



HAL
open science

Non stationnarité dans les modèles de type booléen : application à la simulation d'unités sédimentaires

Marta Benito Garcia-Morales

► **To cite this version:**

Marta Benito Garcia-Morales. Non stationnarité dans les modèles de type booléen : application à la simulation d'unités sédimentaires. Mathématiques [math]. École Nationale Supérieure des Mines de Paris, 2003. Français. NNT : 2003ENMP1180 . pastel-00001187

HAL Id: pastel-00001187

<https://pastel.hal.science/pastel-00001187>

Submitted on 6 Aug 2010

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

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

THESE

pour obtenir le grade de

Docteur de l'Ecole des Mines de Paris

Spécialité « **Géostatistique** »

qui sera présentée et soutenue publiquement

par

Marta Benito García-Morales

en décembre 2003

**Non stationnarité dans les modèles de type booléen :
application à la simulation d'unités sédimentaires**

Directeur de thèse : Hélène Beucher

Jury

MM. Riccardo	BERSEZIO	Rapporteur externe
Hélène	BEUCHER	Examinateur
Pierre	BIVER	Examinateur
Emile	ELEWAUT	Expert invité
Lin Ying	HU	Rapporteur externe
Christian	RAVENNE	Examinateur
Jacques	RIVOIRARD	Examinateur

Remerciements

Ce travail a été réalisé grâce à l'encadrement de ma directrice de thèse, Hélène Beucher. Sans son soutien et sa disponibilité, ce projet n'aurait pas pu aboutir.

Mes remerciements iront aussi à tous les membres du jury, à Riccardo Bersezio et Lin-Ying Hu en tant que rapporteurs, ainsi qu'à Pierre Biver, Emile Elewaut, Christian Ravenne et Jacques Rivoirard.

Je voudrais remercier aussi à toutes les personnes du Centre de Géostatistique qui m'ont offert leur aide et leurs conseils, Christian Lantuéjoul, Didier Renard, François Geoffroy, Nicolas Bez, Chantal de Fouquet, Christian Lajaunie et Hans Wackernagel. Merci à Nathalie et Isabelle pour leur aide "administrative" (c'est n'est pas toujours facile de remplir correctement les formulaires !) et surtout leur gentillesse. Un grand merci à Françoise pour venir bavarder de temps en temps et pour sa patience et sa disponibilité à l'heure de corriger mon français et mon anglais.

Je ne peux pas terminer sans remercier tous ceux qui m'ont accompagné au cours de ces trois années : Thomas, Etienne, Bea, Stéphane, Brigitte, Jesús, Costin, Mireille, Christophe, Simon, Julien, David, Caroline, ... les pauses café n'auraient pas été la même chose sans votre bonne humeur. Un grand merci spécialement à Sarah après quatre ans pendant lesquels nous avons partagé appartement, bureau, des bons et des moins bons moments ... ça fait beaucoup !

Une pensée spéciale pour Eric, qui m'a soutenue dans les moments difficiles et qui a fait que les bons moments soient encore meilleurs. Je ne peux pas oublier sa famille, qui est devenue ma «famille française».

Pour finir, je remercie ma famille qui m'a soutenue pendant tout ce temps et qui a toujours été très proche malgré la distance.

Table des matières

Introduction	9
I Introduction à la théorie des ensembles aléatoires. Modèle booléen	11
1 Ensembles aléatoires	15
1.1 Ensembles fermés aléatoires (EFA)	16
1.2 Fonction de répartition d'un ensemble fermé aléatoire	17
1.2.1 Quelques propriétés des ensembles fermés aléatoires	18
1.2.2 Translation, dilatation, érosion	19
1.2.3 Mesures et transformations sur les ensembles fermés aléatoires	21
1.3 Ensembles fermés aléatoires stationnaires et isotropes	22
2 Processus de Poisson	25
2.1 Processus ponctuels	25
2.1.1 Ensemble fermé aléatoire versus mesure de comptage	26
2.1.2 Moments et mesures des moments	27
2.1.3 Stationnarité et isotropie	28
2.1.4 Ergodicité	28
2.2 Processus de Poisson	28
2.2.1 Processus de Poisson stationnaire	28
2.2.2 Processus de Poisson général : non stationnaire ou régionalisé	30
3 Modèle booléen	33
3.1 Définition du modèle booléen	34
3.1.1 Stationnarité, isotropie et ergodicité	34
3.1.2 Propriétés de stabilité	34
3.2 Capacité de Choquet du modèle booléen	35
3.2.1 Une propriété fondamentale du modèle booléen	36
3.2.2 Applications de la propriété fondamentale : cas stationnaire	37
3.3 Simulation d'un modèle booléen	38
II Inférence des paramètres du modèle	43
4 Description des variables expérimentales	47
4.1 Proportions	47
4.1.1 Proportion expérimentale d'un faciès	48

4.1.2	Probabilité ponctuelle associée au modèle booléen : capacité de Choquet ponctuelle .	49
4.2	Probabilité ponctuelle associée à l'objet	52
4.2.1	Deux cas élémentaires : parallélépipède et demi-ellipse	56
4.3	Conclusions	63
5	Inférence de l'intensité de Poisson	65
5.1	Méthode de déconvolution pour l'obtention de l'intensité	65
5.1.1	Filtre de déconvolution : filtre de Wiener	66
5.1.2	Interprétation du bruit - Paramètre $\Gamma(\nu)$	67
5.1.3	Étude de sensibilité	69
5.2	Aspects pratiques	83
5.2.1	Traitement des valeurs négatives de l'intensité	83
5.2.2	Calcul du nombre d'objets à simuler	88
5.2.3	Définition du domaine de calcul	91
5.3	Conclusions. Avantages et limitations de la méthode.	92
6	Information sur les objets	95
6.1	Cas stationnaire	95
6.1.1	Étude des variogrammes verticaux du modèle	96
6.1.2	Étude des variogrammes verticaux expérimentaux : indicatrices des simulations et indicatrices aux puits	99
6.1.3	Conclusions	102
6.2	Cas non stationnaire	104
6.2.1	Étude des variogrammes verticaux expérimentaux des proportions	105
6.2.2	Étude des variogrammes verticaux des indicatrices aux puits	107
6.2.3	Conclusions	109
III	Mise en œuvre	111
7	Application à un cas fictif 3D	113
7.1	Description des données	114
7.1.1	Proportions	114
7.1.2	Description des objets	120
7.2	Obtention de l'intensité. Simulations	122
7.3	Analyse des composantes connexes	137
7.4	Résultats et conclusions	138
8	Application à un cas réel : Permien de l'Utah	141
8.1	Cadre du travail	142
8.1.1	Contexte géologique	142
8.1.2	Transect régional	144
8.1.3	Description des données	145
8.1.4	Définition de l'unité de travail	148
8.1.5	Calcul de la grille de proportion	152
8.1.6	Définition des objets	154
8.2	Calcul de l'intensité	155
8.3	Simulations	158

8.4 Discussion et conclusions	165
Conclusions	173
9 Conclusions et perspectives	173
Annexes	181
A Calcul de la probabilité associée à une demi-ellipse	181
B Convolution et corrélation	187
B.1 Convolution	187
B.2 Corrélation	188
C Rappels de la théorie de Fourier	191
C.1 La transformée de Fourier (TF)	191
C.1.1 Propriétés de la TF	191
C.1.2 Théorème de convolution	192
C.1.3 Formule de Parseval	193
C.1.4 Échantillonnage	193
C.2 Transformée discrète de Fourier (DFT)	194
C.2.1 Transformée rapide de Fourier (FFT)	194
D Développement du calcul pour l'obtention du filtre de Wiener	195
E Description du programme	197
Bibliographie	201

Introduction

Introduction

Les enjeux économiques de l'industrie pétrolière imposent un besoin de décrire la structure géologique des réservoirs d'hydrocarbures et les propriétés des roches qui les constituent. Ce chapitre a pour but de présenter brièvement la modélisation géologique des réservoirs, notamment à travers les approches probabilistes dans lesquelles le sujet de cette thèse s'inscrit. Avant de décrire ce que l'on cherche avec la modélisation (ou caractérisation) des réservoirs, une définition des réservoirs est donnée, avec les facteurs principaux qui contribuent à leur formation. Puis, dans un second temps l'approche probabiliste est introduite dans ce contexte et les principales familles de modèles utilisés sont décrites. Une de ces familles, les modèles de type objet, nous intéresse tout particulièrement puisque le modèle étudié dans ce travail, le modèle booléen, en fait partie. Pour finir, une description du fil conducteur de cette thèse et des différentes parties qui la constituent est présentée.

Modélisation des réservoirs

Un réservoir est très généralement un dépôt sédimentaire ou une série de dépôts interconnectés qui contient des fluides (huile, gaz, eau...). Ces dépôts sont constitués de roches poreuses et perméables à l'intérieur desquelles les fluides circulent et peuvent éventuellement s'accumuler pour former un gisement s'il y a des barrières qui les retiennent comme le ferait un barrage. Ces barrières peuvent être de natures différentes : une faille, un piège anticlinal, une variation de perméabilité par changement de faciès, etc. (Sallé et al., 1976).

La perméabilité de la roche détermine sa capacité à laisser passer le fluide à travers elle. La porosité¹ est le pourcentage d'espace vide à l'intérieur de la roche et donne le volume de fluide que celle-ci peut contenir. Ce sont les deux propriétés pétrophysiques fondamentales que l'on cherche à déterminer pour décrire le réservoir et sa qualité. Ces propriétés ne sont pas uniformes dans tout le réservoir, mais dépendent des structures géologiques qui le constituent. La connaissance du réservoir passe par la détermination de telles hétérogénéités, ce qui implique l'étude des facteurs qui gèrent la nature et la porosité des roches (la pression, la diagenèse, la fracturation, l'environnement de dépôt ... (Sallé et al., 1976)).

L'objectif de la **caractérisation des réservoirs** est de décrire le plus précisément possible les caractéristiques pétrophysiques du milieu poreux ainsi que l'intensité et la direction des écoulements des fluides qui s'y produisent. Dans le cas de réservoirs d'hydrocarbures, cette description représente un enjeu économique important pour différentes raisons : elle doit permettre une bonne estimation des réserves exploitables, elle fournit des renseignements pour une meilleure localisation des puits d'exploitation, en résumé elle aide à la décision sur l'évolution du développement du gisement.

¹Il faut distinguer entre la **porosité effective** qui correspond au volume des vides interconnectés, à travers lesquels les fluides peuvent circuler, et la **porosité totale** qui comprend non seulement les pores interconnectés mais aussi ceux qui sont fermés par un ciment (Moureau et Brace, 1979).

Le processus de caractérisation de réservoir est indispensable tout au long de la vie d'un réservoir, de sa découverte à sa phase de déclin, et le ou les modèles construits évoluent au fur et à mesure que le gisement est exploité et que plus de données sont acquises. L'échelle de la modélisation dépend de la phase de la vie du réservoir et des données disponibles à chaque étape. Ainsi, pendant les premières phases, on cherche à déterminer la distribution spatiale des hétérogénéités et leur connexité tandis que pour choisir un scénario pour la phase de développement, il faut disposer d'un modèle de porosité et de perméabilité des réservoirs (Chessa, 1995). Ceci implique une description et une modélisation du réservoir à différentes échelles.

Dans le travail de Chessa (1995), on trouve une classification hiérarchique des hétérogénéités dans un réservoir, à partir de la classification de Weber (1986). Ces hétérogénéités ont, à chaque échelle, une influence sur la récupération des hydrocarbures. Par exemple, à l'échelle du champ (distances de l'ordre du kilomètre horizontalement, $\sim 1 - 10$ km, et de ~ 100 m en profondeur), des corps sableux connectés ou non sont identifiés. La connexité entre ces corps est, à cette échelle, le premier facteur dans la détermination de la quantité de fluide qu'il est possible d'extraire : avec un seul puits on peut extraire les hydrocarbures de plusieurs réservoirs s'ils sont connectés. À l'échelle du réservoir (~ 100 m horizontalement, ~ 10 m en profondeur), les frontières entre les unités génétiques peuvent marquer des forts contrastes de perméabilité, pouvant ainsi générer à l'intérieur du réservoir des régions inexploitable. Le modèle que nous étudions ici est utilisé principalement à l'échelle d'une **unité² génétique** ($\sim 1 - 10$ m d'épaisseur) et à l'**échelle granulométrique** ($\sim 10 - 100$ mm). À l'échelle de l'unité, nous sommes intéressés à modéliser une variété de structures sédimentaires pour obtenir un profil de perméabilité déterminant le mouvement de fluide à l'intérieur de l'unité. À l'échelle granulométrique, ce sont les différences de granulométrie que nous voulons estimer, puisqu'elles peuvent générer de fortes variations de perméabilité et créer des pièges pour le fluide à cette échelle. La détermination de ces variations est de grande importance pour la récupération des fluides résiduels dans la phase finale de la vie du réservoir.

Les données acquises pour décrire le gisement sont de nature très différente. Chacune d'entre elles apporte une information à ces différentes échelles en quantifiant, de façon plus ou moins directe, les propriétés pétrophysiques des réservoirs. Comme exemple nous pouvons citer l'étude des carottes, des diagraphies, les tests des puits, qui fournissent une information pétrophysique de haute résolution dans un voisinage très local autour des puits. À une échelle plus grande, les études sismiques permettent d'identifier des structures géologiques à l'échelle du gisement. L'étude d'affleurements dans des environnements de dépôt similaires à celui du réservoir étudié (analogues) a aussi une grande importance pour comprendre l'architecture du réservoir et pour définir, de façon cohérente, les règles géologiques dans l'élaboration du modèle de réservoir (Hu, 2002). La précision de ces données est aussi très différente. L'information aux puits est très précise mais, à cause des coûts très élevés des forages et des carottages, elle est peu abondante. De plus, elle peut être biaisée, puisque l'emplacement des puits est choisi de préférence dans les régions les plus productives. L'information sismique est, au contraire, plus abondante mais moins précise puisqu'elle ne fournit que des mesures indirectes des propriétés pétrophysiques des réservoirs. Il est donc important d'intégrer toutes les informations à disposition dans le modèle de réservoir. Puisque l'étude du réservoir est réalisée tout au long de sa vie, il est aussi nécessaire de pouvoir intégrer dans le modèle les nouvelles informations acquises au fur et à mesure de l'évolution de l'exploitation. Une mise à jour du modèle devrait être continuellement réalisée.

Cependant, et bien qu'il y ait, au cours du développement du réservoir, de nouvelles et plus nombreuses informations sur ses caractéristiques, la réalité n'est jamais parfaitement connue et il existe toujours une incertitude dans la connaissance du réservoir. Cette incertitude est particulièrement importante dans les phases

²Unité : terme désignant tout ensemble de terrains que l'on peut individualiser pour des raisons tectoniques et/ou stratigraphiques (Foucault et Raoult, 1984).

d'exploration où les données sont peu abondantes. Il est donc nécessaire de construire de multiples scénarios qui puissent représenter toutes les réalités possibles et tenir ainsi compte de l'incertitude (Hu, 2002).

L'**approche probabiliste** est adaptée à ce problème et présente plusieurs avantages : il existe un nombre important de modèles probabilistes, ce qui permet de choisir le modèle le mieux adapté à l'environnement de dépôt ; il est possible d'utiliser des modèles différents pour les différentes échelles d'étude, de l'échelle du gisement à l'échelle granulométrique et, enfin, la nature probabiliste de ces approches permet d'évaluer l'incertitude du modèle en construisant de multiples scénarios possibles.

En d'autres termes, bien qu'un réservoir soit quelque chose d'intrinsèquement déterministe, puisqu'il résulte de toute une série de processus physiques bien définis, il est possible de parler de «nature probabiliste» dans la modélisation à cause de l'information insuffisante qui introduit une importante incertitude dans la connaissance du réservoir. Haldorsen et MacDonald (1987) expliquent cette double nature déterministe - stochastique dans la modélisation des réservoirs.

Modèles stochastiques pour la caractérisation des réservoirs

Expérimentalement la modélisation des réservoirs consiste à construire des modèles numériques qui représentent les hétérogénéités à différentes échelles en termes de propriétés pétrophysiques sur les blocs d'une grille. On assigne à chaque bloc une valeur de la porosité et de la perméabilité. Nous pouvons distinguer trois étapes principales dans la modélisation d'un réservoir (Chessa, 1995) et chacune desquelles implique une échelle de travail différente :

- Modélisation de l'**architecture interne** du réservoir : dans cette étape la structure à l'échelle du gisement est modélisée en termes de lithofaciès³. Les volumes des roches constituant des formations géologiques plus ou moins homogènes lithologiquement sont identifiés, ainsi qu'éventuellement, la modélisation des failles et fractures dans le réservoir. Ces hétérogénéités à l'intérieur du gisement sont dépendantes des processus sédimentaires qui les ont constituées et peuvent donc être de nature très différente : dépôts de chenaux fluviaux ou marins, dépôts d'origine éolienne, dépôts en environnement deltaïque, etc. Chaque environnement présente des caractéristiques géométriques et des variabilités de porosité et de perméabilité particulières.
- Modélisation des **propriétés pétrophysiques** des roches à l'intérieur du réservoir : une fois la distribution des hétérogénéités modélisée, il faut déterminer les propriétés des roches et leur variabilité à l'intérieur de ces corps géologiques. Les principales variables d'intérêt sont, encore une fois, la perméabilité et la porosité, et aussi le degré de saturation en eau ou en hydrocarbure.
- Modélisation **hydrodynamique** : les réalisations ainsi obtenues servent de base pour la simulation des écoulements, qui seront utilisés dans la prédiction de la production du gisement. Dans cette phase il faut résoudre le problème de l'homogénéisation des propriétés pétrophysiques. Les modèles géologiques obtenus sont calculés sur des grilles assez fines pour mieux caractériser la variabilité de ces propriétés. Par souci d'économie de temps de calcul, les modèles hydrodynamiques sont utilisés sur des grilles plus larges et il faut donc uniformiser les valeurs des propriétés des roches obtenues sur un support plus large. Une description plus détaillée de ce problème est présentée dans les travaux de Grindeheim et Aasen (1992), Gelhar (1993), Journel (1996) et aussi dans Fisher et Knipe (2001) et

³**Facies** : catégorie dans laquelle on peut ranger une roche ou un terrain, et qui est déterminée par un ou plusieurs caractères lithologiques (**lithofaciès**) ou paléontologiques (biofaciès). Le terme faciès est aussi employé pour désigner une catégorie correspondant à un milieu ou à un domaine de sédimentation (Foucault et Raoult, 1988). Les **lithotypes** sont définis par association de faciès correspondant à un même type lithologique et des caractéristiques hydrodynamiques similaires. Dans ce travail un lithotype correspond, généralement, à un seul lithofaciès. Les termes faciès, lithofaciès et lithotype seront utilisés indistinctement.

Castaing et al. (2002) pour le cas de milieux fracturés.

Il existe un grand nombre de modèles probabilistes utilisés pour la caractérisation des réservoirs. Le choix d'un modèle est conditionné par l'environnement du dépôt, par l'échelle de travail et par l'étape de la modélisation dans laquelle on se trouve.

Le **modèle booléen**, qui fait l'objet de ce travail, est utilisé pour construire un modèle lithologique de l'architecture interne des réservoirs, en particulier à l'échelle des unités génétiques. Il est aussi utilisé pour la modélisation des hétérogénéités à l'échelle granulométrique à l'intérieur du réservoir. Il fait partie de la famille d'approches dites d'objet qui constituent, avec les **approches de pixel**, les modèles le plus utilisés dans ce contexte.

Modèles de pixel : Ces modèles sont basés sur les corrélations spatiales des lithofaciès (sables des chenaux, sables éoliens, argiles, etc.) par l'intermédiaire du variogramme de leurs indicatrices. Parmi les modèles appartenant à cette famille, nous pouvons citer comme les plus utilisés la simulation gaussienne tronquée (Matheron et al., 1987) et la simulation séquentielle d'indicatrices (Journel et Alabert, 1990).

La première consiste à utiliser une fonction gaussienne continue pour déterminer l'appartenance d'un pixel à un faciès. En tout point du domaine, la valeur du faciès simulé est fonction de la valeur de la fonction gaussienne par rapport à certains seuils de troncature. Ceux-ci sont déterminés à partir des proportions des lithofaciès. Des variations de cette méthode sont son application à un cas non stationnaire (Beucher et al., 1993 ; Johann et al., 1997) et la simulation pluri-gaussienne tronquée (Galli et al., 1994) où plusieurs fonctions gaussiennes interviennent pour la détermination des faciès, ce qui permet une grande variété dans l'agencement des faciès.

Le principal avantage de cette méthode est sa capacité à reproduire des réservoirs décrits par de nombreux faciès et en présence d'un grand nombre de puits.

La méthode de simulation d'indicatrices consiste à affecter à chaque pixel une catégorie ou classe de faciès à partir des distributions locales de probabilité. Cette affectation tient compte de la classe simulée dans les pixels voisins et peut aussi prendre en compte des variations locales de proportions des faciès (Journel et al., 1998). Cette méthode a aussi été utilisée pour reproduire des variations des directions de continuité (Deutsch, 1999).

Les modèles de pixel sont contraints assez facilement aux données quantitatives, mais ils ne restituent que statistiquement les images des réservoirs. Souvent ces images ne sont pas satisfaisantes dans le sens où elles ne reproduisent pas la forme des objets géologiques présents dans le réservoir.

Modèles d'objets : Ces approches modélisent la géométrie des corps géologiques lithologiquement homogènes par des objets idéalisés distribués en 3D. Ainsi, ils sont spécialement appropriés pour simuler les dépôts sableux d'origine fluviale, les réseaux de fractures, les milieux poreux à l'échelle granulométrique, etc. Chaque objet est défini par sa position, fixée dans un point aléatoire de l'espace, et ses caractéristiques géométriques (leur forme, leur taille, leur orientation, etc.). L'utilisation des modèles d'objet dans la caractérisation de l'architecture des réservoirs en environnement fluviale est très répandue. Nous pouvons citer quelques exemples : Haldorsen et MacDonald (1987), Chessa et Martinus (1992), Henriquez, Tyler et Hurst (1990), Schmitt et Beucher (1997), Viseur (1999), etc. Des applications à la modélisation des fractures dans les réservoirs ont été faites par Munthe et al.

(1993) et Cacas et al. (2001), par exemple.

Parfois il y a dans la littérature une certaine confusion entre les termes «modèle d'objet» et «modèle booléen». Par modèles d'objets on désigne toute une famille de modèles de différentes constructions, qui ont pour points communs l'utilisation de figures géométriques plus ou moins complexes pour modéliser les hétérogénéités dans un réservoir, (e.g. chenaux, crevasses, barrières, etc.) et l'implantation plus ou moins aléatoire de ces objets dans l'espace. Le modèle booléen appartient à cette famille, mais tous les modèles d'objets ne sont pas des modèles booléens. Celui-ci implique des contraintes assez fortes : l'indépendance entre les points d'implantation des objets, générés par un processus de Poisson, et l'indépendance entre les objets. D'autres modèles créent les objets selon des règles de génération de façon à ce qu'ils vérifient le conditionnement aux puits (Shmaryan et Deutsch, 1999 ; Viseur et al., 1998).

Le principal avantage de ce type d'approches est qu'elles créent des images assez intuitives de la géométrie des réservoirs. Les principaux inconvénients qu'ils présentent sont, d'une part, le conditionnement aux données des puits (indicatrices aux puits) qui devient difficile et lourd du point de vue informatique quand le nombre de puits commence à être important ; d'autre part, l'observation des proportions verticale et latérale des lithofaciès est aussi une contrainte critique de la modélisation souvent difficile à respecter.

D'autres méthodes utilisées pour la modélisation des lithofaciès sont les **modèles génétiques** («process-based»), en plein développement à présent. Contrairement à d'autres modèles probabilistes qui ne modélisent que le résultat des processus sédimentaires, la modélisation aléatoire génétique tient compte de ces processus de dépôt et donc de la nature déterministe du réservoir. Puisque les paramètres qui contrôlent le modèle de sédimentation sont, en général, très incertains, ils sont représentés par des variables ou des fonctions aléatoires (Hu, 2002), ce qui rejoint la nature probabiliste de la modélisation. Depuis les travaux initiaux de Matheron (1969a, 1969b) et de Jacod et Joathon (1971) proposant divers modèles génétiques pour différents types de processus sédimentaires, de nombreuses études ont été développées dans cette direction. Nous pouvons citer, entre autres Hu, Joseph et Dubrule (1992) pour la modélisation des dépôts deltaïques, Lopez, Galli et Cojan (2001) et Lopez (2003) pour la modélisation des chenaux méandriques. Leur principal inconvénient est le conditionnement aux données des puits, qui est très difficile.

Puisque la caractérisation des réservoirs implique la modélisation de structures et de propriétés géologiques à différentes échelles, un seul modèle ne suffira pas pour satisfaire toutes les caractéristiques. Dans la pratique, la combinaison de deux ou plusieurs modèles appartenant à des familles différentes est le plus souvent inévitable pour construire des modèles complexes de réservoir (e.g. Damsleth et Holden (1994), Damsleth et al. (1992)).

De plus amples synthèses sur les différentes approches utilisées dans la caractérisation de réservoirs se trouvent dans les travaux de Dubrule (1988), Haldorsen et Damsleth (1990) et Damsleth et Holden (1994), Galli et Beucher (1997), Chilès et Delfiner (1999).

Cadre de la thèse

Un des intérêts principaux de la modélisation de l'architecture interne des réservoirs est qu'elle permet d'incorporer toute l'information géologique disponible. Les proportions des lithofaciès constituent un outil très puissant dans ce contexte, puisqu'elles quantifient la présence des lithofaciès tout en indiquant leur distribution spatiale. Ainsi, les proportions sont devenues une contrainte importante à respecter dans la

caractérisation des réservoirs. Dans les approches de type pixel par exemple, notamment dans les modèles des gaussiennes tronquées, l'information géologique contenue dans les proportions a été introduite avec succès.

Par contre, un des problèmes que présentent les approches de type objet consiste précisément en la difficulté à incorporer l'information géologique concernant la distribution des lithofaciès et à respecter les données de proportion.

Il existe différentes approches pour construire des modèles ou des images qui respectent les données de proportion. Certaines se basent sur l'incorporation itérative des objets jusqu'à atteindre les proportions voulues à travers une fonction objectif (Deutsch et Tran, 2002). D'autres utilisent les proportions comme des probabilités de présence des lithofaciès pour favoriser l'implantation d'un objet dans une région plutôt que dans une autre, puis, après l'incorporation d'un nouvel objet, il est vérifié si la simulation obtenue récupère bien les données de proportion (Viseur, 1999).

L'approche que nous avons voulu développer dans ce travail consiste à incorporer l'information des proportions directement dans le modèle. De cette façon la distribution des objets et le nombre à simuler sont déterminés naturellement sans qu'on ait besoin de recourir à un processus itératif de vérification. Le modèle booléen, en tant que modèle bien connu et maîtrisé, semble être le choix approprié.

Les premiers travaux dans cette direction sont dus à Schmitt et Beucher (1997) qui ont établi la relation entre les proportions et le modèle dans une situation simple pour une variation des proportions uniquement suivant la verticale. Une suite à ces travaux était nécessaire, notamment dans le développement de la théorie, l'extension de la méthode à 3D, l'adaptation des données à l'information requise par le modèle, la valorisation des résultats et l'application à une situation réelle. C'est dans ce cadre que cette thèse s'inscrit.

Un double enjeu ...

L'objectif de ce projet est d'élargir le champ d'application du modèle booléen à la modélisation des réservoirs dont la distribution des lithofaciès est non stationnaire, c'est-à-dire ils sont répartis de façon non homogène. Pour cela, l'idée a été de proposer une méthode pour estimer les paramètres qui définissent le modèle booléen à partir des proportions de ces lithofaciès. Cette méthode doit permettre de prendre en compte la non stationnarité dans les simulations des réservoirs mais aussi de proposer aux professionnels de ce domaine un outil de caractérisation fiable. Ainsi, le caractère de ce travail se situe à l'interface entre le développement scientifique et mathématique et l'application industrielle. L'enjeu est donc double.

L'**enjeu scientifique** est d'intégrer l'information des proportions en termes de paramètres du modèle booléen. Ceci relève du problème de l'*inférence*. Une interprétation correcte des paramètres dans le contexte de notre application et réciproquement, une interprétation des données expérimentales en fonction du modèle, sont nécessaires pour la cohérence théorique du processus.

L'**enjeu industriel** consiste à construire une interface entre l'information d'entrée, les proportions des faciès, et les résultats attendus, les simulations. Il existe déjà un grand nombre d'algorithmes de simulation, conditionnelle et non conditionnelle, du modèle booléen qui ont été largement étudiés. Avec ce travail nous voulons ajouter une phase intermédiaire préalable à ces algorithmes permettant de traiter la présence d'une non stationnarité.

Ce projet a été inscrit à l'origine dans le cadre d'une convention existant entre l'Institut Français du Pétrole et le Centre de Géostatistique, destinée à développer la recherche dans le domaine de la caractérisation des réservoirs.

Le fil conducteur de ce travail

Cette thèse est divisée en trois grandes parties. La première partie (chapitres 1, 2 et 3) est consacrée à la théorie du modèle booléen, ce qui est nécessaire pour comprendre la construction du modèle et ses propriétés. Dans un premier temps, la famille générique des ensembles aléatoires, à laquelle le modèle booléen appartient, est introduite (chapitre 1). Puis, le processus de Poisson qui sert de base à la construction du modèle est présenté (chapitre 2) et, enfin, le modèle booléen et ses propriétés sont présentés dans le chapitre 3. L'algorithme de simulation (conditionnelle et non conditionnelle) utilisé dans ce travail y sera décrit.

La deuxième grande partie (chapitres 4, 5 et 6) concerne l'estimation des paramètres du modèle booléen à partir des informations expérimentales : les proportions des lithofaciès. Le chapitre 4 décrit les données expérimentales qui constituent notre point de départ et son rapport avec les paramètres du modèle. L'établissement de ce rapport est la base de la méthode que nous proposons pour estimer le paramètre du modèle qui gouverne la distribution des objets dans le volume. Cette méthode, présentée dans le chapitre 5, constitue le corps de la thèse. L'étude de l'autre paramètre du modèle booléen concernant la géométrie des objets à simuler se trouve dans chapitre 6, dans lequel nous étudions la possibilité d'obtenir une information sur ces objets à partir des proportions.

Dans la troisième partie (chapitres 7 et 8), deux applications pratiques servant à valider la méthode proposée sont décrites. La première (chapitre 7) concerne un cas fictif dont le faciès modélisé par le modèle booléen présente une distribution non stationnaire à 3D. Cet exemple pourrait être assimilé à la modélisation d'hétérogénéités à l'échelle granulométrique. Malheureusement nous n'avons pas disposé de données réelles pour tester la méthode dans ce cadre. Le chapitre 8 décrit un exemple, cette fois-ci réel, de modélisation d'un réservoir fluviatile à partir des données d'affleurement. La distribution du faciès modélisé, correspondant aux chenaux, est ici considérée comme non stationnaire dans un plan vertical.

Enfin, le chapitre 9 établit un bilan du travail réalisé et présente les conclusions et une discussion des résultats. Quelques perspectives pour une éventuelle suite à ce projet sont également présentées.

Première partie

Introduction à la théorie des ensembles aléatoires. Modèle booléen

L'étude quantitative de nombreux phénomènes naturels est conditionnée par leur variabilité dans l'espace et dans le temps. Ces variables sont étudiées à partir de mesures expérimentales parfois rares et bruitées, obtenues par un échantillonnage souvent préférenciel et peu exhaustif pour des raisons pratiques diverses : économiques, d'accessibilité au phénomène ou même à cause des limitations dans les instruments de mesure.

Un tel échantillonnage permet, selon les cas, d'avoir une connaissance partielle des objets étudiés (par exemple, les gisements de minerais connus par sondages) ou d'obtenir une information sur une population à partir d'un groupe d'objets (l'étude des populations de plantes, les villes dans une région, les cellules biologiques, ...). L'extrapolation de cette information partielle à tout le phénomène est affectée d'une grande incertitude, puisqu'un grand nombre d'interprétations, toutes cohérentes avec les observations, sont possibles. Une des approches utilisées pour répondre à cette question est l'*approche probabiliste*.

Les modèles probabilistes essaient de tenir compte de cette incertitude en donnant l'opportunité de créer plusieurs réalisations, c'est à dire différentes «réalités possibles». Chacune de ces réalisations est une *simulation*. Elles sont qualifiées de *conditionnelles* si elles respectent les données expérimentales. L'utilisation de ce type de modèles permet ainsi de résoudre certains problèmes liés à la prédiction et à l'estimation : l'interpolation des données, le filtrage, l'optimisation, le changement de support et d'échelle, l'intégration de plusieurs paramètres, la simulation, etc.

Les *ensembles aléatoires* constituent une classe particulière des modèles probabilistes. La notion d'ensemble aléatoire est introduite par la nécessité de modéliser certains phénomènes naturels à nature géométrique qui présentent une hétérogénéité plus ou moins importante à différentes échelles et, en même temps, un certain caractère structuré. Des exemples de tels phénomènes sont les milieux poreux, les structures granulaires dans les matériaux, les cellules biologiques ou certaines structures géologiques, etc. Ces modèles cherchent alors à représenter les propriétés de la structure, de la morphologie ou les propriétés physiques de ces milieux hétérogènes.

La classe des ensembles aléatoires est d'une grande richesse : les processus ponctuels, les réseaux des droites, les partitions polyédriques de l'espace, les agrégats, Mais les principaux modèles utilisés sont les processus de points, le modèle booléen et les partitions polyédriques de l'espace. De différents phénomènes physiques sont associés à des modèles variés. Ainsi, par exemple, des inclusions non métalliques dans un acier sont modélisées par des processus stochastiques ponctuels, les structures granulaires par des partitions aléatoires de l'espace et les milieux poreux par le modèle booléen, etc.

Du fait du grand nombre de propriétés que la notion d'ensemble aléatoire met en jeu (la géométrie des réalisations, la topologie, la variabilité dans la forme et taille des objets...) la caractérisation de la loi d'un tel ensemble n'est pas immédiate. La formalisation utilisée dans ce travail pour définir un ensemble aléatoire a été établie par Matheron (1969, 1972). Cette approche est basée sur une idée principale : à partir d'un certain ensemble connu, S , on analyse si l'événement " S rencontre l'ensemble aléatoire étudié" est vérifié ou non.

Dans notre étude nous nous sommes intéressés à un modèle concret d'ensemble aléatoire, le modèle booléen pour la caractérisation des réservoirs pétroliers. Plus précisément, le modèle booléen sera utilisé pour modéliser le faciès sableux ou gréseux correspondant aux chenaux à l'intérieur d'un réservoir.

Les trois chapitres qui composent cette première partie essaient de présenter de façon résumée les bases théoriques qui permettent de décrire le modèle booléen. Celui-ci est construit à partir d'un autre modèle d'ensemble aléatoire plus simple, le processus de Poisson. Premièrement, les caractéristiques générales des ensembles aléatoires sont présentées dans le premier chapitre. Le principal outil d'analyse, la capacité de

Choquet, sera ici présenté ainsi que quelques propriétés fondamentales qui seront utilisées tout au long de ce travail. Puis, dans le second chapitre le processus de Poisson est introduit et, pour finir, le troisième chapitre sera consacré au modèle booléen.

Un développement approfondi de la théorie des ensembles aléatoires peut se trouver les ouvrages de Matheron (1969, 1972), Serra (1982), Lantuéjoul (1993, 2002), et Stoyan et al. (1995).

Chapitre 1

Ensembles aléatoires

Les ensembles aléatoires constituent un modèle mathématique général utilisé depuis longtemps pour reproduire des phénomènes naturels présentant des formes géométriques plus ou moins irrégulières. Ces phénomènes peuvent présenter des structures très complexes qu'il est nécessaire de modéliser à partir d'une information expérimentale souvent limitée.

Un ensemble est défini mathématiquement à travers la totalité des relations existant entre ses parties constitutives. Expérimentalement il faudrait tester, pour déterminer complètement cet ensemble, toutes les relations possibles et examiner si elles sont vérifiées ou non. Ceci est généralement impossible. Dans la pratique, nous ne disposons que de quelques paramètres mesurables et nous ne pouvons tester qu'un nombre limité de relations pour définir le modèle. Ceci a été le problème abordé par Matheron quand, en 1967, il donne une première interprétation des milieux poreux en termes d'ensembles aléatoires et "résume les propriétés essentielles de leurs structures sous la forme d'un petit nombre de paramètres statistiquement représentatifs" (Matheron, 1972). Pour caractériser ces ensembles aléatoires, il s'est servi d'un outil structural qui joue le même rôle que la fonction de répartition¹ pour les variables aléatoires : la *capacité de Choquet* ou *fonction de répartition d'un ensemble aléatoire*.

Pour définir cette application, dénommée par T , deux éléments sont nécessaires : un ensemble «test» et une opération élémentaire entre ensembles. Nous allons illustrer ceci à l'aide d'un exemple. Prenons, dans notre application particulière, le cas des chenaux sableux dans un milieu non réservoir. Nous voulons modéliser l'ensemble de tous les chenaux par l'ensemble aléatoire X , et la matrice argileuse dans laquelle ils sont contenus par l'ensemble complémentaire de X , X^c . Pour caractériser cet ensemble aléatoire, nous nous servons d'une famille d'ensembles «tests» S , encore appelées figures structurantes, et analysons leurs relations d'interaction avec l'ensemble aléatoire. En d'autres termes, nous allons déplacer cet ensemble S à l'intérieur du réservoir et nous allons regarder s'il rencontre les chenaux. Les événements les plus simples et immédiats à considérer sont l'intersection non vide ($S \cap X \neq \emptyset$: S rencontre le faciès chenal) où l'inclusion ($S \subset X$: S est contenu dans le faciès chenal). Ces deux événements élémentaires génèrent toute une famille de relations, appelée mathématiquement σ -algèbre², quand on y ajoute les opérations logiques «et», «ou» et

¹Une variable aléatoire Z est définie par sa fonction de répartition F , i.e., par la probabilité pour que le résultat d'un tirage au sort de Z soit inférieure à une certaine valeur z : $F(z) = P\{Z < z\}$

²Une σ -algèbre $\sigma_{\mathcal{A}}$ est une famille d'ensembles qui est fermée sous les opérations élémentaires d'union, d'intersection, de complémentation, et d'union et d'intersection infinies mesurables. Techniquement $\sigma_{\mathcal{A}}$ est une famille d'ensembles $A \in \sigma_{\mathcal{A}}$ satisfaisant :

- $\Omega \in \sigma_{\mathcal{A}}$; $\emptyset \in \sigma_{\mathcal{A}}$;
- $\forall A \in \sigma_{\mathcal{A}}, A^c \in \sigma_{\mathcal{A}}$ et
- pour toute famille finie ou dénombrable $A_i (i \in I)$: $A_i \in \sigma_{\mathcal{A}}$ implique que $\bigcup_{i \in I} A_i \in \sigma_{\mathcal{A}}$ et $\bigcap_{i \in I} A_i \in \sigma_{\mathcal{A}}$

Un exemple est la σ -algèbre borélienne, la plus petite σ -algèbre qui contient les ensembles ouverts d'un espace topologique, comme \mathbb{R}^d . Elle inclut les fermés et les compacts.

«non». Ainsi, la structure de l'ensemble aléatoire X sera connue, du point de vue probabiliste, si on connaît la probabilité pour que chacune de ces relations soit vérifiée. On dit alors que l'ensemble aléatoire est défini par une probabilité sur la σ -algèbre des relations.

Comme nous le verrons par la suite, la famille d'ensembles «test» retenue sera celle des ensembles compacts K^3 de \mathbb{R}^d , et l'opération élémentaire choisie est l'intersection. De cette façon, il est possible de définir un ensemble aléatoire X par l'application T , la *fonction de répartition de l'ensemble aléatoire*, qui donne la probabilité pour que l'ensemble aléatoire X rencontre le compact K : $T_X(K) = P\{K \cap X \neq \emptyset\}$. La logique de l'opération d'intersection conduit naturellement à la notion d'ensemble aléatoire *fermé* : étant donné que $K \cap X \neq \emptyset$ si et seulement si K rencontre l'adhérence de X : $K \cap \bar{X} \neq \emptyset^4$, nous pouvons limiter l'étude à celui des ensembles aléatoires fermés.

Les ouvrages utilisés pour la rédaction ce chapitre sont ceux de Matheron (1969, 1972), Serra (1982), Ayala (1988), Lantuéjoul (1993), Stoyan et al. (1995), Jeulin (1999) et Chilès et Delfiner (1999).

1.1 Ensembles fermés aléatoires (EFA)

Comme il a été vu dans le paragraphe précédent, les ensembles aléatoires peuvent être caractérisés à partir d'une famille d'ensembles «tests» en analysant leurs intersections avec les ensembles aléatoires. Pour un ensemble aléatoire arbitraire X , nous pouvons regarder si l'intersection $S \cap X$, pour un ensemble test S , est vide ou non. Le choix de considérer si S rencontre l'ensemble aléatoire X , est à la base des théories développées par Matheron (Matheron, 1969, 1972) et Kendall (Kendall, 1974). La logique de l'intersection, engendrée par les événements du type $S \cap X \neq \emptyset$, nous amène à considérer la famille \mathcal{F} des sous-ensembles fermés de \mathbb{R}^d . Par ailleurs, cette famille va être munie d'une famille de relations entre ensembles, $\sigma_{\mathcal{F}}^5$ engendrée par les sous-ensembles de \mathcal{F} tels que leur intersection avec un compact K de \mathbb{R}^d est non vide :

$$\mathcal{F}_K = \{F \in \mathcal{F} | K \cap F \neq \emptyset\}$$

Le choix de la famille des compacts comme éléments structurants est introduit par le besoin de rendre mesurables les opérations entre ensembles : union, intersection, érosion, dilatation, etc. L'ensemble «test» générique S est donc restreint à un ensemble compact, K , et les ensembles aléatoires génériques aux ensembles aléatoires fermés.

Définition 1 *Un ensemble fermé aléatoire X est une variable aléatoire en $[\mathcal{F}, \sigma_{\mathcal{F}}]$. En d'autres termes, si $(\Omega, \sigma_{\mathcal{A}}, P)$ est un espace probabilisé⁶, un ensemble aléatoire fermé est une application mesurable de $(\Omega, \sigma_{\mathcal{A}})$ dans $[\mathcal{F}, \sigma_{\mathcal{F}}]$.*

³Un ensemble compact est un ensemble fermé et borné. Il est mesurable.

⁴Un ensemble est fermé s'il contient sa frontière : $\partial X \subset X$

⁵ $\sigma_{\mathcal{F}}$ est la σ -algèbre borélienne associée à la topologie engendrée par les ouverts de \mathcal{F} de la forme (Matheron, 1969) :

$$\mathcal{V}^K = \{F \in \mathcal{F} | F \cap K = \emptyset\}$$

$$\mathcal{V}_G = \{F \in \mathcal{F} | F \cup G \neq \emptyset\}$$

pour tout ouvert G et tout compact K .

⁶Un espace probabilisé $(\Omega, \sigma_{\mathcal{A}}, P)$ est constitué de trois éléments : un espace d'événements Ω , une σ -algèbre $\sigma_{\mathcal{A}}$ et une mesure positive ou probabilité P .

Ainsi, X génère une distribution P en $[\mathcal{F}, \sigma_{\mathcal{F}}]$. Définie de cette façon, la loi du fermé aléatoire induite par P n'est pas facile à manipuler. Cependant, il est possible de se servir d'un outil similaire à la fonction de répartition de la variable aléatoire : la *fonction de répartition d'un ensemble fermé aléatoire*, introduite par la suite (Matheron 1969, 1972).

1.2 Fonction de répartition d'un ensemble fermé aléatoire

La caractérisation de la distribution P d'un ensemble fermé aléatoire X peut se faire à travers une application T qui associe à tout compact K la probabilité pour qu'il rencontre l'ensemble X :

$$T_X(K) = P\{X \cap K \neq \emptyset\}$$

Cette application vérifie les propriétés suivantes :

1. T_X est **borné** : aucun fermé ne rencontre l'ensemble vide :

$$T_X(\emptyset) = 0$$

2. $0 \leq T_X \leq 1$ puisqu'il s'agit d'une probabilité ;
3. T_X est **croissante** : si K et K' sont deux sous-ensembles compacts de \mathbb{R}^d et $K \subseteq K'$, alors :

$$T_X(K) \leq T_X(K')$$

4. T_X est **semi continue supérieurement** sur la famille des sous-ensembles compacts de \mathbb{R}^d . C'est à dire si $(K_n, n = 1, 2, \dots)$ est une suite décroissante de compacts d'intersection K :

$$K_{n+1} \subset K_n, \quad \Rightarrow \quad \bigcap_{n=1}^{\infty} K_n = K$$

alors $T_X(K_n)$ converge vers $T_X(K)$:

$$\lim_{n \rightarrow \infty} T_X(K_n) = T_X(K)$$

5. Si K, K_1, K_2, \dots sont des compacts, nous pouvons définir les fonctions S_1, S_2, \dots comme :

$$S_1(K; K_1) = T_X(K \cup K_1) - T_X(K)$$

$$S_n(K; K_1, \dots, K_n) = S_{n-1}(K; K_1, \dots, K_{n-1}) - S_{n-1}(K \cup K_n; K_1, \dots, K_{n-1})$$

La fonction $S_n(K; K_1, \dots, K_n)$ représente la probabilité pour que l'ensemble aléatoire X rencontre les compacts K_1, K_2, \dots, K_n et soit disjoint du compact K :

$$P(X \cap K = \emptyset, X \cap K_i \neq \emptyset \quad \forall i = 1, \dots, n)$$

Cela implique que les fonctions S_n doivent être strictement positives ou nulles.

Les propriétés 3, 4 et 5 ci-dessus définissent T_X comme une **capacité de Choquet alternée d'ordre infini**, ou tout simplement, une capacité de Choquet. Dans beaucoup de cas on utilisera aussi la fonctionnelle Q_X définie à partir de T_X :

$$Q_X(K) = 1 - T_X(K)$$

Théorème 1 (G.Choquet) Soit T_X une fonction définie sur la famille des sous-ensembles compacts de \mathbb{R}^d . Il existe une distribution P sur $\sigma_{\mathcal{F}}$, nécessairement unique, vérifiant :

$$P\{X \in \mathcal{F} | X \cap K \neq \emptyset\} = T_X(K)$$

si et seulement si T_X est une capacité de Choquet alternée d'ordre infini et telle que $T_X(\emptyset) = 0$ et $0 \leq T_X(K) \leq 1$, pour K un sous-ensemble compact.

Le théorème de Choquet montre que la fonctionnelle T_X a les mêmes propriétés qu'une fonction de répartition. Elle joue le même rôle pour les ensembles aléatoires fermés que la fonction de distribution pour les variables aléatoires. Quelques propriétés des ensembles fermés aléatoires, comme celles présentées par la suite, peuvent être exprimées au moyen de cette capacité de Choquet.

1.2.1 Quelques propriétés des ensembles fermés aléatoires

1.2.1.1 Indépendance

Deux ensembles fermés aléatoires X_1, X_2 sont *indépendants* si pour deux ensembles compacts quelconques K_1, K_2 il se vérifie :

$$P\{X_1 \cap K_1 \neq \emptyset, X_2 \cap K_2 \neq \emptyset\} = T_{X_1}(K_1)T_{X_2}(K_2)$$

1.2.1.2 Stationnarité et isotropie

Un ensemble aléatoire fermé X est *stationnaire* si sa capacité de Choquet est invariante par translation :

$$T_X(K) = T_X(K_h)$$

pour tout compact K et toute translation $h \in \mathbb{R}^d$.

De la même façon, l'ensemble aléatoire fermé X est *isotrope* si, pour toute rotation r autour de l'origine :

$$T_X(K) = T_X(rK)$$

1.2.1.3 Ergodicité

En général, une fonction aléatoire est *ergodique* si, en moyenne, la moyenne spatiale de la fonction sur un domaine $V \subset \mathbb{R}^d$ converge vers son espérance quand $V \rightarrow \infty$. Ceci est aussi applicable aux ensembles aléatoires. Cette propriété est très importante dans les applications puisqu'elle permet la détermination de la moyenne à partir d'une seule réalisation du processus aléatoire. Cela implique que les espérances statistiques peuvent être exprimées comme des limites des moyennes arithmétiques ou spatiales. Le modèle booléen avec des grains convexes et compacts est un exemple d'ensemble aléatoire ergodique.

1.2.1.4 Infinie divisibilité

Un ensemble aléatoire fermé X est *indéfiniment divisible par rapport à l'union* si, pour tout entier n , X est équivalent à la réunion $X_1 \cup X_2 \cdots \cup X_n$ de n ensembles aléatoires fermés indépendants et équivalents⁷ X_1, \dots, X_n . Soit la probabilité pour que X n'intersecte pas l'ensemble compact K , définie comme :

$$Q_X(K) = 1 - T_X(K) = P\{X \cap K = \emptyset\}$$

⁷Deux ensembles sont équivalents s'il existe entre eux une relation biunivoque, élément à élément (Théorie des groupes, G. Cantor)

X est indéfiniment divisible si, et seulement si, les conditions du théorème de Choquet sont satisfaites par les fonctionnelles :

$$T_{X_n} = 1 - Q_X^{1/n}$$

En définitive, un ensemble aléatoire est indéfiniment divisible s'il peut être décomposé en l'union d'un certain nombre d'ensembles fermés aléatoires indépendantes et équivalents de la même famille.

S'il n'existe pas de points fixes, c.a.d. des points $x \in \mathbb{R}^d$ tels que $P\{x \in X\} = 1$, alors :

Théorème 2 (Matheron, 1972) Une fonction T sur l'ensemble des compacts est associée à un ensemble fermé aléatoire indéfiniment divisible sans points fixes si, et seulement si, il existe une capacité de Choquet d'ordre infini Ψ telle que, pour K compact :

1. $\Psi(\emptyset) = 0$ et
2. $Q(K) = 1 - T(K) = e^{-\Psi(K)}$

La démonstration du théorème est dans Matheron (1972).

Ce théorème est à la base de la définition du modèle booléen. En effet, nous verrons dans le chapitre 3 que la capacité de Choquet du modèle booléen s'exprime en termes de l'exponentielle d'une fonction sur l'ensemble des compacts.

1.2.1.5 Stabilité

Un ensemble fermé aléatoire est *stable par rapport à l'union* si, pour tout entier n , il existe une constante positive λ_n telle que la réunion $X_1 \cup X_2 \cdots \cup X_n$ des n ensembles fermés aléatoires équivalents et indépendants est équivalente à l'ensemble aléatoire fermé $\lambda_n X$ ⁸. Un ensemble fermé aléatoire stable est indéfiniment divisible, mais l'inverse n'est pas forcément vraie.

La stabilité est le concept mathématique pour décrire la similitude entre ensembles. Cette propriété est plus forte que la précédente, puisqu'elle implique que l'ensemble résultante de l'union est similaire à ceux qui l'ont construit, c'est à dire ils sont identiques sauf un facteur d'échelle. Les processus de Poisson sont stables par rapport à l'union.

1.2.2 Translation, dilatation, érosion

Etant donné que nous travaillons dans l'espace euclidien \mathbb{R}^d nous pouvons définir quelques transformations d'ensembles, pour $K \subset \mathbb{R}^d$. Elles seront fondamentales le long de ce travail.

1.2.2.1 Translation

La *translation* K_x du compact K est définie comme :

$$K_x = \{x + y : y \in K\}$$

Nous pouvons aussi définir l'ensemble *symétrique* de K par rapport à l'origine :

$$\check{K} = \{-y : y \in K\}$$

⁸ $\lambda_n X$ représente l'homothétique de X par λ_n . Une homothétie est une transformation géométrique qui agrandit ou réduit une figure tout en conservant sa forme initiale. Mathématiquement on décrit l'homothétie de la façon suivante : soit O un point de l'espace et λ une constante positive. L'homothétie de centre O et de rapport λ est la transformation qui à tout point M associe le point M' défini par la relation : $\overrightarrow{OM'} = \lambda \overrightarrow{OM}$. La même opération est possible pour ensembles en $\mathbb{R}^2, \mathbb{R}^3, \mathbb{R}^d$

1.2.2.2 Dilatation

On appelle le *dilaté* de l'ensemble aléatoire X par le compact K l'ensemble des points $x \in \mathbb{R}^d$ tels que K_x rencontre X :

$$X \oplus K = \{x : X \cap \check{K}_x \neq \emptyset\}$$

En termes de capacité de Choquet :

$$T_X(K) = P\{K \cap X \neq \emptyset\} = P\{o \in X \oplus \check{K}\}$$

où o représente l'origine. D'une autre façon :

$$T_X(K_x) = P\{K_x \cap X \neq \emptyset\} = P\{x \in X \oplus \check{K}\}$$

1.2.2.3 Erosion

L'*érosion* d'un ensemble aléatoire X par un ensemble K compact est l'ensemble des points $x \in \mathbb{R}^d$ tels que K_x soit contenu dans X :

$$X \ominus K = \{x : X \cap \check{K}_x \subset X\}$$

De façon similaire que dans le cas précédant, à l'aide de la fonctionnelle Q :

$$Q(K_x) = P\{K_x \subset X^c\} = P\{x \in X^c \ominus \check{K}\}$$

Les opérations \oplus et \ominus sont appelées respectivement l'addition et la soustraction de Minkowski. Ces transformations permettent d'estimer $T_X(K)$ à partir des réalisations de X . Dans le cas général, nous pouvons estimer au plus une fréquence pour chaque point x à partir de plusieurs réalisations. Pour un ensemble aléatoire *stationnaire*, $T_X(K_x) = T_X(K)$, et pour le cas d'un ensemble aléatoire *ergodique*, $T_X(K)$ peut être estimé à partir d'une seule réalisation et, donc :

$$T_X(K)^* = P\{x \in X \oplus \check{K}\}^* = V_V(X \oplus \check{K})^*$$

où V_V représente la fraction de volume occupée par l'ensemble aléatoire et $*$ un estimateur.

Chaque ensemble compact K apporte une information différente sur la loi spatiale de l'ensemble aléatoire étudié :

– Si $K = \{x\}$:

$$T_X(K) = T_X(\{x\}) = P\{x \in X\}$$

– Si $K = \{x, x+h\}$:

$$T_X(K) = T_X(\{x, x+h\}) = P\{x \in X \cup X_{-h}\}$$

$$Q_X(K) = Q_X(\{x, x+h\}) = P\{x \in X^c \cup X_{-h}^c\}$$

$Q_X(\{x, x+h\})$ est la *covariance non centrée* de X^c .

- Si $K = \{x_1, x_2, \dots, x_n\}$, la loi spatiale de X est :

$$T_X(K) = 1 - P\{x_1 \in X^c, x_2 \in X^c, \dots, x_n \in X^c\}$$

Un ensemble fermé aléatoire X qui admet T comme *loi spatiale* n'est jamais unique. S'il existe un autre EFA X' indépendant de X dont la loi spatiale est nulle (un processus ponctuel stationnaire, par exemple), alors l'EFA $X \cup X'$ admet la même loi spatiale T . Donc, la loi spatiale ne suffit pas pour caractériser complètement un ensemble fermé aléatoire X .

- Si $K = B(r)$, la boule fermée centrée à l'origine et de rayon r :

$$T_X(K) = T_X(B_x(r)) = P\{x \in X \oplus B(r)\}$$

$T_X(B_x(r))$ permet d'estimer la loi de la variable aléatoire $R(x, X)$, qui donne la distance qui sépare x de X :

$$R(x, X) = \text{Sup}\{r; B_x(r) \subset X^c\}$$

La loi de répartition de $R(x, X)$:

$$F_x(r) = P\{R(x, X) \leq r : x \in X^c\} = \frac{T_X(B_x(r)) - T_X(x)}{1 - T_X(x)}$$

Pour quelques cas particuliers (cas stationnaire, le modèle booléen, ...) ces expressions peuvent être obtenues explicitement, mais ce n'est pas le cas général.

1.2.3 Mesures et transformations sur les ensembles fermés aléatoires

1.2.3.1 Mesures sur des EFA

Tout ensemble fermé aléatoire est mesurable dans le sens suivant : soit μ une mesure positive et $\mathbb{I}_X(x)$ l'application indicatrice :

$$\mathbb{I}_X(x) = \begin{cases} 1 & \text{si } x \in X \\ 0 & \text{si } x \notin X \end{cases}$$

Si $\mathbb{I}_X(x)$ est mesurable, la mesure de X peut s'exprimer comme :

$$\mu(X) = \int_{\mathbb{R}^d} \mathbb{I}_X(x) \mu(dx)$$

et $\mu(X)$ est une variable aléatoire d'espérance $E[\mu(X)]$:

$$E[\mu(X)] = \int_{\mathbb{R}^d} P\{x \in X\} \mu(dx)$$

1.2.3.2 Transformations sur des EFA

Toute transformation Ψ semi-continue supérieurement (s.c.s.) ou semi-continue inférieurement (s.c.i.)⁹ peut être appliquée à un ensemble aléatoire X .

Quelques exemples de transformations :

⁹Une application $\Psi : \Omega \rightarrow \mathcal{F}$ est s.c.s. $\Leftrightarrow \Psi^{-1}(\mathcal{V}^K)$ est ouvert dans $\Omega, \forall K$ compact. Ψ est s.c.i. $\Leftrightarrow \Psi^{-1}(\mathcal{V}_G)$ est ouvert dans $\Omega, \forall G$ ouvert. Une application s.c.s. et s.c.i. est continue.

Union : $X, X' \longrightarrow X \cup X'$

Transposition dans \mathbb{R}^d : $X \longrightarrow \check{X} = \{-x : x \in X\}$

Homothétie : $X \longrightarrow \lambda X$ ($\lambda > 0$)

Dilatation par un compact : $X \longrightarrow X \oplus \check{K}$

Intersection : $X, X' \longrightarrow X \cap X'$

Erosion par un compact : $X \longrightarrow X \ominus \check{K}$

Ouverture par un compact : $X_K = X \ominus \check{K} \oplus K$

Fermeture par un compact : $X^K = X \oplus \check{K} \ominus K$

D'autres opérations sont présentées dans Jeulin (1999).

1.3 Ensembles fermés aléatoires stationnaires et isotropes

Dans cette section nous particularisons les éléments définis auparavant pour le cas d'un EFA stationnaire et isotrope. Ici, la capacité de Choquet T est invariante pour toutes les translations et rotations de K . En particulier, elle reste la même pour tout point de l'espace x . Dans ce cadre, nous pouvons définir :

Proportion et porosité La *proportion* ou *fraction de volume* de l'espace, p , occupée par l'ensemble aléatoire fermé X est donnée, en moyenne, par la probabilité associée à T . Dans ce cas, $T_X(\{x\})$ est la même pour tous les points de l'espace et égale à l'espérance de l'indicatrice de X en x :

$$T_X(\{x\}) = P\{x \in X\} = E[\mathbb{I}_X(x)] = p$$

Également, la proportion q de l'espace occupée par le complémentaire de X est donnée par l'espérance de l'indicatrice de X^c :

$$Q_X(\{x\}) = 1 - T_X(\{x\}) = P\{x \in X^c\} = E[\mathbb{I}_{X^c}(x)] = q = 1 - p$$

De plus,

$$Var[\mathbb{I}_X(x)] = Var[\mathbb{I}_{X^c}(x)] = q(1 - q)$$

Covariance La fonctionnelle $Q_X(\{x, x + h\})$ donne la probabilité pour que deux points distants h appartiennent au complémentaire de l'ensemble fermé aléatoire X . Dans le cas stationnaire, elle ne dépend que de h , et correspond à la *covariance non centrée* de l'indicatrice de X^c :

$$\begin{aligned} Q_X(\{x, x + h\}) &= P\{x, x + h \in X^c\} = P\{x \in X^c \cup X_{-h}^c\} = \\ &= E[\mathbb{I}_{X^c}(x)\mathbb{I}_{X^c}(x + h)] = C(h) \end{aligned}$$

Variogramme Le *variogramme* de l'indicatrice de X et de X^c est :

$$\gamma(h) = C(0) - C(h) = q - C(h)$$

Dans ce chapitre nous avons présenté de façon brève et générale les ensembles aléatoires. Pour caractériser ces modèles, nous avons défini un outil qui permet de décrire les ensembles aléatoires de façon similaire aux variables aléatoires. Cet outil est une application qui détermine une probabilité associée à ces ensembles aléatoires, dénommée *capacité de Choquet*. Elle est définie à partir d'un ensemble test et de l'opération d'intersection : on regarde si cet ensemble test rencontre ou non l'ensemble aléatoire. Chaque ensemble test (un point, deux points, une boule, etc.) apporte une information différente sur l'ensemble aléatoire. À travers cette capacité de Choquet nous avons décrit toute une série de propriétés des ensembles aléatoires qui seront très utilisées tout au long de cette étude.

Nous avons aussi introduit deux concepts importants pour la suite de ce travail : le concept de mesure d'un ensemble aléatoire, qui sera utilisé notamment dans la définition du processus de Poisson, et le concept de proportion comme la fraction de volume occupé par l'ensemble aléatoire. Cette notion de proportion sera très utilisée dans la description de l'information disponible lors de la simulation de réservoirs.

Chapitre 2

Processus de Poisson

Le processus de Poisson fait parti d'une famille plus générale : les modèles de points aléatoires. Ces modèles constituent la classe la plus étudiée d'ensembles aléatoires. De la même façon que la structure élémentaire de la géométrie est le point, les modèles de points aléatoires sont à la base de la géométrie stochastique. Malgré leur simplicité géométrique, ils permettent la modélisation de nombreux phénomènes aléatoires qui se présentent sous la forme d'une population de points : des réserves forestières, des inclusions non métalliques dans un acier, des étoiles, etc. De plus, ils sont le point de départ pour la construction de modèles plus complexes. En particulier, le modèle booléen auquel nous nous intéressons est généré à partir d'un processus de Poisson.

Pour cette raison, avant d'introduire les bases théoriques du modèle booléen, dans ce chapitre nous étudions de façon très générale les processus ponctuels, puis quelques éléments mathématiques du processus de Poisson sont établis.

Par la suite, la notation sera Φ pour un processus ponctuel général, \wp pour le processus de Poisson et la notation X est réservée pour l'ensemble booléen.

Ce chapitre est fortement inspiré de Serra (1982), Lantuéjoul (1993) et Stoyan (1995) où d'autres processus de points (processus de Cox, de Neyman-Scott, etc.) sont également étudiés, et aussi de Chessa (1995), Molchanov (1997) et Jeulin (1999).

2.1 Processus ponctuels

Un processus de points aléatoires dans \mathbb{R}^d est un ensemble aléatoire dont chacune des réalisations est constituée d'une population de points. Ici, nous sommes concernés par les processus fermés (i.e. la limite de toute suite de points de la réalisation appartient aussi à la réalisation) qui constituent des ensembles fermés aléatoires. Plus particulièrement, nous nous intéressons aux processus *localement finis* pour lesquels le nombre de points du processus tombant dans un sous-ensemble mesurable est presque sûrement fini. Dans ce cadre, les caractéristiques statistiques du processus peuvent être définies en dénombrant le nombre de points qui tombent dans un ensemble "test" mesurable. Cette restriction nous permet d'introduire la notion de *mesure de comptage*, qui va définir la loi spatiale d'un processus ponctuel de façon analogue à la loi qui caractérise les fonctions aléatoires¹.

¹La distribution spatiale de la fonction aléatoire Z est décrite par la famille des fonctions de distribution multivariée et cumulées :

$$F_{x_1, \dots, x_n}(z_1, \dots, z_n) = P\{Z(x_1) < z_1, \dots, Z(x_n) < z_n\}$$

Si Z prend des valeurs discrètes, comme dans le cas des processus ponctuels, alors il est possible de définir :

$$F_{x_1, \dots, x_n}(z_1, \dots, z_n) = P\{Z(x_1) = z_1, \dots, Z(x_n) = z_n\}$$

Définition 2 *Mathématiquement, un processus ponctuel Φ en \mathbb{R}^d est une variable aléatoire qui prend ses valeurs dans un espace mesurable $[\mathcal{N}, \sigma_{\mathcal{N}}]$, où \mathcal{N} est la famille de toutes les suites φ de points de \mathbb{R}^d qui vérifient les deux conditions suivantes :*

1. φ est **localement finie** : tout sous-ensemble mesurable borné de \mathbb{R}^d renferme un nombre fini de points de φ ,
2. φ est **simple** : $x_i \neq x_j$ si $i \neq j$.

et $\sigma_{\mathcal{N}}$ est la σ -algèbre définie en \mathcal{N} .

C'est à dire, un processus ponctuel est une suite aléatoire de points tel que le nombre de ces points qui tombent dans un sous-ensemble mesurable est fini : il est possible de «compter» le nombre de points.

2.1.1 Ensemble fermé aléatoire versus mesure de comptage

Le processus ponctuel Φ peut donc être interprété aussi bien comme un *ensemble aléatoire de points*, $\Phi = \{x_1, x_2, \dots\} = \{x_n\}$, que comme une *mesure de comptage* du nombre de points de l'ensemble aléatoire qui tombent dans un sous-ensemble mesurable B^2 .

En tant qu'**ensemble aléatoire**, nous utiliserons la notation $x_i \in \Phi$ pour indiquer que le point x_i appartient à l'ensemble Φ . Puisque Φ est localement fini, il est un ensemble aléatoire fermé, et la théorie des ensembles aléatoires fermés vue dans le chapitre précédent peut être appliquée aux processus de points.

La loi spatiale du processus ponctuel Φ est donnée par sa distribution en $[\mathcal{N}, \sigma_{\mathcal{N}}]$. Elle est déterminée mathématiquement par le système de toutes les valeurs :

$$P\{N(B_1) = n_1, \dots, N(B_k) = n_k\}$$

pour $k = 1, 2, \dots$, B_1, \dots, B_k ensembles mesurables bornés et $n_1, \dots, n_k \geq 0$. Plus particulièrement, la distribution de Φ en $[\mathcal{N}, \sigma_{\mathcal{N}}]$ est déterminée par le sous-système pour lequel les B_i sont disjoints deux à deux.

Si l'on considère un sous-ensemble compact K , il est possible d'obtenir la fonctionnelle Q associée au processus ponctuel Φ à partir de sa distribution spatiale :

$$Q_{\Phi}(K) = P\{K \cap \Phi = \emptyset\} = P\{N(K) = 0\}$$

En tant que **mesure de comptage**, nous représentons par $N(B)$ le nombre de points de Φ qui tombent dans l'ensemble B et nous pouvons, donc, l'écrire comme :

$$N(B) = \sum_{x_i \in \Phi} \mathbb{I}_B(x_i)$$

avec

$$\mathbb{I}_B(x_i) = \begin{cases} 1 & \text{si } x_i \in B \\ 0 & \text{sinon} \end{cases}$$

²Plus précisément B appartient à la familles des ensembles boréliens, c'est à dire la famille qui contient tous les sous-ensembles de \mathbb{R}^d qui peuvent être construits à partir des sous-ensembles ouverts. Elle contient aussi les sous-ensembles fermés et les sous-ensembles compacts.

$N(B)$ est une variable aléatoire d'espérance :

$$E[N(B)] = E \left[\sum_{x_i \in \Phi} \mathbb{I}_B(x_i) \right]$$

Puisque le modèle de points aléatoires Φ est localement fini, $N(B)$ est une mesure localement finie et σ -additive, c.a.d :

$$N(B_1 \cup B_2) + N(B_1 \cap B_2) = N(B_1) + N(B_2)$$

pour tout couple d'ensembles mesurables bornés B_1, B_2 .

2.1.2 Moments et mesures des moments

Il est possible de définir les moments (moyenne, variance, etc.) pour les processus de points de façon similaire aux moments des variables aléatoires. Cependant, tandis que les moments de celles-ci prennent des valeurs réelles, les moments des processus ponctuels sont des *mesures*, comme définies dans la section 1.2.3.

Moment d'ordre 1 : Espérance

$$\mu^{(1)}(B) = E[N(B)] = \vartheta(B) \quad (2.1)$$

Un processus ponctuel Φ est caractérisé par une mesure ϑ en \mathbb{R}^d , qui est une caractéristique similaire à la moyenne d'une variable aléatoire. Elle est définie comme l'espérance de la mesure de comptage du processus et représente le nombre moyen de points tombés dans l'ensemble B . Il se vérifie que :

– si Φ est stationnaire : $\vartheta(B) = \theta|B|$, avec $|B|$ le volume de B dans \mathbb{R}^d au sens de la mesure de Lebesgue (voir section 2.1.3) ;

– si Φ est non stationnaire : $\vartheta(B) = \int_B \theta(x) dx$.

Dans ces expressions θ est dénommée l'*intensité* du processus Φ . Ce paramètre détermine la distribution des points du processus Φ dans l'espace. Il est une constante non négative dans le cas stationnaire et une fonction de la position dans le cas non stationnaire. Dans la littérature il est fréquent de trouver les termes intensité ou densité pour définir le même paramètre θ . Dans ce travail nous gardons le terme intensité.

Moment d'ordre 2 : covariance non centrée entre $N(B_1), N(B_2)$ où les B_i sont des ensembles mesurables de \mathbb{R}^d

$$\mu^{(2)}(B_1 \times B_2) = E[N(B_1)N(B_2)] \quad (2.2)$$

Et la variance et la covariance du processus Φ peuvent, donc, s'exprimer comme :

Variance

$$\text{var}[N(B)] = E[N(B)^2] - (E[N(B)])^2 = \mu^{(2)}(B \times B) - (\vartheta(B))^2 \quad (2.3)$$

Covariance

$$\begin{aligned} \text{cov}[N(B_1), N(B_2)] &= E[N(B_1)N(B_2)] - E[N(B_1)]E[N(B_2)] \\ &= \mu^{(2)}(B_1 \times B_2) - \vartheta(B_1)\vartheta(B_2) \end{aligned} \quad (2.4)$$

2.1.3 Stationnarité et isotropie

Comme tout autre ensemble aléatoire, un processus de points arbitraire $\Phi = \{x_n\}$ est *stationnaire* si son translaté par h , $\Phi_h = \{x_n + h\}$ a la même distribution, pour tout $h \in \mathbb{R}^d$. Il est *isotrope* si la même condition se vérifie pour tout processus transformé par rotation $r\Phi = \{rx_n\}$, où r est une rotation autour de l'origine.

Si Φ est stationnaire :

$$\vartheta(B) = E[N(B)] = E[N_h(B)] = E[N_{-h}(B)] = \vartheta(B_h)$$

Cette propriété d'invariance par translation en \mathbb{R}^d implique que :

$$\vartheta(B) = \theta|B|$$

pour $|B|$ le volume en \mathbb{R}^d de B , et θ l'intensité du processus Φ . Si B est de volume unité, θ peut être interprété comme le nombre moyen de points de Φ tombés par unité de volume.

2.1.4 Ergodicité

La propriété d'ergodicité permet d'exprimer les espérances statistiques en termes des limites des moyennes arithmétiques ou spatiales. Dans d'autres termes, cette propriété autorise l'utilisation de moyennes spatiales à la place des espérances statistiques. Si Φ est un processus de points ergodique, alors :

$$\lim_{B_n \uparrow \mathbb{R}^d} \frac{N(B_n)}{|B_n|} = \theta$$

Par $B_n \uparrow \mathbb{R}^d$ nous représentons une suite B_n d'ensembles convexes et compacts qui vérifient :

- $B_n \subset B_{n+1}$
- $\sup\{r \geq 0 : b(x, r) \subset B_n\} \xrightarrow{n \rightarrow \infty} \infty$

où $b(x, r)$ est la boule de rayon r centrée en x . Un exemple de processus ergodique est le processus de Poisson stationnaire.

2.2 Processus de Poisson

Le processus de Poisson est le plus simple et le plus important des processus de points, d'une part parce qu'il répond à l'idée intuitive de points distribués au hasard et d'autre part parce qu'il est à la base d'autres modèles plus compliqués. De plus, de nombreuses techniques de simulation utilisent souvent ce processus de Poisson. En particulier, nous nous intéressons à ce processus parce qu'il sert de point de départ pour la construction du modèle booléen. On note le processus de points de Poisson par \wp .

2.2.1 Processus de Poisson stationnaire

Définition 3 Un processus de Poisson \wp en \mathbb{R}^d stationnaire est caractérisé par deux propriétés fondamentales :

1. Le nombre de points du processus, $N(B)$, tombant dans un ensemble mesurable B est une variable aléatoire qui suit une loi de Poisson de moyenne $\theta|B|$, pour une certaine constante non négative θ :

$$P\{N(B) = n\} = \frac{(\theta|B|)^n}{n!} e^{-\theta|B|}$$

Le paramètre θ est l'**intensité** du processus de Poisson. Dans le cas stationnaire il est constant et nous assumons que $0 < \theta < \infty$. Dans ce cas il représente le nombre moyen de points de Poisson tombés par unité de volume.

2. Pour toute famille d'ensembles mesurables deux à deux disjoints B_1, \dots, B_k , les nombres de points du processus, $N(B_1), \dots, N(B_k)$, sont mutuellement indépendants, pour k arbitraire.

2.2.1.1 Espérance, variance et covariance

Soit le processus de Poisson stationnaire \wp considéré comme une mesure de comptage aléatoire $N(B)$ sur les ensembles $B \in \mathbb{R}^d$.

Espérance L'espérance pour le processus de Poisson est, comme définie en 2.1.3 :

$$\wp(B) = E[N(B)] = \theta|B| \quad (2.5)$$

Variance et covariance Soient B_1, B_2 deux ensembles mesurables de \mathbb{R}^d . Ils peuvent être décrits comme l'union de deux ensembles disjoints :

$$B_1 = (B_1 \cap B_2) \cup (B_1/B_2)$$

$$B_2 = (B_1 \cap B_2) \cup (B_2/B_1)$$

et, donc :

$$N(B_1) = N(B_1 \cap B_2) + N(B_1/B_2)$$

$$N(B_2) = N(B_1 \cap B_2) + N(B_2/B_1)$$

En utilisant l'indépendance de la *mesure de comptage* des domaines disjoints (propriété (2) de la définition 3), la **covariance non centrée** (eq. 2.2) peut s'exprimer comme :

$$\begin{aligned} \mu^{(2)}(B_1 \times B_2) &= E[N(B_1)N(B_2)] = \\ &= E[N(B_1)]E[N(B_2)] + \text{var}[N(B_1 \cap B_2)] \end{aligned}$$

et puisque $N(B_1 \cap B_2)$ suit une loi de Poisson de moyenne et variance $\theta|B_1 \cap B_2|$, il est possible d'écrire :

$$\mu^{(2)}(B_1 \times B_2) = \theta^2|B_1||B_2| + \theta|B_1 \cap B_2|$$

La **variance et covariance** du processus de Poisson sont donc :

$$\text{var}[N(B)] = \mu^{(2)}(B \times B) - \theta^2|B|^2 = \theta|B| \quad (2.6)$$

$$\begin{aligned} \text{cov}[N(B_1), N(B_2)] &= \mu^{(2)}(B_1 \times B_2) - \theta^2|B_1||B_2| = \\ &= \theta|B_1 \cap B_2| = \text{var}[N(B_1 \cap B_2)] \end{aligned} \quad (2.7)$$

2.2.1.2 Quelques propriétés générales du processus de Poisson stationnaire

- Si B_1, \dots, B_k sont des ensembles mesurables bornés disjoints, alors $N(B_1), \dots, N(B_k)$ sont des variables aléatoires de Poisson indépendantes de moyennes $\theta|B_1|, \dots, \theta|B_k|$. Leur probabilité conjointe est :

$$\begin{aligned} P\{N(B_1) = n_1, \dots, N(B_k) = n_k\} &= \\ &= \frac{\theta^{n_1 + \dots + n_k} (|B_1|^{n_1} \dots |B_k|^{n_k})}{n_1! \dots n_k!} e^{-\sum_{i=1}^k \theta|B_i|} \end{aligned}$$

- Les propriétés du processus φ et de θ sont invariantes par translation et rotation.
- Si φ est un processus de Poisson stationnaire sur un ensemble compact $B \subset \mathbb{R}^d$ avec la condition $N(B) = n$, les n points sont indépendants et uniformément distribués en B . La démonstration est disponible dans Lantuéjoul (1993) et Stoyan (1995).
- La probabilité pour qu'aucun point ne tombe dans l'ensemble B est :

$$P\{N(B) = 0\} = e^{-\theta|B|}$$

2.2.2 Processus de Poisson général : non stationnaire ou régionalisé

Dans les cas modélisés par un processus de Poisson stationnaire, les points sont distribués dans le milieu de façon homogène, c'est à dire, le nombre moyen de points tombés par unité de volume ne varie pas dans l'espace. Ceci est représenté par une mesure ϑ qui est, dans ce cas, proportionnelle au volume de la région étudiée (eq. 2.5). La constante de proportionalité θ , appelée intensité, donne le nombre moyen de points par unité de volume.

Cependant, de nombreux phénomènes présentent des variations impossibles à modéliser par un modèle stationnaire. Dans ce cas, le processus de Poisson général fournit un modèle stochastique plus général, où le nombre moyen de points par unité de volume n'est plus constant, mais il peut varier dans l'espace. Par la suite, nous allons introduire brièvement la généralisation des concepts, présentés dans cette section (la mesure, l'intensité du processus, etc.), qui caractérisent le modèle général.

Le processus de Poisson général est caractérisé par une intensité θ qui n'est plus constante. Elle est une fonction de l'espace et on va la supposer positive et intégrable localement. La mesure ϑ s'exprime, dans ce cas, comme :

$$\vartheta(B) = \int_B \theta(x) dx < \infty$$

pour tout ensemble mesurable B de \mathbb{R}^d . La paramètre θ est maintenant une fonction, la **fonction d'intensité** du processus de Poisson général. Le terme $\theta(x)dx$ est la probabilité pour qu'un point de φ tombe dans une région de volume infinitésimal placée en x .

Définition 4 Un processus de Poisson général, φ , avec la mesure ϑ vérifie les propriétés suivantes :

1. Le nombre de points tombés dans un ensemble mesurable borné B , $N(B)$ est une variable aléatoire qui suit une distribution de Poisson de paramètre $\vartheta(B)$:

$$P\{N(B) = n\} = \frac{(\vartheta(B))^n}{n!} e^{-\vartheta(B)}$$

2. Les nombres de points tombant en k ensembles mesurables deux à deux disjoints constituent k variables aléatoires indépendantes.

Théorème 3 Soit B un sous-ensemble mesurable de \mathbb{R}^d tel que $\vartheta(B) < +\infty$. Si $N(B) = n$, les n points sont distribués indépendamment dans B selon la densité $\theta(x)/\vartheta(B)$.

2.2.2.1 Espérance, variance et covariance

Nous pouvons généraliser les expressions 2.5, 2.6 et 2.7 pour le processus de Poisson général de la façon suivante :

$$\begin{aligned}E[N(B)] &= \vartheta(B) \\var{ar}[N(B)] &= \vartheta(B) \\cov(N(B_1), N(B_2)) &= \vartheta(B_1 \cap B_2)\end{aligned}$$

Le processus de Poisson est à la base du modèle booléen que nous introduirons ultérieurement. Dans ce chapitre nous avons posé les bases de la théorie des processus ponctuels en particulierisant pour le processus de Poisson, dans les cas stationnaire et non stationnaire. Les processus ponctuels en général peuvent être décrits aussi bien comme des ensembles fermés aléatoires, que comme une mesure de comptage. Cette seconde notion est celle que nous avons utilisée dans les processus de points puisqu'elle nous permet de les décrire d'une façon intuitive : en dénombrant le nombre de points tombés dans une certaine région de l'espace. Dans le cas du processus de Poisson, ce nombre suit une loi de Poisson. Ainsi, nous avons introduit un paramètre qui sera très important dans ce travail : la *fonction d'intensité de Poisson* qui va déterminer le nombre moyen de points tombant dans un volume et aussi leur distribution dans ce volume. La détermination de ce paramètre à partir des informations concernant les chenaux dans un réservoir constitue une grande partie du corps de cette thèse.

Chapitre 3

Modèle booléen

Le modèle booléen est sûrement le plus important des modèles d'ensembles fermés aléatoires. Il correspond à la réunion d'objets implantés au hasard. Il s'agit d'une famille de modèles très flexibles, utilisés principalement quand les objets à simuler ont une interprétation physique ou génétique.

Ce modèle est un processus de jetons ou de type objet («marked point processes») basé sur un processus de Poisson. Dans chaque point du processus ponctuel, on implante un jeton ou objet qui est une réalisation indépendante d'un sous-ensemble aléatoire. Le modèle booléen est caractérisé par le fait que ces sous-ensembles sont identiquement et indépendamment distribués (i.i.d.). Un exemple très élémentaire de construction d'un modèle booléen est le suivant : supposons des points dispersés dans un plan selon un processus de Poisson stationnaire d'intensité θ . Sur chacun de ces points on place un disque de rayon constant. L'union X de tous ces disques est un exemple de modèle booléen. Les points du processus de Poisson sont appelés les *germes* du modèle et les disques, les *grains primaires* ou *objets*. Le schéma booléen divise ainsi l'espace en deux parties : le milieu des *grains* et son complémentaire, le milieu des *pores*.

Cette construction peut être généralisée pour obtenir le modèle booléen général pour lequel le processus de Poisson n'est plus stationnaire mais de fonction d'intensité $\theta(x)$. Soient les points du processus de Poisson général qui se trouvent aux points $x_i \in \mathbb{R}^d$. Prenons plusieurs réalisations A de l'ensemble aléatoire des grains primaires et implantons-les aux points x_i du processus de Poisson général $\varphi : A(x_i)$. Les différents $A(x_i)$ sont, donc, mutuellement indépendants et indépendants du processus φ . Le modèle booléen est la réunion de tous les $A(x_i)$. Encore une généralisation du modèle booléen consiste à prendre la loi des objets dépendant de leur point d'implantation.

Une propriété élémentaire de ce type de modèle est l'**infinie divisibilité** (voir section 1.2.1.4) qui implique que le modèle peut être considéré comme l'union d'un certain nombre d'ensembles aléatoires indépendants de la même famille. De plus, si un modèle booléen est défini en \mathbb{R}^d , ses restrictions en \mathbb{R}^2 ou en \mathbb{R}^3 sont toujours des modèles booléens en \mathbb{R}^2 ou en \mathbb{R}^3 .

Le modèle booléen a été très largement utilisé, surtout dans les sciences naturelles. Des exemples d'applications de ce modèle concernent la microstructure du papier, la cristallisation dans les métaux, l'étude des matériaux en général, l'étude des systèmes des gouttes d'eau dans les transition de phase ..., et plus récemment l'étude des structures géologiques générées par sédimentation. C'est à ce dernier cas que nous nous intéressons.

Dans ce chapitre nous introduisons les éléments de base de la théorie du modèle booléen qui seront utilisés pour la suite de ce travail. Les principaux ouvrages utilisés dans ce chapitre sont ceux de Serra (1982), Lantuéjoul (1993,2002), Stoyan (1995), Molchanov (1997) et Chilès et Delfiner (1999).

3.1 Définition du modèle booléen

Les éléments de base pour construire un modèle booléen dans \mathbb{R}^d sont :

1. Un ensemble de points x_i , les *germes* du modèle : ils appartiennent à un processus de Poisson, \wp , caractérisé par sa fonction d'intensité $\theta(x)$, positive et intégrable.
2. Une famille de sous-ensembles compacts indépendants et non vides de \mathbb{R}^d , A . Le sous-ensemble $A(x_i)$ est l'*objet* ou *grain primaire* implanté au point de Poisson x_i .

Définition 5 Soit \wp un processus de Poisson ponctuel en \mathbb{R}^d . Soit A une famille de compacts aléatoires non vides mutuellement indépendants et de même loi. Le **schéma booléen** est la réunion de tous les compacts A implantés aux points x_i du processus de Poisson :

$$X = \bigcup_{x_i \in \wp} A(x_i)$$

Le modèle booléen dépend, donc, de deux paramètres : l'intensité du processus de Poisson, $\theta(x)$, et la loi des grains primaires, caractérisée par leur capacité de Choquet, $T_{x_i}(K) = P\{A(x_i) \cap K \neq \emptyset\}$.

3.1.1 Stationnarité, isotropie et ergodicité

Si le processus de Poisson \wp des germes est stationnaire, et les grains primaires sont identiquement distribués, le modèle booléen défini est **stationnaire**. C'est à dire, sa distribution est invariante par translation. Si les grains primaires ont en plus une distribution isotrope (une distribution invariante par rotation autour de l'origine o) alors le modèle booléen est aussi **isotrope**.

Un modèle booléen avec des grains primaires convexes et compacts est **ergodique**. Une conséquence de l'ergodicité est que :

$$\lim_{r \rightarrow +\infty} \frac{|X \cap b(o, r)|}{|b(o, r)|} = p \quad (3.1)$$

où $b(o, r)$ est la boule de rayon r centrée à l'origine et $p = P\{o \in X\}$ est la fraction du volume du modèle booléen. De façon simplifiée, cette propriété permet d'exprimer l'espérance statistique du modèle comme une limite spatiale (voir section 1.2.1.3).

Quelques applications utilisent des modèles booléens non stationnaires. Un cas important est celui où les grains sont indépendants et identiquement distribués, comme dans le cas stationnaire, mais le processus de Poisson des germes est régionalisé avec une fonction d'intensité $\theta(x)$.

3.1.2 Propriétés de stabilité

Le schéma booléen satisfait les propriétés de stabilité suivantes :

1. la réunion de deux modèles booléens indépendants est un modèle booléen ;
2. le dilaté d'un modèle booléen par un sous-ensemble compact non vide de \mathbb{R}^d est un modèle booléen ;
3. l'intersection entre un modèle booléen et un sous-ensemble compact de \mathbb{R}^d est un modèle booléen et
4. l'intersection entre un modèle booléen et un i -plan est un ensemble booléen.

La démonstration de ces propriétés est présentée dans Lantuéjoul (1993) et Lantuéjoul (2002).

3.2 Capacité de Choquet du modèle booléen

Comme pour tout autre ensemble fermé aléatoire, la distribution du modèle booléen est déterminée uniquement par sa capacité de Choquet, T_X :

$$T_X(K) = P\{K \cap X \neq \emptyset\} \quad \forall K \text{ compact} \quad (3.2)$$

Dans le cas d'un modèle booléen, T_X peut s'exprimer d'une façon relativement simple. À partir du processus de Poisson \wp , nous pouvons définir un nouveau processus \wp' en éliminant les x_i de \wp si $A(x_i) \cap K = \emptyset$, avec K un compact :

$$\wp' = \{x_i \in \wp : A(x_i) \cap K \neq \emptyset\}$$

En d'autres termes, \wp' comprend uniquement les points de \wp tels que les grains implantés en eux rencontrent K . Le fait qu'un germe x_n soit effacé ou non de \wp à travers cette procédure est indépendant de ce qui se passe aux autres germes. Cette propriété vient de la propriété d'indépendance du modèle booléen : l'«élimination» dépend uniquement de la localisation du germe et du grain associé. Cette indépendance implique que \wp' est aussi un processus de Poisson de mesure \wp' :

$$\wp'(B) = \int_B \theta(u)p(u)du \quad \forall B \text{ ensemble mesurable}$$

avec $p(u) = P\{A(u) \cap K \neq \emptyset\} = T_u(K)$.

Notons que $p(u)$ est la probabilité pour qu'un germe au point u ne soit pas éliminé. Chaque élément du volume du de l'espace apporte sa contribution indépendamment des autres. En du , centré en u , trois événements mutuellement exclusifs peuvent arriver :

1. aucun germe en du avec une probabilité $1 - \theta(du)$ ou
2. un germe en du mais le grain $A(u)$ n'intersecte pas K avec une probabilité $\theta(du)[1 - T_u(K)]$
3. un germe en du et le grain $A(u)$ intersecte K avec une probabilité $\theta(du)T_u(K)$

Du fait des propriétés du processus de Poisson, la distribution du nombre total de points de Poisson dans \wp' dépend seulement de la mesure :

$$\wp'(\mathbb{R}^d) = \int_{\mathbb{R}^d} \theta(u)T_u(K)du$$

Le terme $\theta(u)du$ représente la probabilité pour que u soit un point du processus \wp et $T_u(K)$ est la probabilité pour que l'objet implanté en u intersecte le compact K .

La valeur $T_X(K)$ est donnée par :

$$1 - T_X(K) = Q_X(K) = P\{\wp' \text{ ne contient pas de points}\}$$

Si $\wp'(\mathbb{R}^d)$ est fini, alors la probabilité de \wp' de ne pas contenir de point est, comme indiqué dans la section 2.2.2 :

$$e^{-\wp'(\mathbb{R}^d)} = e^{-\int_{\mathbb{R}^d} \theta(u)T_u(K)du}$$

et donc,

$$\begin{aligned} T_X(K) &= 1 - \exp \left[- \int_{\mathbb{R}^d} \theta(u)T_u(K)du \right] \\ Q_X(K) &= \exp \left[- \int_{\mathbb{R}^d} \theta(u)T_u(K)du \right] \end{aligned} \quad (3.3)$$

Deux modèles booléens ayant la même capacité de Choquet, présentent la même distribution. *Schmitt (1991) a montré que si la capacité de Choquet d'un modèle booléen est connue, alors elle détermine uniquement l'intensité θ de son processus de Poisson. La distribution des grains est connue aux translations aléatoires près.*

Nombre de grains A qui intersectent K : Nous avons supposé, dans le chapitre précédent, que l'intensité du processus de Poisson vérifie : $0 \leq \theta \leq \infty$. C'est à dire le nombre de points de Poisson tombant dans un domaine est fini. Alors, l'ensemble des réalisations de la famille des grains primaires A qui sont implantés dans ces points ($A(x_i); x_i \in \wp$) prend aussi un nombre fini de valeurs. Cette condition assure que seulement un nombre fini de grains rencontre un ensemble compact : le modèle booléen est d'ordre fini. En particulier elle assure que l'ensemble booléen est fermé.

Le nombre de grains primaires qui rencontrent un compact K , $N(K)$, est égal à :

$$N(K) = \sum_{x_i \in \wp} \mathbb{I}_{K \cap A(x_i) \neq \emptyset}$$

$N(K)$ est une variable aléatoire qui suit une loi de Poisson d'espérance :

$$E[N(K)] = \vartheta(K) = \int_{\mathbb{R}^d} \theta(u) T_u(K) du \leq \infty$$

$$N_\theta(K) \sim \text{Poisson}(\vartheta(K))$$

La démonstration de ce résultat est disponible dans Lantuéjoul (2002) et Serra (1982).

3.2.1 Une propriété fondamentale du modèle booléen

Soit le grain primaire centré à l'origine $A(x_i = 0)$. En tant qu'ensemble aléatoire, $A(0)$ est caractérisé par la famille de fonctions :

$$T_0(K) = P\{A(0) \cap K \neq \emptyset\}$$

où $T_0(K)$ est la capacité de Choquet d'un objet implanté à l'origine.

Puisque les objets A sont identiquement distribués, si $A(0)$ est déplacé de l'origine selon un vecteur $u \in \mathbb{R}^d$, la nouvelle fonction est :

$$T_u(K) = P\{A(u) \cap K \neq \emptyset\} = T_0(K_{-u})$$

La probabilité (voir eq. 3.3) pour que l'ensemble K n'intersecte pas l'ensemble booléen X , c'est à dire, la probabilité pour que l'ensemble K soit dans X^c , les pores, s'exprime alors :

$$Q_X(K) = P\{K \subset X^c\} = \exp \left[- \int_{\mathbb{R}^d} \theta(u) T_0(K_{-u}) du \right] \quad (3.4)$$

Cette expression lie la capacité de Choquet ($T = 1 - Q$) de l'ensemble booléen X à celle du grain primaire implanté à l'origine, à travers la fonction d'intensité du processus de Poisson. Si l'ensemble compact K est réduit à un point, nous avons :

$$Q_X(\{x\}) = P\{x \in X^c\} = \exp \left[- \int_{\mathbb{R}^d} \theta(u) T_0(\{x\}_{-u}) du \right] \quad (3.5)$$

Cette expression est à l'origine de la méthode développée dans ce travail et qui nous permettra d'obtenir la fonction d'intensité de Poisson. Cette méthode sera détaillée plus en profondeur dans les chapitres suivants.

3.2.2 Applications de la propriété fondamentale : cas stationnaire

L'utilisation directe des formules (3.3) nous conduit à calculer les fonctionnelles usuelles. D'abord il faut noter que le côté droit de l'expression est constant quand θ est constant (modèle booléen stationnaire).

3.2.2.1 Interprétation géométrique de la propriété fondamentale

Nous allons, tout d'abord, interpréter géométriquement l'expression (3.4). Par besoin de simplicité nous nous limiterons au cas stationnaire, quand la fonction θ est constante dans l'espace. Dans ce contexte cette expression est égale à :

$$Q_X(K) = P\{K \subset X^c\} = \exp \left[-\theta \int_{\mathbb{R}^d} T_0(K_{-u}) du \right] \quad (3.6)$$

Par hypothèse, les objets implantés sont distribués de façon indentique et indépendante dans l'espace. Ceci implique que la capacité de Choquet T de l'objet implanté en u est égale à la fonctionnelle d'un objet implanté à l'origine et déplacé en u . Si on dénote par $A(u)$ l'objet implanté en u , par A l'objet implanté à l'origine et par A_u son translaté en u , nous avons la relation suivante :

$$T_u(K) = P\{A(u) \cap K \neq \emptyset\} = P\{A_u \cap K \neq \emptyset\} = P\{A \cap K_{-u} \neq \emptyset\} = T_0(K_{-u}) \quad (3.7)$$

Soit $\mathbb{I}_{A \cap K_{-u}}$ l'indicatrice de l'ensemble aléatoire $A \oplus \check{K}$:

$$\mathbb{I}_{A \cap K_{-u}} = \begin{cases} 1 & \text{si } A \cap K_{-u} \neq \emptyset \\ 0 & \text{sinon} \end{cases}$$

L'intégrale dans (3.6) peut être exprimée comme :

$$\begin{aligned} \int_{\mathbb{R}^d} T_0(K_{-u}) du &= E \left[\int_{\mathbb{R}^d} \mathbb{I}_{A \cap K_{-u}} du \right] \\ &= E \left[\int_{\mathbb{R}^d} \mathbb{I}_{-u \in A \oplus \check{K}} du \right] \\ &= E[|A \oplus \check{K}|] \end{aligned}$$

et, donc, la probabilité pour que le compact K ne rencontre pas l'ensemble booléen X est :

$$Q(K) = P\{K \subset X^c\} = e^{-\theta E[|A \oplus \check{K}|]} \quad (3.8)$$

Cette expression lie la fonctionnelle $Q = 1 - T$ de X à la mesure de Lebesgue (le volume moyen) $E[|A \oplus \check{K}|]$ du grain primaire dilaté. Ce résultat a une valeur uniquement théorique, puisqu'il ne fournit pas un moyen expérimental pour calculer la probabilité pour qu'un ensemble compact K rencontre l'ensemble aléatoire X , sauf dans certains cas. Si A est convexe et pour quelques K particuliers (un point, un segment, par exemple), nous sommes capables de déterminer $P\{K \subset X^c\}$, mais pas en général.

3.2.2.2 $K = \{x\}$: porosité - fraction de volume

Si l'élément structurant K est réduit à un point x , alors l'expression (3.8) donne la probabilité q pour qu'un point de l'espace appartienne aux pores, appelée **porosité** :

$$q = e^{-\theta E[|A|]}$$

q est la proportion du complémentaire de X . Alors, nous pouvons définir la **proportion** p comme le complémentaire de q , c'est à dire la fraction moyenne du volume occupé par X dans une région de volume unité :

$$p = 1 - q$$

Cette valeur ne dépend pas du choix de la région, en vertu de la stationnarité de X .

3.2.2.3 $K = \{x, x + h\}$: covariance - variogramme

Si $K = \{x, x + h\}$ un couple de points :

$$E[|A \oplus \check{K}|] = E[2|A| - |A \cap A_{-h}|] = 2g(0) - g(h)$$

avec g le covariogramme géométrique moyen des grains primaires, c'est à dire le volume moyen de l'intersection de A avec son translaté par $-h$:

$$g(h) = E[|A \cap A_{-h}|] = g(-h)$$

La **covariance non centrée des pores** X^c , $C(h)$, est donnée par :

$$C(h) = P\{x, x + h \in X^c\} = Q_X(K) = e^{[-\theta(2g(0) - g(h))]}$$

La porosité $q = C(0)$ est égale à $\exp[-\theta g(0)]$ et

$$C(h) = q^2 e^{\theta g(h)}$$

Quant à la **covariance non centrée des grains** X , $C_1(h)$, à partir de la relation :

$$P\{x \in X\} = P\{x, x + h \in X\} + P\{x \in X, x + h \in X^c\}$$

on obtient :

$$C_1(h) = P\{x, x + h \in X\} = 1 - 2q + q^2 e^{\theta g(h)}$$

Les **variogrammes** de X et de X^c (ou, plus exactement, de leurs indicatrices) sont identiques :

$$\gamma(h) = C(0) - C(h) = q - C(h)$$

et peuvent être exprimés comme :

$$\gamma(h) = q(1 - q e^{\theta g(h)}) \quad (3.9)$$

3.3 Simulation d'un modèle booléen

Les algorithmes de simulation conditionnelle utilisés lors de ce travail ont été proposés par Gedler et Lantuéjoul en 1991 (Gedler, 1991) pour le cas d'un modèle booléen stationnaire. Puis, en 1995, Lantuéjoul a présenté l'extension de cet algorithme pour un modèle booléen général (Lantuéjoul, 1995). Ici nous présentons schématiquement ces algorithmes, pour une description plus détaillée voir Lantuéjoul (2002).

L'algorithme de simulation conditionnelle est basé sur une simulation non conditionnelle, celle-ci est donc présentée en premier lieu.

Simulation non conditionnelle

Tout d'abord, nous allons définir le modèle booléen restreint au domaine de simulation V . Soit V un sous-ensemble compact non vide de \mathbb{R}^d . Suivant la propriété stabilité du modèle booléen (sec. 3.1.2), l'intersection d'un tel sous-ensemble avec l'ensemble booléen X est aussi un modèle booléen. Commençons par calculer la probabilité pour que l'ensemble $X \cap V$ ne contienne aucun objet :

$$\begin{aligned}
 P\{(X \cap V) \cap K = \emptyset\} &= P\{X \cap (V \cap K) = \emptyset\} \\
 &= \exp \left[- \int_{\mathbb{R}^d} \theta(u) T_u(V \cap K) du \right] \\
 &= \exp \left[- \int_{\mathbb{R}^d} \theta(u) T_u(V) \frac{1}{T_u(V)} T_u(V \cap K) du \right] \quad (3.10)
 \end{aligned}$$

Le terme $\frac{T_u(V \cap K)}{T_u(V)}$ représente la probabilité pour que l'objet implanté en u , $A(u)$, intersecte l'ensemble $V \cap K$ sous la condition que $V \cap A(u) \neq \emptyset$. Alors, l'ensemble $X \cap V$ est un modèle booléen d'intensité $\theta(u) T_u(V)$ et de capacité de Choquet associée à l'objet $\frac{T_u(V \cap K)}{T_u(V)}$.

Le nombre d'objets de $X \cap V$ suit une loi de Poisson d'espérance :

$$\vartheta(V) = \int_{\mathbb{R}^d} \theta(u) T_u(V) du$$

Un objet de $X \cap V$ est implanté en x selon la fonction de densité $\theta(x) T_x(V)$ et sa capacité de Choquet est $\frac{T_x(V \cap K)}{T_x(V)}$.

La capacité de Choquet de l'ensemble $X \cap V$ est donnée par le nombre d'objets qui intersectent $V \cap K$ parmi ceux qui sont implantés en V :

$$\begin{aligned}
 T_{X \cap V}(K) &= P\{(X \cap V) \cap K \neq \emptyset\} = \\
 &= \int_{\mathbb{R}^d} \frac{\theta(u) T_u(V) T_u(V \cap K)}{\vartheta(V) T_u(V)} du \\
 &= \frac{1}{\vartheta(V)} \int_{\mathbb{R}^d} \theta(u) T_u(V \cap K) du \\
 &= \frac{\vartheta(V \cap K)}{\vartheta(V)} \quad (3.11)
 \end{aligned}$$

Le nombre d'objets de V suit une loi de Poisson d'espérance $\vartheta(V)$ et il est possible de simuler cette distribution de façon itérative selon un processus de naissance et de mort. Supposons que la simulation soit une population Θ de N_t grains à l'instant t . N_t reste constant est égale à n pendant une période de temps qui suit une distribution exponentielle d'espérance : $\frac{1}{\vartheta(V)+n}$. Alors, à chaque étape nous pouvons soit ajouter un nouvel objet, soit retirer un objet parmi ceux déjà implantés. L'algorithme choisi est le suivant :

Algorithme 1 : simulation de la population d'objets d'un modèle booléen

1. Initialisation : $\Theta = \emptyset$;
2. simuler une variable aléatoire U de valeurs $+1, -1$ ou 0 avec les probabilités respectives :

$$\begin{aligned}
 p_{+1} &= \frac{\vartheta(V)}{\vartheta(V) + n + 1} \\
 p_{-1} &= \frac{n}{\vartheta(V) + n} \\
 p_0 &= 1 - p_{+1} - p_{-1}
 \end{aligned}$$

3. si $U = 0$, ne rien faire ;
4. si $U = +1$, simuler un objet A et alors $\Theta = \Theta \cup A$;
5. si $U = -1$, tirer un objet A de façon uniforme dans les grains déjà simulés. Alors : $\Theta = \Theta \setminus A = \Theta - A$
et
6. retourner à 2.

À l'instant t , l'ensemble simulé est la réunion des N_t objets de la population. Le nombre d'objets effectivement simulés suit une loi de Poisson de moyenne $\vartheta(V)(1 - e^{-t})$ (Feller, 1968) et quand $t \rightarrow \infty$, ce valeur converge vers $\vartheta(V)$. En pratique, on peut seulement dire que l'on reste autour de cette valeur (Gedler, 1991). Selon cet algorithme la population initiale d'objets Θ est vide, mais il est tout à fait possible commencer à partir d'une population quelconque d'objets de $X \cap V$: une simulation ou image initiale. Les pas 5 et 6 de l'algorithme assurent la disparition des objets de cette image initiale au cours de la simulation.

Au pas 4 de l'algorithme précédent, nous devons simuler un objet de $X \cap V$. Nous avons supposé que le domaine de simulation est déjà V , c'est à dire que les objets simulés intersectent forcément la région $X \cap V$. Dans un cas plus général, si cette condition n'est pas imposée a priori, un algorithme pour simuler un objet implanté en un point x de l'espace et qui intersecte l'ensemble $X \cap V$, est le suivant :

Algorithme 2 : simulation d'un objet

1. Générer $x \sim \theta(u)T_u(V)$;
2. générer $A \sim \frac{T_x(V \cap K)}{T_x(V)}$;
3. si $A \cap V = \emptyset$, retourner à 2
4. rendre $A \cap V$.

Simulation conditionnelle

Le problème qui se pose dans la plupart des situations et notamment dans la caractérisation des réservoirs, est le conditionnement. C'est à dire comment réaliser une simulation dans V sous la condition que deux sous-ensembles finis de points C_0 et C_1 soient contenus dans l'ensemble des pores et l'ensemble des grains respectivement. L'algorithme que nous utilisons dans ce travail est un algorithme d'acceptation et rejet dû à Lantuéjoul (voir Lantuéjoul, 1995 ; Lantuéjoul, 1997 et Lantuéjoul, 2002).

Algorithme 3 : simulation conditionnelle Soit Ω_c l'ensemble des populations d'objets qui vérifient les conditions C_0, C_1 .

1. Générer $\Theta \in \Omega_c$, une population d'objets qui satisfait les conditions sur les pores et les grains ;
2. Simuler une variable aléatoire U qui prenne les valeurs $+1, -1, 0$ avec les probabilités respectives :

$$p_{+1} = \frac{\vartheta(V)}{\vartheta(V) + n + 1}$$

$$p_{-1} = \frac{n}{\vartheta(V) + n}$$

$$p_0 = 1 - p_{+1} - p_{-1}$$

3. si $U = 0$, ne rien faire ;
4. si $U = +1$, simuler un objet A . Si $A \cap C_0 = \emptyset$, alors $\Theta = \Theta \cup A$, sinon aller à 2 ;

5. si $U = -1$, enlever un objet tiré uniformément parmi ceux déjà simulés. Si $C_1 \subset \Theta \setminus A$, alors $\Theta = \Theta \setminus A$, sinon aller à 2 ;
6. retourner à 2.

Dans le cas conditionnel, toutes les simulations initiales ne sont pas autorisées. Le conditionnement doit être respecté : $\Theta \in \Omega_c$. Un moyen de construire l'image (ou la simulation) initiale est de générer une séquence d'objets indépendants. Chaque fois qu'un objet rencontre C_0 il est éliminé. Le processus continue jusqu'à ce que les objets restants recouvrent C_1 . L'algorithme utilisé est le suivant :

Algorithme 4 : initialisation de la simulation conditionnelle

1. Faire $\Theta = 0$ et $C = C_1$;
2. simuler un objet A selon l'algorithme 2 ;
3. si $A \cap C_0 \neq \emptyset$ ou $A \cap C = \emptyset$, retourner à 2 ;
4. faire $\Theta = \Theta \cup A$ et $C = C \setminus A$;
5. si $C \neq \emptyset$, retourner à 2 ;
6. rendre Θ

En pratique, cependant, pour accélérer le processus la génération des germes est limité au voisinage du conditionnement.

La démonstration détaillée de ces algorithmes et de la capacité de Choquet du modèle booléen conditionnel sont dans Gedler (1991), Lantuéjoul (1995) et Lantuéjoul (2002).

En résumé, le modèle booléen permet de reproduire un phénomène naturel possédant une structure géométrique définie. L'idée est d'implanter dans l'espace des objets plus ou moins réalistes qui reflètent cette géométrie. Dans le cas des réservoirs de type fluviale, les chenaux seront représentés par des objets de forme plus ou moins sinusoïdale.

Le modèle booléen est basé sur un processus de Poisson qui génère les germes sur lesquels les objets seront implantés. Ces objets sont des ensembles aléatoires distribués de façon indépendante et identique, c'est à dire que leur loi ne dépend pas de leurs points d'implantation.

Dans ce projet, nous nous sommes intéressés aux modèles booléens non stationnaires pour essayer de reproduire une réalité non stationnaire. Dans ce modèle le processus de Poisson des germes a une intensité non constante, mais variable dans l'espace. Le cas stationnaire a été vu lors de ce chapitre comme un cas particulier du modèle booléen général.

Nous avons vu une propriété fondamentale du modèle booléen qui lie la capacité de Choquet de l'ensemble booléen à la capacité de Choquet des objets et à l'intensité du processus de Poisson, c'est à dire aux caractéristiques des objets et à leur distribution dans l'espace respectivement. Dans le cas particulier où l'élément structurant K est réduit à un point, il est possible d'estimer l'intensité de Poisson θ si on connaît la loi des objets et si nous sommes capables d'exprimer la capacité de Choquet de l'ensemble en fonction de l'information expérimentale disponible. Ceci est la base de la méthode d'inférence développée dans ce travail et qui sera détaillée dans les chapitres suivants.

Deuxième partie

Inférence des paramètres du modèle

Dans le domaine de la caractérisation des réservoirs, les approches de type objet sont principalement utilisées pour simuler des réservoirs pétroliers en environnement fluvial (chenaux, crevasses). Ces méthodes créent des modèles de réservoirs basés sur des objets possédant une interprétation géologique. Ainsi, elles sont utilisées dans les phases initiales d'évaluation d'un réservoir, quand il n'y a pas beaucoup d'information de puits et que l'intégration de l'information géologique est très importante. À titre d'exemple, dans les réservoirs en environnement fluvial, les sables déposés par des anciennes rivières forment une structure géologique bien définie, un chenal, qui sera représentée directement par un objet sinusoïdal de section semi-ellipsoïdale. Il est donc nécessaire de choisir les objets, leur forme et leur taille ainsi que le nombre d'objets à simuler, et, éventuellement, les relations existant entre eux.

Un des problèmes posés par ces approches auquel nous nous sommes intéressés, concerne la distribution non stationnaire des objets dans la région étudiée. En effet, dans la réalité, les structures géologiques que nous essayons de modéliser ne sont pas réparties de façon uniforme dans le réservoir. Cette non stationnarité est mise en évidence par la distribution, dans le volume étudié, de la proportion du lithofaciès associé à ces structures. Pour obtenir des images réalistes du réservoir il est nécessaire de reproduire cette caractéristique dans les simulations, c'est-à-dire d'implanter plus d'objets dans les endroits susceptibles d'avoir une proportion plus élevée du lithofaciès correspondant. Nous avons choisi de modéliser ce lithofaciès par un modèle booléen.

La proportion du lithofaciès en question donne une information partielle sur la loi spatiale de l'ensemble booléen, soit la capacité de Choquet $T_X(K)$ du modèle, notamment dans le cas où K est réduit à un point. En termes des paramètres du modèle, la non stationnarité dans la distribution de ce lithofaciès peut être relative soit à la distribution spatiale des objets, soit à la variation des paramètres des objets, soit aux deux. Chacun de ces aspects concerne un paramètre du modèle booléen : la distribution spatiale des objets fait référence à la distribution des germes et donc, à la fonction d'intensité du processus de Poisson, θ ; les caractéristiques des objets sont introduites à travers leur capacité de Choquet, $T_d(K)$.

Le but de notre travail est double : d'une part, il faut déterminer la distribution des objets dans l'espace. Le problème à résoudre est l'inférence de la fonction d'intensité de Poisson pour une famille d'objets choisie compte tenu de la distribution de la proportion du lithofaciès. Conjointement, il faut déterminer quelle information sur les objets peut être obtenue à partir des données. En particulier, il faut tester la cohérence du choix de la famille d'objets avec la distribution de la proportion. Il est, par exemple, impossible de reproduire un faciès dont la proportion présente des fluctuations de faible dimension avec des objets très grands.

Cette deuxième partie est divisée en trois chapitres. Dans le premier, nous décrivons l'information disponible et nous établissons le lien entre celle-ci et les paramètres du modèle booléen. Dans le second chapitre nous proposons une méthode pour inférer l'intensité de Poisson à partir des proportions du faciès modélisé. Pour finir, le troisième chapitre de cette seconde partie traite de l'adaptabilité du choix des objets par rapport à la proportion du faciès.

Chapitre 4

Description des variables expérimentales

Nous disposons le plus souvent de deux informations expérimentales concernant le réservoir : la proportion du faciès d'intérêt dans une région, et des informations sur la taille et la forme des hétérogénéités à simuler. Dans le modèle, ces informations sont liées, respectivement, à la capacité de Choquet de l'ensemble booléen et à la capacité de Choquet de l'objet à simuler. Dans ce chapitre, nous décrivons l'information expérimentale disponible et la façon dont elle est obtenue. Le lien entre ces variables expérimentales et les paramètres théoriques du modèle est établi.

Le point de départ de notre démarche est l'utilisation des proportions des lithotypes par analogie avec les méthodes basées sur les gaussiennes seuillées. Classiquement, dans ces méthodes (Beucher et al., 1993), la répartition du lithotype étudié est donnée par la distribution 3D de sa proportion dans le volume considéré. Cette proportion est calculée sur une *matrice ou grille de proportion* en 3D et reproduit le comportement moyen, éventuellement non stationnaire, de ce lithotype.

4.1 Proportions

Les courbes de proportion constituent un outil très répandu dans la simulation et la description des environnements sédimentaires, que ce soient des dépôts fluviaux ou marins. Elles représentent la fraction sur un certain support de chaque lithofaciès présent dans l'unité d'étude. Ces courbes permettent de décrire, d'une façon assez intuitive, l'évolution des lithofaciès dans le réservoir. À l'origine elles ont été développées pour introduire dans les simulations géostatistiques une information relative à la distribution des hétérogénéités dans le réservoir en quantifiant la géologie. De plus, elles ont été appliquées à la description et l'interprétation géologique et sédimentaire des réservoirs, puisqu'elles apportent une analyse qualitative des séquences stratigraphiques en décrivant l'évolution latérale et verticale des dépôts (Volpi et al., 1997 ; Ravanne et al., 2002).

Dans Volpi et al. (1997) il est montré, à travers quelques exemples, l'importance des courbes de proportion pour quantifier la géologie et contraindre les simulations, ainsi que pour l'analyse haute résolution des séquences stratigraphiques à différents ordres. La stratigraphie séquentielle permet d'établir des hypothèses sur l'environnement de dépôt et de choisir les horizons chronostatigraphiques qui divisent les dépôts en séquences. Les proportions traduisent ces hypothèses en termes quantitatifs et aident à tester et contrôler celles-ci.

Les courbes de proportions permettent donc de résumer quantitativement la réalité géologique de l'environnement sédimentaire. L'importante charge d'information géologique contenue dans ces courbes nous a amené à vouloir utiliser cet outil dans les simulations de type objet, et plus particulièrement dans le modèle booléen. Pour cela il faut établir les rapports entre ces données expérimentales, qui reflètent la géologie du milieu, et les variables qui interviennent dans le modèle.

Il est nécessaire de distinguer deux termes : la **probabilité ponctuelle du modèle** (capacité de Choquet ponctuelle de l'ensemble booléen), c'est-à-dire la probabilité qu'un point appartienne au modèle booléen et donc au lithotype modélisé par cet ensemble, et la **proportion expérimentale du lithotype**, calculée sur un certain volume. Les rapports existant entre ces éléments font le sujet de la présente section.

4.1.1 Proportion expérimentale d'un faciès

Le processus suivi pour obtenir les proportions (courbes ou matrice) est le suivant (Galli et al., 1994 ; Beucher et al., 1997) : dans un premier temps, les lithofaciès et les unités stratigraphiques présents dans le réservoir sont identifiés aux puits grâce à l'étude des logs, et à d'autres informations telle que la sismique. Ces lithofaciès peuvent être éventuellement regroupés en plusieurs lithotypes. Puis, ces lithotypes, que l'on notera par F_i , sont transformés en indicatrices : $\mathbb{I}_{F_1}, \mathbb{I}_{F_2}, \dots$, avec :

$$\mathbb{I}_{F_i}(x) = \begin{cases} 1 & \text{si } x \in F_i \\ 0 & \text{sinon} \end{cases}$$

Théoriquement, la **proportion du lithotype F au point x** est définie par :

$$p_F(x) = E[\mathbb{I}_F(x)] = P\{x \in F\} \quad (4.1)$$

Le terme «proportion» fait ici référence à l'espace des réalisations et exprime la probabilité, sur toutes les réalisations possibles, pour que le point x soit recouvert par le faciès F . Dans la pratique, les proportions ponctuelles ne sont pas disponibles et on calcule les pourcentages, dans un volume donné, de chaque lithotype présent. Cette proportion expérimentale correspond à la régularisation, sur un support B , des proportions ponctuelles. La taille de ce support reflète l'échelle de travail et implique un lissage plus ou moins fort des proportions ponctuelles. Cette **proportion expérimentale**, que nous appelons simplement **proportion** par la suite et que l'on dénote par P_F , représente la «fraction de volume» du lithotype, et donne, en moyenne, la partie du volume du support B occupée par F .

$$P_F = \frac{1}{|B|} \int_B p_F(x) dx$$

On peut calculer deux types de proportions expérimentales (fig. 4.1) : soit le pourcentage de faciès présent est calculé niveau par niveau, ce qui donne des *courbes verticales* (CPV), soit le calcul est effectué puits par puits et les courbes obtenues sont *horizontales* (CPH). Les courbes horizontales sont utilisées pour l'analyse de la non stationnarité horizontale dans l'unité considérée. Les verticales décrivent le comportement des lithotypes dans la séquence étudiée.

Si le faciès est réparti de façon stationnaire, la probabilité ponctuelle est la même pour tous les points de l'espace et elle coïncide avec la proportion expérimentale du faciès dans le volume.

S'il existe une non stationnarité horizontale, il est possible de construire, à partir de courbes de proportion verticales régionales, une grille 3D de proportions, appelée **matrice ou grille de proportion**, et ainsi reproduire cette complexité géologique. Cette matrice apporte une information sur la variabilité latérale et verticale des proportions des lithotypes et permet l'obtention d'images assez réalistes dans les simulations.

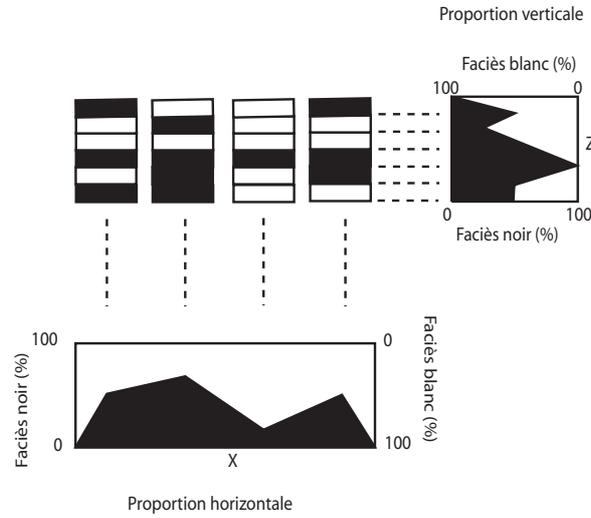


FIG. 4.1 – Exemple de calcul, pour deux faciès, d'une courbe de proportion verticale (CPV), calculée niveau par niveau verticalement, et d'une courbe de proportion horizontale (CPH), calculée puits par puits. Les informations aux puits correspondent aux indicatrices de chaque faciès à chaque niveau. Les valeurs de proportion entre les puits sont arbitrairement dessinées.

Dans Beucher et al. (1997) il est montré comment cette matrice de proportion 3D est construite et quel est son rôle dans les simulations géostatistiques basées sur les gaussiennes seuillées. Il y a deux approches principales pour construire la matrice des proportions : si le réservoir peut être considéré comme homogène latéralement par zones, alors la distribution du lithotype est décrite par une CPV par zone ; si, au contraire, il existe un passage latéral de faciès plus continu, la matrice de proportion est obtenue par *estimation*, et reproduit ainsi cette non stationnarité. Dans les deux situations, la matrice de proportion est construite sur une grille régulière dont chaque maille est informée par une courbe de proportion verticale. Il y a ici une idée d'homogénéité horizontale dans chacune de ces mailles : la proportion, pour un niveau donnée, est calculée au point central de chaque maille, et sa valeur est attribuée à tout le bloc ; elle est donc constante à l'intérieur de chaque bloc et représente la fraction de volume occupé par le lithotype présent dans le bloc.

4.1.2 Probabilité ponctuelle associée au modèle booléen : capacité de Choquet ponctuelle

Dans la section 3.2.1 nous avons vu une propriété fondamentale du modèle booléen qui permet d'établir une liaison entre la capacité de Choquet du modèle, T_X , la fonction d'intensité correspondant au processus de Poisson, θ , et la capacité de Choquet associée au grain primaire, T_u . Cette propriété, en termes de la fonctionnelle Q_X du modèle booléen, est exprimée par :

$$1 - T_X(K) = Q_X(K) = P\{K \subset X^c\} = \exp \left[- \int_{\mathbb{R}^d} \theta(u) T_u(K) du \right] \quad (4.2)$$

La méthode développée dans ce travail est basée sur cette propriété fondamentale quand K est réduit à un point ($K = \{x\}$). Cette méthode sera décrite dans le chapitre 5.

Les données expérimentales disponibles sont les caractéristiques des objets à simuler (leur forme et leur taille) et la matrice de proportion expérimentale du lithotype. Les premières sont introduites dans le modèle booléen à travers $T_u(\{x\})$. L'étude de ce rapport fait le sujet de la section 4.2. La matrice de proportion,

quant à elle, est liée à la capacité de Choquet du modèle, $T_X(\{x\})$, qui est une probabilité ponctuelle. Cette fonctionnelle donne, comme il a été déjà expliqué, la probabilité que le point x soit recouvert par l'ensemble booléen X . Elle est donc liée à la distribution du lithotype modélisé par X . Le problème à résoudre est *comment estimer ces probabilités ponctuelles à partir des proportions expérimentales*.

Seuls deux lithotypes sont considérés : celui modélisé par l'ensemble booléen X et le regroupement de tous les autres lithotypes présents, modélisé par X^c . Ainsi, nous pouvons définir, dans un cas ponctuel, la fonction suivante :

$$T_X(\{x\}) = P\{x \in X\} = E[\mathbb{I}_X(x)] = p(x) \quad (4.3)$$

Classiquement, dans la littérature (voir Matheron (1972), p. 52), $p(x)$ est appelée la «proportion du modèle booléen» au point x et ainsi elle a été présentée dans la section 3.2.2.2. Dans ce travail, nous réservons le terme proportion pour le contexte expérimental et la fonction $p(x)$ sera appelée simplement la **probabilité ponctuelle du modèle**.

Dans notre approche, si F est le lithotype modélisé par l'ensemble booléen X , pour un point x de l'espace nous avons :

$$p_F(x) = P\{x \in F\} = P\{x \in X\} = T_X(\{x\}) = p(x)$$

Alors, la probabilité au point x du lithotype F est égale à la probabilité du modèle au point x . Cependant, nous avons vu que la donnée disponible n'est pas p_F , mais une valeur estimée sur un support, P_F . Dans le cadre du modèle, cette proportion représente, en moyenne, la fraction de l'espace occupée par l'ensemble booléen, et donc par F . Si l'on considère une région de l'espace B , de volume $|B|$, la fraction de volume occupée par le lithotype est :

$$\begin{aligned} P_F &= \frac{E[|X \cap B|]}{|B|} = \frac{1}{|B|} E \left[\int_{\mathbb{R}^d} \mathbb{I}_{X \cap B}(x) dx \right] \\ &= \frac{1}{|B|} E \left[\int_B \mathbb{I}_X(x) dx \right] \\ &= \frac{1}{|B|} \int_B E[\mathbb{I}_X(x)] dx \\ &= \frac{1}{|B|} \int_B P\{x \in X\} dx \\ &= \frac{1}{|B|} \int_B T_X(\{x\}) dx \\ &= \frac{1}{|B|} \int_B p(x) dx \end{aligned} \quad (4.4)$$

S'il y a stationnarité, la probabilité qu'un point appartienne à X est la même pour tout point et pour tout B et la fraction de volume occupée par F est égale à la probabilité ponctuelle du modèle dans un point quelconque :

$$P_F = p(x) = cte \quad \forall x \quad (4.5)$$

Si, au contraire, la situation est non stationnaire, $p(x)$ n'est plus constante et l'expression (4.4) dépend de la région B :

$$P_F |_{B=} = \frac{1}{|B|} \int_B p(x) dx \quad (4.6)$$

En d'autres termes, l'information disponible expérimentalement est la fraction de volume occupée par le faciès F dans une région B de l'espace. Cette fraction de volume peut être interprétée comme la moyenne, sur cette région, de la capacité de Choquet ponctuelle du modèle. En utilisant $P_F|_B$ comme une estimation de la moyenne de $T_X(\{x\})$ nous effectuons déjà un premier lissage des probabilités. La figure 4.2 montre cet effet.

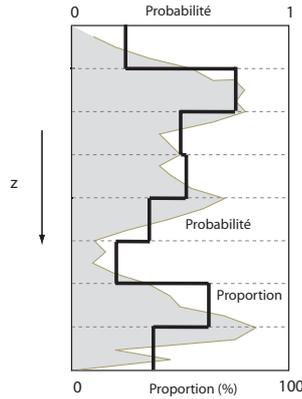


FIG. 4.2 – Lien entre la proportion expérimentale (ligne noir gras) et la probabilité ponctuelle du modèle (gris). La première est une moyenne de la seconde sur un support.

Ici apparaît la différence entre probabilité et proportion. Le fait que la proportion prenne une valeur constante à l'intérieur du support B n'implique pas que la probabilité pour tout point de B d'appartenir à l'ensemble booléen soit identique. De plus, il y a un effet de «discontinuité» dans les bords des mailles de proportion qui n'est pas forcément représentative du comportement de la probabilité. Du point de vue pratique, il se pose aussi le problème de la représentativité des proportions dans la région d'étude et l'intervalle de confiance à lui accorder, étant donné le nombre souvent limité de données disponibles pour calculer cette proportion.

4.1.2.1 Hypothèses expérimentales : «stationnarité par bloc».

Les valeurs expérimentales de p sont donc obtenues par estimation des valeurs régularisées sur un certain support. Les données, les proportions du lithotype étudié, sont estimées sur une grille régulière et prennent des valeurs constantes sur chacun des blocs de cette grille, C_i . C'est-à-dire :

$$P_F|_{C_i} = p|_{C_i}$$

Nous associons cette valeur au point central x_i du bloc C_i . Comme nous l'avons précisé auparavant, en pratique ceci implique une certaine «stationnarité par bloc». L'intensité du processus de Poisson correspondante sera donc aussi constante par bloc : $\theta(C_i) = \theta_i$. Avec cette hypothèse, il serait tentant de considérer chaque bloc indépendamment comme un modèle stationnaire :

$$p|_{C_i} = 1 - e^{-\theta_i E[A]}$$

Prenons C_i le bloc i de la grille des proportions et supposons pour l'instant qu'il soit un ensemble compact (les bords de la maille appartiennent à C_i)¹. Selon les propriétés de stabilité des ensembles booléens, l'intersection entre un modèle booléen et un sous-ensemble compact de \mathbb{R}^d est aussi un modèle booléen. Nous pouvons alors regarder la fonctionnelle Q associée à $X \cap C_i$ (voir Lantuéjoul, 2002, p.156) :

$$\begin{aligned} P\{(X \cap C_i) \cap K = \emptyset\} &= P\{X \cap (C_i \cap K) = \emptyset\} = \\ &= \exp\left(-\int_{\mathbb{R}^d} \theta(u) T_u(C_i \cap K) du\right) = \\ &= \exp\left(-\int_{\mathbb{R}^d} \theta(u) T_u(C_i) \frac{T_u(C_i \cap K)}{T_u(C_i)} du\right) \end{aligned} \quad (4.7)$$

Le terme $\frac{T_u(C_i \cap K)}{T_u(C_i)}$ est la capacité de Choquet de l'«objet» $A(u) \cap C_i$, en supposant que cette intersection soit non vide. En effet :

$$\begin{aligned} P\{(A(u) \cap C_i) \cap K \neq \emptyset | A(u) \cap C_i \neq \emptyset\} &= \frac{P\{A(u) \cap (C_i \cap K) \neq \emptyset\}}{P\{A(u) \cap C_i \neq \emptyset\}} \\ &= \frac{T_u(C_i \cap K)}{T_u(C_i)} \end{aligned}$$

Alors, $X \cap C_i$ est un ensemble booléen de densité $\theta(u) T_u(C_i)$ et $\frac{T_u(C_i \cap K)}{T_u(C_i)}$ est la capacité de Choquet de l'«objet» $A(u) \cap C_i$. Le nombre d'objets de $X \cap C_i$ suit alors une distribution de Poisson d'espérance :

$$\begin{aligned} \vartheta(C_i) &= \int_{\mathbb{R}^d} \theta(u) T_u(C_i) du \\ &= \underbrace{\theta_i \int_{C_i} T_u(C_i) du}_{(a)} + \underbrace{\int_{\mathbb{R}^d \setminus C_i} \theta(u) T_u(C_i) du}_{(b)} \end{aligned}$$

Le terme (a) représente le nombre moyen d'objets dont les germes tombent dans le bloc C_i ($u \in C_i$) et sont distribués selon l'intensité θ_i . Le terme (b) tient compte des objets intersectant C_i qui germent dans les blocs adjacents ($u \in \mathbb{R}^d \setminus C_i$, i.e., $u \in \mathbb{R}^d - \{C_i\}$) avec une intensité qui dépend de la position du point de germination (figure 4.3). Il n'est donc pas possible de traiter chaque cellule indépendamment. Il faut procéder autrement pour estimer θ , ce qui fait le sujet de ce travail.

4.2 Probabilité ponctuelle associée à l'objet

Le second paramètre du modèle à définir est lié aux objets qui vont modéliser le faciès étudié. L'information disponible pour définir ces objets provient de différentes sources : géologiques, sismiques, variographiques ... Ces informations donnent des estimations, parfois avec une grande incertitude, de la taille et de la forme des objets qui vont modéliser le faciès étudié. La forme de ces objets est souvent suggérée par les structures sédimentaires elles-mêmes : les chenaux, les crevasses, par exemple, présentent, schématiquement, des formes géométriques relativement bien définies. Leurs dimensions doivent être estimées à

¹En toute rigueur les cellules C_i ne sont pas des compacts. Cela voudrait dire que leurs bords appartiennent à deux blocs consécutifs : $C_i \cap C_j \neq \emptyset$. Il y aurait donc des objets générés dans un bord de C_i , avec l'intensité θ_i , mais en même temps il y aurait des objets implantés avec l'intensité θ_j . Bien que cela ne pose pas un problème majeur puisque la probabilité qu'un germe tombe sur un bord (de volume nul) est infime, du point de vue pratique il faudra établir une règle d'attribution des bords aux cellules.

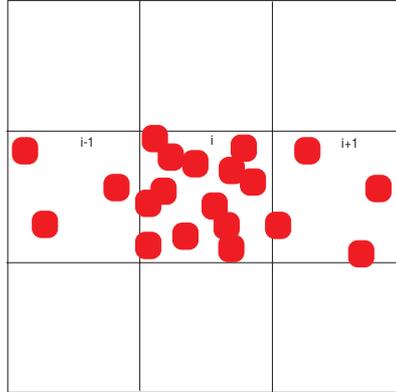


FIG. 4.3 – L'intensité de Poisson dans le bloc i est clairement plus importante que dans les blocs autour : il y a plus d'objets qui y ont germé. Cependant, la proportion du faciès correspondant à l'intérieur de ce bloc dépend non seulement des objets germés dans le bloc i avec l'intensité θ_i , mais aussi des objets qui ont germé dans les blocs $i - 1$ et $i + 1$ (et, en général, dans tous les blocs autour du bloc i) avec leurs intensités respectives et qui intersectent le bloc i .

partir des données disponibles. Les objets choisis dans le modèle, bien que respectant la forme induite par la structure sédimentaire, restent de formes relativement simples.

Une fois la taille et la forme des objets choisies, il est nécessaire de définir ceux-ci en termes de paramètres du modèle. La fonctionnelle $T_u(\{x\})$ exprime la probabilité qu'un grain implanté en $u \in \mathbb{R}^d$, recouvre le point «cible» x . Puisque les objets sont indépendamment et identiquement distribués, la fonctionnelle d'un objet implanté en u est égale à la fonctionnelle de l'objet implanté à l'origine et déplacé en u :

$$\begin{aligned} T_u(\{x\}) &= P\{A(u) \cap \{x\} \neq \emptyset\} = P\{A_u(0) \cap \{x\} \neq \emptyset\} \\ &= P\{A(0) \cap \{x\}_{-u} \neq \emptyset\} = T_0(\{x\}_{-u}) \\ &= P\{x - u \in A(0)\} = p_0(x - u) \end{aligned}$$

Le terme p_0 est dénoté par la suite comme la **probabilité ponctuelle associée au grain primaire** et dépend de la forme et la taille des objets.

Nous cherchons donc à calculer la probabilité qu'un point de l'espace tombe à l'intérieur d'un objet implanté à l'origine : $p_0(x) = T_0(\{x\})$. Un objet général est un ensemble aléatoire dont la taille et la forme sont tirées au hasard. Dans notre cas particulier, les objets appartiennent à une même famille (cercles, rectangles, demi-ellipsoïdes, etc.) dont les dimensions peuvent varier aléatoirement. Par la suite, nous illustrons en détail comment ce paramètre est calculé en pratique.

Dans un cas très simple, si la famille des objets à simuler est, à 2D, constituée de *cercles* C de rayon R constant (figure 4.4), la probabilité qu'un point $M(x, y)$ soit recouvert par un objet, $p_0(x, y)$ est tout simplement donnée par :

$$p_0(x, y) = \mathbb{I}_C(x, y) = \begin{cases} 1 & \text{si } M \in C; \text{ i.e. si } d_o(x, y) < R \\ 0 & \text{sinon} \end{cases}$$

où o représente le centre du cercle (origine) que nous avons fait coïncider avec le point d'implantation du grain primaire, et $d_o(x, y) = d_o$ est la distance entre le point M et l'origine. Cette définition implique déjà deux choses : une orientation du grain, dans le sens de l'emplacement du germe par rapport à l'objet, et l'introduction de la distance entre le point cible et la position du germe comme paramètre de calcul de la probabilité.

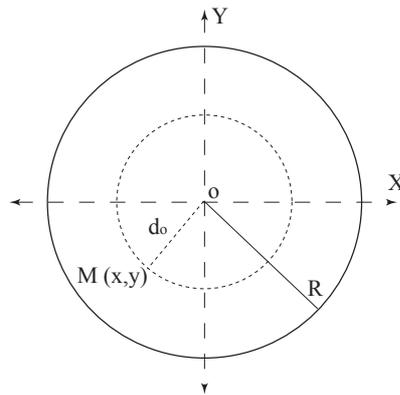
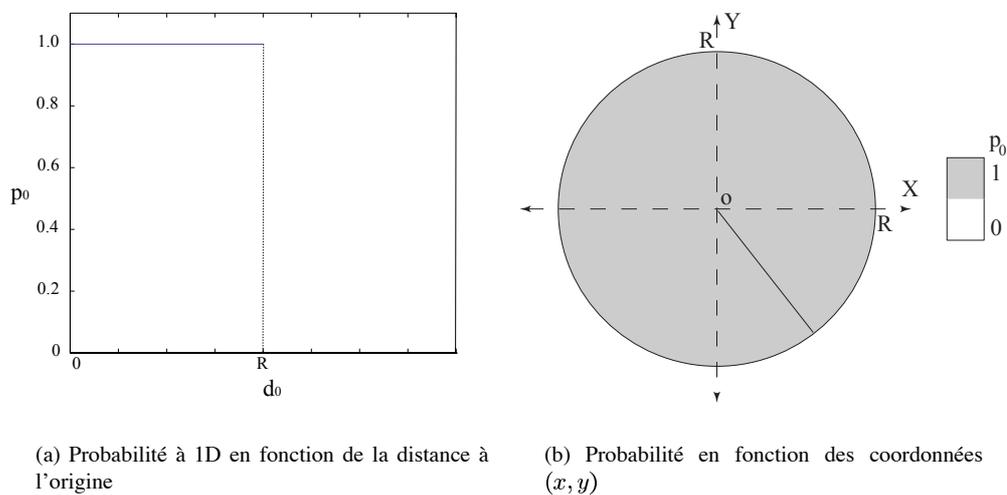


FIG. 4.4 – Représentation des objets à simuler. Tous les cercles sont de rayon R constant.

La figure 4.5(a) montre graphiquement cette probabilité en fonction de la distance au point de germination de l'objet. La figure 4.5(b) montre cette même probabilité en fonction des coordonnées x et y .



(a) Probabilité à 1D en fonction de la distance à l'origine

(b) Probabilité en fonction des coordonnées (x, y)

FIG. 4.5 – Représentations de la probabilité associée aux cercles de rayon R constant.

Supposons maintenant que les *cercles* ne soient pas de taille constante mais de *taille r variable* de façon aléatoire et uniforme entre R_1 et R_2 (fig. 4.6). La probabilité qu'un point $M(x, y)$ appartienne au cercle de rayon r fixé est :

$$p_0(x, y) = P\{M \in C_r\} = \mathbb{I}_{C_r}(x, y) = \begin{cases} 1 & \text{si } d_o < r \\ 0 & \text{sinon} \end{cases}$$

avec d_o la distance du point à l'origine. Si on rend aléatoire r , l'expression précédente s'exprime comme :

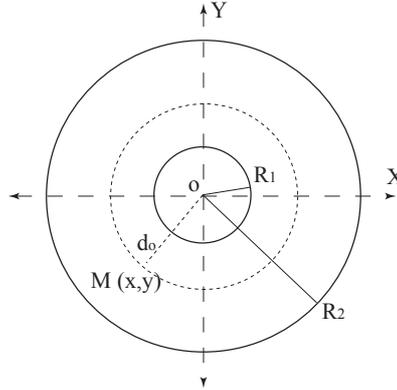


FIG. 4.6 – Représentation des objets à simuler. Le rayon du cercle varie uniformément entre R_1 (cercle minimal autorisé) et R_2 (cercle maximal).

$$p_0(x, y) = \int_{R_1}^{R_2} P\{d_o < r\} f(r) dr \quad (4.8)$$

Le terme $f(r)$ correspond à la probabilité que le cercle soit de rayon r . Il faut alors intégrer pour toutes les valeurs possibles de r . Si r suit une loi uniforme, $r \sim U_{[R_1, R_2]}$, le terme $f(r)$ est :

$$f(r) = \frac{1}{R_2 - R_1}$$

Si le point x se trouve à une distance plus petite que le rayon mineur R_1 , alors la probabilité que x soit recouvert par un cercle de taille quelconque est égale à 1 : tous les cercles possibles vont le recouvrir. De la même façon, si x se trouve à une distance plus grande que le rayon maximum autorisé R_2 , alors $p_0(x) = 0$. Dans le cas où x se trouve à une distance intermédiaire entre R_1 et R_2 , la probabilité d'être recouvert par un grain sera de plus en plus faible à mesure que cette distance augmente : les cercles de rayon plus petit que cette distance ne le recouvriront pas. Nous pouvons exprimer ces trois situations possibles comme suit :

- si $d_o > R_2$ alors $p_0(x) = 0$;
- si $d_o \leq R_1$ alors

$$p_0(x, y) = \int_{R_1}^{R_2} 1 \cdot \frac{1}{R_2 - R_1} dr = 1$$

- si $R_1 < d_o \leq R_2$ alors

$$p_0(x, y) = \int_{d_o}^{R_2} 1 \cdot \frac{1}{R_2 - R_1} dr = \frac{R_2 - d_o}{R_2 - R_1}$$

La figure 4.7 montre la probabilité correspondant à ce cas. Nous observons que cette probabilité n'est plus constante mais qu'elle dépend du point d'implantation du grain, de l'origine, et de la loi des paramètres de l'objet. Ces expressions sont similaires à 1D (cas d'un segment) ou 3D (cas d'une boule).

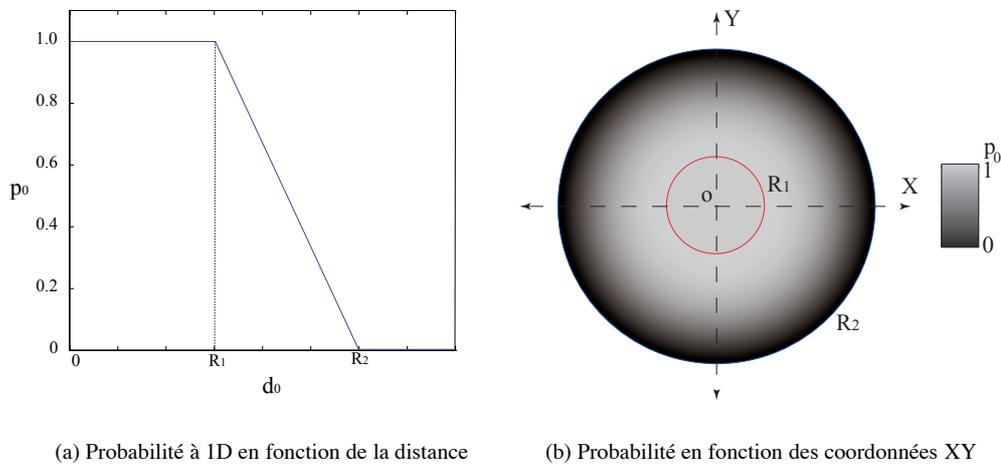


FIG. 4.7 – Représentations de la probabilité associée aux cercles de rayon R variable uniformément entre R_1 et R_2 .

L'exemple présenté ici (une famille de cercles de taille aléatoire suivant une loi uniforme) est assez élémentaire. Il est possible d'envisager des objets avec une plus grande complexité concernant plusieurs aspects :

1. Rendre aléatoire la forme des objets : la forme des grains implantés deviendrait une variable aléatoire dont il faudrait déterminer la loi. Ceci introduirait un nouveau terme à l'expression donnée par 4.8.
2. Changer la loi de la taille des objets : au lieu de suivre une loi uniforme, elle pourrait suivre une loi normale, binomiale, etc.
3. Introduire des objets de forme arbitraire : en général, la forme des grains pourrait ne pas s'exprimer de façon paramétrique, comme dans le cas d'une forme géométrique élémentaire. Dans ce cas, il ne serait pas possible d'exprimer l'événement « x est à l'intérieur du grain» analytiquement, comme dans le cas du cercle. Il faudrait donc obtenir cette probabilité numériquement.

En général, l'expression 4.8 ne sera calculable analytiquement que dans certains cas. Nous verrons dans la section 5.2.2 comment il est possible de calculer une probabilité p_0 pour une forme relativement simple et de «corriger» ultérieurement pour introduire des grains plus complexes.

4.2.1 Deux cas élémentaires : parallélépipède et demi-ellipse

Les deux cas élémentaires que nous allons considérer au cours de ce travail sont des parallélépipèdes (à 3D) de taille variable et des demi-ellipses (à 2D) de taille variable. La première de ces figures est choisie parce qu'elle est un objet très élémentaire et sera utilisée dans un premier temps pour tester la méthode proposée. Le deuxième objet, à 2D, est choisi parce qu'il correspond, de manière simplifiée, à la section verticale d'un objet chenal.

Parallélépipèdes

La figure 4.8 montre l'objet à simuler. Nous avons choisi comme point de germination o , origine du système, le point central de la surface supérieure du parallélépipède D . Les demi-longueurs selon les directions

X , Y et la longueur selon la direction Z prennent des valeurs aléatoires indépendamment les unes des autres suivant les lois uniformes :

$$\begin{aligned} l_x &\sim U_{[a,b]} \\ l_y &\sim U_{[c,d]} \\ l_z &\sim U_{[e,f]} \end{aligned}$$

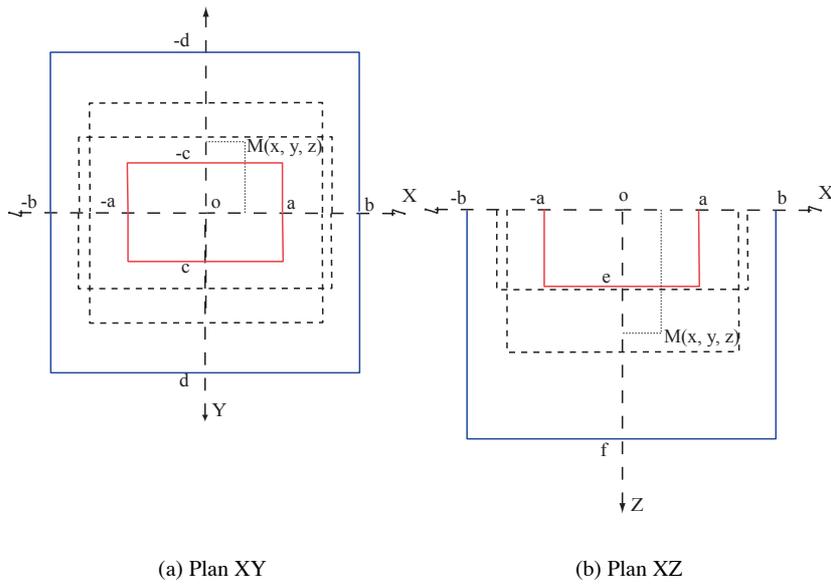


FIG. 4.8 – Sections XY et XZ de parallélépipèdes possibles. En rouge, le plus petit des objets possibles et en bleu le plus grand.

La probabilité qu'un point de coordonnées (x, y, z) soit recouvert par un objet D de dimensions $(2l_x, 2l_y, l_z)$ est :

$$p_0(x, y, z) = P\{(x, y, z) \in D\} = \mathbb{I}_D(x, y, z)$$

Autrement dit, un point (x, y, z) est à l'intérieur de l'objet D si :

$$\begin{aligned} x &\in [-l_x, l_x] \\ y &\in [-l_y, l_y] \\ z &\in [0, l_z] \end{aligned}$$

Ces événements sont, par construction, indépendants entre eux. Alors, $p_0(x, y, z)$ peut s'exprimer comme le produit des probabilités individuelles de chaque coordonnée. Nous pouvons suivre la même démarche que dans l'exemple du cercle. Il y a trois cas :

- le cas général :

$$p_0(x, y, z) = P\{x \in [-l_x, l_x]\}P\{y \in [-l_y, l_y]\}P\{z \in [0, l_z]\}$$

et les cas limites :

- si $|x| \leq a$ et $|y| \leq c$ et $|z| \leq e$ alors $p_0(x) = 1$;
- si $|x| > b$ et/ou $|y| > d$ et/ou $|z| > f$ alors $p_0(x) = 0$;

Étant données les symétries de la figure, nous pouvons restreindre l'étude à une seule coordonnée et aux axes de coordonnées positives. La probabilité dans les autres régions de l'espace s'obtient de façon totalement analogue.

- Si $x > 0$:
 - si $x \leq a$: $P\{x \in [0, l_x]\} = 1$;
 - si $x > b$: $P\{x \in [0, l_x]\} = 0$;
 - dans les cas restants, en appliquant l'expression 4.8 :

$$P\{x \in [0, l_x]\} = \int_{l_x}^b \mathbb{I}_{[0, l_x]}(x) \cdot \frac{1}{b-a} \cdot \mathbb{I}_{[a, b]}(l_x) dl_x = \frac{b-l_x}{b-a}$$

La probabilité pour le point de coordonnée x soit recouvert par l'objet est la même que la probabilité que le point se trouve à une distance l_x de l'origine. L'expression précédente peut alors s'écrire :

$$P\{x \in [0, l_x]\} = \frac{b-x}{b-a}$$

Le processus est similaire pour les coordonnées y et z . Le résultat est :

- Si $x > 0, y > 0, z > 0$:

$$p_0(x, y, z) = \frac{b-x}{b-a} \cdot \frac{d-y}{d-c} \cdot \frac{e-z}{f-e}$$

La figure 4.9 montre une représentation graphique des valeurs que cette probabilité prend pour un cas simple où : $l_x \sim U_{[2m, 4m]}$, $l_y \sim U_{[1m, 3m]}$ et $l_z \sim U_{[1m, 2m]}$.

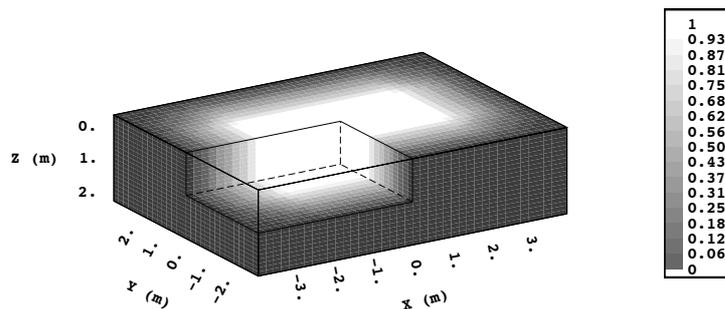


FIG. 4.9 – Probabilité associée à un objet parallélépipède de demi-côtés variant uniformément : $l_x \sim U_{[2m, 4m]}$, $l_y \sim U_{[1m, 3m]}$ et $l_z \sim U_{[1m, 2m]}$.

Demi-ellipses

Supposons des demi-ellipses dans le plan XZ de demi-axes a et b , $E(a, b)$. Ces axes seront tirés de façon indépendante selon les lois uniformes :

$$\begin{aligned} a &\sim U_{[a_0, a_1]} \\ b &\sim U_{[b_0, b_1]} \end{aligned}$$

La figure 4.10 montre les demi-ellipses possibles avec ce choix de paramètres. Le point de germination o a été choisie comme étant le point central des ellipses.

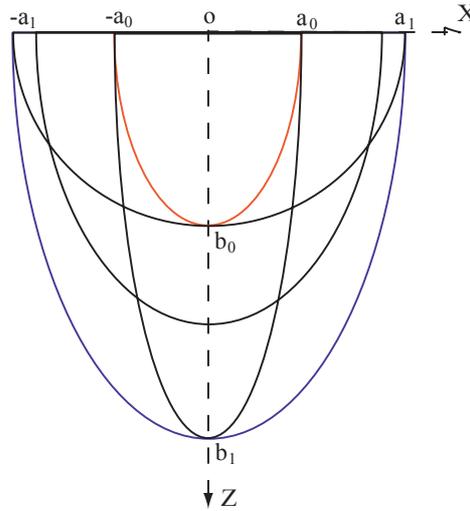


FIG. 4.10 – Exemple des demi-ellipses possibles. En rouge la plus petite et en bleu la plus grande des demi-ellipses autorisées.

Comme toujours, la probabilité qu'un point (x, z) du plan soit recouvert par une demi-ellipse E de paramètres (a, b) est :

$$p_0(x, z) = \begin{cases} 1 & \text{si } (x, z) \in E(a, b) \\ 0 & \text{sinon} \end{cases}$$

C'est-à-dire, le point (x, z) est à l'intérieur de E si :

$$x \in [-a, a], \quad z \in [-b, b] \quad \text{et} \quad \frac{x^2}{a^2} + \frac{z^2}{b^2} \leq 1$$

À la différence du cas précédent, l'événement $(x, z) \in E(a, b)$ ne dépend pas uniquement des probabilités des coordonnées x et z , mais aussi de la relation de dépendance entre elles. Pour le calcul de la fonction de probabilité associée à cet objet, il faut distinguer plusieurs cas selon la position du point considéré, $M(x, z)$ (fig. 4.11). Par symétrie, nous ne considérons que le cas où les coordonnées de M sont positives, le calcul étant similaire pour les cas restants.

Nous pouvons distinguer dans la figure 4.11 six cas possibles. Cette division est introduite par le fait qu'il existe des combinaisons de valeurs des demi-axes pour lesquelles les demi-ellipses résultantes ne sont pas autorisées. L'expression générale pour $p_0(x, z)$ est :

$$p_0(x, z) = \int_a \int_b 1 \cdot \frac{1}{a_1 - a_0} \cdot \frac{1}{b_1 - b_0} da db$$

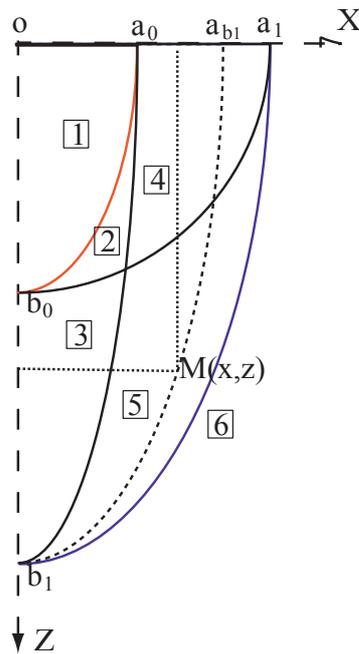


FIG. 4.11 – Différentes régions à étudier selon la position du point $M(x, z)$. Les régions sont délimitées par les courbes en trait continu : la région 1 est délimitée par $E(a_0, b_0)$; la région 2 par $E(a_0, b_0)$, $E(a_1, b_0)$ et $E(a_0, b_1)$; la région 3 par $E(a_1, b_0)$ et $E(a_0, b_1)$; la région 4 par $E(a_0, b_1)$ et $E(a_1, b_0)$; la région 5 par $E(a_1, b_0)$, $E(a_0, b_1)$ et $E(a_1, b_1)$ et la région 6 correspond à tout l'espace extérieur à $E(a_1, b_1)$.

Selon la position du point $M(x, z)$, la probabilité qu'il soit recouvert par une demi-ellipse est donnée par les cas suivants :

1. Si $\frac{x^2}{a_1^2} + \frac{z^2}{b_1^2} > 1$ (zone 6), aucune demi-ellipse ne recouvre M et :

$$p_0(x, z) = 0$$

2. Si $\frac{x^2}{a_0^2} + \frac{z^2}{b_0^2} \leq 1$ (zone 1), toutes les demi-ellipses recouvrent M et :

$$p_0(x, z) = 1$$

3. Si $\frac{x^2}{a_1^2} + \frac{z^2}{b_0^2} \leq 1$ et $\frac{x^2}{a_0^2} + \frac{z^2}{b_1^2} \leq 1$ (zone 2) :

$$p_0(x, z) = \frac{a_1 b_1 - a_1 b_0 - a_0 b_1 + z \sqrt{a_0^2 - x^2} + x \sqrt{b_0^2 - z^2}}{(a_1 - a_0)(b_1 - b_0)}$$

4. Si $\frac{x^2}{a_1^2} + \frac{z^2}{b_0^2} > 1$ et $\frac{x^2}{a_0^2} + \frac{z^2}{b_1^2} \leq 1$ (zone 3) :

$$p_0(x, z) = \frac{b_1(a_1 - a_0) - z(\sqrt{a_1^2 - x^2} - \sqrt{a_0^2 - x^2})}{(a_1 - a_0)(b_1 - b_0)}$$

5. Si $\frac{x^2}{a_1^2} + \frac{z^2}{b_0^2} \leq 1$ et $\frac{x^2}{a_0^2} + \frac{z^2}{b_1^2} > 1$ (zone 4) :

$$p_0(x, z) = \frac{a_1(b_1 - b_0) - x(\sqrt{b_1^2 - z^2} - \sqrt{b_0^2 - z^2})}{(a_1 - a_0)(b_1 - b_0)}$$

6. Si $\frac{x^2}{a_1^2} + \frac{z^2}{b_0^2} > 1$ et $\frac{x^2}{a_0^2} + \frac{z^2}{b_1^2} > 1$ (zone 5) :

$$p_0(x, z) = \frac{a_1 b_1 - z \sqrt{a_1^2 - x^2} - x \sqrt{b_1^2 - z^2}}{(a_1 - a_0)(b_1 - b_0)}$$

Naturellement, les cas 3 à 6 doivent aussi vérifier que $\frac{x^2}{a_0^2} + \frac{z^2}{b_0^2} > 1$ et $\frac{x^2}{a_1^2} + \frac{z^2}{b_1^2} \leq 1$. Les détails du calcul sont présentés en annexe A. Nous pouvons constater que cette expression est notamment plus compliquée que celle du parallélépipède du fait de la dépendance entre les coordonnées. Ceci met en évidence la difficulté à définir analytiquement la probabilité associée à des objets plus complexes.

La figure 4.12 montre cette fonction de probabilité graphiquement pour le cas particulier $a \sim U_{[2m, 4m]}$ et $b \sim U_{[1m, 2m]}$.

Toutes les probabilités associées aux objets présentés ici sont calculées de façon continue. En pratique, nous sommes dans un cadre discret : une fois la probabilité obtenue, il est nécessaire de prendre des valeurs discrètes correspondant aux points centraux de la grille de travail. La figure 4.13 montre l'effet de ce processus de discrétisation. La courbe bleue continue montre un exemple d'une fonction possible de probabilité associée à l'objet. Plus la maille de discrétisation est large et plus la précision dans la définition de l'objet sera faible. Ce point sera important dans la mise en œuvre de la méthode d'estimation de l'intensité que nous proposons et sera traité plus en détail dans la section 5.2.3.

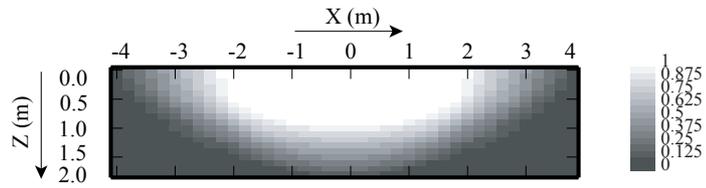


FIG. 4.12 – Exemple de probabilité associée à un objet demi-ellipse d'axes de taille variable selon des lois uniformes : $a \sim U_{[2m,4m]}$ et $b \sim U_{[1m,2m]}$.

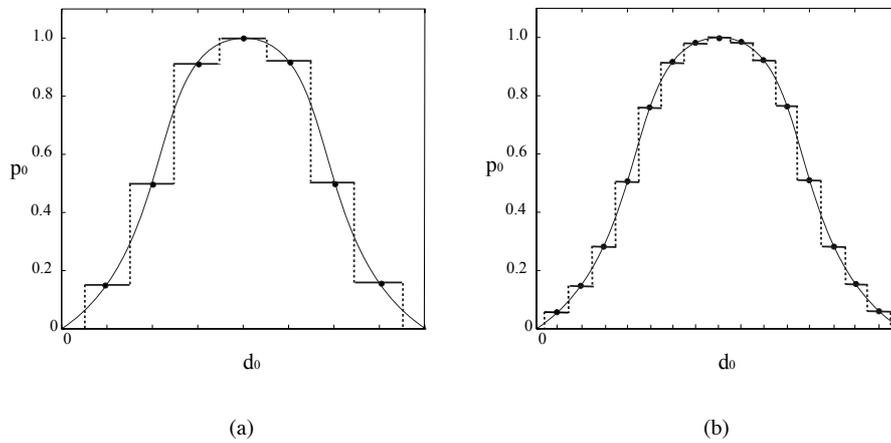


FIG. 4.13 – Exemples de l'effet de discrétisation. En continu une fonction de probabilité ; en gras et pointillé la discrétisation réalisée.

4.3 Conclusions

Dans la caractérisation des réservoirs, il est très fréquent de rencontrer des phénomènes non stationnaires, ainsi la distribution des sables et argiles à l'intérieur du réservoir est, le plus souvent, non homogène. Cette non stationnarité est très rapidement mise en évidence dans l'étude des proportions de ces faciès. En raison de l'importante charge d'information géologique contenue dans les proportions, il nous a semblé assez intuitif de les exploiter dans le cas du modèle booléen. Ainsi les proportions sont utilisées comme l'information expérimentale à partir de laquelle les paramètres du modèle seront estimés.

Pour cela il faut d'abord établir le rapport entre la proportion expérimentale du faciès qui sera modélisé par l'ensemble booléen, et la probabilité associée au modèle booléen. Il est possible de considérer la première comme une régularisation ou valeur moyenne de la seconde sur un certain support. Ainsi, à chaque bloc de la grille correspond une valeur constante de la proportion. La première approche que nous pourrions suivre serait de considérer chaque bloc comme un modèle booléen stationnaire et de le traiter indépendamment des autres. Mais ceci mène à de faux résultats, puisque la valeur de la proportion dans un bloc ne tient pas compte uniquement des objets qui y germent, mais aussi des objets qui germent dans les blocs adjacents. La méthode que nous proposons pour étudier le modèle booléen dans ce contexte tient compte de cet effet de bord.

Le fait d'utiliser les proportions comme information dans le modèle introduit deux effets liés :

- Implicitement, la probabilité pour tout point d'un bloc de la grille d'appartenir au modèle booléen est considérée comme constante. Ceci est faux, puisque à l'intérieur d'un même bloc les objets peuvent être distribués non uniformément, mais cet effet à petite échelle est masqué par le lissage introduit lors du calcul de la proportion. Le problème est que le plus souvent la matrice de proportion est calculée à partir de très peu de données, ce qui impose une contrainte très forte sur la région de représentativité des proportions.
- Le fait d'avoir des valeurs de proportion constantes sur chaque bloc de la grille introduit des discontinuités parfois importantes entre un bloc et ses voisins. Ces discontinuités ne sont pas forcément représentatives de la réalité puisque la variation dans la présence d'un faciès est le plus souvent graduelle. Naturellement, il est possible de rencontrer des situations extrêmes, comme la présence d'une faille, où cette variation est réellement très brusque. Il faut alors établir jusqu'à quel point nous pouvons considérer les proportions comme indicatives du comportement du faciès.

Quant au paramètre du modèle relatif aux objets, p_0 , il est dépendant de la forme et la taille de l'objet choisi. Les exemples présentés ont donné un aperçu de la complexité de ce calcul même dans un cas très simple (voir les demi-ellipsoïdes). Si les objets deviennent plus réalistes, et donc plus complexes, il se peut qu'il soit même impossible de calculer une expression analytique de p_0 . Ce problème sera traité dans le chapitre suivant.

Chapitre 5

Inférence de l'intensité de Poisson

L'**inférence** consiste à obtenir les paramètres d'un modèle à partir des mesures expérimentales. Nous avons choisi d'utiliser les proportions des faciès comme information pour déterminer les deux paramètres du modèle booléen, les caractéristiques des objets et l'intensité de Poisson. Dans le chapitre suivant nous verrons quelle information sur les objets peut être extraite des données. Le présent chapitre a pour but l'obtention de l'intensité de Poisson en supposant que les objets sont connus.

Dans ce travail, la non stationnarité étudiée sera reproduite par une distribution spatiale régionalisée des objets appartenant à une même famille, dont la loi restera constante dans toute la région. Le choix de la loi des objets est primordial dans ce problème puisqu'il est possible de retrouver la même distribution de proportion avec peu d'objets de grand taille ou avec un nombre important d'objets petits. En effet, du fait que la loi spatiale de l'ensemble booléen est connue seulement en partie et qu'elle ne définit pas complètement l'ensemble aléatoire fermé (section 1.2.2), il existe plusieurs couples intensité - objet qui peuvent aboutir au même résultat, le but final étant de restituer les proportions du faciès étudié dans les simulations. Comme il a été déjà précisé, seulement deux faciès, celui modélisé par le modèle booléen, et son complémentaire seront considérés.

5.1 Méthode de déconvolution pour l'obtention de l'intensité

La méthode développée dans ce travail a été initialement proposée à 1D verticalement par Schmitt et Beucher (1997). Elle est basée sur la propriété fondamentale du modèle booléen (section 3.2.1) qui lie la capacité de Choquet de l'ensemble booléen à l'intensité de Poisson et à la capacité de Choquet du grain primaire (eq. 3.4)). Si l'ensemble compact K est réduit à un point, la capacité de Choquet du modèle et du grain deviennent des probabilités ponctuelles. Sous cette condition et si les objets sont indépendants de leur position, l'équation fondamentale peut s'exprimer sous forme d'une convolution :

$$p(x) = 1 - e^{[\theta * p_0](x)} \quad (5.1)$$

Nous avons établi dans la section 4.1 le lien entre les données disponibles, la proportion du faciès à modéliser, et la probabilité ponctuelle associée au modèle. Si, en plus des proportions, les caractéristiques des objets et donc p_0 sont connues, il est alors possible d'obtenir l'intensité de Poisson du modèle à l'aide d'un processus de déconvolution. Nous proposons, pour résoudre ce problème, une méthode classique en traitement du signal, le filtre de Wiener, qui est détaillée par la suite.

5.1.1 Filtre de déconvolution : filtre de Wiener

L'intensité peut être inférée à partir des proportions par une procédure de déconvolution du logarithme de $p(x)$ et d'une certaine fonction inverse de $p_0(x)$. Le processus de déconvolution est un problème courant dans le domaine de la théorie du signal. Nous nous servons de techniques très élémentaires de ce domaine pour résoudre notre problème.

L'opération de déconvolution est difficile à réaliser dans le domaine réel, mais en appliquant le *théorème de convolution* (voir annexe B) elle est immédiate dans l'espace des fréquences. Il faut donc travailler dans le domaine des transformées de Fourier. Les transformées de Fourier ne sont, en définitive, que des transformations des variables en termes de fréquences (annexe C). Dans cette approche, les variables impliquées sont considérées comme le résultat d'une superposition des fonctions sinusoïdales de différentes fréquences, dont chacune d'elles transmet une partie de l'information relative à la variable en question. Les transformées de Fourier donnent la décomposition de ces variables en leurs fréquences caractéristiques. Typiquement, les composantes de basses fréquences transmettent l'information relative à la grande échelle, comme la forme générale de la courbe de proportions, et les composantes de haute fréquence transmettent l'information relative aux petites échelles.

Dans l'équation que nous essayons de résoudre (eq. 5.1) les données sont p et p_0 et l'inconnue est θ . Par souci de simplification nous réalisons un changement de variable :

$$s(x) = -\ln(1 - p(x)) = [\theta * p_0](x) \quad (5.2)$$

Il faut alors chercher une fonction f telle que, appliquée à la convolution, elle rende l'intensité cherchée :

$$[f * s](x) = [f * \theta * p_0](x) \quad (5.3)$$

En prenant la transformée de Fourier (notée TF) de l'équation (5.3) :

$$[f * s](x) = [f * \theta * p_0](x) \xrightarrow{TF} \hat{f}(\nu) \cdot \hat{s}(\nu) = \hat{f}(\nu) \cdot \hat{\theta}(\nu) \cdot \hat{p}_0(\nu)$$

Dans le domaine des fréquences, la fonction cherchée est :

$$\hat{f}(\nu) = \frac{1}{\hat{p}_0(\nu)} \quad (5.4)$$

appelée le *filtre inverse* dans le domaine de la théorie du signal. L'intensité cherchée est donc :

$$\hat{\theta}(\nu) = \frac{\hat{s}(\nu)}{\hat{p}_0(\nu)} \quad (5.5)$$

Mais cette expression n'est définie que si $\hat{p}_0(\nu) \neq 0$. En particulier, il apparaît des problèmes d'instabilité quand $\hat{p}_0(\nu)$ prend des valeurs très faibles pour certaines fréquences et fait diverger la fonction \hat{f} . Une méthode pour stabiliser cette opération est d'appliquer un **filtre de Wiener**, une méthode classique de la théorie du signal et en analyse d'images (González et Woods, 1992 ; Pratt, 1978).

Dans le cadre de la théorie du signal, ce filtre a pour but de séparer deux signaux aléatoires éventuellement en présence d'un bruit. Pour notre application, nous allons considérer que la convolution (eq. 5.2) est affectée par un bruit, $n(x)$, que l'on va supposer additif et non corrélé avec l'intensité que l'on cherche à calculer. Cette théorie suppose une connaissance a priori des caractéristiques statistiques du bruit : elle considère que le bruit est une réalisation d'un processus stochastique de moyenne et densité spectrale (TF)

de son autocorrélation) connues. Comme première approximation, nous supposons que le bruit est de moyenne nulle et de densité spectrale constante : il s'agit donc d'un bruit blanc.

Dans cette situation les proportions s'expriment, à travers $s(x)$, par :

$$s(x) = [\theta' * p_0](x) + n(x) \quad (5.6)$$

Le **filtre de Wiener** (voir Réfrégier, 1993 et Réfrégier, 1994) est la fonction f qui, selon les hypothèses déjà mentionnées, minimise l'écart quadratique moyen entre la fonction cherchée, $\theta(x)$, et son estimation, $\hat{\theta}'(x)$:

$$EQM = E \left[(\theta'(x) - \hat{\theta}'(x))^2 \right]$$

Il faut noter que, selon ces expressions, nous allons trouver un estimateur non pas de l'intensité θ mais d'une certaine fonction θ' que l'on supposera très proche de θ . Pour cela, le bruit considéré doit être très faible. Dans la section suivante nous détaillons la signification du bruit et ses implications. Le bruit, dans cette approche, est considéré simplement comme un artefact mathématique pour stabiliser l'expression 5.5. Nous allons noter désormais θ et $\hat{\theta}'$ comme θ et $\hat{\theta}$ respectivement pour simplifier la notation.

Le processus peut donc s'écrire :

$$s(x) = [\theta * p_0](x) + n(x) \longrightarrow \theta(x) = [f * s](x) \quad (5.7)$$

Cette équation devient, dans le domaine de Fourier :

$$\hat{\theta}(\nu) = \hat{f}(\nu) \cdot \hat{\theta}(\nu) \cdot \hat{p}_0 + \hat{f}(\nu) \cdot \hat{n}(\nu)$$

et la fonction cherchée est :

$$\hat{f}(\nu) = \frac{\hat{p}_0^*(\nu)}{|\hat{p}_0(\nu)|^2 + \frac{S_{nn}(\nu)}{S_{\theta\theta}(\nu)}} \quad (5.8)$$

où $\hat{p}_0^*(\nu)$ est la complexe conjuguée de $\hat{p}_0(\nu)$ et $S_{\theta\theta}$, S_{nn} représentent, respectivement, les densités spectrales de l'intensité du processus de Poisson et du bruit. Dans cette extension du filtre de Wiener, il est supposé que l'intensité cherchée est aussi une réalisation d'un processus stochastique de densité spectrale et moyenne connues. En général cette information n'est pas disponible. Alors dans une première approximation, le rapport $\Gamma(\nu) = S_{nn}(\nu)/S_{\theta\theta}(\nu)$ est considéré constant et le plus petit possible comparé à l'autre terme du dénominateur de façon à se rapprocher du filtre inverse (5.5).

La démonstration de l'obtention de l'expression 5.8 est présentée dans l'annexe D.

5.1.2 Interprétation du bruit - Paramètre $\Gamma(\nu)$

Nous avons vu, dans la section précédente, que s'il n'y avait pas de problèmes de divergence, le filtre inverse (eq. 5.5) nous permettait de récupérer exactement l'intensité de Poisson («intensité réelle») à partir des proportions. Or, le fait que le terme $\hat{p}_0(\nu)$ puisse prendre des valeurs très faibles ou nulles nous mène à introduire un facteur de stabilisation. Pour cela, nous avons supposé que la convolution qui lie les proportions à l'intensité de Poisson et aux grains primaires (eq. 5.2) est affectée d'un bruit.

Dans notre approche, le bruit n'est qu'un *artefact mathématique stabilisateur* et il a été pris comme un bruit blanc non corrélé avec l'intensité que l'on cherche à estimer. Il est supposé faire partie intégrante de l'information contenue dans les proportions, donc on considère que les proportions sont légèrement bruitées. De plus, puisque le but final est d'obtenir une estimation la plus proche possible de l'«intensité réelle», ce bruit doit être le plus faible possible.

Le bruit est introduit à travers le facteur $\Gamma(\nu) = S_{nn}(\nu)/S_{\theta\theta}(\nu)$, comme il apparaît dans l'expression 5.8. Ce terme représente le rapport bruit sur signal et donne une idée de la quantité de bruit présent par rapport à la fonction cherchée. Plus ce terme est important et plus la variable initiale est bruitée. En général Γ est une fonction de la fréquence : le bruit peut être plus important à certaines fréquences qu'à d'autres. En théorie du signal, il est très courant de considérer Γ comme constant quand il n'y a pas, a priori, d'information statistique disponible sur la variable cherchée. Cette hypothèse peut être aussi adoptée dans notre application.

Mais, quelle valeur faut-il assigner à Γ ? En pratique elle sera prise arbitrairement très faible. Les tests de sensibilité présentés dans la section suivante donneront une idée des valeurs que ce paramètre peut prendre.

Nous avons supposé que le bruit était contenu dans l'information des proportions. Dans une autre approche il serait envisageable de l'associer à l'*incertitude dans ces proportions*. En effet, dans le chapitre précédent, on a établi le rapport entre la probabilité associée au modèle et les proportions expérimentales. Ces dernières sont des régularisées des probabilités sur un certain support. Le bruit pourrait donc être lié à l'erreur de régularisation et à l'incertitude avec laquelle les proportions sont connues. Dans ce cas l'hypothèse de bruit blanc non corrélé ne serait plus valable puisque le bruit serait probablement lié à la structure des proportions et donc, à l'intensité. Malheureusement, ces incertitudes sur les proportions et le support de régularisation sont inaccessibles et l'analyse de la structure du bruit devient impossible. Cependant, en géostatistique, si ce support est connu, son effet pourrait être estimé, ainsi que l'évaluation de l'incertitude dans les estimations si l'on connaît le processus d'estimation suivi. Il serait alors possible de faire des hypothèses sur la nature du bruit contenu dans les proportions et intégrer celui-ci dans l'estimation de l'intensité de Poisson.

D'un autre côté, c'est le *paramètre* Γ qui nous intéresse dans cette approche. Il représente le rapport entre les transformées de Fourier des autocorrélations du bruit et de l'intensité que l'on cherche à estimer. Puisque tant le bruit que l'intensité sont des variables aléatoires, Γ peut être interprétée en définitive comme le rapport entre les transformées de Fourier des covariances non centrées du bruit et de l'intensité. En effet, l'autocorrélation de l'intensité peut s'exprimer par :

$$C_{\theta\theta}(x) = \int_{-\infty}^{+\infty} \theta^*(u)\theta(x+u)du$$

et sa covariance non centrée par :

$$C(h) = \int_h \theta^*(u)\theta(u+h)du$$

Le lien entre l'une et l'autre est immédiat et il en est de même pour le bruit.

Pour déterminer Γ , si le bruit est connu ou du moins peut être modélisé, il faut pouvoir encore déterminer la structure spatiale de l'intensité. Pour cela nous pouvons nous servir de la proportion. Celle-ci peut être interprétée comme le résultat d'une moyenne glissante et éventuellement pondérée (voir eq. 5.2) sur un support correspondant à l'objet, à travers \mathcal{D} . Une estimation de la covariance non centrée de θ pourrait donc être obtenue à partir de l'analyse de $s(x)$ si l'impact de l'effet de support de régularisation est connu.

La différence entre les deux approches réside dans le fait que dans la première (le bruit est un artefact) nous considérons que les proportions sont «correctes» et cherchons à les récupérer le plus fidèlement possible dans les simulations. Dans la seconde, où le bruit est lié à l'incertitude sur les proportions, nous sommes intéressés à ne pas le retrouver dans les simulations. Ainsi, si l'erreur dans les proportions est connue, cette méthode de déconvolution nous permettrait d'en tenir compte dans le modèle.

Nous avons choisi de suivre la première méthode, très facile à mettre en œuvre et qui ne requiert pas d'hypothèses sur les statistiques de l'intensité et du bruit. Par ailleurs, elle s'est avérée donner de bons résultats. Cependant, il serait intéressant de réaliser une analyse plus fine de l'impact dans le processus de déconvolution d'une connaissance de la structure de l'intensité, en fixant des modèles de covariance pour cette intensité, ou en modélisant cette covariance à partir de celle de la proportion initiale.

5.1.3 Étude de sensibilité

Dans cette section nous étudions l'impact sur le résultat de déconvolution de chacun des paramètres impliqués : les proportions, le choix de l'objet et le terme stabilisateur Γ . Pour ceci des exemples élémentaires à 1D seront étudiés. Le processus suivi dans l'étude de sensibilité est le suivant : pour une courbe de proportion, appelée *proportion initiale*, une famille d'objets définie à travers \mathcal{O} et une valeur de Γ , on applique le processus de déconvolution décrit plus haut pour obtenir la courbe d'intensité. Puis, pour tester la qualité du résultat obtenu, le processus de convolution directe (expression 5.2) est utilisé avec l'intensité calculée ; ensuite, le résultat, dénommé *proportion estimée*, est comparé avec les proportions initiales. L'intensité calculée par notre méthode sera d'autant mieux estimée que les proportions initiale et estimée sont proches.

Nous allons distinguer l'*intensité mathématique*, c'est-à-dire le résultat obtenu par déconvolution, de l'*intensité de Poisson*, en tant que paramètre du modèle booléen. Nous verrons par la suite que ces deux termes ne sont pas équivalents puisque certaines des valeurs mathématiques obtenues, notamment les valeurs négatives, ne sont pas interprétables en termes d'une intensité de Poisson. Le traitement de ces valeurs fera le sujet de la section 5.2. Dans cette section nous nous intéressons à l'analyse du processus de déconvolution et donc à l'intensité mathématique.

5.1.3.1 Sensibilité aux proportions initiales

Dans la section précédente nous avons vu que les proportions, à travers la relation de convolution, peuvent être interprétées comme le résultat d'une moyenne glissante et pondérée de l'intensité sur un support correspondant à l'objet, i.e. le support de convolution \mathcal{O} . Il est facile de supposer que si la proportion est très lisse, l'intensité qui l'a générée doit être aussi, a priori, relativement lisse. À l'inverse, si les proportions présentent des variations importantes à petite échelle, ceci doit se refléter dans l'intensité. A priori, on peut penser que la proportion et l'intensité partagent la même structure spatiale au changement de support près.

Pour analyser le lien entre la proportion et l'intensité, deux proportions ont été testées : une assez lisse (variations à grande échelle) et une très variable à petite échelle. L'objet et le paramètre Γ sont les mêmes dans les deux cas :

- Objet : segment de longueur constante égale à 2.5 m ;
- $\Gamma = 0.005$

Pour un même objet, l'intensité résultante est très différente selon le type de proportion initiale (figures 5.1 et 5.2). Quand celle-ci est plus ou moins lisse (fig.5.1), sans variations importantes à petite échelle (plus petites que la taille du grain, par exemple), l'intensité présente le même comportement à l'échelle du champ de travail, avec plus de variations à faible longueur d'onde.

Quand la proportion de départ présente des variations importantes à petite échelle (fig. 5.2) l'intensité résultante ne ressemble plus à la proportion et présente des variations périodiques d'une longueur d'onde correspondant à la taille de l'objet. Ce phénomène est dû, comme nous allons le voir par la suite, au rapport entre la proportion initiale et l'objet, ou, plus précisément, à l'adaptabilité de l'objet à l'échelle de variabilité des proportions.

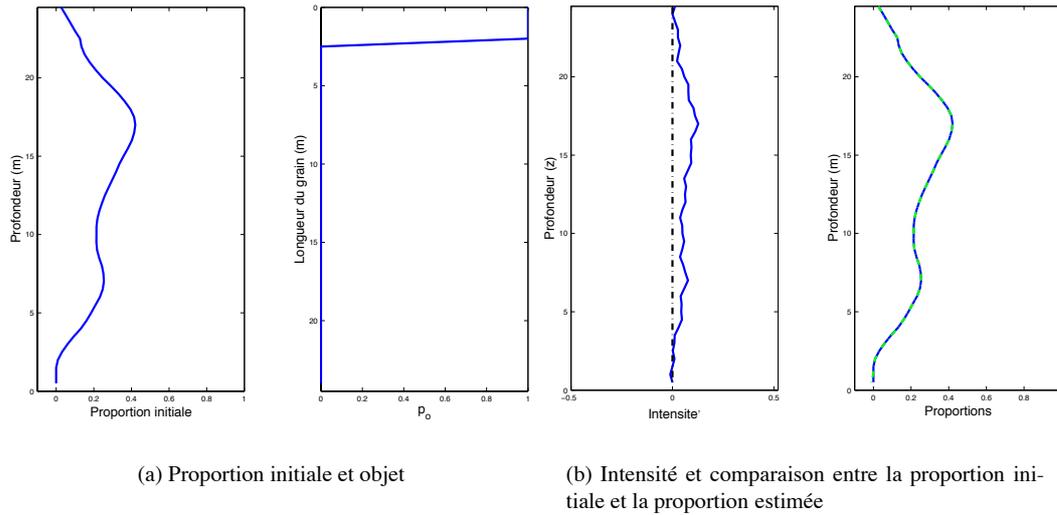


FIG. 5.1 – Proportion initiale peu variable. En vert pointillé la proportion estimée.

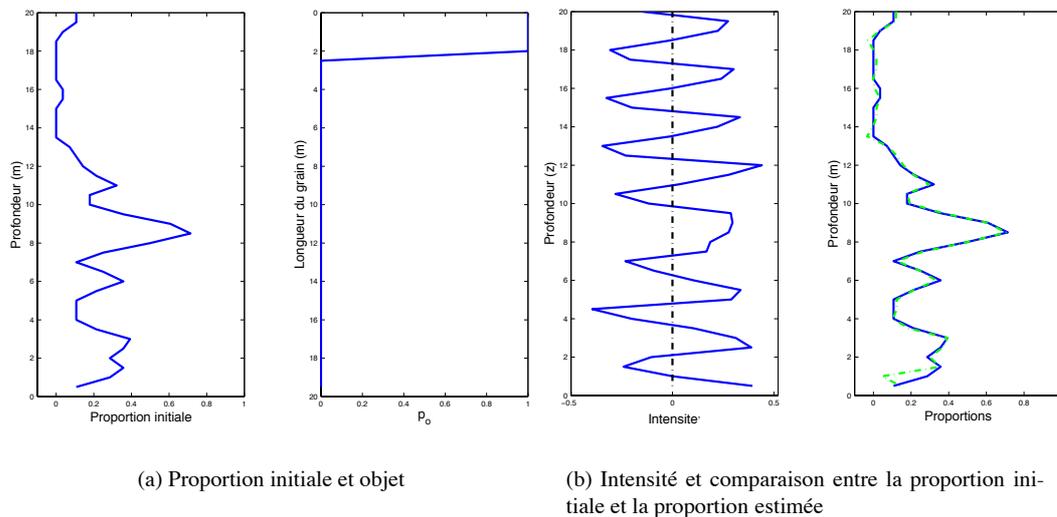


FIG. 5.2 – Proportion initiale très variable. En vert pointillé la proportion estimée.

Ce simple test a montré que, pour une même famille d'objets, des proportions à faibles valeurs et fortement variables à petite échelle induisent des intensités avec des valeurs négatives très importantes et une composante de période égale à la taille de l'objet. La différence entre la proportion initiale et la proportion estimée est plus importante (~ 0.02 en valeur absolue moyenne) que dans le cas où la proportion initiale est peu variable (~ 0).

5.1.3.2 Sensibilité à l'objet

Pour tester l'influence du choix de l'objet, on utilise une proportion initiale lisse et une valeur de Γ faible (0.005). De cette façon seules les variations dues à p_0 sont prises en compte. Nous avons choisi de tester deux familles d'objets différentes : des segments dont la taille reste constante et des segments de taille variable selon une loi uniforme. Les probabilités p_0 associées à ces deux familles sont représentées dans la figure 5.3. Des objets de tailles différentes sont testés pour chaque cas.

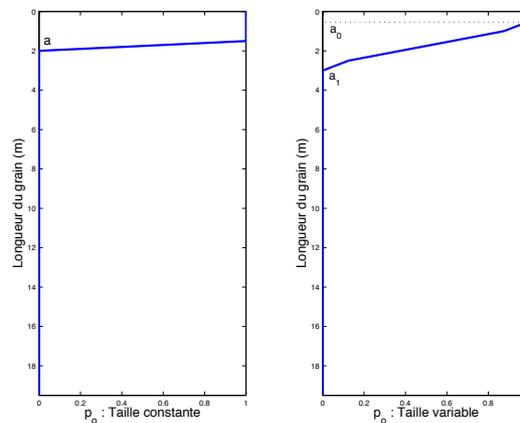


FIG. 5.3 – Probabilités associées aux deux familles d'objets utilisées. À gauche, des segments de taille constante a . À droite, des segments de taille variable aléatoirement entre a_0 et a_1 selon une loi uniforme : $a \sim U_{[a_0, a_1]}$.

Dans le cas où les objets sont de taille constante (Cas 1 : fig. 5.4) pour des tailles petites par rapport à la taille du champ (cas a, b, c), l'intensité mathématique obtenue reflète les caractéristiques générales de la proportion et la présence des valeurs négatives est faible. Le nombre de valeurs négatives, ainsi que les valeurs elles-mêmes, augmentent au fur et à mesure que la taille augmente, spécialement à partir de $a = 5$ m et la forme générale de l'intensité s'éloigne de plus en plus de celle de la proportion. Cependant, la courbe de proportion estimée (en vert pointillé dans les figures) est, dans tous les cas, très proche ou presque identique à la proportion initiale, sauf pour les effets de bord qui sont plus importants pour les objets plus grands.

Le deuxième cas (fig. 5.5) concerne des grains de taille variable selon des lois uniformes. Cette loi est plus flexible que la précédente, dans le sens où elle autorise l'introduction de grains plus grands sans que l'intensité résultante soit affectée par une augmentation des valeurs négatives.

L'analyse des résultats obtenus est présentée à la figure 5.6. Nous regardons d'un côté la ressemblance entre les proportions initiale et estimée et de l'autre côté, les valeurs négatives de l'intensité mathématique puisqu'elles ont une signification particulière. En effet, l'implantation d'un objet dans un point de l'espace a un impact sur un voisinage dont la taille dépend des dimensions du grain. Ainsi, la proportion en un point

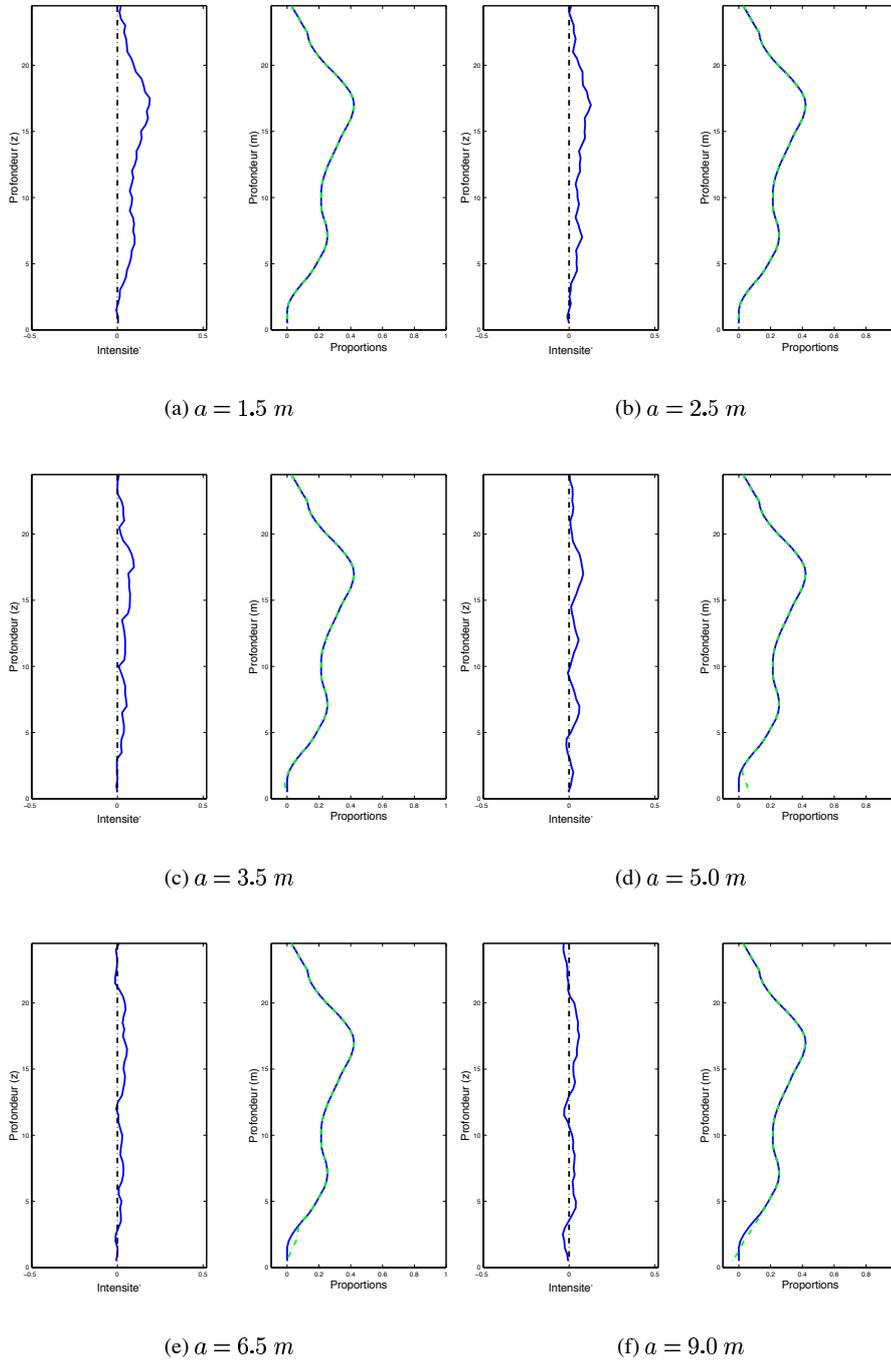


FIG. 5.4 – Cas 1 : Intensité mathématique et proportions initiales (en bleu) et estimées (en pointillé vert). Les objets sont de taille constante a dans chaque cas. Le paramètre Γ est constant et égal à 0.005.

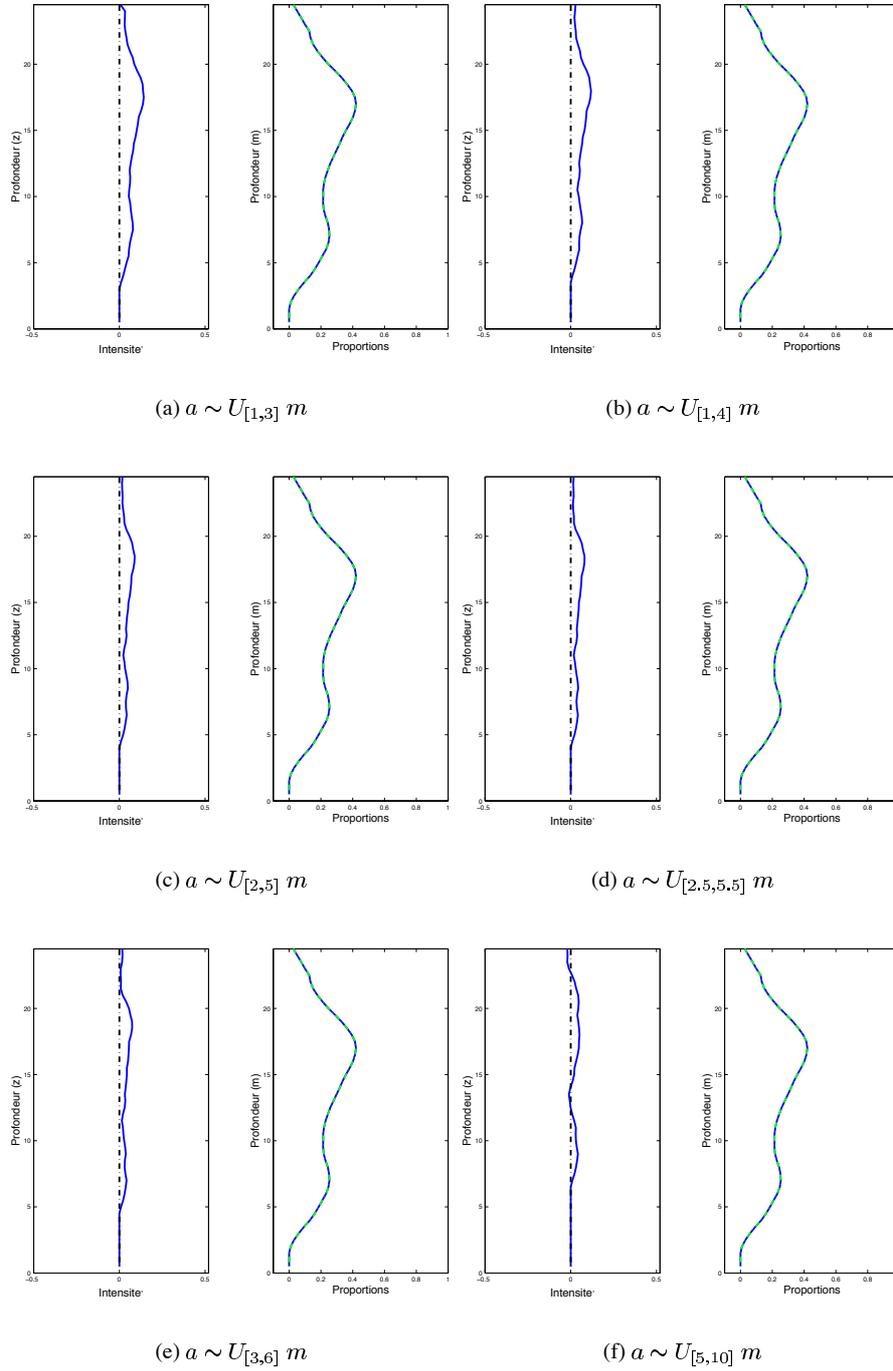
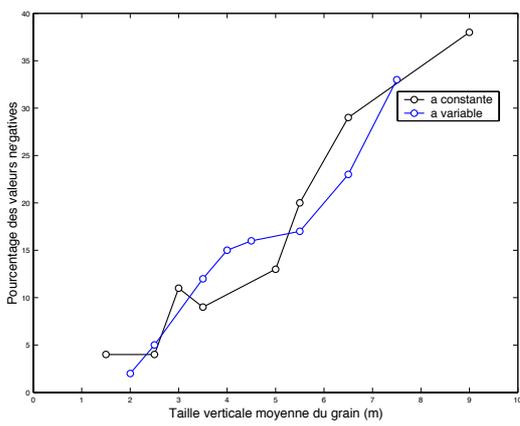
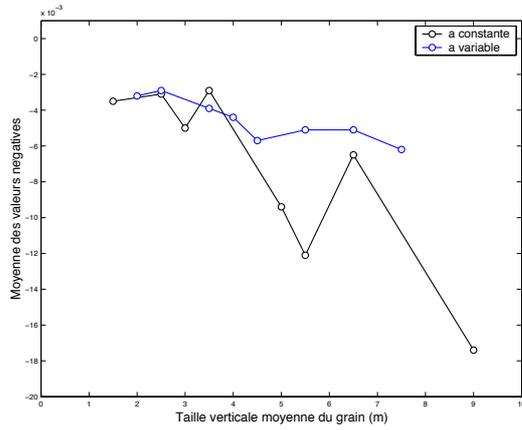


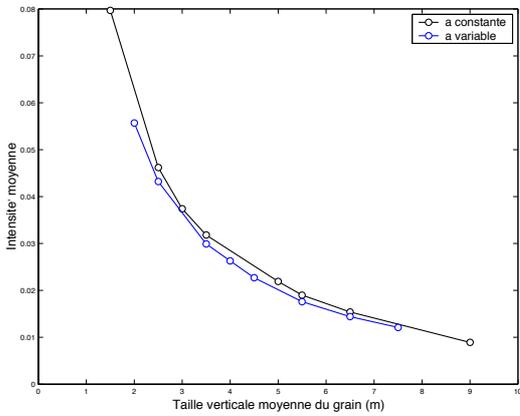
FIG. 5.5 – Cas 2 : Intensité mathématique et proportions initiales (en bleu) et estimées (en pointillé vert). Les objets sont de taille variable selon des lois uniformes. Le paramètre Γ est constant et égale à 0.005.



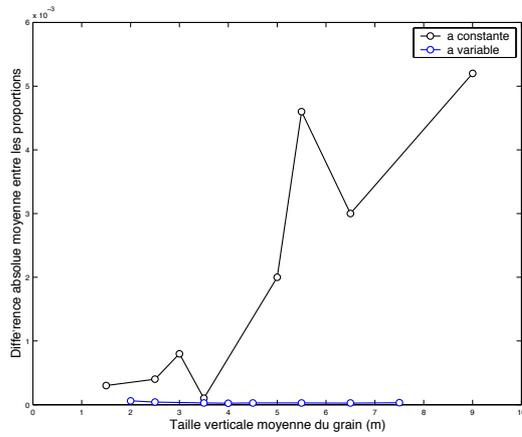
(a) Pourcentages des valeurs négatives dans l'intensité obtenue



(b) Valeur moyenne des valeurs négatives de l'intensité



(c) Valeur moyenne de l'intensité



(d) Valeur moyenne de la différence, en valeur absolue, entre la proportion initiale et la proportion estimée avec l'intensité

FIG. 5.6 – Résumé des statistiques en fonction de la taille verticale moyenne des objets, a .

tient compte de tous les objets qui ont germé dans son voisinage. Supposons que la proportion dans un point x_A est faible et que les proportions autour de ce point sont plus importantes. Pour respecter les proportions fortes, il sera nécessaire d'implanter beaucoup d'objets dans ces points (intensité fortement positive), mais cela se traduira par un «excès» de proportion en x_A . Pour compenser cet effet, il faudra éliminer des objets : la valeur de l'intensité en x_A sera négative. Cet effet est d'autant plus marqué que les objets sont grands et la proportion variable, puisque le domaine d'influence des objets est très important et qu'ils ne peuvent pas reproduire la variabilité de la proportion. On dira alors que l'objet n'est pas «adapté» aux proportions, et l'intensité présentera des valeurs négatives importantes.

La figure 5.6(a) montre l'évolution, en fonction de la taille moyenne de l'objet, de la présence des valeurs négatives, exprimée en pourcentage sur le nombre total de points. Ce pourcentage augmente, pour les deux familles d'objets, avec la taille moyenne du grain. Cette augmentation est assez rapide : elle peut passer de $\sim 5\%$ pour un segment de 2.5 m de taille moyenne à $\sim 15\%$ pour un segment de 5 m de taille moyenne. Si la taille de l'objet se réduit à 0 m, i.e. l'objet est réduit à un point, l'intensité mathématique résultante sera identique à la proportion initiale, donc sans points négatifs. Dans notre application, cependant, les objets auront toujours une certaine taille parfois assez importante.

Ce résultat est général pour toute courbe de proportion initiale. D'autres tests avec une proportion initiale très variable à petite échelle ont montré que le nombre de valeurs négatives dans le résultat de la déconvolution augmente rapidement avec la taille des grains.

La valeur moyenne de ces valeurs négatives (fig. 5.6(b)) tend à diminuer avec la taille des objets. Ceci veut dire que quand la taille des grains augmente, ceux-ci sont donc de moins en moins adaptés à la proportion initiale et l'intensité mathématique non seulement présente de plus en plus de valeurs négatives mais ces valeurs sont plus importantes en valeur absolue. Cet effet est moins marqué dans le cas où les grains sont de taille variable.

La valeur moyenne expérimentale de l'intensité diminue avec la taille de l'objet (fig. 5.6(c)). En d'autres termes, pour une proportion donnée, il faut peu d'objets dans la simulation quand ceux-ci sont grands.

Pour finir, la différence moyenne, en valeur absolue, entre la proportion initiale et la proportion estimée est représentée à la figure 5.6(d). Pour le cas où les objets sont de taille constante, il apparaît que la proportion estimée avec l'intensité mathématique s'éloigne d'autant plus de la proportion initiale que les objets sont grands. Cependant, cette différence est minime, de l'ordre de 10^{-3} , et est due principalement aux effets de bords (voir fig. 5.4). Dans le cas où la taille des objets suit une loi uniforme, la différence moyenne entre les deux proportions est pratiquement nulle.

Ce test a montré que, pour un couple proportion - objet quelconque, nous pouvons calculer une intensité mathématique qui nous permet de reconstruire la proportion initiale presque exactement. Cependant l'intensité présente le plus souvent des valeurs négatives qui peuvent être très importantes, notamment quand les objets ne sont pas «adaptés» à la proportion initiale. Ces valeurs ne sont pas acceptables comme intensité de Poisson qui est, par définition, strictement positive ou nulle.

5.1.3.3 Sensibilité au paramètre Γ

Nous allons étudier ici l'influence sur le processus de déconvolution de la valeur choisie pour le paramètre Γ . Les tests seront effectués sur les deux courbes de proportion utilisées précédemment, l'une peu variable et l'autre très variable à petite échelle. Seules des familles d'objets à taille constante seront utilisées. Quatre cas sont analysés :

- Cas 1 : La proportion est peu variable et l'objet est de taille constante égale à 2.5 m.
- Cas 2 : La proportion est peu variable et l'objet est de taille constante égale à 5.0 m.

- Cas 3 : La proportion est très variable et l'objet est de taille constante égale à 2.5 m.
- Cas 4 : La proportion est très variable et l'objet est de taille constante égale à 1.5 m.

Les figures 5.7 à 5.10 montrent ces exemples pour différentes valeurs du paramètre Γ . Dans chacune de ces figures sont représentées la proportion de départ, i.e. les données, et la probabilité associée à l'objet \mathcal{P} (figure indécelée (a)). Les figures (b) à (f) montrent, pour des valeurs différentes de Γ , l'intensité mathématique obtenue et la proportion estimée correspondante (lignes de traits verts).

De façon générale, pour une proportion et un objet donnés, l'augmentation de la valeur de Γ se traduit par l'obtention d'une intensité de plus en plus lisse, avec une plus faible présence de valeurs négatives. Le fait de prendre des valeurs fortes de Γ , donc du rapport bruit sur signal, implique, selon notre hypothèse, qu'une partie importante de l'information contenue dans la proportion initiale est due au bruit et elle est, en conséquence, filtrée dans le processus de déconvolution. Pour cette raison, quand Γ augmente, la proportion estimée sera de plus en plus éloignée des données de départ. Puisque pour le moment nous nous intéressons à respecter autant que possible les données de départ, la valeur de Γ doit alors être la plus petite possible mais, dans quelle fourchette de valeurs ? Nous avons testé différentes valeurs de Γ , choisies arbitrairement, pour avoir une idée de l'ordre de grandeur de ce paramètre.

Une proportion peu variable (cas 1, fig. 5.7, et cas 2, fig. 5.8) nous permet d'utiliser des valeurs de Γ dans une fourchette relativement large, entre 0.005 et 1, tout en respectant la proportion initiale (la différence moyenne entre la proportion initiale et la proportion estimée est de moins de 0.01).

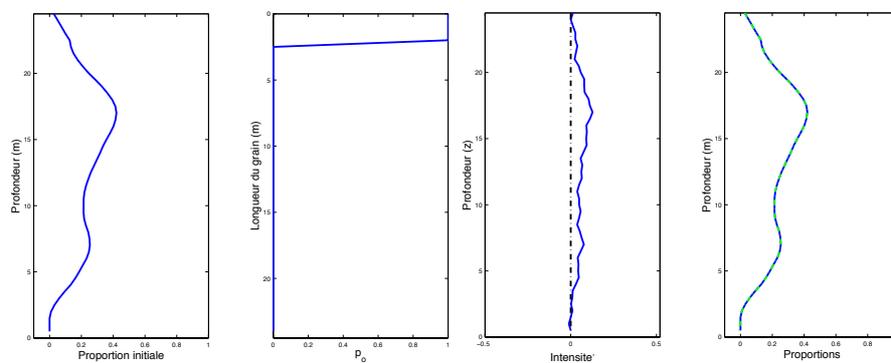
Dans le cas d'une proportion très variable à petite échelle (cas 3 et 4, fig. 5.9 et 5.10 respectivement) la fourchette des valeurs est plus restreinte, entre 0.005 et 0.1, la différence entre les proportions initiale et estimée augmentant rapidement.

Les objets interviennent aussi dans la détermination de la fourchette des valeurs de Γ appropriées. Dans le cas d'objets de petite taille (cas 1 et 4), la différence entre la proportion initiale et la proportion estimée augmente très rapidement avec Γ : dans le cas 1, cette différence passe de moins de 0.008 pour $\Gamma = 1$ à 0.055 pour $\Gamma = 10$; dans le cas 4, la différence passe de 0.02 à 0.085 pour la même fourchette de valeurs de Γ . Par contre, les cas avec des objets de grande taille (cas 2 et 3) se montrent plus stables par rapport aux changements de Γ : si celui-ci varie entre 1 et 10, par exemple, la différence entre les proportions initiale et estimée augmente de 0.004 à moins de 0.02 pour le cas 2 et de 0.043 à 0.062 pour le cas 3. Dans tous les cas, si Γ prend des valeurs inférieures à 1 cette différence est toujours plus faible.

Il faut signaler que le fait de considérer un objet comme grand ou petit dépend de la proportion à laquelle il est associé : si la taille de l'objet est plus grande que l'échelle de variabilité de la proportion, alors il est considéré comme grand et inversement s'il est plus petit. Ainsi le segment de longueur 2.5 m est grand quand la proportion est très variable (cas 3) et petit quand la proportion est très lisse (cas 2).

Les graphiques de la figure 5.11 résument les résultats de ce test. Dans les quatre cas étudiés, le pourcentage des valeurs négatives tend à diminuer avec Γ (fig. 5.11(a)). Cette diminution est plus évidente pour les cas 3 et 4, correspondant à la proportion très variable. Pour une proportion donnée, à Γ fixé faible (≤ 1) le pourcentage de valeurs négatives est plus élevé si les objets sont grands. Quand Γ est grand ($\Gamma = 10$), le pourcentage de valeurs négatives dans tous les cas tend à se stabiliser autour de 5 – 10%.

La valeur moyenne des valeurs négatives de l'intensité (fig. 5.11(b)) se rapproche de 0 quand Γ augmente, c'est-à-dire que les valeurs négatives tendent non seulement à diminuer en nombre mais aussi en valeur. Cet effet est aussi plus visible pour les cas 3 et 4 (proportion très variable). À Γ faible (< 1), les valeurs négatives sont très importantes dans ces deux cas, avec une moyenne qui peut être plus petite que -0.15 , donc très importante par rapport à la valeur moyenne globale de l'intensité mathématique, qui est de l'ordre de 10^{-2} . Dans les cas 1 et 2 (proportion peu variable), la moyenne des valeurs négatives reste assez stable et très proche de 0. Quand $\Gamma = 10$, la moyenne des valeurs négatives est pratiquement nulle, montrant



(a) Cas 1 : Proportion initiale et probabilité associée à l'objet

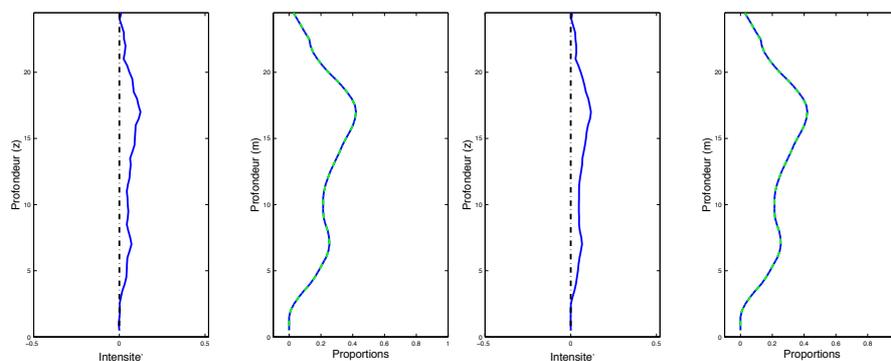
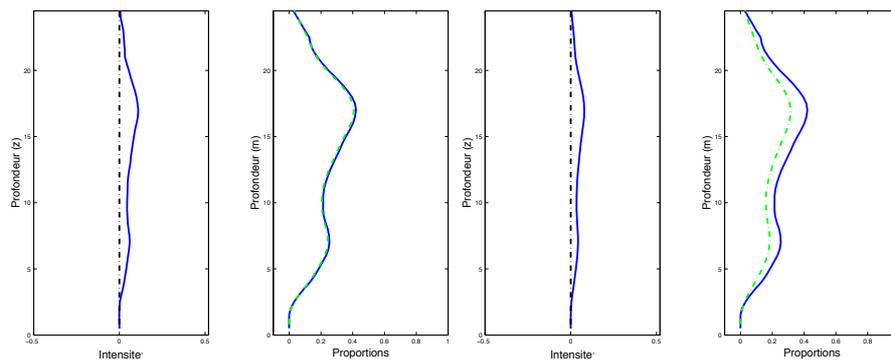
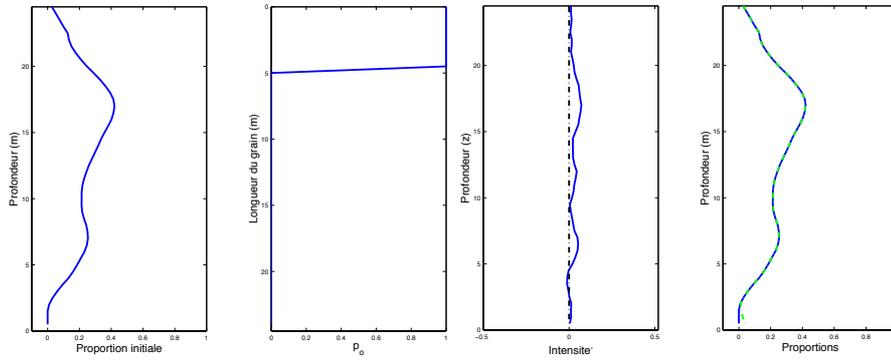
(b) $\Gamma = 0.005$ (c) $\Gamma = 0.01$ (d) $\Gamma = 0.1$ (e) $\Gamma = 1$ (f) $\Gamma = 10$

FIG. 5.7 – Cas 1 : Intensité mathématique et proportions initiales (en bleu) et estimées (en pointillé vert). La proportion initiale est peu variable et l'objet utilisé est de taille verticale constante égale à 2.5 m.



(a) Cas 2 : Proportion initiale et probabilité associée à l'objet

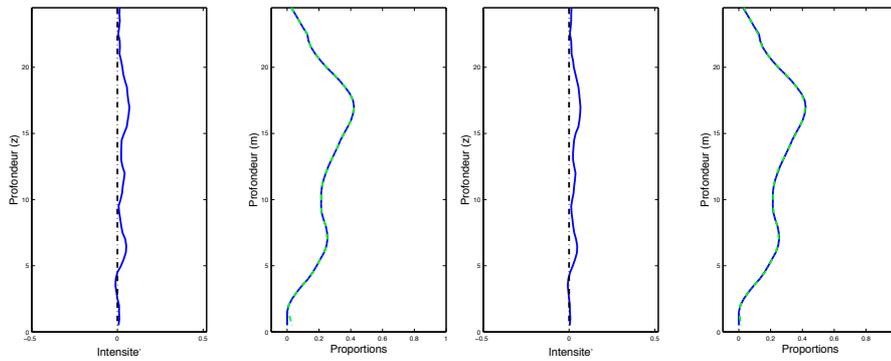
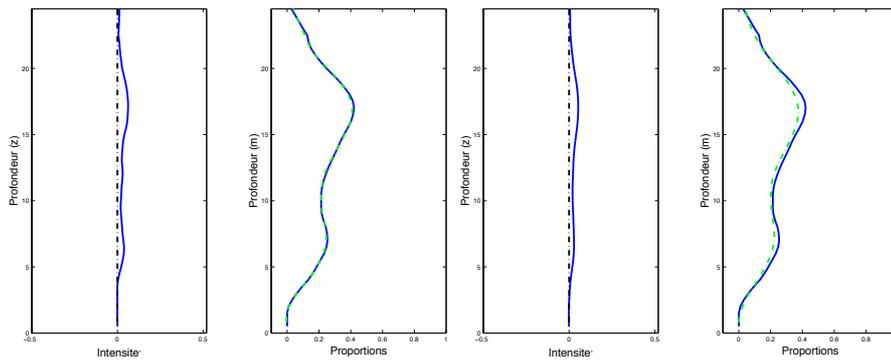
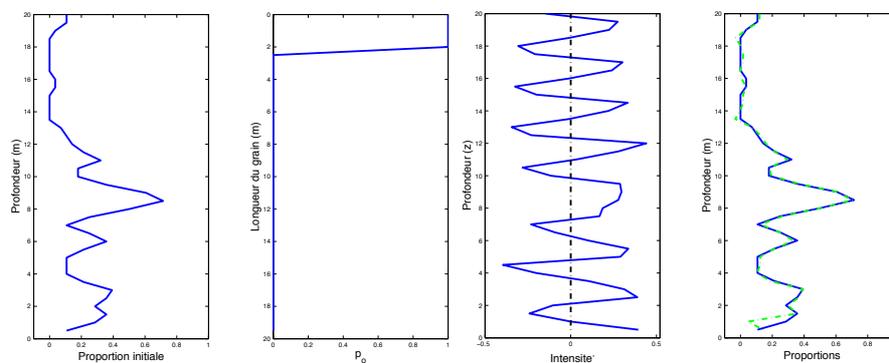
(b) $\Gamma = 0.005$ (c) $\Gamma = 0.01$ (d) $\Gamma = 0.1$ (e) $\Gamma = 1$ (f) $\Gamma = 10$

FIG. 5.8 – Cas 2 : Intensité mathématique et proportions initiales (en bleu) et estimées (en pointillé vert). La proportion initiale est peu variable et l'objet utilisé est de taille verticale constante égale à 5 m.



(a) Cas 3 : Proportion initiale et probabilité associée à l'objet

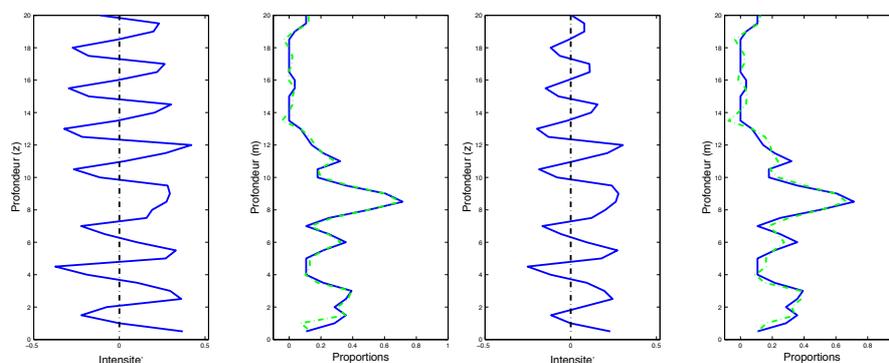
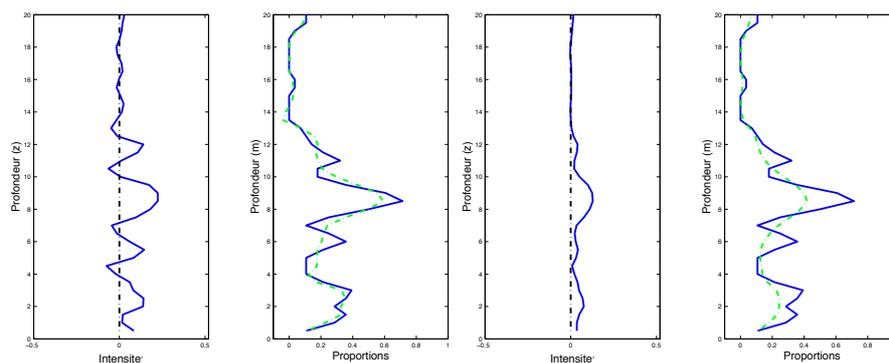
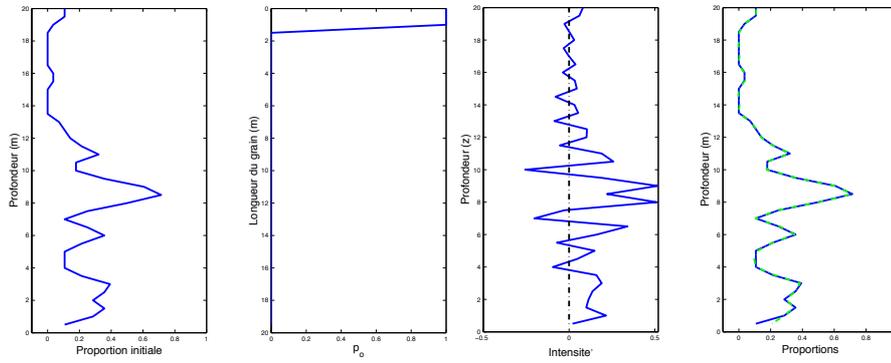
(b) $\Gamma = 0.005$ (c) $\Gamma = 0.01$ (d) $\Gamma = 0.1$ (e) $\Gamma = 1$ (f) $\Gamma = 10$

FIG. 5.9 – Cas 3 : Intensité mathématique et proportions initiales (en bleu) et estimées (en pointillé vert). La proportion initiale est très variable et l'objet utilisé est de taille verticale constante égale à 2.5 m.



(a) Cas 4 : Proportion initiale et probabilité associée à l'objet

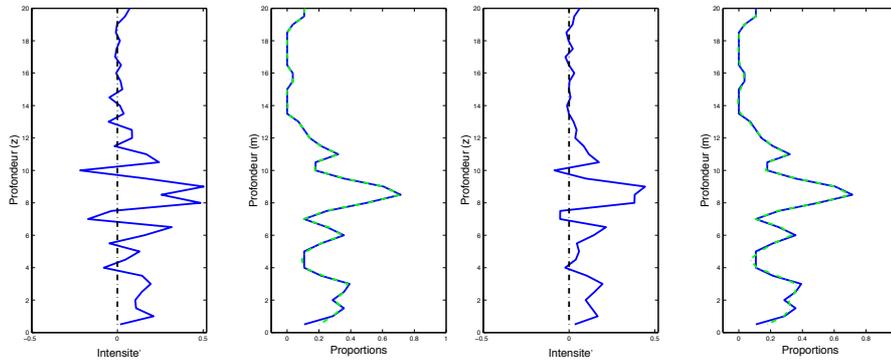
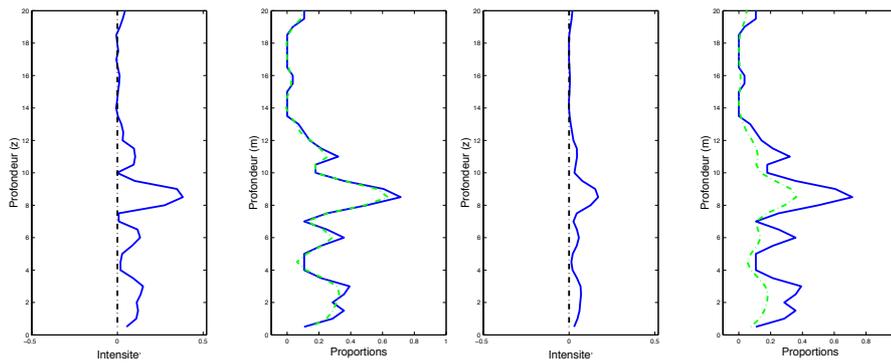
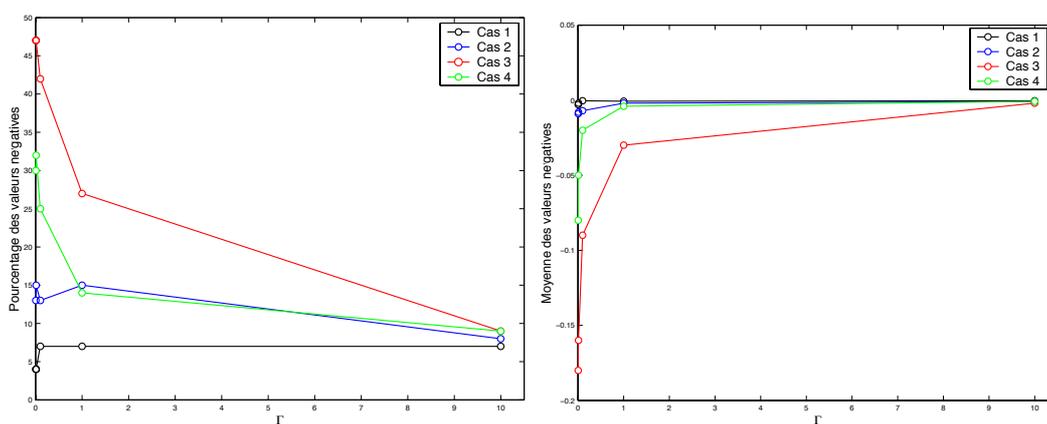
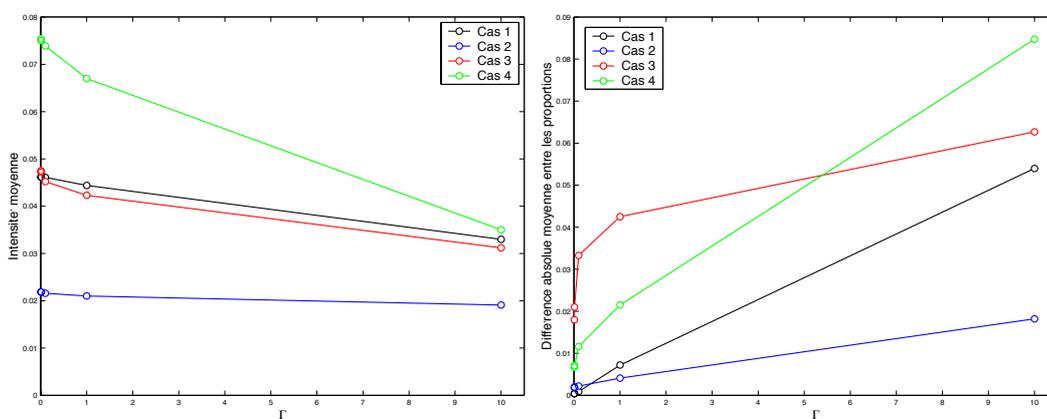
(b) $\Gamma = 0.005$ (c) $\Gamma = 0.01$ (d) $\Gamma = 0.1$ (e) $\Gamma = 1$ (f) $\Gamma = 10$

FIG. 5.10 – Cas 4 : Intensité mathématique et proportions initiales (en bleu) et estimées (en pointillé vert). La proportion initiale est très variable et l'objet utilisé est de taille verticale constante égale à 1.5 m.



(a) Pourcentages des valeurs négatives dans l'intensité obtenue

(b) Valeur moyenne des valeurs négatives de l'intensité



(c) Valeur moyenne de l'intensité

(d) Valeur moyenne de la différence, en valeur absolue, entre la proportion initiale et la proportion estimée avec l'intensité

FIG. 5.11 – Résumé des statistiques en fonction de Γ .

que les valeurs négatives sont à peine représentatives.

La moyenne de l'intensité mathématique (fig. 5.11(c)) obtenue diminue au fur et à mesure que Γ augmente, avec une tendance à se rapprocher dans tous les cas pour Γ très grand. Cela montre la perte d'information remarquée auparavant : en augmentant Γ , on suppose que les proportions initiales sont affectés par un bruit de plus en plus important et qui va être filtré lors du processus de déconvolution. Cela se reflète aussi dans le graphique 5.11(d) puisque la différence entre la proportion initiale et la proportion estimée avec l'intensité est de plus en plus grande quand Γ augmente.

Le fait d'augmenter Γ fait diminuer en nombre et en valeur les points négatifs de l'intensité mathématique et, en même temps, perdre de l'information contenue dans la proportion initiale. Si cette information, et donc le bruit, est effectivement relative à l'erreur des proportions, alors nous sommes capables de la filtrer et de ne pas la reproduire dans le modèle. Le problème est que nous ne sommes pas capables de quantifier le bruit et il est fixé arbitrairement à travers Γ . Ainsi, nous avons choisi de le considérer comme très faible de façon à récupérer les proportions initiales de façon satisfaisante.

5.1.3.4 Conclusions

Les trois paramètres qui interviennent dans le processus de déconvolution sont la proportion initiale, les caractéristiques de l'objet et le paramètre Γ , i.e., le rapport bruit sur signal. Ce paramètre Γ représente la quantité de bruit contenu dans les proportions initiales. Lors des tests de sensibilité de la méthode à ces trois paramètres nous avons pu mettre en évidence les conclusions suivantes :

- Pour tout couple proportion initiale - objet et pour des valeurs faibles de Γ , il est possible d'obtenir par le processus de déconvolution une *intensité mathématique* qui respecte presque exactement les proportions initiales.
- Le nombre de points à valeur négative dans l'intensité mathématique augmente avec la taille du grain, pour une proportion initiale donnée. Ces valeurs deviennent en même temps de plus en plus fortes. Cet effet est plus important quand les objets sont de taille constante. Cette famille est, dans ce sens, plus contraignante que celle qui autorise des tailles différentes. En général, plus les objets sont petits et moins il y a de valeurs négatives dans l'intensité mathématique.
- À objet et Γ fixés, il y a moins de valeurs négatives d'intensité pour des proportions peu variables que pour des proportions très variables à petite échelle.
- Dans tous les cas, des valeurs fortes de Γ induisent une intensité très lisse avec moins de valeurs négatives et à plus faible valeur absolue, mais ne garantissent pas la non existence de ces valeurs.
- Une augmentation de Γ implique, par hypothèse, que la proportion initiale contient de plus en plus de bruit. Ce bruit est filtré dans le processus de déconvolution. Le problème est que, en l'absence d'autres informations, nous sommes contraints de choisir ce bruit arbitrairement. De ce fait, nous avons supposé, dans un premier temps, que les données disponibles sont sans erreur et la valeur de Γ utilisée, bien que arbitrairement choisie, doit être très faible pour récupérer la proportion initiale.

Les valeurs négatives de l'intensité mathématique, parfois très fortes et en grand nombre, ne peuvent pas être interprétées en terme d'intensité de Poisson, qui est par définition strictement positive ou nulle. Ces valeurs apparaissent principalement quand les objets sont trop grands pour reproduire la variabilité de la proportion. Il est donc nécessaire de corriger le résultat de la déconvolution. Dans la section suivante, dédiée aux aspects pratiques de la mise en œuvre de la méthode de déconvolution proposée, ce problème sera étudié.

5.2 Aspects pratiques

Cette section est consacrée aux aspects pratiques concernant la mise en œuvre de la méthode d'inférence proposée. Plusieurs points vont être analysés : la définition du domaine de calcul, la correction des valeurs négatives de l'intensité mathématique et la détermination du nombre d'objets à introduire dans la simulation. Une description rapide du logiciel mettant en œuvre cette méthode est présentée dans l'annexe E.

5.2.1 Traitement des valeurs négatives de l'intensité

La méthode proposée pour inférer l'intensité de Poisson est basée sur un processus purement mathématique. Ce processus, tel qu'il a été développé ici, ne tient pas compte du fait que l'intensité (le nombre moyen de points de Poisson à simuler par unité de volume) prend des valeurs strictement positives ou nulles. De ce fait, il est courant d'obtenir des intensités à valeurs négatives qui peuvent être assez importantes et qui ne peuvent pas être interprétées comme une intensité de Poisson. Ces valeurs négatives doivent donc être corrigées. Plus les valeurs négatives sont importantes et plus la correction à faire sera grande, avec, en conséquence, une perte d'information et de fiabilité dans les résultats. Cette fiabilité sera mesurée par la similitude entre la proportion initiale et la proportion «estimée» après correction (proportion corrigée).

Mais, tout d'abord, que signifie la présence de valeurs négatives d'intensité ? Si les proportions présentent des variations à petite échelle, l'implantation d'objets qui recouvrent une région plus grande que cette échelle de variation ne permettra pas de respecter cette proportion. Le «lissage» de l'intensité par l'objet lors de la convolution est trop fort pour restituer la variabilité des proportions. Pour compenser cet effet il faut, mathématiquement, enlever des objets dans le voisinage. Ceci se traduit par la présence de valeurs négatives dans le résultat de déconvolution qui doivent être éliminées pour pouvoir interpréter cette intensité «mathématique» comme une intensité de Poisson.

Dans la section précédente il a été montré qu'une augmentation de la valeur du paramètre Γ implique une réduction significative des valeurs négatives. Il a été donc envisagé d'utiliser ce paramètre pour éliminer ces valeurs. Cependant, l'augmentation de ce paramètre entraîne une réduction de l'information des proportions qui sera restituée. L'intensité mathématique se lisse donc significativement en réduisant les valeurs négatives mais en éloignant, en même temps, la proportion estimée des données initiales. De plus, pour une valeur donnée de Γ , il n'y a pas de garantie de la complète disparition de ces valeurs négatives. La difficulté à contrôler ces valeurs par ce paramètre réside dans le fait que le bruit est difficile à interpréter et à modéliser en termes d'erreur d'estimation de la proportion ou d'erreur de mesure, puisque nous ne disposons pas de cette information. Le but de ce travail étant alors d'obtenir une intensité qui puisse reproduire les proportions initiales le plus fidèlement possible, nous ne pouvons pas augmenter sa valeur arbitrairement sans risque de perdre d'information significative. Le paramètre Γ est donc fixé à une valeur arbitraire très faible.

Outre les considérations sur Γ , plusieurs options ont été testées pour éliminer ces valeurs négatives. Nous avons vu précédemment que le résultat obtenu par déconvolution est mathématiquement correct puisqu'il nous rend presque exactement la proportion initiale. Ainsi, pour évaluer les différentes options proposées nous comparons la proportion estimée, rendue par l'intensité mathématique, avec la *proportion corrigée*, rendue par l'intensité corrigée par chaque méthode. De plus, nous allons aussi comparer la valeur de l'intégrale de chacune des intensités sur tout le volume (tableau 5.1). Cette valeur est liée au nombre moyen d'objets qu'il faudra introduire dans la simulation. Par la suite nous appellerons cette valeur *intensité globale*.

Les exemples utilisés pour illustrer les différentes approches de correction sont construits à 1D avec une même courbe de proportions et une famille d'objets de longueur constante égale à 4 m, assez grands

par rapport au domaine (~ 20 m). Nous fixons $\Gamma = 0.005$ dans le processus de déconvolution. L'intensité mathématique obtenue est donc identique dans tous les cas et l'intensité corrigée varie selon l'approche utilisée. Les figures 5.12(a) à 5.12(d) montrent des exemples graphiques de chacune de ces corrections. La notation employée est la suivante : θ l'intensité mathématique, θ' l'intensité après correction (intensité corrigée), p la proportion initiale (les données), p_e la proportion estimée avec l'intensité mathématique et p'_e la proportion estimée avec l'intensité corrigée (proportion corrigée).

Option 1 : Ramener les valeurs négatives à zéro C'est la première approche utilisée et la plus simple à mettre en oeuvre. Les résultats sont acceptables si le nombre de valeurs négatives est petit et si ces valeurs ne sont pas trop importantes en valeur absolue. Sinon la correction est trop grossière et l'erreur introduite est grande. De façon générale, la transformation des valeurs négatives en valeurs nulles fait augmenter artificiellement, d'une part les valeurs des proportions corrigées par rapport aux valeurs initiales (fig. 5.12(a)), et d'autre part la valeur de l'intensité globale, donc le nombre d'objets à simuler.

Option 2 : Translation vers des valeurs positives La courbe d'intensité est déplacée vers l'axe positif d'une distance égale à la valeur minimale. La valeur négative la plus forte prend ainsi une valeur nulle et le reste de la courbe est déplacé en conséquence. Les valeurs de proportion obtenues avec cette intensité corrigée sont aussi plus fortes que celle de départ (fig. 5.12(b)) puisque le nombre d'objets à simuler est plus grand : la valeur globale d'intensité a augmenté. Il faut corriger cette valeur en déduisant l'aire sous la courbe d'intensité correspondant à la translation. Cette correction fera diminuer les valeurs des proportions : la courbe corrigée se rapprochera de la courbe de proportion initiale mais la tendance générale ne sera pas respectée.

Option 3 : Distribuer les valeurs négatives entre les points voisins Dans cette approche nous partons de l'idée que les valeurs négatives apparaissent pour compenser l'introduction de trop d'objets ou d'objets excessivement grands en certains points. Pour éliminer une valeur négative en un point donné, nous réduisons les valeurs positives d'un certain pourcentage aux points du voisinage dont la taille est liée à celle du grain primaire. Ainsi, l'élimination des valeurs négatives est compensée par une diminution des valeurs positives autour de celles-ci. Cependant, ce procédé ne garantit pas le fait que certains points à faible valeur d'intensité positive ne deviendront pas négatifs après correction. Cette approche de traitement local s'est avérée donner d'assez bons résultats pour des intensités mathématiques avec peu de valeurs négatives, en conservant la valeur de l'intégrale de l'intensité sur tout le volume et en respectant la proportion initiale. Par contre, si l'intensité présente beaucoup de valeurs négatives, les résultats sont loin d'être acceptables (fig. 5.12(c)). Ceci sera encore plus marqué à 2D et 3D.

Option 4 : Lissage par moyenne sur un voisinage glissant L'idée est, dans le fond, la même que dans l'approche précédente : éliminer les valeurs négatives par compensation avec une diminution des valeurs fortement positives. Nous réalisons une moyenne de l'intensité sur un voisinage de taille égale à la région d'influence de l'objet. Cette moyenne peut être éventuellement pondérée, avec des points de pondération qui dépendent de la probabilité associée à l'objet, p_0 . De cette façon les valeurs d'intensité excessivement fortes et excessivement faibles disparaissent, en lissant considérablement mais tout en respectant la proportion initiale et la valeur globale de l'intensité (fig. 5.12(d)). C'est une approche simple à mettre en oeuvre et qui donne des résultats assez satisfaisants. Le fait de réaliser une moyenne glissante revient à calculer une convolution entre l'intensité mathématique et un noyau égal à p_0 mais normalisé. Comme dans l'option 3, nous ne pouvons pas garantir la complète disparition des valeurs négatives, mais celles-ci, si elles existent, sont rares et à très faible valeur. Dans ce cas nous pouvons tout simplement les tronquer à zéro sans introduire trop de modifications.

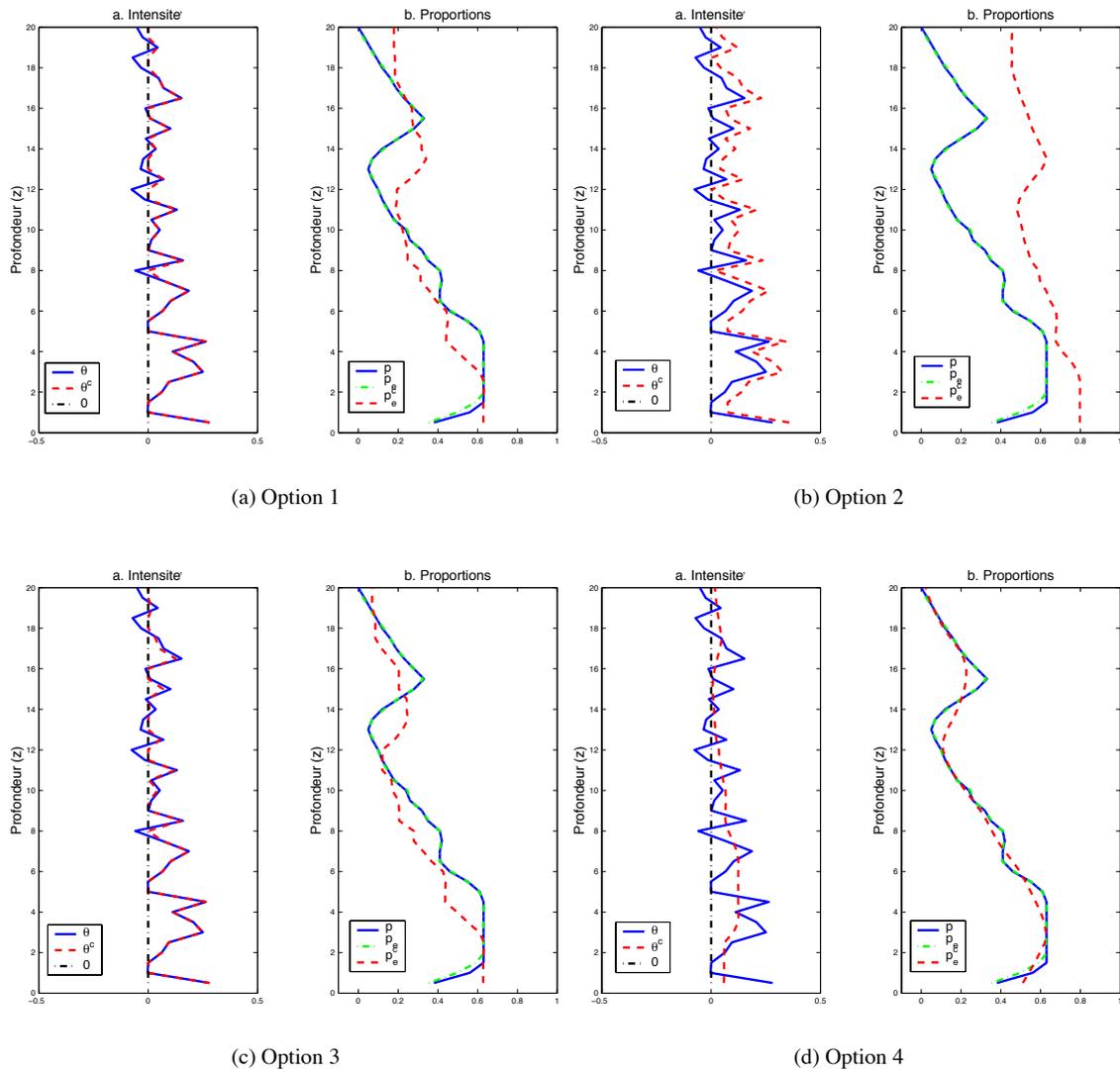


FIG. 5.12 – Comparaison graphique des quatre approches testées pour la correction de l'intensité mathématique.

Les approches 1 et 3 sont locales, c'est-à-dire, qu'elles ne corrigent l'intensité mathématique qu'à certains endroits précis. Ceci implique une modification dans le comportement global de l'intensité, et donc dans la distribution des objets, qui peut être très significative, comme le montrent ces exemples.

Il existe un effet de décalage vers les niveaux inférieurs des courbes de proportion corrigées obtenues avec les méthodes 1 et 2 par rapport à la courbe de proportion initiale. Dans ces deux approches, la signification physique des valeurs négatives n'a pas été prise en compte. L'élimination d'une valeur négative n'est pas compensée par une réduction du nombre d'objets générés dans son voisinage. Toutes les valeurs sont donc décalées, dans ce cas, vers les niveaux inférieurs puisque le point d'implantation correspond au point supérieur du segment. L'approche 3, bien que tenant compte de la signification des valeurs négatives, n'arrive pas à toutes les compenser, et un décalage est aussi présent dans la courbe de proportion corrigée.

L'approche 4, qui est une approche globale dans le sens que la correction par lissage est réalisée sur toute la courbe des proportions, donne les meilleurs résultats et sera utilisée dans la suite de ce travail. Ici, l'élimination des valeurs négatives est compensée par une réduction du nombre d'objets à introduire à d'autres niveaux, et les proportions initiale et corrigée sont en phase.

Le tableau 5.1 compare les valeurs globales de l'intensité mathématique et celles obtenues après correction. Les approches 3 et 4 respectent assez bien cette intensité globale, comme cela a été imposé par construction, tandis que les approches 1 et 2 sont nettement plus éloignées. Cela se traduira par une importante différence dans le nombre d'objets à simuler dans le volume total.

Correction	Intégrale sur le domaine		Comparaison des proportions
	θ	θ^c	$ p - p_e^c $
Option 1	2.30	2.90	0.088
Option 2	2.30	2.98 ^(*)	0.281
Option 3	2.30	2.49	0.080
Option 4	2.30	2.40	0.032

TAB. 5.1 – Comparaison des valeurs globales de l'intensité mathématique et de l'intensité après correction pour chaque approche proposée. Le terme «global» fait référence à l'intégrale sur tout le domaine de calcul de l'intensité. Ce tableau montre aussi la comparaison, en termes de différence absolue, entre la proportion initiale et la proportion estimée avec l'intensité corrigée. (*) La valeur de l'option 2 est obtenue après la correction due à la translation réalisée.

Dans la section précédente il a été montré que si le choix de la loi des objets est approprié par rapport à la proportion initiale, la courbe d'intensité ne devrait pas présenter de valeurs négatives trop importantes et donc les corrections n'entraîneraient pas de fortes modifications dans les valeurs de l'intensité. Nous pouvons utiliser ce critère comme une indication de la qualité des résultats. Avec les paramètres introduits lors du test de sensibilité, notamment le pourcentage de valeurs négatives, la valeur moyenne de ces valeurs négatives, et la moyenne en valeur absolue entre la proportion initiale et la proportion estimée, il est possible d'avoir une idée, bien que qualitative, de l'erreur commise au cours du processus. Dans cet exemple le pourcentage de valeurs négatives dans l'intensité mathématique est de 35%, avec une valeur moyenne de -0.353 par rapport à une valeur moyenne d'intensité sur tout le volume de 0.047 . La courbe de proportion corrigée fournit une indication du meilleur résultat que l'on peut attendre avec cette intensité corrigée par rapport aux proportions initiales : elle constitue la **proportion du modèle**.

Proportions à forte variabilité et à valeurs nulles Nous allons maintenant regarder ce qui se passe quand la courbe de proportions présente des variations très brusques et atteint des valeurs nulles. Ce sont deux caractéristiques typiques qu'il faut respecter dans le modèle. Les paramètres du processus sont les mêmes qu'auparavant. La figure 5.13 montre le résultat obtenu. La courbe de proportion est respectée de façon globale mais les sauts d'une valeur à l'autre sont très lissés. Ceci est une conséquence directe du lissage effectué dans la courbe d'intensité, puisqu'il est réalisé sur un support égal à la taille du grain qui est très grand par rapport au volume total. Les valeurs nulles de proportions sont restituées sauf dans la zone d'impact du lissage.

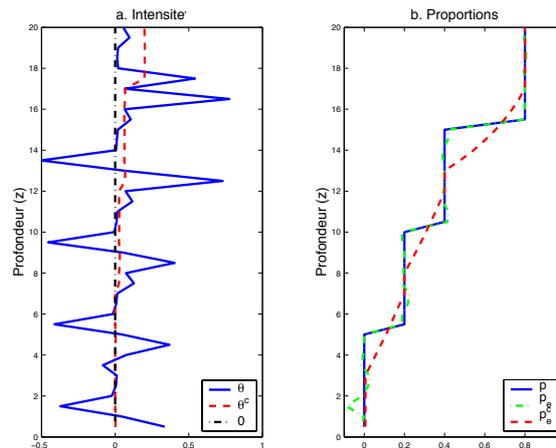


FIG. 5.13 – Résultats obtenus avec l'approche 4 dans un cas où la courbe de proportions présente de brusques changements et atteint des valeurs nulles.

Ces deux aspects (variations brusques et valeurs strictement nulles) sont délicats à analyser. D'une part nous pouvons nous interroger sur la représentativité de ces caractéristiques : les proportions ne sont que des estimations de la présence des faciès sur un certain support et sont, par définition, lissées. D'autre part, s'il y a des informations supplémentaires assurant ces changements brusques (la présence d'une faille, par exemple), on introduit une information déterministe et le problème devrait alors être abordé autrement. La présence de valeurs strictement nulles en est un exemple typique. Parfois on sait que dans une certaine région, le faciès étudié n'est pas présent et nous ne voulons donc pas introduire d'objets dans cette région lors des simulations.

Malheureusement, l'imposition de valeurs déterministes dans le résultat de déconvolution n'est pas un problème mineur. Il pourrait être envisageable d'introduire une contrainte dans le processus de déconvolution à travers les multiplicateurs de Lagrange, par exemple, comme c'est le cas pour certains problèmes en analyse d'image (González et Woods, 1992, p. 268-269). L'idée serait d'imposer que le résultat de la déconvolution soit positif et qu'il soit tel qu'en certains points il nous permette de restituer les sauts brusques des proportions initiales. Le problème réside dans la difficulté à trouver un moyen pour imposer ces contraintes, étant donné que l'on ne connaît pas a priori les points de l'intensité à valeurs négatives, ni les valeurs qu'elle doit prendre pour respecter les sauts des proportions.

En raison de ces difficultés il faudrait peut-être traiter les proportions avant la déconvolution. En effet, puisque l'on connaît l'impact du lissage sur l'intensité, lié à l'objet, et sur la proportion du modèle, on pourrait penser à modifier les valeurs de proportion dans la région d'influence du lissage et garantir ainsi la présence de valeurs nulles et de sauts dans la proportion du modèle (proportion corrigée).

5.2.2 Calcul du nombre d'objets à simuler

Une fois que l'intensité obtenue par déconvolution a été corrigée des valeurs négatives, elle peut être utilisée comme une intensité de Poisson qui détermine la distribution des objets dans le volume de simulation. Mais il faut aussi calculer le nombre d'objets à introduire pour restituer la proportion dans le volume. Dans un modèle booléen, le nombre d'objets $N(V)$ à implanter dans un volume V suit une loi de Poisson de moyenne :

$$N = E[N(V)] = \int_V \theta(x) dx$$

où $\theta(x)$ représente l'intensité de Poisson au point x du volume et V est le volume de simulation élargi pour tenir compte de la taille des objets et du processus de déconvolution (la définition de ce domaine V est détaillée dans la section suivante). Alors, si on connaît l'intensité dans tout le volume d'étude, le nombre moyen d'objets qu'il faut introduire dans la simulation s'en déduit directement. Pour simplifier les notations, θ fera par la suite référence à l'intensité corrigée.

Inférence 1D, 2D à 3D Supposons que pour représenter un certain faciès nous ne disposons que de l'information des proportions sur un plan à 2D, par exemple le long d'un affleurement, et que ce phénomène est considéré comme stationnaire selon la direction perpendiculaire. Dans ce cas le processus de déconvolution peut se réaliser sur le plan 2D. L'intensité à 2D (θ_{2D}) obtenue donne alors la distribution des objets selon ce plan et sera identique sur tous les plans parallèles le long de la direction perpendiculaire à ce plan. Cependant, puisque la simulation doit se réaliser sur un volume, il faut connaître le nombre moyen d'objets à simuler sur tout le volume. Dans Lantuéjoul et Simaku (1994), il est montré comment inférer ce nombre à partir d'une information 2D. À partir des formules stéréologiques et des propriétés de stabilité du modèle booléen selon lesquelles l'intersection d'un modèle booléen avec un plan (où une droite) est aussi un modèle booléen, il est possible d'inférer le modèle 3D à partir du modèle 2D. Le rapport entre θ_{2D} et θ peut s'exprimer par :

$$\theta_{2D}(x') = \theta(x) L_y^{obj}$$

où x' est un point du plan 2D de travail, XZ par exemple, x est un point du volume de simulation et L_y^{obj} est la traversée moyenne des objets selon la direction perpendiculaire à ce plan, Y , direction de stationnarité. Alors, le nombre moyen de points à simuler dans le volume est :

$$N = E[N(V)] = \frac{1}{L_y^{obj}} \int_V \theta_{2D}(x) dx = \frac{L_y^V}{L_y^{obj}} \int_{XZ} \theta_{2D}(x') dx'$$

où L_y^V est la longueur moyenne du champ selon la direction Y . Le nombre moyen d'objets peut donc être obtenu à partir de l'intensité 2D et en tenant compte de la taille des objets selon la troisième direction. L'intérêt de ces inférences est par exemple de permettre de réduire le problème de déconvolution de 3D à 2D (si l'hypothèse de stationnarité selon la direction perpendiculaire est valable) lorsque le processus est trop lourd en temps de calcul. Le raisonnement est analogue pour l'inférence du modèle à 2D ou 3D à partir d'un modèle 1D (voir Lantuéjoul et Simaku, 1994).

Objets réalistes Dans la section 4.2.1 nous avons détaillé le calcul de ρ pour deux familles d'objets très élémentaires. Nous avons vu que, même quand les objets sont géométriquement simples et peuvent être décrits selon des expressions paramétriques (cas des demi-ellipsoïdes), l'expression de ρ n'est pas nécessairement élémentaire. Dans certains cas il n'est pas possible de calculer une expression analytique.

Pour résoudre ce problème, nous essayons d'approcher les objets par des formes géométriques plus simples, des parallélépipèdes par exemple, de façon à pouvoir utiliser une expression relativement simple de p_0 . Le nombre d'objets à simuler doit être donc corrigé par le rapport des volumes entre l'objet à simuler et son approximation. Il est évident que l'intensité obtenue avec cette approximation ne correspond pas exactement à celle que nous aurions obtenu avec l'objet réel, mais nous pouvons espérer que l'erreur commise est acceptable.

Nous pouvons examiner un exemple à 1D pour étudier cette approximation. Nous disposons d'une courbe de proportion à 1D montrant une non stationnarité selon la direction verticale. La proportion présente un comportement stationnaire selon le plan horizontal. Puisque le faciès est modélisé par des objets 3D, il faut faire l'inférence du modèle à 1D en calculant le p_0 correspondant. Les objets 1D sont donc les traversées moyennes des objets 3D selon la direction verticale. Si le faciès est modélisé par de parallélépipèdes D de taille constante, de section carrée et de profondeur L , leurs traversées moyennes sont des segments verticaux, h_z , tous identiques et de longueur égale à L . Leur probabilité associée p_0 est très simple : $p_0(h_z) = 1$, pour $h_z \in [0, L]$ et zéro ailleurs.

Si ce même faciès est modélisé par des demi-ellipsoïdes A de taille constante et de section horizontale circulaire, les traversées moyennes ne sont plus constantes et leur probabilité associée est, dans ce cas : $p_0(h_z) = 1 - \frac{h_z}{L}$ si $h_z \in [0, L]$ et zéro ailleurs. La figure 5.14 montre la différence entre les deux familles d'objets.

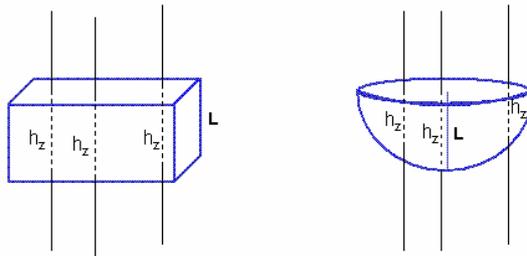


FIG. 5.14 – Intersections des objets parallélépipède (à gauche) et demi-ellipsoïde (à droite) avec des droites verticales. Le calcul de leur p_0 correspondant à 1D dépend de ces traversées. La traversée verticale maximale est égale à L . Les traversées du parallélépipède sont toutes identiques : $h_z = L$; dans le cas d'un demi-ellipsoïde ces traversées sont variables.

Supposons maintenant que le faciès est modélisé par les demi-ellipsoïdes mais en utilisant une probabilité p_0 correspondant aux plus petits parallélépipèdes qui contiennent ces demi-ellipsoïdes. La figure 5.15 montre graphiquement l'erreur commise en substituant un objet par un autre dans le calcul de l'intensité de Poisson. Les intensités obtenues avec l'objet demi-ellipsoïde et avec l'objet parallélépipède sont notées θ^A et θ^D respectivement. Ce sont des intensités corrigées. Les courbes d'intensité sont naturellement différentes puisque les objets sont de forme différente. La courbe θ^A a des valeurs plus fortes que θ^D : le volume du parallélépipède étant supérieur au volume du demi-ellipsoïde, le nombre d'objets nécessaires pour restituer les proportions initiales est moindre. Cependant, les deux courbes présentent le même comportement ; la distribution des objets est similaire dans les deux cas. Ceci se reflète dans les courbes de proportions estimées p^A et p^D qui sont très semblables et proches de la courbe de proportion initiale.

Il est alors acceptable d'approcher les objets réalistes qui vont modéliser le faciès par des objets relativement simples dans le calcul de l'intensité, à condition que cette approximation soit raisonnable : par exemple, approximer les demi-ellipsoïdes par de très petits parallélépipèdes conduirait à des résultats aber-

rants.

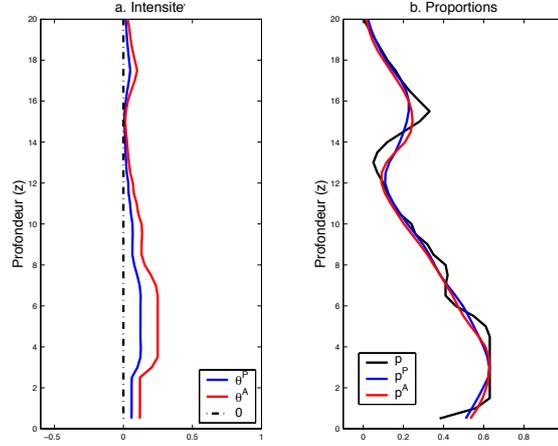


FIG. 5.15 – Intensités corrigées (à gauche) obtenues avec des parallélépipèdes (D) et des demi-ellipsoïdes (A). À droite une comparaison entre les proportions estimées avec chacune de ces intensités et la proportion initiale. Les parallélépipèdes sont les plus petits qui contiennent les demi-ellipsoïdes.

Pour déterminer le rapport entre le nombre d'objets à simuler associé à deux familles différentes, nous plaçons dans un cas général. Soit un faciès de proportion p qui va être modélisé par une famille d'objets A . Nous allons remplacer chaque objet A par un objet D . Si les objets A sont de taille variable, les objets D le seront aussi selon la même loi. Selon notre modèle, la proportion s'exprime en fonction de l'intensité de Poisson et de la capacité de Choquet associée à l'objet :

$$p(x) = T_X(\{x\}) = 1 - \exp\left(-\int_{\mathbb{R}^d} \theta^A(u) T_u^A(\{x\}) du\right)$$

où θ^A et T_u^A sont respectivement l'intensité de Poisson et la capacité de Choquet associées à l'objet A . Dans notre approche, nous substituons l'objet A par l'objet D en imposant que $\forall x$:

$$p(x) = 1 - \exp\left(-\int_{\mathbb{R}^d} \theta^A(u) T_u^A(\{x\}) du\right) = 1 - \exp\left(-\int_{\mathbb{R}^d} \theta^D(u) T_u^D(\{x\}) du\right)$$

Ceci se vérifie si, pour tout point u , les termes à l'intérieur de l'intégrale sont identiques. Dans ce cas :

$$\theta^A(u) = \theta^D(u) \frac{T_u^D(\{x\})}{T_u^A(\{x\})} = \theta^D(u) \frac{P\{x \in D(u)\}}{P\{x \in A(u)\}} = \theta^D(u) \frac{V_u^D}{V_u^A}$$

avec V_u^D le volume moyen de l'objet D implanté en u et V_u^A le volume de l'objet A implanté en u . Puisque ce qui nous intéresse est de déterminer le nombre moyen d'objets A à simuler dans le volume de simulation V , nous avons :

$$N^A = E[N^A(V)] = \int_V \theta^A(u) du = \int_V \theta^D(u) \frac{V_u^D}{V_u^A} du$$

Par hypothèse, les lois des objets ne dépendent pas de leur point d'implantation, donc nous pouvons exprimer le nombre moyen d'objets A à simuler en fonction du nombre moyen d'objets D et du rapport entre leurs volumes :

$$N^A = N^D \frac{V^D}{V^A} \quad (5.9)$$

Ce rapport est valable pour deux familles quelconque, mais, en pratique, pour que l'approximation dans le calcul de l'intensité soit acceptable, il est nécessaire que l'approximation de l'objet réel soit raisonnable.

5.2.3 Définition du domaine de calcul

Un autre aspect concernant la mise en œuvre de ce processus de déconvolution concerne la définition correcte du domaine de calcul et des supports des différentes variables. Il faut, d'une part, tenir compte des effets de bord et d'autre part, assurer la cohérence entre les différents supports. Ces deux points nécessitent un traitement des variables avant le processus de déconvolution. Ils sont détaillés par la suite.

Discretisation Les variables qui interviennent dans le problème (proportions, objets et intensité) sont des variables continues mais qui sont discrétisées dans la pratique. Ces valeurs se trouvent donc sur une grille. La discrétisation de la probabilité associée à l'objet, p_0 , est fixée lors du calcul de cette fonction (voir section 4.2.1, fig. 4.13) et détermine la précision dans la description de l'objet. La discrétisation de la proportion, p , dépend du domaine d'étude et correspond à la matrice de proportions qui est souvent calculée pour de grands blocs.

Afin de pouvoir réaliser le processus de déconvolution décrit précédemment, les grilles de p_0 et de p doivent avoir la même maille de discrétisation. Pour avoir une bonne définition des objets, il est nécessaire de travailler sur une grille fine et il faut donc discrétiser plus finement la grille de proportions. L'interpolation des valeurs des proportions est réalisée en considérant la proportion comme constante à l'intérieur de chaque bloc, mais il est tout à fait envisageable de réaliser un processus d'interpolation plus fin, par krigeage ou autre. L'intensité sera alors obtenue sur cette grille fine.

La figure 5.16 illustre à 2D ce processus de discrétisation fine. Il est complètement analogue pour le cas à 1D ou 3D. Le plus souvent cette discrétisation n'est nécessaire que sur le plan horizontal, puisque les proportions sont déjà initialement très finement calculées à la verticale.

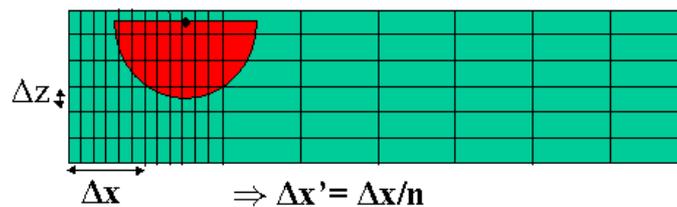


FIG. 5.16 – Discretisation fine de la grille de proportion. La maille de proportion est initialement $\Delta x \times \Delta z$. La discrétisation est donc réduite à $\Delta x' \times \Delta z$, avec $\Delta x' = \frac{\Delta x}{n}$ pour une valeur de n qui permet une précision suffisante dans le calcul de p_0 .

Effets de bord Sous le nom d'effets de bord, nous analysons deux contributions différentes. L'une tient compte de la signification physique des proportions et l'autre des effets mathématiques liés au processus de déconvolution.

La première de ces contributions concerne le fait que les proportions calculées sur un certain volume tiennent compte non seulement des objets contenus dans ce volume mais aussi des objets qui germent en dehors de la région définie par la grille des proportions et qui l'intersectent. Il est donc nécessaire de simuler sur un volume plus grand que celui de la grille de proportions pour prendre en compte cet effet. Afin d'obtenir l'intensité dans cette région étendue, les proportions ont été extrapolées par une valeur constante égale à la valeur des proportions aux limites de la zone d'intérêt (figure 5.17). La définition de la zone d'extrapolation dépend de la taille du grain et de son origine, i.e., de la localisation du germe à l'intérieur de l'objet. Dans l'exemple de la figure, seuls les objets qui germent à une distance maximale L_1 du bord supérieure de la région et à une distance $L_2/2$ des bords latéraux peuvent contribuer à la proportion.

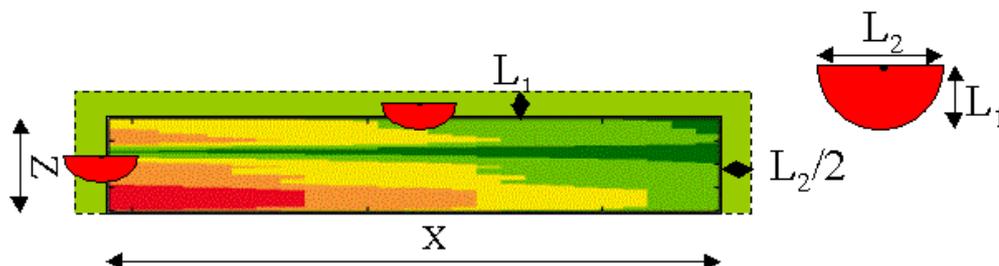


FIG. 5.17 – Extrapolation due aux objets qui germent en dehors de la région d'étude. Dans cet exemple à 2D, L_1 et L_2 sont les dimensions maximales de l'objet. La région d'étude initiale est $X \times Z$ et la région étendue à cause de cet effet est $(X + 2 \cdot \frac{L_2}{2}) \times (Z + L_1)$.

Le second effet de bord à prendre en compte est dû au fait que, dans le processus de déconvolution, il y a une diminution de la taille de la matrice d'intensité par rapport à la matrice des proportions (figure 5.18). Supposons que la proportion est sur une grille 2D de $M_1 \times M_2$ mailles, la grille de p_0 est de taille $L_1 \times L_2$ et celle de l'intensité est de taille $N_1 \times N_2$. Puisque la proportion est exprimée comme le résultat d'un processus de convolution entre l'intensité et p_0 , le rapport qui existe entre les dimensions des grilles est :

$$M_i = N_i + L_i - 1 \quad ; \quad i = 1, 2$$

Les dimensions de la grille d'intensité sont donc plus petites que celles de la grille de proportion. Comme nous avons besoin des valeurs d'intensité sur la grille dilatée précédente, il faut de nouveau extrapoler la grille des proportions dans chaque bord d'une distance de $\frac{L_i-1}{2}$. Dans un cas plus général, cette distance n'est pas forcément la même pour chaque bord puisque l'objet n'est pas nécessairement symétrique par rapport à son origine.

5.3 Conclusions. Avantages et limitations de la méthode.

Le rapport entre les proportions et les paramètres du modèle booléen peut s'exprimer, sous certaines conditions, comme une convolution. Nous pouvons donc obtenir l'intensité de Poisson du modèle par un processus de déconvolution.

Nous avons choisi, pour résoudre ce processus, de travailler dans le domaine de Fourier. L'utilisation des transformées de Fourier permet de traiter ce problème rapidement et efficacement. Ainsi, nous avons été amenés à intégrer des outils du domaine de la théorie du signal, le **filtre de Wiener**, pour obtenir le résultat de déconvolution. Dans ce contexte, nous avons supposé que les proportions sont affectées d'un

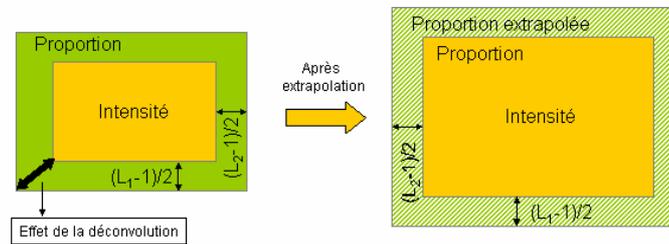


FIG. 5.18 – Extrapolation due à l'effet de la déconvolution. À gauche la diminution du support de l'intensité par rapport à la proportion, due à la déconvolution. À droite l'extrapolation introduite dans la proportion pour obtenir l'intensité dans le support qui nous intéresse, celui de la proportion.

certain bruit qui peut être interprété de deux façons : soit il est lié à l'erreur d'estimation et de régularisation des proportions, soit il est considéré comme un artefact stabilisateur dans la déconvolution.

Si on considère que le **bruit** est associé à l'erreur des proportions cette méthode permettra de le filtrer lors de la construction du modèle. Mais il faut encore pouvoir quantifier et modéliser le bruit. En l'absence d'informations autres que les proportions, l'estimation du bruit est impossible et il faut le fixer arbitrairement. Dans ce contexte nous avons préféré le considérer comme un élément fictif stabilisateur du processus de déconvolution et, en tant que tel, nous l'avons fixé le plus petit possible de façon à perturber le moins possible l'information contenue dans les proportions.

Dans cette hypothèse, nous sommes capables, pour toute proportion et à objet fixé, d'obtenir un résultat de déconvolution (**intensité mathématique**) qui nous permettra de restituer les proportions initiales. Cependant, il paraît naturel de penser que tout objet n'est pas en adéquation avec les caractéristiques de la proportion. Ainsi, des objets très grands ne seront pas utilisables pour reproduire une variabilité à petite échelle de la proportion. La proportion obtenue par cette nouvelle intensité (proportion « corrigée » dans le texte) sera légèrement plus lisse que la proportion initiale et constituera la **proportion du modèle booléen**.

Il aurait été envisageable de traiter le problème de déconvolution en imposant la contrainte de positivité de l'intensité, via le krigeage déconvolutif, par exemple (Daly, 1991 ; Jeulin et Renard, 1992). Des résultats satisfaisants ont été obtenus dans le domaine de l'analyse d'images microscopiques. Mais la principale difficulté de cette méthode réside dans le fait qu'elle requiert la modélisation du variogramme de la variable recherchée sous-jacente, à travers les variogrammes de la variable expérimentale (Seguret, 1988). Dans notre cas cela veut dire qu'il faut pouvoir modéliser le variogramme de l'intensité à travers le variogramme expérimental des proportions. Nous verrons dans le chapitre suivant que la modélisation des variogrammes (des proportions ou des indicatrices aux puits) pose d'importantes difficultés d'autant plus qu'il existe une non stationnarité et que les données sont peu nombreuses. Par ailleurs, un des enjeux de cette thèse étant une mise en œuvre robuste de la méthode, nous avons considéré qu'il était préférable de choisir une approche opérationnelle.

Le niveau de ressemblance que l'on doit exiger entre la proportion du modèle et la proportion initiale ouvre le débat sur le **dégré de représentativité** à accorder à cette dernière. Les proportions des faciès sont obtenues par estimation (krigeage, par exemple) à partir des données lithologiques aux puits. Bien que cette estimation soit réalisée à l'aide d'autres sources d'information (la géologie, la sismique, les essais de puits), elle ne représente que le comportement moyen du faciès dans le volume. Est-il alors raisonnable d'essayer

de la récupérer exactement dans notre processus ? On peut se demander si la présence d'une variabilité très forte à petite échelle ou de certaines valeurs nulles dans la proportion, par exemple, sont très représentatives de la réalité. Il est clair que nous voulons respecter l'information que les proportions apportent, mais l'utilisateur doit décider de la représentativité de ses proportions et de la qualité du modèle par rapport à ces données.

Pour finir, nous avons introduit l'idée d'utiliser des **objets différents** lors de la déconvolution et lors de la simulation. En effet, le faciès géologique à modéliser peut correspondre à des formes géométriques complexes. Ceci ne pose pas de problème théorique dans l'application du modèle booléen, en dehors des problèmes de conditionnement. Mais, dans notre approche d'inférence, il faut décrire ces objets à travers leur probabilité associée p_0 , ce qui n'est pas forcément un calcul élémentaire. Ainsi, pour simplifier le processus, nous avons proposé d'utiliser des objets relativement simples lors du calcul de p_0 bien que les objets à simuler soient plus réalistes.

Pour que cette approximation soit acceptable, il faut que les intensités correspondant à chaque objet ne soient pas trop différentes et pour cela les objets doivent être raisonnablement choisis. Ainsi par exemple, l'utilisation d'objets très petits comme approximation pour des objets très grands peut aboutir à des résultats non satisfaisants. De même il est important que le point d'implantation dans les deux familles d'objets soit similaire, puisque l'intensité résultante dépend de ce point.

Cette approche semble donner des résultats assez satisfaisants pour des objets élémentaires (un exemple sera présenté dans le chapitre suivant) et, si l'approximation est convenable, des objets plus réalistes pourraient ainsi être approximés.

La méthode proposée semble donner des bons résultats, comme nous allons le vérifier par la suite sur deux cas. Elle présente comme principal avantage le fait d'être simple à implémenter et rapide à calculer. Ceci est important puisque ce processus d'inférence a été conçu comme une étape préalable au processus de simulation qu'il ne doit pas alourdir.

Du point de vue **pratique**, les approximations introduites dans le calcul de p_0 et la correction des valeurs négatives constituent les sources d'erreur les plus importantes, mais nous allons voir dans les cas d'étude suivants qu'elles sont tout à fait acceptables.

Du point de vue **théorique**, la principale limitation est le fait d'utiliser uniquement des probabilités ponctuelles, ce qui introduit une incertitude dans la définition du modèle booléen. Mais cette incertitude peut aussi être considérée comme un avantage dans le sens où elle permet une certaine flexibilité dans le choix des objets : il est possible de restituer les mêmes proportions avec différents couples intensité - objet.

Il serait envisageable d'étudier des statistiques multipoints ou sur de supports plus complexes pour définir plus précisément le modèle. Le chapitre suivant est consacré à l'étude des variogrammes, donc à un support biponctuel. D'autres méthodes basées sur des aspects morphologiques (la méthode de Steiner (Serra, 1982), de Weil (Weil, 1988), de Schmitt (Schmitt, 1991)) ont été proposées pour estimer l'intensité poissonnienne. Le support (l'élément structurant K) utilisé dans l'étude de la capacité de Choquet n'est pas ponctuel mais doté d'un certain volume. Ces méthodes sont très appropriées en analyse d'images, où la réalité est considérée comme l'image, dans une fenêtre, d'une réalisation du modèle. L'information disponible est très abondante, ce qui permet de calculer les capacités de Choquet pour des éléments structurants complexes. Une application à un réservoir pétrolier (l'affleurement de Mesa Verde) a été aussi réalisée (Schmitt et Beucher, 1997) avec de bons résultats mais toujours avec la contrainte de disposer d'une image représentative (interprétation «complète» d'un affleurement analogue, par exemple), ce qui n'est pas le cas dans la plupart des situations.

Chapitre 6

Information sur les objets

Ce chapitre est consacré à l'analyse de l'information que les données expérimentales peuvent apporter sur la définition des objets. Le variogramme est utilisé pour quantifier l'information contenue dans les puits et les proportions.

Nous allons distinguer le cas stationnaire du cas non stationnaire. Les hypothèses de travail dans chaque cas ne sont pas les mêmes, ce qui va induire des démarches différentes.

6.1 Cas stationnaire

Dans le cas stationnaire la proportion du faciès est constante sur tout le volume et de valeur connue. Dans ce cas, le variogramme de l'indicatrice du modèle est accessible analytiquement, à partir de la valeur de la proportion et pour un couple intensité - objet donné. Le variogramme de la proportion, celle-ci étant une valeur constante, n'apporte aucune information.

La démarche suivie pour cette étude est la suivante : le variogramme du modèle est calculé pour différents couples intensité - objet. De cette façon on peut mettre en évidence l'impact des deux paramètres. Puis on se place dans une situation expérimentale et, à partir d'une réalisation d'un modèle donné, plusieurs puits sont extraits. Le but est d'étudier les variogrammes expérimentaux moyens de l'indicatrice du faciès aux puits et de rechercher l'information relative au modèle qui peut être obtenue. En d'autres termes, nous aimerions pouvoir modéliser les variogrammes expérimentaux par le variogramme du modèle.

Le cas étudié est très élémentaire par souci de simplification : un milieu stationnaire où le faciès est modélisé par des parallélépipèdes de taille constante. Seule la direction verticale sera considérée, l'étude des autres directions étant similaire.

Tout d'abord, rappelons (section 3.2.2.2) que le variogramme du modèle booléen X (et de son complémentaire X^c) peut s'exprimer, dans le cas stationnaire, par :

$$\gamma(h) = q(1 - qe^{\theta g(h)}) \quad (6.1)$$

avec q la porosité et $g(h)$ le covariogramme géométrique de l'objet A . Cette relation permet de déterminer $\theta g(h)$ à partir du variogramme :

$$\theta g(h) = \ln\left(\frac{1}{q} - \frac{\gamma(h)}{q^2}\right) \quad (6.2)$$

Puisque $\theta g(h) \geq 0$, on a $\left(\frac{1}{q} - \frac{\gamma(h)}{q^2}\right) \geq 1$ et, donc $\gamma(h) \leq q(1 - q)$ (variance des indicatrices). Le variogramme du modèle booléen ne peut donc pas dépasser la variance. L'expression (6.2) nous permettra

d'obtenir de l'information sur le covariogramme géométrique des objets donc à partir du variogramme. Si on observe une direction, le covariogramme géométrique correspond aux traversées moyennes des objets selon cette direction.

6.1.1 Étude des variogrammes verticaux du modèle

Le but de cette étude est la description du variogramme vertical moyen d'un schéma booléen et la détermination de l'influence de l'objet sur le variogramme du modèle. Il sera possible de comparer le comportement des variogrammes pour différents modèles par rapport à un modèle de référence.

Le modèle de référence correspond à un volume de $3750 \times 3750 \times 20 \text{ m}^3$. L'objet de référence est un parallélépipède, dont les axes sont parallèles aux axes des coordonnées, de taille $200 \times 200 \times 2.5 \text{ m}^3$. Notons que l'objet est relativement grand par rapport au domaine. Dans ce cas, la traversée moyenne de l'objet est égale à sa taille verticale. La probabilité associée au modèle, et donc la proportion du faciès sur le volume, est fixée à 25% ; l'intensité correspondante est de $2.88 \cdot 10^{-6}$. Le variogramme du modèle est donc parfaitement déterminé : on connaît la porosité, q , l'intensité θ et l'objet. Dans ce cas simple, nous pouvons calculer la covariance géométrique des objets $g(h)$. Celle-ci indique le volume de l'intersection des objets translétés d'une distance h selon la direction verticale :

$$g(h) = S \cdot (a - h)$$

où a est la taille verticale de l'objet, S est sa surface horizontale et h la distance verticale. Dans la situation de référence, le covariogramme géométrique de l'objet et le variogramme du modèle sont représentés à la figure 6.1. La portée du variogramme donne directement la taille verticale de l'objet.

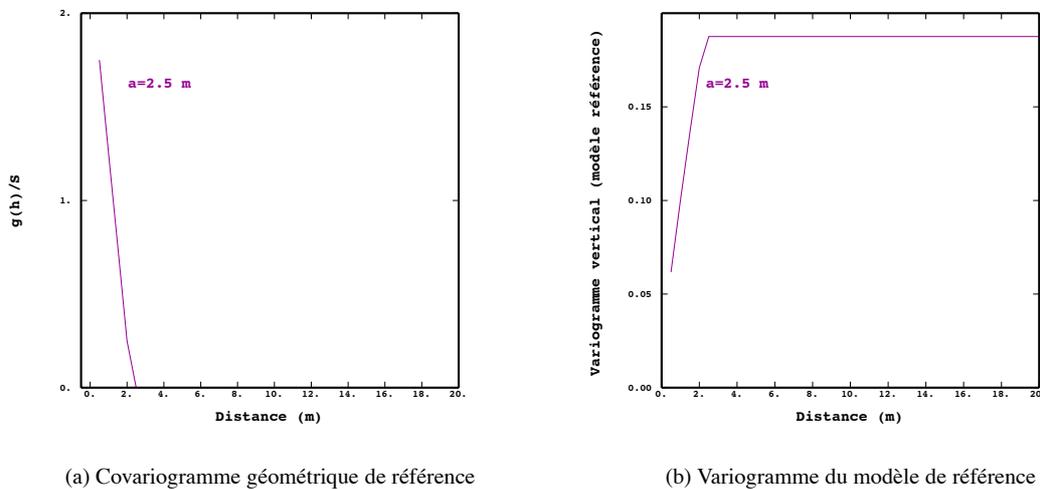


FIG. 6.1 – Covariogramme géométrique de l'objet de référence et variogramme du modèle de référence.

Afin d'étudier l'influence de la taille de l'objet sur le variogramme vertical du modèle, nous comparons plusieurs modèles créés avec différents couples intensité - objet. Pour tester la sensibilité du variogramme du modèle aux deux paramètres, deux situations différentes seront considérées :

1. Pour une même valeur de la proportion (proportion de référence) et des objets de tailles différents.

2. Pour une même valeur de l'intensité (intensité de référence) et des objets de tailles différentes.

6.1.1.1 Cas 1 : Même proportion du modèle et différents couples intensité - objet

Dans ce cas, la proportion est la même pour tous les modèles et égale à la proportion de référence (25%). On fait varier la taille verticale de l'objet, puisque seulement la direction verticale est étudiée. Une fois l'objet déterminé, l'intensité correspondante est calculée. Le tableau 6.1 présente les informations pour les différents modèles. Il est évident que, si la proportion du faciès est la même pour tous les modèles, quand la taille de l'objet augmente, moins d'objets doivent être introduits dans le modèle, ce qui se traduit par une valeur plus faible de l'intensité.

<i>Modèle</i>	<i>Taille verticale a (m)</i>	<i>Intensité</i>
Modèle 1.5	1.5	$4.79 \cdot 10^{-6}$
Modèle 2	2	$3.60 \cdot 10^{-6}$
Référence 2.5	2.5	$2.88 \cdot 10^{-6}$
Modèle 3.5	3.5	$2.06 \cdot 10^{-6}$
Modèle 5	5	$1.44 \cdot 10^{-6}$
Modèle 6.5	6.5	$1.11 \cdot 10^{-6}$
Modèle 8	8	$8.99 \cdot 10^{-7}$

TAB. 6.1 – Paramètres des modèles avec la même proportion de référence.

La figure 6.2 montre les covariogrammes géométriques correspondant à ces différents objets (fig. 6.2(a)) et les variogrammes verticaux des modèles (fig. 6.2(b)). Ils ont été obtenus en appliquant directement l'expression 6.1. Le changement dans la taille verticale des objets se traduit dans les variogrammes par une augmentation dans la portée et donc une diminution de la pente à l'origine, puisque le palier est le même dans tous les cas.

Chaque couple intensité - objet a un variogramme associé bien différencié, même si les proportions des modèles sont identiques. Le palier du variogramme indique la proportion moyenne du modèle et la portée indique la taille de l'objet selon la direction verticale. En pratique, si on se fixe la proportion dans le domaine étudié, la pente à l'origine du variogramme donne une indication de la taille des objets. Nous pouvons donc obtenir aussi l'intensité.

6.1.1.2 Cas 2 : Même intensité et objets de tailles différentes

Dans ce cas l'intensité reste identique pour tous les modèles et égale à l'intensité de référence ($2.88 \cdot 10^{-6}$). Nous faisons varier les tailles verticales des objets, et les modèles résultants ont des proportions associées différentes (tab. 6.2). Quand la valeur de l'intensité reste constante (ce qui se traduit par un nombre moyen d'objets à simuler identique dans tous les cas), plus les objets sont grands et plus la proportion du modèle est importante.

Les variogrammes verticaux des modèles sont calculés (fig. 6.3). Comme dans le cas précédent, il y a une augmentation de la portée du variogramme et une diminution de la pente à l'origine quand les objets augmentent en taille. Maintenant les paliers sont différents pour chaque modèle, ce qui traduit les différentes proportions. À partir des paliers des variogrammes, il est possible de calculer la proportion correspondant à chaque modèle, étant donné que la variance de l'indicatrice du modèle booléen dans le cas stationnaire est égale à $q(1 - q)$. Notons que, même si les proportions augmentent (modèle 8, $a = 8$ m), le palier ne dépasse

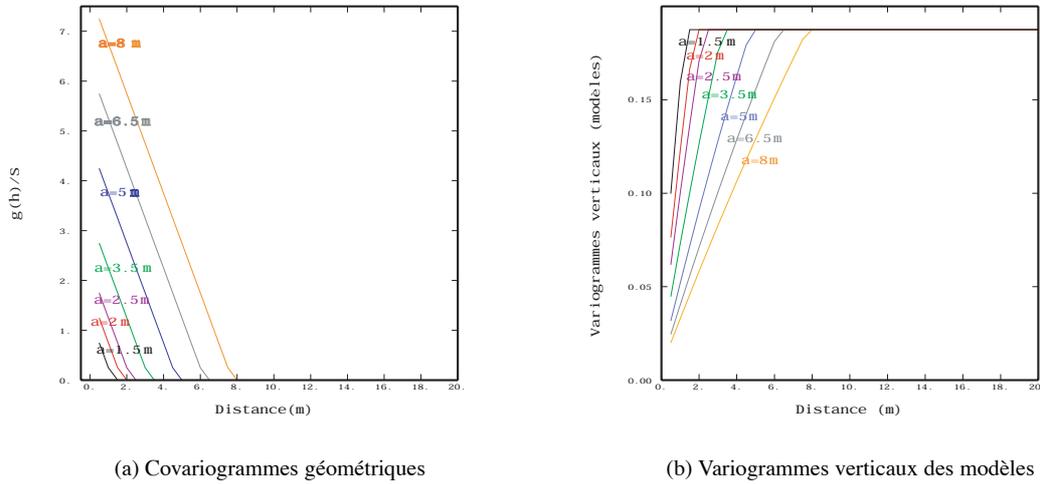


FIG. 6.2 – Covariogrammes géométriques des différents objets et variogrammes verticaux des modèles correspondants. La proportion reste constante pour tous les modèles, donc le palier est égal à $q(1 - q)$.

Modèle	Taille verticale a (m)	Proportion modèle
Modèle 1.5	1.5	16%
Modèle 2	2	21%
Référence 2.5	2.5	25%
Modèle 3.5	3.5	33%
Modèle 5	5	44%
Modèle 6.5	6.5	53%
Modèle 8	8	60%

TAB. 6.2 – Paramètres des modèles avec la même valeur d'intensité. Les objets ne varient que suivant la direction verticale.

pas le seuil de 0.25, valeur maximale de $q(1-q)$, la limite supérieure pour les variogrammes des indicatrices.

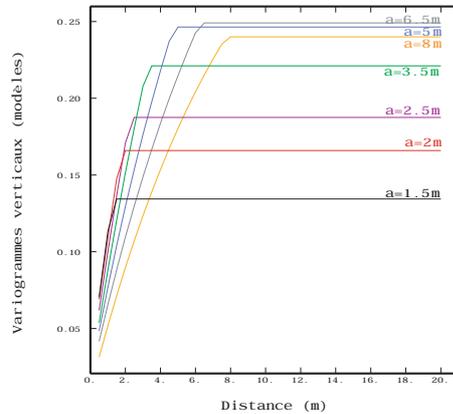


FIG. 6.3 – Variogrammes verticaux des modèles correspondant à une même intensité et à différentes tailles d'objets. La proportion varie pour chaque modèle.

Dans le cas stationnaire il devrait donc être possible, à partir des variogrammes expérimentaux, d'estimer les paramètres du modèle : la taille des objets et l'intensité. C'est l'objet de la prochaine section.

Dans un cas réel nous ne disposons en général que des indicatrices, en quelques puits, des faciès du dépôt sédimentaire. Nous avons choisi de modéliser ce dépôt par un modèle pour lequel on cherche des paramètres plausibles. Dans ce cas, l'idéal serait de pouvoir modéliser les variogrammes expérimentaux par le variogramme d'un modèle booléen. Il serait ainsi possible d'avoir une information supplémentaire sur les paramètres du modèle. Pour analyser ceci, nous allons nous placer dans une situation «pseudo-réelle», et extraire quelques puits d'une réalisation de chacun des modèles présentés ci dessus. Les variogrammes expérimentaux des indicatrices aux puits et aussi des indicatrices dans toute la réalisation seront étudiés.

On parle de situation «pseudo-réelle» puisqu'on est sûr que les données résultent d'un seul schéma booléen. On se place alors dans un cas favorable par rapport à la réalité puisque celle-ci peut être modélisée par différents schémas. Cette incertitude liée à la détermination des paramètres n'existe pas dans notre cas.

6.1.2 Étude des variogrammes verticaux expérimentaux : indicatrices des simulations et indicatrices aux puits

6.1.2.1 Cas 1 : Même proportion et différents couples intensité - objet

Pour cette analyse, il est nécessaire de réaliser des simulations de chaque modèle et de les discrétiser. La maille de simulation choisie est de $50 \times 50 \times 0.5\text{ m}^3$. Le tableau 6.3 montre les résultats obtenus. On remarque une diminution de la proportion moyenne et de la variance calculées dans les réalisations des modèles quand la taille des objets augmente. L'influence d'un seul grain, quand il est grand par rapport au volume total, est très importante sur une réalisation, et on est donc dans la limite de l'hypothèse d'ergodicité : plus les objets sont grands, plus leur influence sur la proportion globale moyenne est importante et plus les réalisations s'éloignent de la situation stationnaire.

Dans la pratique nous ne disposons pas d'une réalisation du modèle, mais des indicatrices aux puits. Dix puits ont été extraits d'une réalisation de chaque modèle. Ils sont distribués de façon non régulière dans le volume. L'information correspondante aux puits extraits des réalisations est dans le tableau 6.4. La proportion moyenne et la variance sont encore plus variables puisque l'information a été considérablement réduite.

<i>Réalisation</i>	<i>Proportion réalisation</i>	<i>Variance</i>
Modèle 1.5	24%	0.18
Modèle 2	24%	0.18
Référence 2.5	23%	0.18
Modèle 3.5	22%	0.17
Modèle 5	22%	0.17
Modèle 6.5	21%	0.17
Modèle 8	18%	0.15

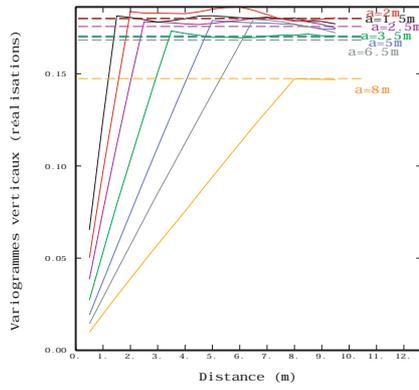
TAB. 6.3 – Information obtenue dans une réalisation de chaque modèle, la proportion théorique initiale étant la même (25%).

<i>Réalisation</i>	<i>Proportion puits</i>	<i>Variance</i>
Modèle 1.5	28%	0.20
Modèle 2	25%	0.19
Référence 2.5	25%	0.21
Modèle 3.5	22%	0.17
Modèle 5	23%	0.18
Modèle 6.5	25%	0.19
Modèle 8	23%	0.17

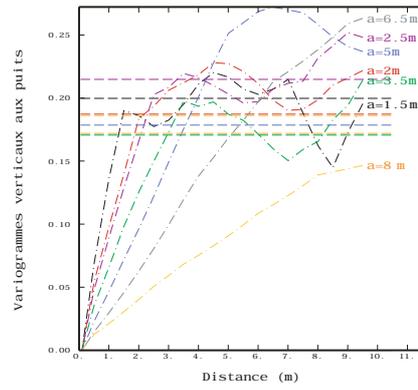
TAB. 6.4 – Information obtenue sur les dix puits extraits d'une réalisation de chaque modèle, la proportion théorique initiale étant la même (25%).

Les variogrammes des simulations (figure 6.4(a)) atteignent plus ou moins le même palier (excepté le cas du modèle 8), puisqu'il leur correspond la même valeur de proportion théorique. Les portées indiquent les tailles des différents objets utilisés. Les variogrammes expérimentaux des indicatrices des réalisations peuvent alors être modélisés assez bien par les variogrammes du modèle correspondant. Cependant, quand les objets sont relativement grands (modèle 8), une seule réalisation peut être assez différente du modèle et donc le variogramme expérimental sera plus éloigné de celui du modèle.

Les variogrammes expérimentaux des puits (figure 6.4(b)) sont très fluctuants et les paramètres sont difficiles à déterminer. Cependant, leur pente à l'origine pourrait donner une idée de la taille des objets selon la direction verticale : la pente diminue avec la taille de l'objet. La première rupture brusque peut servir d'indication de la taille de l'objet selon la direction verticale. Cette rupture indique la distance maximale d'influence du grain. Quand les objets sont grands (modèles 6.5 et 8), la rupture de pente est beaucoup moins marquée et les variogrammes ne se stabilisent pas autour de la variance. Cela s'explique par le fait que, à proportion constante, le nombre d'objets de grand taille est faible et certains puits ne rencontrent pas d'objet (figure 6.5).



(a) Indicatrices des simulations



(b) Indicatrices aux puits

FIG. 6.4 – Variogrammes verticaux expérimentaux des indicatrices dans les simulations et aux puits pour les différents modèles. La proportion moyenne théorique des modèles est la même pour tous les cas.

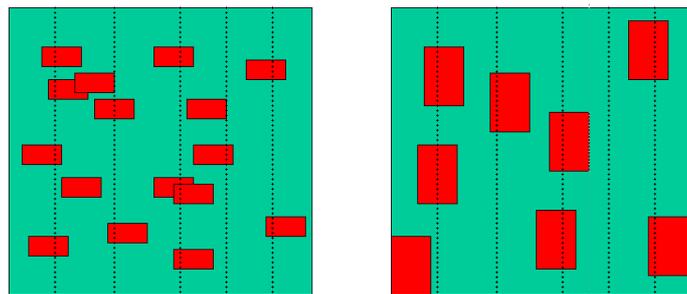


FIG. 6.5 – Effet du nombre d'objets.

6.1.2.2 Cas 2 : Même intensité et différents objets

Les tableaux 6.5 et 6.6 montrent l'information obtenue sur les réalisations et les puits pour ce cas.

Réalisation	Proportion réalisation	Variance
Modèle 1.5	16%	0.14
Modèle 2	20%	0.16
Référence 2.5	23%	0.18
Modèle 3.5	28%	0.20
Modèle 5	37%	0.23
Modèle 6.5	45%	0.25
Modèle 8	47%	0.25

TAB. 6.5 – Information obtenue pour une réalisation de chaque modèle. L'intensité théorique des modèles est la même ($2.88 \cdot 10^{-6}$).

Réalisation	Proportion puits	Variance
Modèle 1.5	15%	0.12
Modèle 2	24%	0.18
Référence 2.5	25%	0.21
Modèle 3.5	23%	0.18
Modèle 5	41%	0.24
Modèle 6.5	31%	0.22
Modèle 8	42%	0.24

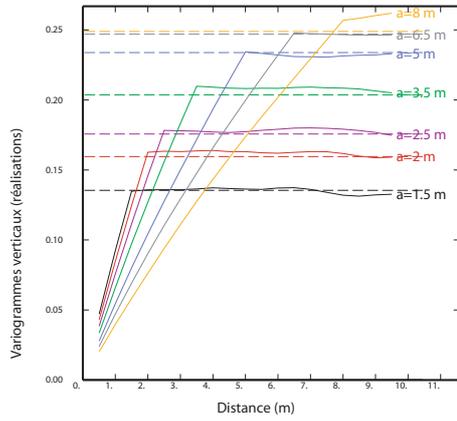
TAB. 6.6 – Information obtenue sur les dix puits extraits d'une réalisation de chaque modèle. L'intensité théorique des modèles est la même ($2.88 \cdot 10^{-6}$).

Les variogrammes calculés à partir des simulations (figure 6.6(a)) montrent des différences de palier dues aux variations de proportions. Le changement de portée correspond aux variations de la taille verticale des objets.

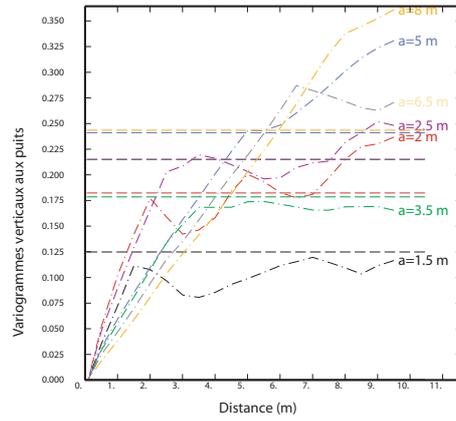
Les variogrammes des indicatrices aux puits (figure 6.6(b)), par contre, ne montrent pas si clairement ces différences de palier. Les portées sont plus difficiles à déterminer, spécialement quand les objets sont grands. Par contre on retrouve les premières ruptures de pente aux distances correspondant aux tailles des objets. On observe aussi une diminution des pentes à l'origine quand la taille verticale des objets augmente. Pour des objets de grande taille (modèles 6.5, 8), les variogrammes ne se stabilisent pas autour de la variance, et montrent un certain comportement non stationnaire. Plus les objets sont grands et plus leur recouvrement est important, puisque l'intensité est la même dans tous les cas (figure 6.7). Les traversées des objets ont plus de chances de correspondre à la superposition d'objets, ce qui fait que la première rupture de pente, correspondant à la région d'influence du grain, est moins marquée.

6.1.3 Conclusions

Les situations analysées ici se placent dans un cadre favorable : une situation stationnaire (proportions et intensités constantes) où les objets sont des parallélépipèdes de taille constante dans chaque cas. Les variogrammes des modèles, accessibles analytiquement, reflètent les deux paramètres : si la proportion moyenne



(a) Indicatrices des simulations



(b) Indicatrices aux puits

FIG. 6.6 – Variogrammes verticaux expérimentaux des indicatrices dans les simulations et aux puits pour les différents modèles. L'intensité des modèles est la même pour tous les cas.

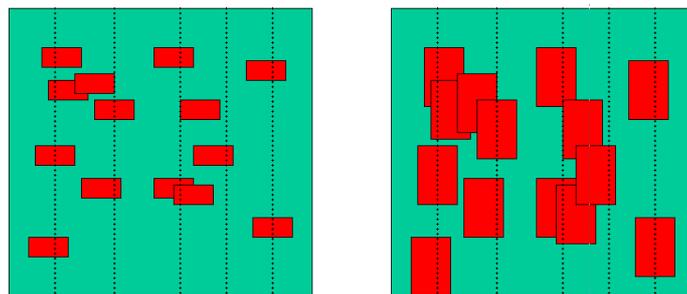


FIG. 6.7 – Illustration de l'effet du recouvrement des objets.

du modèle est connue, le palier et la portée nous permettent de calculer l'intensité de Poisson et la taille des objets respectivement.

Les **variogrammes des réalisations** reproduisent les caractéristiques des modèles utilisés dans chaque cas : la portée est liée à la dimension verticale des objets et le palier à la proportion du faciès. Il est possible alors de déduire les paramètres du modèle à partir d'une réalisation. Mais dans ce cas on dispose d'une grande quantité d'information. Ce n'est pas le cas dans une situation réelle où on ne dispose que des informations en quelques puits. Dans ce cas, l'échantillonnage n'est pas assez dense pour obtenir des **variogrammes expérimentaux des puits** assez stables et ils peuvent parfois présenter un comportement non stationnaire. Les paliers et les portées des variogrammes expérimentaux ne sont pas facilement identifiables. Cependant, dans tous les cas nous avons noté la diminution des pentes avec l'augmentation de la taille des objets, ce qui était aussi présent dans les variogrammes des modèles. Ceci permet de distinguer deux modèles avec des objets différents. La première rupture de pente semble aussi indiquer la région d'influence verticale d'un objet, c'est à dire la traversée verticale moyenne de l'objet.

À partir des données des puits, surtout quand ils sont peu nombreux, il est très difficile de remonter jusqu'aux paramètres du modèle, même dans un cas favorable où ces paramètres sont connus a priori et il existe une stationnarité. L'**analyse des variogrammes** atteint donc ses limites, d'autant plus qu'ils sont très sensibles à la taille des objets : à proportion fixée, quand ils sont grands par rapport au volume total, il sont peu nombreux et il se peut que quelques puits ne les rencontrent pas, donnant ainsi une fausse impression de non stationnarité. De plus, l'utilisation d'objets grands, bien que courant dans la modélisation des structures géologiques, est à la limite de l'hypothèse d'ergodicité du modèle, puisque deux réalisations d'un même modèle peuvent être très différentes entre elles, mais aussi très différentes du modèle, du fait que le volume étudié est fini.

6.2 Cas non stationnaire

À la différence du cas stationnaire, il n'est plus possible ici d'obtenir une expression analytique simple du variogramme du modèle booléen à partir de la proportion du modèle, de l'intensité et du covariogramme géométrique de l'objet. Dans le cas non stationnaire, la proportion varie dans le volume, donc l'intensité de Poisson est régionalisée. Le variogramme est maintenant dépendant non seulement de la distance entre deux points mais aussi de la valeur de la proportion associée au modèle en chaque point. La covariance non centrée s'exprime comme :

$$\begin{aligned} C(x, h) &= E[\mathbb{I}_{x \in \bar{X}} \mathbb{I}_{x+h \in \bar{X}}] = Q_X(\{x, x+h\}) = \\ &= \exp \left[- \int_{\mathbb{R}^n} \theta(u) T_0(\{x, x+h\}_{-u}) du \right] \end{aligned}$$

Il faut donc procéder d'une façon expérimentale. L'information disponible maintenant est, en plus des indicatrices aux puits, une grille ou matrice de proportion qui n'est plus une valeur constante. Nous pouvons donc calculer les variogrammes sur cette matrice de proportion et voir l'information sur le modèle qui peut en être obtenue.

Il est clair que, à intensités identiques, la cartographie des proportions d'un modèle construit avec des objets très grands sera plus lisse que celle des proportions d'un modèle construit avec des objets plus petits. Dans le premier cas, la région d'influence de l'objet sera plus importante que dans le deuxième, avec un effet de recouvrement également plus fort. Les proportions devraient donc apporter une information sur la

taille des objets. Cependant, il faudra tenir compte du fait que les proportions sont affectées d'un autre «effet de support» : leur support de calcul. Ces deux contributions ne sont pas facilement dissociables. Plus ces deux effets sont grands et plus la proportion sera lissée, ainsi que son variogramme. Il est aussi possible qu'un effet masque l'autre : si le support de calcul de la proportion est très grand, il peut cacher l'influence de l'objet, par exemple, et masquer l'existence d'une non stationnarité.

La démarche suivie dans cette étude est la suivante : calculer les variogrammes expérimentaux de différents modèles, avec plusieurs couples intensité - objet, et se placer ensuite dans un cas «réel» où, en plus de la matrice de proportions, quelques puits sont disponibles. Comme dans le cas précédent, seule la direction verticale est étudiée. La matrice de proportion est considérée comme un vecteur vertical et les objets ne varient que selon cette direction. Les paramètres des objets et de la simulation sont les mêmes que dans le cas stationnaire. Les mêmes situations sont analysées : pour une courbe de proportions fixée et pour une courbe d'intensité fixée.

6.2.1 Étude des variogrammes verticaux expérimentaux des proportions

6.2.1.1 Cas 1 : Même proportion et différents objets

L'étude des variogrammes est réalisée à 1D, comme dans le cas stationnaire. Les proportions sont donc sur une courbe verticale. À partir d'une courbe initiale et pour différents objets (tab. 6.7), on calcule les intensités par déconvolution. L'intensité est obtenue après un lissage correcteur, et la proportion estimée correspondante à chaque intensité sera alors différente de la proportion initiale, et correspondra à la proportion du modèle.

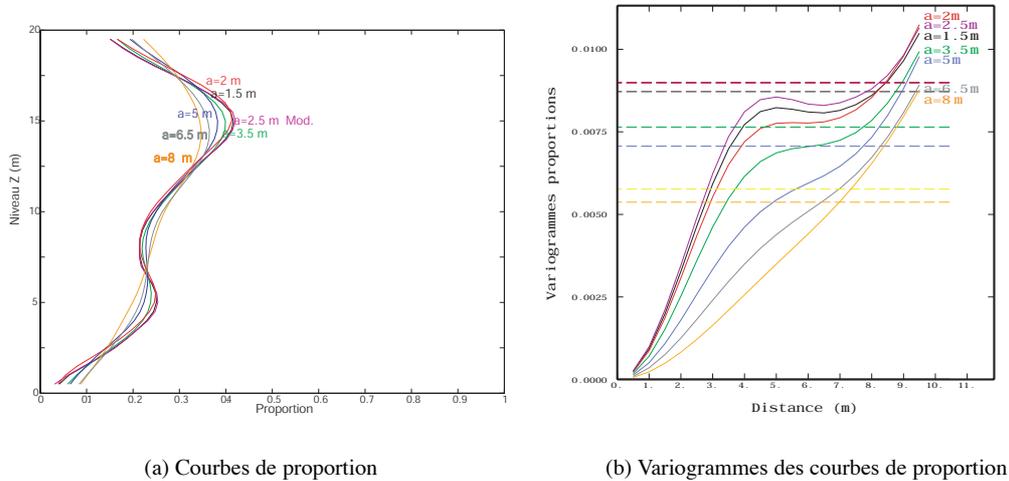
<i>Modèle</i>	<i>Taille verticale a (m)</i>
Modèle 1.5	1.5
Modèle 2	2
Référence 2.5	2.5
Modèle 3.5	3.5
Modèle 5	5
Modèle 6.5	6.5
Modèle 8	8

TAB. 6.7 – Tailles des objets pour les différents modèles. La courbe de proportion associée est identique dans tous les cas, et égale à celle de référence.

La figure 6.8(a) montre les courbes de proportions à partir des intensités obtenues avec chaque type d'objet. Nous avons donc différents modèles, selon le choix de l'objet, construits à partir de la même courbe de proportion initiale. A priori ces proportions estimées devraient être proches de la proportion initiale ($a = 2.5$ m), mais les effets de lissage et de l'influence de l'objet écartent les proportions des différents modèles de la proportion de référence.

Les proportions correspondant aux modèles construits avec des objets de taille allant jusqu'à $a = 5$ m sont très similaires ; au-delà de cette taille, les modèles résultants sont loin de la proportion initiale de référence. Il est alors possible de trouver une fonction d'intensité qui respecte la proportion de référence à une erreur près, pour toute une gamme d'objets.

L'étude des variogrammes de chacune des proportions (fig. 6.8(b)) montre leur dépendance vis à vis



(a) Courbes de proportion

(b) Variogrammes des courbes de proportion

FIG. 6.8 – Courbes et variogrammes des proportions obtenues avec différents couples intensité - objet. La proportion initiale de référence correspond à $a = 2.5$ m.

des paramètres des modèles respectifs. Il y a trois éléments qui interviennent dans les variogrammes des proportions : le volume sur lequel les proportions sont calculées (volume de régularisation), la taille de l'objet et le lissage introduit par la méthode de déconvolution. Ce dernier fait que les variogrammes des proportions correspondant aux intensités calculées sont systématiquement plus lisses que le variogramme de référence.

Au fur et à mesure que la taille de l'objet augmente, les variogrammes des proportions des modèles s'éloignent, dans leur forme, du variogramme de la proportion initiale ($a = 2.5$ m Mod.), notamment à partir de $a = 5$ m : ils deviennent plus lisses, ce qui implique que les objets sont trop grands pour être adaptés à la variabilité de la courbe de proportions. C'est justement à partir de $a = 5$ m que les courbes de proportion des modèles ressemblent moins à la courbe initiale.

Si l'on considère que le premier changement de pente dans le variogramme donne une indication de la zone d'influence moyenne de l'objet dans le comportement de la proportion, on pourrait placer cette rupture dans le variogramme de la proportion de référence autour de 4 – 4.5 m, ce qui nous donnerait une indication de la taille maximale de l'objet qui serait encore utilisable avec cette courbe de proportions. Ceci est cohérent avec les résultats montrés dans la figure 6.8(a) et avec le fait de respecter la forme du variogramme de la proportion initiale. Plus les objets sont petits et plus ils peuvent restituer une courbe de proportion quelconque. Plus les objets sont grands, au-dessus d'une taille limite indiquée par le variogramme de la proportion initiale, et moins ils peuvent restituer la courbe de proportion puisqu'ils entraîneraient un lissage plus grand que celui qu'on observe dans la courbe de proportion. Le problème principal consiste à fixer cette taille limite à partir des variogrammes des proportions. Les notions de portée et de palier dans un cas non stationnaire n'étant pas adéquates, fixer cette distance de changement de pente devient très approximatif.

6.2.1.2 Cas 2 : Même intensité et différents objets primaires

Nous allons comparer maintenant différents modèles obtenus avec une même intensité et des objets de différentes tailles. Puis, nous calculons les variogrammes expérimentaux des proportions correspondantes. Les proportions de chaque modèle (fig. 6.9(a)) reflètent la différence pour chaque objet utilisé. Plus celui-ci est grand et plus la proportion prend des valeurs fortes. La forme de la courbe de proportion devient aussi plus lisse.

Les variogrammes associés (fig. 6.9(b)) montrent aussi cet effet de lissage. La distance à laquelle le changement de pente se produit est plus importante pour les objets de grande taille. Ceci voudrait dire que, dans le cas où la courbe de proportions présente une variabilité moins marquée ou à plus grande échelle, la taille de l'objet maximal «compatible» avec la proportion est plus grande : une fourchette plus large de tailles est autorisée. Pour les cas présentés dans la figure 6.9(b) on pourrait placer ces distances à plus ou moins 9 m pour le cas $a = 8$ m, 7 – 7.5 m pour le cas $a = 6.5$ m, 5.5 – 6 m pour le cas $a = 5$ m, 4.5 m pour le cas $a = 3.5$ m, 4 – 4.5 m pour le cas $a = 2.5$ m, 3.5 – 4 m pour le cas $a = 2$ m et 3 – 3.5 m pour le cas $a = 1.5$ m. Ces distances sont très approximatives et peuvent être influencées par l'échelle de la représentation et par l'observateur. D'un autre côté, il faut tenir compte du volume de régularisation des proportions, qui interviendra fortement dans le lissage de la courbe de proportions : si ce volume est très grand alors il sera reflété dans le variogramme en tolérant des objets plus grands.

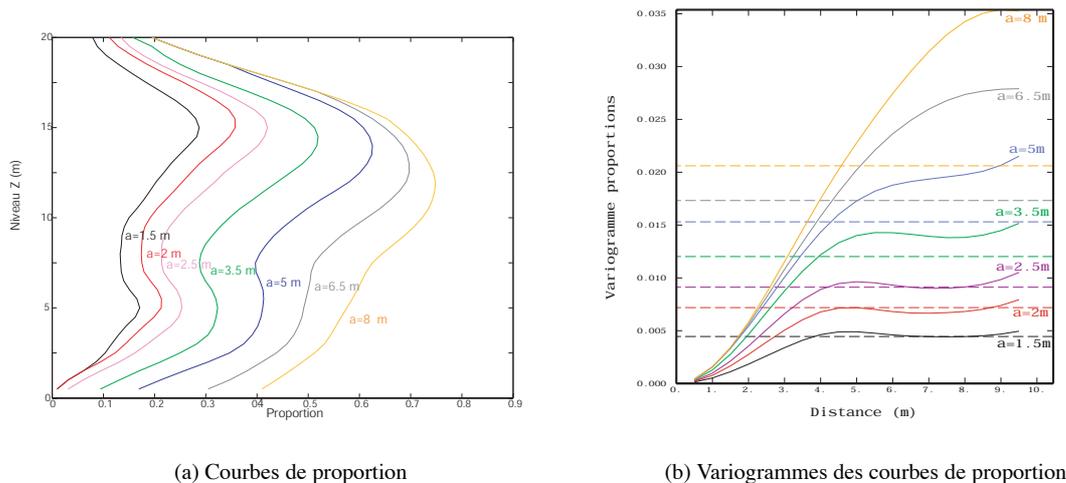


FIG. 6.9 – Courbes de proportion et variogrammes verticaux correspondant aux proportions obtenues avec une même intensité et différents objets.

6.2.2 Étude des variogrammes verticaux des indicatrices aux puits

Comme dans le cas stationnaire, si on dispose des indicatrices de faciès aux puits il est possible d'étudier les variogrammes verticaux expérimentaux de ces indicatrices. Dans notre étude particulière, nous calculons pour chaque modèle une réalisation sur laquelle quelques puits sont extraits. Les variogrammes expérimentaux pour toute la réalisation et pour les puits sont donc disponibles. Les deux cas décrits précédemment (même proportion et même intensité) sont considérés. Le volume et la grille de simulation restent inchangés.

6.2.2.1 Cas 1 : Même proportion et différents couples intensité - objet

Pour chaque couple intensité - objet nous avons une réalisation et un jeu de puits. Puisque les réalisations doivent toutes respecter la même proportion initiale, le nombre d'objets à simuler diminue quand leur taille augmente. Les variogrammes verticaux expérimentaux des réalisations sont dans la figure 6.10(a). Il est difficile d'identifier clairement dans ces variogrammes le comportement non stationnaire des modèles, puisqu'ils semblent se stabiliser autour de leurs variances respectives. Comme dans le cas stationnaire, la taille des objets se reflète dans les portées des variogrammes, et donc dans la diminution de la pente à l'origine avec l'augmentation de la taille.

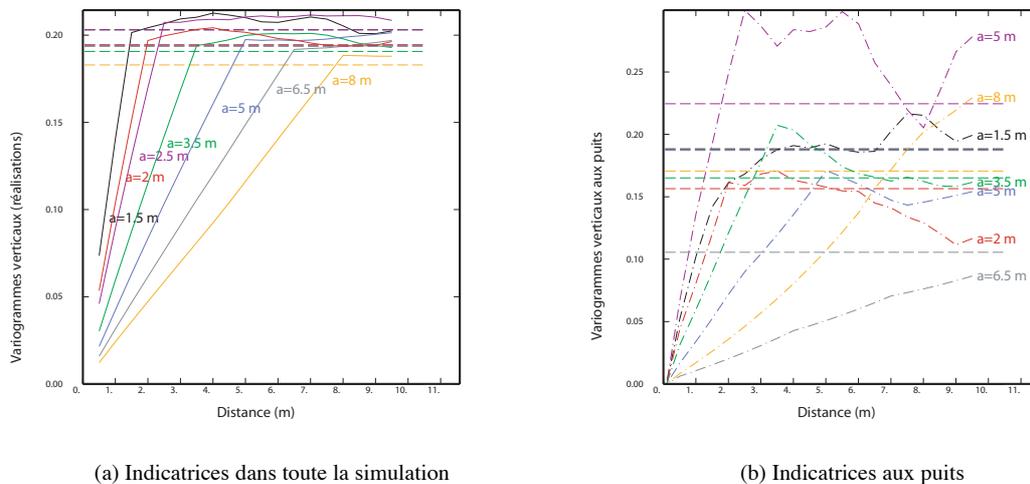


FIG. 6.10 – Variogrammes verticaux des indicatrices dans toute la réalisation et aux puits pour les différents modèles. La proportion moyenne théorique dans tous les modèles est la même.

Les variogrammes des indicatrices aux puits (fig. 6.10(b)) sont beaucoup plus fluctuants. Ils présentent tous de comportements très divers. Déterminer l'existence ou non d'une non stationnarité est difficile à partir de ces variogrammes, d'autant plus que dans un cas stationnaire les variogrammes des indicatrices aux puits (fig. 6.4(b)) peuvent être faussement associés à une non stationnarité. Tout de même, il est possible d'identifier la taille verticale de l'objet avec la première rupture brusque de pente dans chaque cas. L'effet de la diminution de la pente à l'origine avec la taille de l'objet est, ici, beaucoup moins évident.

6.2.2.2 Cas 2 : Même intensité et différents objets

La courbe d'intensité restant maintenant identique, le nombre d'objets à simuler et leur répartition sont les mêmes pour toutes les réalisations. La proportion atteindra des valeurs plus importantes avec les objets plus grands. Les variogrammes des indicatrices des réalisations (fig. 6.11(a)) donnent des résultats similaires à ceux du cas stationnaire : une augmentation des «paliers» pour les proportions globales plus fortes et une diminution de la pente pour les objets plus grands.

Les variogrammes des indicatrices aux puits (fig. 6.11(b)) présentent une rupture de pente à une distance plus ou moins égale à la taille de l'objet correspondant. La différence entre les différentes proportions est ici beaucoup moins marquée.

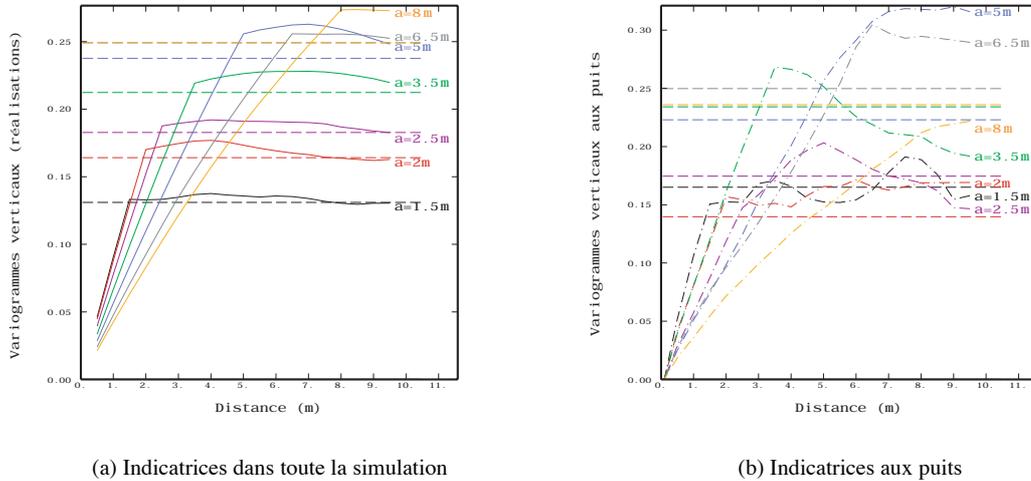


FIG. 6.11 – Variogrammes verticaux des indicatrices dans toute la réalisation et aux puits pour les différents modèles. L'intensité dans tous les modèles est la même.

6.2.3 Conclusions

Dans un cas non stationnaire, nous ne disposons pas d'une expression analytique simple du modèle de variogramme mais il est possible de calculer expérimentalement le **variogramme de la proportion du modèle**. Ces variogrammes sont très sensibles à la taille de l'objet et à l'influence du volume sur lequel les proportions sont calculées. Plus ces effets sont importants et plus le variogramme, ainsi que les proportions (lisses par construction), sont lisses. Souvent, le volume de régularisation n'est pas connu et son effet sur le variogramme ne peut pas être distingué de celui lié à l'objet. Il n'est donc pas possible de déterminer la taille de l'objet correspondant au modèle à partir du variogramme. Cependant il est possible de donner une taille maximale approximative à partir de laquelle l'objet ne nous permettra plus de retrouver une intensité correspondant à la courbe de proportions donnée. En d'autres termes, toute proportion peut être reproduite avec des objets petits, à la limite des points, qui permettront de reproduire la variabilité à petite échelle de la proportion. Plus les objets sont grands et moins ils sont utilisables pour restituer des proportions très variables. Il est donc intéressant de pouvoir donner une limite à la taille des objets à partir des proportions, ce qui constitue une information supplémentaire à celle dont on dispose par d'autres moyens. Toutefois, il faut tenir compte du fait que, si le support de régularisation des proportions est très important, comme c'est le cas souvent dans le plan horizontal, son effet cachera l'influence des objets et donnera des proportions et des variogrammes très lisses : si le support de régularisation est très grand par rapport à la taille des objets, alors la non stationnarité peut être masquée en donnant une fausse impression de stationnarité.

D'autre part nous disposons parfois de quelques puits dont les **variogrammes des indicatrices** pourront servir de guide pour la détermination de la taille des objets. Nous avons vu que tant dans le cas stationnaire que dans le non stationnaire, la distance d'influence de l'objet, et donc sa traversée moyenne (sa hauteur dans nos exemples), est marquée dans les variogrammes par une brusque rupture de pente. Ici, on s'est placé dans un cas où seulement la direction verticale est étudiée. Les données selon cette direction sont souvent assez nombreuses et permettent de calculer les variogrammes à petites distances. Dans le plan horizontal, ce

n'est pas le cas. Les données sont parfois rares et souvent distantes entre elles et ne nous permettent pas de déduire beaucoup d'information sur la taille des objets, d'autant moins quand il existe une non stationnarité horizontale.

L'utilisation des **variogrammes pour l'estimation des paramètres** du modèle booléen est très limitée, en particulier quand les données sont peu abondantes ce qui est, malheureusement, la situation la plus courante dans la modélisation de structures sédimentaires. Les variogrammes des indicatrices aux puits sont difficilement modélisables, même dans un cas stationnaire où le modèle est connu. La présence d'une non stationnarité complique encore le problème. Cependant, nous pouvons utiliser les variogrammes (aux puits dans le cas stationnaire, des proportions dans le cas non stationnaire) comme un test qualitatif et préliminaire sur les objets à utiliser dans le modèle.

Bien que l'analyse ait été réalisée expérimentalement sur quelques modèles et soit loin d'être exhaustive, nous pouvons **généraliser les résultats** dans le sens que, dans tous les cas, les changements brusques dans la pente des variogrammes (des proportions et aux puits) correspondent à de forts changements de la variable étudiée. Ainsi, nous pouvons extraire au moins une information sur la zone d'influence des objets dans le volume et donc avoir une idée a priori de la taille maximale qu'il sera possible d'utiliser. Si les traversées des objets selon une certaine direction sont variables (des demi-ellipsoïdes, des objets de taille variable), nous aurons une information sur leurs dimensions moyennes.

Troisième partie
Mise en œuvre

Chapitre 7

Application à un cas fictif 3D

Le premier cas expérimental pour tester la méthode proposée est la simulation d'hétérogénéités possédant une distribution non stationnaire à 3D. Ces hétérogénéités peuvent correspondre par exemple à des crevasses à l'intérieur d'un réservoir, à des hétérogénéités à l'intérieur d'un chenal, etc. Les figures 7.1 et 7.2 montrent deux exemples, à différentes échelles, du type d'images que l'on essaie de reproduire.



FIG. 7.1 – *Fragments de roches déposés, formant un «mosaïque», pendant l'écoulement d'un torrent d'eau (D'après [http : //course1.winona.msus.edu/csumma/FieldTrips/Death_Valley/DV4.html](http://course1.winona.msus.edu/csumma/FieldTrips/Death_Valley/DV4.html))*

Dans notre cas particulier nous allons supposer que les hétérogénéités modélisées par l'ensemble booléen sont des formations sableuses à l'intérieur d'une matrice argileuse, mais le cas contraire est aussi parfaitement envisageable : des lentilles argileuses à l'intérieur d'un faciès réservoir. Avec cet exemple nous voulons, d'un côté, tester la méthode proposée pour réaliser des simulations d'un modèle booléen dans un cadre non stationnaire, et de l'autre côté, illustrer l'importance de la prise en compte de la distribution des objets dans les simulations.

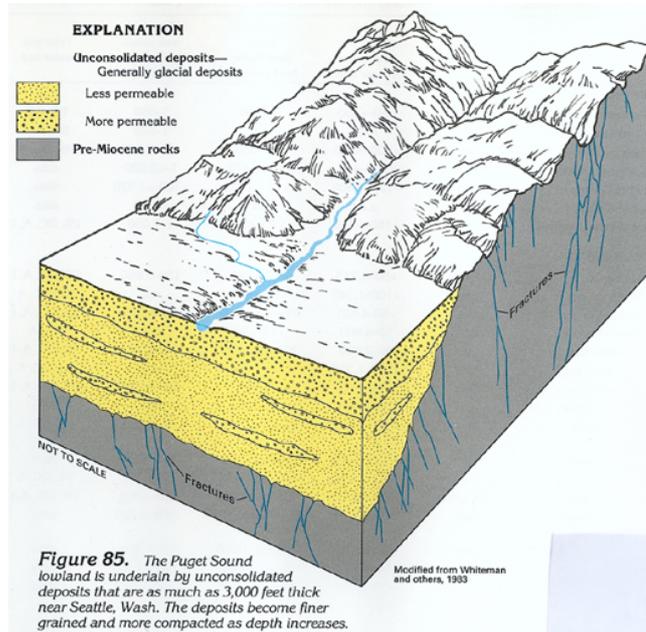


FIG. 7.2 – Dépôts de sable et des graviers constituant des sédiments non consolidés en forme de lentilles de différente perméabilité par rapport à la matrice qui les contient. (R. L. Whitehead, U.S. Geological Survey, dans http://capp.water.usgs.gov/gwa/ch_h/H-text2.html).

7.1 Description des données

Les données dont on dispose sont une grille de proportions à 3D indiquant la distribution du faciès sableux et de son complémentaire, la matrice argileuse. Nous fixons a priori les caractéristiques des objets pour simuler ce faciès réservoir. Dans ce cas fictif nous ne disposons pas de puits, les simulations seront donc non conditionnelles.

7.1.1 Proportions

Les paramètres de la grille de proportion sont :

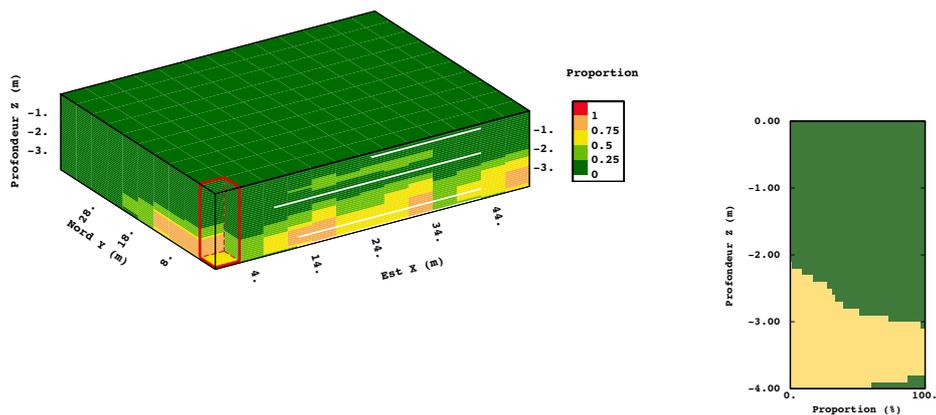
- Maille : $\Delta x = 4.00$ m, $\Delta y = 4.00$ m, $\Delta z = 0.10$ m ;
- Nombre de mailles : $N_x = 13$, $N_y = 10$, $N_z = 40$.

À chacun des nœuds de cette grille correspond une valeur de proportion qui donne le pourcentage du volume occupé par le faciès étudié sur chaque maille de volume $4 \times 4 \times 0.10$ m³. Le volume total de la région étudiée est de $52 \times 40 \times 4$ m³.

La figure 7.3(a) montre les données de proportions sur un bloc 3D. Pour mieux identifier la distribution du faciès, nous adoptons la représentation suivante : chaque colonne du bloc de proportion (en rouge dans la figure 7.3(a)) donne la distribution verticale de la proportion du faciès et peut être définie comme une courbe de proportion verticale (7.3(b)). Cette courbe (en jaune) représente la proportion du faciès présent dans chaque niveau sur une surface horizontale égale à la maille de la grille de proportion. La proportion est donnée en pourcentage (0% – 100%) ou comme une variable comprise entre 0 – 1. En même temps,

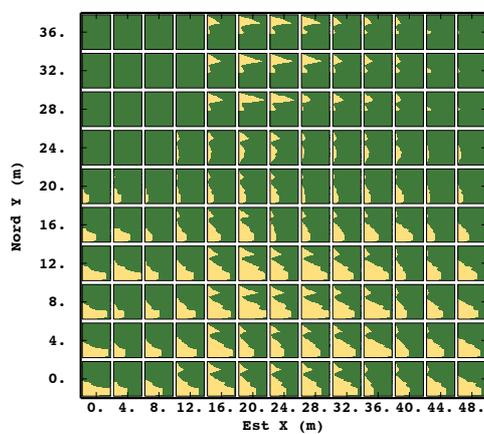
la proportion du faciès complémentaire (en vert) est aussi visible. Ainsi, il est possible de construire une «matrice» de courbes de proportions (7.3(c)) où chaque carré est informé par une courbe de proportion verticale. Dans cette image nous pouvons identifier les différents comportements aussi bien dans le plan horizontal XY que verticalement. Dans la moitié Sud du champ d'étude, la proportion du faciès sableux est plus élevée dans les niveaux profonds que dans les niveaux superficiels. La proportion tend à augmenter de l'Ouest vers l'Est, avec l'apparition progressive de ce faciès dans les niveaux supérieurs de la région centrale, puis une diminution de la proportion vers l'Est qui finit par disparaître dans ces niveaux superficiels. Dans la moitié Nord du champ la proportion du faciès est globalement inférieure à celle de la moitié Sud. De plus, la proportion de ce faciès est pratiquement concentrée dans une zone centrale et dans les niveaux supérieurs, avec une diminution des proportions vers l'Est. Au Nord Ouest il y a une région dans laquelle aucune présence de faciès réservoir n'a été détectée.

Pour illustrer un peu plus en détail la distribution de la proportion du faciès étudié, nous réalisons des sections sur lesquelles nous regardons la distribution de la proportion moyenne verticale (fig. 7.4 à 7.6). Celle-ci augmente de l'Ouest (axe négatif de X) vers les zones centrales, puis diminue vers l'Est (axe positif de X) (fig. 7.4). Cette évolution est plus accentuée au Nord du champ (fig. 7.4(c)). Les sections Nord - Sud, à $x = cte$, (fig. 7.5) montrent une diminution assez régulière des proportions moyennes du Sud (axe négatif de Y) vers le Nord du champ (axe positif de Y), jusqu'à pratiquement disparaître dans les extrêmes Est et Ouest du champ (fig. 7.5(c) et 7.5(a) respectivement). Dans la région centrale, les proportions moyennes se stabilisent au Nord autour d'une valeur presque constante. Quant aux sections diagonales (fig. 7.6), elles montrent une diminution générale du Sud vers le Nord (du SO vers le NE et du SE vers le NO).



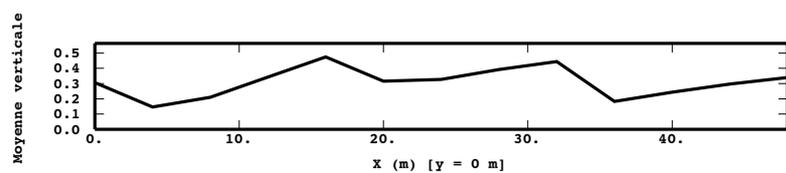
(a) Bloc 3D de proportion

(b) Courbe verticale de proportion

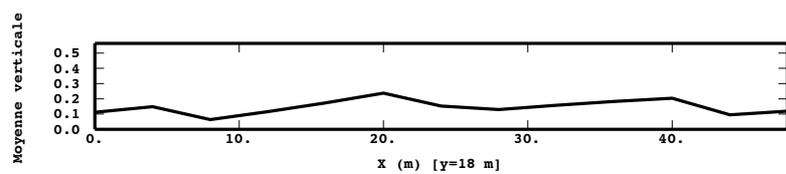


(c) Matrice des courbes verticales de proportion

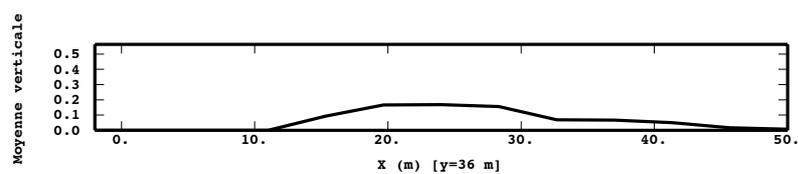
FIG. 7.3 – Description de la proportion initiale en 3D (figure 7.3(a)). Chaque colonne du bloc, comme celle détachée en rouge, peut être représentée comme une courbe de proportion verticale (7.3(b)). Cette figure montre, niveau par niveau, la distribution de la proportion, en pourcentage, du faciès étudié (en jaune) et du complémentaire (en vert). Nous pouvons donc représenter le bloc 3D de proportion comme une matrice de courbes de proportion verticale (7.3(c)).



(a) y = 0 m (Sud)



(b) y = 18 m (Centre)



(c) y = 36 m (Nord)

FIG. 7.4 – Moyennes verticales de la proportion le long de différentes sections Est - Ouest ($y = \text{cte}$) dans le plan horizontal.

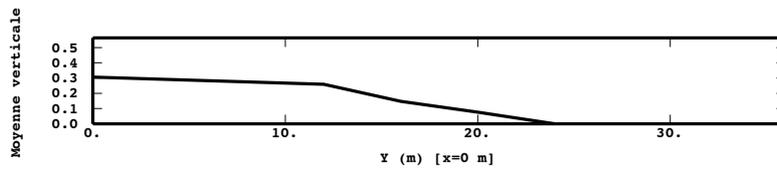
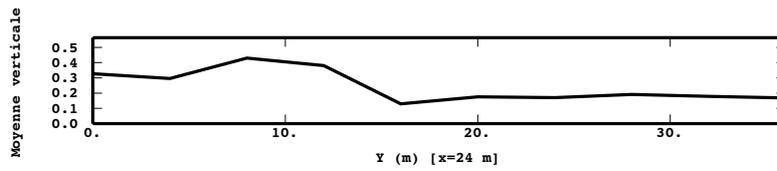
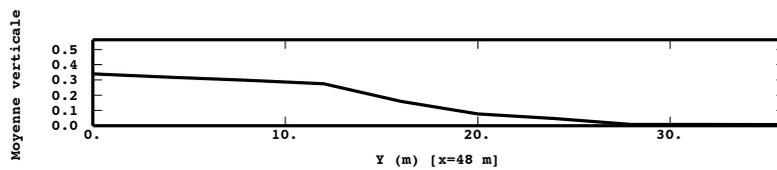
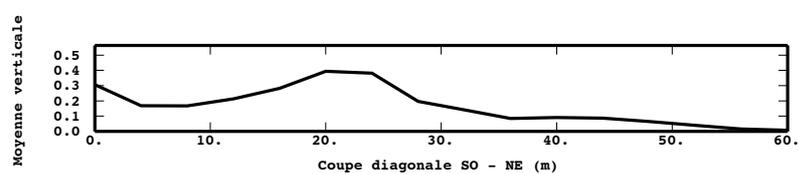
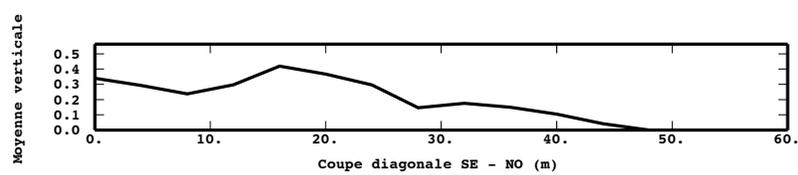
(a) $x = 0$ m (Ouest)(b) $x = 24$ m (Centre)(c) $x = 48$ m (Est)

FIG. 7.5 – Moyennes verticales de la proportion le long de différentes sections Nord - Sud ($x = cte$) dans le plan horizontal.



(a) Section diagonale SO - NE



(b) Section diagonale SE - NO

FIG. 7.6 – Moyennes verticales de la proportion le long de deux sections diagonales dans le plan horizontal.

7.1.2 Description des objets

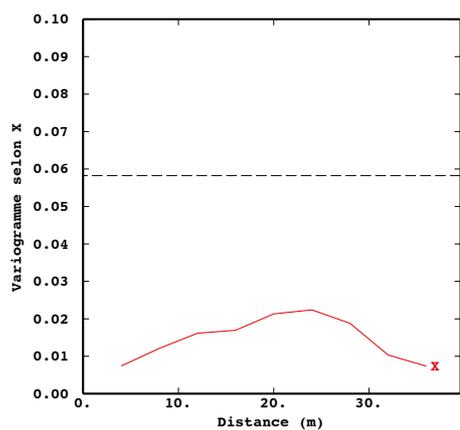
Dans un cas réel, l'information sur les objets qui modélisent le faciès est obtenue à partir des puits et éventuellement d'affleurements analogues, et/ou d'images sismiques. Dans notre approche nous avons fait l'hypothèse que les caractéristiques des objets sont fixées a priori. Nous avons vu dans le chapitre 6 qu'il est possible de vérifier si le choix de la taille des objets est compatible avec les proportions.

Dans notre exemple, le faciès étudié est modélisé par des demi-ellipsoïdes dont les demi-axes sont de taille variable selon des lois uniformes et indépendantes :

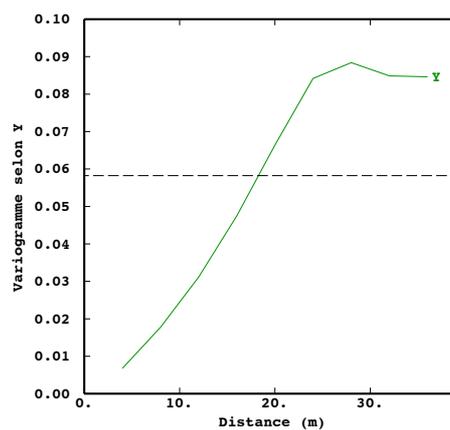
$$\begin{aligned}l_x &\sim U_{[1\text{ m}, 2\text{ m}]} \\l_y &\sim U_{[1\text{ m}, 3\text{ m}]} \\l_z &\sim U_{[0.2\text{ m}, 0.5\text{ m}]}\end{aligned}$$

Ces dimensions, bien que relativement importantes par rapport au volume total du champ, sont compatibles avec les variogrammes expérimentaux de la grille de proportion (fig. 7.7). Ceux-ci présentent un comportement non stationnaire, notamment selon les directions Y et Z , où la différence entre les valeurs des proportions augmente assez rapidement avec la distance. Selon la direction X , cette augmentation est plus lente. Les premières ruptures de pente des variogrammes se produisent autour de 1 m dans la direction verticale ; selon la direction X , elle se situe plus ou moins à 12 m et selon Y , elle se situe vers la moitié du champ. Nous pouvons ainsi avoir une idée de la région maximale d'influence que peut avoir un objet pour restituer la variabilité indiquée par les proportions. En effet, les changements brusques dans la distribution de la proportion présentée dans les différentes coupes horizontales (10 – 15 m selon X (fig. 7.4), jusqu'à 15 m selon Y (fig. 7.5)) ne pourront pas être reproduits avec des objets plus grands.

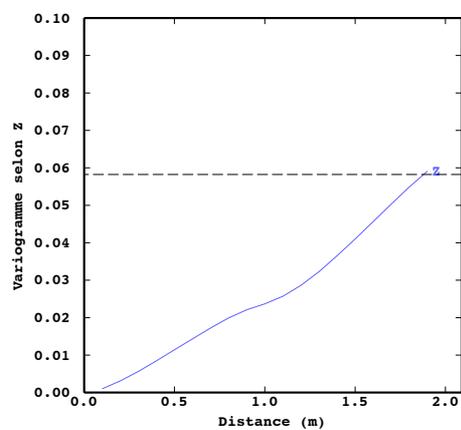
Nous avons vu dans le chapitre 5 que l'expression analytique de η_0 peut être très compliquée, voire inexistante, même pour des objets relativement simples, comme c'est le cas dans cet exemple. Nous avons simplifié ce problème en calculant la probabilité η_0 correspondant aux parallélépipèdes les plus petits qui contiennent les demi-ellipsoïdes et en corrigeant le nombre moyen d'objets à simuler par le rapport entre les volumes respectifs. Les dimensions des demi-côtés des parallélépipèdes suivent les mêmes lois que les demi-axes des demi-ellipsoïdes. L'expression de η_0 a été déjà calculée dans la section 4.2.1.



(a) Direction X (EO)



(b) Direction Y (NS)



(c) Direction Z (Profondeur)

FIG. 7.7 – Variogrammes expérimentaux de la proportion initiale selon les trois axes principaux de la grille.

7.2 Obtention de l'intensité. Simulations

Comme il a été expliqué dans le chapitre 5, lors du processus de déconvolution permettant d'obtenir l'intensité de Poisson, il faut modifier la grille de proportion initiale par interpolation et extrapolation successives. La grille dans laquelle l'intensité est obtenue est :

- Maille : $\Delta x = 1.00$ m, $\Delta y = 1.00$ m, $\Delta z = 0.10$ m ;
- Nombre de mailles : $N_x = 57$, $N_y = 47$, $N_z = 45$

et le volume réel de simulation est de $57 \times 47 \times 4.5$ m³.

L'intensité obtenue avec le processus de déconvolution décrit et après correction est représentée figure 7.8. Ce «cube» d'intensité sera utilisé dans le processus de simulation comme l'intensité de Poisson, et déterminera la distribution des objets dans le volume. Le nombre moyen d'objets parallélépipèdes à introduire dans la simulation est de 832. Une fois cette valeur corrigée, le nombre moyen d'objets demi-ellipsoïdes à simuler est de 1589.

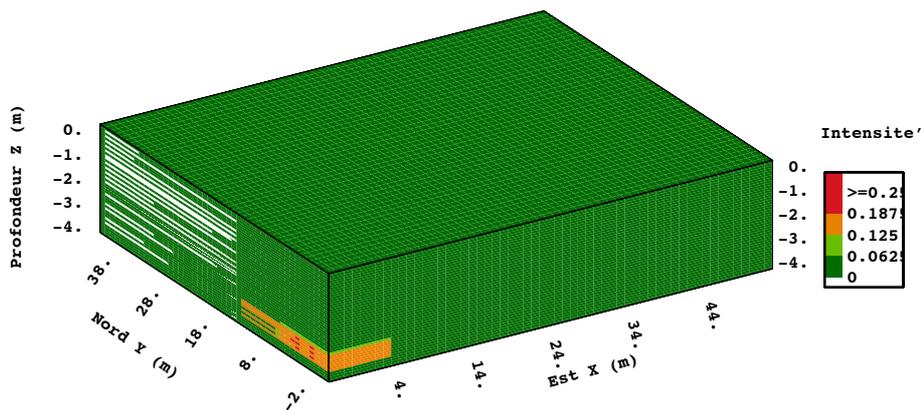


FIG. 7.8 – Bloc 3D de l'intensité obtenue par déconvolution et après correction. La fonction p_0 utilisée correspond à des parallélépipèdes.

L'intensité utilisée est affectée principalement de deux approximations par rapport à l'intensité théorique :

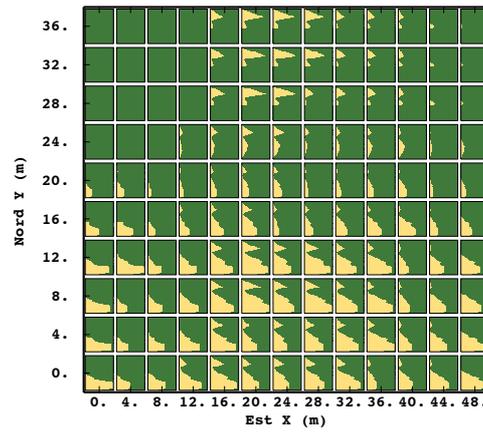
- un effet dû au lissage pour la correction des valeurs négatives et
- une approximation due au fait que la déconvolution et les simulations ont été réalisées avec des objets différents.

Le pourcentage des valeurs négatives qui ont été corrigées est de 48%, ce qui est assez important. Ces nombreuses valeurs négatives sont induites par l'importante présence dans les données de proportion de valeurs strictement nulles (34% des valeurs) ou très faibles (41% des valeurs de proportion sont inférieures à 4%).

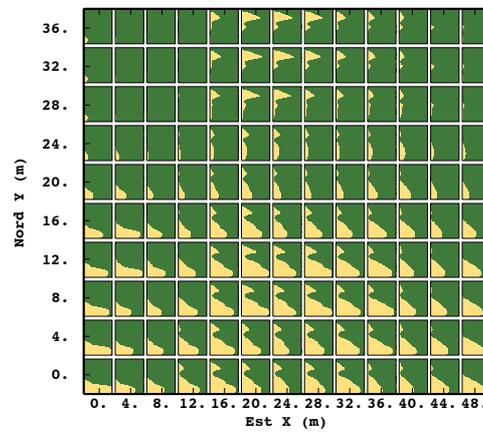
La proportion effective qui correspond à cette intensité corrigée et qui constitue la proportion de notre modèle est très proche de la proportion initiale calculée sur les données (figure 7.9). La différence moyenne ponctuelle, en valeur absolue, entre les deux est de 5%. Comme espéré, la proportion du modèle est légèrement plus lisse que la proportion initiale, avec une tendance à surestimer les valeurs très faibles et à sous-estimer les valeurs très fortes. Ce résultat est aussi reflété dans les histogrammes des proportions (fig. 7.10) et dans le nuage de corrélation (fig. 7.11) entre les deux variables. Ce dernier tend à se situer au dessus de la première bissectrice pour les valeurs de la proportion initiale proches de 0% (le modèle donne entre 0% et 17%) et en dessous pour les valeurs proches de 100% (entre 99% et 68% dans le modèle).

Il faut souligner que les valeurs nulles de proportion dans la région Nord Ouest ne sont pas complètement récupérées dans le modèle. L'existence de valeurs de proportion strictement nulles rejoint le problème de la représentativité des proportions évoqué dans le chapitre 5 et sera traité ultérieurement.

Les caractéristiques principales de la proportion initiale décrites précédemment se retrouvent dans la proportion du modèle, et nous pouvons donc considérer que les approximations réalisées sont acceptables.



(a) Données



(b) Proportion du modèle

FIG. 7.9 – Grilles de proportion correspondant aux données et à la proportion résultant des paramètres du modèle (intensité « corrigée » et objet parallélépipède).

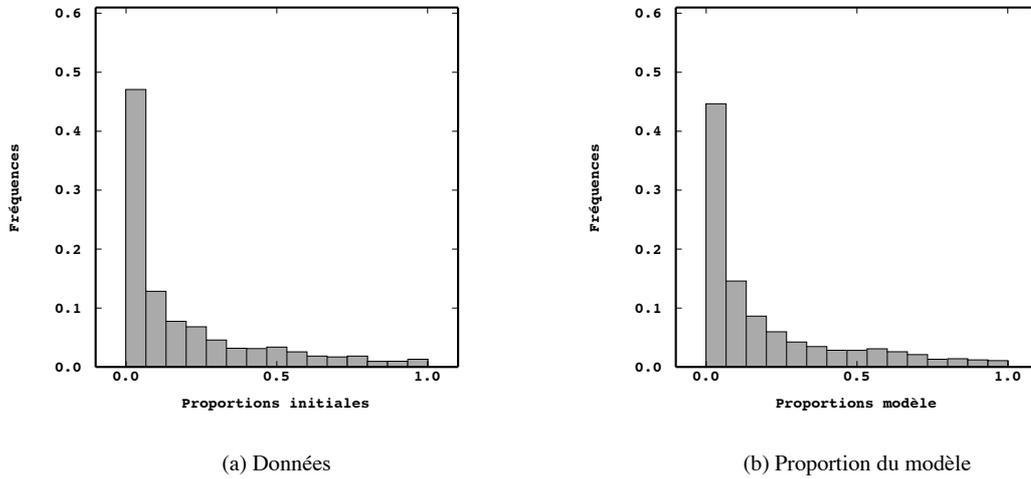


FIG. 7.10 – Histogrammes des valeurs de la proportion initiale (a.) et de la proportion du modèle (b.).

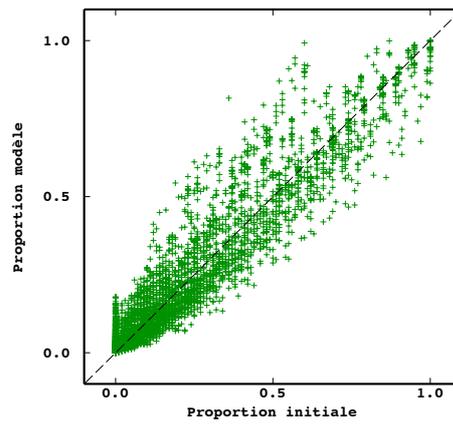


FIG. 7.11 – Nuage de corrélation entre la proportion initiale et la proportion du modèle.

Simulations. Objets demi-ellipsoïdes

Maintenant tous les paramètres nécessaires pour réaliser les simulations sont connus : les caractéristiques des objets qui modélisent le faciès, l'intensité de Poisson qui détermine la façon dont les objets vont être distribués, et le nombre moyen d'objets à introduire dans la simulation. Les paramètres de la grille de simulation sont :

- Maille : $\Delta x = 0.50$ m, $\Delta y = 0.50$ m, $\Delta z = 0.05$ m ;
- Nombre de mailles : $N_x = 104$, $N_y = 80$, $N_z = 80$.

La figure 7.12 montre un exemple de simulation obtenue.

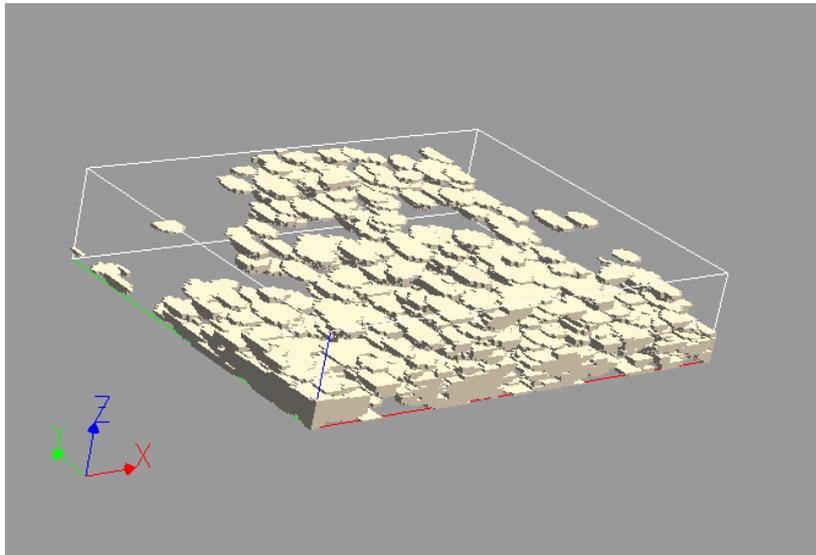


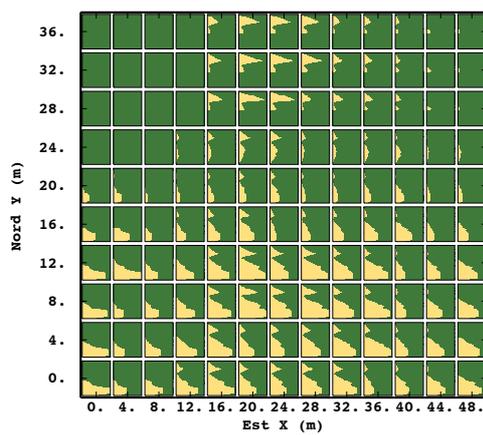
FIG. 7.12 – Exemple de simulation obtenue avec l'intensité calculée par déconvolution.

Pour déterminer la qualité des simulations, nous allons comparer les proportions initiales (les données) aux proportions calculées sur les simulations. Cette comparaison est faite d'une part sur la ressemblance graphique et d'autre part sur les valeurs des proportions «globales». On appelle proportion «globale» la fraction totale du volume occupée par le faciès considéré. Cette valeur détermine la quantité globale, en pourcentage, de faciès que l'on s'attend à trouver dans la région et constitue parfois une contrainte importante qu'il faut reproduire dans les simulations.

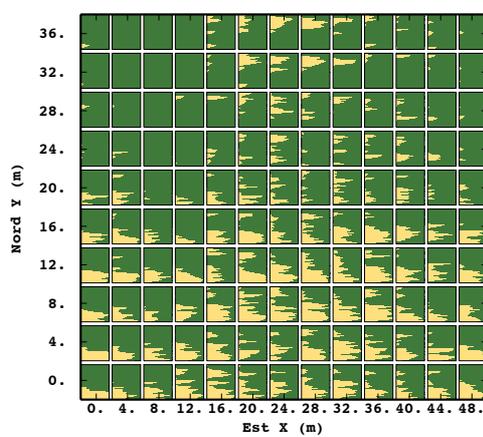
Les matrices de proportion correspondant à un exemple de simulation et à la proportion initiale sont présentées à la figure 7.13.

Comme on s'y attendait, les proportions de la simulation sont assez proches des proportions du modèle (fig. 7.9(b)) mais les proportions initiales sont aussi assez bien respectées, en dehors de la variabilité intrinsèque associée aux simulations. Les mêmes caractéristiques générales, hormis les valeurs nulles de proportion dans la région NO, se retrouvent dans la simulation.

La valeur globale de la proportion est relativement bien retrouvée : 19.0% dans la simulation pour 18.3% des données et 19.4% de la proportion du modèle. L'histogramme de la proportion calculée sur la simulation (fig. 7.14) récupère assez bien l'histogramme de la proportion initiale, sauf pour deux aspects : les valeurs très faibles et très fortes sont plus abondantes dans la proportion de la simulation. Les valeurs intermédiaires, notamment entre 0 et 25% sont moins présentes dans la simulation. Cela s'explique par le fait que les objets



(a) Données



(b) Proportion de la simulation

FIG. 7.13 – Matrices de proportion correspondant aux données (a.) et à la proportion calculée sur une simulation (b.).

simulés sont de taille comparable à la maille des proportions. La présence ou l'absence d'un objet a une grande influence sur la proportion calculée sur une maille.

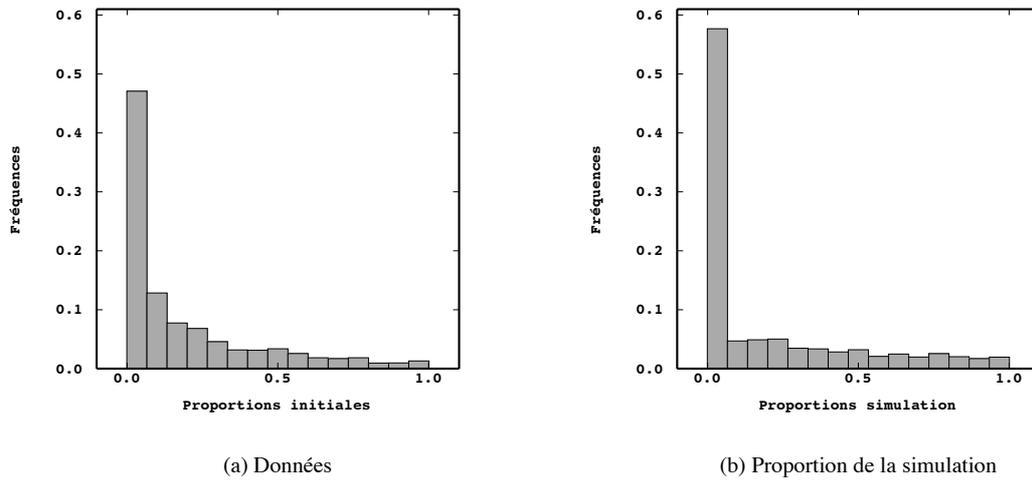


FIG. 7.14 – Histogrammes des valeurs de la proportion initiale et de la proportion calculée sur la simulation.

Sur un échantillon de 30 simulations, nous avons pu mettre en évidence les mêmes comportements. Les figures 7.15 à 7.17 montrent les proportions moyennes verticales pour différentes sections horizontales pour quelques simulations. Nous avons ainsi vérifié le lissage introduit par la correction de l'intensité : les changements brusques de la proportion initiale (en noir dans les images) sont moins abrupts dans les simulations et les valeurs maximales et minimales sont lissées. Naturellement, les courbes correspondant aux simulations se situent autour des courbes associées à la proportion du modèle (en rouge dans les images) qui les a générées, malgré l'approximation introduite dans le calcul de ρ .

Les proportions globales obtenues sur ces simulations sont assez proches de la proportion globale initiale, variant entre 17% et 19% pour une proportion globale initiale de 18.3%. La figure 7.18 montre la distribution des valeurs globales des proportions calculées sur les simulations. Elles se situent pratiquement entre les valeurs globales de la proportion initiale et de la proportion du modèle, avec une tendance à surestimer légèrement la proportion globale initiale, à cause des valeurs nulles de la proportion initiale qui sont devenues positives dans le modèle et les simulations.

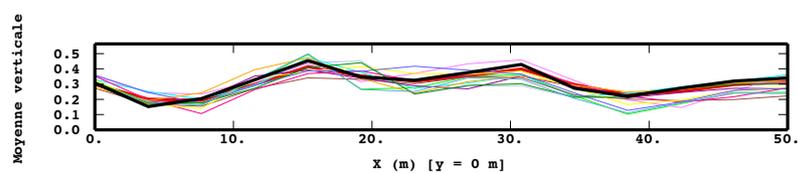
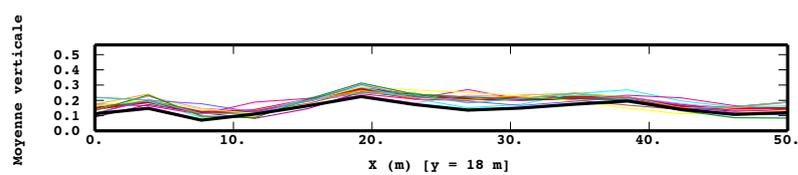
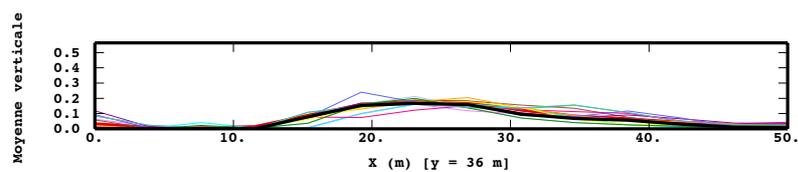
(a) $y = 0$ m (Sud)(b) $y = 18$ m (Centre)(c) $y = 36$ m (Nord)

FIG. 7.15 – Moyennes verticales de la proportion le long de différentes sections Est - Ouest ($y = cte$) dans le plan horizontal. En noir la proportion initiale ; en rouge la proportion du modèle.

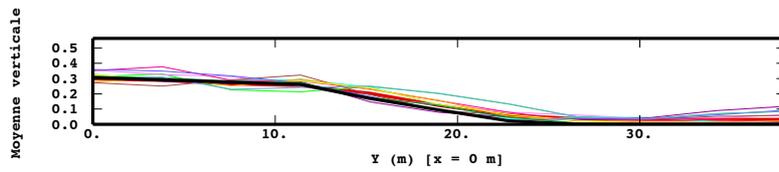
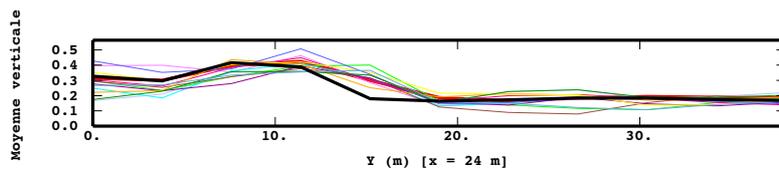
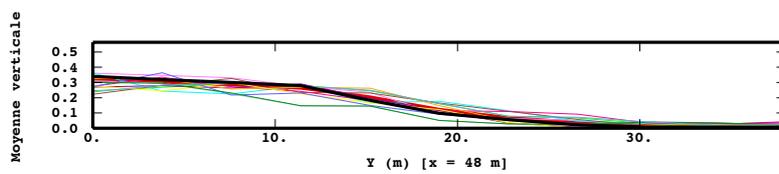
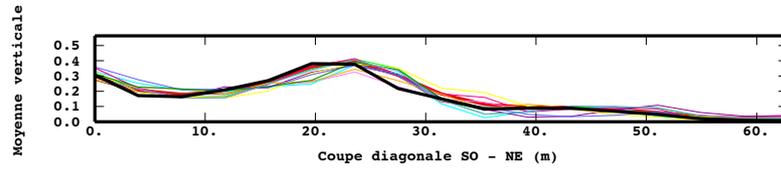
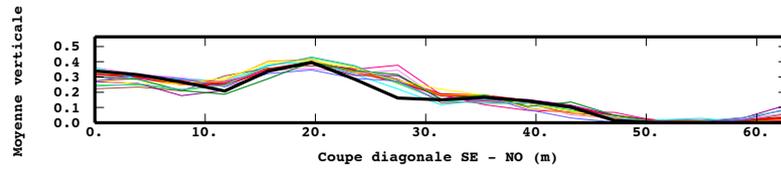
(a) $x = 0$ m (Ouest)(b) $x = 24$ m (Centre)(c) $x = 48$ m (Est)

FIG. 7.16 – Moyennes verticales de la proportion le long de différentes sections Nord - Sud ($x = cte$) dans le plan horizontal. En noir la proportion initiale ; en rouge la proportion du modèle.



(a) Section diagonale SO - NE



(b) Section diagonale SE - NO

FIG. 7.17 – Moyennes verticales de la proportion le long de différentes sections diagonales dans le plan horizontal. En noir la proportion initiale ; en rouge la proportion du modèle.

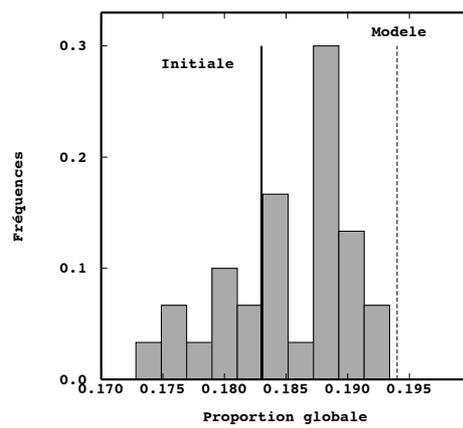


FIG. 7.18 – Histogramme de la distribution des valeurs de proportion globale obtenues à partir de 30 simulations. En noir la proportion globale initiale et en rouge la proportion globale du modèle (objet demi-ellipsoïde).

Simulations. Objets parallélépipèdes

Dans la section précédente, les objets choisis pour modéliser le faciès étudié sont des demi-ellipsoïdes. Cependant, par souci de simplification, nous avons calculé l'intensité de Poisson avec des objets de loi plus simple, des parallélépipèdes. Sur les résultats obtenus, cette approximation peut être considérée comme acceptable.

Nous allons étudier ici brièvement les résultats obtenus lorsque les simulations sont réalisées avec les objets effectivement utilisés dans le calcul de l'intensité, les parallélépipèdes. Comme précédemment, cette étude est basée sur la comparaison des proportions initiale, du modèle et des simulations. Naturellement, la proportion initiale et la proportion du modèle restent les mêmes.

Les proportions globales calculées sur trente simulations varient entre 17% et 20%, comme dans le cas précédent, et se placent aussi entre les valeurs globales de la proportion initiale et du modèle, 18.3% et 19.4% respectivement (fig. 7.19). Les valeurs globales correspondant aux simulations avec les parallélépipèdes varient dans une fourchette légèrement plus large que dans le cas des demi-ellipsoïdes (les objets sont plus grands) et tendent à se rapprocher davantage de la valeur du modèle, comme espéré. Le fait que les proportions globales soient, en général, plus faibles que celle du modèle est dû au grand nombre de valeurs de proportion très faibles ou nulles qui sont sur-représentées dans les simulations. Tandis que le modèle tend à lisser les proportions initiales, les simulations accentuent la présence des valeurs extrêmes.

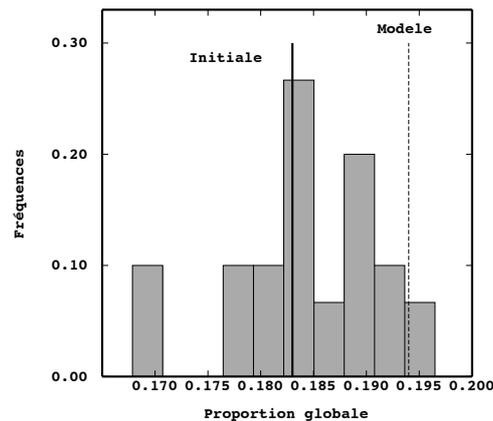


FIG. 7.19 – Histogramme de la distribution des valeurs de proportion globale obtenues à partir de 30 simulations. En noir la proportion globale initiale et en rouge la proportion globale du modèle (objet parallélépipède).

Sur les simulations on retrouve les mêmes caractéristiques que pour les simulations avec les demi-ellipsoïdes : une forte présence des valeurs faibles et fortes dans les simulations, tandis que les valeurs intermédiaires sont moins bien récupérées. Sur les courbes des valeurs verticales moyennes des proportions, nous avons aussi observé plus de ressemblance entre les proportions des simulations et la proportion du modèle qu'avec la proportion initiale, notamment pour les valeurs maximales et minimales et dans les variations brusques. Cependant, nous pouvons conclure que, dans ses caractéristiques générales, les proportions initiales sont assez bien retrouvées.

Si on compare les courbes moyennes verticales (fig. 7.20 à 7.22) de la proportion initiale (en noir dans

les figures), de la proportion du modèle (en rouge), de la proportion moyenne sur les 30 simulations réalisées avec l'objet demi-ellipsoïde (en bleu) et avec l'objet parallélépipède (en vert), on observe que les simulations avec les deux familles d'objets sont, en moyenne, très similaires et respectent la proportion du modèle. La proportion initiale est aussi bien récupérée sauf dans les changements très brusques où le modèle et les simulations ont un comportement plus graduel, comme on s'y attendait.

Ce petit test a mis en évidence le fait que nous ne pouvons pas distinguer l'effet de l'approximation faite lors du calcul de l'intensité, en utilisant des objets plus simples que dans la simulation, de l'effet du lissage après déconvolution et de la variabilité intrinsèque du processus de simulation. Naturellement, il est nécessaire que cette approximation par des objets simples soit acceptable : les objets utilisés dans la déconvolution et dans les simulation doivent être choisis de façon cohérente, la proportion initiale et la proportion du modèle doivent être semblables et les valeurs de la proportion globale doivent être aussi similaires. Dans ce cas, par exemple, la différence moyenne entre les deux est seulement de 1%.

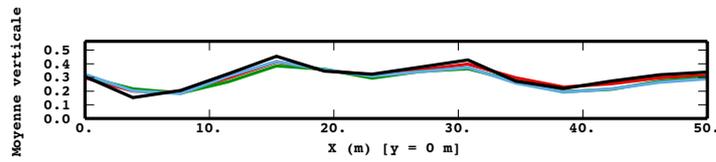
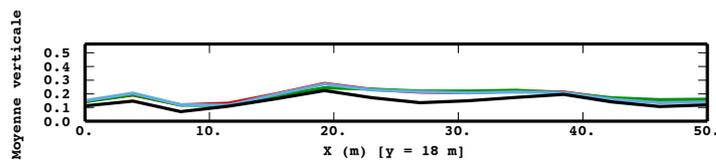
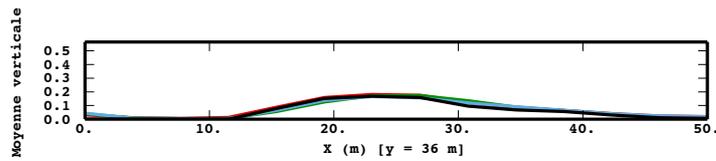
(a) $y = 0$ m (Sud)(b) $y = 18$ m (Centre)(c) $y = 36$ m (Nord)

FIG. 7.20 – Moyennes verticales de la proportion le long de différentes sections Est - Ouest ($y = cte$) dans le plan horizontal. En noir la proportion initiale, en rouge la proportion du modèle, en bleu la proportion moyenne sur 30 simulations avec l'objet demi-ellipsoïde, en vert la proportion moyenne sur 30 simulations avec l'objet parallélépipède.

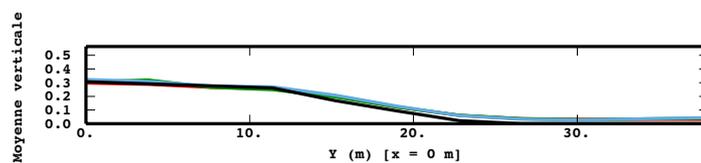
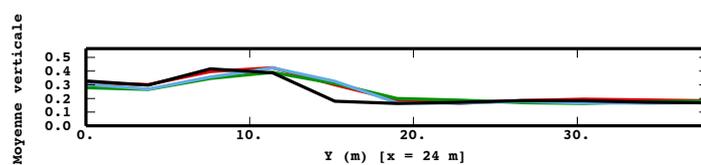
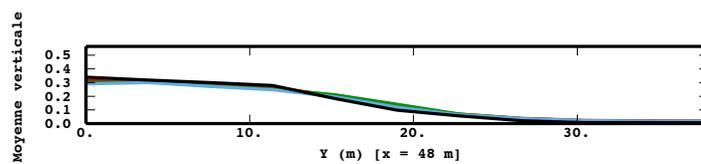
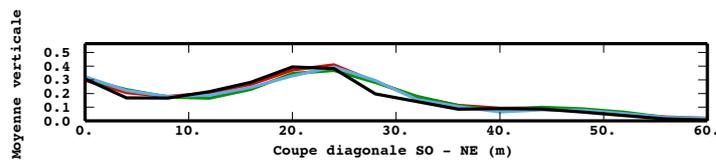
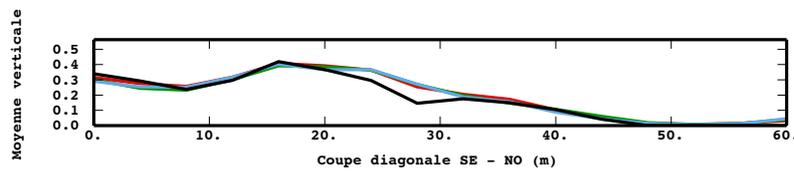
(a) $x = 0$ m (Ouest)(b) $x = 24$ m (Centre)(c) $x = 48$ m (Est)

FIG. 7.21 – Moyennes verticales de la proportion le long de différentes sections Nord - Sud ($x = cte$) dans le plan horizontal. En noir la proportion initiale, en rouge la proportion du modèle, en bleu la proportion moyenne sur 30 simulations avec l'objet demi-ellipsoïde, en vert la proportion moyenne sur 30 simulations avec l'objet parallélépipède.



(a) Section diagonale SO - NE



(b) Section diagonale SE - NO

FIG. 7.22 – Moyennes verticales de la proportion le long de deux sections diagonales dans le plan horizontal. En noir la proportion initiale, en rouge la proportion du modèle, en bleu la proportion moyenne sur 30 simulations avec l'objet demi-ellipsoïde, en vert la proportion moyenne sur 30 simulations avec l'objet parallélépipède.

7.3 Analyse des composantes connexes

L'importance de la distribution d'un faciès dans un volume est mise en évidence lors de l'étude de l'écoulement d'un fluide à l'intérieur d'un faciès poreux, par exemple. Si ce faciès poreux est modélisé par un ensemble booléen, les propriétés de l'écoulement seront dépendantes de la façon dont les objets sont distribués dans le volume de simulation. Pour illustrer l'importance de la distribution des objets, nous étudions les composantes connexes dans un cas simple.

Une composante connexe est l'union de tous les pixels dans la simulation qui appartiennent à l'ensemble booléen et qui sont connectés à travers une ligne complètement contenue dans l'ensemble (fig. 7.23) et donc dans le faciès modélisé. Chaque composante connexe est ensuite numérotée par ordre croissant, de la plus grande à la plus petite.

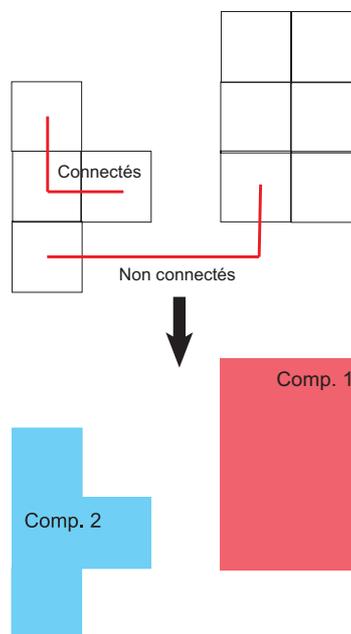


FIG. 7.23 – Composantes connexes : seuls les pixels connectés à travers un chemin contenu dans l'ensemble booléen appartiennent à la même composante connexe.

Prenons comme exemple une simulation dont la distribution du faciès est non stationnaire (cas présenté précédemment) et une simulation dont la distribution du faciès est stationnaire mais avec la même proportion globale (18.3%), ou la même porosité dans le cas d'un faciès poreux. Nous analysons les cinq premières composantes connexes pour chaque simulation (fig. 7.24), considérées comme les plus représentatives.

Dans le cas non stationnaire (fig. 7.24(a)), nous avons pu identifier 101 composantes connexes. La première composante (en rouge) représente les 75% du volume total occupé par le faciès. Les cinq premières composantes constituent déjà 85% du volume du faciès. Par contre, dans le cas stationnaire (fig. 7.24(b)), nous avons identifié 324 composantes différentes, la première composante n'étant que 4% du volume total du faciès. Dans ce cas, les composantes diminuent en taille très lentement, ainsi les cinq premières composantes n'occupent que 16% du volume du faciès.

Cet exemple illustre clairement l'importance de la connaissance de la distribution du faciès, et donc des objets dans les simulations, pour reproduire le comportement des fluides dans le milieu. Dans l'exemple

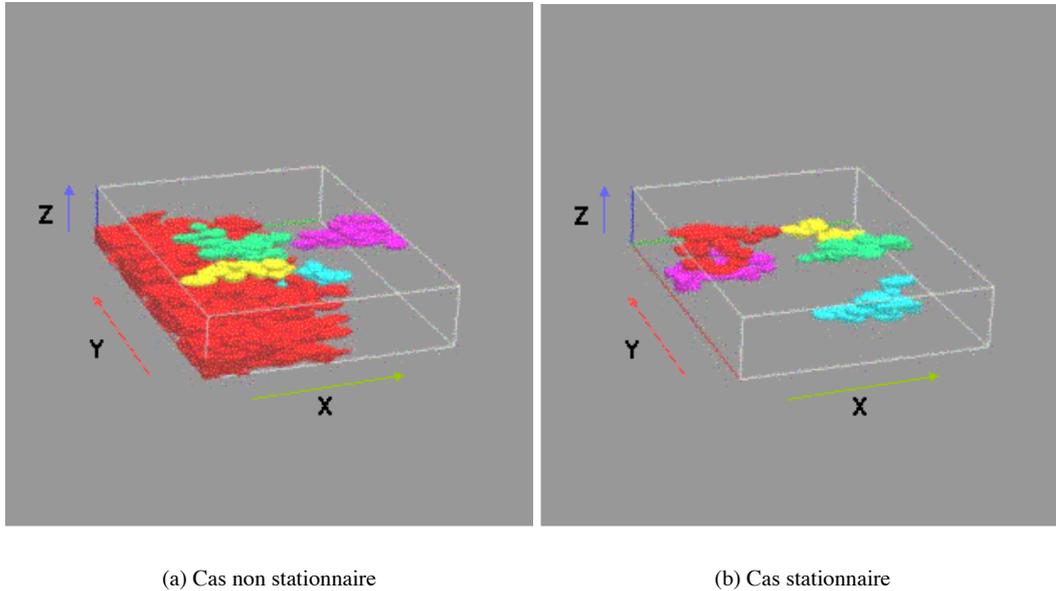


FIG. 7.24 – Composantes connexes d'une simulation non stationnaire et d'une simulation stationnaire avec la même proportion globale du faciès simulé (18%). Seules les cinq premières composantes sont représentées, en rouge la plus grande.

stationnaire, ce fluide sera confiné à l'intérieur des lentilles non connectées les unes aux autres alors que dans l'exemple non stationnaire, il pourra traverser toute l'unité.

7.4 Résultats et conclusions

L'exemple présenté dans ce chapitre nous a servi d'une part à tester la méthode et déterminer la qualité des résultats et d'autre part à mettre en évidence l'importance de la prise en compte de la distribution des objets dans la simulation lors de l'interprétation des résultats en termes de porosité, par exemple.

Globalement on peut conclure que la méthode proposée donne d'assez bons résultats selon les deux critères que nous nous sommes fixés : la ressemblance entre les données, i.e. la proportion initiale, et la proportion calculée sur les simulations, et la ressemblance entre les proportions globales initiale et des simulations.

Pour valider le processus de déconvolution, trois étapes sont nécessaires :

1. La validation du modèle.
2. La validation des simulations par rapport au modèle.
3. La validation des simulations par rapport aux données initiales.

La **validation du modèle** est réalisée en comparant les proportions initiales aux proportions associées à l'intensité calculée et à l'objet choisi. Les approximations réalisées lors du calcul de l'intensité introduisent

une différence entre les données initiales et le modèle obtenu. Ainsi, nous exigeons que notre modèle présente le même comportement que la proportion initiale (dans cet exemple, la différence ponctuelle moyenne en valeur absolue est de 5%) et que les valeurs globales soient proches, dans ce cas la différence entre les proportions globales moyennes du modèle et des données est de 1%. Nous avons considéré que les proportions sont modélisées de façon satisfaisante par le modèle obtenu, mais c'est à l'utilisateur de déterminer si le modèle est acceptable ou non.

Il existe néanmoins deux différences inévitables et significatives entre la proportion initiale et le modèle :

- Les maxima et les minima ainsi que les variations brusques dans les proportions initiales sont lissés dans le modèle. Le lissage réalisé pour corriger l'intensité mathématique impose cet effet par construction. Ainsi, nous ne pourrions pas modéliser exactement un saut brusque dans les proportions initiales, mais il sera modélisé par une variation graduelle dans la proportion sur une distance qui est liée à la taille de l'objet. Cela semble assez réaliste par rapport à la nature, sauf en présence de structures telles que des failles qui peuvent marquer ces changements brusques. Un traitement séparé des différentes zones peut alors s'imposer.

- Les valeurs de proportion strictement nulles sont seulement récupérées en partie dans le modèle. Cet effet est lié directement au point précédent : l'effet de bord associé à l'objet ne nous permet pas de retrouver les proportions nulles strictement dans la région où elles ont été identifiées. Il faut ici se poser de nouveau la question sur la représentativité des proportions. Celles-ci étant une valeur d'estimation sur un support relativement grand, elles ne représentent qu'une tendance dans la présence des faciès. Une valeur nulle de proportion peut indiquer que la présence du faciès est très faible, mais pas forcément inexistante. Néanmoins, si les valeurs nulles sont strictement vraies, c'est-à-dire si elles sont déterminées et confirmées par d'autres sources, alors elles deviennent des valeurs déterministes, et leur traitement doit se réaliser autrement. On pourrait, par exemple, traiter cette région séparément, ou se servir de l'effet de bord dû à l'objet pour modifier les proportions initiales avant le processus de déconvolution en créant une région autour des valeurs strictement nulles. De cette façon, on pourrait forcer le modèle à diminuer ses proportions progressivement jusqu'à atteindre les valeurs nulles dans la zone voulue. Il faut tout de même signaler que ce problème se situe à la limite d'application de cette approche probabiliste.

La comparaison des proportions calculées sur les simulations obtenues nous ont permis de **valider les simulations par rapport au modèle**, notamment par rapport à l'approximation introduite en utilisant des objets plus simples dans le modèle que dans les simulations. Si le choix des objets est fait de façon cohérente, c'est-à-dire si les deux familles d'objets ne sont pas trop différentes, il n'est pas possible de différencier cet effet des autres approximations introduites dans le processus. Ainsi, l'approximation réalisée est acceptable et les simulations obtenues sont cohérentes avec le modèle utilisé. Ce résultat est intéressant pour l'utilisation d'objets plus réalistes, dont l'expression géométrique n'est pas élémentaire.

Finalement, la comparaison entre les proportions calculées sur les simulations et les proportions initiales ont servi à **valider les simulations par rapport aux données de proportion**, ce qui est en somme le but du processus. Ainsi, nous avons observé que les simulations, individuellement, reproduisent le comportement indiqué par les proportions initiales. Les valeurs extrêmes de proportion sont sur-représentées dans les simulations par rapport à la proportion initiale, tandis qu'il y a une fréquence plus faible des valeurs intermédiaires. Ceci s'explique par le fait que, dans l'exemple montré, un objet a une taille comparable à la maille des proportions, ce qui fait que la proportion résultante soit très sensible à la présence ou l'absence d'un objet. Sur le nombre de simulations réalisées on observe que, en moyenne, elles respectent la proportion initiale dans la même mesure que le modèle. La proportion globale est aussi respectée dans toutes les

simulations à une différence d'environ 1%.

Cette méthode peut être appliquée par exemple à la simulation de crevasses dans un réservoir, ou, à une échelle différente, de lentilles argileuses à l'intérieur d'un faciès poreux. En général, il est possible de reproduire la distribution non stationnaire d'un faciès qui peut être modélisé par des objets relativement petits par rapport à la taille du champ de travail, et ceci à différentes échelles de travail. Mais, que se passe-t-il quand les objets sont «infinis» par rapport à la taille du champ ? C'est le cas de la simulation des chenaux à l'intérieur d'un réservoir où les objets sont infinis selon une direction, la direction d'écoulement, dans le champ de travail. Dans ce cas, comment définir la probabilité associée aux objets ? et comment les simuler ? De plus, un même objet contribue à la proportion dans deux mailles de la grille qui peuvent se trouver très loin l'une de l'autre, alors comment tenir compte de cette corrélation dans le calcul de l'intensité ? Le chapitre suivant essaie de répondre à ces questions dans un cas d'étude particulier où les chenaux sont sub-parallèles et traversent tout le champ de travail.

Chapitre 8

Application à un cas réel : Permien de l'Utah

Dans ce second exemple, la simulation d'objets concerne les dépôts fluviaux. La géométrie de ces dépôts est bien identifiable. En effet ceux-ci sont principalement composés de chenaux, dont la figure 8.1 montre des exemples.



FIG. 8.1 – Quelques images de chenaux. Image extraite de : <http://www.geosc.psu.edu/~sedweb>

Il s'agit d'un cas réel de modélisation géostatistique d'un réservoir potentiel. Ce travail suit les travaux réalisés au cours du stage «Modélisation géostatistique de la série continentale du Permien d'Utah» (Clément, 1999), qui s'est déroulé à l'IFP (décembre 1998 - février 1999) sous la responsabilité de O. Lerat et B. Doligez. L'objectif de ces travaux était de modéliser, selon des approches géostatistiques, la distribution des hétérogénéités dans des réservoirs potentiels éoliens-fluviaux à partir de données d'affleurement. Les

unités considérées sont les séries continentales permienes du bassin du Paradox (Utah).

Dans ces études, les simulations géostatistiques réalisées étaient basées sur deux méthodes différentes selon le phénomène géologique à mettre en évidence : la méthode des gaussiennes seuillées et le modèle booleen. Ce dernier avait été utilisé pour modéliser le lithofaciès des dépôts fluviaux présentant une géométrie de chenal. Cependant, les outils disponibles lors de l'étude ne permettaient pas de simuler la distribution des chenaux par zones. Il avait été possible de reproduire une non stationnarité dans la distribution verticale des objets, mais non une variation latérale. La solution proposée avait consisté à diviser l'unité en sous-unités et à les traiter séparément. Le nombre d'objets à simuler a dû être adapté à la suite de tests successifs.

Dans ce travail nous allons répondre à ce problème. La méthode proposée pour estimer l'intensité de Poisson permet de reproduire dans les simulations une non stationnarité dans toute direction signalée par la proportion du faciès. Le nombre moyen d'objets à simuler est donné directement à partir de cette intensité et il est cohérent avec les caractéristiques géométriques des objets et leur distribution.

8.1 Cadre du travail

La description paléogéographique des séries continentales du Permien de l'Utah présentée ici (sections 8.1.1 et 8.1.2) a été extraite de Eschard et al. (1999) et Lerat et al. (1999). Ces études avaient pour objectifs la caractérisation des faciès sédimentaires en environnements éoliens et fluviaux et la détermination d'une séquence type qui puisse servir de guide à des corrélations en subsurface dans des environnements comparables. Le premier de ces travaux visait en outre à identifier les facteurs climatiques et tectoniques qui contrôlent la conservation des sédiments éoliens et fluviaux pour mieux prédire la localisation des réservoirs potentiels. Il a été ainsi proposé un modèle stratigraphique de haute-résolution qui est adopté dans notre étude. Le second de ces travaux a donné une proposition de modèle géostatistique du réservoir.

8.1.1 Contexte géologique

Les séries continentales du Permien de l'Utah se sont déposées dans le bassin intracratonique du Paradox sous une tectonique en extension peu active. La sédimentation est à dominance continentale, fluviale et éolienne, avec néanmoins des dépôts carbonatés à la base dus à des épisodes transgressifs pendant lesquels la mer inondait la région.

Les figures 8.2 et 8.3 montrent respectivement la carte de l'emplacement et le schéma paléogéographique de la région étudiée. Les caractéristiques générales sont les suivantes :

- Au Nord Est du bassin, l'érosion du horst de l'Uncompaghre a mis en place de grands systèmes de cônes alluviaux qui alimentent des systèmes fluviaux riches en sables. Les directions de transport dominantes sont vers l'Ouest et le Nord Ouest.
- Au Sud un grand erg éolien (Formation de Cedar Mesa) s'étend et passe latéralement vers le Nord aux séries mixtes de dépôts fluviaux et éoliens (Formation de Cutler). Les vents dominants dans cet erg soufflaient du Nord vers le Sud.
- Au Sud Ouest (en Arizona et Nouveau Mexique) une mer peu profonde a induit des épisodes transgressifs qui ont déposé des dépôts marins carbonatés, riches en faune marine.
- En même temps que la sédimentation, un fluage des séries salifères plus profondes du Pennsylvanien a induit la formation de larges anticlinaux.
- La base de la série Permienne montre une transition avec les séries sous-jacentes du Pennsylvanien, où les influences marines sont mieux enregistrées.
- Le sommet de la série est marquée par une grande discordance régionale triasique, due probablement au fluage du sel Pennsylvanien.

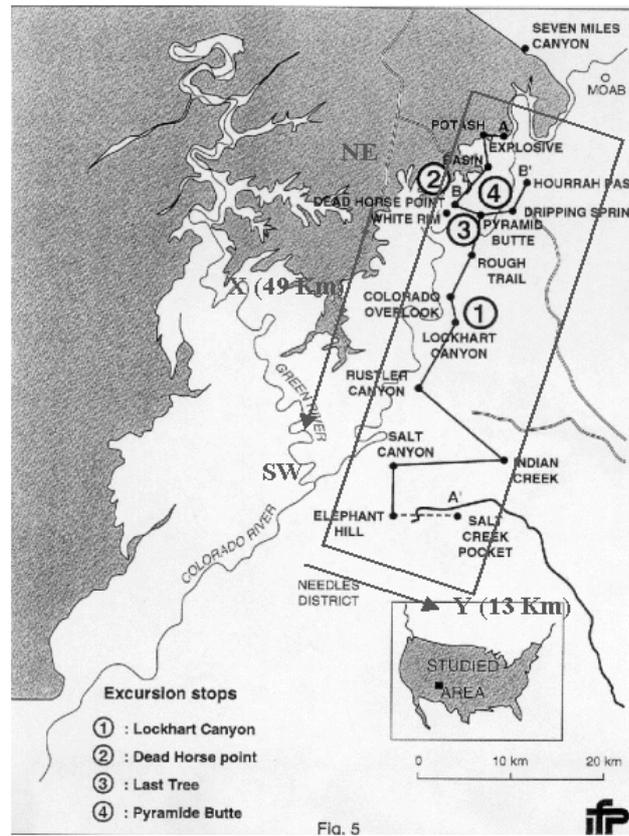


FIG. 8.2 – Définition de la région étudiée, localisation des principales coupes et du transect régional (d'après Eschard et al. (1999)).

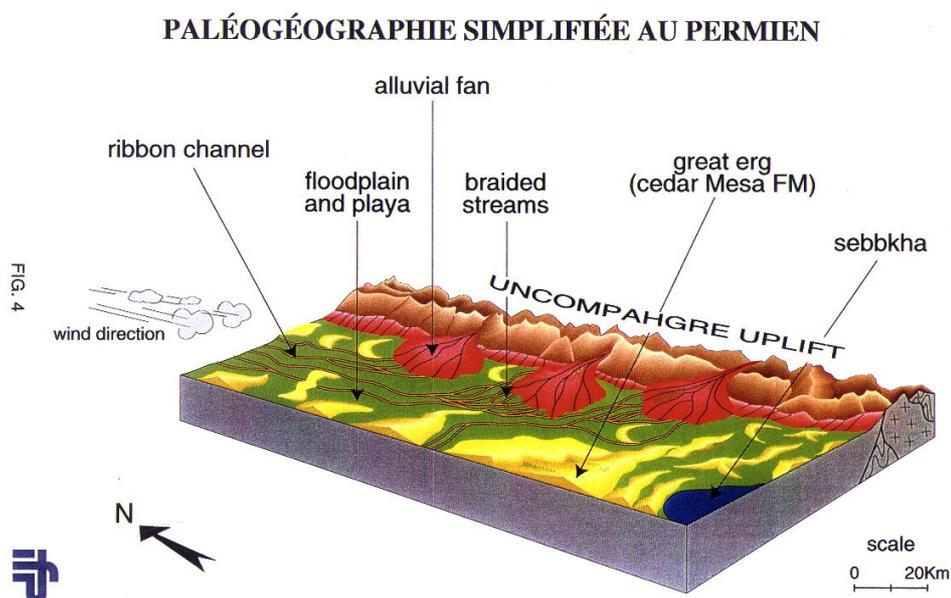


FIG. 8.3 – Schéma paléogéographique de la région étudiée (extrait de Eschard et al. (1999)).

8.1.2 Transect régional

L'étude de terrain a été effectuée sur le flanc Est du canyon du Colorado. Un transect régional NE-SO de 50 km de long environ a été établi à partir de 15 coupes sédimentologiques et de photos-composées (fig. 8.4)¹. Environ 280 m de sédiments permien sont représentés à partir d'un horizon repère marin à la base jusqu'à la discordance Triasique au sommet, qui a été utilisée comme marqueur pour l'horizontalisation de la section. Ces séries sont composées essentiellement de dépôts éoliens alternant avec des séries fluviales et lacustres. Les caractéristiques principales de ces séries sont :

- Une douzaine de cycles éoliens-fluviaux d'environ 20 à 40 m d'épaisseur ont été identifiés et corrélés à l'échelle régionale. Un cycle complet est composé des phases suivantes :
 1. Accumulation de matériel éolien, sous forme de grands ergs avec des dunes bien développées, en période sèche et bas niveau des nappes phréatiques.
 2. Préservation partielle des dunes éoliennes en période humide lors d'une montée des nappes.
 3. Réactivation d'un système fluvial en période humide et montée des nappes.
 4. Incision des dunes par des chenaux fluviaux remplis de matériel grossier.
 5. Aggradation des systèmes alluviaux avec le développement des faciès argileux de plaine d'inondation et de petits chenaux méandriformes.
 6. Phase de pédogenèse, dans un climat sec les systèmes fluviaux deviennent inactifs en raison de l'assèchement des rivières. Fin du cycle.

¹Dans cette étude l'axe positif des X a été pris du NE vers le SO. Les représentations graphiques des figures 8.4 et 8.5 sont inversées par rapport à ce critère.

- À l'intérieur des séquences il existe un passage latéral de faciès Nord - Sud : les dépôts à dominance fluviatile passent vers le Sud à des dépôts lacustres et à l'extrême Sud à de dépôts principalement éoliens.
- On note aussi la présence de grands anticlinaux salifères contemporains de la sédimentations du Permien, induits par le fluage du sel Pennsylvanien sous-jacent. Ces anticlinaux causent des changements d'épaisseur très marqués à l'intérieur d'une même unité génétique.

8.1.3 Description des données

Les informations dont on dispose pour la modélisation géostatistique sont celles décrites par Clément (1999) et Lerat et al. (1999) :

Unités La série totale a été divisée en 9 litho-unités de 20 à 50 m d'épaisseur (fig. 8.5 ; voir note 1). Chacune d'entre elles correspond en partie aux différents cycles génétiques, de façon à obtenir un modèle fin de la distribution des hétérogénéités.

Marqueurs Les marqueurs indiquent un niveau de référence dans un profil géologique. Dans toute la série, 19 marqueurs ont été identifiés et numérotés en ordre décroissant de la base vers la surface. Seuls 9 de ces marqueurs correspondent aux surfaces des limites d'unités. Dans notre étude, nous ne nous intéressons qu'à l'unité limitée par les surfaces 10 et 11 détachée en bleu dans la figure (fig. 8.5). Cette unité correspond à l'union des unités 4 et 5.

Lithofaciès et lithotypes Neuf lithofaciès correspondant à des associations de faciès ont été distingués. Les lithotypes sont définis par des associations de lithofaciès qui possèdent une signification sédimentaire et pétrophysique. Nous ne nous sommes intéressés qu'au faciès «chenal», modélisé par les objets du modèle booléen, les autres faciès étant regroupés dans un seul lithotype complémentaire.

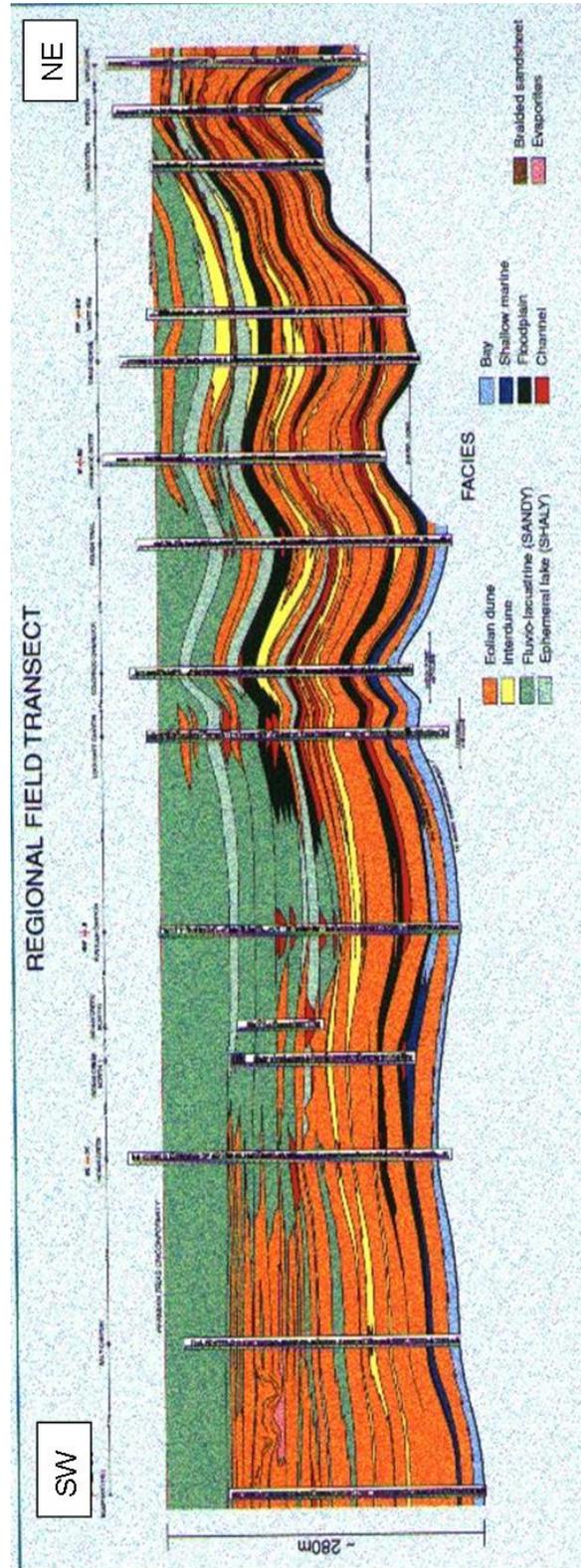


FIG. 8.4 – Transect régionale de la zone (extrait de Lerat et al. (1999))

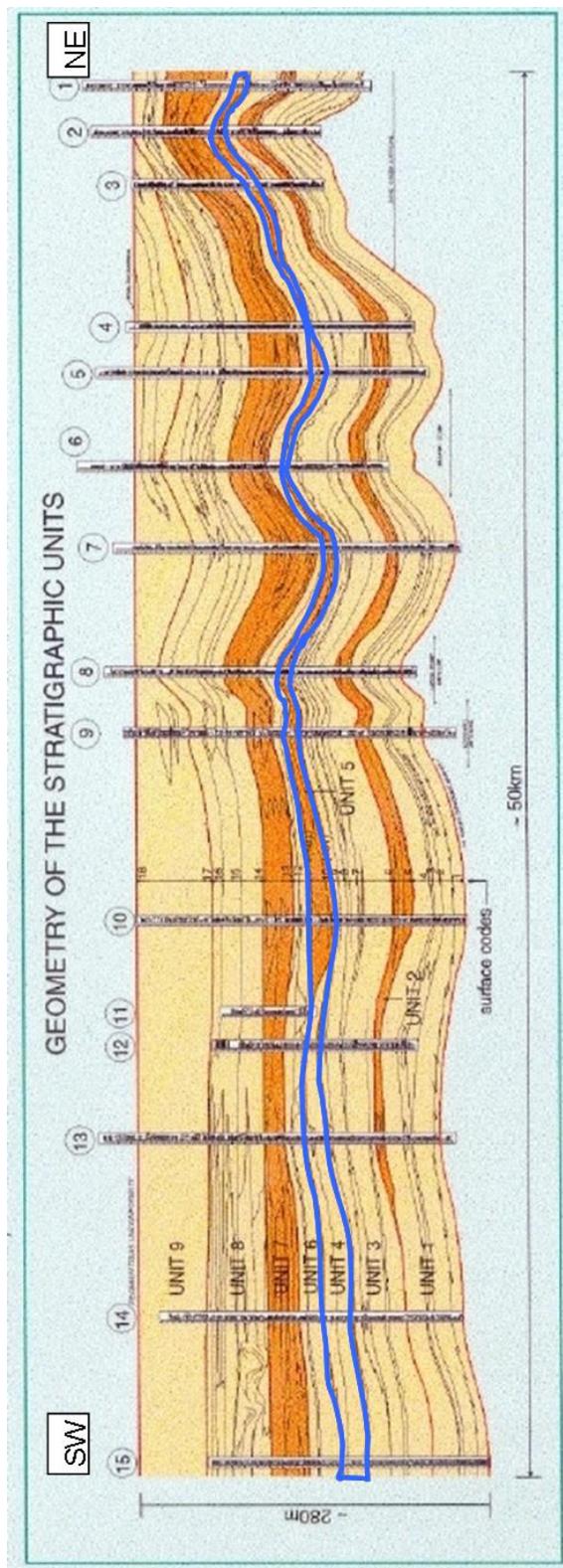


FIG. 8.5 – Découpage de la série en 9 unités lithostratigraphiques. En bleu l'unité étudiée, réunion des unités 4 et 5 (extrait de Eschard et al. (1999)).

8.1.4 Définition de l'unité de travail

L'unité choisie pour appliquer notre méthode d'inférence est donc celle qui est limitée au toit par la surface 11 et à la base par la surface 10. Elle constitue un exemple très net du type de non stationnarité que nous voulons reproduire dans les simulations. La non stationnarité latérale du faciès chenal est très marquée : très abondant dans le NE, il diminue progressivement vers le SO jusqu'à pratiquement disparaître. Ceci s'explique par le contexte paléogéographique décrit précédemment : l'érosion du horst au NE a généré des systèmes fluviaux très riches en sables ; au SO, par contre, l'erg éolien de la formation de Cedar Mesa a favorisé la présence des dépôts éoliens.

La paléogéographie et la dimension de la région étudiée nous permettent de faire une hypothèse importante sur l'architecture de l'unité. Cette zone s'étend sur une dizaine de kilomètres selon la direction estimée d'écoulement des chenaux (N-NO). On peut supposer que les chenaux restent sub-parallèles sur cette distance et qu'ils traversent pratiquement toute la région d'étude. De plus, toutes les coupes qui ont servi à construire le transect régional se trouvent, à l'échelle régionale, sur un même plan vertical (sauf la coupe «Indian Creek») de direction NE-SO à peu près perpendiculaire à la direction d'écoulement des chenaux. Ainsi, nous pouvons considérer que toute l'information apportée par les différentes coupes sont représentatives d'une même section verticale dans la région d'étude.

On peut donc faire l'hypothèse de non stationnarité dans la distribution des faciès selon la direction d'écoulement (direction Y dans la figure 8.5). Cette hypothèse nous permet de ramener le calcul de la proportion du faciès chenal à un plan vertical. Dans ce contexte, on admet que les proportions calculées sur un plan vertical XZ quelconque, perpendiculaire à la direction d'écoulement, sont représentatives de toute la région, i.e., que les proportions calculées sur ces différents plans sont statistiquement similaires.

Jusqu'à présent, notre hypothèse de travail a été d'admettre que les données disponibles sont la proportion du faciès à modéliser et les caractéristiques des objets qui vont le simuler. C'est la démarche suivie dans le cas d'étude du chapitre précédent. Ici, nous ne disposons que des informations lithologiques des coupes et il nous faut calculer la proportion correspondante, guidés par les travaux préalables et l'information géologique. Nous avons ainsi l'opportunité d'étudier l'influence du processus d'estimation des proportions sur notre méthode de calcul de l'intensité de Poisson puis sur les simulations.

La première étape de la simulation d'une unité lithostratigraphique consiste à définir les surfaces qui la délimitent et à choisir un repère de travail tel que les statistiques effectuées dans ce repère soient représentatives des sédiments lors du dépôt.

8.1.4.1 Calcul des surfaces

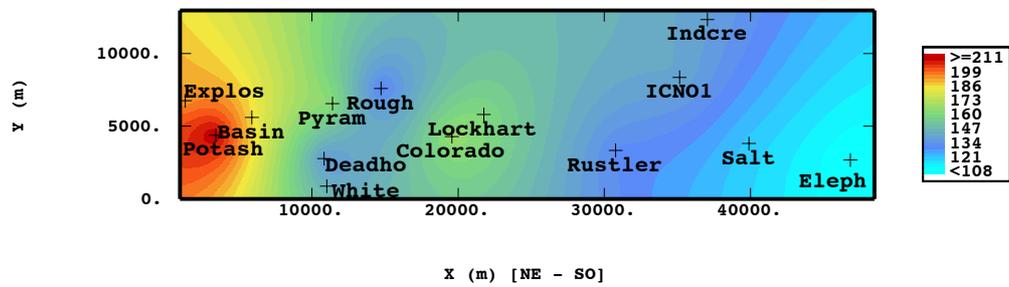
À partir des marqueurs identifiés dans les coupes sédimentologiques il est possible d'estimer les surfaces 10 et 11 qui déterminent respectivement la base et le toit de l'unité. Le nombre très limité d'informations disponibles horizontalement (15), qui sont de plus approximativement alignées, ne permet pas une estimation très rigoureuse d'un modèle de variogramme. Nous avons donc choisi de réaliser une interpolation avec un modèle linéaire (fig. 8.6). Les paramètres de la grille de calcul se trouvent au tableau 8.1. Une représentation 3D est à la figure 8.7. Cette estimation met bien en évidence la grande variabilité des données : par exemple, le marqueur correspondant à la surface 10 dans les coupes Basin et Potash, éloignées de 2500 m selon l'axe X , est localisé sur chacune d'elles avec une différence de profondeur de 26 m.

Selon les estimations, l'épaisseur de l'unité, de 11.85 m en moyenne, varie fortement entre 3.52 m et 20.87 m. Les valeurs calculées à partir des coupes sont similaires : une valeur moyenne de 11.83 m, les épaisseurs maximum et minimum étant respectivement de 19.80 m et 3.50 m. Il faut remarquer un amincissement important au NE, région de concentration du lithotype chenal, puis un épaissement progressif

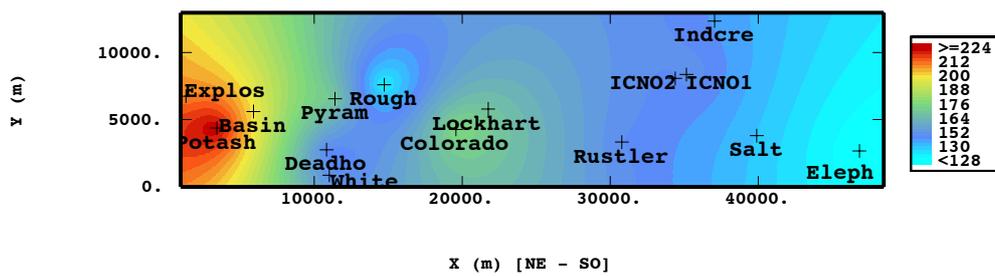
vers les zones de dépôts éoliens (SO).

	X	Y
Origine	1000 m	0 m
Maille	50 m	50 m
Nombre de mailles	950	260

TAB. 8.1 – Grille des surfaces 10 et 11.



(a) Surface 10 : base de l'unité



(b) Surface 11 : toit de l'unité

FIG. 8.6 – Projection horizontale des surfaces de la base et du toit de l'unité. En noir les emplacements des coupes sédimentologiques. La coupe nommée ICNO2 n'intersecte pas la surface 10.

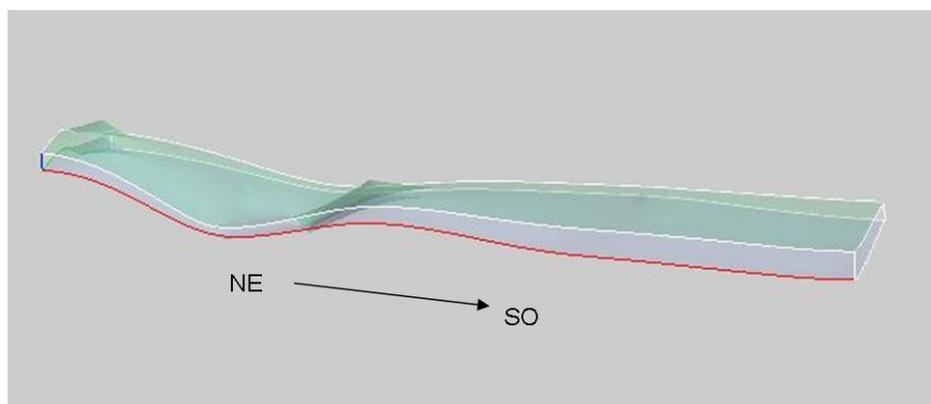


FIG. 8.7 – Représentation, dans le repère structural, des surfaces limitant l'unité.

8.1.4.2 Discrétisation des coupes sédimentologiques

Une fois les surfaces définies, il est nécessaire de discrétiser les coupes en lithofaciès dans le repère de travail. Dans ce repère il faut générer une grille 3D complète, c'est-à-dire que chacune de ses mailles doit être informée par une valeur de proportion.

En général, les puits, ou en l'occurrence les coupes sédimentologiques, sont discrétisés avec un pas régulier choisi en fonction de la variabilité lithologique et de la résolution de la diagraphie. Dans cette étude le pas de discrétisation a été fixé à 0.5 m, ce qui correspond à l'épaisseur des mailles des grilles de simulation et de proportion.

Les repères de travail sont actuellement de deux types (fig 8.8) :

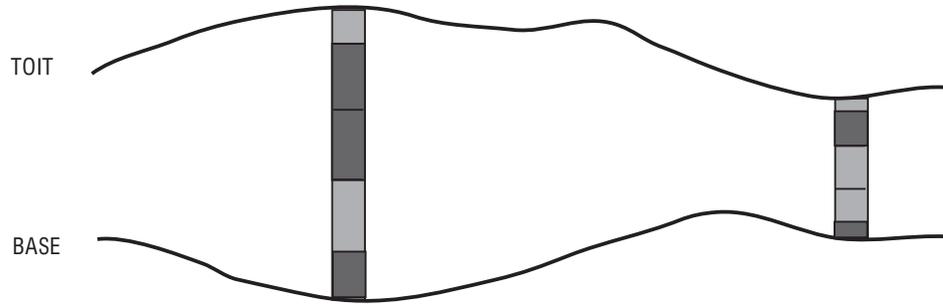
- **maillage parallèle**, où toutes les couches de la grille sont de même épaisseur et parallèles à une surface de dépôt de référence ; l'unité ainsi discrétisée est le plus souvent incomplète dans la grille de travail puisque la taille verticale de la grille correspond à l'épaisseur maximale de l'unité estimée (fig. 8.8(b)) ;
- **maillage proportionnel**, où l'unité est divisée en un nombre constant de couches entre le toit et la base de l'unité, ce nombre étant choisi de façon à avoir une discrétisation représentative de la variabilité (fig. 8.8(c)) ; le volume des mailles de la grille est variable.

Le repère de travail est choisi selon des considérations géologiques : un effet d'érosion peut être mieux reproduit par un maillage parallèle suivi d'une érosion, par exemple. Pour représenter une unité déformée par des processus de compaction, le maillage proportionnel peut être plus approprié.

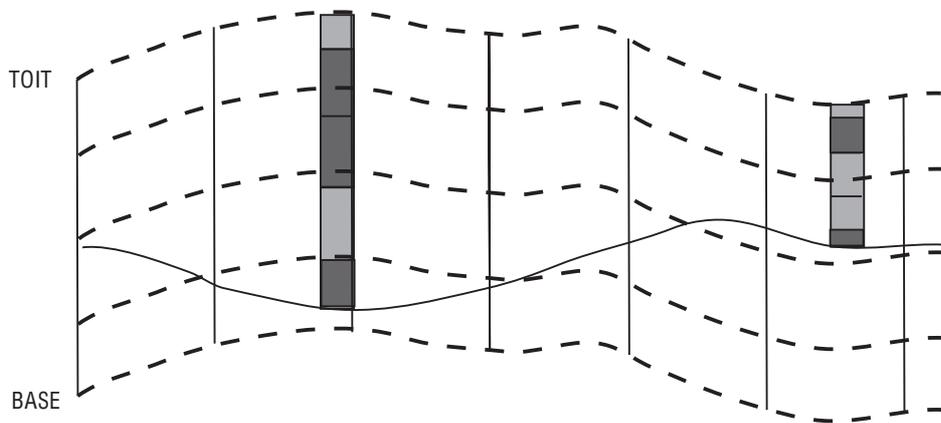
Pour l'unité étudiée, nous avons utilisé un maillage parallèle à la surface 11, le toit de l'unité. Les couches de la grille de simulation sont donc parallèles et ont 0.5 m d'épaisseur. Ce choix a été guidé par le rapport IFP initial (Clément, 1999) dans lequel le maillage pour l'unité limitée par les surfaces 101-11 (au NE, où les chenaux sont présents) est parallèle au sommet.

Dans ce repère de travail, la grille a une épaisseur d'une vingtaine de mètres. Les coupes discrétisées n'arrivent pas toutes à la base de l'unité et n'apportent donc pas d'information sur toute la grille de travail (fig. 8.9). Notons deux comportements très marqués :

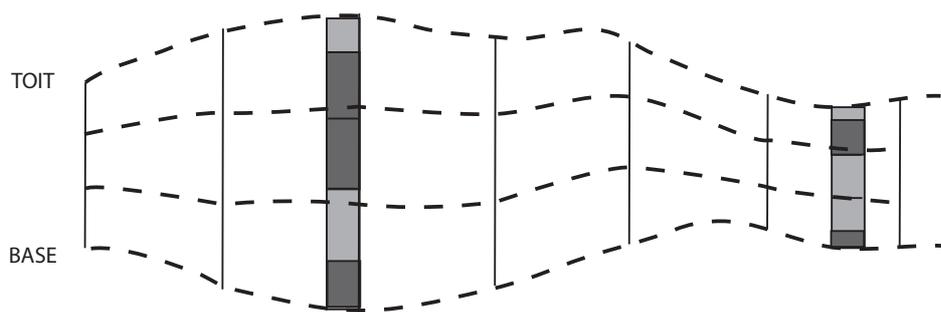
- au Nord Est : une présence importante du lithotype chenal (en rouge) associée à des coupes plus courtes dues à la faible épaisseur de l'unité dans cette région ;
- au Sud Ouest : une prédominance du lithotype complémentaire (en vert) associée à des coupes plus longues, à l'exception des coupes Indcre, ICN1 qui sont plus courtes en raison d'un nouveau rétrécis-



(a) Unité dans son repère structural



(b) Maillage parallèle au toit de l'unité



(c) Maillage proportionnel de l'unité

FIG. 8.8 – Comparaison des deux types de maillages possibles pour la discrétisation d'une unité.

sement de l'unité, et de la coupe ICN2 qui n'arrive pas à la base de l'unité sédimentaire.

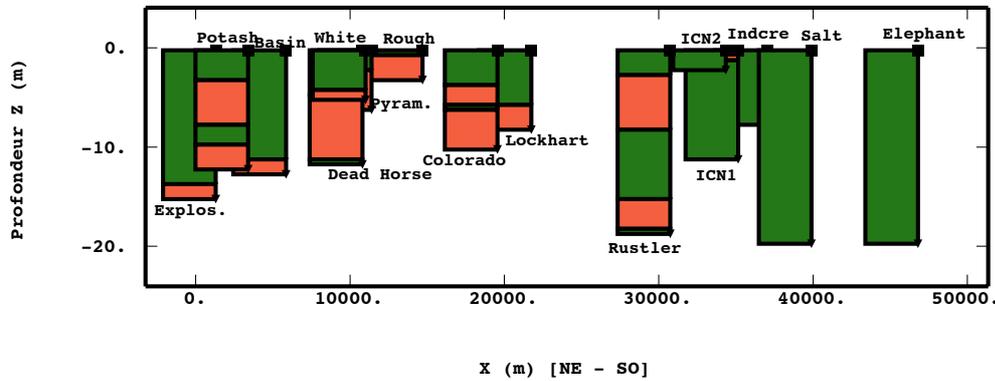


FIG. 8.9 – Discretisation des puits avec l'identification du faciès chenal (en rouge) et son complémentaire.

8.1.5 Calcul de la grille de proportion

Nous avons posé comme hypothèse que la répartition du faciès chenal dans l'unité est stationnaire selon l'axe Y , la direction d'écoulement des chenaux. Le plan perpendiculaire à cette direction, XZ , est appelé ici le plan de non stationnarité, puisque c'est sur ce plan que l'éventuelle non stationnarité dans la distribution du faciès est mise en évidence. Il est donc possible de restreindre le calcul de la proportion à un plan localisé au milieu de l'unité et qu'on estime représentatif de toute l'unité.

Puisque dans le processus d'inférence de l'intensité de Poisson proposé, la proportion du faciès doit être calculée sur une grille relativement fine, dans un premier temps l'estimation des proportions sera faite à partir des CPV avec un voisinage de calcul très fin en X et recouvrant toute la région en Y . Les paramètres de cette grille de proportion sont dans le tableau 8.2.

Proportions			
	X	Y	Z (profondeur)
Origine	1225 m	6475 m	-20.75 m
Maille	500 m	13000 m	0.5 m
Nombre de mailles	95	1	42

TAB. 8.2 – Grille des proportions 2D

À partir des 15 coupes sédimentologiques et de la grille de proportion définie, 14 courbes de proportion verticale ont été obtenues (fig. 8.10). Il y a eu un seul regroupement de coupes (Dead Horse et White Rim) pour construire une CPV (CPV 4), les CPV restantes ont été calculées à partir d'une seule coupe. Ceci rejoint le problème de représentativité évoqué plus haut : chaque CPV symbolise un comportement très local sur un petit domaine ; il est possible de trouver un groupe de CPV proches (e.g. CPV 4, 5 et 6) présentant

des comportements différents. Ceci peut induire des discontinuités dans l'estimation de la proportion d'un faciès.

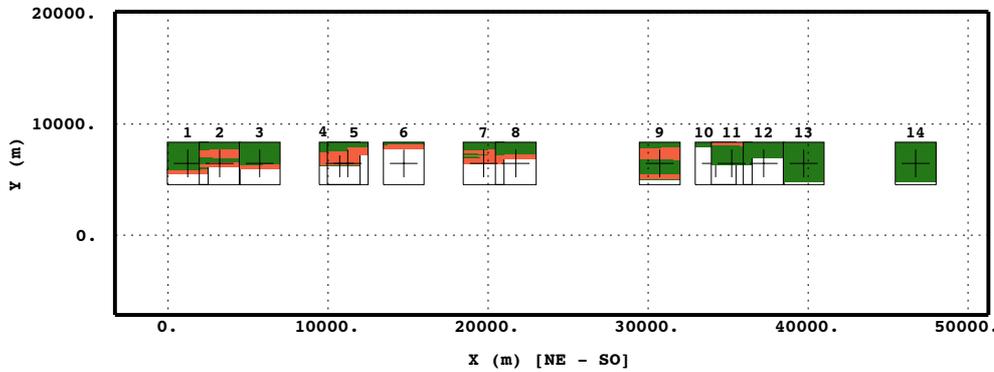


FIG. 8.10 – Courbes de proportion verticales (CPV) du faciès chenal (en rouge) et du faciès complémentaire. Ces courbes sont incomplètes à cause de la différence d'épaisseur à l'intérieur de l'unité et du type de maillage choisi.

Les courbes de proportion obtenues ne sont pas complètes sur toute l'unité puisque les coupes discrétisées n'arrivent pas à la base de l'unité dans le repère de travail. Pour pouvoir estimer la proportion sur toute la région, il a été nécessaire de compléter ces courbes. Cela implique d'introduire une hypothèse sur la présence des faciès dans les niveaux inférieurs de l'unité. Le processus de remplissage réalisé est le suivant : après un lissage simple sur quatre points, par souci de continuité, les niveaux inférieurs de chaque CPV ont été complétés avec le lithotype du dernier niveau connu (fig. 8.11). Ainsi, les CPV de la moitié SO du champ ont été complétées avec le lithotype vert (non chenal) et les CPV de la moitié NE avec le lithotype chenal. Le problème de continuité évoqué plus haut est ici mis en évidence lors du remplissage de la CPV 4. Bien que très proche des CPV 5 et 6, son comportement dans les derniers niveaux est différent de celles-ci : les CPV 5 et 6 ont été remplies avec le lithotype chenal tandis que la CPV 4 devrait l'être avec le lithotype complémentaire. Nous avons choisi de respecter le même comportement à la base et donc la CPV 4 a été complétée avec le faciès chenal. Nous avons ainsi introduit une légère discontinuité qui se reflète dans l'estimation de la carte de proportion (fig. 8.12) autour de $x \approx 10000$ m.

Cette estimation est nécessaire vu qu'il n'existe pas d'information sur un grand nombre de mailles de la grille de proportion. Puisque dans le plan horizontal les informations disponibles sont peu nombreuses, les variogrammes correspondants sont très fluctuants et ne permettent pas de déterminer une structure spatiale claire. L'estimation sera réalisée très simplement par krigeage en imposant un modèle de variogramme exponentiel de très longue portée horizontale (20 km).

Le choix d'un maillage parallèle a conditionné le calcul des proportions et a obligé à remplir les CPV. Une conséquence directe du choix de remplissage des CPV est une présence artificiellement élevée, dans l'estimation de la proportion, du lithotype chenal dans les niveaux inférieures de la région NE. Ceci implique que le nombre d'objets à simuler sera très important du fait que la simulation est réalisée sur toute une région «ajoutée» lors de la discrétisation et qui sera éliminée a posteriori quand la simulation sera transposée dans le repère réel.

L'utilisation d'un maillage proportionnel aurait limité la simulation strictement à la région d'intérêt sans introduire de problèmes liés au remplissage de l'unité incomplète. En échange, la définition des objets doit

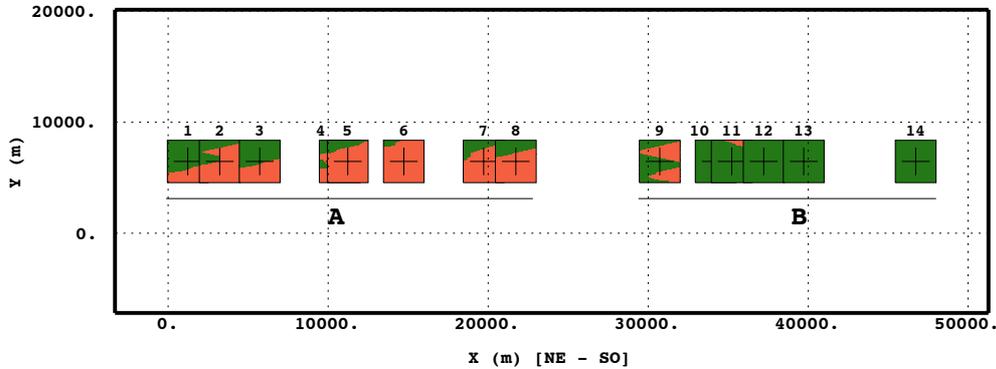


FIG. 8.11 – Courbes de proportion verticales complétées. Le groupe A a été complété avec le faciès chenal et le groupe B avec le faciès complémentaire.

tenir compte de la déformation que ceux-ci subissent dans le passage du repère de travail au repère structural.

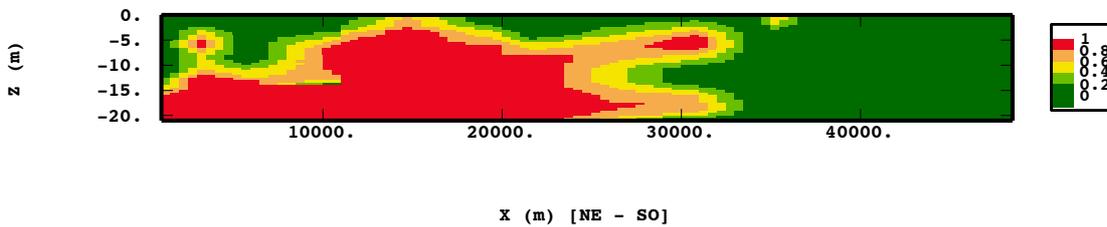


FIG. 8.12 – Carte de proportion initiale obtenue par krigeage à partir des CPV complétées. Une discontinuité dans l'estimation à $x \approx 10000$ m est présente.

8.1.6 Définition des objets

Le lithotype étudié correspond à des dépôts fluviaux possédant une géométrie de chenaux faiblement sinueux qui traversent le champ de travail selon la direction Y (NO). Ces chenaux sont associés à des demi-cylindres de génératrice sinusoidale et de section verticale semi-ellipsoïdale évoluant selon cette direction perpendiculaire au plan de non stationnarité. Faute d'informations supplémentaires, les dimensions des objets dans le plan de la non stationnarité ont été choisies en se basant sur le rapport IFP initial (Clément, 1999) selon lequel les caractéristiques des objets ont été déduites à partir de la géologie. Nous avons introduit une incertitude de $\sim 20\%$ dans chacun des paramètres :

- orientation : perpendiculaire au plan XZ , avec une tolérance de $\pm 10^\circ$;
- largeur (selon X) : $500 \pm 20\%$ m ;

- longueur (selon Y) : la profondeur du champ, 28000 m ;
- hauteur (selon Z) : $4 \pm 25\%$ m ;
- période : $3000 \pm 20\%$ m ;
- amplitude : $300 \pm 20\%$ m.

Il faut traduire ces caractéristiques en termes de paramètres du modèle, i.e. de ρ . Puisque le processus d'inférence est réalisé à 2D, les objets sont des demi-ellipses. La probabilité associée a été calculée dans la section 4.2.1. La maille de discrétisation choisie, dans le plan de non stationnarité est de $\Delta x = 100$ m, $\Delta z = 0.5$ m, assez fine pour pouvoir décrire les objets avec une précision suffisante.

8.2 Calcul de l'intensité

Les variogrammes expérimentaux de la proportion dans le plan XZ (fig. 8.13) montrent que les dimensions des objets s'adaptent à l'échelle de variabilité de la proportion. Les paramètres de krigeage utilisés pour le calcul de cette matrice de proportion, notamment selon la direction X , ont donné comme résultat une proportion très lisse qui permet d'utiliser des objets d'assez grande taille.

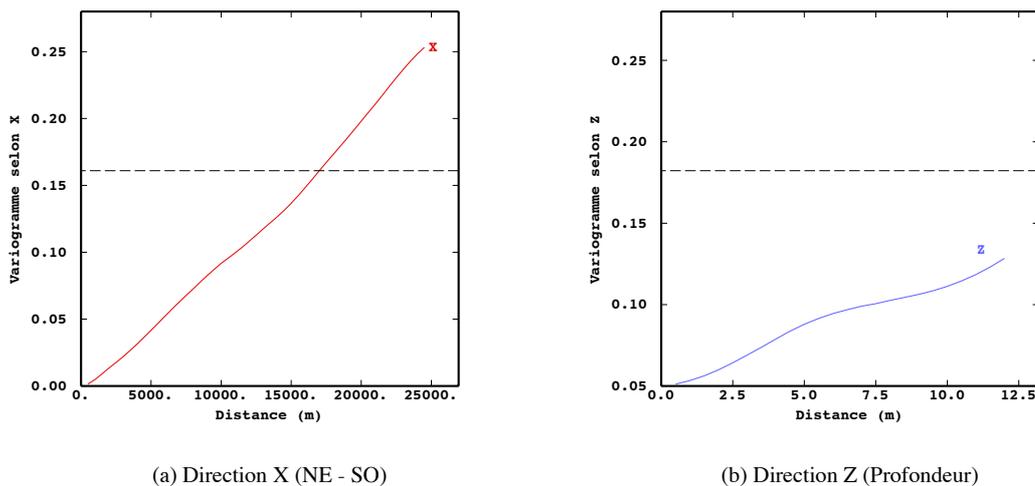


FIG. 8.13 – Variogrammes expérimentaux de la proportion initiale selon les axes principaux de la grille.

Une fois la proportion du lithotype chenal et les objets définis, nous pouvons calculer l'intensité par le processus de déconvolution. Cette matrice donne la distribution des objets dans le volume. Nous calculons de même le nombre moyen d'objets qu'il est nécessaire d'introduire dans la simulation. Comme indiqué dans la section 5.2.3, l'intensité doit tenir compte des objets qui sont générés en dehors de la région étudiée mais qui l'intersectent. Il faut alors simuler sur une zone plus grande qui va dépendre de la taille des objets. L'intensité doit être connue dans toute cette zone. Pour cela nous avons extrapolé la grille de proportion.

Une interpolation a aussi eu lieu pour égaliser les mailles de discrétisation de la grille de proportion et de la grille de ρ . Même si la proportion a déjà été estimée sur une grille très fine par rapport à la dimension du volume de simulation, elle n'est pas encore assez fine pour permettre de décrire les objets avec la précision requise. Un affinement de la grille de proportion par interpolation est nécessaire : elle passe de $\Delta x = 500$

m à $\Delta x' = 100$ m. Cette étape est inévitable dans un cas réel puisque le choix de la maille de la grille de proportion n'est a priori déterminé que par des raisons géologiques et de précision des données et non par la dimension des objets de simulation. La grille d'intensité ainsi obtenue est définie dans le tableau 8.3.

	X (NE)	Y	Z (profondeur)
Origine	775 m	6475 m	-21.75 m
Maille	100 m	13000 m	0.5 m
Nombre de mailles	482	1	53

TAB. 8.3 – Grille d'intensité

L'intensité obtenue par déconvolution, après lissage, est présentée à la figure 8.14. Le nombre moyen d'objets à introduire dans la simulation est de $N = 1266$ objets, ce nombre étant supérieur à celui de 787 estimé dans le travail de Clément (1999). Il y a deux raisons pour cela : nous avons considéré toute l'unité, au lieu de la moitié NE et, par ailleurs, le maillage choisi et la façon dont les CPV ont été complétées (avec le faciès chenal dans les niveaux profonds) ont induit un volume fictif de simulation et donc, un nombre de chenaux superflus. Une fois que la simulation sera dans le repère structural, ces chenaux superflus seront éliminés.

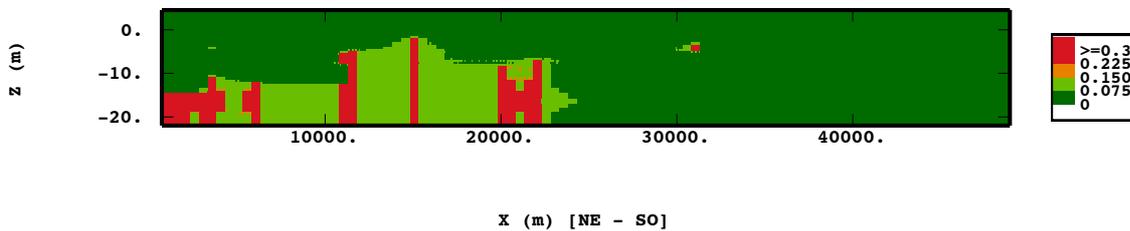


FIG. 8.14 – Carte de l'intensité obtenue par déconvolution et après correction.

La proportion du modèle correspondant à ce couple intensité - objet (fig. 8.15) présente une configuration plus lisse que la proportion initiale, notamment avec une légère diminution des valeurs maximales (fig. 8.16), mais tout en respectant les caractéristiques générales et la valeur moyenne sur tout le volume : 43.7% pour la proportion initiale et 43.9% pour la proportion du modèle.

La valeur moyenne des valeurs absolues des différences entre les données de proportion et la proportion du modèle est de 3%. Les écarts les plus importants correspondent à la région de discontinuité ($x \approx 10000$ m) dans les données, qui a été considérablement lissée dans le modèle. Cette discontinuité correspond aux cinq points isolés dans le nuage de corrélation de la figure 8.17, de faible valeur dans la proportion initiale ($< 50\%$) et de forte valeur dans le modèle ($> 70\%$). Le nuage de corrélation montre aussi que le modèle répond assez bien aux données, les points qui s'en éloignent le plus correspondant à des régions de transition (par exemple le passage du rouge à l'orange dans la carte des proportions). Ces régions de transition sont plus importantes dans le modèle que dans les données.

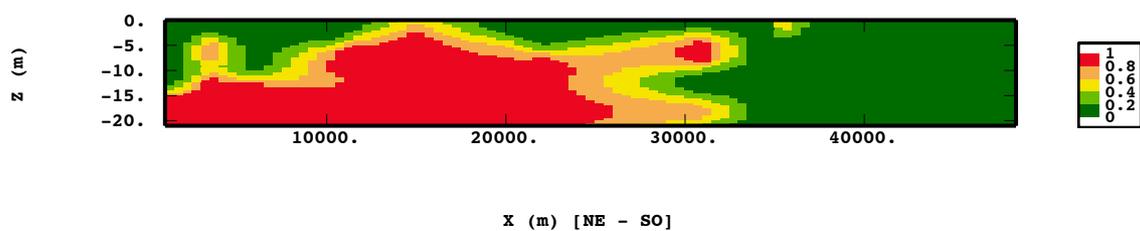


FIG. 8.15 – Proportion du modèle correspondant à l'intensité de Poisson calculée avec des objets demi-ellipsoïdes.

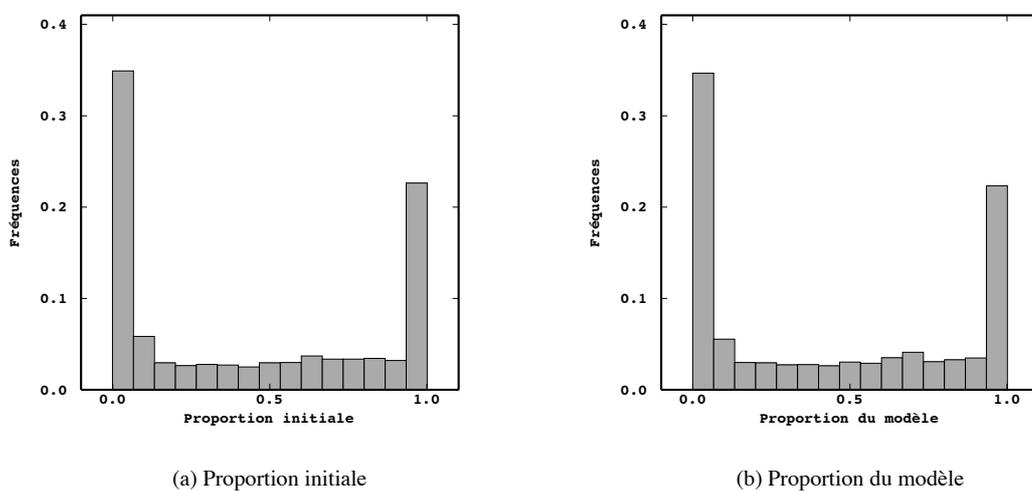


FIG. 8.16 – Histogrammes des valeurs de la proportion initiale (à gauche) et de la proportion correspondant au modèle.

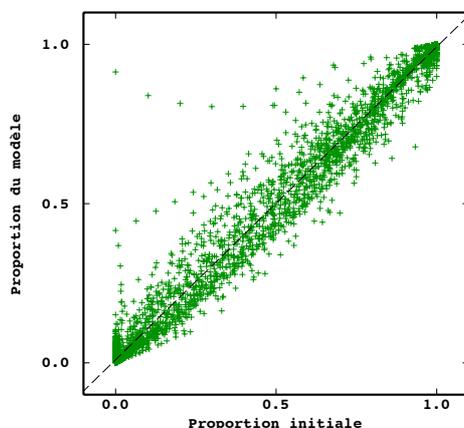


FIG. 8.17 – Nuage de corrélation entre les valeurs de la proportion initiale et la proportion du modèle.

8.3 Simulations

Les études paléogéographiques effectuées sur la région considérée et les dimensions de cette région nous ont permis de faire deux hypothèses fondamentales pour la réalisation des simulations (voir section 8.1.4) :

- les chenaux sont sub-parallèles le long de la direction d'écoulement (N-NO), perpendiculairement au plan du transect régional,
- tous les chenaux traversent, en moyenne, l'unité considérée.

Ces hypothèses nous ont permis de considérer une stationnarité selon la direction d'écoulement des chenaux. Le processus que nous avons envisagé, dans ce contexte, permet de restreindre le problème à la simulation booléenne 2D. Il consiste à simuler les intersections des chenaux avec un plan vertical perpendiculaire à leur direction d'écoulement. Ce plan est appelé le *plan de germination* et correspond au plan de la proportion et de l'intensité de Poisson. Puis, les chenaux vont évoluer à partir de ce plan de façon à traverser l'unité. La direction d'évolution (ou d'écoulement) de chacun des chenaux est tirée aléatoirement selon une loi uniforme ayant pour moyenne la direction perpendiculaire au plan de germination et avec une faible tolérance ($\leq 10\%$). Une tolérance plus importante risquerait de faire diverger les chenaux qui ne respecteraient plus l'hypothèse de chenaux sub-parallèles.

Dans ces conditions, on admet que l'intersection d'un plan quelconque, parallèle au plan de germination, avec une réalisation du modèle est aussi un modèle booléen à 2D des mêmes paramètres. Le plan de germination peut alors être localisé aléatoirement dans le volume de simulation, puisqu'il est représentatif de toute la simulation.

Dans un cas plus général où les chenaux ne sont pas sub-parallèles entre eux (cas d'un delta, par exemple), l'hypothèse de stationnarité selon la direction moyenne d'écoulement n'est plus valable. Les intersections des chenaux avec différents plans ne correspondent pas au même modèle booléen. Un seul plan de germination n'est donc pas représentatif de toute la simulation et son emplacement joue maintenant un rôle important dans le processus de simulation. La construction mise en œuvre n'est pas applicable dans ce cas.

Résultats

Les paramètres de la grille de simulation utilisée pour cette unité se trouvent dans le tableau 8.4.

Simulation			
	X	Y	Z (<i>profondeur</i>)
<i>Origine</i>	1000 m	0 m	-20.75 m
<i>Maille</i>	50 m	50 m	0.5 m
<i>Nombre de mailles</i>	950	260	42

TAB. 8.4 – Grille de simulation

Un exemple de simulation obtenue avec cette approche est présenté figures 8.18 et 8.19 correspondant respectivement au repère de travail et au repère structural. Les simulations sont conditionnelles aux valeurs de l'indicatrice du faciès chenal dans les coupes sédimentologiques. Notons la forte présence des chenaux dans la région NE et leur diminution progressive vers le SO, comme le montre la proportion du faciès chenal (cf. fig. 8.12).

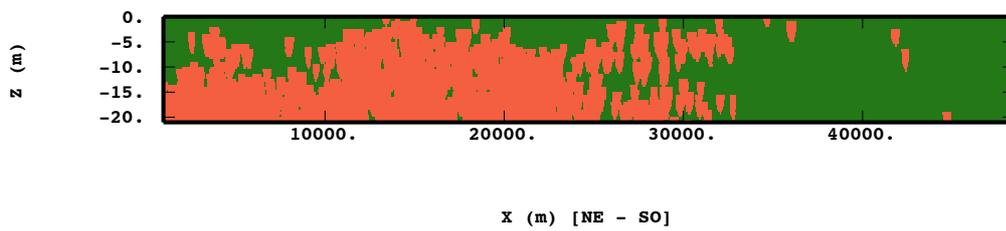
La validation des résultats de simulation est réalisée comme dans le cas 3D : en comparant les cartes de proportion calculées sur les simulations avec la carte de proportion initiale, et en comparant les proportions globales moyennes sur tout le volume de simulation. Ce sont les deux critères que notre méthode doit vérifier. L'analyse s'effectue à deux échelles : à l'échelle d'une seule simulation et à l'échelle de toutes les réalisations calculées (~ 30). Les caractéristiques générales remarquées sur la simulation présentée sont communes à toutes les simulations.

Pour cette simulation, la carte de proportion (fig. 8.20) montre que la variabilité dans la présence du faciès chenal est respectée. On retrouve les caractéristiques générales de la carte de proportion initiale, bien que la variabilité à l'échelle de la maille de simulation soit plus importante.

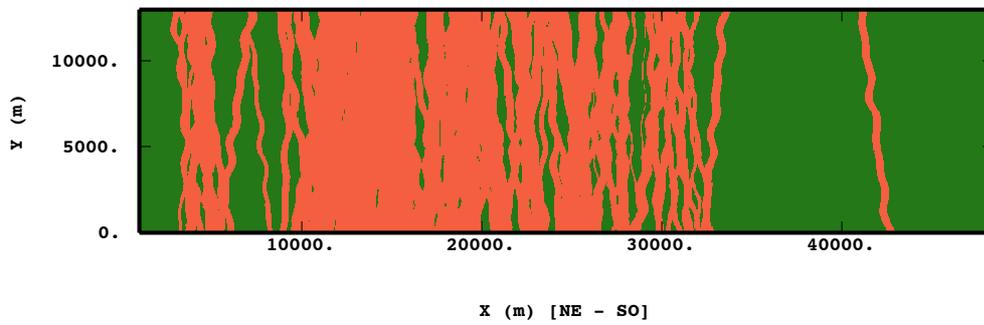
Comme dans le cas 3D précédent, on remarque une présence plus importante des valeurs extrêmes (proches de 0% et de 100%) de la proportion dans la simulation par rapport à la proportion initiale et à la proportion du modèle (voir les histogrammes de la figure 8.21). Par contre, les valeurs intermédiaires apparaissent avec des fréquences plus faibles dans la simulation. Dans la carte de proportion de la simulation, cela se traduit par des variations brusques : les fourchettes de valeurs entre 40% – 80% (couleurs jaune et orange dans la figure 8.20) sont peu représentées par rapport à la proportion initiale. Ainsi, la proportion globale de la simulation est plus forte que celles du modèle et des données (44.7% dans la simulation par rapport à 43.9% du modèle et 44.3% de la proportion initiale).

La raison à cela est que la taille des objets, notamment dans la direction X (400 – 600 m), est comparable à la maille de calcul de la proportion dans cette direction (500 m). La présence ou l'absence d'un objet dans la simulation peut faire varier de 100% à 0% sa proportion dans une maille. De plus, la carte de proportion initiale montre que les régions de transition entre les valeurs très fortes et très faibles sont très étroites ce qui, dans les simulations, se traduit par des variations encore plus brusques et une tendance à augmenter la valeur de la proportion globale par rapport à celle du modèle. Pour les simulations réalisées, les proportions moyennes globales sur tout le volume sont comprises entre 43% et 47% (histogramme fig. 8.22).

La carte de proportion moyenne calculée sur toutes les simulations réalisées (fig 8.23(a)) montre qu'on retrouve bien le modèle et que, en moyenne, la méthode proposée nous permet de récupérer de façon très satisfaisante les données initiales. Les valeurs de l'écart type (fig. 8.23(b)) des proportions des 30 simulations par rapport à leur proportion moyenne sont comprises entre 0 et 0.33, les écarts maximaux correspondant aux régions des valeurs de proportion intermédiaire.



(a) Section XZ



(b) Section XY

FIG. 8.18 – Exemple de simulation dans le repère de travail.

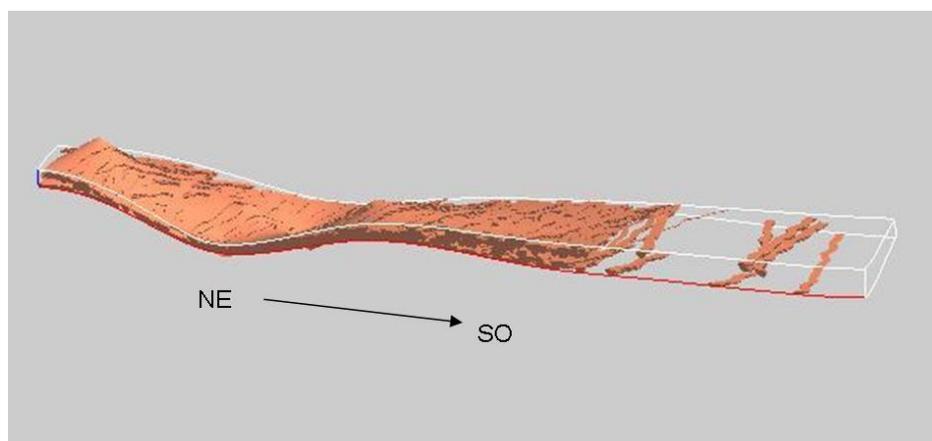


FIG. 8.19 – Exemple de simulation dans le repère structural de l'unité.

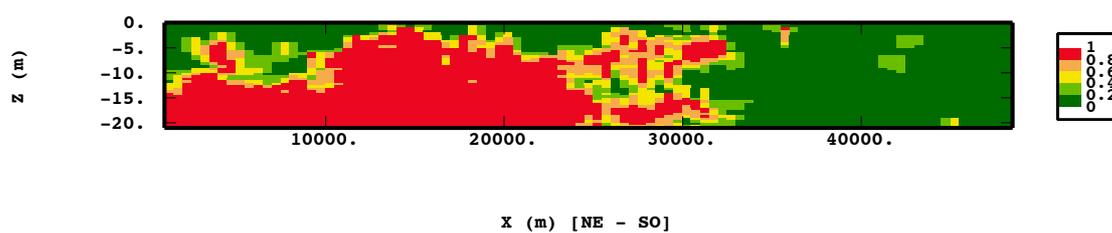


FIG. 8.20 – Proportion calculée sur un exemple de simulation.

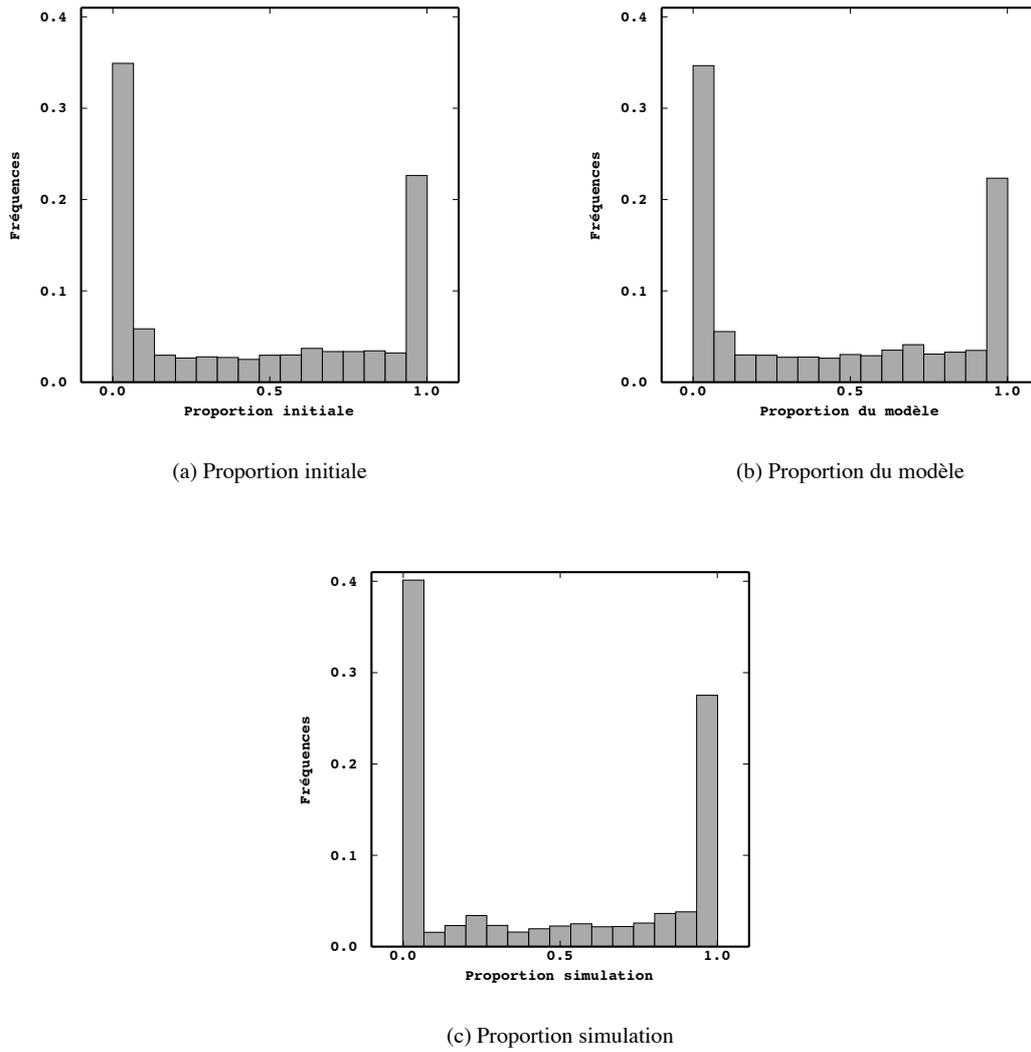


FIG. 8.21 – Histogrammes des valeurs de la proportion initiale, du modèle et de la proportion calculée sur la simulation exemple.

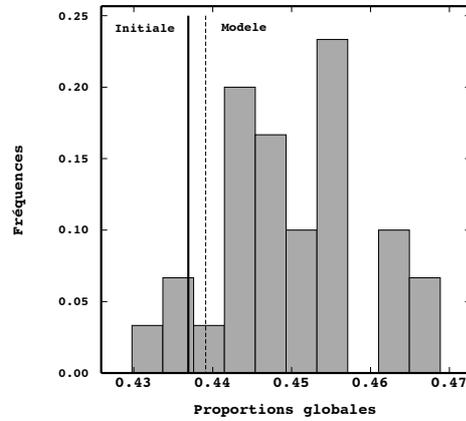
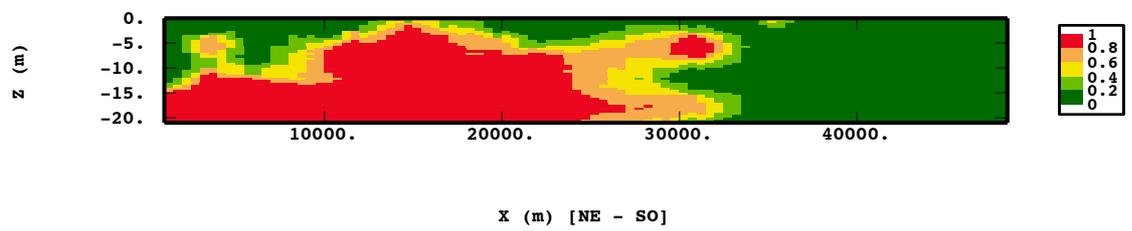


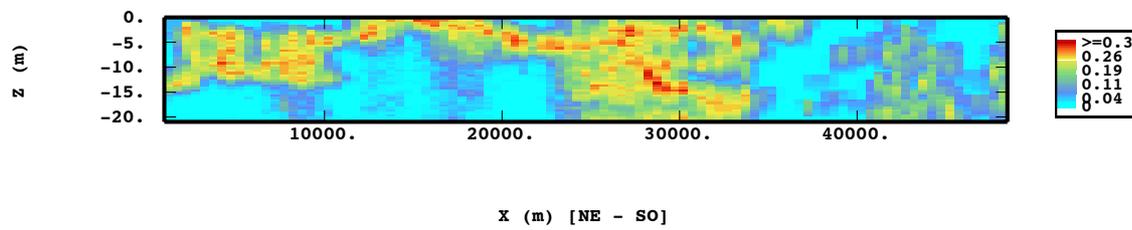
FIG. 8.22 – Histogramme des valeurs des proportions globales (moyennes sur tout le volume de simulation) obtenues dans toutes les simulations réalisées. Le trait noir continu représente la valeur globale de la grille initiale. Le trait pointillé est la valeur globale moyenne de la grille modèle.

Le nuage de corrélation (fig. 8.24) entre cette carte de proportion moyenne et la proportion initiale reflète ce résultat : les valeurs intermédiaires des proportions initiales correspondent, dans les simulations, à des valeurs comprises dans une fourchette très large ; par exemple, à une valeur de proportion initiale de 50%, peuvent correspondre des valeurs entre 30% et 80% dans les simulations. Les cinq valeurs isolées du nuage de corrélation correspondent aux points de discontinuité dans l'estimation de la carte de proportion initiale.

Pour pouvoir comparer plus en détail les proportions initiales et celles des simulations, nous représentons les proportions cumulées verticalement et horizontalement (cf. figures 8.25(a) et 8.25(b)). Toutes les simulations reproduisent, en moyenne, les comportements de la proportion du modèle (en rouge dans les figures) et de la proportion initiale (en noir) avec, comme on l'a vu précédemment, une légère surestimation de la proportion initiale.



(a) Proportion moyenne sur toutes les simulations



(b) Écart type

FIG. 8.23 – Cartes de la proportion moyenne et de l'écart type de toutes les simulations calculées.

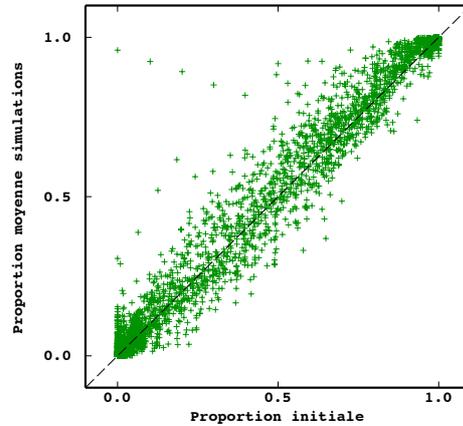


FIG. 8.24 – Nuage de corrélation entre les valeurs de la proportion initiale et la proportion moyenne de toutes les simulations.

8.4 Discussion et conclusions

L'unité choisie dans ce cas d'étude présente une évidente non stationnarité horizontale et verticale dans la répartition du faciès chenal et elle constitue un exemple très représentatif du type de non stationnarité que nous voulons reproduire dans les simulations booléennes.

L'analyse des résultats obtenus ici montre les mêmes caractéristiques que dans le cas présenté dans le chapitre précédent. Les proportions du modèle obtenu sont plus lisses que celles des données, notamment les maxima et minima locaux sont moins marqués ainsi que les passages brusques entre valeurs fortes et valeurs faibles.

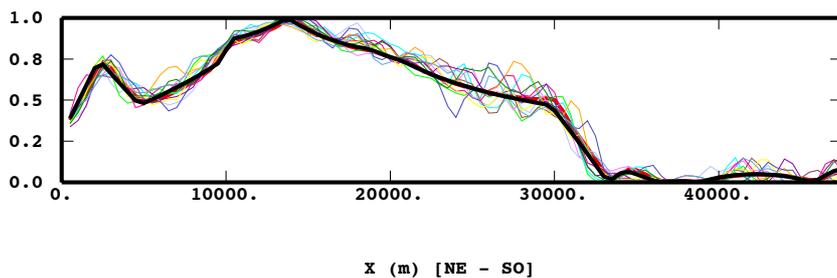
Dans les proportions calculées sur les simulations, par contre, la présence des valeurs extrêmes est plus importante au détriment des valeurs intermédiaires. C'est un effet inhérent à l'application du modèle booléen au problème géologique où les objets sont de taille considérable par rapport à la maille de la grille des proportions. Un seul objet peut alors faire varier la proportion dans une maille de 0% à 100%. Cependant, aussi bien les proportions du modèle que celles des simulations restituent, de façon globale, les caractéristiques générales des données de proportion.

Il y a deux points importants évoqués tout au long de ce chapitre sur lesquels il est important de revenir : la définition des objets et le calcul des proportions. Bien qu'ils n'aient pas une influence directe sur la méthode d'estimation de l'intensité, ils jouent un rôle très important dans le processus de simulation et d'interprétation des résultats.

Nous cherchons à simuler des **objets**, les chenaux, qui traversent tout le volume de simulation : ils sont de taille infinie. Ceci pose un double problème : d'abord, comment définir un objet infini sur une grille discrète et finie et ensuite, comment le décrire en terme de la probabilité associée p .

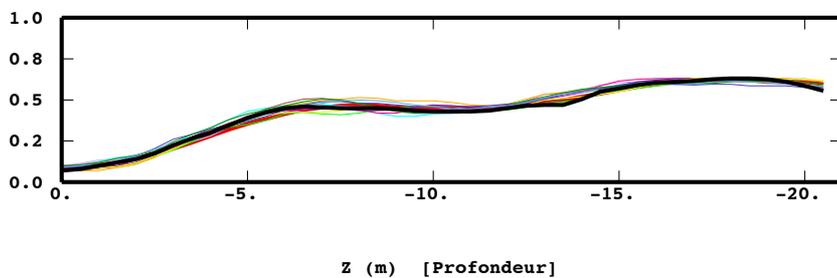
Dans ce cas particulier nous avons choisi de poser une hypothèse de stationnarité (chenaux sub-parallèles perpendiculaires au plan de la falaise) pour contourner ce problème. En effet, la propriété du modèle booléen selon laquelle son intersection avec un plan vertical est aussi un modèle booléen a été exploitée ici pour réduire le problème 3D à un problème 2D. Ainsi, le processus de Poisson des germes des objets est réalisé sur un plan 2D et les chenaux progressent de façon à traverser toute l'unité le long d'une direction

Proportions cumulées en Z



(a) Proportion cumulée verticalement

Proportions cumulées en X



(b) Proportion cumulée horizontalement

FIG. 8.25 – Courbes de proportion cumulées verticalement (a) et horizontalement (b). En noir la courbe correspondant à la proportion initiale. En rouge la courbe correspondant au modèle. L'axe des profondeurs Z dans la figure (b) est dilaté pour mieux observer les courbes.

tirée aléatoirement selon une loi uniforme dont la moyenne correspond à la direction estimée d'écoulement et avec une faible tolérance. Naturellement il s'agit d'une solution singulière pour un cas particulier. L'étape suivante consisterait à proposer une autre démarche pour un cas plus général où l'hypothèse de parallélisme des chenaux n'est plus applicable.

Quant au **calcul des proportions**, dans notre processus de déconvolution pour l'obtention de l'intensité de Poisson nous avons supposé que la matrice (ou grille) de proportion est une donnée de départ. Le support de calcul des proportions initiales, bien que n'ayant pas une influence directe sur le processus d'estimation de l'intensité, détermine l'échelle de variabilité que notre modèle devra reproduire.

Dans le cas proposé ici, les proportions verticales ont une représentativité à petite échelle. La matrice de proportion a été obtenue par estimation à partir des CPV, avec une incertitude associée. Dans ce cas nous exigeons de notre modèle qu'il reproduise ce comportement très local, ce qu'il arrive à faire avec un certain lissage.

On pourrait s'interroger sur le résultat de l'estimation de la proportion si une maille plus large avait été choisie. La figure 8.26(a) montre une carte de proportion calculée sur une maille 20 fois plus large en X . Dans ce cas, seulement 5 CPV ont été obtenues, une pour chaque nœud de la grille de proportion. Celle-ci est alors complètement informée et l'étape d'estimation n'est pas nécessaire. La proportion ainsi obtenue ne présente pas de variations à petite échelle : seules les grandes tendances sont mises en évidence. De plus, la proportion moyenne globale a diminué par rapport au cas étudié dans ce chapitre : 34.5% ici par rapport à 43.7% précédemment. La valeur de la proportion globale est donc aussi très dépendante de l'estimation des proportions.

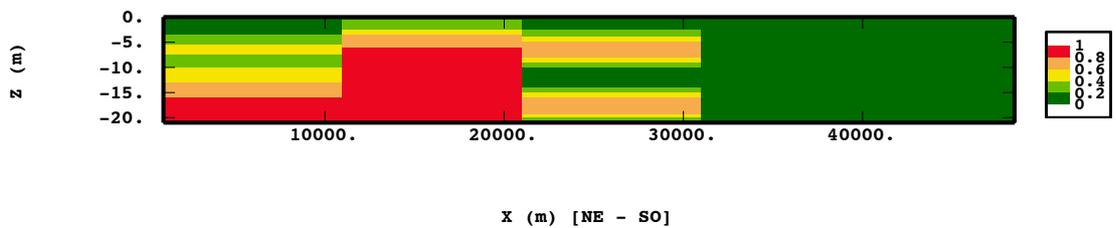
La proportion du modèle obtenu (fig. 8.26(b)) reproduit très bien cette variabilité à grande échelle et l'effet de lissage est moins évident. Si on observe la proportion sur une des simulations obtenues (fig. 8.26(c)), il apparaît qu'elle est plus proche de la proportion initiale que le cas étudié dans ce chapitre. Le fait de calculer les proportions sur un support large induit une dissolution des variations à petite échelle. La présence ou absence d'un objet est ici moins critique et de ce fait les valeurs intermédiaires de proportions, déjà peu nombreuses dans la proportion initiale, sont mieux reproduites.

Cependant, le processus de déconvolution mis en œuvre requiert une grille de proportion assez fine pour pouvoir décrire les objets avec une précision suffisante. Idéalement nous aimerions travailler avec la probabilité ponctuelle d'appartenance aux faciès en question. Les proportions sont, comme il a été déjà expliqué, les moyennes de ces probabilités sur un certain support. On pourrait se demander s'il est possible d'estimer ces probabilités ponctuelles à partir des proportions.

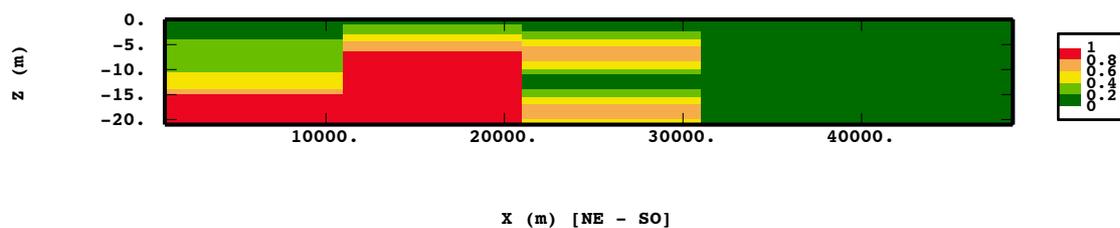
Hu (2002) propose une méthode pour estimer les proportions lithologiques sur un support quelconque à partir des données d'indicatrices aux puits et des essais de puits. En particulier, quand ce support est réduit à un point, il s'agit des probabilités discrètes d'appartenance à un certain faciès. La méthode est basée sur le *krigeage itératif sous contrainte d'agrégation* et requiert des informations supplémentaires aux données de puits. Dans notre cas, nous nous sommes intéressés aux probabilités ponctuelles discrètes mais avec un pas de discrétisation très petit, ce qui peut alourdir notablement le système de krigeage impliqué.

Une autre approche envisageable serait de simuler ces probabilités de façon à ce qu'elles respectent les proportions comme une fonction moyenne tout en introduisant une variabilité à l'intérieur de chaque maille.

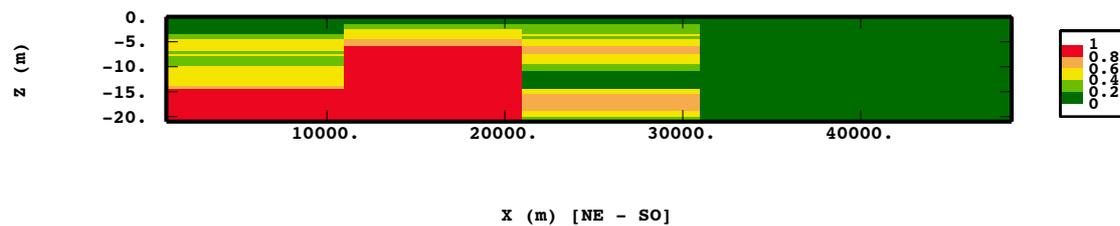
Mis à part ces considérations et pour conclure, la méthode proposée donne de bons résultats par rapport à l'objectif recherché. En effet, les simulations obtenues respectent les proportions latérales et verticales, la proportion globale et les données aux puits (il s'agit de simulations conditionnelles). À partir d'une matrice de proportion et les caractéristiques des objets étant fixées a priori, nous sommes capables d'estimer une matrice d'intensité qui détermine la répartition des objets dans le volume et le nombre moyen d'objets à



(a) Proportion initiale



(b) Proportion du modèle



(c) Proportion sur une simulation

FIG. 8.26 – Carte de proportion obtenue à partir de 5 courbes de proportion verticale (a) ; proportion correspondant au modèle calculé (b) et proportion calculée sur un exemple de simulation de ce modèle (c).

simuler, qui se déduit naturellement de l'intensité de Poisson. Nous avons ainsi répondu à la question posée au début de ce chapitre. Les principales difficultés résident en une estimation correcte de la matrice de proportion et une définition adéquate des objets à simuler.

Conclusions

Chapitre 9

Conclusions et perspectives

Cadre du projet

Le modèle booléen présente un important problème dans son application à la modélisation des réservoirs pétroliers. En effet, la réalité est souvent non stationnaire, dans le sens où les structures à reproduire dans la simulation sont réparties de façon non homogène dans le gisement et cette information est difficile à intégrer dans le modèle booléen.

Les proportions des faciès qui constituent ces structures fournissent une information sur leur distribution spatiale et constituent un outil très efficace pour déterminer la présence d'une non stationnarité. Cet outil permet de quantifier l'information géologique des dépôts représentant ainsi une importante contrainte à respecter pour obtenir des simulations géologiquement réalistes. Ainsi, les proportions ont été utilisées avec succès pour intégrer cette information dans les simulations géostatistiques de gaussiennes seuillées (Ravenne et Beucher, 1988 ; Beucher et al., 1997 ; Doligez et al., 1999). Dans cette démarche, la question qui se pose de façon naturelle est de savoir si cet outil peut être aussi introduit dans les simulations de type objet et en particulier dans le modèle booléen.

Différentes approches existent déjà pour contraindre les simulations d'objets avec les données de proportions. De façon générale, elles visent à minimiser une fonction objectif relative à la différence entre les proportions cherchées et les proportions des simulations. Ainsi, des objets sont incorporés successivement dans les simulations jusqu'à ce que la proportion globale moyenne du faciès dans le volume («net-to-gross ratio») soit atteinte (Shmaryan and Deutsch, 1999 ; Henriquez et al., 1990 ; Lanzarini et al., 1997 ; Tavares et al., 2002). Dans certaines approches, le respect des distributions latérale et verticale des proportions est réalisé en adaptant l'intensité de Poisson aux proportions et, par des processus d'essais et erreurs, en ajoutant ou en enlevant des objets, en modifiant leur taille ou en variant leur position (Deutsch et Wang, 1996 ; Journel et al., 1998 ; Deutsch et Tran, 2002).

Le problème a été ici abordé autrement : on a essayé d'intégrer directement dans le modèle l'information contenue dans les proportions. Le problème abordé est donc l'*inférence de paramètres* du modèle booléen en tenant compte des proportions. Le problème de l'inférence dans le modèle booléen a été largement abordé dans la littérature et différentes méthodes ont été proposées (Steiner (dans Serra, 1982), Weil (1988) ou Schmitt (1991), et Lantuéjoul et Schmitt (1991) pour une révision des trois méthodes) mais les premiers travaux sur l'inférence appliquée au cas particulier de la modélisation de lithofaciès sont dus à Schmitt et Beucher (1997) et Schmitt (1997a). Ces travaux ayant abouti à des résultats très prometteurs à 1D (verticalement), il a été décidé de les prolonger par ce travail de thèse.

La contribution de cette thèse est centrée sur deux aspects principaux : d'une part, la définition du problème d'inférence à travers l'étude théorique du modèle et des rapports existant entre ses paramètres

et les informations expérimentales et, d'autre part, le développement d'un outil informatique permettant d'automatiser le passage entre les données de proportions et les paramètres du modèle dans un cadre non stationnaire et généralisé à 3D. Ce dernier aspect nous a contraint à regarder le problème du point de vue de l'utilisateur et à le poser dans la perspective de sa mise en pratique. Ainsi, nous avons essayé de rendre le processus directement applicable en proposant une méthode simple et rapide à implémenter et qui puisse être automatisée. L'application à deux cas nous a permis de valider le processus, d'évaluer les résultats obtenus et de mettre en évidence les limites de la méthode.

La résolution du problème d'inférence a demandé la coordination d'éléments appartenant à différentes disciplines : la géologie des réservoirs pour interpréter les informations disponibles et valider les résultats, la géostatistique pour décrire les paramètres du modèle booléen en fonction des informations et, dans une certaine mesure, l'analyse du signal pour résoudre le problème lié à la déconvolution impliquée dans le processus d'inférence.

Les résultats ...

Le résultat principal de ce travail est l'estimation des deux paramètres du modèle booléen à partir des proportions du faciès modélisé :

- L'intensité de Poisson a été inférée à partir de ces proportions en fixant les caractéristiques des objets. La distribution des objets dans le volume est ainsi déterminée, de même que le nombre moyen d'objets à simuler pour respecter les données des proportions.
- Les variogrammes des proportions ont fourni un test qualitatif sur l'adéquation des objets choisis par rapport à l'échelle de variabilité des proportions ; il est ainsi possible de tester, avant le processus d'inférence, si les objets sont appropriés pour reproduire les proportions données dans les simulations.

Les proportions du faciès analysé sont ainsi restituées dans le modèle et dans les simulations de façon naturelle pour les deux aspects fondamentaux suivants : leur distribution spatiale (verticale et latérale) et leur valeur moyenne globale sur le volume d'étude. Le nombre moyen d'objets à simuler est déduit automatiquement du modèle obtenu, et il n'est donc pas nécessaire de comparer les proportions sur les simulations et les données à chaque nouvel objet ajouté, celles-ci étant respectées par construction.

Dans le processus de déconvolution proposé, nous avons introduit un bruit en tant qu'élément stabilisateur. Dans le cadre de ce travail nous avons vu que, pour des raisons pratiques et par manque d'information, ce bruit n'a été considéré que comme un artefact mathématique. Les proportions ont alors été considérées comme «exactes» et donc comme un paramètre important à restituer. Cependant, le bruit peut être associé à l'incertitude sur les proportions et, si celle-ci peut être estimée et modélisée, elle peut, à l'aide de cette méthode, être intégrée au calcul des paramètres du modèle. De ce fait les proportions obtenues ne correspondront pas aux proportions de départ, ce qui est logique puisqu'elles sont entachées d'une incertitude. Les proportions obtenues seront d'autant plus différentes des proportions de départ que l'incertitude sur ces dernières est importante.

La méthode proposée est facilement et rapidement implémentée. Le fait que le processus d'inférence soit automatisé dans une étape préalable aux algorithmes de simulations (conditionnelles ou non conditionnelles) ne les alourdit pas, d'autant plus qu'aucun processus itératif n'est nécessaire. L'intervention de l'utilisateur dans cette phase consiste à introduire les proportions, choisir les objets et, éventuellement, valider le modèle par rapport aux proportions initiales, sans avoir à spécifier la présence ou non d'une non

stationnarité puisque le cas stationnaire correspond à une proportion constante.

Les objets utilisés dans les exemples présentés ici sont assez élémentaires, mais il est tout à fait possible, du point de vue théorique, d'utiliser des objets plus réalistes, donc plus complexes. Des exemples de paramétrisation de géométries typiques dans la simulation de réservoirs pétroliers (des lobes, des chenaux, des dunes, etc.) se trouvent dans Lanzarini et al. (1997) mais il faut encore pouvoir les exprimer en fonction de leur probabilité associée p_0 . En pratique, l'expression de cette probabilité peut être très complexe ou ne pas exister analytiquement. Nous avons proposé une simplification du problème en approximant les objets réalistes par des objets plus simples dans le processus de déconvolution (par exemple un lobe par le parallélépipède qui le contient). Le nombre moyen d'objets à simuler est alors calculé à partir du nombre moyen d'objets simples corrigé par le rapport entre leurs volumes. De cette façon, l'utilisation d'objets complexes est facilement envisageable du point de vue pratique, notamment si l'approximation d'une famille d'objets par une autre est raisonnable : l'approximation d'un ellipsoïde par de petits cubes, par exemple, conduirait à des résultats absurdes.

En conclusion, la méthode proposée donne de bons résultats dans la modélisation de structures «finies», c'est-à-dire plus petites que le volume de simulation. Elle est applicable à différentes échelles : pour la modélisation des crevasses ou des barres de méandres à l'échelle du gisement, pour modéliser des hétérogénéités à l'intérieur d'un réservoir, à l'échelle granulométrique ... L'importance de bien reproduire la distribution de ces hétérogénéités est mise en évidence dans l'étude du comportement des fluides, par exemple. Le mouvement des fluides à l'intérieur d'un faciès poreux est très dépendant de sa distribution, si bien que son comportement sera différent selon que le faciès forme des structures peu interconnectées ou très interconnectées. La connexité entre ces éléments est très importante lors de l'étude d'emplacement des puits pour l'extraction des réserves, par exemple.

Dans le cas de la modélisation des chenaux dans un réservoir pétrolier, l'application directe de cette approche est réalisée dans le cas où les chenaux sont sub-parallèles et traversent l'unité d'étude. Dans ce cas, une simplification du problème d'inférence 3D à un problème 2D est possible. Les résultats, comme il a été vu dans le chapitre 8, sont aussi très satisfaisants dans le sens où la distribution de la proportion et sa valeur globale sont bien restituées.

Nous avons ainsi donné une réponse à la question posée au début de ce travail sur l'incorporation des proportions de faciès dans le modèle booléen. Cependant, certains points doivent être approfondis, notamment ceux qui concernent les limitations théoriques et pratiques de la méthode mises en évidence au cours du projet.

... et les limitations

Ainsi, les principales restrictions de notre approche du point de vue **théorique** concernent, d'une part, les hypothèses utilisées et, d'autre part, la connaissance partielle du modèle.

Les deux principales hypothèses utilisées pour poser le problème d'inférence sont :

- la présence d'une non stationnarité dans la distribution des faciès est attribuée à la distribution des objets dans l'espace, et non à la variation de leurs caractéristiques, et
- la loi des objets est indépendante de leur point d'implantation.

La distribution des proportions dans le modèle booléen est due à l'effet conjoint de la distribution des objets, θ , et de leurs caractéristiques (leur taille et leur forme), p_0 , qui peuvent aussi être variables dans

l'espace. Pour déterminer le modèle il est donc nécessaire d'estimer ces deux paramètres à partir d'une information unique, les proportions, ce qui constitue un problème indéterminé. Par ailleurs, ces deux paramètres sont liés, puisque les objets sont distribués différemment selon leur taille et leur forme. Nous devons donc faire une hypothèse sur un des deux paramètres, et nous avons choisi de le faire sur les objets, puisque leurs caractéristiques peuvent être plus facilement déterminées à partir des informations extérieures aux proportions, comme celles extraites de la sismique, des affleurements ou de réservoirs analogues, par exemple.

L'hypothèse d'indépendance de la loi des objets par rapport à leur point d'implantation nous a permis d'exprimer le problème de l'inférence de l'intensité de Poisson en terme d'une opération relativement simple comme l'est la convolution. Évidemment, il est tout à fait réaliste de supposer que les objets suivent une loi qui dépend de leur point d'implantation, par exemple qu'ils varient en taille avec la profondeur. Dans ce cas, le lien entre les proportions et les paramètres du modèle n'est plus une simple déconvolution, et le problème de l'inférence devient beaucoup plus complexe. Il faudrait ici résoudre une forme de convolution où le noyau (la probabilité associée à l'objet) varie avec la position.

Une autre limite d'ordre théorique concerne le degré de connaissance du modèle. En effet, nous déterminons les paramètres du modèle uniquement à partir des probabilités ponctuelles, contenues dans les proportions, alors que de nombreux éléments structurants plus complexes que le point sont nécessaires pour définir complètement le modèle. Cette limitation est principalement due au nombre de données disponibles. Si nous disposons de toute une image du réservoir, comme une image d'entraînement par exemple, ou d'un affleurement complet, il est possible de calculer les probabilités associées au modèle booléen pour des éléments structurants complexes. Dans ce cas, on peut obtenir l'intensité en utilisant des approches morphologiques comme celles de Steiner (dans Serra, 1982), Weil (1988) ou Schmitt (1991). Mais, dans notre application, nous sommes le plus souvent soumis à la contrainte d'avoir peu de données, ce qui impose déjà une certaine incertitude sur le modèle construit.

Du point de vue **pratique**, la méthode est limitée précisément par la source d'information utilisée, les données de proportions. En effet, les proportions ont été introduites dans le modèle comme étant les régularisées des probabilités ponctuelles d'appartenance au modèle sur un certain support, la maille de proportions. Expérimentalement nous disposons alors d'une valeur constante sur chaque maille. Dans notre approche, ceci implique que les probabilités ponctuelles à l'intérieur de ce support prennent aussi une valeur constante, ce qui est faux. Une amélioration dans la méthode concernant ce problème consisterait à simuler une carte ou un bloc de probabilités quasi-ponctuelles (des grilles à mailles très fines) qui respecte les proportions comme leur valeur moyenne. De cette façon il serait possible d'éviter les discontinuités entre les valeurs de proportion de deux mailles contiguës.

Ces discontinuités entre mailles voisines constituent une deuxième limitation pratique de notre approche. Le modèle construit ne peut pas reproduire des changements très brusques dans les proportions, en forme de marches par exemple. Ces passages d'une valeur très forte à une très faible sont plus lisses dans le modèle que nous proposons, ce qui, par ailleurs, semble cohérent avec la réalité. Cependant, il est possible de trouver dans la réalité ces changements brusques dans la présence d'un faciès, par exemple à cause d'une faille. Notre modèle ne pourra pas reproduire exactement cette rupture, mais dans ce cas il pourrait être envisageable de traiter les deux zones séparément.

Les valeurs nulles de proportion constituent aussi un point important de discussion qui rejoint les deux questions précédentes. Les valeurs nulles marquent souvent les limites des séquences ou des surfaces particulières et pour cela il est important de les retrouver dans les simulations. Comme dans le cas des dis-

continuités très marquées, la présence de proportions strictement nulles dans une région ne pourra pas être exactement reproduite par notre modèle, mais il apparaîtra une zone, déterminée par la taille des objets, des valeurs de plus en plus faibles de proportion jusqu'à atteindre les valeurs nulles. Il existe aussi ici un *effet de bord*. Si, toutefois, à cause de contraintes géologiques, il faut absolument respecter dans le modèle ces valeurs, le problème devient déterministe et il doit être abordé autrement. Simuler les probabilités ponctuelles, comme on l'a suggéré précédemment, peut aider à résoudre ce problème. On pourrait également traiter les proportions avant le processus de déconvolution de façon à créer une région de garde autour de ces valeurs nulles pour tenir compte de cet effet de bord, ou encore, imposer que l'intensité de Poisson prenne des valeurs strictement nulles dans cette région à travers les multiplicateurs de Lagrange. Dans ce dernier cas, le problème de déconvolution est traité comme un problème d'opérateurs linéaires discrets et on pourrait imposer que l'intensité prenne des valeurs nulles en certains points à l'aide des multiplicateurs de Lagrange. Il faut, toutefois, déterminer a priori ces points.

Le statut particulier attribué aux valeurs de proportion nulles peut être parfaitement étendu aux autres valeurs de proportion. Dans ce cas il ne s'agit plus de valeurs moyennes, mais de valeurs fixées et déterministes et, dans ces conditions, l'approche probabiliste n'est plus appropriée.

Ces aspects concernant les proportions ouvrent le débat sur leur représentativité et la signification que l'on donne aux proportions. Celles-ci ne sont, en définitive, que des estimations, plus ou moins précises, sur la présence des faciès dans une certaine région. Comme ces estimations sont souvent réalisées à partir d'un nombre limité de données, quelle précision devons-nous exiger de notre modèle dans le respect des proportions ? C'est à l'utilisateur de décider, guidé par son expérience et les informations dont il dispose.

La dernière limitation importante concerne l'application de notre méthode à la simulation d'objets infinis, tels que les chenaux dans un environnement fluvial. Nous avons vu qu'il est possible, sous certaines conditions (chenaux sub-parallèles, stationnarité selon la direction d'écoulement), d'appliquer le processus de déconvolution en réduisant le problème à deux dimensions. Mais, que se passe-t-il si ces hypothèses ne sont pas applicables ? Dans le cas d'un delta, par exemple, il est clair que les proportions calculées sur deux sections verticales ne sont pas représentatives du même comportement. Il existe une non stationnarité selon les trois directions. Le problème ici est de définir le point d'implantation et la probabilité associés à un objet infini par rapport au volume de simulation et qui apporte une contribution à la proportion en des points très éloignés de la grille de proportion. Ce problème se situe probablement à la limite d'une approche purement booléenne.

Perspectives

Outre les améliorations à introduire dans notre méthode déjà mentionnées (comme la prise en compte de l'incertitude dans les proportions, le traitement des valeurs nulles, l'utilisation des probabilités ponctuelles...) les suites naturelles de ce travail seront son application à la modélisation des chenaux non parallèles et la simulation multifaciès.

La méthode proposée dans ce travail ne semble pas très appropriée pour inférer l'intensité correspondant à la simulation de **structures infinies** telles que les chenaux. La localisation du point d'implantation est délicate, puisque les objets peuvent avoir leur point d'implantation très loin du volume de simulation et pourtant le traverser. De plus, la signification du point d'implantation ou d'origine d'un objet infini n'est pas définie. De même, on peut se demander si la probabilité associée aux objets de dimension infinie a un sens, puisque un seul objet a une influence sur tout le volume de travail.

Dans le cas des chenaux sub-parallèles nous avons contourné le problème à l'aide des hypothèses de stationnarité, mais dans le cas des chenaux non parallèles cette simplification n'est plus possible. Une démarche hybride de différents modèles peut être plus adaptée à ce problème.

En pratique, notre approche sera appliquée à la simulation de **deux faciès ou plus** si chaque faciès est étudié indépendamment. Cependant, ceci pose des problèmes théoriques puisque les proportions des faciès ne sont pas indépendantes. De plus, les faciès sont spatialement intercorrélés. Ainsi, par exemple, les crevasses se trouvent au voisinage des chenaux. L'introduction d'une hiérarchie lors des simulations des différentes familles d'objets pourrait résoudre le problème et un processus de simulation basé sur cette approche est en cours de développement au Centre de Géostatistique.

D'autres applications ?

Le modèle booléen a été largement utilisé dans la simulation de réseaux de fractures (Chilès et Delfiner, 1999). Celles-ci sont représentées par des disques ou des rectangles 2D dans le cas d'une simulation à 3D et leur point d'implantation suit un processus de Poisson éventuellement non stationnaire. L'application du processus d'inférence proposé ici ne pose pas de problèmes théoriques, sauf dans la définition des ces objets en terme de leur probabilité associée, qui va dépendre de leur taille (longueur, hauteur), de leur orientation (azimut, pendage) et de leur ouverture. Le nombre de paramètres impliqués complique significativement le calcul de p_0 . D'autre part, il est aussi nécessaire de disposer d'une matrice de proportion des failles, de façon analogue à la matrice de proportion des faciès, et il faut déterminer ce que cela représente.

Cette approche peut également être utilisée dans la construction d'images d'entraînement des réservoirs pour des études ultérieures de statistiques multipoints (Strebelle et al., 2002). Cela contribuerait à créer des images initiales assez réalistes, notamment en introduisant la distribution des structures sédimentaires.

Bien que nous ayons centré notre étude sur la simulation des réservoirs pétroliers, cette approche peut être utilisée dans d'autres domaines d'application, comme en hydrogéologie, où le problème de modélisation est essentiellement le même, ou en études des matériaux, pour créer des simulations permettant d'analyser leurs propriétés.

Annexes

Annexe A

Calcul de la probabilité associée à une demi-ellipse

Dans cette section nous explicitons les calculs pour l'obtention de la probabilité associée à une ellipse $E(a, b)$ dont les axes sont tirés aléatoirement et de façon indépendante selon des lois uniformes : $a \sim U_{[a_0, a_1]}$ et $b \sim U_{[b_0, b_1]}$. Pour les axes positifs, nous analysons les différents cas possibles (fig. A.1). Par symétrie, les calculs sont identiques pour les axes négatifs.

Cas 1 : zone 6

Si $\frac{x^2}{a_1^2} + \frac{z^2}{b_1^2} > 1$, aucune demi-ellipse ne recouvre M et :

$$p_0(x, z) = 0$$

Cas 2 : zone 1

Si $\frac{x^2}{a_0^2} + \frac{z^2}{b_0^2} \leq 1$, toutes les demi-ellipses recouvrent M , alors :

$$p_0(x, z) = 1$$

Les cas 1 et 2 sont immédiats. Pour les cas restants nous utilisons l'expression générale (voir section 4.2.1) :

$$p_0(x, z) = \int_a \int_b 1 \cdot \frac{1}{a_1 - a_0} \cdot \frac{1}{b_1 - b_0} da db$$

où il faudra bien établir les limites de chaque cas particulier. À part les restrictions concernant chaque situation, les cas 3 à 6 doivent aussi vérifier que :

$$\frac{x^2}{a_1^2} + \frac{z^2}{b_1^2} \leq 1 \quad \text{et} \quad \frac{x^2}{a_0^2} + \frac{z^2}{b_0^2} > 1$$

Cas 3 : zone 2

Dans cette zone :

$$\frac{x^2}{a_1^2} + \frac{z^2}{b_0^2} \leq 1 \quad \text{et} \quad \frac{x^2}{a_0^2} + \frac{z^2}{b_1^2} \leq 1$$

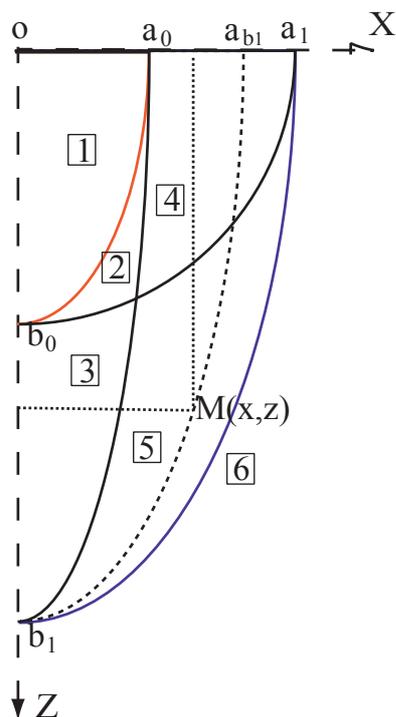


FIG. A.1 – Différentes régions à étudier selon la position du point $M(x, z)$. Les régions sont délimitées par les courbes en trait continu : la région 1 est délimitée par $E(a_0, b_0)$; la région 2 par $E(a_0, b_0)$, $E(a_1, b_0)$ et $E(a_0, b_1)$; les régions 3 et 4 par $E(a_1, b_0)$ et $E(a_0, b_1)$; la région 5 par $E(a_1, b_0)$, $E(a_0, b_1)$ et $E(a_1, b_1)$ et la région 6 correspond à tout l'espace extérieur à $E(a_1, b_1)$.

Si $a_0 \leq a < a_{b_0}$ (figure A.2) les ellipses dont l'axe $b < b_1$ interviennent dans le calcul. Si $a_{b_0} \leq a$ toutes les ellipses comptent. Alors, la valeur de p_0 pour les points se trouvant dans cette zone est donnée par :

$$p_0(x, z) = \int_{a_0}^{a_{b_0}} \int_{\frac{az}{\sqrt{a^2-x^2}}}^{b_1} \frac{1}{b_1-b_0} \cdot \frac{1}{a_1-a_0} db da + \int_{a_{b_0}}^{a_1} \int_{b_0}^{b_1} \frac{1}{b_1-b_0} \cdot \frac{1}{a_1-a_0} db da$$

avec $a_{b_0} = \frac{b_0}{\sqrt{b_0^2-z^2}}$ et $b = \frac{az}{\sqrt{a^2-x^2}}$. Cette intégrale donne comme résultat :

$$p_0(x, z) = \frac{a_1 b_1 - a_1 b_0 - a_0 b_1 + z \sqrt{a_0^2 - x^2} + x \sqrt{b_0^2 - z^2}}{(a_1 - a_0)(b_1 - b_0)}$$

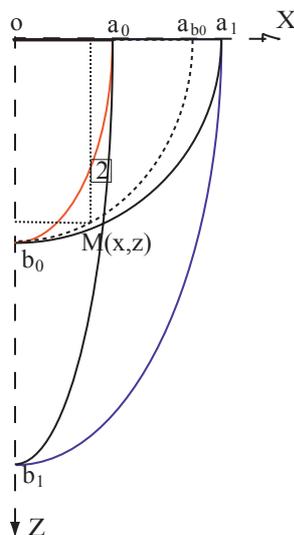


FIG. A.2 – Zone 2 : limitée par les ellipses $E(a_0, b_0)$, $E(a_0, b_1)$ et $E(a_1, b_0)$

Cas 4 : zone 3

Ici :

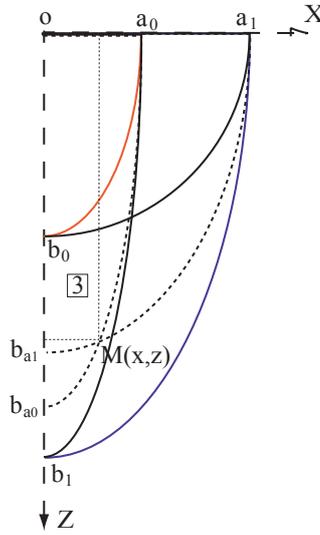
$$\frac{x^2}{a_1^2} + \frac{z^2}{b_0^2} > 1 \quad \text{et} \quad \frac{x^2}{a_0^2} + \frac{z^2}{b_1^2} \leq 1$$

Si $b < b_{a_1}$ (figure A.3) aucune ellipse ne recouvre le point M . Si $b_{a_1} \leq b \leq b_{a_0}$, alors toutes les ellipses dont le demi-axe a est compris entre a_0 et a_1 interviennent dans l'intégration :

$$p_0(x, z) = \int_{a_0}^{a_1} \int_{\frac{az}{\sqrt{a^2-x^2}}}^{b_1} \frac{1}{b_1-b_0} \cdot \frac{1}{a_1-a_0} db da$$

avec $b_{a_1} = \frac{a_1 z}{\sqrt{a_1^2-x^2}}$. Le résultat est :

$$p_0(x, z) = \frac{b_1(a_1 - a_0) - z(\sqrt{a_1^2 - x^2} - \sqrt{a_0^2 - x^2})}{(a_1 - a_0)(b_1 - b_0)}$$

FIG. A.3 – Zone 3 : limitée par les ellipses $E(a_0, b_1)$ et $E(a_1, b_0)$ **Cas 5 : zone 4**

Dans ce cas :

$$\frac{x^2}{a_1^2} + \frac{z^2}{b_0^2} \leq 1 \quad \text{et} \quad \frac{x^2}{a_0^2} + \frac{z^2}{b_1^2} > 1$$

Par symétrie (figure A.4) avec le cas précédent :

$$p_0(x, z) = \frac{a_1(b_1 - b_0) - x(\sqrt{b_1^2 - z^2} - \sqrt{b_0^2 - z^2})}{(a_1 - a_0)(b_1 - b_0)}$$

Cas 6 : zone 5

Pour finir, dans ce cas :

$$\frac{x^2}{a_1^2} + \frac{z^2}{b_0^2} > 1 \quad \text{et} \quad \frac{x^2}{a_0^2} + \frac{z^2}{b_1^2} > 1$$

Si $a < a_{b_1}$ (fig. A.5), aucune ellipse ne contient le point M . Pour $a_{b_1} < a \leq a_1$, il existe un demi-axe $b = \frac{az}{\sqrt{a^2 - x^2}}$ tel que toute ellipse d'axes a et $b < b_1$ contient M . L'intégrale à résoudre est, donc :

$$p_0(x, z) = \int_{a_{b_1}}^{a_1} \int_{\frac{az}{\sqrt{a^2 - x^2}}}^{b_1} \frac{1}{b_1 - b_0} \cdot \frac{1}{a_1 - a_0} db da$$

avec $a_{b_1} = \frac{b_1 x}{\sqrt{b_1^2 - z^2}}$. Après intégration, la probabilité du grain primaire dans cette région est :

$$p_0(x, z) = \frac{a_1 b_1 - z\sqrt{a_1^2 - x^2} - x\sqrt{b_1^2 - z^2}}{(a_1 - a_0)(b_1 - b_0)}$$

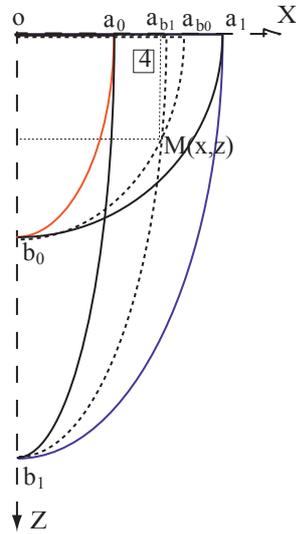


FIG. A.4 – Zone 4 : limitée par $E(a_0, b_1)$ et $E(a_1, b_0)$

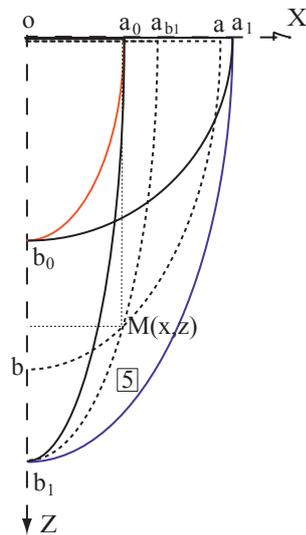


FIG. A.5 – Zone 5 : limitée par $E(a_0, b_1)$, $E(a_1, b_0)$ et $E(a_1, b_1)$

Annexe B

Convolution et corrélation

Dans cette section on fera un bref rappel à la théorie du signal en introduisant les concepts de convolution, corrélation et leurs propriétés. Les références principales utilisées sont Albiol et al., Papoulis (1962), Réfrégier (1993) et Smith (1999).

B.1 Convolution

La convolution est une opération mathématique qui utilise deux signaux et en renvoie un troisième. Formellement, l'opération de convolution de deux fonctions $f(x)$ et $g(x)$ est définie par :

$$\begin{aligned}h(x) &= \int_{-\infty}^{+\infty} f(x)g(x-y)dy \\ &= (f * g)(x) \\ &= f(x) * g(x)\end{aligned}\tag{B.1}$$

et elle vérifie les propriétés de commutativité, associativité et distributivité :

commutativité

$$f(x) * g(x) = g(x) * f(x)$$

associativité

$$f(x) * (g(x) * h(x)) = (f(x) * g(x)) * h(x)$$

distributivité

$$(f * g)(x) + (f * h)(x) = f(x) * (g(x) + h(x))$$

Par définition, la convolution est un système linéaire : elle vérifie les conditions d'**homogénéité** et d'**additivité**. La première signifie qu'un changement dans l'amplitude du signal d'entrée se traduit par un changement similaire dans l'amplitude du signal de sortie. La seconde condition, l'additivité, représente le fait que la somme de signaux à l'entrée du système donne comme résultat la somme de signaux à la sortie.

Un théorème important est le **théorème de convolution** : *La transformée de Fourier de la convolution de deux fonctions $f(x)$ et $g(x)$ s'exprime comme le produit des transformées de Fourier (TF) de ces deux fonctions, $\hat{f}(\nu)$ et $\hat{g}(\nu)$:*

$$h(x) = f(x) * g(x) \xrightarrow{TF} \hat{h}(\nu) = \hat{f}(\nu)\hat{g}(\nu)\tag{B.2}$$

Ces définitions sont valables pour les fonctions ou les signaux continus et définis dans tout l'espace, entre $-\infty$ et $+\infty$. Cependant, dans la réalité les signaux disponibles sont finis et constitués par des valeurs discrètes. Dans un cas réel, l'intégrale de l'expression (B.1) devient :

$$h[i] = \sum_{j=1}^M f[i-j]g[j] \quad (\text{B.3})$$

où $f[n]$ est un signal de N points, $g[n]$ un signal de M points (M et N indépendants) et i indique le point de sortie calculé. Cette expression implique que chaque point de la convolution est calculé indépendamment des autres.

Dans cette opération chaque point du signal de sortie est obtenu à partir de plusieurs points du signal d'entrée. D'une autre façon, chaque point de l'entrée contribue à plusieurs points de la sortie et, inversement, chaque point de la sortie reçoit de l'information de plusieurs points de l'entrée. Il est donc, possible de considérer chaque point résultant de la convolution comme une valeur moyenne pondérée sur un certain nombre de points du signal d'entrée, les poids de pondération et le nombre de points intervenants étant donnés par le noyau de déconvolution.

Dans notre application, la probabilité pour qu'un point appartienne à un faciès est donnée par le résultat de la convolution et est dépendante de l'intensité de Poisson dans ce point et les points voisins, dont l'influence sera déterminée par le type de grain primaire introduit.

Différence de longueurs D'un point de vue pratique il faut tenir compte de la différence de longueurs entre le signal d'entrée (N), la réponse du système (M) et le signal de sortie (L) : $L = N + M - 1$. Il faudra en tenir compte pour la réalisation du programme.

Effets de bord Le fait d'utiliser des signaux finis et discrets entraîne une information incomplète quant aux valeurs des points des bords du résultat de convolution : selon l'expression (B.3), si la longueur de la réponse du système est M , les $M - 1$ premières valeurs et les $M - 1$ dernières valeurs du signal de sortie, de longueur $N + M - 1$, contiennent moins d'information que les points au milieu et ne sont pas utilisables. Une des techniques pour éviter cela est de compléter le signal d'entrée en ajoutant $M - 1$ zéros à gauche du premier point et à droite du dernier.

B.2 Corrélation

Une autre opération intéressante est la corrélation (ou intercorrélation) entre deux signaux. Mathématiquement, la fonction de **corrélation** de deux fonctions $f(x)$, $g(x)$ est définie comme :

$$C_{f,g}(x) = \int_{-\infty}^{+\infty} f^*(y)g(x+y)dy \quad (\text{B.4})$$

où f^* représente la fonction complexe conjuguée de f .

Si $f(x) = g(x)$ la fonction résultante est appelée **autocorrélation**. Physiquement, la corrélation représente la ressemblance entre deux signaux et l'autocorrélation représente la ressemblance d'un signal en deux points différents. D'une autre façon, la corrélation fait référence à la dépendance entre deux signaux en deux points et l'autocorrélation fait référence à la dépendance d'un même processus entre deux points.

La différence entre corrélation et convolution est que dans celle-ci, une des deux fonctions est inversée avant le calcul.

La fonction de corrélation vérifie la propriété suivante :

$$C_{f,g}(-x) = [C_{g,f}(x)]^*$$

Il est nécessaire, pour les calculs ultérieurs, de définir la **densité spectrale d'interaction** entre deux signaux $f(x)$ et $g(x)$, qui est donnée à partir de leurs transformées de Fourier $\hat{f}(\nu)$ et $\hat{g}(\nu)$:

$$S_{f,g}(\nu) = \hat{f}^*(\nu)\hat{g}(\nu) \quad (\text{B.5})$$

d'où on peut en déduire que la densité spectrale d'interaction des fonctions $f(x)$ et $g(x)$ est la transformée de Fourier de leur fonction d'intercorrélation $C_{f,g}$.

$$C_{f,g}(x) \stackrel{TF}{\rightleftharpoons} S_{f,g}(\nu)$$

De même, la densité spectrale $S_{f,f}$ du signal $f(x)$ est la transformée de Fourier de sa fonction d'auto-corrélation.

$$C_{f,f}(x) \stackrel{TF}{\rightleftharpoons} S_{f,f}(\nu)$$

Dans cette section les opérateurs ou transformées de Fourier ont été introduits. Dans l'annexe suivant des concepts très généraux de la théorie de Fourier sont introduits pour mieux comprendre ces opérations.

Annexe C

Rappels de la théorie de Fourier

Dans cette section la théorie et les transformées de Fourier seront schématiquement présentées. Il a été montré auparavant comment déduire la fonction d'intensité du processus de Poisson à partir uniquement des proportions du modèle par un procédé de déconvolution. Ceci n'est pas tout à fait évident dans l'espace réel des données mais, par contre, en appliquant le théorème de convolution (B.2), cette opération devient immédiate. Pour cela il faut travailler dans le domaine de Fourier. Puisque les données sont discrètes, la méthode à utiliser est la Transformée Discrète de Fourier (DFT) mais, pour rendre compréhensible cette notion, il faut introduire la théorie de Fourier.

Les références utilisées sont Albiol et al., Papoulis (1962), Réfrégier (1993) et Smith (1999).

C.1 La transformée de Fourier (TF)

Cette transformée décompose ou sépare une fonction ou un signal en une infinité de sinusoides de fréquences différentes dont la somme est la fonction originale. Il existe plusieurs définitions de l'intégrale de Fourier, celle que l'on va utiliser est :

$$\hat{f}(\nu) = \int_{-\infty}^{+\infty} f(x)e^{-2\pi i\nu x} dx \quad (\text{C.1})$$

et la transformée inverse :

$$f(x) = \int_{-\infty}^{+\infty} \hat{f}(\nu)e^{+2\pi i x \nu} d\nu \quad (\text{C.2})$$

Il y a des fonctions dont la TF n'existe pas ; cependant, ce n'est pas le cas pour la plupart des fonctions à signification physique.

C.1.1 Propriétés de la TF

Quelques théorèmes simples qui caractérisent la transformée de Fourier sont listés ici.

1. Linéarité. Étant donnés

$$\begin{aligned} \hat{f}_1(\nu) &\Rightarrow f_1(x) \\ \hat{f}_2(\nu) &\Rightarrow f_2(x) \end{aligned}$$

et a_1, a_2 , deux constantes arbitraires,

$$a_1 f_1(x) + a_2 f_2(x) \Leftrightarrow a_1 \hat{f}_1(\nu) + a_2 \hat{f}_2(\nu)$$

2.Symétrie. Étant donné

$$\hat{f}(\nu) \Leftrightarrow f(x)$$

alors,

$$\hat{f}(x) \Leftrightarrow f(\nu)$$

3.Similitude. Si a est une constante réelle ou complexe,

$$f(ax) \Leftrightarrow \frac{1}{|a|} \hat{f}\left(\frac{\nu}{a}\right)$$

$$\frac{1}{|a|} f\left(\frac{x}{a}\right) \Leftrightarrow \hat{f}(a\nu)$$

4.Translation. Si la fonction $f(x)$ est traduite d'une constante x_0 , sa transformée de Fourier reste la même sauf pour un terme linéaire qui s'ajoute à la phase :

$$f(x - x_0) \Leftrightarrow \hat{f}(\nu) e^{-2\pi i x_0 \nu}$$

Dans le domaine des fréquences :

$$f(x) e^{-2\pi i \nu_0 x} \Leftrightarrow \hat{f}(\nu - \nu_0)$$

5.Différentiation. La relation entre la n -ième dérivée de la fonction et sa TF est

$$\frac{d^n f}{dx^n} \Leftrightarrow (i\nu)^n \hat{f}(\nu)$$

$$(-ix)^n f(x) \Leftrightarrow \frac{d^n \hat{f}(\nu)}{d\nu^n}$$

6.Fonction conjuguée. En général, il est possible de décomposer une fonction complexe en

$$f(x) = f_1(x) + i f_2(x)$$

La transformée de Fourier de la conjuguée de $f(x)$ est alors

$$f^*(x) = f_1(x) - i f_2(x) \Leftrightarrow \hat{f}^*(-\nu)$$

C.1.2 Théorème de convolution

La transformée de Fourier $\hat{f}(\nu)$ de la convolution de deux fonctions $f_1(x)$ et $f_2(x)$ est égale au produit des transformées de Fourier $\hat{f}_1(\nu)$, $\hat{f}_2(\nu)$ de ces fonctions :

$$f(x) = f_1(x) * f_2(x) \Leftrightarrow \hat{f}(\nu) = \hat{f}_1(\nu) \hat{f}_2(\nu)$$

La **densité spectrale d'interaction** de deux fonctions est la transformée de Fourier de leur corrélation :

$$C_{f_1 f_2}(x) \Leftrightarrow S_{f_1 f_2}(\nu) = \hat{f}_1^*(\nu) \hat{f}_2(\nu)$$

La **densité spectrale** d'une fonction $f(x)$ est la transformée de Fourier de sa fonction d'autocorrélation :

$$C_{ff}(x) \Leftrightarrow S_{ff}(\nu) = \hat{f}^*(\nu) \hat{f}(\nu) = |\hat{f}(\nu)|^2$$

C.1.3 Formule de Parseval

Produit scalaire : Le produit scalaire de deux fonctions $f_1(x)$ et $f_2(x)$ est l'intégrale, si elle existe :

$$f_1^\dagger(x) \cdot f_2(x) = \int_{-\infty}^{+\infty} f_1^*(x) f_2(x) dx \quad (\text{C.3})$$

avec \dagger le transposé complexe conjugué.

Si les deux fonctions $f_1(x)$, $f_2(x)$ sont identiques, l'intégrale (C.3) s'appelle la **norme** de la fonction : $\|f(x)\|^2$.

Théorème de Parseval : Le produit scalaire de deux fonctions $f_1(x)$, $f_2(x)$ est égal au produit de leurs transformées de Fourier

$$\int_{-\infty}^{+\infty} f_1^*(x) f_2(x) dx = \int_{-\infty}^{+\infty} \hat{f}_1^*(\nu) \hat{f}_2(\nu) d\nu \quad (\text{C.4})$$

Si les deux fonctions sont identiques, le théorème dit que la norme de la fonction est la même calculée dans l'espace de la fonction ou calculée dans l'espace de Fourier : la transformée de Fourier conserve la norme de la fonction.

C.1.4 Échantillonnage

Une fonction seuillée dans le domaine de Fourier est une fonction qui n'a pas de composantes spectrales au-delà d'une certaine fréquence B : $\hat{f}(\nu) = 0$ si $|\nu| > 2\pi B$. Le **théorème d'échantillonnage** établit qu'une fonction réelle limitée dans l'espace des fréquences par la fréquence maximale B peut être reconstruite sans erreur à partir des échantillons pris uniformément avec une fréquence d'échantillonnage supérieure au double de la fréquence maximale du signal. Cette fréquence minimale d'échantillonnage, $\nu N = 2B$ Hz, est la fréquence de Nyquist et l'intervalle d'échantillonnage correspondant dans le domaine du signal est $T = 1/2B$. Le fait d'échantillonner génère quelques problèmes ou artefacts, notamment "aliasing" et le phénomène de Gibbs.

"Aliasing" (recouvrement de spectre) Le théorème précédent assume que la fonction est limitée dans l'espace de Fourier. Dans la pratique, cependant, les signaux sont limités dans leur propre domaine puisque les mesures disponibles sont discrètes et forcément finies. Pour cette raison il est difficile de choisir une fréquence d'échantillonnage adéquate qui ne fasse pas perdre d'information. Quand un signal est sous-échantillonné, c'est à dire, échantillonné sur une intervalle d'échantillonnage $T > 1/2B$, son spectre (transformée de Fourier) présente des superpositions de fréquences : $\hat{f}(\nu)$ n'apporte pas une information complète et il n'est plus possible de récupérer $f(x)$ à partir des échantillons. Dans ce cas il apparaît quelques basses fréquences dans le spectre qui perturbent la reconstruction du signal. Théoriquement il est impossible d'échantillonner une fonction sans aliasing. L'erreur dans la reconstruction du signal sera inversement proportionnelle au nombre d'échantillons.

Phénomène de Gibbs. Un signal fini $f_R(x)$ peut être interprété comme la restriction d'une fonction infinie $f(x)$ sur une fenêtre $[-T/2, T/2[$, où T est la durée du signal. Si $f_R(-T/2)$ est différent de $f_R(T/2)$, (discontinuité parasite), certaines hautes fréquences parasites perturberont la récupération du signal. Une façon de réduire cet effet est de multiplier le signal échantillonné initial par une fenêtre de valeurs nulles en $-T/2$ et en $T/2$, par exemple le pulse rectangulaire (fenêtre de Dirichlet).

Convolution circulaire. Le fait de calculer une corrélation ou une convolution par transformées discrètes de Fourier implique un échantillonnage dans le domaine de Fourier. Cela se traduit par une périodicité du signal. Cet effet peut être diminué en ajoutant des zéros à droite et à gauche du signal initial.

C.2 Transformée discrète de Fourier (DFT)

En pratique, seules les valeurs discrètes, f_k , sont disponibles, et la transformée $\hat{f}(\nu)$ ne peut être calculée que pour des points discrets : des échantillons de la transformée, \hat{f}_r . Si $f(kT)$ et $\hat{f}(r\nu_0)$ sont les points numéro k et r de $f(x)$ et $\hat{f}(\nu)$ respectivement, et N_0 est le nombre d'échantillons de la fonction dans une période T_0 , alors

$$\begin{aligned} f_k &= T f(kT) = T_0 N_0^{-1} f(kT) \\ \hat{f}_r &= \hat{f}(r\nu_0) \end{aligned} \quad (\text{C.5})$$

où $\nu_0 = 2\pi T_0^{-1}$

La DFT est définie comme :

$$\hat{f}_r = \sum_{k=0}^{N_0-1} f_k \exp(-i \frac{2\pi}{N_0} kr) \quad (\text{C.6})$$

et son inverse

$$f_k = \frac{1}{N_0} \sum_{r=0}^{N_0-1} \hat{f}_r \exp(i \frac{2\pi}{N_0} rk) \quad (\text{C.7})$$

La propriété de linéarité des systèmes convolutifs permet de les résoudre en décomposant un signal compliqué (le résultat de la convolution, par exemple) dans plusieurs signaux plus simples. Une des décompositions les plus importantes est la décomposition ou transformée de Fourier. D'une façon simple il s'agit, dans le cas de signaux discrets, de :

- séparer un signal de N points en deux signaux de $N/2 + 1$ points contenant les amplitudes des composantes sinus et cosinus ;
- faire passer ces composantes par le système et
- regrouper les sorties en un seul signal.

où N est le nombre de points ou échantillons dans le domaine initial. Il est normalement choisi comme une puissance de 2 (les ordinateurs travaillent en code binaire et l'algorithme le plus utilisé pour le calcul des DFT est la FFT (Fast Fourier Transform), transformée rapide de Fourier, qui travaille en puissances de 2). La DFT considère les N points de départ comme appartenant à une seule période d'un signal infiniment long. Cela implique que les extrêmes du signal sont connectés par une période. La conséquence de cela est le phénomène d'aliasing.

C.2.1 Transformée rapide de Fourier (FFT)

La FFT est un algorithme développé par Tukey et Cooley en 1965 qui diminue le nombre de calculs pour la transformée de Fourier de N^2 à $N \log_2(N)$. Cet algorithme est simplifié si N est une puissance de 2, mais ce n'est pas indispensable. Il est basé sur les DFT complexes, une méthode un peu plus sophistiquée que celle expliquée auparavant dans laquelle les données sont représentées par de nombres complexes. Voir Smith (1999) pour plus d'information.

Annexe D

Développement du calcul pour l'obtention du filtre de Wiener

Soit

$$s(x) = -\ln(1 - p(x)) = [\theta * p_0](x) + n(x)$$

l'équation à résoudre (section 5.1.1). Une estimation de l'intensité θ cherchée est :

$$\tilde{\theta}(x) = [f * s](x) = \underbrace{[f * \theta * p_0](x)}_{a(x)} + \underbrace{[f * n](x)}_{b(x)}$$

Le filtre de Wiener est une fonction f telle que l'écart quadratique moyen entre l'intensité réelle θ et son estimateur $\tilde{\theta}$ soit minimal :

$$EQM = E[(\theta(x) - \tilde{\theta}(x))^2] \quad \text{minimum}$$

Désormais, par un souci de simplification, nous allons supprimer dans les expressions les dépendances de x ou ν . Ces dépendances seront sous-entendues.

$$\begin{aligned} EQM &= E[(\theta - \tilde{\theta})^2] = E[(\theta - \tilde{\theta})^* (\theta - \tilde{\theta})] = \\ &= E[(\theta - a - b)^* (\theta - a - b)] = \\ &= E[\theta^* \theta - \theta^* a - \theta^* b - a^* \theta + a^* a + a^* b - b^* \theta + b^* a + b^* b] \end{aligned}$$

Chaque terme de cette expression sera analysé séparément en tenant compte du *théorème de Parseval* (voir Annexe C), selon lequel le produit scalaire de deux fonctions est égale au produit scalaire de leurs transformées de Fourier. Le bruit, n et l'intensité, θ , sont considérés comme des variables aléatoires.

$$\bullet E[\theta^* \theta] = \int_{-\infty}^{\infty} \theta^* \theta dx = C_{\theta\theta}(0) \stackrel{\text{Th. Parseval}}{=} \int_{-\infty}^{\infty} \hat{\theta}^* \hat{\theta} d\nu = S_{\theta\theta}(0)$$

Le terme $C_{\theta\theta}(0)$ représente l'autocorrélation de l'intensité, et $S_{\theta\theta}(0)$, sa densité spectrale. $S_{\theta\theta}(0)$ est la transformée de Fourier de l'autocorrélation.

$$\bullet E[\theta^* a] = \int_{-\infty}^{\infty} \theta^* a dx = \int_{-\infty}^{\infty} \hat{\theta}^* \hat{a} d\nu = \int_{-\infty}^{\infty} \hat{\theta}^* \hat{f} \hat{\theta} \hat{p}_0 d\nu = \int_{-\infty}^{\infty} \hat{f} \hat{p}_0 S_{\theta\theta} d\nu$$

$$\bullet E[\theta^*b] = \int_{-\infty}^{\infty} \theta^* b dx = \int_{-\infty}^{\infty} \hat{\theta}^* \hat{b} d\nu = \int_{-\infty}^{\infty} \hat{f} S_{\theta n} d\nu = 0$$

puisque l'intensité et le bruit sont non corrélés. Alors, $S_{\theta n} = 0, \forall \nu$. Pour la même raison les termes suivants sont aussi égales à zéro : $E[a^*b] = E[b^*\theta] = E[b^*a] = 0$

$$\bullet E[a^*\theta] = \int_{-\infty}^{\infty} \hat{f}^* \hat{p}_0^* \hat{\theta}^* \hat{\theta} d\nu = \int_{-\infty}^{\infty} \hat{f}^* \hat{p}_0^* S_{\theta\theta} d\nu$$

$$\bullet E[a^*a] = \int_{-\infty}^{\infty} \hat{f}^* \hat{p}_0^* \hat{\theta}^* \hat{f} \hat{p}_0 \hat{\theta} d\nu = \int_{-\infty}^{\infty} \hat{f}^* \hat{f} |\hat{p}_0|^2 S_{\theta\theta} d\nu$$

$$\bullet E[b^*b] = \int_{-\infty}^{\infty} \hat{f}^* \hat{n}^* \hat{f} \hat{n} d\nu = \int_{-\infty}^{\infty} \hat{f}^* \hat{f} S_{nn} d\nu$$

L'écart quadratique moyen peut donc s'exprimer comme :

$$EQM^2 = S_{\theta\theta}(0) + \int_{-\infty}^{\infty} \left[-\hat{f} \hat{p}_0 S_{\theta\theta} - 0 - \hat{f}^* \hat{p}_0^* S_{\theta\theta} + \hat{f}^* \hat{f} |\hat{p}_0|^2 S_{\theta\theta} + \hat{f}^* \hat{f} S_{nn} \right] d\nu$$

Puisque l'on cherche la fonction f qui minimise EQM , nous devons résoudre $\frac{\partial EQM^2}{\partial f} = 0$. Cette équation se vérifie, pour toute fréquence ν , si :

$$-\hat{p}_0 S_{\theta\theta} - (\hat{p}_0^* S_{\theta\theta})^* + \hat{f}^* |\hat{p}_0|^2 S_{\theta\theta} + (\hat{f} |\hat{p}_0|^2 S_{\theta\theta})^* + \hat{f}^* S_{nn} + (\hat{f} S_{nn})^* = 0$$

et, donc :

$$-2\hat{p}_0 S_{\theta\theta} + 2\hat{f}^* |\hat{p}_0|^2 S_{\theta\theta} + 2\hat{f}^* S_{nn} = 0$$

du fait que les autocorrélations $S_{\theta\theta}$ et S_{nn} sont réelles et paires. Le résultat cherché est :

$$\hat{f}^* = \frac{\hat{p}_0 S_{\theta\theta}}{|\hat{p}_0|^2 S_{\theta\theta} + S_{nn}} \implies \hat{f} = \frac{\hat{p}_0^*}{|\hat{p}_0|^2 + \frac{S_{nn}}{S_{\theta\theta}}}$$

L'estimation de l'intensité que l'on obtient est, dans le domaine des fréquences :

$$\hat{\theta}(\nu) = \hat{s}(\nu) \cdot \frac{\hat{p}_0^*(\nu)}{|\hat{p}_0(\nu)|^2 + \frac{S_{nn}(\nu)}{S_{\theta\theta}(\nu)}}$$

Annexe E

Description du programme

Dans cette section le programme écrit pour l'obtention de l'intensité de Poisson selon le processus de déconvolution est expliqué pas à pas. Il a été implémenté en Matlab.

1. *Introduction des paramètres du programme.* Les données initiales sont la proportion initiale, qui est associée à la proportion du modèle, p , la probabilité associée à l'objet, p_0 et la valeur du paramètre Γ . Les deux premières sont introduites comme les éléments des matrices (1D, 2D ou 3D) et le paramètre Γ est une valeur constante fixée a priori à 0.005, mais avec la possibilité d'être modifiée. La proportion initiale subit deux transformations : une première interpolation de façon à égaliser sa maille avec la maille de la matrice de p_0 et une extrapolation dans les bords qui tient compte de la définition du domaine de calcul, comme il a été expliqué dans la section 5.2.3. Puis nous transformons la variable p pour obtenir une dépendance plus simple entre les proportions et la convolution (voir 5.2) :

$$s(x) = -\ln(1 - p(x)) = [\theta * p_0](x)$$

2. *Calcul des transformées de Fourier.* Les transformées discrètes de Fourier des variables impliquées dans le processus (p_0, s) sont calculées à travers un algorithme de transformées rapides de Fourier (FFT). La matrice de p_0 est complétée par des zéros jusqu'à la taille de la matrice de s , de façon à pouvoir réaliser les produits des matrices point par point.
3. *Calcul du filtre.* Dans cette partie le filtre donné par l'expression (5.8) est implémenté et l'intensité de Poisson calculée. Nous calculons deux intensités : l'intensité mathématique, c'est à dire celle que l'on obtient directement par le processus de déconvolution, et l'intensité corrigée des valeurs négatives. Cette intensité corrigée sera utilisée comme l'intensité de Poisson dans le modèle booléen. La correction utilisée est calculer une moyenne glissante, éventuellement pondérée, sur une fenêtre de taille égale à la taille du grain. Une moyenne glissante n'est qu'un processus de convolution, avec un noyau de convolution égale à la normalisation de p_0 : l'intégrale du noyau sur son support est égale à l'unité.
4. *Calcul du nombre d'objets.* Une fois l'intensité obtenue nous calculons le nombre moyen d'objets à simuler. Si nous avons utilisé une expression de p_0 correspondant à un objet différent à celui de la simulation, il faut tenir compte du rapport entre les volumes des objets. De même, il faut considérer l'inférence à 3D (voir section 5.2.2) s'il y a lieu.
5. *Processus de validation.* Le dernier pas consiste à vérifier si l'intensité obtenue est acceptable. Pour cela on calcule la proportion du modèle par convolution entre l'intensité calculée et la loi de l'objet puis on compare le résultat avec les données de départ.

Bibliographie

Références

- Albiol, A., Naranjo, V., Prades, J. *Tratamiento digital de la señal. Teoría y aplicaciones*. Departamento de comunicaciones. Universidad Politécnica de Valencia. Notes de cours.
- Ayala Gallego, G. (1988). *Inferencia en modelos booleanos*. Thèse, Departamento de Estadística e Investigación Operativa. Universitat de Valencia.
- Beucher H., Galli A., Le Loc'h G., Ravenne, Ch. & Heresim Group (1993). Including a regional trend in reservoir modelling using the truncated Gaussian method. A. Soares (Ed.), *Geostatistics Tróia'92*, Volume 2, p. 555–566. Kluwer.
- Beucher H., Geffroy, F., Doligez, B. (1997). Introducing more geology in reservoir stochastic modelling using truncated gaussian approach. *Memorias del I Congreso de Sedimentología*, Volume I, p. 89–95. Sociedad Venezolana de Geólogos.
- Cacas, M. C., Daniel, J. M., Letouzey, J. (2001). Nested geological modelling of naturally fractured reservoirs. *Petroleum Geoscience* 7, S43–S52.
- Castaing, C., Genter, A., Bourguine, B., Chilès, J. P., Wendling, J., Siegel, P. (2002). Taking into account the complexity of natural fracture systems in reservoir single-phase flow modelling. *Journal of Hydrology* 266, 83–98.
- Chessa, A. G. (1995). *Conditional Simulation of Spatial Stochastic Models for Reservoir Heterogeneity*. Thèse, Delft University of Technology.
- Chessa, A. G., Martinius, A. W. (1992). Object based modelling of the spatial distribution of fluvial sandstone deposits. M. A. C. et al. (Ed.), *Proceedings of the 3rd European Conference on the Mathematics of Oil Recovery*, p. 5–14. Delft University Press.
- Chilès, J. P., Delfiner, P. (1999). *Geostatistics. Modeling Spatial Uncertainty* : Wiley. Wiley series in probability and statistics.
- Clément, I. (1999). Modélisation géostatistique de la série continentale du Permien d'Utah : application de la méthode gaussienne tronquée et du modèle booléen à l'aide du logiciel Heresim3D. Rapport Technique 45524, Institut Français du Pétrole.
- Daly, C. (1991). *Applications de la géostatistique à quelques problèmes de filtrage*. Thèse, Centre de Géostatistique et de Morphologie Mathématique, ENSMP.
- Damsleth, E., Holden, L. (1994). Mixed reservoir characterization methods. *Proceedings of the Centennial Petroleum Engineering Symposium, Tulsa*. Society of Petroleum Engineers. SPE 27969.
- Damsleth, E., Tjølsen, Ch. B., Omre, H., Haldorsen, H. H. (1992). A two-stage stochastic model applied to a north sea reservoir. *Journal of Petroleum Technology* (April). SPE 20605.
- Deutsch, C. V. (1999). Reservoir modeling with publicly available software. *Computers and Geoscience* 25, 355–363.
- Deutsch, C. V., Tran, T. T. (2002). Fluvsim : a program for object-based stochastic modeling of fluvial depositional system. *Computer & Geoscience* 28, 525–535.

- Deutsch, C. V., Wang, L. (1996). Hierarchical object-based modelling of fluvial reservoirs. *Journal of Mathematical Geology* 28(7), 857–880.
- Doligez, B., Granjeon, D., Joseph, Ph., Eschard, R., Beucher, H. (1999). How can stratigraphic modeling help constrain geostatistical reservoir simulations ? *SEPM Special Publications* 62, 239–244.
- Dubrulle, O. (1988). A review of stochastic models for petroleum reservoirs. *BSRG meeting on quantification of sediment body geometries and their internal heterogeneities, London*.
- Eschard, R., Lerat, O., Desaubliaux, G., Doligez, B., Granjeon, D., Lafont, F., Didier, B. (1997). Stratigraphie séquentielle haute-résolution dans les séries continentales du Permien de l'Utah. Rapport Technique 44030, Institut Français du Pétrole.
- Feller, W. (1968). *An Introduction to Probability Theory and its Applications* (3rd ed.) : Wiley. Wiley Series in Probability and Mathematical Statistics I.
- Fisher, Q. J., Knipe, R. J. (2001). The permeability of faults within siliciclastic petroleum reservoirs of the north sea and norwegian continental shelf. *Marine and Petroleum Geology* 18, 1063–1081.
- Foucault, A., Raoult, J. F. (1980-1984). *Dictionnaire de géologie* (seconde ed.) : Masson.
- Galli, A., Beucher, H. (1997). Stochastic models for reservoir characterization : a user-friendly review. *Proceedings of 5th Latin American and Caribbean Petroleum Engineering Conference and Exhibition, Rio de Janeiro*. Society of Petroleum Engineers. SPE 38999.
- Galli, A., Beucher, H., Le Loc'h, G., Doligez, B. (1994). The pros and cons of the truncated gaussian method. M. Armstrong et P. A. Dowd (Eds.), *Geostatistical simulations*, p. 217–233. Kluwer.
- Gedler, G. (1991). Algorithme de simulation d'un schéma booléen. Rapport Technique S-281, Centre de Géostatistique, ENSMP.
- Gelhar, L. W. (1993). *Stochastic Subsurface Hydrology* : Prentice Hall.
- Gonzalez, R. C., Woods, R. E. (1992). *Digital Image Processing* : Addison-Wesley.
- Grindeheim, A. O., Aasen, J. O. (1992). An evaluation of homogenization techniques for absolute permeability. *Proceedings of the Lerkendal Petroleum Engineering Workshop, Trondheim*.
- Haldorsen, H. H., Damsleth, E. (1990). Stochastic modeling. *Journal of Petroleum Technology* 42, 404–412.
- Haldorsen, H. H., MacDonald, C. J. (1987). Stochastic modeling of underground reservoir facies (smurf). *Proceedings of 62nd SPE Annual Technical Conference, Dallas*. Society of Petroleum Engineers. SPE 16751.
- Henriquez, A., Tyler, K. J., Hurst, A. (1990). Characterization of fluvial sedimentology for reservoir simulation modeling. *SPE Formation Evaluation*. Society of Petroleum Engineers. SPE 18323.
- Hu, L. Y. (2002). *Modélisation Stochastique des Réservoirs Hétérogènes*. Université Pierre et Marie Curie (Paris VI). Mémoire pour obtenir l'habilitation à diriger des recherches.
- Hu, L. Y. (2003). History matching of object-based stochastic reservoir models. *Proceedings of 13th SPE Middle East Oil Show & Conference, Bahrain*. Society of Petroleum Engineers. SPE 81503.
- Hu, L. Y., Joseph, P., Dubrulle, O. (1992). Random genetic simulation of the internal geometry of deltaic sandstone bodies. *Proceedings of 67th SPE Annual Technical Conference and Exhibition, Washington*. Society of Petroleum Engineers. SPE 24714.
- Jacod, J., Joathon, O. (1971). Use of random genetic models in the study of sedimentary process. *Mathematical Geology* 3, 265–279.
- Jeulin, D. (1999). *Modèles de structures aléatoires*. Centre de Morphologie Mathématique, ENSMP. CT-01/99/MM.
- Jeulin, D., Renard, D. (1992). Practical limits of the deconvolution of images by kriging. *Microscopy Microanalysis Microstructure* 3, 333–361.

- Johann, P., Fournier, F., Souza Jr., O., Doligez, B. (1997). 3-d stochastic reservoir modeling : a non stationary approach guided by seismic data. *Proceedings of 5th Latin American and Caribbean Petroleum Engineering Conference and Exhibition, Rio de Janeiro*. Society of Petroleum Engineers. SPE 38997.
- Journel, A. G. (1996). Conditional simulation of geologically averaged block permeabilities. *Journal of Hydrology* 183, 23–35.
- Journel, A. G., Alabert, F. G. (1990). New method for reservoir mapping. *Journal of Petroleum Technology* (February), 212–223. SPE 18324.
- Journel, A. G., Gunderso, R., Gringarten, E., Yao, T. (1998). Stochastic modelling of a fluvial reservoir : a comparative review of algorithms. *Journal of Petroleum Science & Engineering* 21, 95–121.
- Kendall, D. G. (1974). Foundations of a theory of random sets. E. Harding et D. Kendall (Eds.), *Stochastic Geometry*. Wiley.
- Lantuéjoul, Ch. (2002). *Geostatistical Simulation : models and algorithms* : Springer-Verlag.
- Lantuéjoul, Ch. (1993). *Les ensembles aléatoires*. Centre de Géostatistique, ENSMP. Cours C-149.
- Lantuéjoul, Ch., (1995). Conditional simulation of a boolean model. Rapport Technique N-1/95/G, Centre de Géostatistique. ENSMP.
- Lantuéjoul, Ch. (1997). Iterative algorithms for conditional simulations. E. Baafi et N. A. Schofield (Eds.), *Geostatistics Wollongong '96*, Volume 1, p. 27–40. Kluwer.
- Lantuéjoul, Ch. (1998). Conditional simulation of random sets. *Proceedings of the Section on Statistics and the Environment, annual meeting of the American Statistical Association*, p. 37–45.
- Lantuéjoul, Ch., Schmitt, M. (1991). Use of two new formulae to estimate the poisson intensity of a boolean model. *Proceedings Treizième Colloque Gretsi, Juan-les-Pins*, Volume 2, p. 1045–1048.
- Lantuéjoul, Ch., Simaku, A. (1994). Simulation conditionnelle d'un schéma booléen. Rapport Technique N-24/94/G, Centre de Géostatistique. ENSMP.
- Lanzarini, W. L., Poletto, C. A., Tavares, G., Pesco, S., Lopes, H. (1997). Stochastic modeling of geometric objects and reservoir heterogeneities. *Proceedings of 5th Latin American and Caribbean Petroleum Engineering Conference and Exhibition, Rio de Janeiro*. Society of Petroleum Engineers. SPE 38953.
- Lerat, O., Doligez, B., Clément, I., Desaubliaux, G., Eschard, R., Granjeon, D., Laffont, F., (1999). Geostatistical modelling of eolian and fluvial reservoirs (Cutler Fm., Permien Utah). EAGE Conference, Helsinki. Best poster.
- Lopez, S. (2003). *Modélisation des réservoirs chenalisés méandriques, Approche génétique et stochastique*. Thèse, École des Mines de Paris.
- Lopez, S., Galli, A., Cojan, I. (2001). Fluvial meandering channelized reservoir : a stochastic and process-based approach. *Proceedings of the IAMG Annual Conference, Cancun*. International Association for Mathematical Geology. CD-ROM, 16 p.
- Matheron, G. (1969a). Le processus d'Ambarzoumian et leur application en géologie. Rapport Technique N-131, Centre de Géostatistique, ENSMP.
- Matheron, G. (1969b). Structures aléatoires et géologie mathématique. Rapport Technique N-111, Centre de Géostatistique, ENSMP.
- Matheron, G. (1969c). Théorie des ensembles aléatoires. *Les cahiers du Centre de Morphologie Mathématique* 4. ENSMP.
- Matheron, G. (1972). Ensembles aléatoires et géométrie intégrale. *Les cahiers du Centre de Morphologie Mathématique* 6. ENSMP.

- Matheron, G., Beucher, H., De Fouquet, Ch., Galli, A., Guérillot, D., Ravenne, Ch. (1987). Conditional simulation of the geometry of fluvio-deltaic reservoirs. *Proceedings of the SPE Annual Technical Conference and Exhibition, Dallas*. Society of Petroleum Engineers. SPE 18234.
- Molchanov, I. (1997). *Statistics of the Boolean Model for Practitioners and Mathematicians* : Wiley. Wiley Series in Probability and Statistics.
- Moureau, M., Brace, G. (1979). *Dictionnaire technique du pétrole* (2ème ed.) : Technip.
- Munthe, K. L., Omre, H., Holden, L., Damsleth, E., Heffer, K., Olsen, T. S., Watterson, J. (1993). Sub-seismic faults in reservoir description and simulation. *Proceedings of 68th SPE Annual Technical Conference and Exhibition, Houston*. Society of Petroleum Engineers. SPE 26500.
- Papoulis, A. (1962). *The Fourier integral and its applications* : McGraw-Hill.
- Pratt, W. K. (1978). *Digital Image Processing* : Wiley.
- Ravenne, Ch., Beucher, H. (1988). Recent development in description of sedimentary bodies in a fluvio deltaic reservoir and their 3-D conditional simulation. *Proceedings of 63rd SPE Annual Technical Conference and Exhibition, Houston*. Society of Petroleum Engineers. SPE 18310.
- Ravenne, Ch., Galli, A., Doligez, B., Beucher, H., Eschard, R. (2000). Quantification of facies relationships via proportion curves. M. Armstrong, C. Bettini, N. Champigny, A. Galli, et A. Remacre (Eds.), *Geostatistics Rio 2000*, p. 19–39. Kluwer.
- Réfrégier, Ph. (1993). *Théorie du signal. Signal-Information-Fluctuations* : Masson.
- Réfrégier, Ph. (1994). *Filtrage et déconvolution. Le traitement numérique d'images : analyse d'images, fondements*. Collège de Polytechnique. Notes de cours.
- Rivoirard, J., Beucher, H. (1993). Validité du modèle booléen et tests sur simulations. Rapport technique, Centre de Géostatistique, ENSMP. N-28/93/G.
- Sallé, C., Debyser, J. (et al.) (1976). *Formation des gisements de pétrole* : Technip.
- Schmitt, M. (1991). Estimation of the density in a stationary boolean model. *Journal of Applied Probability* 28, 702–708.
- Schmitt, M. (1997). Estimation of intensity and shape in a non-stationary boolean model. D. Jeulin (Ed.), *Advances in Theory and Applications of Random Sets : Proceedings of the 1996 International Symposium in*, p. 251–267. World Scientific.
- Schmitt, M., Beucher, H. (1997). On the inference of the boolean model. E. Baafi et N. Schofield (Eds.), *Geostatistics Wollongong '96*, Volume 1, p. 200–210. Kluwer.
- Seguret, S. (1988). Pour une méthodologie de déconvolution de variogrammes. Rapport Technique N-51/88/G, Centre de Géostatistique, ENSMP.
- Serra, J. (1982). *Image Analysis and Mathematical Morphology* : Academic Press.
- Shmaryan, L. E., Deutsch, C. V. (1999). Object-based modeling of fluvial/deepwater reservoirs with fast data conditioning : methodology and case studies. *Proceedings of SPE Annual Technical Conference and Exhibition, Houston*. Society of Petroleum Engineers. SPE 56821.
- Smith, S. W. (1999). *The Scientist and Engineer's Guide to Digital Signal Processing* (2nd ed.) : California Technical Publishing.
- Stoyan, D., Kendall, W. S., Mecke, J. (1995). *Stochastic Geometry and its Applications* (2nd ed.) : Wiley.
- Strebelle, S., Payrazyan, K., Caers, J. (2002). Modeling a deepwater turbidite reservoir conditional to seismic data using multiple-point geostatistics. *Proceedings of the SPE Annual Technical Conference and Exhibition, San Antonio*. Society of Petroleum Engineers. SPE 77425.
- Tavares, G., Lopez, H., Pesco, S., Poletto, C. A. (2002). Petbool : a software for stochastic modeling of geological objects. *Proceedings of the 8th IAMG Annual Conference, Berlin*, Volume I, p. 203–208. International Association for Mathematical Geology.

- Viseur, S. (1999). Stochastic boolean simulation of fluvial deposits : a new approach combining accuracy with efficiency. *Proceedings of SPE Annual Technical Conference and Exhibition, Houston*. Society of Petroleum Engineers. SPE 56688.
- Viseur, S., Shtuka, A., Mallet, J. L. (1998). New fast, stochastic, boolean simulation of fluvial deposits. *Proceedings of SPE Annual Technical Conference and Exhibition, New Orleans*. Society of Petroleum Engineers. SPE 49281.
- Volpi, B., Galli, A., Ravenne, Ch. (1997). Vertical proportions curves : a qualitative and quantitative tool for reservoir characterization. *Memorias del I Congreso de Sedimentología*, Volume I, p. 351–357. Sociedad Venezolana de Geólogos.
- Weber, K. J. (1986). How heterogeneity affects oil recovery. L. Lake et H. C. Jr. (Eds.), *Reservoir Characterization*. Academic Press.
- Weil, W. (1988). Non-isotropic boolean models. *Journal of Microscopy* 151-3, 235–246.

Summary

Given the important economic stakes involved in the oil industry, there is a real need for a description of the geological structure of oil reservoirs and of the petrophysical properties of the constituting rocks. Reservoir modelling consists in making up numerical models representing the heterogeneities of the reservoir at different scales. The **Boolean model** is frequently used to design a lithological model of the reservoir internal architecture both at the scale of the genetic unit and at the granulometric scale inside the reservoir. However, Boolean models present a major problem when applied to the modelling of reservoirs having a **non stationary distribution** of their constituting lithofacies.

Lithofacies proportions give information about the spatial distribution of these lithofacies. They are an efficient tool for detecting a non stationarity. They also make it possible to quantify the geological information of the deposits. They are therefore an important constraint to respect in order to obtain simulations that are geologically realistic.

In this work, we have integrated into the Boolean model the information contained in the proportions. The problem to be solved is the **inference of the model parameters**. Two aspects have been considered. First, we have defined the inference problem from a theoretical point of view by establishing the relationship between experimental information and the model parameters. Secondly, we have implemented a computing tool to automatically pass from the proportion data to the model parameters.

To do that, we propose a method of inference based on a **deconvolution process**. This method enables us to introduce the information of the proportions in the Boolean model. As a result, simulations can reproduce a non stationary distribution of the heterogeneities in the sedimentary unit. By construction, these simulations then respect the lateral and vertical distribution of the proportions into the unit as well as the global mean value of the proportion over the studied volume. As this global value is related to petrophysical parameters, it is important to recover it in simulations.

Keywords :

Boolean model, facies, geostatistics, inference of parameters, non stationarity, object based process, proportions, reservoir, simulation, stochastic.

Résumé

Les enjeux économiques de l'industrie pétrolière imposent un important besoin de décrire la structure géologique des réservoirs d'hydrocarbures et des caractéristiques pétrophysiques des roches qui les constituent. La modélisation des réservoirs consiste à construire des modèles numériques qui représentent les hétérogénéités dans le réservoir à différentes échelles. Le **schéma booléen** est un modèle très utilisé pour établir un modèle lithologique de l'architecture interne du réservoir, que ce soit à l'échelle d'une unité génétique ou à l'échelle granulométrique à l'intérieur de ce dernier. Ce modèle présente néanmoins un problème important lors de son application à la modélisation de réservoirs présentant une distribution **non stationnaire** des lithofaciès qui les composent.

Les **proportions** de ces lithofaciès fournissent une information sur leur distribution spatiale et constituent un outil très efficace pour déterminer la présence d'une non stationnarité. Elles permettent de quantifier l'information géologique des dépôts et constituent ainsi une importante contrainte à respecter pour obtenir des simulations géologiquement réalistes.

Dans cette thèse, nous avons intégré dans le modèle booléen cette information contenue dans les proportions. Le problème abordé est l'**inférence des paramètres** du modèle à partir des proportions. Deux aspects ont été traités. D'une part, la définition du problème d'inférence du point de vue théorique, en établissant le rapport entre les variables expérimentales et les paramètres constitutifs du modèle. D'autre part, le développement d'un outil informatique opérationnel permettant d'automatiser le passage entre les données de proportion et les paramètres du modèle.

Nous proposons une méthode d'inférence, basée sur un processus de **déconvolution**, qui permet d'introduire l'information des proportions dans le modèle et ainsi de reproduire dans les simulations une distribution non stationnaire des hétérogénéités de l'unité sédimentaire. Ces simulations respectent ainsi par construction la distribution latérale et verticale des proportions dans l'unité de même que la valeur moyenne globale de la proportion sur tout le volume étudié. Celle-ci est associée à des paramètres pétrophysiques, et il est important de la retrouver dans les simulations.

Mots clefs :

Caractérisation des réservoirs, faciès, géostatistique, inférence des paramètres, modèle booléen, non stationnarité, proportions, simulation, stochastique.