Intégration de modèles approchés pour mieux transmettre l'impact des incertitudes statiques sur les courbes de réponse des simulateurs d'écoulements

Gaétan Bardy

▶ **To cite this version:**

Gaétan Bardy. Intégration de modèles approchés pour mieux transmettre l'impact des incertitudes statiques sur les courbes de réponse des simulateurs d'écoulements. Sciences de la Terre. Université de Lorraine, 2015. Français. NNT : 2015LORR0253 . tel-01754538

HAL Id: tel-01754538

https://hal.univ-lorraine.fr/tel-01754538

Submitted on 30 Mar 2018

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

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

AVERTISSEMENT

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

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

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

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

LIENS

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

INTEGRATION DE MODELES APPROCHES POUR MIEUX TRANSMETTRE L'IMPACT DES INCERTITUDES STATIQUES SUR LES COURBES DE REPONSE DES SIMULATEURS D'ECOULEMENTS

THESE

pour obtenir le grade de

Docteur de l'Université de Lorraine

Ecole doctorale RP2E

Spécialité Géosciences

par

Gaétan BARDY

Thèse soutenue le 27 Octobre 2015 devant le jury :

Mme. Mickaele Le Ravalec	IFPEN	Rapporteur, Président
M. Jef Caers	Professeur, Stanford University	Rapporteur
M. Daniel Busby	Total	Examinateur
M. Guillaume Caumon	Professeur, Université de Lorraine	Directeur de Thèse
M. Peter King	Professeur, Imperial College	Directeur de Thèse
M. Pierre Biver	Total	Directeur de Thèse
M. Philippe Renard	Professeur, Université de Neuchâtel	Invité

UMR 7359 – GeoRessources

ENSG – Université de Lorraine

TSA 70605, 54518 Vandœuvre-lès-Nancy Cedex – France

Remerciements

Mes tous premiers remerciements vont aux membres de mon comité de thèse qui ont accepté de suivre mon travail, de me guider par leurs remarques et de relire à de nombreuses reprises mes quelques productions littérales. Plus particulièrement : **Pierre Biver**, qui a cru en moi dès mon stage de fin d'études, m'a ensuite proposé de continuer à travailler ensemble avec ce sujet de thèse, et dont le soutien pour la reconnaissance de celle-ci a été sans faille. **Guillaume Caumon**, qui après trois années en tant que professeur, a bien voulu rempiler pour trois nouvelles années en tant que directeur de thèse et m'a accueilli chaque année au sein de son laboratoire. **Philipe Renard**, qui a toujours su poser les bonnes questions me permettant de proposer une solution innovante, notamment pour la minimisation. **Peter King**, qui a accepté de prendre part à ces recherches, et m'a permis de prendre du recul dans la perception de mon travail. Enfin **Vincent Corpel**, qui m'a permis de garder un œil vigilant concernant les besoins et les attentes de l'industrie sur ce sujet.

Ma reconnaissance va ensuite à **Total S.A.**, qui a accepté de financer ces travaux de thèse et de m'héberger au sein de son centre de recherche Jean Feger à Pau pendant ces trois ans. J'ai eu de ce fait l'occasion de discuter avec différentes personnes et découvrir plus en profondeur le travail au sein d'une major pétrolière.

Je remercie également l'**ANRT** pour son financement dans ces travaux grâce à une convention Cifre qui a permis le bon déroulement de ces années de recherches.

Enfin je remercie chaleureusement **Daniel Busby**, ingénieur de recherche réservoir à Total S.A. qui a accepté d'être examinateur, ainsi que **Jef Caers**, Professeur à l'université de Stanford, et **Mickaele Le Ravalec**, Chef du département de Géologie de l'IFPEN, qui ont bien voulu être les rapporteurs de mon travail.

En ce qui concerne les remerciements moins professionnels mais tout aussi importants, je suis reconnaissant envers mes différents **collègues de Pau** qui m'ont permis de passer trois grandes années : l'ensemble du service Géostatistiques et Incertitudes, et tout particulièrement **Tatiana** pour m'avoir supporté dans son bureau pendant deux ans et avoir partagé son eau chaude pour le thé. **Victor**, co-thésard du service avec lequel nous avons eu de nombreuses discussions passionnantes sur les différences franco-russes, les bizarreries de la langue française et comment des chapeaux colorés peuvent sauver des vies ! **Christophe**, **Grégoire** et **Jérôme**, pour m'avoir accompagné lors de toutes nos réunions à la machine à café (et spécialement Jérôme pour avoir si souvent payé !).

Je tiens aussi à remercier tous **les thésards de Nancy** pour les bons moments passés lors de mes séjours au laboratoire, ainsi que **les non thésards** présents également. Remerciement tout particulier à **Fatima** pour son aide lors de toutes mes inscriptions à distance.

Pour finir, ma reconnaissance va à **ma famille** pour son soutien sans faille tout au long de mes études, et tout particulièrement à ma femme **Marie** qui m'a suivi à l'autre bout de la France !

27/10/2015

Table des matières

Partie I : Introduction 4
1.1 Contexte de la recherche 5
1.2 Principales méthodes pour quantifier les incertitudes 6
1.3 Limites des approches actuelles et objectifs de la thèse 11
1.4 Organisation de la thèse 13
Partie II : Théorie 15
Chapitre 1 : Sélection des modèles 17
1.1 La sélection *a priori* 17
1.2 La sélection *a posteriori* 19
Chapitre 2 : Les simulations proxys 22
2.1 Les proxys scalaires 22
2.2 Les proxys par propagations de front 23
2.3 Les proxys plus complexes 26
Chapitre 3 : Comparaison des proxys 28
3.1 Critères basés sur la distance entre les modèles 28
3.2 Critères basés sur l'accessibilité et le gain de temps 33
3.3 Critères basés sur la stabilité vis-à-vis du calcul de quantiles de production 34
Chapitre 4 : Réduction de l'incertitude sur les profils de production à partir des simulations proxys et dynamiques ..... **Error! Bookmark not defined.**
4.1 Relation entre les distances proxys et les distances vraies 37
4.2 Approximation du profil de FOPT par une fonction paramétrique 39
4.3 Algorithmes d'optimisation pour la détermination des paramètres des modèles mathématiques 42
4.4 Correction de l'erreur systématique 45

Partie III : Application **48**
Chapitre 1 : Développement informatique 50
1.1 Mise en place d'un modèle et d'une base de données 50
1.2 Mise en place d'une interface graphique 55
Chapitre 2 : Application au champ X 60
2.1 Présentation générale 60
2.2 Comparaison des différents proxys 62
2.3 Amélioration du calcul des quantiles de production 71
Chapitre 3 : Application au champ Y 79
3.1 Présentation générale 79
3.2 Comparaison des différents proxys 81
3.3 Amélioration du calcul des quantiles de production 87
Partie IV : Conclusions **94**
Partie V : Références **98**

Partie I : INTRODUCTION

1.1 Contexte de la recherche

Dans la société actuelle, les besoins en énergie sont en constante augmentation. La part des énergies fossiles (charbon, hydrocarbures et nucléaire) est encore très importante et leur approvisionnement constitue un enjeu majeur. Etant donné leur période de renouvellement très longue, leur quantité est finie à l'opposé des énergies renouvelables. Il est donc primordial de pouvoir évaluer les quantités en place, ainsi que celles qui sont récupérables. Dans le cas de l'industrie pétrolière, domaine d'application de cette thèse, il s'agit d'hydrocarbures qui sont des molécules formées de grandes chaînes d'atomes de carbone hydratés. Ces molécules proviennent de la décomposition de matière organique sous des conditions de haute pression et haute température. Les hydrocarbures forment des fluides (liquides ou gaz) qui sont capables de se déplacer à l'intérieur de la roche, grâce aux pores qu'elle contient. Chercher du pétrole (ou du gaz), revient à chercher des caractéristiques géologiques particulières : une roche riche en matière organique, qui a subi un enfoncement et un réchauffement suffisant pour produire des hydrocarbures. La présence de pièges géologiques (failles ou plis imperméables) capables de retenir les fluides est également nécessaire. Ce sont ces pièges qui sont appelés des réservoirs pétroliers. Ces derniers sont formés d'une roche poreuse, c'est-à-dire capable de contenir des fluides et perméable pour pouvoir laisser circuler ces fluides. Afin de pouvoir repérer la présence de réservoirs d'hydrocarbures à exploiter, l'industrie pétrolière dispose de plusieurs outils comme la sismique (image du sous-sol obtenue par réflexion d'onde sur des réflecteurs) ou les puits d'exploration.

Une fois un réservoir détecté, il faut le modéliser informatiquement pour pouvoir déterminer les quantités d'hydrocarbures présentes. Il faut pour cela suivre plusieurs étapes clés d'un processus, décrites dans la Figure 1 :

- L'interprétation des données à notre disposition afin de les faire coïncider pour que, par exemple, un contact fluide visible sur la sismique se retrouve sur les données de puits à la bonne profondeur.
- La construction des surfaces géologiques caractéristiques comme les horizons, qui sont une interface entre deux couches aux propriétés différentes ou des failles.
- Le maillage du réservoir, de telle sorte que l'ensemble de la zone d'intérêt soit constituée de petits volumes.
- La modélisation des propriétés pétrophysiques afin de remplir l'ensemble des cellules grâce à différents algorithmes.

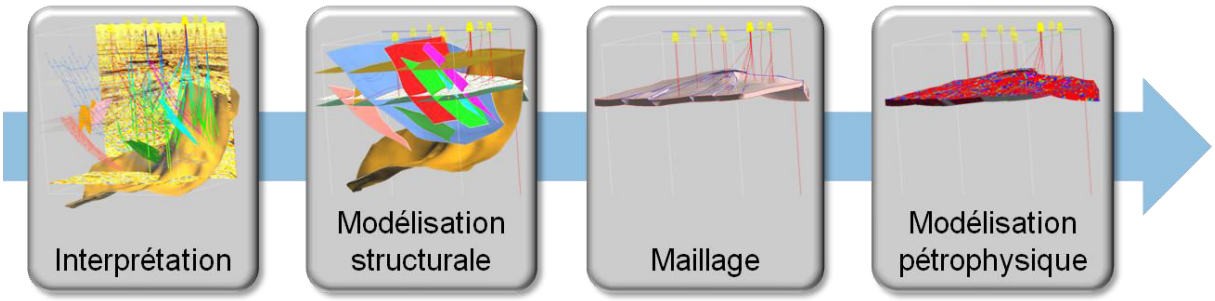

Figure 1 : Processus de construction d'un géomodèle.

Une fois le géomodèle construit il est possible d'évaluer les quantités d'hydrocarbures en place pour se faire une idée de la rentabilité d'un champ. Pour pouvoir récupérer les hydrocarbures présents dans les réservoirs, qui se trouvent le plus souvent enfouis à quelques kilomètres de profondeurs, il faut forer des puits de production afin de permettre aux différents fluides de remonter à la surface où ils vont ensuite être traités. Les réservoirs étant de plus en plus difficilement accessibles et nécessitant de lourds investissements pour être exploités, il est nécessaire d'évaluer les quantités qui vont pouvoir en être extraites et qui sont appelées réserves. Pour cela des simulations dynamiques d'écoulements sont effectuées, c'est-à-dire que, grâce à un logiciel

informatique, le comportement dynamique des fluides présents dans le réservoir au cours du temps va être simulé en fonction de différents paramètres (puits, fluides, propriétés des roches, etc.). C'est seulement une fois les résultats de cette simulation obtenus qu'il est possible de dire si le projet sera rentable ou non.

1.2 Principales méthodes pour quantifier les incertitudes

L'observation directe du sous-sol étant physiquement impossible, la représentation numérique d'un réservoir et son exploitation ne peut être certaine. Il est donc également important de connaitre les incertitudes sur ces réserves en matière première, car ce sont elles qui aident à estimer si le risque associé n'est pas trop grand.

Dans le domaine pétrolier, les incertitudes sont présentes à toutes les étapes :

- Lors de la collecte des mesures : netteté des réflecteurs sur la sismique, précision des machines utilisées lors des tests sur les échantillons en laboratoire, ou encore la trajectoire des puits dont la longueur ne correspond pas forcément à la profondeur et peut impacter sur la détermination de la position des contacts fluides mesurés.
- Lors de la création du géomodèle : environnement de dépôt géologique, répartition des faciès, fluides en place ou propriétés des failles (scellantes ou passantes), cellules du maillage irrégulières.
- Lors de la simulation dynamique d'écoulement dans le réservoir : fluides présents, perméabilités relatives, performance des puits, ou encore modèle utilisé pour représenter l'aquifère entourant le réservoir.

Quelque-soit l'origine de ces incertitudes, elles peuvent avoir un impact important à deux niveaux distincts :

- Sur l'évaluation des quantités d'hydrocarbures en place : c'est le cas pour la répartition des propriétés pétrophysiques ou la position des contacts. Nous les appellerons par la suite incertitudes sur la statique.
- Sur l'évaluation des réserves (quantités d'hydrocarbures qu'il est possible de récupérer) : c'est le cas des perméabilités relatives ou des propriétés des failles mais aussi encore une fois de la répartition de la pétrophysique. Nous les appellerons par la suite incertitudes sur la dynamique.

Il a donc fallu développer des méthodes qui permettent d'estimer accumulations et réserves dans une zone particulière du sous-sol, à partir de données ponctuelles (forages) et indirectes (sismiques), tout en tenant compte des incertitudes : c'est la géostatistique.

Le premier algorithme développé à ces fins, dans les années 50, est le krigeage (Krige (1951), Matheron (1955)). Il permettait historiquement à l'industrie minière d'estimer les teneurs en éléments par bloc rocheux, afin d'optimiser l'exploitation. Pour fonctionner, le krigeage nécessite de connaître le comportement spatial des données et d'avoir des points de contrôle. Afin de caractériser le comportement spatial des données, un variogramme expérimental (Figure 2a) est construit puis approché ensuite avec un modèle mathématique (Figure 2b).

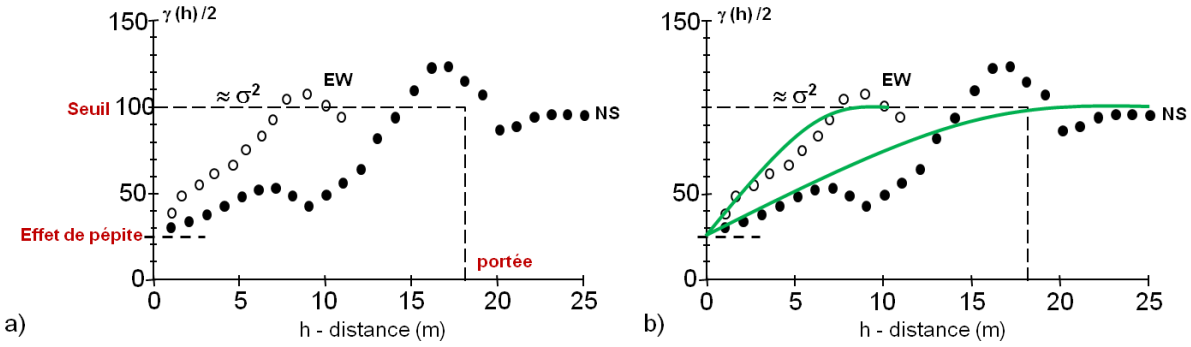

Figure 2 : Variogrammes (a) expérimentaux, (b) expérimentaux et analytiques

Les variogrammes illustrent la corrélation spatiale des données en fonction de la distance qui les sépare. Ils permettent de visualiser rapidement les caractéristiques statistiques suivantes :

- La portée : la distance maximale à partir de laquelle on considère que les points ne sont plus corrélés.
- Le seuil : la valeur de stabilisation du variogramme, qui correspond à la variance de la variable.
- L'effet de pépite : la discontinuité à l'origine qui peut être liée à de fortes variations à petite échelle ou à des erreurs de mesure.

Les points de contrôle sont les positions de données (forage). Il existe une large variété de krigeages dont les estimateurs s'écrivent de façons différentes (Chilès and Delfiner (1999)) Envisageons par exemple le krigeage simple, celui-ci revient à estimer en tout point l'écart de la variable à une moyenne connue. Cet écart est lui-même une combinaison linéaire des écarts à la moyenne au voisinage de ce point :

$$\hat{Z}_x = m + \sum_{i=1}^{N} \lambda_i \left(Z_{x_i} - m\right) \tag{1}$$

où N est le nombre de données au voisinage du point estimé, m la moyenne connue de la variable et λ_i les poids de krigeage. Le krigeage est considéré comme le meilleur estimateur linéaire, car on cherche à minimiser la variance de l'erreur par rapport à la vraie valeur (2). Il est aussi dit non biaisé car l'espérance des erreurs est nulle (3).

$$\min\left(Var\left(Z_x - \hat{Z}_x\right)\right) = \min\left(C(0) - 2\sum_{i=1}^{N} \lambda_i C(x - x_i) + \sum_{i,j=1}^{N} \lambda_i \lambda_j C\left(x_i - x_j\right)\right) \tag{2}$$

$$E\left(Z_x - \hat{Z}_x\right) = 0 \tag{3}$$

Où C représente la covariance. Pour déterminer les valeurs de λ_i, il faut résoudre le système linéaire suivant comportant N équations et N inconnues :

$$\begin{bmatrix} C(x_1 - x_1) & \cdots & C(x_1 - x_N) \\ \vdots & \ddots & \vdots \\ C(x_N - x_1) & \cdots & C(x_N - x_N) \end{bmatrix} \times \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{bmatrix} = \begin{bmatrix} C(x_1 - x) \\ \vdots \\ C(x_N - x) \end{bmatrix} \tag{4}$$

La Figure 3a montre un exemple de krigeage simple de porosité obtenue sur une grille avec plusieurs dizaines de puits de conditionnement.

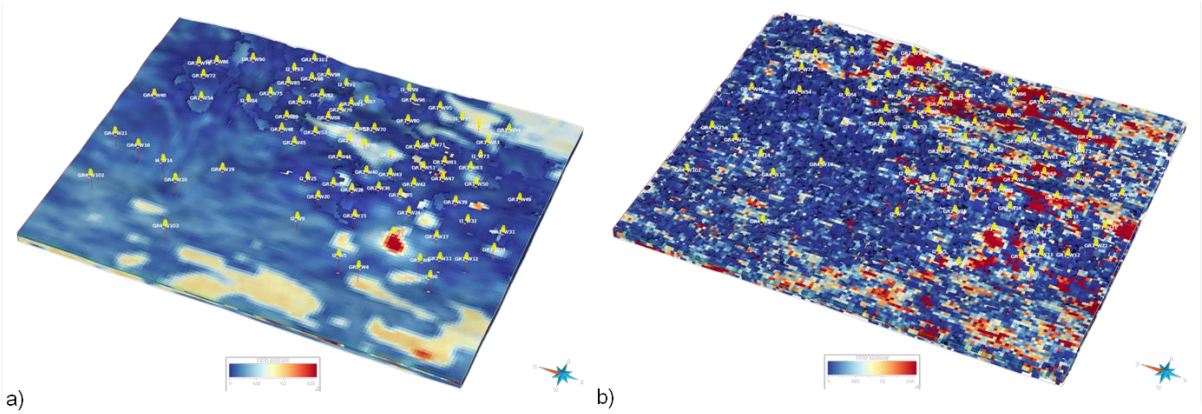

Figure 3 : Interpolation d'une propriété dans un géomodèle. (a) par krigeage. (b) par simulation.

Un des principaux problèmes du krigeage est qu'il a tendance à lisser les variations spatiales de la variable simulée (il s'agit d'un estimateur au sens des moindres carrés). Ceci est particulièrement vrai loin des données conditionnantes où l'on retrouve mathématiquement la moyenne de la variable. Or les variations locales peuvent avoir une forte influence sur les résultats d'une simulation d'écoulements. De plus, le krigeage étant un interpolateur moyen, il ne fournit qu'une seule représentation possible du sous-sol ce qui n'est pas l'idéal pour étudier les incertitudes.

Pour pallier ces limites, d'autres algorithmes géostatistiques ont été développés : les simulations conditionnelles. Tout comme le krigeage elles remplissent la grille du géomodèle, mais elles déterminent la valeur de la variable par un tirage aléatoire selon une loi de distribution permettant de fournir un nombre quasi illimité de représentations possibles. Il existe plusieurs algorithmes de simulation différents, regroupés en deux catégories principales (Caers (2011)) :

- les ***simulations pixels*** : elles simulent la valeur de la variable aléatoire cellule après cellule (pixel après pixels). Parmi les différents algorithmes on peut mentionner les SGS (Simulations Gaussiennes Séquentielles) pour les variables continues, les SIS (Simulations par Indicatrices Séquentielles) (Figure 4a) et TGS (Simulations par Gaussiennes Tronquées) (Figure 4b) pour les variables discrètes, ou bien encore les MPS (Simulations MultiPoints) (Figure 4d) qui n'utilisent pas explicitement le variogramme mais une image d'entraînement pour définir la corrélation spatiale entre les cellules.
- les ***simulations objets*** (Figure 4c) : elles permettent de générer des structures géologiques en positionnant des objets élémentaires dans la grille du géomodèle. Ce type de simulation est utilisé pour simuler des variables discrètes, telles que les faciès. Il faut alors déterminer la forme des objets à générer, leur nombre ainsi que les différents contacts entre eux. L'avantage de telles simulations est de garantir la préservation de l'aspect géologique des résultats obtenus. Par contre, il est parfois assez difficile d'obtenir un bon conditionnement aux données de puits.

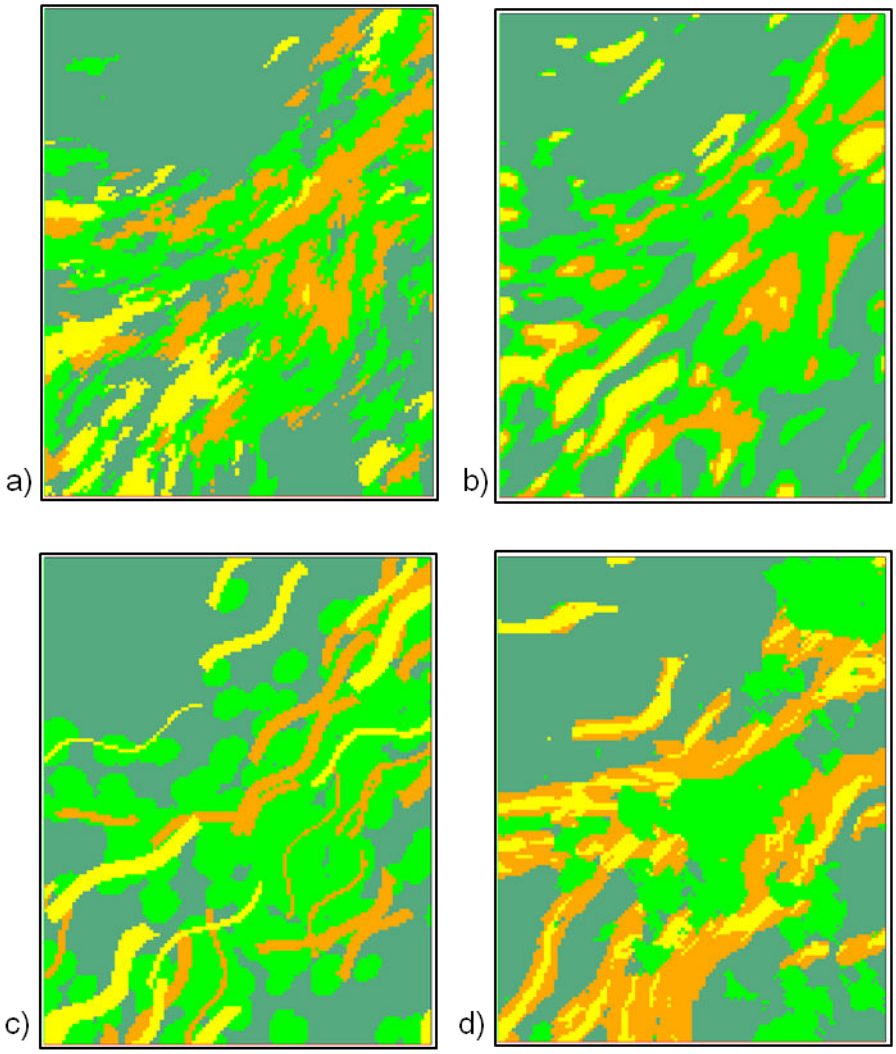

Figure 4 : Exemples de simulations géostatistiques de faciès. (a) SIS, (b) TGS, (c) Simulation objet et (d) MPS

Comme on le voit sur les différents exemples de simulation (Figure 3 et Figure 4), et plus particulièrement pour la variable continue, le résultat présente des fluctuations plus réalistes que pour le krigeage. Mais le résultat n'est pas unique. Ceci a un effet sur le comportement dynamique du modèle.

Considérons un modèle dont les propriétés pétrophysiques ont été obtenues par krigeage et comparons-le, du point de vue de l'écoulement des fluides à un ensemble de modèles dont les propriétés pétrophysiques ont été obtenues par simulation (SIS par exemple). Le cas krigé ne représente pas le comportement moyen de l'ensemble des modèles simulés : dans notre cas (Figure 5), la déplétion est plus rapide et la percée d'eau plus précoce ; dans d'autres cas, si le maintien de la pression est meilleur, la percée d'eau peut-être plus tardive pour le cas krigé.

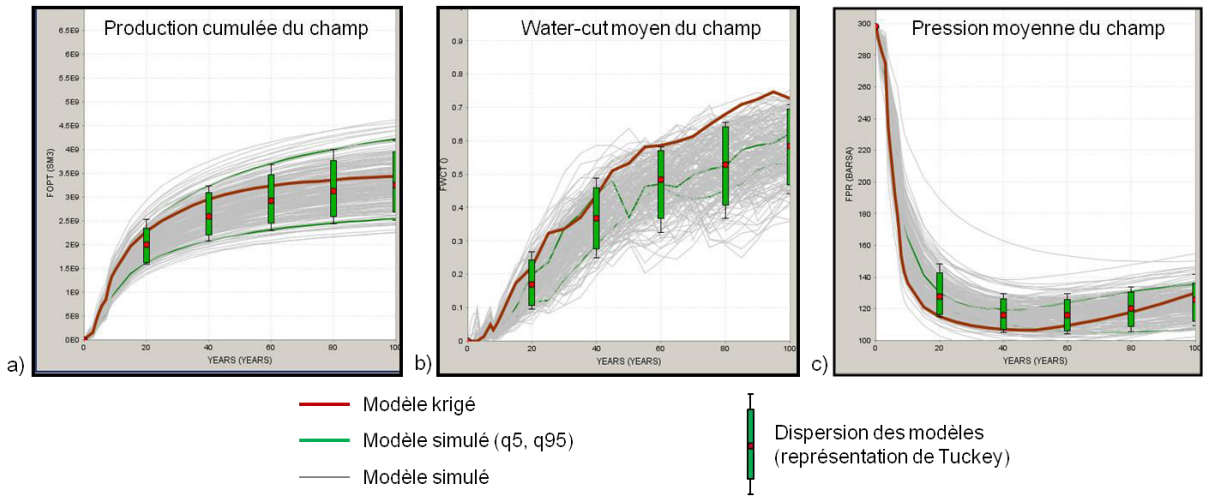

Figure 5 : Profils de production dynamiques pour des modèles construits par krigeage et des modèles construits par simulation. (a) Production d'huile cumulée du champ. (b) Pourcentage d'eau moyen parmi les fluides produits (water-cut) du champ. (c) Pression moyenne du champ.

Pour avoir des comportements dynamiques réalistes, il est donc le plus souvent nécessaire de modéliser le réservoir avec des algorithmes de simulations (pour les variables catégoriques, c'est même crucial). Comme ceux-ci ne produisent pas de résultat unique, il convient d'étudier les incertitudes qu'ils apportent.

Pour atteindre cet objectif, plusieurs centaines de réalisations sont produites afin de représenter au mieux les différentes possibilités (Corre et al. (2000), Charles et al. (2001)). Ces réalisations permettent de déterminer l'incertitude pour une fonction de réponse (hydrocarbures en place, réserves, etc.) calculée sur chaque modèle construit. Il faut en effet plusieurs dizaines (voire centaines) de modèles pour que la moyenne des volumes d'huile en place se stabilise (Figure 6).

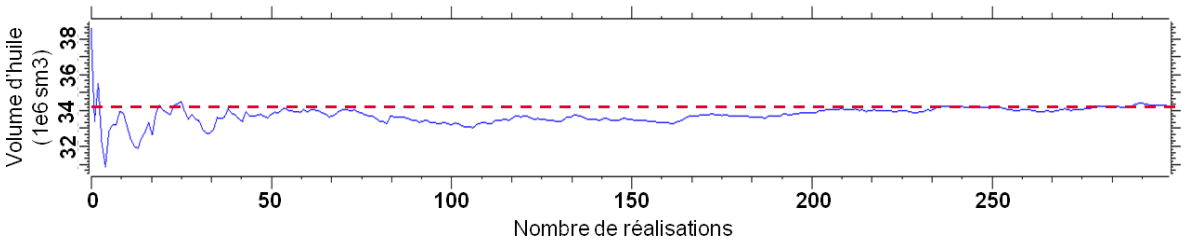

Figure 6 : Evolution de la moyenne d'huile en place en fonction du nombre de modèles générés

Une fois les incertitudes sur les hydrocarbures en place déterminées, c'est au tour des incertitudes sur les réserves. Pour cela, c'est habituellement un modèle appelé « cas de base », qui est utilisé comme référence. Le modèle généralement choisi est le cas médian de l'ensemble des paramètres statiques étudiés. Un schéma de développement de puits est ensuite déterminé, afin de pouvoir effectuer la simulation d'écoulements (Figure 7). Il est alors possible de faire varier certains paramètres dynamiques, tels que les perméabilités relatives ou le type de modèle analytique utilisé pour représenter un aquifère présent dans le sous-sol. En effectuant à chaque fois la simulation d'écoulements correspondant, il est possible de déterminer les incertitudes sur les réserves engendrées par ces mêmes paramètres. Mais il convient également de juger de l'impact des paramètres ayant servi à construire le géomodèle.

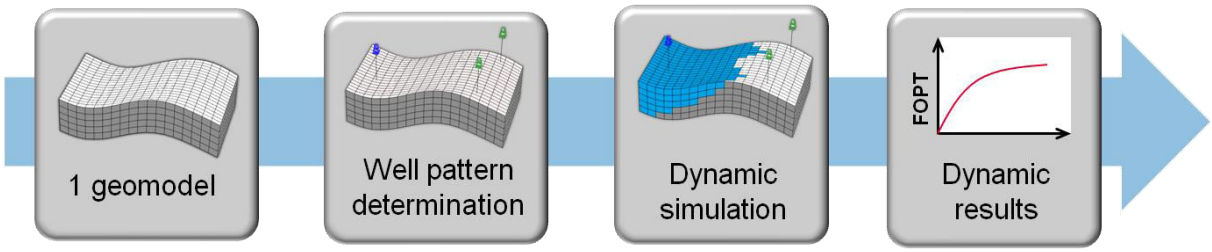

Figure 7 : Processus d'utilisation d'un cas de base pour étudier les incertitudes dynamiques. FOPT signifie « *Field Oil Production Total* » soit la quantité d'huile cumulée produite à l'échelle du champ.

1.3 Limites des approches actuelles et objectifs de la thèse

Avec l'évolution des capacités de calcul des ordinateurs, il est devenu très rapide de construire un modèle à l'aide de simulations géostatistiques (pour un modèle de plusieurs millions de cellules il ne faut que quelques minutes). De plus certains algorithmes sont parallélisables, ce qui accélère encore le temps de calcul. Le problème est que le temps d'une simulation dynamique reste, lui, beaucoup plus long. En effet, il est lié au nombre de cellules dans le géomodèle et celui-ci a tendance à augmenter. Il n'est donc pas possible de simuler tous les modèles produits. Si l'on peut générer plusieurs centaines de modèles pour être sûr d'avoir une bonne évaluation des incertitudes sur les volumes d'hydrocarbures en place, il est rare dans un contexte opérationnel de pouvoir simuler plus d'une vingtaine de modèles différents (certaines simulations les plus complexes avec des fluides à composition variable, des effets de double milieu ou des couplages géomécaniques durent plus d'une semaine). Afin de pallier cette limitation, mais pouvoir toujours étudier les incertitudes dynamiques, le processus du cas de base (Figure 7) peut être modifié de deux façons différentes :

- Soit une étape de sélection de modèles représentatifs est ajoutée, afin de réduire le nombre de simulations d'écoulements à réaliser (Figure 8a). Dans ce cas seules les informations des modèles simulés sont utilisées, au contraire du reste des modèles.
- Soit la simulation dynamique est remplacée par une simulation plus rapide sur l'ensemble des modèles (Figure 8b). Dans ce cas les réponses obtenues sont approchées. Nous parlerons de simulation proxy pour qualifier ce type de méthodes.

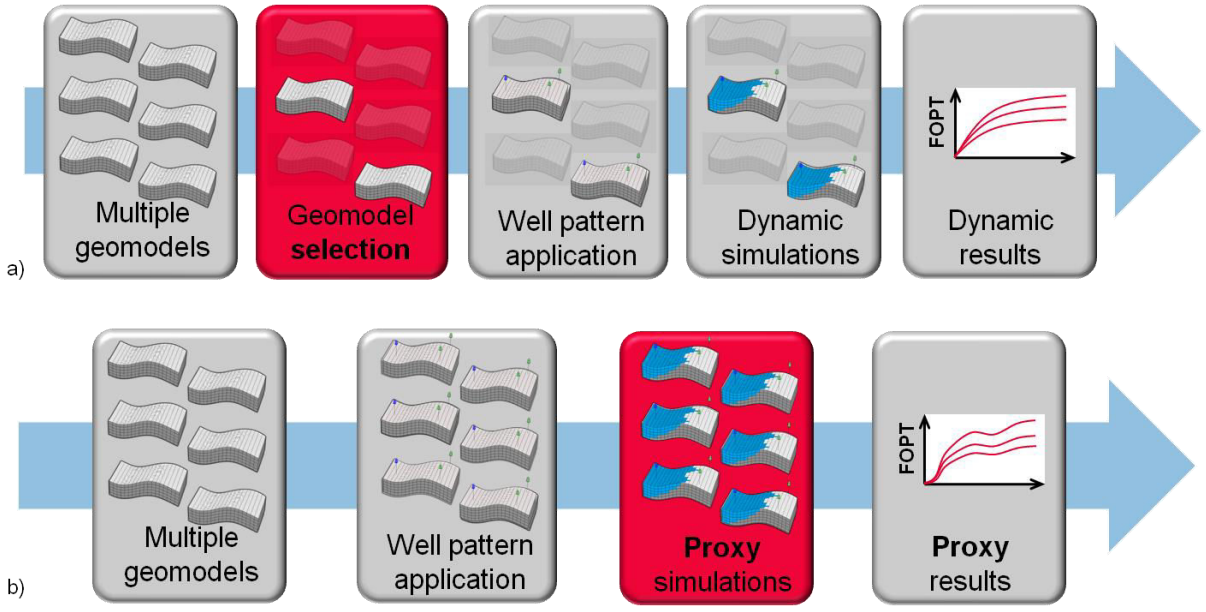

Figure 8 : Modification du processus de simulations dynamiques pour pouvoir utiliser plusieurs géomodèles. (a) On effectue une sélection des modèles caractéristiques. (b) On remplace la simulation dynamique par une simulation proxy.

Dans le cas d'une sélection de modèles, la principale question à se poser est « Comment choisir ces quelques modèles pour qu'ils puissent être suffisamment représentatifs et permettre d'évaluer des incertitudes sur les réserves ? » Pour répondre à cette question, il existe traditionnellement deux méthodes différentes :

- Soit on construit les modèles dont on a besoin spécifiquement pour étudier la sensibilité de différents paramètres : c'est l'approche par plan d'expérience (Figure 9) (Box and Draper (1987), Damsleth et al. (1992)).
- Soit on échantillonne parmi l'ensemble des modèles disponibles pour n'en retenir qu'un petit nombre qui peut être simulé dynamiquement (Scheidt and Caers (2007))(Figure 10).

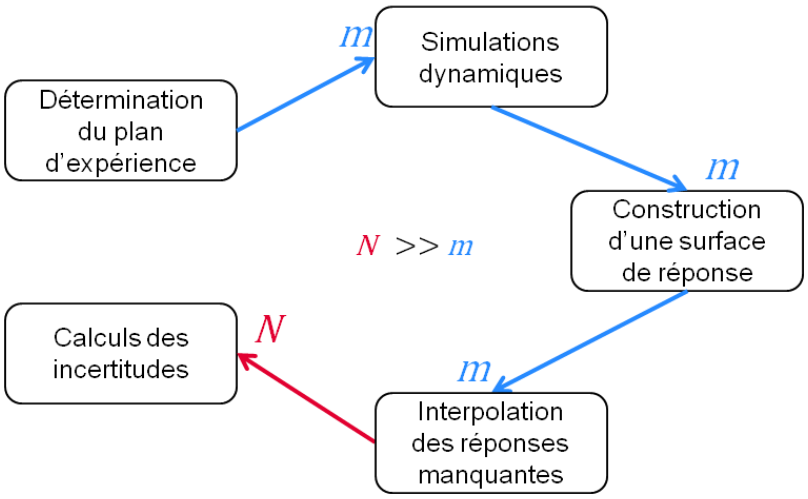

Figure 9 : diagramme de fonctionnement d'un plan d'expérience, *N* étant le nombre de modèles total et *m* le nombre de modèles réellement simulés.

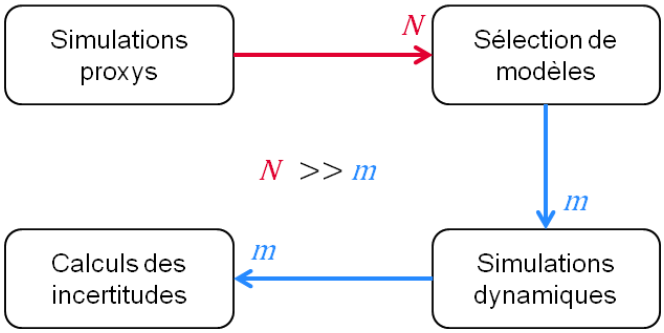

Figure 10: diagramme de fonctionnement de la sélection de modèles, *N* étant le nombre de modèles initiaux et *m* le nombre de modèles réellement simulés.

La première méthode permet de garantir un échantillonnage optimal pour étudier les paramètres voulus. Elle est très utile pour étudier l'influence de tous les paramètres globaux : modèle de saturation en eau, débit d'injection des puits, perméabilité relative. Par contre cette méthode n'est pas robuste pour les paramètres locaux comme la position d'une hétérogénéité géologique ou la variation locale de la perméabilité qui se rapportent à des variables spatiales tridimensionnelles dans le géomodèle. C'est encore plus difficile quand il faut construire le géomodèle à partir des valeurs de chaque paramètre. Pour cette raison, il peut sembler plus facile de partir d'une collection de modèles existants pour n'en sélectionner qu'un petit nombre le plus représentatif possible de l'ensemble. Plusieurs méthodes existent pour cela. La plus utilisée est le *space filling design* qui nécessite de représenter les modèles dans un espace de paramètres scalaires que l'on veut utiliser pour distinguer nos modèles. L'algorithme emploie alors des sphères

d'exclusion pour ne retenir que quelques modèles distincts les uns des autres. Le problème de cette technique est qu'elle ne peut prendre en compte que des paramètres scalaires tels que les accumulations. Quelle que soit la technique utilisée, le biais principal de la sélection de modèles est que toutes les informations contenues dans les géomodèles non sélectionnés sont perdues.

Comment est-il possible d'intégrer l'information des modèles restants non simulés pour améliorer l'évaluation des incertitudes ?

En ce qui concerne l'utilisation des proxys pour approximer la réponse dynamique des modèles (Figure 8b), il existe de très nombreux proxys qui représentent autant de compromis entre la fiabilité de la réponse par rapport à la référence dynamique, la vitesse de simulation et la facilité d'utilisation.

Comment faut-il choisir le proxy à utiliser ?

La contribution principale de cette thèse est de proposer une nouvelle méthode pour évaluer l'impact des paramètres servant à définir les géomodèles (comme les variables pétrophysiques) sur l'incertitude autour des réserves. Nous nous limiterons dans cette thèse au cas particulier des champs pétroliers non opérés, c'est-à-dire pour lesquels il n'existe pas encore de données de production. La Figure 11 illustre la méthode proposée. Elle combine les deux techniques classiquement utilisées par l'industrie pétrolière pour répondre à cette question : l'utilisation de simulations proxys et la sélection de géomodèles représentatifs (grâce aux réponses du proxy). Mais en plus, nous avons rajouté une étape qui permet de déterminer des profils dynamiques possibles pour les modèles non simulés et de les ajouter aux modèles vraiment simulés. C'est cette étape qui présente un caractère nouveau et permet de préciser l'évaluation des incertitudes autour des réserves. En parallèle de ce travail, nous présenterons aussi une méthodologie de comparaison des proxys sans avoir besoin de réellement simuler tous les modèles pour calibrer la réponse, car ce n'est pas envisageable en contexte opérationnel.

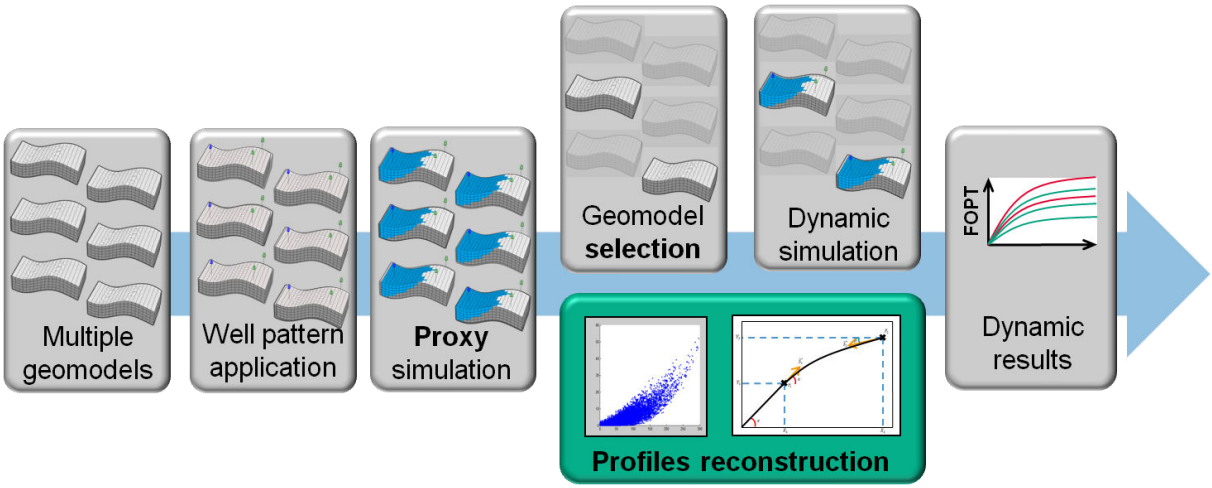

Figure 11 : diagramme de fonctionnement de la méthodologie proposée afin d'améliorer l'étude de l'impact des paramètres statiques sur l'évaluation des incertitudes dynamiques.

1.4 Organisation de la thèse

Le manuscrit sera divisé en deux parties. Une première partie théorique, qui expliquera le principe de la sélection de modèles, présentera plusieurs proxys de simulateur d'écoulements et décrira en détails les méthodologies pour comparer les proxys et pour améliorer la précision des calculs de quantiles de production. La deuxième partie sera plus applicative et exposera le développement informatique réalisé pour implémenter les travaux, ainsi que l'application des méthodologies sur deux cas réels fournis.

Cette thèse a bénéficié d'un financement CIFRE entre Total S.A et l'Université de Lorraine. Certains choix, que nous préciserons par la suite, ont donc fait l'objet de contraintes particulières imposées par la société et ne seront pas débattus.

Partie II : THEORIE

Nous allons détailler dans cette partie les motivations de ce travail, ainsi que les développements originaux qui ont été proposés. Comme nous l'avons vu dans l'introduction, le but de cette thèse est de développer une nouvelle méthode pour étudier l'impact des paramètres statiques du géomodèle (existence et position des failles, distribution des variables pétrophysiques, position des contacts) sur le résultat des simulations dynamiques d'écoulements, et plus particulièrement sur les prévisions de production. L'idée est d'être capable d'obtenir, à partir d'un ensemble de modèles représentant les variations des paramètres statiques, un modèle d'incertitudes sur l'écoulement dynamique. Dans l'industrie pétrolière, celui-ci est généralement exprimé sous la forme de trois quantiles de la variable étudiée : Q10, Q50 et Q90. Dans notre cas cette variable, que nous appellerons **FOPT** (pour *Field Oil Production Total*), est la production cumulée d'huile au niveau du champ au cours du temps. Elle forme donc une courbe. Notre modèle d'incertitude sera ainsi constitué de trois courbes statistiques qui ne sont pas de vrais profils dynamiques, car elles sont formées des quantiles calculés à partir de l'ensemble des courbes considérées à chaque pas de temps :

$$q_X(t) = q_X\big(FOPT_i(t)\big) \quad i \in [1, N] \tag{1}$$

où X correspond au quantile calculé et N au nombre de modèles disponibles. Plus ces trois courbes seront resserrées, plus l'impact sera faible, et inversement. Les simulations dynamiques d'écoulement étant souvent très longues, il n'est pas réellement possible de travailler avec l'ensemble des géomodèles obtenus pour illustrer la variation des paramètres statiques. Plusieurs méthodes permettant de pallier ce problème existent. Nous allons donc voir quelles sont les solutions de sélection des modèles qui existent dans la littérature, avec leurs avantages et leurs inconvénients, pour ne simuler qu'un nombre acceptable de modèles. Les différents proxys qu'il est possible de trouver seront aussi passés en revue, ainsi que les buts pour lesquels ils ont été développés. Nous détaillerons plus particulièrement les proxys de propagation de front utilisant l'algorithme de *Fast Marching*. Les deux méthodologies développées pendant cette thèse seront ensuite présentées dans les chapitres 3 et 4.

Chapitre 1 : Sélection des modèles

Plusieurs méthodes permettant de sélectionner des modèles peuvent être trouvées dans la littérature pour le domaine pétrolier. Nous allons présenter une revue bibliographique afin d'expliquer pourquoi il existe de telles variations et dans quels buts ces méthodes ont été développées. Nous exposerons aussi les principales limitations qui motivent ce travail de thèse.

Comme expliqué en introduction, la sélection des modèles est nécessaire lors de l'étude des incertitudes dynamiques d'un réservoir pétrolier. Cette sélection est difficile car il s'agit de ne retenir qu'un petit nombre de modèles pour effectuer les simulations dynamiques d'écoulements, c'est-à-dire qu'une partie de l'information disponible va être perdue. Plusieurs questions majeures se posent : quel est le nombre optimal de modèles à conserver ? Comment faire en sorte que les modèles soient suffisamment représentatifs pour permettre de caractériser les incertitudes correctement ? Il existe deux principales approches pour répondre à ces questions :

- la technique des plans d'expériences initialement développée par R. A. Fisher en 1924. Elle permet de déterminer les valeurs optimales des paramètres à étudier. Ainsi, le nombre d'expériences numériques à effectuer est le plus petit possible (Sacks et al., 1989).
- la décimation représentative au sein de l'ensemble initial des différents modèles possibles. Cet ensemble est souvent très large car les modèles sont peu coûteux à générer et nécessaires pour caractériser les incertitudes statiques.

Dans le premier cas le choix des paramètres se fait avant la génération des modèles, nous pouvons donc parler de sélection *a priori*, alors que dans le second cas nous effectuons la sélection parmi des modèles déjà en notre possession, c'est donc une sélection *a posteriori.*

1.1 La sélection *a priori*

Une sélection à priori repose sur trois étapes distinctes :

- la détermination des paramètres incertains qui peuvent influer sur l'évaluation des incertitudes ainsi que leur gamme de variation ;
- l'analyse de sensibilité de ces paramètres vis-à-vis de la simulation d'écoulement ;
- la construction d'un modèle analytique de réponse (ou surface de réponse) permettant de prévoir le comportement des autres modèles.

La première étape consiste simplement à effectuer un listing de tous les paramètres qui pourraient influer sur le résultat d'une simulation dynamique d'écoulement (Damsleth et al., 1992, Egeland et al., 1992). Il faut aussi déterminer la gamme de variation de ces paramètres, c'est-à-dire l'ensemble des valeurs possibles que peut prendre la variable (Zabalza-Mezghani et al., 2004, Jourdan and Zabalza-Mezghani, 2004).

Le nombre de paramètres peut vite devenir élevé, c'est pourquoi la deuxième étape d'analyse de sensibilité (aussi appelée *screening*) est ensuite effectuée. Cette dernière permet de retenir tous les paramètres incertains vraiment impactant sur le résultat parmi tous ceux listés (Venkataraman, 2000). En effet, si des paramètres non influents sont conservés dans notre étude d'incertitude, le nombre de simulations à effectuer ainsi que le temps de calcul vont être plus élevés sans qu'il n'y ait un intérêt. A l'inverse, si des paramètres influents sont oubliés dans la liste initiale, nous risquons d'obtenir des réponses biaisées.

La méthode classique pour effectuer une analyse de sensibilité consiste à déterminer trois niveaux de valeur pour chaque paramètre, qui correspondent aux cas pessimistes, médians et optimistes (typiquement Q10, Q50 et Q90). Chaque paramètre varie ensuite successivement en lui attribuant la valeur P10 ou P90, alors que tous les autres sont à leur valeur médiane. Le problème de cette méthodologie est qu'elle suppose que tous les paramètres sont indépendants, ce qui n'est pas toujours vérifié dans le cas d'un réservoir pétrolier (notamment il existe très souvent une relation entre porosité et perméabilité) (Franco, 2008).

Pour remédier à ce problème, il est possible d'utiliser des plans d'expériences. Ceux-ci permettent de déterminer le minimum de combinaisons de paramètres nécessaires pour obtenir le maximum d'informations. Le fait d'utiliser des combinaisons de paramètres plutôt que de les faire varier un par un permet de prendre en compte les interactions entre ceux-ci. Un plan d'expériences est matérialisé par une matrice D, où chaque colonne représente un paramètre et chaque ligne une expérience (une combinaison de valeurs de paramètres). Afin d'être plus compréhensibles, les paramètres sont généralement normalisés entre -1 et 1, ce qui permet une représentation graphique. Il existe plusieurs plans d'expériences. Le plus simple est appelé plan factoriel 2^k. Dans ce plan, chaque paramètre peut prendre deux valeurs possibles (-1 et 1) qui sont les valeurs extrêmes.

$$D = \begin{bmatrix} -1 & -1 \\ 1 & -1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix} \quad (2)$$

L'équation (2) illustre le cas d'un plan 2^k factoriel à deux paramètres. Nous pouvons voir que ce plan peut très vite amener à effectuer un grand nombre d'expériences (des simulations dynamiques dans notre cas), ce qui ne peut pas être envisagé à cause du temps de calcul.

Afin de mieux prendre en compte la variabilité de la réponse dynamique aux combinaisons de paramètres, il est possible d'utiliser plutôt des plans d'expériences qui sont des fractions du plan de type factoriel 3^k, où les paramètres peuvent prendre trois valeurs distinctes : -1, 0 et 1. Nous citerons notamment le plan de Box-Behnken (Box and Behnken, 1960) et les hypercubes latins (Mc Kay et. Al., 1979) (Figure 12).

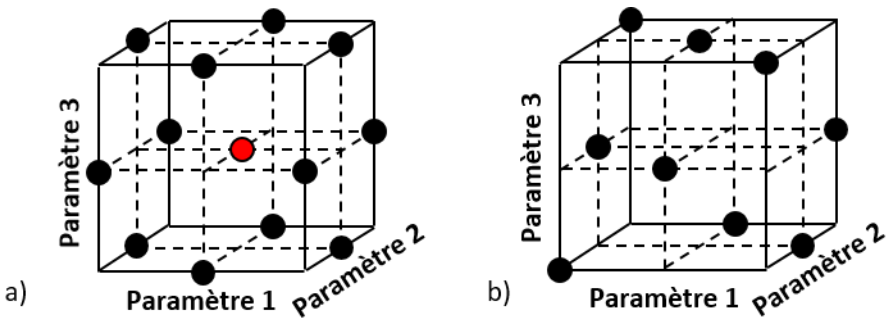

Figure 12 : Exemples de plans d'expériences à trois paramètres et trois niveaux. (a) Plan de Box-Behnken. (b) Plan par hypercubes latins

Une fois les expériences déterminées grâce à l'analyse de sensibilité, les simulations dynamiques d'écoulement correspondantes sont effectuées. Les réponses de celles-ci nous permettent de construire un modèle analytique, qui nous permet d'estimer la réponse dynamique pour toutes les autres combinaisons de paramètres. C'est la phase de modélisation. Parmi les méthodes utilisées pour construire le modèle analytique se trouvent la régression multi-variée (Dejean and Blanc, 1999), les réseaux de neurones (Lechner and Zangl, 2005) ou encore le krigeage. Les thèses de Franco (2008) et Scheidt (2006) présentent un grand nombre de références bibliographiques sur les applications de ces méthodes au domaine pétrolier, tels Damsleth et al. (1992). Malgré ces nouvelles techniques, le nombre d'expériences devient vite très grand avec le

nombre de paramètres à étudier. Ainsi nous devons souvent nous restreindre à une dizaine de paramètres pour être capables de simuler tous les modèles. Dans sa thèse, Santiago (2013) présente un autre algorithme permettant d'étudier davantage de paramètres (vingt-neuf pour l'application pratique présentée).

L'autre inconvénient majeur des plans d'expériences est qu'ils ne sont pas bien adaptés à l'étude des paramètres spatiaux, tels que les distributions des paramètres pétrophysiques (porosité, perméabilité, etc.). En effet, bien que la valeur moyenne sur l'ensemble de la grille puisse être calculée, plusieurs répartitions spatiales peuvent correspondre à la même moyenne. Or, la distribution spatiale dans le modèle géologique de ces propriétés influe considérablement sur le comportement dynamique du réservoir dans son ensemble, mais aussi sur le comportement individuel des puits.

1.2 La sélection *a posteriori*

Une autre méthodologie pour sélectionner les modèles consiste non pas à travailler sur les paramètres qui servent à les générer, comme le font les plans d'expériences, mais directement sur un ensemble de modèles existants et leur réponse en termes d'écoulements dynamiques. Un des avantages majeurs est que les modèles utilisés ont déjà été générés, car ils ont servis à étudier l'impact des différents paramètres sur des résultats statiques comme les volumes de fluides en place ou la fraction de ces volumes connectés aux puits. On peut citer, parmi les méthodes existantes utilisées au sein de l'industrie pétrolière, les filtres de Kalman ou *EnKF*, qui mettent à jour le géomodèle à chaque pas de temps de la simulation dynamique d'écoulement à partir d'un ensemble de modèles. Moins récemment, Ballin et al. (1992) présentent une technique qui permet de construire les distributions statistiques d'impact de paramètres sur les réserves d'un modèle. Ces courbes leur servent ensuite à déterminer les modèles P10, P50 et P90 en fonction de paramètres choisis pour étudier les incertitudes. Leur méthode n'utilise que la réserve comme réponse dynamique, c'est-à-dire uniquement la quantité d'huile effectivement produite au bout d'une durée définie, ce qui est assez simple. Pour aller plus loin, Scheidt and Caers (2009) présentent une nouvelle approche pour estimer les quantiles de production Q10, Q50 et Q90 :

- Un proxy est utilisé sur chaque modèle pour en estimer la réponse en écoulement.
- Une distance est calculée entre les réponses proxys des modèles deux à deux.
- Un algorithme de *Multi Dimensional Scaling* (MDS) (Cox and Cox, 2001) est utilisé pour déterminer les grandeurs caractéristiques des modèles.
- Un algorithme de Clustering, couplé avec une méthode à noyaux, permet de regrouper les modèles semblables afin de sélectionner un sous-ensemble représentatif.
- Les quantiles de productions sont calculés sur les modèles représentatifs, préalablement simulés.

Cette méthode utilise donc une réponse approchée du modèle pour ensuite calculer une distance entre chaque paire de modèles. La notion de distance a été employée à de nombreuses reprises en géologie, par Suzuki and Caers (2008) entre autres. Elle correspond à une mesure de non-similarité entre deux réponses (dynamiques dans notre cas). Une distance $\boldsymbol{d}$ doit respecter trois principes : **séparation** – une distance est toujours strictement positive exceptée la distance entre un modèle et lui-même qui est nulle (3) – **symétrie** – une distance d'un modèle A à un modèle B est la même que celle d'un modèle B au modèle A (4) – et **inégalité triangulaire** – une distance d'un modèle A à un modèle B ne peut pas être supérieure à la somme des distances d'un modèle A à un modèle C et d'un modèle C à un modèle B (5) :

$$d \geq 0 \;\&\; d_{A \to A} = 0 \tag{3}$$

$$d_{A \to B} = d_{B \to A} \tag{4}$$

$$d_{A\to B} \leq d_{A\to C} + d_{C\to B} \tag{5}$$

Le calcul des distances entre tous les modèles permet de construire une matrice des distances ($\boldsymbol{D}$) qui est positive, symétrique et nulle sur la diagonale :

$$D = \begin{bmatrix} 0 & d_{1\to 2} & \cdots & d_{1\to n} \\ d_{2\to 1} & 0 & \cdots & d_{2\to n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n\to 1} & d_{n\to 2} & \cdots & 0 \end{bmatrix} \tag{6}$$

Il existe plusieurs façons de calculer la distance entre des courbes. Par exemple, avec une distance de Hausdorff (le plus grand écart entre un point de chacune des courbes) calculée entre les courbes de production (Belogay et al., 1997) ou une norme L2 (racine carrée de la somme des écarts au carré). Tout dépend de la variable considérée. Une fois la matrice des distances construite, les auteurs utilisent l'algorithme de MDS permettant de représenter les modèles en tant que points dans un espace euclidien en respectant aux mieux les différentes distances (Cox and Cox, 2001, Borg and Groenen, 2005). Cet espace est formé par les composantes principales de la matrice des distances, ce qui permet de pouvoir réduire le nombre de dimensions facilement (souvent 95% de la distance s'explique par seulement deux ou trois valeurs propres). Après avoir effectué le MDS, les auteurs effectuent la sélection à proprement parler. Pour cela ils utilisent un algorithme de *Kernel Clustering,* c'est-à-dire un regroupement dans un espace non euclidien calculé par transformation par noyaux. En effet les méthodes par noyaux permettent, grâce à une fonction mathématique, de plonger l'espace euclidien $\boldsymbol{R}$ dans un espace, appelé *feature space* $\boldsymbol{F}$, dans lequel les données sont plus linéaires et où il est donc plus aisé de classifier les données (Shawe-Taylor and Christianini, 2004). Il existe plusieurs fonctions de transformation par noyaux, dont le noyau Gaussien (7) (fonction à base radiale) utilisé par Scheidt et Caers dans leur article.

$$k(x,y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \quad with\ \sigma > 0 \tag{7}$$

Une fois les modèles translatés dans l'espace F, il est possible de regrouper les modèles les plus proches entre eux en utilisant un algorithme de clustering. Ce dernier va commencer par effectuer un regroupement aléatoire des modèles en un nombre donné de classes. Puis, par une méthode d'optimisation itérative, l'algorithme améliore progressivement le regroupement (distances maximales entre les classes et minimales à l'intérieur de chacune d'elles). Il s'arrête quand les permutations ne permettent plus d'amélioration. Chaque cluster est alors caractérisé par les modèles qui le constituent et par son centre. Il existe de nombreux algorithmes possibles dont l'un des plus connus est le *K-Means* (Jain, 2012). Statistiquement, il est impossible qu'un des modèles de départ corresponde au centre du cluster. Comme il est difficile de reconstruire un modèle à partir de ses coordonnées dans l'espace du MDS, il est préférable d'utiliser un algorithme de clustering appelé *K-Medioïd* qui caractérise les clusters par les modèles qui le composent, son medioïd, c'est-à-dire le modèle le plus proche du centre ainsi que le poids de celui-ci. En effet, les clusters ne contiennent pas le même nombre de modèles, or ils ne sont représentés que par un seul modèle. Le poids de celui-ci (un scalaire compris entre 0 et 1) est dont proportionnel au nombre de modèles dans le cluster.

Scheidt et Caers utilisent les modèles médioïdes comme modèles représentatifs de l'ensemble initial, pondérés de leurs poids, pour étudier les incertitudes sur la production du réservoir. L'enjeu majeur est ici de trouver le nombre optimal de modèles à garder, c'est-à-dire le nombre de classes du *K-Means*. Contrairement à la sélection *a priori*, il n'existe pas d'algorithme donnant le nombre de modèles en fonction des paramètres utilisés. Ce nombre doit être suffisamment petit pour qu'il soit possible de simuler tous les modèles, mais aussi suffisamment

grand pour pouvoir représenter l'ensemble des variations. Nous illustrerons plus loin les erreurs qui peuvent venir d'une mauvaise sélection.

La sélection *a posteriori* possède aussi des inconvénients. En particulier, il n'est pas certain que l'ensemble initial illustre suffisamment les variations possibles. Dans ce cas, la sélection ne s'effectuant que sur les modèles existants, le sous-ensemble retenu sera également non représentatif. Il existe tout de même des méthodes essayant de résoudre ce problème (Satija and Caers, 2012).

Chapitre 2 : Les simulations proxys

Ce chapitre est dédié aux proxys. Après un bref rappel des équations qui régissent les écoulements, nous présenterons une revue bibliographique non exhaustive des proxys existants. Nous détaillerons tout particulièrement les proxys de propagation de front de saturations qui présentent un bon compromis entre rapidité et précision. Ils offrent en plus de grandes possibilités en termes d'évolution et ont fait partie intégrante de ce travail de thèse.

La simulation d'écoulement dynamique caractérise le comportement fluide en résolvant les équations de conservation, qui régissent le déplacement dans un milieu poreux, grâce à différentes méthodes numériques de résolution des équations aux dérivés partielles (comme les volumes finis). Par exemple, dans le cas d'un système bi-phasique huile-eau incompressible avec des fluides immiscibles, l'équation de conservation est celle de la masse qui spécifie que pour un volume de contrôle donné, la masse de matière échangée avec l'extérieur est la même que celle s'accumulant à l'intérieur. Les trois équations aux dérivées partielles utilisées sont :

L'équation de la pression :

$$\nabla \mathrm{u_t} = -\nabla K\left(\lambda_t \nabla P + \lambda_g \nabla D\right) = q_s \tag{8}$$

où u_t est la vitesse totale de Darcy, K le tenseur de perméabilité intrinsèque, P la pression totale des fluides, D la profondeur du réservoir, q_s une possible injection ou production de fluide, λ_t la mobilité totale et λ_g la mobilité gravitaire.

L'équation de Buckley-Leverett (1942) :

$$\Phi \frac{\partial S_w}{\partial t} + u_t \nabla f_w + \nabla g_w = f_{w,s} q_s \tag{9}$$

où S_w représente la saturation en eau, Φ la porosité effective, f_w la fraction du flux lié à l'eau, g_w la fraction du flux lié à la gravité et $f_{w,s}$ la fraction du flux global lié à l'eau à saturation donnée.

La répartition du volume poreux entre les différents fluides en présence :

$$S_o + S_w = 1 \tag{10}$$

où S_o représente la saturation en huile.

En géologie, et plus particulièrement dans l'exploitation pétrolière, le modèle se compose généralement de mailles hexaédriques représentant une partie du réservoir. Les équations doivent être résolues pour tous les pas de temps et dans toutes les mailles, ce qui explique la lenteur de la résolution numérique. C'est pourquoi en simplifiant le problème, il est possible d'arriver à diminuer le temps de calcul. C'est la base des proxys. Il en existe de très nombreuses versions, de la plus simple à la plus complexe selon le problème à résoudre. Comme nous l'avons vu dans le chapitre précédent, certaines méthodes de sélection de modèles utilisent des simulations proxys. Nous allons tenter de présenter un large éventail de ce qui existe, puis nous détaillerons plus particulièrement une certaine catégorie de proxys basés sur la propagation de front, car ils ont été l'objet d'une partie du travail réalisé pendant cette thèse.

2.1 Les proxys scalaires

Les proxys les plus simples se contentent de caractériser les modèles par une grandeur scalaire (0D). Un des plus utilisés dans le domaine pétrolier est basé sur l'hydrocarbure en place,

c'est-à-dire que pour chaque modèle la quantité d'hydrocarbures présente dans chaque cellule est additionnée (HIP). Pour plus de réalisme, il est possible d'utiliser le **STOOIP** (pour *Stock Tank Original Oil In Place* ou huile originellement en place en condition de surface) car, les hydrocarbures formant des fluides compressibles, la différence de volume entre un fluide dans un réservoir sous plusieurs centaines de bars et le même fluide en surface à la pression atmosphérique est non négligeable. La limitation d'un tel proxy est évidente : il ne tient compte d'aucun paramètre dynamique excepté le facteur volumique de surface. Il est toutefois possible de lui appliquer un coefficient de récupération global à l'échelle du champ, mais cela reste extrêmement frustre. Or, les hydrocarbures présents dans le réservoir ne sont jamais récupérés en totalité, car certaines zones ne peuvent pas être drainées. Pour augmenter la précision de ce proxy lorsque la position des puits est connue, il peut être intéressant de ne calculer que la quantité d'huile uniquement présente dans une zone autour des puits dont la perméabilité est supérieure à une valeur seuil. C'est ce que nous appellerons le **volume connecté** aux puits (injecteurs ou producteurs). Ce proxy considère que seules les zones dont la perméabilité est suffisamment grande seront drainées au cours d'un temps de production raisonnable. Par contre, ce proxy ne fournit toujours qu'un résultat scalaire et les informations concernant les pressions et débits des puits ne sont pas intégrées.

2.2 Les proxys par propagations de front

Afin d'améliorer la qualité du proxy tout en gardant un temps de simulation très court, Alabert et Modot (1992) présentent une méthode très similaire au volume connecté mais permettant d'obtenir une courbe de réponse (2D). Pour cela, ils propagent un front à partir des puits producteurs. Ce front se propage cellule après cellule à partir des perforations du puits, avec pour seule contrainte que la cellule courante, ainsi que sa voisine où le front doit se propager, aient une perméabilité supérieure à une valeur seuil (Figure 13). Quand le front s'est propagé dans toute la zone possible, la quantité d'hydrocarbures est sommée par indice de front, ce qui donne une courbe. Bien que plus évolué qu'une simple grandeur scalaire, ce proxy peut encore être amélioré sans pour autant augmenter le temps de simulation. En effet, le front propagé par Alabert et Modot s'incrémente de 1 en 1 pour chaque cellule traversée, indépendamment des propriétés pétrophysiques, alors que celles-ci ont une influence sur la vitesse des fluides dans la cellule. De plus il y a un aspect binaire dans la propagation, le front se propageant au-dessus d'un seuil et pas en dessous. Le comportement des écoulements est différent, un front se propage vite dans un milieu très perméable et plus lentement dans un milieu peu perméable.

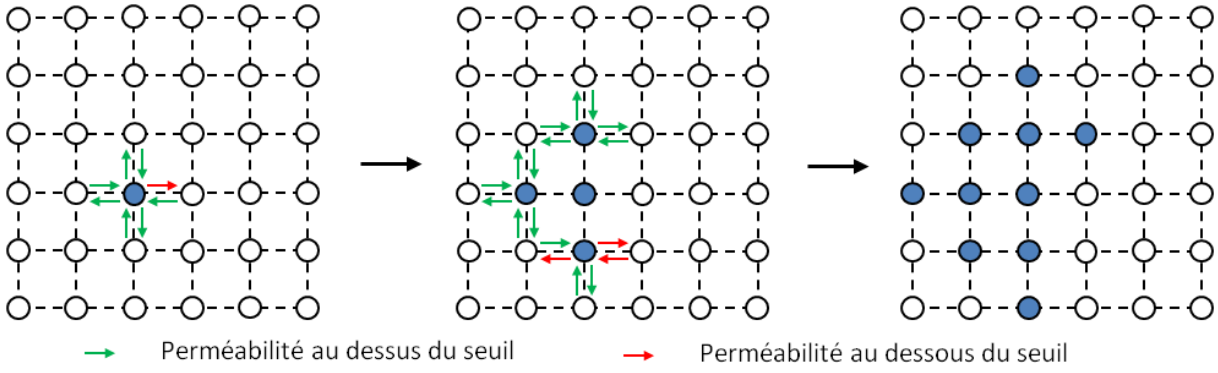

Figure 13 : Propagation d'un front de connectivité selon la méthode d'Alabert et Monod

Pour remédier à cela, il est possible d'utiliser des algorithmes de propagation de front plus complexes basés sur les méthodes de *Fast Marching* (Datta-Gupta et al., 2011, Hovadik and Larue, 2011). Les deux plus populaires ont été développés par Sethian (1999, 2001) et Tsitsiklis (1995).

2.2.a *Le* Fast Marching *isotrope*

Nous allons présenter ici la version de base de Sethian. Elle a été développée initialement pour résoudre l'équation Eikonale, afin d'approcher la propagation d'ondes dans un milieu isotrope :

$$|\nabla T|F = 1 \tag{11}$$

où ∇T représente le temps d'arrivée du front et F la vitesse du front. Dans le cas d'une grille régulière à six voisins par cellule (Figure 14) et en utilisant une résolution à l'aide d'une méthode par différences finies, l'équation (11) peut alors s'écrire :

$$\begin{bmatrix} \max\left(D_{ijk}^{-x}T, 0\right)^2 + \min\left(D_{ijk}^{+x}T, 0\right)^2 \\ + \max\left(D_{ijk}^{-y}T, 0\right)^2 + \min\left(D_{ijk}^{+y}T, 0\right)^2 \\ + \max\left(D_{ijk}^{-z}T, 0\right)^2 + \min\left(D_{ijk}^{+z}T, 0\right)^2 \end{bmatrix}^{\frac{1}{2}} = \frac{1}{F_{ijk}} \tag{12}$$

Où :

$$D_{ijk}^{+x} = \frac{\left(\Psi_{i+1,j,k} - \Psi_{i,j,k}\right)}{\Delta x} \tag{13}$$

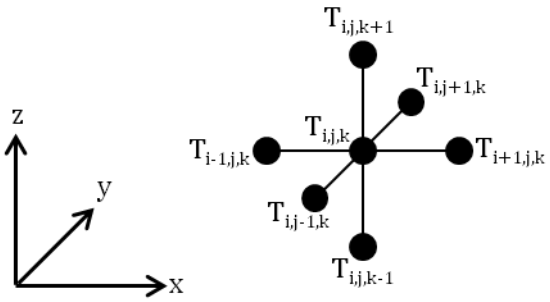

Figure 14 : Schéma grille à six voisins

Il est possible de faire l'analogie avec le domaine pétrolier en remplaçant le front d'ondes par un front de saturation fluide et le milieu isotrope par un réservoir dont la perméabilité moyenne est identique quelle que soit la direction. Nous allons donc chercher à déterminer le temps d'arrivée du front dans chaque cellule du réservoir. Pour cela, l'algorithme constitue trois piles de cellules :

- les cellules estimées (*FROZEN*), où le temps d'arrivée a été calculé ;
- les cellules en cours d'estimation (*TRIAL*), qui sont en train d'être calculées et dont l'ensemble constitue la *NARROW BAND* ;
- les cellules non traitées (*AWAY*), pour lesquelles aucun temps d'arrivée n'a encore été calculé.

Nous commençons par mettre les cellules qui contiennent les perforations des puits dans la pile *FROZEN* avec pour temps d'arrivée du front 0, les autres étant dans la pile *AWAY*. Le temps d'arrivée du front est ensuite calculé pour chacune des cellules voisines des puits, qui vont être mises dans la *NARROW BAND*. L'algorithme va ensuite itérer jusqu'à ce que toutes les cellules aient été traitées et se situent dans la pile *FROZEN* de la manière suivante (Figure 15):

- la cellule de la *NARROW BAND* qui a le plus petit temps d'arrivée du front passe dans la pile *FROZEN*.
- le temps d'arrivée du front dans ses cellules voisines est calculé (ou mis à jour) et les cellules qui n'appartiennent pas encore à la *NARROW BAND* sont ajoutées.

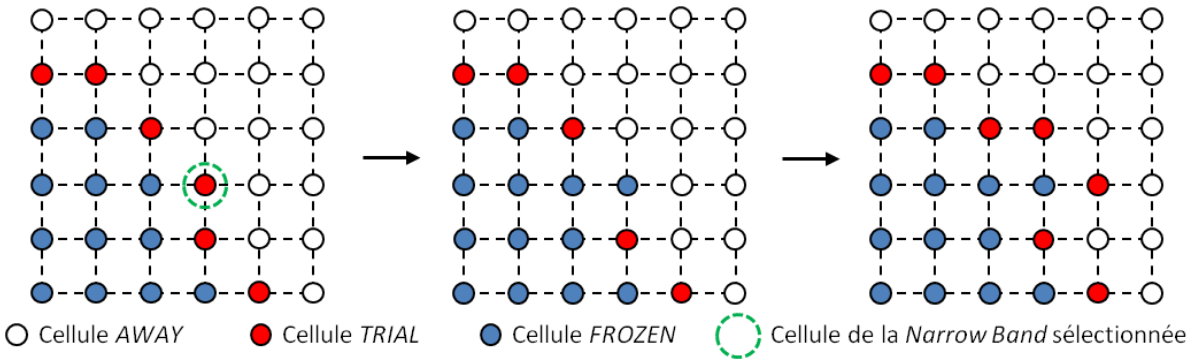

Figure 15 : Boucle de propagation générale d'un front à l'aide de l'algorithme de *Fast Marching*.

Pour obtenir des courbes de réponse, il suffit de discrétiser la gamme des valeurs des arrivées de front puis de sommer la quantité d'hydrocarbures pour l'ensemble des cellules dont le front est compris entre deux valeurs. Ce pas de front est assez difficile à choisir. Soit il est fixé manuellement et alors le nombre de pas dans la plage de valeur varie pour chaque courbe. Soit il est calculé de manière à obtenir un nombre voulu de pas. C'est cette deuxième méthode que nous utilisons.

L'avantage de cet algorithme est qu'il utilise une fonction de vitesse pour estimer le temps d'arrivée du front. Cette fonction peut englober différents paramètres pour la rendre plus ou moins complexe. C'est ce que proposent Hovadick et Larue (2011). Ils utilisent l'équation de la vitesse de Darcy d'un fluide en milieu poreux (14) qu'ils vont ensuite approximer de différentes manières : tout d'abord comme un rapport porosité sur perméabilité qui est la plus simple expression (15), puis ils ajoutent un gradient de pression (16) issu d'un seul pas de temps calculé à l'aide d'un simulateur à différences finies, et enfin l'effet des perméabilités relatives grâce à l'équation de Buckley-Leverett (Buckley and Leverett, 1942) :

$$F = -\frac{k}{\mu\Phi}(\nabla P + \rho g e_z) \tag{14}$$

$$F \approx cst \times \frac{k}{\Phi} \tag{15}$$

$$F \approx cst \times \frac{k}{\Phi} \times \nabla P \tag{16}$$

où k représente la perméabilité moyenne de la cellule, Φ la porosité, μ la viscosité de l'eau, ρ la masse volumique de l'eau, ∇P le gradient de pression, g l'accélération de la gravité et e_z le vecteur unitaire.

2.2.b *Le Fast Marching anisotrope*

Une des limitations principales de l'algorithme de *Fast Marching* de Sethian est qu'il ne fonctionne que dans un milieu isotrope, or ce n'est souvent pas le cas d'un milieu poreux tel qu'un réservoir pétrolier. En effet, il existe très souvent un facteur dix entre les perméabilités horizontale et verticale en raison de la stratification des dépôts sédimentaires. Plusieurs améliorations de cet algorithme présentes dans la littérature permettent de pallier cela (Alton and Mitchell, 2007, Sethian and Vladimirsky, 2001). Nous allons présenter ici une méthode basée sur l'algorithme de Tsitsiklis (Jbabdi et al., 2007). Cette version de l'algorithme n'utilise plus seulement les six voisins directs de chaque cellule mais les vingt-six présents tout autour (Figure 16a). Cet ensemble de points est ensuite triangulé pour former quarante-huit triangles (Figure 16b). Nous pouvons alors affirmer que pour arriver au point central x, le front va passer par un de ces triangles. Le temps d'arrivée du front au point x est donc approximé par la formule suivante :

$$t(x) = \underbrace{t(g)}_{(I)} + \underbrace{\frac{\|\overrightarrow{gx}\|}{v(x)}}_{(II)} \tag{17}$$

où g est un barycentre du triangle $\widehat{x_1x_2x_3}$ (Figure 16c), le terme (I) correspond au temps d'arrivée du front en g et le terme (II) au temps pour aller de g à x. $v(x)$ correspond au tenseur de vitesse du front en x et peut donc être anisotrope. Calculer le temps d'arrivée du front en x revient donc à déterminer le point g qui minimise ce temps. L'équation (17) peut donc s'écrire sous la forme :

$$t(x) = \min\left(\sum_{i=1}^{3} w_i t_{x_i} + \frac{\left\|x - \sum_{i=1}^{3} w_i x_i\right\|}{v(x)}\right) \tag{18}$$

Nous cherchons ainsi à déterminer les w_i qui sont les poids du barycentre g tels que : $\sum_i w_i = 1$ et $w_i \in [0\ 1]$.

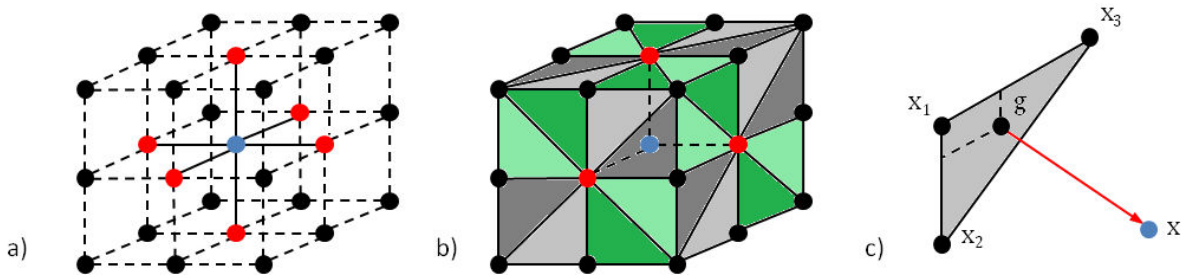

Figure 16 : Fast Marching anisotrope. (a) Les vingt-six voisins pris en compte. (b) Triangulation des faces. (c) Utilisation du centre de gravité d'une face pour calculer le temps d'arrivée du front.

Le fonctionnement de l'algorithme est le même que celui du *Fast Marching* isotrope de Sethian avec trois piles de cellules. Bien que plus complexe que la version isotrope, cet algorithme de *Fast Marching* reste assez simple à mettre en œuvre et rapide à exécuter car seule l'équation du calcul de l'arrivé du front est différente.

2.3 Les proxys plus complexes

Des proxys beaucoup plus complexes ont aussi été développés pour approcher de manière plus précise la simulation d'écoulement. Les simulateurs par lignes de courant ou l'utilisation d'un changement d'échelle peuvent notamment être cités.

2.3.a Les lignes de courant

Dans le premier cas, le but est de ne calculer l'évolution de la saturation que le long de lignes de courant, passant d'un problème 3D à un problème 1D pour cette variable (Batycky et al., 1997, Thiele et al., 1996, Thiele et al., 2010). Le calcul du champ de pression reste, lui, un problème 3D. Pour cela, il faut déterminer le déplacement du fluide le long de lignes de courant comme étant tangent au champ de vitesse. Dans le cas du déplacement d'un fluide en milieu poreux, cette dernière correspond à la vitesse de Darcy. L'évolution de la saturation est ensuite résolue le long de ces lignes de courant. Le gradient de pression est d'abord calculé à l'intérieur du réservoir grâce aux conditions aux limites (pression initiale, présence d'aquifère, hauteur du contact et pression dans les puits). Le gradient de pression donne alors le champ de vitesse qui permet de tracer les lignes de courant. Les propriétés dynamiques de la grille sont ensuite transférées aux lignes de courant. Il est ainsi possible de calculer l'évolution de la saturation le long de ces mêmes lignes, ce qui se rapporte à un problème unidimensionnel. Nous transférons alors dans l'autre sens la saturation vers la grille. S'il le faut, les phénomènes gravitaires ou de pression capillaire sont ensuite ajoutés (Blunt et al., 1996). Les avantages de cette méthode sont : qu'elle permet de résoudre la saturation dans un cas unidimensionnel au lieu de tridimensionnel, et que le gradient de pression n'a pas besoin d'être

recalculé à chaque pas de temps. En effet, tant que celui-ci reste valide, il n'est pas recalculé. Il est considéré invalide si les conditions aux limites changent (fermeture de puits ou ouverture de nouveaux) ou si l'évolution de la saturation est suffisamment significative pour nécessiter une mise à jour du champ de pression.

Pour les cas les plus complexes, tels que les transferts de chaleur, les écoulements tri-phasiques ou avec une forte compressibilité fluide, les proxys par lignes de courant restent difficilement applicables ou peu efficaces car le champ de pression doit être mis à jour très fréquemment.

2.3.b Le changement d'échelle

L'*upscaling*, au lieu de simplifier la physique de l'écoulement comme tous les précédents proxys, vise à simplifier le modèle, c'est-à-dire que nous allons diminuer le nombre de cellules dans le modèle afin de réduire le temps de simulation (Durlofsky, 2005, Farmer, 2001, Christie, 1996). Pour cela, les cellules sont agrégées pour en former de plus grosses (Figure 17). Les propriétés pétrophysiques du modèle sont alors homogénéisées ainsi que d'autres propriétés dynamiques, telles que les perméabilités relatives ou le skin factor des puits.

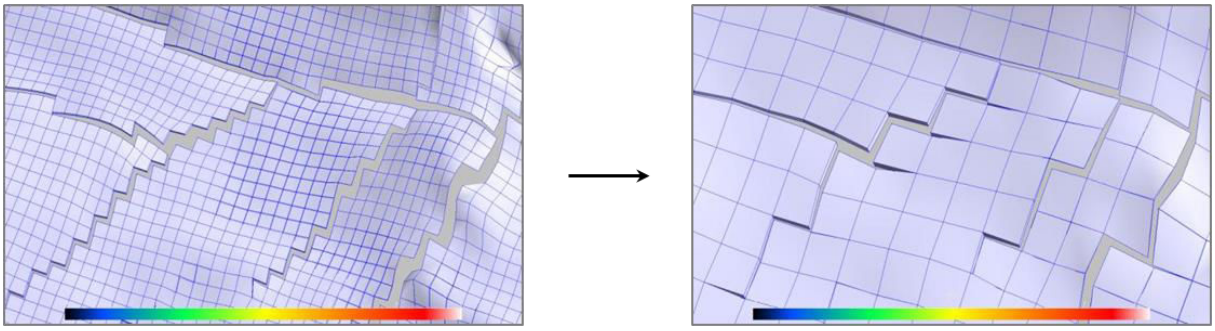

Figure 17 : Exemple d'*upscaling* sur une coupe de modèle géologique.

Différentes méthodes sont utilisées selon le type de propriété : pour la porosité et le Net-to-Gross (fraction de roche réservoir dans la cellule), par exemple, l'important est de conserver le volume initial d'huile en place dans la zone de drainage des puits afin de rester comparable. Dans ce cas, de simples moyennes arithmétiques pondérées suffisent. En revanche, pour la perméabilité, il faut cette fois respecter le comportement dynamique (i.e. si une cellule fait office de barrière à l'écoulement, la nouvelle cellule qui l'englobe doit conserver ce caractère). Le modèle ainsi obtenu par homogénéisation est simulé dynamiquement et, comme il contient moins de cellules et que celles-ci sont plus grosses, le temps de calcul est réduit et la convergence facilitée.

Les deux derniers types de proxys que nous avons décrits sont plus précis et permettent d'obtenir plus d'informations, mais sont aussi plus difficiles à implémenter car ils nécessitent quasiment tous les paramètres requis pour une simulation dynamique classique. Or ces paramètres sont très spécifiques et peuvent provenir d'études en laboratoire. Cela exclue toute utilisation du proxy dans un cadre industriel par des personnes ne maîtrisant pas ce genre de données.

Ces dernières années de nouveaux proxys ont vu le jour, à l'intérieur desquels la physique des écoulements est remplacé par des méthodes numériques de data mining. Ces proxys ne s'appliquent qu'à des champs matures car ils ont besoin de données d'historique. Il est par contre difficile de trouver parmi la littérature scientifique la présentation des méthodes utilisées[1].

[1] Exemples : http://www.foroil.com/, http://www.intelligentsolutionsinc.com/

Chapitre 3 : Comparaison des proxys

Dans ce chapitre, nous allons présenter notre méthodologie de comparaison de proxys afin d'être capable de juger de la qualité d'un proxy dans l'étude des incertitudes dynamiques. Cette méthodologie est basée sur trois points : le respect des distances dynamiques, la facilité de mise en œuvre et la stabilité lors d'une sélection de modèles.

Comme nous venons de le voir, il existe de très nombreux proxys ayant chacun leurs avantages et leurs inconvénients. Tous ces proxys ont été développés dans un but particulier. La simplification opérée se focalise souvent sur quelques critères spécifiques (le STOOIP ne conserve que l'apport des réserves à la production d'huile). C'est pourquoi certains proxys ne « voient » pas la même chose, mais peuvent être détournés s'ils fournissent l'information nécessaire. Dans notre cas, nous effectuons une sélection de modèles. Plusieurs questions se posent alors pour choisir un bon proxy :

- Le proxy permet-il de reproduire la distance entre les réponses au simulateur d'écoulements des modèles ?
- Le proxy est-il suffisamment rapide pour pouvoir être calculé sur tous les modèles disponibles ?
- Possédons-nous toutes les informations pour pouvoir effectuer les simulations proxy ?
- La sélection de modèles obtenue à l'aide du proxy est-elle précise et stable ?

Répondre à ces questions est important car le but d'un proxy est de gagner du temps et il n'est pas toujours possible de pouvoir en tester plusieurs. De même, si les résultats du proxy ne sont pas bons, les conclusions qui en seront tirées peuvent être fausses, ce qui peut engendrer de graves conséquences (surestimation de la rentabilité, mauvaise implantation des puits…). Afin de pouvoir répondre rapidement à ces questions et ainsi aider à prendre une décision, nous proposons une méthode basée sur plusieurs critères. Cette méthode peut être utilisée en contexte de recherche ou d'évaluation méthodologique (pour dire si un nouveau proxy développé améliore les résultats), mais aussi dans un contexte opérationnel (avec un nombre d'opérations restreint et des limites de temps imparties). Cette méthode, publiée lors de la conférence ECMOR XIV (Bardy et al., 2014), utilise plusieurs critères basés sur trois caractéristiques distinctes :

- la cohérence entre les distances calculées entre les réponses proxys et les distances calculées entre les réponses du simulateur d'écoulements à haute fidélité ;
- la facilité de mise en œuvre du proxy et son gain de temps ;
- la stabilité du proxy vis-à-vis du calcul de quantiles de production.

3.1 Critères basés sur la distance entre les modèles

Cette première caractéristique, qui permet d'étudier la cohérence du proxy, est basée sur les distances entre modèles. La notion de distance permet en effet de quantifier la similitude entre les comportements de deux modèles vis-à-vis d'une variable. Dans notre cas, la variable utilisée est la réponse dynamique du proxy et plus particulièrement la production cumulée d'huile à l'échelle du champ : FOPT. Calculer une distance revient donc à calculer une différence de réponses au simulateur d'écoulements dynamiques. Plusieurs raisons orientent à choisir cette variable. Tout d'abord, il s'agit de la production d'huile qui est le premier critère économique d'un réservoir pétrolier. Ensuite, elle correspond à la somme des productions de chaque puits ce qui rend le signal plus stable car, même si un puits ne produit pas, la production totale peut continuer à augmenter.

Enfin, il s'agit d'une production cumulée ce qui stabilise encore un peu le signal car celui-ci ne peut que croître. Tout cela donne donc plus de sens au calcul d'une distance. Il existe plusieurs moyens de calculer une distance. L'important est que cette dernière respecte les trois principes donnés précédemment : séparation, symétrie et inégalité triangulaire.

Dans notre cas, nous utilisons une norme L2 entre deux courbes de FOPT. :

$$d_{A\to B} = \sqrt{\sum_{i=0}^{N} \left(FOPT_A(i) - FOPT_B(i)\right)^2} \tag{19}$$

avec N le nombre de pas de temps dans la simulation. Ces courbes sont continues et strictement croissantes sans valeurs aberrantes, l'utilisation de la norme L2 est donc adéquate. De plus dans la plupart des cas de simulations dynamiques, les profils utilisent tous les mêmes pas de temps (paramétrage de la simulation).

Pour des proxys comme les lignes de courant ou l'homogénéisation, qui fournissent le même type de résultat (courbes), nous utiliserons la même distance (norme L2). Pour les proxys de propagations de front, les courbes de réponses peuvent avoir une longueur différente car, le front se propageant tant que les cellules ont une perméabilité supérieure au seuil, plus la courbe est longue et plus il a fallu traverser de cellules ou plus le front a été lent. Plusieurs solutions sont possibles pour prendre cela en compte lors du calcul de distance :

- Prolonger la courbe la plus courte de sa dernière valeur pour chaque pas de temps manquant (Figure 18a). Chaque courbe correspondant à une production d'huile cumulée, cela correspond à une production nulle lors des pas de temps manquants. Par contre, cette méthode fait perdre l'information amenée par la longueur de la courbe en ajoutant de la « fausse » information.
- Arrêter le calcul quand une des courbes est finie, ce qui revient à tronquer la courbe la plus longue (Figure 18b). Cette fois encore, l'information apportée par la différence de longueur n'est pas prise en compte. Par contre, à la place d'ajouter de la « fausse » information, une partie est perdue.
- Normaliser les courbes entre 0 et 1 ou entre 0 et le temps final des simulations de références, afin d'obtenir des courbes de même longueur (Figure 18c). Mais dans ce cas les distances calculées n'ont plus vraiment de sens car elles ne correspondent plus au même front ou pseudo-temps.
- Ajouter une composante au calcul de la distance, correspondant à la différence de longueur entre les courbes de réponses (Figure 18**Error! Reference source not found.**d). Cela peut se faire en utilisant n'importe laquelle des trois techniques précédentes comme première composante, puis la différence de longueur comme deuxième composante, chacune d'elle étant pondérée par un poids.

Le choix de poids à attribuer à la différence de longueur entre deux courbes n'ayant pas été traité dans cette thèse, nous avons choisi la deuxième possibilité car cela correspond aux prévisions à court terme. En effet celles-ci sont les plus importantes, car plus elles sont grandes plus elles permettent d'amortir rapidement l'investissement. Pour les proxys plus simples (STOOIP par exemple), un simple écart en valeur absolue est utilisé car les proxys ne fournissent pas de courbes.

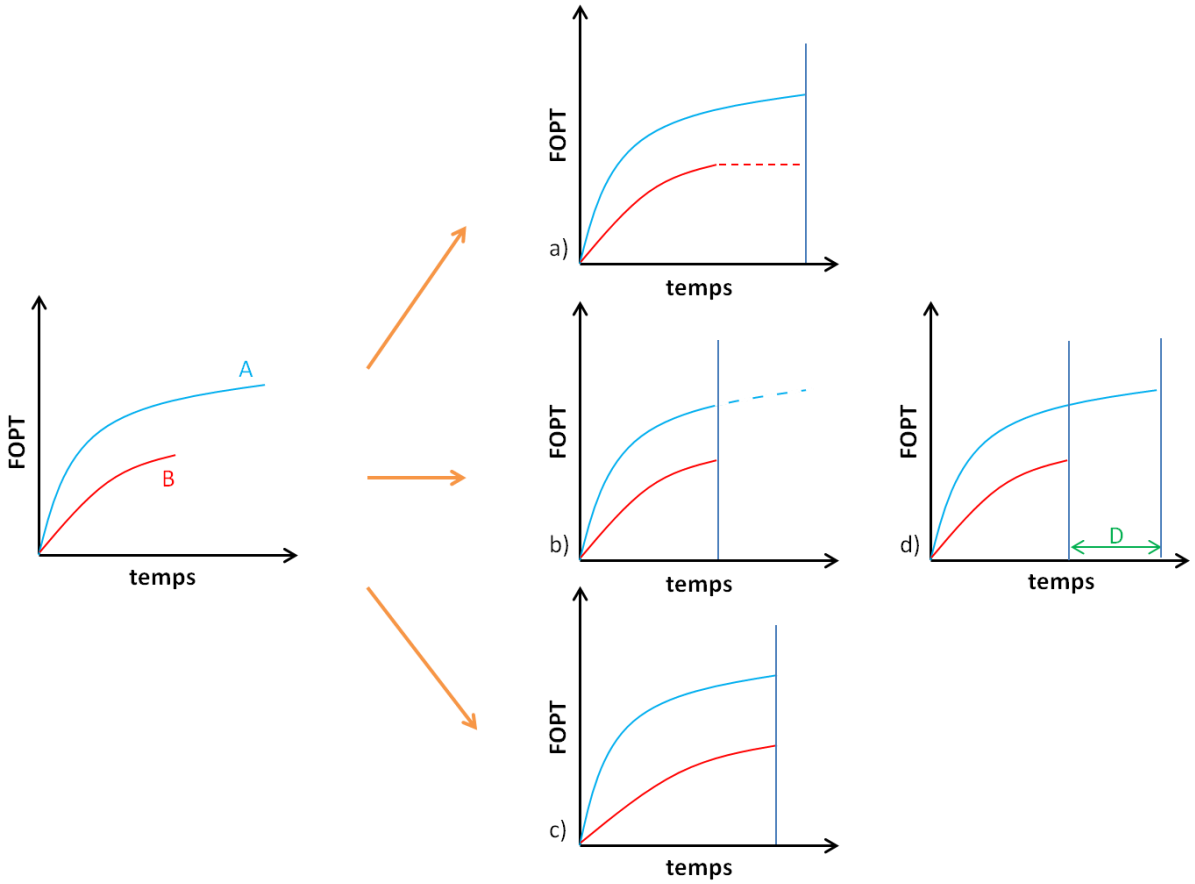

Figure 18 : Technique de calcul de distance entre deux réponses de durées différentes. (a) la courbe la plus courte est prolongée. (b) la courbe la plus longue est raccourcie. (c) les courbes sont normalisées. (d) on ajoute un poids avec la différence de longueur.

Une fois les distances calculées entre les modèles, vis-à-vis des réponses dynamiques et des réponses proxys, nous construisons un graphique avec ces distances dans l'espace distances proxys/distances dynamiques (Figure 19a). Pour un cas où nous disposons de N modèles distincts, K distances différentes non nulles sont simplement obtenues par :

$$K = \frac{\left(N(N-1)\right)}{2} \tag{20}$$

Ce graphique permet de visualiser l'existence d'une relation entre les distances calculées sur la réponse proxy et les distances calculées sur la réponse dynamique. Nous pouvons ensuite calculer un **coefficient de corrélation** pour quantifier cette relation. C'est lui qui sera notre premier critère. Dans le cas d'une relation linéaire, le coefficient de corrélation de Pearson est estimé :

$$\widehat{r_p} = \frac{(\widehat{\sigma_{XY}})}{\widehat{\sigma_X} * \widehat{\sigma_Y}} \tag{21}$$

avec :

$$\widehat{\sigma_{XY}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) * (y_i - \bar{y}) \tag{22}$$

$$\widehat{\sigma_X} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{23}$$

$$\widehat{\sigma_Y} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (24)$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (25)$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (26)$$

qui sont respectivement les estimateurs de la covariance (22), des écarts-types (23 – 24) et des espérances (25 – 26) des deux variables X et Y. Ces dernières sont ici respectivement la distance calculée sur les réponses proxys et la distance calculée sur les réponses dynamiques. Afin de pouvoir aussi quantifier les relations non linéaires entre les distances, il est préférable de présenter le coefficient de corrélation de Spearman ou coefficient de corrélation de rang, qui s'écrit :

$$r_s = 1 - \frac{6\sum_{i=1}^{n}\big(R(x_i) - R(y_i)\big)^2}{n(n^2-1)} \quad (27)$$

avec R le rang.

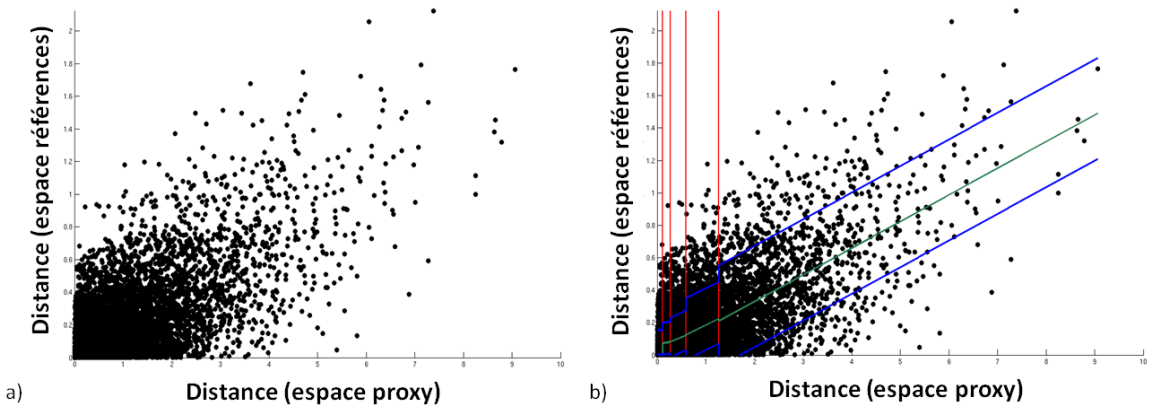

Figure 19 : (a) Graphique des distances calculées sur les réponses dynamiques en fonction des distances calculées sur les réponses proxys. (b) Régression linéaire par morceaux dans l'espace des distances. Les lignes rouges représentent la délimitation des zones, les lignes vertes la régression linéaire et les lignes bleues les quantiles Q10 et Q90 des résidus autour de la régression.

En plus du coefficient de corrélation, il est important d'analyser la dispersion des valeurs. En effet, de petites distances indiquent que les réponses des deux modèles sont semblables ou inversement de grandes distances indiquent des comportements différents. Cela doit se traduire par une faible distance proxy quand la distance dynamique est petite et vice versa. Pour caractériser ce comportement d'un proxy, nous subdivisons le graphique des distances en plusieurs zones comportant approximativement le même nombre de points et à l'intérieur desquelles la corrélation est la plus linéaire possible. Le nombre de zones obtenues est donc fonction de ces deux contraintes. Il doit aussi tenir compte du fait que nous effectuons une régression linéaire dans chaque zone, le nombre de points à l'intérieur de chacune d'elles doit donc être statistiquement représentatif. Cette régression par morceaux permet de calculer les résidus (différence entre l'ordonnée du point et l'ordonnée du point de même abscisse sur la droite de régression) de chaque point avec la droite de régression linéaire de la zone dans laquelle il est situé (Figure 19b). Cette régression est un peu particulière dans la première zone (la plus près de l'axe des ordonnées), car nous forçons le passage par l'origine (une distance nulle signifie des modèles identiques). Les **distributions des résidus** ainsi obtenues dans chaque zone sont une autre façon de caractériser les proxys et forment un deuxième

critère de comparaison. Plus elles sont centrées sur 0 avec un faible écart-type et plus le comportement du proxy est capable d'expliquer celui de la simulation dynamique. Du fait que chaque distribution est calculée indépendamment, il n'y a pas de transition d'une zone à l'autre. La Figure 20 montre une autre représentation de la dispersion des résidus. Nous pouvons ainsi observer un biais des résidus qui se résorbe uniquement à partir de la dernière zone. Cela traduit que, pour de petites distances entre les modèles observés sur les réponses proxys (modèles semblables), il est possible d'obtenir de plus fortes variations en regardant les distances réelles en dynamique. Le proxy ne représente donc pas correctement le comportement dynamique des modèles.

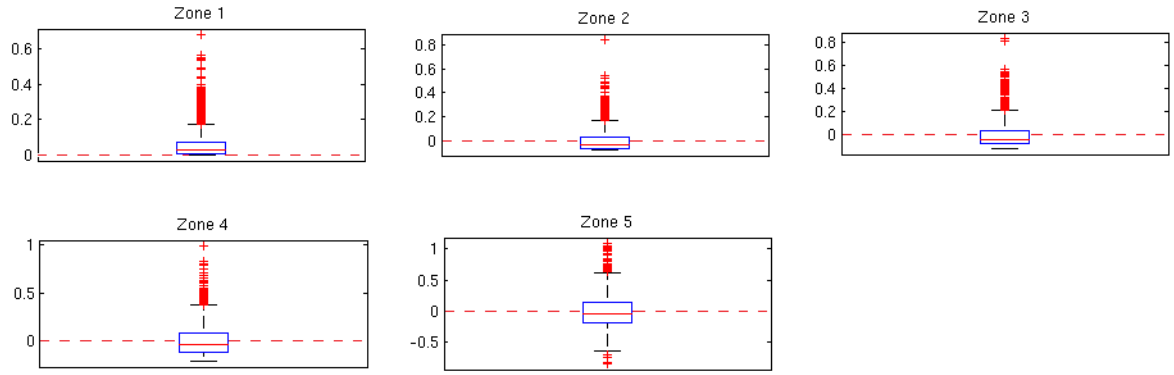

Figure 20 : Boxplots représentant la dispersion des résidus autour de la régression linéaire dans chaque zone.

En plus de ces deux critères quantifiables, il peut être intéressant de regarder les deux autres zones suivantes des crossplots :

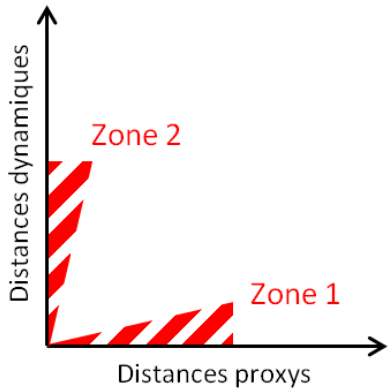

Figure 21 : Autres zones d'intérêt sur les crossplots.

En effet cette part du graphique concerne les très faibles distances entre réponses, ce qui signifie que les modèles ont un comportement similaire. Si la dispersion est grande autour de l'axe des abscisses (Figure 21 Zone 1), cela indique que notre proxy peut faire apparaître une forte différence de comportement alors qu'il n'y en a pas. Inversement (Figure 21 Zone 2), une grande dispersion autour de l'axe des ordonnées signifie que le proxy n'est pas capable de différencier des modèles qui n'ont pas la même réponse dynamique.

Tous ces critères de comparaison nécessitent de connaître les réponses dynamiques des modèles. Cependant, comme nous l'avons vu grâce à l'équation (20), avec peu de modèles nous obtenons de nombreuses distances. Ces critères peuvent donc déjà donner un bon aperçu de la relation entre les distances proxy et du simulateur d'écoulement dynamique, qu'il s'agisse d'un contexte recherche ou opérationnel.

3.2 Critères basés sur l'accessibilité et le gain de temps

Après des critères portants sur la fiabilité du proxy à reproduire la distance entre modèles, nous proposons d'utiliser des critères plus qualitatifs : **la facilité à mettre en place le proxy** et **le gain de temps** d'une simulation proxy par rapport à une simulation dynamique totale. Le gain de temps est très facile à mesurer, il s'agit du ratio du temps moyen d'une simulation dynamique par rapport à celui d'une simulation proxy :

$$g = \frac{t_{simulation\ dynamique}}{t_{simulation\ proxy}} \tag{28}$$

Le ratio doit être supérieur à 1 pour qu'il soit possible de parler de gain de temps (et donc de proxy), mais il peut être très variable d'un proxy à l'autre. La prise en compte seulement du temps CPU n'est pas suffisante. En effet certains proxys sont très simples à utiliser, comme les simulations de front de propagation par exemple. Dans ce cas, il suffit de sélectionner les données d'entrée puis de lancer la simulation proxy. Par contre, pour d'autres proxys il faut effectuer un travail de pré-processing qui peut être long. Ainsi, pour le proxy basé sur les lignes de courant, il faut convertir le paramétrage utilisé pour les simulations de référence en un paramétrage utilisable par le simulateur à lignes de courant. Par exemple, il faut redéfinir le fonctionnement des puits (le fonctionnement de chaque puits doit être explicité et défini en fonction d'une pression ou d'un débit, et le temps de simulation se définit en volume injecté). Pour le proxy basé sur l'*upscaling*, il faut reconstruire de nouveaux modèles. Cela implique des choix :

- il faut définir le ratio d'agrégation des mailles du modèle, c'est-à-dire le nombre de cellules regroupées ensemble dans chaque direction. Celui-ci dépend de la géologie du réservoir (si notre modèle comporte plusieurs barrières à l'écoulement relativement rapprochées, celles-ci doivent se retrouver dans le modèle grossier), de l'implantation des puits (il faut souvent respecter un espacement de trois cellules entre deux puits pour assurer la convergence numérique du simulateur) ou bien des contacts (le volume initial de chaque fluide présent doit être conservé) ;
- il faut choisir les algorithmes d'homogénéisation à utiliser pour les différentes propriétés pétrophysiques du modèle. Pour la porosité ou le net-to-gross, la conservation du volume grâce à un algorithme de volume moyen est souvent privilégiée. Pour les faciès, c'est un algorithme de valeur la plus probable qui est souvent utilisé. Par contre, pour des propriétés comme la perméabilité il est important de respecter le comportement du modèle vis-à-vis du simulateur, et il est ainsi préférable d'utiliser un algorithme basé sur un calcul de pression.

Ces différentes étapes préliminaires, bien que mises en place sur quelques modèles puis appliquées à l'ensemble des modèles à notre disposition, sont longues car elles nécessitent l'intervention d'un utilisateur et de nombreux calculs qui peuvent avoir un coût CPU important et font que certains proxys ne peuvent être utilisés que par un utilisateur expérimenté.

Pour ce qui est de la facilité de mise en place du proxy, comme nous l'avons vu précédemment, certains sont plus simples que d'autres. Les proxys basés sur les volumes d'huile en place ou la propagation ne demandent pas d'informations spécifiques autres que les données initiales de pétrophysique (porosité, perméabilité...) et la position des puits. Par contre, les proxys basés sur les lignes de courant requièrent des propriétés dynamiques telles que la compressibilité et la mobilité des différents fluides, les courbes de saturation et des données initiales de pression ou de flux pour les puits et les bords du modèle. En ce qui concerne les proxys basés sur l'*upscaling*, étant donné qu'une simulation dynamique complète est effectuée sur le modèle grossier, tous les paramètres sont nécessaires tels que le type d'aquifère utilisé, les tables PVT des fluides (Pression Volume Température) ainsi que le pilotage des puits (débits surface maximaux, groupes de contrôle).

Le tableau suivant présente la comparaison du proxy basé sur les hydrocarbures en place (HIP) avec les hydrocarbures en places connectés aux puits par une perméabilité supérieure à un seuil pour un même modèle :

Proxy	Paramètres	Réponses	Gain de temps	
			Mise en œuvre	simulation
HIP	Φ, NTG, GRV, S_w	HIP	0	> 1000
HIP connectés	*id.* + K, seuil, puits	HIP connecté aux puits	*id.*	*id.*

Tableau 1 : Paramètres, types de réponses et gain de temps par proxy. K correspond à la perméabilité, Φ à la porosité effective, NTG au *Net-to-gross*, GRV au Gross rock volume, S_w à la saturation initiale en eau et Bo au ratio volume de l'huile entre le fond du réservoir et la surface.

3.3 Critères basés sur la stabilité vis-à-vis du calcul de quantiles de production

Le but ici est de quantifier la qualité d'un proxy vis-à-vis de l'étude des impacts des paramètres statiques sur les incertitudes dynamiques. Comme nous l'avons expliqué en début de partie, ce sont les quantiles de production qui sont utilisés comme représentation de cette incertitude et sont calculés sur les profils de production. Classiquement, ce sont les quantiles Q10, Q50 et Q90, qui donnent des informations sur le comportement dynamique du réservoir dans les cas respectivement pessimiste, de base et optimiste, qui sont utilisés. Dans la pratique il est impossible de calculer ces quantiles sur l'ensemble des modèles, car il n'est pas faisable de tous les simuler dynamiquement. Une méthode de sélection des modèles est donc nécessaire afin de réduire le nombre de simulations tout en essayant de conserver la représentativité de l'ensemble initial. Dans notre cas nous utilisons la méthode de sélection de modèle DKM, usant d'une matrice des distances entre réponses proxys pour sélectionner quelques modèles représentatifs de l'ensemble à simuler (Scheidt and Caers, 2009). L'algorithme de regroupement de cette méthode se sert d'un processus stochastique à l'initialisation, ce qui signifie que si la fonction coût de l'algorithme ne converge pas suffisamment, en le répétant plusieurs fois, il est possible d'obtenir des résultats différents. Or la convergence de l'algorithme est induite par plusieurs facteurs dont la répartition des modèles dans l'espace. Cette répartition provient de la matrice des distances proxys utilisée, c'est pourquoi si les distances proxys ne correspondent pas à celles des réponses dynamiques, les modèles sélectionnés ne seront pas représentatifs et donc les quantiles de production calculés ne donneront pas de bonnes indications.

En répétant cette procédure, plusieurs quantiles de production différents peuvent être obtenus. Ces différentes réalisations nous permettent de calculer des intervalles de confiance sur les profils statistiques :

$$IC_{q_X}(t) = \left[q_{10}\left(q_{X_i}(t)\right) ; q_{90}\left(q_{X_i}(t)\right)\right] \quad i \in [1; N] \tag{29}$$

Deux critères peuvent en être extraits (Figure 22a) :

- la **largeur des intervalles de confiance** : plus les intervalles seront larges et moins le proxy sera précis ;
- la position des **vrais quantiles à l'intérieur des intervalles** : cela permet d'indiquer que le proxy n'est pas biaisé. Cette caractéristique n'est possible qu'en contexte de recherche si l'on veut évaluer un nouveau proxy. En effet, elle nécessite de posséder l'ensemble des simulations de références.

Une vérification supplémentaire peut être réalisée en effectuant une sélection aléatoire, c'est-à-dire en choisissant les modèles représentatifs au hasard. Cela donne les intervalles de confiance les moins précis ; le proxy doit fournir des intervalles plus étroits, sinon le proxy ainsi que la distance utilisée seront considérés comme non informatifs pour sélectionner des modèles représentatifs. La Figure 22b présente des intervalles obtenus par sélection aléatoire pour un ensemble de modèles. Nous remarquons que comparés à une sélection basée sur les distances proxys (Figure 22a), les intervalles sont plus larges et se superposent. La position des vrais quantiles étant à l'intérieur de chaque intervalle respectif, le proxy est considéré comme valide vis-à-vis de notre problème, c'est-à-dire que sa réponse permet de distinguer suffisamment les modèles pour pouvoir en sélectionner un sous-ensemble représentatif.

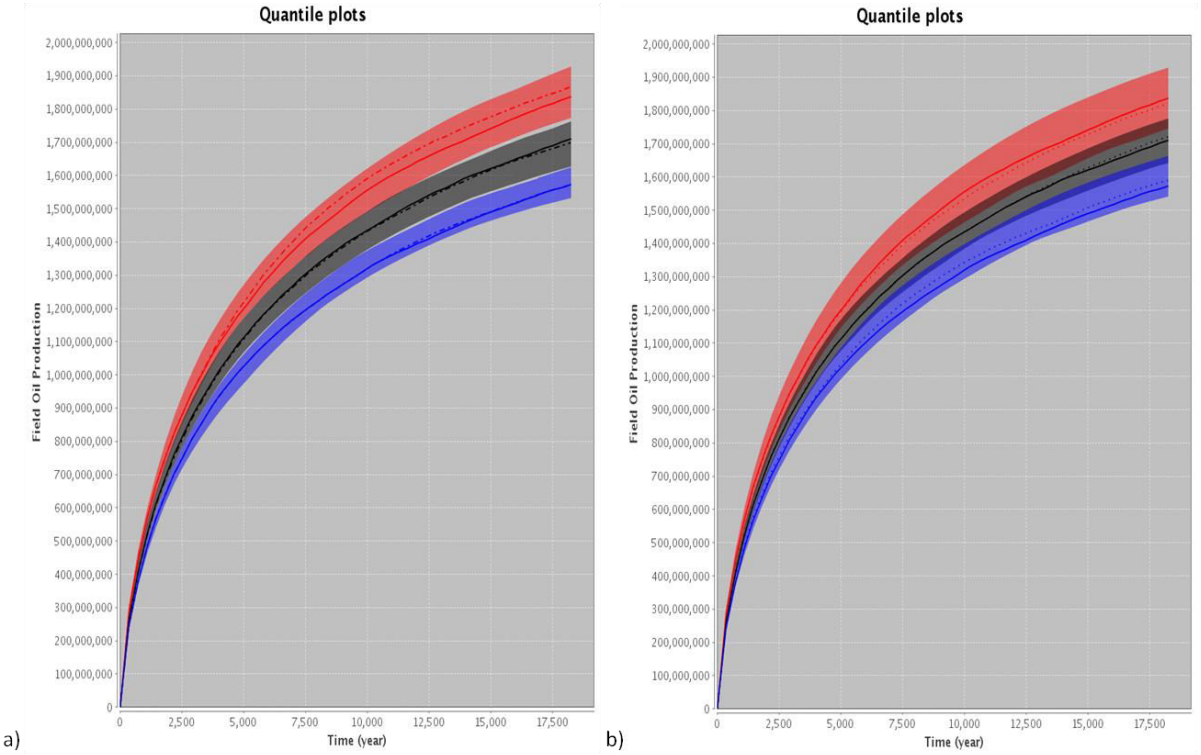

Figure 22 : Intervalles de confiance autour des quantiles de production Q10, Q50 et Q90. (a) Après sélection de modèles représentatifs avec DKM. (b) Après sélection aléatoire de modèles représentatifs.

Cette méthode de sélection, à force de répétitions et si le proxy n'est pas informatif, donne un très grand nombre de modèles différents à simuler voire l'ensemble des modèles disponibles. Elle n'est donc pas envisageable dans un contexte opérationnel, où le nombre de simulations d'écoulement est limité. Dans ce cas, il faudra se contenter des critères issus des deux premières caractéristiques à retenir pour une sélection de modèles. Elle permet toutefois de comparer les proxys de façon objective à des fins de recherche.

L'ensemble des critères proposés ci-dessus nous permet de comparer les proxys entre eux dans l'optique d'évaluer les incertitudes sur la production dynamique. Cela permet d'aider à choisir un proxy en faisant un bon compromis entre sa précision et les contraintes opérationnelles : gain de précision sur l'évaluation des incertitudes versus temps/facilité de mise en œuvre.

Chapitre 4 : Amélioration de la prise en compte des réponses proxys dans l'étude des incertitudes sur les profils de production dynamiques

Dans ce dernier chapitre, nous proposons une méthodologie pour reconstruire des profils de FOPT pour des modèles non simulés avec le simulateur d'écoulements dynamiques. Pour cela nous présenterons un modèle analytique à cinq paramètres que nous optimiserons grâce à un algorithme de minimisation, de manière à respecter les distances proxys connues.

Notations :

- $\boldsymbol{\mathcal{S}}$ correspond à l'ensemble des $\boldsymbol{n}$ modèles à notre disposition simulés grâce à un simulateur proxy.
- $\boldsymbol{\mathcal{S}_1}$ correspond à un sous-ensemble représentatif de $\mathcal{S}$ de taille $\boldsymbol{m}$ sélectionné par la méthode DKM et aussi simulé grâce au simulateur d'écoulements à haute fidélité.
- $\boldsymbol{\mathcal{S}_2}$ correspond au sous-ensemble des modèles restants, complémentaire de $\mathcal{S}_1$.

Ce chapitre répond à la deuxième question de la problématique de cette thèse : « Comment pouvons-nous améliorer l'évaluation des incertitudes sur les prévisions de production ? ». Nous proposons ici une nouvelle méthodologie qui, combinée à une sélection de modèles, détermine un profil analytique pour chacun des modèles non simulés afin de stabiliser et améliorer le calcul des quantiles de production. Contrairement à Josset et al. (2015) qui déterminent l'erreur permettant de passer de la réponse proxy à la réponse théorique du simulateur, nous proposons de directement déterminer les paramètres des profils analytiques de manière à respecter les distances entre modèles observées sur les réponses proxys.

Comme nous l'avons vu précédemment, si nous considérons un large ensemble de modèles $\mathcal{M}_i \in \mathcal{S}, i = [1, 2, \dots, n]$, il n'est pas possible de tous les simuler dynamiquement à cause du temps CPU prohibitif que cela impliquerait. Une solution classique consiste à ne sélectionner qu'un sous ensemble de modèles $\mathcal{S}_1 \subset \mathcal{S}$, plus petit et le plus représentatif possible, que nous allons alors simuler pour calculer les quantiles de production.

Afin de pallier les limites de cette sélection (cf. section 3.3), nous proposons de reconstruire les profils dynamiques de production d'huile des modèles restants $\mathcal{S}_2 = \mathcal{S} - \mathcal{S}_1$. Pour cela, nous allons commencer par déterminer les distances manquantes dans l'espace des simulations dynamiques à l'aide des distances proxys. Nous utiliserons ensuite un algorithme de minimisation pour déterminer les paramètres optimaux d'une fonction paramétrique, qui permet d'estimer les profils manquants (Figure 23). Cette méthodologie a été présentée dans un article soumis à *SPE Journal*.

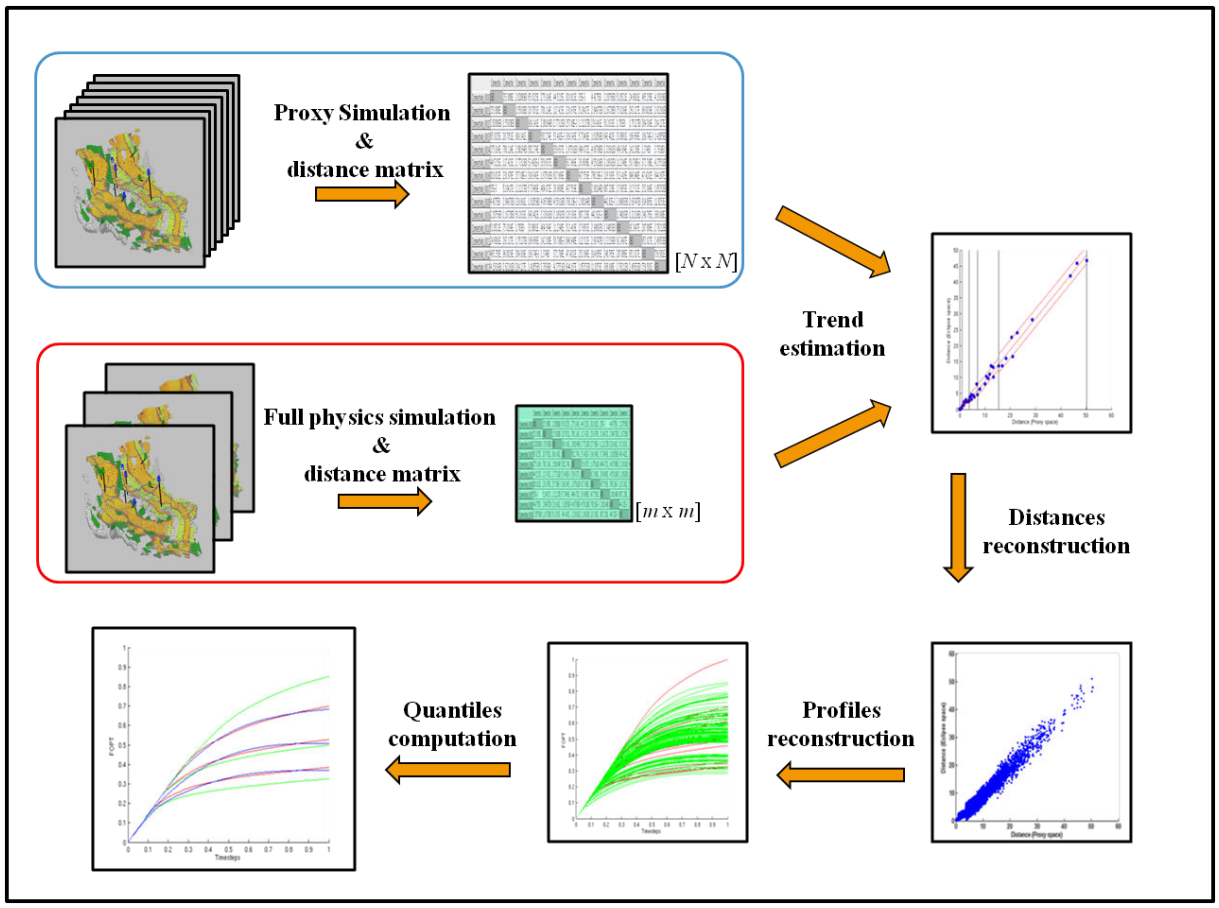

Figure 23 : Méthodologie proposée pour améliorer la précision du calcul de quantile de production

4.1 Relation entre les distances proxys et les distances vraies

La première étape de notre méthodologie consiste à déterminer les distances manquantes entre les modèles en termes de simulations dynamiques. Pour cela, nous allons utiliser le graphique présenté dans le chapitre précédent (3.1) qui montre la relation entre distances proxys et distances aux réponses du simulateur haute fidélité ($d_{vraie} = f(d_{proxy})$). A partir des m modèles sélectionnés, dont les réponses aux deux simulateurs (haute fidélité et proxys) sont connues, nous déterminons plusieurs zones dans le graphique. Dans chacune d'entre elles, nous calculons une régression linéaire ainsi que la loi de distribution des résidus (Figure 24). Nous allons ensuite utiliser ces lois pour estimer les distances dynamiques manquantes ($d_{vraie} = f(d_{proxy}) + r$).

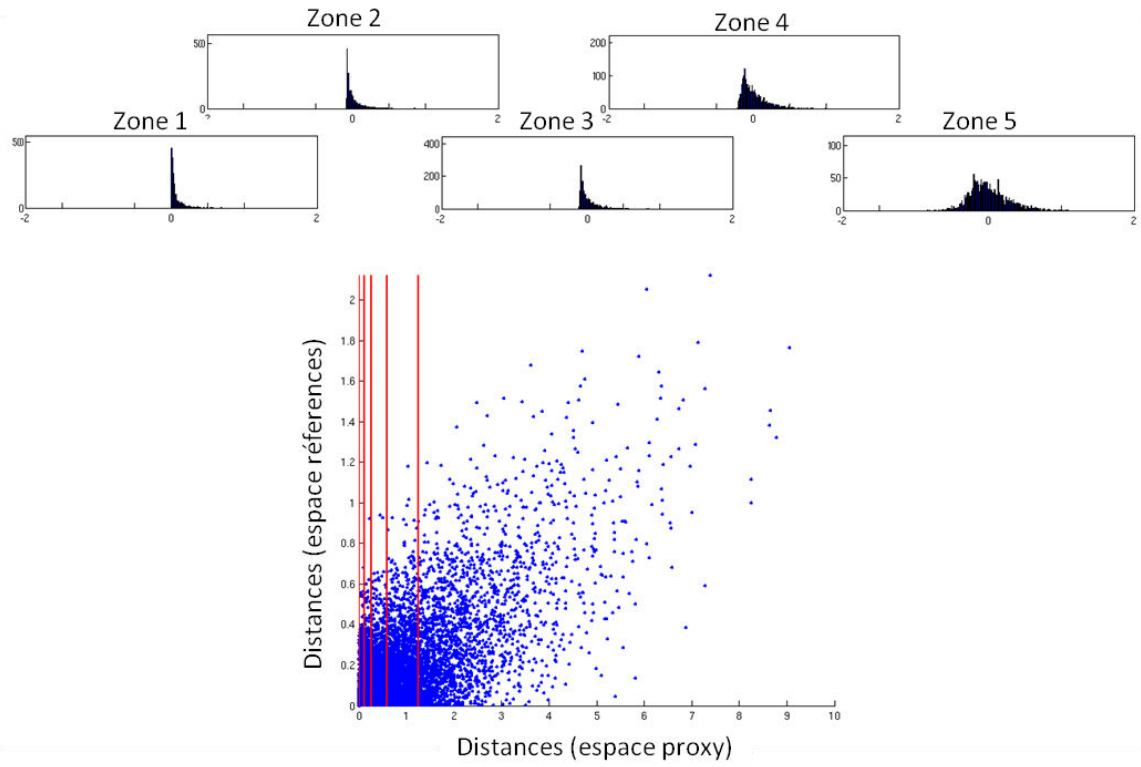

Figure 24 : Loi de distribution des résidus par zone de régression linéaire

Pour cela, chaque loi de distribution des résidus par zone est modélisée par une loi de distribution statistique :

- beta pour les zones proches de l'origine afin d'être sûr de ne pas obtenir de distance négative. En effet cette loi peut être bornée ;
- normale pour les zones à plus grandes valeurs car la probabilité d'obtenir des distances négatives est négligeable et ce genre de loi correspond exactement aux résidus d'une régression linéaire.

En déterminant dans quelle zone se situe la distance proxy, il est possible grâce à un tirage aléatoire du résidu de déterminer la distance dynamique correspondante. Ce processus permet de disposer de l'ensemble des distances entre modèles en termes de réponses à l'écoulement dynamique. Etant donné que le tirage s'effectue de manière aléatoire, il est possible d'obtenir différentes distances dynamiques pour une même distance proxy.

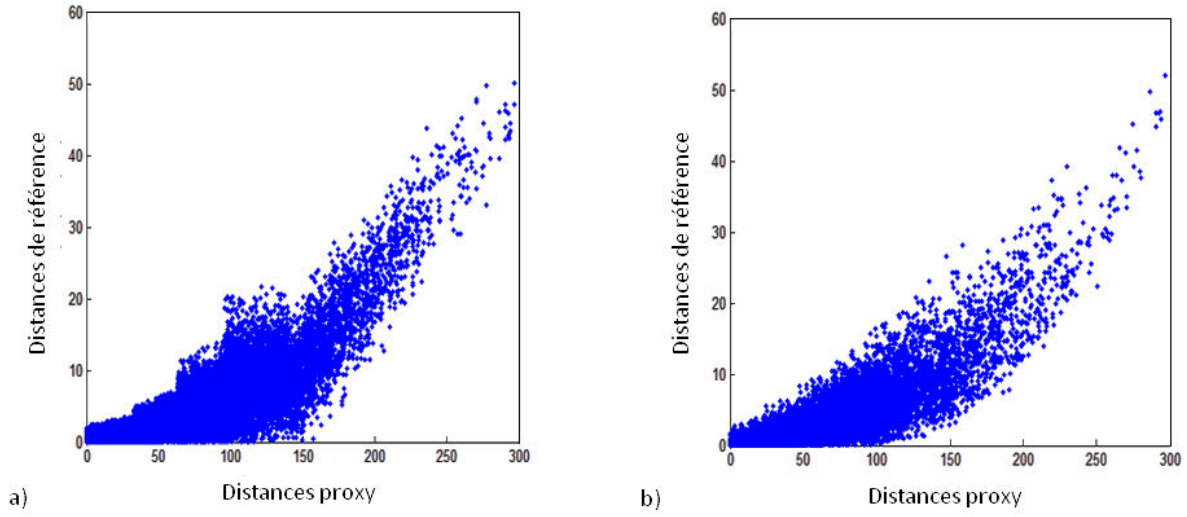

Figure 25 : Nuage de points distances-distances pour l'ensemble des modèles pour lesquels les distances de références proviennent (a) du tirage aléatoire proposé ou (b) du calcul direct sur les profils issus du simulateur haute fidélité (non connus en opérationnel)

Comme le montre avec un exemple de tirage la Figure 25, le nuage de points obtenus ressemble à celui que l'on a si l'ensemble des profils de référence est connu. Les différentes zones de la régression linéaire sont visibles mais ne gênent pas dans la reproduction des distances de référence. C'est pour cela que nous ne présenterons pas, dans ce manuscrit, d'amélioration avec un lissage de la distribution d'une zone à l'autre par exemple.

4.2 Approximation du profil de FOPT par une fonction paramétrique

Le but ici est de déterminer une fonction paramétrique qui pourrait correspondre à nos profils FOPT. La démarche envisagée est similaire à celle employée en exploration pétrolière, où un profil de production est estimé par une courbe analytique issue d'une analyse sur une base de données d'analogues (approche *type curves*). Pour cela, il faut d'abord connaître le contexte opérationnel utilisé pour notre réservoir. Dans notre cas, il s'agit de champs qui ne subiront pas de modification du plan de développement au cours du temps, c'est-à-dire que le nombre de puits restera le même tout au long de leur vie, que ceux-ci seront mis en service dès le début et qu'ils ne changeront pas d'affectation (un producteur ne deviendra pas injecteur par exemple). Cette hypothèse de travail correspond à ce que l'on appelle un champ *green field*, par opposition à un champ *brown field* pour lequel il existe des données de production et qui nécessite un calage d'historique. Il faut ensuite déterminer un comportement commun à partir des quelques profils disponibles (les résultats des simulations des modèles de $\mathcal{S}_1$).

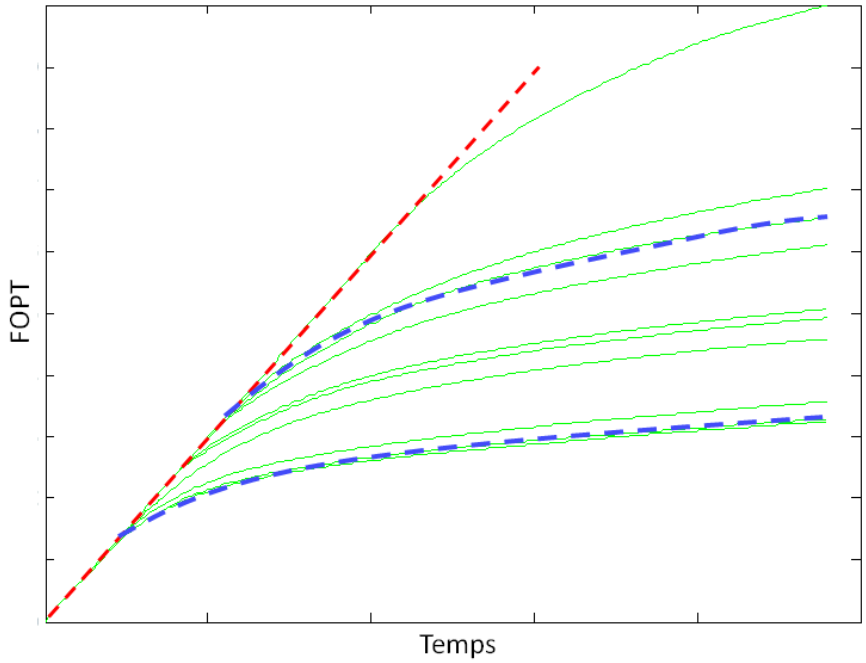

Figure 26 : Exemples de profils dynamiques issus du simulateur haute fidélité. Les lignes pointillées représentent le comportement général à reproduire.

La Figure 26 montre que tous ces profils ont la même forme : un comportement linéaire au début (ligne pointillée rouge) puis une décroissance continue (lignes pointillées bleues). Cela s'explique physiquement par le fait qu'au début de la vie du champ les puits produisent à un débit fixe (limité par les systèmes d'extraction, pompe et tuyaux en sortie et la capacité de traitement des infrastructures), avec une saturation en huile relativement constante à l'échelle du champ (Bourgeois and Couillens, 1994). Au bout d'un certain temps (appelé *water breakthrough*), des percées d'eau apparaissent au niveau des puits producteurs, faisant diminuer la quantité d'huile produite. Afin de respecter ce comportement, nous avons choisi une fonction paramétrique constituée de deux parties : une première linéaire puis une seconde représentée par une spline cubique d'Hermite (Figure 27a).

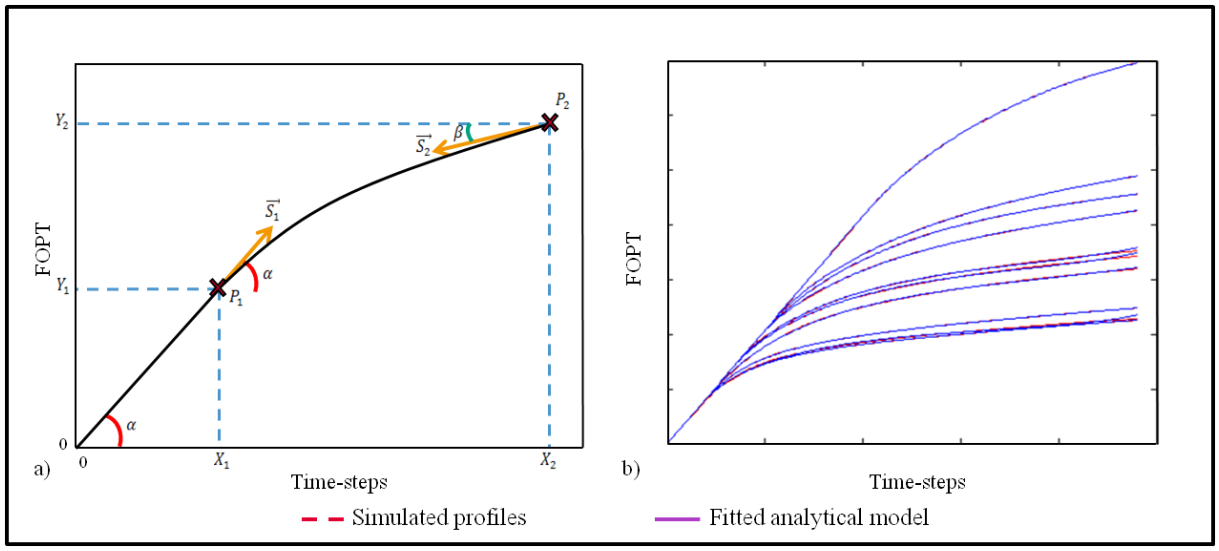

Figure 27 : (a) Modèle analytique pour les profils de production. (b) Calage des modèles sur les vrais profils de production.

La fonction est paramétrisée à l'aide de trois points caractéristiques : l'origine, le point P_1 (qui correspond à la transition entre la partie linéaire et la courbe de décroissance) et le point P_2 (qui correspond au dernier point de la courbe de décroissance). La partie linéaire relie le point $P_1 = \begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}$ à l'origine par une droite affine d'angle α :

$$Y_1 = \alpha X_1 \tag{30}$$

Le point P_2 est relié au point P_1 par une spline cubique d'Hermite, qui ne dépend que des points et des tangentes à chacune des extrémités (31).

$$\Gamma(u) = h_{00}(u)P_1 + h_{10}(u)\overrightarrow{S_1} + h_{01}(u)P_2 + h_{11}(u)\overrightarrow{S_2} \text{ pour } u \in [0,1] \tag{31}$$

$$h_{00}(u) = 2u^3 - 3u^2 + 1 \tag{32}$$

$$h_{10}(u) = u^3 - 2u^2 + u \tag{33}$$

$$h_{01}(u) = -2u^3 + 3u^2 \tag{34}$$

$$h_{11}(u) = u^3 - u^2 \tag{35}$$

Avec S_1 et S_2 les tangentes aux points respectifs P_1 et P_2, et h_{00}, h_{10}, h_{01} et h_{11} les fonctions d'Hermite. En utilisant les équations (30) et (31), notre modèle peut s'écrire :

$$C(t) = \begin{cases} \alpha t & \text{si } t < X_1 \\ \tilde{\Gamma}(t) & \text{si } t > X_1 \end{cases} \tag{36}$$

où $\tilde{\Gamma}(t)$ est l'expression de $\Gamma(u)$ dans l'espace cartésien, obtenu pratiquement en évaluant cent valeurs de u régulièrement espacées.

Les fonctions d'Hermite étant connues, ainsi que l'origine, les seules paramètres variables de notre modèle sont : P_1, P_2, $\overrightarrow{S_1}$ et $\overrightarrow{S_2}$. Dans un repère cartésien, ces variables possèdent chacune deux composantes et peuvent s'écrire comme suit:

$$P_1 = \begin{pmatrix} X_1 \\ \alpha X_1 \end{pmatrix} \tag{37}$$

$$P_2 = \begin{pmatrix} t_f \\ Y_2 \end{pmatrix} \tag{38}$$

$$\overrightarrow{S_1} = \begin{pmatrix} \|\overrightarrow{S_1}\| \cos(\alpha) \\ \|\overrightarrow{S_1}\| \sin(\alpha) \end{pmatrix} \tag{39}$$

$$\overrightarrow{S_2} = \begin{pmatrix} \|\overrightarrow{S_2}\| \cos(\beta) \\ \|\overrightarrow{S_2}\| \sin(\beta) \end{pmatrix} \tag{40}$$

t_f, le temps final de la simulation est considéré comme étant le même pour tous les profils et est donc connu. Il en est de même pour α l'angle initial des modèles (le même pour tous les profils), qui peut être déterminé à partir des profils réellement simulés de $\mathcal{S}_1$. Il ne reste pour notre modèle de profil analytique qu'un ensemble de cinq paramètres inconnus :

$$p = \begin{pmatrix} X_1 \\ Y_2 \\ \|\overrightarrow{S_1}\| \\ \|\overrightarrow{S_2}\| \\ \beta \end{pmatrix} \tag{41}$$

Afin de s'assurer que ce modèle peut être utilisé pour approximer nos profils dynamiques, nous effectuons un calage en déterminant un jeu de paramètres qui permet de représenter les

profils dynamiques de l'ensemble $\mathcal{S}_1$. La Figure 27b montre que la fonction paramétrique choisie (en bleu) se superpose parfaitement avec les vrais profils dynamiques (en rouge), à l'exception de certaines fins de profils. Ce petit écart, qui n'est présent que sur certains profils, provient de la méthode utilisée pour effectuer le calage qui nécessite de discrétiser la partie Spline en un ensemble de points dont le dernier peut ne pas correspondre exactement à un pas de temps du profil dynamique.

4.3 Algorithmes d'optimisation pour la détermination des paramètres des modèles mathématiques

Maintenant que nous avons défini la fonction paramétrique à utiliser, nous allons chercher à optimiser ses paramètres.

4.3.a Formulation de notre problème

Dans notre méthodologie, nous cherchons à déterminer les paramètres optimaux de telle façon que les distances calculées, dans l'espace euclidien des profils avec les modèles de $\mathcal{S}_1$, correspondent aux distances calculées grâce au tirage aléatoire des résidus.

(i) La fonction objectif

Le but dans cette partie est de construire une approximation de la courbe de réponse du modèle dynamique, à partir des données de distances pour les modèles géologiques sur lesquels seulement le proxy a été évalué ($\mathcal{S}_2$). Le calcul se fait en deux temps.

Il faut tout d'abord utiliser la régression présentée en 3.1 pour simuler les distances possibles $\widetilde{d}_\iota$ entre un modèle de $\mathcal{S}_2$, dont on veut reconstruire la courbe, et tous les modèles géologiques ($i \in [1, m]$) pour lesquels une courbe a été obtenue avec le modèle dynamique précis ($\mathcal{S}_1$).

Ensuite, il est possible d'exprimer les distances $\hat{d}_i(p)$ entre l'approximation analytique de la réponse dynamique (fonction paramétrique) et les réponses calculées avec le simulateur d'écoulement à haute fidélité, pour chaque géomodèle i appartenant à $\mathcal{S}_1$.

$$\widehat{d}_\iota(p) = \sqrt{\sum_{t=0}^{Nb_{timesteps}} \left(FOPT_i(t) - C(p,t)\right)^2} \tag{42}$$

avec $FOPT_i(t)$ représentant la réponse du modèle dynamique précis, t symbolisant les pas de temps, et $C(p,t)$ la fonction analytique permettant d'approximer une réponse dynamique.

Le problème à résoudre consiste alors à trouver le vecteur de paramètres p tel que les écarts entre ces deux distances soit minimum, simultanément pour tous les modèles appartenant à $\mathcal{S}_1$. Il faut donc minimiser la fonction objectif suivante :

$$fobj(p) = \sum_{i=0}^{m} \left(\widehat{d}_\iota(p) - \widetilde{d}_\iota\right)^2 \tag{43}$$

avec i l'index des différents modèles de $\mathcal{S}_1$.

(ii) Les contraintes

Il faut ensuite rajouter des contraintes pour garantir que notre solution correspondra bien à un profil cumulé d'huile. En effet, la spline cubique d'Hermite ne garantit pas d'obtenir un profil strictement croissant ni une décroissance constante (tangente qui diminue constamment). Cela correspond à une dérivée seconde du modèle strictement négative :

$$C(t)'' \leq 0 \tag{44}$$

De plus, pour accélérer le processus et faciliter la convergence de l'algorithme, nous rajoutons des bornes pour la recherche de nos paramètres :

$$g(t) = \begin{cases} 0 < X_1 < t_f \\ 0 < Y_2 < \alpha t_f \\ 0 < \beta < \alpha \\ 0 < \|\overrightarrow{S_1}\| \\ 0 < \|\overrightarrow{S_2}\| \end{cases} \quad (45)$$

Bien qu'il ne garantisse pas la convergence, ce domaine a toujours permis, lors de nos différentes études de cas, de trouver un ensemble de paramètres correspondant à un modèle analytique satisfaisant et dont le calcul de la fonction objectif est suffisamment petit.

4.3.b Choix de l'algorithme et implémentation

(i) Etat de l'art

La minimisation, et plus généralement l'optimisation mathématique, est employée dans de très nombreux domaines et a donc fait l'objet de nombreux travaux (Nocedal and Wright, 2006, Boyd and Lieven, 2004). En géologie, nous pouvons notamment citer la construction des surfaces des géomodèles (Mallet, 2014), les différents algorithmes de krigeage (Chilès and Delfiner, 1999) ou encore l'inversion sismique.

La forme générale d'un problème de minimisation s'écrit classiquement de la manière suivante :

$$\begin{cases} \min\limits_{x} f(x) \\ g(x) \leq 0 \\ h(x) = 0 \\ x \in \Re^n \end{cases} \quad (46)$$

où $f(x)$ représente la fonction à minimiser (fonction objectif), $g(x)$ les contraintes d'inégalité, $h(x)$ les contraintes d'égalité et x le vecteur de paramètres à minimiser. Dans la pratique nous distinguons les cas d'optimisation sans contrainte, qui utilisent des algorithmes comme la méthode de Newton, Levenberg-Marquardt, les régions de confiance ou Gauss-Newton, des cas d'optimisation sous contraintes utilisant le gradient réduit, le Lagrangien ou encore les conditions de Karush-Kuhn-Tucker (KKT). Tous ces algorithmes peuvent être regroupés en trois catégories différentes :

- Les algorithmes de descente de gradient
- Les algorithmes d'ordre 0 et évolutionnaires
- Les algorithmes hybrides

Pour **les algorithmes de descente de gradient**, il est nécessaire que la fonction objectif soit continue et deux fois dérivable. Notons ∇f le gradient de cette fonction et H sa matrice hessienne, la condition d'optimalité d'un problème sans contrainte peut alors s'écrire :

$$\begin{cases} \nabla f = 0 \\ H \text{ définie positive} \end{cases} \quad (47)$$

Le principe de tous ces algorithmes est de descendre le gradient à partir d'un point de départ jusqu'à arriver au minimum, en déterminant à chaque étape une direction de descente et un pas. La grande variété d'algorithmes s'explique par les différentes techniques employées pour ces étapes.

Dans le cas de la méthode de Newton, la direction de descente est déterminée grâce à une approximation quadratique locale en utilisant un développement en série de Taylor du second ordre :

$$f(x + \Delta x) \approx f(x) + (\nabla f)^T \Delta x + \frac{1}{2} \Delta x^T H \Delta x = \tilde{f}(\Delta x) \quad (48)$$

En appliquant la condition suffisante du premier ordre $\nabla \tilde{f}(\Delta x) = 0$, nous obtenons :

$$\nabla f + H\Delta x = 0 \Rightarrow \Delta x = -H^{-1}\nabla f \quad (49)$$

Cette méthode est idéale lorsque la fonction objectif est elle-même quadratique, mais peut poser des problèmes de convergence lorsque le point de départ se situe loin de l'optimum. Il est alors possible d'employer la méthode des régions de confiance, qui restreint la zone de validité de l'approximation quadratique et ainsi l'empêche de diverger. Une autre limitation de la méthode de Newton vient de l'utilisation de la matrice Hessienne. Quand celle-ci existe, mais est singulière ou n'est pas définie positive, il est possible d'utiliser la méthode de Levenberg-Marquardt qui modifie H comme suit :

$$\widetilde{H} = H + \beta I \quad (50)$$

en choisissant β de telle manière que H soit définie positive.

Lorsque la matrice Hessienne n'est pas disponible, les méthodes dites quasi-Newton qui approximent directement H^{-1} (souvent appelé B) à chaque pas de temps seront préférées. Dans le cas de l'algorithme BFGS (Broyden, Fletcher, Goldfarb, Shannon), B est mise à jour en utilisant l'équation suivante :

$$B_{k+1} = B_k + \frac{(\nabla f_{k+1} - \nabla f_k)(\nabla f_{k+1} - \nabla f_k)^T}{(\nabla f_{k+1} - \nabla f_k)^T \Delta x} - \frac{B_k \Delta x \Delta x^T B_k}{\Delta x^T B_k \Delta x} \quad (51)$$

Dans le cas où des contraintes sont présentes, l'approche la plus simple consiste à éliminer des variables grâce aux contraintes d'égalité puis de résoudre le problème comme s'il s'agissait d'un problème sans contrainte avec moins d'inconnues. Si ce n'est pas possible, il faut intégrer directement les contraintes sous la forme d'un Lagrangien :

$$L(x, \lambda, \mu) = f(x) + \lambda^T h(x) + \mu^T g(x) \quad (52)$$

où μ et λ sont les variables de Lagrange. La condition d'optimalité devient alors :

$$\frac{\partial L}{\partial x} = \frac{\partial f}{\partial x} + \sum \mu_i \frac{\partial g_i}{\partial x} + \sum \lambda_i \frac{\partial h_i}{\partial x} = 0 \quad (53)$$

Dans de nombreux cas d'optimisation, l'expression du gradient de la fonction objectif n'est pas disponible de façon explicite ou son coût de calcul est trop élevé. Les **algorithmes d'ordre 0 ou évolutionnaires** seront alors utilisés. La plupart de ces derniers ne nécessitent pas de gradient et fonctionnent selon la méthode essai/erreur. Nous pouvons notamment citer les algorithmes génétiques ou les stratégies d'évolution. Dans sa thèse, Do (2006) présente une revue bibliographique assez complète des différentes techniques existantes.

Les **algorithmes hybrides**, quant à eux, associent des algorithmes de descente de gradient et des techniques évolutionnaires afin de tirer parti des deux et ainsi réduire les contraintes. Par exemple, un algorithme à stratégie évolutionnaire peut être associé avec une descente de gradient. Le premier est employé au début pour effectuer une recherche à grande échelle, puis, quand la zone de recherche est suffisamment restreinte, une recherche avec descente de gradient est utilisée pour trouver la solution optimale plus rapidement.

(ii) Application

Le problème posé ci-dessus appartient à la classe de problème de minimisation non linéaire sous contraintes. Plusieurs méthodes de minimisation peuvent être employées mais, à cause des besoins industriels propres à cette thèse (utilisation d'algorithmes implémentés dans des logiciels précis), nous nous proposons d'utiliser un algorithme de type « point intérieur » où la matrice Hessienne et le gradient sont approchés numériquement. Ce type d'algorithme à fonctionnement itératif tire son nom de son fonctionnement qui tente d'approcher la solution à l'aide d'un point nécessairement à l'intérieur du domaine admissible. Nous serons donc sûrs que la solution fournie respectera nos contraintes.

Après avoir identifié le type de méthode à utiliser pour résoudre notre problème, il faut déterminer le point de départ de l'algorithme. S'il est bien choisi, cela permet de réduire considérablement le temps de calcul et le risque de divergence. Nous proposons donc d'employer, une fois de plus, les profils de production dynamiques de l'ensemble $\mathcal{S}_1$ à disposition. En effet, pour déterminer si la fonction paramétrique choisie est correct, nous avons effectué un calage sur les profils de l'ensemble $\mathcal{S}_1$. Cela permet d'obtenir un ensemble de paramètres acceptables. Le vecteur de paramètres est initialisé avec les paramètres du modèle, dont la distance $\tilde{d}$ avec le profil recherché est la plus petite.

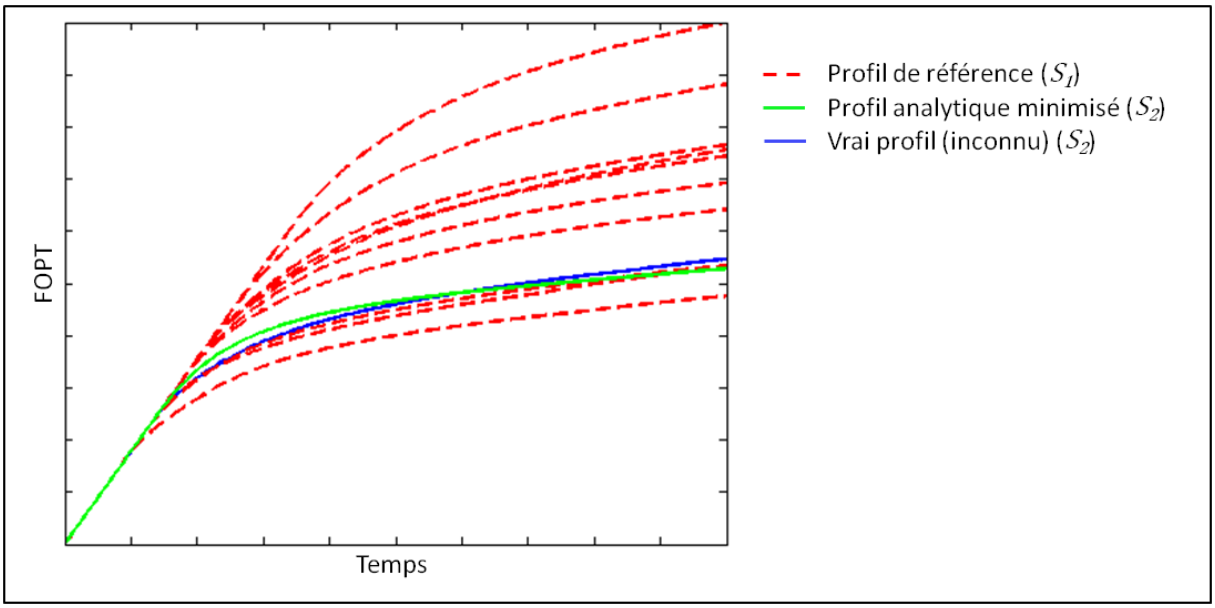

Figure 28 : Exemple de profil minimisé. Les courbes rouges, représentent les courbes de référence de $\mathcal{S}_1$ qui ont servis pour la minimisation, la courbe verte correspond au profil analytique résultat de la minimisation, et la courbe bleue au profil obtenu grâce au simulateur d'écoulements pour le même modèle.

Afin d'évaluer la qualité de cette méthode, nous avons simulé à l'aide du simulateur haute fidélité certains modèles de l'ensemble $\mathcal{S}_2$ pour pouvoir obtenir une courbe de référence et comparer les deux profils (Figure 28). On constate que la fonction paramétrique est très proche de la courbe de référence. Il correspond bien à un profil dynamique avec une partie constante au début semblable aux autres profils, puis une décroissance lente et constante. Le léger décalage toutefois constaté entre les deux courbes motive l'étape de correction de l'erreur, qui est décrite dans la section suivante.

4.4 Correction de l'erreur systématique

Pour s'assurer que la méthode de minimisation ne produit pas de biais systématique, une correction post-minimisation est ajoutée en construisant un modèle d'erreurs. Elle consiste à appliquer la méthodologie proposée sur les profils de référence afin d'obtenir un profil analytique à partir des $m-1$ autres profils dynamiques. Pour tous les modèles de $\mathcal{S}_1$, nous possédons donc un profil dynamique de référence et un profil analytique. Nous pouvons alors calculer la différence pour chaque pas de temps (Figure 29a).

$$\epsilon(t) = C(t) - FOPT(t) \tag{54}$$

Cette erreur est projetée sur les profils analytiques dans l'espace production d'huile en fonction du temps (Figure 29b). L'erreur dans le reste de l'espace est interpolée grâce à un krigeage unidimensionnel pour chaque pas de temps, ce qui donne une carte d'erreur (Figure 29c). Le krigeage choisi est un krigeage ordinaire car la moyenne des erreurs à chaque pas de temps n'est pas connue. Cette carte est ensuite utilisée pour corriger chaque profil.

$$C'(t) = C(t) - \epsilon(t) \quad (55)$$

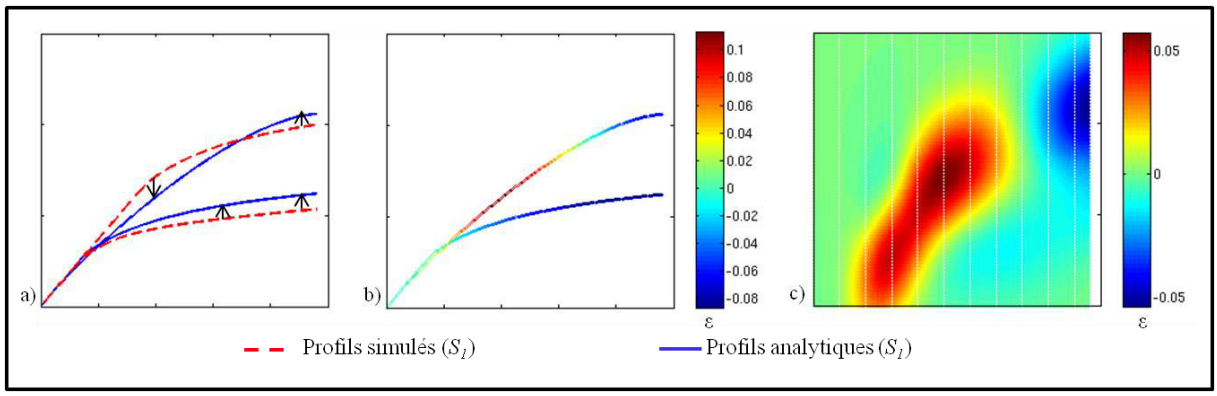

Figure 29 : Correction de l'erreur systématique des profils analytiques. (a) Calcul de la différence entre profil analytique et profil vrai. (b) Projection de l'erreur sur les profils analytiques. (c) Interpolation de l'erreur dans le reste de l'espace.

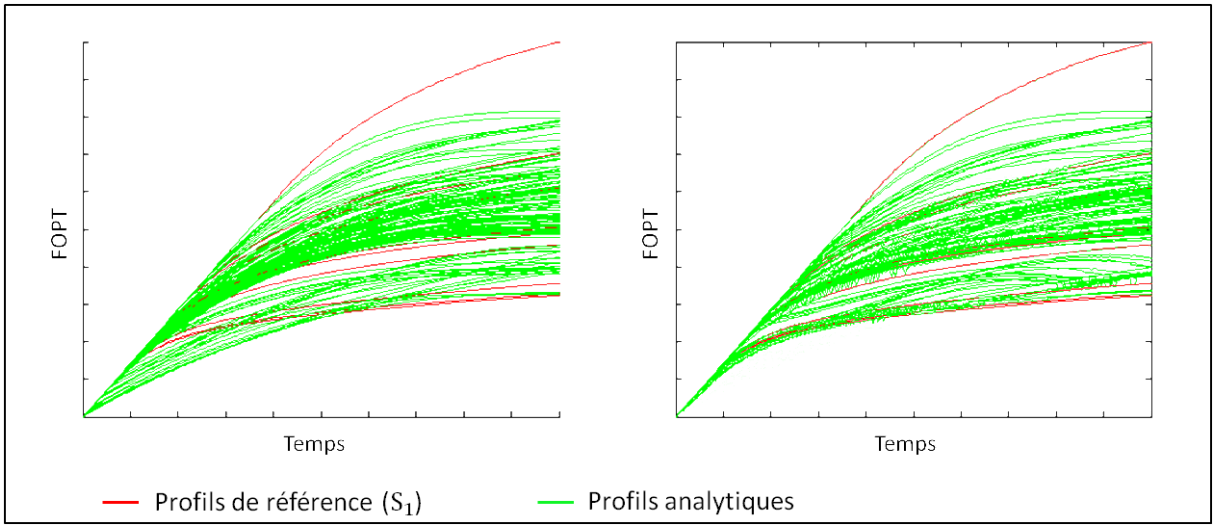

Figure 30 : Ensemble des profils analytiques (a) avant correction, (b) après correction.

L'application d'une telle correction peut entraîner la perte de la monotonicité des fonctions paramétriques. Pour éviter cela, il serait possible de corriger les profils *a posteriori* grâce à un lissage par exemple. Cependant, la correction améliore le résultat de l'estimation en diminuant le biais (clairement visible pour les temps courts sur l'exemple de la Figure 30). De plus, le but final de cette méthodologie est d'améliorer le calcul des quantiles de productions or, ceux-ci étant des profils statistiques, leur comportement restera donc plus facilement monotone. C'est pourquoi nous ne présenterons pas cela ici. La Figure 31 présente un exemple de quantiles obtenus en utilisant notre méthodologie, comparés au quantiles théoriques. Ceux-ci sont très proche exceptés pour les temps court du Q10. Cela peut s'expliquer par le fait que nous utilisons une distance globale pour l'ensemble du profil ainsi que par le fait que plusieurs paramètres des modèles analytiques peuvent influencer leur forme au niveau des temps courts.

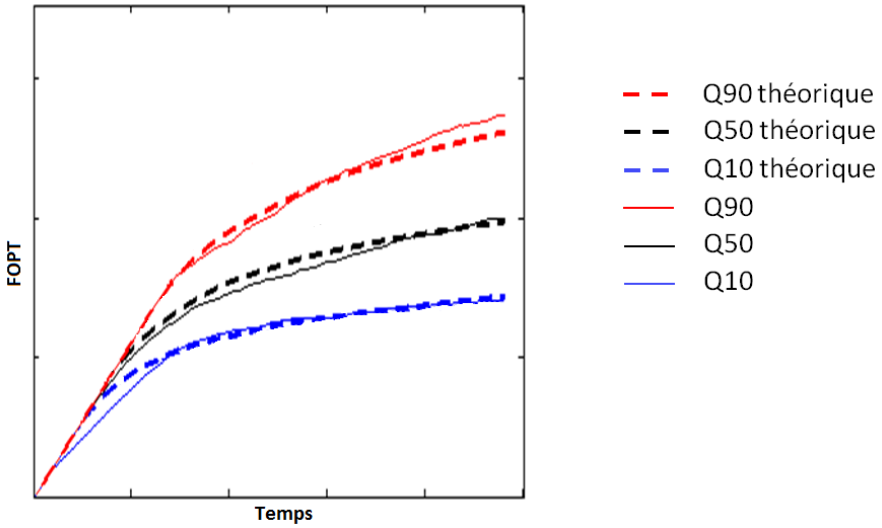

Figure 31 : Quantiles statistiques calculés sur l'ensemble des modèles de $\mathcal{S}$ (théoriques) et sur l'union des profils de référence de $\mathcal{S}_1$ avec l'ensemble des profils analytiques de $\mathcal{S}_2$

Grâce à l'ensemble de la méthode proposée ci-dessus (corrélation entre distances proxys et distances sur les profils issus de la simulation à haute fidélité, simulation des distances, approximation des courbes de réponses pour tous les géomodèles et correction), il devient possible de calculer les quantiles de production non pas sur un petit nombre de profils dynamiques mais sur un très large ensemble de profils, dont certains sont vrais ($\mathcal{S}_1$) et d'autre analytiques ($\mathcal{S}_2$). Ceci permet de rendre les calculs plus stables.

Le point clé de cette méthodologie est que les profils obtenus ne sont pas utilisés en tant qu'approximation du vrai profil dynamique, mais pour apporter de l'information à partir des modèles dont seule la réponse approchée par un proxy a été calculée dans la quantification des incertitudes sur les réserves. Ainsi, même si le profil analytique surestime ou sous-estime localement la solution dynamique, l'ensemble des profils, lui, apporte une information utile pour l'évaluation des incertitudes.

Nous avons présenté dans cette thèse une nouvelle méthode pour améliorer l'évaluation des incertitudes sur les réserves dans le cas d'un réservoir *green field*. Cette méthode utilise une sélection de modèles représentatifs, ainsi que des simulations proxys. Cela nous a naturellement amené à concevoir un moyen de comparer les différents proxys à notre disposition, afin de choisir le plus adapté à notre but. Nous avons donc également mis au point plusieurs critères de comparaison permettant de révéler certains comportements des proxys, pour pouvoir faire le bon choix.

Partie III : APPLICATION

Dans cette partie sont présentés deux cas d'étude, pour lesquels nous avons comparé différents proxys grâce à notre méthodologie. Nous avons aussi utilisé notre procédure de reconstruction des profils de production d'hydrocarbure, afin d'estimer l'incertitude sur le comportement à l'écoulement des fluides d'un grand nombre de réservoirs. Nous y exposons également les développements informatiques qui ont mené à un prototype possédant une interface graphique permettant la mise en œuvre des méthodes par un utilisateur informé mais non spécialiste.

Chapitre 1 : Développement informatique

Dans ce premier chapitre, nous allons développer les différents aspects informatiques traités pendant cette thèse. Nous commencerons par présenter les différentes données disponibles, ainsi que la manière de les stocker et les problèmes que cela pose. Nous exposerons ensuite la solution de base de données employée. Nous finirons par une présentation de certaines des interfaces graphiques développées pour faciliter l'utilisation.

L'ensemble des travaux réalisés pendant cette thèse a fait l'objet d'un développement informatique sous la forme d'un programme indépendant pouvant être directement utilisé au sein de la société Total S.A. Nous allons ici détailler certains points importants de la conception de ce logiciel, ayant nécessité un travail et une réflexion particulière.

Plusieurs principes et objectifs ont guidé ce développement :

- Ergonomie simple : réalisation d'une interface graphique (IHM) afin de faciliter la prise en main par un utilisateur.
- Intégration facilitée dans le logiciel interne : utilisation du langage de programmation Java déjà employé dans la société.
- Interconnexion avec les différents logiciels utilisés : prise en charge de formats propriétaires tels que les fichiers du simulateur haute fidélité Eclipse®(Schlumberger) ou les feuilles de calcul Excel®(Microsoft).
- Rapidité d'exécution : utilisation du parallélisme quand celui-ci est possible.
- Optimisation de l'espace de stockage et rapidité d'accès aux informations : déploiement d'une base de données.

1.1 Mise en place d'un modèle et d'une base de données

La totalité des travaux repose sur un ensemble de géomodèles générés pendant une étude de l'incertitude autour de l'impact des paramètres statiques sur les réserves. Ces géomodèles sont composés d'une grille géométrique et d'un ensemble de propriétés 3D. Il existe plusieurs types de grilles :

- Les « *pillard grids* », ou **grilles piliers**, sont les premières historiquement employées. L'ensemble des cellules cubiques qui composent la grille sont alignées, formant des piliers plus ou moins verticaux (Figure 32a). Ces grilles sont les plus simples à utiliser car elles forcent les cellules à toujours posséder une partie de face commune, ce qui facilite la simulation dynamique d'écoulement.
- Les « *stair step grids* », ou **grilles en marches d'escalier**, sont une évolution des grilles piliers. Elles sont aussi formées de cellules cubiques alignées, mais dont certaines sont désactivées. Elles permettent ainsi de mieux s'ajuster à certaines structures géologiques complexes, telles que les failles.
- Les « *unstructured grids* », ou **grilles non structurées,** sont apparues plus récemment. Il en existe de nombreux types, selon la forme des cellules (tétraèdre, maillage de Voronoï,…) (Figure 32b). De telles grilles permettent de s'adapter au mieux aux contraintes géologiques ou à la position des puits, mais ne sont pas encore très répandues, surtout en ce qui concerne la simulation dynamique d'écoulement.

Pour cette thèse nous utiliserons des grilles piliers, car elles constituent encore le standard dans le domaine pétrolier.

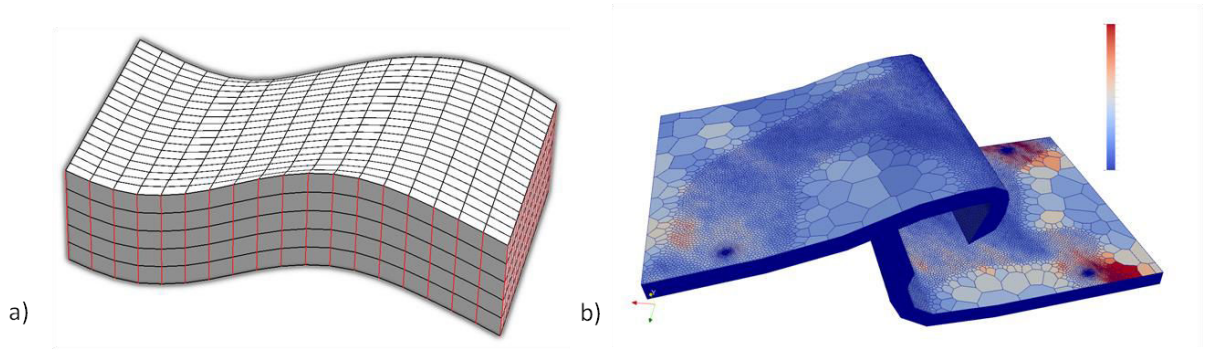

Figure 32 : Différentes grilles réservoir. (a) Grille piliers. L'alignement vertical des cellules forme lesdits piliers (lignes rouges). (b) Grille non structurée, constituée de cellules de Voronoï de tailles différentes et affichant la porosité.

Informatiquement les grilles sont souvent stockées de façon identique, quel que soit le logiciel employé (Figure 33). L'objet « grille » contient deux listes :

- Une liste des coordonnées des sommets, ou points, formée de la suite des coordonnées x, y et z pour chaque point distinct.
- Une liste des sommets formant chaque cellule.

L'avantage d'une telle représentation est qu'elle évite le stockage d'informations redondantes. Par exemple dans la Figure 33, les coordonnées du sommet 2 qui est commun à la cellule 1 et à la cellule 2 ne sont stockées qu'une seule fois. Dans le cas d'un géomodèle sans faille, la majorité des sommets sont communs à huit cellules, ce qui permet de fortement réduire l'espace nécessaire au stockage de la grille.

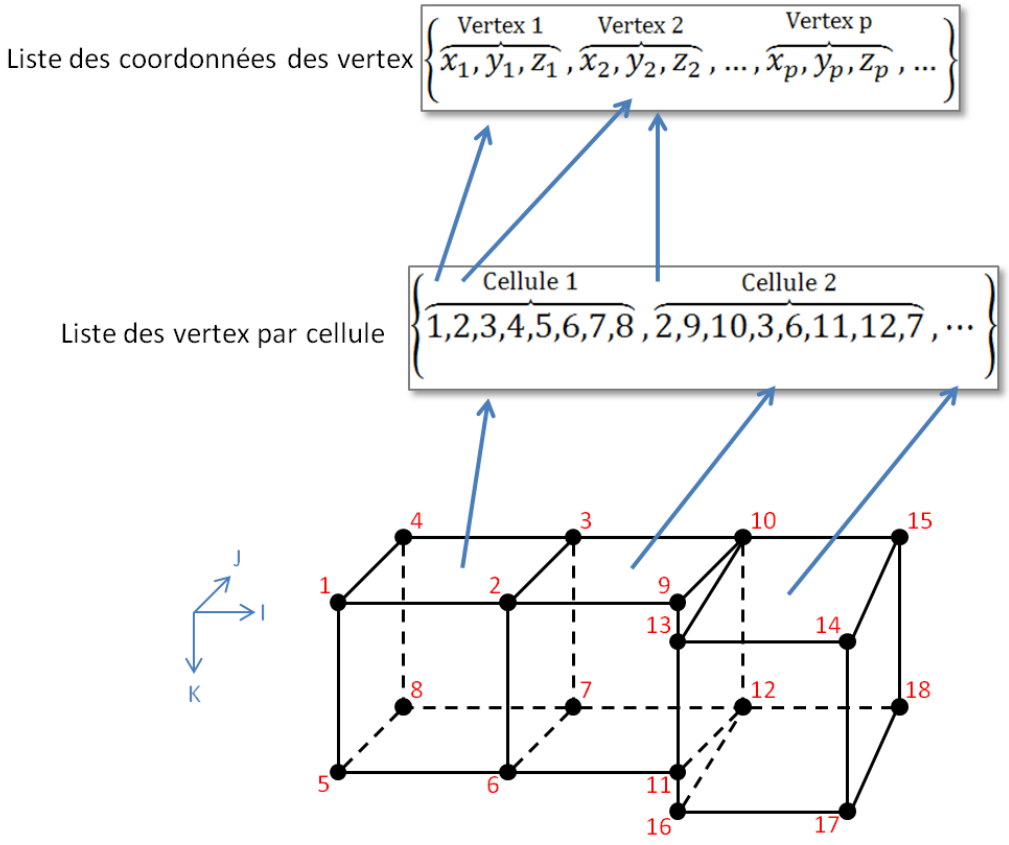

Figure 33 : Stockage informatique des grilles hexaédriques.

Dans notre cas, les géomodèles ont été obtenus à l'aide du plugin JACTA® du géomodeleur Gocad®, permettant d'étudier l'impact des incertitudes statiques sur les volumes d'hydrocarbure en

place et de générer les géomodèles associés. Les simulations proxys, quant à elles, sont effectuées dans notre programme indépendant. Il faut donc transférer les modèles du géomodeleur au code de simulation proxy. Il n'existe pas de standard officiel pour cela, nous utiliserons donc des fichiers ASCII formatés de manière à être pris en charge par le simulateur d'écoulements Eclipse®. Cette méthode confère différents avantages dont les plus importants sont :

- Le format ASCII peut être lu par un simple éditeur de texte.
- Ce format est déjà supporté par la plupart des logiciels commerciaux.
- Le simulateur Eclipse®, que nous avons utilisé comme simulateur de référence, est un des plus répandus dans l'industrie pétrolière.

Ces fichiers fonctionnent sur le principe mots-clés/valeurs. Les grilles piliers sont définies par trois mots-clés principaux (Figure 34) :

- SPECGRID, qui définit le nombre de cellules dans la grille dans les trois directions de l'espace.
- COORD, qui définit l'ensemble des piliers constituant la grille. Ceux-ci sont formés par deux extrémités situées sur le toit et la base de la grille, ayant chacune trois coordonnées spatiales.
- ZCOORN, qui définit la coordonnée verticale (z) de chaque sommet le long des piliers. Les deux autres coordonnées sont calculées par interpolation linéaire entre les extrémités des piliers.

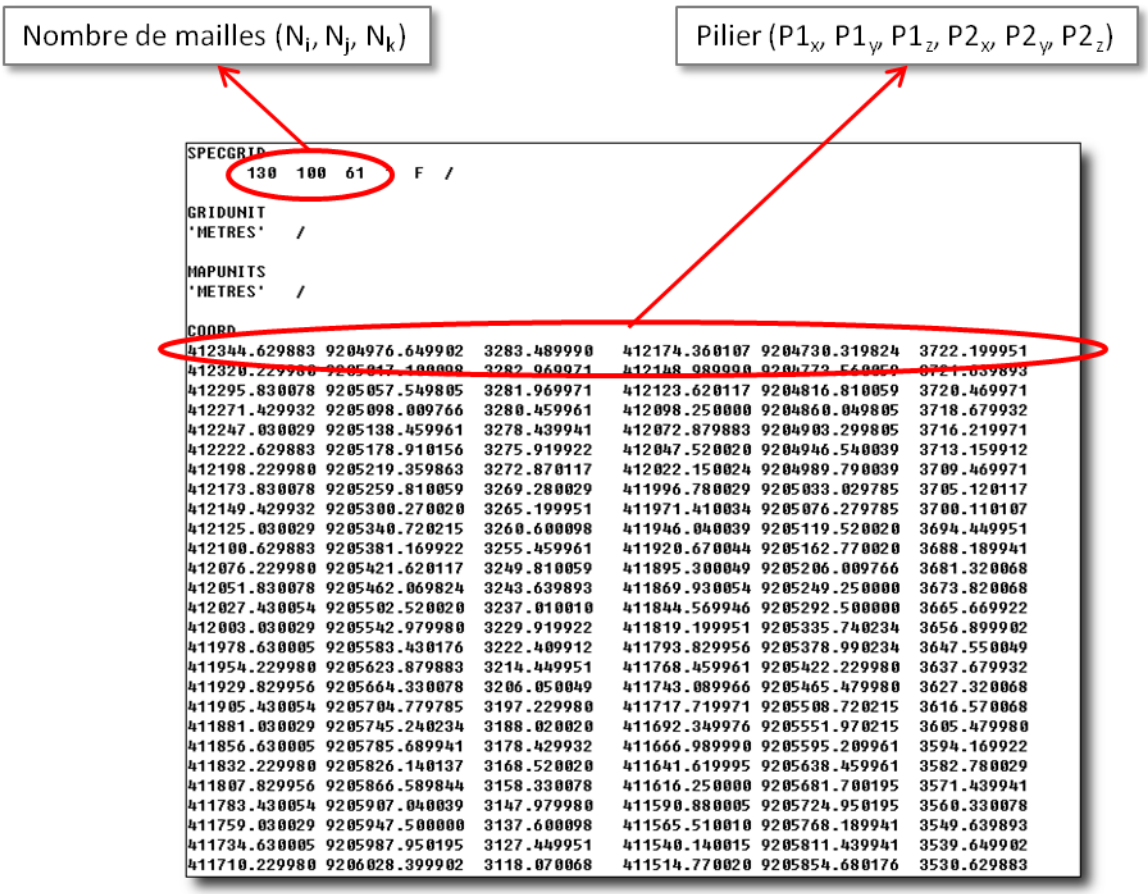

Figure 34 : Exemple de fichier ASCII de définition de grille.

Les propriétés du géomodèle ne sont formées que d'un seul mot clé, puis de l'ensemble des valeurs pour chacune des cellules de la grille (Figure 35).

Figure 35 : Exemple de fichier de propriété 3D du géomodèle.

Ce type de format pose un problème majeur : l'espace de stockage. En effet, il contient de nombreux espaces. Un fichier de définition de grille de 800 000 mailles, ce qui constitue une taille moyenne, dépasse les 80 Mo et les fichiers de propriété font presque 10Mo. De plus, la lecture et l'écriture de tels fichiers peuvent s'avérer longues si elles sont répétées de nombreuses fois. Afin de s'affranchir de ces problèmes, une solution est d'utiliser une base de données.

Une base de données est constituée de plusieurs tables (une pour chaque type d'objet stocké). Chaque table est constituée de plusieurs colonnes (une par attribut) et de relations entre elles. Il faut donc commencer par définir un modèle de données. La Figure 36 représente, sous forme d'un diagramme UML, les relations qui existent entre les différentes tables de notre base de données.

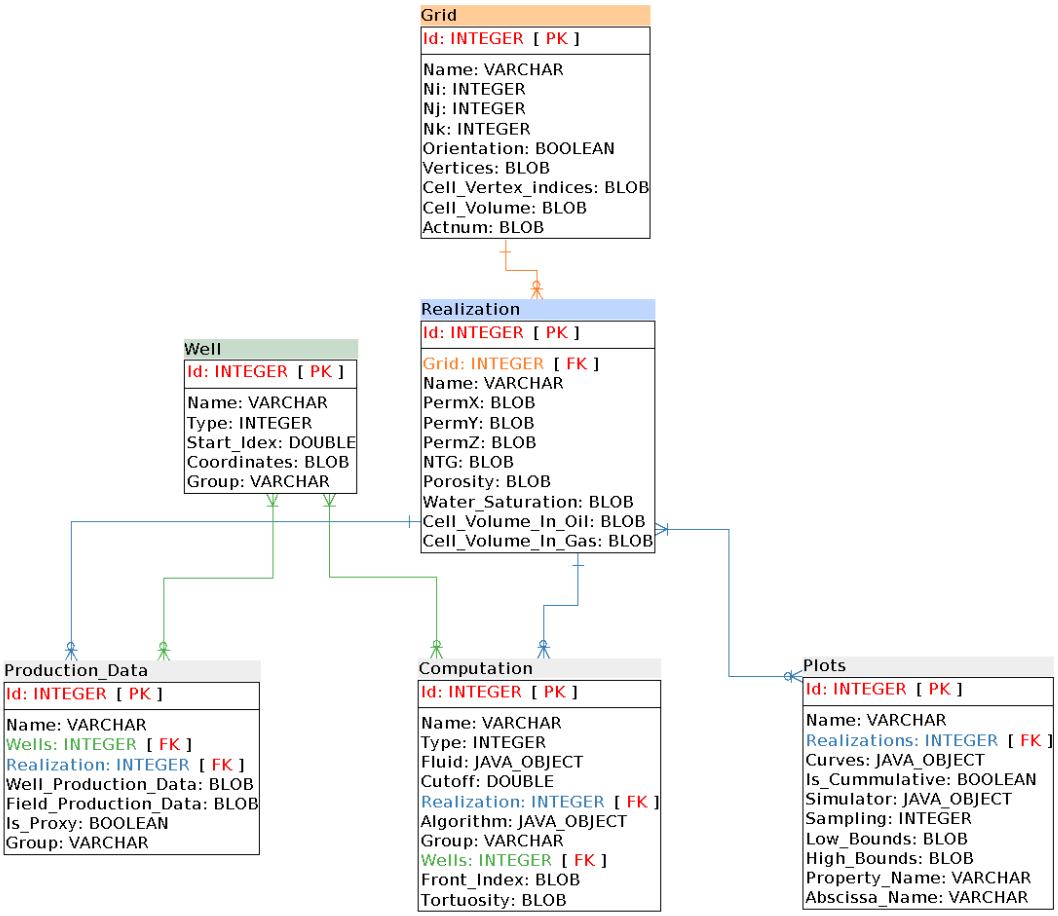

Figure 36 : Diagramme UML représentant les tables de la base de données ainsi que les relations qui la régissent. Pour chaque variable de la table, le type utilisé est renseigné (BLOB : *Byte Long OBject*, correspond informatiquement à un tableau de byte) ainsi que si elle correspond à la clé primaire (PK : *Primary Key*) ou à une clé étrangère (FK : *Foreign Key*).

Nous utilisons six tables distinctes :

- Les **Grilles**, pour stocker la géométrie des géomodèles, ainsi que la propriété ACTNUM qui définit quelles sont les cellules actives.
- Les **Réalisations**, pour stocker l'ensemble des propriétés pétrophysiques, ainsi que la saturation en eau initiale et les volumes de fluides en place.
- Les **Puits**, pour stocker les coordonnées des perforations de chaque puits, ainsi que le type de puits (injecteur ou producteur).
- Les **Données de production**, qui stockent les résultats de simulations dynamiques d'écoulements.
- Les **Calculs de propagation de front**, pour stocker les résultats de simulations proxys.
- Les **Graphes**, pour stocker les configurations d'affichage et de calcul de distances pour un ensemble de courbes.

Afin de relier ces tables entre elles, nous définissons plusieurs relations comme indiqué sur le diagramme :

- Relation « **Un-à-Plusieurs** » pour relier des réalisations à une grille, ainsi que des graphes, des simulations proxys et des simulations dynamiques à une réalisation.
- Relation « **Plusieurs-à-Plusieurs** » pour relier les différents puits impliqués dans les différentes simulations (proxys ou dynamiques). Ce type de relation est en fait réalisé grâce à des tables de jointures non représentées sur le diagramme.

L'emploi de cette base de données apporte de nombreux avantages. Tout d'abord, l'espace de stockage est grandement réduit car les informations ne sont plus stockées sous format ASCII mais en binaire. De plus, l'utilisation de la propriété définissant les cellules actives permet, pour l'objet Réalisation, de ne stocker que les valeurs de ces cellules. Or dans certains cas, elles représentent moins de la moitié des cellules du géomodèle. Dans le premier cas que nous présenterons par exemple, un gain de place de l'ordre de 73% est obtenu (taille des fichiers ASCII : 110Go, taille de la base de données : 30Go) pour 200 géomodèles différents et 82% pour le second cas d'étude (1.2To de fichiers ASCII et 220Go de Base de données) composé de 100 géomodèles.

Un autre avantage majeur de l'emploi d'une base de données est la possibilité d'utilisation des requêtes. Elles permettent :

- Soit de ne récupérer que les informations nécessaires, comme le nom des réalisations stockées, sans pour autant récupérer l'ensemble des propriétés pétrophysiques. Cela s'avère très utile lors de la construction des interfaces.
- Soit de récupérer les données reliées entre elles. Il est notamment possible de récupérer rapidement les résultats de simulations dynamiques correspondant à une simulation proxy grâce à la réalisation à laquelle elles font référence.

En pratique nous avons utilisé une base de données H2[2] qui permet de fonctionner en local. Ce type de base de données est un peu plus lente qu'une base hébergée sur un serveur indépendant, mais permet de s'affranchir des problèmes réseaux (serveur allumé, trafic, multi-utilisateurs, etc.). Les requêtes se font en SQL, mais pour plus de commodité nous avons utilisé la bibliothèque Hibernate[3] qui permet une meilleure intégration avec le code Java®. En effet, la librairie construit elle-même les requêtes à partir du code informatique.

1.2 Mise en place d'une interface graphique

Afin de rendre notre programme utilisable par le plus grand nombre d'utilisateurs possible, il était nécessaire de développer une interface Homme-Machine. Pour cela nous avons utilisé principalement la bibliothèque Swing, fournie avec les packages de base du langage Java.

Figure 37 : Menu principal du programme développé.

La Figure 37 présente le menu principal de notre programme. Le premier onglet (**File**) offre la possibilité à l'utilisateur de créer un nouveau projet (base de données) ou d'en ouvrir un déjà existant afin de récupérer des résultats précédents. Le deuxième onglet (**Settings**) permet d'effectuer toutes les actions de gestion des données de la base. Il est possible de rentrer de nouvelles données comme des grilles, des réalisations, des puits ou bien des données de production. Les données obsolètes ou fausses peuvent être supprimées afin de libérer de la place et d'accélérer les requêtes. Les relations de type clés secondaires doivent être définies lors de l'ajout des objets en base

[2] Projet H2 : http://www.h2database.com/
[3] Projet Hibernate : http://hibernate.org/

(spécification de la grille à laquelle se rapporte une réalisation par exemple), car il n'est plus possible de les modifier par la suite.

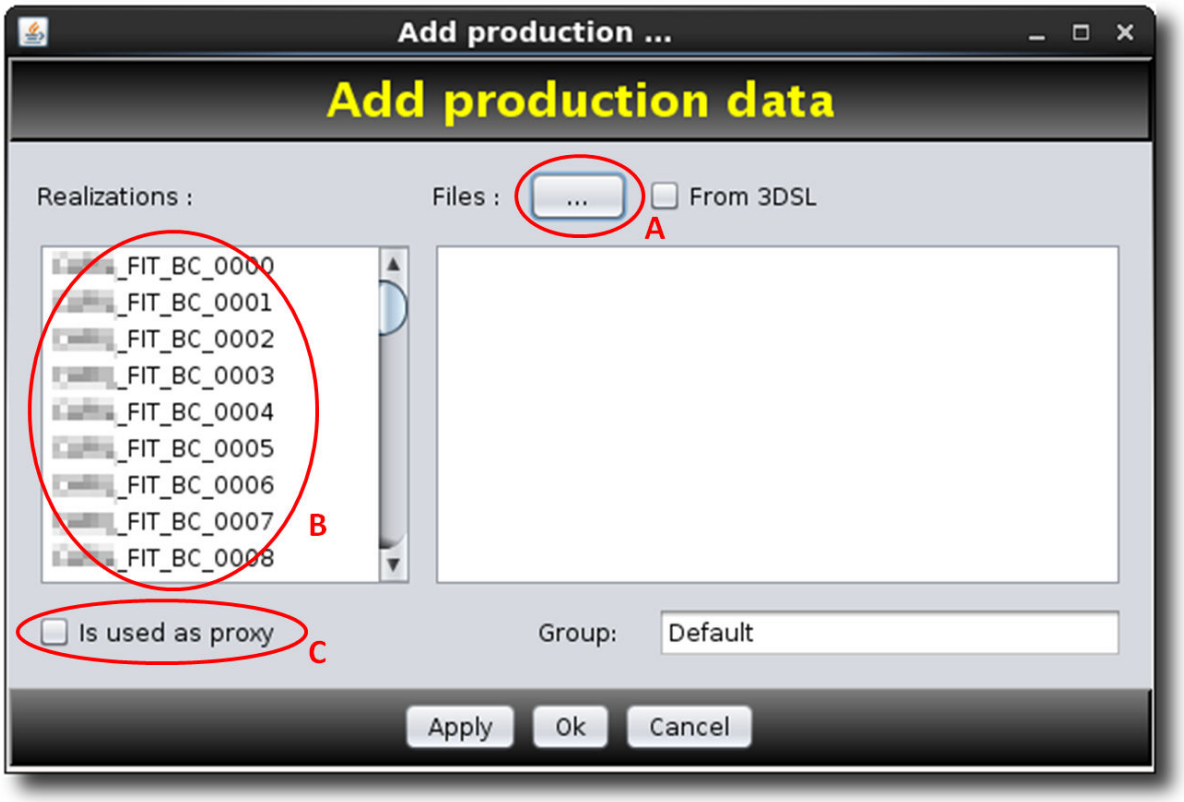

Figure 38 : Fenêtre permettant l'ajout de données de production dynamiques à la base de données.

La Figure 38 montre l'interface servant à ajouter les données de production. Le bouton A permet à l'utilisateur d'aller chercher les fichiers à ajouter. La liste B affiche l'ensemble des réalisations disponibles (récupérées grâce à une requête sur la base), l'utilisateur peut ainsi choisir celles qui sont destinées à recevoir les données de production. Grâce à la case C, il est possible de préciser si les résultats sont utilisés à des fins de proxy (StreamLines, Upscaling). Enfin, le troisième onglet du menu principal (**Proxies**) permet d'effectuer les simulations proxys de propagation de front.

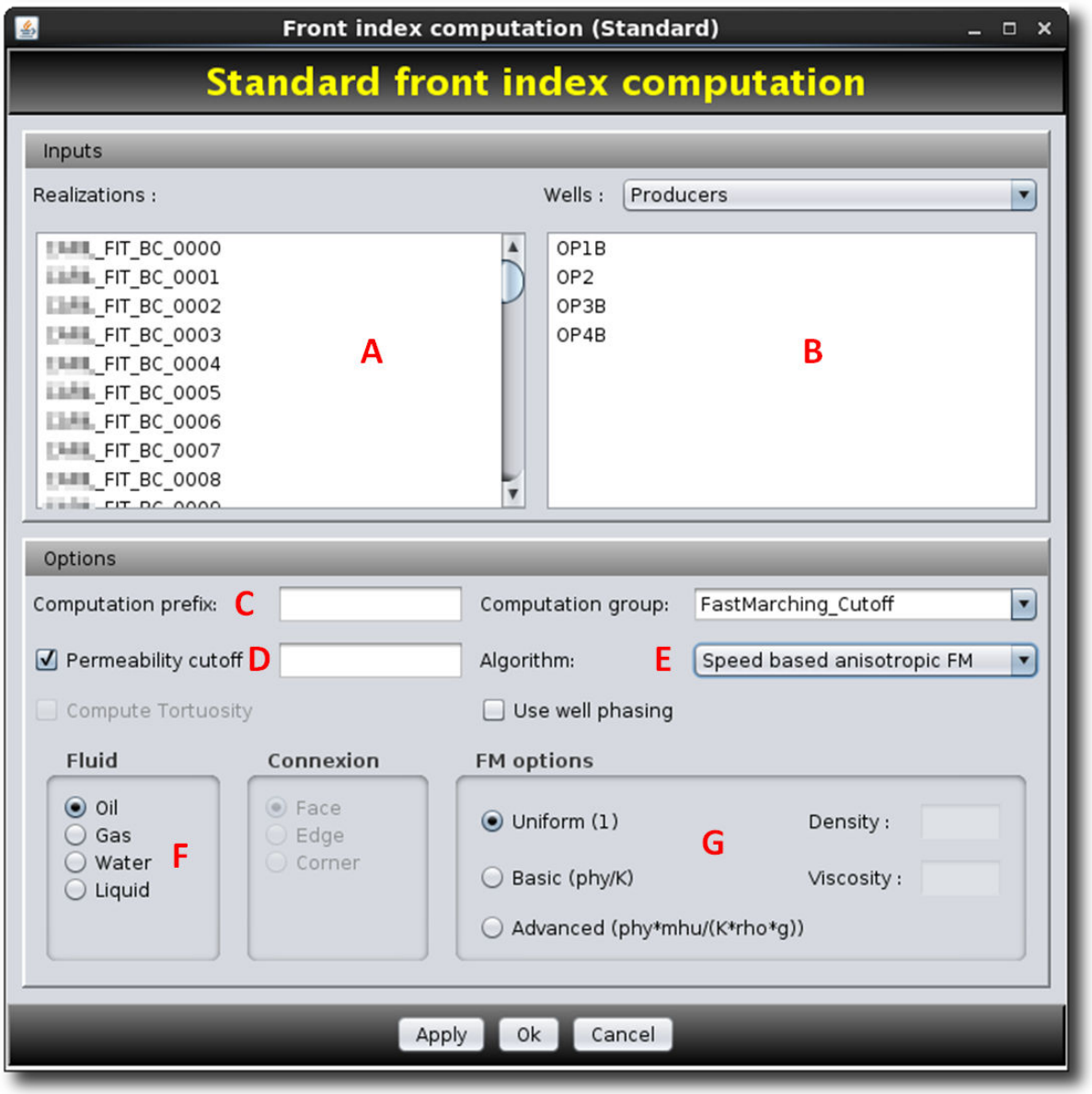

Figure 39 : Interface pour effectuer les simulations proxys de propagation de front.

Comme nous pouvons le voir sur la Figure 39, plusieurs informations sont nécessaires pour effectuer ces simulations proxys :

- A : les réalisations sur lesquelles sont effectuées les simulations.
- B : les puits utilisés comme points de départ du front.
- C : les options de nommage des résultats.
- D : le seuil de perméabilité (facultatif pour les algorithmes basés sur le *Fast Marching*).
- E : l'algorithme utilisé pour la simulation (Connectivité, *Fast Marching* isotrope, *Fast Marching* anisotrope).
- F : le choix du fluide présent dans le réservoir et comptabilisé lors de la propagation.
- G : la fonction de vitesse (uniquement pour les algorithmes de *Fast Marching*).

Une fois les paramètres réglés, l'utilisateur lance les calculs. Les résultats sont alors automatiquement ajoutés à la base de données, ainsi que les relations correspondantes. Le quatrième onglet du menu principal (**Post-Processing**) permet de réaliser la comparaison des proxys, grâce à la méthode exposée en I.4, et ne sera pas présenté ici. Dans le dernier onglet (**Special**), il est

possible de visualiser les résultats. Pour cela nous avons utilisé, en plus de la bibliothèque Swing, la bibliothèque JFreeChart qui permet de réaliser des graphiques en deux dimensions.

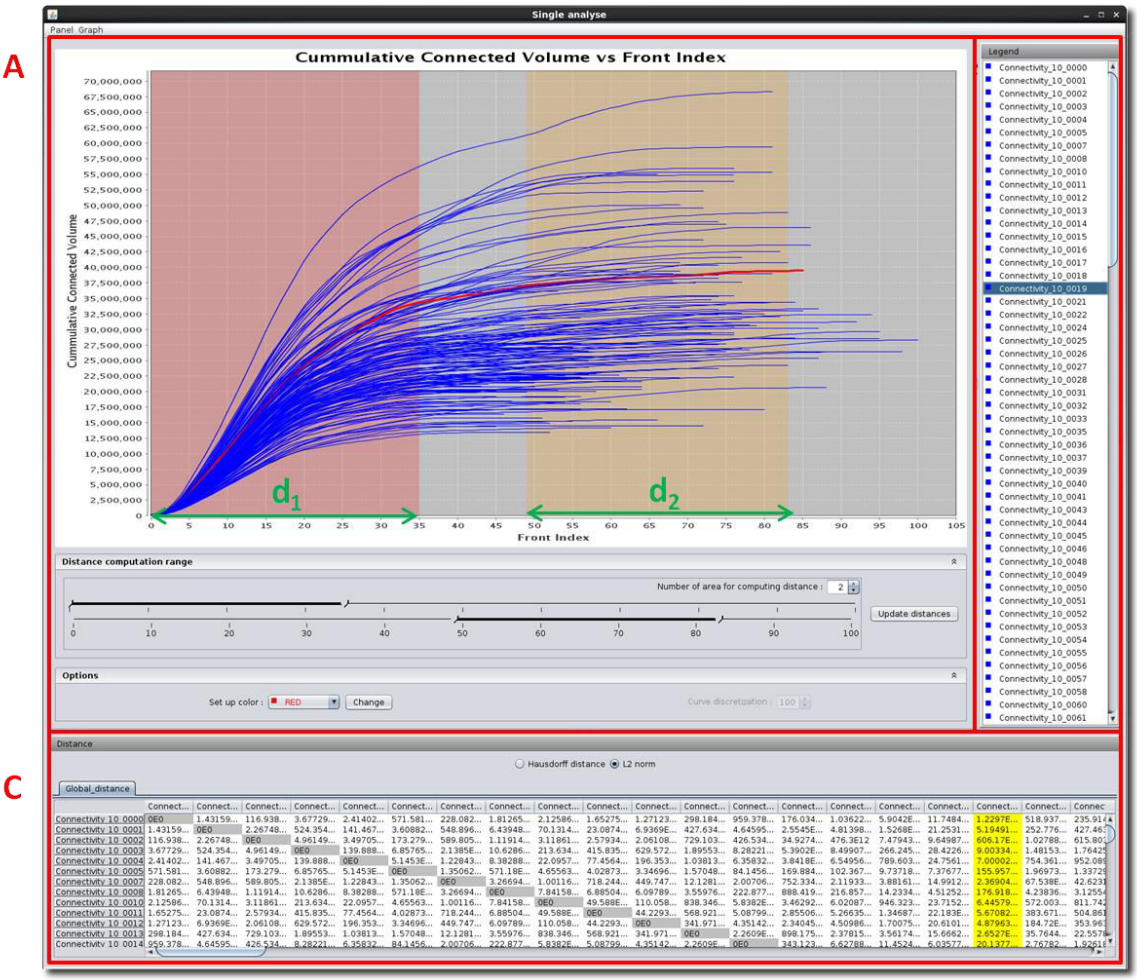

Figure 40 : Interface permettant de définir les zones de calcul des distances.

Une fois les données à afficher choisies, la Figure 40 est obtenue. Elle présente une fenêtre découpée en plusieurs zones :

- A : les **courbes de simulations** (proxys ou dynamiques) avec leurs options. Il est possible de changer la couleur des courbes sélectionnées, ou bien leur échantillonnage pour permettre de changer le nombre de points qui forment les courbes.
- B : la **légende** servant à identifier les courbes. Une ou plusieurs de ces courbes peuvent être sélectionnées et apparaîtront alors en rouge avec une plus grande largeur de trait dans le graphique.
- C : la **matrice des distances** entre chaque courbe de réponse. Il est possible de choisir le type de distance calculée (Norme L2 ou distance de Hausdorff). Les profils sélectionnés sont surlignés afin d'augmenter la lisibilité des résultats. En utilisant les options des graphiques, la zone sur laquelle la distance est calculée peut être modifiée, en ajustant la plage de calcul et en ajoutant de nouvelles plages de calcul. Ce changement permet de ne calculer les distances que dans certaines zones d'intérêt.

Une fois notre graphique configuré avec la ou les plages de distances ajustées, nous pouvons le sauvegarder dans la base de données afin de le réutiliser pour effectuer une comparaison de ce

proxy avec un autre. Il est aussi possible d'exporter la matrice distance et les données graphiques aux formats texte et Excel®, afin de les utiliser dans d'autres programmes.

Chapitre 2 : Application au champ X

Ce chapitre présente les résultats du premier cas d'étude utilisé pour tester notre méthodologie de comparaison des proxys et notre procédure de reconstruction de profil de production cumulée d'huile. Nous commencerons par exposer le contexte géologique et les techniques de modélisation. Nous comparerons ensuite les sept proxys à notre disposition. Nous finirons par l'optimisation avec l'ensemble des proxys.

Nous allons à présent mettre en œuvre nos différentes méthodes, grâce au logiciel présenté précédemment, sur un premier cas d'étude réel proposé par la société Total S.A. et que nous appellerons X.

2.1 Présentation générale

Il s'agit d'un réservoir formé de dépôts turbiditiques et encore non opéré (*green field*). Le géomodèle est constitué de presque 800 000 cellules ($130 \times 100 \times 61$) de $\pm 50m$ de côté. Il s'agit d'une taille fréquemment rencontrée pour un réservoir de taille modeste. En moyenne, sur l'ensemble des modèles représentant ce réservoir, seulement 10% des cellules sont actives, c'est-à-dire utilisées pour la simulation dynamique d'écoulement. Ce faible pourcentage s'explique par le fait que les chenaux turbiditiques qui forment le réservoir sont inclus dans une matrice argileuse, dont les propriétés pétrophysiques n'en font pas une bonne roche réservoir. Afin de simplifier la simulation numérique, les cellules qui constituent cette matrice sont désactivées. Cette matrice correspond à l'élément architectural[4] 1 dans la Figure 41a.

Le schéma de développement du champ est constitué de trois puits injecteurs d'eau (en bleu) et de quatre puits producteurs (en vert). Ceux-ci sont principalement situés dans la partie haute du modèle car un aquifère actif est présent dans la partie basse, permettant de maintenir la pression. Ce schéma de développement est identique pour tous les modèles et a été construit à partir d'un cas de base. C'est pourquoi lors de l'étude des incertitudes statiques sur ce cas, qui a permis d'obtenir un ensemble de 200 géomodèles différents (Figure 41), certains puits peuvent se situer dans une zone non réservoir. Cet ensemble de géomodèles possède des propriétés pétrophysiques ainsi que des géométries distinctes. Les propriétés pétrophysiques ont été obtenues à l'aide de simulations géostatistiques conditionnées par une donnée sismique 3D sur l'ensemble de la grille et des puits d'exploration (faciès, porosité, logarithme de perméabilité notamment) ponctuels. Les géométries distinctes, quant à elles, sont obtenues en faisant varier les différents horizons géologiques. Ainsi le nombre de cellules ne change pas, au contraire des coordonnées des sommets des cellules, permettant de faire varier les volumes de chaque maille. Cette méthode permet de prendre en compte des incertitudes structurales, car chaque horizon est défini par une position moyenne ainsi qu'une zone de variation de part et d'autre (Abrahamsen, 1993, Holden et al., 2003).

Pour l'ensemble des géomodèles, les écoulements ont été simulés grâce au simulateur haute fidélité Eclipse® afin de pouvoir disposer d'un ensemble complet de référence. Pour ces simulations, la présence d'eau, d'huile, de gaz et de gaz dissouts a été considérée ainsi qu'une certaine compressibilité pour tous les fluides. De plus, les proxys suivants ont été calculés sur chacun des géomodèles :

- le **STOOIP** ;

[4] Un élément architectural est une des hétérogénéités de grande échelle définies dans le réservoir, comme par exemple le lit du fleuve.

- les **volumes connectés** aux puits producteurs avec une perméabilité au-dessus d'un seuil de $10mD$;
- la propagation de fronts par **connectivité**, à partir des puits producteurs et avec une perméabilité au-dessus d'un seuil de $10mD$;
- la propagation de fronts par ***Fast Marching***, à partir des puits producteurs et en utilisant une fonction de vitesse incluant uniquement le ratio porosité/perméabilité ;
- les **lignes de courant** grâce au logiciel 3DSL®, après avoir converti les données des fichiers de configuration Eclipse® ;
- deux résultats de simulations Eclipse® sur des modèles **homogénéisés** avec deux ratios d'agrégation différents :
  - un faible $(2 \times 2 \times 2)$, qui agrège les cellules de manière identique dans toutes les directions et permet de diviser le nombre de cellules totales du modèle par huit, mais tout en gardant les formes de ces dernières ;
  - un très grossier $(2 \times 2 \times 31)$, qui permet d'obtenir finalement un modèle à seulement deux niveaux de cellules verticalement.

Cet ensemble de proxys représente un large panel du plus simple (STOOIP) au plus complexe (homogénéisation).

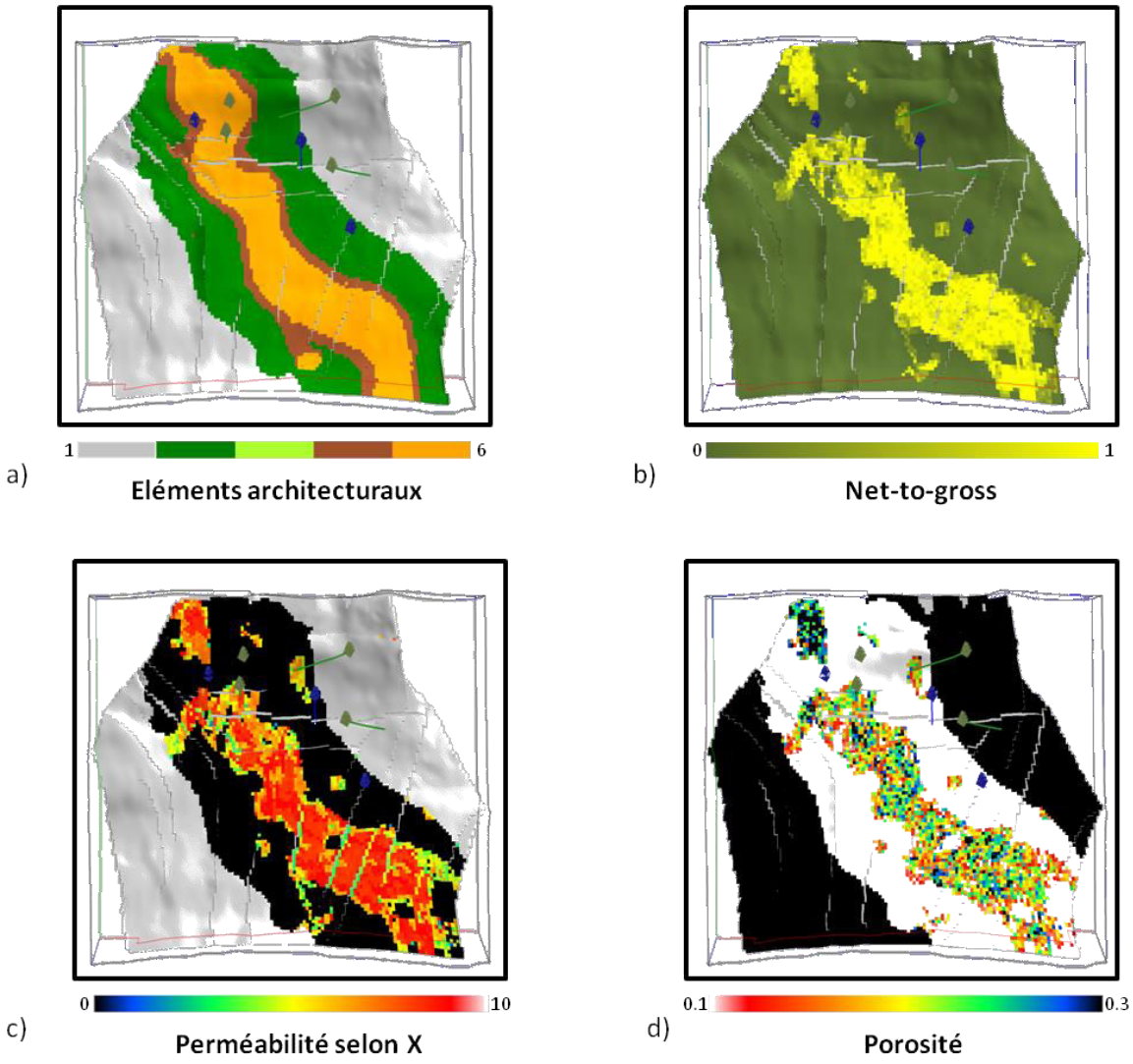

Figure 41 : Répartition des propriétés pétrophysiques (a) Eléments architecturaux, (b) *Net-to-gross*, (c) Perméabilité selon l'axe X de la grille et (d) Porosité dans un des modèles du cas d'étude X. La zone 1 des éléments architecturaux a été désactivée pour les simulations de la perméabilité et de la porosité. Les puits verts correspondent aux puits producteurs et les bleus aux puits injecteurs.

2.2 Comparaison des différents proxys

Après avoir fait une brève présentation des différents proxys utilisés, nous allons commencer par les comparer en utilisant notamment la méthodologie décrite précédemment (I.3).

2.2.a Mise en place, réponse et gain de temps.

Les deux premiers proxys ne fournissent qu'une valeur unique pour caractériser les modèles. La Figure 42 montre leurs résultats présentés sous forme d'histogrammes de distribution. Leurs formes générales sont assez similaires, ce qui indique que la majorité de l'huile présente dans les modèles se situe dans la zone de drainage (perméabilité > 10mD) des puits producteurs.

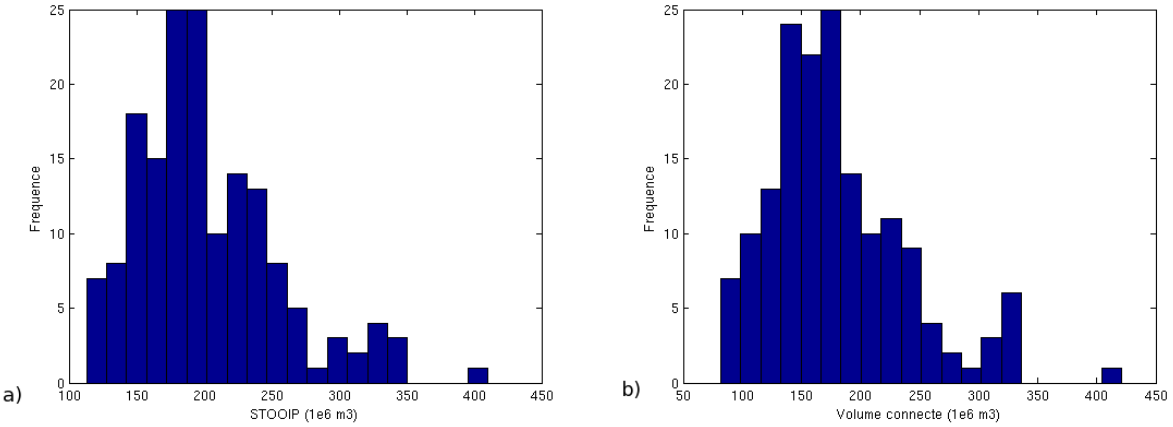

Figure 42 : Histogramme des résultats des proxys (a) STOOIP et (b) Volume connecté.

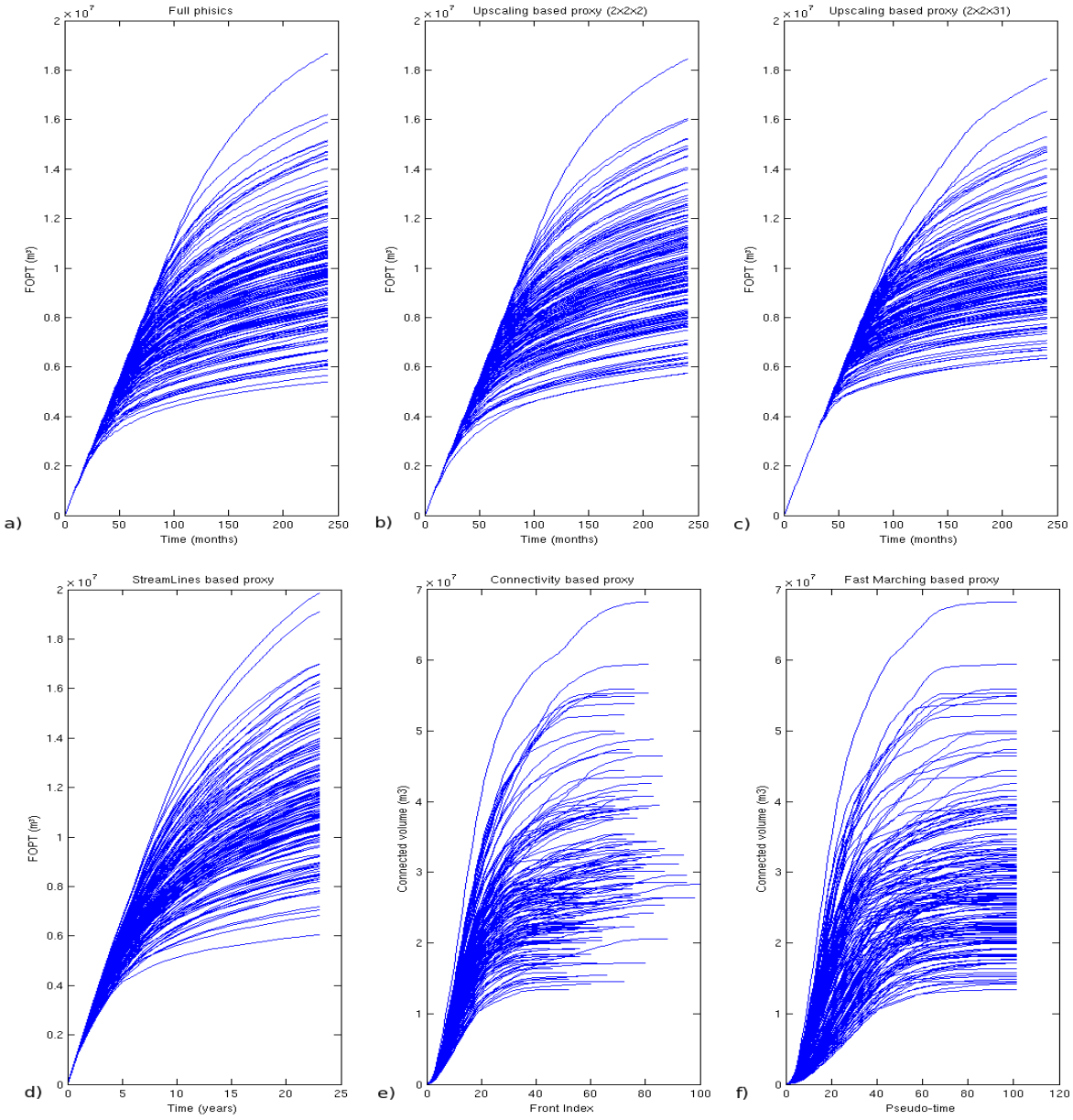

Figure 43 : Profils de réponses des simulations (a) de référence (haute fidélité), (b) upscaling faible, (c) upscaling grossier, (d) lignes de courant, (e) connectivité et (f) *Fast Marching*.

Pour tous les autres proxys, une courbe de réponse est obtenue pour chaque modèle. Celle-ci correspond à un volume en fonction du temps pour les simulations de référence, les lignes de courant ou les résultats d'homogénéisation (Figure 43a-d), des volumes en fonction d'un indice de front pour la connectivité (Figure 43e) ou bien des volumes par pseudo-temps (variable homogène à un temps à une constante près) pour la propagation de front par *Fast Marching* (Figure 43f). Bien que tous les proxys donnent des réponses différentes, nous pouvons tout de même remarquer que pour les proxys les plus complexes (upscaling et lignes de courant) les réponses sont assez proches, en termes de formes, des courbes de réponses des simulations de référence. L'aspect est plus différent pour les deux autres proxys (connectivité et *Fast Marching*) bien que nous observons, comme sur les références, un modèle qui produit plus que les autres. En termes de dispersion des courbes, nous constatons qu'à part l'homogénéisation de faible ratio tous les autres proxys montrent un léger écart par rapport à la dispersion observée sur les courbes de référence : une plus grande dispersion pour les proxys lignes de courant, connectivité et *Fast Marching*, ainsi qu'une plus faible dispersion pour l'upscaling grossier.

Prenons maintenant les trois critères de notre méthodologie. Le Tableau 2 présente les différences de paramètres entre les proxys utilisés.

Proxy	Paramètres	Réponses	Gain de temps	
			Mise en œuvre	Simulation
STOOIP	Φ, NTG, GRV, S_w, Bo	STOOIP	+	> 1000
Volumes connectés	*id.* + K, seuil, puits	STOOIP connecté aux puits	+	*id.*
Connectivité	*id.*	Volume d'huile/indice de front	+	240
Fast Marching	*id.*	Volume d'huile/pseudo-temps	+	180
Lignes de courant	*id.* + propriétés fluides	Fluides produits/temps	–	10
Upscaling grossier	*id.* + aquifère, perméabilité relative, contrôles puits	*id.*	–	60
Upscaling fin	*id.*	*id.*	–	20

Tableau 2 : Paramètres, types de réponses et gain de temps par proxy. K correspond à la perméabilité, Φ à la porosité effective, NTG au *Net-to-gross*, GRV au Gross rock volume, S_w à la saturation initiale en eau et Bo au ratio volume de l'huile entre le fond du réservoir et la surface.

Nous remarquons que pour les quatre premiers proxys le nombre de paramètres est relativement limité (huit maximum), car ce ne sont que des propriétés pétrophysiques ainsi que la position des puits. Leur gain de temps est lui très important, car ils ne nécessitent quasiment pas de traitement préalable, et leur temps de simulation est très court (inférieur à dix secondes pour les propagations de front). Pour les proxys plus complexes, les résultats sont plus surprenants. Bien que les paramètres soient plus restreints dans le cas des lignes de courant, le gain de temps est le plus faible. Cela peut s'expliquer de deux manières :

- Le temps de travail préalable est important, car il faut utiliser un logiciel commercial qui nécessite un paramétrage particulier. Même si celui-ci est capable de lire un fichier de configuration de simulation dynamique, il faut le redéfinir, certains paramètres n'étant pas pris en compte. C'est notamment le cas du contrôle des puits. Lors d'une simulation dynamique classique, ceux-ci sont contrôlés par des paramètres physiques tels que la pression minimale en bas du puits et le débit maximal pour un groupe de puits. Dans le cas de la simulation par lignes de courant, il faut définir manuellement ces paramètres de telle sorte que cela corresponde, ce qui demande du temps.
- La simulation par lignes de courant n'est pas particulièrement adaptée lorsque du gaz peut apparaître, car dans ce cas le champ de pression doit être recalculé plus souvent et la simulation peut devenir très longue.

Il est aussi important de noter que l'homogénéisation a été réalisée de manière assez simple. Certains paramètres, tels que les perméabilités relatives ou les aquifères, n'ont pas été mis à l'échelle mais conservés. Or, ceux-ci peuvent avoir un impact important et leur homogénéisation est un processus plus complexe et long. Dans ce premier cas d'étude, cette approximation ne semble pas avoir trop d'influence sur la qualité du proxy au niveau de l'huile produite (Figure 43).

2.2.b *Comparaison distances proxys et distances de référence.*

Après avoir mis en évidence les différences en termes de réponses, de facilité de mise en œuvre et de gain de temps pour chaque proxy, nous nous intéressons à la capacité de chacun d'eux à reproduire les distances entre les modèles du point de vue des simulations d'écoulements à haute fidélité. Pour cela, nous utilisons le graphique distances en termes de réponses de référence en fonction des distances en termes de réponses proxys pour l'ensemble des 200 modèles (*Crossplot Distances-Distances*), défini dans la méthodologie de comparaison (section I.3.1) pour chaque proxy (Figure 44).

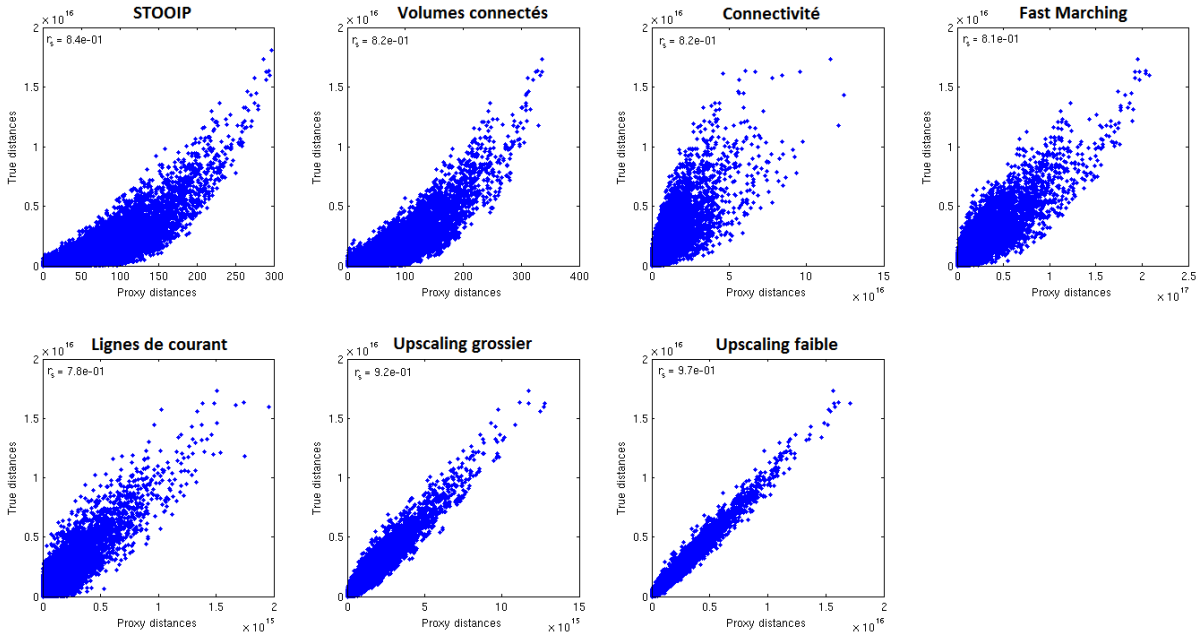

Figure 44 : Nuages de corrélation entre les distances estimées pour chaque proxy et pour chaque modèle et la distance de référence correspondante.

La Figure 44 met en évidence une corrélation – parfois non linéaire – pour chacun d'entre eux, ce qui signifie que les distances calculées grâce aux réponses proxys donnent de l'information sur les distances calculées sur les réponses de référence. Pour les deux proxys les plus simples (STOOIP et Volumes connectés), cette relation n'est pas linéaire. La dispersion des nuages de points n'est pas non plus similaire et augmente visuellement pour les proxys de propagation de front par connectivité et par *Fast Marching*.

Proxy	Coefficient de corrélation linéaire (Pearson)	Coefficient de corrélation de rang (Spearman)
STOOIP	0.86	0.84
Volumes connectés	0.86	0.82
Connectivité	0.81	0.82
Fast Marching	0.89	0.81
Lignes de courant	0.90	0.79
Homogénéisation grossière	0.96	0.92
Homogénéisation fine	0.99	0.97

Tableau 3 : Coefficients de corrélation pour les nuages de points distances-distances par proxy.

Le calcul des coefficients de corrélation (Tableau 3) permet lui aussi de tirer quelques conclusions. Tout d'abord la corrélation pour l'ensemble des proxys est bonne quel que soit le coefficient choisi, ce qui confirme notre première impression suite à l'observation seule des Nuages de points. Ensuite il apparaît que la plus mauvaise corrélation linéaire correspond au proxy de propagation de fronts par connectivité, alors que selon le coefficient de corrélation de rang il s'agit de celui par lignes de courant. Dans le premier cas, cela est également visible sur le nuage de points avec la dispersion pour les grandes valeurs. En revanche, pour le proxy par lignes de courant, ce résultat est assez surprenant et n'est pas identifiable avec le nuage de points car la corrélation semble linéaire et les points peu dispersés. Cela veut dire qu'une distance entre deux réponses proxys plus grande ne signifie pas forcément une plus grande distance en termes de réponse dynamique.

Enfin, en regardant l'origine des graphiques qui concerne les très faibles distances entre réponses, on remarque que les proxys STOOIP, volumes connectés et lignes de courant sont plus dispersés que les autres. Cela signifie que ces proxys ne permettent pas de détecter deux modèles au comportement semblable, en termes de simulation haute fidélité (Figure 45).

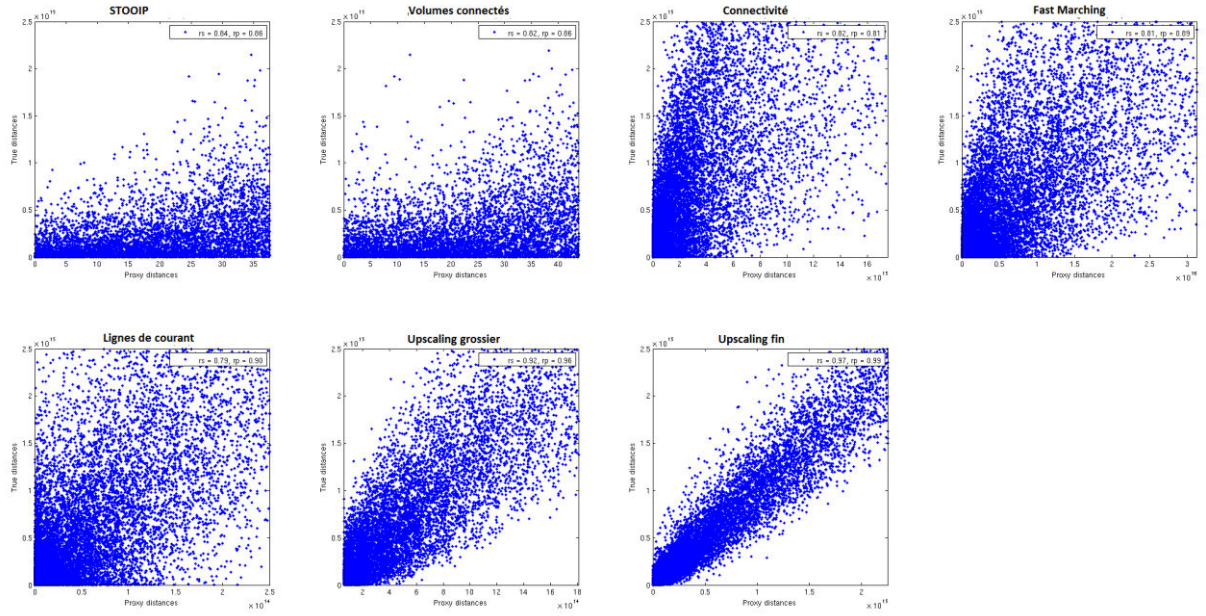

Figure 45 : Zoom sur les nuages de corrélation au niveau de l'origine.

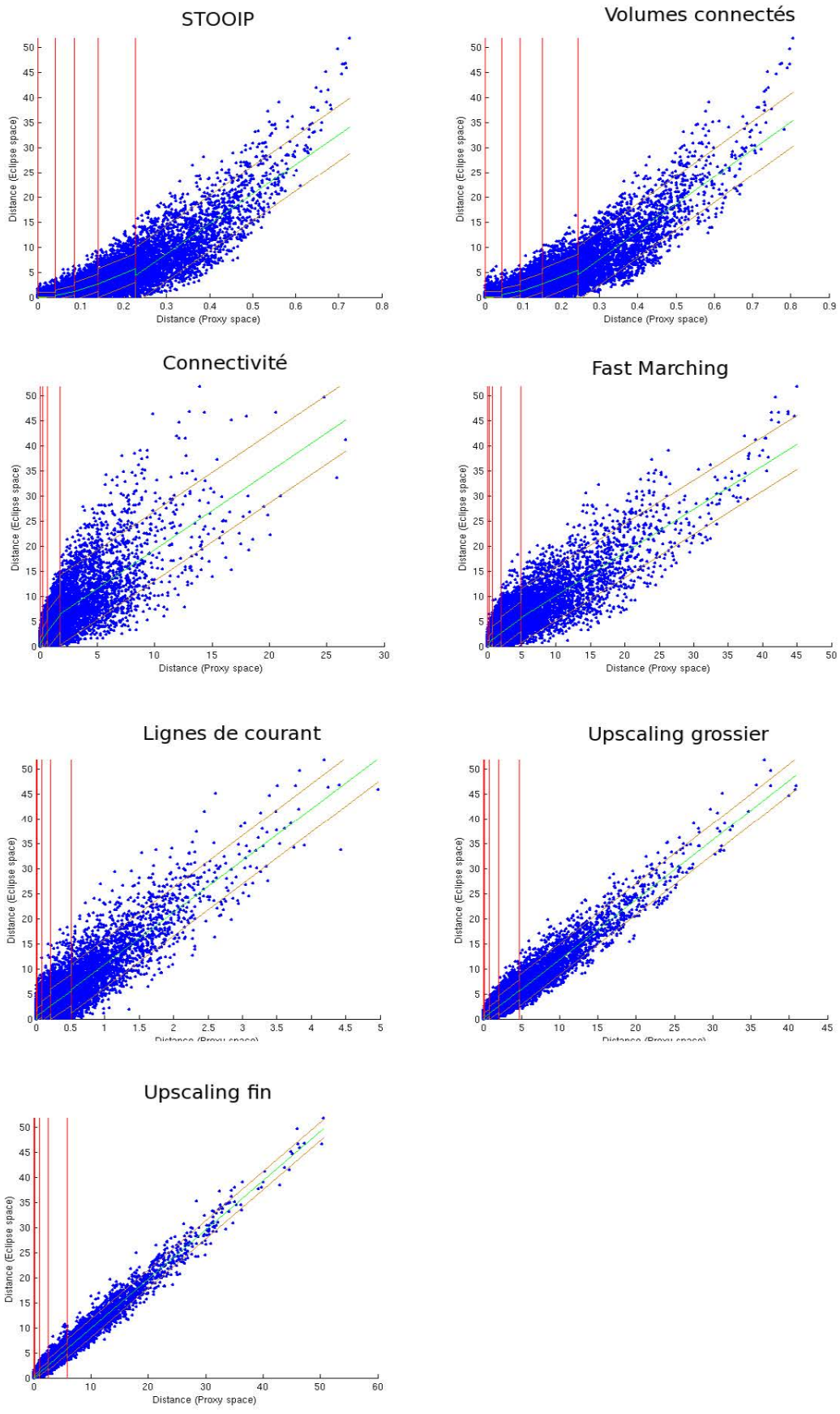

Figure 46 : Régression par morceaux pour chaque proxy appliqué sur le cas d'étude. Les lignes rouges verticales représentent les bornes de chaque zone.

L'étape suivante de la méthodologie consiste à effectuer une régression linéaire par parties sur les nuages de points. Chacun d'entre eux est alors divisé en cinq zones contenant le même nombre de points (quasiment 4000). Ceci nous permet de calculer un résidu pour chaque point, qui correspond à la différence de distance dynamique entre un point observé et le point correspondant sur la droite de régression linéaire de la zone. Il est possible de voir sur la Figure 46 les zones obtenues pour chaque proxy. Pour les deux premiers proxys, les quatre premières zones sont à peu près aussi larges les unes que les autres. Par contre, pour les autres proxys, la première zone n'est quasiment pas visible. Ce comportement signifie que les deux proxys les plus simples sont les plus sensibles aux faibles distances entre les réponses de référence. Cette hypersensibilité explique la forme exponentielle de l'ensemble des points.

Si nous regardons ensuite plus en détails la distribution des résidus par zone pour tous les proxys, nous observons un biais positif dans la première zone excepté pour le proxy d'upscaling fin. En effet, la ligne en pointillés verts sur la Figure 47 matérialise un résidu nul. Or, au lieu d'être centrées sur cette ligne, les distributions sont toujours positives. Ce comportement se corrige dès la deuxième zone. Cet effet peut s'expliquer par la troncature effectuée dans cette zone pour obtenir des distances négatives.

A partir de la comparaison des proxys vis-à-vis de leur capacité à reproduire les distances de référence entre les modèles, nous pouvons dire que les upscalings sont largement meilleurs. Pour les autres, à part le caractère non linéaire des nuages de points pour les proxys basés sur l'huile en place, aucun ne semble se distinguer.

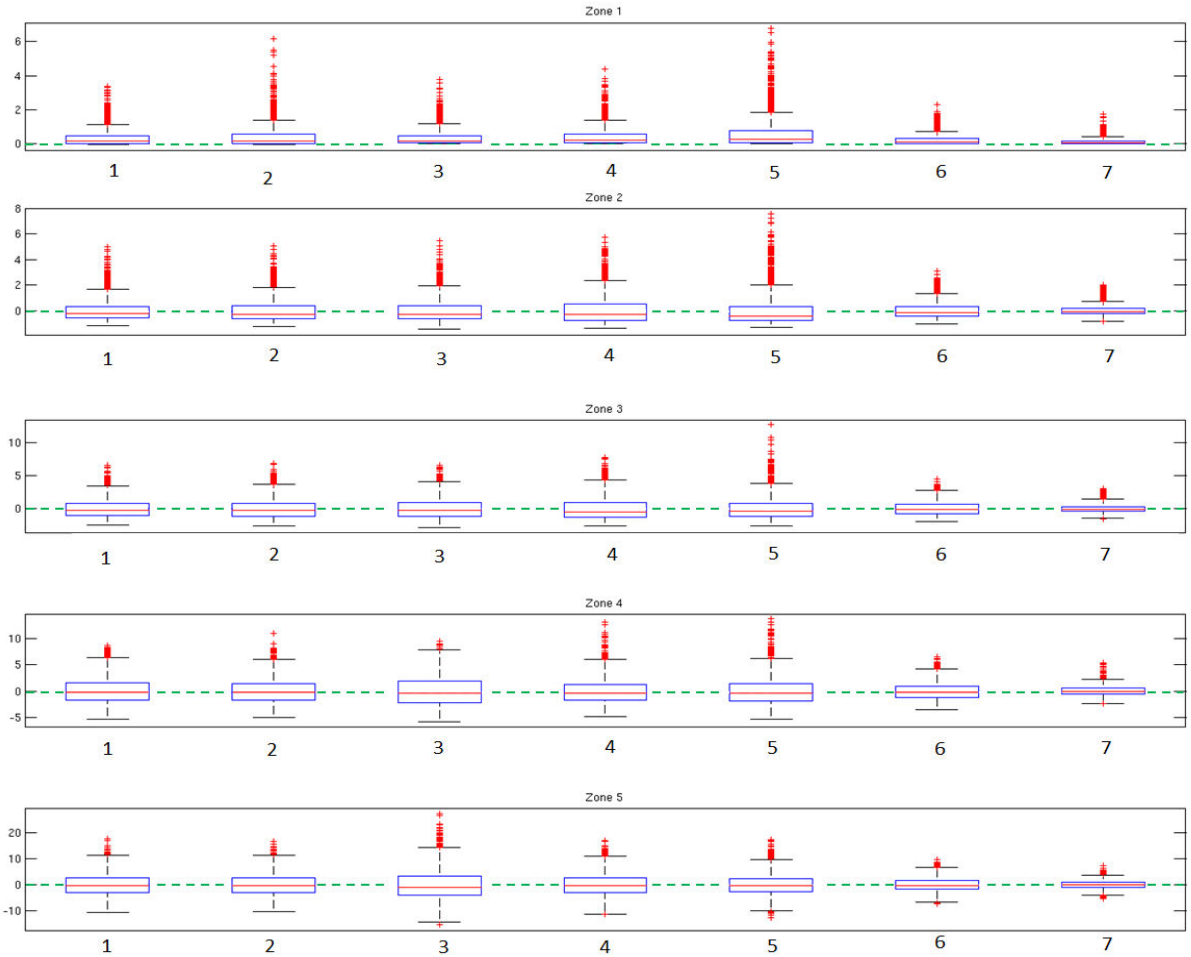

Figure 47 : Distributions des résidus par zone pour les proxys (1) STOOIP, (2) Volumes connectés, (3) Connectivité, (4) *Fast Marching*, (5) Lignes de courant, (6) Homogénéisation grossière et (7) Homogénéisation fine. La ligne rouge dans les boîtes représente la médiane, les barres extrêmes les quantiles Q5 et Q95.

2.2.c Comparaison de l'efficacité des proxys pour effectuer une sélection de modèles

Ce point est le dernier de notre méthodologie de comparaison. Son objectif est de différencier les proxys en fonction de leur capacité à sélectionner efficacement des modèles représentatifs, pour étudier l'impact des paramètres statiques des géomodèles sur les incertitudes de production. Comme expliqué précédemment (Parie I section 3), il faut d'abord constituer un sous-ensemble de modèles représentatifs à partir des simulations proxys de tous les modèles à disposition. Les simulations dynamiques de ces modèles sont ensuite utilisées pour déterminer les quantiles de production Q10, Q50 et Q90. Enfin, ces étapes sont répétés plusieurs fois afin d'obtenir autant de variantes de chacun des quantiles. Cela permet de construire un intervalle de confiance autour de chacun d'eux.

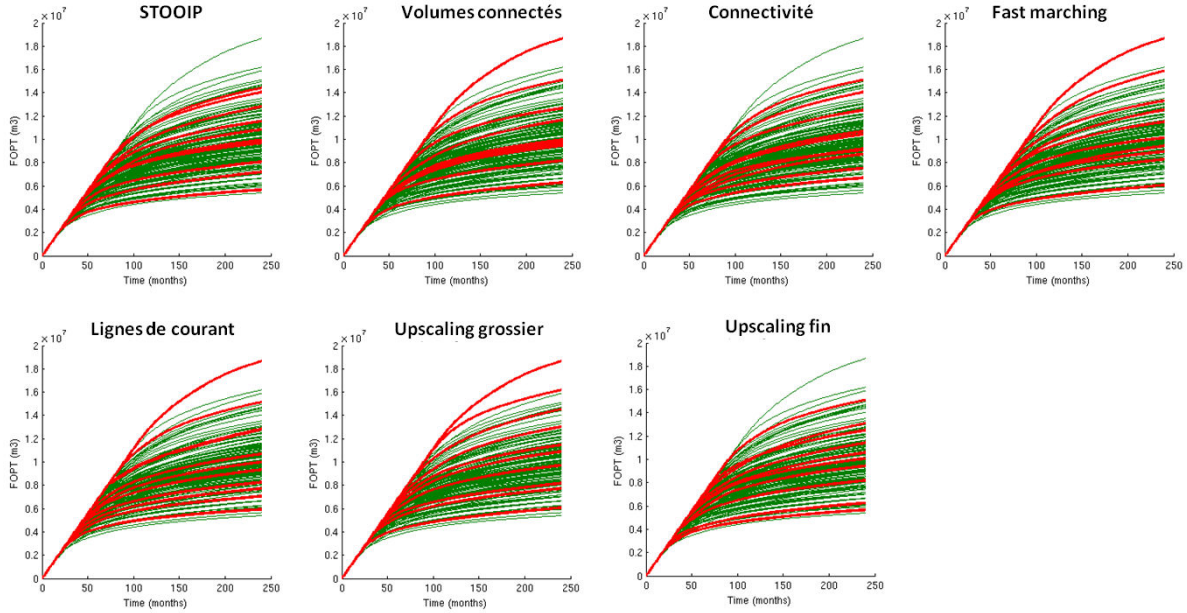

Figure 48 : Exemples de sélections de dix modèles parmi l'ensemble initial en fonction des réponses proxys. Les courbes rouges représentent les modèles sélectionnés, les vertes les autres.

La sélection de modèles par la méthode DKM (I.1) est un processus stochastique, c'est pourquoi la Figure 48 ne présente qu'une possibilité de sélection pour chaque proxy. Visuellement, cette sélection semble échantillonner correctement l'ensemble initial de profils. Il n'est toutefois pas possible de conclure sur cette base qu'un proxy est meilleur qu'un autre. Nous allons donc comparer les quantiles et leurs intervalles de confiance respectifs. Nous effectuerons en parallèle une sélection aléatoire de modèles afin de servir de référence à améliorer.

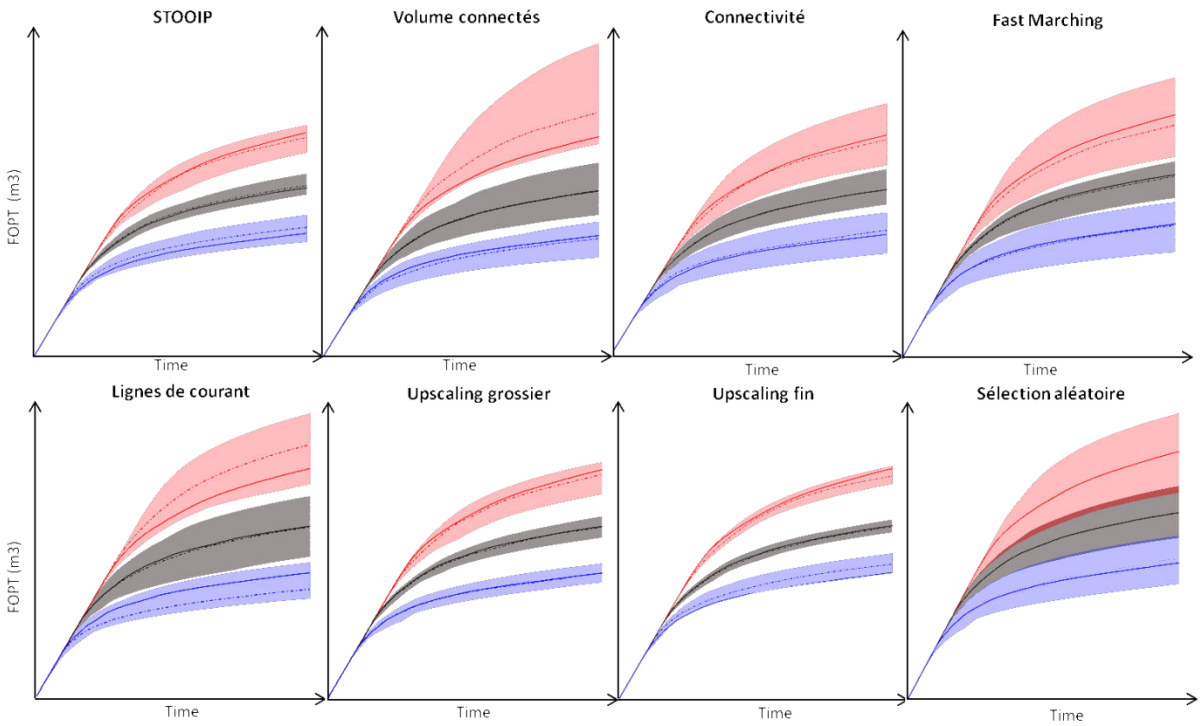

Figure 49 : Intervalles de confiance autour des différents quantiles (Q90 en rouge, Q50 en noir et Q10 en bleu) après une sélection pour chaque proxy. Les traits pleins correspondent aux vrais quantiles calculés sur l'ensemble des modèles, alors que les traits pointillés correspondent au quantile moyen après 200 sélections.

La Figure 49 montre les intervalles ainsi obtenus. Cette fois, les proxys se comportent très différemment. En effet, si tous les proxys permettent d'obtenir un intervalle de confiance qui englobe les vrais quantiles (calculés sur l'ensemble des modèles disponibles), ces mêmes intervalles ont des tailles très variées. Pour les proxys basés sur l'*homogénéisation* ou bien le *STOOIP*, les intervalles sont très resserrés, ce qui signifie que la sélection de modèles effectuée est représentative de l'ensemble des modèles. En revanche, pour les autres proxys les intervalles sont aussi larges, voire plus larges, qu'une sélection purement aléatoire, ce qui indique que ces proxys ne sont pas informatifs. Il faut aussi noter que pour les proxys *volumes connectés* et *lignes de courant*, il existe un écart significatif entre les vrais quantiles et la moyenne des quantiles obtenus par sélections successives. La sélection est donc non seulement non informative, mais aussi biaisée et donc dangereuse. Notons encore que, quel que soit le proxy utilisé, la sélection de modèles par *Distance Kernel Method* n'augmente pas les intervalles de confiance par rapport à une sélection aléatoire.

2.2.d Synthèse du choix des proxys sur le cas d'étude X

Il ressort de l'étude comparative présentée ci-dessus que le meilleur proxy pour étudier l'impact des paramètres statiques du géomodèle sur la production d'hydrocarbure est une homogénéisation fine. En effet, ce proxy reproduit précisément les distances entre modèles en termes de comportement dynamique. De plus, il réduit considérablement l'incertitude sur les quantiles de production s'il est utilisé à des fins de sélection de modèles. Enfin, il permet d'obtenir de nombreux résultats tels que la production d'huile et l'évolution de la pression. Par contre, il s'agit d'un proxy relativement lourd à mettre en place et qui ne donne pas de gain de temps très significatif. C'est pour cette raison que si le temps est un facteur très important, l'utilisation du proxy STOOIP représente pour notre réservoir X un très bon compromis, car il fournit aussi une bonne évaluation des distances entre modèles et de faibles intervalles autour des quantiles de production. Il est d'ailleurs étonnant de voir à quel point ce proxy est bon, alors que d'autres tels que les volumes connectés ou bien les lignes de courant, qui semblent pourtant mieux représenter le comportement physique du réservoir, fournissent de plus mauvais résultats. Nous avons donc effectué un dernier

graphique très courant dans l'industrie pétrolière, représentant les réserves en fonction des accumulations d'huiles présentes dans le géomodèle, soit la quantité d'huile produite à la fin de la simulation de référence (vingt ans) par rapport au STOOIP. La Figure 50 présente les résultats obtenus. Nous pouvons voir que la corrélation est très forte et linéaire (ce qui n'est pas toujours le cas). L'huile en place, à elle seule, permet d'expliquer fortement la réserve à vingt ans, ce qui peut expliquer les résultats précédents mais pose la question de la capacité des autres proxys (lignes de courant, propagation de front et volumes connectés) à approcher le comportement dynamique réel du champ.

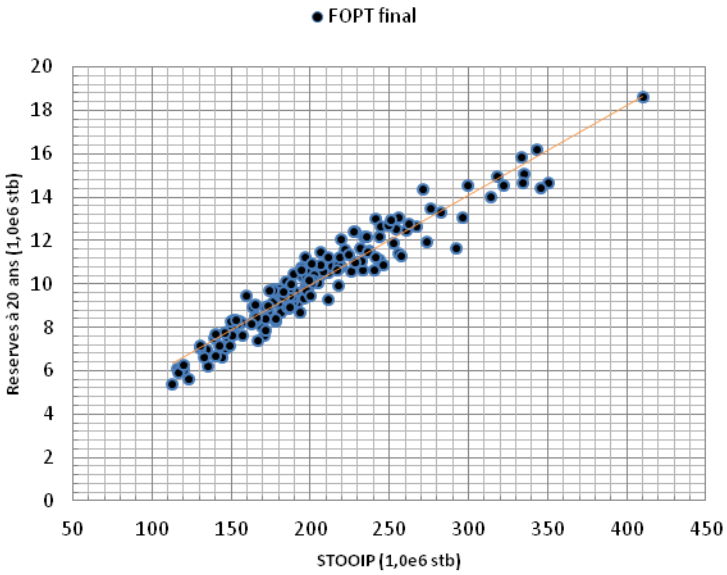

Figure 50 : Corrélation accumulations/réserves à vingt ans pour l'ensemble des modèles à notre disposition.

Maintenant que les différents proxys ont été comparés, nous allons appliquer la méthodologie de reconstruction des profils manquants. Cela va permettre d'améliorer l'étude de l'incertitude autour des quantiles de production, afin de pouvoir mieux étudier l'impact des paramètres statiques sur le comportement dynamique des modèles.

2.3 Amélioration du calcul des quantiles de production

Nous appliquons ici la méthodologie présentée dans la section I.4, afin de déterminer les paramètres d'une fonction paramétrique pouvant approcher un profil non simulé.

2.3.a Reconstruction des profils non simulés

Le terme non simulé signifie que le profil n'a pas été retenu lors de la sélection des modèles et que son comportement dynamique réel n'est théoriquement pas connu. Nous commençons par effectuer un calage du modèle analytique sur chacune des courbes de production issues de la simulation d'écoulements de référence des modèles sélectionnés. Cela nous permet d'obtenir plusieurs vecteurs de paramètres valides, parmi lesquels le point de départ à la procédure de minimisation pourra être choisi. Pour chaque proxy utilisé afin d'effectuer une sélection de modèles, la Figure 51 montre le résultat du calage du modèle analytique. Nous remarquons que, comme présenté dans la partie I chapitre 4, le modèle analytique permet de représenter parfaitement le profil de production. Nous appliquons ensuite la procédure de minimisation sur les dix modèles sélectionnés pour les reconstruire les uns par rapport aux autres (étape de cross-validation). Nous obtenons ainsi un profil analytique et un profil réel pour chacun d'entre eux. Comme la Figure 52 le montre, les profils analytiques sont visiblement différents des profils dynamiques et ceci quel que soit le proxy utilisé. Cela peut s'expliquer par le fait que l'algorithme de minimisation choisi n'accepte pas de tolérance vis-à-vis des contraintes. Il peut donc converger vers un minimum local. Il serait

donc intéressant de tester un autre algorithme (par exemple évolutionnaire) permettant de « ressortir » de ce genre de situation. Malgré cela, à quelques exceptions près, le profil analytique est suffisamment proche du profil de référence correspondant pour être identifié visuellement.

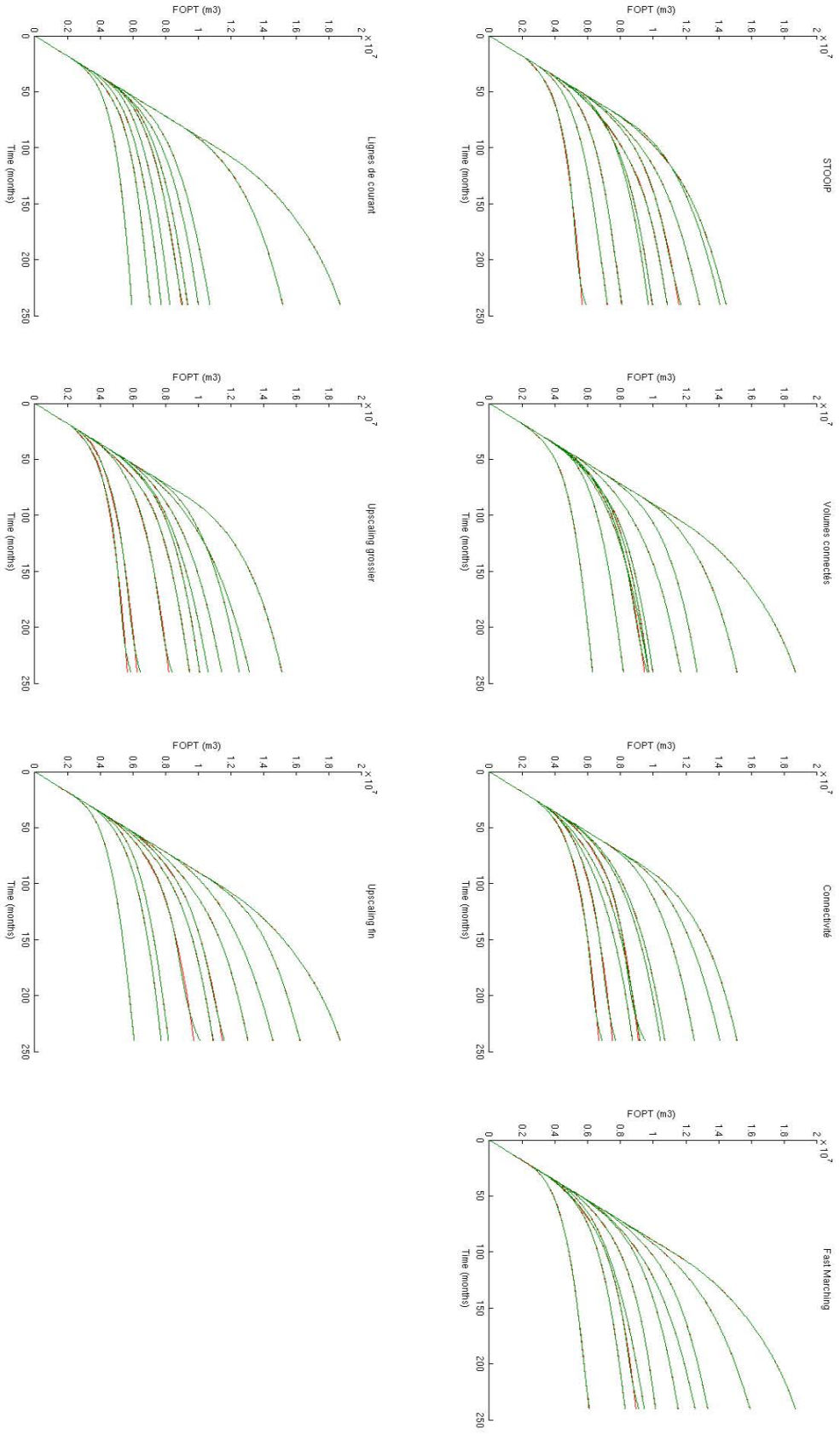

Figure 51 : Calage des courbes de production (en rouge) à l'aide du modèle analytique choisi (en vert) pour chaque proxy.

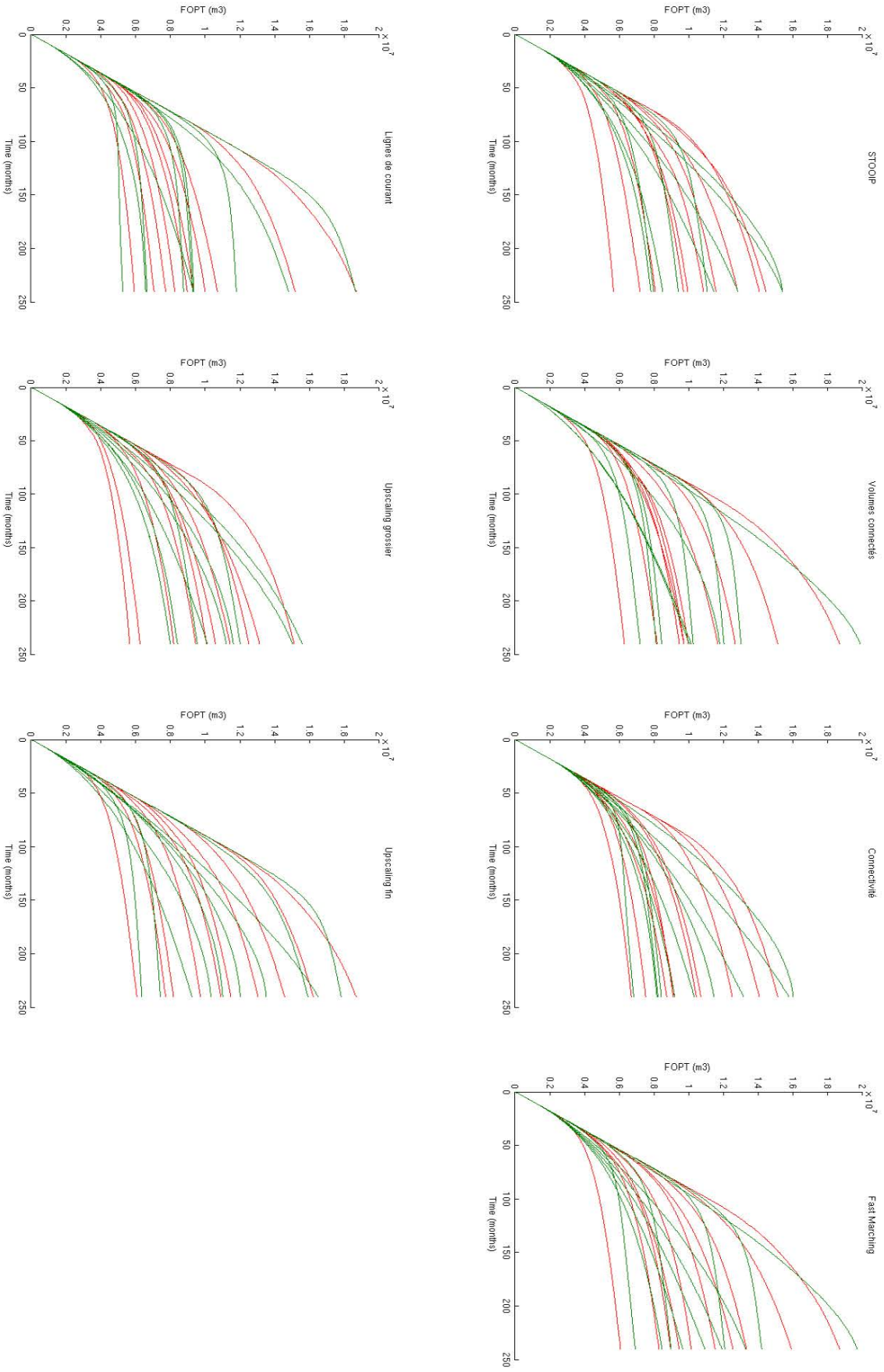

Figure 52 : Profils réels (en rouge) et profils analytiques (en vert) pour l'ensemble des modèles sélectionnés par proxy.

En calculant la différence entre les deux profils de chaque modèle et en la projetant sur les profils analytiques, nous obtenons les cartes d'erreurs qui permettent de corriger les profils analytiques des modèles non simulés que nous allons estimer. Cette procédure permet d'obtenir un ensemble de courbes de natures différentes : 10 profils calculés par le simulateur d'écoulement haute fidélité correspondant aux géomodèles sélectionnés et 190 profils analytiques estimés correspondants aux autres géomodèles. La Figure 53 présente un exemple de profils obtenus pour chaque géomodèle et avec chaque proxy. Cette fois encore, il est assez difficile de différencier visuellement les proxys. Or, la procédure de minimisation utilise les réponses proxys et nous avons vu précédemment que leur comportement (les corrélations linéaires et de rang, ainsi que la distribution des résidus) était significativement différent. Notre procédure semble donc pouvoir être utilisée pour autant que le proxy ait un minimum de relation avec le simulateur haute fidélité en termes de distances entre modèles. En effet. Avec seulement dix modèle sélectionnés, les nuages de points utilisés pour les régressions linéaires ne comportent que 45 points et chaque zone que 9 d'entre eux.

Enfin, nous calculons les quantiles de production Q10, Q50 et Q90 pour chaque ensemble de modèles. Nous avons répété deux cents fois l'ensemble des étapes pour chaque proxy. Nous avons ainsi pu de nouveau construire des intervalles autour des trois quantiles. La Figure 54 montre les résultats obtenus pour chaque proxy. Nous constatons que ces résultats sont similaires d'un proxy à l'autre pour la taille des intervalles. Par contre, un léger biais est observé pour le quantile pessimiste (Q10) qui se trouve toujours légèrement surévalué pour les proxys de volumes connectés, connectivité et *Fast Marching*. Ces intervalles sont à confronter à ceux obtenus lors de la comparaison des proxys (2.2.c). Ils sont beaucoup plus étroits quand les quantiles sont calculés avec l'ensemble des informations des modèles.

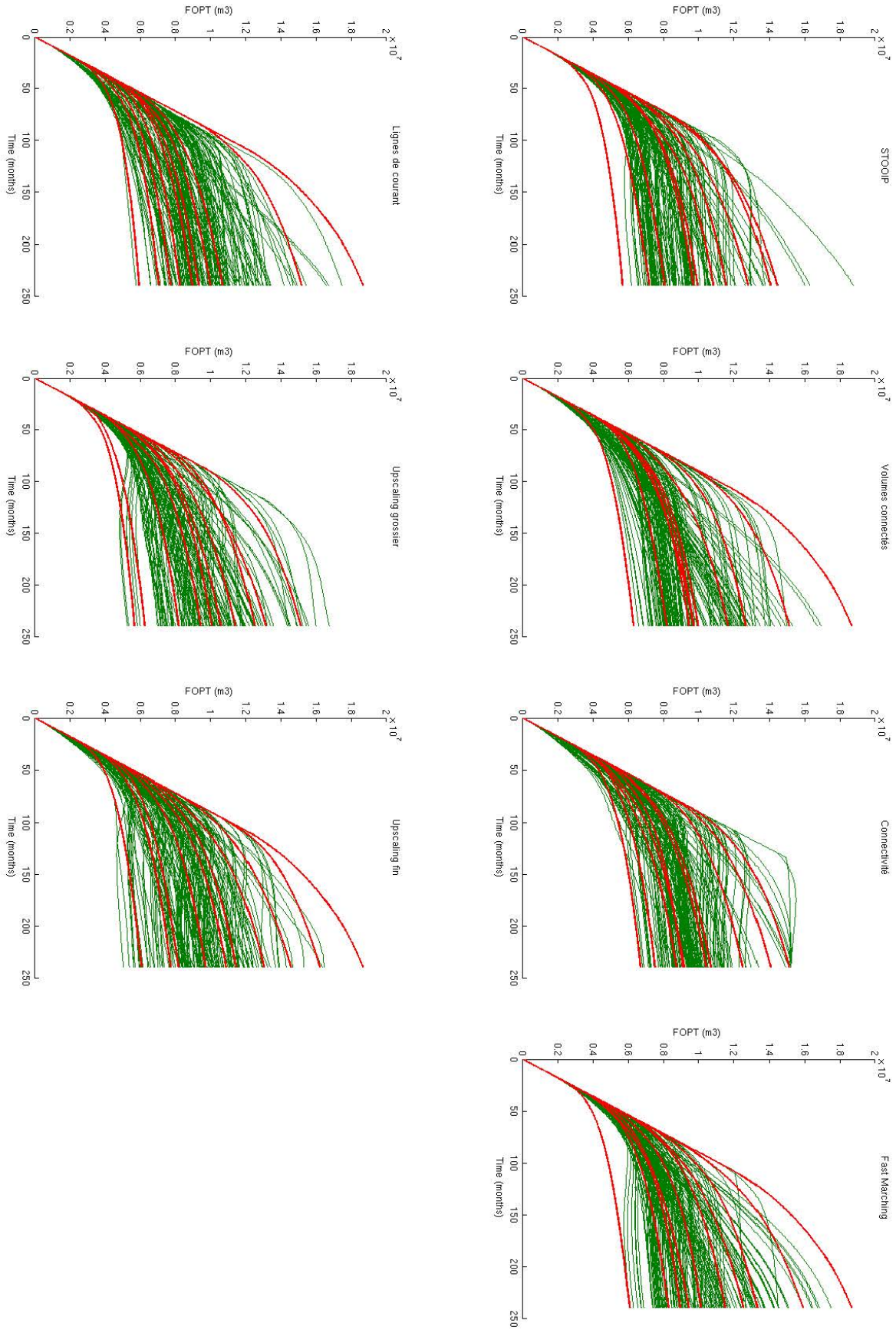

Figure 53 : Ensemble de profils obtenus pour chaque proxy. Les courbes rouges représentent ceux obtenus par simulation dynamique et les courbes vertes, ceux obtenus par minimisation.

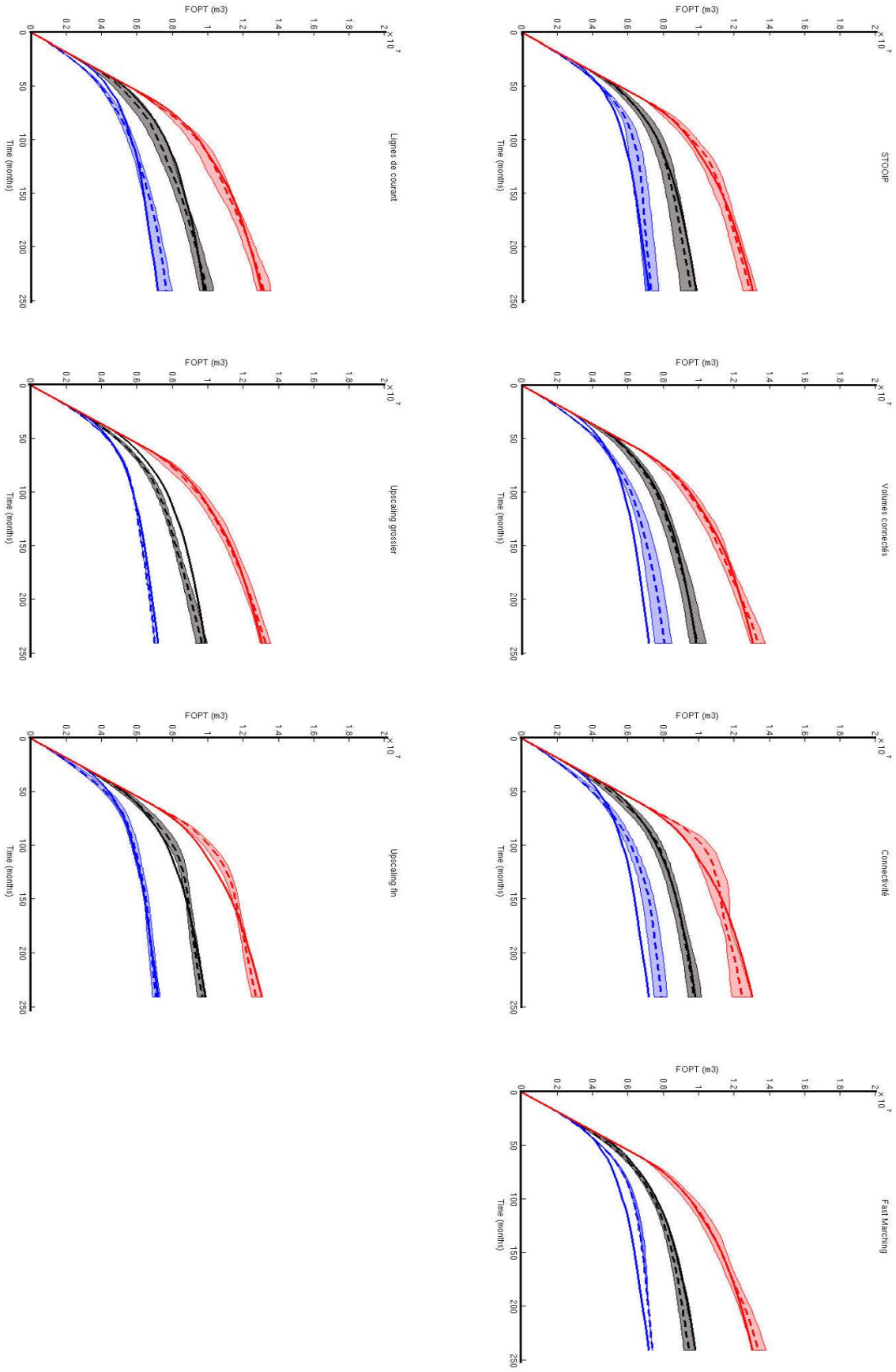

Figure 54 : Intervalles de confiance autour des quantiles de productions après deux cents répétitions de minimisation pour chaque proxy.

2.3.b Synthèse de la procédure de minimisation sur le champ X

Ce cas d'étude démontre l'intérêt d'une telle procédure par rapport à une évaluation des quantiles de production basée uniquement sur les modèles effectivement simulés, quel que soit la qualité du proxy. En effet, les quantiles obtenus étant plus proches de la réalité, il est plus facile d'évaluer l'impact des incertitudes statiques des géomodèles sur les prévisions de production. Il pourrait être intéressant ici de réduire le nombre de modèles sélectionnés pour voir si cela a un impact sur la précision des quantiles statistiques estimés. Si c'est le cas il s'agit d'une limite de notre procédure, sinon le nombre de simulations haute fidélité à réaliser peut encore être réduit pour gagner du temps. La méthode appliquée ci-dessus fonctionne quel que soit le proxy utilisé, il se pose donc toujours la question de la pertinence des proxys plus complexes que le STOOIP mais qui ne donnent pas de meilleurs résultats. Afin d'essayer de répondre à cette question, il serait intéressant de voir ce qui se passe pour un cas d'étude plus complexe où le choix des proxys à utiliser est plus limité et où la capacité de ces proxys à reproduire le comportement du simulateur haute fidélité est réduite.

Chapitre 3 : Application au champ Y

Ce dernier chapitre est consacré au deuxième cas d'étude testé, qui présente des caractéristiques différentes. Après une description du nouveau champ, nous comparerons cinq proxys différents et appliquerons la procédure de reconstruction des profils non simulés avec chacun d'entre eux. Cela nous permettra de justifier et d'approfondir nos conclusions.

Ce deuxième cas d'application également fourni par Total S.A., bien que proche géologiquement du précédent, a été choisi car sa corrélation accumulation d'huile en place/réserves à vingt ans est plus mauvaise (Figure 55). Nous allons de nouveau effectuer une comparaison de différents proxys et une reconstruction de profils dynamiques manquants, afin de voir si les méthodes que nous avons développées fonctionnent toujours et si nous pouvons répondre à certaines des questions posées concernant la validité de nos méthodes ou la pertinence de certains proxys.

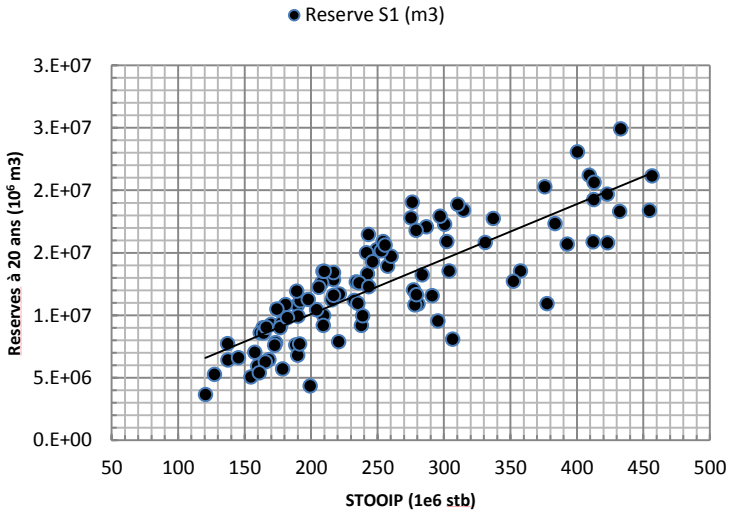

Figure 55 : Corrélation accumulation/réserves à vingt ans pour le deuxième cas d'étude.

3.1 Présentation générale

Le cas d'étude Y est aussi un réservoir turbiditique. Le nombre de mailles est cette fois-ci beaucoup plus grand ($184 \times 398 \times 160$), soit presque douze millions de cellules dont deux millions environ sont actives. Bien que le nombre de cellules du géomodèle soit beaucoup plus important, le volume total de chaque cas d'étude est assez semblable. Cela s'explique par la taille des cellules qui sont cette fois beaucoup plus petites ($\pm\ 20m$ à l'horizontale). Pour produire ce volume, le schéma de développement est lui aussi relativement proche du cas X, avec quatre puits injecteurs d'eau et cinq puits producteurs. Cette fois encore, nous disposons aussi d'un ensemble de 200 géomodèles construits pour étudier les incertitudes des paramètres statiques. Chaque modèle diffère des autres par sa géométrie, du fait de l'incertitude sur la position des surfaces géologiques et sa pétrophysique, car les propriétés sont générées à l'aide de simulations géostatistiques contraintes aux données de puits et aux données sismiques (La Figure 56). On remarquera l'organisation plus complexe des corps géologiques.

La grosse différence entre les deux cas d'étude est l'information apportée par l'imagerie sismique. Dans le cas X nous avons une image de bonne qualité, alors que dans le cas Y l'image est

nettement moins bonne. Ceci explique la moins bonne corrélation entre les accumulations et les réserves. C'est pour cette même raison que les simulations géostatistiques utilisées pour construire les géomodèles sont différentes.

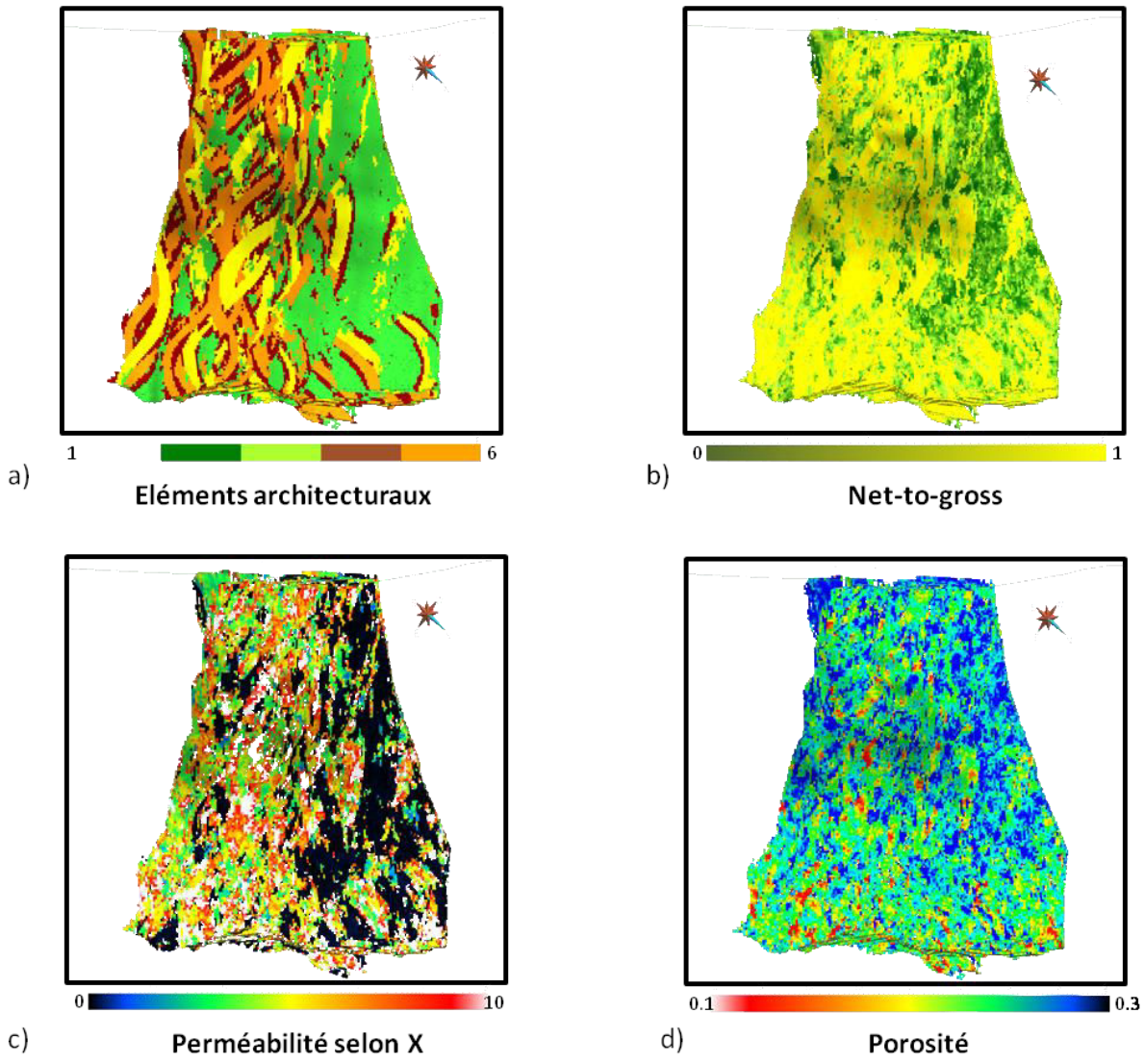

Figure 56 : Répartition des propriétés pétrophysiques dans un des modèles du deuxième cas d'étude.

Du fait du très grand nombre de cellules et de la taille des fichiers d'export correspondants, il n'a pas été possible cette fois de travailler avec plus de cent modèles. De plus, il a fallu adapter le code informatique pour passer outre une limitation de la base de données mise en œuvre qui n'autorise pas de stocker un objet de plus de 2Go. La dernière différence majeure entre ce cas d'étude et le précédent est que les simulations de référence ont été obtenues en utilisant le logiciel Intersect®, aussi développé par Schlumberger, car celui-ci est plus rapide (quarante minutes au lieu de plus de deux heures). Ce simulateur n'avait pas été retenu lors du premier cas d'étude, car il n'était pas encore accessible dans la société à ce moment de la thèse. Les résultats obtenus sont quasiment identiques à ceux fournis par Eclipse® et le gain de temps est appréciable.

Cette fois-ci, nous n'utiliserons que cinq proxys différents :

- Le STOOIP
- Les volumes connectés aux puits producteurs,
- La propagation de front par connectivité
- La propagation de front par *Fast Marching*

- Un seul ratio d'upscaling

Ce plus petit nombre de proxys s'explique par un manque de temps ne permettant pas de calculer tous les proxys, notamment ceux avec un important travail préalable (homogénéisation et lignes de courant). L'environnement géologique étant le même que celui du cas précédent, le seuil de perméabilité utilisé pour les volumes connectés ou la propagation de front par connectivité est identique à celui employé pour le cas X, à savoir 10mD. Pour le proxy de propagation de front par *Fast Marching*, la même fonction de vitesse que dans le premier cas a également été choisie (rapport porosité sur perméabilité). Enfin, le seul ratio d'homogénéisation utilisé est de $3 \times 3 \times 3$.

3.2 Comparaison des différents proxys

Comme pour le cas d'étude précédent, nous allons d'abord confronter les proxys les uns aux autres en utilisant notre méthodologie de comparaison en trois points.

3.2.a Mise en place, réponse et gain de temps

Les proxys utilisés dans ce nouveau cas d'étude sont pour la plupart identiques aux précédents. Seul le ratio du proxy d'upscaling changeant, les mises en place des différents proxys sont donc similaires. Nous allons de ce fait directement présenter les résultats des simulations proxys. La Figure 57 présente les histogrammes de distribution des proxys ne retournant qu'une seule valeur par modèle (STOOIP et volumes connectés aux puits). Contrairement au premier champ, la ressemblance visuelle entre les deux distributions est plus faible. Il y a notamment une baisse des fréquences pour les valeurs fortes dans le cas des volumes connectés. Cela peut s'expliquer par l'existence d'une barrière à l'écoulement (zone de faible perméabilité) dans un certain nombre de modèles, créant ainsi une accumulation d'huile non connectée aux puits producteurs qui diminue la réponse au proxy volume connecté de ces modèles

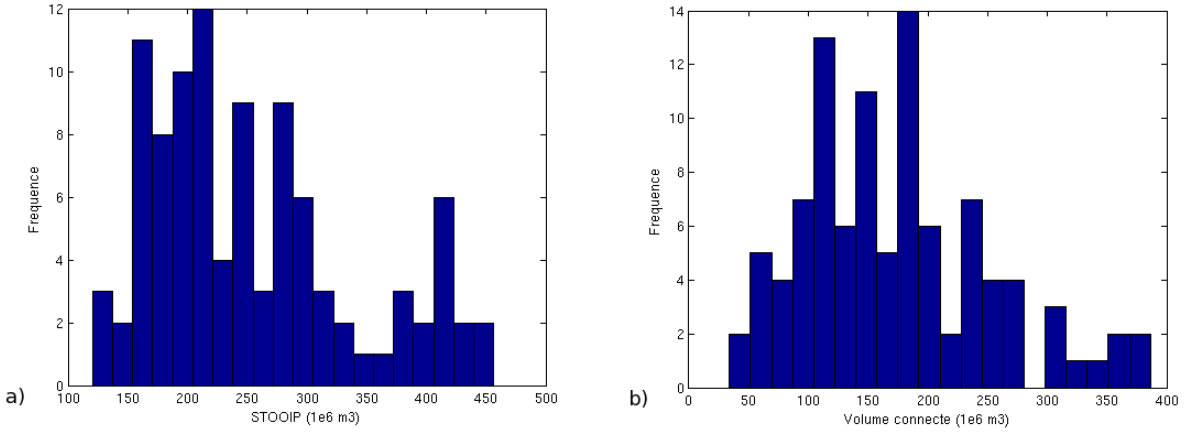

Figure 57 : Histogrammes de distribution des résultats de simulation pour les proxys (a) STOOIP et (b) volumes connectés aux puits.

En ce qui concerne les proxys permettant d'obtenir une courbe de réponse (Figure 58), comme pour le cas d'étude X, l'homogénéisation fournit des réponses dont la forme est la plus proche de celle des simulations d'écoulements de référence. Par contre, cette fois-ci une différence de dispersion plus importante entre les réponses est visible, même dans le cas de l'homogénéisation.

Il n'est pas nécessaire de présenter un tableau indiquant les paramètres utilisés ainsi que le gain de temps, car celui-ci serait identique à celui obtenu auparavant (Tableau 2). Ici le gain de temps de l'homogénéisation est plus proche de celle de l'homogénéisation fine précédente. En effet, bien que le temps de simulation soit grandement réduit (9min au lieu de 45min en moyenne), le temps de traitement préliminaire est lui supérieur à cause de la taille du modèle (la grande quantité de mémoire RAM nécessaire pour le logiciel ne permet de ne traiter qu'un seul modèle à la fois).

Comme dans le premier cas d'étude, il est difficile de choisir un proxy avec ces seuls résultats. Par contre, étant donnés la taille des géomodèles et les temps de simulations dynamiques, déterminer le bon proxy est particulièrement important.

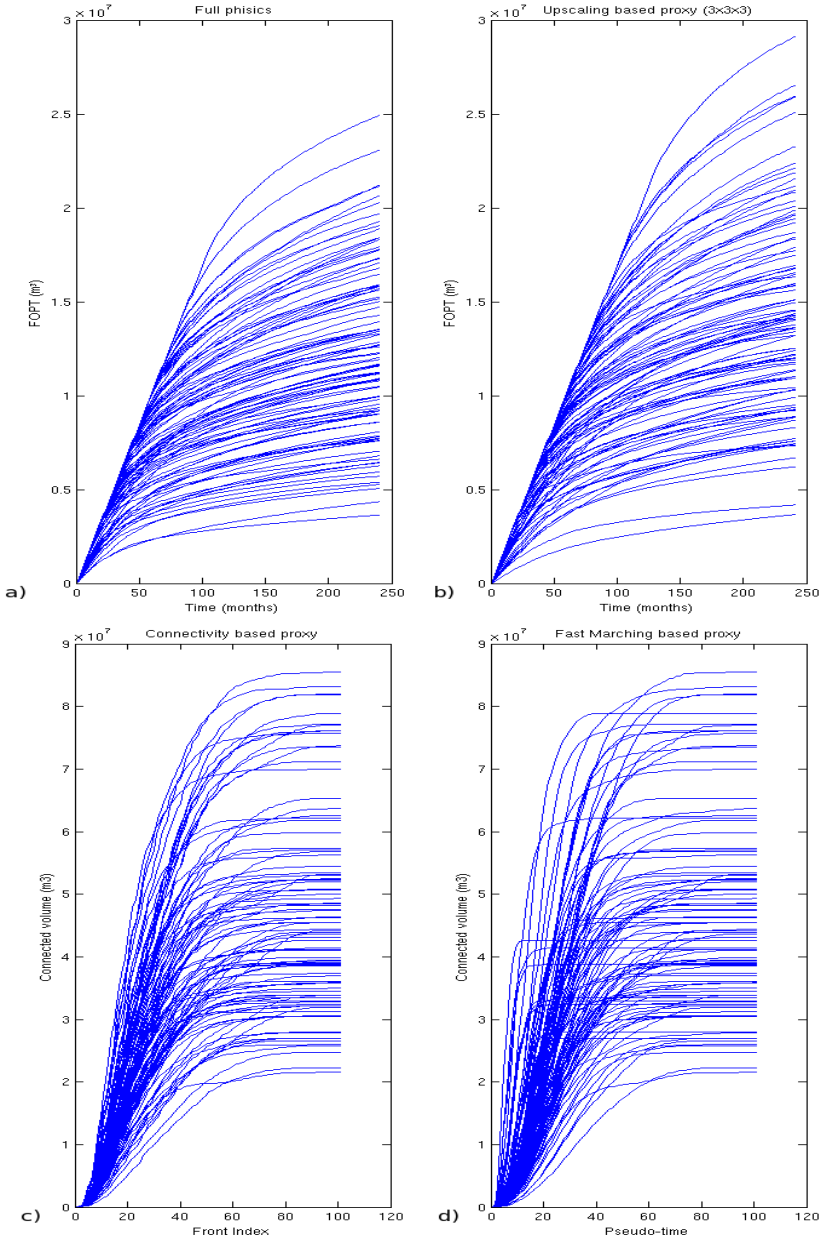

Figure 58 : Réponses sur les modèles du deuxième cas d'étude pour (a) les simulations dynamiques, et les proxys (b) d'upscaling, (c) de connectivité et (d) de Fast Marching.

3.2.b Comparaison distances proxys et distances de référence

Nous comparons maintenant les nuages de corrélation distances-distances pour chacun des proxys, afin de voir si dans le champ Y les proxys choisis permettent de reproduire les distances entre les modèles en termes de simulations d'écoulements haute fidélité. Après une première analyse des graphiques (Figure 59), plusieurs points méritent d'être relevés. Tout d'abord, il semble toujours y avoir une corrélation entre les distances proxys et les distances dynamiques. En revanche, contrairement au cas d'étude précédent, ces corrélations sont plus faibles excepté pour le proxy basé sur l'homogénéisation des modèles. Ce point est illustré par les données du Tableau 4, où tous les coefficients sont supérieurs à 0.50 mais ne dépassent pas 0.77 sauf pour l'homogénéisation.

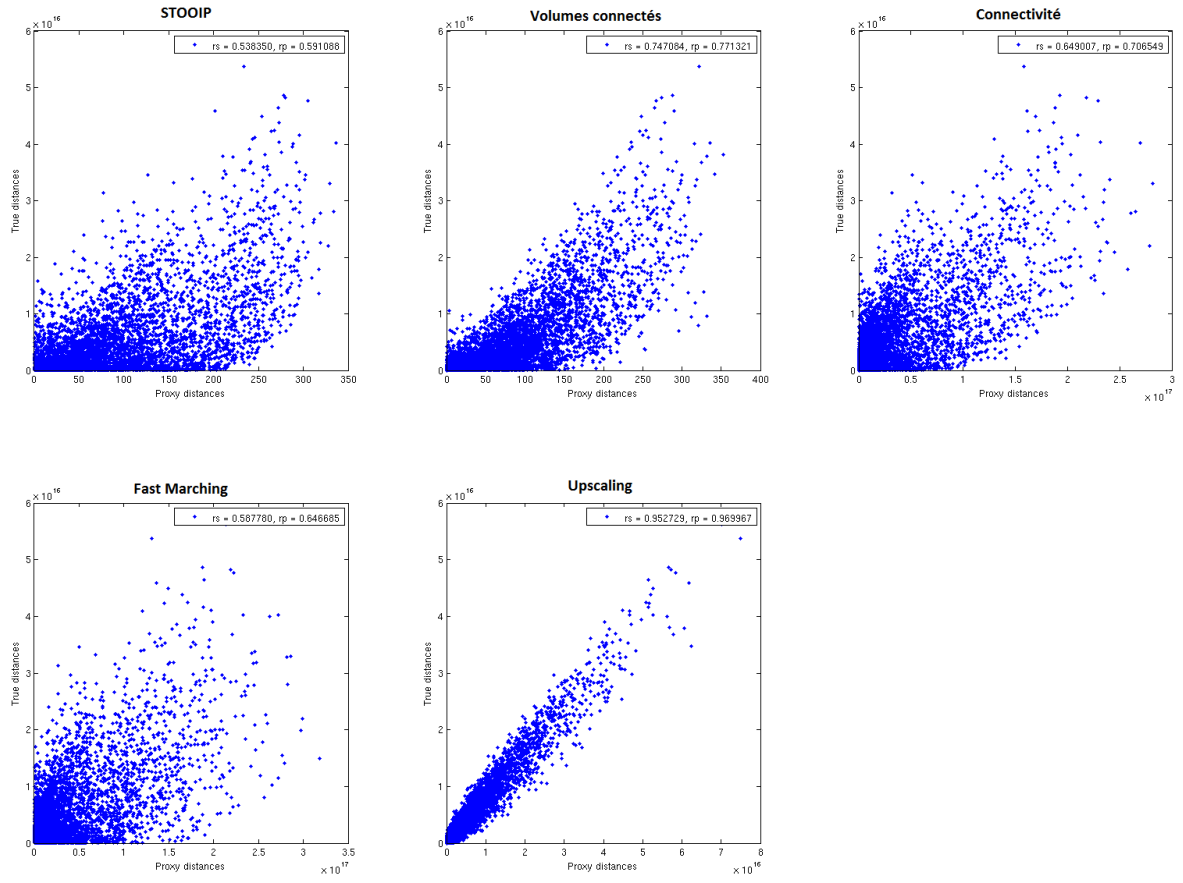

Figure 59 : Nuages de corrélation distances-distances pour l'ensemble des proxys appliqués au cas d'étude.

Comme dans le cas précédent, il est intéressant de remarquer que la corrélation entre les distances n'est pas linéaire pour les deux proxys les plus simples.

Proxy	Coefficient de corrélation de Pearson	Coefficient de corrélation de Spearman
STOOIP	0.59	0.54
Volumes connectés	0.77	0.75
Connectivité	0.71	0.65
Fast Marching	0.65	0.59
Upscaling	0.97	0.95

Tableau 4 : Coefficients de corrélation distances dynamiques - distances proxys pour les différents proxys utilisés.

On remarque cette fois-ci que le proxy volumes connectés aux puits est nettement plus informatif que le STOOIP. Cela répond donc à la question de la pertinence de certains proxys que nous nous posions.

Afin de continuer l'analyse des graphiques, nous effectuons la régression linéaire par morceaux pour chacun des proxys (Figure 60) en divisant chaque graphique en cinq zones d'environ mille points. Nous pouvons donc calculer pour chaque point son résidu par rapport à la droite de régression (Figure 61). Ces nouveaux résultats confirment nos observations sur les nuages de corrélation : seul le proxy homogénéisation donne une très faible dispersion des résidus quelle que

soit la zone observée. Parmi les autres proxys, les volumes connectés et la connectivité semblent avoir une dispersion légèrement plus faible que les deux autres.

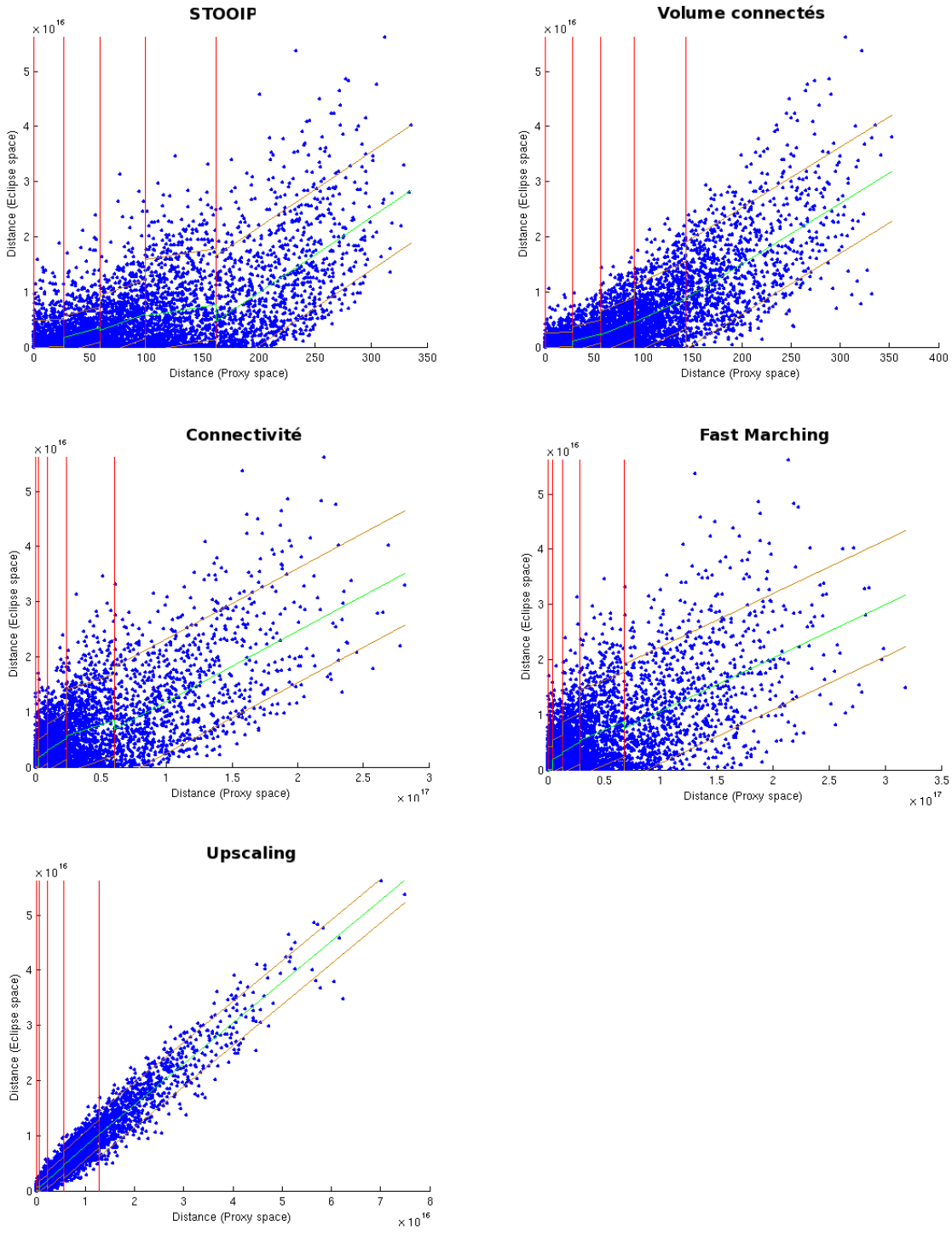

Figure 60 : Régression par morceaux pour chacun des crossplots distances-distances.

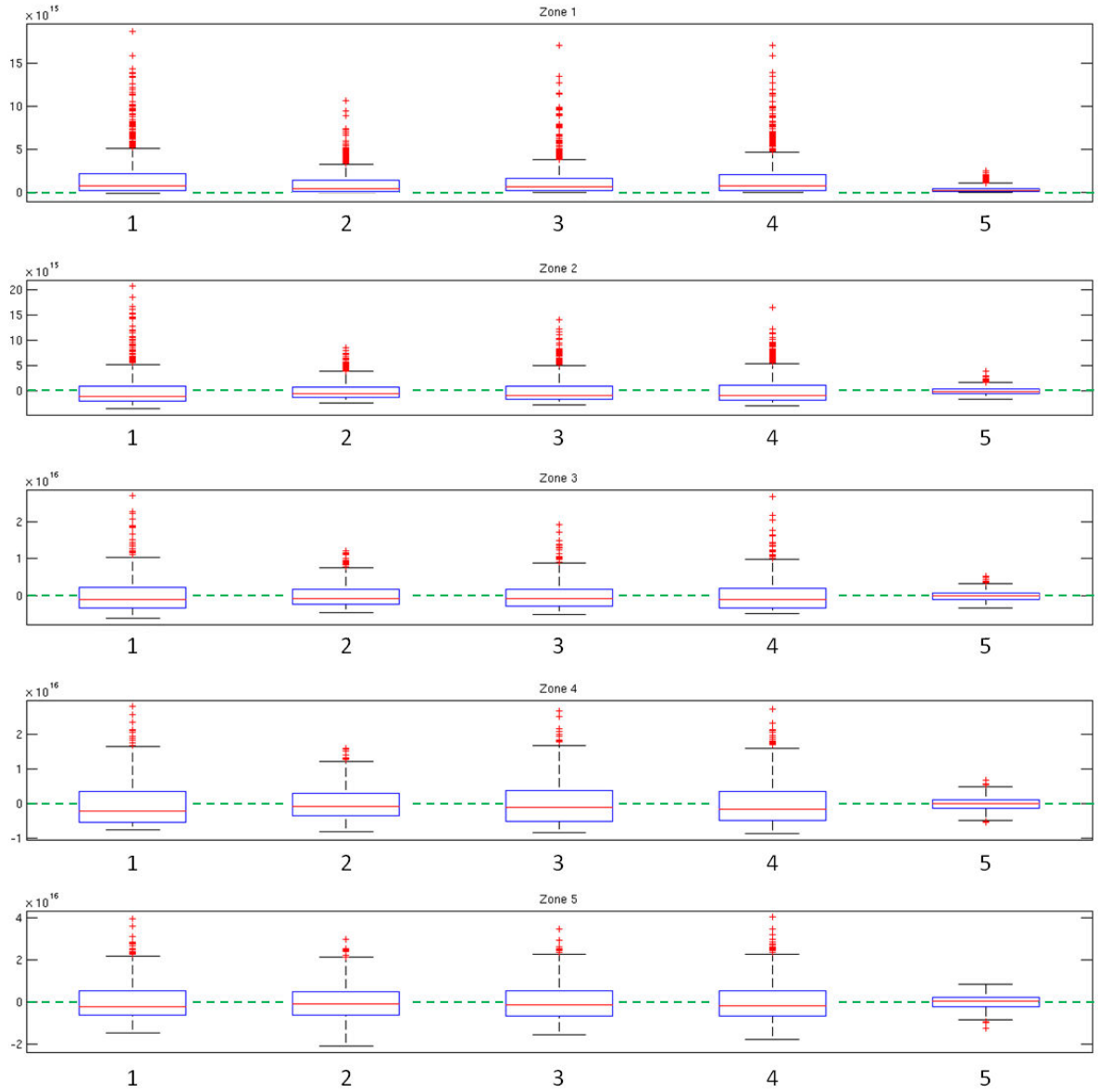

Figure 61 : Distributions des résidus par zone pour les proxys (1) STOOIP, (2) Volumes connectés, (3) Connectivité, (4) Fast Marching et (5) Upscaling.

A partir de ces résultats, nous pouvons déjà tirer deux conclusions. Le proxy par homogénéisation reste bien meilleur que les autres. En revanche, s'il faut choisir un autre proxy plus simple, nous retiendrons plutôt les volumes connectés.

3.2.c Comparaison de l'efficacité des proxys pour effectuer une sélection de modèles

Le dernier point de notre méthode de comparaison est la pertinence à utiliser nos proxys pour effectuer une sélection de modèles grâce à la technique *DKM*. Comme dans le cas d'étude précédent, nous cherchons à ne retenir que dix modèles parmi ceux à notre disposition pour ensuite calculer les quantiles de production.

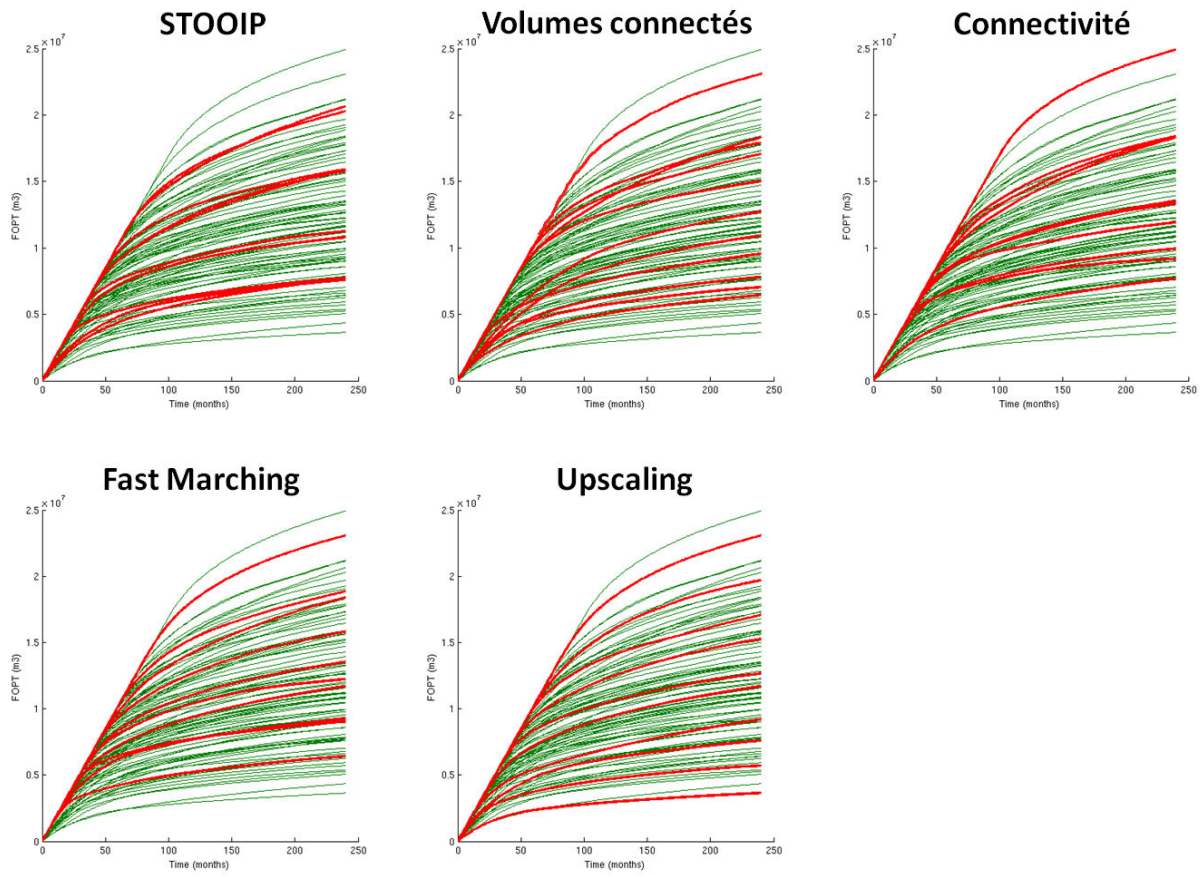

Figure 62 : Exemple de sélection de modèles pour chaque proxy. Les courbes rouges représentent les réponses de modèles sélectionnés et les vertes les autres.

La Figure 62 montre un exemple de sélection de modèles pour chaque proxy. Nous constatons une fois de plus que les proxys n'ont pas le même comportement. En effet, pour les deux proxys STOOIP et propagation de front par connectivité, l'échantillonnage ne semble pas représentatif. Comme le processus est stochastique, nous effectuons cette sélection deux cents fois afin d'obtenir des intervalles de confiance autour des trois quantiles de production (Figure 63). Si nous comparons la largeur des intervalles, nous nous apercevons que cette fois-ci il n'y a pas de proxy nettement meilleur que les autres. Nous pouvons donc en déduire que les exemples de sélection que nous avons obtenus ne sont pas généralisables sur plusieurs sélections. Il est aussi intéressant de remarquer que la sélection aléatoire est biaisée au niveau du quantile Q90. Pour celui-ci le quantile moyen est loin du quantile théorique, ce qui est surprenant pour une sélection aléatoire. Cela peut s'expliquer par un plus petit nombre de profils disponibles (100 contre 200 dans le cas X) et une plus grande variabilité. Un nombre de répétitions plus important devrait diminuer ce biais. Quoi qu'il en soit, aucun des proxys utilisés n'est moins informatif qu'une sélection aléatoire, mais aucun ne permet d'obtenir un sous-ensemble de modèles représentatifs de manière systématique avec la méthode DKM.

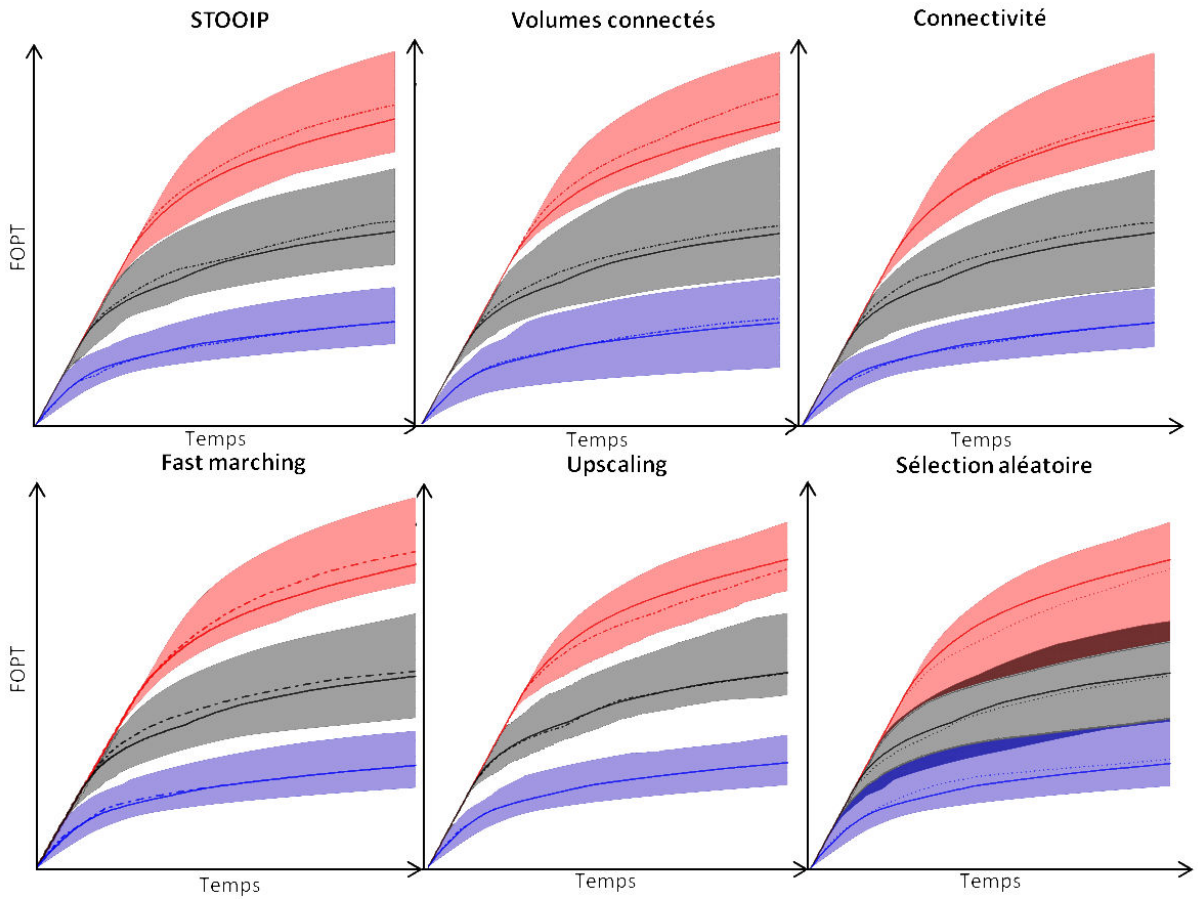

Figure 63: Intervalles de confiance autour des différents quantiles (Q90 en rouge, Q50 en noir et Q10 en bleu) après une sélection pour chaque proxy. Les traits pleins correspondent aux vrais quantiles calculés sur l'ensemble des modèles, alors que les traits pointillés correspondent au quantile moyen après deux cents sélections.

3.2.d Synthèse de la comparaison des proxys sur le champ Y

La comparaison des proxys sur ce deuxième cas d'étude nous permet de tirer plusieurs conclusions :

- Parmi tous les proxys que nous avons testés, celui basé sur l'homogénéisation semble donner les meilleurs résultats.
- S'il n'est pas possible de mettre en place une homogénéisation convenable, l'emploi d'un proxy beaucoup plus simple et rapide tels que les volumes connectés ou la propagation de front semble être une bonne alternative.

En comparaison avec le cas d'étude précédent, nous pouvons ajouter que la bonne corrélation accumulation/réserves du champ X semble être une explication plausible quant au fait que le proxy STOOIP fournisse d'aussi bons résultats. Dans le cas où cette corrélation est moins évidente, l'intérêt d'un proxy plus évolué comme les volumes connectés ou les propagations de front est indéniable.

3.3 Amélioration du calcul des quantiles de production

3.3.a Reconstruction de profils non simulés

Pour finir nous appliquons la technique de reconstruction des profils dynamiques par minimisation sur ce nouveau cas d'étude, bien que les résultats proxys semblent moins convaincants. Comme précédemment, le but est de sélectionner dix modèles représentatifs parmi tous ceux disponibles en utilisant la méthode *DKM*. Les profils des autres modèles sont ensuite déterminés grâce à la méthode d'optimisation proposée. D'après la forme de quelques profils dynamiques

obtenus et étant donné que le schéma de développement du réservoir est le même que pour le cas d'étude précédent, nous pouvons utiliser le modèle analytique à deux structures. Les résultats du calage sont présentés sur la Figure 64 et confirment notre choix.

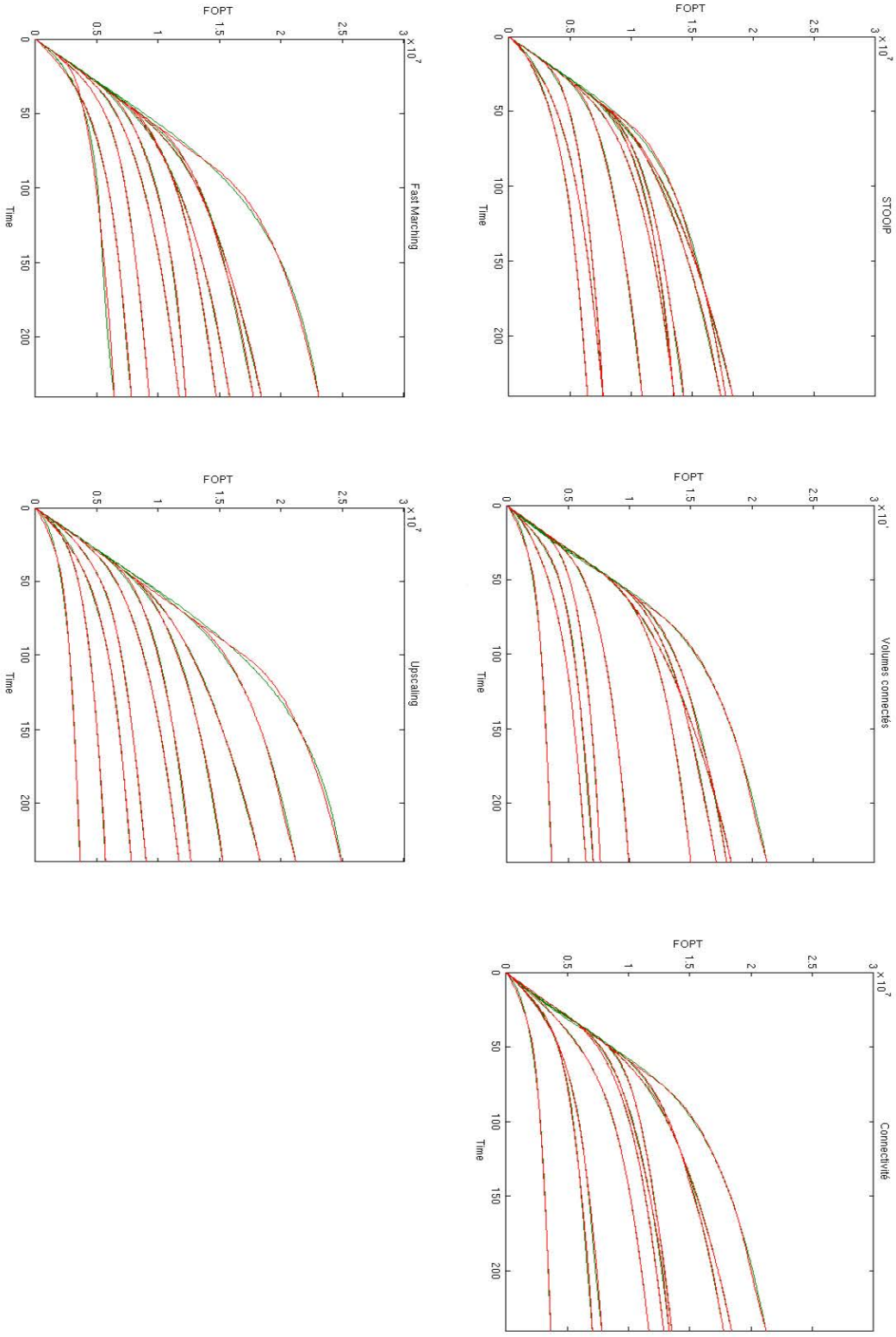

Figure 64 : Calage de modèles analytiques (en vert) sur les profils dynamiques (en rouge) des modèles sélectionnés pour chaque proxy.

En effet, à part de très légers écarts au niveau de la courbure sur quelques profils, les deux familles de courbes se superposent parfaitement. Après cela, nous déterminons les profils analytiques qui correspondent aux modèles sélectionnés (Figure 65) afin de pouvoir construire les cartes d'erreur et de posséder des ensembles de paramètres de départ.

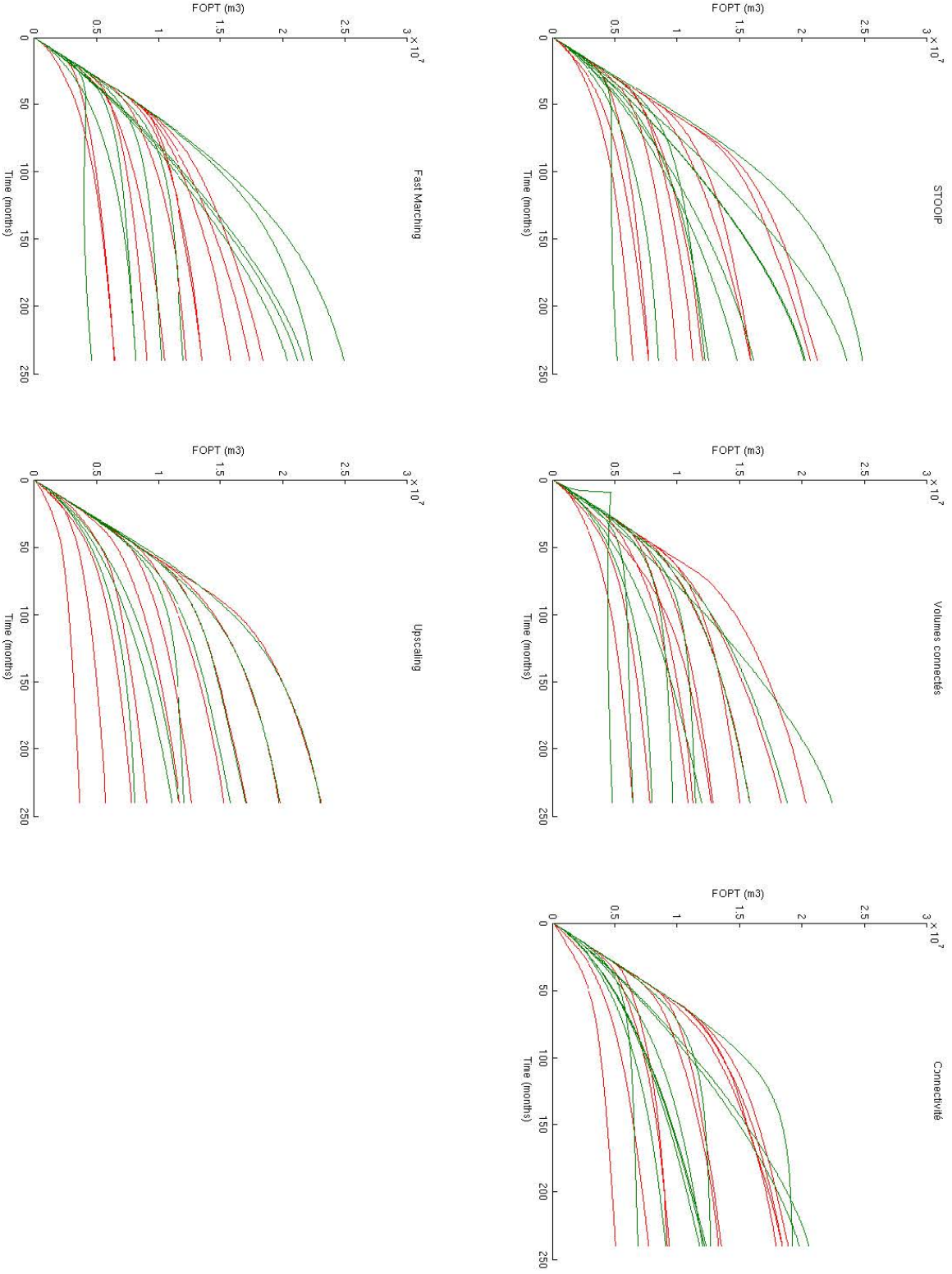

Figure 65 : Détermination des paramètres des modèles analytiques (en vert) pour les modèles sélectionnés dont les profils dynamiques sont connus (en rouge).

Comparé aux résultats que nous avions obtenus pour le premier cas d'étude, il est plus difficile cette fois-ci d'associer les deux courbes de chaque modèle. De plus, nous pouvons voir apparaître des profils qui ne respectent plus les contraintes et décroissent localement (les profils les plus bas pour les proxys STOOIP et *Fast Marching,* ainsi que le profil haut pour la connectivité).

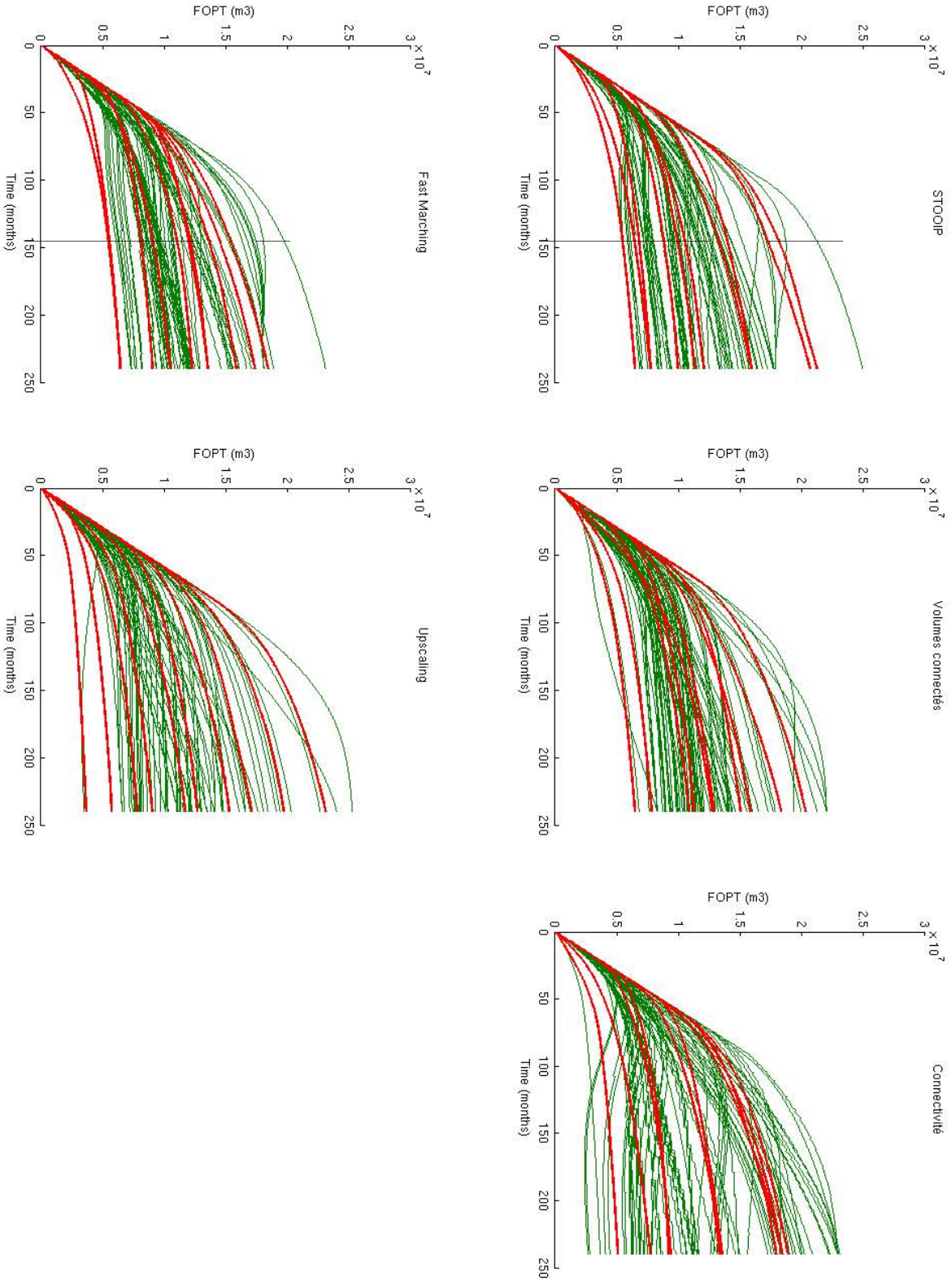

Figure 66 : Exemples de profils (en vert) obtenus après minimisation et correction pour chaque proxy. Les courbes rouges représentent les dix modèles réellement simulés.

Malgré ces écarts, une reconstruction de profils pour l'ensemble des modèles non simulés est effectuée. La Figure 66 présente un exemple de profils qu'il est possible d'obtenir pour chacun des proxys utilisés. Cette fois encore, nous retrouvons des profils qui ne sont plus strictement croissants. Bien que ce type de profils fût déjà présent dans le cas précédent, ce phénomène est plus important cette fois-ci. Cela peut s'expliquer par le cumul de la correction et des profils déjà légèrement décroissants. Il faut tout de même noter que, quel que soit le proxy utilisé, l'ensemble des profils finalement obtenus semble bien se répartir dans l'espace comme cela doit être le cas (Figure 58a). Nous pouvons donc calculer les quantiles de production dynamique à partir des ensembles de modèles à notre disposition (vrais et reconstruits) pour chaque proxy. Comme précédemment, nous répétons deux cents fois l'opération afin de pouvoir construire des intervalles de confiance et déterminer un comportement général.

La Figure 67 présente les résultats ainsi obtenus. Ces derniers illustrent l'apport de la reconstruction des profils manquants pour réduire considérablement la largeur des intervalles de confiance autour des quantiles et donc en même temps l'incertitude autour de leur calcul. En revanche, il existe cette fois-ci une réelle différence entre les proxys. La largeur des intervalles est beaucoup plus étroite pour l'upscaling que pour tous les autres proxys. De plus, il existe un biais au niveau du Q10 que seuls l'upscaling et la connectivité permettent de réduire en donnant un intervalle qui encadre le vrai quantile calculé sur l'ensemble des profils dynamiques. Cette erreur souligne la difficulté à obtenir des profils analytiques justes pour les petites réserves, comme constaté sur la Figure 65 et la Figure 66.

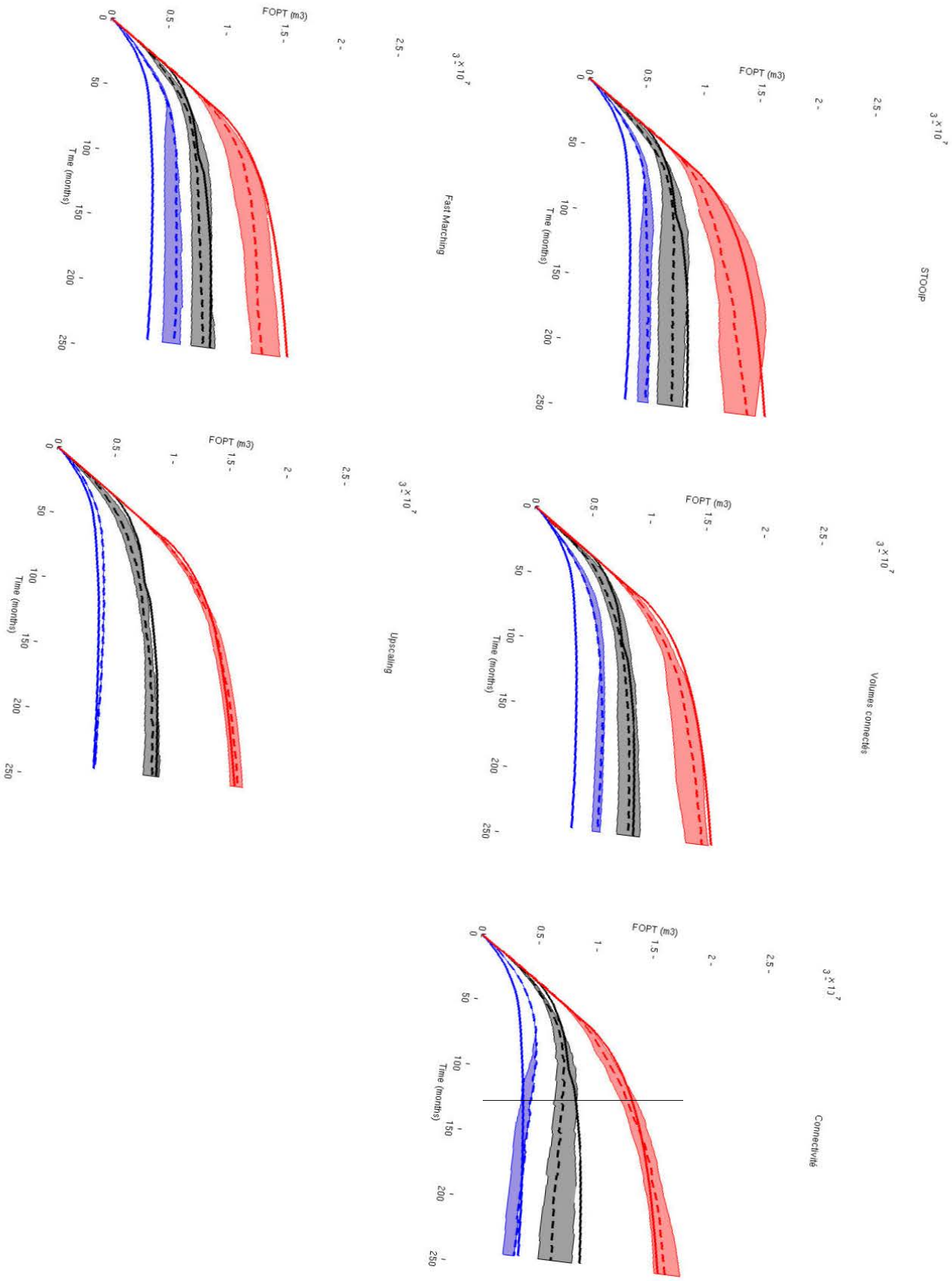

Figure 67 : Quantiles de production et intervalles de confiance obtenus en utilisant notre méthodologie de reconstruction de profils manquants pour chaque proxy. Les profils pleins correspondent aux vrais quantiles calculés sur l'ensemble des simulations dynamiques.

3.3.b Synthèse de la procédure d'optimisation sur le champ Y

L'utilisation de ce nouveau cas d'étude nous conforte à propos de l'efficacité de notre méthodologie de réduction des incertitudes sur le calcul des quantiles de production. L'emploi de profils analytiques couplés à une sélection de modèles permet d'améliorer le calcul des quantiles utilisés pour déterminer l'impact des paramètres statiques sur le comportement dynamique des modèles. Plus le proxy est bon pour respecter les distances entre réponses dynamiques et pour sélectionner des modèles représentatifs, plus la reconstruction des profils manquants apportera d'informations. Par contre, si le proxy n'est vraiment pas informatif, la procédure d'optimisation atteint sa limite et n'améliore pas le calcul des quantiles.

Sur les deux cas d'étude, l'homogénéisation a toujours été un bon proxy pour étudier l'impact des paramètres statiques sur la production d'hydrocarbure. En revanche, sa complexité et son faible gain de temps peuvent être rédhibitoires. Les champs X et Y illustrent que des proxys plus simples peuvent alors être utilisés et fournir de très bonnes informations. L'utilisation de la méthodologie de comparaison des proxys peut alors nous guider dans le choix de ce proxy.

Partie IV : Conclusions

L'étude des incertitudes constitue une étape indispensable avant toute décision concernant un réservoir pétrolier (achat, développement, forage de nouveau puits, etc.), car elle permet d'évaluer les risques associés à la décision. En fonction de l'étape de la vie du champ et du domaine d'application, les techniques utilisées sont différentes. Pour le cas particulier d'un champ encore non exploité (*green field*) et de l'étude de l'impact des paramètres statiques sur le comportement dynamique et les réserves disponibles, deux techniques sont habituellement employées. D'une part, la sélection de modèles représentatifs qui fournit une réponse physique fiable, mais ne donne pas de garantie théorique sur l'évaluation des incertitudes. D'autre part, l'approximation de la simulation d'écoulements par un proxy qui permet d'estimer le comportement d'un plus grand nombre de modèles de réservoirs, mais peut mener à des biais dus aux approximations de la physique des

écoulements. Les travaux réalisés dans le cadre de cette thèse ont permis de passer outre ces limitations en tirant parti des deux méthodes.

Principaux résultats

Le travail s'est axé à la fois sur l'approximation de l'écoulement dynamique d'une part et sur le processus d'étude des incertitudes d'autre part. Plus précisément, le premier point a consisté à développer des proxys de simulateur dynamique d'écoulements basés sur la propagation de front de saturation dans les réservoirs d'hydrocarbure. Nous sommes partis des travaux d'Alabert et Modot (1992) sur la connectivité, auxquels nous avons ajouté une meilleure prise en compte de la perméabilité ainsi que d'autres propriétés pétrophysiques grâce à l'emploi d'un algorithme de *Fast Marching*. En effet cet algorithme, dont différentes versions ont été développées dans les années 90 (Tsitsiklis, 1995, Sethian, 1999), a été utilisé à plusieurs reprises en ingénierie de réservoir ces dernières années (Datta-Gupta et al., 2011, Hovadik and Larue, 2011). Nous avons donc modifié un proxy de connectivité pour que la propagation du front de saturation emploie l'algorithme de *Fast Marching* et représente mieux le comportement réel d'un réservoir. Cela nous a permis de garder les avantages du proxy par connectivité, comme la prise en compte des contacts fluides en tant que limite de propagation, ou la possibilité d'utiliser plusieurs phases de démarrage des puits. Pour finir, nous avons utilisé une version du *Fast Marching* permettant de prendre en compte un milieu anisotrope. Les résultats de ce premier axe de recherche sont en demi-teinte. En effet, le champ X nous a montré que dans certains cas ce proxy, bien que plus sophistiqué qu'une simple accumulation, ne donne pas de meilleurs résultats pour l'étude de l'impact des paramètres statiques sur le comportement dynamique des réservoirs. Par contre, grâce au champ Y nous avons vu que cela n'était pas généralisable et que ce nouveau proxy pouvait être utile pour notre méthodologie d'optimisation. C'est pourquoi il a été décidé de l'industrialiser. De plus, nous avons mis en évidence qu'un proxy beaucoup plus réaliste, comme l'homogénéisation, pouvait améliorer les résultats dans les deux cas d'étude que nous avons présentés. Cela constitue une motivation suffisante pour continuer d'améliorer notre proxy basé sur le *Fast Marching*, en intégrant un solveur de pression qui pourrait calculer la pression (en régime permanent ou pour différents pas de temps) et ainsi perfectionner la fonction de vitesse utilisée.

En ce qui concerne l'amélioration de la prise en compte des proxys dans l'étude des incertitudes dynamiques, nous avons mis au point une méthodologie originale qui permet de reconstruire, à l'aide d'un modèle analytique et d'un algorithme de minimisation, les profils dynamiques de modèles non simulés. Cette partie de la recherche peut elle-même être divisée en plusieurs sous-parties :

- la mise au point d'un modèle analytique capable de représenter un profil de production d'huile cumulée classique. Pour cela, il a fallu décomposer les différents processus qui contrôlent l'écoulement dans un réservoir ;
- la mise au point d'une procédure de minimisation mathématique ;
- l'étude de la relation qui associe les distances entre modèles calculées grâce aux réponses proxys à celles calculées entre les modèles en termes de réponses dynamiques.

L'application de cette méthodologie sur les deux cas d'étude nous a permis de montrer son efficacité pour réduire les incertitudes sur l'estimation des quantiles de production, afin de représenter l'impact des paramètres statiques sur le comportement dynamique. L'étude du champ Y a cependant mis en évidence la nécessité d'utiliser un proxy suffisamment informatif. Il serait donc intéressant d'essayer de l'améliorer. Pour cela, nous pourrions revoir notre choix de l'algorithme de minimisation qui n'a pas été optimal. Il serait ainsi envisageable d'essayer un autre algorithme basé sur une stratégie évolutionnaire, qui est normalement plus adapté aux problèmes non linéaires comme le nôtre. La difficulté d'utilisation d'un tel algorithme est qu'il est souvent compliqué de lui

ajouter des contraintes sans reformuler la fonction objectif, or ces contraintes sont nécessaires pour forcer notre fonction paramétrique à rester croissante.

D'un autre côté, il faut être capable de déterminer si un proxy peut être utilisé pour notre procédure de reconstruction de profils. C'est pour cette raison que nous avons développé plusieurs critères de comparaison des différents proxys à notre disposition (ceux que nous avons développés et ceux déjà existants). Nous avons ainsi mis en place une méthodologie systématique permettant d'évaluer chaque proxy et de choisir le plus adéquat pour l'étude des incertitudes. Son utilisation sur les champs X et Y nous a effectivement montré son intérêt, en mettant notamment en avant que l'amélioration du modèle physique présenté par certains proxys peut être minime vis-à-vis de leur complexité et de leur temps de mise en place. Certains points de comparaison de cette méthode peuvent cependant être améliorés, telle que la régression par parties qui pourrait être couplée à une transformation afin d'essayer d'améliorer la corrélation. Il serait également intéressant de l'essayer sur d'autres proxys régulièrement utilisés, comme les réseaux de neurones ou le *Data mining*.

Perspectives

La méthodologie proposée de reconstruction de profils de production grâce à une fonction paramétrique pourrait être utilisée autrement que pour étudier directement l'impact des paramètres statiques sur la production des réservoirs. Par exemple, elle pourrait servir pour l'amélioration des distributions bi-variées accumulation/réserves. En effet, dans l'industrie pétrolière, ces dernières sont très utilisées pour décrire le comportement d'un réservoir (nous l'avons fait pour différencier les champs X et Y). Comme nous l'avons vu avec nos deux cas d'étude, ce comportement a son importance. Or, comme la création d'un tel graphique nécessite des simulations dynamiques très coûteuses en temps, seuls peu de points sont disponibles, ce qui rend l'interprétation incertaine. En utilisant la reconstruction de profils, il est possible d'obtenir des points de données supplémentaires, qui, même s'ils sont approximatifs, aident à approcher le comportement réel du réservoir lors de son exploitation. En effet, la corrélation mesurée avec un nuage de points intégrant les fonctions paramétriques aux simulations de référence est plus proche de la réalité que celle mesurée sur un nuage ne contenant que les quelques modèles sélectionnés (Figure 68).

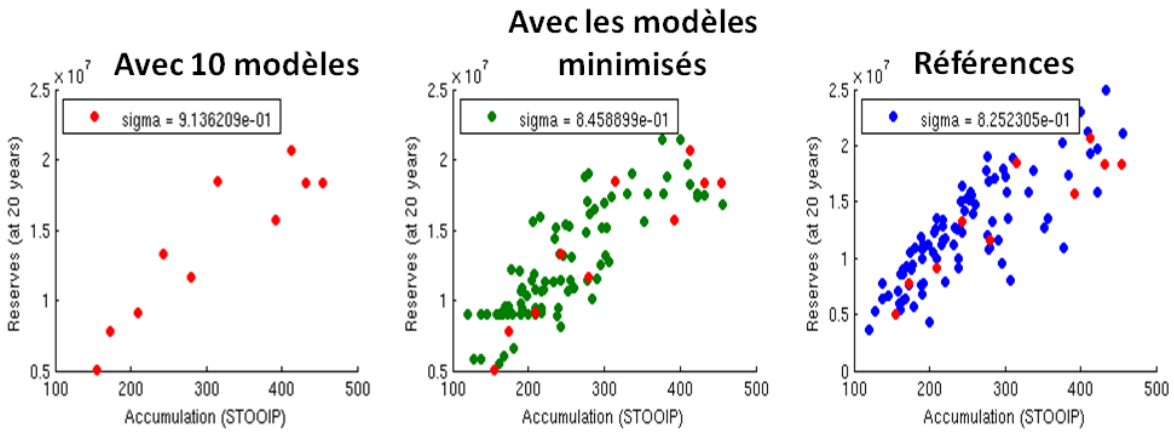

Figure 68 : Amélioration des nuages de points accumulation/réserves grâce à la reconstruction des profils manquants. Les points rouges correspondent aux modèles sélectionnés (communs aux trois graphiques).

Nous pourrions aussi employer la procédure de reconstruction pour optimiser le placement des puits dans un réservoir. Ce problème est un vaste sujet et de nombreuses méthodes ont été proposées, toutes ayant pour point commun d'utiliser de très nombreux critères afin de caractériser une trajectoire de puits, comme par exemple l'aire de drainage ou la porosité de la roche perforée (Bouzarkouna et al., 2013, Chang et al., 2014) ; ces méthodologies nécessitent souvent de simuler

dynamiquement un grand nombre de configurations. Nous pourrions imaginer remplacer notre proxy par une fonction qui combine le résultat de plusieurs critères donnant une note à chaque configuration. À partir de quelques simulations de référence avec des configurations de puits choisies, nous pourrions ainsi déterminer le profil analytique de chaque configuration non simulée et donc évaluer le meilleur schéma de développement de puits à moindre coût.

Par ailleurs, il est important de noter que l'ensemble de la méthode proposée dans cette thèse est générale. Elle pourrait être appliquée à d'autres courbes de réponse du simulateur d'écoulements haute fidélité, comme le pourcentage d'eau du fluide extrait, la production de gaz ou encore l'évolution de la pression. La méthodologie pourrait aussi être employée dans d'autres domaines pour lesquels l'étude des incertitudes est importante. Par exemple, dans le domaine de l'hydrogéologie, de la climatologie ou des sciences du sol. Il faudrait alors simplement envisager que les courbes approchées ne présenteront pas forcément les mêmes caractéristiques de forme et que les fonctions paramétriques correspondantes devront être adaptées. Il faudra également mettre en place de nouvelles contraintes et de nouvelles méthodes de correction. Il serait alors intéressant de comparer nos résultats à ceux obtenus, par exemple, par la méthode proposé par Josset et al. (2015) en hydrogéologie. En effet, ces derniers proposent de créer un modèle d'erreur pour pouvoir prédire directement la réponse dynamique à partir de la réponse proxy des modèles non sélectionnés, alors que nous proposons de prédire la distance entre les réponses dynamiques à partir des distances entre réponses proxys afin d'en déduire les paramètres d'une fonction paramétrique qui reproduira la réponse dynamique.

Partie V : REFERENCES

ABRAHAMSEN, P. 1993. Bayesian Kriging for Seismic Depth Conversion of a Multi-Layer Reservoir. *In:* SOARES, A. (ed.) *Geostatistics Tróia '92.* Springer Netherlands.

ALABERT, F. G. & MODOT, V. 1992. Stochastic Models of Reservoir Heterogeneity: Impact on Connectivity and Average Permeabilities. *SPE 67th Annual Technical Conference and Exhibition.* Washington D.C.: Society of Petroleum Engineers.

ALTON, K. & MITCHELL, I. M. 2007. Fast Marching Methods for a Class of Anisotropic Stationary Hamilton-Jacobi Equations. *SIAM Journal on Numerical Analysis*, 38.

BALLIN, P. R., JOURNEL, A. G. & AZIZ, K. 1992. Prediction of uncertainty in reservoir performance forecast. *Journal of Canadian Petroleum Technology,* 31, 11.

BARDY, G., BIVER, P., CAUMON, G., RENARD, P., CORPEL, V. & KING, P. 2014. Proxy Comparison For Sorting Models and Assessing Uncertainty on Oil Recovery Profiles. *ECMOR XIV 14th European Conference on the Mathematics of Oil Recovery.* Canania, Sicily: EAGE.

BATYCKY, R. P., BLUNT, M. J. & THIELE, M. R. 1997. A 3D Field-Scale Stremline-Based Reservoir Simulator. *SPE Resrvoir Engineering***,** 246-254.

BELOGAY, E., CABRELLI, C., MOLTER, U. & SHONKWILER, R. 1997. Calculating the Hausdorff distance between curves. *Information Processing Letters,* 64.

BLUNT, M. J., LIU, K. & THIELE, M. R. 1996. A Generalized Streamline Method to Predict Reservoir Flow. *Petroleum Geoscience,* 2**,** 256-269.

BORG, I. & GROENEN, P. J. F. 2005. *Modern Multidimensional Scaling: Theory and Applications*, Springer.

BOURGEOIS, M. & COUILLENS, P. 1994. Use of Well Test Analytical Solution for Production Prediction. *European Petroleum Conference.* London, U.K.: SPE.

BOUZARKOUNA, Z., DING, D. Y. & AUGER, A. 2013. Partially Separated Metamodels With Evolution Strategies for Well-Placement Optimization. 18.

BOX, G. E. P. & BEHNKEN, D. W. 1960. Some New Three Level Designs for the Study of Quantitative Variables. *Technometrics,* 2**,** 21.

BOX, G. E. P. & DRAPER, N. 1987. *Empirical Model-Building and Response Surfaces,* New York, New York, Wiley & Sons.

BOYD, S. & LIEVEN, V. 2004. *Convex Optimization*, Cambridge University Press.

BUCKLEY, S. E. & LEVERETT, M. C. 1942. Mechanism of fuid displacement in sands. *Transaction of the American Institute of Mining Metallurgical and Petroleum Engineers,* 146**,** 107-116.

CAERS, J. 2011. *Modeling Uncertainty in the Earth Sciences*, Wiley-Blackwell.

CHANG, Y., BOUZARKOUNA, Z. & DEVEGOWDA, D. Handling Geological Uncertainty in Oil Field Development with Multi-objective Optimization Technique. ECMOR XIV-14th European conference on the mathematics of oil recovery, 2014.

CHARLES, T., GUEMENE, J. M., VINCENT, G. & DUBRULE, O. Experience with the Quantification of subsurface uncertainties. *In:* SPE, ed. SPE Annual Technical Conference and Exhibition, September 26-29 2001 Houston, Texas.

CHILÈS, J. P. & DELFINER, P. 1999. *Geostatistics, Modeling Spatial Uncertainty*, John Wiley & Sons.

CHRISTIE, M. A. 1996. Upscalling for Reservoir Simulation. *JPT***,** 1004-1010.

CORRE, B., THORE, P., DE FERAUDY, V. & VINCENT, G. Integrated Uncertainty Assessment For Project Evaluation and Risk Analysis. *In:* SPE, ed. SPE European Petroleum Conference, October 24-25 2000 Paris, France.

COX, T. F. & COX, M. A. A. 2001. *Multidimensional Scaling, 2nd Ed.*, Chapman & Hall/CRC.

DAMSLETH, E., A., H. & VOLDEN, R. 1992. Maximum Information at Minimum Cost : A North Sea Field Development Study with an Experimental Design. *Journal of Petroleum Technology***,** 7.

DATTA-GUPTA, A., XIE, J., GUPTA, N., KING, M. J. & LEE, W. J. 2011. Radius of Investigation and its Generalization to Unconventional Reservoir. *Journal of Petroleum Technology.*

DEJEAN, J. P. & BLANC, G. 1999. Managing Uncertainties on Production Predictions Using Integrated Statistical Methods. *SPE Annual Technical Conference and Exhibition.* Houston, Texas: Society of Petroleum Engineers.

DO, T. T. 2006. *Optimisation de Formes en Forgeage 3D.* Doctor, Ecole des Mines de Paris.

DURLOFSKY, L. J. Upscaling and gridding of fine scale geological models for flow simulation. 8th International Forum on Reservoir Simulation, 2005 2005 Stresa, Italy.

EGELAND, T., HATLEBAKK, E., HOLDEN, L. & LARSEN, E. A. 1992. Designing Better Decisions. *In:* SPE (ed.) *SPE European Petroleum Conference.* Stavanger, Norway.

FARMER, C. L. 2001. Upscaling: a review. *International Journal for Numerical Methods in Fluids***,** 63-78.

FRANCO, J. 2008. *Planification d'expériences numériques en phase exploratoire pour la simulation des phénomènes complexes.* Ecole Nationale Supérieure des Mines de Saint-Etienne.

HOLDEN, L., MOSTAD, P., NIELSEN, B., GJERDE, J., TOWNSEND, C. & OTTESEN, S. 2003. Stochastic Structural Modeling. *Mathematical Geology,* 35**,** 899-914.

HOVADIK, J. & LARUE, D. 2011. Predicting Waterflood Behavior by Simulating Earth Models with No or Limited Dynamic Data: From Model Ranking to Simulating a Billion-Cell Model. *Uncertainty analysis and reservoir modeling: AAPG Memoir 96.*

JAIN, A. K. 2012. Data Clustering: 50 years beyond K-means. *Pattern Recognition Letters,* 31**,** 651-666.

JBABDI, S., BELLEC, P., TORO, R., DAUNIZEAU, J., PÉLÉGRINI-ISSAC, M. & BENALI, H. 2007. Accurate Anisotropic Fast Marching for Diffusion-Based Geodesic Tractography. *International Journal of Biomedical Imaging,* 2008**,** 12.

JOSSET, L., GINSBOURGER, D. & LUNATI, I. 2015. Functional error modeling for uncertainty quantification in hydrogeology. *Water Resources Research,* 51**,** 19.

JOURDAN, A. & ZABALZA-MEZGHANI, I. 2004. Response surface designs for scenario management and uncertainty in reservoir production. *Mathematical Geology,* 36**,** 965-985.

KRIGE, D. G. 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa***,** 21.

LECHNER, J., P. & ZANGL, G. 2005. Treating Uncertainties in Reservoir Performance Prediction with Neural Networks. *SPE Europec/EAGE Annual Conference.* Madrid, Spain: SPE.

MALLET, J.-L. 2014. *Elements of Mathematical Sedimentary Geology: the GeoChron Model*, EAGE Publication.

MATHERON, G. 1955. Applications des méthodes statistiques a l'évaluation des gisements. *Annales des Mines***,** 25.

MC KAY, M.D., BECKMAN R.J., CONOVER W.J. 1979. A comparison of three metods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21, 239-245.

NOCEDAL, J. & WRIGHT, S. J. 2006. *Numerical Optimization*.

SACKS, J., SCHILLER, S. & WELCH, W. 1989. Designs for Computer Experiments. *Technometrics,* 31**,** 8.

SANTIAGO, J. 2013. *DEveloppement de nouveaux plan d'expériences uniformes adaptés à la simulation numérique en grande dimension.* Université d'Aix-Marseille.

SATIJA, A. & CAERS, J. Preconditioning discrete geostatistical models in flow inverse problems. Stanford Center for Reservoir Forecasting Annual Meeting, 2012 Stanford University.

SCHEIDT, C. 2006. *Analyse statistique d'expériences simulées : Modélisation adaptative de réponses non-régulières par krigeage et plans d'expériences.* Docteur, Université Louis Paster, Strasbourg I.

SCHEIDT, C. & CAERS, J. A Workflow for Spacial Uncertainty Quantification using Distances and Kernels. SCRF procedings, 2007. 34.

SCHEIDT, C. & CAERS, J. 2009. Representing Spatial Uncertainty Using Distances and Kernels. *Mathematical Geosciences,* 41**,** 23.

SETHIAN, J. A. 1999. *Level Set Methods and Fast Marching Methods; Evolving interfaces in computational geometry, fluid, mechanics, computer vision, and materials science*, Cambridge University Press.

SETHIAN, J. A. 2001. Evolution, Implementation, and Application of Level Set and Fast Marching Methods for Advancing Front. *Journal of Computationnal Physics***,** 503-555.

SETHIAN, J. A. & VLADIMIRSKY, A. 2001. Ordered upwind methods for static Hamilton-Jacobi equations. *Proceedings of the National Academy of Sciences of the United States of America,* 98.

SHAWE-TAYLOR, J. & CHRISTIANINI, N. 2004. *Kernel Methods for Pattern Analysis*, Cambridge University Press.

SUZUKI, S. & CAERS, J. 2008. A Distance-based Prior Model Parameterization for Constraining Solutions of Spatial Inverse Problems. *Mathematical Geosciences,* 40**,** 445-469.

THIELE, M. R., BATYCKY, R. P., BLUNT, M. J. & ORR, F. M. J. 1996. Simulating Flow in Heterogeneous Systems Using Streamtubes and Streamlines. *SPE Resrvoir Engineering***,** 5-12.

THIELE, M. R., BATYCKY, R. P. & FENWICK, D. H. 2010. Streamline Simulation for Modern Reservoir-Engineering Workflows. *JPT,* 62**,** 64-70.

TSITSIKLIS, J. N. 1995. Efficient algorithms for globally optimal trajectories. *IEEE Transactions on Automatic Control,* 40**,** 1528-1538.

VENKATARAMAN, R. 2000. Application of the Method of Experimental Design to Quantify Uncertainty in Production Profiles. *In:* SPE (ed.) *SPE Asia-Pacific Conference.* Yokohama, Japan.

ZABALZA-MEZGHANI, I., MANCEAU, E., FERAILLE, M. & JOURDAN, A. 2004. Uncertainty Management: From Geological Scenarios to Production Sceme Optimization. *Journal of Petroleum Science and Engineering special issue: Risk Analysis Applied to Petroleum Exploration and Production*.

Intégration de modèles approchés pour mieux transmettre l'impact des incertitudes statiques sur les courbes de réponse des simulateurs d'écoulements

Résumé : Alors que l'on utilise couramment de nombreux modèles numériques différents pour la description statique des réservoirs souterrains et des incertitudes associées, les incertitudes sur les écoulements des fluides à l'intérieur de ces réservoirs ne peuvent, pour des raisons de performance, que s'appuyer que sur quelques simulations d'écoulements. Les travaux de cette thèse ont donc pour objectif d'améliorer la transmission de l'impact des incertitudes statiques sur les réponses du simulateur dynamique d'écoulements sans augmenter le temps de calcul, grâce à des modèles approchés (proxy). Pour cela deux axes de recherche ont été menés :

- L'implémentation de nouveaux proxys basés sur le *Fast Marching*, afin de modéliser la propagation d'un fluide dans un réservoir avec seulement quelques paramètres. Cela permet d'obtenir des courbes de réponse similaires à celles fournit par le simulateur d'écoulement pour un temps de calcul très court ;
- La mise en place d'une procédure de minimisation mathématique afin de prédire les courbes de réponses du simulateur d'écoulement à partir d'un modèle analytique et des distances entre les modèles calculées avec les réponses des proxys.

Les méthodes développées ont été appliquées sur deux cas d'études réels afin de les valider face aux données disponibles dans l'industrie. Les résultats ont montrés que les proxys que nous avons implémentés apportent de meilleures informations que les proxys disponibles bien que les nôtres soient toujours perfectibles. Nous avons aussi mis en évidence l'intérêt de notre procédure de minimisation pour mieux évaluer les incertitudes dynamiques à partir du moment où le proxy utilisé est suffisamment fiable.

Mots-clés : incertitudes, réservoir, proxy, minimisation

Integration of approximated models in order to better assess impact of static uncertainties on flow simulator's response curves

Abstract: Although it is common to use many different numerical models for the static description of underground reservoirs and their associated uncertainties, for fluid flow uncertainties through these reservoirs only few dynamic simulations can be used due to performance reasons. The objective of this thesis' work is to better transmit the impact of static uncertainties on flow simulator's responses without increasing computation time, using approximated models (proxies). Research has been undertaken in 2 directions:

- Implementation of new proxies based on Fast Marching in order to better approach fluid propagation behavior in a reservoir using only a few parameters. This allows to obtain response curves close to those provided by the flow simulator in a very short period of time
- Set up a mathematical minimization's procedure in order to predict flow simulator's response curves using an analytical model and distances between proxy responses computed on every model.

The methods developed during this PhD have been applied on two different real cases in order to validate them with industry data. Results have shown that our new proxy improve the quality of the information about fluid behavior compared to the available proxy even though ours can still be improved. We also highlight that our minimization procedure better assesses dynamic uncertainties if the proxy used is reliable enough.

Keywords: uncertainties, reservoir, proxy, minimization