies aux puits et sur le bord extérieur du modèle discret.

Modèle de puits L'approche envisagée pour représenter les relations entre un puits et le milieu poreux est le modèle de Peaceman [Peaceman, 1983]. Celui-ci est relativement simple comparé à ceux proposés par les simulateurs de type conventionnel et n'est applicable qu'aux puits verticaux. Toutefois, il est à la fois facile à mettre en place et à contrôler. Considérons un puits vertical traversant n_l cellules, le débit total du puits, q_s est égal à :

$$q_s = \sum_{k=1}^{n_l} T_k^w \ (P_k^w - P_k) \tag{A.14}$$

où, P_k^w est la pression fluide à l'intérieur du puits, P_k, la pression dans la cellule traversée et T_k^w la transmissibilité entre le puits et le milieu poreux définie telle que :

$$T_k^w = \frac{2\pi h}{\ln \frac{r_{o,k}}{r_{w,k}} + s_k} \lambda_{t,k} \tag{A.15}$$

où, s_k est le facteur de "skin", qui représente l'état du réservoir à proximité immédiate du puits, celui-ci pouvant être colmaté ou fracturé lors du forage, h, la hauteur de la cellule considérée, $r_{w,k}$ le rayon du puits et, enfin, $r_{o,k}$ le rayon de Peaceman [Peaceman, 1983] évalué en fonction de la forme de la cellule traversée.

Dans le cas où un puits traverse plusieurs cellules, la pression P_k^w dans une cellule donnée doit être reliée à la pression de la première cellule de la zone de complétion, $P_{k^*}^w$, en appliquant un gradient dû à la gravité :

$$P_k^w = P_{k^*}^w + \frac{1}{2} \sum_{i=k^*+1}^{k} \left(\frac{\lambda_{t,i-1}}{\lambda_{g,i-1}} + \frac{\lambda_{t,i}}{\lambda_{g,i}}\right)(D_i - D_{i-1}) \tag{A.16}$$

Bord extérieur L'écoulement sur le bord extérieur du réservoir est supposé nul. Cette condition est relativement simpliste car elle suppose que les bords du réservoir sont étanches ce qui est rarement le cas. Il eut été préférable de considérer la présence d'aquifères à l'aide de modèles analytiques.

Résolution numérique du système linéaire

La méthode de discrétisation transforme, en chaque nœud du maillage, l'EDP en une équation algébrique (équations (A.9) à (A.13)) faisant intervenir la valeur, inconnue, de la pression au nœud et celle de certains de ses voisins. Exprimé sur l'ensemble du maillage, ce processus conduit à un système linéaire qu'il faut résoudre en fonction de l'état initial et des conditions aux limites (équations (A.14) à (A.16)). Ce système s'écrit sous la forme matricielle suivante :

$$\mathbf{T}\ \mathbf{p} = \mathbf{b} \tag{A.17}$$

où

1. **T** est une matrice contenant les transmissibilités, que ce soit entre cellules voisines ou entre les cellules traversées par un puits et celui-ci. Dans le cadre de la formulation IMPES, cette matrice est carré, symétrique, définie positive et a pour taille le nombre de cellules de la grille plus le nombre de puits ouverts et contrôlés en débit ;

2. **p** est un vecteur contenant les pressions inconnues, c'est-à-dire celles dans toutes les cellules de la grille ainsi que celles des puits contrôlés en débit ;

3. enfin, **b** est un vecteur contenant les termes dus à la gravité et les conditions imposées aux puits (qu'ils soient contrôlés en débit ou en pression) et aux frontières du réservoir.

L'équation (A.17) représente donc un système matriciel de très grande taille. Le choix, le détail et l'implantation de la méthode de résolution de ce système est au delà de la portée de ce travail. Dans ce manuscrit, les résultats présentés (par exemple, figure (A.9)) ont été obtenus à l'aide d'un gradient conjugué avec un préconditionnement de type SSOR. Tous deux sont inclus dans la librairie dédiée au calcul matriciel LASPACK-1.12.2 [Skalicky, 1996]. Pour plus de détails sur ces méthodes, le lecteur est invité à consulter l'abondante littérature sur le sujet, en particulier Barret et al. [1994] et Press et al. [2002].

Calcul du champ de vitesse

Le calcul de la vitesse totale de Darcy pour chacune des faces de cellule est ensuite réalisé explicitement à partir des valeurs de pression et à l'aide des équations (A.11) à (A.13). Cette vitesse est alors divisée par la porosité pour obtenir la vitesse intersticielle des particules fluides nécessaire au calcul des lignes de courant.

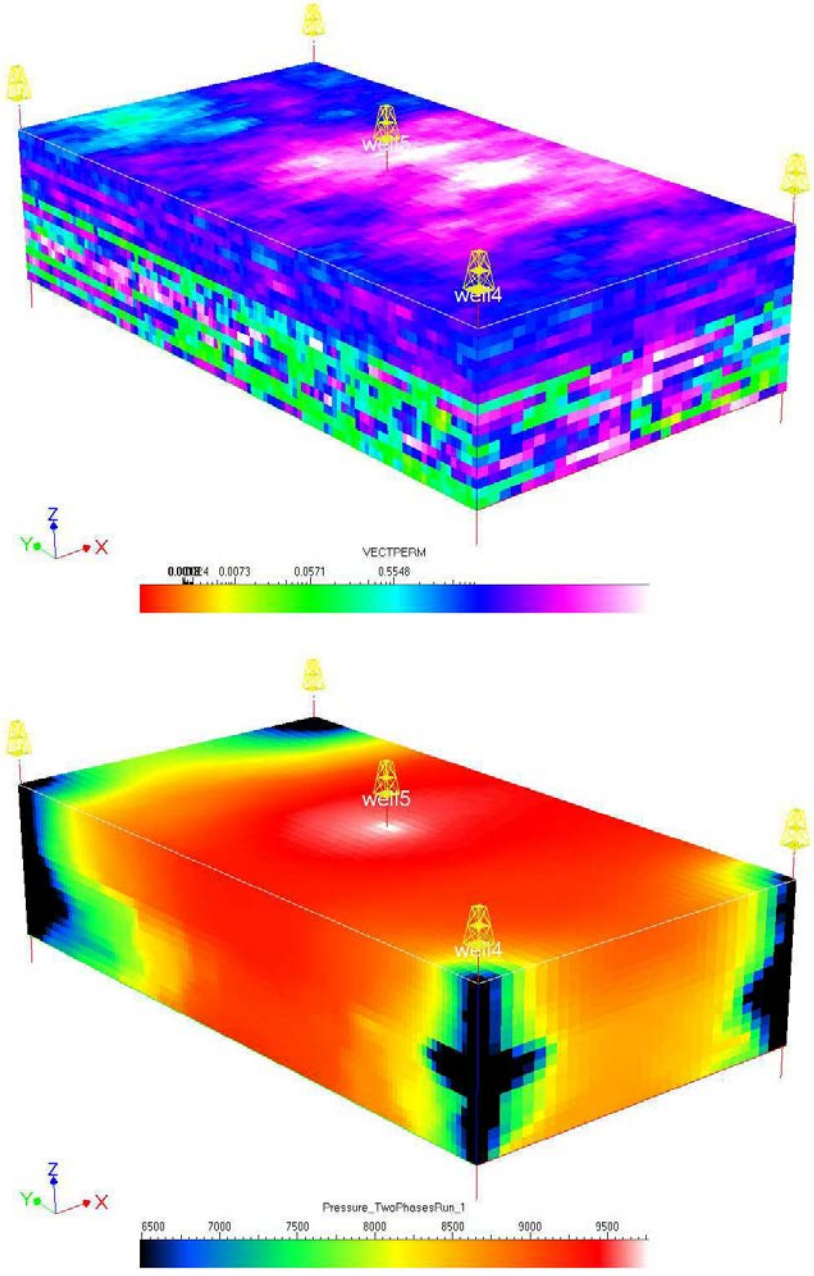


FIG. A.4 – Modèle synthétique de réservoir. En haut : champ de perméabilité (les valeurs varient de 0.005 mD à 500 mD) et configuration des puits : un injecteur au centre et quatre producteurs aux extrémités du modèle. En bas : champ de pression calculé avec la méthode des volumes finis. Les conditions aux limites sont : aucun flux aux frontières du modèle et des puits contrôlés en pression à 4000 psi pour l'injecteur et 10000 psi pour les producteurs.

Validité du champ de pression

L'une des idées fondamentales de la technique de simulation sur lignes de courant est de supposer que le champ de pression, et donc la géométrie des lignes de courant, reste constant et valide sur de grands pas de temps. Les conditions qui requièrent une nouvelle estimation de la pression sont : (1) un changement du type de contrôle des puits ou (2) une modification du profil de saturation affectant de façon non-négligeable la pression. L'estimation de cette dernière condition se base sur un bilan global de la production du réservoir. Le pas de temps, Δt_p, de validité du champ de pression est calculé tel que :

$$\Delta t_p = \frac{PV\ DPV_{max}}{max(Q_{inj}, Q_{prod})} \tag{A.18}$$

où, PV représente le volume poreux total du réservoir, DPV_{max} est un facteur de contrôle compris entre 0 et 1 représentant le ratio entre la quantité maximale de fluides injectée ou produite et le volume poreux durant le pas de temps et Q_{inj} et Q_{prod} sont les débits en fluide, respectivement, injectés et produits sur l'ensemble du champ.

A.2.3 Méthode de tracé des lignes de courant

Une ligne de courant est définie comme une courbe qui en tout point du domaine d'étude est tangente au vecteur vitesse [Bear, 1972]. Il existe ainsi une infinité de lignes de courant et idéalement il serait possible de toutes les calculer de façon analytique à l'aide des fonctions paramétriques de courant et de potentiel (équations de Cauchy-Reimann) [Thiele et al., 1994; Sato, 2001]. Cependant, comme souligné dans Batycky [1997], dans un réservoir pétrolier, en trois dimensions et compte-tenu des hétérogénéités du milieu poreux, cette approche est irréaliste. La principale difficulté venant du fait que les lignes de courants sont des objets géométriques complexes difficiles à suivre. Par conséquent, la méthode choisie est un algorithme de tracé de particule : les lignes sont calculées en suivant la trajectoire d'un nombre fini de particules neutres dans le réservoir depuis les zones d'injection (puits ou aquifère) vers les zones de production (puits) ou les limites du réservoir. En pratique, ce tracé s'effectue de façon itérative : la position d'un point sur la ligne est déterminé à partir de celle du point précédent. Cette procédure est répétée tant qu'une condition d'arrêt n'est pas remplie (nombre d'itérations ou atteinte d'une zone de sortie). Enfin, le tracé s'effectue indifféremment dans le sens du courant (tracé de type *forward*) ou dans le sens inverse (tracé de type *backward*) [Voillemont et Royer, 2001].

Définition des points de départ

La première étape pour calculer une ligne de courant à l'aide d'un algorithme de tracé de particule consiste à définir les points de départ. Ceux-ci sont placés aux niveaux des zones d'injection si le tracé est de type *forward* ou des zones de production dans le cas contraire. La méthode de placement est la même quel que soit le cas. L'approche choisie est celle proposée dans Batycky [1997], où les lignes de courant commencent directement sur les faces des cellules injectrices ou productrices (figure (A.5)). Cette approche,

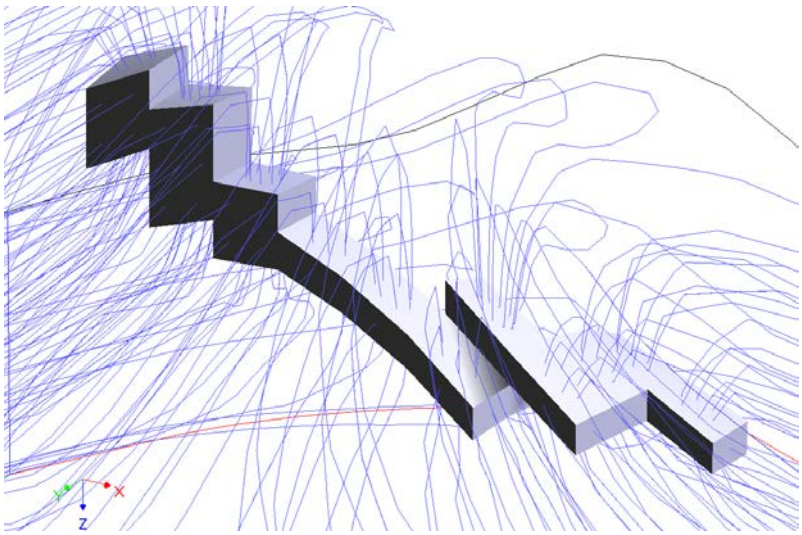


FIG. A.5 – Vue détaillée du départ des lignes de courant dans le cas d'un puits dévié.

- permet de distribuer les lignes de courant proportionnellement au flux de chacune des faces et ainsi d'obtenir des lignes représentatives de l'écoulement en terme de densité. Le débit total en fluides d'une ligne est égal au débit traversant la face d'où elle part divisé par le nombre de lignes associées à la dite face.
- permet de définir des points de départs en tout point de la grille, par exemple les cellules intersectées par les puits, les frontières du modèle, mais aussi, le cas échéant, les zones où la compressibilité du milieu agit comme une source ou un piège à fluides ;
- et est, comme nous le verrons par la suite, adaptée à la méthode de tracé de Pollock car les points sont directement placés sur les faces de cellule.

Cependant, elle est aussi sensible aux effets de rastérisation dans les cas de géométrie de puits complexes ou de grilles non-structurées.

Une autre alternative [Voillemont et Royer, 2003] considère que l'écoulement est radial autour des puits. Par conséquent, les points de départs sont placés uniformément sur un cylindre à une distance donnée des puits. Cette approche permet de s'affranchir du type de grille (structurée ou non) sur laquelle les calculs sont effectués et ne souffre pas d'un effet de rastérisation. Cependant, la distribution uniforme des lignes de courant n'est pas adaptée aux hétérogénéités du champ de vitesse, les débits associés avec chaque ligne sont alors très hétérogènes.

Algorithmes de tracé de particules

Après avoir déterminé la position des points de départ des futures lignes de courant celles-ci sont construites à l'aide d'un algorithme de tracé de particule. Comme décrit précédemment, ce type d'algorithme est itératif : à partir d'un point donné, x_i, l'objectif est de calculer la position du point suivant, x_{i+1}, le long de la ligne de courant. Cette opération est ensuite répétée autant

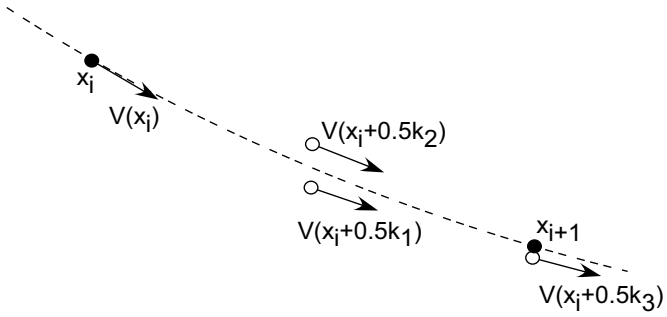


FIG. A.6 – Principe de l'estimation d'une ligne de courant à l'aide de la méthode de Runge Kutta.

de fois que nécessaire afin de construire l'ensemble de la ligne. Différentes approches peuvent être considérées : la méthode numérique de Runge-Kutta, déjà implantée dans les versions précédentes du plugin [Voillemont et Royer, 2003] et la méthode semi-analytique de Pollock [Chiang et Kinzelbach, 2001; Batycky et al., 1996].

Méthode de Runge-Kutta La technique de Runge-Kutta est une méthode d'intégration numérique, ici, de la vitesse pour le calcul de la trajectoire d'une ligne de courant. En pratique, elle s'appuie sur l'évaluation du vecteur vitesse en quatre points intermédiaires le long de la ligne de courant pour estimer la position du point x_{i+1} à partir de x_i, tel que (figure (A.6)) :

$$x_{i+1} \;=\; x_i \;+\; \frac{1}{6}(\mathbf{k}_1 + 2\ \mathbf{k}_2 + 2\ \mathbf{k}_3 + \mathbf{k}_4) \tag{A.19}$$

avec

$$\begin{aligned} \mathbf{k}_1 = h\ .\mathbf{v}(x_i), \quad \mathbf{k}_2 = h\ .\mathbf{v}(x_i + 1/2\ \mathbf{k}_1), \\ \mathbf{k}_3 = h\ .\mathbf{v}(x_i + 1/2\ \mathbf{k}_2), \quad \text{et,} \quad \mathbf{k}_4 = h\ .\mathbf{v}(x_i + \mathbf{k}_3) \end{aligned} \tag{A.20}$$

où, h est le pas d'espace entre x_i et x_{i+1} et, enfin $\mathbf{v}(x)$ le vecteur vitesse unitaire en x, point quelconque du domaine d'étude où l'on peut interpoler le champ de vitesse. Cette approche permet de construire des lignes de courant dont les points sont espacés d'un pas régulier. Il est aussi possible d'étendre l'approche pour construire des lignes à pas de temps constant [Voillemont et Royer, 2003].

D'une façon générale, la méthode de Runge-Kutta est relativement rapide et stable numériquement. Cependant, elle est aussi sensible aux fortes variations du champ de vitesse. En effet, dans de telles conditions, la ligne de courant présente une forte courbure et n'est plus parfaitement tangente aux estimations de la vitesse utilisés lors de l'intégration [Press et al., 2002]. Ceci peut entraîner un non-respect de la loi de conservation de la masse qui est l'une des contraintes fondamentales des simulations d'écoulement. Pour palier cet effet, une solution est de réduire le pas d'espace ou de temps considéré lors du tracé, mais au prix d'un volume plus important de calculs.

Méthode de Pollock Cette méthode, basée sur la formule analytique développée par Pollock [1988], calcule le point de sortie d'une particule dans une cellule hexaédrique orthogonale en

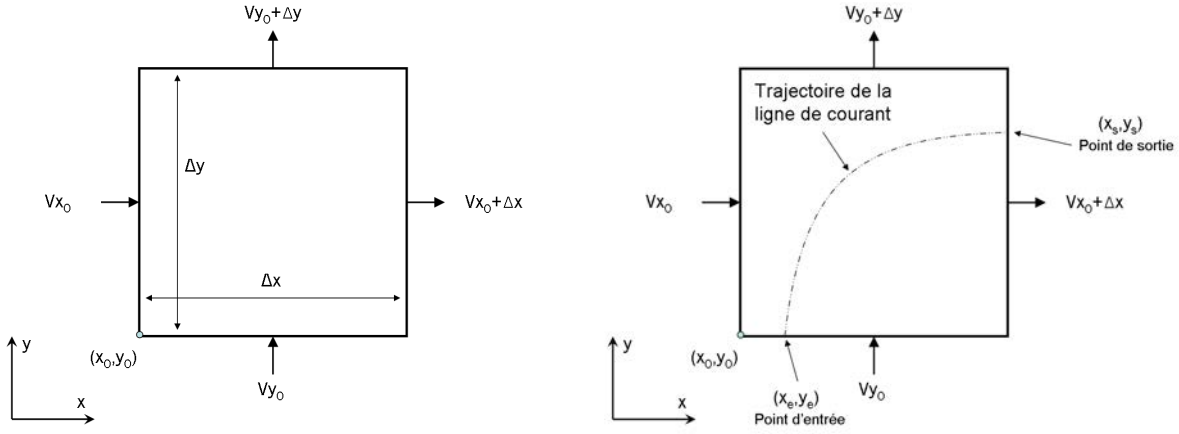


FIG. A.7 – Notations utilisées dans le cadre de la méthode de tracé de particules de Pollock en deux dimensions.

fonction de : (1) son point d'entrée et (2) des vitesses interstitielles moyennes des fluides sur chacune des faces de la cellule.

En pratique, cette méthode suppose que les composantes du vecteur vitesse sont indépendantes et qu'elles varient linéairement à l'intérieur d'une cellule. Le temps requis par une particule injectée au point(x_e, y_e, z_e) pour atteindre un face donnée, par exemple $x = x_0$, est alors :

$$\Delta_{t_{x_e \to x_0}} = \frac{1}{m_x} \ln \left(\frac{v_{x_0} + m_x(x_0 - x_0)}{v_{x_0} + m_x(x_e - x_0)} \right) \tag{A.21}$$

où

- m_x est le gradient de la composante de la vitesse dans la direction x :

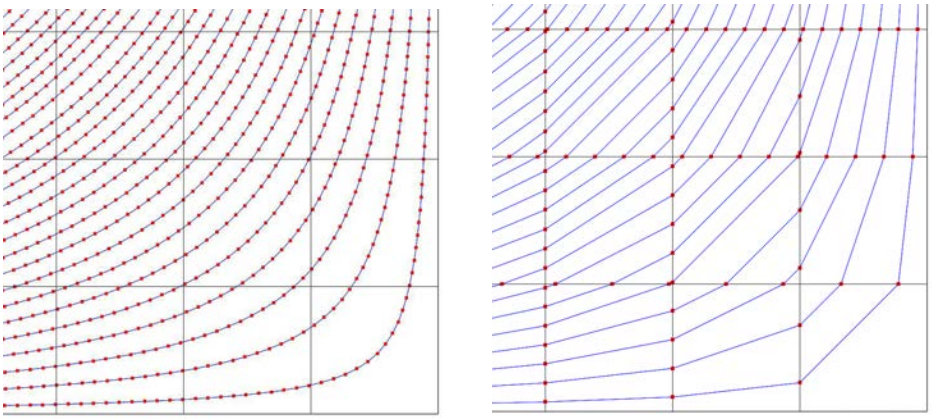
$$m_x = (v_{x_0+\Delta_x} - v_{x_0})/(\Delta_x) \tag{A.22}$$

- x_0 est la coordonnée du point d'origine de la cellule ;
- v_{x_0} est la composante de la vitesse dans la direction x en $x = x_0$;
- $v_{x_0+\Delta_x}$ correspond à la même composante sur la face opposée en $x = x_0 + \Delta_x$;
- Δ_x est la dimension de la cellule dans la direction x.

L'approche est identique pour les autres directions de l'espace. À partir de (A.21), le temps nécessaire pour atteindre chaque face de la cellule est calculé, la face de sortie étant alors celle pour laquelle cette valeur est minimale. Il est ensuite possible de calculer explicitement la position du point de sortie telle que :

$$\left| \begin{array}{rcl} x_s &=& \frac{1}{m_x}(v_{x_e} e^{m_x \Delta t} - v_{x_0}) \\ y_s &=& \frac{1}{m_y}(v_{y_e} e^{m_y \Delta t} - v_{y_0}) \\ z_s &=& \frac{1}{m_z}(v_{z_e} e^{m_z \Delta t} - v_{z_0}) \end{array} \right. \tag{A.23}$$

L'application de cette technique sur l'ensemble des cellules traversées par une particule donnée permet ainsi de construire la ligne de courant correspondante. Il faut cependant noter que cette



(a) Tracé de Runge-Kutta (b) Tracé de Pollock

FIG. A.8 – Exemple de lignes de courant en deux dimensions tracées avec la méthode de Runge-Kutta (a) et la méthode de Pollock (b). En noir maillage de la grille sous-jacente, en bleu trajectoire suivie par la ligne de courant et en rouge points calculés lors du tracé . La méthode de Runge-Kutta permet ainsi de suivre précisément la trajectoire des lignes de courant dans les cellules mais au prix d'une plus grande densité de points. Par opposition, la méthode de Pollock permet de suivre les trajectoires uniquement d'une face de cellule à la suivante.

trajectoire n'est effectivement connue qu'au niveau des faces tandis que, dans une cellule, elle est interpolée linéairement.

Cette technique est attractive et a été choisie pour ces travaux car : (1) elle s'adapte au type de grille considérée, (2) elle utilise directement le champ de vitesse calculé avec la méthode des volumes finis, (3) elle est semi-analytique, ce qui permet une plus grande précision lors de l'estimation de la trajectoire des particules, et enfin (4) elle permet le respect strict de la loi de conservation de la masse. Cette méthode initialement décrite dans Chiang et Kinzelbach [2001] est adaptée aux grilles régulières à mailles orthogonales. Dans notre cas, les cellules n'étant pas nécessairement orthogonales, une transformation iso-paramétrique telle que proposée dans Prévost et al. [2001] est considérée.

A.2.4 Résolution de l'équation de saturation

Une fois les lignes de courant obtenues, l'objectif est de résoudre l'équation de saturation le long de celles-ci. Cette équation s'écrit alors :

$$\phi \frac{\partial S_w}{\partial t} + |\mathbf{u}_t| \frac{\partial f_w}{\partial s} = 0 \tag{A.24}$$

où, s est l'abscisse curvilinéaire le long de la ligne. Cette équation peut alors être mise sous forme adimensionelle , à l'aide du temps de vol, noté τ [Batycky, 1997]. Il est défini comme le temps nécessaire à une particule neutre pour parcourir une distance donnée, s, le long d'une ligne de

courant [Datta-Gupta et King, 1995] :

$$\tau(s) = \int_0^s \frac{\phi(\zeta)}{\|\mathbf{u}_t(\zeta)\|} d\zeta \tag{A.25}$$

Pratiquement, le temps de vol est intégré numériquement lors du tracé de la ligne de courant. Il est possible ensuite de réécrire (A.24) sous la forme :

$$\frac{\partial S_w}{\partial t} + \frac{\partial f_w}{\partial \tau} = 0 \tag{A.26}$$

L'équation (A.26) est ensuite résolue à l'aide de méthodes purement analytiques (solution de Buckley-Leverett [Dake, 2001]) ou bien à l'aide de méthodes numériques. Les méthodes analytiques sont relativement rapides et efficaces car non-sujettes à la diffusion numérique. Cependant, elles requièrent une saturation initiale uniforme le long de la ligne de courant. Cela limite donc les phénomènes pouvant être modélisés, en particulier, les changements dans la configuration des puits ou la ségrégation des fluides due à la gravité. Par conséquent, les solutions numériques ont été préférées, en particulier le schéma décentré à gauche, encore appelé schéma *Single Point Upstream* (SPU). Celui-ci sera comparé au schéma à variation totale décroissante, encore appelé *Total Variation Diminishing* (TVD) proposé dans Voillemont et Royer [2003].

Schéma de résolution

La résolution de l'équation de saturation le long des lignes de courant l'équation(A.26) suivant une approche numérique s'appuie sur la discrétisation de l'équation(A.26) à la fois dans l'espace (le support sera alors les points de la ligne) et dans le temps. Le schéma SPU pour *Single Point Upstream*, choisi dans ces travaux, est de type différences finies et se caratérise par (figure (A.9)[1] :

– la discrétisation décentrée à gauche du terme d'espace au temps t :

$$\frac{\partial f_w}{\partial \tau} \simeq \frac{f_{w,i}^t - f_{w,i-1}^t}{\Delta\tau} \tag{A.27}$$

– la discrétisation explicite du terme temporel entre $t+1$ et t à la position i :

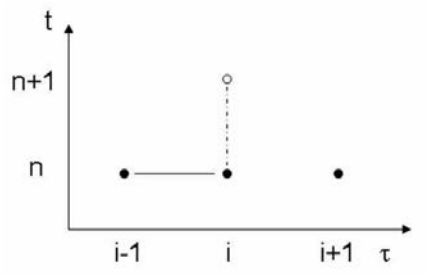
$$\frac{\partial S_w}{\partial t} \simeq \frac{S_{w,i}^{t+1} - S_{w,i}^t}{\Delta t} \tag{A.28}$$

Le schéma de calcul s'écrit alors :

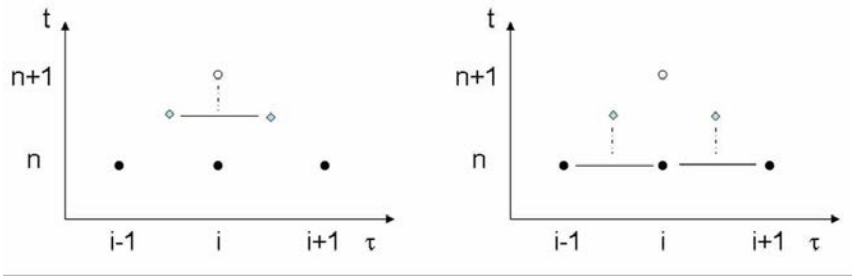
$$S_{w,i}^{t+1} = S_{w,i}^t + \frac{\Delta t}{\Delta \tau} \left(f_{w,i}^t - f_{w,i-1}^t\right) \tag{A.29}$$

Contrairement à la résolution de l'équation de pression pour laquelle la procédure de discrétisation conduisait à un ensemble d'équations algébriques impliquant les valeurs inconnues de pression à un nœud et à ses voisins, ici la seule inconnue est $S_{w,i}^{t+1}$. On parle alors de schéma explicite. Ainsi, pour estimer l'évolution de la saturation au cours du temps, il suffit d'effectuer

[1]Dans la suite, l'indice i indique la position dans l'espace et l'indice t dans le temps. On suppose les valeurs au temps t connues, la valeur inconnue à déterminer étant à la position $(i,\, t+1)$



(a) Schéma SPU



(b) Schéma TVD

FIG. A.9 – Schémas numériques de résolution de l'équation de saturation en eau le long des lignes de courant. Le schéma de type *Single Point Upstream* (SPU) est un schéma décentré à gauche explicite en temps. À l'opposé, le schéma à variation totale décroissante ou *Total Variation Diminishing* (TVD) est explicite en temps mais centré, avec une discrétisation au niveau de point intermédiaires (en bas à gauche) suivie d'une discrétisation finale au niveu des points de données (en bas à droite). Sur l'ensemble des schémas, l'axe vertical représente la temps et l'axe horizontal le temps de vol. Les points noirs représentent les valeurs connues de la saturation et les point blancs la valeur que l'on cherche à déterminer.

ce calcul (A.29) sur l'ensemble de la grille jusqu'à atteindre la limite de validité du champ de pression.

Ce schéma a l'avantage d'être décentré à gauche, donc conservatif [Press et al., 2002], c'est-à-dire qu'il limite la diffusion numérique, et d'être explicite en temps, c'est-à-dire que l'estimation de la valeur inconnue en $(i,\, t+1)$ ne dépend que de valeurs connues au temps précédent en $(i-1,\, t)$ et $(i,\, t)$. Ce schéma est ainsi simple à mettre en œuvre, robuste et ne requiert qu'une quantité limitée de calculs. Il est donc parfaitement adapté à la simulation de l'évolution d'un front de saturation à l'échelle d'un réservoir pétrolier.

Comparativement, le schéma TVD combiné avec une méthode de compression artificielle utilisée par Voillemont et Royer [2003] est plus précis. Dans ce schéma, les termes de dérivées spatiales et temporelles à des pas d'espace, respectivement de temps, intermédiaires correspondants à la moitié des pas initiaux (figure (A.9)). C'est une méthode numérique d'ordre deux qui permet d'améliorer la précision et la stabilité des calculs car elle limite fortement la diffusion numérique. Cependant, l'utilisation de points intermédiaires augmente aussi la quantité de calcul nécessaire. En conclusion, dans notre cas d'étude, (1) au vu du volume de calcul à effectuer

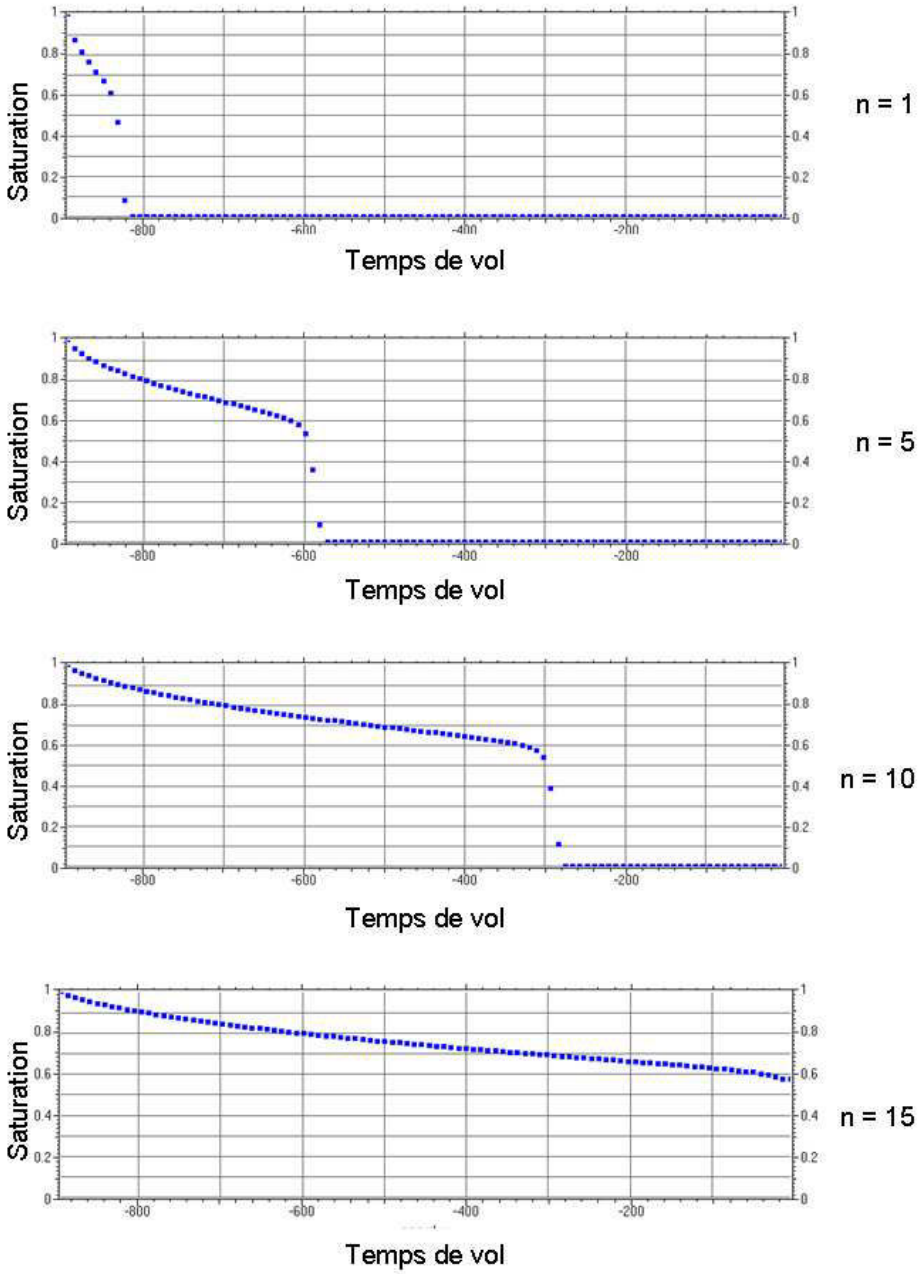


FIG. A.10 – Résolution de l'équation de saturation (équation de Buckley-Leverett) le long d'une ligne de courant. Initialement, la ligne de courant est remplie d'huile avec une injection continue d'eau à une extrémité. Les fluides sont non-miscibles, incompressibles et avec des courbes de perméabilité relative de Corey ($kr_w = S_w^2$ et $kr_o = (1 - S_w)^2$). Les figures représentent les résultats pour différentes itérations du solveur 1D : de haut en bas, itération 1, 5, 10 et 15. Sur l'ensemble des figures, l'axe vertical est la saturation en eau et l'axe horizontal est le temps de vol.

à l'échelle d'un réservoir pétrolier et (2) sachant que le schéma de type SPU est conservatif et limite relativement bien la diffusion en avant du front de saturation, ce dernier a été préféré.

Critère de stabilité numérique

Afin de pouvoir calculer une solution stable à l'équation (A.29), il est nécessaire de prendre en compte un critère de stabilité pour contraindre le pas de temps numérique Δt. Dans le cas d'un schéma de calcul explicite en temps, ce pas doit être inférieur au temps requis par le front de saturation le plus rapide pour parcourir la distance entre deux nœuds. Dans le cas contraire, il serait impossible de suivre exactement sa position ce qui conduirait à des résultats erronés. Ceci forme la base du critère de stabilité de Courant-Friedrich-Lewy aussi appelé condition de Courant. Mathématiquement, il s'écrit sous la forme :

$$\frac{\Delta t \; v_{max}}{\Delta \tau} \; \leq \; N_c \tag{A.30}$$

où, N_c est un facteur, appelé nombre de Courant choisi entre 0 et 1, il permet d'équilibrer la précision des calculs et le temps nécessaire. D'autre part, v_{max} représente la vitesse maximale des fronts de saturation le long de la ligne de courant et est égal à :

$$v_{max} = \left(\frac{\partial f_w(S_w)}{\partial S_w} \right)_{max} \tag{A.31}$$

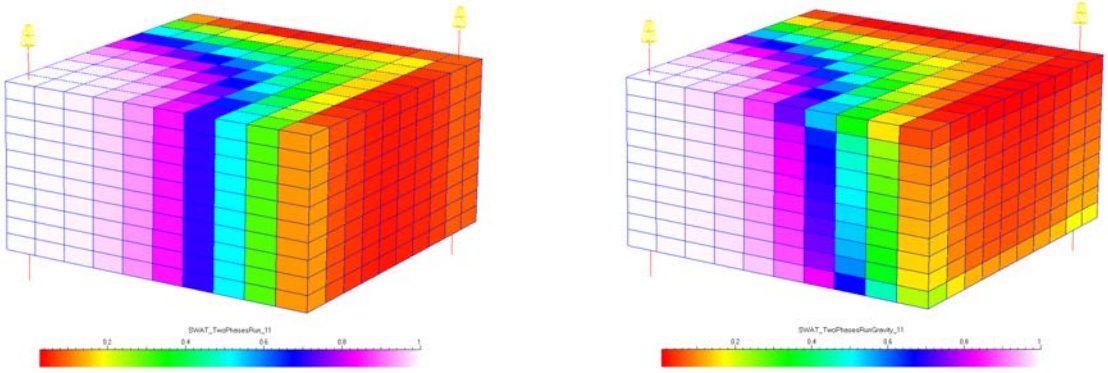
Pour une ligne de courant donnée, l'évaluation du pas de temps numérique s'effectue en prenant le minimum des pas de temps locaux, noté $\Delta t_i, \forall i \in [1, n_{sl}]$, calculés en chaque point de la ligne tel que :

$$\Delta t_i = \frac{N_c \; \Delta \tau_i}{v_{max,i}} = \frac{N_c \; (\tau_{i-1} - \tau_i) \; (S_{w,i-1} - S_{w,i})}{f_w(S_{w,i-1}) - f_w(S_{w,i})}, \forall i \in [1, n_{sl}] \tag{A.32}$$

Re-échantillonnage des lignes de courant

Le critère de stabilité présenté ci-dessus dépend fortement du temps de vol entre les points d'une ligne de courant. Or, à cause de la méthode de tracé de particule, les valeurs de $\Delta \tau_i$ sont très hétérogènes ce qui conduit dans la plupart des cas à des pas de temps ridiculement petits. Comme suggéré dans Batycky et al. [1996], une solution consiste à re-échantillonner la ligne de courant de telle façon que ses points soient espacés par un pas à temps de vol constant. Celui-ci est choisi tel que le nombre de points après re-échantillonnage soit le double du nombre de points initial.

Ceci permet de calculer pour une ligne de courant donnée un pas de temps numérique unique et optimal pour la résolution de l'équation de saturation réduisant, ainsi, fortement le volume de calcul. Cela se fait au prix d'approximations et d'imprécisions lors du re-échantillonnage, néanmoins celles-ci sont jugées négligeables par rapport à celles liées aux multiples transferts entre la grille tri-dimensionnelle représentant le réservoir et les lignes de courant.



(a) (b)

FIG. A.11 – Distributions de la saturation en eau dans un modèle de réservoir synthétique sans (a) et avec (b) l'effet de la gravité.

A.2.5 Transfert des résultats des lignes de courant vers la grille

Une fois le temps de validité du champ de pression atteint, la distribution de la saturation dans le réservoir est transférée depuis les lignes de courant vers la grille. Pour cela, lorsque plusieurs lignes traversent la même cellule, les saturations sont moyennées. En pratique, l'effet de chaque ligne de courant est pondéré par (1) le temps de vol des particules suivant sa trajectoire dans la cellule et (2) le débit qui lui est associé. Soit une cellule $\mathcal{C}$ traversée par $n_{sl,\mathcal{C}}$ lignes de courant, la saturation dans la cellule, $S_{w,\mathcal{C}}$, est alors égale à :

$$S_{w,\mathcal{C}} = \sum_{i=0}^{n_{sl,\mathcal{C}}} w_i \; S_{w,i} \quad avec \quad w_i = \frac{q_i \Delta\tau_{i,\mathcal{C}}}{\sum_{j=0}^{n_{sl,\mathcal{C}}} q_j \Delta\tau_{j,\mathcal{C}}} \tag{A.33}$$

où, pour une ligne de courant donnée i, $S_{w,i}$ est sa saturation en eau au niveau de la cellule considérée, q_i son débit total en fluide et $\Delta\tau_{i,\mathcal{C}}$ le temps de vol mis par une particule suivant sa trajectoire pour traverser $\mathcal{C}$.

A.2.6 Modélisation de l'effet de la gravité

Comme souligné précédemment, l'écoulement et la ségrégation des fluides dus à la gravité ne s'effectuent pas le long des lignes de courant et n'a donc pas été inclus dans le solveur numérique uni-dimensionel. Afin de pouvoir modéliser cet effet, il est nécessaire de procéder à un post-processus à la fin du pas de temps de pression une fois que les saturations ont été transférées sur la grille. Celui-ci s'effectue le long de pseudo "lignes" de gravité alignées avec l'accélération de la pesanteur $\mathbf{g}$.

Dans le cas présent, ces pseudo lignes sont supposées confondues avec l'axe vertical de la grille. Le schéma de calcul considéré pour approcher l'équation (A.8) est alors le suivant :

$$S_{w,i}^{t+1} = S_{w,i}^{t} + \frac{\Delta t}{\phi_i \Delta z_i}(G_{i+\frac{1}{2}} - G_{i-\frac{1}{2}}) \tag{A.34}$$

où, pour une cellule donnée de la grille et indiquée par l'indice i, $S_{w,i}^{t}$ représente sa saturation en eau au temps t, ϕ_i sa porosité, Δz_i sa hauteur et $G_{i+\frac{1}{2}}$, respectivement $G_{i-\frac{1}{2}}$, le débit dû à la gravité au travers de sa face supérieure, respectivement inférieure. Ces valeurs sont estimées, suivant les formules données dans Batycky [1997] égales à :

$$G_{i+\frac{1}{2}} = \frac{(k_{z,i+1}\ kr_{o,i+1})\ (k_{z,i}\ kr_{w,i})\ (\rho_w\ -\ \rho_o)\ g}{(\mu_w\ k_{z,i+1} kr_{o,i+1}) + (\mu_o\ k_{z,i} kr_{w,i})} \tag{A.35}$$

et

$$G_{i-\frac{1}{2}} = \frac{(k_{z,i}\ kr_{o,i})\ (k_{z,i-1}\ kr_{w,i-1})\ (\rho_w\ -\ \rho_o)\ g}{(\mu_w\ k_{z,i} kr_{o,i}) + (\mu_o\ k_{z,i-1} kr_{w,i-1})} \tag{A.36}$$

Enfin, Δt est un pas de temps numérique tel qu'une solution stable au schéma de calcul (A.34) puisse être obtenue. De même que pour l'équation (A.29), le critère choisi est le nombre de Courant, pour une cellule donnée, Δt_i vaut :

$$\Delta t_i = \frac{N_c \phi_i \Delta z_i}{G_{i+\frac{1}{2}} - G_{i-\frac{1}{2}}} \tag{A.37}$$

Δt est alors le minimum des Δt_i. Finalement, de même que pour l'équation de saturation, ce schéma (A.34) est appliqué de façon itérative sur l'ensemble de la grille jusqu'à atteindre la durée de validité du champ de pression. La figure (A.11) illustre la prise en compte l'effet de la gravité lors de la simulation de répartition des saturations dans un réservoir.

A.3 Apport de StreamLab-2.0.0

Comme précisé dans l'introduction, le plugin StreamLab a été initialement développé pour effectuer des simulations d'écoulements mono-phasiques en milieu granitique à l'aide de lignes de courant [Voillemont et Royer, 2003]. Les principales applications étaient la simulation d'échange de chaleur entre un fluide et la roche ainsi que la simulation d'une injection ponctuelle de traceurs chimiques. Les contraintes principales de ces travaux étaient de pouvoir tracer de façon précise un nombre modeste de lignes de courant et surtout de disposer de schémas de résolution numériques uni-dimensionels de grande qualité limitant le plus possible la dispersion numérique. En outre, ces outils avaient pour objectif de pouvoir traiter tout type de grille : cartésienne, stratigraphique et non-structurée.

Dans le cadre de simulations d'écoulements bi-phasiques en milieu poreux et à l'échelle d'un réservoir pétrolier, le code existant a été transformé sur différents points. En effet, ce type de simulation requiert le respect strict de la loi de conservation de la masse et un volume de calcul plus important : (1) les grilles traitées sont de grande taille ($\geq$ 1 million de cellules), (2) le

nombre de lignes de courant à manipuler est considérable (≥ 1000) et, (3) il est nécessaire de procéder à des mises à jour régulières du champ de pression et donc de la géométrie des lignes de courant. Pour atteindre ces objectifs, les principales modifications du code de calcul ont été :

- l'implantation d'un solveur numérique basé sur la méthode des volumes finis pour le calcul du champ de pression et de la vitesse des fluides. Celui-ci remplace le simulateur Thermass et le plugin FlowLab basés sur la méthodes des éléments finis. Les avantages sont (1) le respect strict de la loi de conservation de la masse et (2) l'intégration complète du code dans StreamLab. Néanmoins, l'implantation courante se limite aux grilles structurées à mailles hexaédriques ;
- l'ajout d'un modèle de Peaceman pour représenter la relation entre les puits et le milieu poreux lors du calcul du champ de pression qui permet un meilleur contrôle des puits que ce soit en débit ou en pression ;
- la construction des lignes de courant suivant la méthode semi-analytique de Pollock plutôt que l'interpolation de Runge-Kutta. Ceci (1) diminue le volume de calcul nécessaire et (2) améliore le strict respect de la loi de conservation de la masse. Mais, de même qu'au point précédent, l'approche choisie reste limitée aux grilles structurées à mailles hexaédriques ;
- la résolution de l'équation de transport le long des lignes de courant à l'aide d'un schéma numérique explicite de type décentré à gauche (schéma SPU). Comparé au schéma à variation totale décroissante (schéma TVD), celui-ci est certes plus sujet à la diffusion numérique, mais est beaucoup plus rapide car il ne nécessite pas l'utilisation de points intermédiaires lors de la discrétisation ;
- enfin, la prise en compte de la gravité qui joue un rôle majeur dans le cas d'écoulements biphasiques avec des fluides de densités différentes. Cette prise en compte s'effectue à la fois au niveau du calcul du champ de pression, mais aussi au niveau de l'estimation de l'évolution des saturations à l'aide d'un traitement post-processus.

En résumé, pour pouvoir effectuer des simulations biphasiques en milieu poreux, d'une part, le code de calcul a été spécialisé sur les grilles structurées à mailles hexaédriques (cartésienne ou stratigraphique) et, d'autre part, les algorithmes mis en œuvre ont été sélectionnés pour leur rapidité au prix, parfois, d'un manque de précision. Une seconde amélioration concerne la structure de donnée pour représenter et manipuler les lignes de courant. Celle-ci est détaillée dans la section suivante.

A.4 Structure de données pour la représentation et la manipulation de lignes de courant

Comme décrit précédemment, une ligne de courant correspond à la trajectoire suivie au cours du temps par une particule dans un écoulement. Une telle ligne a les caractéristiques suivantes (voir la figure A.12) :

- sa géométrie est unique, constante dans le temps et n'est jamais discontinue ;

- elle est définie par une liste ordonnée de points à quatre dimensions : les coordonnées géométriques (x, y, z) ainsi que le temps de vol (τ) ;
- elle est orientée, généralement du puits injecteur vers le puits producteur ;
- plusieurs lignes peuvent être regroupées sous forme de réseau ;
- différentes variables lui sont associées : son débit en fluides mais aussi, pour chacune de ses extrémités, l'index de la cellule correspondante dans la grille ainsi que son type (puits producteur, puits injecteur, frontière du modèle, etc.) ;
- et, pour finir, c'est à la fois un objet géométrique mais aussi un support de calculs et de propriétés.

Les sections suivantes décrivent les structures de données **StreamLine** et **StreamLineGroup** considérées pour représenter, respectivement, une ligne de courant et un réseau de lignes de courant.

A.4.1 Structures de données pré-existantes

Précédemment, les objets représentant les lignes de courant étaient basés sur des structures de données semblables, **ILine** et **PLine** qui représentent, respectivement, une courbe continue unique tri-dimensionnelle et un groupe de courbes dans le géomodeleur Gocad. D'un point de vue pratique, (figure A.13), une **ILine** est formée à partir d'une liste doublement chaînée dont chaque nœud correspond à un point. Le stockage des propriétés est effectué au niveau des points à l'aide d'un tableau rassemblant de façon contiguë en mémoire les valeurs de propriétés qui lui sont associées. Cependant, cette forme de stockage présente les inconvénients suivants :

- elle présente un coût mémoire important, en effet, sa topologie étant explicite. En particulier, les nœuds ne sont pas stockés de façon contiguë sous forme de tableau et chacun doit conserver des pointeurs vers ses deux voisins ;
- elle est également peu efficace en terme de calcul car le parcours le long de la ligne s'effectue à partir d'une extrémité, élément par élément ;
- elle n'est pas orientée comme doit l'être une ligne de courant. Ainsi, lors des calculs, si l'on considère la liste des points d'une **Iline**, cela oblige à tester son orientation et à l'inverser, le cas échéant ;
- le stockage des propriétés est effectué nœud par nœud, et non de façon contiguë sous forme de tableau pour l'ensemble de la ligne de courant. Ceci se révèle coûteux en temps et en mémoire lors de la résolution des équations de transport le long de la ligne.

L'avantage de cette structure est qu'elle offre une grande flexibilité au niveau de l'ajout ou du retrait de points. Cependant, si l'on considère que les lignes de courant ont une géométrie fixe, une forme de stockage plus compacte et plus efficace est possible.

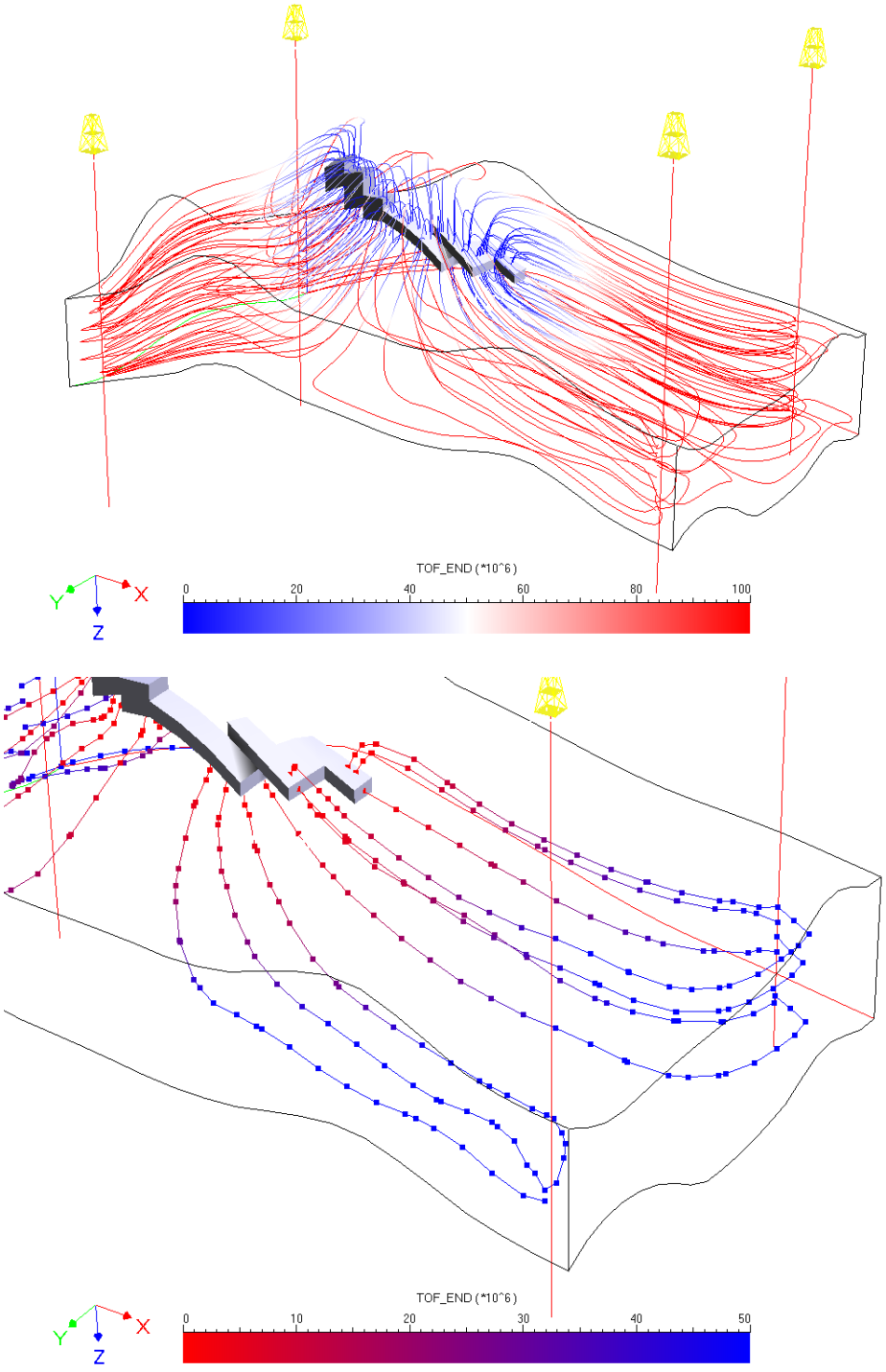


FIG. A.12 – Haut : exemple de réseau complexe de lignes de courant (objet **StreamLineGroup**) peintes avec la propriété de temps de vol au producteur. Le réservoir est un anticlinal avec un puits producteur dévié central et quatre puits injecteurs verticaux aux coins du modèle. Bas : Détail du réseau de lignes de courant précédent.

A.4.2 Implantation proposée

La nouvelle architecture d'implantation proposée reprend la structure globale précédente avec la séparation en deux objets : Ll'objet **StreamLine** représentant une ligne de courant unique et l'objet **StreamLineGroup** représentant un réseau de lignes de courant. Cette organisation est non seulement naturelle, mais, s'intègre aussi au sein du géomodeleur Gocad en se basant sur les structure pré-existantes que sont les **GObj**, ou "geological object" et **GObjGroup** ou groupe de **GObj**. Ainsi grâce à cette intégration, il est possible de bénéficier d'un environnement global comprenant le graphisme, la gestion d'une librairie d'objets, la possibilité d'effectuer des sauvegardes sous forme de fichiers ascii ou binaire, la gestion de propriétés, etc. Au contraire, les structures internes sont entièrement revues avec pour objectif d'optimiser le coût mémoire et l'accès aux données représentant la géométrie et les propriétés des lignes de courant.

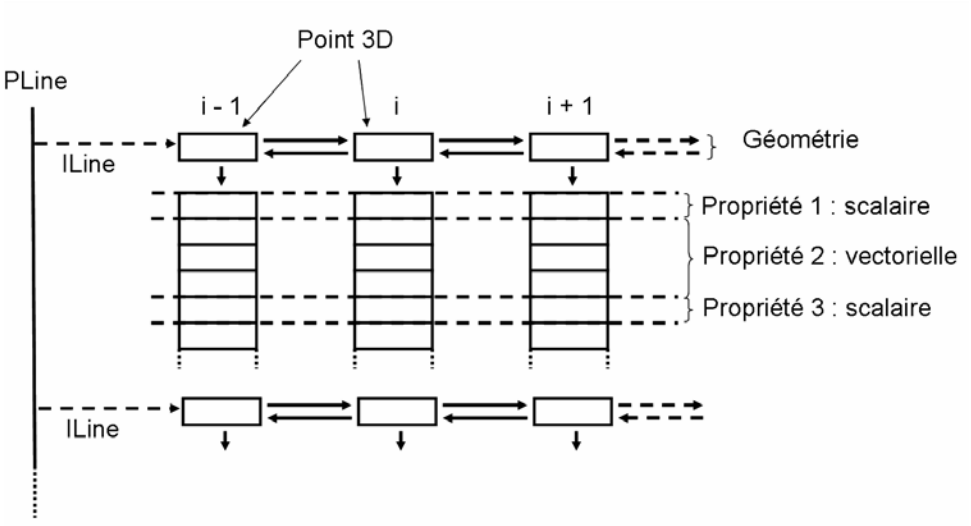
Géométrie D'un point de vue géométrique, l'objet **StreamLine** est composé des éléments suivants (figure A.13) :

- le nombre de points constituant la ligne de courant ;
- un tableau de points ordonnés depuis les zones d'injection de fluide (puits, frontière du domaine, etc.) vers les zones de production. Ces points sont stockés sous forme de **SLPoint3d** qui sont des **Point3D** auxquels une valeur de temps de vol est associée ;
- la valeur du débit total en fluides traversant la ligne de courant ;
- les deux extrémités qui sont aussi des **SLPoint3d** auxquels s'ajoute le type et l'index de la cellule correspondante.

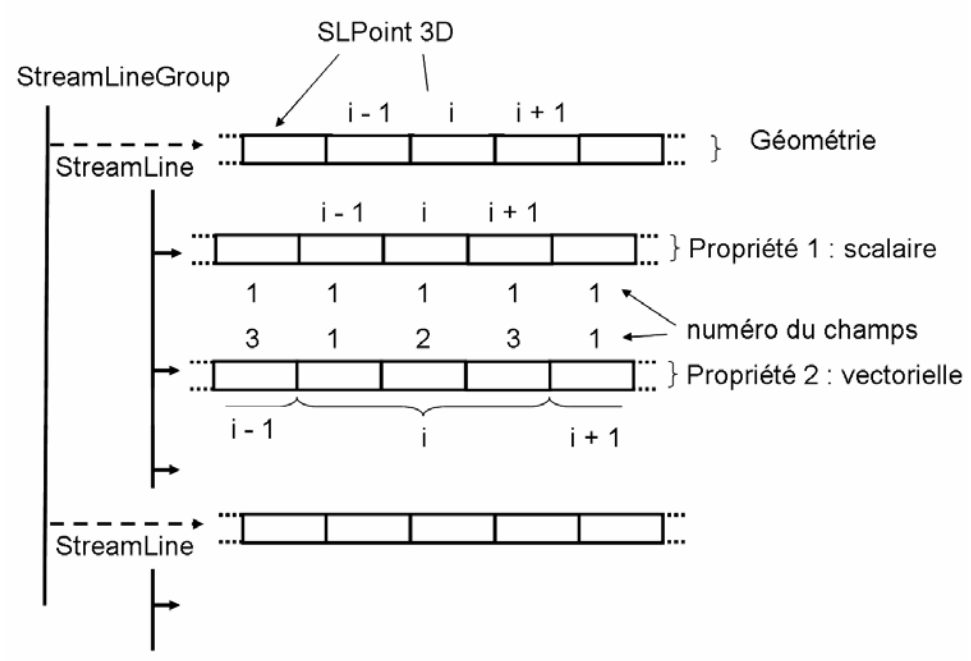
Gestion et stockage des propriétés La gestion et le stockage des propriétés est plus complexe, car (1) leur nombre et taille (i.e nombre de champs) varient, (2) elles doivent être définies pour l'ensemble du réseau et, enfin, (3) elles peuvent être associées à différents niveaux topologiques, soit aux points (dimension 0) soit aux segments (dimension 1). Ici, seul le niveau topologique de dimension 0 est considéré et l'architecture d'implantation a été découplée telle que :

- la gestion globale (nom, nombre de champs, ajout/retrait etc..) est effectuée au sein d'une base de donnée, appelée **SLDataPack**, associée à l'objet **StreamLineGroup** qui s'appuie sur la **PropertyDB** déjà intégrée dans le géomodeleur Gocad.
- le stockage effectif des données est réalisé au niveau de l'objet **StreamLine**. D'une façon similaire à la géométrie, chaque propriété est stockée de façon contiguë en mémoire sous forme d'un tableau de taille équivalente au nombre de points de la ligne.

Différentes extensions de cette implantation sont possibles : d'une part la gestion de régions sur les lignes de courant qui se ferait, le cas échéant, de façon similaire à celle des propriétés et, d'autre part, l'implantation de propriétés associées au niveau topologique de dimension 1, correspondant aux segments, comme par exemple le flux entre deux points.



(a) Structures de données **PLine** et **ILine**



(b) Structures de données **StreamLineGroup** et **StreamLine**

FIG. A.13 – Comparaison des structures de données considérées pour le stockage de la géométrie et des propriétés des objets courbes (PLine/ILine) et lignes de courant (StreamLineGroup/StreamLine) dans le géomodeleur Gocad.

Implantation	Nombre de points	Temps de tracé (s)	Mémoire occupée (Mo)	Gain en Mémoire (%)
ILine/Pline	296 758	24,2	46	-
StreamLine/StreamLineGroup	296 758	25	15	67
ILine/Pline	590 193	47,4	101	-
StreamLine/StreamLineGroup	590 193	47,5	10	71

TAB. A.1 – Comparaison du coût mémoire de l'architecture d'implantation de l'objet lignes de courant basée sur les objets pré-existants **ILine** et **PLine** avec celle proposée.

Les avantages de cette structure de données sont les suivants :
- elle intègre un certains nombre d'éléments déjà présents au sein du géomodeleur Gocad ;
- la géométrie ainsi que les propriétés d'une ligne de courant sont stockées de façon contiguë en mémoire. Le coût mémoire de la structure et le temps d'accès aux informations sont fortement diminués, les calculs qui en dépendent s'en trouvent donc accélérés (table A.1) ;
- le découplage **StreamLine** / **StreamLineGroup** permet une plus grande flexibilité dans la gestion du réseau, en particulier dans l'ajout ou le retrait de lignes de courant.

A.5 Validation du calcul

Le modèle physique et les techniques de résolution implantées dans StreamLab ont été validés en les comparant aux résultats obtenus à l'aide du simulateur sur lignes de courant commercial 3DSL [Streamsim, 2004].

A.5.1 Description du modèle

Description du réservoir Le réservoir est une grille 2D structurée régulière de dimensions 1000 ft × 1000 ft × 10 ft et divisée en 100 × 100 × 1 cellules. Les propriétés de la roche sont la porosité ϕ et la perméabilité horizontale k_h (le tenseur est alors supposé diagonal avec $k_x = k_y = k_h$). La porosité est estimée homogène dans tout le réservoir et égale à 20 % alors que la perméabilité horizontale est simulée à l'aide d'une méthode de simulation gaussienne séquentielle non conditionnée basée sur un variogramme sphérique isotrope. L'histogramme de la propriété est de type triangulaire entre 50 et 1000 mD avec un mode à 500mD (figure A.14).

Position et contrôle des puits Deux puits verticaux traversent le réservoir et sont modélisés à l'aide de la formule de Peaceman. Un puits injecteur d'eau est localisé au centre de la cellule (4,4,1), tandis qu'un puits producteur l'est dans le coin opposé en (96,96,1). Les deux puits sont contrôlés en pression à, respectivement, 1000 et 400 psi. Enfin, chaque puits a un diamètre de 0.15 ft, un facteur de "skin" de 0 et un facteur de complétion de 1.

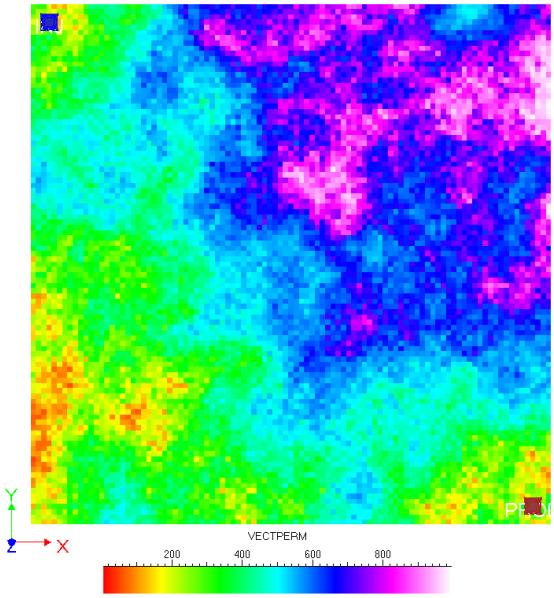


FIG. A.14 – Champ de perméabilité horizontale utilisé pour la validation du plugin StreamLab. Celui-ci provient d'une simulation gaussienne séquentielle non conditionnée avec un variogramme sphérique isotrope et pour cible un histogramme triangulaire entre 50 et 1000 mD et un mode à 500mD

Propriétés des fluides Ce réservoir ne contient que deux fluides, l'huile et l'eau, qui sont considérés à la fois immiscibles et incompressibles. Leurs masses volumiques sont, respectivement, de 53 et 64 lb.ft^3, tandis que leurs densités sont, de 2.0 and 0.5 cP. Les courbes de perméabilité relative considérées dans cette étude sont de type Corey sans aucune saturation résiduelle que ce soit en huile ou en eau, c'est-à-dire : $kr_w(S_w) = S_w^2$, et $kr_o(S_w) = (1 - S_w)^2$. Enfin, le réservoir est supposé entièrement rempli d'huile à l'état initial et subit une injection d'eau.

A.5.2 Résultats

La production du réservoir a été simulée pendant 1000 jours. Les résultats obtenus par StreamLab et 3DSL sont montrés sue les figures suivantes. La figure (A.15) montre différentes cartes de saturations à différents pas de temps, tandis que la figure (A.16) compare l'évolution de la production cumulée ainsi que les débits de production en huile et en eau.

D'une façon générale, les résultats des deux simulateurs sont similaires ce qui est très satisfaisant. En terme de performance, les deux simulateurs requièrent un nombre semblable d'itérations pour le calcul de la pression et des lignes de courant (72 pour StreamLab et 73 pour 3DSL). Les temps de calcul sont approximativement de 2 min pour 3DSL et 6 min pour StreamLab sur un Pentium 4 à 3.2 GHz avec 1 Go de RAM.

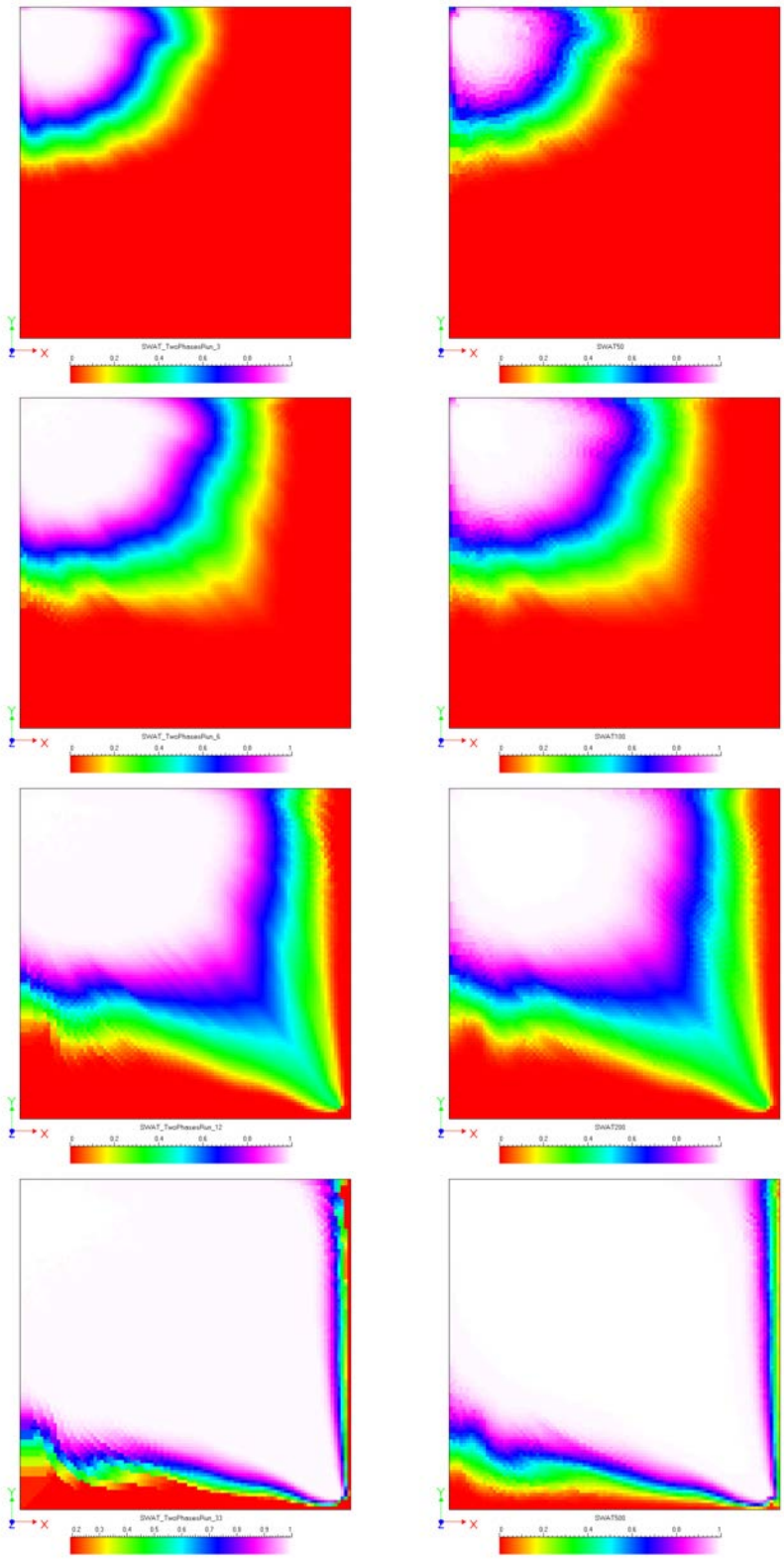
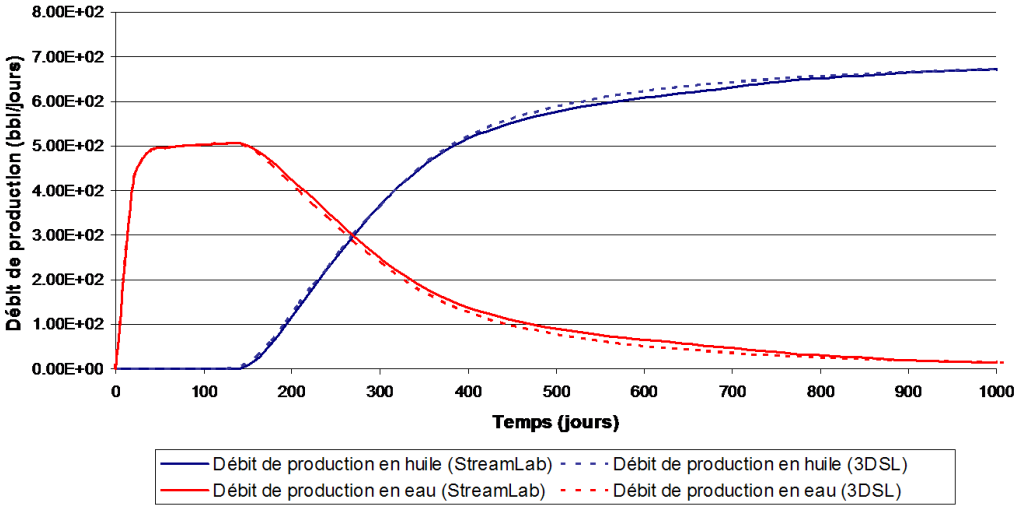


FIG. A.15 – Profils de saturation à différents pas de temps, de haut en bas : 50, 100, 200 et 500 jours. Les résultats montrés proviennent, à gauche, du plugin StreamLab et à droite du simulateur sur ligne de courant 3DSL.



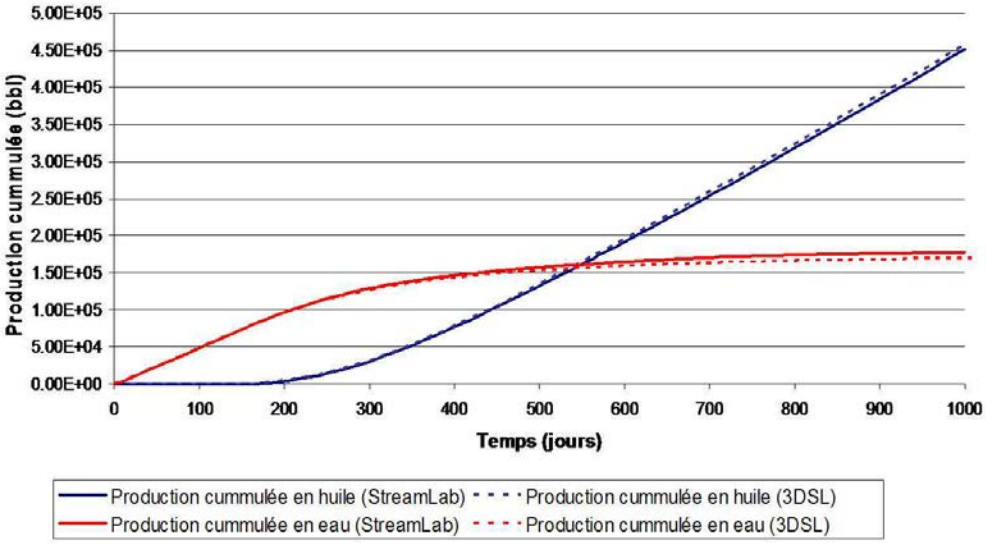


FIG. A.16 – Comparaison des courbes de production obtenues à l'aide des simulateurs sur ligne de courant StreamLab et 3DSL. En haut évolution du débit en huile et en eau au cours du temps et, en bas, courbes de production cumulées.

A.6 Conclusions

Dans cette annexe, nous avons présenté le modèle physique ainsi que les techniques implantées dans le plugin StreamLab pour réaliser des simulations d'écoulements biphasiques sur lignes de courant en milieu poreux. Les principales caractéristiques de ce simulateur sont :

1. la prise en compte de grilles 3D irrégulières structurées et non faillées (aussi appelées grilles stratigraphiques) ;
2. la modélisation de deux phases l'huile et l'eau, qui sont de plus supposées non miscibles et incompressibles ;
3. la prise en compte de l'effet de la gravité, tant au niveau de la résolution de l'équation de pression que lors de la simulation de l'évolution des saturations ;
4. la modélisation des puits à l'aide du modèle de Peaceman et leur contrôle à la fois en pression et en débit.

L'objectif de ce travail était de pouvoir réaliser ces simulations à l'échelle d'un réservoir pétrolier, c'est à dire sur des modèles discrets de plusieurs million de cellules, avec un milieu poreux présentant de fortes hétérogénéités et éventuellement des conditions de puits variables. Par rapport aux travaux précédents le code de calcul a été spécialisé sur les grilles structurées hexaédriques (cartésienne ou stratigraphique) et les algorithmes mis en œuvre ont été sélectionnés pour leur rapidité au prix, parfois, d'un manque de précision. Une seconde amélioration concerne l'implantation d'une structure de donnée, plus compacte en mémoire et plus efficace en terme de calculs, pour représenter et manipuler les lignes de courant. Naturellement, divers développements peuvent être envisagés, notamment : (1) l'extension du code pour prendre en compte les failles, (2) l'implantation d'un modèle de fluides de type Black-Oil, avec la présence d'une phase gazeuse et la dissolution possible de gaz dans l'huile et vice-versa, (3) la modélisation de la compressibilité de la roche et éventuellement des fluides, (4) l'ajout de modèles analytiques pour représenter un ou plusieurs aquifères aux frontières du réservoir.

L'objectif de ce travail n'est pas de fournir un simulateur d'écoulement aussi précis et robuste qu'un simulateur commercial. Il convient plutôt de considérer ce plugin comme un ensemble d'outils pour effectuer une évaluation rapide du potentiel de production d'un réservoir pétrolier. Diverses applications pourraient être basées sur celui-ci, telles que, de nouvelles techniques de changement d'échelle (*upscaling*), la prise en compte d'un historique de production dans la modélisation d'un réservoir par le biais de méthodes d'inversion (*history matching*) et plus généralement la quantification des incertitudes liées aux écoulements comme cela a été réalisée durant ces travaux de thèse.

Annexe B

Description des structures de données NdPointSet et NdCube

Sommaire

B.1 Implantation d'objets dans le géomodeleur Gocad 169
B.2 La structure de donnée NdPointSet 170
B.3 La structure de donnée NdCube 172
B.4 Remarques finales 174

Les objets **NdPointSet** et **NdCube** sont deux structures de données développées et implantées durant ces travaux de thèse pour permettre la représentation et l'interpolation d'une fonction dans un espace de dimension quelconque. Dans ce cadre, ces structures représentent, respectivement, un ensemble de points et une grille régulière structurée plongés dans un espace de dimension quelconque. Elles ont été implantées en langage *C++* dans le géomodeleur Gocad. Cette annexe décrit les principaux éléments de l'organisation et de l'architecture d'implantation de ces deux structures.

B.1 Implantation d'objets dans le géomodeleur Gocad

Le stockage et la représentation d'objets géologiques dans le géomodeleur Gocad tant au niveau de la géométrie que des propriétés s'organisent autour des deux classes suivantes :

– la classe **GObj** pour "*Geological Object*" : cette classe a pour objectif d'offrir une interface et un ensemble de fonctionnalités communes à tous les objets. En outre, il est possible de rassembler plusieurs **Gobj**s sous la forme d'un groupe au sein de la classe **GObjGroup**, dans ce cas un certain nombre de fonctions sont disponibles pour faciliter la manipulation (c'est-à-dire l'ajout ou le retrait) des différents sous-objets. Enfin, c'est autour de cette classe que la plupart des fonctionalités de Gocad s'organisent que ce soit pour le graphisme, la sauvegarde des objets dans un fichier ascii ou binaire, ou bien la gestion d'une base de données d'objets en général, etc.

– la classe **PropertyDB** : cette classe est responsable de la gestion des propriétés sur les différents objets. Une propriété correspondant à la classe **Property** qui stocke son nom et diverses informations sur son graphisme ou ses statistiques (valeurs minimale et maximale, moyenne, etc.). En pratique, la classe **PropertyDB** assure l'accès, le retrait, l'ajout et la modification des données. Elle n'est cependant pas responsable de leur stockage effectif qui s'effectue au niveau de la classe dérivant de **GObj**. Là encore, l'objectif est d'offrir une interface globale pour l'ensemble des objets afin de favoriser le développement de fonctions qui ne dépendent pas du type effectif de l'objet.

En outre, ces deux classes sont accompagnées de classes annexes qui sont chargées, par exemple, de gérer les différents niveaux topologiques d'un objet (classe **TopologicalObject**) ou bien l'accès proprement dit aux valeurs de propriétés (classe **PropertyAcessor**), etc. Dans la suite pour chacun des deux objets **NdPointSet** et **NdCube** nous détaillerons leurs structures ainsi que leurs liens avec les deux classes décrites ci-dessus.

B.2 La structure de donnée NdPointSet

Le **NdPointSet** est une structure de données représentant un ensemble de points plongés dans un espace de dimension quelconque. L'un des critères qui guide son implantation est de favoriser l'ajout ou le retrait de points. L'organisation globale est semblable à celle de l'objet **VSet** qui représente dans Gocad un ensemble de points en 3D et en constitue l'extension en nD. L'architecture d'implantation du **NdPointSet** est décrite plus en avant ci-dessous et est illustrée sur le schéma (B.1).

Géométrie La base de cette structure est la classe **NdPointSet** proprement dite. Cette classe dérive directement de **GObj** et est composée des éléments suivants :

– la dimension et le nom de chacune des directions (encore appelées axes) de l'espace dans lequel est plongé l'objet ;

– l'ensemble des points contenus dans l'objet. Ces points sont de type **NdNode** et sont stockés dans le **NdPointSet** sous forme d'une liste doublement chaînée. Chaque **NdNode** représente un point de donnée unique et dérive, en pratique, de la classe **NdPoint** qui contient un tableau de float correspondants à ses coordonnées dans l'espace ;

– un pointeur vers une classe de type **NdDataPack** qui est responsable de la gestion des différentes propriétés (voir ci-dessous).

Gestion et stockage des propriétés La manipulation des propriétés est plus complexe et est divisée en deux niveaux : d'une part, le stockage effectif des données et, d'autre part, la gestion globale de l'ensemble (nombre, nom, informations, etc.) :

– le stockage des données proprement dit ainsi que l'accès aux différentes informations s'effectuent au niveau de chaque **NdNode**. Pour cela, ceux-ci dérivent de la classe **DataRec**

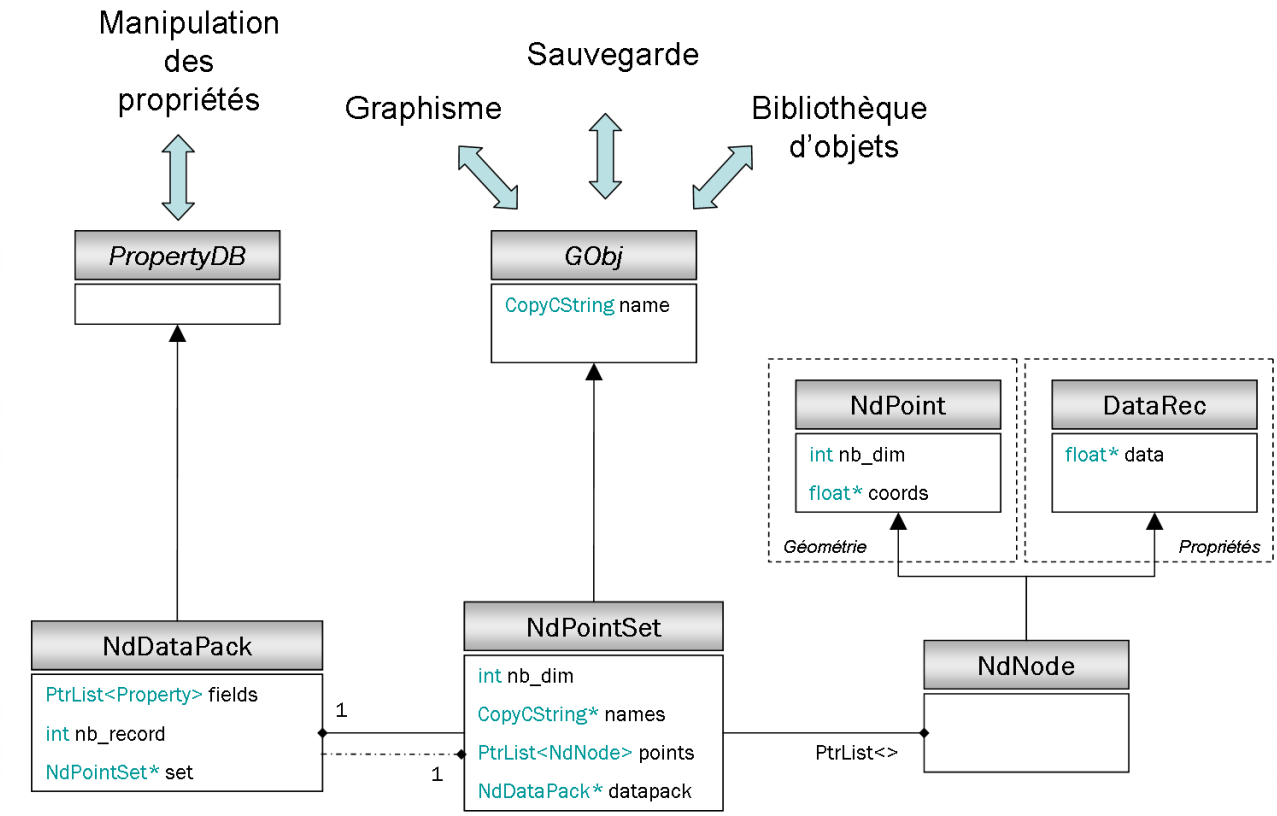


FIG. B.1 – Schéma illustrant l'organisation de l'objet **NdPointSet** et son intégration dans le géomodeleur Gocad

qui est un tableau de valeurs de dimension variable. Cependant à aucun moment, la classe **NdNode** n'a la maîtrise de la gestion de la taille de ce tableau ;

– en parallèle, la gestion globale des propriétés s'effectue dans un **NdDataPack** qui dérive du **PropertyDB**. Cette classe contient la liste des propriétés (classes **Property**) associées à l'objet courant et gère de façon effective la taille du **DataRec**. En particulier, elle est chargée de le re-dimensionner dans le cas où l'on ajoute ou retire une ou plusieurs propriétés.

Dans le détail, la structure d'un **NdPointSet** est relativement simple et s'intègre facilement dans le géomodeleur Gocad. Les deux points importants sont, d'une part, le stockage dans une même sous-structure compacte de toutes les informations (que ce soit la géométrie ou les propriétés) nécessaire pour définir un point et, d'autre part, l'organisation de l'ensemble des points sous forme d'une liste doublement chaînée où chaque point est indépendant, ce qui en favorise la manipulation.

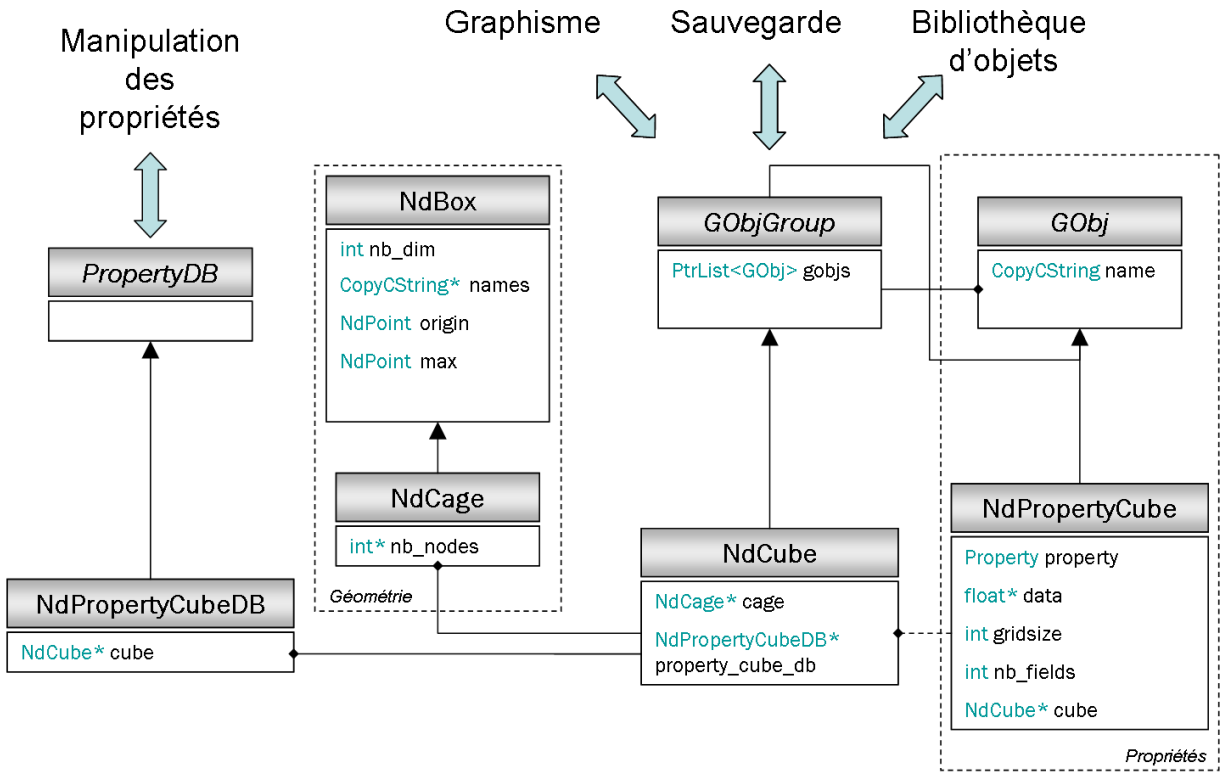


FIG. B.2 – Schéma illustrant l'organisation de l'objet **NdCube** et son intégration dans le géomodeleur Gocad

B.3 La structure de donnée NdCube

La structure de type **NdCube** représente une grille structurée régulière plongée dans un espace de dimension quelconque. Cet objet sert de support de calcul pour l'interpolation lisse discrète et doit permettre le stockage et la représentation du résultat. La contrainte principale de cet objet est la gestion de la mémoire. En effet, il faut pouvoir représenter la surface avec précision ce qui requiert un maillage fin et ce, dans un espace de grande dimension ($n > 3$). L'organisation des différentes classes nécessaires est décrite ci-dessous et est illustrée en détails sur le schéma B.2.

Structure de grille Comme expliqué précédemment, l'objet **NdCube** est une grille de type structurée régulière. "*Structurée*" signifie que que les connections entre les nœuds du maillage obéissent à une règle unique et sont donc prévisibles, tandis que "*régulière*" indique que la géométrie de l'ensemble des cellules du maillage est identique, c'est-à-dire, ici, de forme "hypercubique". L'avantage principal de ce choix est que la géométrie de la grille est entièrement implicite et peut être décrite à partir des quelques éléments clefs suivants :

- la dimension de l'espace dans lequel l'objet est plongé ;
- les deux **NdPoint** (voir ci-dessus) correspondants, respectivement, aux points de la grille de coordonnées minimales et maximales dans toutes les directions de l'espace ;
- le nombre de nœuds pour chaque axe de la grille ;
- et, pour finir, un tableau de chaînes de caractères correspondants aux noms des différents axes de la grille.

L'ensemble de ces informations sont regroupées dans les classes **NdBox** et **NdCage** qui forment donc la base géométrique du **NdCube**. Elle fournissent aussi un ensemble de fonctions qui effectuent les transformations nécessaires pour passer de l'index d'un point dans la grille (index ijk) aux indexes correspondant dans chacune des directions de la grille (indexes (i, j, k)) puis aux coordonnées dans l'espace (coordonnées (x, y, z)).

Manipulation des propriétés Dans le cas d'un **NdCube**, le stockage et la gestion des propriétés s'appuient aussi sur la structure de grille structurée régulière. En effet, comme le nombre et la disposition des nœuds sont prévisibles, les propriétés peuvent être stockées de façon contiguë en mémoire sous forme d'un tableau unique. Il est donc possible de les séparer de la partie géométrique. D'autre part, au vue de la taille des tableaux à manipuler, ceux-ci sont stockés sous la forme d'une liste chaînée qui permet plus facilement d'en ajouter ou d'en retirer. Ainsi l'implantation s'organise de la façon suivante :

- chaque propriété correspond à une classe de type **NdPropertyCube**. Celle-ci contient le tableau des données proprement dites, mais aussi, le nombre de champs de la propriété, et par le biais d'un membre de type **Property**, le nom et diverses informations sur son graphisme ou ses statistiques (valeurs moyennes, minimales et maximales, par exemple).
- l'architecture globale s'organise de façon similaire à celle du **Voxet** déjà présent dans le géomodeleur GOcad. En effet, la classe **NdPropertyCube** est considérée comme étant un **GObj** et en dérive. Cela permet alors de considérer le **NdCube** comme un **GObjGroup** de propriétés et donc de bénéficier des outils et fonctions déjà implantés pour la gestion d'un telle liste d'objets.
- enfin, d'une façon similaire aux **NdPointSet**, la gestion globale des propriétés est laissée à une classe, **NdPropertyCubeDB**, qui dérive de **PropertyDB**.

Les avantages de cette structure sont principalement :

- qu'elle reprend des éléments déjà présent dans Gocad, ce qui favorise sont intégration ;
- que sa géométrie est implicite et donc ne nécessite que peu d'informations ;
- et que les propriétés sont découplées de la géométrie et stockées de manière compacte en mémoire.

Le choix de cette architecture d'implantation permet de rendre cette structure de données la moins coûteuse en mémoire tout en facilitant sa manipulation. En outre, le découplage permet une extension plus facile de l'objet. Par exemple, il a été possible de redéfinir une grille de

type “structurée irrégulière” en modifiant le **NdCage** sans avoir à toucher aux propriétés et inversement, il est possible de redéfinir la notion de propriété, sous la forme d'une fonction analytique par exemple, sans modifier la cage.

B.4 Remarques finales

Les structures de données **NdPointSet** et **NdCube** permettent donc de représenter dans le géomodeleur Gocad, respectivement, un ensemble de points et une grille structurée régulière, tous deux plongés dans un espace de dimension quelconque. D'une façon générale, durant l'implantation de ces structures de données, l'attention a été portée principalement sur la distinction et le découplage entre la géométrie et les propriétés. En effet comme souligné dans la section (2.1.1), dans le cadre de l'interpolation de surfaces de réponse, la géométrie est constituée des coordonnées des points de données ou des nœuds du maillage dans l'espace des paramètres incertains tandis que les propriétés représentent les résultats de simulation ou d'autres informations nécessaires à l'interpolation. Ainsi une fois le problème posé, c'est-à-dire les paramètres incertains définis, la géométrie est considérée fixe tandis le nombre des propriétés doit pouvoir évoluer facilement. En outre, ces structures s'appuient largement sur les objets Gocad pré-existants et en sont en quelque sorte l'extension dans un espace de dimension quelconque. Cette intégration a permis d'utiliser de nombreuses fonctions dont le développement n'était pas essentiel, mais utiles, pour ces travaux de thèse.

AUTORISATION DE SOUTENANCE DE THESE
DU DOCTORAT DE L'INSTITUT NATIONAL
POLYTECHNIQUE DE LORRAINE

o0o

VU LES RAPPORTS ETABLIS PAR :
Monsieur Jean-Paul CHILÈS, Professeur, ENSMP, Fontainebleau

Monsieur Richard SINGING-LARSEN, Professeur, NTNU, Norvège

Le Président de l'Institut National Polytechnique de Lorraine, autorise :

Monsieur FETEL Emmanuel

NANCY BRABOIS
2, AVENUE DE LA
FORET-DE-HAYE
BOITE POSTALE 3
F - 54501
VANDŒUVRE CEDEX

à soutenir devant un jury de l'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE, une thèse intitulée :

"Quantification des incertitudes liées aux simulations d'écoulement dans un réservoir pétrolier à l'aide de surfaces de réponse non linéaires".

en vue de l'obtention du titre de :

DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE

Spécialité : « **Géosciences** »

Fait à Vandoeuvre, le 21 décembre 2006
Le Président de l'I.N.P.L.,
L. SCHUFFENECKER



Résumé

Quantification des incertitudes liées aux simulations d'écoulement dans un réservoir pétrolier à l'aide de surfaces de réponse non linéaires

La quantification des incertitudes constitue une étape clé de la modélisation du sous-sol. Les données disponibles sont limitées, éparses et hétérogènes par conséquent le modèle obtenu reste une approximation de la réalité. Pour mener une étude de réservoir complète, il est nécessaire de disposer d'outils pour appréhender des incertitudes pour un coût minimal (temps, mémoire, etc.). Dans ce cadre, les simulations d'écoulement posent des problèmes spécifiques : elles sont coûteuse en temps et l'exploitation des résultats est difficile dû à la nature complexe de la relation entre les variables de production et les paramètres du modèle.

Le cœur des travaux présentés dans cette thèse concerne le développement de méthodologies basées sur l'utilisation de surfaces de réponse pour analyser cette relation :

- le développement et l'évaluation de l'interpolation lisse discrète et du krigeage dual pour la construction de surface de réponse. En complément, ont été développées une méthode d'analyse de sensibilité basée sur la variance de la production vis-à-vis des paramètres incertains ainsi qu'une technique d'inversion bayésienne de données historiques de production.
- la combinaison de résultats de simulations d'écoulement conventionnelles avec ceux de simulations rapides alternatives afin d'améliorer la prédictivité des surfaces de réponse.
- la prise en compte de paramètres incertains de type stochastique caractérisés par un effet aléatoire sur la production et ne pouvant être inclus dans une approche déterministe classique.

Le choix des techniques implantées s'est concentré sur les trois aspects suivants : (1) le nombre important de paramètres incertains qui doivent être pris en compte, (2) l'existence de bruit dans les données provenant le plus souvent de paramètres jugés peu influents et (3) la forme complexe de la relation entre les paramètres incertains et la production du réservoir.

Mots-clés: Surface de réponse, analyse de sensibilité, inversion bayésienne, simulations secondaires, paramètres stochastiques, simulation sur lignes de courant.

Abstract

Uncertainty Quantification on Reservoir Flow Simulation using Non-Linear Response Surfaces

Uncertainty Quantification is a key step in any subsurface modeling pipeline. Available data are limited, sparse and heterogeneous, therefore the geomodel itself is an approximation of the reality. This requires to provide tools to assess these uncertainties at a minimal cost (memory, time, etc.). In this context, flow simulation adds its own burden : it is time and memory consuming and the results are difficult to interpret due to the complex relationship between reservoir uncertain parameters and production variables.

This work develops several methodologies based on response surfaces to analyze this relationship :

- construction of response surfaces using the discrete smooth interpolation and the dual kriging approach ;
- analysis of non-linear response surfaces using variance based sensitivity analysis and bayesian inversion of production history ;
- integration of fast flow simulation results within the response surface construction step ;
- handling of stochastic uncertain parameters characterized by a random effect on reservoir production and which cannot be included in a classic deterministic framework.

The selection of these methodologies is based on the following considerations : (1) the number of uncertain parameters to handle is usually large, (2) data may be noisy, especially if some parameters has been estimated unimportant and neglected and (3) the complex relationship between uncertain parameters and reservoir production.

Keywords: Response surface, sensitivity analysis, bayesian inversion, fast alternative simulation, stochastic parameters, streamline simulation

Centre de Recherches Pétrographiques et Géochimiques

Laboratoire d'Infographie et d'Analyse de Données

Rue du Doyen Marcel Roubault - 54500 Vandœuvre