



HAL
open science

Analyse statistique d'expériences simulées : Modélisation adaptative de réponses non régulières par krigeage et plans d'expériences, Application à la quantification des incertitudes en ingénierie des réservoirs pétroliers

Céline Scheidt

► **To cite this version:**

Céline Scheidt. Analyse statistique d'expériences simulées : Modélisation adaptative de réponses non régulières par krigeage et plans d'expériences, Application à la quantification des incertitudes en ingénierie des réservoirs pétroliers. Mathématiques [math]. Université Louis Pasteur - Strasbourg I, 2006. Français. NNT : 2006STR13102 . tel-00127750

HAL Id: tel-00127750

<https://theses.hal.science/tel-00127750>

Submitted on 29 Jan 2007

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

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

Thèse présentée pour obtenir le grade de
Docteur de l'Université Louis Pasteur
Strasbourg I

Discipline : Mathématiques Appliquées
par Céline Scheidt

Analyse statistique d'expériences simulées :
Modélisation adaptative de réponses non-régulières
par krigeage et plans d'expériences
Application à la quantification des incertitudes en
ingénierie des réservoirs pétroliers

Soutenue publiquement le 25 septembre 2006

Membres du jury

Directeur de thèse : COLLOMBIER Dominique, Professeur, IRMA, Strasbourg
Rapporteur Interne : GALTCHOUK Leonid, Professeur, IRMA, Strasbourg
Rapporteur Externe : JOURNEL André, Professeur, Université de Stanford
Rapporteur Externe : PRONZATO Luc, Directeur de recherche CNRS, Université de Nice
Examinatrice : ZABALZA-MEZGHANI Isabelle, Docteur, IFP, Rueil-Malmaison
Membre invité : CURTIS Alan, Manager, Bhpbilliton, Houston

Remerciements

Cette thèse a été réalisée au sein de la direction Ingénierie de Réservoir de l'Institut Français du Pétrole. Au terme de ces trois années, je souhaite adresser mes sincères remerciements :

A Monsieur Dominique Collombier,
qui m'a fait l'honneur d'être mon directeur de thèse. Son expertise dans le domaine des plans d'expériences a été déterminante dans les orientations prises dans mon travail. Je tiens à le remercier pour sa grande disponibilité, ses qualités humaines et pour tout le savoir qu'il a pu me transmettre. J'ai pu tirer parti de son expérience et des nombreuses idées qu'il m'a suggérées, et qui figurent dans ce mémoire.

A Madame Isabelle Zabalza-Mezghani,
qui est à l'origine de cette thèse et a encadré mon travail à l'Institut Français du Pétrole. Je tiens à la remercier pour la confiance qu'elle m'a accordé jour après jour, pour m'avoir donné goût à la recherche scientifique. Par son expérience, elle a su guider mon travail vers des problèmes suscitant un vif intérêt industriel. Qu'elle soit remerciée ici pour ses qualités humaines et scientifiques, pour son aide précieuse, pour la gentillesse et le soutien dont elle a fait preuve à mon égard pendant ses trois années.

A Monsieur André Journal,
qui m'a fait l'honneur d'être rapporteur de cette thèse, pour ses remarques pertinentes et le temps qu'il a consacré à examiner mon travail. Je suis également très reconnaissante à la remarquable hospitalité dont il a fait preuve en m'accueillant pendant 2 mois à l'Université de Stanford. Je tiens à lui exprimer ici toute ma gratitude, pour sa disponibilité et sa gentillesse à mon égard. L'ensemble des discussions que nous avons eu ont fortement contribué à enrichir ma connaissance sur les méthodes géostatistiques.

A Monsieur Luc Pronzato,
qui m'a fait l'honneur de rapporter cette thèse et de présider le jury. Je tiens à le remercier pour le temps passé à la lecture de mes travaux, pour l'intérêt porté à mon travail et pour ses précieuses remarques.

A Monsieur Leonid Galtchouk,
qui m'a fait l'honneur de bien vouloir juger mes travaux de thèse. Pour le temps qu'il a aimablement consacré à une lecture avisée de mon travail. Ses enseignements à l'Université Louis Pasteur m'ont permis d'élargir le spectre de mes connaissances mathématiques.

A Monsieur Alan Curtis,
qui a accepté de faire partie du jury de cette thèse. Je le remercie sincèrement pour sa collaboration et pour sa disponibilité. Son soutien et sa vision en ingénierie pétrolière ont été d'une grande aide tout au long de la thèse.

Mes très sincères remerciements vont à Mathieu Feraille pour ses idées et remarques fort instructives, pour les discussions et réflexions menées sur mes travaux. Également un grand merci à David Chazalviel pour sa précieuse aide pour élucider bien des problèmes logistiques lors de la phase de validation de la méthodologie.

C'est aussi pour moi l'occasion de remercier toutes les personnes avec qui j'ai eu l'opportunité de travailler, pour leur bonne humeur et leur sympathie. Les membres de la division « Ingénierie Réservoir », particulièrement les hôtes du bâtiment Laurier : qu'ils trouvent ici l'expression de ma gratitude pour la bonne humeur et la qualité des rapports que j'ai pu avoir avec chacun d'eux.

A tous mes collègues thésards, tout particulièrement les deux Alexandre, Carine, Charles, Thomas et Ludovic. Leur amitié et leur bonne humeur a été d'une aide précieuse tout au long de la thèse.

Je remercie chaleureusement mes parents, Delphine et Thierry pour leur soutien, leurs encouragements pendant ces trois années. Ce travail de thèse leur est dédié.

Enfin, un remerciement particulier à Darryl, pour tout ce qu'il a pu apporter à cette thèse. Son soutien sans faille, ses nombreux conseils et encouragements, et son engagement dans mon travail avec ses relectures attentives de mes travaux ont été d'une aide très précieuse.

Table des matières

CHAPITRE 1 INTRODUCTION	1
1.1 L'INGÉNIERIE DE RÉSERVOIR	2
1.1.1 PHASES DE DÉVELOPPEMENT D'UN RÉSERVOIR – DESCRIPTION GÉNÉRALE	3
1.1.2 CARACTÉRISATION DU RÉSERVOIR – MODÉLISATION	5
1.1.3 PRÉVISION DE PRODUCTION ET SCHÉMA DE DÉVELOPPEMENT	11
1.2 GESTION DES INCERTITUDES EN INGÉNIERIE DE RÉSERVOIR	12
1.2.1 NOTION D'INCERTITUDES	12
1.2.2 GESTION DES INCERTITUDES : PROBLÉMATIQUE	14
1.2.3 GESTION DES INCERTITUDES : MÉTHODOLOGIE	15
1.3 OBJECTIF DE LA THÈSE	16
1.4 CONCLUSIONS	18
CHAPITRE 2 GESTION DES INCERTITUDES EN INGÉNIERIE DE RÉSERVOIR - ÉTAT DE LA QUESTION	19
2.1 MÉTHODOLOGIE POUR LA GESTION DES INCERTITUDES EN INGÉNIERIE DE RÉSERVOIR	19
2.1.1 POSITION DU PROBLÈME	20
2.1.2 LES MÉTHODES DE MONTE-CARLO EN INGÉNIERIE DE RÉSERVOIR : RECOURS À UN MODÈLE APPROCHÉ	20
2.2 PLANS D'EXPÉRIENCES POUR L'ÉCHANTILLONAGE DU DOMAINE INCERTAIN	24
2.2.1 CHOIX DU NOMBRE ET DE L'EMPLACEMENT DES SIMULATIONS À RÉALISER	24
2.2.2 PLANS D'EXPÉRIENCES EUCLIDIENS	25
2.3 MÉTHODES D'AJUSTEMENT DU MODÈLE APPROCHÉ	28
2.3.1 MODÈLES DE RÉGRESSION	29
2.3.2 RÉSEAUX DE NEURONES	32
2.3.3 LE KRIGEAGE	34
2.4 LIMITES DES APPROCHES EXISTANTES	35
2.5 PROPOSITION MÉTHODOLOGIQUE	37
CHAPITRE 3 MODÉLISATION D'EXPÉRIENCES DÉTERMINISTES	41
3.1 INTRODUCTION	41
3.2 PRÉDICTION DES EXPÉRIENCES DÉTERMINISTES : POSITION DU PROBLÈME	42
3.3 EXPRESSION DU MODÈLE	44
3.3.1 MODÈLE COMPOSITE	44
3.3.2 AJUSTEMENT DE LA TENDANCE MOYENNE : RÉGRESSION PAR MOINDRES CARRÉS	46
3.3.3 AJUSTEMENT DE LA PARTIE RÉSIDUELLE : ESTIMATEUR PAR KRIGEAGE	49
3.3.4 MODÈLE COMPOSITE ET APPROCHE BAYÉSIENNE	51
3.3.5 MODÈLE COMPOSITE : SYNTHÈSE	52
3.4 PROBLÈME LIÉ À LA MODÉLISATION : ESTIMATION DE LA FONCTION DE COVARIANCE ENTRE LES OBSERVATIONS	54
3.4.1 DÉFINITIONS ET PROPRIÉTÉS DU VARIOGRAMME	54
3.4.2 ESTIMATION DU VARIOGRAMME DANS LA MODÉLISATION ADAPTATIVE	57
3.5 REMARQUES COMPLÉMENTAIRES : APPROCHES ALTERNATIVES OU COMPLÉMENTAIRES	62
3.5.1 COKRIGEAGE	63

3.5.2 KRIGEAGE ORDINAIRE AVEC VOISINAGES GLISSANTS	64
3.5.3 BILAN	67
3.6 CONCLUSIONS.....	68
CHAPITRE 4 RECHERCHE DE L'ESTIMATEUR DE PLUS GRANDE PRÉDICATIVÉ – AJOUT DE	
POINTS PILOTES	69
4.1 INTRODUCTION	69
4.2 MÉTHODE DES POINTS PILOTES - PRINCIPE.....	70
4.3 EMLACEMENT DES POINTS PILOTES	74
4.4 VALEUR FICTIVE DE LA RÉPONSE AUX POINTS PILOTES	77
4.4.1 CALCUL DU COEFFICIENT PRESS À MOINDRE COÛT	78
4.4.2 OPTIMISATION DE LA VALEUR FICTIVE AUX POINTS PILOTES.....	80
4.5 CONCLUSIONS.....	89
CHAPITRE 5 STRATÉGIE D'EXPÉRIMENTATION POUR RÉPONSES SIMULÉES NON-RÉGULIÈRES	91
5.1 INTRODUCTION	91
5.2 ETUDE DE LA DÉPENDANCE DE LA RÉPONSE AUX PARAMÈTRES	94
5.2.1 DIMINUTION DE LA DIMENSION DU PROBLÈME : AGRÉGATION DE PARAMÈTRES	95
5.2.2 ANALYSE DE SENSIBILITÉ	96
5.3 AUGMENTATION DU DISPOSITIF EXPÉRIMENTAL	97
5.3.1 RECHERCHE D'EXTREMA	98
5.3.2 RECHERCHE DES POINTS STATIONNAIRES	104
5.3.3 SENSIBILITÉ DU MODÈLE À L'APPORT D'INFORMATION DANS LES ZONES NON-ÉCHANTILLONNÉES..	108
5.3.4 ÉVALUATION DE LA PRÉDICATIVITÉ DU MODÈLE : AJOUT DE SIMULATIONS DE CONFIRMATION.....	113
5.4 CONCLUSIONS.....	115
CHAPITRE 6 MODÉLISATION ADAPTATIVE APPLIQUÉE À DES FONCTIONS TYPES.....	117
6.1 DÉMARCHE ET NOTATIONS	117
6.2 AJUSTEMENT DE LA FONCTION NON-RÉGULIÈRE « CHAMEAU »	118
6.3 EXEMPLES DE MODÉLISATION DE DIFFÉRENTES FONCTIONS À DEUX VARIABLES	122
6.3.1 FONCTION CHAMEAU DÉCALÉ.....	122
6.3.2 FONCTION SINUS.....	123
6.3.3 FONCTION SINUS DÉCALÉE.....	125
6.3.4 FONCTION BRANIN	126
6.3.5 FONCTION ANALYTIQUE NON-RÉGULIÈRE 1	127
6.3.6 FONCTION ANALYTIQUE NON-RÉGULIÈRE 2	128
6.4 INDÉPENDANCE AU DISPOSITIF EXPÉRIMENTAL INITIAL	130
6.5 ÉCHANTILLONNAGE INFORMATIF : MÉTHODE ÉVOLUTIVE, PLAN COMPLET ET HYPERCUBE LATIN	134
6.6 INTÉRÊT DES DIFFÉRENTS CRITÈRES D'AJOUT DE POINTS	139
6.6.1 INTÉRÊT DE L'AJOUT DES EXTREMA AU DISPOSITIF EXPÉRIMENTAL COURANT.....	140
6.6.2 INTÉRÊT DES POINTS DE DÉRIVÉE PARTIELLE NULLE	142
6.6.3 INTÉRÊT DES POINTS PILOTES.....	143
6.7 RÉSULTATS DE MODÉLISATION POUR PLUS DE DEUX PARAMÈTRES	148
6.7.1 EXEMPLE DE MODÉLISATION POUR TROIS PARAMÈTRES	148
6.7.2 TABLEAU ET GRAPHES RÉCAPITULATIFS DES RÉSULTATS OBTENUS	149
6.8 CONCLUSIONS.....	152

CHAPITRE 7 APPLICATION À L'INGÉNIERIE DE RÉSERVOIR.....	155
7.1 CAS PUNQ - S3.....	155
7.1.1 PRÉSENTATION DU CAS PUNQ-S3.....	155
7.1.2 ANALYSE D'INCERTITUDE.....	157
7.1.3 APPROXIMATION PAR MODÉLISATION ADAPTATIVE.....	159
7.1.4 MODÉLISATION ADAPTATIVE ET PLANIFICATION CLASSIQUE.....	161
7.1.5 CONCLUSIONS.....	165
7.2 CAS PBR.....	165
7.2.1 PRÉSENTATION DU CAS PBR.....	165
7.2.2 ANALYSE D'INCERTITUDES.....	167
7.2.3 APPROXIMATION PAR MODÉLISATION ADAPTATIVE.....	170
7.2.4 MODÉLISATION ADAPTATIVE ET PLANIFICATION CLASSIQUE.....	175
7.2.5 CONCLUSION.....	183
7.3 CONCLUSIONS.....	184
CHAPITRE 8 BILAN ET PERSPECTIVES.....	185

Table des illustrations

Figure 1.1 : Modélisation statique : Modèle Géologique	6
Figure 1.2 : Modélisation dynamique du réservoir	9
Figure 1.3 : Caractérisation du modèle dynamique.....	10
Figure 1.4 : Variations du cumulé d'huile produit en 1825 jours en fonction de différents paramètres	13
Figure 2.1 : Méthodologie employée pour la gestion des incertitudes.....	21
Figure 2.2 : Echantillonnage du domaine incertain.....	25
Figure 2.3 : Plan composite centré pour trois paramètres : Plan normalisé et simulations associée.....	26
Figure 2.4 : Plan composite centré pour trois paramètres : Représentation spatiale dans l'espace des paramètres.....	27
Figure 2.5 : Plans à trois niveaux : plan complet, composite ou fractionnaire.....	28
Figure 2.6 : Représentation d'un neurone formel selon Mc Culloch et Pitts	32
Figure 2.7 : Représentation d'un réseau de neurones multicouche.....	33
Figure 2.8 : Ajustement d'une fonction analytique avec les plans classiques.....	36
Figure 2.9 : Principe de la méthodologie adaptative	38
Figure 3.1 : Principe de la méthodologie adaptative	41
Figure 3.2 : Ajustement de la réponse par superposition de l'ajustement moyen et du terme correctif	45
Figure 3.3 : Interprétation géométrique de MCO	47
Figure 3.4: Ajustement de la moyenne par moindres carrés ordinaires et pondérés	48
Figure 3.5: Ajustement par régression puis par prédicteur composite	53
Figure 3.6: Ajustement d'un variogramme expérimental.....	56
Figure 3.7 : Ajustement d'un variogramme : Méthode de moindres carrés	60
Figure 3.8 : Ajustement de variogrammes exponentiel, gaussien et sphérique.....	61
Figure 3.9 : Ajustement d'un variogramme : combinaison linéaire de variogrammes	62
Figure 3.10 : Approximation d'une fonction analytique par krigeage et cokrigeage.....	64
Figure 3.11 : Estimation d'un voisinage : deux points par quadrant.....	66
Figure 3.12 : Approximation d'une fonction analytique par régression et krigeage résiduel puis krigeage ordinaire avec voisinages glissants	67
Figure 4.1 : Principe de la méthodologie adaptative	69
Figure 4.2 : Influence d'un point pilote sur l'approximation (1).....	72
Figure 4.3 : Influence d'un point pilote sur l'approximation (2).....	73
Figure 4.4 : Position des points pilotes sur le domaine incertain en 2D.....	74
Figure 4.5 : Variance de krigeage dans le domaine incertain.....	75
Figure 4.6 : Approximation et variance de krigeage associée.....	82
Figure 4.7 : Position séquentielle des points pilotes : points de variance maximale.....	84
Figure 4.8 : Estimation de la statistique Q_2 pour les points pilotes	85
Figure 4.9 : Approximation de prédictivité maximale	86
Figure 4.10 : Fonction Sinus Décalée - Référence	86
Figure 4.11 : Approximation avec et sans Points Pilotes	87
Figure 4.12 : Évolution du Q_2 en fonction du nombre de points pilotes	88

Figure 5.1 : Principe de la méthodologie adaptative	92
Figure 5.2 : Echantillonnage spécifique à chaque réponse.....	94
Figure 5.3 : Influence d'un paramètre linéaire et un paramètre non-régulier.....	97
Figure 5.4 : Maximisation à partir de plusieurs points initiaux.....	98
Figure 5.5 : Recherche des extrema de l'approximation : Principe.....	100
Figure 5.6 : Définition du voisinage ϑ_j	101
Figure 5.7 : Méthode « Iso-simplexe » : Emplacement du nouveau point à simuler	102
Figure 5.8 : Recherche d'un maximum	103
Figure 5.9 : Recherche des extrema	104
Figure 5.10 : Optimisation de n valeurs de dérivées	106
Figure 5.11 : Recherche d'un point d'inflexion	107
Figure 5.12 : Variance de krigeage dans le domaine expérimental.....	109
Figure 5.13 : Traitement des différents points pilotes.....	111
Figure 5.14 : Valeur de la statistique Q_2 pour différentes valeurs fictives.....	112
Figure 5.15 : Principe d'enrichissement du dispositif expérimental	115
Figure 6.1 : Fonction Chameau : Référence et approximation polynomiale.....	119
Figure 6.2 : Différentes étapes de l'ajustement de la fonction Chameau	120
Figure 6.3 : Approximation de la fonction Chameau Décalé.....	123
Figure 6.4 : Approximation de la fonction Sinus	124
Figure 6.5 : Approximation de la fonction Sinus Décalée	125
Figure 6.6 : Ajustement de la fonction Branin	127
Figure 6.7 : Ajustement de la « fonction analytique 1 ».....	128
Figure 6.8 : Ajustement de la « fonction analytique 2 ».....	129
Figure 6.9 : Approximation de la fonction Sinus décalée pour différents plans d'expériences initiaux	132
Figure 6.10 : Fonction Chameau – hypercube latin 42 simulations	135
Figure 6.11 : Fonction Chameau Décalée – hypercube latin 42 simulations	136
Figure 6.12 : Fonction Sinus – hypercube latin 41 simulations	136
Figure 6.13 : Fonction Sinus Décalée – hypercube latin 36 simulations.....	137
Figure 6.14 : Fonction Branin – hypercube latin 27 simulations	137
Figure 6.15 : « fonction analytique 1 » – hypercube latin 62 simulations.....	138
Figure 6.16 : « fonction analytique 2 » – hypercube latin 58 simulations.....	138
Figure 6.17 : Approximation de la fonction Sinus Décalée pour 21 simulations.....	140
Figure 6.18 : Approximation de la « fonction analytique 2 » pour 40 simulations.....	140
Figure 6.19 : Approximation de la fonction Sinus Décalée sans extrema.....	141
Figure 6.20 : Approximation de la « fonction analytique 2 » sans extrema.....	141
Figure 6.21 : Approximation de la fonction Sinus Décalée sans les points de dérivée nulle	142
Figure 6.22 : Approximation de la « fonction analytique 2 » sans les points de dérivée nulle	143
Figure 6.23 : Approximation de la fonction Sinus Décalée – Méthode partielle 1	144
Figure 6.24 : Approximation de la fonction Sinus Décalée – Méthode partielle 2	145
Figure 6.25 : Approximation de la fonction Sinus Décalée – Méthode partielle 3	145
Figure 6.26 : Approximation de la « fonction analytique 2 » – Méthode partielle 1	146

Figure 6.27 : Approximation de la « fonction analytique 2 » – Méthode partielle 2	147
Figure 6.28 : Approximation de la « fonction analytique 2 » – Méthode partielle 3	147
Figure 6.29 : Ajustement d'une fonction non-régulière à 3 paramètres	149
Figure 6.30 : Graphes de prédiction d'une fonction non-régulière à trois paramètres	150
Figure 6.31 : Graphes de prédiction d'une fonction non-régulière à quatre paramètres	150
Figure 6.32 : Graphes de prédiction d'une fonction non-régulière à cinq paramètres	150
Figure 6.33 : Évolution du nombre de simulations en fonction du nombre de paramètres	152
Figure 7.1 : Modèle PUNQ	156
Figure 7.2 : Vue en perspective de différentes coupes du débit d'huile au puits PRO-1 après 2 992 jours	158
Figure 7.3 : Approximation de QOil par modélisation adaptative au cours des itérations.....	160
Figure 7.4 : Évolution des densités de probabilité du débit d'huile au cours des itérations.....	161
Figure 7.5 : Prédiction du débit d'huile par plans factoriel et composite.....	162
Figure 7.6 : Densité de probabilité du débit d'huile au puits PRO-1 pour différents dispositifs expérimentaux	163
Figure 7.7 : Prédiction du débit d'huile par hypercubes latins.....	163
Figure 7.8 : Densité de probabilité du débit d'huile pour différents dispositifs expérimentaux	164
Figure 7.9 : Vue en 3 dimensions du champs PBR	165
Figure 7.10 : Vue de dessus du champs PBR.....	166
Figure 7.11 : Vue en perspective de graphes partiels du Cumulé d'huile à 2 192 jours	169
Figure 7.12 : Vue en perspective de graphes partiels du Cumulé d'eau à 2 192 jours.....	170
Figure 7.13 : Approximation du cumulé d'huile à 2 192 jours par modélisation adaptative au cours de itérations	172
Figure 7.14 : Évolution des densités de probabilité de l'huile cumulée au cours des itérations	173
Figure 7.15 : Approximation du cumulé d'eau à 2 192 jours par modélisation adaptative au cours des itérations	174
Figure 7.16 : Évolution des densités de probabilité de l'eau cumulée au cours des itérations.....	175
Figure 7.17 : Prédiction du cumulé d'huile par plans factoriel et composite	176
Figure 7.18 : Densité de probabilité de Cumoil pour les différents dispositifs expérimentaux	177
Figure 7.19 : Résultats de modélisation pour des hypercubes latins	178
Figure 7.20 : Densité de probabilité de la réponse Cumoil pour les différents dispositifs expérimentaux	179
Figure 7.21 : Résultats de modélisation pour les plans factoriel et composites	180
Figure 7.22 : Densité de probabilité pour les plans factoriel et composites et modélisation adaptative	181
Figure 7.23 : Résultats de modélisation pour des hypercubes latins	182
Figure 7.24 : Densité de probabilité pour les hypercubes latins et modélisation adaptative.....	183

Liste des tableaux

Tableau 6-1 : Résultats de modélisation pour différents échantillonnages.....	134
Tableau 6-2 : Erreur relative et nombre de simulations nécessaires pour une modélisation à partir de plans évolutifs ou d'hypercubes latins.....	151
Tableau 7-1 : Propriétés pétrophysiques par couches.....	156
Tableau 7-2 : Intervalle de variation des paramètres incertains.....	157
Tableau 7-3 : Propriétés pétrophysiques par faciès.....	167
Tableau 7-4 : Intervalle de variation des paramètres incertains.....	168

Chapitre 1

Introduction

L'étude des réservoirs à hydrocarbures a pour objectif l'exploitation d'un réservoir de pétrole ou de gaz afin d'optimiser les rendements économiques d'une compagnie pétrolière. Les gisements les plus rentables économiquement sont les gisements les plus facilement accessibles et les plus simples à exploiter. Les technologies actuelles associées à une bonne compréhension du réservoir, suffisent généralement à extraire les hydrocarbures présents dans de tels gisements, leur exploitation est donc de coût optimal. Mais ces champs sont généralement en cours de production ou épuisés. Les gisements restant à exploiter sont découverts dans des terrains difficiles d'accès, comme par exemple, dans de grandes profondeurs sous la mer, et demandent des technologies d'exploitation très poussées. Des centaines de millions de dollars d'investissement peuvent ainsi être nécessaires pour évaluer le potentiel du réservoir avant même de commencer la production d'hydrocarbures. Ces gisements pétroliers sont également de plus en plus complexes et de géologie difficile à appréhender. Cette complexité supplémentaire entraîne une plus grande incertitude sur l'estimation des réserves en place et des coûts de production accrus, particulièrement les coûts d'accès au réservoir et de forage de puits sont exorbitants. De ce fait, la quantification d'incertitude et l'analyse de risque sont d'importance cruciale en ingénierie de réservoir car permettant l'intégration des risques dans les prises de décision.

Ce premier chapitre présente brièvement les différentes étapes de la vie d'un gisement pétrolier ainsi que les principaux enjeux de l'ingénierie de réservoir, à savoir : comprendre, caractériser et prévoir le comportement dynamique des réservoirs. Pour cela, il est indispensable d'acquérir des données, de les analyser et les interpréter, ainsi que de construire un modèle numérique d'écoulement. La description, dans ce chapitre, des techniques d'acquisition des données et de modélisation est indispensable à l'introduction de la notion d'incertitude, celle-ci étant très présente dans l'élaboration de ces modèles. Il est donc nécessaire de quantifier l'incertitude pour éviter des prises de décisions erronées, ou du moins d'évaluer la probabilité de telles décisions. Enfin, nous précisons l'objectif de la thèse et les différentes étapes, divisées en chapitres, qui nous ont permis de répondre à la problématique posée.

1.1 L'ingénierie de réservoir

Un gisement pétrolier est formé d'un ou plusieurs réservoirs poreux, souterrains, contenant des hydrocarbures d'origine sédimentaire sous forme liquide et/ou gazeuse. Un réservoir est en fait une roche comportant un grand nombre de minuscules cavités, appelées pores, dans laquelle la circulation des fluides est possible. On dit ainsi qu'une roche réservoir est poreuse et perméable. La structure du gisement ou du réservoir est limitée par des barrières imperméables qui piègent les hydrocarbures.

L'étude d'un gisement a pour but, à partir de la découverte d'un champ pétrolier, d'établir un schéma de développement et un scénario de production cherchant à optimiser la récupération des hydrocarbures dans le cadre d'une politique économique donnée. Un *schéma de développement* consiste principalement à définir l'implantation des puits : le nombre, l'emplacement et le type de puits (horizontal, vertical, longueur, ect.). Par *scénario de production*, on désigne les conditions de production possibles parmi les différentes options qui se présentent lorsque l'on veut produire un réservoir : type de récupération (injection d'eau ou de gaz, production primaire sous l'effet d'un aquifère actif), débits de production et d'injection, contraintes de fonctionnement de puits etc. Le scénario de production est régulièrement changé au cours du temps.

Dans un souci d'accroissement continu des réserves d'huile et de gaz, l'ingénieur doit améliorer constamment sa connaissance du gisement pétrolier et adapter son schéma de développement à cette nouvelle information. Les techniques de développement et de production d'un réservoir doivent donc évoluer afin d'autoriser la récupération des hydrocarbures non conventionnels pour des gisements difficiles à caractériser. Les spécialistes étudient, pendant toute la durée de vie du champ, les différentes caractéristiques de ce gisement pour élaborer un modèle de réservoir et un modèle d'écoulement. Ces modèles sont construits en utilisant toute la connaissance disponible, à savoir les données réservoir et les interprétations des experts sur ces données. Mais ces données sont la plupart du temps incertaines. La construction de ces deux modèles est nécessaire à une exploitation optimale du gisement sous contraintes économiques.

A ceci s'ajoute un grand nombre de possibilités de développement ou de production d'un réservoir. Par conséquent, pour étudier le comportement d'un gisement pétrolier de manière précise, une démarche scientifique adaptée et rigoureuse est requise. La mise en place d'une telle étude va permettre à l'ingénieur de prendre les meilleures décisions pour le développement d'un gisement que ce soit en terme de production ou de coût, afin d'exploiter le réservoir de manière optimale.

Toutes ces étapes constituent un défi, puisqu'elles doivent être réalisées à moindres coûts. De ce fait, plusieurs phases de développement sont indispensables, elles fournissent des

données diverses et variées sur le réservoir pour effectuer une étude approfondie des caractéristiques du gisement. Chaque phase a pour but de mieux comprendre et donc réduire les incertitudes propres à la phase, afin de mieux rentabiliser l'investissement et de réduire les risques associés à la mauvaise connaissance du gisement.

1.1.1 Phases de développement d'un réservoir – Description générale

- **Première phase de développement : Phase de délinéation**

La première phase de développement d'un gisement est appelée phase de délinéation, elle est réalisée dès sa découverte. Cette phase a pour objectif d'acquérir le plus de données possible sur le réservoir. Pour ce faire, des prélèvements sur carottes sont effectués en divers endroits du réservoir, des diagraphies et des études sismiques sont réalisées. Les mesures sur carottes vont permettre de déterminer les caractéristiques pétrophysiques (saturations initiales, saturations résiduelles, perméabilité, porosité, etc...) de ces échantillons de roches. Les diagraphies donnent des éléments tels que la lithologie ou la saturation en fonction de la profondeur. La sismique, dite de réflexion, consiste en une émission d'ondes vers le sous-sol qui vont, selon le principe physique de l'optique, se réfléchir sur les différentes interfaces géologiques du sous-sol pour ensuite être enregistrées par des capteurs situés en aval de la source d'émission. L'interprétation de ces données sismiques permet d'obtenir une image structurale des couches géologiques du réservoir. Lors de la phase de délinéation, avant la production d'huile ou de gaz, le gisement est « à l'équilibre », la pression ne dépendant que de la profondeur. L'ensemble des données obtenues, associé à l'expertise des géologues, va permettre :

- une estimation du volume global d'hydrocarbures dans le réservoir. Cette estimation est d'importance cruciale car très incertaine et jouant énormément sur les calculs de rentabilité économique du gisement.
- une estimation des réserves potentiellement récupérables à partir de plusieurs modes d'exploitation.
- la création d'un modèle statique de réservoir, qui détermine entre autres les propriétés physiques, les réseaux de fractures et l'emplacement du réservoir dans le gisement.

L'ensemble de ces caractéristiques va déterminer si ce gisement est rentable économiquement ou non. Dans le cas où l'exploitation du réservoir est décidée, le choix de l'emplacement des puits initiaux est effectué en tenant compte de critères économiques, tels que la production d'hydrocarbures et le nombre de puits que l'on souhaite forer.

- **Mise en production du réservoir**

Après ou pendant la phase de délinéation, une première mise en production du réservoir est effectuée. Généralement, celle-ci est réalisée en ouvrant simplement quelques puits. La pression interne du réservoir est la plupart du temps suffisante pour permettre une production de pétrole. Cette phase de récupération primaire, dite de déplétion, est de très courte durée, et ne permet de drainer qu'une partie de l'huile initialement en place (~10%). Pendant cette phase, le réservoir est soumis à une dépression et une expansion de l'huile due à une baisse de pression est observée. Un suivi de l'évolution de la pression du champ est effectué par essai de puits et, si le maintien de la pression n'est pas correct, on peut envisager d'injecter de l'eau ou du gaz dans le réservoir pour assurer ce maintien de pression. Le fluide est ainsi balayé vers le puits producteur. Pendant cette phase, l'ingénieur effectue des analyses de pression, volume et température des fluides. Il essaye, par ailleurs, de comprendre les différents mécanismes de déplacement des fluides dans le réservoir, afin de réduire l'incertitude dans l'élaboration d'un modèle de l'écoulement dynamique du fluide. Cette phase permet, entre autres, d'identifier la présence éventuelle d'un aquifère actif dans le système.

- **Deuxième phase de développement : Injection de fluide**

Une seconde phase de développement intervient dès qu'on décide d'injecter de l'eau ou du gaz dans le réservoir afin de maintenir la pression du champ. En effet, la récupération d'huile est directement liée au maintien de pression dans le réservoir : la production d'un fluide dans le réservoir est une fonction directe des gradients de pression. Lors de cette phase, on s'intéresse plus particulièrement à :

- la détection des hétérogénéités (failles, fissures, etc.) du réservoir et aux connections entre les puits,
- la géologie du terrain et aux propriétés de l'écoulement pour améliorer les modèles statique et dynamique de réservoir,
- la ré-estimation du volume d'huile en place.

Une grande attention est portée à la détection des hétérogénéités et des connections, car traduisant des incertitudes de grande influence sur les estimations de production futures. Par exemple, la présence d'hétérogénéités peut se manifester par l'existence de chenaux de fluides. Dans ce cas, une bonne caractérisation des connections entre les puits injecteurs et les puits producteurs est essentielle pour améliorer la connaissance du comportement du réservoir et les prédictions de performance.

Cette phase va permettre une meilleure réflexion sur les scénarii de production et sur l'emplacement de nouveaux puits et donc, *in fine*, une meilleure exploitation du gisement.

- **Autres phases de développement**

Il peut y avoir plusieurs autres phases de développement, et de nouveaux modèles de réservoir sont alors ajustés, basés sur l'obtention de nouvelles données. Au fur et à mesure du développement du gisement, les ingénieurs ont une idée plus précise des éventuels endroits où l'huile ne sera pas récupérée du fait de la configuration du schéma de développement existant combiné avec les effets de l'hétérogénéité du réservoir. Le placement de nouveaux puits peut alors être amélioré.

Ainsi les phases de développement successives permettent de déterminer les phénomènes régissant dans le réservoir et de construire un modèle géologique statique du réservoir et un modèle dynamique, dont la discrétisation donne un modèle numérique d'écoulement contraint par l'ensemble des informations disponibles. Le recours à la simulation de réservoir est donc envisageable, ces simulations sont généralement basées sur la modélisation mathématique du comportement dynamique du gisement par un système d'équations aux dérivées partielles non-linéaires.

1.1.2 Caractérisation du réservoir – Modélisation

La phase de caractérisation du réservoir permet la compréhension des propriétés du réservoir. Elle englobe toute l'étude du réservoir, en commençant par la construction d'un modèle géologique, puis par la modélisation de l'écoulement des fluides (modélisation dynamique) et enfin avec la modélisation numérique de l'écoulement prenant en compte ces deux modèles et reproduisant, à l'échelle du gisement, les données de production. Grâce aux progrès techniques et informatiques, les modèles de réservoir sont de plus en plus détaillés, le réservoir peut être discrétisé en plusieurs millions/milliers de blocs élémentaires ou mailles. Une maille représente un volume d'un gisement qui est souvent de l'ordre de 50-100m horizontalement et 0.1-10m verticalement. Les équations fondamentales d'écoulement sont alors appliquées entre chaque maille.

- **Modélisation statique du réservoir : Modèle géologique**

La modélisation statique du réservoir est réalisée pendant la phase de délimitation présentée précédemment, et au fur et à mesure de l'acquisition des données lors du développement du champ. Le modèle décrit les propriétés du réservoir avant la mise en production, c'est-à-dire lorsque le réservoir est à l'équilibre de pression. Cette modélisation est complexe du fait de l'importance des coûts relatifs à l'acquisition des données, le spécialiste ne disposant d'information que sur une infime partie du gisement qu'il étudie. Ces informations proviennent généralement de mesures sur carottes, de diagraphies, ou de la sismique. L'ingénieur réservoir doit donc extrapoler ces quelques données ponctuelles sur la totalité du champ pétrolier afin d'évaluer l'ensemble des caractéristiques non plus à quelques endroits donnés mais sur l'ensemble du réservoir.

La construction de ce modèle géologique est réalisée en trois grandes étapes :

- construction d'un modèle structural. Cette première modélisation provient principalement de la sismique associée aux mesures aux puits. Elle permet d'obtenir une image des différentes couches sédimentaires de réservoir, d'estimer l'emplacement des réservoirs dans le gisement en localisant les fonds et couvertures des réservoirs, de caractériser les principaux événements de fracturation, à savoir les failles principales, et enfin les contacts entre les fluides, notamment le contact huile-eau.
- construction d'un modèle stratigraphique. Ce modèle est une discrétisation du modèle structural sur un maillage cartésien fin.
- construction d'un modèle de faciès. Ce modèle détermine la géométrie et les proportions de chaque faciès et leur attribue des propriétés pétrophysiques. Il est obtenu à partir de la sismique, des mesures aux puits et des propriétés des carottes. Rappelons qu'un faciès caractérise les conditions sédimentaires et environnementales du sous-sol, chaque faciès ayant des propriétés spécifiques (type de roche, perméabilité, porosité, etc.).

Une schématisation des étapes de la construction du modèle géologique est représentée en Figure 1.1. Notons que pour cette figure, ainsi que les figures suivantes dans cette partie, nous avons représenté en jaune les données servant à la construction du modèle et en bleu le modèle considéré.

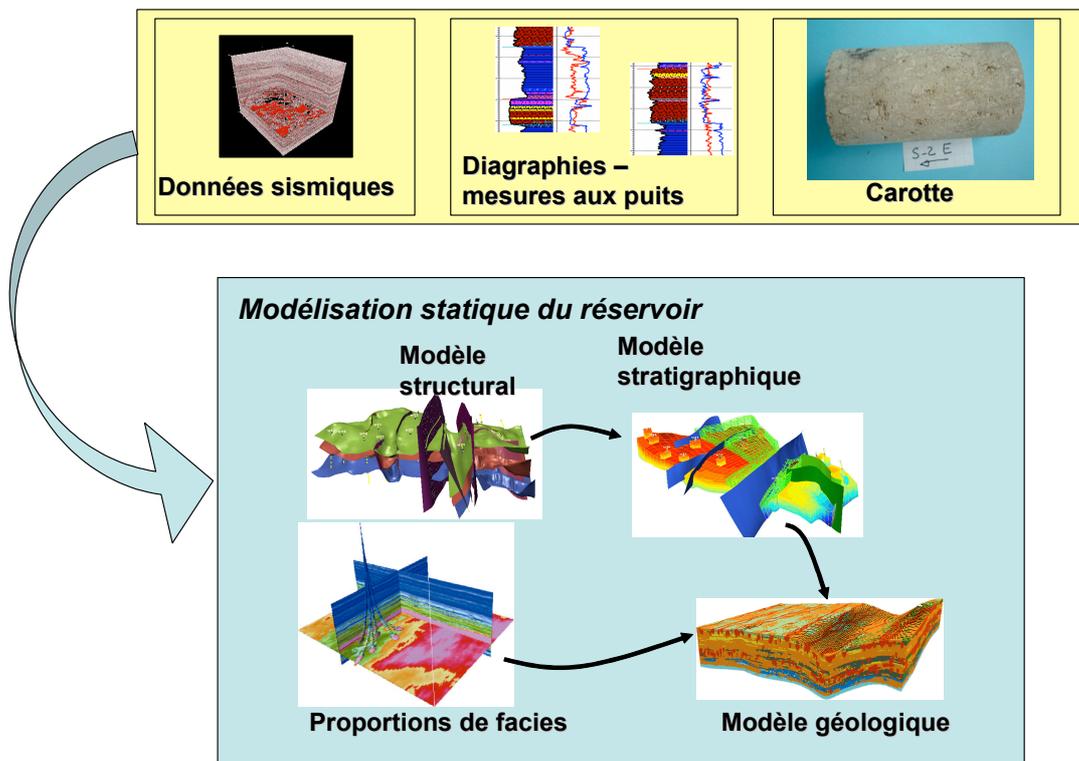


Figure 1.1 : Modélisation statique : Modèle Géologique

Remarque : Il existe une différence de plusieurs ordres de grandeur entre l'échelle d'observation des données sismiques, typiquement de l'ordre de la dizaine de mètres, l'échelle d'observation des mesures sur carottes de l'ordre de quelques centimètres et enfin la taille des mailles du modèle. Par conséquent l'interprétation des données sismiques et des mesures sur carottes est sujette à l'introduction d'incertitudes sur le modèle statique.

La construction d'un modèle statique de réservoir revient donc à déterminer, entre autres, la structure géologique du sous-sol, les limites du réservoir, le type de sédimentation (éolien, fluvial), le type de piège (faille), etc. La caractérisation du réservoir est uniquement statique, elle ne considère pas la dynamique des fluides et le type de déplacement, car la production du gisement n'a pas encore commencée. Ce modèle peut être mis à jour ou remis en cause lors de futures phases de développement, particulièrement lorsque le forage de nouveaux puits ou l'acquisition de nouvelles données suggèrent un comportement anormal. Il doit alors intégrer les nouvelles données statiques. Cette mise à jour du modèle statique est loin d'être systématiquement effectuée, elle est très coûteuse en temps de calcul, car réalisée sur un maillage cartésien fin.

- **Modélisation dynamique : Modèle de l'écoulement**

Une fois que le réservoir est mis en production et produit par simple déplétion, une étude des propriétés des fluides et du type de déplacement des fluides est nécessaire. Dans un gisement, entre autres phénomènes, il peut y avoir un écoulement simultané d'une ou plusieurs phases fluides : eau, huile, gaz. En début de production, il n'y a parfois qu'un seul fluide produit, l'huile, et lorsque l'on injecte de l'eau et/ou du gaz, 2 ou 3 phases fluides peuvent être imbriquées les unes dans les autres. Déterminer le type d'écoulement dans un réservoir est donc une tâche complexe. Ainsi, devons nous nous intéresser à l'écoulement d'un fluide polyphasique dans un milieu poreux. L'objet de ce chapitre n'est pas de faire une description physique complète du phénomène étudié. On pourra se reporter notamment aux deux ouvrages de K. Aziz et A. Settari [1] ou D.-W. Peaceman [45] pour une description plus détaillée de ces problèmes.

Ce sont les propriétés pétrophysiques d'une roche qui permettent de déterminer sa capacité à constituer un réservoir à hydrocarbures. Une roche est dite réservoir si elle possède une certaine capacité de stockage, si les fluides peuvent y circuler, et enfin si elle contient une quantité suffisante d'hydrocarbures avec une saturation satisfaisante. La capacité de stockage de la roche est caractérisée par sa porosité : plus la porosité de la roche est importante et plus la capacité de stockage est grande. Le déplacement des fluides à travers les pores de la roche est déterminé par la perméabilité de cette roche. Ainsi, le fluide se déplace d'autant plus facilement que la perméabilité de la roche est importante. Enfin, la quantité d'hydrocarbures et leur concentration sont prises en compte grâce au volume imprégné et à la saturation de chaque fluide (rapport entre le volume de fluide et le volume poreux).

Dans le cas le plus général, la modélisation de l'écoulement en milieux poreux fait intervenir trois phases fluides : l'eau, l'huile et le gaz. L'écoulement est décrit par la variation spatiale et temporelle des pressions et saturations de ces trois phases sous l'effet de l'injection ou de la production d'un ou plusieurs fluides par les puits. La modélisation de l'écoulement dans le réservoir est fondée sur la prise en compte de différentes lois physiques : loi de conservation de la masse, loi de Darcy, les équations d'état des fluides, ainsi que les conditions d'équilibre thermodynamique. Nous rappelons ici les deux équations cruciales pour la modélisation de l'écoulement, leur démonstration est donnée dans [1].

La première relation fondamentale est la loi de conservation de la masse. Elle s'obtient en écrivant que la variation de la masse d'une phase (eau/huile/gaz) dans un volume donné du milieu poreux est égale à la différence des quantités entrantes et sortantes de cette phase, pendant un temps donné. Pour chacune des phases p , la conservation de la masse s'écrit :

$$\frac{\partial}{\partial t}(\rho_p S_p \phi) + \nabla \cdot (\vec{U}_p \rho_p) = q_p \quad (1.1)$$

où q_p désigne le débit d'injection ou de production de la phase p (terme source), ρ_p et S_p respectivement la densité et la saturation de la phase p , ϕ la porosité de la roche et enfin \vec{U}_p la vitesse d'écoulement de cette même phase.

Outre l'équation de conservation de la masse, une relation liant la vitesse d'écoulement de chacune des phases \vec{U}_p au gradient de pression est nécessaire. Il s'agit de la loi de Darcy dont la formulation est donnée par (1.2).

$$\vec{U}_p = -\frac{k_{rp}}{\mu_p} \mathbf{k} \nabla (P_p - \rho_p \vec{g}z) \quad (1.2)$$

où \vec{g} représente l'accélération de la pesanteur, \mathbf{k} le tenseur de perméabilité de la roche, k_{rp} la perméabilité relative de la phase p , μ_p la viscosité et P_p la pression de la phase p .

La perméabilité relative k_{rp} d'une phase p quantifie l'aptitude de cette phase à circuler dans la roche en présence des autres phases. Elle est fonction des saturations des différentes phases dans le volume poreux.

Un facteur important pour la modélisation d'écoulement est la taille du maillage. Il est toujours préférable de travailler sur une grille de simulation aussi fine que possible afin de capturer les hétérogénéités du réservoir. Cependant, pour des raisons pratiques évidentes, il convient de faire un compromis entre la finesse de modélisation et les capacités de temps de calcul. Ainsi, les simulations d'écoulement sont généralement effectuées sur un maillage relativement grossier faisant appel à des techniques de mise à l'échelle pour calculer les propriétés équivalentes du réservoir sur les mailles grossières. En effet, la construction du

modèle statique est réalisée sur un maillage fin, d'échelle environ 10 mètres horizontalement et 0.1 mètres verticalement, comprenant jusqu'à 100 millions de mailles, alors que la modélisation de l'écoulement est effectuée sur un maillage d'échelle de l'ordre de la centaine de mètres et comprenant de 10 000 à 100 000 mailles. Lors de ce changement d'échelle, de l'information est agrégée, ce qui contribue à augmenter l'incertitude sur la caractérisation des propriétés du réservoir.

L'introduction des données dynamiques dans le modèle de réservoir permet une meilleure compréhension des phénomènes présents dans le réservoir. Ce modèle dynamique est ensuite discrétisé pour construire un modèle numérique d'écoulement qui va ensuite servir à optimiser l'exploitation du gisement. On peut envisager l'utilisation de différents types de modèles de simulation d'écoulement afin de rendre compte de la complexité de chaque gisement. Ces modèles prennent en considération la diversité des réservoirs (géologie, pétrophysique, etc.), le type d'écoulement (monophasique, polyphasique) et les différents modes d'exploitation (déplétion, récupération secondaire, etc.). Le principe de l'introduction de la dynamique des fluides dans le modèle de réservoir est présenté en Figure 1.2.

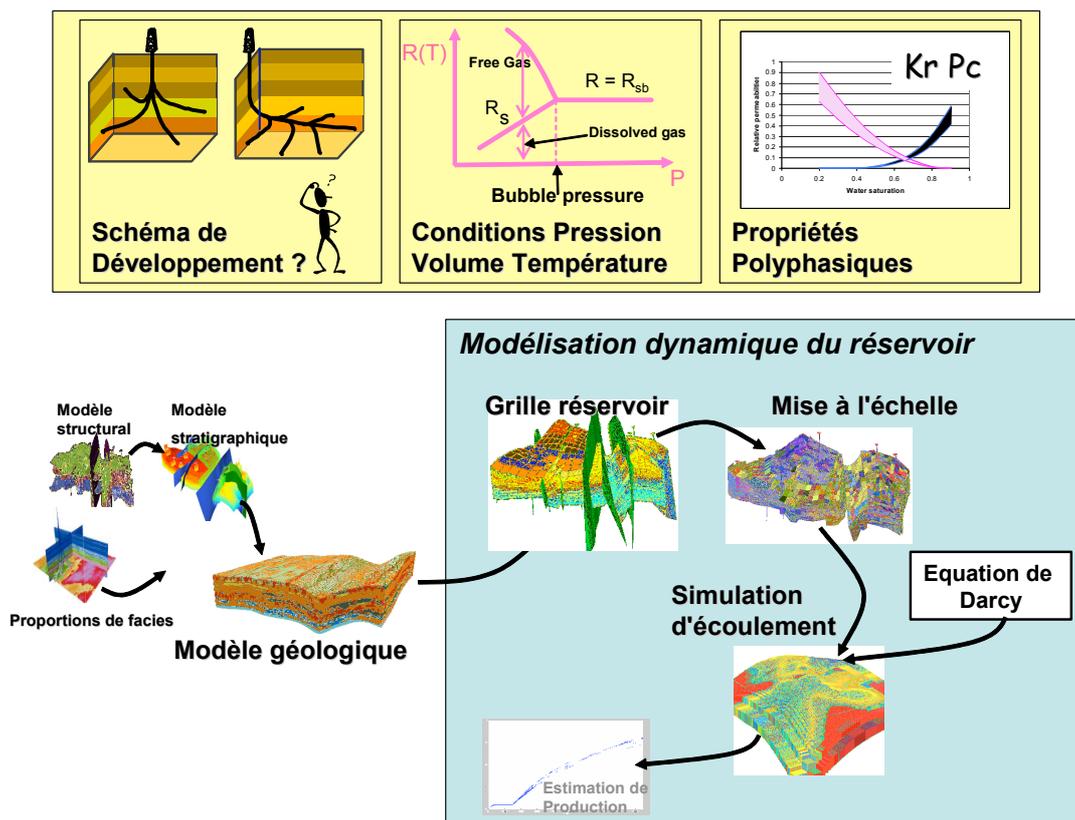


Figure 1.2 : Modélisation dynamique du réservoir

Hormis dans quelques cas très simples, les systèmes d'équations régissant les modèles mathématiques ci-dessus n'ont pas de solutions explicites. Il faut donc faire appel à des méthodes numériques pour obtenir des solutions approchées. L'accroissement des performances informatiques notamment la vitesse de calcul et l'amélioration des algorithmes

de résolution des systèmes d'équations aux dérivées partielles ont grandement contribué au développement de la simulation numérique pour l'étude des réservoirs pétroliers.

- **Prise en compte des données de production**

Le modèle dynamique présenté dans le paragraphe précédent tient uniquement compte des propriétés du réservoir et de l'écoulement des fluides, il est maintenant nécessaire de prendre en considération les données de production. Les décisions des ingénieurs ne peuvent en effet être basées sur des modèles numériques n'ajustant pas les données de production déjà observées. L'intégration des données de production dans le modèle d'écoulement est désigné sous le terme de « calage d'historique ». Le calage d'historique a pour but de calibrer le modèle numérique d'écoulement afin qu'il restitue au mieux le comportement et les données du gisement et qu'il reproduise l'historique de production du gisement. En pratique, le modèle est « perturbé » jusqu'à ce qu'il permette de reproduire les données observées. L'estimation des paramètres de ce modèle peut être vue comme un problème d'optimisation visant à minimiser l'écart entre l'historique réel de production et les estimations obtenues par le simulateur d'écoulement. La minimisation est alors réalisée soit par « essais et erreurs », soit en utilisant des techniques d'optimisation non-linéaires [30]. Le principe du calage d'historique est présenté sur la Figure 1.3.

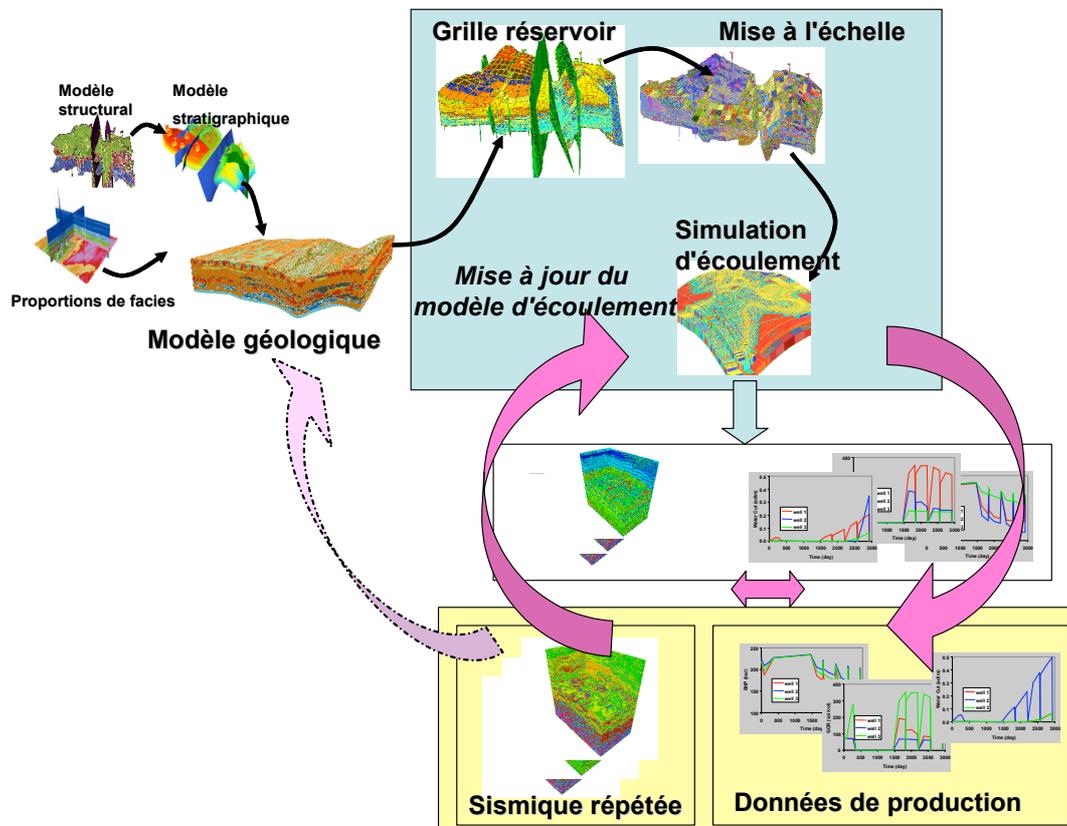


Figure 1.3 : Caractérisation du modèle dynamique

Remarque : Dans certains cas, le calage d'historique aboutit à une mise à jour du modèle géologique, mais ceci est loin d'être systématique.

Ainsi, dès lors que le gisement est en phase de production, il convient de procéder à un calage de l'historique de production pour contraindre les paramètres du modèle numérique d'écoulement. Les paramètres de ce modèle numérique de simulation sont réajustés au cours de l'exploitation du gisement, afin de pouvoir intégrer les nouvelles données de production et ainsi aboutir à une meilleure compréhension des caractéristiques du réservoir.

Ce modèle de simulation est fonction d'un grand nombre de paramètres d'entrée : perméabilité, porosité, saturation, emplacement des puits, scénario envisagé pour produire, etc. Dans la suite de ce manuscrit, le terme *paramètres* désigne les variables d'entrée du simulateur, il s'agit du terme communément employé en ingénierie de réservoir. Ce terme ne doit pas être confondu avec les paramètres du modèle statistique.

Les réponses en production sont évaluées sur des milliers de mailles, de ce fait, une simulation d'écoulement peut durer de quelques heures à plusieurs jours en temps de calcul, en fonction de la taille du réservoir et de la complexité du modèle d'écoulement résultant. Ce modèle de simulation est ensuite utilisé pour prévoir les productions futures du gisement sous différents schémas d'exploitation pour enfin aboutir à un développement optimal du gisement en prenant en compte les risques liés au choix du schéma d'exploitation.

1.1.3 Prévision de production et schéma de développement

Lors de la mise en exploitation du réservoir, il est indispensable de déterminer un schéma de développement optimisant la récupération à la surface des hydrocarbures présents dans le réservoir, tout en respectant certains critères économiques. Le plan de développement doit, par exemple, déterminer le nombre de puits à mettre en place, la configuration de ces puits (emplacement dans le gisement, longueur, etc.), etc. L'élaboration du schéma de développement requiert une bonne compréhension des phénomènes à l'œuvre dans le gisement. Toutes les caractéristiques du gisement ainsi que les conditions d'exploitation étant intégrées dans le modèle d'écoulement, il va permettre de prédire le plus précisément possible les productions du réservoir lorsqu'il est soumis à différentes contraintes et ainsi envisager une exploitation optimale. Il est alors, par exemple, possible d'étudier l'impact sur la production d'une modification du scénario de production et ainsi de déterminer un schéma de développement plus performant ou d'améliorer le schéma de production en cours. Ceci a pour but une gestion optimisée du réservoir et donc une meilleure évaluation du risque présent lors de l'exploitation du gisement.

Toutefois, il ne faut pas oublier que la phase de caractérisation du réservoir est effectuée grâce à des données nombreuses et variées, mais ne provenant que d'une infime partie du gisement. Ainsi, bien que le modèle de simulation soit calibré à partir d'un grand nombre de données, beaucoup d'incertitudes subsistent. Dans toute étude de réservoir, il est donc indispensable de quantifier les incertitudes.

1.2 Gestion des incertitudes en ingénierie de réservoir

1.2.1 Notion d'incertitudes

La gestion des incertitudes apparaît à chaque pas de la vie d'un gisement, en commençant par l'exploration puis par le développement et la production du gisement. Les premières incertitudes apparaissent dès la découverte du réservoir, lors de la phase de délimitation avec, par exemple, des incertitudes sur la taille du réservoir ou sur les estimations du volume d'hydrocarbures et des réserves productibles. A ces premières incertitudes s'ajoutent les incertitudes relatives à l'acquisition des données (quantité et fiabilité des prélèvements) et aux interprétations qui en résultent. Ceci est notamment le cas pour les données sismiques, la caractérisation des fractures du réservoir et pour déterminer les propriétés des fluides présents dans le gisement. Ces incertitudes sont d'autant plus grandes en début d'exploitation lorsque le nombre de puits et de données de production est faible. L'incertitude résulte alors d'une connaissance incomplète des caractéristiques du réservoir et laisse donc de nombreuses interprétations possibles. Il est par exemple possible de construire plusieurs modèles géologiques respectant les données acquises.

Le système physique du réservoir est modélisé par des équations mathématiques appropriées. L'utilisation de ces modèles nécessite l'introduction d'hypothèses indispensables pour la résolution du problème. Par exemple, pour la modélisation géologique, il faut tenir compte des caractéristiques du sous-sol, en particulier le nombre et la proportion des faciès, ceux-ci caractérisant la perméabilité, la porosité, la sédimentation, etc. Pour le modèle numérique de simulation, les hypothèses concernent principalement les dimensions du réservoir, la présence et la localisation de failles ou de fissures, la profondeur du contact huile/eau, le type d'écoulement, etc. De plus, le modèle numérique de simulation fait en général intervenir des dizaines voire des centaines de paramètres d'entrée parmi lesquels un grand nombre sont incertains : les paramètres géométriques (taille du réservoir, emplacement des failles, etc.), les paramètres d'exploitation (type de récupération, position et nombre de puits, débits de production, injection d'eau ou de gaz, etc.), les paramètres pétrophysiques (perméabilités, porosités), etc. La solution de ce modèle numérique, c'est-à-dire l'estimation de la production d'hydrocarbure en fonction des différents paramètres d'entrée, est donc déduite de données incertaines. Une faible variation de l'un des paramètres du modèle de simulation d'écoulement peut, par exemple, entraîner des prévisions de production très différentes.

Nous donnons ici plusieurs exemples pour illustrer l'influence sur la réponse d'une incertitude sur les paramètres, ou variables d'entrée. Ainsi, le premier exemple, très simple, permet d'illustrer l'effet des incertitudes lors de la phase de délimitation. Supposons que l'étude porte sur un gisement à exploiter de taille 3km par 3km. L'épaisseur du gisement est

mal connue, on la suppose entre 50m et 100m. La porosité moyenne varie entre 15% et 25% et la saturation en huile entre 60% et 75%. Enfin la récupération d'huile est estimée entre 20% et 40%. En considérant un prix d'un baril de brut à \$50, la valeur économique du gisement peut différer d'environ \$3 milliards de dollars entre le meilleur et le moins bon des cas. Il est évident qu'une telle incertitude sur l'économie du système aura un impact considérable sur la prise de décision de développer le champ ou non.

Trois autres exemples sont présentés sur la Figure 1.4.

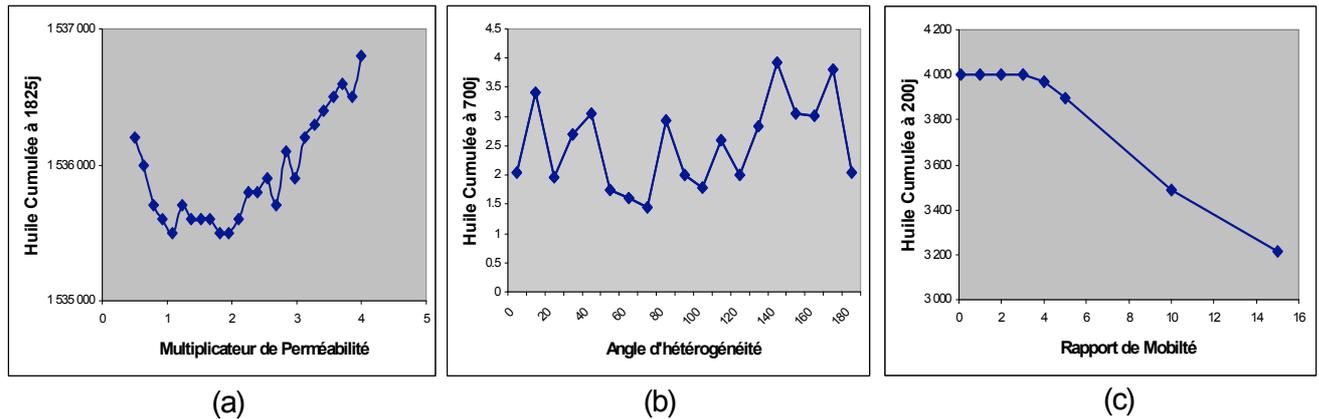


Figure 1.4 : Variations du cumulé d'huile produit en 1825 jours en fonction de différents paramètres

Le premier exemple, illustré sur le graphe (a) de la Figure 1.4, décrit l'impact d'une incertitude sur la perméabilité. D'après l'équation de Darcy (1.2), cette incertitude va se propager sur la valeur du gradient de pression et donc sur la production du réservoir. Par conséquent, une faible variation de la perméabilité peut entraîner une variation dans l'estimation de la production. Ce graphe représente les variations d'une réponse en production, la production cumulée d'huile après 5 ans, en fonction d'un coefficient multiplicateur de la perméabilité. Nous voyons que le coefficient multiplicateur de perméabilité a une influence globalement quadratique sur la production cumulée d'huile, mais que cette tendance générale est affectée d'un grand nombre d'irrégularités. Une estimation précise de la production, dans un contexte d'incertitude sur la perméabilité, n'est donc pas possible, une faible variation du coefficient multiplicateur pouvant avoir un effet non négligeable sur la production.

Le graphe (b) représente la valeur de la production d'huile cumulée à 700 jours en fonction de l'angle d'hétérogénéité. En effet, nous considérons un réservoir 2D, avec deux faciès de perméabilités différentes et nous faisons varier l'angle d'hétérogénéité dans le variogramme. Nous constatons ici une grande irrégularité de la production d'huile cumulée, une incertitude sur l'estimation de l'angle d'hétérogénéité peut entraîner de grandes variations sur la production d'huile cumulée.

Le graphe (c) représente la valeur de la production d'huile cumulée à 200 jours en fonction du rapport de mobilité entre les fluides, i.e. du rapport entre la mobilité de l'huile et la mobilité de l'eau. Ainsi, on constate que pour certains rapports, l'eau n'arrive pas avant 200 jours de production, alors que pour des rapports plus élevés, i.e. lorsque l'huile est plus visqueuse, l'eau arrive au puits producteur. Il y a percée d'eau et donc la production d'huile cumulée diminue. Il est donc important de bien estimer ce rapport de mobilité, celui-ci ayant une grande influence sur les estimations de production.

Ainsi, l'ingénieur doit faire face à une multitude d'incertitudes, chacune d'entre elles ayant une influence sur les prévisions de production. La prise en compte des incertitudes en ingénierie de réservoir a donc pour but de répondre aux deux questions suivantes :

- L'incertitude sur les paramètres du réservoir a-t-elle un impact sur les résultats de production?
- Comment quantifier et/ou réduire cet impact sur les prévisions de production?

La gestion des incertitudes devient donc essentielle en ingénierie de réservoir, elle permet non seulement d'accroître la production, mais précise surtout les enjeux économiques. Les coûts d'accès au réservoir et de forage de puits étant très importants, il est crucial d'investir dans des zones peu risquées et de ne pas effectuer des forages dans des endroits où les réserves sont incertaines. Les prises de décision doivent donc être envisagées en intégrant le contexte incertain inhérent à la connaissance partielle des réservoirs.

1.2.2 Gestion des incertitudes : problématique

Afin d'obtenir des prévisions de production réalistes, il est essentiel de tenir compte des caractéristiques des paramètres incertains influençant les réponses en production. Dans un premier temps, il est indispensable d'identifier, parmi tous les paramètres incertains spécifiés, ceux qui ont une influence réelle sur les évaluations de production, c'est-à-dire ceux dont une faible variation entraîne une grande variabilité dans la production. Il s'agit ainsi de mener une étude de sensibilité. Cette étude permettra aussi d'identifier les paramètres qui, bien qu'incertains, n'induisent pas de variations significatives de la production.

Une fois les paramètres incertains les plus influents déterminés, une phase de quantification du risque associé à ces paramètres est envisagée. L'ingénieur peut alors juger de réduire les risques, s'ils sont trop importants, par une étude plus approfondie des paramètres les plus influents. Pour cela, il peut par exemple, préconiser de nouvelles mesures ou réaliser des expériences en laboratoire ou directement sur champ.

Par ailleurs, cette phase de quantification des risques va permettre à l'ingénieur réservoir de considérer l'incertitude lors de ses prises de décision durant l'exploitation du gisement. Il peut ainsi effectuer une étude d'optimisation du schéma de production. Celle-ci

consiste à déterminer les valeurs des paramètres influents (et contrôlables) permettant de maximiser la production (nombre de puits, injection d'eau ou de gaz, débit d'injection, etc.), tout en minimisant le risque associé aux incertitudes. Il réduira ainsi l'impact des incertitudes sur la production en privilégiant les décisions qui aboutiront au compromis entre une production maximale et le risque.

Les deux objectifs principaux de la quantification des incertitudes sont donc :

- acquérir une meilleure connaissance du système étudié, notamment en terme de risque et éventuellement préconiser un retour à la caractérisation pour réduire le risque. Cette dernière phase n'est rien d'autre qu'un problème d'identification.
- la calibration des paramètres contrôlés par l'ingénieur afin d'optimiser la production du gisement.

La gestion des incertitudes est une des principales difficultés de l'industrie pétrolière, elle est devenue nécessaire pour la prise de décision. Ceci est d'autant plus important en début de production, lorsque le nombre de données servant à la calibration des modèles est peu important.

1.2.3 Gestion des incertitudes : méthodologie

Afin de bien caractériser l'impact de chaque paramètre incertain sur la production, un nombre important de simulations numériques d'écoulement est nécessaire [75]. Compte tenu du coût important requis par une simulation d'écoulement, à savoir de quelques heures à plusieurs jours, il est évident qu'on ne peut pas envisager de tester tous les scénarii envisageables directement via le modèle d'écoulement. Une solution pour remédier à ce problème est de construire un modèle approché du comportement du modèle numérique d'écoulement. Ce modèle approché traduit le comportement d'une réponse en production donnée, par exemple, le cumulé d'huile produit à 10 ans, en fonction de quelques paramètres considérés incertains. Les modèles approchés peuvent être évalués très rapidement et donc la question des coûts de calcul n'est plus un obstacle à la quantification des incertitudes, les simulations de Monte-Carlo sont, par exemple, envisageables.

Au cours de l'étude de quantification d'incertitude ou de schéma d'exploitation, le modèle approché est utilisé parce qu'il est simple et explicite et donc chaque « simulation » via ce modèle est immédiate, ce qui constitue une réduction de coût considérable. Ce modèle a pour paramètres d'entrée les différentes caractéristiques à optimiser, il reste donc le même pour chaque scénario. L'utilisation de ce modèle autorise ainsi l'ingénieur réservoir à tester autant de scénarii qu'il le souhaite, sans se soucier du coût effectif d'une simulation numérique d'écoulement.

Une gestion des incertitudes basée sur la construction d'un modèle approché du simulateur d'écoulement est donc incontournable pour l'obtention de prévisions de production fiables et pour l'optimisation des schémas de production tout en tenant compte du risque associé.

La construction de ce modèle approché pose toutefois un problème clairement identifié : comment traduire correctement le comportement de la réponse en production. Aussi, le modèle doit satisfaire aux deux caractéristiques suivantes :

- reproduire le comportement déjà observé via les simulations : le modèle devra ajuster les simulations réalisées directement via le simulateur d'écoulement. On parlera ici de qualité d'ajustement du modèle.
- prédire de manière fiable le comportement de la réponse en production pour des valeurs des paramètres du modèle pour lesquels aucune simulation n'a été réalisée. On parlera ici de qualité de prédictivité du modèle.

1.3 Objectif de la thèse

L'objectif de cette thèse est la mise en place d'une méthodologie pour conduire à une gestion maîtrisée des incertitudes lors de l'exploitation d'un gisement pétrolier. Il s'agit d'établir un formalisme cohérent de modélisation du comportement des réponses en production en tenant compte de la dépendance complexe, souvent fortement non-régulière, de ces réponses (water-cut, cumulé d'huile, etc.) par rapport aux paramètres physiques (perméabilité, porosité, etc.). Le problème consiste à construire un modèle approché du modèle numérique d'écoulement qui soit le plus prédictif possible, tout en contrôlant le nombre de simulations numériques d'écoulement requises pour calibrer ce modèle approché.

De récents travaux, [32] et [33], ont permis d'aborder le problème de la quantification des incertitudes en ingénierie de réservoir. Ces travaux sont basés sur la construction d'un modèle approché du simulateur numérique d'écoulement par une fonction polynomiale des paramètres, généralement du premier ou second degré. Cependant, dans la majorité des cas, cette hypothèse de régularité aux paramètres est mise en défaut. Lorsque nous parlerons de régularités, il s'agit de la possibilité de modéliser les relations paramètres – réponse à l'aide de fonctions polynomiales de degré fixé, non seulement un (cas de la linéarité), mais aussi deux, complets ou tronqués (par exemple de degré un avec interactions), voire trois ou plus.

Afin d'aboutir à une gestion maîtrisée des incertitudes, il est nécessaire de mettre en place une méthodologie peu coûteuse en terme de simulations numériques permettant de prédire de manière efficace des comportements de réponses non-régulières. Le meilleur compromis entre la qualité de la modélisation de l'incertitude et le coût informatique est recherché. La construction de ce modèle approché, à la fois fiable et prédictif, permettra une

meilleure analyse des scénarii et l'optimisation des schémas de production par une prise de décision visant à minimiser les risques liés aux incertitudes.

Le caractère prédictif du modèle approché est principalement lié à la forme du modèle choisi. Il est par exemple difficile d'obtenir une bonne qualité de prédiction si l'on cherche à prédire, par un polynôme du premier degré, un phénomène régi par un polynôme du second degré. Dans ce contexte, une attention particulière a été portée à l'étude des modèles d'approximation autres que linéaires ou quadratiques par rapport aux paramètres. Il est en effet fréquent d'avoir à étudier des réponses en production non-régulières par rapport aux paramètres dans le cadre d'une étude réservoir. Étant donné que nous ne disposons d'aucune information *a priori* sur le type de non-régularités de la réponse, les méthodes de statistique non-linéaire (moindres carrés non-linéaire, etc...) sont généralement délicates à mettre en œuvre. Nous envisageons donc une autre approche, plus empirique, basée sur un principe d'apprentissage.

Ce manuscrit décrit une méthodologie de prédiction de réponses non-régulières pour des réservoirs pétroliers. Nous commençons par présenter l'expression du modèle d'approximation que nous préconisons puis nous proposons une approche originale de recherche de l'estimateur de plus grande prédictivité. Nous établissons ensuite une stratégie d'échantillonnage du domaine expérimental.

Dans le chapitre deux, nous faisons un rapide bilan des travaux mathématiques sur la gestion des incertitudes en ingénierie de réservoir, celle-ci connaissant un intérêt croissant ces dernières années [32]. Nous introduisons les principales méthodes utilisées pour la quantification des incertitudes, notamment à partir de plans d'expériences euclidiens (méthodologie des surfaces de réponse). Nous présentons l'intérêt de ces approches puis indiquons leurs limitations dans le contexte de la modélisation de réponses non-régulières. Nous terminons par formuler une proposition méthodologique, consistant à modéliser la réponse de manière itérative.

Le troisième chapitre est consacré à l'emploi couplé de techniques de régression et de krigeage pour approximer la loi simulée par interpolation d'un échantillon de réponses du simulateur. Plus précisément, nous proposons ici d'ajuster une fonction polynomiale par régression et d'ajouter d'un terme aléatoire. Nous examinons quelles difficultés interviennent dans l'estimation de la fonction de covariance entre observations.

Dans le quatrième chapitre, nous proposons une approche innovante consistant à améliorer la qualité prédictive de l'approximation par ajout d'information fictive. Ces points fictifs, encore appelés points pilotes, sont choisis de manière appropriée pour améliorer la prédictivité de l'approximation.

Le cinquième chapitre est consacré aux problèmes d'échantillonnage du domaine incertain. Nous décrivons en premier lieu une étude de la dépendance de la réponse aux paramètres afin de définir un sous-ensemble du domaine expérimental pour lequel un enrichissement du dispositif est nécessaire. Ensuite, nous proposons différentes stratégies d'ajout de nouvelles simulations au plan d'expériences initial. Différentes options sont présentées ici, en particulier la recherche des points d'inflexion et des extrema. Nous formulons également un critère de sensibilité à la prédictivité. Nous terminons par l'ajout de points de confirmation et d'un critère permettant de juger la qualité de l'approximation.

Le sixième chapitre présente plusieurs exemples de modélisation de fonctions types par l'approche évolutive. Le chapitre sept est consacré à deux cas tests d'application en ingénierie de réservoir envisagés au cours de cette thèse. Nous présentons un cas synthétique et un cas réel appartenant au patrimoine minier de la compagnie nationale brésilienne Petrobras. Nous remercions ici Petrobras pour leur autorisation à exploiter leurs jeux de données.

Nous terminons enfin, dans le huitième chapitre, par présenter les conclusions de ces travaux de recherche et nous décrivons les différentes perspectives envisagées.

1.4 Conclusions

La quantification des incertitudes lors de la récupération d'hydrocarbures est d'importance cruciale en ingénierie de réservoir. Elle a un double enjeu : maximiser la production tout en minimisant les coûts d'exploitation. Elle constitue un défi puisqu'elle requiert une compréhension de l'architecture statique d'un réservoir et de son comportement dynamique lors de la production. Par l'intermédiaire du modèle numérique d'écoulement, la gestion des incertitudes implique un grand nombre de paramètres et peut être réalisée sur plusieurs réponses en production (production d'huile, production d'eau, de gaz, ...). Une simulation d'écoulement pouvant être très coûteuse en temps de calcul, il est nécessaire de disposer d'approches permettant de contrôler le coût d'une étude de quantification d'incertitudes.

Chapitre 2

Gestion des Incertitudes en Ingénierie de Réservoir - État de la question

Pour caractériser et quantifier l'impact des incertitudes en ingénierie de réservoir, les ingénieurs ont recouru à la simulation numérique d'écoulement. Des simulations sont réalisées pour déterminer les effets des hétérogénéités géologiques et de la variabilité des paramètres sur les performances du réservoir en production. Par paramètre, nous considérons par exemple la saturation initiale en eau, la perméabilité de la roche, etc. Les simulations servent également à mieux comprendre les mécanismes physiques qui régissent l'écoulement et à tester différents scénarii de production. Mais plusieurs paramètres peuvent interagir et donc affecter les estimations de production, ainsi un examen exhaustif du comportement de la production pour toutes les combinaisons possibles des paramètres est-il prohibitif en temps de calcul. Les paramètres qui ont une forte influence sur le comportement de la production doivent être identifiés pour se focaliser ensuite sur l'analyse des risques associés à ces paramètres influents. Nous présentons, dans ce chapitre, les différentes techniques employées au cours des dernières années pour la gestion des incertitudes sur les estimations de production de réservoirs pétroliers.

2.1 Méthodologie pour la gestion des incertitudes en ingénierie de réservoir

La gestion des incertitudes d'un réservoir repose sur des informations partielles recueillies lors de la caractérisation du réservoir. La modélisation demande donc l'introduction d'hypothèses qui permettent l'extrapolation des quelques mesures ponctuelles et locales à l'ensemble du réservoir. Le nombre de paramètres qui sont incertains peut dès lors être important, typiquement de l'ordre de la dizaine, voire de quelques milliers. De plus, les incertitudes sur ces paramètres sont d'autant plus grandes que la connaissance du réservoir est imprécise, ce qui est notamment le cas de gisements en développement ou en début

d'exploitation. Dans ce contexte, le recours à des approches probabilistes et stochastiques est rapidement devenu indispensable.

2.1.1 Position du problème

L'introduction dans les années 60 des ordinateurs et les progrès de l'analyse numérique ont permis de recourir à la simulation numérique en ingénierie de réservoir avec l'emploi de méthodes de Monte-Carlo pour l'échantillonnage. Ainsi, en 1967, Walstrom et al. [71] ont été les premiers à appliquer avec succès cette technique à plusieurs problèmes en ingénierie de réservoir. Ils proposent d'attribuer à chaque variable une loi de probabilité uniforme ou triangulaire et de réaliser les simulations directement via le simulateur d'écoulement. Mais le nombre prohibitif de simulations requises par les méthodes de Monte-Carlo, typiquement plusieurs milliers, rend cette approche difficile à mettre en œuvre pratiquement.

2.1.2 Les méthodes de Monte-Carlo en ingénierie de réservoir : recours à un modèle approché

Les performances croissantes des calculateurs n'ont pas permis de surmonter ce problème de coût, d'autant plus que les modèles simulés n'ont cessé de devenir de plus en plus complexes. Il est donc devenu quasiment indispensable de recourir non seulement à des méthodes d'échantillonnage systématique (méthodes de quasi Monte-Carlo), mais aussi à des modèles statistiques approximant les lois simulées. C'est ainsi que les plans d'expériences, en particulier les plans utilisés pour l'ajustement et l'analyse de surfaces de réponse, sont intervenus en ingénierie de réservoir. Ainsi, en 1980, Damsleth et al. [15] proposent d'utiliser cette méthodologie pour la simulation d'écoulement avec l'étude de la production d'un réservoir de la mer du Nord.

Dès lors, les méthodes de quasi Monte-Carlo ne sont plus utilisées directement via le simulateur d'écoulement. La méthodologie de quantification des incertitudes en ingénierie de réservoir se divise ainsi selon les phases suivantes :

- Détermination des paramètres incertains et définition des plages d'incertitudes associées
- Évaluation et analyse de la sensibilité de la réponse à l'incertitude aux paramètres
 - o Réalisation de simulations numériques d'écoulement en vue de déterminer l'impact des paramètres incertains sur la réponse en production étudiée,
 - o Analyse de sensibilité en vue de sélectionner les paramètres influents et de supprimer les paramètres non influents,

- Construction d'un modèle approché du comportement de la réponse en production en fonction des paramètres incertains déterminés comme influents lors de l'analyse de sensibilité,
- Estimation des incertitudes sur la réponse en production par méthode de quasi Monte-Carlo, via le modèle approché.

Ainsi, le modèle approché construit à l'issue de la phase de sensibilité est utilisé comme une version simplifiée du modèle de simulation pour quantifier l'impact des paramètres (de réservoir ou géologiques) qui influent sur les prévisions de production. Une analyse de risque est réalisée : elle consiste à effectuer un échantillonnage quasi Monte-Carlo sur les paramètres du modèle approché afin de prédire la densité de probabilité correspondante de la réponse. A chaque paramètre est associée une loi de probabilité, et grâce aux prédictions du modèle simplifié pour cet échantillonnage, on obtient une estimation de la densité de probabilité des estimations de production, sans avoir, à nouveau, recours au simulateur d'écoulement. La réponse prédite est utilisée pour calculer la densité et les quantiles (P10, P50 et P90) de la production, ce qui est primordial pour prendre des décisions dans un contexte incertain. Les valeurs P10, P50 et P90 caractérisant les valeurs possibles minimale, moyenne et maximale de la production.

La méthodologie utilisée pour la quantification des incertitudes en ingénierie de réservoir peut être illustrée par la Figure 2.1.

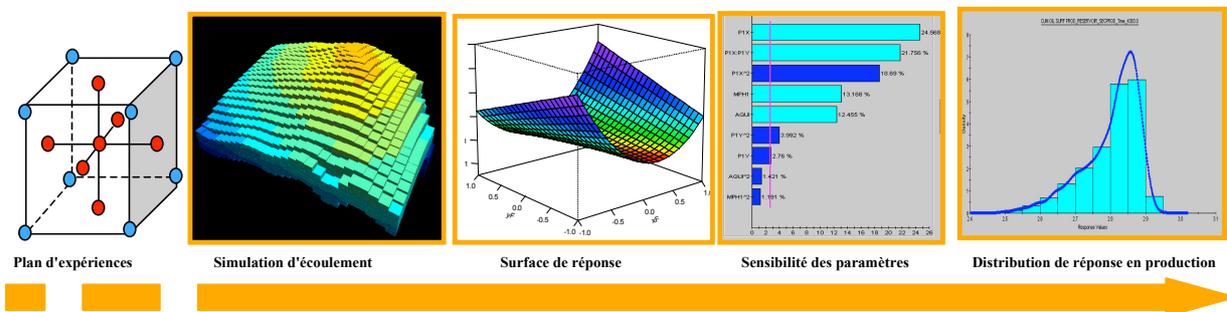


Figure 2.1 : Méthodologie employée pour la gestion des incertitudes.

- **Analyse de sensibilité et réduction du nombre de paramètres incertains**

Pour réduire le temps d'analyse et de construction du modèle approché, une première étape consiste à présélectionner les paramètres du modèle d'écoulement qui pourraient avoir une influence importante sur les estimations de production du réservoir étudié [70]. Cette première analyse va déterminer les paramètres qui sont à prendre en compte dans la construction du modèle approché : c'est l'analyse de sensibilité.

Cette étape est primordiale car si des variables influentes ne sont pas prises en considération pour la construction du modèle approché, cela peut amener à une mauvaise caractérisation de l'incertitude et du risque et donc induire des décisions erronées. A

contrario, le fait de prendre en compte des variables non significatives augmente la durée de l'analyse et le nombre de simulations d'écoulement à réaliser, sans pour autant aboutir à une amélioration des résultats.

L'approche traditionnelle pour l'analyse de sensibilité est souvent désignée par « one-parameter-at-a-time » [15]. Cette méthode est utilisée pour la quantification des incertitudes sur les résultats de simulation en fonction des différents paramètres. Trois niveaux par paramètres (P10, P50, P90) sont définis. Dans cette approche, on attribue la valeur P10 ou P90 à un paramètre, tous les autres restent au niveau P50 pour chaque simulation. Cette méthode permet de réaliser une étude de sensibilité des paramètres en vue d'éliminer les paramètres influents pour les phases d'analyse de risque et ainsi réduire le temps de calcul requis. Cette méthode peut être biaisée parce que son analyse repose sur la comparaison de différentes simulations pour lesquelles tous les paramètres sauf un sont maintenus au niveau P50. Implicitement, elle suppose donc que tous les paramètres sont indépendants et ne sont pas en interaction les uns par rapport aux autres.

Damsleth et al. [15] ont étudié un exemple d'analyse de sensibilité dans un champ de gaz de la mer du Nord. Par le biais de l'approche « one-parameter-at-a-time », cette analyse aurait requis l'équivalent de 15 500 simulations, soit environ une durée de 40 ans !

De ce fait, il est apparu la nécessité d'utiliser des techniques beaucoup moins coûteuses. L'industrie pétrolière a eu recours dans les années 1990 à la méthode des plans d'expériences, technique qui fut introduite par Fisher en 1924, notamment pour des applications à l'expérimentation réelle en agronomie et en biologie. En industrie pétrolière, les premières utilisations des plans d'expériences ont concerné les expériences réelles, réalisées en laboratoire. Toutefois, il y eu peu d'autres exploitations de cette méthode, hormis pour l'analyse de sensibilité, jusqu'à l'introduction par Sacks et al., en 1989, des plans d'expériences pour expériences simulées. Le champ des applications dans l'industrie pétrolière s'est alors largement élargi. Dès 1992, des plans ont été utilisés avec grand succès en simulation de réservoir [15] et [20].

Un plan d'expériences a en effet pour objet de planifier le nombre et l'emplacement des simulations à réaliser afin d'acquérir le maximum d'information à un coût le plus faible possible en faisant varier les paramètres simultanément. Il faut noter qu'en ingénierie de réservoir, le processus physique est étudié par le biais d'un simulateur d'écoulement, les expériences sont donc numériques et par conséquent leur traitement diffère des expériences réelles.

Dans cette approche, les paramètres varient simultanément, et donc les interactions entre les variables sont prises en compte dans les estimations, contrairement à la méthode « one-variable-at-a-time » où les paramètres varient l'un après l'autre. Ainsi, il est possible d'évaluer non seulement l'influence d'un paramètre sur la réponse en production, mais aussi

d'évaluer l'impact d'une interaction entre deux paramètres lors de la phase d'analyse de sensibilité.

En 1992, Damsleth et al. [15] proposent l'utilisation de plan D-Optimaux, qui permettent de gagner 30 à 40% de simulations par rapport à l'approche « one-variable-at-a-time » pour des analyses de sensibilité. De même, en 1999, White [72] et Dejean [16] exposent ainsi une méthode, basée sur l'utilisation de plans d'expériences D-optimaux et de régression linéaire, pour réaliser une étude de sensibilité aux paramètres. Ils proposent d'investiguer la dépendance de la production aux paramètres par une analyse des résidus obtenus par régression. La sensibilité de la réponse à un facteur est représentée par la dérivée initiale de la régression par rapport à ce facteur. L'analyse de sensibilité est basée sur un test de Student, qui permet de déterminer quels sont les termes de plus grande influence sur la production à l'aide de statistiques suivant une loi de Student [73]. Les facteurs qui ne sont pas influents sont supprimés du modèle afin de réduire la variabilité des résultats.

L'introduction de la méthode des plans d'expériences en ingénierie de réservoir a été d'importance cruciale, aussi bien pour une meilleure précision dans les analyses de sensibilité comme nous venons de le voir, que pour la mise en pratique des analyses de risque.

- **Plans d'expériences et construction d'un modèle approché**

Dès lors que les paramètres influents sont déterminés, la construction d'un modèle approché dépendant de ces paramètres nécessite le choix d'un certain nombre de simulations d'écoulement à effectuer. C'est un problème d'échantillonnage dont la solution dépend de la méthode utilisée pour construire le modèle.

Les trois principales méthodes utilisées en ingénierie de réservoir pour la construction d'un modèle approché sont la méthode des réseaux de neurones [23], [31], la régression multivariée [12], [16], [46] et [72], et le krigeage [20] et [74]. Elles peuvent ici intervenir pour obtenir une approximation du modèle numérique d'écoulement.

Depuis 1991, la méthode des plans d'expériences est fréquemment mise en œuvre pour la gestion des incertitudes en ingénierie de réservoir [23] et [33], elle est très utilisée pour évaluer les incertitudes sur la géologie et sur la production [32]. Les plans d'expériences ont, par exemple, montré leur intérêt pour les estimations de production en substituant au modèle numérique de simulation une surface de réponse approchée [69] et [70], puis pour l'optimisation [16] avec la recherche de l'emplacement optimal de nouveaux puits.

Par ailleurs, des plans d'expériences ont été utilisés pour traiter des problèmes spécifiques en ingénierie de réservoir, notamment pour la gestion de paramètres dits stochastiques [76] [79] ou discrets [25] [79].

Dans la suite de ce chapitre, nous présentons plus en détail les différentes étapes de ce processus de quantification des incertitudes. Nous introduisons principalement les notions les plus utilisées pour la gestion des incertitudes avec entre autres, la théorie des plans d'expériences et les méthodes de construction de modèle approché.

Nous verrons dans la suite de ce chapitre les plans d'expériences les plus classiquement utilisés pour la gestion des incertitudes.

2.2 Plans d'expériences pour l'échantillonnage du domaine incertain

Le premier travail à réaliser pour la quantification des incertitudes sur les prévisions de production est donc la construction d'un modèle approché du modèle d'écoulement. Il faut avoir défini au préalable une liste des paramètres susceptibles d'influencer la production et leur domaine de variation, c'est-à-dire le domaine expérimental (aussi appelé domaine incertain). Il faut également choisir la forme *a priori* du modèle approché. L'objectif ici est alors d'estimer les coefficients de ce modèle afin qu'il décrive correctement le comportement de la réponse. Tout ceci passe par un échantillonnage de valeurs de paramètres, pour lesquels les simulations vont être effectuées. Nous allons à présent voir comment choisir le nombre et les conditions des simulations à effectuer.

2.2.1 Choix du nombre et de l'emplacement des simulations à réaliser

La taille de l'échantillonnage résulte d'un compromis : il faut d'une part que le nombre de simulations soit suffisamment important pour estimer tous les coefficients du modèle approché et d'autre part, il est souhaitable de réaliser un nombre réduit de simulations pour limiter le coût de simulation.

Le choix des valeurs des paramètres pour lesquels les simulations doivent être effectuées est le principal problème : les simulations doivent en effet être choisies de manière à être le plus informatives possible pour un coût donné. Pour illustrer cela, nous donnons ci-dessous, sur la Figure 2.2, deux exemples d'échantillonnage pour deux paramètres : le coefficient d'aquifère et la perméabilité. Le coefficient d'aquifère prend ses valeurs entre 0.2 et 0.3, la perméabilité entre 150mD et 250mD. La réponse considérée ici est la production cumulée d'huile à sept ans de production. L'échantillon est constitué de cinq simulations.

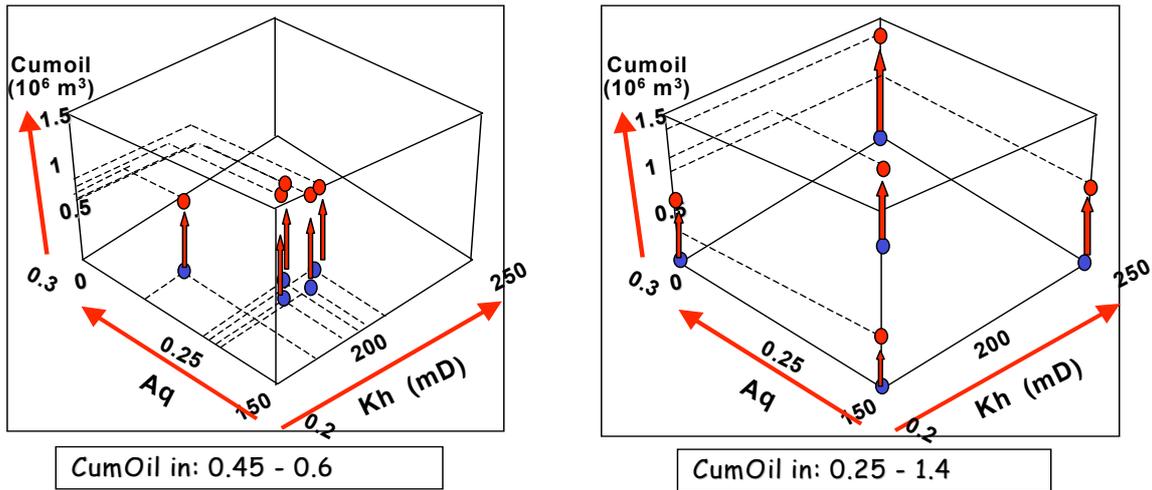


Figure 2.2 : Echantillonnage du domaine incertain

Le premier échantillonnage présenté ici est choisi aléatoirement dans le domaine incertain (méthode de Monte-Carlo). Le second est obtenu par la méthode des plans d'expériences (échantillonnage systématique).

Nous pouvons voir dans cet exemple que les valeurs extrêmes observées de la production cumulée d'huile décrivent une variation plus étendue de la réponse pour le second échantillonnage, à savoir de 0.25 à 1.4 millions m³ que pour le premier, à savoir de 0.45 à 0.6 millions m³. Ainsi, la méthode des plans d'expériences permet, pour un coût donné, de s'assurer que les points choisis pour la simulation décrivent au mieux la plage de variation de la réponse.

2.2.2 Plans d'expériences euclidiens

Dans le cas général, c'est-à-dire sans utiliser la méthode des plans d'expériences, une analyse rigoureuse de N paramètres prenant p niveaux différents demande p^N simulations. Un niveau est une valeur que peut prendre le paramètre. Ce nombre peut rapidement remettre en question la faisabilité de l'étude. Par exemple, l'étude de 12 paramètres prenant chacun trois valeurs différentes, requiert plus d'un demi-million (3^{12}) de simulations. Ces méthodes sont donc inapplicables pour un grand nombre de paramètres. De ce fait, la plupart des méthodes d'approximation, généralement les régressions ou les réseaux de neurones, s'appuient sur la méthode des plans d'expériences, puisqu'elle permet de minimiser le nombre de simulations requises pour la construction d'un modèle, tout en assurant un niveau d'information important.

Un plan d'expériences est représenté par une matrice D donnant les coordonnées de chaque simulation, une ligne représente une simulation, une colonne représente les valeurs d'un paramètre. Quand le domaine expérimental est un pavé, les paramètres sont classiquement normalisés à des valeurs entre -1 et 1, généralement -1 pour la valeur minimale

et 1 pour la valeur maximale d'un paramètre, et ce afin de travailler sans dimensions. La valeur moyenne est alors symbolisée par la valeur 0 [18].

Nous présentons, sur la Figure 2.3, un exemple de plan d'expériences pour l'étude de trois paramètres incertains, ainsi que l'interprétation résultant du plan « normalisé ». Il s'agit d'un exemple classique de plan composite centré à quinze simulations. Ce plan faisant intervenir trois niveaux, -1, 0, 1 par paramètre, suppose généralement une dépendance quadratique de la réponse aux paramètres.

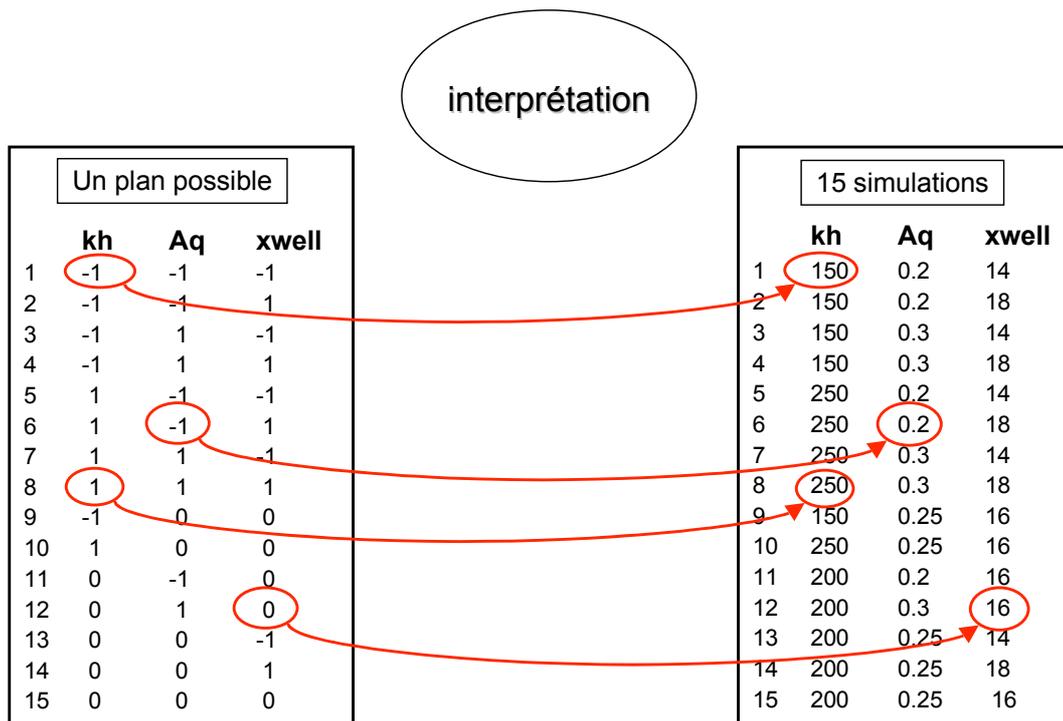


Figure 2.3 : Plan composite centré pour trois paramètres : Plan normalisé et simulations associée

La Figure 2.4 schématise la représentation de ce plan dans le domaine incertain.

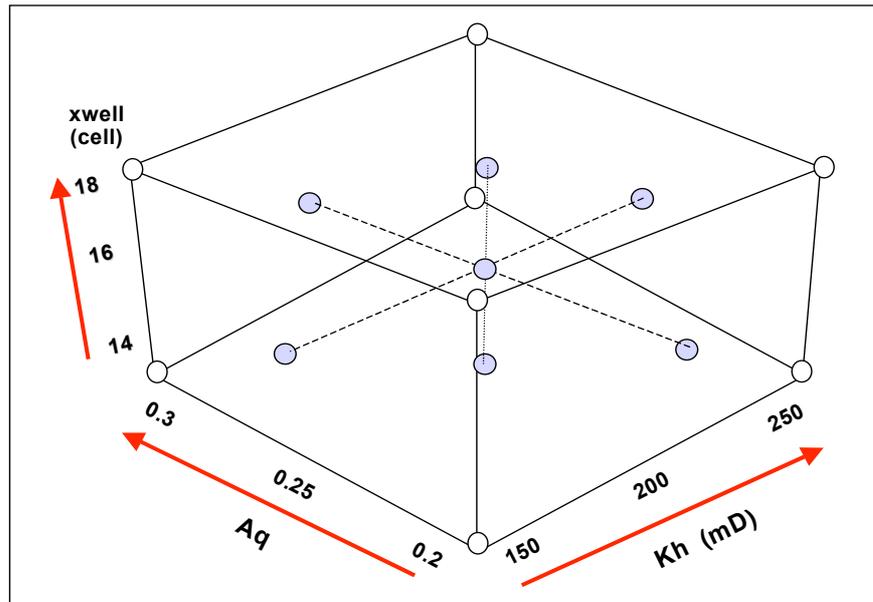


Figure 2.4 : Plan composite centré pour trois paramètres : Représentation spatiale dans l'espace des paramètres

Nous avons présenté ici un exemple de plan à 3 niveaux, mais dans le cas d'une étude de sensibilité, on utilise généralement des plans d'expériences à 2 niveaux. Ce sont des plans où n'interviennent que les valeurs extrêmes (-1,1) des paramètres considérés pour le choix des conditions de simulation. Parmi ces plans, le plus simple est le plan factoriel à deux niveaux, aussi appelé plan factoriel 2^k , k étant le nombre de paramètres étudiés [10]. L'ensemble des combinaisons possibles des paramètres est représenté, et donc le nombre total d'expériences est égal à 2^k . Dans la majorité des cas (lorsque k est grand), il n'est pas possible de simuler le plan factoriel complet du fait de son coût. Ainsi, pour un nombre élevé de paramètres, d'autres plans moins onéreux sont préférés. On citera notamment les plans factoriels fractionnaires de résolution III [10] et les plans Plackett-Burman [23]. Ils requièrent moins de simulations et permettent une étude de sensibilité fiable, mais supposent que les interactions sont négligeables.

Si les plans factoriels complets ou fractionnaires permettent de répondre au problème de l'analyse de sensibilité, la quantification des incertitudes, du fait de la non-régularité de la réponse, peut nécessiter une étude plus détaillée des paramètres en considérant plus de deux niveaux. Ainsi, il est utile de construire des modèles polynomiaux de degré (ou ordre) deux pour améliorer l'ajustement et surtout les prédictions. Pour ce faire, on peut considérer des plans à trois niveaux pour les paramètres. Une classe de plans adaptée à ce type de modèle est celle des fractions régulières des plans du type factoriel 3^k , bien qu'ils soient peu fréquemment utilisés dans la quantification d'incertitudes. Pour ajuster un modèle polynomial de degré 2, les plans les plus classiquement utilisés restent les plans composites centrés [16] [33] et [70] ou les plans de Box-Behnken [23] [72]. Ces plans requièrent un nombre de

simulations moins important que les plans 3^k , comme l'indique la Figure 2.5. Pour une étude de l'ensemble des caractéristiques de ces plans, nous renvoyons à [10] et [18].

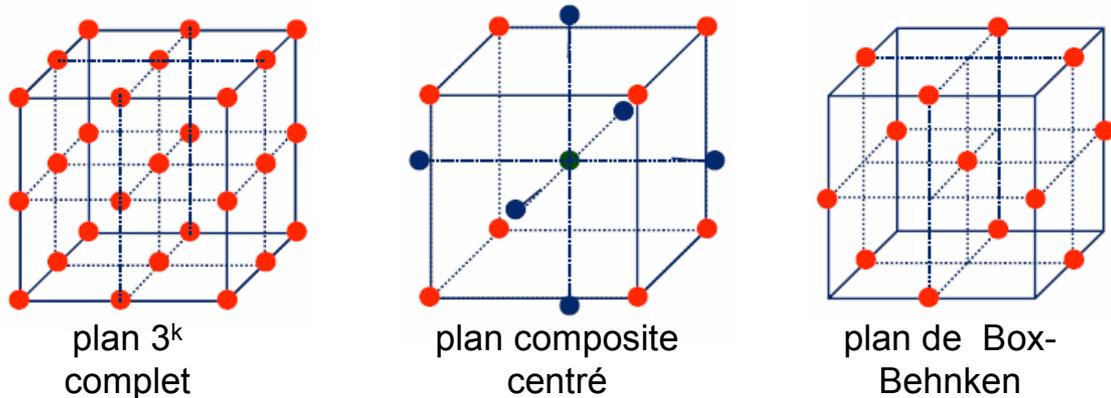


Figure 2.5 : Plans à trois niveaux : plan complet, composite ou fractionnaire

Des plans D-Optimaux ont également été employés, pour l'étude de sensibilité en ingénierie de réservoir par exemple, Damsleth [15] et Corre [12]. Un plan D-optimal est un plan qui maximise le déterminant de $'XX$, X représentant la matrice du modèle, soit pour un polynôme du premier degré :

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{p1} & \cdots & x_{np} \end{bmatrix}$$

où p est le nombre de paramètres du modèle.

Ce type de plan permet de minimiser la variance généralisée des estimations des paramètres. Plus de détails sur les plans D-optimaux sont donnés en annexe A.4.1.

Une fois les simulations d'écoulement effectuées conformément au plan d'expériences retenu, les réponses du simulateur obtenues aux différents points vont permettre de construire un modèle approché de la production.

2.3 Méthodes d'ajustement du modèle approché

Nous présentons dans ce paragraphe les méthodes les plus utilisées en ingénierie de réservoir qui sont les méthodes de régression polynomiale, l'approximation par réseau de neurones et enfin les méthodes géostatistiques de krigeage. Du fait de la prépondérance des méthodes de régressions en ingénierie de réservoir, nous développons cette méthodologie plus en détail. Nous nous contentons de rappeler les bases pour la modélisation par réseau de neurones et de citer quelques références bibliographiques pour le krigeage, cette dernière méthode étant largement décrite dans le chapitre 3.

2.3.1 Modèles de régression

Une des méthodes les plus utilisées pour la construction d'un modèle d'approximation d'un processus sur une région d'intérêt est la Modélisation par Surface de Réponse (RSM) ou régression. Un modèle de régression est un polynôme de degré quelconque (généralement de degré deux) qui ajuste au mieux les données, les coefficients de ce modèle étant obtenus par la méthode des moindres carrés. Les méthodes de régression fournissent des outils pour identifier les variables réellement influentes sur la réponse et pour construire un modèle représentant la réponse en fonction des variables influentes. Le modèle final peut être utilisé pour effectuer des prédictions du processus sur l'ensemble du domaine expérimental [15].

On considère donc une réponse d'intérêt Y qui dépend de variables ξ_1, \dots, ξ_n . La relation entre les données peut s'écrire sous la forme :

$$Y = f(\xi_1, \xi_2, \dots, \xi_n) + \varepsilon \quad (2.1)$$

où f est la fonction réponse inconnue et ε est un terme résiduel ou d'erreur qui représente les différentes sources de variabilité non modélisées par f .

Les variables ξ_1, \dots, ξ_n sont appelées variables naturelles parce qu'elles sont exprimées dans leurs unités de mesure naturelles. Dans la plupart des études de régression, les variables naturelles sont transformées en variables adimensionnées x_1, \dots, x_n , encore appelées facteurs, de valeurs comprises entre -1 et 1. Ces dernières valeurs sont appelées niveaux et représentent respectivement le minimum et le maximum du domaine de variation retenu pour la variable. L'équation (2.1) peut être reformulée de la manière suivante :

$$Y = f^*(x_1, x_2, \dots, x_n) + \varepsilon \quad (2.2)$$

La fonction f^* doit être approximée. Pour des raisons de coût, on utilise à cette fin généralement des polynômes de degré peu élevé, la plupart du temps de degré 2, qui sont supposés convenir sur une petite région de l'espace des variables.

Considérons par exemple le cas de deux facteurs, x_1 et x_2 et un modèle de régression polynomial complet de degré 1 avec interaction. Celui-ci est alors de la forme :

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon \quad (2.3)$$

Le coefficient constant β_0 est un estimateur de la moyenne de Y sur l'ensemble du domaine expérimental. Les termes x_1 et x_2 sont appelés termes linéaires. Les coefficients β_1 (respectivement β_2) représentent les variations attendues de Y lorsque x_1 (respectivement x_2) varie d'une unité : de -1 à 0 ou de 0 à 1. Ce coefficient peut être vu comme un estimateur de la dérivée partielle de Y par rapport à x_1 (respectivement à x_2). Le terme β_{12} est le coefficient lié à l'interaction à deux facteurs x_1 et x_2 . Il exprime le fait que les variations de Y provenant de x_1 (respectivement x_2) dépendent des valeurs de x_2 (respectivement x_1).

Le modèle (2.3) est le modèle interactif le plus généralement utilisé pour deux variables. Ce type de modèle est surtout employé pour des études de sensibilité lorsqu'on suspecte un effet d'interaction des facteurs sur la réponse. Par contre, pour la phase de modélisation, des modèles plus précis sont préférés, par exemple, toujours dans le cas de deux variables, un polynôme complet de degré 2 :

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon \quad (2.4)$$

Considérons donc un modèle complet de degré 2 pour n facteurs x_1, \dots, x_n , il s'écrit :

$$Y = \beta_0 + \sum_{i=1}^n \beta_i x_i + \sum_{i=1}^n \beta_{ii} x_i^2 + \sum_{i < j} \beta_{ij} x_i x_j + \varepsilon \quad (2.5)$$

Appliqué à l'ensemble des simulations effectuées selon le plan d'expériences retenu, on en déduit le modèle matriciel suivant :

$$Y = X\beta + \varepsilon \quad (2.6)$$

avec :

- X la matrice $n \times p$ du modèle. Elle dépend à la fois du modèle de régression et du dispositif expérimental,
- Y le vecteur $n \times 1$ des observations de la réponse aux n points du dispositif expérimental, ε vecteur des résidus ou erreurs inobservables,
- β le vecteur $p \times 1$ des coefficients (paramètres statistiques) du modèle,
- n le nombre de simulations et p le nombre de termes du modèle (en incluant la constante).

Les coefficients β sont inconnus et restent à estimer, usuellement par la méthode des moindres carrés ordinaires [8] [11] et [37].

Les méthodes de régression classiques supposent que les erreurs soient i.i.d., le vecteur des erreurs ε est un vecteur aléatoire normalement distribué de moyenne nulle et de variance uniforme σ^2 (homoscédasticité des erreurs). Sous cette hypothèse, et sous réserve que les colonnes de X soient indépendantes, on a d'après le théorème de Gauss Markov :

$$E(X\hat{\beta}) = XE(\hat{\beta}) = X\beta \quad (2.7)$$

où $\hat{\beta}$ est l'estimateur des moindres carrés, c'est-à-dire que l'estimateur $\hat{\beta}$ minimise la fonction L :

$$L = \sum_{i=1}^n \varepsilon_i^2 = {}^t \varepsilon \varepsilon = {}^t (Y - X\beta)(Y - X\beta) \quad (2.8)$$

Sous réserve que les colonnes de X soient indépendantes et que les erreurs soient i.i.d., l'expression de $\hat{\beta}$ est alors donnée par : $\hat{\beta} = ({}^tXX)^{-1} {}^tXY$. C'est un estimateur sans biais, plus précisément c'est l'estimateur de dispersion minimale dans la classe des estimateurs linéaires sans biais. On a en fait :

$$Cov(\hat{\beta}) = \sigma^2 ({}^tXX)^{-1} \quad (2.9)$$

Quant à l'estimation de l'espérance mathématique de la réponse Y , elle s'écrit de la forme :

$$E(Y) = X\hat{\beta} \quad (2.10)$$

L'estimation de la covariance de $\hat{\beta}$ est très importante parce qu'elle est directement liée à la qualité du modèle. Il apparaît qu'elle est fonction de la matrice du modèle et de la variance σ^2 . On retient en général comme estimateur pour σ^2 :

$$\hat{\sigma}^2 = \frac{SS_E}{n-p} \quad (2.11)$$

où SS_E est la somme du carré des erreurs, c'est-à-dire :

$$SS_E = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2.12)$$

et p est le nombre de termes du modèle.

Cet estimateur est sans biais et a des propriétés d'optimalité étudiées, par exemple, dans Rao et Mitra [49].

Ainsi, pour un dispositif expérimental fixé, si deux modèles sont possibles, il est recommandé de prendre celui qui à la plus petite estimation de σ^2 , ce modèle expliquant mieux la variabilité de la réponse. Cependant, l'estimation de σ^2 ne suffit pas à apprécier la qualité du modèle. Afin de mieux quantifier celle-ci, un certain nombre de tests statistiques sont à réaliser, ainsi qu'une analyse des résidus.

Le vecteur des erreurs ε (ou résidus) obtenu lors de la modélisation (2.10) est représenté par la différence entre les observations et les valeurs estimées, on a :

$$\varepsilon = Y - \hat{Y} \quad (2.13)$$

L'analyse de ces résidus de modélisation est une étape cruciale pour s'assurer de la qualité et de la fiabilité de la modélisation et de son pouvoir prédictif. Pour plus de détails sur le modèle linéaire présenté ci-dessus, nous renvoyons à [8], [19] et [41].

Après avoir construit le modèle linéaire, un test de significativité de la régression est effectué (cf. A.1.1). Ce test a pour but de déterminer s'il y a bien une dépendance linéaire entre la réponse et au moins une des variables explicatives. C'est l'objet de l'analyse de variance. Si le modèle linéaire est validé, une mesure de la qualité d'ajustement et de prédiction de l'approximation est possible. Trois statistiques sont calculées : R^2 et R_{adj}^2 permettent d'apprécier la qualité d'ajustement du modèle aux données et le coefficient Q_2 permet d'estimer la qualité de prédiction du modèle. Globalement, pour être efficace, l'estimation de la qualité d'un modèle doit prendre en considération les trois coefficients : R^2 , R_{adj}^2 et Q_2 . Une description de l'obtention de ces trois statistiques, ainsi que des tests statistiques plus classiques, est décrite en annexe A.1.2.

Une alternative aux techniques classiques de régression en ingénierie de réservoir est l'approximation par réseau de neurones.

2.3.2 Réseaux de Neurones

La technique des réseaux de neurones permet l'approximation d'une fonction à partir de l'observation de variables de sortie pour différentes valeurs de variables d'entrée [47] [52]. De ce fait, elle constitue aujourd'hui une alternative à certaines méthodes d'inférences non paramétriques.

Les réseaux de neurones peuvent être utilisés pour prédire des valeurs de simulation, après apprentissage du réseau sur un certain nombre de résultats de simulation.

Considérons 2 variables aléatoires, X et Y , de dimension respective n et 1, et un neurone qui, à partir des entrées X , produit la sortie Y .

Une représentation d'un neurone formel est donné sur la Figure 2.6 où :

- les x_i représentent les données d'entrées,
- w_i représente le poids de la $i^{\text{ème}}$ connexion,
- f est la fonction d'activation,
- $y = f(w_1 x_1 + \dots + w_n x_n)$ représente la sortie du neurone.

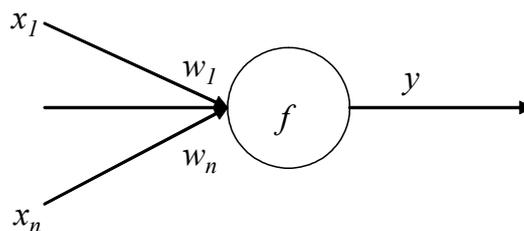


Figure 2.6 : Représentation d'un neurone formel selon Mc Culloch et Pitts

Le principe général de la méthode consiste à construire une fonction constituée de neurones qui échangent des données par des liens. Le réseau ainsi formé comporte des nœuds d'entrée et des nœuds de sortie, comme illustré sur la Figure 2.7.

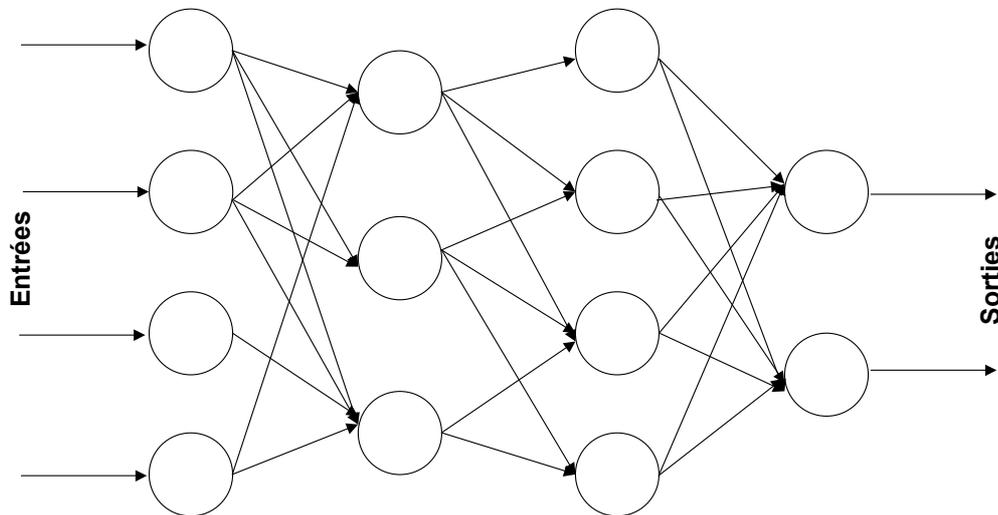


Figure 2.7 : Représentation d'un réseau de neurones multicouche

Supposant qu'on dispose d'un échantillon D , de taille m , d'observations des variables (X,Y) . Étant donné l'échantillon D , on veut trouver la meilleure fonction \hat{f} qui restitue les valeurs Y de l'échantillon pour les valeurs de X .

Les résultats des simulations initiales D sont utilisés pour l'apprentissage du réseau : les poids affectés aux liens sont ajustés par un algorithme de régression pour que le réseau restitue le plus fidèlement possible les résultats connus. Après la phase d'apprentissage, le réseau peut être utilisé pour prédire les résultats de simulation à partir de nouveaux jeux de données [47]. La technique est robuste et peut être appliquée aux modèles non-linéaires et non-stationnaires dans un environnement multivarié.

Les principales étapes d'une approximation par réseaux de neurones sont :

- le choix des paramètres d'entrée et de sortie
- la définition de la topologie du réseau (choix du nombre de neurones et des liens)
- la phase d'apprentissage
- l'optimisation éventuelle de la topologie du réseau
- l'utilisation en restitution ou en prédiction.

Les deux premières étapes sont réalisées grâce à la connaissance de l'ingénieur réservoir. Le principe de fonctionnement d'un réseau de neurones ainsi que les différentes étapes de sa mise en place sont décrits en annexe B.1.

Le principal enjeu dans la mise en place d'un réseau de neurones est le calcul des poids des connections permettant d'obtenir le comportement et les résultats souhaités. Pour obtenir un bon réseau, il faut un bon apprentissage. Cet apprentissage consiste à présenter au réseau un certain nombre de jeux de paramètres d'entrée dont on connaît la réponse afin d'ajuster les poids (w_i) associés aux liens. Une fonction coût mesurant les différences entre les réponses du réseau et les résultats attendus est ensuite minimisée en modifiant les w_i .

Pour apprendre, comprendre ou définir un processus d'apprentissage, il est nécessaire de modéliser l'environnement dans lequel opérera le réseau de neurones, c'est-à-dire qu'il faut savoir quel type d'information sera fournie au réseau. Cette modélisation est le paradigme d'apprentissage. Ensuite, il faut définir les règles d'apprentissage qui gouvernent la mise à jour des poids des connexions. Ces règles sont ensuite utilisées par un algorithme d'apprentissage pour procéder à l'ajustement des poids. On peut considérer qu'il existe trois principaux paradigmes d'apprentissage : supervisé, non supervisé et hybride. Une description détaillée de l'apprentissage supervisé, est donnée en annexe B.1.

L'apprentissage non supervisé modifie les poids du réseau en fonction d'un critère interne, indépendant de l'adéquation entre le comportement du réseau et la réponse qu'il doit caractériser. L'apprentissage supervisé, au contraire, dispose d'un comportement de référence vers lequel il tente de faire converger le réseau. L'apprentissage hybride suppose qu'un comportement de référence précis n'est pas disponible, mais qu'en revanche il est possible d'obtenir des indications qualitatives ou lacunaires sur les performances du réseau.

Ce modèle est validé par le calcul des trois statistiques mentionnées lors de la régression : R^2 , R_{adj}^2 et Q_2 . Tant que ces statistiques ne sont pas satisfaisantes, le modèle n'est pas de bonne qualité et donc la phase d'apprentissage continue.

Les réseaux de neurones sont encore peu utilisés pour la construction d'un modèle approché du simulateur d'écoulement en ingénierie de réservoir parce que la phase d'apprentissage est délicate à mettre en place. Généralement, les techniques géostatistiques de krigeage leurs sont préférés.

2.3.3 Le krigeage

Le terme « Géostatistique », employé par G. Matheron dès 1962, désigne l'étude statistique des processus spatiaux appliqués aux phénomènes naturels [34]. On définit en toute généralité la Géostatistique comme l'étude des variables numériques réparties dans l'espace. Le préfixe « Géo » permet de signifier la prise en compte de la répartition spatiale des données. Nous retiendrons principalement la définition qui précise l'outil probabiliste employé : « La géostatistique est l'application du formalisme des fonctions aléatoires à la reconnaissance et à l'estimation de phénomènes naturels ». Cette approche est très

intéressante parce qu'elle permet de modéliser à la fois la tendance spatiale et les corrélations spatiales. Les méthodes géostatistiques ont été utilisées intensivement pour le traitement statistique des problèmes miniers et en géologie [26], elles sont généralement limitées à des problèmes en 3 dimensions. Toutefois, les algorithmes de krigeage peuvent être étendus à n dimensions. L'apparition de la géostatistique ailleurs dans la bibliographie statistique est plus récente [6] [7] [13] [34] et [50].

Les méthodes de krigeage ont été introduites pour la modélisation d'expériences simulées par Sacks et al. [54] en 1989 avec la résolution numérique d'un système d'équations différentielles issues de la mécanique des fluides. Il s'agissait de modéliser la vitesse de flamme lors de la combustion du méthane. De nombreux travaux de recherche ont été consacrés à l'analyse d'expériences simulées, nous citerons notamment les ouvrages de Santer et al. [57] et plus récemment de Fang K.T et al. [22].

En ingénierie de réservoir, la méthode de krigeage est très utilisée depuis une quinzaine d'années pour la caractérisation de réservoir avec la construction d'un modèle géologique du gisement, avec, entre autres, les travaux de A. Journel [27]. L'utilisation du krigeage pour la construction d'un modèle approché du simulateur d'écoulement a été introduite par Egeland et al. [20] en 1992, ils ont montré que cette méthode améliore les surfaces obtenues par régression. Actuellement, le krigeage est largement utilisé par les compagnies pétrolières lors de la construction de ce modèle approché pour des analyses de risques [5] et [74].

Remarque : Pour certains choix de la structure de covariance, la méthode de krigeage est équivalente à une interpolation de type splines. Une correspondance entre les deux méthodes a été discutée dans la littérature géostatistique [6] et [7]. Une spline étant une famille de fonctions régulières de courbure minimale, particulièrement de polynômes, l'utilisation des splines et a fortiori du krigeage permet de meilleurs ajustements que les approximations polynomiales simples.

La méthodologie de krigeage étant partie intégrante de la thèse, nous ne la présentons pas ici, elle est amplement détaillée dans le chapitre suivant.

2.4 Limites des approches existantes

Nous avons présenté dans ce chapitre les trois méthodes d'approximation généralement employées en ingénierie de réservoir pour la construction d'un modèle approché du simulateur d'écoulement. Chacune de ces méthodes permet une bonne quantification des incertitudes, mais sont d'application limitée.

La méthode des réseaux de neurones permet de prendre en compte les non-régularités, mais elle ne permet pas une compréhension facile du modèle et de la relation entre les données, ce qui peut être utile au chercheur qui souhaite comprendre le phénomène à

modéliser. De plus, ces méthodes sont développées dans un but d'estimation et non de prévision de l'incertitude de la réponse en des points non échantillonnés. C'est principalement pour ces raisons que ces modèles sont peu utilisés en ingénierie de réservoir. De plus, Yeten et al. [74], qui ont utilisé des réseaux neurones pour l'ajustement d'une surface de réponse, ont montré que cette méthode ne donne pas des résultats aussi satisfaisants que les méthodes de krigeage.

A l'opposé, la méthode de krigeage connaît actuellement un succès grandissant. Il existe sous diverses formes : le krigeage simple, le krigeage ordinaire ou le krigeage universel, chacune ayant ses caractéristiques spécifiques. Les méthodes de krigeage varient en fonction de la manière dont est estimée la structure de covariance. Comme dans le cas des régressions, la modélisation par krigeage est basée sur la réalisation de simulations et donc l'échantillonnage est crucial pour optimiser le coût de simulations.

Enfin, la méthode la plus utilisée jusqu'à présent pour une quantification des incertitudes en ingénierie de réservoir est l'utilisation de plans d'expériences associés à une modélisation de type régression [32]. Cette dernière méthode est très répandue puisqu'elle est à la fois peu coûteuse en temps de calcul, facile à mettre en place et surtout, elle permet de disposer d'une expression analytique explicite de la réponse. Elle est très efficace lorsque la production peut être ajustée par des modèles polynomiaux de divers degré : un, deux voire trois ou plus, typiquement lorsque les paramètres incertains peuvent varier dans un domaine relativement restreint. C'est lorsque la dépendance de la réponse par rapport aux paramètres ne peut pas être approchée par un modèle polynomial (après éventuelles transformations des échelles de mesures des paramètres ou/et la réponse) que les plans euclidiens classiques trouvent leurs limites, comme nous le montre la Figure 2.8.

La Figure 2.8 montre la modélisation approchée d'une fonction non-régulière (a), par la méthode des plans d'expériences classiques, avec respectivement un plan factoriel à quatre expérimentations (b) et un plan composite centré à neuf expérimentations (c).

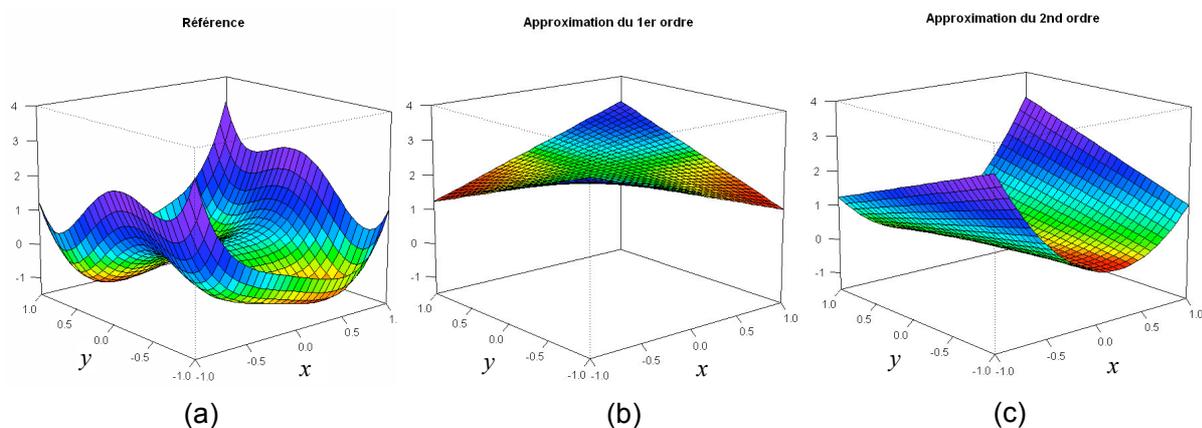


Figure 2.8 : Ajustement d'une fonction analytique avec les plans classiques

Ainsi, certaines réponses en production, de comportement non-régulier aux paramètres, possiblement discontinues, restent difficiles à appréhender. Compte tenu de la piètre qualité prédictive des modèles obtenus, il n'est pas envisageable d'utiliser ces modèles approchés pour évaluer le risque ou optimiser le processus de décision, une telle imprécision pouvant amener à des prises de décision non fondées ou à une évaluation du risque erronée.

Ainsi, une nouvelle méthodologie, à la fois plus flexible et plus robuste, est nécessaire pour permettre une modélisation de phénomènes non-réguliers et tenir compte des contraintes sur le nombre de simulations possibles. Cette nouvelle méthodologie doit adresser les divers problèmes qui ne peuvent être pris en considération par les méthodes de régression classiques : détecter les non-régularités de la réponse, ajuster une surface non-stationnaire et tenir compte du biais potentiel dû à l'échantillonnage non réguliers des n points $\{x\}$.

2.5 Proposition méthodologique

La nécessité de développer une approche de modélisation de réponses en production non-régulière, tout en contrôlant le coût de simulation induit, constitue l'objectif ambitieux de cette thèse. Afin d'atteindre cet objectif, nous proposons une méthodologie de modélisation non-régulière basée sur les plans d'expériences et les méthodes de modélisation géostatistique.

Les phénomènes physiques étudiés, c'est-à-dire la dépendance de la production d'hydrocarbures à différentes options de développement pour des géologies complexes et incertaines, sont très coûteux à modéliser vu les temps de simulation. Ce problème de réduction du temps de calcul et *a fortiori* du nombre simulations, justifie une approche par plans d'expériences.

La solution consistant à modéliser une surface de réponse non-régulière par rapport aux paramètres avec des plans d'expériences adaptés nécessite plus de simulations qu'un plan d'expériences euclidien classique du second ordre (type plan composite centré ou Box-Benhken). On peut alors se demander si le coût supplémentaire est justifié. C'est pourquoi, il nous a semblé judicieux de procéder de manière itérative. Tout d'abord, on ajuste à moindre coût un modèle de régression polynomial à l'aide d'un plan d'expériences peu coûteux. Ensuite, dans le cas où cette surface de réponse n'est pas jugée suffisamment prédictive, on ajoute quelques simulations judicieusement choisies. Nous procéderons par ajustements itératifs, partant d'un plan d'expériences initial et le complétant au fur et à mesure du processus, afin de maîtriser le nombre de simulations et d'intégrer graduellement les non-régularités potentielles des réponses en production. Dès lors que la prédictivité du modèle est satisfaisante, l'échantillonnage est jugé optimal.

La Figure 2.9 représente une illustration du principe de la méthodologie, dite adaptative, proposée dans cette thèse.

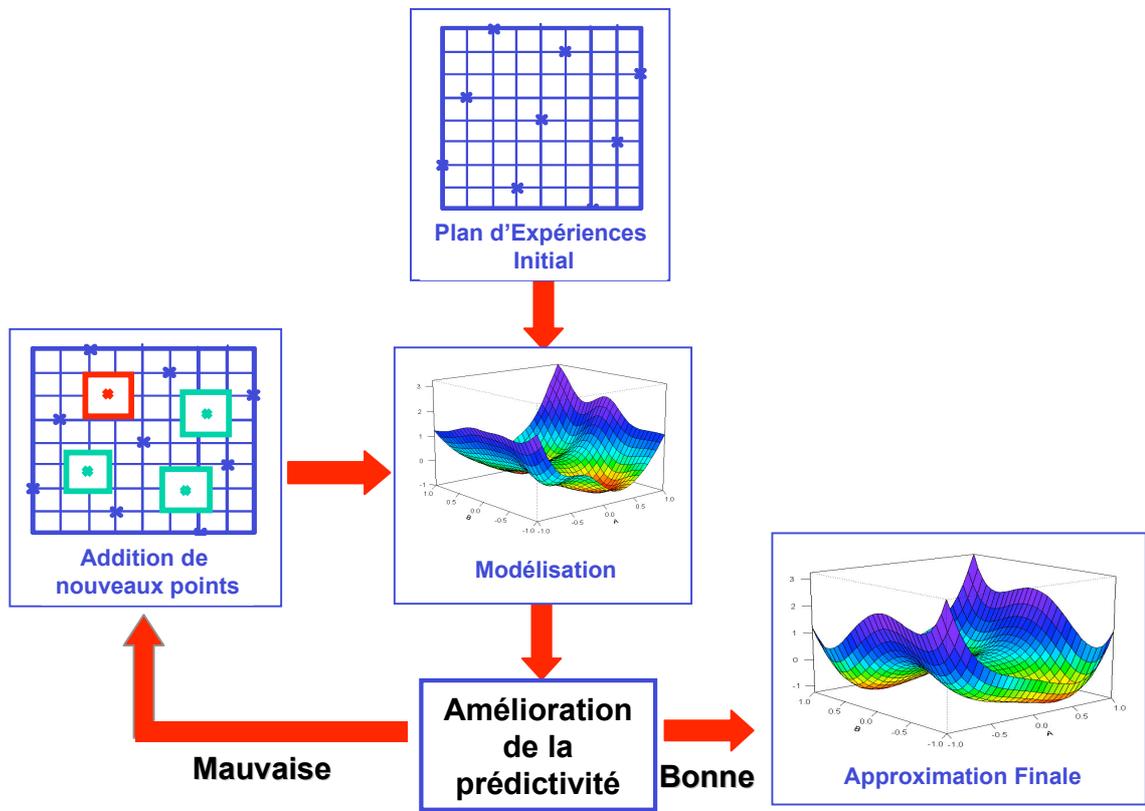


Figure 2.9 : Principe de la méthodologie adaptative

La première étape de la méthodologie consiste à déterminer les valeurs de chaque paramètre de l'étude pour lesquels les simulations numériques d'écoulement vont être effectuées. Puisque les simulations de réservoir sont très coûteuses en temps de calcul, il est important de sélectionner un nombre réduit de simulations informatives. La méthode des plans d'expériences est particulièrement bien adaptée à ce type de problème. Parce que nous ne connaissons pas la nature de la relation entre la réponse et les paramètres, le dispositif expérimental initial doit permettre d'ajuster une variété de modèles et doit donner des informations sur chaque partie du domaine expérimental. Dans ce contexte, l'utilisation de plans hypercubes latins de distance dite « maximin » est recommandée puisque ces plans combinent les propriétés de deux critères. Les hypercubes latins imposent la sélection de points à l'intérieur du domaine, et du fait de leurs propriétés de projection, ils peuvent être utilisés dans des situations pour lesquelles seulement un ou plusieurs paramètres ont un impact important sur la réponse. La distance maximin impose une répartition uniforme des points expérimentaux : on ne peut avoir de points du plan trop proches ou uniquement sur la diagonale du domaine. On gagne donc en information. Le compromis consistant en l'utilisation d'un critère de distance maximin et d'indice minimum dans la classe des hypercubes latins permet l'obtention de plans efficaces pour prédire la réponse, à la fois

quand peu ou beaucoup de paramètres sont importants [55]. Une description détaillée des différents types de plans est présentée en annexe. Remarquons que si nous avons une hypothèse *a priori* sur la régularité de la réponse à ajuster, les plans d'expériences classiques tels que les plans factoriels ou composites peuvent s'avérer tout aussi pertinents.

Dès lors que le dispositif expérimental initial est déterminé, la mise en place de cette méthodologie adaptative demande d'une part la définition d'un modèle d'approximation pour des expériences déterministes, d'un moyen pour améliorer la prédictivité *a priori* de cette approximation, et enfin une stratégie d'échantillonnage optimal du domaine expérimental. Afin de contrôler au mieux le nombre de simulations réalisées, nous proposons de procéder de manière évolutive, d'enrichir graduellement le dispositif expérimental courant. Ce type d'approche va également permettre d'optimiser l'emplacement des nouvelles expérimentations et donc d'augmenter la quantité d'information apportée par ces simulations.

Les différentes étapes de cette méthodologie seront présentées en détails dans les chapitres suivants. Chaque chapitre représentant une étape du processus, nous les présentons dans l'ordre où ils interviennent dans l'approche de modélisation adaptative.

Chapitre 3

Modélisation d'expériences déterministes

3.1 Introduction

Ce chapitre est consacré à la présentation des travaux et réflexions menés autour de la méthode d'approximation que nous proposons dans le cadre d'une modélisation adaptative de réponses déterministes non-régulières. Ainsi, ce chapitre a pour objet la description de la phase centrale schématisée en Figure 3.1.

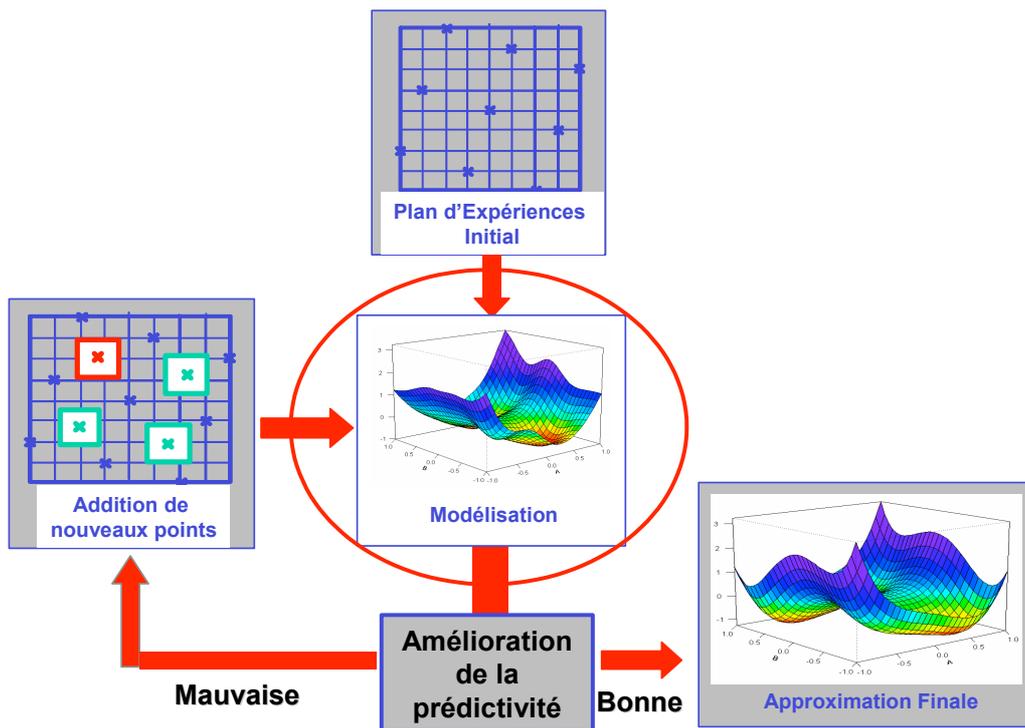


Figure 3.1 : Principe de la méthodologie adaptative

La modélisation de réponses non-régulières est tout à fait possible en faisant appel aux outils de la statistique inférentielle, comme par exemple la méthode des moindres carrés non linéaires, mais sa mise en œuvre pratique est généralement délicate [3]. Il faut en effet disposer de la forme et de la structure *a priori* du modèle que l'on souhaite ajuster. Or, dans le

cadre d'une quantification des incertitudes, nous ne disposons généralement que de très peu d'information *a priori* sur le type de non-régularité, nous n'avons aucune idée de la forme de la fonction de régression. Dans ce cas, les algorithmes d'optimisation pour l'obtention des termes et des coefficients du modèle à ajuster sont très coûteux en temps de calcul et demandent un nombre élevé de simulations. Il paraît donc nécessaire d'envisager une autre approche, plus empirique, celle que nous proposons est basée sur un principe d'apprentissage.

Dans ce chapitre, nous précisons le type d'approximation du modèle numérique d'écoulement que nous préconisons dans le cadre d'une méthodologie de modélisation adaptative de fonctions non-régulières. Ce modèle approché, appelé φ^* , est souvent désigné sous le terme de proxy dans la bibliographie anglo-saxonne. Il s'agit soit d'une fonction analytique des paramètres, soit d'un tableau d'interpolation numérique. Ce modèle approché doit permettre une estimation de la réponse Y en un quelconque point \mathbf{x} du domaine expérimental D où l'on ne dispose pas de mesures. Il est obtenu à partir des réponses issues du plan d'expériences, c'est-à-dire à partir des \mathbf{x}_i et des $Y(\mathbf{x}_i)$ associés. En modifiant de façon continue la position du point \mathbf{x}_i , nous pouvons ainsi estimer la fonction sur l'ensemble du domaine incertain et nous avons : $Y = \varphi^*(\mathbf{x})$.

Nous proposons ici une technique adaptative, basée sur l'utilisation de plans d'expériences évolutifs qui vont permettre d'une part de détecter, pas à pas, les non-régularités de la fonction réponse et d'autre part, de contrôler et maîtriser le coût de simulation. L'emploi des plans d'expériences est ici associé à une technique de modélisation interpolatrice, proposée par Sacks et al. [54] et [55] en 1989, pour tenir compte du fait que les expérimentations sont déterministes.

3.2 Prédiction des expériences déterministes : position du problème

Nous commençons tout d'abord par expliciter les notations employées dans la suite de ce document. Soient :

- k le nombre de paramètres incertains et n le nombre de points d'expérimentation ou simulations,

- $X_j, j=1, \dots, k$, les paramètres incertains,

- \mathbf{x}_i les n points pour lesquels des simulations ont été réalisées, généralement, selon un plan d'expériences :

$$\mathbf{x}_i = (x_{i1}, \dots, x_{ik}), i = 1, \dots, n$$

– $Y(\mathbf{x}_i)$ les réponses observées aux points \mathbf{x}_i : $Y(\mathbf{x}_i)$ est donnée par le simulateur d'écoulement. Chacune des simulations étant très coûteuse en temps de calcul, le nombre total n d'observations doit rester limité, typiquement à quelques dizaines.

La quantification d'incertitude sur les réponses en production nécessite au préalable la définition des différents paramètres à prendre en compte pour la modélisation ainsi que leurs intervalles de valeurs, afin de déterminer le domaine expérimental qui est donc :

$$D = \prod_{j=1}^k [X_j^{\min}, X_j^{\max}]$$

où :

- X_j^{\min} représente la valeur minimale que peut prendre le paramètre X_j ,
- X_j^{\max} la valeur maximale que peut prendre le paramètre X_j .

Rappelons tout d'abord que lors d'études de réservoir, les phénomènes physiques en jeu sont observés par le biais d'expériences simulées via un modèle numérique d'écoulement en milieu poreux. Il faut ainsi noter que, contrairement au cas d'expériences réelles, nous sommes en présence d'expérimentations déterministes. Ainsi, si un même essai est répété plusieurs fois, le simulateur donnera exactement la même valeur de réponse : il n'y a aucune erreur liée aux conditions expérimentales. De plus, pour toute valeur fixée des paramètres \mathbf{x} d'entrée, la réponse Y est reproductible puisqu'elle est entièrement déterminée par les paramètres de simulation. Il faut toutefois rappeler que la réponse Y n'est pas exacte

Par conséquent, les méthodes utilisées pour les plans classiques, tels que les plans en bloc ou la randomisation, basées sur l'interprétation d'expériences répétées, ne sont pas pertinentes. Il est ainsi important de noter que la méthodologie des plans d'expériences simulés diffère des méthodes statistiques utilisées depuis plusieurs décennies pour les expériences réelles.

Dans la méthodologie classique d'ajustement de surface de réponses, nous avons précisé au chapitre précédent que la modélisation de la réponse est généralement donnée par :

$$Y(\mathbf{x}) = f(\mathbf{x}) + \varepsilon \quad (3.1)$$

où ε est l'erreur liée à l'expérimentation.

Dans ces approches classiques, la fonction inconnue est approximée par une fonction paramétrique dont la forme est connue par avance, soit implicitement ou explicitement, par exemple polynomiale ou de courbure minimale. Les paramètres de la fonction paramétrique sont alors déterminés de manière à optimiser un critère donné, afin d'ajuster les observations de la meilleure façon possible. Une fois que la fonction d'approximation est déterminée, il est aisé et peu coûteux de l'évaluer en tout point du domaine de définition.

Dans le cas de simulations déterministes on ne considère pas d'erreur de mesure et de ce fait, on souhaite que la surface de réponse ajuste parfaitement les simulations numériques. Mathématiquement, ce problème peut être vu comme un problème d'interpolation. C'est pourquoi, dans la bibliographie pour la modélisation d'expériences simulées, des techniques d'interpolation sont souvent préconisées. Ainsi, on considère le modèle (3.1) où ε est le résidu de l'approximation introduite à l'aide de f . La méthode proposée par Sacks et al. consiste à considérer que ε est la réalisation d'un processus spatial, ce qui revient à faire appel aux techniques statistiques pour procéder à cette interpolation.

Ainsi, nous proposons ici d'utiliser des techniques d'analyse de processus spatiaux observables, donc aléatoires, issues essentiellement de la géostatistique, mais pour traiter des données de simulations non aléatoires. On introduit en fait ces processus pour prendre en compte la méconnaissance des résidus d'une approximation fonctionnelle, par exemple par une fonction polynomiale de degré deux, d'une fonction simulée certaine. Bien qu'aucune des données observées ne soit aléatoire, il est fait usage des techniques statistiques et du vocabulaire de cette discipline.

3.3 Expression du modèle

3.3.1 Modèle composite

Nous avons vu, dans le chapitre deux, qu'une interpolation à l'aide d'une fonction polynomiale de degré un ou deux n'est pas suffisante dans une majorité de cas. L'idée ici est alors d'ajouter un terme correctif à l'ajustement polynomial de la réponse moyenne, ce terme étant aléatoire. En fait, il s'agit de modéliser la réponse déterministe $Y(\mathbf{x})$ comme la combinaison d'un modèle de régression (linéaire ou quadratique) et d'une réalisation d'un processus stochastique Z ayant une structure de corrélation entre les observations. Ce concept est largement utilisé dans le contexte des expériences simulées [54] et [55], il est illustré sur la Figure 3.2.

La réponse $Y(\mathbf{x})$ est alors décomposée en 2 parties :

$$Y(\mathbf{x}) = \mathbf{X}(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}) \quad (3.2)$$

où :

- $E[Y(\mathbf{x})] = \mathbf{X}(\mathbf{x})\boldsymbol{\beta}$ est l'ajustement de la moyenne qui est la partie déterministe du modèle (3.2), soit ici une fonction polynomiale. $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ est le vecteur des coefficients inconnus de la régression.

- $Z(\mathbf{x})$ est une réalisation d'un processus stochastique gaussien de moyenne nulle et de fonction de covariance entre $Z(\mathbf{t})$ et $Z(\mathbf{u})$: $C(\mathbf{t}, \mathbf{u}) = \text{Cov}(Z(\mathbf{t}), Z(\mathbf{u})) = \sigma^2 R(\mathbf{t}, \mathbf{u})$ où σ^2 est la variance du processus et $R(\mathbf{t}, \mathbf{u})$ est la fonction de corrélation entre $Z(\mathbf{t})$ et $Z(\mathbf{u})$. N'ayant aucune information sur la structure de ce résidu, nous le supposons classiquement gaussien.

Cet estimateur est un excellent candidat en ingénierie de réservoir parce qu'il est sans biais (du fait de l'ajustement polynomial) et que l'ajout d'un terme de correction résiduel assure une interpolation des données (cf. Figure 3.2).

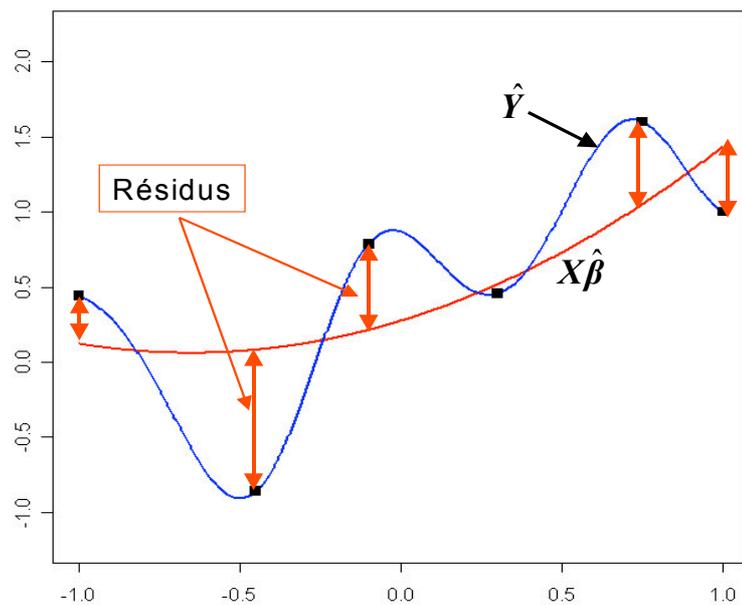


Figure 3.2 : Ajustement de la réponse par superposition de l'ajustement moyen et du terme correctif

Nous parlerons dans la suite de cette thèse de modèle composite pour désigner un modèle du type (3.2).

L'utilisation du prédicteur composite de la forme (3.2) nécessite l'estimation des deux termes du modèle : tendance moyenne et partie résiduelle. Ceci est l'objet de la suite de cette partie consacrée à l'approximation.

Nous commençons par présenter les techniques d'estimation de la tendance moyenne, que nous choisissons de modéliser par régression par moindres carrés ordinaires. Notons que la construction d'un prédicteur de la forme (3.2) peut être adapté à n'importe quelle technique d'approximation pour l'ajustement de la moyenne, elle ne se limite pas seulement au choix d'un modèle de régression polynomial. Il est tout à fait possible de considérer des polynômes de degré supérieur à deux, un réseau de neurones ou une modélisation par machines à support vectoriel. Le choix de ce modèle va dépendre d'une part

du nombre de simulations maximum envisageable par l'utilisateur et d'autre part du plan d'expériences initial utilisé.

Nous présentons ensuite la méthode d'estimation du processus résiduel, celui-ci est déterminé par la méthode du krigeage. En effet, l'estimation de la moyenne de la réponse par un modèle polynomial est sans biais et par conséquent les résidus sont de moyenne nulle. Ceci permet de déterminer, par krigeage simple, un modèle pour les résidus. Si l'estimateur de la moyenne n'est plus sans biais, il faut utiliser le krigeage ordinaire pour interpoler les résidus. Il est toutefois possible d'utiliser d'autres méthodes d'interpolation des résidus, en particulier la méthode des splines. Une fois que les résidus sont interpolés, le modèle est reconstruit en superposant les résultats de l'interpolation résiduelle au modèle polynomial de la moyenne.

Enfin, l'utilisation du krigeage nécessite une bonne estimation de la structure de covariance des résidus, nous présentons également comment estimer cette fonction de covariance.

3.3.2 Ajustement de la Tendence Moyenne : Régression par moindres carrés

Les résultats présentés sur la modélisation de la tendance moyenne par régression sont classiques [37], mais il nous a semblé utile de les rappeler brièvement. Nous décrivons donc l'ajustement d'un modèle de régression par moindres carrés ordinaires.

- **Régression par Moindres Carrés Ordinaires**

On considère le modèle :

$$Y = X\beta + \varepsilon \quad (3.3)$$

avec :

1. X la matrice de régression du modèle,
2. β le vecteur des paramètres du modèle,
3. ε le vecteur des erreurs tels que : $E(\varepsilon) = 0$ et $Cov(\varepsilon) = \sigma^2 I$.

L'objectif est de déterminer les coefficients de la régression, c'est-à-dire le vecteur β à partir des observations de la réponse Y . Pour ce faire, nous estimons l'espérance mathématique $E(Y)$, parce que celle-ci dépend du vecteur β : $E(Y) = X\beta$ avec le vecteur β inconnu à estimer. Au vu de cette relation, l'espérance $E(Y)$ appartient à l'image de X : $Im(X)$; $Im(X)$ désignant le sous-espace vectoriel engendré par les colonnes de X .

Par définition, un estimateur $\hat{\beta}$ est un estimateur des moindres carrés de β si $X\hat{\beta}$ est le vecteur de $Im(X)$ qui est le plus proche de Y pour la métrique euclidienne usuelle. En d'autres

termes, $\hat{\beta}$ est un estimateur des moindres carrés ordinaires de β s'il minimise la fonction distance :

$$L = \sum_{i=1}^n \varepsilon_i^2 = {}^t \varepsilon \varepsilon = {}^t (Y - X\beta)(Y - X\beta)$$

On peut facilement démontrer que le vecteur de $\text{Im}(X)$ qui est le plus proche de Y est le projeté perpendiculaire de Y sur $\text{Im}(X)$ [8]. Ceci est illustré sur la Figure 3.3.

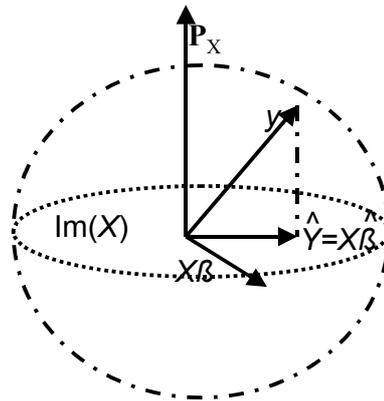


Figure 3.3 : Interprétation géométrique de MCO

En supposant que les colonnes de X soient indépendantes, soit par exemple pour un plan factoriel classiquement utilisé avec les méthodes de régression, nous savons que le projecteur P_X sur l'espace $\text{Im}(X)$ a pour expression $P_X = X({}^t X X)^{-1} {}^t X$ et donc l'estimateur de Y obtenu par moindres carrés ordinaires est, par définition :

$$\hat{Y} = X({}^t X X)^{-1} {}^t X Y$$

De plus, nous cherchons $\hat{\beta}$ tel que $\hat{Y} = X\hat{\beta}$ donc nous pouvons facilement déduire de l'équation (3.3) la forme de l'estimateur de β par moindres carrés ordinaires :

$$\hat{\beta} = ({}^t X X)^{-1} {}^t X Y \quad (3.4)$$

Le vecteur $\hat{\beta}$ est également appelé vecteur des coefficients de régression.

L'application de cette méthode suppose que les erreurs sont non-corrélées et de variance uniforme sur tout le domaine : elles sont homoscédastiques. Ces conditions peuvent, dans certains cas, être trop contraignantes, une alternative consiste à employer un estimateur de la moyenne par moindres carrés pondérés (MCP). Pour justifier l'emploi de moindres carrés pondérés, on fait appel en statistique au théorème de Gauss-Markov en tenant ainsi compte des corrélations entre les résidus. Il est de la forme : $\hat{\beta} = ({}^t X V^{-1} X)^{-1} {}^t X V^{-1} Y$, avec V la matrice de corrélation entre les observations. Mais ce théorème ne s'applique pas ici stricto sensu puisque les observations sont obtenues par un simulateur déterministe.

- **Discussion MCO/MCP**

La principale différence entre les ajustements par moindres carrés ordinaires et pondérés est que la deuxième méthode tient compte des corrélations entre les erreurs, alors que la première suppose les erreurs non-corrélées et de même variance (hypothèse d'homoscédasticité), ce qui est rarement le cas. Comme nous pouvons le voir sur la Figure 3.4, les deux méthodes peuvent donner des ajustements différents, surtout dans le cas de grandes corrélations. Toutefois, nous verrons dans le cadre d'une modélisation adaptative avec enrichissement itératif du dispositif expérimental que le choix de la méthode à employer pour ajuster la moyenne a peu d'influence sur la modélisation finale, la correction des résidus étant déterminante.

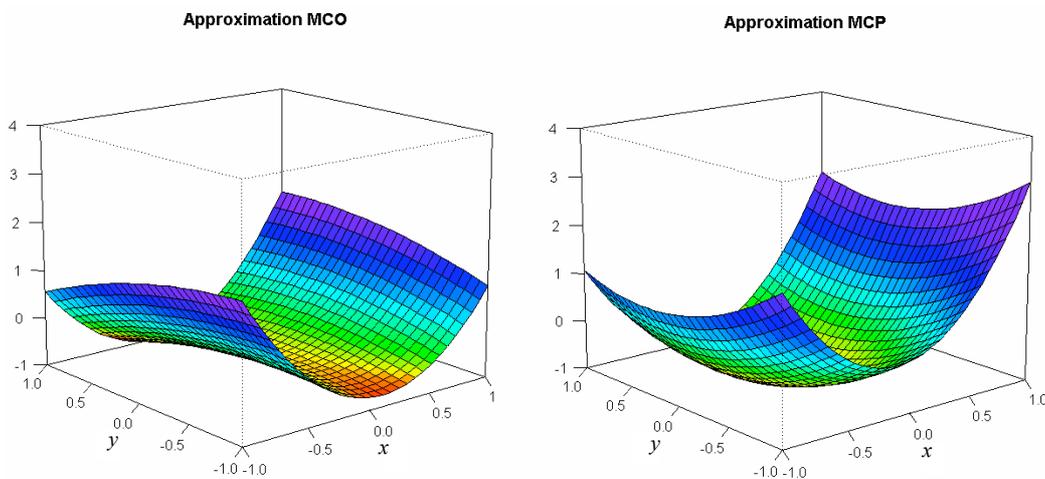


Figure 3.4: Ajustement de la moyenne par moindres carrés ordinaires et pondérés

Remarque : Nous proposons ici de considérer une régression par moindres carrés ordinaires et de faire varier le degré du polynôme d'ajustement en fonction du nombre de données. Par exemple, en début d'analyse, le nombre de points issus du plan d'expériences nous permet généralement d'ajuster un polynôme du degré deux. Dès lors que le dispositif expérimental est suffisamment riche, il semble judicieux de d'augmenter le degré du polynôme, ceci permet d'attacher moins d'importance aux résidus et donc à la fonction de covariance.

Voyons à présent comment traiter le second terme de (3.2) : le processus résiduel Z . Deux approches différentes sont considérées dans la bibliographie. Les statisticiens construisent un prédicteur linéaire sans biais de $Z(x)$, aussi appelé BLUP, alors que les géostatisticiens utilisent le krigeage. Ces deux méthodes diffèrent essentiellement par le vocabulaire et les notations. Nous présentons ici le krigeage. En effet, c'est principalement cette approche qui est exposée dans la bibliographie sur les expériences simulées. De plus, c'est cette méthode qui sera privilégiée lors de la description de la modélisation adaptative. Nous verrons également que cette méthode s'inscrit dans un cadre bayésien. La construction de l'estimateur BLUP est décrite en annexe A.3.

3.3.3 Ajustement de la partie résiduelle : estimateur par krigeage

Les résidus Z obtenus suite à l'ajustement de la tendance moyenne par régression sont calculés aux différents points du plan d'expériences. Par résidus, nous entendons la différence entre les observations et les prévisions obtenues par le modèle de régression. Il s'agit en fait des coordonnées du vecteur aléatoire $Z = (I - P_X)Y$ avec P_X le projecteur sur $\text{Im}(X)$. Les résidus Z sont interpolés afin d'aboutir à une approximation respectant les simulations réalisées. La structure du modèle composite se prête bien à la problématique de modélisation de réponses non-régulières sans information *a priori* car elle se décompose en deux parties : un modèle polynomial et un terme « correcteur » qui permet en quelque sorte de combler l'écart entre la prédiction du modèle polynomial et la simulation. C'est pour cela que nous avons choisi de ne pas effectuer du krigeage sur les observations mais sur les résidus, de manière à obtenir une tendance générale du comportement de la réponse, qui peut être suffisant dans certains cas. Dans le cas où le modèle polynomial est satisfaisant, il n'est pas nécessaire d'ajouter ce terme correcteur. Dans le cas contraire, le modèle composite permettra de corriger la surface de réponse pour ainsi prendre en compte les non-régularités détectées de la réponse. Par ailleurs, ce caractère composite de la modélisation permet une grande robustesse de l'approximation dans les zones non échantillonnées, puisque cette approximation ne résulte pas alors de la seule régression polynomiale de la réponse moyenne.

- **Généralités sur le krigeage**

L'estimation par krigeage consiste à considérer une classe d'estimateurs sans biais, généralement linéaire par rapport aux observations et de choisir celui qui minimise l'incertitude mesurée au travers de la variance de l'erreur. Par rapport à d'autres méthodes d'interpolation classiques telles que les splines ou les inverses de distance, le krigeage tient compte de la corrélation entre les données $Z(x_i)$. Il se distingue par ses caractéristiques d'estimation non biaisée et d'estimation de la variance associée. Le krigeage est une méthode géostatistique de prédiction spatiale minimisant le carré de l'erreur moyenne. C'est une méthode optimale d'estimation, au sens statistique du terme, car elle ajuste de manière parfaite les données et considère l'organisation spatiale des variables étudiées [6]. De plus, le krigeage peut rendre compte de certains types de non-stationnarités, et caractérise explicitement la dépendance spatiale des observations par la définition d'une fonction de covariance.

L'approche de krigeage est une méthode d'interpolation linéaire, basée sur une distribution normale multivariée, et qui prend en compte la covariance spatiale des données.

Le prédicteur de krigeage de la réponse, qui est en fait pour nous le résidu Z de l'ajustement d'un modèle de régression, en un point x_0 pour lequel aucune simulation n'a été

effectuée, est une fonction linéaire des n observations $Z(\mathbf{x}_i)$ aux points du plan \mathbf{x}_i [26] et [34] :

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i(\mathbf{x}_0) Z(\mathbf{x}_i) \quad (3.5)$$

Sous forme matricielle, nous avons $\hat{Z}(\mathbf{x}_0) = {}^t \boldsymbol{\lambda} \mathbf{Z}$, avec :

- ${}^t \boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$
- ${}^t \mathbf{Z} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n))$

Le krigeage consiste à attribuer un poids λ_i à chaque mesure, c'est-à-dire à déterminer les poids qui permettent d'estimer la valeur $\hat{Z}(\mathbf{x}_0)$ de la fonction en tout \mathbf{x}_0 en interpolant les données $\hat{Z}(\mathbf{x}_i)$ déjà observées. Ces poids dépendent de l'endroit où l'on se trouve dans l'espace des paramètres. Ils sont déterminés de manière à minimiser la variance d'estimation résultante, en prenant en compte l'emplacement des points de mesure.

On montre, par minimisation de la variance de l'erreur au point \mathbf{x}_0 , que les poids optimaux λ_i dépendent de la covariance entre les observations. La méthode de calcul des poids de krigeage λ_i est donnée en annexe B.2.

Nous obtenons la relation suivante :

$$\hat{Z}(\mathbf{x}_0) = {}^t \boldsymbol{\lambda} \mathbf{Z} = {}^t \mathbf{c}(\mathbf{x}_0) \mathbf{C}^{-1} \mathbf{Z} = {}^t \mathbf{Z} \mathbf{C}^{-1} \mathbf{c}(\mathbf{x}_0) \quad (3.6)$$

où :

- $\mathbf{C} = [Cov(Z(\mathbf{x}_i), Z(\mathbf{x}_j)), i, j = 1, \dots, n]$
- ${}^t \mathbf{c}(\mathbf{x}_0) = [Cov(Z(\mathbf{x}_0), Z(\mathbf{x}_i)), i = 1, \dots, n]$

Remarque : Il est important de rappeler que lors de l'utilisation du krigeage simple des résidus, on peut indépendamment utiliser la fonction de covariance ou la fonction de corrélation soit $\hat{Z}(\mathbf{x}_0) = {}^t \mathbf{r}(\mathbf{x}_0) \mathbf{R}^{-1} \mathbf{Z}$. En effet, la relation liant la fonction de corrélation $R(\cdot)$ à la fonction de covariance $Cov(\cdot)$ est donnée par :

$$Cov(Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})) = \sigma^2 \mathbf{R}(Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})) \quad (3.7)$$

Le terme σ^2 se simplifie dans le calcul des poids de krigeage, pour donner :

$${}^t \boldsymbol{\lambda}(\mathbf{x}) = {}^t Cov(\mathbf{x}) \mathbf{C}^{-1} = \sigma^2 {}^t \mathbf{R}(\mathbf{x}) \frac{\mathbf{R}^{-1}}{\sigma^2} = \mathbf{R}(\mathbf{x}) \mathbf{R}^{-1} \quad (3.8)$$

où :

- $Cov(\mathbf{x})$ et $\mathbf{R}(\mathbf{x})$ représentent respectivement la covariance et la corrélation entre le point à estimer \mathbf{x} et le plan d'expériences,

- C et R représentent respectivement les matrices de covariance et corrélation entre les points du plan d'expériences.

Enfin, nous pouvons écrire l'estimateur de krigeage comme une combinaison linéaire de n fonctions de covariances, n étant le nombre de points du plan d'expériences.

L'expression du krigeage sous sa forme duale est :

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^n b_i \text{Cov}(Z(\mathbf{x}_0), Z(\mathbf{x}_i)) , \text{ avec } \mathbf{b} = {}^t \mathbf{Z} \mathbf{C}^{-1} .$$

Le krigeage revient ainsi à ajuster localement les n données par une combinaison linéaire de n fonctions de covariance. Les b_i sont déterminés de manière à ajuster les données de manière exacte.

- **Le krigeage résiduel dans la modélisation adaptative**

Dans notre approche de modélisation, le krigeage est un estimateur linéaire et sans biais du résidu \mathbf{Z} de la production du gisement (la variable caractéristique inconnue). Les résidus \mathbf{Z} entre les simulations réalisées et les prédictions du modèle de la moyenne sont supposés constituer une réalisation d'un processus aléatoire.

L'équation (3.6) s'écrit ainsi dans notre cas, pour tout points \mathbf{x}_0 de D :

$$\begin{aligned} \hat{Z}(\mathbf{x}_0) &= {}^t \mathbf{c}(\mathbf{x}_0) \mathbf{C}^{-1} \mathbf{Z} \\ &= {}^t \mathbf{c}(\mathbf{x}_0) \mathbf{C}^{-1} [\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}] \end{aligned}$$

où $\mathbf{X} \hat{\boldsymbol{\beta}}$ représente la régression polynomiale associée à l'échantillon.

3.3.4 Modèle composite et approche bayésienne

Supposer que le processus résiduel est gaussien pour prédire le résidu $Z_0 := Z(\mathbf{x}_0)$, à partir des n autres résidus $\mathbf{Z} = (Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n))$, c'est adopter implicitement une démarche bayésienne [14], [22] et [57].

Ainsi, on suppose *a priori* que z_x est la réalisation d'un processus spatial $\mathbf{Z}(\mathbf{x})$, indicé par $\mathbf{x} \in D$, plus précisément un processus gaussien centré.

Soit alors $(x_0, \mathbf{x} = (x_1, \dots, x_n)) \in D^{n+1}$ puis, le vecteur aléatoire $(Z_0, \mathbf{Z} = (Z_1, \dots, Z_n))$ où $Z_i := Z(\mathbf{x}_i) \forall i$.

Par hypothèse, on déduit la loi du $(n+1)$ -uple (Z_0, \mathbf{Z}) , i.e. :

$$(Z_0, \mathbf{Z}) \square \text{N} \left(0, \sigma^2 \begin{pmatrix} 1 & {}^t \mathbf{r}_0 \\ \mathbf{r}_0 & \mathbf{R} \end{pmatrix} \right)$$

où :

- σ^2 est un paramètre,
- \mathbf{r}_0 (respectivement \mathbf{R}) dépend de \mathbf{x}_0 et \mathbf{x} (respectivement de \mathbf{x}) et d'un paramètre θ scalaire ou vectoriel : $\mathbf{r}_0 = \frac{\mathbf{c}(\mathbf{x}_0)}{\sigma^2}$ et $\mathbf{R} = \frac{\mathbf{C}}{\sigma^2}$.

On construit un prédicteur ponctuel par projection dans l'espace de Hilbert des variables aléatoires de carré sommable, i.e. en utilisant l'espérance conditionnelle $E[Z_0 | \mathbf{Z}]$ qui est linéaire puisque le processus est Gaussien. Plus précisément,

$$E[Z_0 | \mathbf{Z}] = {}^t \mathbf{r}_0 \mathbf{R}^{-1} \mathbf{Z} \sim \text{N} \left(0, \sigma^2 {}^t \mathbf{r}_0 \mathbf{R}^{-1} \mathbf{r}_0 \right)$$

car le processus est centré.

La loi a posteriori de Z_0 est la distribution conditionnelle :

$$[Z_0 | \mathbf{Z} = \mathbf{z}] \sim \text{N} \left({}^t \mathbf{r}_0 \mathbf{R}^{-1} \mathbf{z}, \sigma^2 {}^t \mathbf{r}_0 \mathbf{R}^{-1} \mathbf{r}_0 \right)$$

Soit encore : $\hat{Z}(\mathbf{x}_0) = {}^t \mathbf{r}(\mathbf{x}_0) \mathbf{R}^{-1} [\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}] = {}^t \mathbf{c}(\mathbf{x}_0) \mathbf{C}^{-1} [\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}]$ et on retrouve bien l'expression du modèle de krigeage (3.6).

Remarque : *Un moyen naturel d'éliminer σ^2 du processus a priori gaussien est de lui attribuer une distribution a priori non-informative, comme par exemple un a priori de Jeffreys. Toutefois, Currin et al. [14] suggèrent d'estimer σ^2 par maximum de vraisemblance parce que cet estimateur est plus précis que ceux obtenus par différentes méthodes de validation croisée.*

Ainsi, la méthode de krigeage est équivalente à l'interpolation bayésienne, avec un *a priori* gaussien sur le résidu. L'interpolation bayésienne a été introduite pour la modélisation d'expériences simulées par Currin et al. [14] et Morris et al. [38]. Ces auteurs ont donné une formulation bayésienne pour les modèles de krigeage.

3.3.5 Modèle composite : synthèse

Ainsi, nous proposons dans l'approche de modélisation adaptative d'estimer la réponse par un prédicteur, dit composite, de la forme :

$$\hat{Y}(\mathbf{x}_0) = \mathbf{X}(\mathbf{x}_0) \hat{\boldsymbol{\beta}} + {}^t \mathbf{c}(\mathbf{x}_0) \mathbf{C}^{-1} [\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}] \quad (3.9)$$

Le principal intérêt de ce modèle est que la surface de réponse ajustée passe par les simulations, l'interpolation est exacte. C'est aussi une méthode de prédiction optimale, au sens statistique du terme, car les poids de krigeage sont déterminés de manière à minimiser la variance de prédiction.

Remarque : Ce modèle est désigné sous le terme de krigeage universel par les géostatisticiens. Dans la littérature anglophone, la terminologie employée est *KT (kriging with a trend model)*.

Nous illustrons, sur la Figure 3.5, l'intérêt d'ajouter ce terme correctif. Nous considérons ici une fonction analytique non-régulière et localement non stationnaire, aussi appelée fonction « Chameau », dont l'équation est :

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

La Figure 3.5 représente la fonction référence (a), l'ajustement de la moyenne par moindres carrés ordinaires (b), puis, l'ajustement par le modèle composite (c), la correction résiduelle par krigeage menant à une évolution de l'ajustement ponctuellement, sans fortement perturber la tendance moyenne. Ces deux ajustements sont effectués à partir du même hypercube latin de distance maximin à neuf simulations.

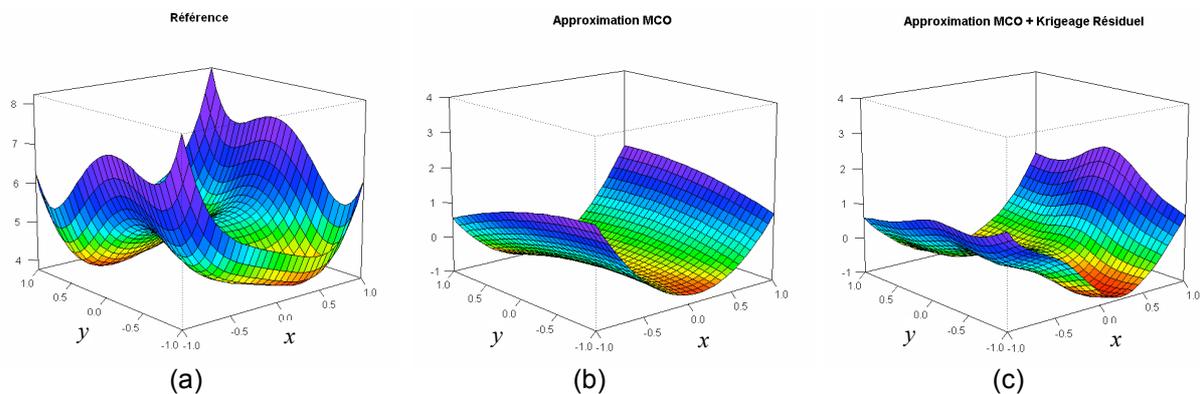


Figure 3.5: Ajustement par régression puis par prédicteur composite

Nous voyons, sur la Figure 3.5, que l'ajout d'un terme correctif à la moyenne par le biais du krigeage simple des résidus améliore la qualité d'estimation et de prédictivité du modèle, la surface de réponse estimée passe par les observations.

La qualité de prédiction d'un estimateur de type (3.9) dépend fortement de la structure de covariance entre les résidus. La covariance caractérise la variabilité spatiale des résidus. L'estimation de la structure de covariance est donc une étape très importante dans la construction du prédicteur, elle est cruciale dans notre approche de modélisation adaptative.

3.4 Problème lié à la modélisation : Estimation de la fonction de covariance entre les observations

L'estimateur de la réponse Y , donné par (3.9), dépend de la fonction de covariance du processus résiduel gaussien $Z = Y - X\hat{\beta}$, et par conséquent les résultats de la modélisation sont très sensibles au choix de la structure de covariance. De plus, le processus Z est entièrement défini par sa fonction de covariance, parce qu'il est gaussien et de moyenne connue nulle. L'estimation de cette fonction de covariance est donc une étape cruciale de la modélisation adaptative.

3.4.1 Définitions et propriétés du variogramme

En géostatistique, la fonction la plus utilisée pour décrire la régularité d'un processus est le variogramme, et ce surtout parce qu'elle est plus simple à estimer que la covariance qui demande l'estimation préalable de l'espérance mathématique, mais également parce qu'elle s'accommode des situations où $Var(Z(\mathbf{x}))$ n'est pas définie. L'estimation de ce variogramme est une des principales difficultés rencontrées au cours de ces travaux parce que nous n'avons aucune connaissance *a priori* sur le type de non-régularité de la réponse à estimer et sur la stationnarité des résidus à kriger.

- **Propriétés du variogramme**

On ne peut résoudre un problème de krigeage, c'est-à-dire calculer effectivement le poids optimal qu'il convient d'attribuer à chaque mesure, qu'à la condition de faire certaines hypothèses sur les caractéristiques du processus à modéliser. Essentiellement, il faut se donner un processus résiduel Z stationnaire d'ordre deux, autrement dit faire l'hypothèse que tous les points séparés d'une même distance \mathbf{h} , dans une direction donnée, ont une même covariance :

$$Cov(Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})) = Cov(\mathbf{h}), \quad \forall \mathbf{x} \in D.$$

Quant à la variance de ce processus stationnaire, puisqu'elle est constante, vérifie :

$$Var(Z(\mathbf{x})) = Cov(Z(\mathbf{x}), Z(\mathbf{x})) = Cov(0) = \sigma^2 \quad (3.10)$$

Une autre propriété indispensable est que la covariance soit définie positive. Cette contrainte assure que la variance de l'estimateur par krigeage soit positive ou nulle. Nous supposons que ces deux propriétés sont vérifiées dans la suite de ce manuscrit. Nous revenons en détail sur ces propriétés en annexe B.4.1.

- **Variogramme théorique**

Le variogramme représente la demi-différence de deux valeurs d'une variable aléatoire séparées de la distance h . L'expression du variogramme théorique est donnée par :

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{Var}[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})] = \frac{1}{2} E[(Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h}))^2],$$

car le processus Z est stationnaire, d'où $E[(Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h}))] = 0$.

Cette fonction, habituellement croissante en fonction de h , apporte beaucoup d'information sur le comportement conjoint des variables aléatoires.

Il existe un certain nombre de variogrammes théoriques qui sont souvent utilisés en géostatistique, et qui vérifient les conditions de stationnarité et de positivité requises. Une liste non exhaustive de différents variogrammes possibles est présentée en annexe B.4.3. Il s'agit des variogrammes les plus couramment employés en ingénierie de réservoir, à savoir les variogrammes exponentiel, gaussien, sphérique et puissance.

Lorsque le variogramme montre un palier, c'est-à-dire si $\lim_{h \rightarrow \infty} \gamma(h) = c$ où c est une constante, il existe une relation entre la valeur du variogramme pour la distance h et la covariance pour deux observations séparées de la distance h . Cette relation est très importante, elle est continuellement utilisée en géostatistique :

$$\begin{aligned} \gamma(\mathbf{h}) &= \frac{1}{2} \text{Var}[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})] \\ &= \frac{1}{2} \{ \text{Var}[Z(\mathbf{x})] + \text{Var}[Z(\mathbf{x} + \mathbf{h})] - 2\text{Cov}(Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})) \} \\ &= \sigma^2 - \text{Cov}(Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})) \end{aligned}$$

Soit enfin, du fait de la stationnarité du processus Z :

$$\gamma(\mathbf{h}) = \sigma^2 - \text{Cov}(\mathbf{h}) \quad (3.11)$$

Il est ainsi possible d'estimer le variogramme, la covariance ou la corrélation pour l'utilisation du prédicteur composite (3.7).

- **Variogramme expérimental**

L'estimation du variogramme théorique passe par la construction d'un variogramme expérimental, c'est-à-dire de points expérimentaux qui vont servir à l'ajustement du variogramme théorique.

Le variogramme expérimental, dans le cas isotrope, est défini par :

$$\gamma_e(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})]^2 \quad (3.12)$$

où $N(\mathbf{h})$ représente le nombre de paires de points séparés de la distance \mathbf{h} .

L'ensemble des points du dispositif expérimental séparés de la distance \mathbf{h} permet le calcul d'une valeur du variogramme empirique $\gamma_e(\mathbf{h})$ associée à la distance \mathbf{h} . La série de points expérimentaux $\gamma_e(\mathbf{h}_i)$ ainsi obtenue servira à la modélisation d'un variogramme théorique. Une illustration de ce principe est présentée sur la Figure 3.6.

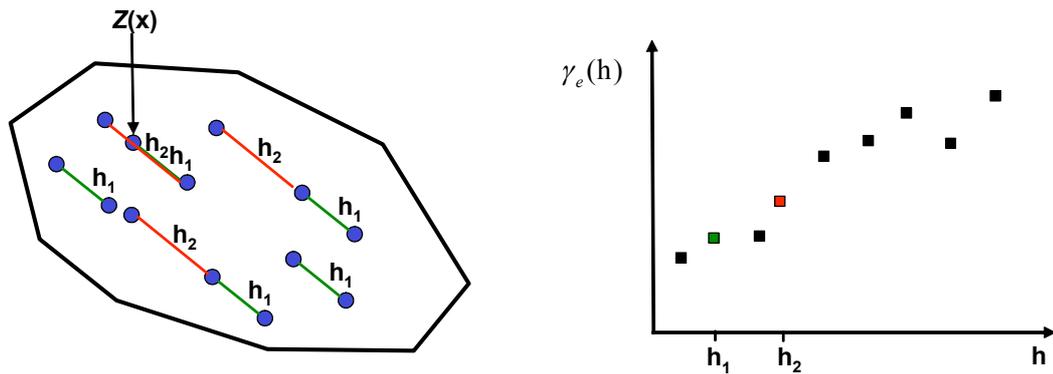


Figure 3.6: Ajustement d'un variogramme expérimental

Pour une réponse Y donnée, rien n'assure que la corrélation spatiale soit identique dans toutes les directions. Certains paramètres peuvent avoir une influence très forte sur la réponse, alors que d'autres non. Ainsi, dès lors que le nombre d'expérimentations est suffisamment grand (typiquement au moins 50, préférablement 100), on peut chercher à évaluer le variogramme expérimental dans les différentes directions.

On peut ainsi calculer le variogramme selon certaines directions :

$$\gamma_e(\mathbf{h}, \boldsymbol{\eta}) = \frac{1}{2N(\mathbf{h}, \boldsymbol{\eta})} \sum_{i=1}^{N(\mathbf{h}, \boldsymbol{\eta})} [Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})]^2$$

où $N(\mathbf{h}, \boldsymbol{\eta})$ représente le nombre de paires de points séparés de la distance \mathbf{h} dans la direction $\boldsymbol{\eta}$.

En pratique, on s'accorde un intervalle de variation autour de \mathbf{h} et $\boldsymbol{\eta}$ afin d'avoir suffisamment de paires pour chaque \mathbf{h} et chaque $\boldsymbol{\eta}$. Pour chacune des classes ainsi formées, on calcule la distance moyenne séparant les extrémités des paires (abscisse) et on évalue le variogramme expérimental pour chaque classe [17].

On obtient donc une série de points expérimentaux auxquels on cherche à ajuster un modèle, c'est-à-dire une expression analytique parmi les plus classiques (gaussien, exponentiel, etc...), permettant de déduire la covariance entre deux points quelconque en

fonction de leur espacement géographique et, éventuellement, de la direction qu'ils définissent.

Une fois le modèle de variogramme adopté, toute la suite des calculs se fait avec les valeurs obtenues du modèle et non avec les valeurs expérimentales.

Nous avons donc naturellement été amené à examiner le problème de l'ajustement d'un modèle de variogramme théorique restituant au mieux les points du variogramme expérimental. Nous présentons différentes méthodes, dépendant du nombre de données disponibles pour la construction d'un variogramme empirique.

3.4.2 Estimation du variogramme dans la modélisation adaptative

Dans le cadre de notre étude, nous disposons généralement d'un nombre insuffisant d'observations pour l'ajustement d'un variogramme. De plus, contrairement à la géostatistique classique où le variogramme permet de caractériser la corrélation spatiale entre les valeurs d'une propriété d'une roche (perméabilité, porosité, etc.), aucun sens physique ne peut être donné au variogramme. Nous n'avons donc aucun *a priori* particulier sur le type de variogramme à utiliser. Ainsi est-il important de le caractériser.

Nous considérons ici que le variogramme est défini par :

$$\gamma(\mathbf{h}) = f(d(\mathbf{x}_i, \mathbf{x}_j))$$

La construction du variogramme requiert d'une part, la définition d'une distance d adaptée au domaine expérimental, et d'autre part, la définition d'une structure de variogramme f . La distance d généralement utilisée est la distance euclidienne usuelle, c'est-à-dire qu'entre les points \mathbf{x}_i et \mathbf{x}_j , nous avons la distance :

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{l=1}^k (\mathbf{x}_{il} - \mathbf{x}_{jl})^2}$$

avec $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})$ et k le nombre de paramètres.

Mais comme toutes les variables d'entrée n'ont pas la même influence sur la réponse, il est préférable de ne pas supposer que la covariance spatiale est la même dans toutes les directions et donc d'utiliser une distance euclidienne pondérée en introduisant des paramètres θ_l , c'est-à-dire de poser :

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{l=1}^k \left(\frac{\mathbf{x}_{il} - \mathbf{x}_{jl}}{\theta_l} \right)^2}$$

De ce fait, on introduit un variogramme anisotrope. Dès lors que la distance d est déterminée, la question du choix de la structure du variogramme se pose, c'est-à-dire le choix de la fonction f introduite ci-dessus.

Nous allons traiter le choix de la structure du variogramme entre les observations Y_i plus en détails dans la suite de ce manuscrit. Nous verrons également que ceci passe par l'estimation des paramètres θ_l de cette structure. Nous préconisons deux cas, en fonction de la taille du plan d'expériences courant. Par plan d'expériences courant, nous désignons ici le plan d'expériences à une itération donnée de la modélisation adaptative.

– **Cas 1 : Peu de points dans le plan d'expériences courant**

Généralement au début de la modélisation adaptative par plans d'expériences, nous disposons d'un nombre de données beaucoup trop faible pour procéder à une estimation empirique robuste de la covariance et donc à la construction du variogramme empirique. C'est pourquoi, nous définissons préalablement la structure de la fonction f , puis nous estimons l'ensemble des paramètres θ_l .

Comme pour le modèle composite (3.9), l'estimateur de $\hat{Z}(\mathbf{x}_0)$ peut s'exprimer en fonction de la structure de corrélation qui dépend uniquement de $\boldsymbol{\theta}$ et qui est indépendant de σ^2 , nous proposons ici d'estimer la fonction de corrélation entre les résidus.

Remarque : *L'estimation de la variance σ^2 est nécessaire pour estimer la variance de krigeage de l'estimateur. Nous détaillerons ce point dans le chapitre suivant.*

Dans ce contexte, c'est-à-dire lorsque peu de points d'observation sont disponibles, nous proposons de faire une hypothèse d'assez forte régularité du modèle de covariance. Ainsi, nous imposons une structure de corrélation gaussienne au processus Z dans les premières phases d'ajustement. Cette hypothèse de régularité permet également de justifier le choix d'un polynôme du premier ou second degré pour l'approximation moyenne de la réponse.

Ainsi, nous définissons une corrélation gaussienne, anisotrope, c'est-à-dire de la forme :

$$\text{Corr}(\mathbf{h}) = \exp \left\{ -\sum_{l=1}^k \left(\frac{h_l}{\theta_l} \right)^2 \right\} \quad \text{avec } h_l = |x_{jl} - x_{kl}| \quad \forall k, j \leq n$$

Nous proposons d'évaluer les paramètres θ_l par maximum de vraisemblance restreint. La méthode du maximum de vraisemblance restreint a été introduite par Patterson et Thompson en 1971 pour réduire le biais des estimateurs du type maximum de vraisemblance. Ces deux méthodes supposent que l'hypothèse de distribution gaussienne des données est vérifiée, c'est-à-dire que :

$$[\mathbf{Y} | \boldsymbol{\theta}, \sigma^2] \sim N_n [\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{R}(\boldsymbol{\theta})]$$

De plus, la méthode du maximum de vraisemblance restreint suppose que la matrice \mathbf{X} soit de rang plein. L'estimateur par maximum de vraisemblance restreint maximise la fonction

de vraisemblance de combinaisons linéaires indépendantes de la réponse Y . Nous préférons utiliser cet estimateur parce qu'il est sans biais, contrairement à l'estimateur par maximum de vraisemblance.

Ainsi, dans le cas d'une estimation par maximum de vraisemblance restreint, une estimation de la variance du processus est donnée par [22] :

$$\tilde{\sigma}_{MVR}^2 = \tilde{\sigma}_{MVR}^2(\boldsymbol{\theta}) = \frac{1}{n-p} {}^t(Y - X\hat{\boldsymbol{\beta}})R^{-1}(Y - X\hat{\boldsymbol{\beta}}) \quad (3.13)$$

De l'expression de la variance (3.13), nous déduisons la valeur de $\tilde{\boldsymbol{\theta}}_{MVR}$, estimateur par maximum de vraisemblance restreint de $\boldsymbol{\theta}$, en minimisant la fonction :

$$\tilde{\boldsymbol{\theta}}_{MVR} = \min_{\boldsymbol{\theta}} \left\{ (n-p) \log(\tilde{\sigma}_{MVR}^2(\boldsymbol{\theta})) + \log(\det(\mathbf{R}(\boldsymbol{\theta}))) \right\} \quad (3.14)$$

avec $p = \text{rang}(X)$.

C'est donc ce modèle de corrélation qui sera utilisé lors du krigeage simple des résidus en début de modélisation, du fait du peu d'observations disponibles. Il a pour équation :

$$\text{Corr}(\mathbf{h}) = \exp \left\{ - \sum_{l=1}^k \left(\frac{h_l}{\tilde{\theta}_{MVR_l}} \right)^2 \right\} \quad \text{avec } h_l = |x_{jl} - x_{kl}| \quad \forall k, j \leq n.$$

Remarque 1 : Dans le cas d'un modèle saturé, c'est-à-dire lorsque le nombre de coefficients de régression est le même que le nombre d'expérimentations du dispositif, il n'y a pas de degrés de liberté pour estimer la variance σ^2 et l'estimateur sans biais restreint des paramètres n'est pas défini. Par conséquent, dans ce cas, nous utilisons l'estimateur du maximum de vraisemblance classique.

Les méthodes du maximum de vraisemblance et maximum de vraisemblance restreint, ainsi que d'autres méthodes possibles pour l'estimation des paramètres θ_i et σ^2 sont décrites en annexe B.4.4.

Ce choix d'estimateur, plutôt que d'un *a priori* non informatif de Jeffreys a notamment été préconisé par Currin *et al.* [14]. En effet, ils ont montré que l'estimateur par maximum de vraisemblance est plus stable pour différents tests de validation croisée effectués.

– Cas 2 : Nombre de points dans le plan d'expériences courant important

Au fur et à mesure de l'enrichissement du dispositif expérimental, il est judicieux de remettre en cause l'hypothèse de régularité gaussienne et donc de la confronter à l'estimation d'un variogramme empirique. Notons que l'hypothèse de stationnarité est conservée. Dès lors que le nombre de points d'observation est suffisant, le choix de la structure de covariance reste ouvert. Étant donné que nous n'avons pas de signification physique de la covariance, il n'est pas possible de définir une structure en fonction du type de problème. Pour ne pas être

pénalisé par le choix d'une structure *a priori*, nous proposons de déterminer une covariance qui minimise une fonction objectif : le carré de l'erreur entre le variogramme théorique et le variogramme empirique.

Pour ce faire, nous construisons le variogramme empirique à partir de l'équation (3.12). Le variogramme théorique que nous cherchons à ajuster est l'un des variogrammes classiques utilisés en ingénierie de réservoir, à savoir exponentiel, gaussien ou sphérique. Étant donné que nous effectuons du krigeage simple, nous ne considérons pas le variogramme puissance, celui-ci n'ayant pas de palier.

Nous cherchons le variogramme minimisant la somme des carrés des écarts avec le variogramme empirique, c'est-à-dire qui minimise la fonction objectif :

$$J = \sum_i \left(\gamma(\mathbf{h}^i, \theta) - \gamma_e(\mathbf{h}^i) \right)^2 \quad (3.15)$$

où :

- \mathbf{h}^i représente les différentes distances qui interviennent dans la construction du variogramme empirique,
- $\gamma_e(\mathbf{h}^i)$ représente le variogramme empirique
- $\gamma(\mathbf{h}^i, \theta)$ est le variogramme à estimer, parmi un variogramme exponentiel, gaussien ou sphérique

Nous illustrons cette approche à l'aide de la Figure 3.7 ci-dessous. Puis nous présentons les variogrammes exponentiels, gaussiens et sphériques obtenus après optimisation de (3.15) à l'aide de la Figure 3.8. Par commodité pour ces représentations graphiques, nous nous plaçons ici dans le cas isotrope.

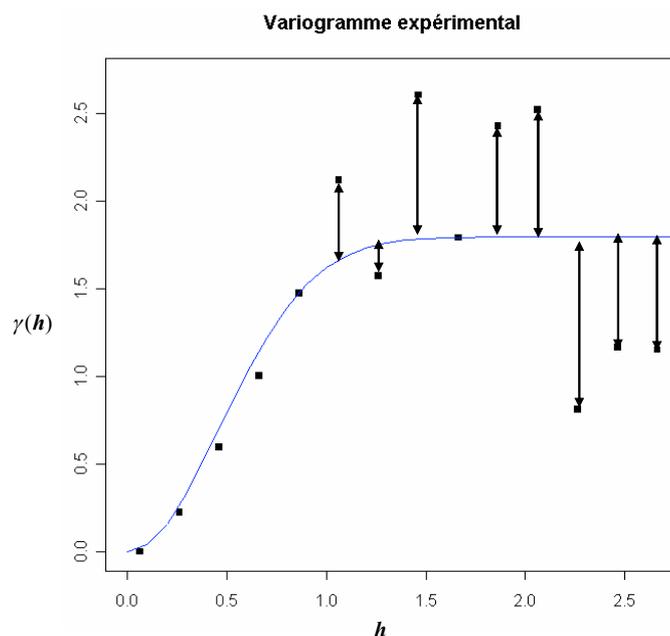


Figure 3.7 : Ajustement d'un variogramme : Méthode de moindres carrés

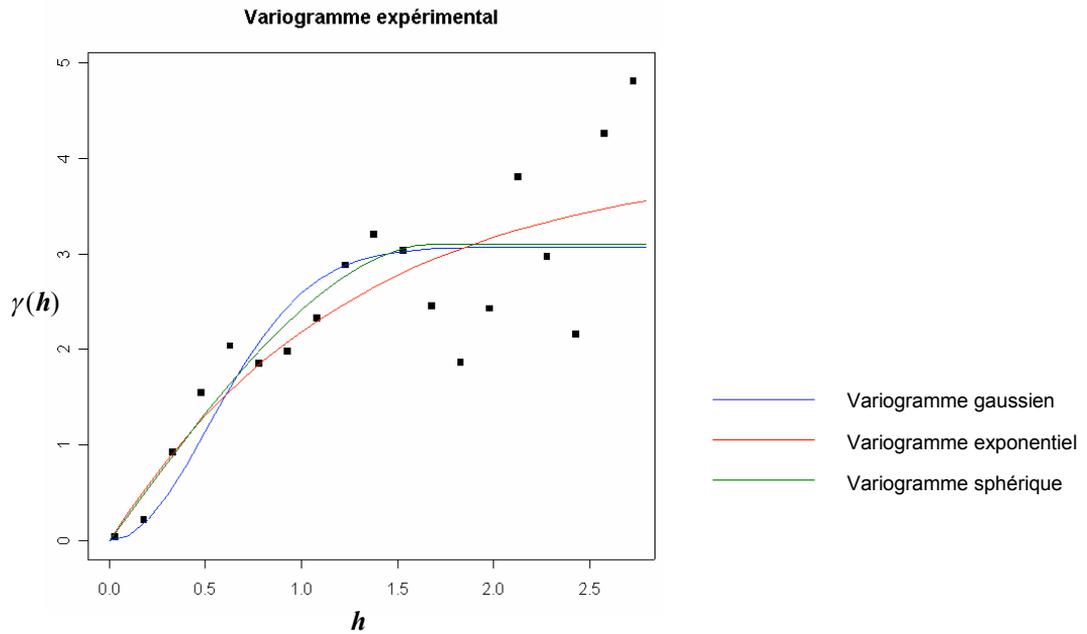


Figure 3.8 : Ajustement de variogrammes exponentiel, gaussien et sphérique

Sur la Figure 3.8, les sommes des carrés d'écarts des variogrammes gaussien, exponentiel et sphérique sont respectivement : 9.04, 7.34 et 8.59. Ainsi, le variogramme exponentiel est celui de meilleure qualité. Dans ce cas bien précis, c'est le variogramme exponentiel qui est utilisé pour la modélisation par krigeage.

Toutefois, le choix d'un seul de ces trois variogrammes n'est pas toujours optimal. Nous proposons donc une généralisation de la formulation de la fonction objectif consistant à ajuster une combinaison linéaire de ces variogrammes. Nous proposons également d'optimiser la valeur de la puissance p du variogramme de type exponentiel. Notons que pour $p = 1$, nous obtenons le variogramme exponentiel et pour $p = 2$, il est gaussien.

Ainsi, le variogramme que nous cherchons à optimiser est de la forme :

$$\gamma(\mathbf{h}, \boldsymbol{\theta}, \mathbf{c}) = \lambda_1 \gamma_{exp}(\mathbf{h}, \boldsymbol{\theta}_1, c_1, p) + \lambda_2 \gamma_{sph}(\mathbf{h}, \boldsymbol{\theta}_2, c_2)$$

avec :

$$\gamma_{exp}(\mathbf{h}, \boldsymbol{\theta}, c, p) = c \left(1 - \exp \left\{ - \left(\frac{\mathbf{h}}{\boldsymbol{\theta}} \right)^p \right\} \right) \text{ et } 0 < p \leq 2.$$

Nous procédons à l'ajustement des paramètres définissant à la fois la structure et les paramètres du variogramme de manière à réduire l'erreur moindres carrées par rapport aux points du variogramme expérimental. Pour ce faire, nous cherchons à évaluer les valeurs des paramètres $\lambda_1, \lambda_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, c$ et p de manière à minimiser la fonction :

$$(\boldsymbol{\theta}, \mathbf{c}, p) = \min \left\{ \sum_i \left(\gamma(\mathbf{h}^i, \boldsymbol{\theta}, \mathbf{c}, p) - \gamma_e(\mathbf{h}^i) \right)^2 \right\} \quad (3.16)$$

Nous illustrons l'ajustement d'un tel variogramme, toujours dans le cas isotrope, à l'aide de l'exemple de la Figure 3.9.

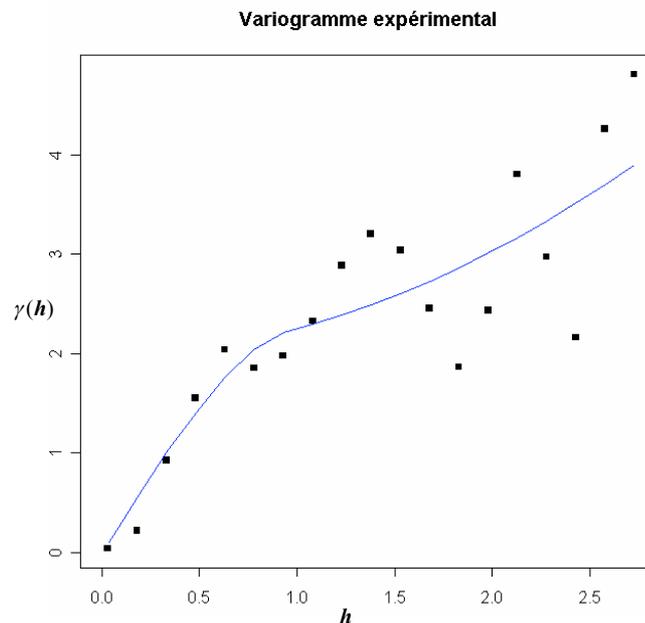


Figure 3.9 : Ajustement d'un variogramme : combinaison linéaire de variogrammes

Le variogramme obtenu ici est de la forme :

$$\gamma(\mathbf{h}, \boldsymbol{\theta}, \mathbf{c}) = 5.98\gamma_{exp}(\mathbf{h}, 10, 4.45, 2) + 1.07\gamma_{sph}(\mathbf{h}, 0.99, 1.86)$$

Le carré des erreurs aux points du variogramme empirique est de 6.15, nous constatons donc qu'il ajuste mieux les données expérimentales que les variogrammes pris individuellement.

Ainsi, nous proposons la définition d'un variogramme qui soit combinaison linéaire de plusieurs variogrammes simples. Ce type de variogramme permet la réduction de l'erreur moindres carrées par rapport aux points d'expérimentation. Dès lors que le nombre de données du plan d'expériences est suffisant, nous proposons de réaliser une telle optimisation du variogramme, s'affranchissant ainsi d'une hypothèse *a priori* sur le choix du modèle de variogramme plutôt qu'un autre.

3.5 Remarques complémentaires : approches alternatives ou complémentaires

Nous avons présenté dans ce chapitre le modèle d'approximation employé dans cette thèse. Par son caractère composite, ce modèle permet une interpolation exacte des données simulées, tout en conservant la robustesse d'une régression dans les zones peu échantillonnées. Le choix de ce modèle d'approximation est suggéré principalement par

l'examen conjoint de la bibliographie sur le sujet et des études de validation permettant la prise en compte des spécificités des phénomènes étudiés en ingénierie de réservoir. Ceci nous a amené à étudier quelques alternatives possibles, notamment deux autres méthodes géostatistiques : le cokrigage et le krigeage ordinaire avec voisinage glissants. Nous présentons brièvement ces méthodes et expliquons leurs avantages et inconvénients par rapport à la méthode d'approximation finalement préférée.

3.5.1 Cokrigage

Le cokrigage est une technique basée sur le principe du krigeage, mais tenant compte de plusieurs données simultanément [9]. Elle est utilisée en géostatistique, dès lors que l'utilisateur dispose de données provenant de différentes sources. Dans le cas de la simulation numérique d'écoulement, une information supplémentaire peut être fournie dans certains cas par le simulateur : les dérivées partielles de la réponse par rapport aux paramètres évaluées en chacun des points expérimentaux.

L'objectif ici est prédire la réponse Y au point \mathbf{x}_0 , en utilisant conjointement les valeurs de simulations et les valeurs des dérivées partielles. Comme précédemment, le cokrigage est appliqué à la modélisation de la partie résiduelle.

Dans le cas où l'on considère deux paramètres \mathbf{x}^1 et \mathbf{x}^2 , pour chaque point du plan d'expériences \mathbf{x}_i , nous disposons d'un résidu Z , et des valeurs de dérivées partielles suivant

les deux directions \mathbf{x}^1 et \mathbf{x}^2 , soit respectivement : $\frac{\partial Z}{\partial x^1}$ et $\frac{\partial Z}{\partial x^2}$.

L'expression du terme résiduel par cokrigage est alors :

$$\hat{Z}(\mathbf{x}) = \sum_{i=1}^n \lambda_{0i}(\mathbf{x})Z(\mathbf{x}_i) + \sum_{i=1}^n \lambda_{1i}(\mathbf{x})\frac{\partial Z}{\partial x^1}(\mathbf{x}_i) + \sum_{i=1}^n \lambda_{2i}(\mathbf{x})\frac{\partial Z}{\partial x^2}(\mathbf{x}_i) \quad (3.17)$$

Ainsi, en supposant que l'information sur les dérivées partielles est disponible par des algorithmes peu coûteux, la méthode de cokrigage permet d'exploiter cette information pour contraindre l'ajustement et ainsi améliorer l'approximation à moindre coût. Utiliser cette approche permet d'approximer une fonction avec un nombre de simulations moins important, sans pour autant sacrifier la prédictivité de l'approximation. Cette méthode peut être utilisée pour l'ajustement de fonctions fortement non-régulières mais cependant continues pour donner un sens aux gradient. Elle contient une information additionnelle très utile permettant de réduire significativement le nombre de simulations requises. Toutefois, cette amélioration en terme de nombre de simulations se fait au détriment du temps de calcul associé à chaque simulation du fait du calcul des dérivées partielles.

La Figure 3.10 donne un exemple en une dimension d'une fonction estimée d'une part, par krigeage simple, représenté en vert, et d'autre part, par cokrigage, en rouge. La fonction

référence est représentée en pointillés. Nous voyons que l'information apportée par les dérivées partielles est capitale ici.

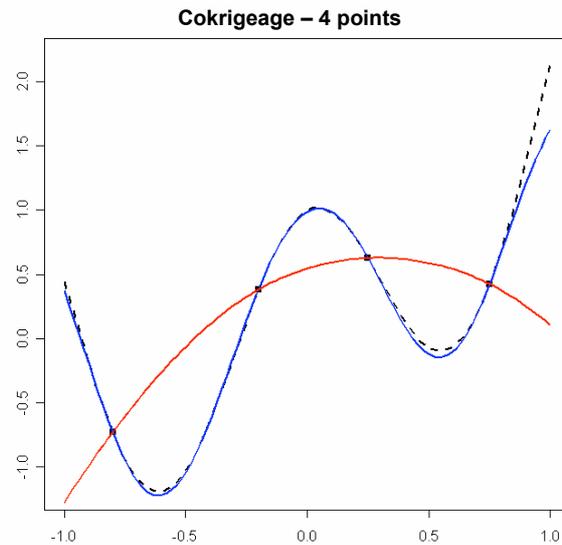


Figure 3.10 : Approximation d'une fonction analytique par krigage et cokrigeage

Supposant que l'information sur les gradients est disponible pour un coût peu élevé, cette approche améliore grandement la précision et l'efficacité de la modélisation obtenue par krigage. Le gain est d'autant plus important que le nombre de paramètres est grand [9]. En utilisant cette approche, on peut approximer une réponse au moyen d'un nombre de simulation grandement diminué, sans pour autant sacrifier la précision de l'approximation. Toutefois, cette information n'est pas toujours disponible, particulièrement pour certains paramètres. Par ailleurs, dans le cas où cette information peut être obtenue, elle peut nécessiter un coût de calcul très important, parfois équivalent à celui de la simulation. L'utilisation de cette méthode peut être envisagée dans le cas où l'on dispose de cette information à un coût raisonnable.

3.5.2 Krigage Ordinaire avec Voisinages Glissants

Nous avons également envisagé une modélisation basée sur une autre technique de krigage, qui consiste à approximer la réponse à l'aide d'un krigage ordinaire avec voisinages glissants. Le principe de construction de cet estimateur de krigage présente en effet l'intérêt de ne pas être basé sur une régression pour modéliser la tendance moyenne. L'interpolation est réalisée directement sur les observations de la réponse et non sur les résidus estimés à partir d'une régression. De plus, contrairement au krigage simple, le krigage ordinaire ne nécessite pas la connaissance de la moyenne du processus, puisqu'on impose à la somme des poids de krigage d'être égale à un.

Pour une estimation par krigeage ordinaire, nous avons :

$$y^*(x) = \sum_{l=1}^L \lambda_l(x) y^l$$

Le système stationnaire de krigeage ordinaire s'écrit :

$$\begin{cases} \sum_{l=1}^L \lambda_l(x) \text{Cov}(x_l - x_{l'}) + \mu(x) = \text{Cov}(x - x_{l'}), \dots l' = 1, \dots, L \\ \sum_{l=1}^L \lambda_l(x) = 1 \end{cases}$$

avec :

- $\lambda_l(x)$ les poids de krigeage ordinaire,
- $\mu(x)$ le paramètre de Lagrange associé à la contrainte $\sum_{l=1}^L \lambda_l(x) = 1$.

Le krigeage ordinaire est généralement utilisé avec des voisinages glissants. La technique des voisinages glissants consiste à définir un voisinage ϑ_x de chaque point x à estimer et de n'utiliser que les points se situant dans ce voisinage pour estimer la valeur du point par krigeage ordinaire. En effet, seules les données proches du point à estimer sont prises en compte lors de l'estimation. Ceci apporte une flexibilité vis-à-vis de l'hypothèse de stationnarité supposée en krigeage simple. En effet, on peut montrer qu'à chaque nouveau point x , le krigeage ordinaire estime une nouvelle moyenne $m(x)$. Comme le krigeage ordinaire est utilisé ici avec voisinages glissants, différentes données permettent d'estimer cette moyenne et par conséquent, celle-ci dépend de la position x du point. On peut montrer que l'estimateur de krigeage ordinaire est de type krigeage simple mais avec une moyenne fonction du point x au lieu d'être contrainte, on a donc :

$$y^*(x) = \sum_{x_l \in \vartheta_x} v_l(x) y(x_l) + \left[1 - \sum_{l=1}^L v_l(x) \right] m(x)$$

Ainsi, la procédure de krigeage ordinaire utilisée avec des voisinages glissants est par construction un algorithme non stationnaire, dans le sens qu'il correspond à un modèle aléatoire non stationnaire avec une moyenne variable mais une covariance stationnaire. Ceci explique la robustesse de cet algorithme.

Les voisinages nécessaires à une estimation par krigeage avec voisinages glissants peuvent être déterminés en fonction de l'échantillonnage. En effet, une non-stationnarité de la fonction φ peut se traduire, par exemple, par des zones où les variations de la réponse sont très lisses (quasi-polynomial) et d'autres zones où la réponse est très variable. Si on considère un échantillonnage optimal d'une fonction non-régulière, nous aurons des zones peu échantillonnées qui correspondent aux endroits où la fonction est lisse et inversement, nous

aurons des zones fortement échantillonnées qui correspondent à de plus grandes fluctuations de la réponse. Il peut être judicieux de construire des voisinages dont la taille dépend du nombre de données : le voisinage est d'autant plus grand que le nombre de données est faible et d'autant plus petit que le nombre de données est important. En effet, pour chaque point à estimer, une recherche des points les plus proches à l'intérieur d'un voisinage est effectuée. On se donne un nombre maximum n_{max} de points par direction (ou quadrant) et seuls les points les plus proches, dans chaque direction, sont conservés, afin d'éviter toute redondance dans l'information. Une illustration de ce principe est présentée sur la Figure 3.11. L'estimation de la valeur du point est ensuite obtenue par krigeage ordinaire, à partir des points conservés lors de la recherche de voisinage.

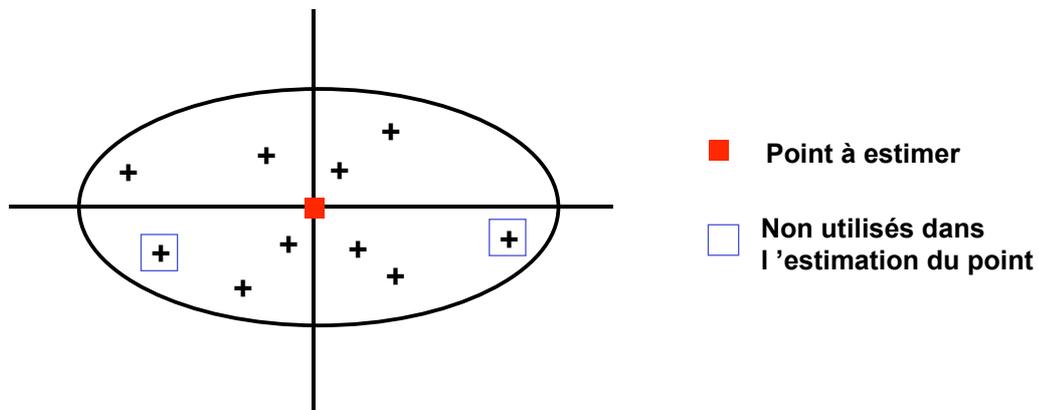


Figure 3.11 : Estimation d'un voisinage : deux points par quadrant

Une étude détaillée sur le cas réservoir synthétique PUNQ [62] a permis de montrer que l'estimation obtenue par krigeage ordinaire avec voisinages glissants donne des résultats très satisfaisants. L'avantage de cette méthode est que l'estimation se fait uniquement à partir des points voisins en imposant un nombre maximal n_{max} de données par quadrants. Ainsi, seule l'information proche est prise en compte, la redondance est évitée et l'on est plus robuste en cas de discontinuités. Cette approche s'est avérée offrir une alternative très intéressante, en particulier pour l'ajustement de fonctions à forte irrégularité.

La Figure 3.12 représente l'ajustement d'une fonction référence (a) par régression et krigeage résiduel (b), puis par krigeage ordinaire avec voisinages glissants (c). Nous voyons que la dernière méthode donne des résultats moins « lisses » que dans la modélisation composite, l'estimation étant effectuée par paliers.

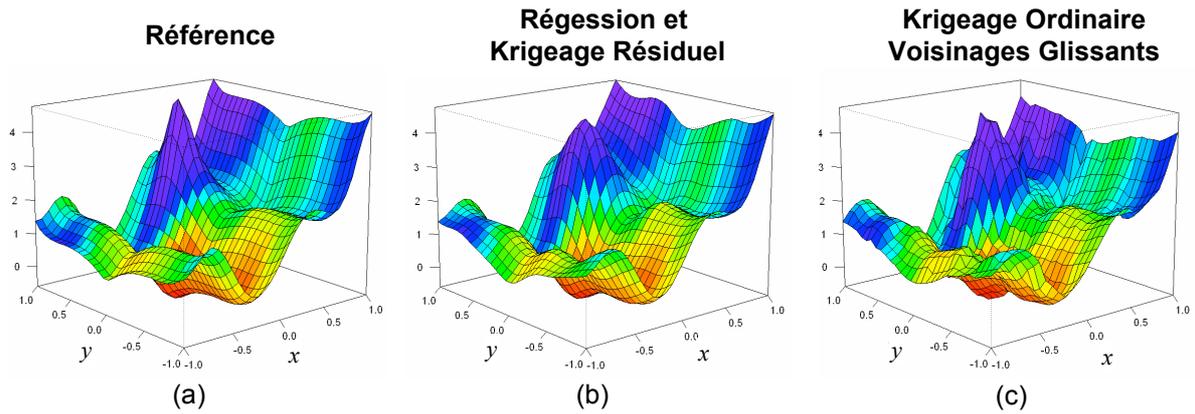


Figure 3.12 : Approximation d'une fonction analytique par régression et krigeage résiduel puis krigeage ordinaire avec voisinages glissants

Le principal inconvénient de cette méthode est qu'elle nécessite un nombre important de points d'observation. Par ailleurs, le caractère local de l'approximation implique une faible robustesse de cette modélisation vis-à-vis de la prédictivité dans les zones où l'on ne dispose pas d'information ou pour l'extrapolation.

3.5.3 Bilan

Trois méthodes d'interpolation ont donc été étudiées : un modèle composite constitué d'une régression et d'un krigeage des résidus, un modèle composite associant une régression et un cokrigeage résiduel et enfin un modèle de krigeage ordinaire avec voisinages glissants. Ces trois techniques se prêtent très bien à notre application de modélisation adaptative à faible coût, parce que la prise en compte de la covariance entre les observations permet une bonne évaluation du comportement de la réponse dans un voisinage des points observés. De plus, ces méthodes permettent d'obtenir une approximation exacte, c'est-à-dire que les valeurs des simulations sont respectées.

A l'issue de l'étude comparative de ces techniques de krigeage, il semble préférable d'utiliser la méthode de cokrigeage dès lors que l'information sur les dérivées partielles peut être obtenue à faible coût. Mais ceci n'est généralement pas le cas.

Ainsi, dans le cas de fonctions réponses assez complexes ou discontinues, nous proposons de débiter le processus itératif par des approximations globales, puis, quand le nombre de données est suffisant, d'utiliser des techniques locales, telle que le krigeage ordinaire avec voisinages glissants. Ce krigeage ordinaire peut être associé à un krigeage avec dérive (KT) aux frontières du domaine, c'est-à-dire là où le krigeage ordinaire donne de moins bons résultats du fait de l'extrapolation. La présence de la dérive permet alors une meilleure estimation là où peu de points sont disponibles, ce qui est particulièrement observé aux frontières du domaine. Nous pouvons également compenser des effets de bord par l'ajout

de points pilotes de l'autre côté des frontières du domaine. Ceux-ci vont permettre une meilleure estimation de la moyenne et donc de contourner le problème de l'extrapolation.

De plus, sans *a priori* particulier, sans hypothèse d'une régularité sur la réponse, ou dans le cas d'un nombre faible de données, l'ajustement d'une moyenne associé à un krigeage des résidus fournit les résultats les plus robustes.

Enfin, l'utilisation du krigeage ordinaire avec voisinages glissants pour l'estimation d'un point présente des similarités avec les méthodes statistiques inférentielles, particulièrement les méthodes semi-paramétriques. Ces méthodes peuvent également être des alternatives de bonne qualité, il serait intéressant d'explorer cette voie.

3.6 Conclusions

Le caractère déterministe de la réponse étudiée justifie le choix d'un interpolateur pour la modélisation de réponses non-régulières. L'approche consistant à modéliser la réponse par la somme d'une régression et d'un krigeage résiduel permet de conserver une tendance moyenne là où il n'y a pas d'observations et que la fonction de covariance est mal définie. L'estimation de la structure de covariance des résidus est essentielle dans la stratégie de modélisation adaptative envisagée ici. En effet, du fait de la prépondérance du terme résiduel dans le prédicteur, les résultats de la modélisation varient énormément en fonction de la valeur de la fonction de covariance. Il s'agit d'une des nombreuses difficultés rencontrées lors de ces travaux car nous ne disposons d'aucune information sur le type de structure de covariance. En s'inspirant des techniques de résolution de problèmes inverses, nous proposons aussi de l'estimer de manière à minimiser l'erreur moindres carrés entre le variogramme expérimental et une combinaison linéaire de variogrammes théoriques.

Les techniques d'interpolation n'ayant pas de solutions uniques, de nombreuses réalisations différentes interpolent les points de simulation. Afin de mieux décrire la classe des interpolateurs possibles, nous proposons, dans le chapitre suivant, une approche innovante d'ajout d'information fictive, visant à construire ainsi un interpolateur de meilleure prédictivité.

Chapitre 4

Recherche de l'estimateur de plus grande prédictivité – Ajout de points pilotes

4.1 Introduction

Dans ce chapitre, nous présentons une approche originale d'optimisation de la prédictivité *a priori* de l'approximation. Comme présenté sur la Figure 4.1, cette étape est réalisée après l'étape de modélisation.

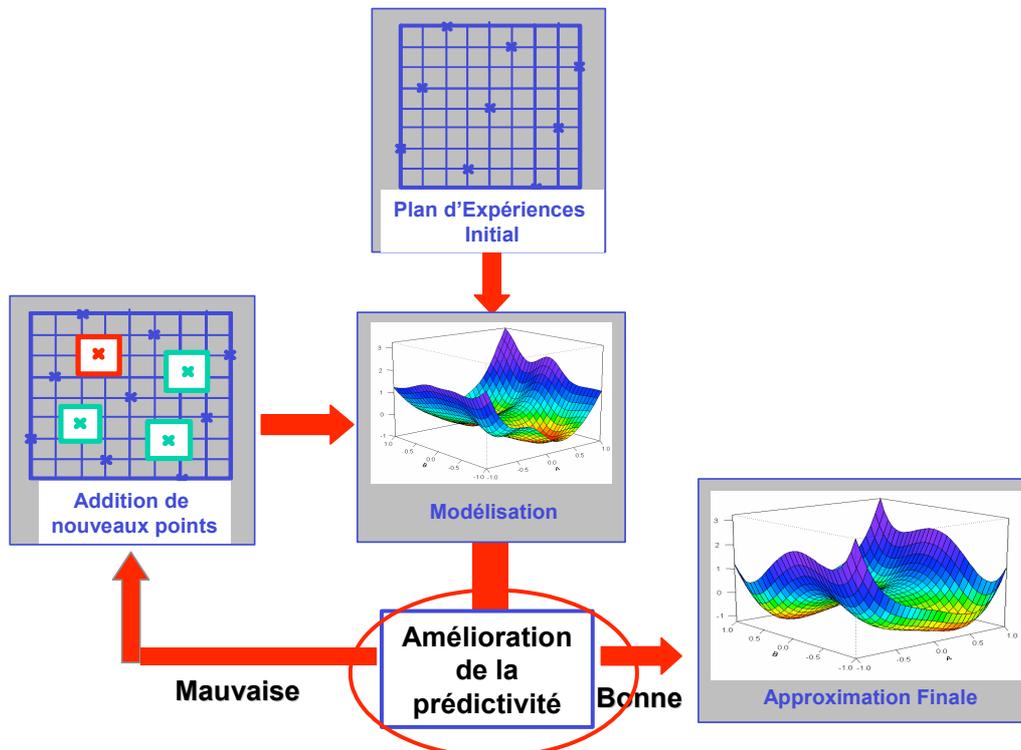


Figure 4.1 : Principe de la méthodologie adaptative

Nous rappelons que l'approximation dont nous voulons améliorer le pouvoir de prédictivité est définie par :

$$Y(\mathbf{x}) = \mu(\mathbf{x}) + Z(\mathbf{x}) \quad (4.1)$$

avec :

- μ l'approximation de la moyenne obtenue par régression à partir de n expérimentations simulées : $\mu(\mathbf{x}) = 'X(\mathbf{x})\hat{\beta}$
- $Z(\mathbf{x})$ un processus aléatoire entièrement défini par ses deux premiers moments : moyenne et variogramme.

Pour un nombre n d'expérimentations données, il existe un grand nombre d'interpolateurs exacts de la forme (4.1) qui respectent la structure spatiale du processus, au travers de la moyenne et de son variogramme. Dans cette classe, nous cherchons un estimateur qui améliore la prédictivité *a priori*. Pour y parvenir, nous introduisons dans ce chapitre une approche innovante d'ajout d'information fictive, visant donc à construire un interpolateur de meilleure prédictivité. Plus précisément, nous procédons à l'ajout de points pilotes selon une terminologie introduite par De Marsily [48], la recherche de l'estimateur de plus grande prédictivité est réalisée en contraignant l'approximation au respect de données dites fictives car non simulées, mais calibrées pour assurer une meilleure robustesse. Ces points pilotes permettent de maintenir la corrélation spatiale indiquée par le variogramme.

Ces points pilotes sont ensuite considérés comme des données bien qu'aucune simulation n'ait été réalisée et permettent de décrire la classe des estimateurs de la forme (4.1) utilisée ici. L'objectif est de sélectionner l'interpolateur qui soit le meilleur au sens de la prédictivité *a priori* du modèle.

4.2 Méthode des Points Pilotes - Principe

Afin de déterminer l'estimateur de prédictivité maximale, nous proposons une mise à jour, à chaque itération, de l'estimateur de la réponse Y . La méthode que nous utilisons, consiste à ajouter aux n points de simulation (points de données dures), n_p points fictifs appelés points pilotes. Nous proposons d'estimer les valeurs inconnues de la réponse aux n_p points pilotes afin de maximiser un critère de prédictivité de l'approximation. Ainsi, chaque interpolateur $\hat{Z}(\mathbf{x})$ est obtenu par krigeage des résidus en prenant en compte les données suivantes :

- n points de données dures pour lesquels les valeurs de réponse $\mathbf{Y} = (Y_1, \dots, Y_n)$ sont obtenues par simulation via le simulateur d'écoulement.

- n_p points fictifs $\mathbf{x}^{pp} = (\mathbf{x}_1^{pp}, \dots, \mathbf{x}_{n_p}^{pp})$, appelés points pilotes, pour lesquels les valeurs $\mathbf{Y}^{pp} = (Y_1^{pp}, \dots, Y_{n_p}^{pp})$ de la réponse sont inconnues. Les couples $(\mathbf{x}_i^{pp}, Y_i^{pp})$, position-réponse, sont déterminés au vu de l'ajustement obtenu à l'étape précédente de façon à améliorer l'interpolation. Les résidus associés à la moyenne $\mu^{pp}(\mathbf{x}^{pp})$ sont désignés par $\mathbf{Z}^{pp} = (Z_1^{pp}, \dots, Z_{n_p}^{pp})$, avec $\mu^{pp}(\mathbf{x})$ l'estimation de la moyenne des données simulées et fictives par moindres carrés ordinaires.

Soit \hat{Y} l'estimateur initial obtenu à partir des n données simulées. L'estimateur en un point \mathbf{x} du domaine expérimental par krigeage est :

$$\hat{Y}(\mathbf{x}) = \mu(\mathbf{x}) + \sum_{i=1}^n \lambda_i(\mathbf{x})Z_i$$

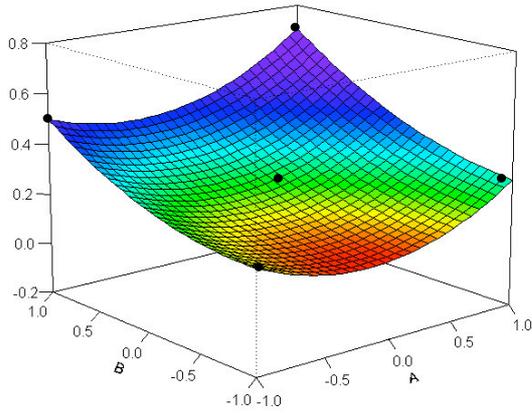
Le conditionnement de la réalisation initiale \hat{Y} par les n_p valeurs estimées et les n valeurs de simulation nous fournit une nouvelle réalisation \hat{Y}_{pp} de la réponse passant par les observations Y_i :

$$\hat{Y}_{pp}(\mathbf{x}) = \mu^{pp}(\mathbf{x}) + \sum_{i=1}^n \lambda_i(\mathbf{x})Z_i + \sum_{j=1}^{n_p} \lambda_j^{pp}(\mathbf{x})Z_j^{pp}$$

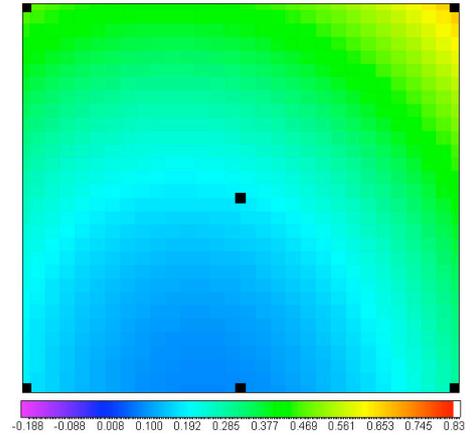
La réalisation \hat{Y}_{pp} est plus précise en terme de prédictivité que la réalisation initiale \hat{Y} . En effet, elle intègre, grâce à ce double conditionnement, d'une part l'information réelle provenant du simulateur d'écoulement et d'autre part, l'information fictive calculée de manière à accroître la prédictivité, tout en respectant la structure spatiale de la réponse à modéliser : la réalisation \hat{Y} et la réalisation \hat{Y}_{pp} sont deux réalisations différentes de la même fonction aléatoire.

Sur la Figure 4.2 et la Figure 4.3 ci-dessous, nous illustrons l'effet de l'ajout d'un point pilote \mathbf{x}_{pp} sur l'interpolation pour diverses valeurs données à Y_{pp} . La position du point pilote est précisée par un triangle, ce point vient s'ajouter aux cinq points d'un plan d'expériences pour lesquels on dispose de cinq simulations.

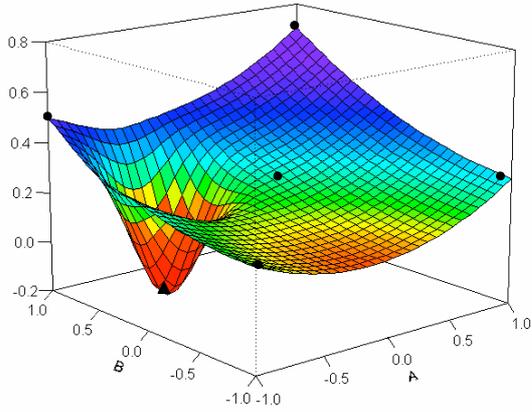
Approximation sans Points Pilotes



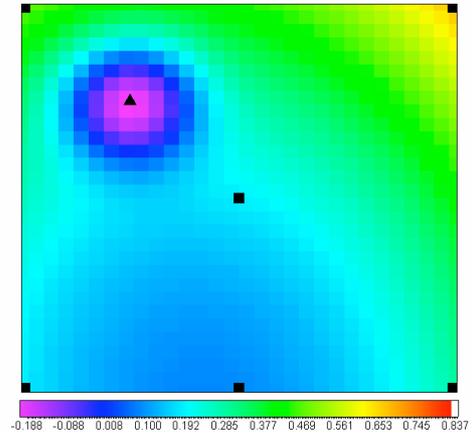
Approximation sans Points Pilotes



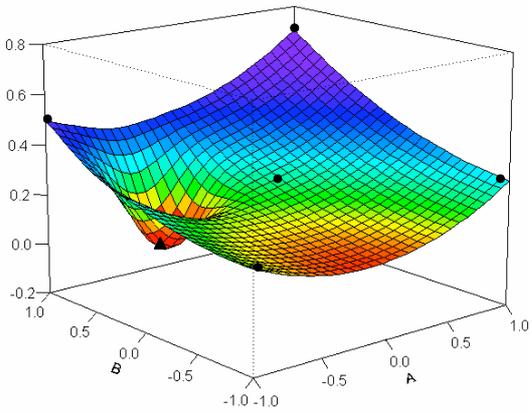
$\gamma^{PP} = -0.2$



Approximation avec Points Pilotes



$\gamma^{PP} = 0$



Approximation avec Points Pilotes

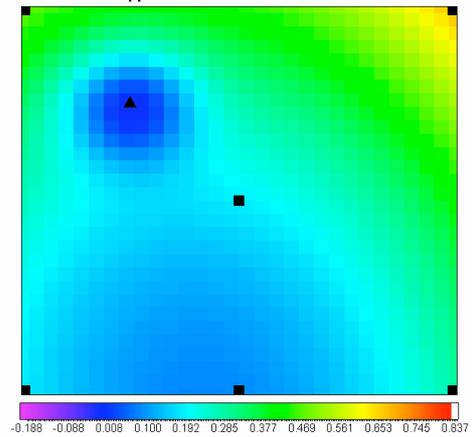


Figure 4.2 : Influence d'un point pilote sur l'approximation (1)

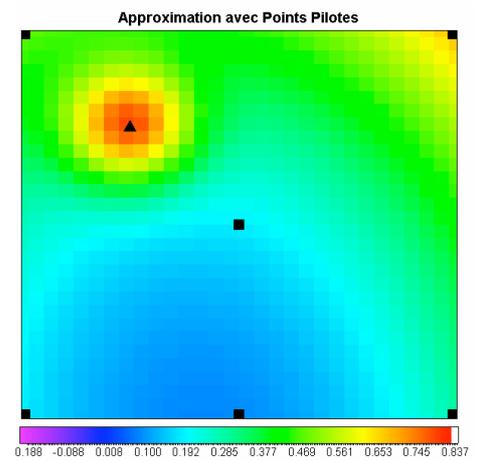
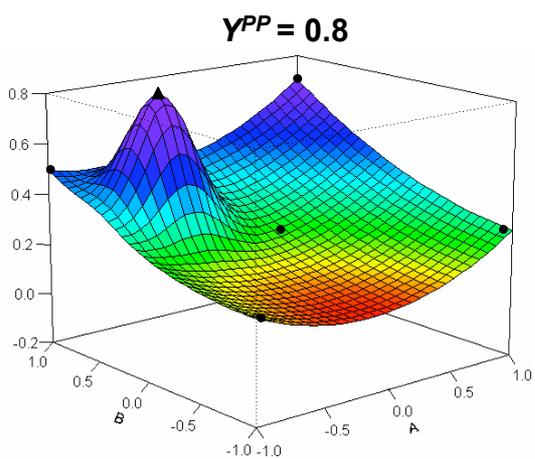
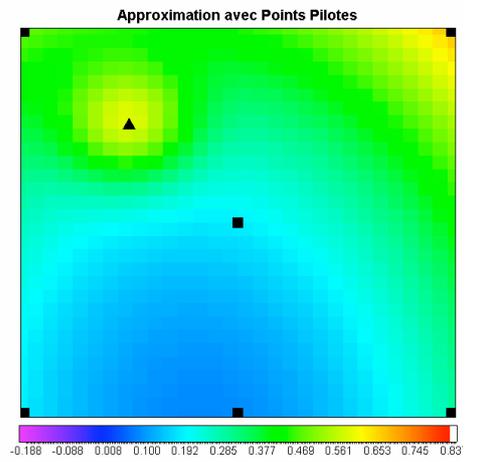
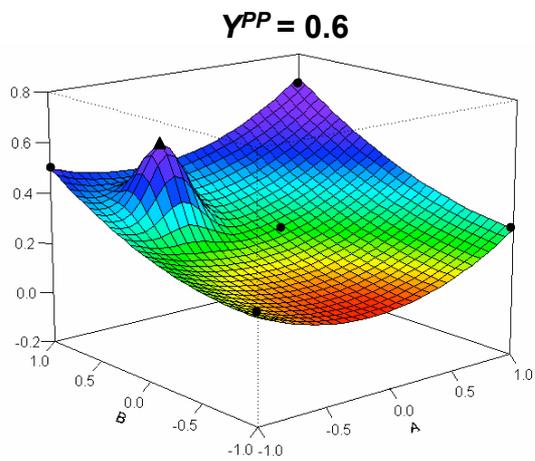
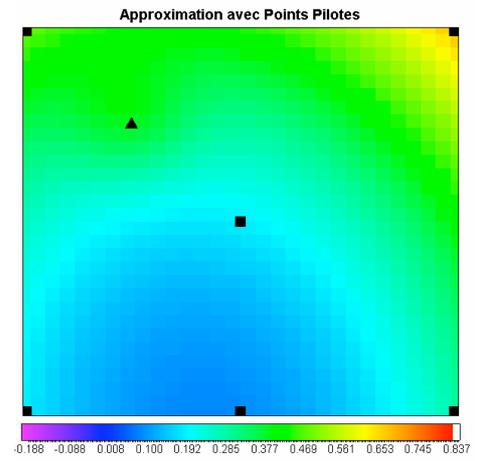
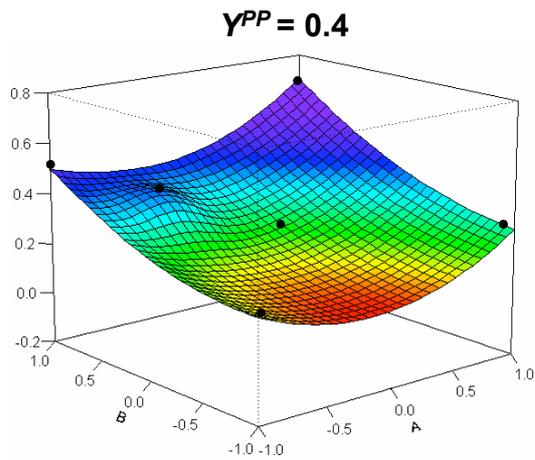


Figure 4.3 : Influence d'un point pilote sur l'approximation (2)

Comme l'illustrent la Figure 4.2 et la Figure 4.3, l'approximation comprenant un ou plusieurs points pilotes est toujours un interpolateur des données simulées. La valeur attribuée à un point pilote permet de contraindre l'approximation dans un voisinage de ce point pilote, et par conséquent la prédictivité de l'approximation est modifiée localement. Ce voisinage est bien évidemment fonction de la longueur de corrélation du variogramme.

Cette approche de pilotage des déformations de l'approximation courante en vue d'une amélioration de sa prédictivité a nécessité d'une part une réflexion sur la manière de positionner ces points pilotes dans le domaine expérimental, et d'autre part sur le choix d'une technique pour déterminer les valeurs de réponses à attribuer à ces points pilotes.

4.3 Emplacement des points pilotes

Généralement en ingénierie de réservoir, l'emplacement des points pilotes est souvent effectué de manière empirique, en adoptant, la plupart du temps, un pas régulier sur le domaine incertain. Par exemple, un point pilote peut être placé toutes les demi-longueurs de corrélation afin de perturber la totalité de l'espace incertain (cf. Figure 4.4).

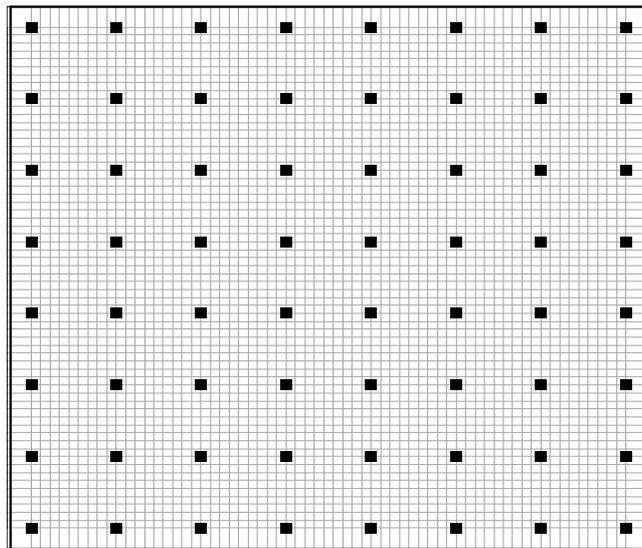


Figure 4.4 : Position des points pilotes sur le domaine incertain en 2D

Mais cette approche pour placer les points pilotes, ne permet pas de prendre en compte les spécificités du modèle courant (emplacement des expériences réelles, etc.) et conduit généralement à un nombre de points pilotes très élevé. Or, on peut naturellement penser que la prise en compte de ces spécificités peut améliorer la prédictivité à l'aide d'un nombre limité de points pilotes, placés de façon plus judicieuse.

Nous envisageons donc de choisir l'emplacement et la valeur d'un point pilote en tenant compte des deux critères suivants :

- la contribution du point pilote dans la réduction des incertitudes sur le modèle d'approximation obtenu précédemment,
- la capacité du point pilote à augmenter la prédictivité de l'approximation courante.

Pour que ce choix soit efficace, il faut pouvoir tout d'abord quantifier l'impact d'un éventuel point pilote sur la robustesse de l'interpolation.

Plaçons nous dans une partie peu échantillonnée du domaine expérimental, supposons que nous appliquons des perturbations locales sur les zones où il n'y a pas d'observations. Leurs effets dépendent de la variance de krigeage. Ainsi, nous voyons sur la Figure 4.5 représentant la variance de l'estimation, que plus on s'éloigne d'un point du plan d'expériences (représenté ici par un carré noir), plus la variance de krigeage augmente. Une variance importante signifie que le manque d'information ne permet pas de contraindre la surface de manière précise. Nous proposons donc de placer un point pilote là où la variance de krigeage est maximale, c'est-à-dire dans les endroits peu échantillonnés du domaine expérimental.

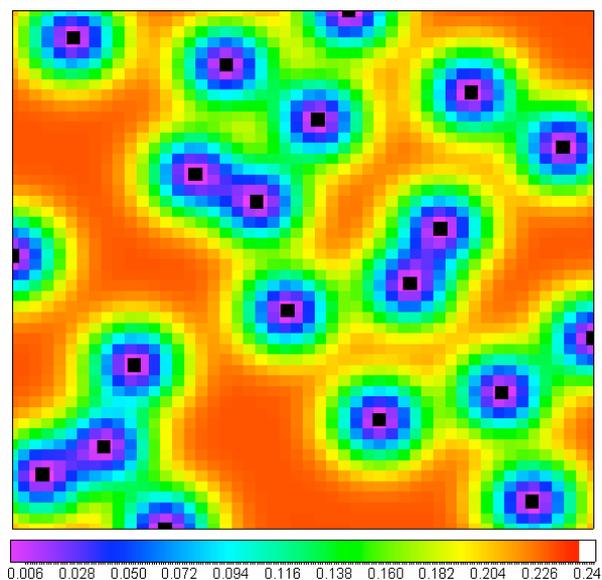


Figure 4.5 : Variance de krigeage dans le domaine incertain

Remarque : Une alternative serait de placer les points pilotes là où l'on observe un gradient important des valeurs estimées ou de possibles discontinuités.

Supposons qu'outre les observations, un certain nombre n_p de points pilotes aient déjà été positionnés dans le domaine incertain D et que l'on cherche à placer de nouveaux points pilotes pour améliorer la prédictivité du modèle. On assimile alors les points pilotes déjà positionnés à des données de variance de krigeage nulle. C'est pour tenir compte de

l'emplacement de points déjà existants que nous optimisons l'emplacement des points pilotes de manière séquentielle.

Pour cela, toujours sous hypothèse de stationnarité, on construit le critère suivant qui représente la variance de krigeage. Celle-ci est présentée en annexe B.2, pour tout point \mathbf{x} du domaine expérimental D , on a :

$$\text{Var}\{Y(\mathbf{x}) - \hat{Y}(\mathbf{x})\} = \text{Var}\{Y(\mathbf{x})\} - {}^t\mathbf{c}(\mathbf{x})\mathbf{C}^{-1}\mathbf{c}(\mathbf{x}) = \tilde{\sigma}^2(1 - {}^t\mathbf{r}(\mathbf{x})\mathbf{R}^{-1}\mathbf{r}(\mathbf{x})) \quad (4.2)$$

où $\tilde{\sigma}^2$ peut être estimé :

- soit à l'aide d'une loi *a priori* non informative de Jeffreys : $\pi(\sigma) \propto \frac{1}{\sigma^2}$, soit

$$\tilde{\sigma}^2 = s^2 = \frac{1}{n-p} \sum_{i=1}^n (Y(\mathbf{x}_i) - \bar{Y})^2,$$

- soit par maximum de vraisemblance restreint, comme suggéré initialement par Currin et al. [14] et recommandé par d'autres auteurs par la suite [22] et [57] :

$$\tilde{\sigma}^2 = \frac{1}{n-p} {}^t(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}).$$

Dans les deux cas, $p = \text{rang}(\mathbf{X})$ et n le nombre d'observations.

Et par conséquent, nous cherchons le point \mathbf{x} qui maximise la fonction J_1 :

$$J_1 = \max_{\mathbf{x} \in D} \left\{ \tilde{\sigma}^2 (1 - {}^t\mathbf{r}(\mathbf{x})\mathbf{R}^{-1}\mathbf{r}(\mathbf{x})) \right\} \quad (4.3)$$

Ce critère quantifie l'impact des incertitudes sur la variation de la fonction d'erreur. La position optimale d'un nouveau point pilote correspond au point \mathbf{x} pour lequel le critère J_1 est maximal.

Les nouveaux points pilotes peuvent donc être positionnés en adoptant la démarche itérative suivante :

1. calcul de la distribution de variance relative au krigeage des mesures réelles et des points pilotes déjà existants,
2. positionnement d'un nouveau point pilote selon le critère J_1 ,
3. retour à l'étape 1 si on veut ajouter un autre point pilote, pour la mise à jour de la distribution de variance de krigeage.

Nous avons choisi d'ajouter un nombre de points pilotes inférieur au nombre de simulations, typiquement au plus de l'ordre de la dizaine, afin de ne pas trop perturber le modèle. Une fois l'emplacement optimal des points pilotes déterminé, il faut attribuer une valeur « fictive » de la réponse en ces points, afin de maximiser la prédictivité *a priori* de l'approximation.

4.4 Valeur fictive de la réponse aux points pilotes

Nous rappelons que la procédure d'ajout d'information fictive a pour intérêt de choisir, parmi la classe d'interpolateur des données réelles complétées par des données fictives aux points pilotes, celui qui est de prédictivité maximale. Il s'agit d'améliorer la prédictivité *a priori* du modèle d'approximation. Dans cette partie, nous montrons comment définir les valeurs fictives $\mathbf{Y}^{pp} = (Y_1^{pp}, \dots, Y_{n_p}^{pp})$ aux n_p points pilotes dont les positions $\mathbf{x}^{pp} = (x_1^{pp}, \dots, x_{n_p}^{pp})$ ont été fixées. Celles-ci sont optimisées à l'aide d'une fonction objectif qui mesure cette prédictivité *a priori*.

Étant donné que le krigeage est une méthode d'interpolation exacte, les résidus « classiques » $e_{(i)} = Y_i - \hat{Y}(\mathbf{x}_i)$ sont nuls. La qualité d'ajustement de tout interpolateur est donc parfaite, par conséquent aucune information sur la prédictivité de l'approximation n'est apportée par ces résidus. Il faut donc recourir à d'autres résidus, que nous appelons ici résidus de prédiction et qui sont définis comme suit.

Le résidu de prédiction en un point du plan d'expériences est obtenu en effectuant l'ajustement d'un nouveau modèle sans ce point. Plus précisément, si on supprime le point \mathbf{x}_i du plan d'expériences, le résidu de prédiction en \mathbf{x}_i est donné par : $e_{(i)} = Y(\mathbf{x}_i) - Y_{-i}(\mathbf{x}_i) = Y_i - \hat{Y}_{(i)}$, où $\hat{Y}_{(i)}$ représente l'estimation au point \mathbf{x}_i d'un modèle ajusté sans le point \mathbf{x}_i [11].

Le fait de supprimer un point \mathbf{x}_i du plan d'expériences et d'estimer un nouveau modèle va permettre de déterminer si le point \mathbf{x}_i est suffisant pour stabiliser l'approximation. En effet, si le nouveau modèle ne permet pas une bonne estimation de la réponse en ce point ou zone du domaine expérimental, alors l'approximation est instable dans un voisinage de \mathbf{x}_i et donc de prédictivité potentiellement faible à cet endroit.

Le calcul des résidus de prédiction en chacun des points du plan d'expériences va ainsi permettre le calcul de la prédictivité *a priori* de l'approximation. On appelle le pouvoir de prédictivité *a priori* du modèle la somme des carrés des erreurs de prédiction, encore appelé coefficient *PRESS* :

$$PRESS = \sum_{i=1}^n e_{(i)}^2 = \sum_{i=1}^n [Y_i - \hat{Y}_{(i)}]^2, \quad (4.4)$$

avec, d'après l'équation du modèle composite (3.9) :

$$\hat{Y}_{(i)} = \mathbf{X}_{(i)} \hat{\boldsymbol{\beta}}_{(i)} + {}^t \lambda_{(i)} (\mathbf{Y}_{(i)} - \mathbf{X}_{(i)} \hat{\boldsymbol{\beta}}_{(i)})$$

où : ${}^t\lambda_{(i)} = \text{Cov}(\mathbf{X}^i, \mathbf{X}_{(i)}) [\text{Cov}(\mathbf{X}_{(i)}, \mathbf{X}_{(i)})]^{-1}$, \mathbf{X}^i est la $i^{\text{ème}}$ ligne de la matrice \mathbf{X} et $\mathbf{X}_{(i)}$ est la matrice \mathbf{X} sans la $i^{\text{ème}}$ ligne.

Maximiser le pouvoir de prédictivité d'une approximation consiste à minimiser les erreurs de prédiction, donc le coefficient *PRESS*.

4.4.1 Calcul du coefficient *PRESS* à moindre coût

Le calcul de la statistique *PRESS* nécessite le calcul des résidus de prédiction pour les n points du plan d'expériences. Pour ce faire, l'ajustement de n modèles de régression et n modèles de krigeage doit être réalisé. En effet, le résidu de prédiction au point i est donné par :

$$\begin{aligned}
 e_{(i)} &= Y_i - \mathbf{Y}_{(i)} \\
 &= Y_i - \mathbf{X}_{(i)} \hat{\boldsymbol{\beta}}_{(i)} - {}^t\lambda_{(i)} \left(Y_{(i)} - \mathbf{X}_{(i)} \hat{\boldsymbol{\beta}}_{(i)} \right) \\
 &= e_{(i)}^{\text{reg}} - \sum_{j \neq i} \lambda_{(i)}^j \left(Y_j - \mathbf{X}_{(i)}^j \hat{\boldsymbol{\beta}}_{(i)} \right) \\
 &= e_{(i)}^{\text{reg}} - e_{(i)}^{\text{krig}} \tag{4.5}
 \end{aligned}$$

Toutefois, il est possible de calculer les résidus de prédiction $e_{(i)}^{\text{reg}}$ à partir de résultats provenant d'un seul ajustement par moindres carrés à n observations. Nous verrons dans le paragraphe suivant comment obtenir les différents termes de l'équation (4.5) à moindre coût. Pour ce faire, nous commençons par expliciter le calcul du résidu de prédiction $e_{(i)}^{\text{reg}}$ pour un modèle de régression, puis nous optimisons le calcul du second terme $e_{(i)}^{\text{krig}}$ de (4.5).

- **Calcul du résidu de prédiction pour un modèle de régression**

Soit $\hat{\boldsymbol{\beta}}_{(i)}$ le vecteur des coefficients de la régression obtenus en supprimant le $i^{\text{ème}}$ point et $e_{(i)}^{\text{reg}}$ le résidu de prédiction issu d'une régression par moindres carrés ordinaires. Par définition, nous avons :

$$\hat{\boldsymbol{\beta}}_{(i)} = \left[{}^t\mathbf{X}_{(i)} \mathbf{X}_{(i)} \right]^{-1} {}^t\mathbf{X}_{(i)} \mathbf{Y}_{(i)}$$

où $\mathbf{X}_{(i)}$ et $\mathbf{Y}_{(i)}$ sont les vecteurs \mathbf{X} et \mathbf{Y} obtenus sans la $i^{\text{ème}}$ observation.

Il existe une relation entre $\left[{}^t\mathbf{X}\mathbf{X} \right]^{-1}$ et $\left[{}^t\mathbf{X}_{(i)} \mathbf{X}_{(i)} \right]^{-1}$, celle-ci est donnée par [37] :

$$\left[{}^t\mathbf{X}_{(i)} \mathbf{X}_{(i)} \right]^{-1} = \left({}^t\mathbf{X}\mathbf{X} \right)^{-1} + \frac{\left({}^t\mathbf{X}\mathbf{X} \right)^{-1} \mathbf{X}^i {}^t\mathbf{X}^i \left({}^t\mathbf{X}\mathbf{X} \right)^{-1}}{1 - h_{ii}} \tag{4.6}$$

avec $h_{ij} = {}^t \mathbf{X}^i ({}^t \mathbf{X} \mathbf{X})^{-1} \mathbf{X}^j$.

En utilisant la relation (4.6), nous pouvons écrire :

$$\begin{aligned}
 e_{(i)}^{reg} &= Y_i - {}^t \mathbf{X}^i \left[({}^t \mathbf{X} \mathbf{X})^{-1} + \frac{({}^t \mathbf{X} \mathbf{X})^{-1} \mathbf{X}^i {}^t \mathbf{X}^i ({}^t \mathbf{X} \mathbf{X})^{-1}}{1 - h_{ii}} \right] {}^t \mathbf{X}_{(i)} \mathbf{Y}_{(i)} \\
 &= Y_i - {}^t \mathbf{X}^i ({}^t \mathbf{X} \mathbf{X})^{-1} {}^t \mathbf{X}_{(i)} \mathbf{Y}_{(i)} - \frac{{}^t \mathbf{X}^i ({}^t \mathbf{X} \mathbf{X})^{-1} \mathbf{X}^i {}^t \mathbf{X}^i ({}^t \mathbf{X} \mathbf{X})^{-1} {}^t \mathbf{X}_{(i)} \mathbf{Y}_{(i)}}{1 - h_{ii}} \\
 &= \frac{(1 - h_{ii}) Y_i - (1 - h_{ii}) {}^t \mathbf{X}^i ({}^t \mathbf{X} \mathbf{X})^{-1} {}^t \mathbf{X}_{(i)} \mathbf{Y}_{(i)} - h_{ii} {}^t \mathbf{X}^i ({}^t \mathbf{X} \mathbf{X})^{-1} {}^t \mathbf{X}_{(i)} \mathbf{Y}_{(i)}}{1 - h_{ii}} \\
 &= \frac{(1 - h_{ii}) Y_i - {}^t \mathbf{X}^i ({}^t \mathbf{X} \mathbf{X})^{-1} {}^t \mathbf{X}_{(i)} \mathbf{Y}_{(i)}}{1 - h_{ii}}
 \end{aligned}$$

De plus, comme ${}^t \mathbf{X} \mathbf{Y} = {}^t \mathbf{X}_{(i)} \mathbf{Y}_{(i)} + \mathbf{X}^i Y_i$, le résidu de prédiction de la régression devient :

$$\begin{aligned}
 e_{(i)}^{reg} &= \frac{(1 - h_{ii}) Y_i - {}^t \mathbf{X}^i ({}^t \mathbf{X} \mathbf{X})^{-1} ({}^t \mathbf{X} \mathbf{Y} - \mathbf{X}^i Y_i)}{1 - h_{ii}} \\
 &= \frac{(1 - h_{ii}) Y_i - {}^t \mathbf{X}^i ({}^t \mathbf{X} \mathbf{X})^{-1} {}^t \mathbf{X} \mathbf{Y} + {}^t \mathbf{X}^i ({}^t \mathbf{X} \mathbf{X})^{-1} \mathbf{X}^i Y_i}{1 - h_{ii}} \\
 &= \frac{(1 - h_{ii}) Y_i - {}^t \mathbf{X}^i \hat{\boldsymbol{\beta}} + h_{ii} Y_i}{1 - h_{ii}}
 \end{aligned}$$

Soit enfin :

$$e_{(i)}^{reg} = \frac{Y_i - {}^t \mathbf{X}^i \hat{\boldsymbol{\beta}}}{1 - h_{ii}} \quad (4.7)$$

Le numérateur de (4.7) est donc le résidu classique e_i d'une approximation par moindres carrés ordinaires des n observations. Ainsi, le $i^{\text{ème}}$ résidu du coefficient *PRESS* est donné par :

$$e_{(i)}^{reg} = \frac{e_i}{1 - h_{ii}} \quad (4.8)$$

Sous cette forme, il est facile de voir que ce coefficient ne dépend que des résidus classiques du modèle de régression ajusté à partir des n données.

- **Calcul du PRESS pour le modèle composite**

Considérons à présent le deuxième terme $e_{(i)}^{reg}$ de l'équation (4.5). D'après [37], nous avons la relation :

$$\hat{\beta}_{(i)} = \hat{\beta} - \frac{({}^t\mathbf{X}\mathbf{X})^{-1} \mathbf{X}^i e_i}{1 - h_{ii}} \quad (4.9)$$

Par conséquent :

$$\begin{aligned} e_{(i)}^{reg} &= \sum_{j \neq i} \lambda_{(i)}^j \left(\mathbf{Y}_j - \mathbf{X}_{(i)}^j \hat{\beta}_{(i)} \right) = \sum_{j \neq i} \lambda_{(i)}^j \left(\mathbf{Y}_j - \mathbf{X}^j \hat{\beta} + \frac{{}^t \mathbf{X}^j ({}^t \mathbf{X} \mathbf{X})^{-1} \mathbf{X}^i e_i}{1 - h_{ii}} \right) \\ &= \sum_{j \neq i} \lambda_{(i)}^j \left(e_j + \frac{h_{ij} e_i}{1 - h_{ii}} \right) \end{aligned}$$

D'où :

$$\begin{aligned} \sum_{i=1}^n (e_{(i)})^2 &= \sum_{i=1}^n (e_{(i)}^{reg} - e_{(i)}^{krig})^2 \\ &= \sum_{i=1}^n \left(e_{(i)}^{reg} - \sum_{j \neq i} \lambda_{(i)}^j \left(e_j + \frac{h_{ij} e_i}{1 - h_{ii}} \right) \right)^2 \\ &= \sum_{i=1}^n \left(\frac{e_i}{1 - h_{ii}} - \sum_{j \neq i} \lambda_{(i)}^j \left(\frac{(1 - h_{ii}) e_j - h_{ij} e_i}{1 - h_{ii}} \right) \right)^2 \end{aligned}$$

Et enfin,

$$PRESS = \sum_{i=1}^n \frac{1}{(1 - h_{ii})^2} \left(e_i - \sum_{j \neq i} \lambda_{(i)}^j \left((1 - h_{ii}) e_j - h_{ij} e_i \right) \right)^2 \quad (4.10)$$

Le coefficient *PRESS* ne dépend donc que des résidus classiques issus de la régression, du coefficient h_{ij} et enfin des poids de krigeage. Ces derniers doivent être recalculés après la suppression de chaque point du plan d'expériences, les autres coefficients étant obtenus directement. C'est cette expression du coefficient *PRESS* que nous utilisons ici pour le calcul de la valeur des points pilotes.

4.4.2 Optimisation de la valeur fictive aux points pilotes

Dans le paragraphe précédent, nous avons présenté un moyen d'obtenir le coefficient *PRESS* à moindre coût. Celui-ci est employé dans la détermination de la valeur des points pilotes pour améliorer la prédictivité *a priori* de l'approximation à partir de valeurs fictives. A présent, concentrons nous sur l'estimation des valeurs fictives aux différents points pilotes \mathbf{x}^{PP} .

La fonction objectif que nous considérons ici est le coefficient Q_2 , il mesure la prédictivité du modèle. Celui-ci est lié au coefficient *PRESS* par la relation :

$$Q_2 = 1 - \frac{PRESS}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

D'après l'équation (4.10), il peut aussi s'écrire de la forme :

$$Q_2 = 1 - \frac{\sum_{i=1}^n \frac{1}{(1-h_{ii})^2} \left(e_i - \sum_{j \neq i} \lambda_{(i)}^j \left((1-h_{ii})e_j - h_{ij}e_i \right) \right)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (4.11)$$

Maximiser la prédictivité *a priori* de l'approximation est équivalente à maximiser le coefficient de prédictivité Q_2 donné en (4.11), c'est-à-dire de réduire les erreurs de prédiction. En effet, une valeur maximale du coefficient de prédictivité Q_2 implique que le niveau d'information obtenu via les simulations pour le modèle donné est nécessaire et suffisant. Dans ce contexte, nous attribuons pour valeur Y_i^{PP} au point \mathbf{x}_i^{PP} , $i = 1, \dots, n$, la valeur de réponse qui maximise le Q_2 , soit :

$$Y_i^{PP} = \max_{Y^{PP}} \{Q_2(Y^{PP})\}$$

L'optimisation de la valeur de la réponse fictive aux différents points est réalisée de manière séquentielle, chaque point pilote \mathbf{x}_i^{PP} n'affectant l'approximation que dans un voisinage de ce point.

Nous utilisons une procédure d'optimisation non-linéaire, de type approche de plus profonde descente pour effectuer la maximisation.

Algorithmes d'optimisation non-linéaire

Les méthodes d'optimisation non-linéaire permettent de calculer les paramètres \mathbf{x} inconnus, c'est à dire de localiser le minimum d'une fonction f , par un processus itératif. L'itération $(k+1)$ d'un algorithme d'optimisation a pour but de déterminer une nouvelle estimation des paramètres $\mathbf{x}^{k+1} = \mathbf{x}^k + t\mathbf{d}_k$, selon le principe suivant :

- Calcul d'une direction : la direction \mathbf{d}_k est solution d'un problème linéarisé en \mathbf{x}_k ,
- Recherche linéaire : le pas t est calculé afin de satisfaire la condition de descente :

$$f(\mathbf{x}^k + t\mathbf{d}_k) < f(\mathbf{x}^k)$$

La mise en œuvre numérique de ces algorithmes repose donc sur le calcul de la direction \mathbf{d}_k . Cette direction est obtenue en calculant les dérivées de la fonction à minimiser f

par rapport aux paramètres. Cette méthode n'est donc envisageable qu'à condition que la fonction f soit différentiable.

- **Construction d'un intervalle de variation possible des valeurs fictives**

Du fait de la difficulté à estimer la covariance et les longueurs de corrélation associées, il nous a semblé judicieux de contraindre les variations des valeurs fictives à un intervalle. En effet, dans le cas d'une faible corrélation donc de grandes fluctuations de la réponse, les valeurs des points optimisés peuvent dépasser les valeurs possibles de la réponse. Afin de limiter la perturbation de l'approximation par de l'information non réelle, nous proposons de limiter les valeurs de la réponse fictive à des intervalles prenant en compte l'incertitude associée au point.

Ce principe est illustré sur la Figure 4.6, où nous avons représenté en bleu l'approximation construite à partir du plan d'expériences et l'intervalle de valeurs possibles de la réponse en pointillés.

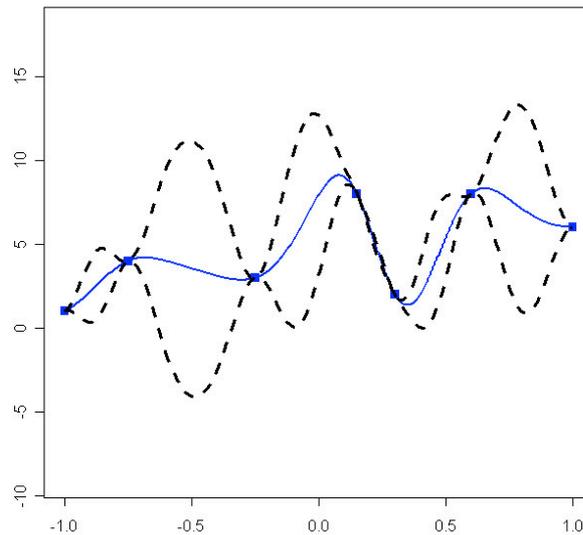


Figure 4.6 : Approximation et variance de krigage associée

Nous avons montré, dans le chapitre précédent, que l'approche employée pour la construction de l'estimateur composite est bayésienne, parce qu'on suppose *a priori* que les résidus sont gaussiens.

Pour obtenir un intervalle de prédiction pour $Z_0 := \hat{Z}(\mathbf{x}_0)$, on utilise les propriétés suivantes qui permettent de construire une application pivotale de cette variable aléatoire [57] :

(i) Z_0 a pour distribution conditionnelle sachant $\mathbf{Z} = \mathbf{z}$:

$$[Z_0 | \mathbf{Z} = \mathbf{z}] \sim N \left({}^t \mathbf{r}_0 \mathbf{R}^{-1} \mathbf{z}, \sigma^2 {}^t \mathbf{r}_0 \mathbf{R}^{-1} \mathbf{r}_0 \right),$$

(ii) la variable aléatoire $Z_0 - {}^t \mathbf{r}_0 \mathbf{R}^{-1} \mathbf{z} \sim N \left(0, \sigma^2 {}^t \mathbf{r}_0 \mathbf{R}^{-1} \mathbf{r}_0 \right)$ est indépendante de \mathbf{Z} ,

(iii) $\frac{1}{\sigma^2} \langle \mathbf{Z}, \mathbf{R}^{-1} \mathbf{Z} \rangle \sim \chi_n^2$ et cette variable aléatoire est indépendante de $Z_0 - {}^t \mathbf{r}_0 \mathbf{R}^{-1} \mathbf{Z}$.

Si le paramètre de corrélation θ est connu et que σ^2 est certain mais inconnu, on déduit de (ii) et (iii) que :

$$\frac{Z_0 - {}^t \mathbf{r}_0 \mathbf{R}^{-1} \mathbf{Z}}{\sqrt{\hat{\sigma}^2 (1 - {}^t \mathbf{r}_0 \mathbf{R}^{-1} \mathbf{r}_0)}} \sim t_n \text{ avec } \hat{\sigma}^2 = \frac{1}{n} \langle \mathbf{Z}, \mathbf{R}^{-1} \mathbf{Z} \rangle$$

d'où une application pivotale pour la prédiction de Z_0 .

Dans ce cas, on déduit, par exemple, les limites suivantes pour un intervalle de prédiction de sécurité $\gamma = 1 - \alpha$:

$$\mathbf{z} \in \square^n \rightarrow {}^t \mathbf{r}_0 \mathbf{R}^{-1} \mathbf{z} \pm t_{1-\alpha/2} \sqrt{\hat{\sigma}^2 (1 - {}^t \mathbf{r}_0 \mathbf{R}^{-1} \mathbf{r}_0)}$$

où $t_{1-\alpha/2}$ désigne le quantile d'ordre $1 - \alpha/2$ de la variable aléatoire de Student à n degrés de liberté.

Dans le cas où les paramètres θ et σ^2 sont inconnus, on peut envisager d'introduire des *a priori* non-triviaux pour tous les paramètres. Mais cette approche présente de nombreuses difficultés théoriques, comme Santner et al. [57] le signalent. Il s'agit de supposer que le paramètre $\phi = (\sigma^2, \theta)$ du modèle est la réalisation d'un aléa $\phi = (\sigma^2, \theta)$ et à introduire une distribution *a priori* pour θ , puis à utiliser le théorème de Bayes pour en déduire, si possible, une application pivotale pour la prédiction de Z_0 .

Toutefois, on peut surmonter, au moins partiellement, les difficultés liées à l'estimation en complétant l'approche bayésienne par l'introduction d'un *a priori* sur tout ou partie des paramètres. On peut ainsi introduire un *a priori* sur σ^2 , par exemple celui de Jeffreys. Dans ce cas, nous avons pour distribution *a priori* pour σ^2 : $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$.

Dès lors que l'on suppose un *a priori* de Jeffreys sur σ^2 , l'application pivotale suit une loi de Student. L'estimation des intervalles de prédiction est donc obtenue en utilisant les quantiles de la loi de Student.

Ainsi, nous avons pour intervalle de prédiction pour chaque point pilote $\mathbf{x}_i^{PP} \in D$:

$$Y_i^{PP} \in I_i^{PP} = \left[\hat{Y}(\mathbf{x}_i^{PP}) - q_\alpha \sqrt{\tilde{\sigma}^2 (1 - {}^t \mathbf{r}(\mathbf{x}_i^{PP}) \mathbf{R}^{-1} \mathbf{r}(\mathbf{x}_i^{PP}))}; \hat{Y}(\mathbf{x}_i^{PP}) + q_\alpha \sqrt{\tilde{\sigma}^2 (1 - {}^t \mathbf{r}(\mathbf{x}_i^{PP}) \mathbf{R}^{-1} \mathbf{r}(\mathbf{x}_i^{PP}))} \right] \quad (4.12)$$

avec :

- $\hat{Y}(\mathbf{x}_i^{PP}) = \mathbf{X}(\mathbf{x}_i^{PP}) \hat{\boldsymbol{\beta}} + {}^t \mathbf{r}(\mathbf{x}_i^{PP}) \mathbf{R}^{-1} \mathbf{Z}$ l'estimation de la valeur de la réponse au point \mathbf{x}_i^{PP} par le modèle composite,

- $q_\alpha = t_{1-\alpha/2}$ désigne le quantile d'ordre $1-\alpha/2$ de la variable aléatoire de Student à n degrés de liberté.

Remarque : Dès lors que l'on ne suppose pas un a priori non informatif de Jeffreys par exemple, pour σ^2 , il n'existe plus d'interprétation paramétrique quant à la distribution de l'application pivotale. Dans ce cas, l'utilisation de techniques de rééchantillonnage telles que le Bootstrap pourrait être mise en œuvre pour estimer les quantiles de l'application pivotale.

En tenant compte de cette contrainte, la valeur optimale au point pilote \mathbf{x}_i^{PP} est obtenue par résolution du problème de maximisation :

$$Y_i^{PP} = \max_{Y^{PP} \in I_i^{PP}} \{Q_2(Y^{PP})\}$$

Supposons à présent que l'on ajoute n_p points pilotes au dispositif expérimental. Le coefficient Q_2 donné en (4.11) est fonction des résidus classiques de la régression, c'est-à-dire des $\mathbf{e} = (\mathbf{e}_j, \mathbf{e}_i^{PP})$, pour $j = 1, \dots, n$ et $i \in \{1, \dots, n_p\}$.

Ainsi, pour chaque point \mathbf{x}_i^{PP} , nous faisons varier la valeur Y_i^{PP} de la réponse dans l'intervalle I_i^{PP} et nous calculons la valeur du coefficient de prédictivité du modèle Q_2 . La valeur attribuée à la réponse fictive Y_i^{PP} est la valeur $Y^{PP} \in I_i^{PP}$ qui maximise la valeur du coefficient Q_2 . Cette procédure est réalisée séquentiellement.

Nous illustrons la méthodologie des points pilotes à l'aide de la Figure 4.7 et la Figure 4.8 ci-dessous.

Tout d'abord, nous rappelons que la variance de krigeage nous a permis d'optimiser la position des points pilotes : les points de variance de krigeage maximale sont choisis pour piloter l'approximation car ils correspondent aux points de plus grande incertitude pour la configuration géométrique des données présentes. La procédure d'ajout séquentiel de trois points pilotes, représentés en rouge, est décrit sur la Figure 4.7.

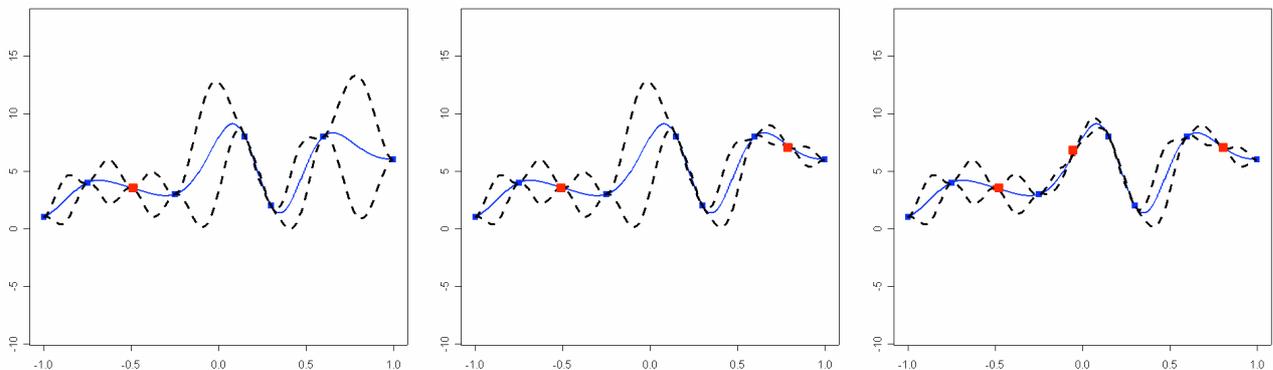


Figure 4.7 : Position séquentielle des points pilotes : points de variance maximale

Une fois que l'emplacement de ces nouveaux points est ajusté, il est important de déterminer un intervalle de valeurs possible pour la valeur de ces points, celui-ci étant fonction de la variance de krigeage (eq.(4.12)). Ainsi, la Figure 4.8 montre l'emplacement des trois points pilotes sélectionnées, représentés en rouge, et l'intervalle de valeur dans lequel la valeur de réponse fictive des points pilotes peut varier, représenté par une flèche rouge. Nous complétons ce premier graphique en représentant les variations du coefficient Q_2 en fonction des valeurs possibles pour Y^{PP} , et ce pour chacun des trois points pilotes. Nous indiquons également la valeur optimale de la réponse fictive pour chaque point, c'est-à-dire celle qui maximise le coefficient Q_2 .

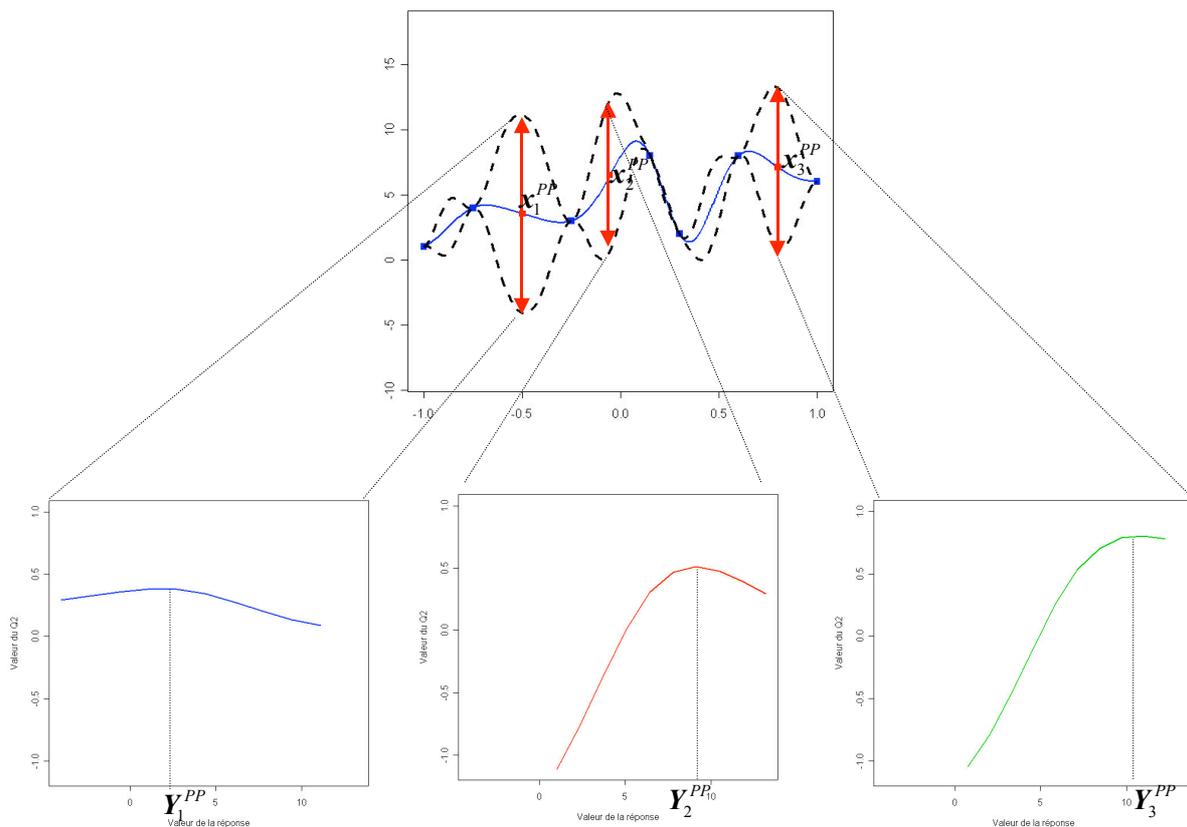


Figure 4.8 : Estimation de la statistique Q_2 pour les points pilotes

Nous disposons à présent des valeurs optimisées des points pilotes, les couples $(\mathbf{x}_i^{PP}, Y_i^{PP})$ sont déterminés de manière à améliorer la prédictivité. La nouvelle approximation, issue d'un modèle composite sur des données réelles et fictives est présentée en rouge sur la Figure 4.9. Nous donnons également l'approximation obtenue uniquement avec des données réelles, en bleu, et enfin la fonction référence.

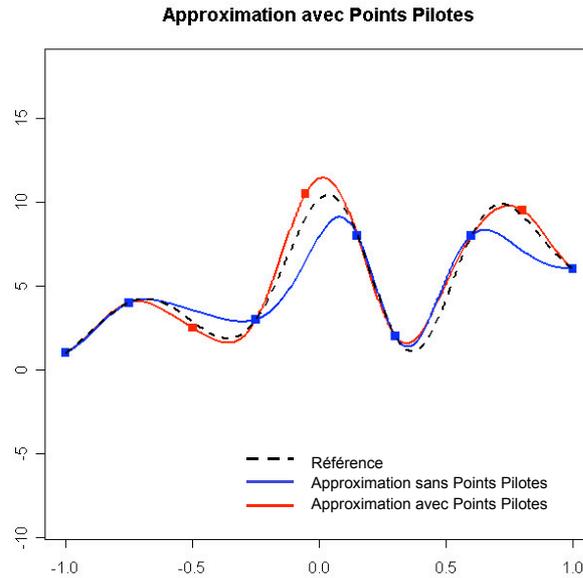


Figure 4.9 : Approximation de prédictivité maximale

Nous constatons dans cet exemple que l'approximation déduite de données fictives associées aux données réelles est plus précise en terme de prédictivité que l'approximation obtenue sans points pilotes. En effet, la valeur du coefficient Q_2 est de -1.54 pour l'approximation sans points pilotes, et de 0.51 pour celle avec 3 points pilotes. De plus, nous constatons que l'approximation avec points pilotes est plus proche de la référence que celle obtenue sans points pilotes.

Nous présentons à présent un exemple d'ajustement avec et sans points pilotes d'une fonction analytique non-régulière en deux dimensions. Une représentation de cette fonction dans le cube unité $[-1,1]^2$ est donnée sur la Figure 4.10, elle a pour expression :

$$F(x,y) = -\frac{\sin(x)}{x} - \frac{\sin(y+2)}{y+2}$$

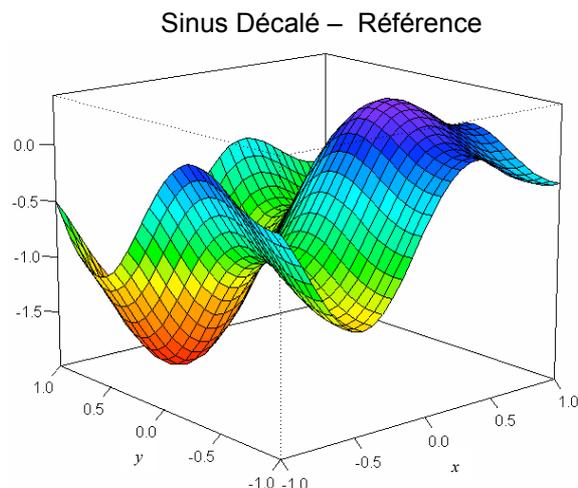


Figure 4.10 : Fonction Sinus Décalée - Référence

Nous considérons dans cet exemple un plan d'expériences à 21 simulations, celui-ci étant issu d'une itération donnée du processus de modélisation adaptative. Dans un premier temps, nous ajustons un modèle à partir des 21 valeurs de réponses simulées. Ensuite, nous ajoutons séquentiellement 6 points pilotes au dispositif expérimental, dont les valeurs sont optimisées comme indiqué dans ce chapitre, c'est-à-dire de manière à maximiser la prédictivité *a priori* de l'approximation courante. Une nouvelle approximation est estimée, celle-ci prenant en compte les 21 valeurs simulées et les 6 valeurs fictives. La Figure 4.11 présente les résultats. Ainsi, les graphes (a) et (c) donnent une illustration de superposition du plan d'expériences utilisé, représenté en gris, des points pilotes rajoutés, en rouge, et les contours de la référence, en noir, et des deux approximations en bleu. Nous y avons également présenté les graphes (b) et (d) de la valeur estimée en fonction de la référence ainsi que le coefficient Q_2 obtenu dans chaque cas. Ces derniers graphes ont été réalisés pour 900 valeurs de la référence et de l'approximation.

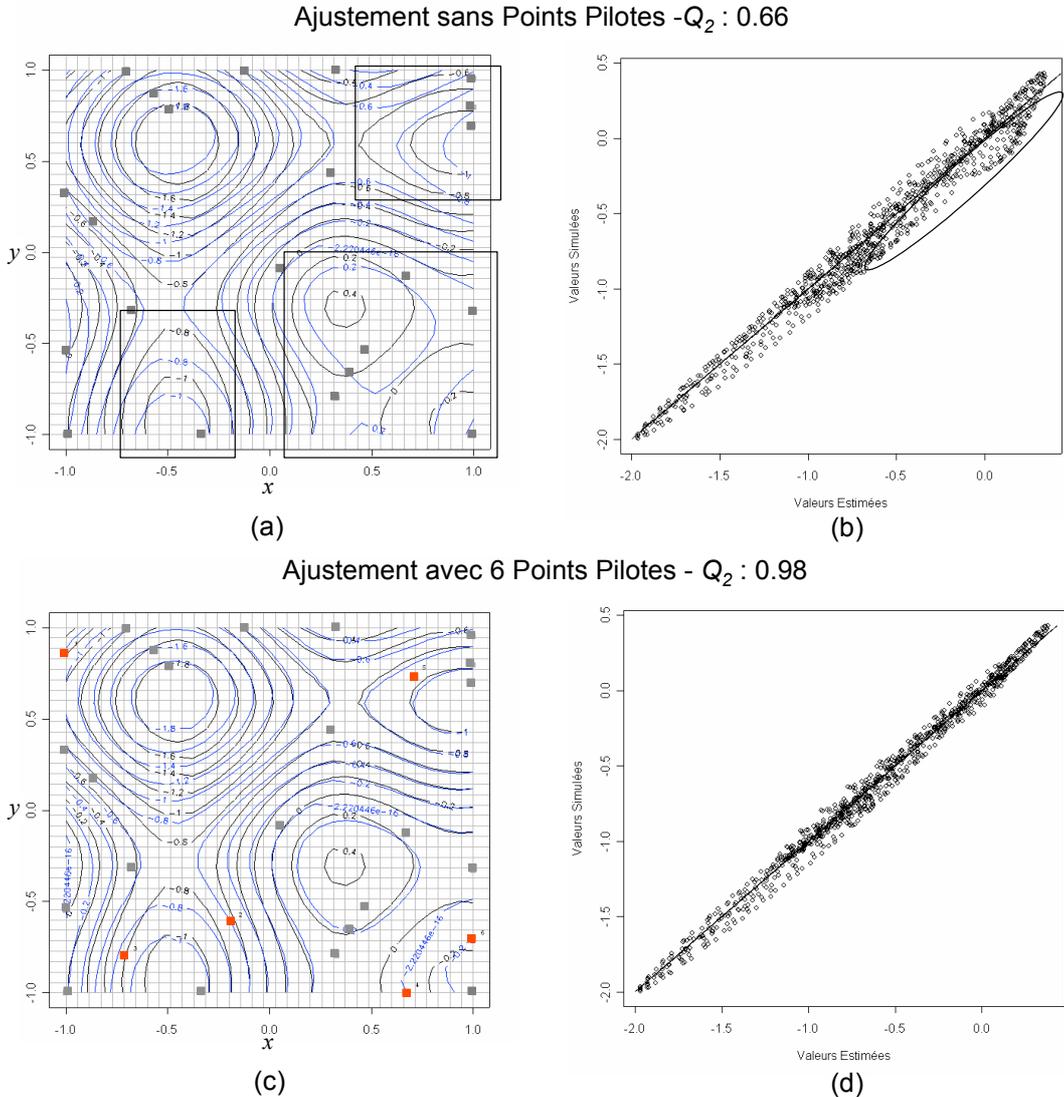


Figure 4.11 : Approximation avec et sans Points Pilotes

Ainsi, d'après la Figure 4.11, nous constatons que l'ajout de 6 points pilotes au plan d'expériences à 21 simulations est informatif. En effet, la prédictivité de l'approximation avec points pilotes est meilleure que celle sans information fictive, la valeur du coefficient Q_2 passant respectivement de 0.66 à 0.98. Nous rappelons que la valeur maximale de ce coefficient est 1. De plus, la comparaison des graphiques (a) et (c) confirme l'intérêt des points pilotes, nous constatons, particulièrement dans la zone en bas à droite du domaine expérimental une nette amélioration de l'approximation, ceci étant du à l'ajout de deux points pilotes dans cette zone. Cet accroissement de la qualité du modèle est également visible sur les graphes (b) et (d), où apparaît clairement une plus faible dispersion des valeurs autour de la bissectrice dans le cas d'un ajout de points pilotes.

A présent, toujours sur le même exemple, nous montrons l'impact du nombre de points pilotes sur la prédictivité de l'approximation. Nous commençons donc par calculer le coefficient Q_2 pour le plan contenant uniquement les valeurs simulées, puis nous ajoutons de un à dix points pilotes à ce plan. Les résultats de cette étude sont présentés sur la Figure 4.12.

Évolution de la Statistique Q_2 en fonction du nombre de Points Pilotes

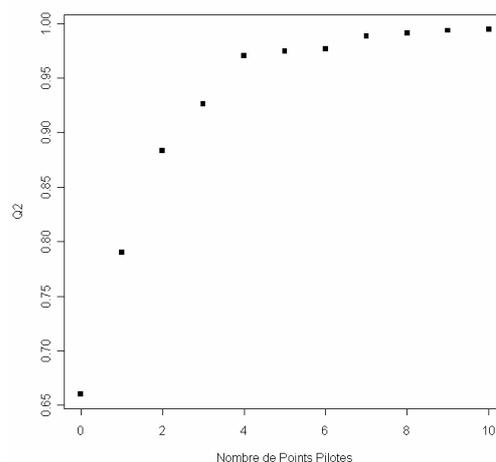


Figure 4.12 : Évolution du Q_2 en fonction du nombre de points pilotes

Ainsi, l'ajout de points pilotes au plan d'expériences courant permet d'augmenter significativement la valeur du coefficient Q_2 sans effectuer de nouvelles simulations. En effet, cette valeur, initialement égale à 0.66, atteint rapidement la valeur de 0.98, soit une excellente prédictivité. Un nombre limité de point pilote suffit donc à améliorer considérablement cette statistique.

Ainsi, cette approche originale d'ajout de points pilotes est efficace car elle permet, non seulement, de sélectionner l'approximation de prédictivité maximale au vu des données simulées, mais également, de réduire le nombre total de simulations à effectuer, car cette approximation est obtenue sans faire de nouvelles simulations.

4.5 Conclusions

Dans ce chapitre, nous avons proposé un moyen original de contrôler et d'accroître la prédictivité de l'approximation à partir d'informations fictives. La technique des points pilotes nous a permis d'améliorer l'approximation en des parties du domaine expérimental sous ou non-échantillonnées, i.e. dans les zones où la variance de krigeage est maximale. Les valeurs des réponses aux points pilotes ont été optimisées de manière à maximiser le coefficient de prédictivité *a priori* de l'approximation : Q_2 . Grâce à cette méthode, de nouvelles valeurs de réponses sont ajoutées aux valeurs réelles sans qu'aucune simulation n'ait été effectuée. L'ajout de points pilotes devant aussi permettre de réduire le nombre de simulations d'écoulement nécessaire à une bonne approximation.

Cette méthode joue donc un rôle essentiel dans notre approche de modélisation, mais ne suffit pas à assurer la prédictivité du modèle. Celle-ci dépend fortement de la stratégie d'échantillonnage du domaine expérimental. Le plan d'expériences initial permettant une première approximation étant de petite taille, il faut maintenant enrichir le dispositif expérimental par des données réelles pour explorer de nouvelles zones et gagner en précision. Ces nouvelles simulations vont permettre d'améliorer la qualité et la stabilité des surfaces obtenues par les méthodes géostatistiques. L'objet du chapitre suivant est de présenter différentes techniques d'ajout de nouvelles simulations : il s'agit d'explorer le domaine expérimental en recherchant les conditions des nouvelles simulations qui pourraient être les plus informatives.

Chapitre 5

Stratégie d'expérimentation pour réponses simulées non-régulières

5.1 Introduction

La définition d'un bon plan d'échantillonnage est essentiel pour construire un prédicteur efficace et précis. La qualité de l'analyse statistique et donc la prédictivité du modèle dépendent fortement de la disposition des points du plan d'expériences, c'est-à-dire de leur nombre et de leur emplacement.

Le plan d'expériences initial, proposé pour la modélisation de réponses non-régulières, est un hypercube latin de distance maximin. Ce plan permet d'obtenir des simulations bien réparties dans le domaine, mais il est généralement constitué d'un nombre d'expériences insuffisant et inadéquat dès lors que la surface à modéliser est complexe. Afin d'améliorer la qualité de l'approximation, au fur et à mesure du processus d'ajustement, de nouvelles simulations sont rajoutées de manière appropriée.

Comme illustré par la Figure 5.1, l'objectif de ce chapitre est de présenter la technique mise en œuvre pour l'augmentation du dispositif expérimental courant, pour améliorer l'approximation de la réponse.

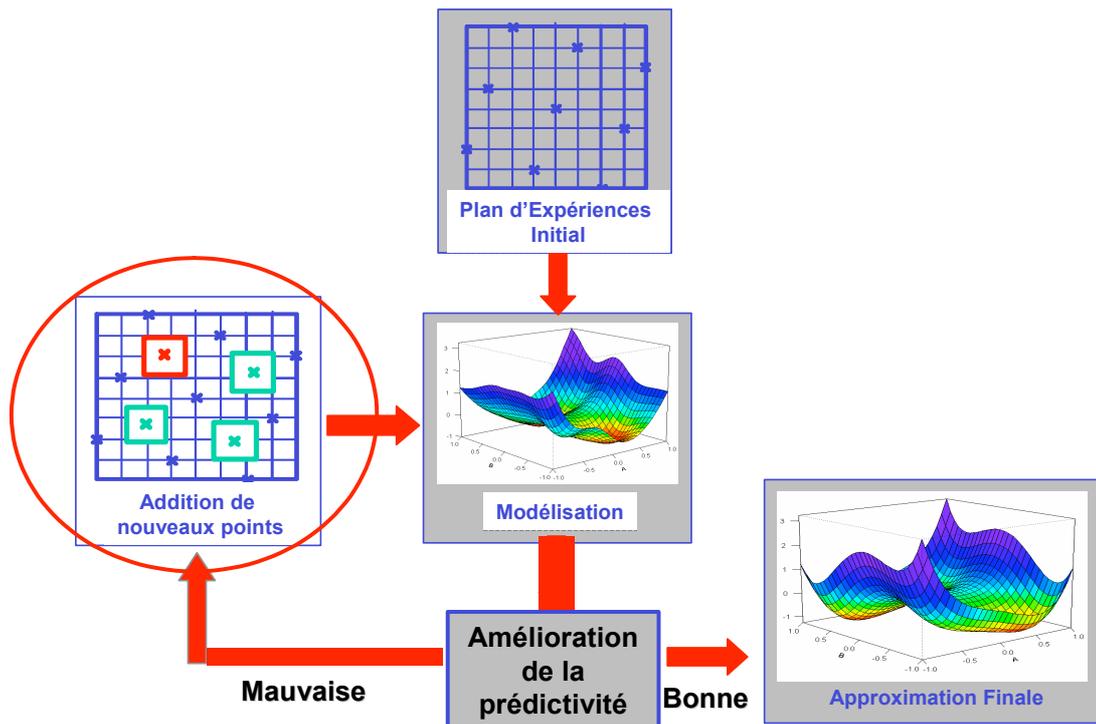


Figure 5.1 : Principe de la méthodologie adaptative

A chaque itération i , soit pour un dispositif expérimental courant $\mathbf{X}^{(i)}$ et une approximation courante issue de $\mathbf{X}^{(i)}$, la phase d'enrichissement comporte deux étapes majeures :

- étude de la dépendance de la réponse aux paramètres. Cette étape a pour objet de déterminer les paramètres pour lesquels l'approximation courante ne permet pas de décrire la complexité réelle de la réponse. En d'autres termes, il s'agit par le biais de techniques d'analyse de sensibilité (ACP, techniques globales par indices de Sobol, etc...) de déterminer le sous-ensemble de paramètres pour lequel un enrichissement du dispositif est nécessaire afin d'améliorer l'approximation en agissant sur ces paramètres.
- augmentation du dispositif expérimental dans le sous-espace défini au préalable par l'étude des dépendances.

Cette organisation en deux étapes est essentielle, car elle vise à maîtriser le nombre de simulations. En effet, par la recherche de nouveaux points de simulation dans une partie du domaine expérimental, on peut réduire le nombre de simulations supplémentaires à effectuer pour améliorer l'approximation.

Dès lors que les paramètres pour lesquels le dispositif courant n'est pas suffisant sont précisés, nous proposons de traiter la deuxième étape de cette stratégie expérimentale, à savoir l'augmentation du dispositif courant à proprement parler, selon quatre critères :

- Recherche des maxima et minima de l'approximation courante : ils permettent de contraindre l'approximation aux extrema locaux et de détecter les non-régularités.
- Simulation d'un ou plusieurs points du domaine expérimental minimisant la norme L_2 des dérivées partielles de l'approximation courante. L'ajout de ce type de point peut aider à détecter des éventuelles irrégularités de la réponse, et plus particulièrement, lorsque la modélisation devient plus précise, déterminer des points d'inflexion.
- Sélection et simulation des points pour lesquels une variation de la réponse affecte grandement le coefficient de prédictivité Q_2 du modèle. Dans ce contexte, les points pilotes utilisés comme information fictive sont d'excellents candidats à la simulation puisque situés dans les zones sous-échantillonnées du domaine correspondant au maximum de la variance de krigeage.
- Ajout de points de confirmation pour évaluer la prédictivité *a posteriori* de l'approximation courante.

Ces différents critères que nous proposons peuvent être mis en œuvre à chaque itération. En effet, nous proposons d'ajouter les nouveaux points de manière séquentielle afin de maîtriser le nombre de simulations et de maximiser la quantité d'information apportée par les simulations. A chaque itération, de nouveaux points sont ainsi simulés et intégrés à l'approximation courante avant de procéder à la recherche des points suivants. Par conséquent, on accroît progressivement la quantité d'information, chaque nouveau point étant placé en fonction du comportement de l'approximation courante.

Pour détailler les deux phases de la procédure d'augmentation du dispositif expérimental, et en particulier les quatre critères choisis pour apporter une information pertinente à chaque itération, il nous paraît important de préciser quelles raisons nous ont conduit à sélectionner ces critères.

Classiquement, l'augmentation de dispositifs expérimentaux repose sur une hypothèse de forte régularité et de continuité de la réponse étudiée. Ainsi, on procède à la sélection de nouvelles simulations en étudiant les propriétés de l'échantillonnage dans l'espace des paramètres. On peut, par exemple, chercher à maximiser l'information *a priori* en se basant sur des critères géométriques de répartition des simulations dans l'espace. On procède ainsi lors de la construction de plans optimaux, en assurant le respect de bonnes propriétés de l'information de Fisher du dispositif. Il en est de même lors du dépliage de plans fractionnaires 2^{k-p} , ou lors de la construction d'hypercubes latins. La qualité du plan est jugée *a priori* sur des propriétés de distance et d'orthogonalité dans le domaine expérimental, indépendamment de la réponse. Ces critères, basés sur une disposition géométrique des points dans le domaine expérimental, sont généralement motivés par le fait que le plan est exploité

pour déterminer la sensibilité de la réponse aux paramètres. Il est alors primordial de construire un dispositif possédant de bonnes propriétés d'orthogonalité dans une base donnée.

Nos travaux s'inscrivent dans un contexte différent. Le plan est en effet construit dans le but de proposer une modélisation approchée de la réponse qui soit fiable et robuste. Ce modèle approché aura ainsi vocation à être utilisé comme « boîte noire », et non pour étudier la sensibilité de la réponse aux paramètres. Les critères d'augmentation de dispositif fondés sur une propriété d'orthogonalité ne répondent donc pas en général à notre objectif. Pour compléter le plan d'expériences, nous aurons le souci d'échantillonner *a priori* les zones peu ou pas explorées, en tenant compte des variations de la réponse et non uniquement celle des paramètres.

Comme nous l'avons précisé dans cette introduction, la première étape de la stratégie d'expérimentation pour la modélisation de réponses non-régulières consiste à étudier les dépendances de la réponse aux paramètres.

5.2 Etude de la dépendance de la réponse aux paramètres

Les différentes stratégies d'enrichissement du dispositif expérimental que nous proposons visent à maximiser la quantité d'information apportée par les simulations. Comme le montre la Figure 5.2, cette méthodologie d'augmentation du dispositif est spécifique à chaque réponse, la recherche des nouveaux points étant basée sur la qualité de l'approximation courante.

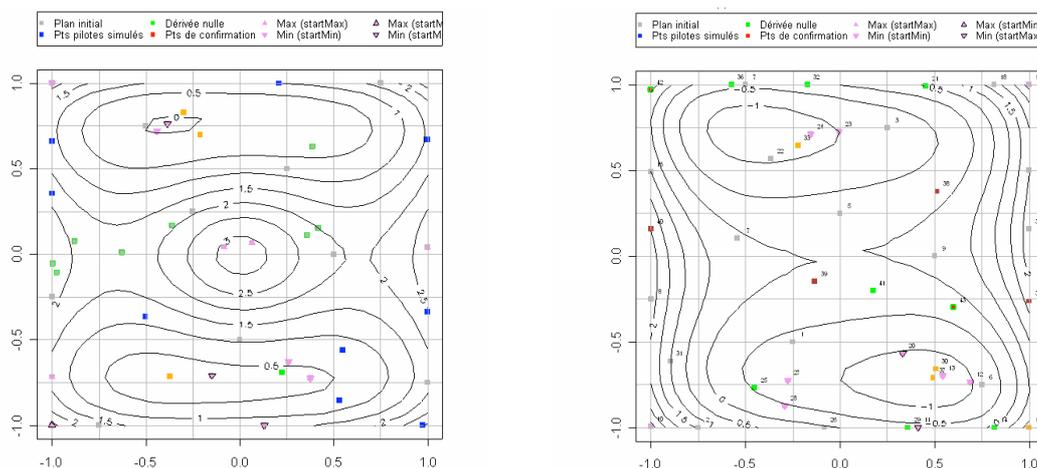


Figure 5.2 : Echantillonnage spécifique à chaque réponse

Cette méthode a pour but de réduire et maîtriser le nombre de simulations en comparaison à la taille requise par des plans classiques tels que les hypercubes latins. Cependant, si la détection des non-régularités est réalisée pour toutes les directions définies

par les paramètres, le nombre de simulations requis pour obtenir une approximation de bonne qualité peut rapidement devenir trop important.

Nous proposons donc dans un premier temps d'étudier, avant tout ajout de points, le type de dépendance entre les paramètres et la réponse. A cette fin, nous proposons l'emploi de techniques où la dimension du domaine de recherche des nouveaux points est réduite, soit pour k paramètres, en travaillant dans un sous-espace $[-1,1]^l$, avec $l \leq k$.

Pour ce faire, nous avons privilégié deux approches :

- l'étude de la dépendance des paramètres entre eux, en procédant par exemple à partir d'un *a priori* physique, ou par analyse en composantes principales, à la représentation de plusieurs paramètres en un attribut.
- l'étude de la dépendance des paramètres à la réponse par le biais de techniques d'analyse de sensibilité.

5.2.1 Diminution de la dimension du problème : agrégation de paramètres

Dans le cas d'un important nombre de paramètres, il peut être nécessaire de réduire la dimension du domaine pour optimiser les temps de calculs. Les méthodes de type analyse en composantes principales ou analyse canonique permettent non seulement de prendre en compte la covariance entre les paramètres, mais aussi de réduire la dimension du problème.

Étant donné que les paramètres peuvent ne pas être indépendants, l'utilisation de la distance euclidienne n'est pas appropriée dans l'espace des paramètres. Ainsi, pour tenir compte des corrélations entre les paramètres, nous proposons de faire un changement de coordonnées en construisant une base orthogonale. Ceci peut être effectué par une analyse en composantes principales classique (ACP), qui détermine une série de nouveaux paramètres, non-corrélés entre eux et qui sont ordonnés en fonction de leur variance.

Soient $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})$, $i = 1, \dots, n$, les n points du plan d'expériences. Dans la base des k composantes principales \mathbf{v}_j , les \mathbf{x}_i de la base initiale sont définis par $\mathbf{cp}_i = (cp_{i1}, \dots, cp_{ik})$ [2]. On a donc :

$$\mathbf{x}_i = \sum_{j=1}^k \mathbf{v}_j^i cp_{ij}$$

Les nouvelles coordonnées cp_{ij} sont combinaisons linéaires des observations \mathbf{x}_i , ceci peut être vu comme une rotation des \mathbf{x}_i .

$$cp_{ij} = \sum_{j=1}^k \mathbf{v}_j^i x_{ij} \quad j = 1, \dots, n$$

Dans le cas où la majeure partie de la variance est expliquée par les premières composantes principales, on peut réduire la dimension de l'espace des paramètres sans perte excessive d'information. Dans ce cas, si on réduit la base à l attributs ($l < k$), seuls les l premiers coefficients cp_i sont conservés. La contribution des autres coefficients est modélisée par une constante b_j et nous avons :

$$\tilde{x}_{ij} \approx \sum_{j=1}^l v_j^i cp_{ij} + \sum_{j=l+1}^k v_j^i b_j$$

Les constantes b_j sont choisies de manière à obtenir la meilleure approximation sur l'ensemble des données \mathbf{x}_i , soit celle qui minimise l'erreur introduite par la réduction de dimension :

$$\tilde{\mathbf{x}}_i - \mathbf{x}_i \approx \sum_{j=l+1}^k v_j^i (b_j - cp_{ij})$$

Ainsi,

$$b_j = \frac{1}{n} \sum_{i=1}^n cp_{ij}$$

Pour que les composantes principales conservent une interprétation physique, on peut, par exemple, envisager de trier les paramètres en trois catégories : variables statiques, variables multiphasiques et variables dynamiques. Une analyse en composantes principales est ensuite réalisée pour chaque catégorie, une ou deux composantes principales sont conservées par catégorie.

L'analyse en composante principale est donc opportune pour la recherche de nouveaux points, elle permet non seulement de limiter les temps de calculs, mais aussi de réduire le nombre total de simulations nécessaires, la dimension de l'espace de recherche de nouvelles simulations étant réduite.

Remarque : *La connaissance préalable du problème physique étudié peut permettre l'inférence de corrélations entre paramètres ou la définition de modèle physique simplifié. Dans une optique d'agrégation des paramètres, l'utilisation de cette connaissance physique, plutôt que de recourir à des méthodes type ACP, permet la prise en compte de la compréhension et l'expertise de l'ingénieur vis-à-vis de la spécificité du problème étudié.*

5.2.2 Analyse de sensibilité

Une mise en œuvre efficace et performante de la modélisation adaptative nécessite d'évaluer, à chaque itération, le type de régularité ou de non-régularité des paramètres vis-à-vis de la réponse.

En effet, si un paramètre a une influence linéaire sur la réponse, seuls deux niveaux pour ce paramètre sont nécessaires, un troisième niveau aidant à confirmer ou infirmer la

dépendance linéaire. Dans ce contexte, réaliser une modélisation adaptative sur l'ensemble des paramètres, sans intégrer cette connaissance sur la dépendance linéaire ou quadratique de la réponse à certains paramètres, conduit à un échantillonnage très coûteux et non-informatif.

La Figure 5.3 illustre ainsi la nécessité d'étudier la dépendance à chaque paramètre avant de réaliser la recherche de nouveaux points. A gauche, l'échantillonnage est obtenu sans considérer la linéarité de la réponse au paramètre A. Le domaine est alors échantillonné indifféremment dans les deux directions A et B. A droite, l'échantillonnage est réalisé en privilégiant l'ajout de nouveaux points dans la direction B.

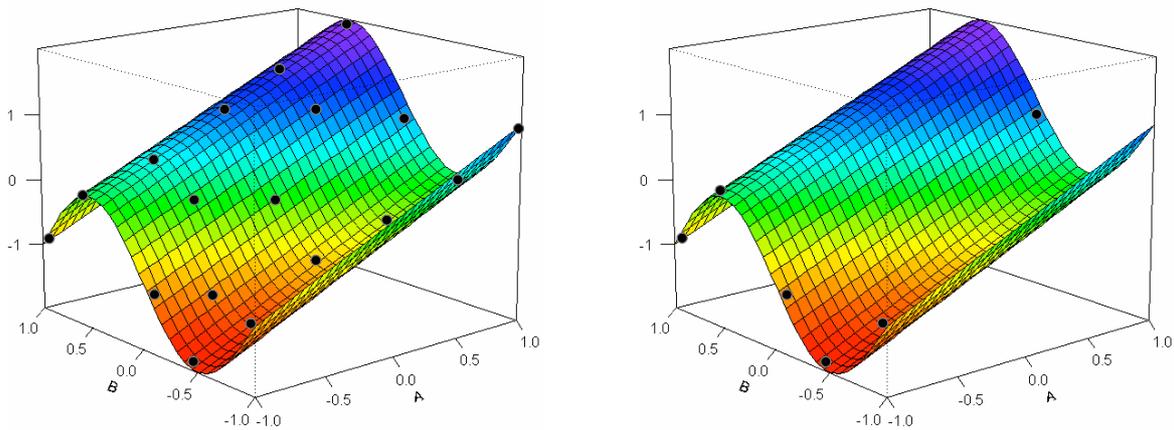


Figure 5.3 : Influence d'un paramètre linéaire et un paramètre non-régulier

Ainsi, la décomposition du problème entre les paramètres ayant un comportement régulier ou suffisamment bien caractérisé par l'approximation courante, et ceux de comportement non-régulier et non expliqué, permet de réduire le nombre de simulations nécessaires pour la construction d'un modèle approché de la réponse.

Ceci est classiquement effectué par des analyses de sensibilité, basées par exemple sur des méthodes issues de la régression, tel que l'analyse de variance ou les tests de Student. Dans la cas d'un échantillon plus important, des techniques d'analyse de sensibilité globales tel que la méthode de Sobol' ou la méthode Fast [56] peuvent être utilisées.

Une fois que le sous-espace du domaine expérimental nécessitant un enrichissement de simulations est défini, il faut fixer les conditions des nouvelles simulations afin d'accroître la prédictivité du modèle dans ce sous-espace.

5.3 Augmentation du Dispositif Expérimental

La modélisation d'une réponse non-régulière est un problème complexe, elle nécessite la détection des principales irrégularités de la réponse à estimer. De plus, le modèle approché a pour vocation d'être utilisé pour les prévisions de production, il doit donc être de

prédictivité maximale. La recherche des nouveaux points vise à optimiser ces deux critères. Ainsi, la recherche des points stationnaires permet une meilleure compréhension des non-régularités et l'ajout des points de sensibilité maximale ou des points de confirmation assure une bonne prédictivité du modèle.

5.3.1 Recherche d'extrema

Le premier critère d'ajout de simulation que nous proposons a pour objectif de détecter des extrema de la réponse. Ainsi, pour compléter le dispositif expérimental, nous allons procéder à la recherche des extrema de l'approximation courante.

Remarque : Nous présentons ici la méthodologie mise en place pour la recherche des maxima de la réponse, la recherche des minima étant réalisée selon le même principe.

La réponse en production étant supposée non-régulière, nous envisageons la recherche de plusieurs extrema. Par le biais de techniques d'optimisation locale, nous déterminons, pour plusieurs points initiaux, les maxima correspondants, i.e. la recherche est réalisée dans un voisinage de chaque point initial. Ce principe est illustré sur la Figure 5.4.

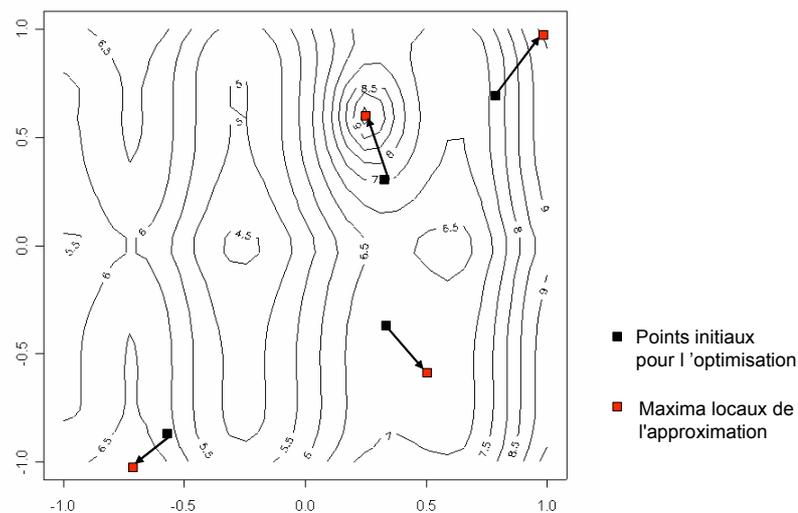


Figure 5.4 : Maximisation à partir de plusieurs points initiaux

A ce stade, il est important de justifier le choix de techniques d'optimisation locale et non globale. Dans notre cas, l'objectif, *in fine*, de cette recherche de maxima, n'est pas la détermination du maximum global de la réponse, les extrema locaux sont pour nous intéressants puisqu'ils nous informent sur les inflexions de la réponse.

Cette première phase de recherche des maxima est basée sur l'approximation pour le dispositif expérimental courant, puisque c'est cette surface de réponse qui est optimisée. Une confiance « *a priori* » est donc accordée à l'approximation courante.

Cette recherche initiale aboutit à la simulation d'un nouveau point, elle est ensuite complétée par une vérification de la pertinence de l'information ainsi obtenue. En effet, si la valeur du nouveau point issu de la maximisation permet la détection d'un nouveau maximum dans un voisinage de ce point, il se peut que ce nouveau maximum ne corresponde pas à un « véritable » maximum de la réponse et donc que la simulation d'un nouveau point soit nécessaire. La valeur de la réponse simulée au point maximisant l'approximation courante est ainsi d'intérêt capital dans la suite du processus de recherche des maxima de la réponse.

Plus précisément, en fonction de la valeur simulée de la réponse au point maximisant l'approximation courante, nous proposons :

- de stopper la recherche de maxima dans un voisinage du point simulé si la valeur de la réponse est inférieure à celle simulée au point initial. En effet, l'hypothèse de croissance constatée sur l'approximation est mise en défaut après simulation de la réponse. Il en est de même dans le cas d'une valeur simulée à peine plus grande que la valeur au point initial, l'accroissement de la réponse n'étant pas significatif.
- de confirmer la présence d'un maximum de la réponse dans ce voisinage, si la valeur de la réponse simulée est significativement supérieure à celle observée au point initial. En effet, si $f(\mathbf{x}_{max}) > f(\mathbf{x}_{init})$, nous proposons de simuler le point symétrique de \mathbf{x}_{init} par rapport à \mathbf{x}_{max} . Cette proposition est inspirée de la technique de recherche de maximum par méthode séquentielle du simplexe proposée initialement par Spendlay, Hext et Himsworth [66]. Elle permet, en ne se basant plus que sur les valeurs de la réponse et non sur l'approximation, d'étudier le comportement de la réponse pour une dérivée directionnelle positive.

Ainsi, comme l'illustre le schéma de la Figure 5.5, la recherche des maxima de la réponse est réalisée à chaque itération en plusieurs étapes :

- sélection de n^{opt} points initiaux pour l'optimisation,
- pour chaque point, recherche du maximum de l'approximation courante dans un voisinage du point initial,
- confirmation de la présence de maxima, par recherche dans la direction d'accroissement positif de la réponse :
 - arrêt de la recherche,
 - recherche dans la direction d'accroissement positif.

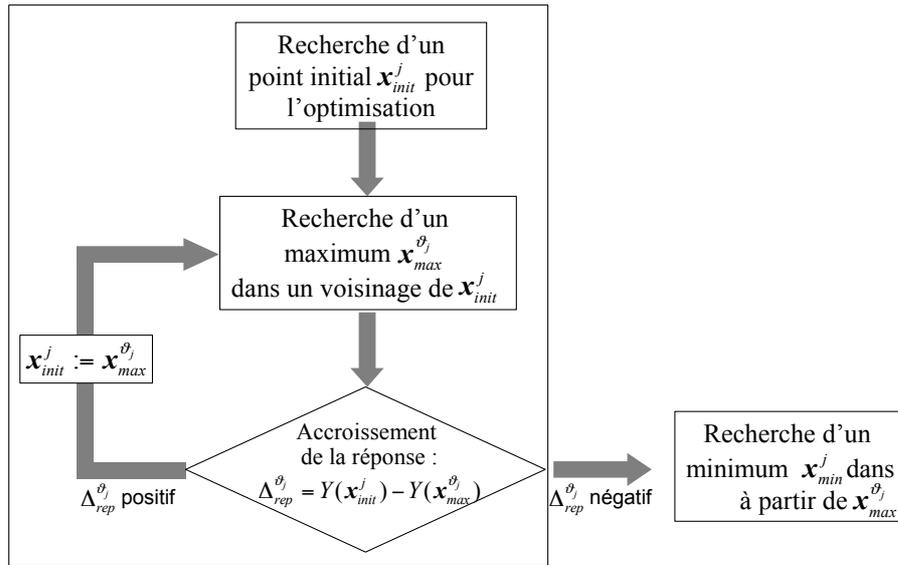


Figure 5.5 : Recherche des extrema de l'approximation : Principe

Les différentes étapes nécessaires à la mise en œuvre de cette technique de maximisation de la réponse sont détaillées dans la suite de cette section.

- Sélection des n^{opt} points initiaux pour l'optimisation

Lors de la procédure d'optimisation, il est nécessaire de définir des points dits initiaux, à partir desquels la recherche des optima sera effectuée.

Nous proposons de sélectionner ces points initiaux parmi les points du dispositif expérimental courant. Ainsi, notre critère de sélection est basé sur la valeur de la réponse aux points du plan. Les points initiaux sont alors déterminés parmi les points d'un ensemble C de points candidats comme les points de réponses les plus élevées.

Ces points vérifient donc :

$$\mathbf{X}_{init}^{opt} = \left\{ \mathbf{x}_{init}^j, j = 1, \dots, n^{opt}, \text{ tq } Y(\mathbf{x}) = \max_{\mathbf{x}_i \in C} \left(\hat{Y}(\mathbf{x}_i) \right) \right\}$$

L'ensemble C des points candidats est initialement le dispositif expérimental initial. Il est ensuite remis à jour à chaque itération. En effet, une fois la recherche d'un maximum à partir d'un des points \mathbf{x}_{init}^j réalisée, nous considérons le voisinage ϑ_j de \mathbf{x}_{init}^j comme exploré. Ce point ainsi que tous les points simulés lors de cette recherche de maximum sont donc supprimés de l'ensemble C des points candidats. Ainsi, au fur et à mesure du processus itératif, des nouvelles zones du domaine sont investiguées et de nouveaux maxima locaux déterminés.

- Recherche du maximum de l'approximation courante $\mathbf{x}_{max}^{\vartheta_j}$ dans un voisinage du point initial \mathbf{x}_{init}^j

L'approximation courante est alors maximisée dans un voisinage ϑ_j du point initial \mathbf{x}_{init}^j , c'est-à-dire que nous cherchons le point tel que :

$$\mathbf{x}_{max}^{\vartheta_j} = \arg \max_{\mathbf{x} \in \vartheta_j} (\hat{Y}(\mathbf{x})) \quad \text{où } \vartheta_j \text{ est un voisinage de } \mathbf{x}_{init}^j .$$

Le voisinage ϑ_j est défini comme un hypercube de taille 0.2 autour de \mathbf{x}_{init}^j , et si une borne est atteinte, on élargit le voisinage, comme illustré sur la Figure 5.6.

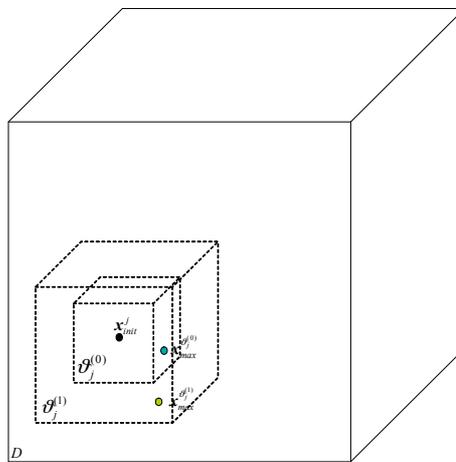


Figure 5.6 : Définition du voisinage ϑ_j

Une simulation est réalisée au point $\mathbf{x}_{max}^{\vartheta_j}$ trouvé lors de la maximisation de l'approximation courante, à condition que celui-ci ne soit pas trop proche d'un point du plan d'expériences courant. Cette dernière condition permet d'éviter des matrices de covariance singulières.

- Confirmation de la présence d'un maximum

A ce stade de la procédure de recherche de maxima, il est intéressant d'étudier l'information apportée par ce nouveau point. Pour ce faire, nous comparons la valeur de la réponse au point $\mathbf{x}_{max}^{\vartheta_j}$ avec celle au point \mathbf{x}_{init}^j : $\Delta_{rep}^{\vartheta_j} = Y(\mathbf{x}_{max}^{\vartheta_j}) - Y(\mathbf{x}_{init}^j)$.

- dans le cas où l'apport d'information $\Delta_{rep}^{\vartheta_j}$ est faible, nous pouvons supposer que nous sommes peu éloignés d'un maximum local de l'approximation courante, la réponse ne connaît pas un accroissement suffisamment important dans cette zone, il n'est pas nécessaire d'ajouter de l'information supplémentaire dans le voisinage du point \mathbf{x}_{init}^j .

- dans le cas où $\Delta_{rep}^{\theta_j}$ est négatif, nous sommes sur une pente ascendante de l'approximation, l'hypothèse de croissance étant mise en défaut, nous ne réalisons pas de nouvelle simulation.
- dans le cas où $\Delta_{rep}^{\theta_j}$ est positif et jugé non négligeable, une augmentation importante de la valeur de la réponse est observée, l'ajout de points dans cette zone est donc pertinent. De ce fait, nous proposons de simuler un nouveau point dans cette zone, afin de poursuivre la recherche de maximum.
 - Recherche de la direction d'accroissement positif

Dans le dernier cas, lorsque $\Delta_{rep}^{\theta_j}$ est jugé non négligeable, nous cherchons un nouveau point $\mathbf{x}_{new}^{\theta_j}$ à simuler dans la direction d'accroissement positif. Disposant de plusieurs valeurs de la réponse dans un même voisinage, nous proposons une recherche de maxima inspirée de la méthode « modified simplex » proposée par Nelder et Mead [40]. Ainsi, nous déterminons le nouveau point à simuler $\mathbf{x}_{new}^{\theta_j}$ comme le symétrique du point \mathbf{x}_{init}^j par rapport à \mathbf{x}_{max}^j et dans le cas d'un accroissement, une extension est réalisée dans la direction d'accroissement.

La Figure 5.7 illustre ce procédé, les coordonnées du nouveau point $\mathbf{x}_{new}^{\theta_j}$ sont données par : $\mathbf{x}_{new}^{\theta_j} = \mathbf{x}_{init}^j + (\mathbf{x}_{max}^j - \mathbf{x}_{init}^j)$.

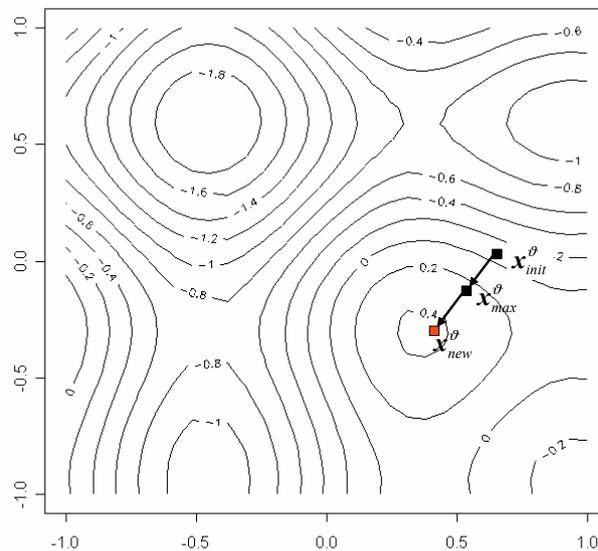


Figure 5.7 : Méthode « Iso-simplexe » : Emplacement du nouveau point à simuler

Une simulation d'écoulement est effectuée en ce point, et nous calculons de nouveau la quantité d'information apportée par cette nouvelle simulation : $\Delta_{rep}^{\theta_j} = Y(\mathbf{x}_{new}^{\theta_j}) - Y(\mathbf{x}_{max}^{\theta_j})$. Si cette valeur est significativement positive, la recherche des nouveaux points est ainsi poursuivie, selon le principe d'extension proposé par Nelder et Mead [40].

Cette procédure itérative, illustrée par la Figure 5.8, est conduite jusqu'à ce que le nouveau point n'apporte pas suffisamment d'information.

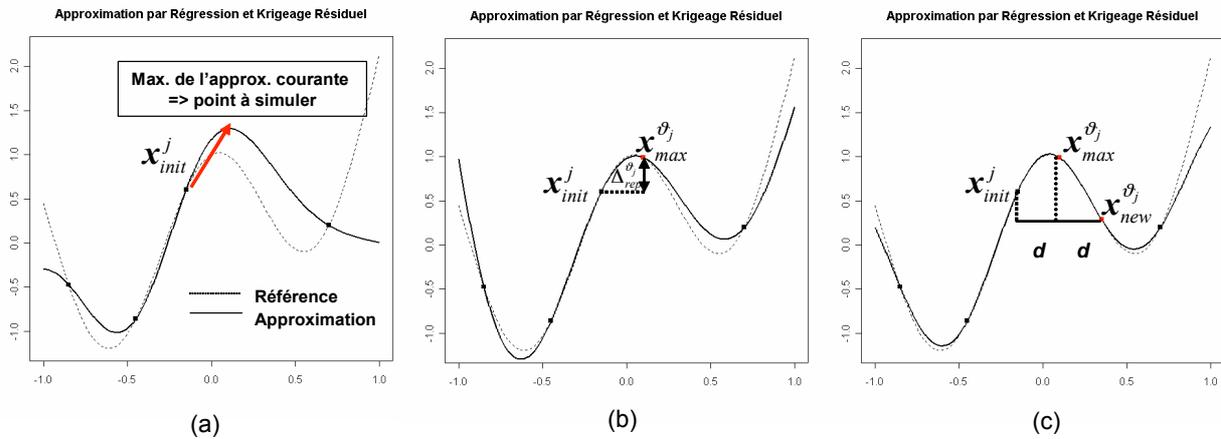


Figure 5.8 : Recherche d'un maximum

Dans le cas présenté sur la Figure 5.8, x_{init}^j est le point du dispositif expérimental de plus grande valeur de réponse. La procédure de maximisation mise en œuvre à partir du point x_{init}^j nous permet d'identifier un maximum local de l'approximation courante le point $x_{max}^{\theta_j}$, situé sur le graphique (b). La valeur de la réponse au point $x_{max}^{\theta_j}$ est plus importante que celle au point x_{init}^j . Par conséquent, nous recherchons le symétrique de x_{init}^j par rapport à $x_{max}^{\theta_j}$, privilégiant ainsi un déplacement dans la direction d'accroissement positif et nous obtenons le point $x_{new}^{\theta_j}$. Le processus de recherche est alors stoppé, car la valeur de $x_{new}^{\theta_j}$ est plus faible que celle de $x_{max}^{\theta_j}$.

- Recherche d'un minimum dans un voisinage du maximum

La procédure de recherche de maxima permet une détermination efficace des valeurs extrêmes de la réponse étudiée, ainsi qu'une caractérisation locale précise du comportement de la réponse. Toutefois, le principe même de cette recherche, basée sur une recherche par voisinage, aboutit à un échantillonnage dense voir à un sur-échantillonnage dans la zone d'étude. C'est pourquoi, à l'issue de cette recherche locale des maxima, nous préconisons une recherche de minima dans le domaine.

- Recherche de minima

Dans une optique de recherche des extrema de la réponse, et pour favoriser l'échantillonnage de zones peu explorées, nous proposons la recherche de minima de l'approximation courante, en considérant pour point initial de cette optimisation les maxima déterminés au préalable.

Ainsi, soit $\mathbf{x}_{max}^{\theta_j}$ un maximum identifié de la réponse. Nous cherchons, en considérant comme point initial le point $\mathbf{x}_{max}^{\theta_j}$, le point tel que \mathbf{x}_{min}^j :

$$\mathbf{x}_{min}^j = \arg \min_{\mathbf{x} \in D} \hat{Y}(\mathbf{x})$$

où D est le domaine expérimental.

Comme l'illustre la Figure 5.9, la procédure de recherche des extrema de la réponse est itérative, et exploite d'une part la connaissance de l'approximation courante, et d'autre part, les valeurs de la réponse dans le voisinage exploré.

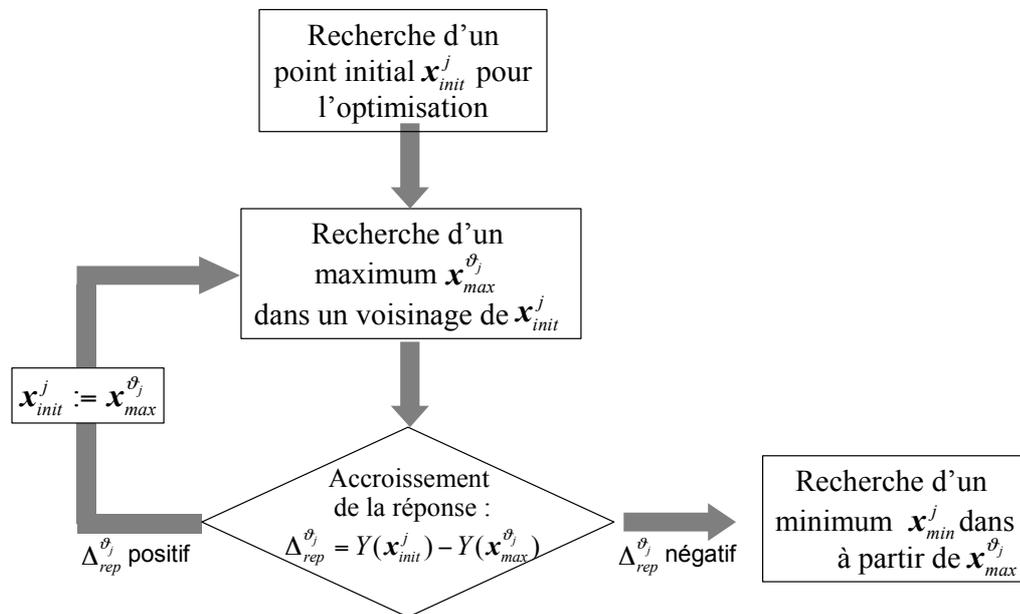


Figure 5.9 : Recherche des extrema

Cette recherche des extrema de la réponse est capitale puisqu'elle permet, *in fine*, de caractériser la « sinuosité » de la réponse et la complexité de sa variabilité. Cependant, ce seul critère d'ajout de points ne permet pas de décrire l'ensemble des caractéristiques variationnelles d'une réponse.

Ainsi, nous proposons d'exposer une recherche des points d'inflexion de la réponse.

5.3.2 Recherche des points stationnaires

Afin de caractériser au mieux les variations de la réponse, nous proposons la recherche de points stationnaires. Ne disposant généralement pas de valeurs simulées des dérivées partielles de la réponse par rapport aux paramètres, cette recherche sera dans un premier temps basée sur l'étude de l'approximation courante, une phase de validation complétant cette recherche dans un deuxième temps.

Ainsi, nous proposons d'ajouter, à chaque itération, le point pour lequel au moins une des l dérivées partielles de l'approximation courante s'annule, l étant la dimension du sous-espace des paramètres considéré. L'approximation de la réponse étant de plus en plus précise au fur et à mesure des itérations, les points ainsi déterminés permettront de caractériser des extrema ou points d'inflexion de la réponse.

Nous proposons de calculer la dérivée analytique du modèle composite de la réponse obtenue à l'aide du plan d'expériences courant. La structure composite du modèle permet en effet ce calcul analytique. La régression est dérivable sur l'ensemble du domaine et le terme de krigeage possède des propriétés de dérivabilité qui ne dépendent que du variogramme. Généralement, le second terme du modèle est dérivable partout, sauf aux points du plan d'expériences.

Ainsi, la dérivée de \hat{Y} en tout point \mathbf{x} de l'espace est donnée par :

$$\frac{\partial \hat{Y}}{\partial X_i}(\mathbf{x}) = \frac{\partial \mathbf{X}}{\partial X_i}(\mathbf{x})\hat{\boldsymbol{\beta}} + \frac{\partial \mathbf{r}}{\partial X_i}(\mathbf{x})\mathbf{R}^{-1}[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}], \quad i = 1, \dots, l \quad (5.1)$$

Sans perte de généralité, la recherche des zéros de la dérivée $\frac{\partial \hat{Y}}{\partial X_i}$ peut être formalisée comme un problème de minimisation :

$$\mathbf{x}_g = \left\{ \mathbf{x}, \exists X_i, i = 1, \dots, l, \quad \text{t.q.} \quad \mathbf{x} = \arg \min_{\mathbf{x} \in D} \left(\frac{\partial \hat{Y}(\mathbf{x})}{\partial X_i} \right)^2 \right\} \quad (5.2)$$

La dépendance de la réponse aux paramètres n'étant *a priori* pas isotrope, nous chercherons à annuler au moins une des dérivées partielles de la réponse.

La mise en œuvre pratique de la recherche des points \mathbf{x}_g est réalisée en considérant n points initiaux pour la minimisation du carré de la dérivée, comme l'illustre la Figure 5.10. Ce choix est motivé par la non-régularité *a priori* de la réponse. En effet, nous nous plaçons dans un contexte où l'unicité de \mathbf{x}_g n'est pas assurée. Nous proposons donc l'utilisation de techniques d'optimisation locale à partir de plusieurs points initiaux.

Comme dans le cas de la recherche des extrema, le choix des points initiaux est essentiel. Pour éviter un sur-échantillonnage local, nous sélectionnons, dans une grille uniforme, les points possédant une grande dérivée, donc *a priori* éloignés des points recherchés, afin de parcourir plus grandement le domaine expérimental.

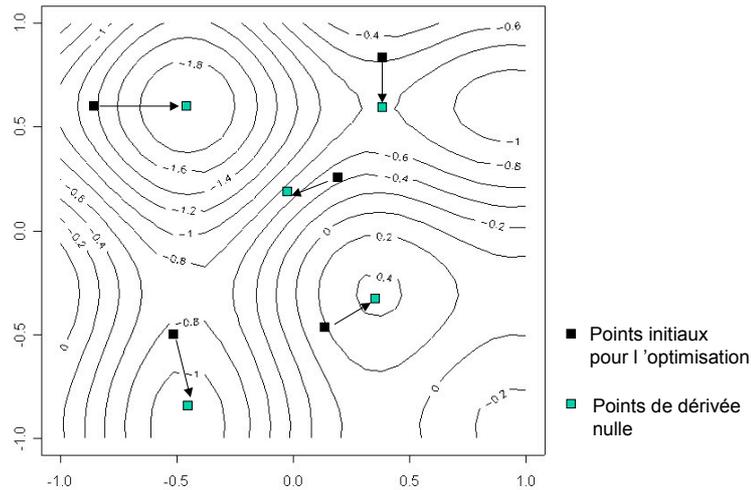


Figure 5.10 : Optimisation de n valeurs de dérivées

Remarque 1 : Dans le cas d'un variogramme non gaussien, le modèle composite n'est pas dérivable aux points du plan d'expériences. Dans ce cas, nous proposons de ne pas prendre des points du plan d'expériences comme points initiaux pour l'optimisation, mais de prendre un point du domaine dans un voisinage de ces points.

Remarque 2 : Cette approche, possible pour le modèle composite, ne peut être mise en pratique dans le cas d'un ajustement par krigeage ordinaire avec voisinages glissants. En effet, du fait des voisinages, ce type d'approximation n'est pas dérivable.

Remarque 3 : Le point \mathbf{x}_g ayant la plus petite valeur de dérivée n'est simulé qu'à condition qu'il soit suffisamment éloigné des autres points du dispositif expérimental. En effet, comme nous l'avons déjà mentionné, après un certain nombre d'itérations, le modèle d'approximation devient plus précis, les extrema et points d'inflexions locaux sont localisés. La recherche des points de dérivées partielles nulles pourrait amener à un sur-échantillonnage de certaines zones. De plus, dans le cas où le point obtenu est trop proche d'un point existant, ce point n'est pas simulé afin d'éviter des matrices de covariances singulières.

La détermination de \mathbf{x}_g étant basée sur l'approximation courante, il est essentiel de passer à une phase de validation, visant à confirmer les propriétés de stationnarité directionnelle de \mathbf{x}_g .

Ainsi, un nouveau modèle d'approximation, intégrant l'information apportée par la simulation en \mathbf{x}_g , est construit.

La valeur de la dérivée directionnelle en \mathbf{x}_g est alors calculée à partir de cette nouvelle approximation :

$$\hat{Y}'^i(\mathbf{x}_g) = \lim_{\varepsilon \rightarrow 0} \frac{\hat{Y}(\mathbf{x}_g + \varepsilon) - \hat{Y}(\mathbf{x}_g)}{\varepsilon}$$

avec $\boldsymbol{\varepsilon} = (0, \dots, 0, \varepsilon_j, 0, \dots, 0)$.

Le recours à une dérivée numérique est ici nécessaire puisque la dérivée analytique du modèle d'approximation n'est pas définie aux points du plan d'expériences.

Cette valeur de la dérivée \mathbf{x}_g est alors étudiée, dans le but de rechercher des points stationnaires de la réponse Y :

- si $\hat{Y}^i(\mathbf{x}_g) \square 0$, la recherche de points stationnaires à cette itération est achevée avec la détermination de \mathbf{x}_g ,
- si $\hat{Y}^i(\mathbf{x}_g) \neq 0$, le point n'est pas stationnaire pour la nouvelle approximation. Un nouveau point stationnaire est recherché à partir de \mathbf{x}_g .

La recherche des points de dérivées partielles nulles est une étape essentielle de la modélisation de fonctions non-régulières. Elle permet en effet la détection des extrema et des points d'inflexion de la réponse, facilitant ainsi la modélisation de la variabilité de la réponse sur le domaine incertain. Une illustration du principe d'ajout de points à dérivée partielle nulle est présentée sur la Figure 5.11. Dans cet exemple, un seul point de dérivée nulle est ajouté, tant que sa valeur de dérivée après simulation n'est pas nulle. Nous pouvons voir qu'un point d'inflexion de l'approximation a été identifié.

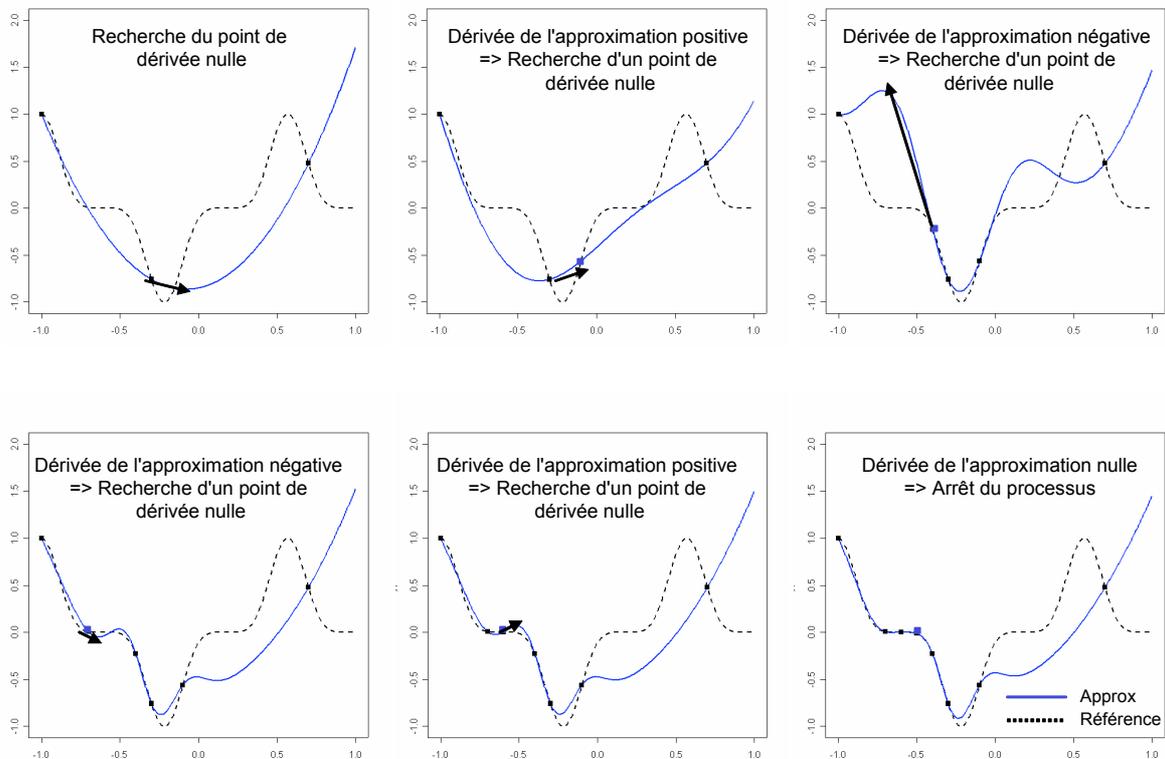


Figure 5.11 : Recherche d'un point d'inflexion

Les deux critères présentés jusqu'ici, à savoir la recherche d'extrema et la recherche des points stationnaires, sont importants pour la modélisation du comportement de la réponse. Toutefois, comme ils sont basés sur une optimisation de l'approximation courante, ils peuvent amener à intensifier l'échantillonnage de zones déjà explorées, négligeant dans le même temps des zones du domaine pour lesquelles aucune information n'est connue. Ainsi, nous proposons d'étudier le comportement de l'approximation courante aux points maximisant la variance de krigeage. Ces points sont en effet les plus éloignés spatialement des points déjà simulés.

5.3.3 Sensibilité du modèle à l'apport d'information dans les zones non-échantillonnées

Le critère introduit ci-dessous ne vise pas le même objectif que les critères précédents. En effet, les critères précédents visaient à améliorer localement l'approximation dans les zones du domaine expérimental déjà échantillonnées pour mieux caractériser les inflexions de la réponse. Ici nous proposons un critère permettant d'initialiser l'échantillonnage dans les zones jusqu'alors non-explorées.

Ce critère est basé sur la méthode des points pilotes présentée dans le chapitre 4. L'ajout de ces points visait à améliorer la prédictivité du modèle dans les zones non-explorées par le plan d'expériences courant. Dans ce chapitre, quelques éléments nouveaux sont introduits à l'utilisation des points pilotes. En effet, afin de sélectionner au mieux les simulations à réaliser dans les zones sous-échantillonnées du domaine expérimental, nous proposons d'étudier *a priori* l'influence de l'apport d'information dans ces zones par l'ajout de points pilotes. Les points pilotes vont ainsi déterminer les zones les plus sensibles et pour lesquelles une simulation est nécessaire.

Notre stratégie comporte donc deux étapes :

- détermination des zones sous-échantillonnées par l'étude de la fonction de variance de krigeage des résidus J_I , introduite dans le chapitre 4. Cette phase aboutit ainsi à la définition d'une série de points candidats à la simulation, ces points maximisant la fonction de variance J_I . L'ensemble des points candidats à la simulation est désigné par C^{PP} .
- étude de la sensibilité de l'approximation à l'apport d'information en ces points. Pour ce faire, pour chacun des points candidats à la simulation \mathbf{x}_i^{PP} , nous proposons d'attribuer une valeur fictive Y_i^{PP} à la réponse, que nous faisons varier dans un intervalle de confiance approximatif. Si la variation de la valeur fictive Y_i^{PP} a un impact important sur l'approximation, le point est simulé, sinon une valeur fictive lui est attribuée.

- Détermination des zones sous-échantillonnées

Nous considérons ici que nous sommes à une itération donnée de la modélisation adaptative. A ce stade de la modélisation, nous proposons d'ajouter n_p nouveaux points pilotes au dispositif expérimental. L'emplacement de ces points dans le domaine expérimental D est choisi afin de maximiser la variance de krigeage (cf. chapitre 4) :

$$J_1 = \max_{x \in D} \left\{ \tilde{\sigma}^2 \left(1 - {}^t r(x) R^{-1} r(x) \right) \right\} \quad (5.3)$$

Avec :

- $\tilde{\sigma}^2 = s^2 = \frac{1}{n-p} \sum_{i=1}^n (Y(x_i) - \bar{Y})^2$ dans le cas d'une estimation à l'aide d'une loi *a priori* non informative de Jeffreys.
- $\tilde{\sigma}^2 = \frac{1}{n-p} {}^t (Y - X\hat{\beta}) R^{-1} (Y - X\hat{\beta})$ pour une estimation de σ^2 par maximum de vraisemblance restreint, avec $p = \text{rang}(X)$ et n le nombre d'observations.

Ces nouveaux points sont potentiellement informatifs car ils sont, du fait du critère J_1 , situés dans les zones non-échantillonnées du domaine. En effet, comme illustré sur la Figure 5.12, la variance de krigeage est nulle aux points du dispositif expérimental et maximale dans les zones non-échantillonnées.

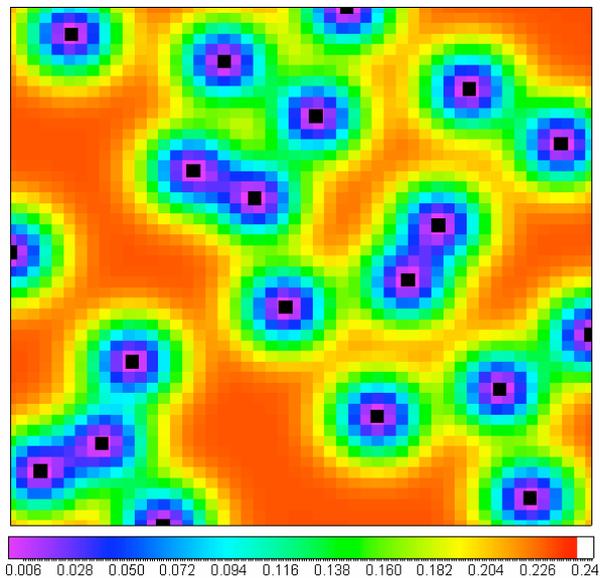


Figure 5.12 : Variance de krigeage dans le domaine expérimental

- Sensibilité de l'approximation à l'apport d'information

Dès lors que l'emplacement x_i^{PP} des points candidats à être simulés est déterminé, nous étudions l'impact d'un ajout d'information en ces points. Nous proposons d'attribuer une

valeur fictive Y_i^{PP} de réponse au point fictif \mathbf{x}_i^{PP} , puis de perturber cette valeur pour étudier la sensibilité de l'approximation à un apport d'information aux points \mathbf{x}_i^{PP} . Ainsi, nous identifions les points pour lesquels un apport d'information pourrait changer significativement le comportement de l'approximation, notamment en terme de prédictivité.

Toujours en nous basant sur les résultats du chapitre 4, pour tout point \mathbf{x}_i^{PP} de l'ensemble C^{PP} , l'intervalle de confiance approximatif que peut prendre la valeur de la réponse fictive est donné par (4.12) :

$$Y_i^{PP} \in I_i^{PP} = \left[\hat{Y}(\mathbf{x}_i^{PP}) - q_\alpha \sqrt{\tilde{\sigma}^2 (1 - {}^t \mathbf{r}(\mathbf{x}_i^{PP}) \mathbf{R}^{-1} \mathbf{r}(\mathbf{x}_i^{PP}))}; \hat{Y}(\mathbf{x}_i^{PP}) + q_\alpha \sqrt{\tilde{\sigma}^2 (1 - {}^t \mathbf{r}(\mathbf{x}_i^{PP}) \mathbf{R}^{-1} \mathbf{r}(\mathbf{x}_i^{PP}))} \right]$$

avec :

- $\hat{Y}(\mathbf{x}_i^{PP}) = \mathbf{X} \hat{\boldsymbol{\beta}} + {}^t \mathbf{r}(\mathbf{x}_i^{PP}) \mathbf{R}^{-1} \mathbf{Z}$ l'estimation de la valeur de la réponse au point \mathbf{x}_i^{PP} par le modèle composite,
- $q_\alpha = t_{1-\alpha/2}$ désigne le quantile d'ordre $1-\alpha/2$ de la variable aléatoire de Student à n degrés de liberté dans le cas d'un *a priori* de Jeffreys sur σ^2 .

Nous proposons de sélectionner, dans cet ensemble de points C^{PP} , tous les points \mathbf{x}_i^{PP} qui soient les plus informatifs vis-à-vis de la prédictivité et qui permettent, de ce fait, une amélioration de cette dernière. Pour ce faire, nous calculons la valeur du coefficient Q_2 en fonction des valeurs de la réponse fictive Y_i^{PP} dans son intervalle de variation I_i^{PP} .

Le coefficient de prédictivité Q_2 mesure la qualité de prédictivité de l'approximation courante. Une valeur de Q_2 proche de 1 signifie que l'approximation est prédictive, et au contraire, une faible valeur de Q_2 indique une mauvaise stabilité de l'approximation. Suivant les valeurs du coefficient Q_2 et de sa variation $\Delta Q_2 = \max(Q_2) - \min(Q_2)$ en fonction des valeurs Y_i^{PP} attribuées à chaque point fictif \mathbf{x}_i^{PP} , il peut être envisagé, comme le schématise la Figure 5.13, de :

- Simuler le point \mathbf{x}_i^{PP} ,
- Maintenir le point \mathbf{x}_i^{PP} dans le dispositif courant, en tant que point fictif et procéder à l'optimisation de la valeur fictive Y_i^{PP} en vue d'améliorer la prédictivité de l'approximation courante,
- Supprimer le point \mathbf{x}_i^{PP} , si les variations de Y_i^{PP} n'induisent pas une grande modification de l'approximation.

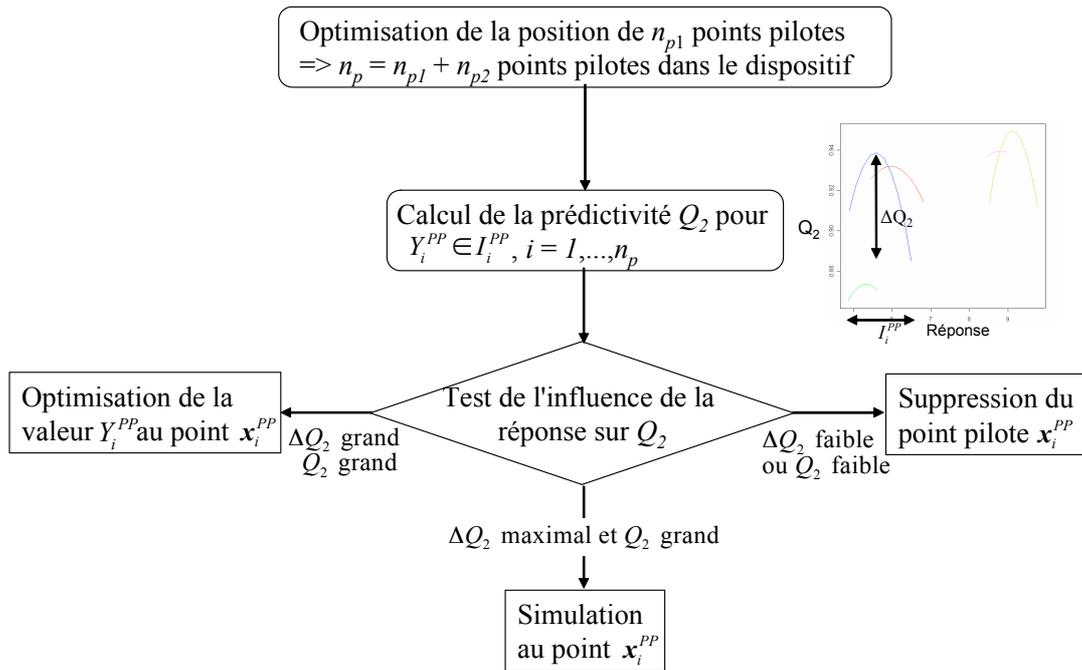


Figure 5.13 : Traitement des différents points pilotes

Ces différentes possibilités sont proposées en fonction de la valeur du coefficient de prédictivité Q_2 et de la variation de cette statistique dans l'intervalle de valeurs I_i^{PP} de la réponse. Nous proposons donc de procéder de la manière suivante, pour chaque $\mathbf{x}_i^{PP} \in C^{PP}$:

- Simulation du point \mathbf{x}_i^{PP} pour lequel la variation du coefficient Q_2 est maximale. L'ajout d'information réelle est justifiée ici par le fait qu'une variation de la réponse fictive au point \mathbf{x}_i^{PP} entraîne une grande variation de la prédictivité. Ainsi, l'approximation est instable dans un voisinage du point \mathbf{x}_i^{PP} , ce point est donc le plus informatif vis-à-vis de la prédictivité, d'où la décision de le simuler.
- Optimisation de la valeur Y_i^{PP} de la réponse au point \mathbf{x}_i^{PP} dans le cas d'une bonne prédictivité, typiquement un coefficient $Q_2 \geq 0.7$, et d'une variation non-négligeable de ce coefficient, c'est-à-dire supérieure à $\Delta Q_2 \geq 0.05$. Dans ce cas, l'approximation est déjà de bonne qualité dans un voisinage du point \mathbf{x}_i^{PP} et nous souhaitons maximiser cette dernière en choisissant pour valeur Y_i^{PP} , la valeur correspondant au maximum du coefficient de prédictivité Q_2 , comme présenté dans le chapitre 4.
- Suppression du point \mathbf{x}_i^{PP} du dispositif expérimental dans les autres cas. En effet, dès lors que la valeur du coefficient Q_2 est quasi-constante et élevée, le point \mathbf{x}_i^{PP} ne permet pas une amélioration significative de la prédictivité. De plus, dans le cas d'une mauvaise prédictivité de l'approximation, il est risqué d'ajouter de l'information

fictive, puisque celle-ci est fondée sur l'approximation courante peu fiable. Nous préférons dans ce cas poursuivre la recherche de points d'inflexion ou d'extrema.

Nous donnons un exemple, sur la Figure 5.14, des variations du coefficient de prédictivité Q_2 pour différents points pilotes, représentés en rouge, et du traitement qui leur est associé.

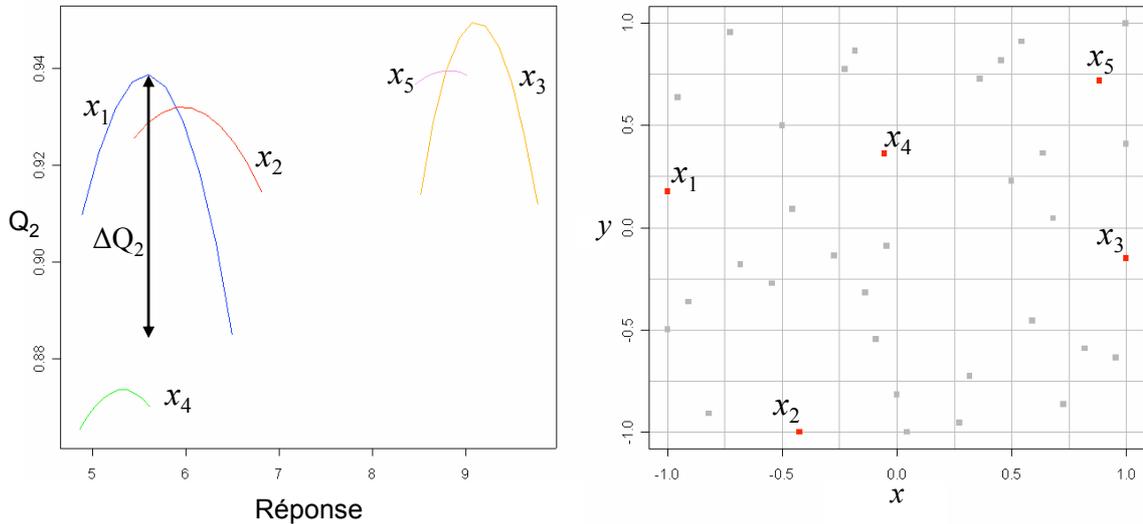


Figure 5.14 : Valeur de la statistique Q_2 pour différentes valeurs fictives

Ainsi, à chaque itération, certains points pilotes sont directement simulés (x_1), certains sont optimisés (x_2 et x_3) et d'autres sont supprimés du plan d'expériences (x_4 et x_5). Cette technique d'ajout de points permet de déterminer les zones les plus sensibles à une variation de la réponse en terme de prédictivité et pour lesquelles une variation de la réponse influe énormément sur la qualité d'approximation.

Remarque : Dès lors que le nombre de paramètres augmente, il est judicieux de simuler, à chaque itération, plusieurs points pilotes de grande influence sur la prédictivité de l'approximation courante.

Ainsi, la méthodologie d'ajout de points pilotes a un double enjeu dans cette thèse. D'une part, comme exposé au chapitre 4, elle nous permet de contrôler la prédictivité *a priori* de l'approximation et donc, à nombre de simulations fixé, d'obtenir une meilleure approximation. D'autre part, elle permet de déterminer des zones du domaine pour lesquelles l'approximation est instable en terme de prédictivité, justifiant ainsi la réalisation d'une simulation numérique d'écoulement supplémentaire. L'ajout de ce type de points a donc une importance cruciale, puisqu'il permet de stabiliser l'approximation.

Dès lors que la prédictivité *a priori* de l'approximation courante est optimale, une évaluation de la prédictivité *a posteriori* du modèle courant peut être réalisée afin de juger de la qualité de celui-ci. Cette évaluation passe par l'ajout de nouvelles simulations dites simulations de confirmation.

5.3.4 Évaluation de la prédictivité du modèle : Ajout de Simulations de Confirmation

Supposons à présent que le modèle courant ajuste la réponse de manière satisfaisante. Ainsi, les phases de recherche d'extrema et de points d'inflexion n'ont pas aboutit à la simulation de nouveaux points.

A ce stade de la modélisation, nous avons besoin de définir un critère permettant de juger de la nécessité d'une amélioration du modèle par l'ajout de nouveaux points au plan courant. Ce critère doit être fonction de la qualité de prédiction du modèle, car c'est principalement celle-ci qui nous intéresse.

Nous proposons ainsi de vérifier la prédictivité du modèle selon deux critères :

- calcul d'une prédictivité *a priori* sur un échantillonnage restreint,
- calcul d'une prédictivité *a posteriori* après simulation de points de confirmation.

L'évaluation de la prédictivité *a priori* que nous considérons consiste à effectuer une validation croisée en certains points du plan tirés aléatoirement. Il s'agit de calculer un coefficient Q_2 pour lequel seuls quelques points du dispositif courant sont utilisés. Nous ne prenons qu'un échantillon du plan d'expériences pour des raisons de coût. Ce coefficient est défini par :

$$Q_2^{prior} = 1 - \frac{\sum_{i \in I} (Y_i - \hat{Y}_{(i)})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

avec :

- I un tirage aléatoire de certains points du plan d'expériences,
- $\hat{Y}_{(i)}$ l'estimation au point supprimé x_i par le modèle composite ajusté sans le point x_i .

Dès lors que ce coefficient confirme la potentielle qualité du modèle d'approximation, une évaluation de la prédictivité *a posteriori* est réalisée avec l'ajout de points de confirmation.

L'ajout de ces points va permettre de conclure quant à la qualité prédictive du modèle d'approximation. Dans ce cas le processus d'échantillonnage évolutif est interrompu. Dans le cas contraire, on procède à une ou plusieurs itérations supplémentaires.

L'emplacement des points de confirmation doit être choisi de manière à détecter les éventuels défauts d'ajustement ou de prédiction de l'approximation. Par conséquent, il semble judicieux de placer ces points dans des zones non-échantillonnées, c'est-à-dire là où l'on ne

dispose que de peu d'information réelle sur la réponse. Pour ce faire, nous proposons de retenir les points pour lesquels la variance de krigeage est maximale.

Suivant le nombre de paramètres, le nombre de points de confirmation peut être plus ou moins important. Le nombre de points de confirmation a une importance non négligeable sur les résultats de la modélisation. C'est en effet sur la base de ces points que l'on va juger l'approximation courante comme satisfaisante ou non.

Lorsque l'emplacement de ces nouveaux points est déterminé, les simulations sont effectuées. Nous comparons alors les valeurs des simulations avec les valeurs prédites par le modèle courant. Nous calculons, un coefficient de type R^2 évaluant la capacité de l'approximation courante à prédire la réponse aux points de confirmation :

$$Q_2^{post} = 1 - \frac{\sum (Y^{conf} - \hat{Y}^{conf})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

avec :

- \mathbf{x}^{conf} les n^{conf} points de confirmation ajoutés,
- Y^{conf} la valeur de la réponse obtenue via le simulateur aux points de confirmation \mathbf{x}^{conf} ,
- \hat{Y}^{conf} la valeur de la réponse prédite par le modèle courant aux points de confirmation.

S'il y a adéquation entre les simulations et les prédictions du modèle courant, c'est-à-dire que $Q_2^{post} \square 1$, le modèle est jugé satisfaisant et le processus itératif est stoppé. C'est cette dernière approximation qui sera utilisée comme modèle approché du simulateur d'écoulement. Dans le cas contraire, un nouveau modèle est construit en intégrant l'information issue des simulations de confirmation. Le processus itératif reprend donc la recherche d'un extrema, d'un point qui annule la dérivée de cette nouvelle approximation ou l'ajout d'information fictive, comme illustré sur la Figure 5.15.

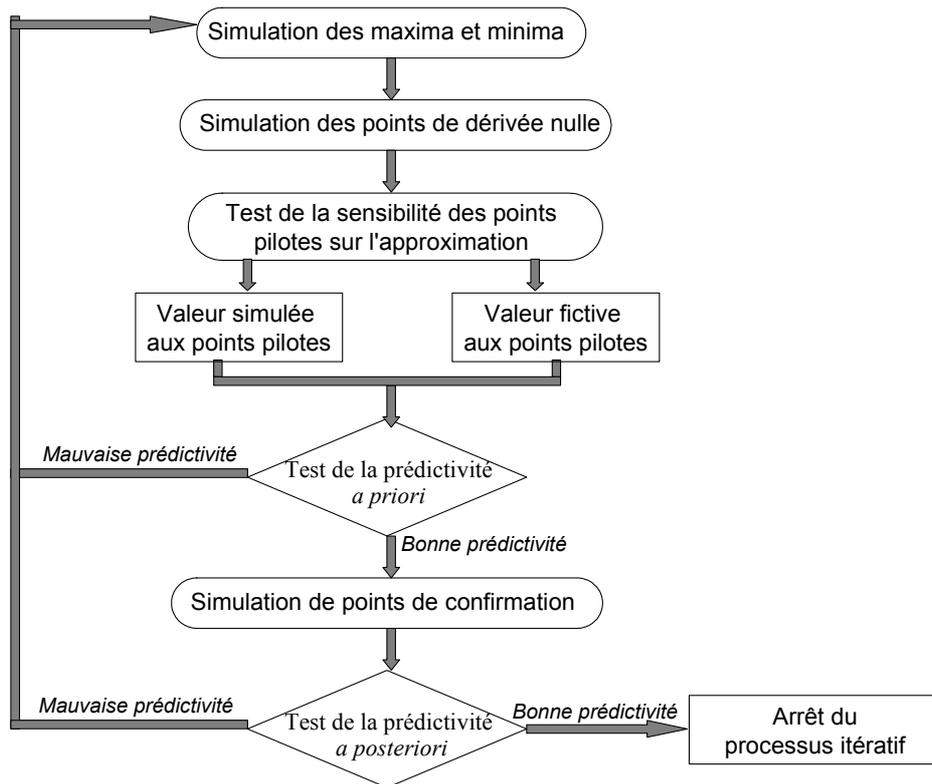


Figure 5.15 : Principe d'enrichissement du dispositif expérimental

Ce processus itératif est réalisé tant que la qualité de l'approximation n'est pas jugée satisfaisante.

5.4 Conclusions

Dans ce chapitre, nous avons montré que le processus itératif de modélisation adaptative peut être envisagé à partir d'un plan d'expériences initial de petite dimension. Nous proposons l'utilisation d'un hypercube latin de distance maximin du fait de ses bonnes propriétés de répartition dans le domaine incertain. Afin d'améliorer la qualité de l'approximation, de nouvelles simulations sont ajoutées successivement.

Nous proposons ici différents critères d'ajout de nouvelles simulations. Chaque critère est essentiel dans l'approche adaptative, chacun ayant des propriétés spécifiques et utiles. Ainsi, l'ajout des points minimum et maximum va déterminer les extrema locaux de la réponse, l'ajout des points de dérivée nulle va permettre une bonne détection des points d'inflexion de la réponse. De plus, la simulation du point de plus grande influence sur le coefficient de prédictivité Q_2 permet de stabiliser l'approximation dans des zones peu échantillonnées et très variables et donc de maximiser la prédictivité *a priori* de l'approximation. Enfin, les points de confirmation vont juger la qualité de la modélisation et diagnostiquer l'arrêt du processus itératif.

Cette technique d'enrichissement du dispositif est novatrice, l'emplacement des nouveaux points est déterminé à partir de l'approximation courante, l'échantillonnage est ainsi spécifique à chaque réponse étudiée.

Afin d'illustrer par des exemples les critères ci-dessus, et particulièrement l'enrichissement du plan d'expériences, le chapitre suivant est consacré à la présentation d'exemples d'application sur des fonctions à plusieurs variables.

Chapitre 6

Modélisation adaptative appliquée à des fonctions types

Dans ce chapitre, nous présentons plusieurs exemples d'estimation de fonctions analytiques non-régulières, modélisées par approche évolutive et plans d'expériences adaptatifs. Tout d'abord, nous présentons toutes les étapes de l'ajustement d'une fonction non-régulière. Ensuite, nous validons cette approche de modélisation adaptative à l'aide de plusieurs exemples en partant de différents dispositifs expérimentaux initiaux. Enfin nous testons l'importance de chaque critère d'ajout de points dans la modélisation. Nous comparons également les approximations obtenues par la méthode de modélisation adaptative à celles obtenues par un hypercube latin ou par un plan complet N^k comportant un nombre comparable de simulations.

6.1 Démarche et notations

Pour chacune des fonctions analytiques non-régulières étudiées, nous présentons différents graphes permettant de juger de la qualité de l'approximation. Ainsi, nous illustrons les résultats par :

- une représentation en perspective de la fonction réponse de référence à ajuster en fonction des paramètres,
- le plan d'expériences final, obtenu par enrichissement successif selon les critères d'ajout de points présentés dans le chapitre 5. Ce graphe permet de situer dans le domaine expérimental, les simulations effectuées tout au long du processus. Chaque type de points : minima, maxima, points de dérivée partielle nulle, point pilote est alors représenté par une couleur spécifique,
- une superposition des iso-contours dans l'espace des paramètres de la fonction de référence, représentée en noir, et de l'approximation finale, représentée en bleue,

- un graphe présentant les prédictions obtenues grâce au modèle d'approximation final comparées aux valeurs de référence, afin d'examiner les propriétés de l'approximation en terme de prédiction. Ce graphe permet de juger de la qualité prédictive, une approximation parfaite correspondant à la première bissectrice.

Ce dernier graphe est obtenu à partir des points d'une grille 30^2 , soit 900 points. Sur ce graphe, nous reportons également la valeur de l'erreur relative ε_{rel} , que nous définissons par :

$$\varepsilon_{rel} = \frac{\sum_{i=1}^n |Y_i - \hat{Y}(\mathbf{x}_i)|}{\sum_{i=1}^n |Y_i|}$$

Dans le cas d'un ajustement parfait, cette erreur est nulle. La valeur de l'erreur relative va nous permettre, d'une part, de juger la qualité de l'approximation et d'autre part, de comparer les qualités des approximations obtenues, pour une même fonction de référence, à partir de différents plans d'expériences ou différents types de points ajoutés.

Dans cette section, nous commençons par détailler les différentes étapes de l'approximation d'une fonction non-régulière, avec les résultats obtenus à chaque itération, puis nous présentons les résultats d'ajustement pour plusieurs exemples.

6.2 Ajustement de la fonction non-régulière « Chameau »

Cette section est consacrée à la présentation détaillée des résultats de modélisation de la fonction « Chameau ». Les différentes fonctions sont nommées suivant une terminologie classique utilisée en modélisation. Nous présentons en effet les résultats d'approximation à chaque itération du processus d'enrichissement. L'expression de la fonction dite « Chameau » est la suivante :

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

Une représentation en perspective de cette fonction dans le cube unité $[-1,1]^2$ est proposée sur la Figure 6.1 (a). La Figure 6.1 (b) illustre les résultats d'un ajustement par un modèle de régression du second ordre, obtenu à partir du plan d'expériences initial. Le recours à la modélisation adaptative est donc justifié, un polynôme de degré 2 ne permettant pas l'obtention d'un prédicteur efficace.

Nous présentons, sur la Figure 6.2, l'évolution au cours des itérations du plan d'expériences courant, des iso-contours de l'approximation, ainsi que le pouvoir de prédiction associé.

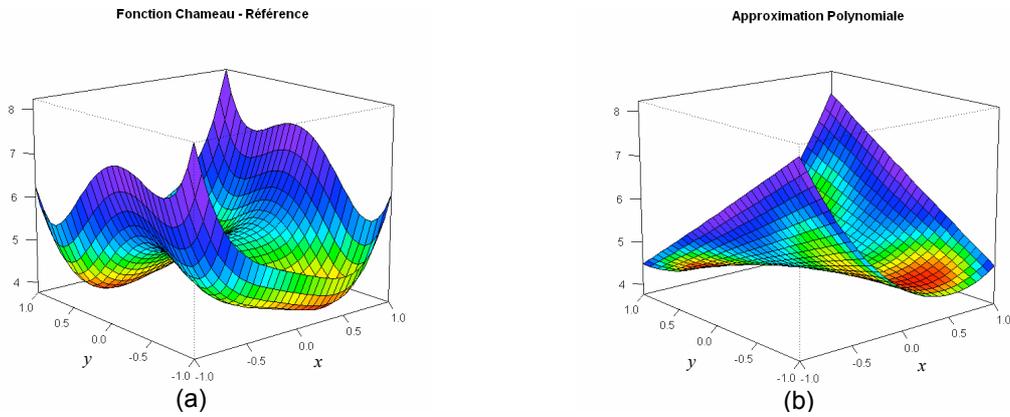
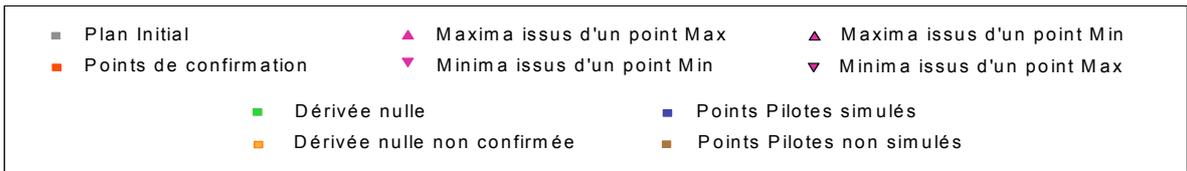
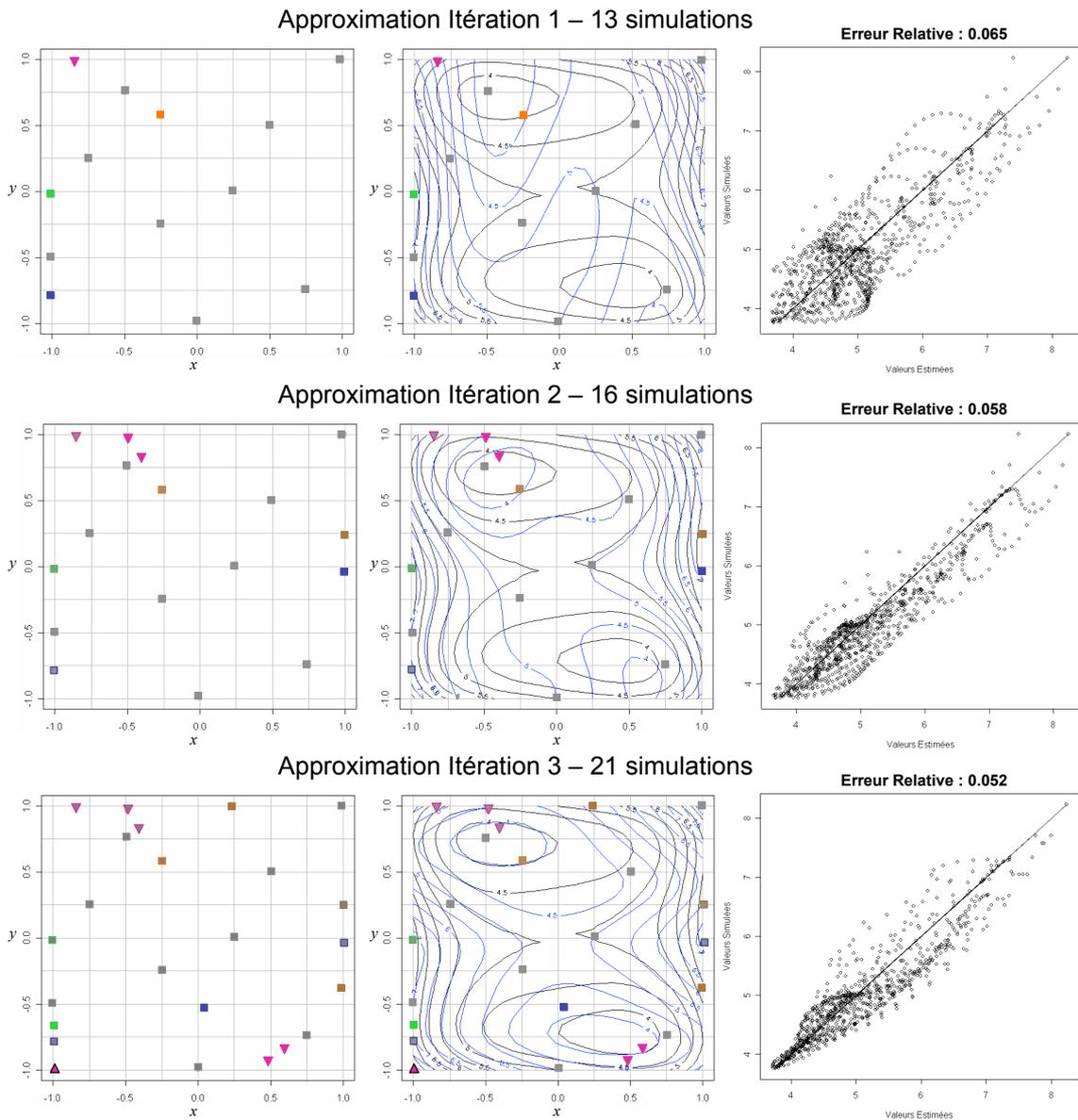
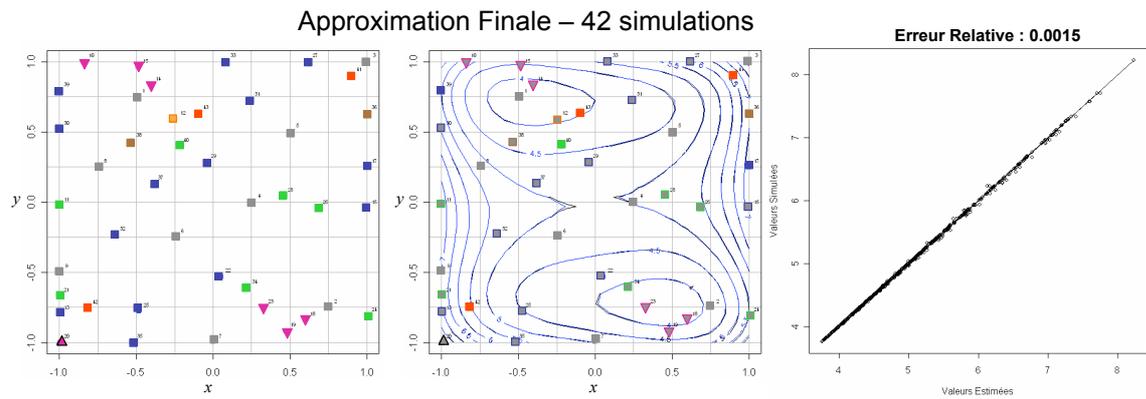
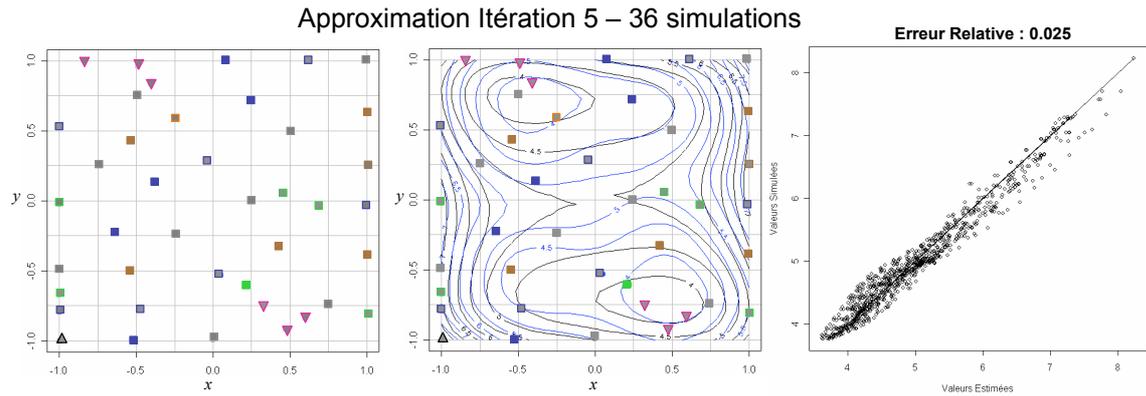
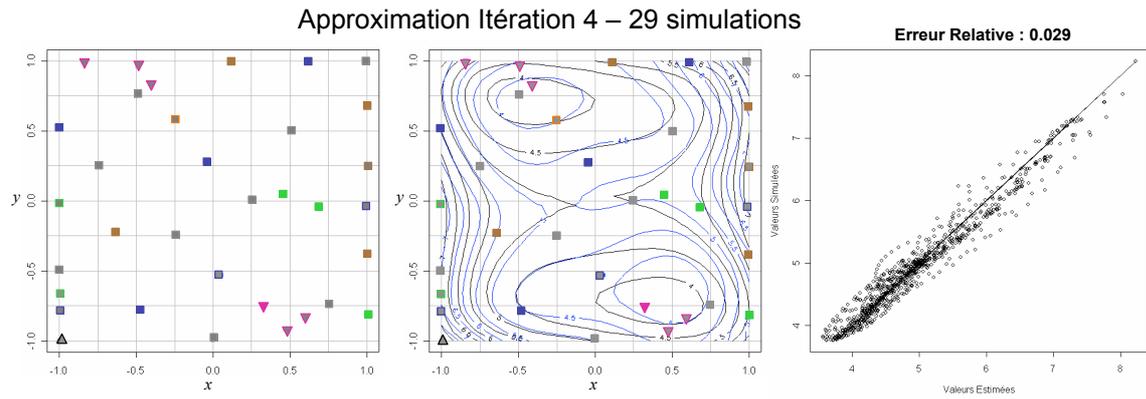


Figure 6.1 : Fonction Chameau : Référence et approximation polynomiale





Fonction Chameau - Référence

Approx finale

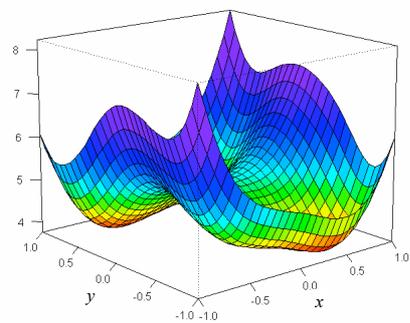
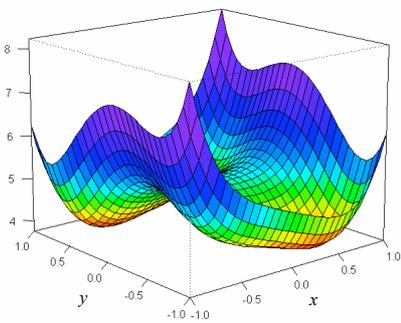


Figure 6.2 : Différentes étapes de l'ajustement de la fonction Chameau dans la méthode de modélisation adaptative

L'ajustement de la fonction Chameau est d'excellente qualité. L'ensemble de la procédure de modélisation adaptative a nécessité un total de 42 simulations, ajoutées au fur et à mesure des itérations.

Dans cet exemple, nous considérons une fonction de covariance gaussienne et anisotrope. Le plan d'expériences initial est un hypercube latin de distance maximin comportant 9 simulations. Une fonction polynomiale de second degré est alors ajusté par régression. A chaque itération, un maximum de six points pilotes est autorisé, deux points de dérivée partielle nulle sont rajoutés, un maximum et un minimum sont optimisés avant l'approche « iso-simplexe ».

Le type de points du dispositif expérimental final est réparti comme suit :

- 9 points constituant le plan d'expériences initial : un hypercube latin de distance maximin,
- 1 maximum,
- 6 minima,
- 7 points de dérivée nulle,
- 1 point de dérivée nulle non-confirmée,
- 13 points pilotes simulés,
- 2 points pilotes ayant des valeurs fictives,
- 3 points de confirmations.

Nous constatons que l'ajout d'information fictive par points pilotes non-simulés est d'importance non négligeable en milieu de processus. En effet, lors des premières itérations, l'ajout de valeurs fictives peut s'avérer risqué car basé sur une information réelle encore insuffisante. Nous avons donc une à deux valeurs fictives au maximum. A l'inverse, lors des dernières itérations, seules deux valeurs fictives sont conservées. En effet, l'approximation est alors représentative de la réponse et l'ajout de points pilotes n'est plus nécessaire pour accroître la prédictivité.

Par contre, nous voyons qu'aux itérations quatre et cinq, nous avons respectivement cinq et six valeurs fictives dans le dispositif expérimental, ce qui contribue grandement à améliorer la qualité de l'approximation.

La qualité de l'approximation est améliorée à chaque itération, pour obtenir, *in fine*, une erreur relative finale de 0.0015.

A présent, nous présentons les résultats obtenus par l'approche de modélisation adaptative sur plusieurs fonctions non-régulières à deux variables.

6.3 Exemples de modélisation de différentes fonctions à deux variables

Nous présentons, pour chaque fonction, son équation et une vue en perspective de la référence, ainsi que les contours de la référence et de l'approximation et enfin le graphique des valeurs prédites par le modèle en fonction de valeurs observées. Pour chaque exemple, nous considérons une fonction de covariance gaussienne anisotrope et nous optimisons les valeurs de six points pilotes.

6.3.1 Fonction Chameau Décalé

Nous souhaitons modéliser à présent une fonction différente de la fonction « Chameau », celle-ci ayant principalement des irrégularités au centre du domaine. L'expression de la fonction Chameau Décalé est la suivante :

$$F(x, y) = 81x^4 - \frac{21}{10}(3x)^4 + \frac{1}{3}(3x)^6 + 9xy - 36y^2 + 81y^4$$

La Figure 6.3 illustre les résultats de modélisation obtenus avec :

- (a) une représentation en perspective de la fonction référence, sur le carré unité $[-1,1]^2$,
- (b) la localisation des essais dans le dispositif expérimental final,
- (c) les contours superposés de la fonction de référence et de l'approximation finale,
- (d) le graphe des valeurs prédites aux 900 points du domaine en fonction des valeurs observées.

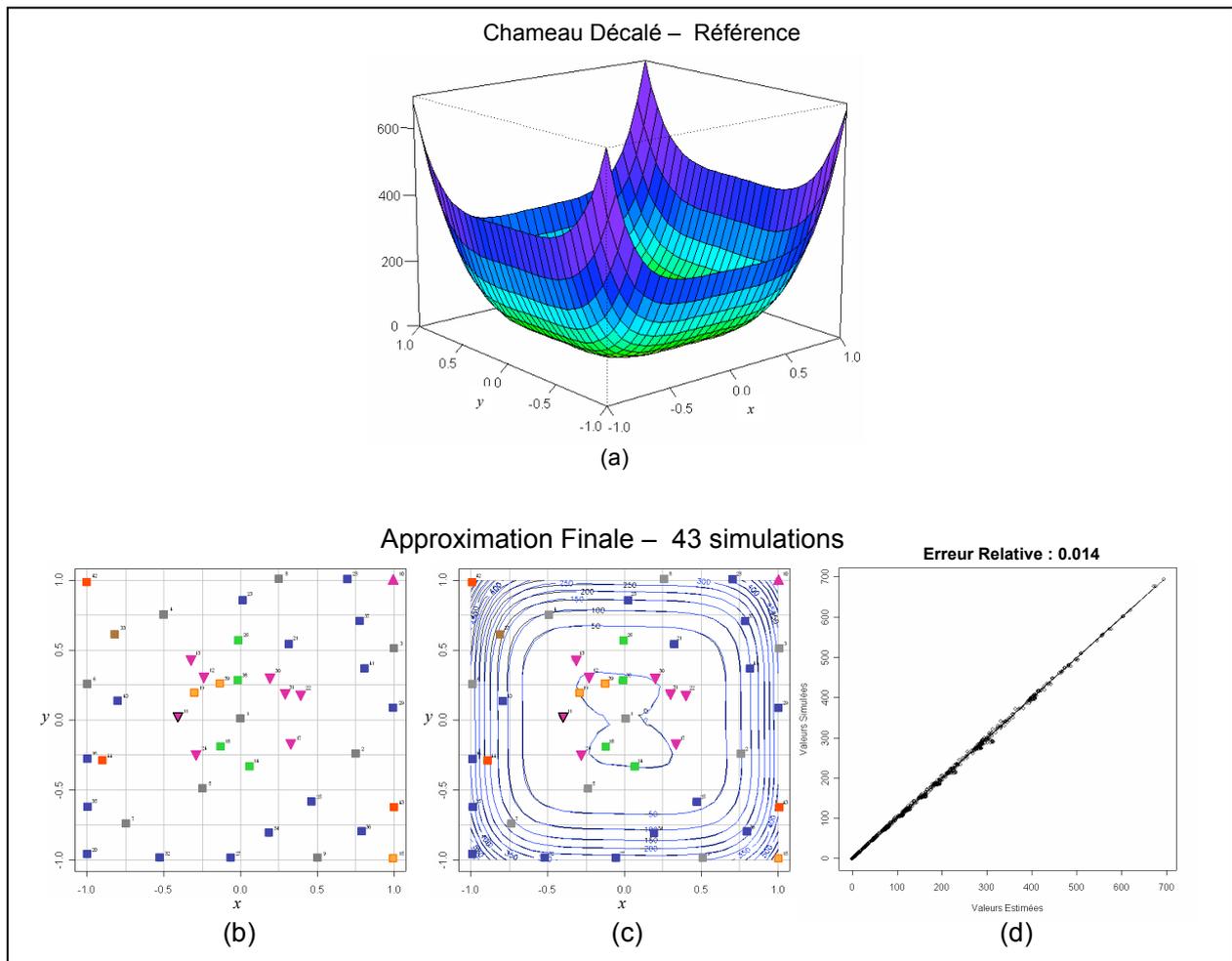


Figure 6.3 : Approximation de la fonction Chameau Décalé

La modélisation de la fonction Chameau Décalé est très satisfaisante, les deux contours sont superposés et l'erreur relative, calculée pour 900 points de prédiction, est de 0.014. Les différences entre la référence et l'approximation sont minimales. L'ensemble de la modélisation a nécessité un total de 43 simulations. On notera par ailleurs que la répartition des points du dispositif expérimental n'est pas uniforme, mais adaptée aux spécifications de la réponse, avec en particulier une concentration d'essais dans la zone centrale du domaine.

6.3.2 Fonction Sinus

La fonction Sinus considérée ici a pour expression :

$$F(x, y) = -\frac{\sin(x-3)}{x-3} - \frac{\sin(y+5)}{y+5}$$

La Figure 6.3 illustre les résultats de modélisation obtenus avec :

- (a) une représentation en perspective de la fonction référence, sur le carré unité $[-1, 1]^2$,

- (b) la localisation des essais dans le dispositif expérimental final,
- (c) les contours superposés de la fonction de référence et de l'approximation finale,
- (d) le graphe des valeurs prédites aux 900 points du domaine en fonction des valeurs observées.

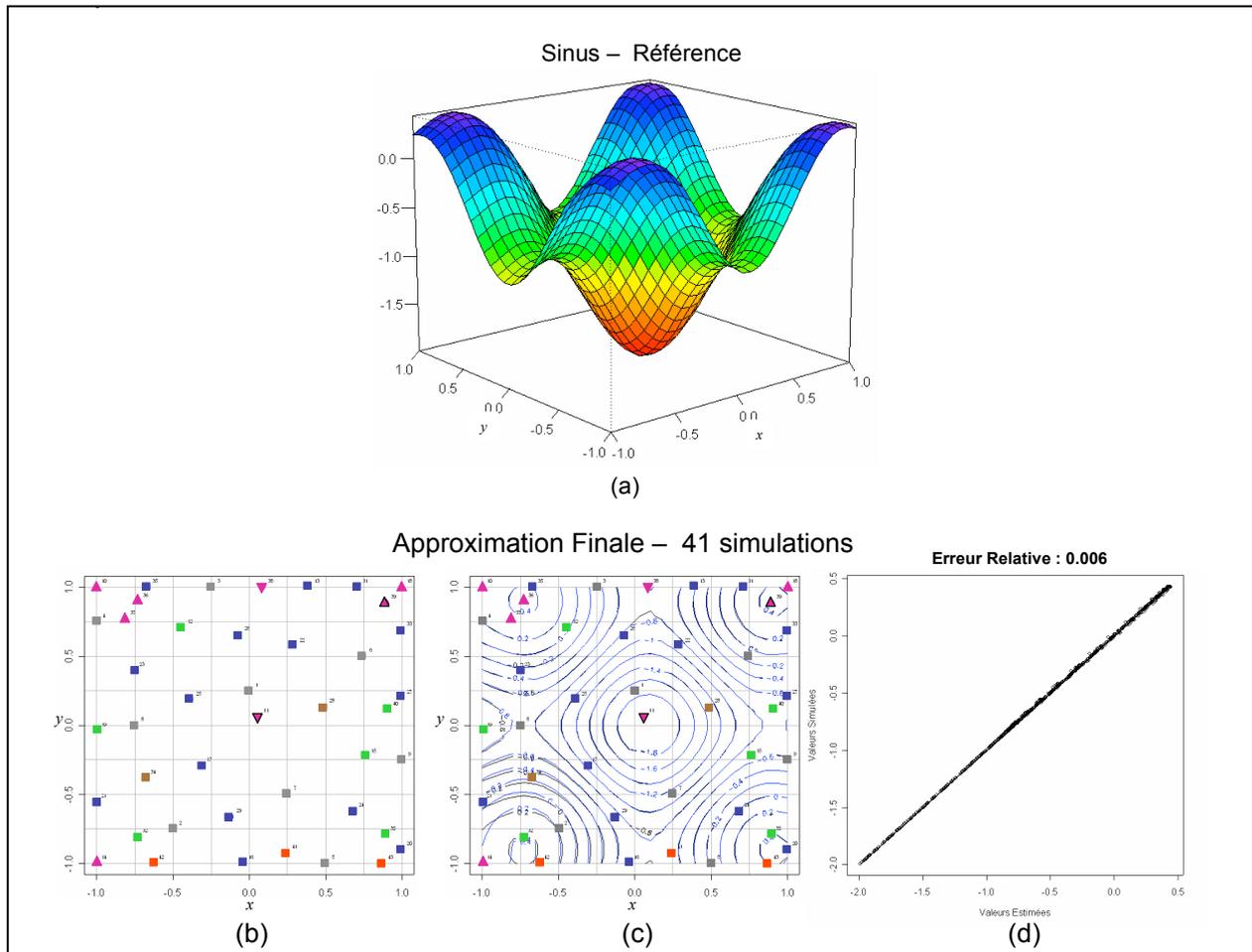


Figure 6.4 : Approximation de la fonction Sinus

Nous pouvons voir que la modélisation de la fonction Sinus est de très bonne qualité : seule une différence, pour les valeurs négatives des deux paramètres, est constatée. L'erreur relative, calculée aux 900 points de prédiction est très faible puisque égale à 0.006. L'ensemble de la modélisation a nécessité un total de 27 simulations.

6.3.3 Fonction Sinus Décalée

La fonction considérée dans cette section est la fonction Sinus Décalée. Son expression est la suivante :

$$F(x, y) = -\frac{\sin(x)}{x} - \frac{\sin(y+2)}{y+2}$$

La Figure 6.5 illustre les résultats de modélisation obtenus avec :

- (a) une représentation en perspective de la fonction référence, sur le carré unité $[-1,1]^2$,
- (b) la localisation des essais dans le dispositif expérimental final,
- (c) les contours superposés de la fonction de référence et de l'approximation finale,
- (d) le graphe des valeurs prédites aux 900 points du domaine en fonction des valeurs observées.

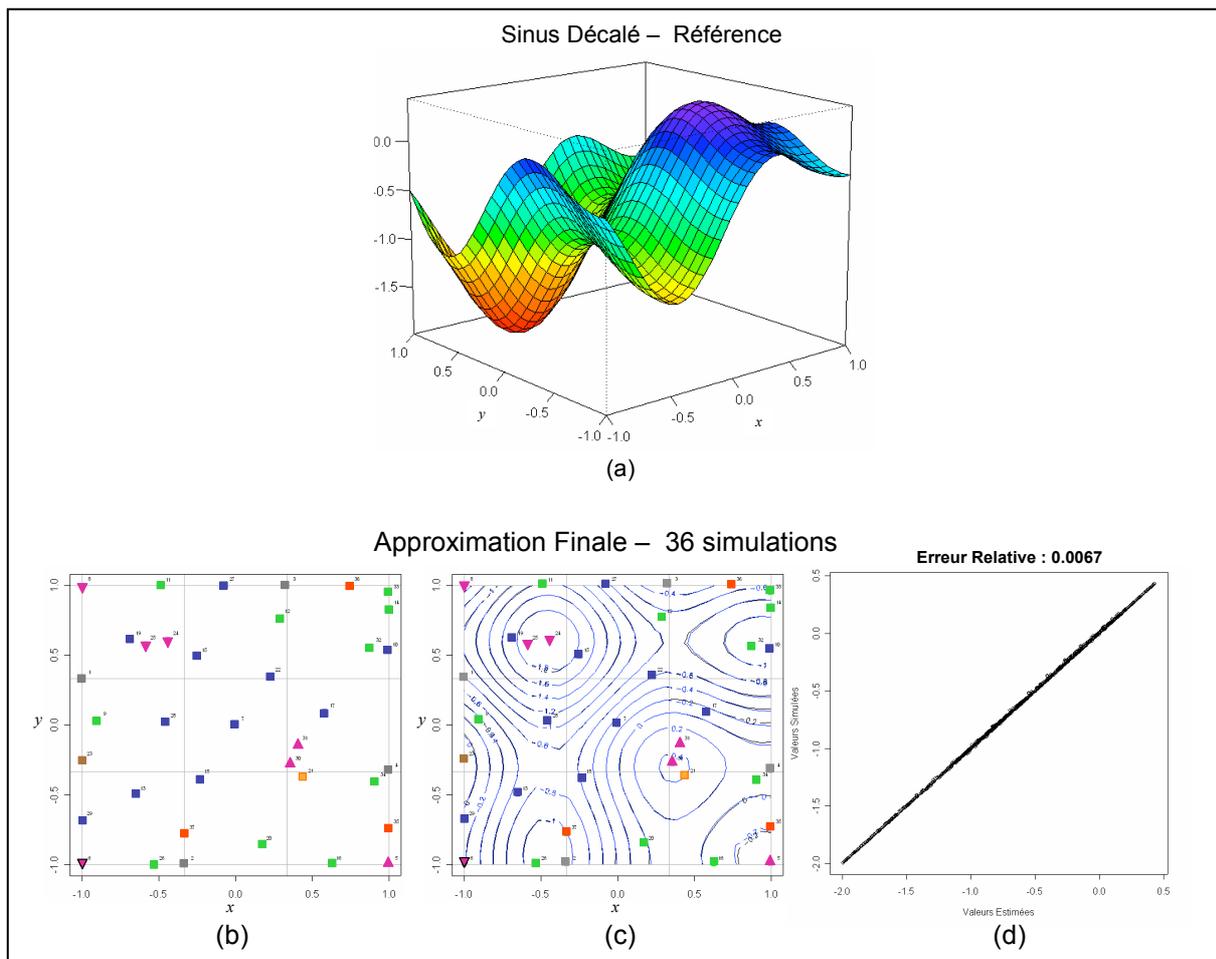


Figure 6.5 : Approximation de la fonction Sinus Décalée

Nous voyons sur ces deux figures que l'approximation de la fonction Sinus Décalée est très précise. Un nombre total de 36 simulations a été nécessaire pour obtenir ces résultats. L'erreur relative, calculée pour 900 points de prédiction, est également très satisfaisante, puisque valant 0.012.

Nous voyons dans cet exemple l'importance capitale de l'ajout des extrema : ils permettent en effet d'identifier les optima de la fonction. Ils sont cruciaux pour les détections des non-régularités de la fonction.

Il est ici important de souligner le caractère réellement spécifique de l'échantillonnage, dont on voit qu'il concentre l'ajout de d'information autour des points d'inflexion de la réponse.

6.3.4 Fonction Branin

L'expression de la fonction Branin est donnée par :

$$F(x, y) = \left(y - \frac{5.1}{4\pi^2} x^2 + \frac{5}{\pi} x - 6 \right)^2 + 10 \left(1 - \frac{1}{8\pi} \cos x \right) + 10$$

La Figure 6.6 illustre les résultats de modélisation obtenus avec :

- (a) une représentation en perspective de la fonction référence, sur le carré unité $[-1, 1]^2$,
- (b) la localisation des essais dans le dispositif expérimental final,
- (c) les contours superposés de la fonction de référence et de l'approximation finale,
- (d) le graphe des valeurs prédites aux 900 points du domaine en fonction des valeurs observées.

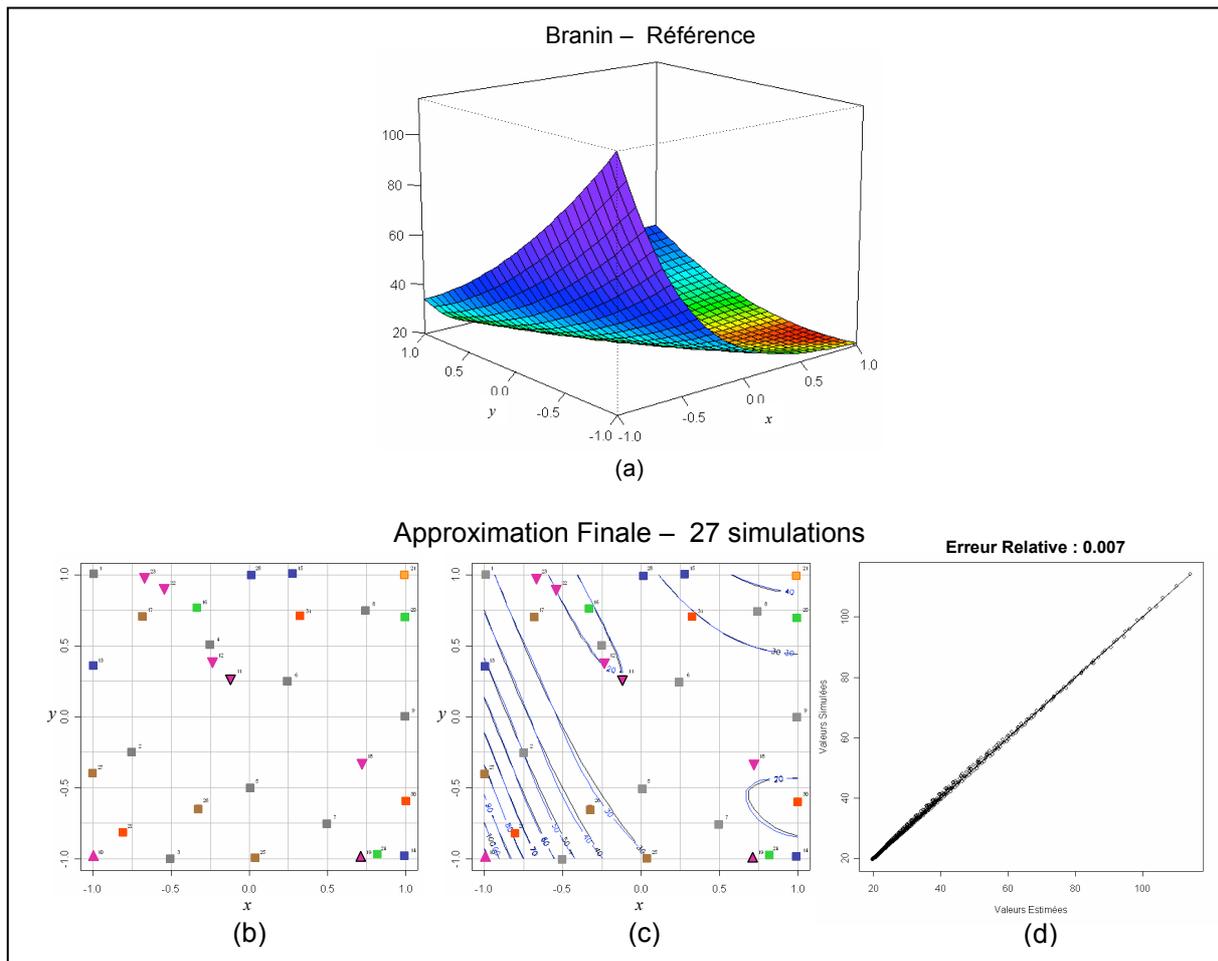


Figure 6.6 : Ajustement de la fonction Branin

Une fois encore, l’ajustement de la fonction Branin est très satisfaisant, les contours de la référence et de l’approximation sont très proches. La modélisation a nécessité un total de 27 simulations, pour une erreur relative calculée en 900 points de prédiction égale à 0.007.

Les zones de non-régularités sont ici encore plus échantillonnées que les autres parties du domaine expérimental.

6.3.5 Fonction analytique non-régulière 1

La fonction que nous présentons ici est la fonction suivante :

$$F(x, y) = 0.2e^x - 0.2y + \frac{1}{3}y^6 + 4y^4 - 4y^2 + \frac{7}{10}x^2 + x^4 + \frac{3}{((2x)^2 + (2y)^2 + 1)}$$

La Figure 6.7 illustre les résultats de modélisation obtenus avec :

- (a) une représentation en perspective de la fonction référence, sur le carré unité $[-1,1]^2$,

- (b) la localisation des essais dans le dispositif expérimental final,
- (c) les contours superposés de la fonction de référence et de l'approximation finale,
- (d) le graphe des valeurs prédites aux 900 points du domaine en fonction des valeurs observées.

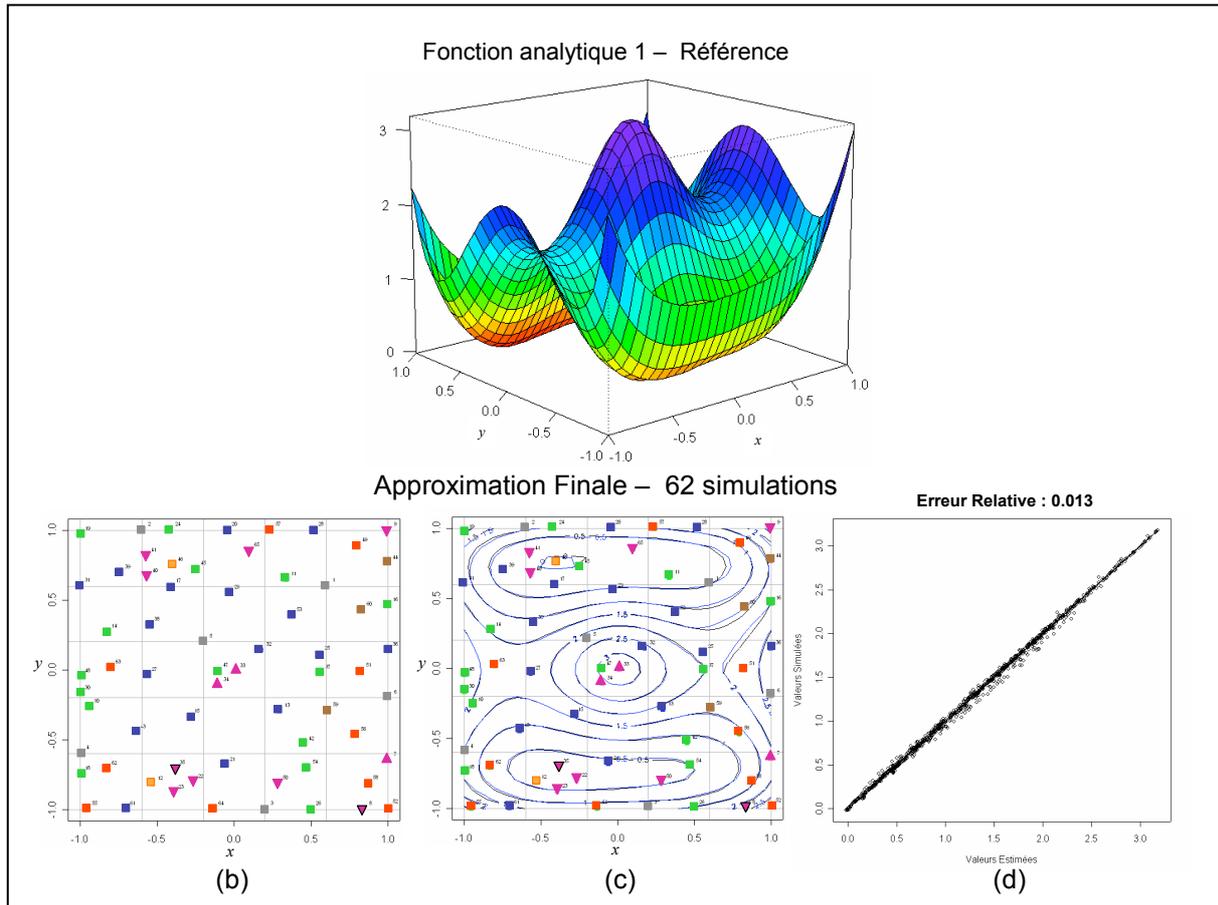


Figure 6.7 : Ajustement de la « fonction analytique 1 »

L'ajustement de la « fonction analytique 1 » a nécessité un nombre total de 62 simulations pour une erreur calculée sur 900 points de prédiction égale à 0.013. L'ajustement est encore une fois de très bonne qualité, les contours de la référence et de l'approximation étant quasiment superposés.

6.3.6 Fonction analytique non-régulière 2

Enfin, nous proposons d'ajuster ici une fonction analytique fortement non-régulière, notamment du fait du caractère non-stationnaire des irrégularités. Cette fonction a pour expression :

$$F(x, y) = 0.2e^{(x-3)} + 2.2|y| + 1.3y^6 - 2y^2 - 0.5y^4 - 0.5x^4 + 2.5x^2 + 0.7x^3 + \frac{3}{(8x-2)^2 + (5y-3)^2 + 1} + \sin(5x)\cos(3x^2)$$

La Figure 6.8 illustre les résultats de modélisation obtenus avec :

- (a) une représentation en perspective de la fonction référence, sur le carré unité $[-1,1]^2$
- (b) la localisation des essais dans le dispositif expérimental final,
- (c) les contours superposés de la fonction de référence et de l'approximation finale,
- (d) le graphe des valeurs prédites aux 900 points du domaine en fonction des valeurs observées.

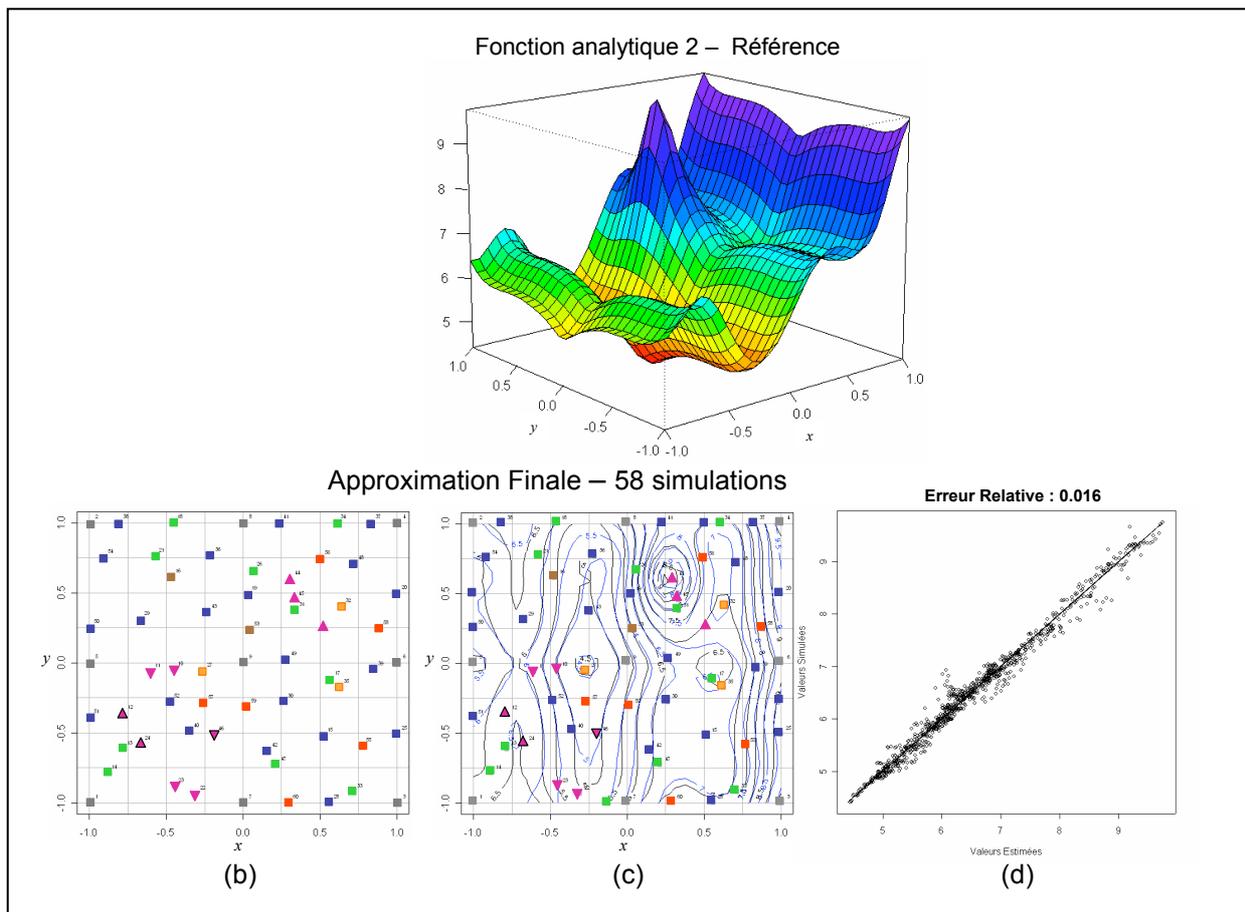


Figure 6.8 : Ajustement de la « fonction analytique 2 »

Sur cet exemple, nous constatons que l'approximation de la « fonction analytique 2 » par la méthode de modélisation adaptative est de bonne qualité, avec une erreur relative calculée pour 900 points de prédiction de 0.016. Un total de 58 simulations a été nécessaire à

l'obtention de cette approximation. Le maximum local de cette approximation, situé en $(0.25, 0.6)$, a été détecté, grâce à la mise en œuvre d'une recherche « iso-simplexe » dans la direction de plus grand accroissement, alors qu'aucun point du dispositif initial ne se trouvait dans un voisinage de ce maximum.

Ainsi, ces six exemples pour deux paramètres montrent que les résultats obtenus par la méthode de modélisation adaptative sont de bonne qualité. En effet, on constate que les non-régularités sont détectées, les courbes des références et les approximations finales étant très proches, dans tous les cas. Nous avons également effectué des tests sur d'autres fonctions à deux variables, ceux-ci donnant des résultats tout aussi satisfaisants.

De plus, chaque type de point a un apport crucial dans la modélisation, nous distinguons bien les points maximum ou minimum ou ceux de dérivée nulle qui sont situées aux extrema de la fonction. La stratégie d'ajout de point est spécifique à la réponse étudiée maximisant ainsi l'information apportée à chaque itération : le plan d'expériences n'est pas uniforme.

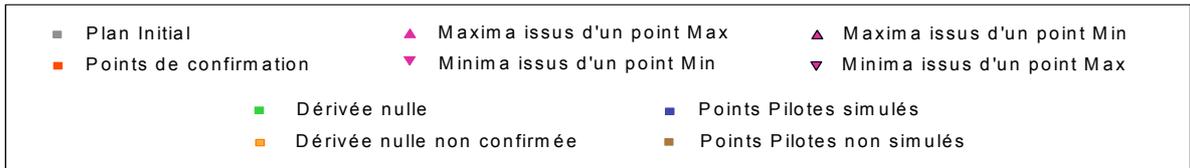
Nous avons donc vu que l'approche itérative permet une modélisation précise de fonctions non-régulières à deux paramètres. La section suivante est consacrée à l'étude de l'influence du plan d'expériences initial sur la qualité des résultats. En effet, il apparaît de première importance de s'assurer que le principe de modélisation par enrichissement évolutif n'est pas dépendant du dispositif expérimental initial.

6.4 Indépendance au dispositif expérimental initial

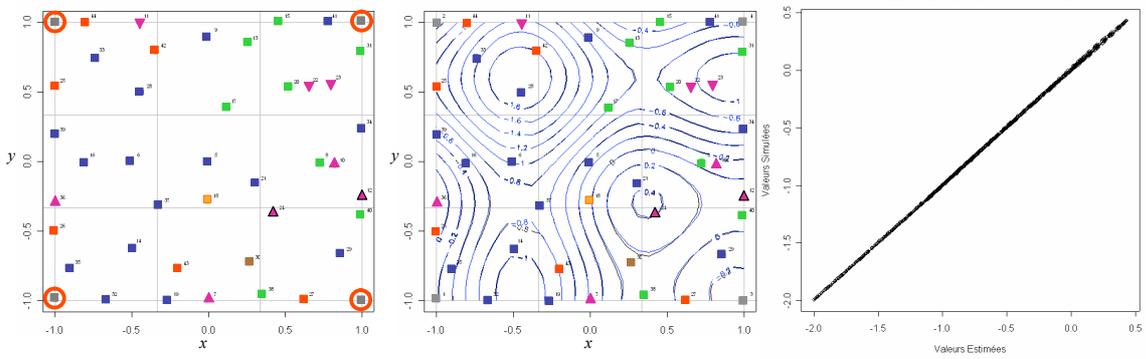
Nous allons à présent donner des exemples d'ajustement d'une fonction pour différents plans d'expériences initiaux. En effet, nous souhaitons juger de l'influence du choix du dispositif initial sur la qualité de la modélisation. Les plans que nous considérons sont :

- le plan factoriel (ff) et le plan composite centré (ccf),
- deux hypercubes latins maximin à quatre simulations (HL41 et HL42),
- deux hypercubes latins maximin à six simulations (HL61 et HL62),
- deux hypercubes latins maximin à neuf simulations (HL91 et HL92).

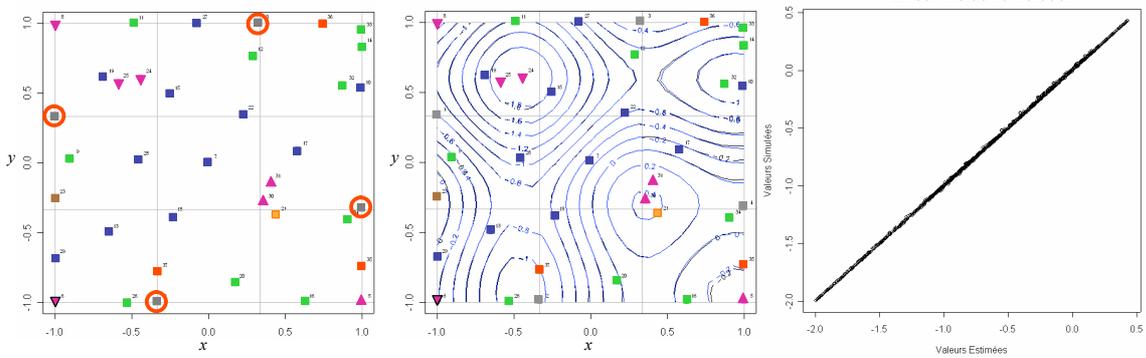
Comme précédemment, nous présentons ici les graphes du plan d'expériences final, les superpositions des contours de la fonction référence Sinus décalée et de son approximation par la méthode évolutive ainsi que le graphe des valeurs estimées en fonction des valeurs prédites. Dans cette partie, pour chaque graphique, seul le dispositif expérimental initial change lors de la modélisation adaptative. Les résultats obtenus sont illustrés sur la Figure 6.9.



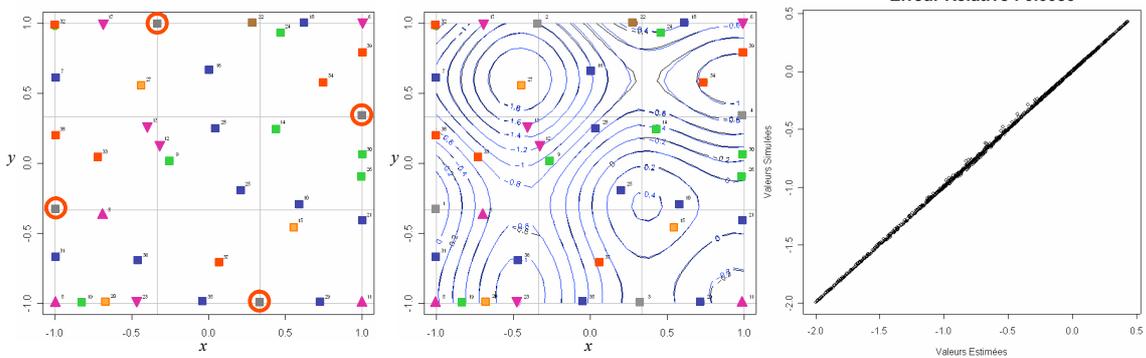
Approximation Finale – ff – 43 simulations



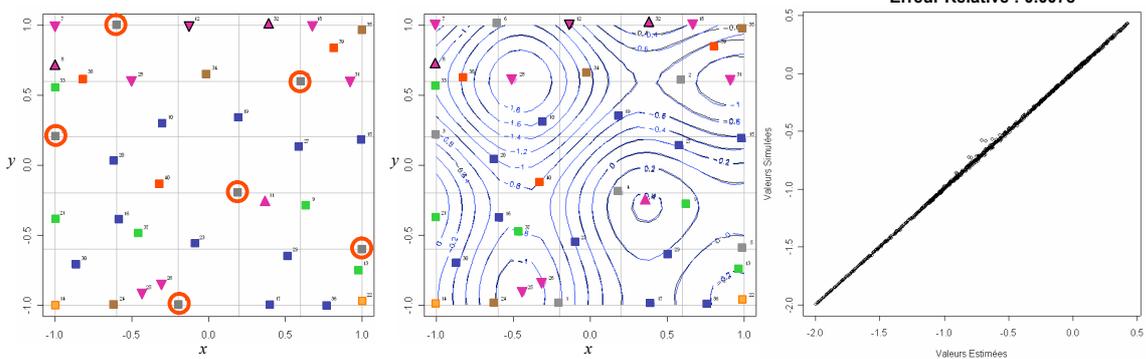
Approximation Finale – HL41 – 36 simulations



Approximation Finale – HL42 – 38 simulations



Approximation Finale – HL61 – 37 simulations



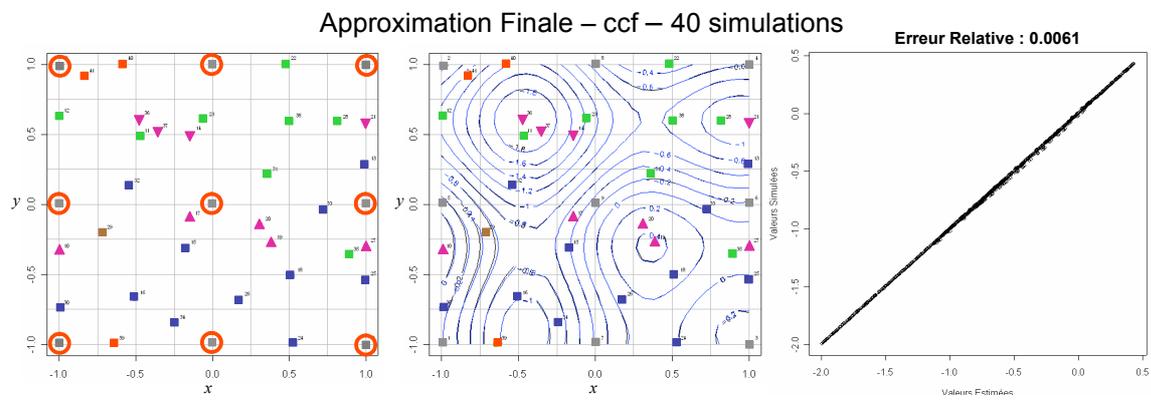
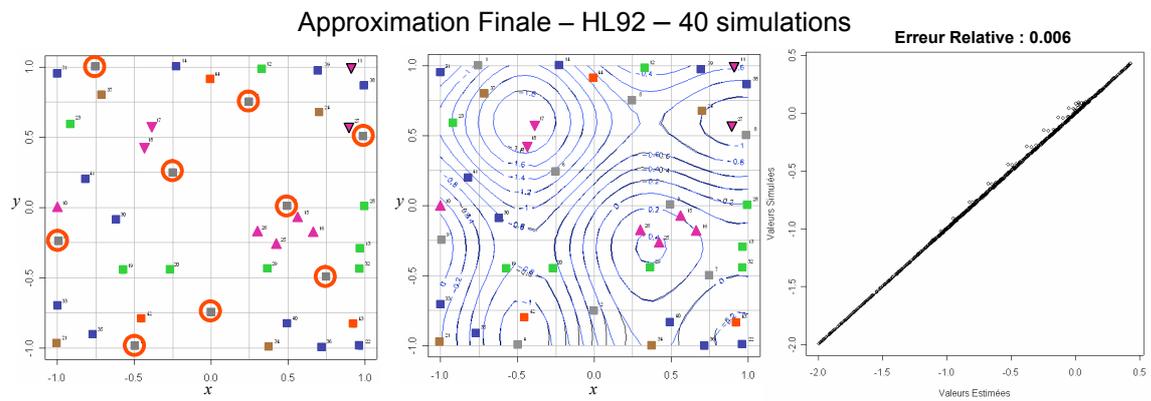
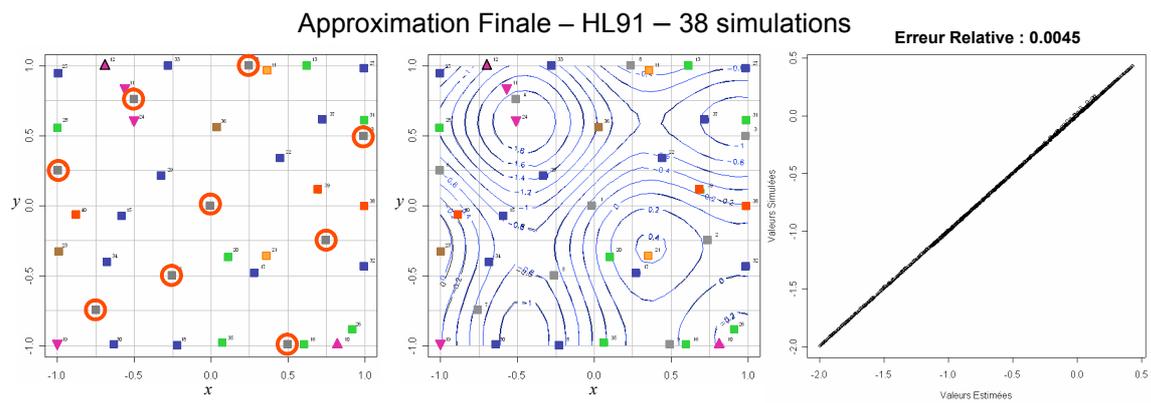
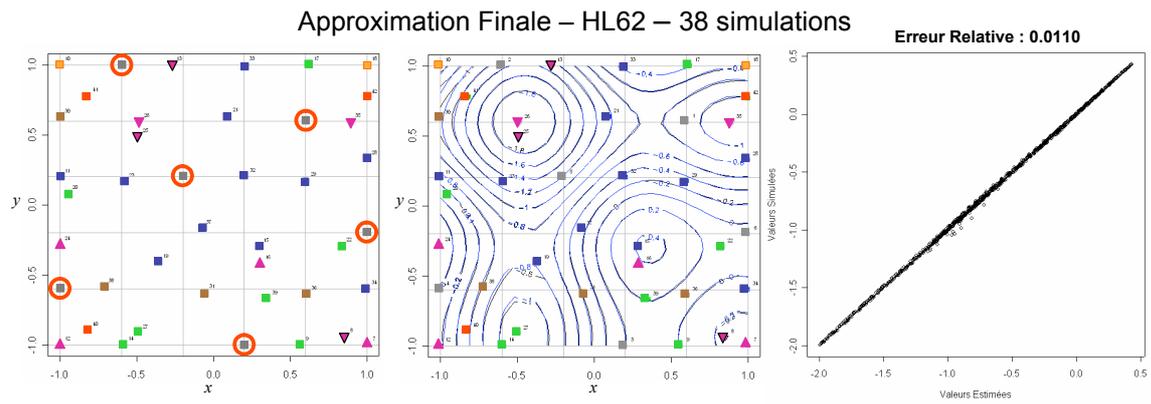


Figure 6.9 : Approximation de la fonction Sinus décalée pour différents plans d'expériences initiaux

Nous avons représenté sur la Figure 6.9 différentes approximations de la fonction Sinus décalée pour différents plans d'expériences initiaux. Les approximations ont été obtenues par l'approche évolutive. Ainsi, les trois premiers graphes représentent les approximations obtenues avec des plans initiaux à quatre expérimentations : respectivement le plan factoriel et deux hypercubes latin de distance maximin. Ensuite, nous avons représenté les résultats obtenus pour deux hypercubes latin de distance maximin contenant chacun 6 expérimentations. Enfin, la dernière série de plans est constituée respectivement d'hypercubes latin de distance maximin à 9 simulations et du plan composite centré.

Cet exemple montre que le choix du plan d'expériences initial, que ce soit en terme de distribution spatiale ou de taille, a très peu d'influence sur la qualité de l'approximation. En effet, dans tous les cas, la modélisation est très précise en terme de prédiction. Nous obtenons des erreurs relatives très proches pour un nombre comparable de simulation, quelque soit le dispositif initial choisi.

Ces tests ont été réalisés sur l'ensemble des fonctions présentées dans la section 6.3. Du fait de la constance des résultats, nous avons choisi de les exposer uniquement sur la fonction Sinus décalée.

Le choix d'ajouter les extrema, les points de dérivée nulle et de variance maximale permettent donc d'échantillonner l'espace de manière optimale. Nous constatons même qu'il est tout à fait possible de partir d'un plan d'expérience classique (factoriel ou composite centré), ce qui est primordial dans l'approche introduite ici. En effet, il est alors possible, dans le cas où l'on ne sait pas si la réponse à ajuster est régulière ou non, de commencer par construire une approximation polynomiale avec la méthodologie des plans d'expériences classiques, puis, dans le cas où les résultats ne sont pas satisfaisants (les résidus issus de l'approximation sont élevés) et non uniformes sur le domaine, de procéder à un enrichissement du plan d'expériences par modélisation adaptative présentée dans ce rapport. Ce caractère évolutif est en effet un des points forts de cette méthodologie, lui assurant ainsi une grande flexibilité.

Nous voulons à présent vérifier que l'échantillonnage issu de la modélisation adaptative de réponses non-régulières est informatif, c'est-à-dire qu'il nous permet de réduire le nombre total de simulations par rapport à l'ajustement de la réponse sur un plan complet N^k ou un hypercube latin de grande taille.

6.5 Echantillonnage informatif : Méthode évolutive, plan complet et hypercube latin

Nous proposons, dans cette partie, de modéliser les différentes fonctions présentées précédemment pour des plans complets de taille N^k et puis $(N+1)^k$ où N est fixé de sorte que le nombre de points requis par modélisation adaptative soit compris entre ces bornes. Nous effectuons également un ajustement pour un hypercube latin de même taille et enfin, nous cherchons la taille nécessaire d'un hypercube latin pour obtenir à la même qualité que celle obtenue par modélisation adaptative. Nous commençons par présenter les résultats sous forme d'un tableau récapitulatif, puis nous verrons plus en détail certains ajustements. Dans le Tableau 6-1 ci-dessous, nous présentons la valeur de l'erreur relative obtenue lors de la prédiction aux 900 points du domaine expérimental pour chacune des fonctions testées dans la partie 6.3 et pour différents échantillonnages. Entre parenthèses figure le nombre de points du dispositif expérimental.

	Modélisation adaptative	Plan complet N^2	Plan complet $(N+1)^2$	HL même nb. sim.	HL même err. rel.
Chameau	0.002 (42)	0.010 (36)	0.007 (49)	0.003 (42)	0.002 (55)
Chameau Décalée	0.014 (43)	0.070 (36)	0.015 (49)	0.032 (43)	0.014 (58)
Sinus	0.006 (41)	0.052 (36)	0.002 (49)	0.011 (41)	0.006 (49)
Sinus Décalée	0.007 (36)	0.023 (36)	0.015 (49)	0.021 (36)	0.007 (42)
Branin	0.007 (27)	0.037 (25)	0.005 (36)	0.010 (27)	0.007 (31)
Fonction analytique 1	0.013 (62)	0.09 (49)	0.054 (64)	0.070 (62)	0.013 (95)
Fonction analytique 2	0.016 (58)	0.015 (49)	0.012 (64)	0.022 (58)	0.016 (74)

Tableau 6-1 : Résultats de modélisation pour différents échantillonnages

D'après les résultats présentés dans le Tableau 6-1, nous constatons que les performances des modélisations adaptatives sont en général supérieures à celles obtenues par échantillonnage systématique par plan complet ou hypercube latin de taille similaire. Les erreurs relatives de la modélisation adaptative sont plus faibles dans la plupart des cas, excepté pour certains plans complets, mais ceci est au prix de simulations supplémentaires (10% à 25%).

De plus, pour une qualité de prédiction égale, nous constatons que le nombre de simulations requises pour un hypercube latin maximin est beaucoup plus important que lors

de l'ajustement par la méthode adaptative. En effet, dans le cas de deux paramètres, le gain de simulation constaté avec la méthode adaptative varie entre 13% et 26%.

Nous détaillons ci-dessous les résultats obtenus avec des hypercubes latins de même taille que les plans issus de la méthode de modélisation adaptative. Ils sont présentés sous forme de graphes de la Figure 6.10 à la Figure 6.15. En effet, les tests présentés dans le tableau ci-dessus montrent que la méthode de modélisation adaptative est plus informative que les méthodes d'échantillonnage systématique par plan complet : avec moins de simulations on obtient au moins la même erreur voire une erreur plus faible. De plus, étant donné que l'utilisation d'échantillonnage systématique par plan complet n'est pas envisageable car trop coûteuse pour de plus grandes dimensions, la comparaison n'est pas étendue et nous nous intéressons plus particulièrement à une comparaison entre la méthode de modélisation adaptative et un échantillonnage par hypercubes latins.

Nous commençons donc par ajuster la fonction Chameau à l'aide d'un hypercube latin constitué de 42 simulations. La Figure 6.10 présente le plan utilisé ainsi que les courbes de niveaux de l'approximation et la référence, puis respectivement les graphes de prédiction obtenus pour la modélisation par hypercube et pour la modélisation adaptative.

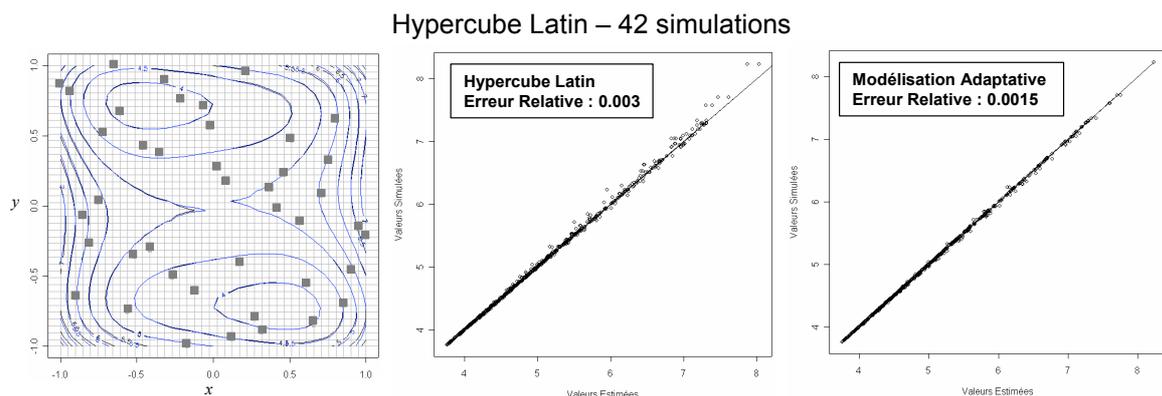


Figure 6.10 : Fonction Chameau – hypercube latin 42 simulations

Dans le cas de l'ajustement de la fonction Chameau à l'aide d'un hypercube latin de 42 points, nous constatons que le modèle obtenu est de bonne qualité, les principales erreurs se trouvent au niveau des extrema de la fonction. L'échantillonnage par hypercube latin mène à une estimation moins précise aux frontières du domaine, particulièrement dans les « coins » du domaine expérimental qui sont peu échantillonnés.

La Figure 6.11 présente les résultats de l'ajustement de la fonction Chameau Décalée à partir d'un hypercube latin de 43 simulations.

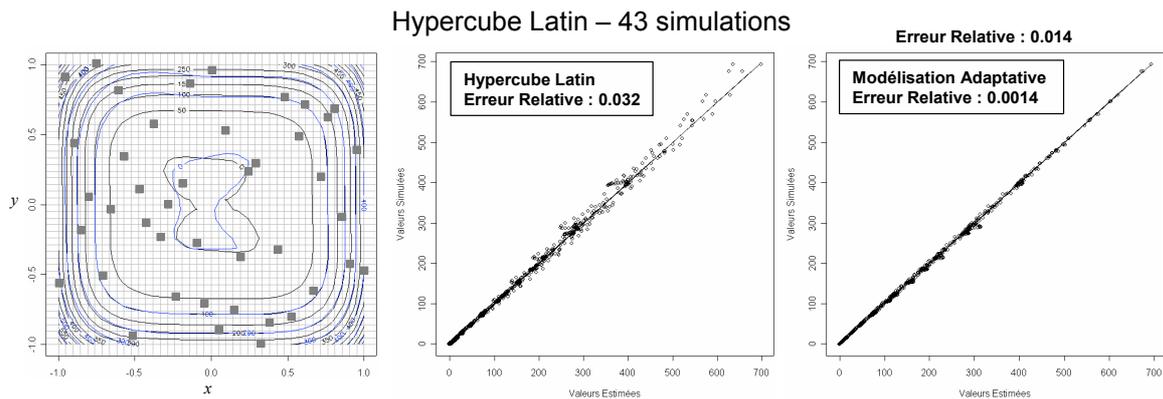


Figure 6.11 : Fonction Chameau Décalée – hypercube latin 42 simulations

Dans le cas de la fonction Chameau Décalée, les principaux défauts d’ajustements se situent au centre du domaine, là où l’on trouve les principales irrégularités. En effet, du fait de l’échantillonnage par hypercube, il n’est pas possible de concentrer les points au centre du domaine, afin de gagner en précision en ces endroits. De plus, comme précédemment, les valeurs maximales de la fonction sont mal ajustées car situées aux frontières du domaine expérimental.

La Figure 6.12 présente les résultats de l’ajustement de la fonction Sinus à partir un hypercube latin de 41 simulations.

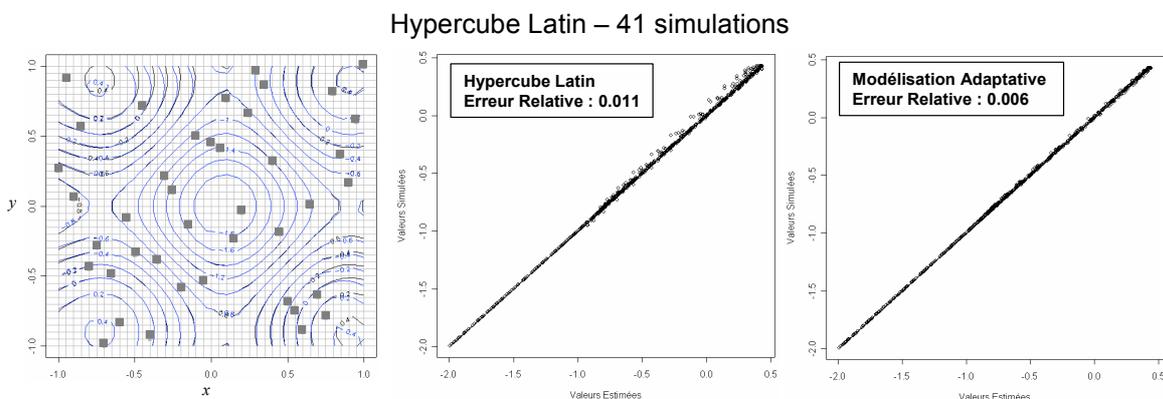


Figure 6.12 : Fonction Sinus – hypercube latin 41 simulations

L’ajustement de la fonction Sinus à partir d’un hypercube latin de 41 simulations donne des résultats satisfaisants, avec toutefois quelques défauts d’ajustement au voisinage des valeurs maximales de la fonction. Encore une fois, du fait des propriétés des hypercubes latins, il n’est pas possible d’échantillonner le domaine expérimental avec précision dans les directions de montée ou de descente de l’approximation et par conséquent, ce type de plan fournit des résultats moins satisfaisants que ceux obtenus par la méthode de modélisation adaptative.

La Figure 6.13 présente les résultats de l’ajustement de la fonction Sinus Décalée pour un hypercube latin de 36 simulations.

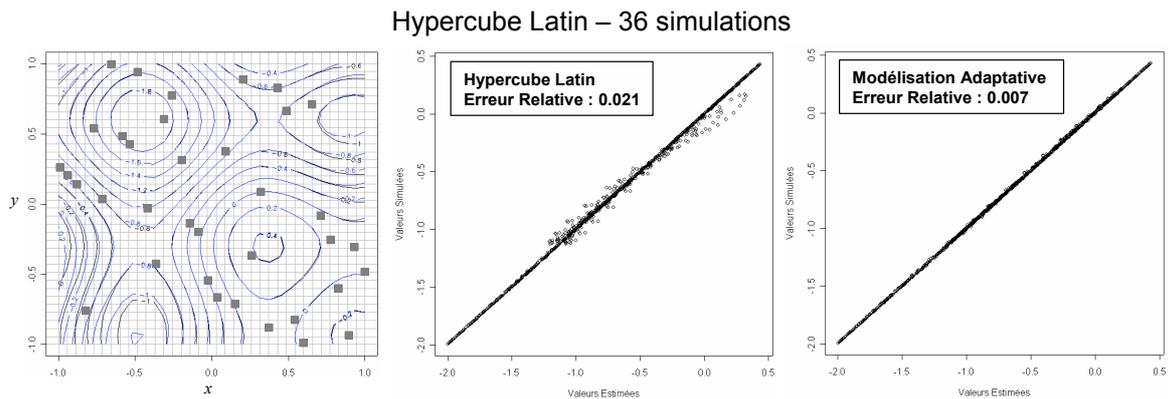


Figure 6.13 : Fonction Sinus Décalée – hypercube latin 36 simulations

Comme le montre la Figure 6.13, nous constatons que les erreurs de l’ajustement de la fonction Sinus Décalée à partir d’un hypercube latin de 36 simulations sont plus importantes que pour la modélisation adaptative. En effet, la dispersion des erreurs relatives est plus grande. De plus, les plus grandes erreurs d’ajustement sont situées pour les valeurs négatives des paramètres, cette partie du domaine ne contenant que très peu de simulation.

La Figure 6.14 présente les résultats de l’ajustement de la fonction Branin par un hypercube latin de 27 simulations.

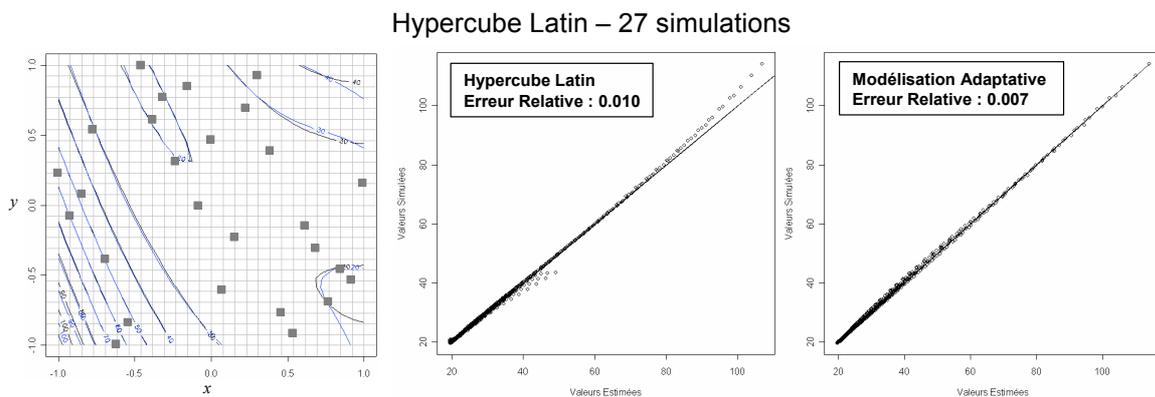


Figure 6.14 : Fonction Branin – hypercube latin 27 simulations

Dans le cas de la fonction Branin, l’ajustement global est de bonne qualité mais une fois encore, les principales dispersions sont situées au voisinage du maximum de la fonction. Nous notons également un défaut d’ajustement pour des valeurs autour de 40. Les défauts d’ajustement se situent dans des zones non-échantillonnées.

La Figure 6.15 présente les résultats de l’ajustement de la « fonction analytique 1 » par un hypercube latin de 62 simulations.

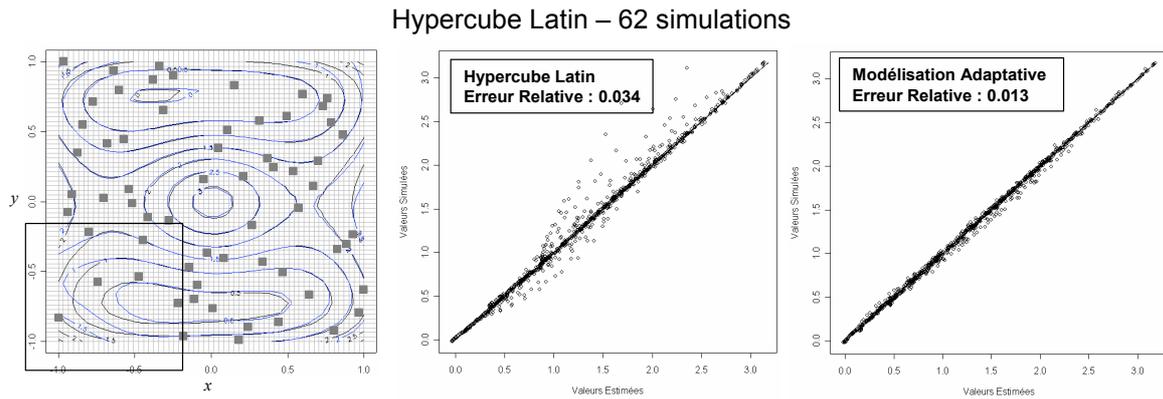


Figure 6.15 : « fonction analytique 1 » – hypercube latin 62 simulations

Dans cet exemple, il apparaît clairement que les principaux défauts d’ajustement sont situés dans la zone peu échantillonnée du dispositif expérimental, à savoir dans le coin en bas à gauche du domaine. Tout comme dans l’exemple précédent, nous constatons que les différences entre les valeurs de références et les valeurs approximées sont beaucoup moins uniformes pour un hypercube latin que pour la modélisation adaptative. En effet, dans ces deux derniers cas, il peut y avoir des erreurs assez conséquentes, ce qui n’est pas le cas de avec la modélisation adaptative où les erreurs sont d’une part plus faibles mais aussi plus réparties dans le domaine de variation de la réponse.

La Figure 6.16 présente les résultats de l’ajustement de la « fonction analytique 2 » par un hypercube latin de 58 simulations.

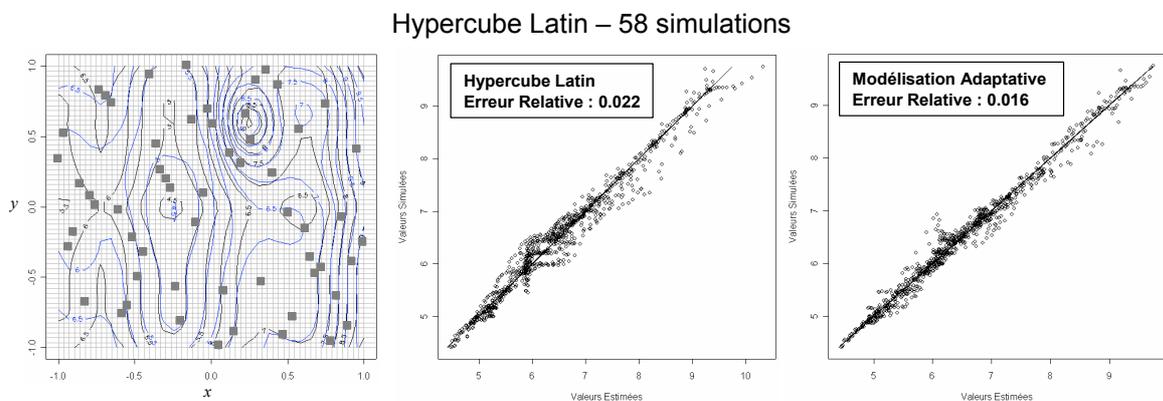


Figure 6.16 : « fonction analytique 2 » – hypercube latin 58 simulations

Pour la « fonction analytique 2 », nous constatons que l’ajustement à partir d’un hypercube latin de 58 points donne une erreur plus élevée que pour la modélisation adaptative. Le maximum local de la fonction est sur-estimé, malgré la présence de deux points de l’hypercube latin à proximité du point où cette fonction atteint son maximum.

Ainsi, nous avons montré que l’utilisation de la méthode de modélisation adaptative permet de réduire le nombre de simulations par rapport aux méthodes d’échantillonnage systématique ou hypercube latin. La stratégie d’enrichissement du plan d’expériences initial

est donc informative. Ces résultats ont été obtenus pour des fonctions à deux paramètres, nous verrons à la fin de ce chapitre que le gain de simulation est d'autant plus grand que le nombre de paramètres est important. En effet, la taille d'une grille croît de façon exponentielle par rapport au nombre de paramètres, celle d'un hypercube latin de façon quadratique. Enfin, la stratégie d'échantillonnage par hypercube latin fournit des résultats variables, du fait du grand nombre d'hypercubes latins possibles pour un nombre donné de simulations. Il est donc possible d'avoir une bonne ou une mauvaise approximation pour un hypercube latin, tout dépend du dispositif retenu.

Nous venons de montrer que la stratégie d'enrichissement du dispositif expérimental dans la modélisation adaptative est pertinente et informative, puisqu'il réduit le nombre total de simulations. Nous exposons à présent l'utilité de chaque type de points : maxima, minima, points de dérivée partielle nulle et points pilotes.

6.6 Intérêt des différents critères d'ajout de points

Dans cette section, nous voulons tester l'intérêt des différents critères d'ajout de points dans la méthodologie de modélisation adaptative. Nous commençons par évaluer la pertinence de l'information apportée par les extrema et les points de dérivée nulle, puis nous montrons que l'ajout de points pilotes au dispositif expérimental est également crucial dans la modélisation adaptative de réponses non-régulières.

Dans cette partie, nous nous limitons à deux fonctions à deux variables, la fonction Sinus Décalée et la « fonction analytique 2 » de la partie 6.3. Pour ces deux fonctions, nous avons présenté dans les sections précédentes un ajustement de très bonne qualité pour, un total de 36 et 58 simulations respectivement, le nombre de simulations ne constituant dans ce cas pas une limitation. Cependant, pour de grandes dimensions, un compromis entre la taille de l'échantillon et la qualité de l'approximation doit être fait. Ainsi, il semble judicieux de présenter les résultats pour une approximation finale de précision moindre mais nécessitant moins de simulations. Ainsi, nous contraignons la taille de l'échantillonnage à un total de 21 et 40 simulations respectivement.

Les approximations ainsi obtenues sont présentées sur la Figure 6.17 et la Figure 6.18. Dans les deux cas, six valeurs fictives sont optimisées à chaque itération pour accroître la qualité de prédiction de l'approximation.

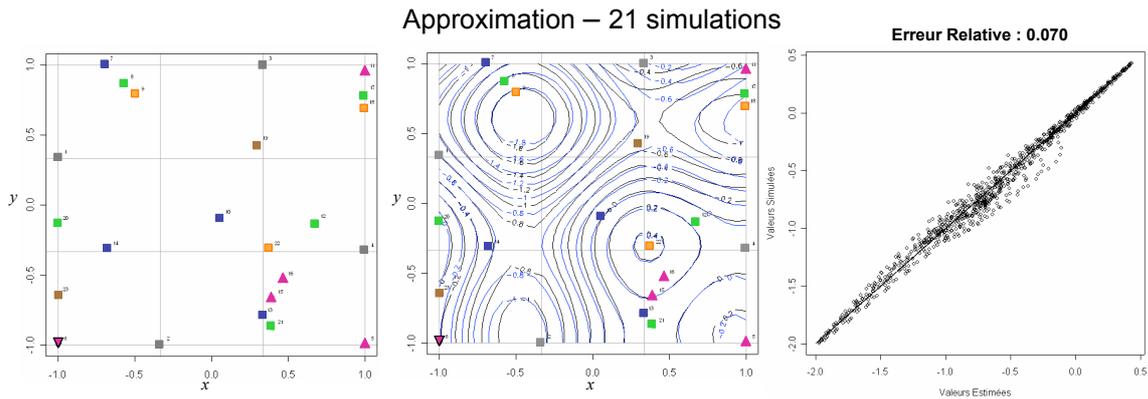


Figure 6.17 : Approximation de la fonction Sinus Décalée pour 21 simulations

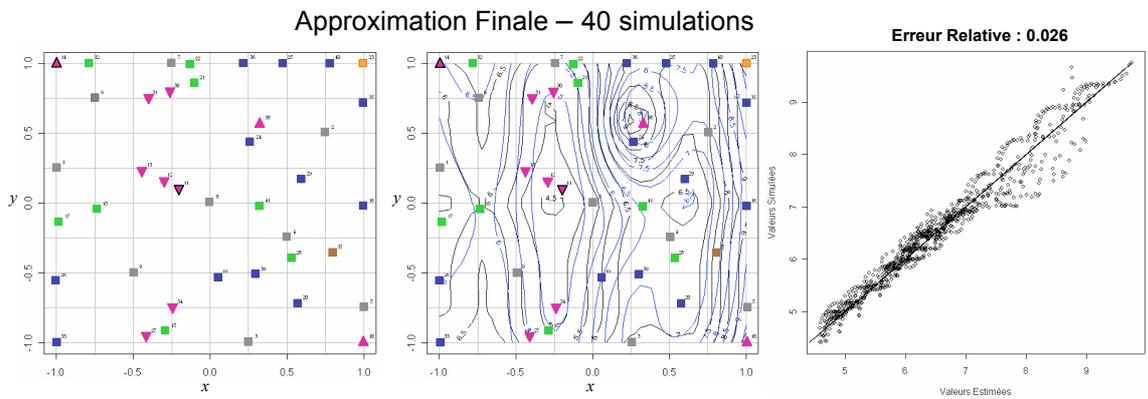


Figure 6.18 : Approximation de la « fonction analytique 2 » pour 40 simulations

Ces deux approximations sont prises pour référence lors de la présentation de l'intérêt de chaque critère de points.

6.6.1 Intérêt de l'ajout des extrema au dispositif expérimental courant

Dans cette partie, nous étudions l'intérêt de considérer comme candidats à la simulation les minima et maxima de l'approximation courante lors de la modélisation adaptative. Pour ce faire, nous considérons la fonction Sinus Décalée, et nous mettons en œuvre un ajustement adaptatif dans les mêmes conditions que dans la partie 6.3 : plan d'expériences initial, nombre de points pilotes, critère d'arrêt, etc., à l'exception du fait que nous n'ajoutons pas les extrema de l'approximation courante. L'approximation obtenue, à partir du plan d'expériences initial enrichi par les points de dérivée partielle nulle et les points pilotes, est présentée sur la Figure 6.19.

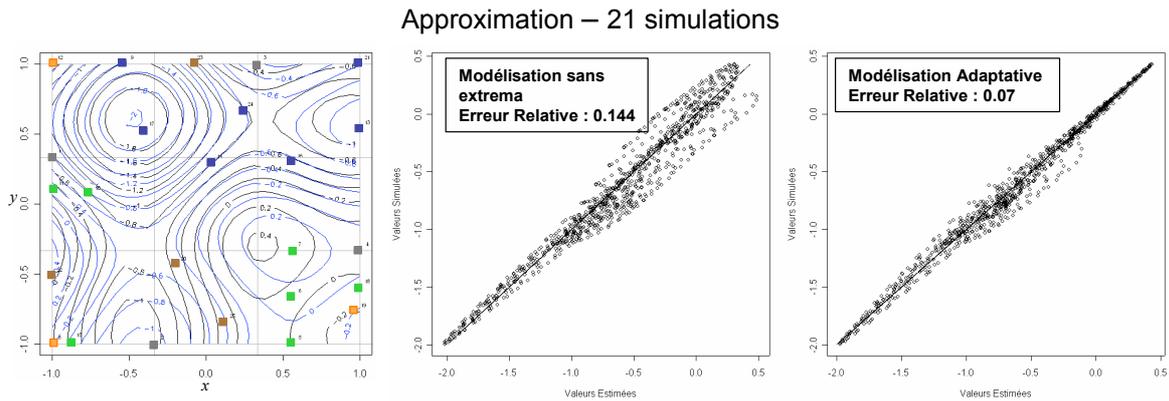


Figure 6.19 : Approximation de la fonction Sinus Décalée sans extrema

Comme nous pouvons le voir sur la Figure 6.19, enrichir le dispositif expérimental par les extrema de l’approximation courante est d’importance cruciale dans notre approche. En effet, pour un dispositif expérimental de même taille que celui issu de la modélisation adaptative, l’erreur relative est beaucoup plus importante. Dans le cas d’un ajustement avec les extrema, un total de 21 simulations étaient nécessaires pour une erreur relative de 0.07, alors que l’erreur relative obtenue ici est de 0.144. Nous observons également une plus grande dispersion des erreurs, principalement pour les grandes valeurs de la réponse.

Nous présentons à présent les résultats de la modélisation adaptative de la « fonction analytique 2 » sans l’ajout d’extrema et pour un total de 40 simulations. Ils sont illustrés sur la Figure 6.20.

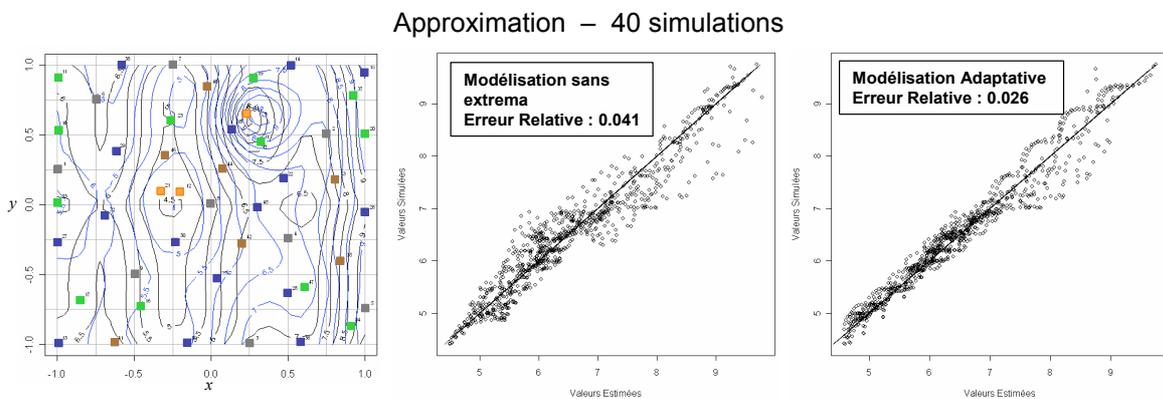


Figure 6.20 : Approximation de la « fonction analytique 2 » sans extrema

Dans le cas de la « fonction analytique 2 », l’erreur relative obtenue par un ajustement avec extrema et sans extrema est respectivement de 0.026 et 0.041. Encore une fois, l’intérêt qu’il y a à ajouter, à chaque itération les points maximum ou minimum de l’approximation courante est confirmé, un plus grand nombre d’irrégularité est détecté, à nombre de simulations fixé.

Remarque : Nous avons effectué les mêmes tests sur les autres fonctions analytiques présentées dans la partie 6.3, et des résultats similaires ont été observés. Nous ne les présentons donc pas ici.

L'ajout des extrema de l'approximation courante à chaque itération est une étape indispensable dans la méthode de modélisation adaptative. Ces points permettent de disposer d'un échantillonnage informatif, assurant la détection des extrema locaux de l'approximation courante, tout en gardant un coût de simulation convenable. Ils contribuent grandement à l'obtention d'un échantillonnage informatif.

Dans le paragraphe suivant, nous étudions l'intérêt d'ajouter des points de dérivée partielle nulle dans la méthode de modélisation adaptative.

6.6.2 Intérêt des points de dérivée partielle nulle

Dans cette partie, nous modélisons les deux mêmes fonctions que ci-dessus, sans l'ajout des points de dérivée partielle nulle. Pour cela, nous réalisons le même nombre de simulations que celui obtenu par la modélisation adaptative et nous travaillons dans les mêmes conditions, excepté pour l'ajout des points de dérivée partielle nulle. Les résultats obtenus, pour la fonction Sinus Décalée et la « fonction analytique 2 » sont présentés sur la Figure 6.21 et la Figure 6.22.

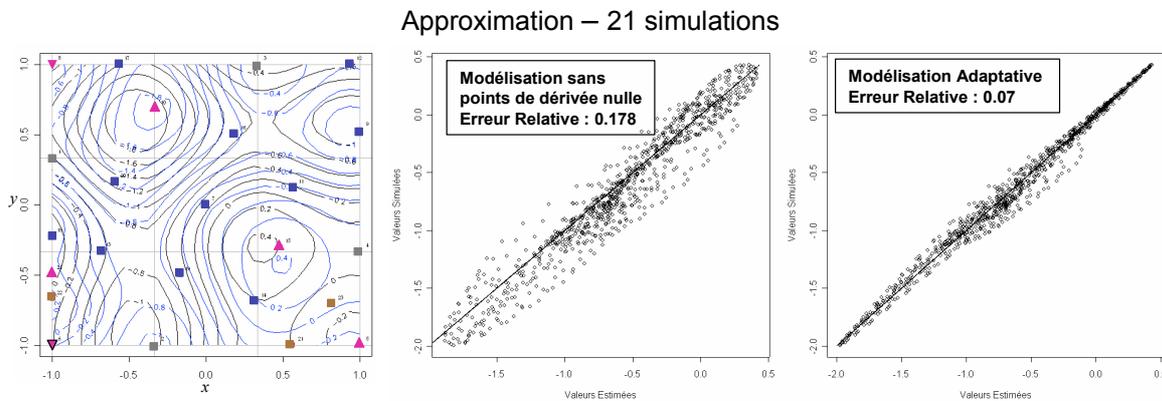


Figure 6.21 : Approximation de la fonction Sinus Décalée sans les points de dérivée nulle

D'après le graphe de la Figure 6.21, nous constatons que l'ajustement de la fonction Sinus Décalée est de moins bonne qualité que celui obtenu par la modélisation adaptative. En effet, pour un même nombre de simulations, l'erreur relative avec et sans les points de dérivée partielle nulle est respectivement de 0.07 et 0.178. Malgré le fait que l'ajout des points maximum et minimum permet de détecter les principales irrégularités de la réponse, la dispersion des erreurs est plus importante, certaines zones ne sont pas précisément modélisées.

La Figure 6.22 présente les résultats de la modélisation de la « fonction analytique 2 » sans l'ajout des points de dérivée partielle nulle.

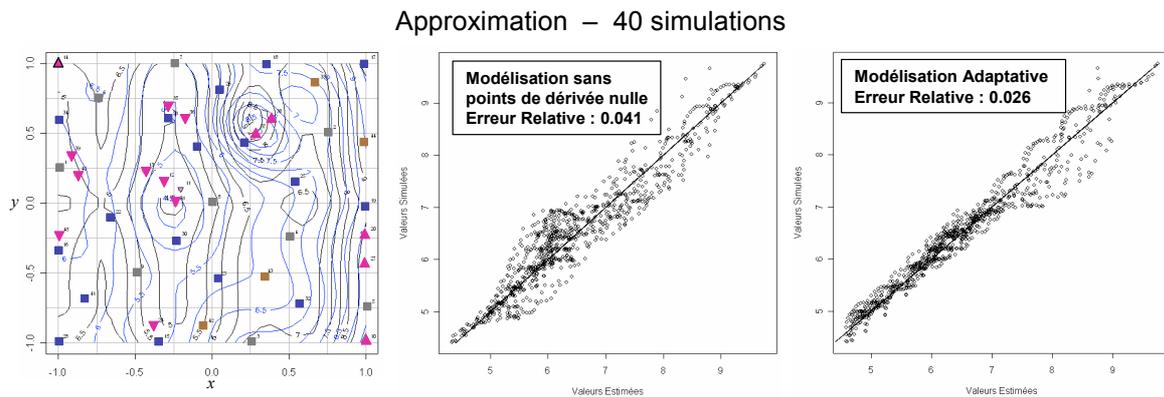


Figure 6.22 : Approximation de la « fonction analytique 2 » sans les points de dérivée nulle

Comme précédemment, nous constatons que, pour un nombre de simulations fixées à 40, la modélisation obtenue sans point de dérivée partielle nulle donne une moins bonne qualité de prédictivité que celle obtenue par la modélisation adaptative avec points de dérivée partielle nulle. En effet, les erreurs relatives sont respectivement de 0.026 et 0.041.

La contribution des points de dérivée partielle nulle est donc de grande importance dans l'approche de modélisation adaptative, la qualité de l'échantillonnage est grandement améliorée. L'ajout de ce type de point permet d'explorer des zones du domaine expérimental non-explorées par l'ajout des extrema de l'approximation courante.

6.6.3 Intérêt des points pilotes

Dans cette partie, nous mettons en évidence l'intérêt d'ajouter des points pilotes pour déterminer les nouveaux points à simuler dans les zones non-échantillonnées du domaine expérimental. Pour ce faire, plusieurs niveaux d'enrichissement du dispositif expérimental sont envisagés :

- niveau d'enrichissement 1 : enrichissement adaptatif du dispositif expérimental avec uniquement les points extrêmes et les points de dérivée partielle nulle, donc sans points pilotes.
- niveau d'enrichissement 2 : enrichissement adaptatif du dispositif expérimental avec les extrema, les points de dérivée partielle nulle et le point de variance de krigeage maximale. Il s'agit d'effectuer une simulation au point de variance de krigeage maximale, sans étude au préalable de l'information apportée par ce point selon la technique des points pilotes.
- niveau d'enrichissement 3 : enrichissement adaptatif du dispositif expérimental avec les extrema, les points de dérivée partielle nulle et le point de variance

maximale ayant la plus grande influence sur la prédictivité. Nous sélectionnons plusieurs points de variance de krigeage maximale, puis par l'approche des points pilotes décrite au chapitre 5, nous réalisons une étude de la sensibilité de chacun des points pour simuler celui dont une variation de la réponse (fictive) induit une variation sur la prédictivité de l'approximation.

Les résultats obtenus avec les trois méthodes partielles sont comparés à ceux obtenus par la modélisation adaptative.

– Fonction Sinus Décalée

Nous commençons ici par ajuster la fonction Sinus Décalée par la méthode partielle 1, à savoir uniquement avec des points maximum et minimum et des points de dérivée partielle nulle. Les résultats d'un tel ajustement sont présentés sur la Figure 6.23.

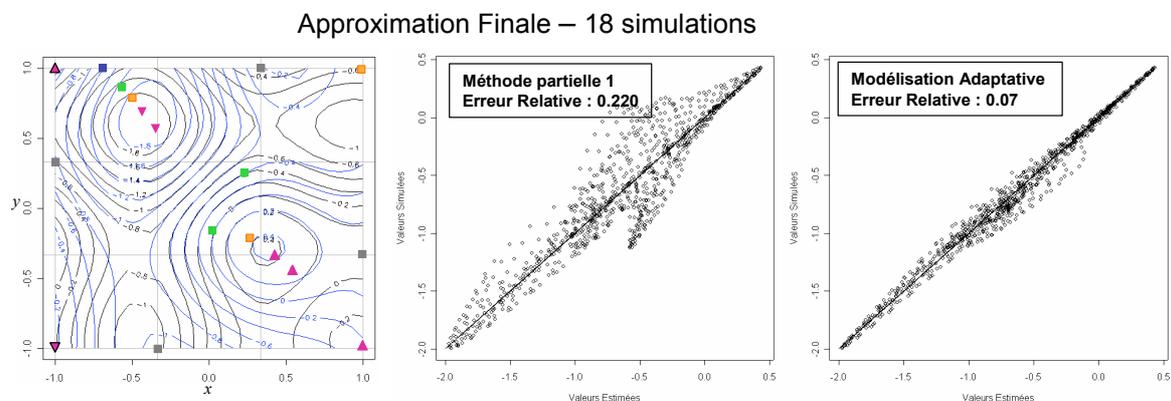


Figure 6.23 : Approximation de la fonction Sinus Décalée – Méthode partielle 1

Ainsi, dans le cas d'un ajustement n'impliquant que les points extrêmes ou de dérivée partielle nulle, nous constatons qu'il y a un sur-échantillonnage de certaines zones, alors que d'autres zones restent totalement inexplorées. En effet, au fur et à mesure de l'avancement du processus, les extrema locaux et points d'inflexion de la réponse sont identifiés et donc la recherche de nouveaux extrema conduit à des points trop proches du dispositif expérimental courant et donc ces points ne sont pas simulés pour éviter un sur-échantillonnage. Dans cet exemple, seulement 18 points ont été simulés. De plus, nous constatons que seuls deux extrema locaux ont été détectés, les autres ne sont pas identifiés.

Afin d'enrichir le dispositif expérimental dans les zones non investiguées du domaine expérimental, nous proposons, selon la méthode partielle 2, de réaliser une simulation au point de variance de krigeage maximale. Ainsi, à chaque itération, nous ajoutons les points extrêmes, les points de dérivée partielle nulle et le point de variance de krigeage maximale. Un tel ajustement, pour la fonction Sinus Décalée est présenté sur la Figure 6.24.

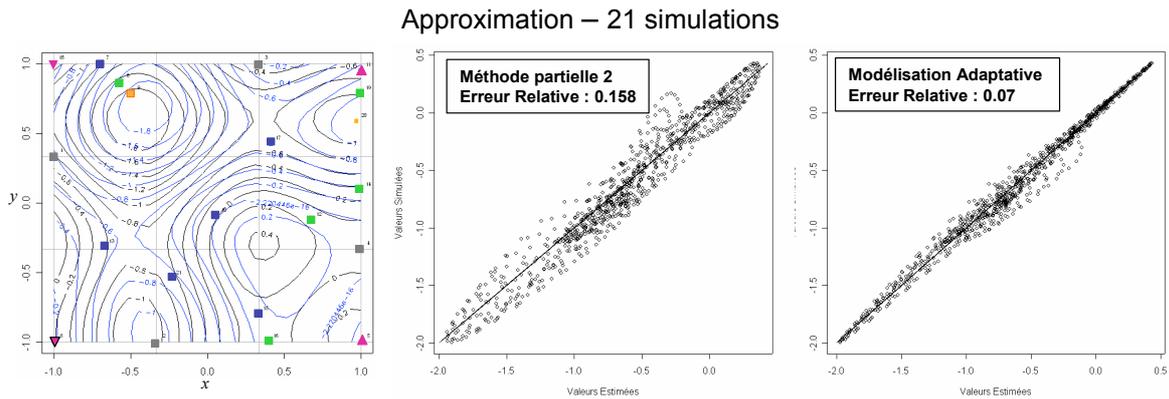


Figure 6.24 : Approximation de la fonction Sinus Décalée – Méthode partielle 2

Nous constatons ici une grande amélioration de la qualité d’approximation par rapport à celle de la méthode partielle 1, l’erreur relative passant de 0.22 à 0.158. Les zones non-échantillonnées sont ainsi explorées et permettent la détection d’un plus grand nombre d’irrégularités de la réponse. Toutefois, l’erreur relative reste bien plus importante que celle obtenue par modélisation adaptative.

Nous voulons à présent comparer les résultats de la méthodologie partielle 2 avec ceux de la méthode partielle 3. Nous rappelons que la méthode partielle 3 consiste à ne pas simuler directement le point de variance de krigeage maximale comme ceci est le cas pour la méthode partielle 2, mais de simuler le point, parmi plusieurs points de plus grande variance de krigeage, qui a la plus grande influence sur la prédictivité du modèle. Les résultats sont présentés sur la Figure 6.25.

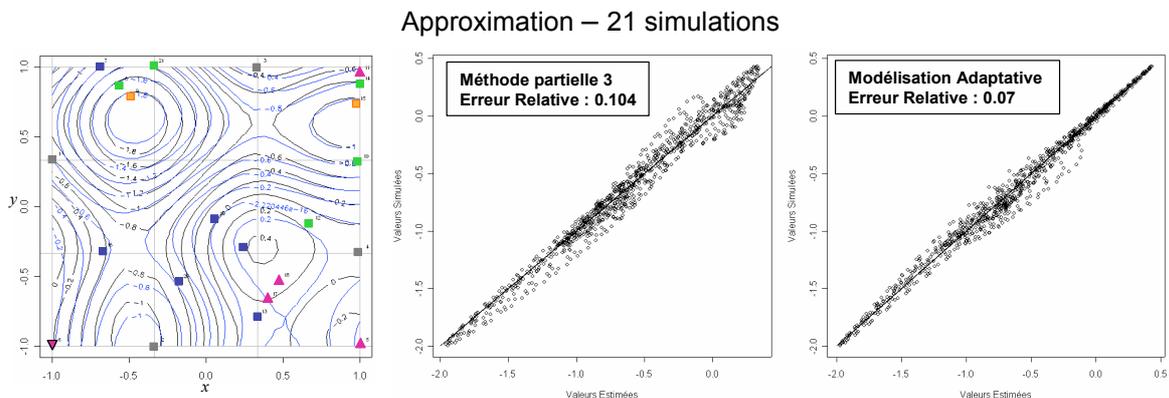


Figure 6.25 : Approximation de la fonction Sinus Décalée – Méthode partielle 3

L’enrichissement du dispositif expérimental par le point de variance de krigeage maximal n’est pas aussi informatif que l’ajout du point de plus grande influence sur la prédictivité de l’approximation. En effet, pour un même nombre de simulations, l’erreur relative est plus faible, passant de 0.158 à 0.104. Ainsi, la détermination d’un nouveau point à simuler par la méthode des points pilotes est capitale ici.

Toutefois, nous constatons que l'erreur relative obtenue par la méthode partielle 3 (0.104) est plus importante que celle issue de la modélisation adaptative (0.07). La seule différence entre les deux approximations est que pour l'ajustement avec modélisation adaptative, 6 points fictifs au maximum sont optimisés à chaque itération, alors que pour la méthode partielle 3, aucune valeur fictive n'est autorisée. Encore une fois, l'intérêt d'enrichir le dispositif expérimental avec de l'information fictive est confirmé. L'ajout de ces points avec valeurs fictives nous a permis d'améliorer la qualité de prédictivité de l'approximation, sans effectuer de simulations supplémentaires.

Montrons à présent les résultats obtenus sur la fonction analytique 2, plus complexe que la précédente.

– « Fonction analytique 2 » à 40 simulations

Comme précédemment, nous effectuons un premier ajustement de cette fonction par enrichissement du dispositif expérimental par la méthode partielle 1, c'est-à-dire en complétant le plan d'expériences uniquement par les extrema et les points de dérivée nulle de l'approximation courante. Les résultats sont présentés sur la Figure 6.26.

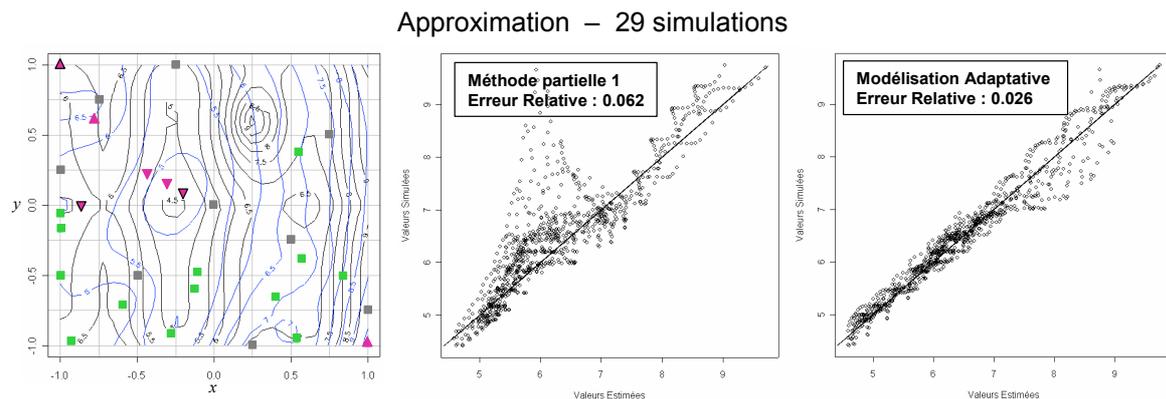


Figure 6.26 : Approximation de la « fonction analytique 2 » – Méthode partielle 1

L'enrichissement du dispositif par les critères prenant en compte uniquement les propriétés de l'approximation courante n'est pas suffisant et aboutit à un sur-échantillonnage de certaines zones alors que d'autres restent inexplorées. Ainsi, dans cet exemple, le maximum local de la fonction n'est pas détecté, aucune simulation n'a été possible dans cette zone, du fait de l'échantillonnage initial. De plus, seules 20 simulations supplémentaires ont pu être ajoutées au dispositif expérimental courant.

Étudions à présent les résultats obtenus par la méthode partielle 2, c'est-à-dire en ajoutant, à chaque itération, le point de variance de krigeage maximale.

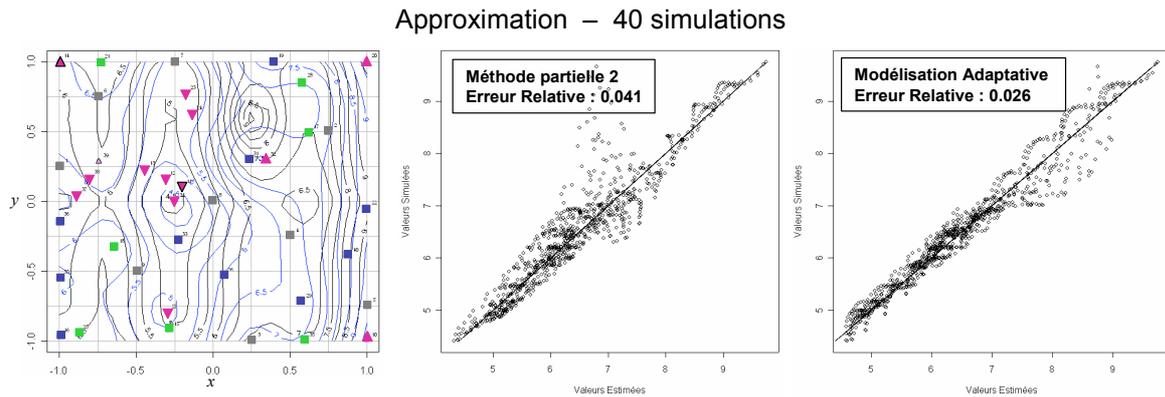


Figure 6.27 : Approximation de la « fonction analytique 2 » – Méthode partielle 2

Comme l'illustre la Figure 6.27, l'exploration des zones non-échantillonnées par le dispositif expérimental courant permet d'accroître grandement la qualité de l'approximation. Toutefois, enrichir le dispositif expérimental par les extrema, les points de dérivée nulle et le point de variance de krigeage maximal ne permet pas d'obtenir d'aussi bons résultats que ceux de la modélisation adaptative. En effet, l'approximation issue de la méthode partielle 2 présente une erreur relative de 0.041 alors que celle obtenue par la modélisation adaptative est de 0.026.

Enfin, nous proposons l'examen des résultats obtenus par la méthode partielle 3, à savoir en ne simulant pas directement le point de variance de krigeage maximale, mais celui, parmi plusieurs points de variance de krigeage maximale, qui a la plus grande sensibilité sur la prédictivité de l'approximation courante. Les résultats d'un tel ajustement sont présentés sur Figure 6.28.

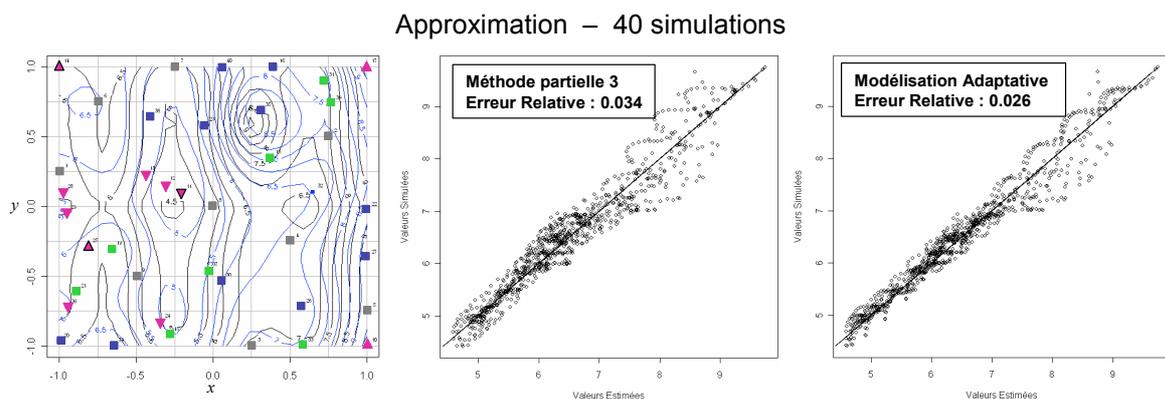


Figure 6.28 : Approximation de la « fonction analytique 2 » – Méthode partielle 3

L'intérêt de l'utilisation de la méthode des points pilotes pour déterminer le point du domaine expérimental à simuler est confirmé ici, puisque la qualité de l'approximation est largement accrue par rapport à celle obtenue par la méthode partielle 2. En effet, l'erreur relative obtenue respectivement avec les méthodes partielles 2 et 3 sont de 0.041 et 0.034. Elles restent toutefois bien supérieures à celle obtenue par la modélisation adaptative qui est de 0.026, cette méthode intégrant l'optimisation de six valeurs fictives à chaque itération.

Ainsi, nous avons prouvé l'intérêt de la méthode des points pilotes dans la méthodologie de modélisation adaptative par plans d'expériences, que ce soit pour la détermination du nouveau point à simuler, ou pour l'amélioration de la prédictivité de l'approximation par optimisation des valeurs fictives aux points pilotes. L'amélioration de la qualité de prédictivité de l'approximation est d'autant plus grande que nous sommes en début de processus itératif. En effet, un des intérêts de la méthode des points pilotes est d'accroître la qualité de prédictivité de l'approximation : il s'agit de stabiliser l'approximation.

L'ensemble des études présentées jusqu'à présent ont été réalisées sur des fonctions à deux paramètres. Nous présentons maintenant des résultats de modélisation pour un nombre de paramètres non-réguliers plus important.

6.7 Résultats de modélisation pour plus de deux paramètres

6.7.1 Exemple de modélisation pour trois paramètres

Nous considérons à présent une fonction non-régulière à trois paramètres. Cette fonction est une généralisation de la fonction « Chameau » présentée précédemment. Elle a pour expression :

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

Nous proposons d'ajuster cette fonction à partir d'un plan initial de type hypercube latin de distance maximin à 15 simulations. A chaque itération, nous ajoutons un maximum et un minimum avant l'éventuelle application de la méthode « iso-simplexe » et deux points de dérivée partielle nulle. De plus, nous effectuons une simulation aux trois points situés au maximum de la variance de krigeage qui sont de plus grande sensibilité vis-à-vis de la prédictivité. Nous optimisons au maximum dix valeurs de points pilotes à chaque itération. Nous considérons une fonction de covariance gaussienne anisotrope.

Afin de présenter les résultats d'un tel ajustement, nous proposons de représenter, sur la Figure 6.29, les courbes de niveaux de la fonction référence et de l'approximation en fixant la côte z . Ainsi, dans cet exemple, nous considérons quatre côtes. Nous présentons également le graphe des valeurs prédites en 1 000 points du domaine expérimental en fonction des valeurs de référence.

Fonction analytique 2 – 3D – 69 simulations

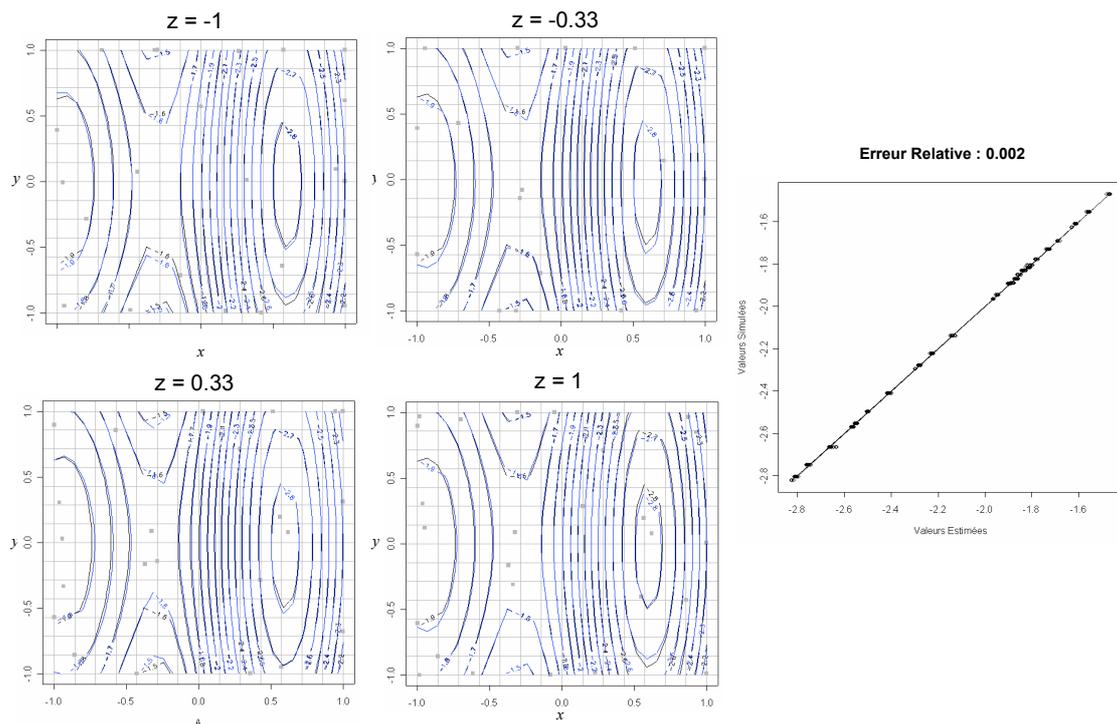


Figure 6.29 : Ajustement d'une fonction non-régulière à 3 paramètres

Ainsi, nous pouvons voir que la modélisation de cette fonction de trois variables est très satisfaisante, les courbes de niveaux sont très proches, pour chaque côte considérée, et l'erreur relative, pour 1000 points de prédiction est égale à 0.002. Un total de 69 simulations a été nécessaire pour aboutir à cette approximation.

6.7.2 Tableau et graphes récapitulatifs des résultats obtenus

Nous présentons dans cette section les résultats de la modélisation de fonctions non-régulières à trois, quatre et cinq paramètres. A cette fin, nous présentons de la Figure 6.30 à la Figure 6.32, les graphes de prédiction, pour respectivement :

- (a) la modélisation adaptative,
- (b) un hypercube latin de taille équivalente à l'approche évolutive,
- (c) un hypercube latin d'erreur relative équivalente à l'approche évolutive.

Fonction analytique 1 - 3D

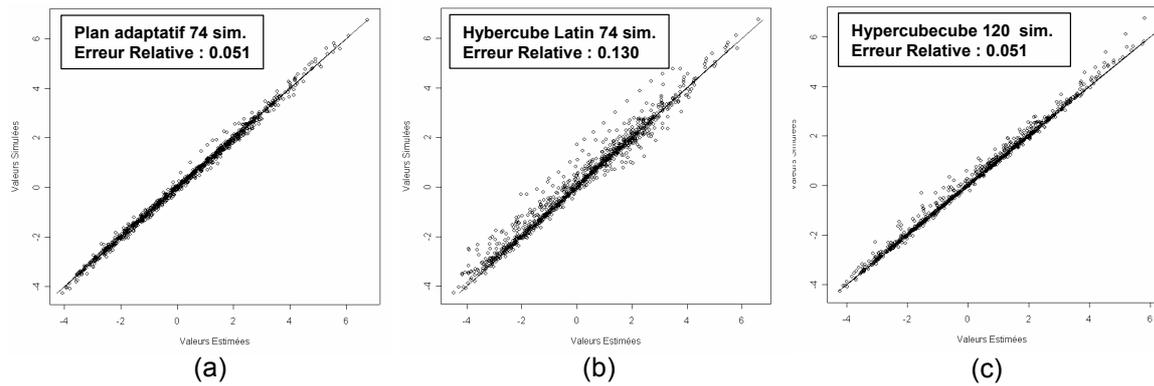


Figure 6.30 : Graphes de prédiction d'une fonction non-régulière à trois paramètres

Fonction analytique 2 - 4D

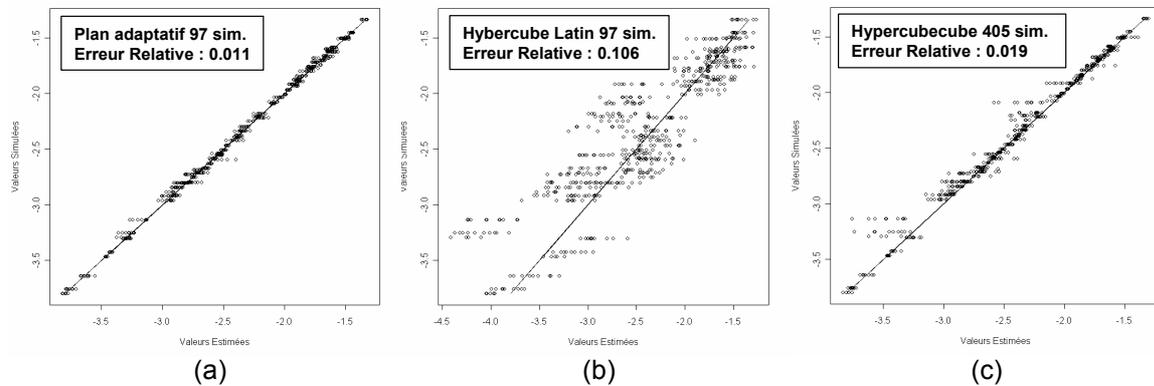


Figure 6.31 : Graphes de prédiction d'une fonction non-régulière à quatre paramètres

Fonction analytique - 5D

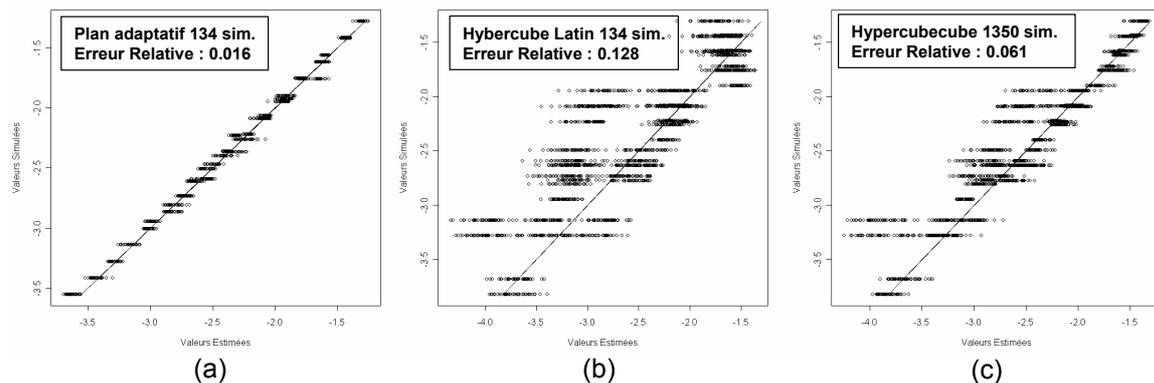


Figure 6.32 : Graphes de prédiction d'une fonction non-régulière à cinq paramètres

Ainsi, nous constatons que quelque soit le nombre de paramètres considérés, l'approximation par modélisation adaptative est de bien meilleure qualité. L'analyse des graphes de prédiction montre une grande dispersion des erreurs pour un hypercube latin de taille équivalente (comparaison graphes (a) et (b)), alors que pour obtenir une qualité de prédiction comparable, le nombre de simulations requis par un hypercube latin est beaucoup plus important (comparaison graphes (a) et (c)). No

Remarque : Au vu du très grand nombre d'hypercubes latin envisageables pour un nombre donné de simulations, nous avons effectué les modélisations sur une série d'hypercubes présentons ici les résultats de meilleure qualité.

Les erreurs relatives obtenues pour plusieurs fonctions sont précisées dans le Tableau 6-2, avec entre parenthèses le nombre de points du plan d'expériences. Nous y avons ajouté, dans la dernière colonne, le pourcentage de simulations gagnées grâce à l'utilisation de la méthode adaptative.

	Modélisation adaptative	Hypercube Latin même # sim.	Hypercube Latin même qualité	Simulations gagnées
Fonction analytique 1 - 3D	0.051 (74)	0.130 (74)	0.051 (120)	38%
Fonction analytique 2 - 3D	0.002 (69)	0.012 (69)	0.002 (110)	37%
Fonction analytique 1 - 4D	0.162 (100)	0.360 (100)	0.161 (345)	71%
Fonction analytique 2 - 4D	0.011 (97)	0.107 (97)	0.019 (405)	76%
Fonction analytique - 5D	0.016 (134)	0.128 (134)	0.061 (1350)	90%

Tableau 6-2 : Erreur relative et nombre de simulations nécessaires pour une modélisation à partir de plans évolutifs ou d'hypercubes latins.

D'après les résultats présentés dans le Tableau 6-2, il est clair que la modélisation d'une réponse non-régulière par la méthode évolutive est bien plus efficace que celle obtenue par échantillonnage par hypercube latin. En effet, à qualité de modélisation équivalente, l'utilisation de la modélisation adaptative permet de réduire le nombre de simulations de respectivement 37% 72% et 90%, pour trois, quatre et cinq paramètres, par rapport aux hypercubes latins. L'efficacité de l'approche adaptative est importante puisque dès que le nombre de paramètres augmente, le gain de simulations est accentué.

Une illustration du nombre de simulations nécessaires pour l'ajustement, à erreur relative donnée, de fonctions non-régulières pour la méthode évolutive et pour un hypercube latin, en fonction du nombre de paramètres, est présenté sur la Figure 6.33.

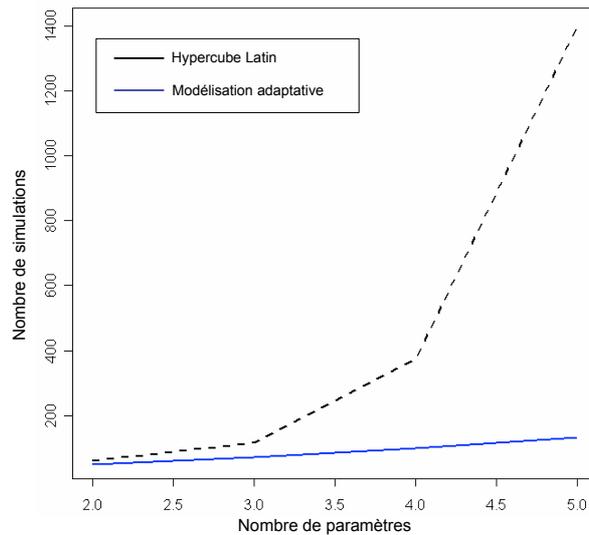


Figure 6.33 : Évolution du nombre de simulations en fonction du nombre de paramètres

Ainsi, plus le nombre de paramètres est important, plus le gain de simulations pour la modélisation de réponses non-régulières est important. En effet, plus on augmente la dimension du problème, plus les hypercubes sont creux, c'est-à-dire qu'il y a des zones non-explorées par le plan d'expériences. Ceci n'est pas le cas de la méthode de modélisation adaptative qui rajoute les points en fonction de l'approximation et donc les zones restées vides correspondent à des zones non-informatives.

6.8 Conclusions

Comme nous avons pu constater sur ces différents exemples, l'approche de modélisation adaptative proposée dans cette thèse donne des résultats très satisfaisants. Les critères de choix de la position des nouveaux points apparaissent tout à fait pertinents : ils assurent la détection des irrégularités de la réponse à estimer et permettent d'obtenir des résultats très précis, quelque soit le dispositif expérimental initial choisi. L'échantillonnage ainsi obtenu est spécifique à la réponse étudiée, il est optimal pour cette réponse. L'échantillonnage est informatif car permet de réduire considérablement le nombre de simulations nécessaires par rapport aux méthodes d'échantillonnage classiques. Le gain de simulation et donc l'efficacité de l'approche est d'autant important que le nombre de paramètres est grand, ce qui est généralement le cas en ingénierie de réservoir.

Chaque type de point proposé dans ce chapitre contribue à enrichir l'information apportée par les nouvelles simulations, l'ajout des extrema de l'approximation courante, et des points de dérivée partielle nulle sont indispensables à la détection des maxima, minima locaux et des points d'inflexions de l'approximation. De plus, la technique des points pilotes, innovante dans la modélisation d'expériences simulées, est d'intérêt crucial, aussi bien pour la détection des points à simuler que par l'ajout d'information fictive. Ainsi, parmi les points

situés dans les zones sous-échantillonnées du domaine, les points de plus grande sensibilité sur la prédictivité de l'approximation sont simulés. Ils apportent plus d'information que la simulation directe du point de variance de krigeage maximale. De plus, l'ajout d'information fictive permet de gagner en précision et prédiction sans avoir à faire de simulations supplémentaires.

Dans ce chapitre, nous avons validé la stratégie de modélisation adaptative sur des fonctions à plusieurs paramètres. Le chapitre suivant a pour objet d'appliquer cette méthode à deux cas spécifiques à l'ingénierie de réservoir.

Chapitre 7

Application à l'ingénierie de réservoir

Dans ce chapitre, nous proposons d'appliquer la méthode de modélisation adaptative à des problèmes rencontrés en ingénierie de réservoir. Nous étudions tout d'abord le comportement de la production pour un cas synthétique largement connu du milieu pétrolier, dérivé d'un cas réel : Punq-S3. Ce cas fut mis en place lors du projet européen PUNQ : Production forecasting with UNcertainty Quantification. La deuxième partie de ce chapitre est consacrée à la présentation d'un cas réel situé dans l'offshore brésilien : PBR. Ce cas appartient au patrimoine minier de la compagnie nationale brésilienne Petrobras.

Pour ces deux cas, nous proposons de construire un modèle approché du simulateur d'écoulement par modélisation adaptative. Il est ainsi possible d'estimer le risque lié à la présence d'incertitudes sur certains paramètres caractérisant la production de ces gisements à moindre coût.

Une modélisation basée sur des plans d'expériences classiques, factoriels et composites centrés, permet tout d'abord de justifier de l'intérêt d'envisager une prise en compte de non-régularités. Afin de juger de la pertinence des résultats obtenus par modélisation adaptative, nous avons par ailleurs procédé à la construction d'un modèle approché basé sur un échantillonnage par hypercube latin de même taille que celle requise lors de la modélisation adaptative. La performance prédictive des modèles adaptatif et par hypercube latin est alors comparée sur une grille de 6^4 points. Il apparaît alors clairement que le caractère spécifique de l'échantillonnage adaptatif permet une meilleure qualité de prédiction.

7.1 Cas Punq - S3

7.1.1 Présentation du cas Punq-S3

Le jeu de données utilisé dans cette étude provient d'un cas test défini dans le cadre du projet européen PUNQ : « Production forecasting with UNcertainty Quantification ». Le cas PUNQ-S3 est un réservoir de petite taille, dérivé d'un gisement exploité par Elf Exploration Production.

Le cas test PUNQ-S3 est un modèle réservoir synthétique 3D. Ce modèle est constitué de $19 \times 28 \times 5$ mailles, soit 2 660 mailles, parmi lesquelles 1 761 sont actives. Une vue de dessus du réservoir est présentée sur la Figure 7.1. Le réservoir est entouré par un fort aquifère au nord et à l'est, par des failles à l'ouest et au sud. Ce réservoir est exploité par 6 puits producteurs localisés au niveau du contact huile/eau. Il n'y a pas de puits injecteurs du fait de la présence d'un aquifère actif.

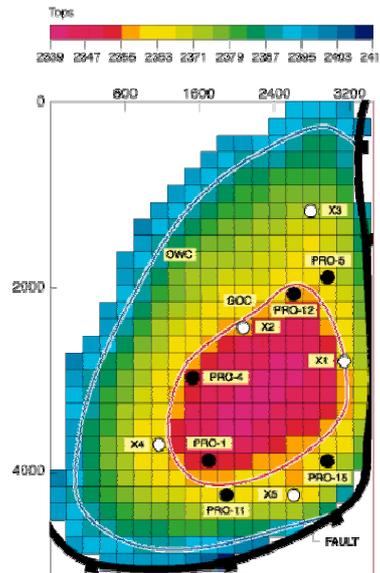


Figure 7.1 : Modèle PUNQ

Les sédiments ont été déposés dans un environnement deltaïque et côtier. Le modèle géologique est composé de cinq couches non communicantes, chaque couche ayant une épaisseur d'environ 5 mètres.

- Les couches 1, 3 et 5 possèdent de bonnes propriétés en terme de perméabilité et porosité. Ces couches sont dites réservoir.
- La couche 2 est de faible porosité, il s'agit d'une couche non-réservoir.
- La couche 4 est de qualité intermédiaire.

La distribution de la porosité dans les 5 couches a été modélisée par simulation géostatistique, basée sur des réalisations aléatoires gaussiennes. Les propriétés pétrophysiques moyennes de chaque couche sont présentées dans le Tableau 7-1.

Couches	Phi - %	Kh - mD	Kv - mD
1 - 3 -5	30	1000	500
2	17	220	50
4	22	500	100

Tableau 7-1 : Propriétés pétrophysiques par couches

7.1.2 Analyse d'incertitude

L'objectif de cette étude est de construire un modèle approché du simulateur d'écoulement, pour une réponse donnée. Ce modèle sera ensuite utilisé pour obtenir des prévisions de production sans faire appel au simulateur d'écoulement, dans le but de réaliser une analyse de risque et donc d'estimer le risque, sur la production d'huile, lié à une incertitude sur les paramètres.

- **Définition des paramètres**

La sélection des paramètres *a priori* influents sur le comportement des réponses en production a résulté d'une analyse de sensibilité faisant intervenir 11 paramètres incertains. Les paramètres pour lesquels l'analyse de sensibilité a mis en évidence une influence non négligeable sont au nombre de quatre :

- le coefficient analytique de la force d'aquifère, noté AQUI,
- la saturation résiduelle en huile après balayage à l'eau, notée SORW,
- la pression au point de bulle, notée BP,
- la profondeur du contact huile/eau, notée ZOW.

Le domaine expérimental est défini par :

$$D = \prod_{j=1}^4 [X_j^{\min}, X_j^{\max}]$$

où X_j^{\min} et X_j^{\max} sont les valeurs minimales et maximales que peuvent prendre les variables AQUI, SORW, BP et ZOW. Ces valeurs sont données dans le Tableau 7-2 :

Paramètre	Min	Max
AQUI (adim)	0.2	0.3
SORW (%)	0.15	0.25
BP (bar)	230.5	238.5
ZOW (m)	2385	2398

Tableau 7-2 : Intervalle de variation des paramètres incertains

- **Définition de la réponse**

Notre objectif est la prédiction du débit d'huile au puits PRO-1 après 2 922 jours de production quand on fait varier les différents paramètres supposés influents sur cette réponse. Ce débit, aussi noté QOil, est exprimé en m³/jours.

Ainsi, nous voulons construire, par modélisation adaptative, une approximation fiable du débit d'huile QOil en fonction des quatre paramètres AQUI, SORW, BP et ZOW.

Avant de chercher à ajuster cette réponse, nous avons effectué des simulations pour un plan complet 6^4 , soit 1 296 simulations, bien que ceci ne soit généralement pas envisageable pour une étude de réservoir. Mais notre objectif est ici de disposer ainsi d'un échantillon « de référence » du comportement de la réponse en fonction des paramètres. Nous présentons sur la Figure 7.2, différentes graphes partiels de la réponse QOil qui illustrent le type de non-régularités présentes dans cette étude.

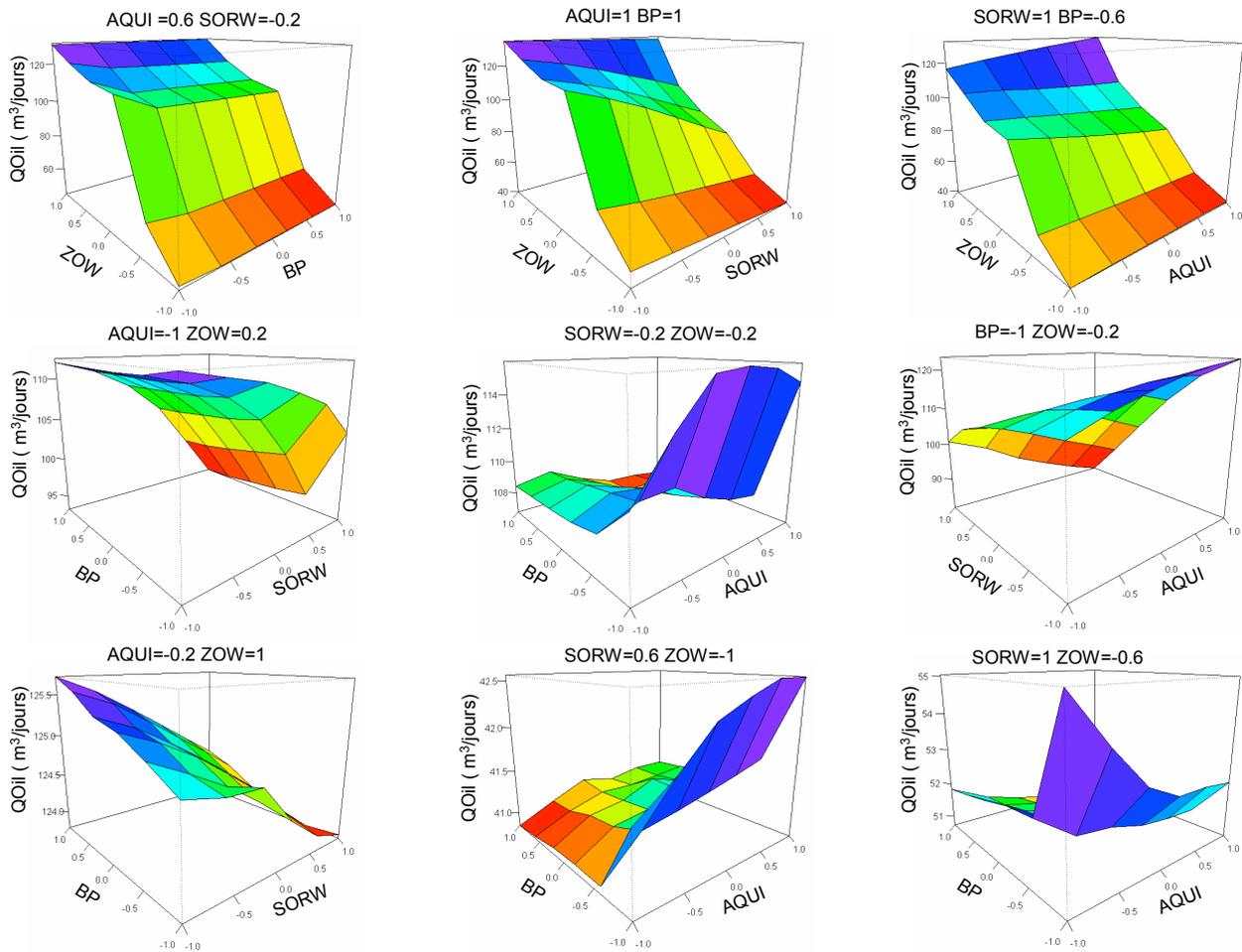


Figure 7.2 : Vue en perspective de différentes coupes du débit d'huile au puits PRO-1 après 2 992 jours

Ainsi, nous constatons que le débit d'huile au puits PRO-1 après 2 922 jours de production est non-régulier par rapport aux paramètres considérés. Nous observons trois niveaux de non-régularité. En effet, sur la première ligne de la Figure 7.2, nous constatons de très grandes variations de la réponse, celle-ci pouvant varier de plus de 90 000 m³ pour deux paramètres fixés. Pour la seconde ligne, nous avons représenté des graphes partiels où la variation de la réponse est intermédiaire, soit environ 10 000 m³. Enfin, il est important de noter les irrégularités de plus faible impact sur la valeur de la réponse, elles sont de l'ordre de

quelques milliers de m³, mais sont sources des principales difficultés pour l'obtention d'une modélisation de bonne qualité.

Nous proposons à présent de modéliser le débit d'huile au puits PRO-1 après 2 922 jours de production par modélisation adaptative.

7.1.3 Approximation par modélisation adaptative

Pour effectuer l'ajustement par modélisation adaptative de QOil, nous partons d'un plan initial qui est un hypercube latin de distance maximin de 25 simulations. Nous supposons ici une covariance gaussienne anisotrope. En ce qui concerne l'enrichissement du dispositif expérimental, nous ajoutons, à chaque itération :

- 1 maximum avant l'éventuelle application de la méthode « iso-simplexe »,
- 1 minimum avant l'éventuelle application de la méthode « iso-simplexe »,
- 3 points de dérivée partielle nulle,
- 3 simulations sélectionnées par la méthode des points pilotes dans les zones maximisant la variance de krigeage.

Remarquons que 10 points pilotes au maximum sont conservés dans le dispositif expérimental.

De plus, dès lors que la qualité de l'approximation apparaît satisfaisante, nous ajoutons cinq points de confirmation afin de juger de la prédictivité de l'approximation et de la nécessité de stopper ou de continuer le processus itératif.

Enfin, nous limitons le nombre maximum de simulations à 150, ce type de contrainte est généralement imposé en ingénierie de réservoir.

Nous présentons, sur la Figure 7.3, les résultats d'approximation du débit d'huile QOil avec l'évolution des graphes des erreurs de prédiction aux 1 296 points du plan complet 6⁴ au cours des itérations. Les points correspondant au plan d'expériences servant à la modélisation sont représentés par des croix bleues, les points de prédiction par des cercles violets. Pour indiquer la qualité du modèle, nous ajoutons, sur ce graphe, l'erreur relative obtenue aux 1 296 points de prédiction, soit :

$$\varepsilon_{rel} = \frac{\sum_{i=1}^{1296} |Y_i - \hat{Y}(\mathbf{x}_i)|}{\sum_{i=1}^{1296} |Y_i|}$$

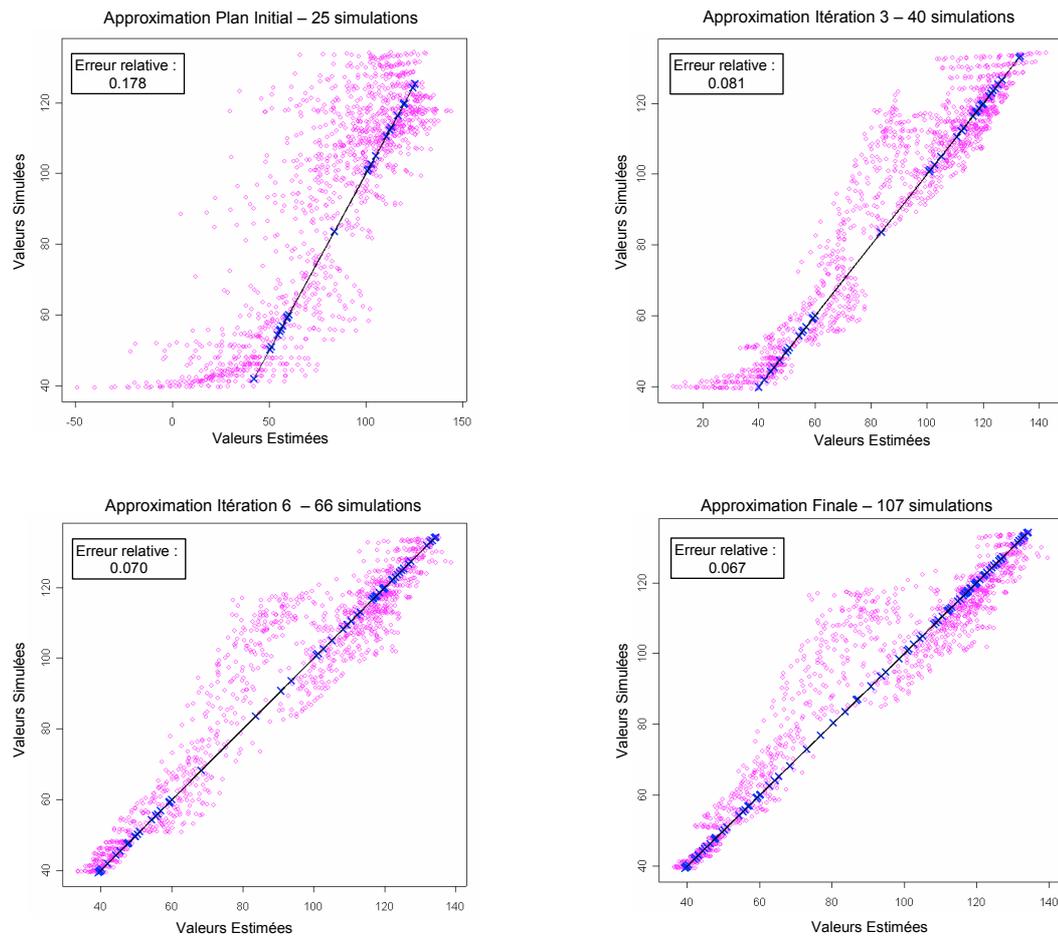


Figure 7.3 : Approximation de QOil par modélisation adaptative au cours des itérations

Ainsi, un nombre total de 107 simulations a été ici nécessaire pour l’ajustement de QOil. Nous constatons, d’après la Figure 7.3, que l’ajustement issu du plan d’expériences initial (un hypercube latin maximin) est de très mauvaise qualité, avec des prédictions de débits négatives. Notons en effet le changement d’échelle pour ce premier graphe. Au fur et à mesure de l’ajout de points, la qualité de l’approximation est améliorée, jusqu’à atteindre une erreur relative aux 1 296 points de prédiction égale à 0.067. Remarquons que la qualité d’approximation est déjà très satisfaisante pour un dispositif de 66 simulations, la prédiction des valeurs du débit d’huile comprises entre 40 et 80 m³/jour reste toutefois meilleure avec un dispositif à 107 simulations.

Nous présentons, sur la Figure 7.4, l’évolution des densités de probabilité au cours des itérations. Les densités de probabilité sont calculées à partir d’un échantillonnage Monte-Carlo résultant du tirage d’une loi uniforme entre -1 et 1 pour les quatre paramètres. L’échantillon comporte 10 000 points. Ainsi, nous représentons en noir la densité estimée à l’aide du plan initial, en bleu et vert celles issues respectivement de la troisième et sixième itération et enfin en rouge les résultats de la modélisation adaptative finale.

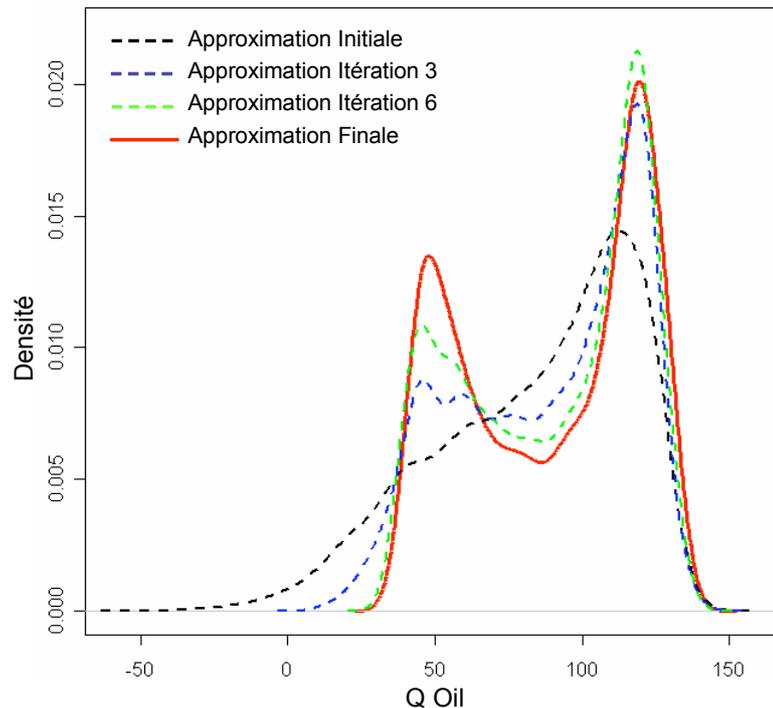


Figure 7.4 : Évolution des densités de probabilité du débit d'huile au cours des itérations

Ainsi, nous constatons que l'approximation déduite du plan initial ne restitue pas la bimodalité de la densité et qu'il y a des estimations de débit négatives. A la troisième itération, le caractère bimodal commence à apparaître dans l'approximation, il devient beaucoup plus précis pour les itérations suivantes. Remarquons enfin que, malgré la faible différence entre les erreurs relatives des résultats issus de la sixième itération et ceux de l'approximation finale, une nette amélioration de l'estimation de la densité de probabilité est observée. Après étude des valeurs des paramètres, il apparaît que l'effet bimodal correspond à la valeur d'un seul paramètre : ZOW. En effet, pour de petites profondeurs du contact huile/eau, le débit d'huile QOil est relativement faible, alors que pour de plus grandes profondeurs, il est plus important.

7.1.4 Modélisation adaptative et planification classique

Dans cette partie, nous proposons de modéliser le débit d'huile au puits PRO-1 à 2 992 jours de production, donc toujours la même réponse QOil, à partir de plans d'expériences classiques. Nous considérons dans un premier temps un plan factoriel et un plan composite centré, puis deux hypercubes latins de même taille que le nombre de simulations auquel aboutit la modélisation adaptative. Pour chaque étude, nous présentons les résultats obtenus pour un modèle de krigeage universel afin de comparer les approximations à modélisation fixée et donc de vérifier l'importance de la stratégie d'échantillonnage du domaine expérimental. Afin de comparer ces différents dispositifs, nous étudions tout d'abord la qualité des prédictions obtenues pour les différents plans avec les simulations d'une grille de

référence 6^4 . Ainsi, nous représentons les valeurs estimées versus les valeurs simulées pour cette grille, pour chacun des dispositifs expérimentaux considérés. Ensuite, nous présentons une étude comparative de la densité estimée de la réponse : d'une part à partir des 6^4 simulations servant de référence et d'autre part par approximation par krigeage universel pour des dispositifs classiques puis par modélisation adaptative, puis échantillonnage quasi Monte-Carlo (hypercubes latins).

Nous présentons, sur la Figure 7.5, les résultats d'approximation obtenus à partir d'un plan factoriel (a) et d'un plan composite (b).

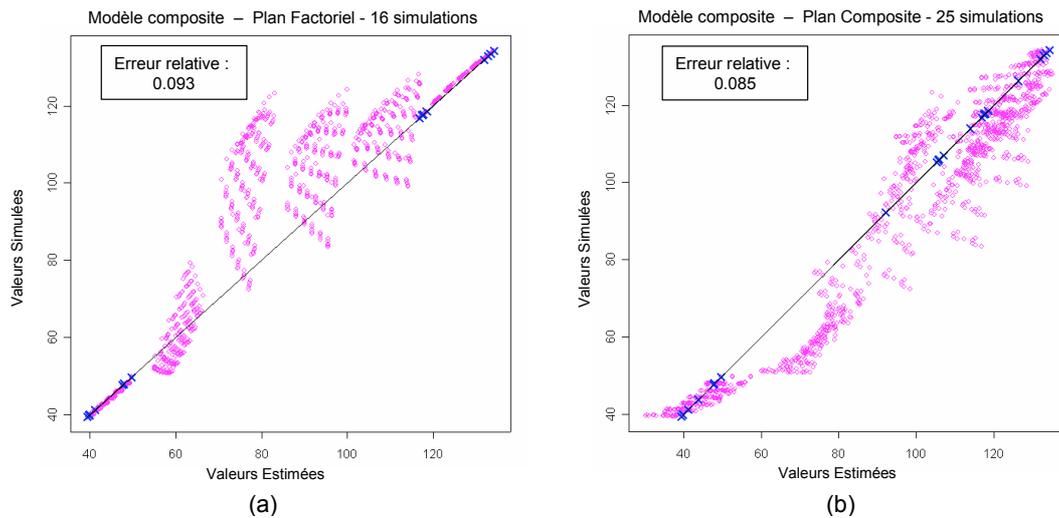


Figure 7.5 : Prédiction du débit d'huile par plans factoriel et composite

Nous constatons que la modélisation du débit d'huile au puits PRO-1 à partir de dispositifs expérimentaux classiques ne permet pas une estimation précise de cette réponse. En effet, dans le cas d'un plan factoriel (a), une importante sous-estimation du débit d'huile est constatée, les points de prédiction étant généralement au dessus de la première bissectrice qui traduit un ajustement parfait. Pour le plan composite centré (b), on observe plutôt une sur-estimation du débit d'huile au puits PRO-1, un grand nombre de points de prédiction étant situés sous la première bissectrice.

Voyons à présent l'impact de la mauvaise qualité des approximations obtenues pour les plans factoriel et composite centré sur les estimations de densité de probabilité. Dans ce but, nous présentons, sur la Figure 7.6, les densités de probabilité déduites du plan complet, i.e. des 6^4 simulations, que l'on considère comme référence (en noir), pour le plan factoriel, en bleu, le plan composite, en vert et enfin pour la modélisation adaptative en rouge. Les densités de probabilité sont obtenues par échantillonnage Monte-Carlo, pour un tirage de 10 000 points issus de lois uniformes sur chaque paramètre.

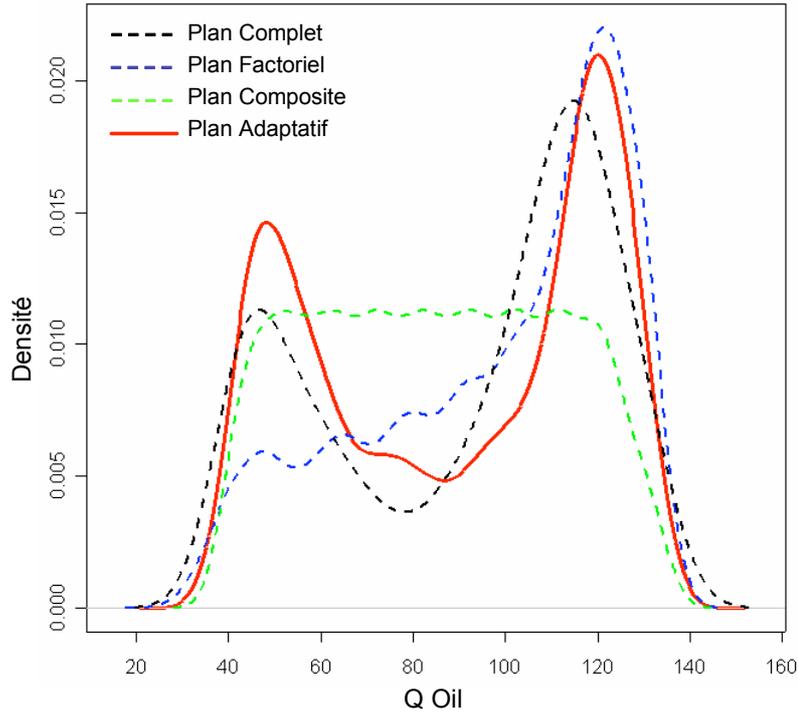


Figure 7.6 : Densité de probabilité du débit d'huile au puits PRO-1 pour différents dispositifs expérimentaux

Au vu de la Figure 7.6 ci-dessus, il apparaît clairement que l'utilisation de plans factoriel ou composite centré ne permet pas une estimation de bonne qualité de la densité de probabilité de la réponse QOil, contrairement aux résultats obtenus par modélisation adaptative. En effet, le caractère bimodal de la densité n'est absolument pas détecté par les plans classiques et les analyses de risque en découlant ne sont donc pas fiables.

Passons maintenant, toujours à titre comparatif, à des plans d'échantillonnage quasi Monte-Carlo classiques, plus précisément à des hypercubes latins de même taille que celle du dispositif déduit de la modélisation adaptative, soit 107 simulations. Le type de modélisation utilisé ici est un krigeage universel, la modélisation adaptative n'intervient pas ici.

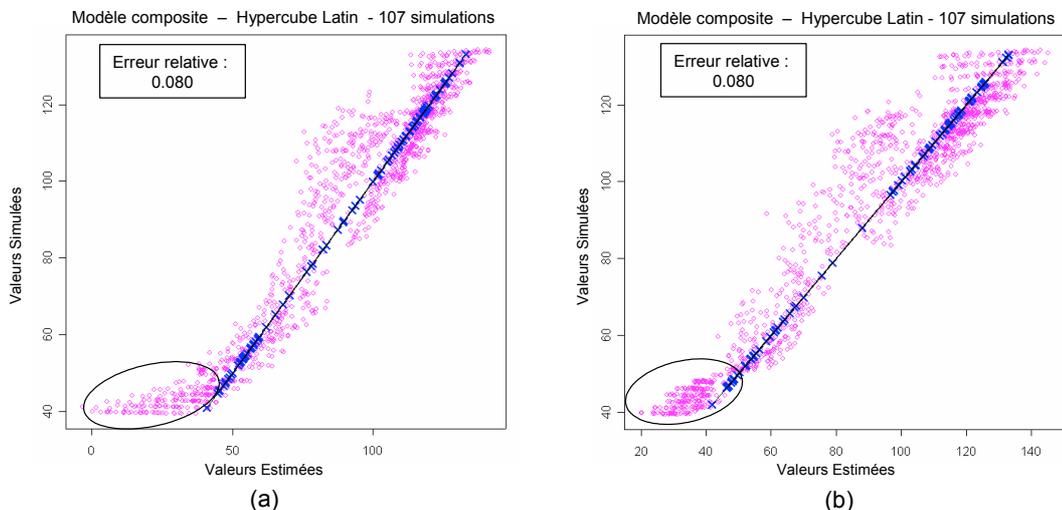


Figure 7.7 : Prédiction du débit d'huile par hypercubes latins

Ainsi, pour les deux hypercubes considérés ici, les approximations obtenues aboutissent à des erreurs d'estimation non négligeables. Nous constatons par exemple que pour les faibles valeurs de débit, des sous-estimations importantes peuvent être obtenues, avec des débits au puits PRO-1 allant de 0 à 20 m³/j, alors que le puits produit au minimum 40 m³/j.

Une estimation des densités de probabilité obtenues à partir des deux hypercubes latins maximin et de la modélisation adaptative sont présentées sur la Figure 7.8. L'échantillonnage considéré contient 10 000 points issus d'un tirage selon des lois uniformes sur chaque paramètre.

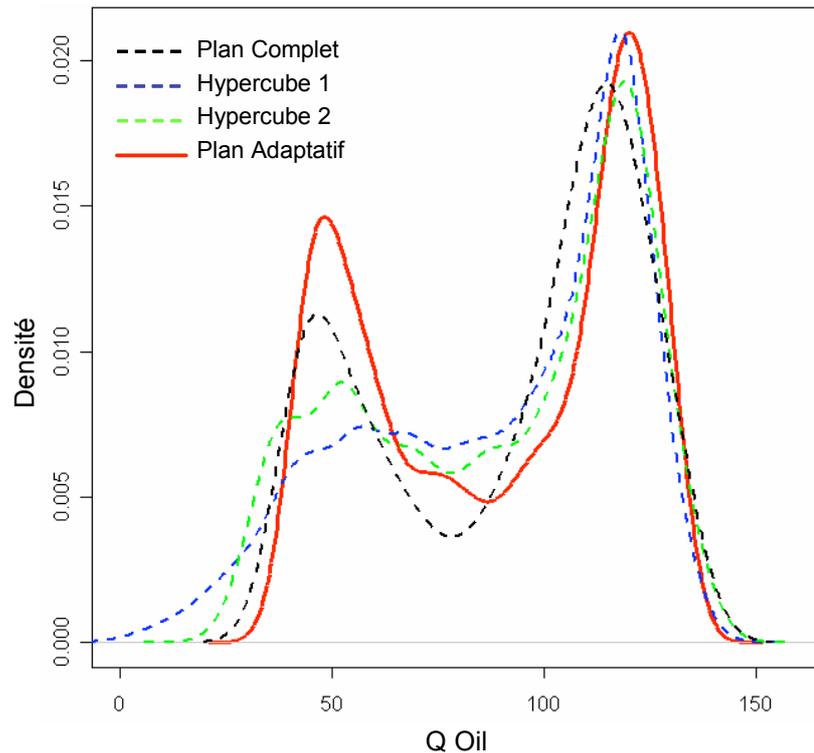


Figure 7.8 : Densité de probabilité du débit d'huile pour différents dispositifs expérimentaux

En se basant sur les résultats obtenus à l'aide du plan complet 6⁴, il apparaît très clairement que les résultats issus de la modélisation adaptative sont plus précis que ceux provenant des deux hypercubes latins. En effet, si on considère l'hypercube 1, en bleu, le caractère bimodal de la densité est à peine détecté, nous avons également des valeurs de débit nulles. Quant à l'hypercube 2, présenté en vert, il permet une meilleure estimation de la densité de probabilité de la réponse QOil, que pour l'hypercube 1, mais confirme ainsi le caractère aléatoire des estimations par hypercubes latin puisqu'il y a non-unicité de ce type d'échantillonnage.

Remarque : Pour confirmer les résultats ci-dessus, nous avons réalisé les mêmes analyses en partant d'hypercubes latin maximin à 300 points. Nous constatons que les résultats diffèrent très peu de ceux obtenus pour des hypercubes latins maximin à 107 points, la modélisation adaptative est de meilleure qualité dans tous les cas.

Ainsi, les plans d'expériences ou d'échantillonnage classiques ne permettent pas une quantification fiable des incertitudes et par conséquent une modélisation adaptative est appropriée pour l'approximation du débit d'huile au puits PRO-1.

7.1.5 Conclusions

Les résultats de l'étude du débit d'huile au puits PRO-1 après 2 922 jours de production ont mis en évidence l'efficacité de la modélisation adaptative pour la quantification des incertitudes et l'analyse de risque. En effet, la comparaison des résultats avec ceux obtenus par des techniques classiques d'échantillonnage a montré une plus grande précision pour les résultats de la modélisation adaptative. Par ailleurs, l'utilisation d'hypercubes latins de même taille que les dispositifs issus de la modélisation adaptative ne permet pas d'obtenir une aussi bonne quantification des risques et reste de plus très aléatoire, du fait du nombre prohibitif d'hypercubes latins existants.

7.2 Cas PBR

7.2.1 Présentation du cas PBR

Le champ PBR est situé dans l'offshore brésilien, dans le bassin de Campos, à 80 km de la côte, sous une tranche d'eau comprise entre 110 et 250 m. La profondeur moyenne du réservoir est de 3100 m. Avec une surface d'environ 20 km² et une épaisseur qui varie entre 90 et 180 m, il possède un volume d'huile estimé à 110 millions de mètres cubes. On peut espérer une récupération finale d'environ 40% de ce volume. Une représentation en 3D du champ PBR est présentée sur la Figure 7.9.

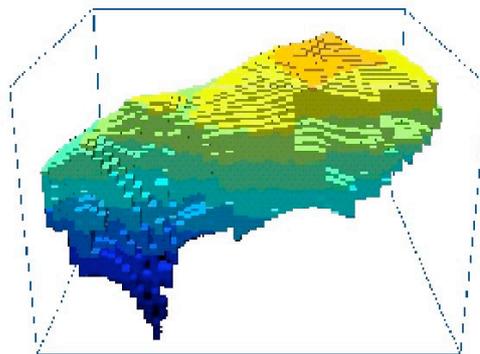


Figure 7.9 : Vue en 3 dimensions du champs PBR

Le réservoir est formé de grès massifs d'origine turbiditique alternant avec des argiles, des marnes et des calcilutites. Ces derniers faciès sont « non-réservoirs » car ils sont de perméabilités très faibles, ils jouent un rôle très important sur le mouvement des fluides dans

Faciès	Phi - %	Kh - mD	Kv – mD
LT-1	30	1100	100
LT-2	24	550	80
LT-3	10	10	0.1
LT-4	6	5	0.1

Tableau 7-3 : Propriétés pétrophysiques par faciès

Les propriétés pétrophysiques des lithotypes sont prises en compte par modélisation géostatistique, plus particulièrement avec la méthode de simulation gaussienne tronquée non-stationnaire.

Un modèle Black-Oil est utilisé pour les simulations d'écoulement. Dans le simulateur d'écoulement, le débit total de fluide (huile + eau) du puits dépend de plusieurs facteurs. Parmi ces facteurs, on trouve la capacité de production du puits. En effet, un débit d'huile est donné au simulateur comme une valeur du débit maximal à atteindre. Dans cette étude, le débit maximal imposé est de 500m³/j.

Le point de bulle initial est de 240 bars et la pression initiale moyenne du réservoir est d'environ 318 bars. La température du réservoir est égale à 104°C. Le contact huile/eau principal du réservoir se situe environ à -3 100 m. Néanmoins, d'autres contacts ont été observés pendant le forage des puits, par exemple, dans la région sud du réservoir, le contact huile/eau a été trouvé à -3 125 m et dans la région nord-est, à -3 152 m.

7.2.2 Analyse d'incertitudes

L'objectif de cette étude est de construire un modèle approché du simulateur d'écoulement, pour une réponse donnée. Ce modèle est construit dans le but de réaliser une analyse de risque afin d'estimer le risque lié à une incertitude sur les paramètres sur la production d'huile et d'eau.

- **Définition des réponses**

Nous nous intéressons ici particulièrement à l'estimation de la production d'huile et d'eau cumulée après 2 192 jours de production, soit environ six ans. Nous désignons désormais ces deux réponses par :

- CumOil pour la production d'huile cumulée à 2 192 jours de production,
- CumWater pour la production d'eau cumulée à 2 192 jours de production.

Les valeurs de ces réponses sont exprimées en millions de mètres cubes (Mm³).

- **Définition des paramètres**

Parmi les 18 paramètres susceptibles d'intervenir dans la modélisation des réponses en production, on a procédé à une sélection à l'issue d'une analyse de sensibilité. On a ainsi retenu quatre paramètres, il s'agit de :

- la profondeur du contact huile/eau dans le secteur 1, noté ZOW1,
- le coefficient d'endommagement du puit, noté SKIN,
- la perméabilité relative maximale pour le gaz, notée KRGM,
- la perméabilité relative maximale pour l'eau, notée KRWM.

Le SKIN, aussi appelé effet pariétal permet de modéliser le fait que la boue de forage provoque un colmatage partiel de la zone envahie, d'où une diminution ou une augmentation de la perméabilité dans cette zone. Ainsi, un SKIN positif signifie que les abords des puits ont été colmatés, un SKIN négatif montre une amélioration de la perméabilité aux abords des puits.

Remarquons que seule la profondeur du contact huile/eau de la zone 1 varie, les autres sont supposées constantes, égales à leur valeurs indiquées plus haut.

Le domaine expérimental est le pavé de \square^4 :

$$D = \prod_{j=1}^4 [X_j^{\min}, X_j^{\max}]$$

où X_j^{\min} et X_j^{\max} sont les valeurs minimales et maximales que peuvent prendre les variables ZOW1, SKIN, KRGM et KRWM. Ces valeurs sont données dans le Tableau 7-4 :

Paramètre	Min	Max
ZOW1 (m)	3 092	3 118
SKIN (adim)	-5	10
KRGM (adim)	0.4	0.6
KRWM (adim)	0.25	0.45

Tableau 7-4 : Intervalle de variation des paramètres incertains

Ainsi, notre objectif est de construire, par modélisation adaptative, une approximation fiable des deux réponses CumOil et CumWater en fonction des quatre paramètres : ZOW1, SKIN, KRGM et KRWM.

Dans un premier temps, nous avons effectué des simulations pour un plan complet 6^4 , soit 1 296 simulations. Bien que ce nombre de simulations ne soit généralement pas envisageable dans le contexte d'une étude de réservoir, nous les avons réalisées ici dans le but

de confirmer la non-régularité des réponses aux paramètres considérés. Nous présentons ci-dessous des exemples de graphes partiels de la production d'huile et d'eau cumulées après 2 192 jours de production en fonction des paramètres considérés.

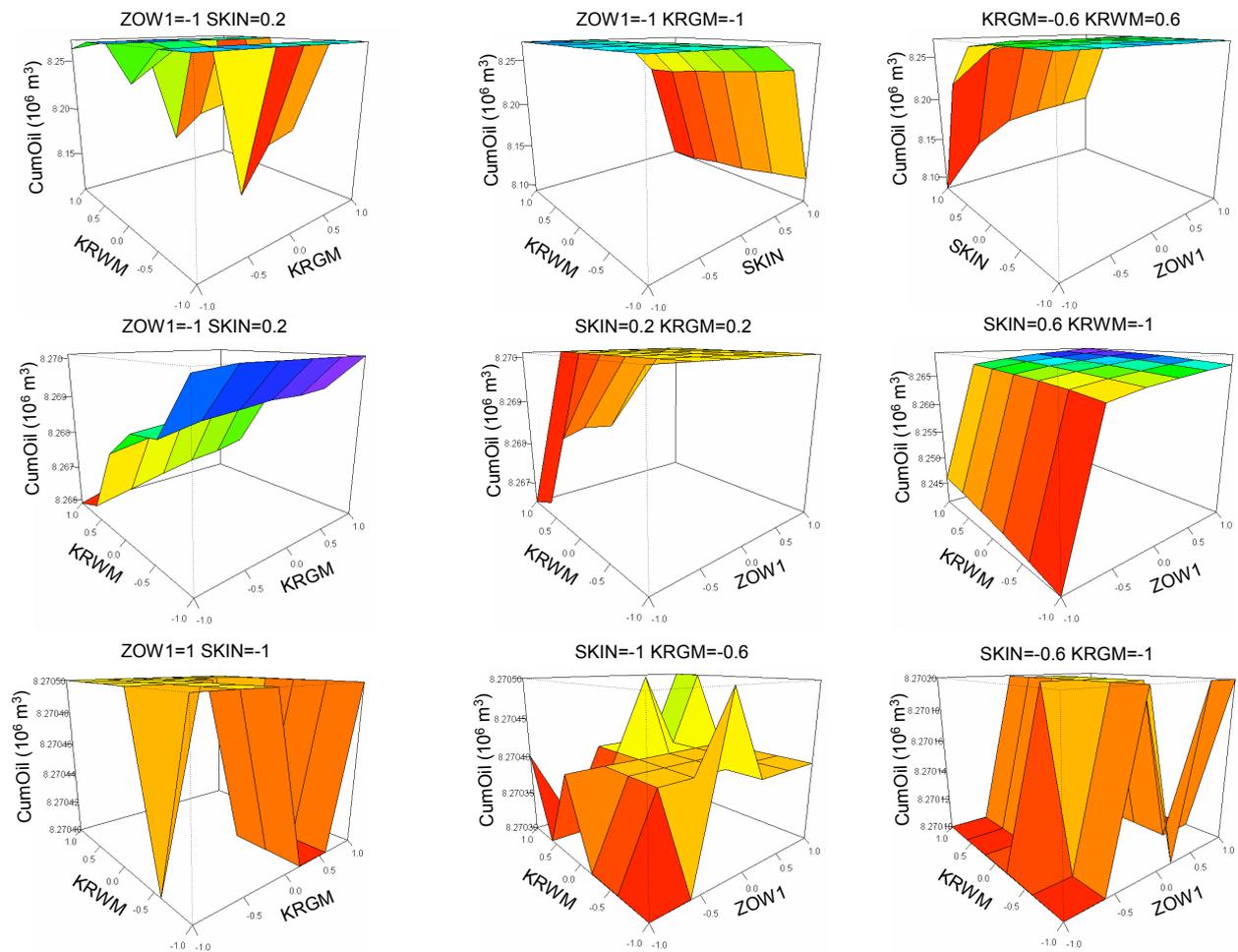


Figure 7.11 : Vue en perspective de graphes partiels du Cumulé d'huile à 2 192 jours

Ainsi, pour la réponse CumOil, nous observons trois types de non-régularité. En effet, sur la première ligne de la Figure 7.11, nous observons des variations globales de la réponse en fonction des paramètres, elles peuvent atteindre $150\,000\text{ m}^3$. Nous observons, sur la deuxième ligne, des variations intermédiaires, c'est-à-dire de l'ordre de $30\,000\text{ m}^3$. Enfin, sur la dernière ligne sont représentées les variations locales de la réponse, on observe ainsi un très grand nombre d'irrégularités.

Passons à présent à la réponse CumWater.

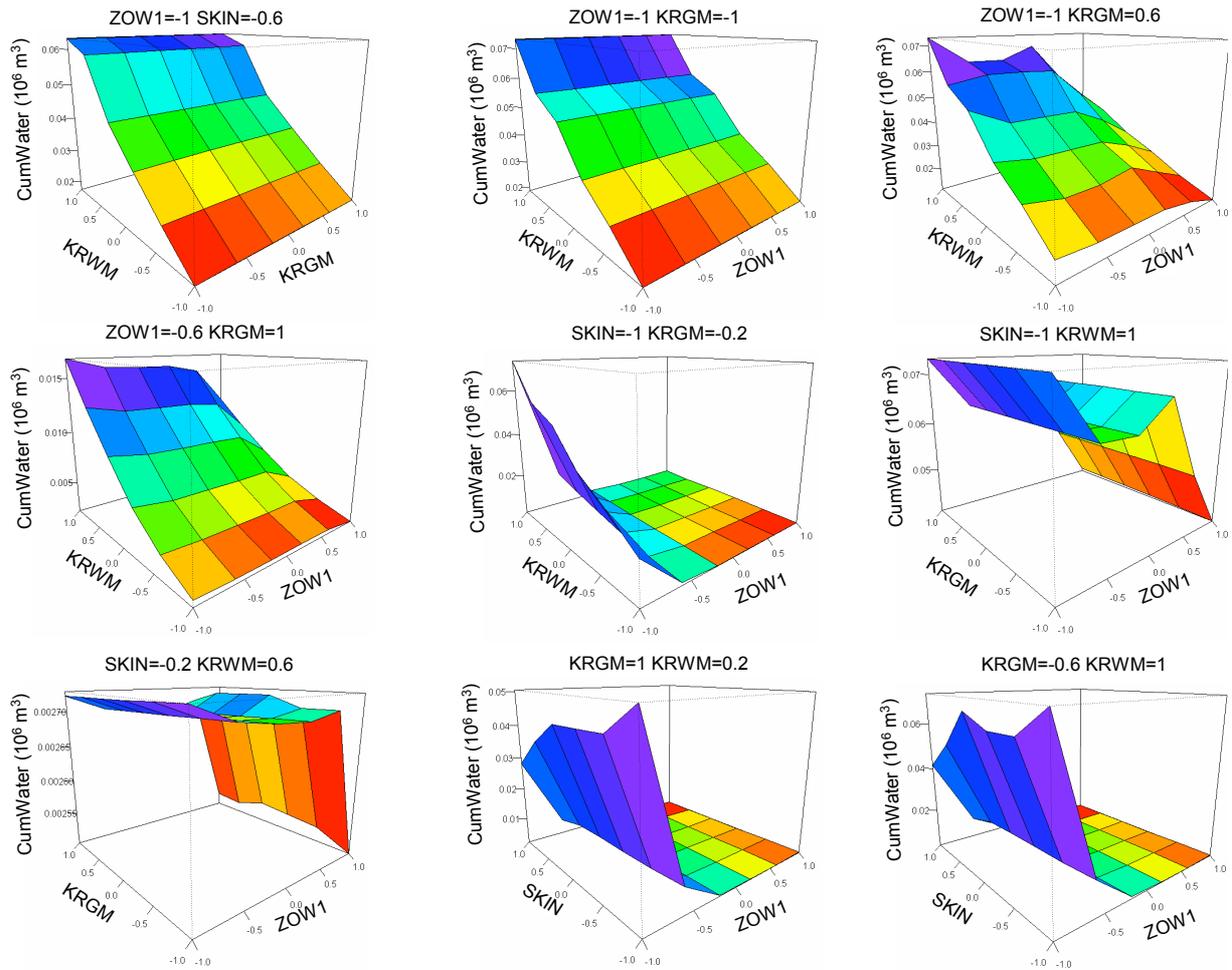


Figure 7.12 : Vue en perspective de graphes partiels du Cumulé d'eau à 2 192 jours

Encore une fois, nous observons des irrégularités de la réponse CumWater aux paramètres la caractérisant.

Ainsi, comme l'illustre la Figure 7.11 et la Figure 7.12, quelle que soit la réponse considérée, son comportement est non-régulier aux paramètres. De plus, les non-régularités ne sont pas de même nature suivant les paramètres. Par exemple nous constatons la présence d'un « plateau », c'est-à-dire que pour certaines valeurs de paramètres, la réponse est quasiment constante, alors que pour d'autres, les fluctuations de la réponse sont beaucoup plus importantes, en général localement. Ainsi, une modélisation polynomiale est insuffisante pour bien caractériser le comportement de la réponse et il est indispensable d'envisager une modélisation adaptative pour les deux réponses considérées.

7.2.3 Approximation par modélisation adaptative

Nous nous plaçons dans les mêmes conditions pour la modélisation adaptative des réponses CumOil et CumWater en fonction des paramètres ZWO1, SKIN, KRGM et KRWM. Le plan initial que nous considérons est un hypercube latin de distance maximin à 25

simulations pour quatre facteurs Nous considérons une fonction de covariance gaussienne anisotrope.

A chaque itération, nous proposons d'ajouter :

- 1 maximum avant l'éventuelle application de la méthode « iso-simplexe »,
- 1 minimum avant l'éventuelle application de la méthode « iso-simplexe »,
- 3 points de dérivée partielle nulle,
- 3 points issus de l'optimisation par points pilotes des points maximisant la variance de krigeage.

A noter que 10 points pilotes au maximum sont conservés dans le dispositif expérimental.

De plus, dès lors que la qualité de l'approximation semble être satisfaisante, nous ajoutons cinq points de confirmation afin de juger de la prédictivité de l'approximation et de la nécessité d'arrêter ou de continuer le processus itératif.

- **Production d'huile cumulée à 2 192 jours**

Les résultats de l'approximation de la production d'huile cumulée à 2 192 jours par modélisation adaptative sont présentés sur la Figure 7.13 et la Figure 7.14. Ainsi, la Figure 7.13 présente l'évolution des graphes des erreurs de prédiction aux 1 296 points du plan complet 6^4 et de l'erreur relative pour différentes itérations du processus. Les points correspondant au plan d'expériences servant à la modélisation sont représentés par des croix bleues, les points de prédiction par des cercles violets. Pour indiquer la qualité du modèle, nous ajoutons sur ce graphe l'erreur relative obtenue aux 1 296 points de prédiction. Celle-ci est donnée par :

$$\varepsilon_{rel} = \frac{\sum_{i=1}^{1296} |Y_i - \hat{Y}(\mathbf{x}_i)|}{\sum_{i=1}^{1296} |Y_i|}$$

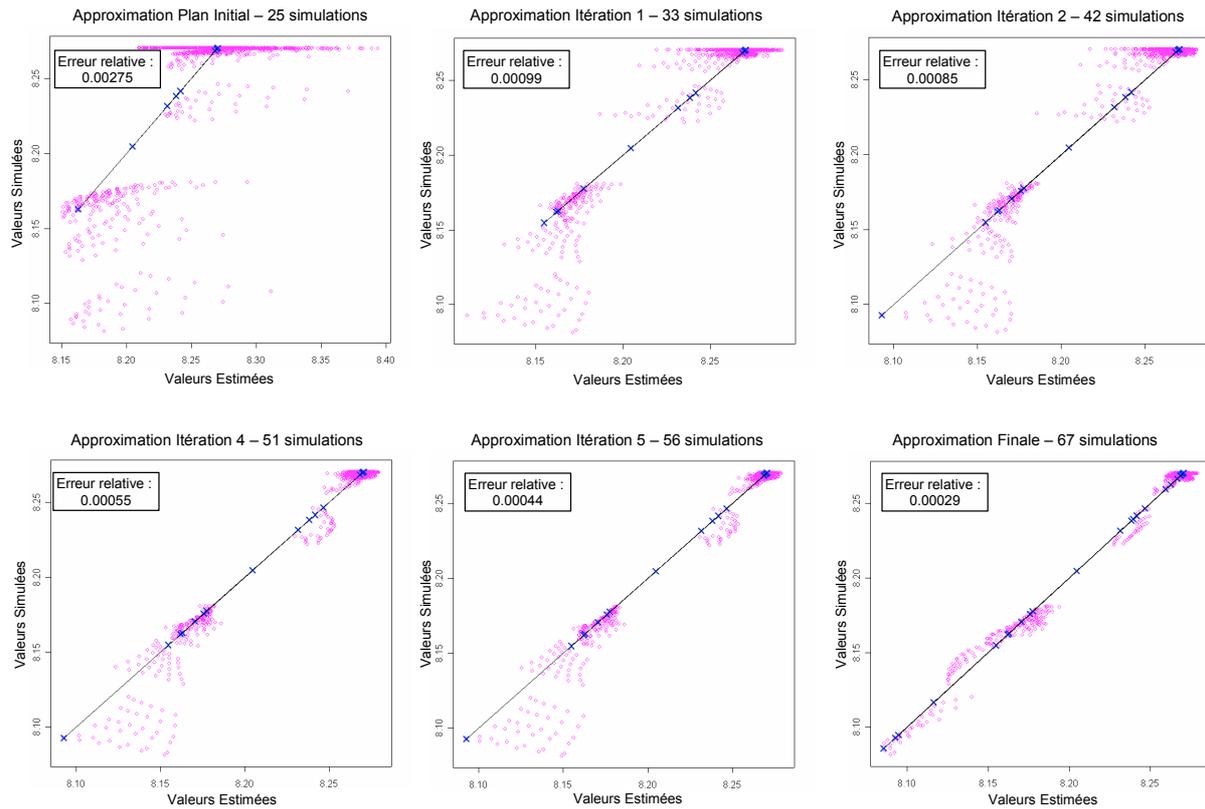


Figure 7.13 : Approximation du cumulé d'huile à 2 192 jours par modélisation adaptative au cours de itérations

Ainsi, un total de 67 simulations a été nécessaire pour la modélisation de CumOil (le cumulé d'huile après 6 ans de production). Pour le plan initial, nous constatons que l'approximation obtenue aboutit à des surestimations très importantes de la réponse. Dès la deuxième itération, l'ajout d'un point minimum a permis de corriger les estimations pour les faibles valeurs de la réponse, réduisant ainsi considérablement la surestimation de la production. Ensuite, à chaque itération, la qualité de l'approximation est améliorée, pour arriver à une approximation finale de très bonne qualité. En effet, l'erreur relative pour 1 296 points de prédiction est très faible puisque valant 0.00029, les points de prédiction sont peu éloignés de la première bissectrice.

La Figure 7.14 précise l'évolution des densités de probabilité de cette réponse CumOil au cours des itérations. Nous avons présenté les densité de probabilité pour l'approximation issue du plan d'expériences initial, en noir, pour l'approximation aux itérations 2, en bleu, et 4, en vert, et enfin pour l'approximation finale, en rouge. Les densités de probabilité sont obtenues par échantillonnage Monte-Carlo, pour un tirage uniforme de 10 000 points.

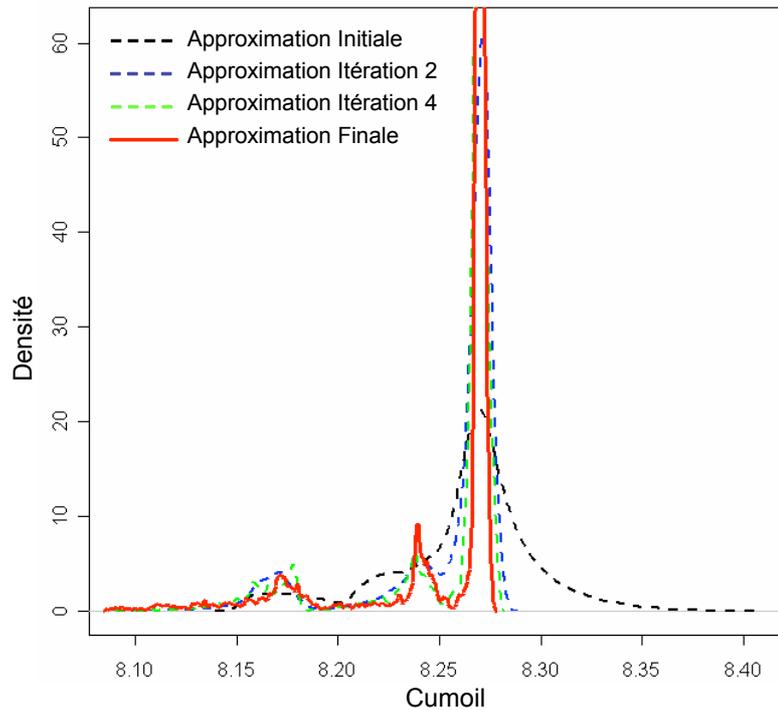


Figure 7.14 : Évolution des densités de probabilité de l'huile cumulée au cours des itérations

En ce qui concerne les densités de probabilité, nous constatons que, au fur et à mesure de l'ajout de points au dispositif, l'incertitude est réduite, les courbes se resserrant à chaque nouvelle itération. Ainsi, le risque lié à l'incertitude sur les paramètres est considérablement diminué du fait de l'enrichissement du dispositif expérimental. Plus la qualité de prédiction du modèle est importante, plus les analyses de risques sont précises, ce qui traduit une meilleure prise de décision dans un contexte incertain.

- **Production d'eau cumulée à 2 192 jours**

Comme dans le cas de la réponse CumOil, nous présentons les résultats sous deux formes :

- la Figure 7.15 décrit l'évolution des graphes de prédiction pour différentes itérations de la modélisation,
- la Figure 7.16 montre les densités de probabilité correspondant à certains ajustements.

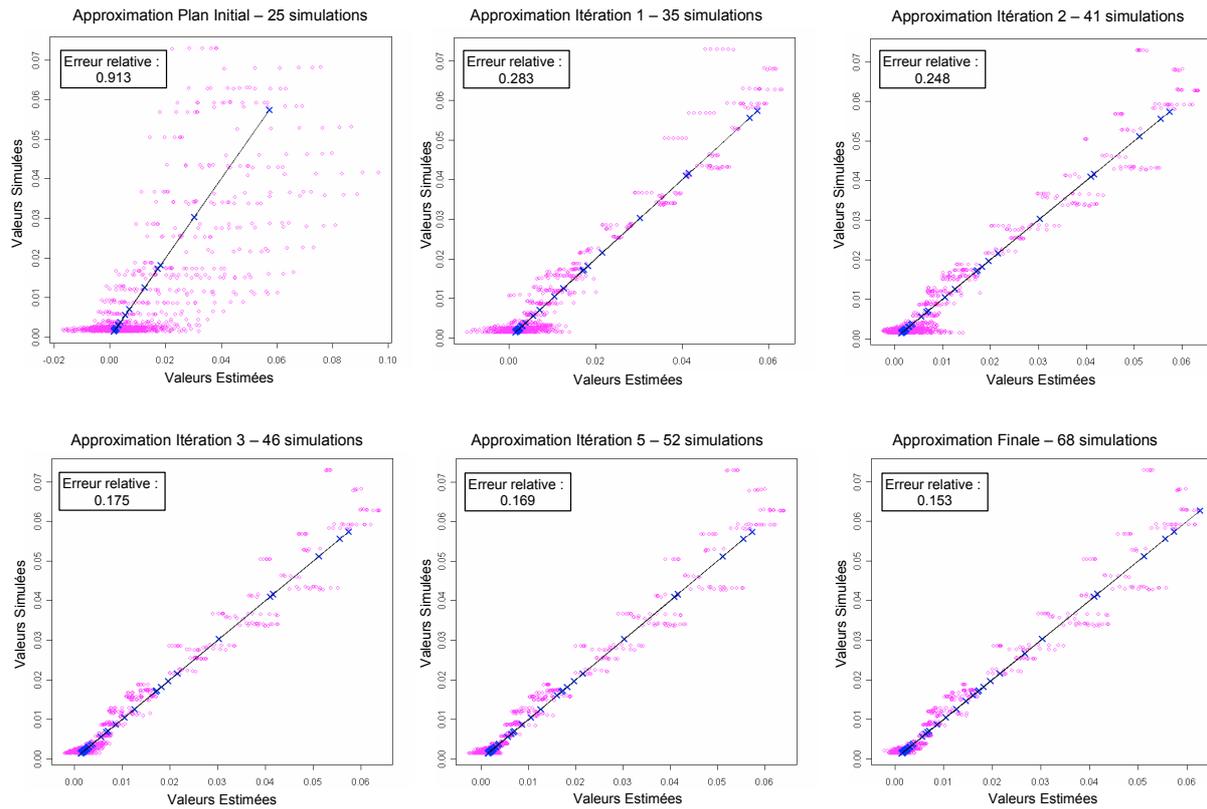


Figure 7.15 : Approximation du cumulé d'eau à 2 192 jours par modélisation adaptative au cours des itérations

La modélisation de la réponse CumWater a nécessité un total de 68 simulations. Les résultats de prédiction obtenus par la modélisation issue du plan d'expériences initial sont de piètre qualité. La simulation de points cruciaux dès la première itération permet une nette amélioration de la prédictivité. Au fur et à mesure des itérations, les faibles valeurs de la réponse sont bien mieux prédites et en fin de modélisation, la simulation d'un point maximum de l'approximation courante permet de mieux caractériser la réponse. L'approximation ainsi obtenue est de bonne qualité, avec une erreur relative aux 1 296 points de prédiction égale à 0.153.

Nous présentons sur la Figure 7.16 les densités de probabilité de la réponse CumWater pour les approximations initiales et finales, ainsi que celles des itérations 3 et 5.

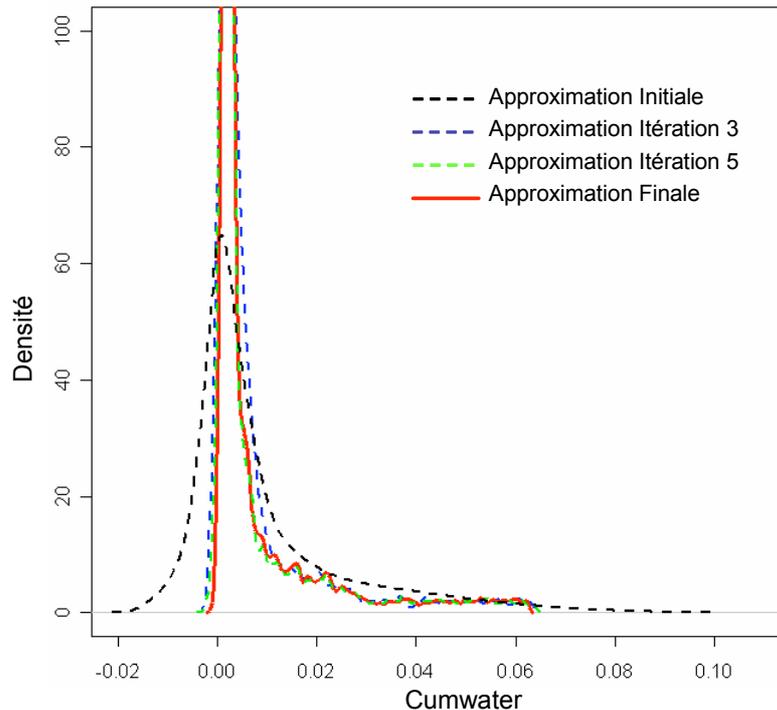


Figure 7.16 : Évolution des densités de probabilité de l'eau cumulée au cours des itérations

Encore une fois, plus la précision de l'approximation est importante, plus l'incertitude sur la réponse est réduite et donc plus fiables sont les analyses de risques.

Nous venons de présenter les résultats de la modélisation des réponses CumOil et CumWater par modélisation adaptative. Nous voulons à présent comparer les résultats obtenus avec ceux issus d'une modélisation par planification classique, soit 68 points.

7.2.4 Modélisation adaptative et planification classique

Dans cette partie, nous présentons, à titre comparatif, les résultats obtenus en modélisant les deux réponses CumOil et CumWater avec un modèle composite à partir de différents plans d'expériences classiques. Pour ce faire, nous considérons le plan factoriel complet 2^4 et un plan composite centré, puis deux hypercubes latins de distance maximin de même taille que le dispositif obtenu par modélisation adaptative.

- **Production d'huile cumulée à 2 192 jours**

Dans un premier temps, nous effectuons les simulations correspondant aux plans d'expériences classiques, soit un plan factoriel à 16 simulations et un plan composite centré à 25 simulations. Pour chaque plan, nous construisons deux types de modèle, un modèle polynomial de degré un ou deux suivant le plan utilisé et un modèle composite constitué d'une régression et d'un krigeage résiduel. Les résultats pour les modèles de krigeage sont présentés sur la Figure 7.17, pour respectivement, le plan factoriel (a) et le plan composite (b).

Nous ne procédons pas ici à un ajustement polynomial comme on le fait classiquement avec ce genre de dispositif pour des expériences réelles.

Comme pour le cas Punq-S3 présenté dans ce chapitre, nous comparons empiriquement la qualité des ajustements ainsi obtenus avec celui que nous donne la modélisation adaptative. Considérons tout d'abord la prédictivité.

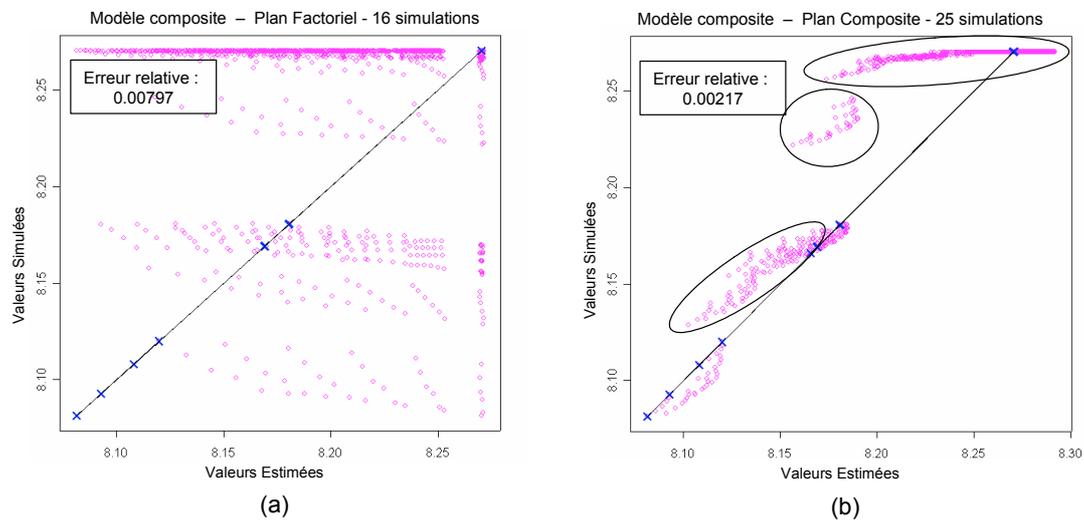


Figure 7.17 : Prédiction du cumulé d'huile par plans factoriel et composite

Comme nous pouvons le voir sur les graphes (a) et (b) de la Figure 7.17, les modèles de krigeage universel pour les plans classiques sont de mauvaise qualité. Il n'est en effet pas possible de prédire de manière précise le comportement de la production d'huile cumulée à 2 192 jours.

Si on considère l'approximation obtenue pour le plan factoriel à 16 simulations (a), il est évident que ce modèle n'est absolument pas prédictif, peu de points de prédiction se trouvent sur la première bissectrice, ils en sont généralement très éloignés.

De même, en considérant le plan composite centré à 25 simulations (b), on observe une nette sous-estimation de la production d'huile cumulée.

Nous présentons, sur la Figure 7.18, l'impact de ces défauts d'ajustement observés pour les plans d'expériences classiques sur les estimations de densité de probabilité de la réponse CumOil. Les densités de probabilité sont obtenues par échantillonnage Monte-Carlo d'une loi uniforme entre -1 et 1 pour les quatre paramètres. La taille de l'échantillon est de 10 000 points. Nous avons superposé, sur la Figure 7.18, les estimations de la réponse Cumoil issues du plan complet à 1 296 simulations, des plans factoriel et composite et du plan provenant de la modélisation adaptative.

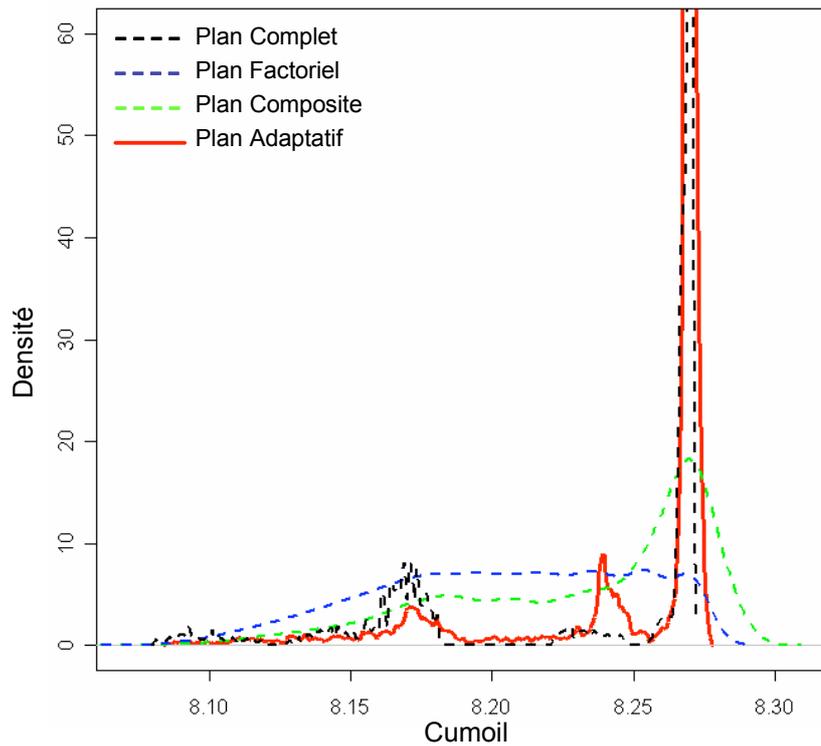


Figure 7.18 : Densité de probabilité de Cumoil pour les différents dispositifs expérimentaux

Ainsi, nous constatons que la modélisation adaptative permet une amélioration très nette des résultats par rapport aux plans classiques, les quantiles P10, P50 et P90 étant bien plus proches de ceux du plan complet 6^4 , considéré ici comme référence. Les analyses de risque issues de la modélisation adaptative seront donc beaucoup plus fiables que celles issues des plans classiques.

Voyons à présent les résultats issus d'une modélisation à partir d'hypercubes latins. Nous avons construit ici deux hypercubes latins de distance maximin comprenant 67 simulations, soit la même taille que le plan issu de la modélisation adaptative. Nous présentons les résultats sur la Figure 7.19.

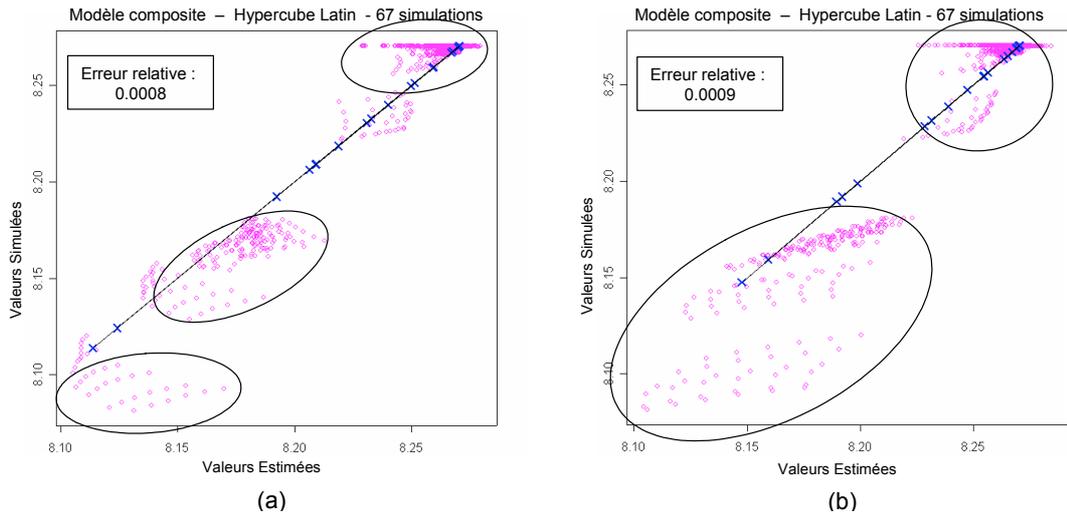


Figure 7.19 : Résultats de modélisation pour des hypercubes latins

Ainsi, nous constatons que les approximations associées à un échantillonnage de type hypercube latin ne permettent pas une bonne prédiction, notamment pour les faibles valeurs de la production d'huile cumulée à 2 192 jours. En effet, les valeurs minimales prédites par les modèles sont respectivement de 8.104 et 8.105 Mm^3 , alors que pour le plan complet le minimum est de 8.08 Mm^3 . Il est évident ici que les valeurs de la réponse CumOil inférieures à 8.20 Mm^3 sont très mal estimées, ce qui peut s'avérer très dommageable pour une prise de décision.

Après analyse des valeurs de paramètres, nous constatons qu'une forte valeur pour le SKIN, soit un colmatage important autour du puit, engendre une faible production d'huile cumulé. Il nous semble donc particulièrement important d'étudier le risque induit pour de grandes valeurs d'endommagement de la perméabilité aux abords des puits. Pour ce faire, nous construisons un échantillonnage Monte-Carlo pour lequel les valeurs de paramètres SOW1, KRGM et KRWM sont issues d'un tirage d'une loi uniforme entre -1 et 1 et nous fixons le paramètre SKIN à 1, sa valeur maximale. Comme précédemment, la taille de l'échantillon est de 10 000 points. Nous présentons sur la Figure 7.20, les densités de probabilité obtenues respectivement par une modélisation basée sur les deux hypercubes latins considérés, en bleu et vert, et par la modélisation adaptative, en rouge.

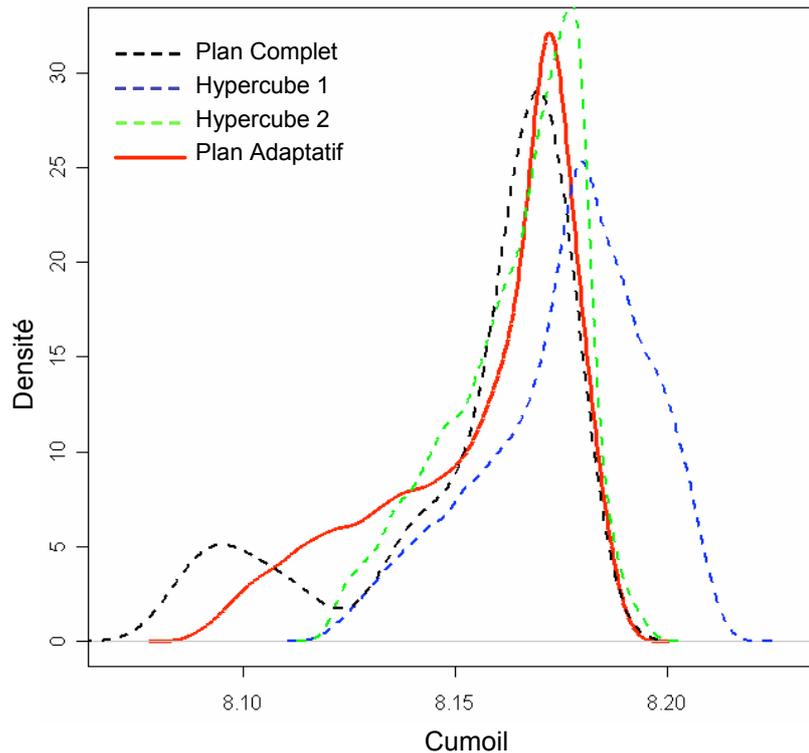


Figure 7.20 : Densité de probabilité de la réponse Cumoil pour les différents dispositifs expérimentaux

Nous observons que s'il y a un fort endommagement de la perméabilité aux abords des puits, les densités de probabilité obtenues par hypercubes latin de 67 points, donc sans recours à la modélisation adaptative, sont très éloignées de celle obtenue par le plan complet. L'utilisation de la modélisation adaptative donne des résultats beaucoup plus satisfaisants, l'estimation des quantiles associés à la modélisation adaptative est bien plus fiable dans la procédure adaptative.

En fait, l'utilisation d'hypercube latin ne permet pas un échantillonnage des faibles valeurs de production, et par conséquent ce type d'échantillonnage ne permet pas l'obtention d'un modèle prédictif et donc d'une analyse de risque précise.

De plus, la comparaison des densités obtenues pour les deux hypercubes latins de 67 points, soit les courbes bleue et verte, montre de grandes différences, aussi bien pour l'estimation de la valeur maximale de production qu'on peut escompter, que pour la production la plus probable. Ainsi, les quantiles P10, P50 et P90 ont des valeurs différentes, ce qui aboutit à des prises de décision différentes suivant l'hypercube considéré. Ainsi, il n'est pas possible d'aboutir de cette manière à une quantification des incertitudes fiable.

- **Production d'eau cumulée à 2192 jours**

Nous avons mené la même étude pour la réponse CumWater. Nous présentons les résultats d'une modélisation par régression polynomiale suivie d'un krigeage des résidus utilisant tout d'abord des plans d'expériences classiques, soit un plan factoriel, un plan composite centré et deux hypercubes latins. Ainsi, la Figure 7.21 représente les graphes de prédiction et de l'erreur relative aux 1 296 points de prédiction pour respectivement un plan factoriel (a) et un plan composite (b).

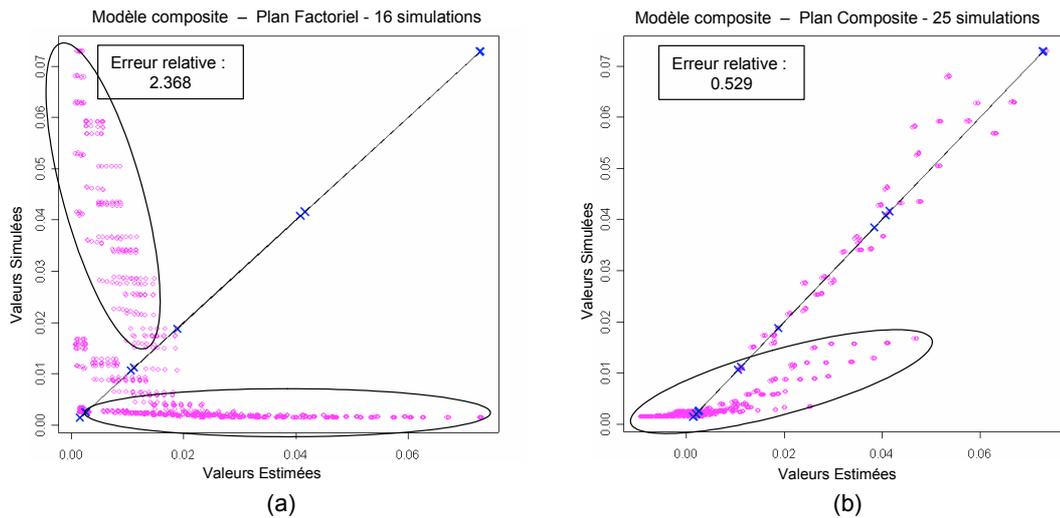


Figure 7.21 : Résultats de modélisation pour les plans factoriel et composites

Comme dans le cas de l'estimation de la production d'huile cumulée, nous constatons, sur la Figure 7.21, que les approximations obtenues à l'aide d'un plan factoriel ou d'un plan composite centré ne sont pas de bonne qualité prédictive, ils nous donnent des prévisions de production très peu fiables. Nous remarquons également sur le graphe (b) qu'un grand nombre de prédictions sont négatives, ce qui est impossible physiquement pour une production cumulée.

Passons maintenant à la densité de CumWater obtenue, comme précédemment, par échantillonnage Monte Carlo impliquant 10 000 points issus d'un tirage selon des lois uniformes sur chaque paramètre.

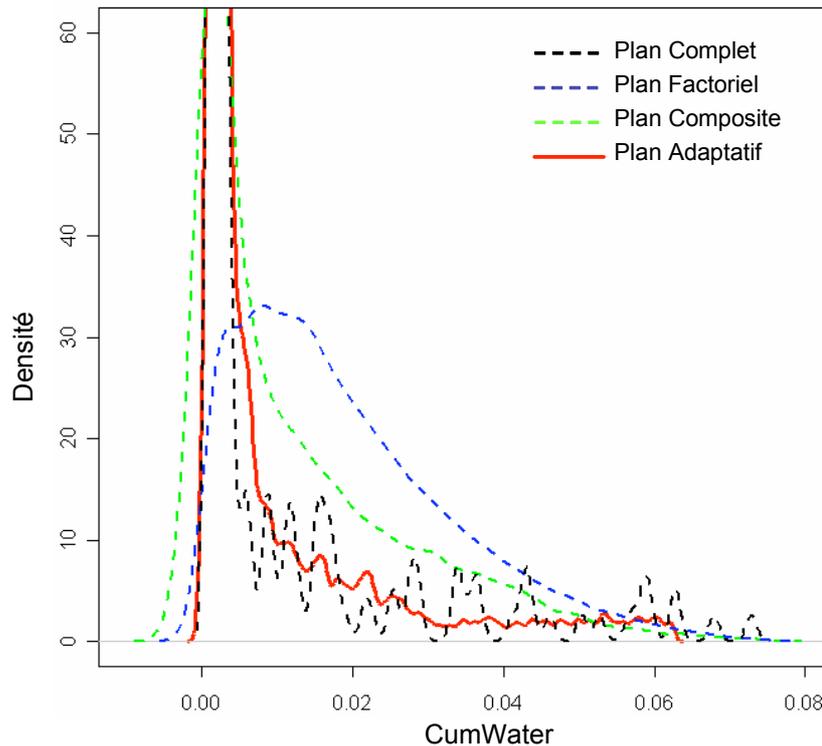


Figure 7.22 : Densité de probabilité pour les plans factoriel et composites et modélisation adaptative

D'après la Figure 7.22, nous constatons que l'approximation du simulateur d'écoulement par le modèle issu de la modélisation adaptative donne des résultats précis quant à la densité de probabilité du cumulé d'eau. En effet, la densité de probabilité et donc les quantiles associés sont très proches de ceux obtenus par le plan complet 6^4 , ce qui n'est pas le cas pour les plans factoriel et composite, pour lesquels l'incertitude reste très importante.

Ainsi, l'approximation par krigeage universel à partir de plans d'expériences classiques ne permet une analyse de risque efficace et fiable, alors que la modélisation adaptative fournit des résultats très satisfaisants.

Comme nous l'avons fait précédemment pour la réponse CumOil, voyons à présent les résultats de la modélisation de la réponse CumWater à partir de deux hypercubes latins comprenant 68 simulations. Les résultats sont présentés sur la Figure 7.23.

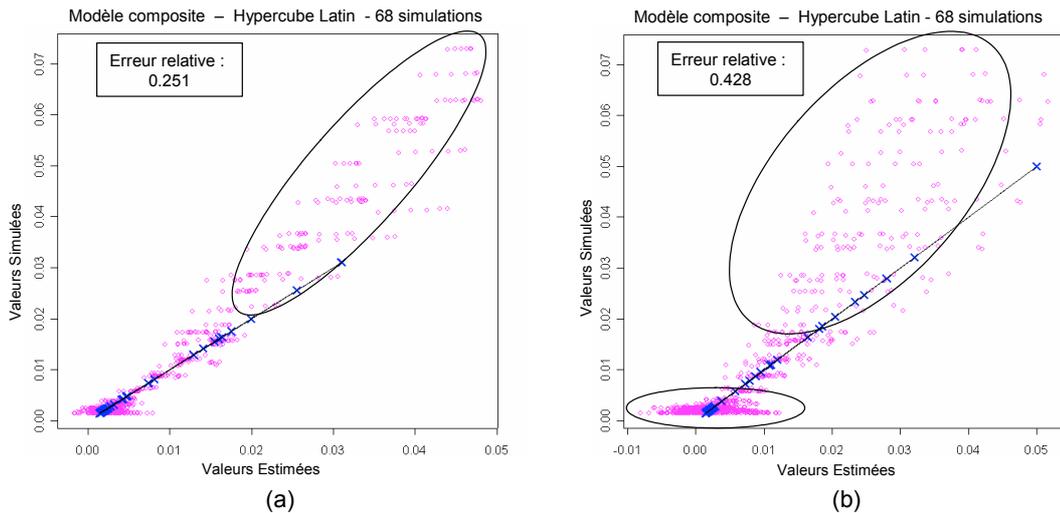


Figure 7.23 : Résultats de modélisation pour des hypercubes latins

S'agissant de la prédictivité, nous constatons la mauvaise qualité de l'approximation issue d'hypercubes latins de même taille que le nombre de simulations retenus pour la modélisation adaptative. Les points de coordonnées (valeur estimée, valeur simulée) pour les 1 269 simulations effectuées sont très éloignés de la première bissectrice. Ces deux modèles aboutissent à une sous-estimation considérable de la production d'eau, voire même à des productions négatives. Notons par ailleurs que la qualité de prédiction est elle même variable d'un hypercube à l'autre.

Les fortes valeurs du cumulé d'eau à 2 192 jours correspondent à une faible valeur du paramètre ZOW1 et à une valeur de KRWM différente de -1, c'est-à-dire à un contact huile/eau situé à -3 092 m et une forte perméabilité relative maximale pour l'eau. Étant donné que ces paramètres sont incertains et non-contrôlables, il est important de voir l'impact de la mauvaise estimation des fortes valeurs du cumulé d'eau sur les analyses de risques. Nous nous plaçons donc dans ce cas pour estimer les densités de probabilités pour respectivement le modèle issu de la modélisation adaptative, les deux hypercubes latins et enfin le plan complet 6^4 . L'échantillonnage considéré ici contient 10 000 points, la valeur du paramètre ZOW1 est constante et égale à -1, celle de SKIN et KRGM sont issues d'une loi uniforme entre -1 et 1, et enfin celle de KRWM provient d'un tirage uniforme entre -0.8 et 1. Les résultats sont présentés sur la Figure 7.24.

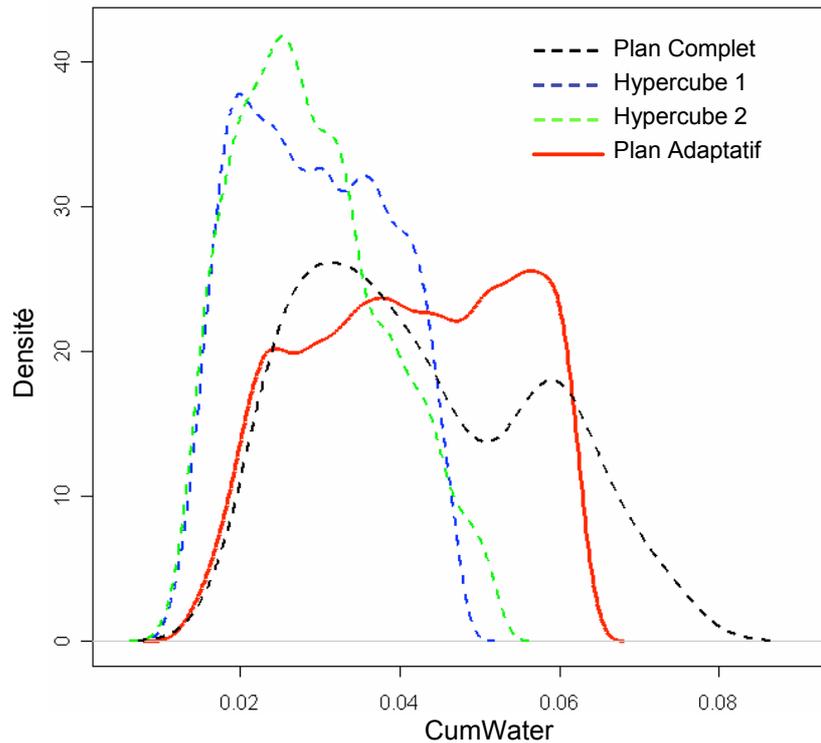


Figure 7.24 : Densité de probabilité pour les hypercubes latins et modélisation adaptative

Encore une fois, les résultats issus de la modélisation adaptative sont plus précis que ceux provenant d’hypercubes latins. En effet, la densité de probabilité obtenue par cette approche et donc les quantiles associés sont beaucoup plus proches de ceux obtenus avec le plan complet que les résultats obtenus grâce aux deux hypercubes latin de distance maximin.

7.2.5 Conclusion

Les résultats des études sur les réponses CumOil et CumWater ont permis de mettre en évidence le fait que la méthode de modélisation adaptative nous donne des approximations beaucoup plus précises que celles issues des techniques classiques d’échantillonnage et de krigeage.

Il est évident que les analyses de risque sont d’autant plus fiables que les approximations sont de bonne qualité prédictive. Par conséquent, la construction d’un modèle approché du simulateur d’écoulement par l’approche adaptative permet une meilleure quantification des incertitudes et donc une estimation des risques plus précise pour aboutir, *in fine*, à une meilleure prise de décision dans un contexte incertain.

7.3 Conclusions

Dans ce chapitre, nous avons présenté deux applications de l'utilisation de la modélisation adaptative pour la construction d'un modèle approché du simulateur d'écoulement en ingénierie pétrolière. La construction de ce modèle approché est en effet nécessaire à la réalisation de quantification d'incertitude et d'analyse de risque à moindre coût.

Ainsi, nous avons montré sur un cas synthétique Punq et un cas réel PBR que l'utilisation de l'approche adaptative permet une plus grande précision et une meilleure fiabilité dans les prises de décision que celles obtenues par krigeage universel (ajustement polynomial suivi d'un krigeage des résidus) à partir de plans d'expériences classiques ou d'échantillonnage quasi Monte-Carlo classiques actuellement utilisés en ingénierie de réservoir.

De plus, la modélisation par hypercubes latins ne permet pas une quantification des incertitudes fiable, du fait, entre autres, de la non-unicité de ce type de plans. Enfin, ne connaissant pas *a priori* la taille du dispositif à employer pour assurer une modélisation de bonne qualité, l'intérêt d'une approche adaptative est d'autant plus grand.

Chapitre 8

Bilan et Perspectives

Cette thèse est consacrée à la planification et l'analyse statistique d'expériences simulées lorsque la réponse du simulateur n'est pas régulière, c'est-à-dire approximable par une fonction polynomiale, sur un domaine expérimental donné. Cette situation intervient fréquemment en industrie pétrolière, plus précisément dans la quantification des incertitudes sur les réponses en production de réservoirs pétroliers. Il faut ici prendre en compte la dépendance complexe, souvent fortement non-régulière, de ces réponses : water-cut, cumulé d'huile, etc. par rapport aux paramètres physiques : perméabilité, porosité, etc. Le problème consiste à construire un modèle approché du modèle numérique d'écoulement qui soit le plus prédictif possible, tout en contrôlant le nombre de simulations numériques d'écoulement requises pour calibrer ce modèle approché.

La modélisation de réponses non-régulières par surfaces de réponses est un problème à ce jour d'une grande difficulté, le nombre de simulations d'écoulements envisageable est réduit du fait de leur coût. L'approche actuellement privilégiée consiste à construire une approximation par régression ou krigeage universel à partir d'un nombre réduit de simulations effectuées dans des conditions fixées *a priori* et organisé selon un plan d'expériences classique (plan factoriel, plan composite centré, etc.) ou un échantillonnage quasi Monte-Carlo usuel (hypercubes latins). Cette approche nécessite une connaissance *a priori* du type de dépendance entre la réponse et les paramètres et atteint ses limites lorsque que la réponse à modéliser est complexe.

L'approche de modélisation de réponses non-régulières que nous présentons dans cette thèse est novatrice, elle combine les avantages de l'échantillonnage quasi Monte-Carlo (hypercubes latins), la méthodologie statistique des plans d'expériences et les techniques géostatistiques d'interpolation. Le caractère déterministe de la réponse étudiée justifie en effet le choix d'un interpolateur pour la modélisation de réponses non-régulières. Ainsi, nous proposons dans ces travaux l'emploi couplé de techniques de régression et de krigeage des résidus pour approximer la réponse par interpolation d'un échantillon de réponses du simulateur. Cette méthode est équivalente à une méthode d'interpolation bayésienne.

Le problème de réduction du nombre de simulations justifie une approche par plan d'expériences, et plus précisément par plans évolutifs. Partant d'un plan initial, nous proposons d'ajouter itérativement de nouvelles simulations judicieusement choisies jusqu'à ce que la qualité de prédiction de l'approximation soit jugée satisfaisante. Le caractère évolutif de l'approche, permet de contrôler le nombre de simulations à chaque itération. L'ingénieur peut décider à tout moment de stopper le processus et sacrifier un peu la précision de l'approximation de la réponse. Cette flexibilité est un atout considérable par rapport aux méthodes classiques. En effet, pour des hypercubes latins ou des plans d'expériences usuels (factoriel ou composite centré), si la qualité de l'approximation n'est pas jugée satisfaisante, l'ajout de nouvelles simulations entraîne la perte des propriétés de ces plans.

Contrairement aux méthodes classiques d'augmentation de dispositif expérimental qui sont fondées sur des critères géométriques de répartition des simulations dans l'espace, nous proposons d'échantillonner *a priori* les zones peu ou pas explorées, en tenant compte des variations de la réponse et pas uniquement de celle des paramètres. Ainsi, la simulation des points extrêmes et des points d'inflexion de l'approximation courante permet une meilleure caractérisation des non-régularités de la réponse. Quant aux zones non explorées par le dispositif expérimental courant, elles sont traitées par l'optimisation et la simulation de points pilotes, ceux-ci ayant pour but d'augmenter la qualité de prédictivité de l'approximation. La technique des points pilotes, introduite ici pour la construction de modèles approchés est originale et joue un rôle majeur dans ces travaux. Elle donne de très bons résultats, aussi bien pour la caractérisation de points pour lesquels une simulation est nécessaire parce qu'informatrice, que pour l'augmentation de la prédictivité de l'approximation par l'ajout d'information fictive, c'est-à-dire sans réaliser de simulations supplémentaires. La validation de cette approche de modélisation adaptative a été réalisée sur plusieurs fonctions types à deux, trois, quatre et cinq paramètres. Dans tout les cas, la qualité prédictive des modèles obtenus par échantillonnage adaptatif est de bien meilleure qualité comparée aux approximations obtenues par krigeage universel à partir de plans d'expériences classiques ou d'hypercubes latins.

Le choix des critères d'ajout de points assure un échantillonnage spécifique et informatif et permet donc de réduire le nombre de simulations nécessaire à l'obtention d'une approximation de même qualité que celle issue de la méthode des plans classiques. Ce gain de simulations est d'autant plus important que le nombre de paramètres augmente. En effet, d'après les différents exemples examinés, le nombre de simulations requises semble augmenter linéairement en fonction du nombre de paramètres pour l'échantillonnage adaptatif, alors qu'il augmente apparemment de manière quadratique, voire exponentielle pour les plans classiques. Ceci est crucial dans notre approche, en particulier pour les applications rencontrées en ingénierie de réservoir. En effet dans ce domaine, les lois simulées

dépendent en général, de manière non-régulière d'un grand nombre de paramètres de simulation.

Nous avons ensuite monté la robustesse de la modélisation adaptative en traitant des exemples rencontrés en ingénierie de réservoir. L'utilisation de la modélisation adaptative pour la modélisation de la production de réservoir à hydrocarbures dans un contexte incertain a permis une meilleure quantification des incertitudes et des risques associés à l'incertitude sur certains paramètres pour aboutir à une prise de décision plus rationnelle et moins risquée, que ce soit en terme opérationnel qu'en terme de coûts.

Une amélioration de la méthode adaptative introduite ici pourrait être obtenue par la recherche systématique des sources de non-régularité liées aux paramètres. En effet, une mise en œuvre efficace et performante de la modélisation adaptative nécessite d'évaluer la régularité ou la non-régularité de la réponse vis-à-vis de chaque paramètre. Dans le but de réduire le nombre de simulations nécessaires à la construction d'un modèle approché du simulateur d'écoulement, il est préférable d'échantillonner plus intensément dans le sous-domaine défini par les paramètres induisant les plus fortes non-régularités.

D'autres améliorations de la procédure adaptative pourraient résulter d'une estimation spécifique de la fonction de covariance en des zones du domaine expérimental. Ainsi, l'hypothèse de forte stationnarité serait affaiblie, par le fait de traiter certaines zones de l'espace expérimental de manière locale, en particulier pour les endroits où il y a beaucoup de simulations et où il est donc pressenti de fortes irrégularités. On pourrait avoir recours à une estimation de la structure de covariance par voisinages, par une modification locale de la longueur de corrélation lors du krigeage.

Par ailleurs, il serait intéressant de tirer avantage d'autres opérateurs de krigeage proposés dans la bibliographie. Par exemple, l'utilisation d'un krigeage avec voisinages glissants dans les zones fortement échantillonnées du domaine expérimental, devrait permettre une modélisation moyenne locale. On pourrait aussi privilégier une régression globale et un krigeage résiduel lorsque le nombre d'observations est faible.

Enfin, l'échantillonnage obtenu par modélisation adaptative est spécifique à la réponse étudiée, et par conséquent la prévision de plusieurs réponses demande l'optimisation d'autant de dispositifs expérimentaux. Il serait donc intéressant de considérer des réponses multiples lors de la construction des plans d'expériences évolutifs.

Il nous apparaît important de préciser que cette thèse s'intègre dans le projet « Gestion des Incertitudes en Exploration Production » et a donné lieu au dépôt d'un brevet. Ce projet de recherche a pour but de fournir aux ingénieurs de réservoir des méthodes basées sur un formalisme statistique rigoureux, pour la prédiction de réponses en production dans un contexte incertain. L'efficacité des méthodes ainsi développées est éprouvée dans le cadre d'un projet collaboratif, le consortium Cougar II, soutenu par six compagnies pétrolières :

Saudi Aramco, Gaz de France, Repsol-YPF, BHP-Billiton, Conoco-Phillips et Petrobras. Ainsi, la méthode adaptative introduite dans cette thèse doit être intégrée sous forme d'un module prototype au logiciel de quantification des incertitudes en ingénierie de réservoir : CougarTM.

Annexe A

Statistique Inférentielle

Nous proposons dans ce premier annexe quelques rappels relatifs aux méthodes de statistique inférentielle. Nous commençons par définir les principaux tests statistiques réalisés sur un modèle de régression. Ils permettent de valider ce modèle par un test de significativité, de tester sa qualité d'estimation et de prédiction et enfin de déterminer les termes influents lors d'un test de Student. Dans une deuxième partie, nous décrivons la manière d'obtenir un estimateur par moindres carrés pondérés (MCP), puis en troisième partie une méthode statistique permettant de définir le meilleur prédicteur linéaire sans biais (BLUP). Enfin, la quatrième partie est consacrée à une description de plusieurs plans d'expériences, particulièrement ceux obtenus à partir d'un critère statistique ou géométrique.

A.1 Analyse du modèle de régression

Dans le deuxième chapitre, nous avons précisé qu'une étude de l'impact des incertitudes se traduisait par la construction d'un modèle approché et l'utilisation de tests statistiques à partir de ce modèle. Nous considérons ici que le modèle est obtenu par régression, c'est-à-dire que nous avons :

$$Y = X\beta + \varepsilon \quad (\text{A.1})$$

avec :

- X la matrice $n \times p$ du modèle. Elle dépend à la fois du modèle de régression et du dispositif expérimental,
- Y le vecteur $n \times 1$ des observations de la réponse aux n points du dispositif expérimental,
- ε le vecteur $n \times 1$ des résidus aux n points du dispositif expérimental,
- β le vecteur $p \times 1$ des coefficients (paramètres) du modèle,
- n le nombre d'expérimentations et p le nombre de termes du modèle (en incluant la constante).

Nous allons à présent présenter plusieurs tests statistiques généralement utilisés pour quantifier l'incertitude associée aux différents paramètres. Ces méthodes sont basées sur l'analyse des résidus $\varepsilon = Y - \hat{Y}$ obtenus suite à l'ajustement du modèle aux simulations. Une description détaillée de ces différents tests sont présentés dans [8] et [11].

A.1.1 Test de significativité de la régression

L'objectif de ce premier test est de déterminer s'il y a bien une relation linéaire entre la réponse Y et une partie des variables x_1, \dots, x_n . Les deux hypothèses à comparer pour ce test sont les suivantes :

$$H_0 : \beta_1 = \beta_2 = \dots \beta_n = 0,$$

$$H_1 : \beta_j \neq 0 \text{ pour au moins un } j.$$

Si l'hypothèse H_0 est rejetée, alors au moins une des variables explique significativement la variabilité observée sur la réponse Y et donc le modèle est valide.

La mise en place de ce test implique une partition de la somme du carré des écarts des observations à la moyenne, par la somme du carré des écarts des estimations du modèle à la moyenne et la somme du carré des erreurs. Ainsi, nous avons :

$$S_{yy} = SS_R + SS_E$$

avec :

- S_{yy} la somme du carré des écarts des observations à la moyenne : $SS_y = \sum_{i=1}^n (Y_i - \bar{Y})^2$

qui est supposée représenter la variabilité totale de la réponse,

- SS_R la somme du carré des écarts des estimations du modèle à la moyenne :

$$SS_R = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 \text{ qui représente la variabilité expliquée par le modèle,}$$

- SS_E la somme du carré des écarts qui est supposée être la variabilité non expliquée par le modèle de régression,

- \bar{Y} l'estimation de la moyenne de la réponse donnée par : $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$.

Sous l'hypothèse H_0 , la statistique

$$F_0 = \frac{SS_R / (p-1)}{SS_E / (p-n)} = \frac{MS_R}{MS_E} \tag{A.2}$$

est une variable de Fischer avec $p-1$ et $n-p$ degrés de liberté.

La probabilité α de choisir H_1 alors que H_0 est vraie est fixée par l'utilisateur (en général 5%), elle est comparée à la probabilité correspondant à F_0 obtenue par les tables de Fischer. Si cette probabilité est inférieure au seuil α , alors H_0 est rejetée et le modèle de régression est significatif.

Cette procédure est souvent appelée analyse de variance. L'analyse de la variance a pour but de calculer des statistiques permettant une meilleure compréhension de la qualité du modèle.

A.1.2 Coefficients de qualité du modèle

- **Qualité d'ajustement du modèle**

Une mesure de la qualité de l'ajustement d'un modèle de régression est donnée par le coefficient de détermination :

$$R^2 = \frac{SS_R}{S_{yy}} = 1 - \frac{SS_E}{S_{yy}} \quad (\text{A.3})$$

R^2 est une mesure du pourcentage de variabilité observée sur une réponse Y et expliquée par les variables x_1, \dots, x_n . Malheureusement, R^2 dépend fortement du modèle, spécialement lorsque le nombre de termes du modèle varie. Par exemple, R^2 croît automatiquement lorsque l'on ajoute des termes au modèle, même si ces termes ne sont pas significatifs. Pour éviter ce problème, nous pouvons construire un coefficient de détermination ajusté donné par :

$$R_{adj}^2 = 1 - \frac{SS_E / (n - p)}{S_{yy} / (n - 1)} \quad (\text{A.4})$$

Ces deux coefficients donnent un estimateur de la qualité d'ajustement du modèle, c'est-à-dire la manière dont le modèle de régression approxime la réponse aux points expérimentaux. Par exemple, une valeur proche de un indique un ajustement aux données de bonne qualité, alors qu'une faible valeur montre que le modèle n'est pas adéquat puisqu'il ne permet même pas la restitution des valeurs simulées.

Mais, le modèle de régression ne doit pas seulement bien ajuster les données, il doit aussi prédire la réponse Y de manière précise sur l'ensemble du domaine. Un modèle de bonne qualité en terme d'ajustement n'est pas forcément un bon prédicteur. Par exemple, les modèles saturés (modèles qui ont autant de termes qu'il y a d'expérimentations) ajustent parfaitement les observations de la réponse ($R^2 = 1$) mais sont en général de très mauvais prédicteurs. De ce fait, une nouvelle statistique doit être construite pour estimer la qualité de prédiction du modèle.

- **Qualité de prédiction du modèle**

Il est intéressant de voir comment le modèle peut prédire la $i^{\text{ème}}$ donnée si celle-ci est enlevée lors de l'estimation du modèle. La différence entre Y_i et la valeur estimée $\hat{Y}_{(i)}$ est appelée résidus de prédiction au point x_i [8] et [11].

L'estimation du pouvoir de prédiction du modèle est basée sur le calcul de la somme du carré des erreurs de prédiction (notée *PRESS*). Le principe consiste à sélectionner une observation i et à la supprimer puis ajuster un modèle de régression par rapport aux $n-1$ données restantes. Ensuite, ce modèle va permettre de prédire la réponse $\hat{Y}_{(i)}$ et de calculer l'erreur de prédiction $e_{(i)}$ pour l'observation i . Cette procédure est répétée pour chaque observation et donne la statistique suivante :

$$PRESS = \sum_{i=1}^n e_{(i)}^2 = \sum_{i=1}^n [Y_i - \hat{Y}_{(i)}]^2 \quad (\text{A.5})$$

où $\hat{Y}_{(i)}$ est la valeur estimée par le modèle pour la $i^{\text{ème}}$ observation.

Le coefficient *PRESS* est alors utilisé pour calculer une approximation de R^2 pour la prédiction, souvent noté Q_2 :

$$Q_2 = 1 - \frac{PRESS}{S_{yy}} \quad (\text{A.6})$$

Les modèles ayant un coefficient Q_2 proche de un, sont meilleurs que ceux ayant une faible valeur de Q_2 . Il est tentant de penser que cette mesure est plus valide que le coefficient de détermination parce que ce coefficient est déterminé à partir de valeurs non présentes lors de l'ajustement du modèle, et par conséquent donne une information sur le pouvoir prédictif du modèle.

A.1.3 Amélioration du modèle : Test de Student

Il est à présent intéressant de tester la significativité de chaque terme du modèle. Les hypothèses pour tester la significativité des coefficients β_j sont :

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

Si est H_0 rejetée, alors le terme en x_j est significatif et peut être maintenu dans le modèle. Sinon, ce terme est supprimé du modèle.

Sous l'hypothèse H_0 , la statistique $t_0 = \frac{\hat{\beta}_j}{\sqrt{\sigma^2 C_{jj}}}$ est une variable de Student avec $n-p-1$

degrés de liberté. Notons que le terme C_{jj} est le $j^{\text{ème}}$ élément de la diagonale $({}^tXX)^{-1}$. Il suffit de chercher la probabilité correspondant à t_0 dans les tables de la loi de Student. Si cette probabilité est inférieure à la valeur $\alpha/2$ alors H_0 est rejetée et le terme x_j n'est pas supprimé du modèle.

Ce test doit être utilisé avec précaution car le résultat dépend de la présence des autres termes dans le modèle. En effet, supprimer un terme du modèle peut changer les résultats pour les autres termes. Ainsi, ce test doit être réalisé séquentiellement, en enlevant un terme après l'autre et en suivant l'évolution des qualités de l'approximation.

A.2 Ajustement par moindres carrés pondérés

Plus généralement que pour l'ajustement par moindres carrés ordinaires, nous supposons que les erreurs soient corrélées [37], c'est-à-dire que :

$$Y = X\beta + \varepsilon, \quad (\text{A.7})$$

avec :

1. X la matrice de régression du modèle,
2. β le vecteur des paramètres du modèle,
3. ε le vecteur des erreurs tels que : $E(\varepsilon) = 0$ et $Cov(\varepsilon) = \sigma^2 R$ où R est une matrice définie positive.

Soit Q l'inverse généralisé de R , c'est à dire que $Q^{-1}R {}^tQ^{-1} = I$. Au lieu d'analyser le modèle (A.7), nous considérons le modèle équivalent :

$$Q^{-1}Y = Q^{-1}X\beta + Q^{-1}\varepsilon \quad (\text{A.8})$$

Pour ce modèle (A.8), nous avons :

$$E(Q^{-1}\varepsilon) = 0 \text{ et } Cov(Q^{-1}\varepsilon) = \sigma^2 Q^{-1}R {}^tQ^{-1} = I.$$

Ce modèle transformé satisfait donc les hypothèses du modèle (3.3) défini pour une estimation par moindres carrés ordinaires que l'on a présenté au chapitre 3.

Par le même raisonnement que celui appliqué dans le cadre d'une régression par moindres carrés ordinaires, $\hat{\beta}$ est un estimateur de β par moindres carrés pondérés si et seulement si il minimise l'équation :

$$L = {}^t(Q^{-1}Y - Q^{-1}X\beta)(Q^{-1}Y - Q^{-1}X\beta) = {}^t(Y - X\beta) {}^tQ^{-1}Q^{-1}(Y - X\beta)$$

Or, nous savons que $\mathbf{Q}^{-1}\mathbf{R}'\mathbf{Q}^{-1} = \mathbf{I}$ donc $\mathbf{R}^{-1} = {}^t\mathbf{Q}^{-1}\mathbf{Q}^{-1}$ et par conséquent :

$$L = {}^t(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \quad (\text{A.9})$$

En fait, l'estimateur de $\boldsymbol{\beta}$ qui minimise cette fonction est appelé estimateur des moindres carrés pondérés parce qu'au lieu de minimiser le carré de la distance entre \mathbf{Y} et $\mathbf{X}\boldsymbol{\beta}$, il minimise le carré de la distance pondérée. Les poids sont déterminés par l'inverse de la matrice de corrélation entre les erreurs : \mathbf{R}^{-1} .

Comme pour l'ajustement par moindres carrés pondérés, nous cherchons $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ tel que l'équation (A.9) soit minimale, ce qui revient à chercher le projeté \mathbf{R}^{-1} orthogonal de \mathbf{Y} sur $\text{Im}(\mathbf{X})$. Ainsi, par définition de ce projecteur, nous avons :

$$\hat{\mathbf{Y}} = \mathbf{X}({}^t\mathbf{X}\mathbf{R}^{-1}\mathbf{X})^{-1}{}^t\mathbf{X}\mathbf{R}^{-1}\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}} \quad (\text{A.10})$$

Les différentes propriétés de cet estimateur de $\boldsymbol{\beta}$ par moindres carrés pondérés sont présentées dans le théorème de Gauss-Markov. On trouve, par exemple, une démonstration de ce théorème dans [8].

Théorème de Gauss Markov:

- $\hat{\boldsymbol{\beta}}$ est l'estimateur des moindres carrés pondérés de $\boldsymbol{\beta}$ si et seulement si :

$$\hat{\boldsymbol{\beta}} = ({}^t\mathbf{X}\mathbf{R}^{-1}\mathbf{X})^{-1}{}^t\mathbf{X}\mathbf{R}^{-1}\mathbf{Y} \quad (\text{A.11})$$

- Si $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{R})$, alors, l'estimateur obtenu par moindres carrés est un estimateur sans biais qui est une combinaison linéaire des Y_i . Nous parlons souvent, pour qualifier les estimateur type moindres carrés de meilleur estimateur linéaire sans biais (BLUP) où meilleur implique de variance minimale. C'est aussi l'estimateur du maximum de vraisemblance de $\boldsymbol{\beta}$.

Nous présentons à présent une méthode statistique pour la construction d'un interpolateur de la réponse Y , qui soit le meilleur estimateur linéaire sans biais (BLUP).

A.3 Estimateur BLUP

Nous considérons initialement le problème général de prédiction d'une variable aléatoire arbitraire Y_0 basée sur les données $\mathbf{Y} = {}^t(Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))$. Comme les variables Y_0 et \mathbf{Y} sont des quantités dépendantes, il semble naturel que \mathbf{Y} contienne des informations sur Y_0 et donc que les prédicteurs de Y_0 dépendent de la distribution jointe de Y_0 et \mathbf{Y} [57].

Si on considère que le prédicteur recherché est donné par la formule :

$$Y(\mathbf{x}) = \sum_{j=1}^p f_j(\mathbf{x})\beta_j + Z(\mathbf{x}) = {}^t\mathbf{X}(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}) \quad (\text{A.12})$$

alors, la distribution jointe de $Y_0 = Y(\mathbf{x}_0)$ et $\mathbf{Y} = {}^t(Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))$ est la loi normale multivariée :

$$\begin{pmatrix} Y_0 \\ \mathbf{Y} \end{pmatrix} \sim N_{1+n} \left[\begin{pmatrix} {}^t \mathbf{X}_0 \\ \mathbf{X} \end{pmatrix} \boldsymbol{\beta}, \sigma^2 \begin{pmatrix} 1 & \mathbf{r}_0 \\ \mathbf{r}_0 & \mathbf{R} \end{pmatrix} \right],$$

où $\mathbf{X}_0 = f(\mathbf{x}_0)$ est le vecteur $p \times 1$ des fonctions de régression pour $Y(\mathbf{x}_0)$, \mathbf{X} est la matrice $n \times p$ de régression dont l'élément (i, j) est $f_j(\mathbf{x}_i)$ pour $1 \leq i \leq n$ et $1 \leq j \leq p$, $\boldsymbol{\beta}$ est le vecteur $n \times p$ des coefficients inconnus de la régression et \mathbf{r}_0 est le vecteur $n \times 1$ $\mathbf{r}_0 = {}^t(R(\mathbf{x}_0 - \mathbf{x}_1), \dots, R(\mathbf{x}_0 - \mathbf{x}_n))$ et \mathbf{R} est la matrice $n \times n$ des corrélations entre les observations $\mathbf{Y} : \mathbf{R} = (R(\mathbf{x}_i - \mathbf{x}_j))$.

Supposons que la matrice \mathbf{X} soit de rang maximum p et que \mathbf{R} est définie positive, alors on peut montrer que : $\hat{Y}_0 = E\{Y_0 | \mathbf{Y}\}$ et donc :

$$\hat{Y}_0 = {}^t \mathbf{X}_0 \boldsymbol{\beta} + {}^t \mathbf{r}_0 \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}) \quad (\text{A.13})$$

est le meilleur prédicteur MPSE (Mean Squared Predictor Error) de Y_0 [57].

Bien sûr, ce prédicteur dépend de l'estimateur $\hat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$, celui-ci est obtenu par la méthode des moindres carrés ordinaires ou pondérés, il est donné par (A.11) :

$$\hat{\boldsymbol{\beta}} = ({}^t \mathbf{X} \mathbf{R}^{-1} \mathbf{X})^{-1} {}^t \mathbf{X} \mathbf{R}^{-1} \mathbf{Y}$$

De plus, la construction de ce prédicteur \hat{Y} impose un critère classique de la prédiction, il est linéaire et sans biais, d'où sa popularité [51].

Le terme correcteur de la moyenne est, d'après (A.13), ${}^t \mathbf{r}_0 \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})$, qui est une combinaison linéaire des résidus $\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}$, dont les coefficients de prédiction aux points spécifiques sont les suivants :

$${}^t \mathbf{r}_0 \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}) = \sum_{i=1}^n c_i(\mathbf{x}_0) (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})_i$$

où le poids $c_i(\mathbf{x}_0)$ est le $i^{\text{ème}}$ élément de ${}^t \mathbf{r}_0 \mathbf{R}^{-1}$ et $(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})_i$ est le $i^{\text{ème}}$ résidu.

Remarque : L'ajout de ce terme de correction force le prédicteur $\hat{Y}(\mathbf{x}_0)$ à interpoler les données. Ainsi, supposons que $\mathbf{x}_0 = \mathbf{x}_i$ pour un i fixé tel que $1 \leq i \leq n$. Alors, $\mathbf{X}_0 = {}^t f(\mathbf{x}_i)$ et ${}^t \mathbf{r}_0 = (R(\mathbf{x}_i - \mathbf{x}_1), \dots, R(\mathbf{x}_i - \mathbf{x}_n))$ qui est la $i^{\text{ème}}$ colonne de la matrice \mathbf{R} . Ainsi, ${}^t \mathbf{r}_0 \mathbf{R}^{-1} = (0, \dots, 0, 1, 0, \dots, 0)^T = \mathbf{e}_i$, qui est le $i^{\text{ème}}$ vecteur unité parce que son produit est la $i^{\text{ème}}$ colonne de $\mathbf{R}^{-1} \mathbf{R} = \mathbf{I}_n$, matrice identité. Ainsi, ${}^t \mathbf{r}_0 \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}) = {}^t \mathbf{e}_i (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}) = Y_i - {}^t f(\mathbf{x}_i) \hat{\boldsymbol{\beta}}$ et donc $\hat{Y}(\mathbf{x}_i) = {}^t f(\mathbf{x}_i) \hat{\boldsymbol{\beta}} + (Y_i - {}^t f(\mathbf{x}_i) \hat{\boldsymbol{\beta}}) = Y_i$ et nous observons bien les données.

La stratégie employée pour prédire $Y(\mathbf{x}_0)$ est donnée par la formule (A.13). Or, nous ne connaissons pas la valeur et la nature de la fonction de corrélation du processus résiduel Z . Par conséquent, l'utilisation de la formule précédente passe par une estimation de la structure de covariance. Ainsi, nous avons :

$$\hat{Y}(\mathbf{x}_0) = \hat{Y}_0 \equiv {}^t \mathbf{X}_0 \hat{\boldsymbol{\beta}} + {}^t \hat{\mathbf{r}}_0 \hat{\mathbf{R}}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}) \quad (\text{A.14})$$

avec, d'une part $\hat{\boldsymbol{\beta}} = ({}^t \mathbf{X} \hat{\mathbf{R}}^{-1} \mathbf{X})^{-1} {}^t \mathbf{X} \hat{\mathbf{R}}^{-1} \mathbf{Y}$, l'estimateur de Gauss-Markov de $\boldsymbol{\beta}$ et d'autre part, $\hat{\mathbf{R}}$ et $\hat{\mathbf{r}}_0$ déterminés par un estimateur de la fonction de corrélation $R(\cdot)$. De tels prédicteurs sont appelés EBLUP (Empirical Best Linear Unbiased Predictors) de $Y(\mathbf{x}_0)$ malgré le fait qu'ils ne sont généralement pas linéaires avec la réponse \mathbf{Y} [51]. Différents EBLUP correspondent à différents estimateurs de $R(\cdot)$.

A.4 Plans d'expériences

Cette partie traite des différents dispositifs expérimentaux possibles pour modéliser la tendance de la réponse à moindre coût. Nous décrivons, sans prétendre être exhaustif, plusieurs types de plans d'expériences, peu coûteux et particulièrement adaptés à la modélisation d'une réponse potentiellement non-régulière aux paramètres.

La théorie des plans d'expériences a largement été étudiée, principalement dans le cas d'expérimentations réelles (i.e. non reproductible). Les spécificités des expériences simulées ont toutefois été étudiées lors des dernières années [38], [53], [54] et [65]. Comme il s'agit ici d'expérimentations simulées déterministes (absence de bruit), il n'y a pas de gain d'information lorsque deux simulations sont répétées. Par conséquent, les méthodes classiques consistant à répéter plusieurs fois les mêmes essais pour réduire la variabilité ou la stabiliser ne sont pas pertinentes. De plus, les plans classiques ont des propriétés de symétrie qui sont intéressantes dans le cas d'expériences réelles, mais qui n'apportent rien pour des expériences déterministes [53]. Ainsi, pour les expériences simulées, il est judicieux de s'intéresser à toutes les parties de la région de manière équiprobable. Nous axons notre présentation sur deux grandes classes de plans :

- les plans optimisant un critère statistique (tel que la D-optimalité, la IMSE, l'entropie)
- les plans optimisant un critère géométrique (plans basés sur la distance entre les points)

A.4.1 Construction de plans basés sur un critère statistique d'optimalité

La première grande classe de dispositifs expérimentaux dont nous allons parler ici est la classe de plans construits à partir d'un critère statistique d'optimalité. Un grand nombre de critères ont été présentés dans la bibliographie, nous donnerons ici les plus utilisés.

Nous allons présenter les différents critères, donner leur définition, les avantages et inconvénients de chacun.

Le critère d'optimalité le plus couramment utilisé pour les plans d'expériences est le critère de D -optimalité [64].

- **D -optimalité**

Le critère de D -optimalité a été établi par J. Kiefer en 1961. Ce critère est fondé sur le calcul du déterminant de la matrice de variance-covariance associée au modèle. Cette matrice est égale à X^tX où X est la matrice du modèle, soit, pour un polynôme du premier ordre :

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{p1} & \cdots & x_{np} \end{bmatrix} \quad (\text{A.15})$$

où p est le nombre de paramètres du modèle.

Kiefer a montré qu'un plan continu qui maximise $\det(X^tX)$, minimise aussi la variance généralisée de chaque valeur prédite (obtenue en utilisant la fonction de régression) dans l'espace expérimental. C'est à partir de cette constatation qu'il a proposé le critère de D -optimalité.

Définition : Un plan est D -optimal (D pour « déterminant ») si, et seulement si, il minimise la fonction $\det((X^tX)^{-1})$ ou, ce qui est équivalent, maximise $\det(X^tX)$ où X représente la matrice du modèle pour le dispositif expérimental (A.15).

La principale caractéristique d'un plan D -optimal est qu'il permet de minimiser la variance généralisée des estimations des paramètres et par conséquent d'obtenir une région de confiance pour les paramètres de volume le plus petit possible dans l'espace des paramètres. [64]. De ce fait, ce critère d'optimalité a tendance à favoriser la localisation des expérimentations sur les frontières du domaine.

Le critère de D -optimalité est certainement celui le plus utilisé dans le cadre des expériences simulées après les plans dit classiques (orthogonaux). Il existe cependant d'autres critères d'optimalité, tels que la G -optimalité ou la A -optimalité.

- Un plan est dit G-optimal si, et seulement si, la plus grande variance d'estimation sur l'ensemble des essais possibles du domaine expérimental est minimale, c'est-à-dire s'il minimise : $\text{Max} \{ \forall i, X_i ('XX)^{-1} X_i \}$
- Un plan est A-optimal (A comme « *Average* ») s'il minimise la trace de $('XX)^{-1}$

Mais ces critères ne possèdent pas une propriété très importante du critère d'optimalité D : l'invariance par transformation linéaire des facteurs et interactions (ou relations de centrage). En d'autres termes, quelque soit la transformation linéaire effectuée sur les paramètres, le plan considéré reste D-optimal.

L'utilisation de plan optimaux reste conditionnée à une forme de modèle donné, le plan étant optimal pour une matrice X fixée. Dans notre optique, à savoir une augmentation adaptative du dispositif expérimental, les plans optimaux ne semblent donc pas être très adaptés.

- **Entropie**

Un autre critère statistique, proposé par Lindley en 1956, est souvent utilisé : la maximisation de l'entropie. Shewry & Wynn [63] appliquent ce critère pour l'ajustement de données spatiales, Currin et al. [14] et Mitchell & Scott [36] s'en servent pour sélectionner des points lors d'expérimentations simulées.

L'entropie est une mesure de l'imprévisibilité d'une variable aléatoire, c'est aussi une mesure de l'uniformité et de la dispersion de la variable aléatoire. Les plans d'entropie maximale ont pour objectif de maximiser la quantité d'information apportée par les expérimentations.

Définition : Soit Z une variable aléatoire continue de densité $f(x)$. Alors, l'entropie de la variable Z est définie par :

$$\text{Ent}(Z) = -E(\ln(f(Z))) = \int f(z)(\ln(f(z)))dz$$

Suivant la distribution de la variable, il existe un grand nombre d'entropies différentes. Dans le cas particulier d'un processus gaussien Z , Currin et al. [14] ont montré que l'entropie de Z est égale à :

$$\text{Ent}(Z) = \ln(\det(\text{cov}(Z))) + cte$$

Par conséquent, dans ce cas précis, un plan D est d'entropie maximale si et seulement si il maximise la fonction $\det(\mathbf{R}_D)$ où \mathbf{R}_D est la matrice de corrélation du processus Z aux points du plan d'expériences D .

Notons que la quantité d'information d'un plan, et particulièrement l'entropie, est une fonction de la corrélation du processus Z . On peut qualifier l'entropie de « mesure de l'information ».

La construction de ce type de plan reste cependant délicate puisqu'elle dépend d'éléments inconnus, les paramètres de la matrice de corrélation \mathbf{R}_D , qui sont à estimer et sont difficilement estimables. De plus, les plans d'entropie maximale sont onéreux à construire parce qu'ils demandent une optimisation à $n \times p$ dimension pour un plan à n simulations.

- **IMSE (Integrated Mean Squared Error)**

Enfin, le dernier critère statistique cité dans ce paragraphe est proposé par Box et Draper en 1959. Il a pour but de minimiser l'erreur quadratique moyenne intégrée (IMSE) de $\hat{Y}(\mathbf{x})$ dans l'espace $[0,1]^p$. Ce critère semble particulièrement bien adapté à notre cas, car pour des expériences simulées, il n'y a pas d'erreurs de mesures. Par conséquent, notre objectif est de trouver un plan qui minimise l'erreur d'ajustement, c'est-à-dire la différence entre la réponse du simulateur d'écoulement et la valeur prédite par le modèle, sans se préoccuper de la variance d'estimation.

Pour les expériences simulées, les plans construits à partir du critère de IMSE sont présentés en détail par Sacks et al. [55]. Nous donnons ici la définition de ce type de plans d'expériences.

Définition : Pour une valeur du coefficient de corrélation θ fixée, un plan D est un plan de IMSE minimale s'il minimise la fonction objectif suivante :

$$J_{\theta}(D) = \frac{1}{\sigma^2} \int_{[0,1]^p} E_{\theta} \left[Y(\mathbf{x}) - \hat{Y}(\mathbf{x}) \right]^2 d\mathbf{x}$$

où \hat{Y} représente le meilleur estimateur linéaire sans biais de Y .

Remarque : Il est possible d'inclure une fonction de poids dans l'intégrale de l'erreur quadratique moyenne.

Ce critère implique que le modèle et la longueur de corrélation θ du processus résiduel Z soient connus. Comme pour le critère d'entropie, la minimisation de $J_{\theta}(D)$ est une optimisation à $n \times p$ dimensions. Sacks et al. [55] proposent une optimisation de Quasi-Newton, mais celle-ci est très longue en temps de calcul. De plus, du fait des nombreux minima locaux, plusieurs données initiales sont nécessaires pour obtenir un résultat d'optimisation fiable.

Notons qu'une alternative au critère IMSE est d'utiliser le critère MMSE (Maximum Mean Squared Error). Au lieu d'intégrer la MSE (Mean Squared Error) de prédiction, la MMSE est un critère minimax qui choisit le plan qui minimise la fonction :

$$\max_{x \in D} \text{MSE} \left[\hat{Y}(x) \right].$$

Ce critère est plus simple à mettre en place car ne nécessite pas d'intégration à plusieurs dimensions. Une comparaison entre les critères IMSE et MMSE est présentée par Sacks & Schiller [53].

A.4.2 Les « Space Filling Designs »

Plusieurs critères, basés sur des notions géométriques ont été introduits pour construire des plans dits « space-filling designs ». Nous présentons dans ce paragraphe quatre méthodes de sélection de points dans la région expérimentale pour la construction de tels dispositifs.

Dans un premier temps, nous évoquons des plans construits à partir d'une mesure de distance entre les points, ce qui permet de ne pas avoir deux points trop proches dans le domaine expérimental. Ensuite, nous considérons deux types de plans construits à partir de méthodes d'échantillonnages du domaine expérimental, contraintes à des propriétés de projection spécifiques. Une autre possibilité de construction de plan est de sélectionner les points se rapprochant le plus possible de l'uniformité.

- **Les plans de distance maximin et minimax**

Étant donné que la fonction de corrélation du processus résiduel est stationnaire, nous pouvons l'écrire sous la forme :

$$R(x, u) = g(d(x, u))$$

où g est une fonction définie sur \mathbb{R} et $d(.,.)$ une distance.

Par conséquent, seuls les points les plus proches deux à deux contribuent de façon significative au déterminant de la matrice de corrélation. L'idée est donc de trouver un plan tel que la distance minimale entre deux points soit la plus grande possible. C'est ce qu'on appelle la distance maximin. En 1990, Johnson et al. [24] proposent l'utilisation de plans de distance minimax et maximin pour les expériences simulées.

Définitions : Soit $d(.,.)$ une distance (en général distance rectangulaire ou euclidienne) et T un échantillon de points.

Un plan maximin D° est un plan qui maximise la distance minimale entre deux points :

$$\max_D \min_{s, s' \in D} d(s, s') = \min_{s, s' \in D^\circ} d(s, s')$$

Un plan minimax D^* est un plan qui minimise la distance maximale entre deux points :

$$\min_D \max_{t \in T} d(s, t) = \max_{t \in T} d(s, D^*)$$

Les plans de distance maximin sont beaucoup plus connus que ceux de distance minimax, principalement pour des raisons de construction. Le principal avantage des plans maximin est qu'ils imposent une distance minimale entre deux points du plan, il n'y a donc pas de zones où plusieurs points sont très proches. Les points sont relativement distants les uns des autres. En fait, dans le cas d'une distance euclidienne (on considère des sphères de rayon r autour de chaque point), l'idée est de remplir l'espace expérimental avec des points et les sphères associées, tout en maximisant le rayon de la sphère. Les sphères peuvent sortir du domaine, mais pas les points. La construction de ce type de plan est illustrée sur la Figure A.1.

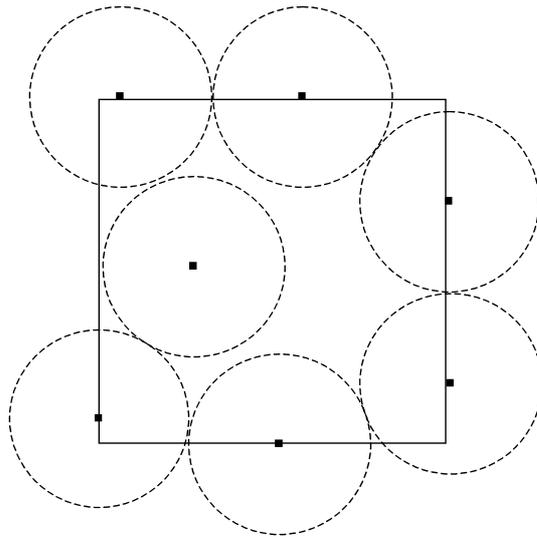


Figure A.1 : Plan maximin pour la distance euclidienne pour $n = 7$ points dans $[-1, 1]^2$

Malgré le fait que ces points sont relativement bien répartis dans l'espace des paramètres, nous voyons sur la Figure A.1 qu'ils ont tendance à rester souvent concentrés aux frontières de la région expérimentale.

De plus, de tels plans ont des caractéristiques asymptotiques optimales. En effet, Johnson et al. [24] ont montré l'équivalence entre la D-optimalité et les plans maximin pour de petites corrélations. Ils ont aussi montré que si la corrélation est une fonction décroissante de la distance (par exemple une corrélation gaussienne), alors un plan d'entropie maximale est, dans un certain sens, un cas limite des plans de distance maximin, i.e. pour \mathbb{R}^k quand k tend vers l'infini.

Par contre, l'inconvénient de ces plans et que, particulièrement dans une région hypercubique, ils ne sont pas bien représentés lorsqu'on fait une projection sur une plus petite dimension de l'espace du plan. Ils dépendent aussi de la fonction de corrélation mais seulement au travers de la distance utilisée, la connaissance du paramètre de corrélation θ n'est donc pas nécessaire. De plus, ils sont souvent très longs et difficiles à construire, si une sélection de points n'a pas été établie au préalable.

Passons à présent aux plans d'expériences qui sont certainement les plus connus et les plus utilisés dans le contexte d'expériences simulées : les hypercubes latins.

- **Les hypercubes latins**

En 1979, McKay, Beckman et Conover [35] ont été les premiers à considérer les plans d'expériences pour des expériences déterministes. Ils utilisent alors des hypercubes latins. Cette méthode a été mise en place pour gagner en précision par rapport aux tirages aléatoires ou stratifiés afin d'estimer la moyenne, la variance et la fonction de la distribution d'une réponse. Les hypercubes latins sont devenus populaires en ingénierie à partir du moment où ils ont été utilisés pour la planification et l'analyse d'expériences simulées par Sacks et al. [55].

Définition : *Un hypercube latin est une matrice de n lignes et k colonnes où n est le nombre de niveaux examinés et k est le nombre de variables d'entrées. Chacune des k colonnes contient les niveaux $1, 2, \dots, n$ permutés aléatoirement et les k colonnes sont disposées aléatoirement pour former l'hypercube latin.*

Plus généralement, le tableau $P=(p_{ij})$ d'ordre $n \times k$ à valeurs dans (\mathbb{Z}/n) est un hypercube latin si quelque soit J , $|J|=1$ toutes les cellules de la J -marge sont représentées une fois chacune par les lignes de P .

Les hypercubes latins assurent que chaque niveau des paramètres soit représenté une seule fois sur chaque marge et par conséquent, ils permettent d'avoir des expérimentations non concentrées sur les bords du domaine expérimental.

Les plans basées sur l'échantillonnage type hypercubes latins ont très connus et utilisés dans le cas des expériences simulées parce qu'ils ont trois grandes qualités. La première qualité de ces plans est qu'ils permettent généralement de remplir au mieux le domaine expérimental, ils vont donc être utile pour détecter les irrégularités de la surface. De plus, ils sont très faciles et rapides à construire, même pour un nombre élevé de facteurs et enfin, tous les niveaux de chaque facteur sont représentés une unique fois.

Par contre, pour un nombre donné N d'expérimentations et p paramètres, il existe $(N!)^p$ possibilités de construction d'un hypercubes latins et par conséquent, les résultats de la modélisation sont très variables suivant le plan considéré. Par exemple, la Figure A.2 montre un exemple de deux Hypercubes Latin obtenus pour deux paramètres et neuf expérimentations.

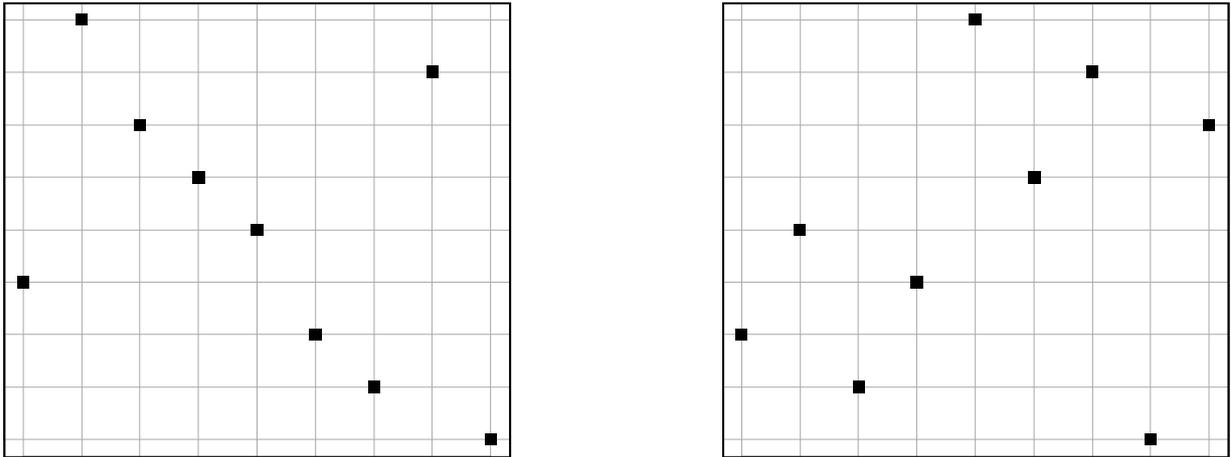


Figure A.2 : Hypercubes latins pour 2 paramètres - 9 expérimentations

- **Les tableaux orthogonaux**

Une extension des hypercubes latins est donnée par les tableaux orthogonaux. Un hypercube latin impose aux points d'échantillonnage des contraintes sur les marges de dimension un. Cette fois, on tient compte des projections des points du plan sur des espaces de plus faible dimension. Ces plans sont présentés dans la bibliographie comme appropriés aux expériences simulées [42].

Définition : Soit T un tableau d'ordre $N \times d$ à éléments dans (\mathbb{Z}/q) . T est un tableau orthogonal de force t , $1 \leq t \leq d$ à q symboles et d'index λ si et seulement si dans tout bloc formé de t colonnes de T , tous les éléments de $(\mathbb{Z}/q)^t$ figurent un même nombre de fois λ chacun. On le note $OA_\lambda(N, d, q, t)$ où $N = \lambda q^t$.

Remarque : Un hypercube latin est un tableau orthogonal de force un.

Le principal avantage des plans orthogonaux est l'orthogonalité des colonnes. Cette caractéristique va permettre une meilleure estimation des effets simples du modèle et des interactions jusqu'à l'ordre $\lambda-1$, ceux-ci seront indépendants. Ceci est principalement intéressant pour les analyses de sensibilité.

De plus amples détails sont donnés par Wu et Hadama [73].

En 1991, Tang [67] propose l'utilisation de tableaux orthogonaux pour la construction d'hypercubes latins. Ce sont les tableaux orthogonaux les plus utilisés dans le cas d'expériences simulées. Ces plans contiennent q^2 expérimentations et deviennent des tableaux orthogonaux de force deux après regroupement des observations. Ceci permet une amélioration des propriétés des hypercubes latins, du fait de l'uniformité de la projection sur les marges de dimensions inférieure à t .

Le principal inconvénient de ce type de plan est que le nombre d'expérimentations du tableau orthogonal n'est pas complètement libre ($N=\lambda q^l$). Ils sont généralement beaucoup plus coûteux que les autres plans cités ici, surtout pour de grandes dimensions.

Notons que ces hypercubes latins, construits à partir de tableaux orthogonaux, sont très proches des travaux issus des méthodes de quasi-Monte Carlo. Mais cette méthode dispose d'un avantage supplémentaire que celle de quasi-Monte Carlo parce qu'elle prend en compte les marges de plus faible dimensions et donc assurent l'uniformité des projections sur chaque marge.

Nous présentons à présent des plans issus de méthodes de quasi-Monte Carlo, aussi appelés plans de discrédance minimale.

- **Les plans de discrédance minimale**

La méthode que nous présentons ici consiste à comparer la distribution des points d'un plan d'expériences P à la distribution uniforme, pour cela la notion de discrédance est nécessaire.

La discrédance d'un plan P mesure l'ampleur à laquelle le plan diffère d'une distribution complètement uniforme. Les plans de faible discrédance sont considérés comme « plus uniforme » que les plans de grande discrédance. Par plan uniforme, nous considérons un plan dont les points sont uniformément dispersés dans le domaine [22].

Définition : Soit D l'espace dans lequel x peut prendre ses valeurs. La mesure de discrédance D_p est définie par :

$$D_p(P) = \left[\int_D |F_n(x) - F(x)|^p dx \right]^{1/p}$$

où $F_n(x)$ est la distribution empirique des points de P , soit :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq x\}$$

La mesure de discrédance la plus populaire est obtenue pour $p = \infty$, elle vaut :

$$D_\infty(P) = \sup_{x \in D} |F_n(x) - F(x)|$$

L'utilisation de la discrédance peut être envisagée pour générer des séquences de points uniformément distribuées dans le contexte de méthodes de quasi-Monte Carlo. Les méthodes de type quasi-Monte Carlo sont des versions déterministes des méthodes de Monte Carlo pour lesquelles les nombres aléatoires sont remplacés par des nombres déterministes qui forment des ensembles ou des suites à faible discrédance, celle-ci ayant une meilleure distribution uniforme. Les suites de quasi-Monte-Carlo ou suites à faible discrédance, permettant de choisir de manière efficace la position des points parmi un espace de points possibles.

On peut utiliser plusieurs types de discrédance qui donnent l'uniformité, celles-ci aboutissent à plusieurs constructions de plans différentes. Suivant les méthodes, les plans obtenus sont plus ou moins orthogonaux.

L'inconvénient ici est la discrédance est très délicate à évaluer pour de grandes dimensions et nous n'avons aucun *a priori* pour choisir un type de discrédance plutôt qu'un autre.

A.4.3 Combinaison de plusieurs critères

Aucun de ces critères, statistique ou géométrique, n'est complètement satisfaisant pour sélectionner un plan, même s'ils produisent des plans ayant des propriétés attractives. Pour certains, les coûts de construction sont énormes quand le nombre de variables augmente et les dispositifs sont très difficiles et longs à construire. De plus, certains critères demandent des connaissances *a priori* sur la forme du modèle, ce qui n'est pas le cas en ingénierie de réservoir. Du fait de la bonne répartition spatiale des expérimentations, l'utilisation de plans « space-filling design » semble être appropriée parce que ces plans permettent de déterminer les irrégularités de la surface et d'améliorer les méthodes d'interpolation telles que le krigeage ou le cokrigeage. De plus, l'erreur de prédiction en chaque point du plan est nulle et par conséquent, les plans qui ne sont pas des « space-filling design » ou les plans pour lesquels les points sont proches des frontières de la région expérimentale, peuvent donner des prédicteurs de moindre qualité à l'intérieur du domaine. Il semble donc que les « space-filling design » et plus particulièrement les hypercubes latins, sont à privilégier pour la modélisation de réponses non-régulières, ils permettent d'imposer un minimum de points à l'intérieur du domaine expérimental.

Les plans basés sur un échantillonnage de type hypercube latin sont les plus connus dans le cas des expériences simulées du fait de trois de leurs propriétés :

- Ils permettent de couvrir au mieux le domaine expérimental, ils sont donc utiles pour détecter les irrégularités de la surface,
- Ils sont de construction simple et rapide,
- Les niveaux de chaque facteurs sont représentés une unique fois.

Le choix d'un hypercube latin parmi tous ceux possibles a une grande influence sur la qualité de la modélisation de la réponse. C'est pour cela qu'il est nécessaire d'imposer une condition supplémentaire pour diminuer l'espace des possibilités et ainsi réduire l'impact du choix du plan. Cette condition revient, par exemple, à combiner les avantages de plusieurs critères simultanément. Une idée pour construire de tels plans peut être de réduire les points candidats à une classe finie à l'aide d'un premier critère et ensuite à appliquer à cet ensemble de plans possibles un second critère pour sélectionner le plan final.

Des critères combinés pour construire des hypercubes latins optimaux ont été présentés par plusieurs auteurs. Nous les citons ci-dessous.

Park [43] a étudié les hypercubes latins optimaux basés sur le critère IMSE. Mais là encore, ces plans dépendent des paramètres de corrélations qui sont inconnus. De plus, ils sont souvent difficiles à construire et très coûteux en temps de calcul.

Schonlau, Hadama, et Welch (1996) proposent l'utilisation d'hypercubes latins qui maximisent la distance minimale entre les points dans le cas de projection sur un espace de faible dimension. Owen [42] suggère l'utilisation d'hypercubes latins avec de faibles corrélations « extra-diagonales ». Mais ces deux critères semblent difficiles à mettre en œuvre et eux aussi, dépendent des paramètres inconnus de la fonction de corrélation.

Kenny [28] propose des hypercubes latins orthogonaux et Tang. [67] des hypercubes latins basés sur des tableaux orthogonaux. Ces plans n'ont pas forcément de bonnes propriétés de « space-filling design ». Ils permettent de garder les mêmes propriétés que les plans factoriels, c'est-à-dire l'orthogonalité des colonnes et par conséquent, ces plans sont très utiles pour effectuer une analyse de sensibilité car l'indépendance des effets est préservée. De plus, ils se prêtent bien au krigeage. Mais, le nombre de simulations de tels plans n'est pas flexible, il est souvent très grand, ce qui dans notre cas, pose problème au vu du coût des simulations.

Pour terminer, Morris & Mitchell [38] introduisent des hypercubes latins optimaux pour les critères d'entropie et de distance maximin. L'utilisation de ces plans est motivée par un résultat de Johnson et al. [24] qui établit une équivalence entre les propriétés des Hypercubes Latins de distance maximin et une sorte de D-optimalité, dans une limite de petites corrélations locales (i.e. pour un coefficient de corrélation θ petit). En fait, ce type de plan assure un bon compromis entre les critères d'entropie et maximin et a de bonnes propriétés de projections sur chaque dimension. De plus, ils permettent une bonne qualité de prédiction.

Dans toute la bibliographie citée précédemment, les plans sont d'abord restreints à la classe des hypercubes latins et ensuite un second critère est appliqué à cette classe dans le but de choisir un hypercube latin avec des propriétés désirables. Cependant, d'autres critères peuvent être utilisés pour réduire la classe des plans candidats [4]. Par exemple, on peut la réduire aux plans uniformes et ensuite appliquer un second critère à cette classe restreinte.

La liste des plans d'expériences présentés ici est loin d'être exhaustive, mais une étude détaillée des différents plans d'expériences existants n'est pas l'objet de cette thèse. Nous nous sommes contenté de citer les plans les plus classiquement utilisés en ingénierie de réservoir ou pour les expériences simulées.

Annexe B

TECHNIQUES DE MODELISATION

L'objectif de cet annexe est de présenter différentes techniques de modélisation, principalement issues de la géostatistique. Nous commençons tout d'abord par définir un réseau de neurones et comment l'ajuster par la phase d'apprentissage. Cette technique est encore peu utilisée en ingénierie de réservoir, mais reste toujours d'actualité car les méthodes d'apprentissage progressent. Ensuite, nous détaillons la manière de construire un estimateur par krigeage simple (ou à moyenne connue). Nous introduisons également les équations régissant un estimateur par cokrigeage. Enfin, nous rappelons les principales propriétés d'un variogramme géostatistique (ou covariance) et donnons des méthodes permettant d'ajuster les paramètres du variogramme.

B.1 Réseaux de Neurones

B.1.1 Définition d'un réseau de neurones

Un réseau de neurones typique contient plusieurs niveaux de neurones, chaque neurone est connecté à tous les neurones du niveau précédent par un simple lien pondéré. Le nombre de neurones d'entrée et de neurones de sortie dépend du domaine d'application.

Un neurone est une sorte d'automate qui se compose d'entrées, d'une sortie et d'une fonction des entrées vers la sortie. Il peut être schématisé de la façon suivante :

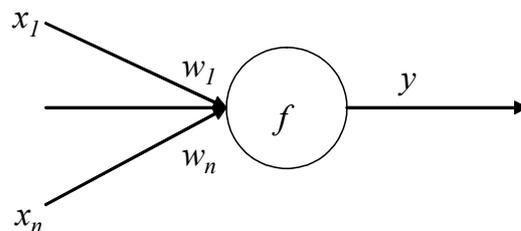


Figure B.1 : Représentation d'un neurone formel selon Mc Culloch et Pitts

où :

- les x_i représentent les données d'entrées,
- w_i représente le poids de la $i^{\text{ème}}$ connexion,
- f est la fonction d'activation,
- $y = f(w_1 x_1 + \dots + w_n x_n)$ représente la sortie du neurone.

Dans le modèle d'origine, la fonction d'activation f est indifféremment une fonction de Heaviside :

$$f(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

ou, une fonction « signe » :

$$f(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ -1 & \text{si } x < 0 \end{cases}$$

Le neurone de Mc Culloch et Pitts a été généralisé en choisissant d'autres fonctions d'activations, telles que des fonctions linéaires par morceaux, des sigmoïdes ou autres gaussiennes (cf. Figure B.2). Les sigmoïdes sont de loin les fonctions les plus utilisées dans les réseaux de neurones, ce sont des fonctions continues, strictement croissantes et qui possèdent les comportements asymptotiques désirés. Et particulièrement la fonction logistique définie par :

$$g(x) = (1 + e^{-\beta x})^{-1}$$

où β est le paramètre de pente de la sigmoïde.

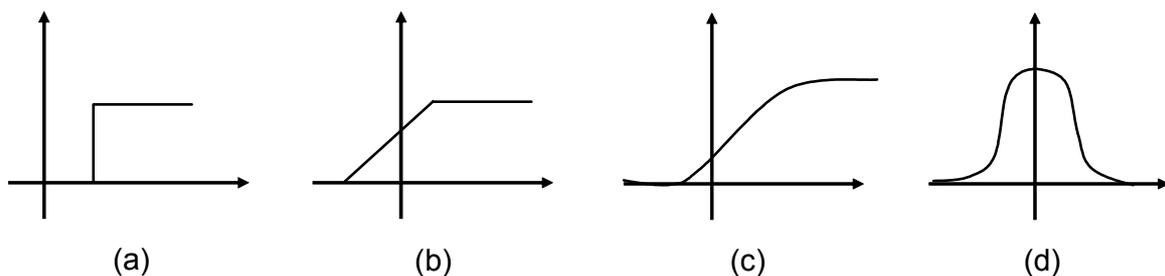


Figure B.2 : Fonctions d'activation : (a) seuil, (b) linéaire par morceau, (c) sigmoïde, (d) Gaussienne

Les neurones formels que l'on vient de définir doivent être assemblés pour former un réseau. Comme son nom l'indique, un réseau de neurones est constitué de plusieurs couches de neurones entièrement connectées entre elles (cf. Figure B.3). En fait, un réseau de neurones est constitué d'unités élémentaires, les nœuds, et de liens directionnels affectés de poids

reliant ces nœuds. Les nœuds peuvent être du type entrée, sortie ou caché par des nœuds intermédiaires. Chaque nœud reçoit en entrée une valeur calculée à partir des valeurs restituées par ses nœuds amonts et pondérées par les poids associés aux liens.

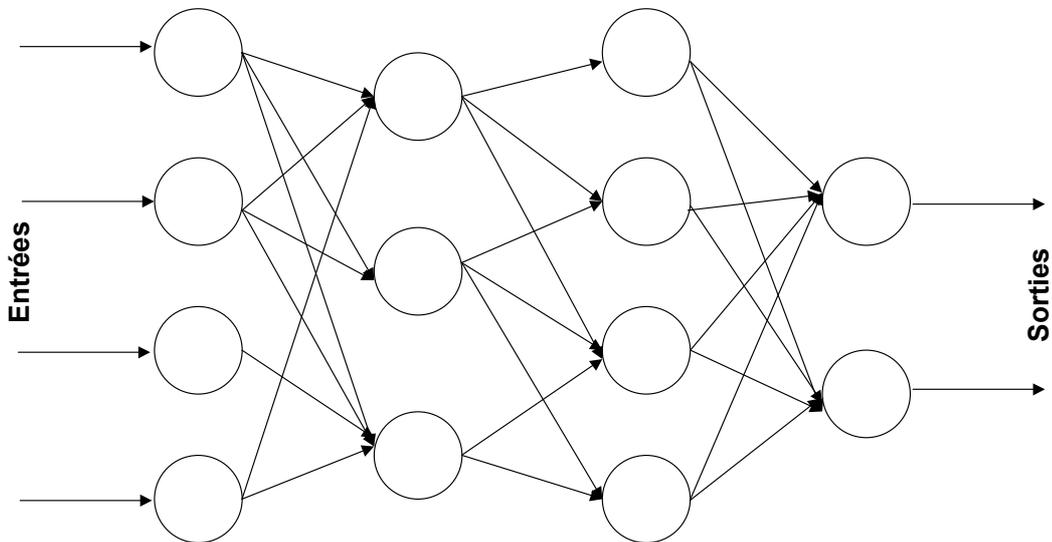


Figure B.3 : Représentation d'un réseau de neurone multicouche

Les réseaux de neurones déterminent la nature de la relation entre les données à travers les interconnexions entre les neurones. Le réseau est nourri par les valeurs des paramètres en entrée et dans le cas d'un réseau multicouche, tel celui représenté en Figure B.3, les résultats en sortie sont obtenus par propagation des valeurs suivant la direction des liens et en classant les nœuds suivant la topologie du réseau (les nœuds de la première couche sont activés avant les nœuds de la deuxième, etc.).

Afin de construire un bon réseau, une phase d'apprentissage est nécessaire. Il existe plusieurs type d'apprentissage d'un réseau de neurone, nous ne présentons que celui le plus employé en ingénierie de réservoir : l'apprentissage supervisé.

B.1.2 Apprentissage d'un réseau de neurone

Le but de l'apprentissage supervisé est d'inculquer un comportement de référence au réseau. Ce comportement de référence est défini à partir d'un ensemble représentatif d'exemples appelé base d'apprentissage. Cet ensemble est composé de couples d'apprentissage formés de vecteurs d'entrée appelé patron d'entrée et d'un vecteur de sortie appelé patron de référence.

L'apprentissage consistera alors à déterminer les poids des différentes liaisons du réseau de manière à obtenir, lorsqu'on présente au réseau un patron d'entrée, une valeur en sortie aussi proche que possible du patron de référence correspondant.

L'apprentissage comprend quatre étapes de calcul :

1. Initialisation des poids de réseau,
2. Présentation des patrons d'entrée de la base d'apprentissage et calcul de l'état du réseau,
3. Calcul de l'erreur. L'erreur est une fonction de la différence entre la sortie obtenue et la sortie désirée. On choisira en général l'erreur quadratique,
4. Modification des poids synaptiques. Cette modification s'effectue grâce à un algorithme d'apprentissage, comme par exemple, la rétro propagation de l'erreur. Cette technique calcule les variations de poids pour chaque jeu de paramètres des cas d'apprentissage avec la règle suivante :

$$\Delta w_{ij} = \eta \delta_j o_i,$$

$$\delta_j = \begin{cases} a_j(\text{net}_j)(t_j - o_j) & \text{si le noeud } j \text{ est une sortie} \\ a_j(\text{net}_j) \sum \delta_k w_{jk} & \text{si le noeud } j \text{ est intermédiaire} \end{cases}$$

où :

- η est une constante,
- t_j est la mesure au nœud de sortie j ,
- w_{jk} est le poids d'un lien du nœud j vers le nœud k .

Ce calcul de perturbation des poids suit la règle de Hebb qui tend à augmenter le poids d'un lien quand ses deux extrémités sont actives en même temps.

Les étapes 2-3-4 sont répétées jusqu'à la fin de l'apprentissage : on s'arrête lorsque l'erreur descend en dessous d'une certaine valeur ou après un certain nombre d'itération.

Une fois que la convergence des poids w_i est obtenue, il est préférable de tester le réseau de neurones pour vérifier si l'on obtient bien les sorties voulues. Si les résultats fournis par le réseau de neurones ne sont pas de qualité suffisante, une nouvelle phase d'apprentissage est réalisée avec de nouveaux poids initiaux, et un échantillon d'apprentissage plus grand, et ainsi de suite.

Une description détaillée du fonctionnement d'un réseau de neurones se trouve dans [68].

Ces deux phases, d'apprentissage et de test du réseau sont très importantes, elles assurent la robustesse du réseau et donc la fiabilité des estimations.

B.2 Le krigeage simple ou à moyenne connue

La méthode de krigeage est très populaire en géostatistique, elle consiste à construire un interpolateur qui soit fonction linéaire des observations et tenant compte de la structure spatiale des données. Soit Z le processus à modéliser et $\mathbf{x}_1, \dots, \mathbf{x}_n$ les points d'observation de ce processus. Une condition à l'utilisation du krigeage est que le processus Z soit considéré stationnaire. Comme nous ne manipulons ici que des espérances et des variances de combinaisons linéaires (le krigeage est un estimateur linéaire), la fonction aléatoire Z est entièrement définie par ses deux premiers moments. Posons :

- $\mu(\mathbf{x}_i) = E(Z(\mathbf{x}_i)) \quad 1 \leq i \leq n$
- $C(\mathbf{h}) = E(Z(\mathbf{x}_i + \mathbf{h})Z(\mathbf{x}_i)) - \mu(\mathbf{x}_i)^2$

Le prédicteur de krigeage de la réponse Z au point \mathbf{x}_0 du domaine expérimental est une fonction linéaire des observations $Z(\mathbf{x}_i)$ aux points du plan : tout estimateur linéaire de $Z(\mathbf{x})$ à partir des $(Z(\mathbf{x}_i))_{1 \leq i \leq n}$ s'écrit :

$$\hat{Z}(\mathbf{x}_0) = \sum_i \lambda_i Z(\mathbf{x}_i) + k \quad (\text{B.1})$$

L'objectif est de déterminer les poids λ_i qui permettent d'estimer la valeur de la fonction en \mathbf{x}_0 tout en interpolant les données déjà observées. Ces poids dépendent de l'endroit où l'on se trouve dans le domaine.

La constante k et les poids optimaux λ_i , sont sélectionnés de manière à minimiser l'erreur d'estimation, ce qui revient à minimiser la variance de l'erreur au point \mathbf{x}_0 .

Ainsi, on veut minimiser :

$$E\left(Z(\mathbf{x}_0) - \hat{Z}(\mathbf{x}_0)\right)^2 = \text{Var}\left(Z(\mathbf{x}_0) - \hat{Z}(\mathbf{x}_0)\right) + \left(E\left(Z(\mathbf{x}_0) - \hat{Z}(\mathbf{x}_0)\right)\right)^2 \quad (\text{B.2})$$

L'équation (B.2) est la somme de deux termes positifs, et par conséquent, minimiser (B.2) revient à minimiser les deux termes.

Concentrons nous tout d'abord sur l'estimation de la constante k , ce qui revient à minimiser le deuxième terme de l'équation (B.2).

D'après l'équation (B.2), nous avons $\mu(\mathbf{x}_0) = \sum_i \lambda_i \mu(\mathbf{x}_i) + k$ avec $\mu(\mathbf{x}_0) = E(Z(\mathbf{x}_0))$.

Minimiser le second terme de (B.2), revient à l'annuler, ce qui assure un estimateur sans biais pour Z . On obtient donc facilement la valeur de la constante k :

$$k = \mu(\mathbf{x}_0) - \sum_i \lambda_i \mu(\mathbf{x}_i) \quad (\text{B.3})$$

Posons $\boldsymbol{\lambda} = {}^t(\lambda_1, \dots, \lambda_n)$, $\mathbf{Z} = {}^t(Z(\mathbf{x}_1), \dots, Z(\mathbf{x}_n))$ et $\boldsymbol{\mu} = {}^t(\mu_1, \dots, \mu_n)$.

En remplaçant la valeur de k dans l'équation (B.1) par l'expression (B.3), l'estimateur de Z en \mathbf{x}_0 devient donc, sous forme matricielle :

$$\begin{aligned}\hat{Z}(\mathbf{x}_0) &= {}^t\lambda\mathbf{Z} + k \\ &= {}^t\lambda\mathbf{Z} + \mu(\mathbf{x}_0) - {}^t\lambda\boldsymbol{\mu}\end{aligned}$$

$$\text{Soit, } \hat{Z}(\mathbf{x}_0) = \mu(\mathbf{x}_0) + {}^t\lambda(\mathbf{Z} - \boldsymbol{\mu}) \quad (\text{B.4})$$

Passons à présent à la minimisation du premier terme de l'équation (B.2). Nous avons :

$$\begin{aligned}\text{Var}(Z(\mathbf{x}_0) - \hat{Z}(\mathbf{x}_0)) &= \text{Var}\left\{Z(\mathbf{x}_0) - \sum \lambda_i Z(\mathbf{x}_i)\right\} \\ &= \text{Var}\{Z(\mathbf{x}_0)\} - 2\sum_i \lambda_i \text{Cov}(Z(\mathbf{x}_0), Z(\mathbf{x}_i)) + \sum_{i,j} \lambda_i \lambda_j \text{Cov}(Z(\mathbf{x}_i), Z(\mathbf{x}_j))\end{aligned} \quad (\text{B.5})$$

Posons $C_{ij} = \text{Cov}(Z(\mathbf{x}_i), Z(\mathbf{x}_j))$

Pour minimiser l'équation précédente nous devons annuler la dérivée de (B.5) par rapport à λ_i , soit :

$$\frac{\partial}{\partial \lambda_i} \text{Var}(Z(\mathbf{x}_0) - \hat{Z}(\mathbf{x}_0)) = \sum_{i,j} \lambda_i \lambda_j C_{ij} - 2\sum_i \lambda_i C_{0i} \quad (\text{B.6})$$

Ainsi, les poids λ_i sont les solutions du système linéaire suivant :

$$2\sum_j \lambda_j C_{ij} - 2C_{0i} = 0 \quad (\text{B.7})$$

Par conséquent, en écriture matricielle, les poids optimaux λ_i sont solution de l'équation :

$$\mathbf{C}\boldsymbol{\lambda} = \mathbf{c}(\mathbf{x}_0) \text{ avec } \mathbf{C} = (C_{ij})_{i,j} \text{ et } \mathbf{c}(\mathbf{x}_0) = (C_{0i})_i \quad (\text{B.8})$$

Et on obtient :

$$\hat{Z}(\mathbf{x}_0) = \mu(\mathbf{x}_0) + {}^t\mathbf{c}(\mathbf{x}_0)\mathbf{C}^{-1}(\mathbf{Z} - \boldsymbol{\mu}) \quad (\text{B.9})$$

Dans le cas de la modélisation adaptative par modèle composite, du fait de l'ajustement de la moyenne par un estimateur sans biais, les résidus sont de moyenne nulle ($\boldsymbol{\mu} = 0$), d'où :

$$\hat{Z}(\mathbf{x}_0) = {}^t\mathbf{c}(\mathbf{x}_0)\mathbf{C}^{-1}\mathbf{Z} \quad (\text{B.10})$$

Il s'agit du krigeage simple. Pour une description plus détaillée du krigeage simple ou pour connaître les fondements du krigeage ordinaire et universel, nous renvoyons à [7] et [26].

De plus, dans le cadre de cette étude, nous avons $\hat{Z}(\mathbf{x}_0) = \hat{Y}(\mathbf{x}_0) - \mathbf{X}(\mathbf{x}_0)\hat{\boldsymbol{\beta}}$, donc le prédicteur de Y au point \mathbf{x}_0 obtenu par le krigeage est :

$$\hat{Y}(\mathbf{x}_0) = \mathbf{X}(\mathbf{x}_0)\hat{\boldsymbol{\beta}} + {}^t\mathbf{c}(\mathbf{x}_0)\mathbf{C}^{-1}[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}] \quad (\text{B.11})$$

avec :

- \mathbf{X} la matrice du modèle aux points du plan d'expériences,
- \mathbf{C} la matrice de covariance aux points du plan d'expériences,
- $\mathbf{c}(\mathbf{x}_0)$ le vecteur des covariances entre \mathbf{x}_0 et les points du plan d'expériences.

Un autre grand avantage du krigeage par rapport à d'autres méthodes d'interpolation est qu'il permet de calculer la variance associée à l'estimateur. D'après (B.5), nous avons

$$\begin{aligned} \text{Var}(Y(\mathbf{x}_0) - \hat{Y}(\mathbf{x}_0)) &= \text{Var}\{Y(\mathbf{x}_0) - {}^t\boldsymbol{\lambda}\mathbf{Y}\} \\ &= \text{Var}\{Y(\mathbf{x}_0) - {}^t2\boldsymbol{\lambda}\mathbf{c}(\mathbf{x}_0) + {}^t\boldsymbol{\lambda}\mathbf{C}\boldsymbol{\lambda}\} \end{aligned}$$

Or, d'après (B.8), ${}^t\boldsymbol{\lambda} = \mathbf{c}(\mathbf{x}_0)\mathbf{C}^{-1}$

D'où :

$$\begin{aligned} \text{Var}(Y(\mathbf{x}_0) - \hat{Y}(\mathbf{x}_0)) &= \text{Var}\{Y(\mathbf{x}_0) - 2{}^t\mathbf{c}(\mathbf{x}_0)\mathbf{C}^{-1}\mathbf{c}(\mathbf{x}_0) + {}^t\mathbf{c}(\mathbf{x}_0)\mathbf{C}^{-1}\mathbf{C}^t\mathbf{C}^{-1}\mathbf{c}(\mathbf{x}_0)\} \\ &= \text{Var}\{Y(\mathbf{x}_0) - {}^t\mathbf{c}(\mathbf{x}_0)\mathbf{C}^{-1}\mathbf{c}(\mathbf{x}_0)\} \end{aligned}$$

Soit encore :

$$\text{Var}\{Y(\mathbf{x}_0) - \hat{Y}(\mathbf{x}_0)\} = \sigma^2(1 - {}^t\mathbf{r}(\mathbf{x}_0)\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}_0)) \quad (\text{B.12})$$

Nous remarquons que la variance du prédicteur est plus faible que celle de $Y(\mathbf{x}_0)$, le dernier terme étant toujours positif.

B.3 Le Cokrigeage

Cette session présentera rapidement la théorie du cokrigeage en deux dimensions. Pour chaque point du plan d'expériences, nous disposons d'un résidu Z , et de deux valeurs pour la dérivée, suivant deux directions \mathbf{x}^1 et \mathbf{x}^2 , soit respectivement : $\frac{\partial Z}{\partial x^1}$ et $\frac{\partial Z}{\partial x^2}$. Pour une description détaillée, nous renvoyons à [7] et [26].

L'objectif ici est prédire la réponse Z au point \mathbf{x}_0 , en utilisant conjointement les valeurs de simulations et les valeurs des gradients dans chaque direction.

Nous avons par cokrigage :

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^n \lambda_{0i} Z(\mathbf{x}_i) + \sum_{i=1}^n \lambda_{1i} \frac{\partial Z}{\partial x^1}(\mathbf{x}_i) + \sum_{i=1}^n \lambda_{2i} \frac{\partial Z}{\partial x^2}(\mathbf{x}_i) \quad (\text{B.13})$$

Dans le cas du krigeage, on considérait les observations des résidus qui sont de moyenne nulle. Ici, la moyenne des résidus Z est toujours nulle ($\mu_Z = 0$), mais celle de ses dérivées, respectivement μ_1 , μ_2 , sont inconnues. Par conséquent, nous effectuons du cokrigage ordinaire ou à moyenne inconnue. D'après l'équation (B.13), nous avons

$$\mu(\mathbf{x}_0) = \sum_{i=1}^n \lambda_{0i} \mu_Z(\mathbf{x}_i) + \sum_{i=1}^n \lambda_{1i} \mu_1(\mathbf{x}_i) + \sum_{i=1}^n \lambda_{2i} \mu_2(\mathbf{x}_i) \quad (\text{B.14})$$

Par définition, l'estimateur de cokrigage est un estimateur sans biais et par conséquent, la condition de non-biais peut se traduire par les contraintes suivantes :

$$\left\{ \begin{array}{l} \sum_{i=1}^n \lambda_{0i} = 1 \\ \sum_{i=1}^n \lambda_{1i} = 0 \\ \sum_{i=1}^n \lambda_{2i} = 0 \end{array} \right.$$

En calculant la valeur de la covariance de Z_0 en fonction de Z et de ses deux dérivées et en ajoutant la condition de non-biais précédente, trouver l'estimateur de cokrigage de Z au point \mathbf{x}_0 revient à résoudre le système suivant :

$$\left\{ \begin{array}{l} \text{Cov}(Z_0, Z_i) = \sum_{j=1}^n \lambda_{0j} \text{Cov}(Z_i, Z_j) + \sum_{j=1}^n \lambda_{1j} \text{Cov}\left(Z_i, \frac{\partial Z_j}{\partial x^1}\right) + \sum_{j=1}^n \lambda_{2j} \text{Cov}\left(Z_i, \frac{\partial Z_j}{\partial x^2}\right) \\ \text{Cov}\left(Z_0, \frac{\partial Z_i}{\partial x^1}\right) = \sum_{j=1}^n \lambda_{0j} \text{Cov}\left(\frac{\partial Z_i}{\partial x^1}, Z_j\right) + \sum_{j=1}^n \lambda_{1j} \text{Cov}\left(\frac{\partial Z_i}{\partial x^1}, \frac{\partial Z_j}{\partial x^1}\right) + \sum_{j=1}^n \lambda_{2j} \text{Cov}\left(\frac{\partial Z_i}{\partial x^1}, \frac{\partial Z_j}{\partial x^2}\right) \\ \text{Cov}\left(Z_0, \frac{\partial Z_i}{\partial x^2}\right) = \sum_{j=1}^n \lambda_{0j} \text{Cov}\left(\frac{\partial Z_i}{\partial x^2}, Z_j\right) + \sum_{j=1}^n \lambda_{1j} \text{Cov}\left(\frac{\partial Z_i}{\partial x^2}, \frac{\partial Z_j}{\partial x^1}\right) + \sum_{j=1}^n \lambda_{2j} \text{Cov}\left(\frac{\partial Z_i}{\partial x^2}, \frac{\partial Z_j}{\partial x^2}\right) \\ \sum_{i=1}^n \lambda_{0i} = 1 \\ \sum_{i=1}^n \lambda_{1i} = 0 \\ \sum_{i=1}^n \lambda_{2i} = 0 \end{array} \right.$$

Soit, sous forme matricielle : $\mathbf{K}\boldsymbol{\lambda}=\mathbf{k}$, c'est-à-dire:

$$\begin{bmatrix} K_{11} & K_{12} & K_{13} & 1 & 0 & 0 \\ K_{21} & K_{22} & K_{23} & 0 & 1 & 0 \\ K_{31} & K_{32} & K_{33} & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_0 \\ \lambda_1 \\ \lambda_2 \\ \mu_Z \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} k_{10} \\ k_{20} \\ k_{30} \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Les termes K_{ij} sont les covariances directes et croisées entre Z et ses dérivées, et les termes k_{j0} , $1 \leq j \leq 3$ sont les termes de la covariance entre, respectivement les vecteurs Z , $\frac{\partial Z}{\partial x^1}$ et $\frac{\partial Z}{\partial x^2}$. Tous ces termes peuvent être obtenus par différentiation de la fonction de covariance de Z , en supposant que les dérivées existent. Ainsi, nous avons :

$$\begin{aligned} - \quad \mathbf{K}_{11} &= \left(\text{Cov}(Z(\mathbf{x}_i), Z(\mathbf{x}_j)) \right)_{i,j} = \sigma^2 R(Z(\mathbf{x}_i), Z(\mathbf{x}_j)) \\ - \quad \mathbf{K}_{1k} &= \left(\text{Cov}(Z(\mathbf{x}_i), \frac{\partial Z}{\partial x^k}(\mathbf{x}_j)) \right)_{i,j} = \sigma^2 R\left(Z(\mathbf{x}_i), \frac{\partial Z}{\partial x^k}(\mathbf{x}_j)\right)_{i,j} \quad k \neq 1 \\ - \quad \mathbf{K}_{kl} &= \left(\text{Cov}\left(\frac{\partial Z}{\partial x^k}(\mathbf{x}_i), \frac{\partial Z}{\partial x^l}(\mathbf{x}_j)\right) \right)_{i,j} = \sigma^2 R\left(\frac{\partial Z}{\partial x^k}(\mathbf{x}_i), \frac{\partial Z}{\partial x^l}(\mathbf{x}_j)\right)_{i,j} \quad k \neq 1 \\ - \quad \mathbf{k}_{01} &= \sigma^2 R(Z(\mathbf{x}_0), Z(\mathbf{x}_i))_i, \quad \mathbf{k}_{0k} = \sigma^2 R\left(Z(\mathbf{x}_0), \frac{\partial Z}{\partial x^k}(\mathbf{x}_i)\right)_i \quad k \neq 1 \end{aligned}$$

De plus, la variance associée au cokrigage au point \mathbf{x}_0 est donnée par:

$$\begin{aligned} \sigma_{ck}^2 &= \text{Var}(Z_0) + \sum_{i=1}^n \lambda_{0i} \text{Cov}(Z(\mathbf{x}_0), Z(\mathbf{x}_i)) + \sum_{i=1}^n \lambda_{1i} \text{Cov}\left(Z(\mathbf{x}_0), \frac{\partial Z}{\partial x^1} Z(\mathbf{x}_i)\right) \\ &\quad + \sum_{i=1}^n \lambda_{2i} \text{Cov}\left(Z(\mathbf{x}_0), \frac{\partial Z}{\partial x^2} Z(\mathbf{x}_i)\right) - \mu_Z \end{aligned}$$

Le lecteur trouvera plus d'information sur les techniques du cokrigage dans [7] et [9].

B.4 Estimation du Variogramme

Dans cette partie, nous décrivons dans un premier temps les hypothèses que doivent vérifier le variogramme afin de rendre possible la construction d'un modèle de krigeage. Ensuite, nous présentons plus en détails le principe de construction d'un variogramme, nous précisons quelques variogrammes théoriques existants et enfin nous terminons en proposant différentes méthodes possibles pour l'estimation des paramètres du variogramme.

B.4.1 Hypothèses de base

Nous verrons ici les hypothèses indispensables à l'utilisation d'un estimateur par krigeage. Les propriétés que nous présentons sont des résultats classiques de la théorie des processus stochastiques. Leur démonstration ne présente aucun intérêt pour la pratique géostatistique, nous nous contentons ici de citer les propriétés en les commentant.

- **Hypothèse de Stationnarité :**

La première hypothèse indispensable à l'utilisation d'un modèle de krigeage est l'hypothèse de stationnarité du second ordre. Elle vise essentiellement à « détacher » les deux premiers moments de localisations précises en permettant des translations des emplacements \mathbf{x} et $\mathbf{x} + \mathbf{h}$. La covariance devient donc une fonction dépendant uniquement de la distance séparant les points d'observation et non plus de leur localisation exacte.

Définition : Soit X le domaine définissant un processus Z . Le processus Z est stationnaire au sens large s'il vérifie les équations suivantes :

$$- E(Z(\mathbf{x})) = \mu \quad \forall \mathbf{x} \in X$$

$$- Cov(Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})) = C(\mathbf{h}) \quad \forall \mathbf{x} \in X$$

Évidemment, ces hypothèses supposent une certaine régularité de la réponse.

Dans la suite de cet annexe, nous allons supposer que le processus résiduel Z est stationnaire d'ordre 2, autrement dit, tous les points séparés de la même distance ont la même covariance. Nous donnons à présent une autre condition nécessaire sur la covariance pour l'utilisation d'un krigeage sur les résidus.

- **Condition de positivité :**

Nous partons du fait que l'estimateur de Y par krigeage est une combinaison linéaire des observations, et par conséquent, quelque soit la combinaison linéaire, la variance de l'estimateur doit être positive ou nulle.

Soit un processus aléatoire stationnaire du 2nd ordre Z , d'espérance m et de covariance $Cov(\mathbf{h})$. Soit une combinaison linéaire finie quelconque, définie par :

$$Y = \sum_{i=1}^n \lambda_i Z(\mathbf{x}_i) \text{ avec les } \lambda_i \text{ des poids quelconques.}$$

Cette combinaison linéaire Y est un processus aléatoire, sa variance ne doit jamais être négative. Explicitons cette variance :

$$Var(Y) = E(Y^2) - (E(Y))^2$$

Nous calculons chaque terme de cette équation :

$$E(Y) = E\left\{\sum_{i=1}^n \lambda_i Z(\mathbf{x}_i)\right\} = \sum_{i=1}^n \lambda_i E\{Z(\mathbf{x}_i)\} = m \sum_{i=1}^n \lambda_i$$

$$E(Y^2) = E\left\{\left(\sum_{i=1}^n \lambda_i Z(\mathbf{x}_i)\right)^2\right\} = E\left\{\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j Z(\mathbf{x}_i) Z(\mathbf{x}_j)\right\} = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j E\{Z(\mathbf{x}_i) Z(\mathbf{x}_j)\}$$

Or, par définition de la covariance, nous avons :

$$Cov(Z(\mathbf{x}_i), Z(\mathbf{x}_j)) = E\{Z(\mathbf{x}_i) Z(\mathbf{x}_j)\} - E\{Z(\mathbf{x}_i)\} E\{Z(\mathbf{x}_j)\} = E\{Z(\mathbf{x}_i) Z(\mathbf{x}_j)\} - m^2$$

Par conséquent :

$$Var(Y) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j (Cov\{Z(\mathbf{x}_i), Z(\mathbf{x}_j)\} - m^2) - \left(m \sum_{i=1}^n \lambda_i\right)^2$$

Soit enfin, en simplifiant par le terme de moyenne m :

$$Var(Y) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j Cov\{Z(\mathbf{x}_i), Z(\mathbf{x}_j)\} \geq 0 \quad (\text{B.15})$$

La fonction de covariance $Cov(\mathbf{h})$ doit être telle qu'elle assure le caractère toujours positif ou nul de la variance de toute combinaison linéaire finie. La fonction $Cov(\mathbf{h})$ est alors, par définition, dite définie positive.

Ainsi, toute fonction réelle $g(\mathbf{h})$ définie positive peut être considérée comme la covariance d'un processus aléatoire stationnaire. L'équation (B.15) montre que, sous hypothèse stationnaire d'ordre deux (existence de la covariance), toutes combinaisons linéaires finies ou tous produits ont une variance définie, et par conséquent leur usage est légitime [26].

Nous supposons à présent que la covariance est stationnaire du 2nd ordre et définie positive. Cette condition sur la fonction de covariance assure l'existence et l'unicité d'une solution à l'équation de krigeage : $\hat{Y}(\mathbf{x}_0) = \mathbf{X}(\mathbf{x}_0)\hat{\boldsymbol{\beta}} + {}^t\mathbf{c}(\mathbf{x}_0)\mathbf{C}^{-1}[\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}]$.

Le caractère défini positif de la fonction de covariance $Cov(\mathbf{h})$ entraîne les propriétés suivantes :

- $Cov(0) \geq 0$, la variance ne peut pas être négative
- $Cov(\mathbf{h}) = Cov(-\mathbf{h})$, la fonction de covariance est paire
- $|Cov(\mathbf{h})| \leq Cov(0)$ Inégalité de Schwartz

La fonction de covariance décroît en général à partir de sa valeur à l'origine $Cov(0)$, le degré de covariance entre deux variables $Z(\mathbf{x})$ et $Z(\mathbf{x}+\mathbf{h})$ décroît lorsque \mathbf{h} grandit.

B.4.2 Définition du variogramme

En géostatistique, la fonction la plus utilisée pour décrire la continuité d'un processus est le variogramme, et ce surtout parce qu'elle est plus simple à estimer que la covariance (qui demande l'estimation préalable de l'espérance mathématique), mais également parce qu'elle permet d'accommoder des situations où $Var(Z(\mathbf{x}))$ n'est pas définie.

Le variogramme théorique est défini comme :

$$\gamma(\mathbf{h}) = \frac{1}{2} Var[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})] = \frac{1}{2} E[(Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h}))^2] \quad (\text{B.16})$$

car le processus est stationnaire et donc $E[(Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h}))] = 0$.

Cette fonction, habituellement croissante en fonction de \mathbf{h} , synthétise beaucoup d'information concernant le comportement conjoint des variables aléatoires. Si la valeur du variogramme converge vers une valeur limite, appelée pallier c , cela signifie qu'il y a une distance \mathbf{h} à partir de laquelle $Z(\mathbf{x})$ et $Z(\mathbf{x} + \mathbf{h})$ sont non corrélés. On définit alors [17] :

- Portée a : il s'agit de la distance à partir de laquelle deux observations ne se ressemblent plus du tout en moyenne, elles ne sont plus liées (covariance nulle) linéairement :
 - Si $\mathbf{h} > a$, il n'y a pas de dépendance entre les points séparés de la distance \mathbf{h}
 - Si $\mathbf{h} < a$, il y a dépendance entre les points et à cette distance, la valeur du variogramme correspond à la variance de la variable aléatoire.
- Pallier c : c'est la variance de la variable aléatoire $Var(Z(\mathbf{x}))$: $c = C_0 + C$
- Effet de pépité C_0 : il s'agit de variations à courte échelle, erreur de localisation, erreurs d'analyse et précision analytique. La pépité traduit la discontinuité du processus à très courtes distances : $\lim_{h \rightarrow 0} \gamma(h) = C_0$ et $\gamma(0) = 0$.
- Longueur de corrélation θ : Il s'agit d'un paramètre d'échelle dans la définition du variogramme, celui-ci est étroitement lié à la portée a .

Une illustration est présentée sur la Figure B.4.

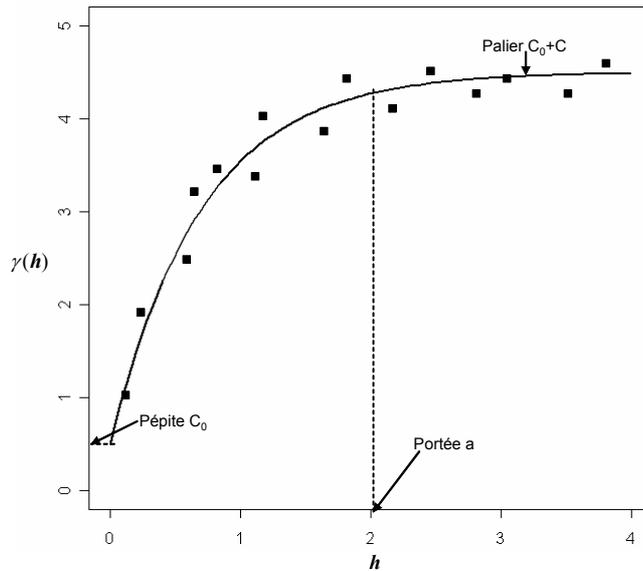


Figure B.4 : Variogramme expérimental : Portée, Pallier et Pépite

Lorsque le variogramme montre un palier, alors on peut facilement établir un lien entre la valeur du variogramme pour la distance h et la covariance pour deux observations séparées de la distance h :

$$\begin{aligned}
 \gamma(\mathbf{h}) &= \frac{1}{2} \text{Var}[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})] \\
 &= \frac{1}{2} [\text{Var}[Z(\mathbf{x})] + \text{Var}[Z(\mathbf{x} + \mathbf{h})] - 2\text{Cov}(Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h}))] \\
 &= \sigma^2 - \text{Cov}(Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})) \\
 &= \sigma^2 - \text{Cov}(\mathbf{h})
 \end{aligned}
 \tag{B.17}$$

Cette relation est très importante et elle est continuellement utilisée en géostatistique. Graphiquement, cette relation peut être représentée par le schéma suivant :

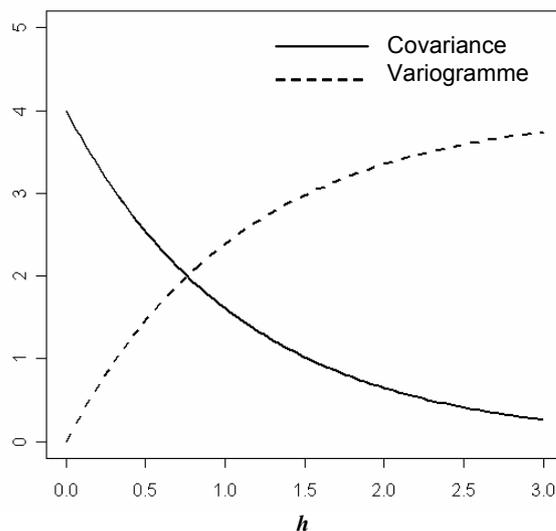


Figure B.5 : Représentation graphique d'un variogramme et de la covariance associée

Parfois, les variogrammes ne montrent pas de pallier, c'est-à-dire que la valeur du variogramme peut augmenter indéfiniment si la variabilité du phénomène n'a pas de limite pour de grandes distances. Dans ce cas, la covariance et la variance n'existent pas.

Le variogramme anisotrope

Quand le variogramme ne varie pas suivant les directions, il est dit isotrope. Il est alors une fonction du module du vecteur \mathbf{h} , représentant la distance entre les points :

$$\gamma(\mathbf{h}) = \gamma(\|\mathbf{h}\|_2) = \gamma\left(\sqrt{\sum_{i=1}^n h_i^2}\right)$$

Pour les modèles isotropes, chaque paire de points séparée de la même distance a la même covariance, peu importe la direction. Par exemple, sur la Figure B.6 (a), l'origine à la même corrélation avec tous les points du cercle.

Dans le cas opposé, le variogramme est dit anisotrope (b). Les deux cas typiques d'anisotropie sont :

- Géométrique : le pallier c est constant, mais la portée a varie avec la direction
- Zonale : le pallier c varie.

Nous parlerons ici du cas le plus simple qui est l'anisotropie géométrique.

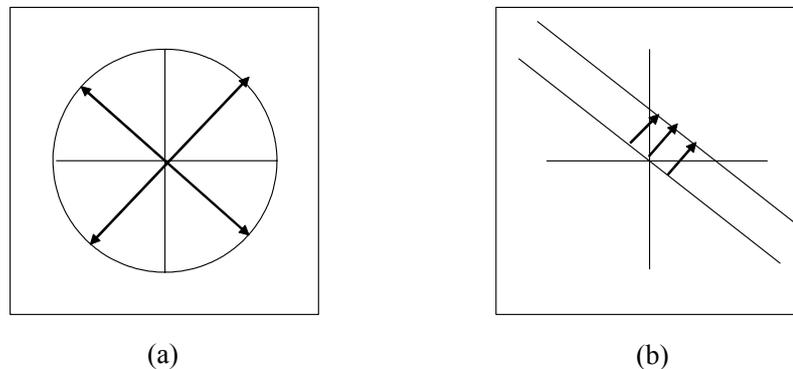


Figure B.6 : Isotropie (a) - Anisotropie (b)

Par définition, un variogramme dans \mathbb{R}^n est dit anisotrope (géométrique) s'il peut s'écrire de la forme :

$$\gamma(\mathbf{h}) = \gamma_0(\|\mathbf{A}\mathbf{h}\|)$$

où :

- γ_0 est un variogramme isotrope
- \mathbf{A} la matrice de transformation de l'espace initial à l'espace isotrope.

Dans le cas où il n'y a pas de rotation d'axe, i.e. les axes principaux correspondent aux paramètres, nous avons, pour la distance euclidienne : $\gamma(\mathbf{h}) = \gamma_0 \left(\sqrt{\sum_{i=1}^n \frac{h_i^2}{\theta_i^2}} \right)$ avec θ_i un paramètre d'échelle.

L'anisotropie géométrique est représentée sur la Figure B.7, il s'agit ici d'un variogramme gaussien. Pour cet exemple, dans la direction N-S, la portée est plus grande que dans la direction O-E.

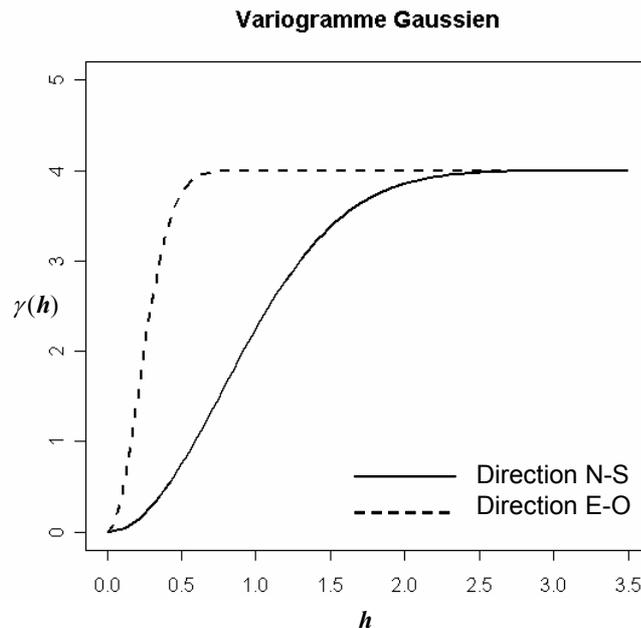


Figure B.7 : Variogramme anisotrope

La construction du variogramme passe par la construction d'un variogramme empirique, à partir duquel un modèle est ajusté : le variogramme théorique. Étant donné que le modèle de variogramme doit assurer des variances positives quelque soit la combinaison linéaire du processus Z utilisée et qu'il doit être stationnaire, il existe un certain nombre de variogramme type vérifiant ces deux propriétés.

B.4.3 Exemples de variogrammes existants

Dans cette partie, nous représentons les variogrammes (ou covariances) les plus utilisés en géostatistique et plus particulièrement en ingénierie de réservoir.

A titre d'exemple, nous citons ici les variogrammes les plus courants, valides dans \mathbb{R}^n . Pour les variogrammes associés à une covariance, nous donnons la forme analytique de la covariance, le variogramme étant déduit de la covariance par la relation : $\gamma(\mathbf{h}) = C(0) - C(\mathbf{h})$. Cette liste est évidemment loin d'être exhaustive.

Les fonctions de covariances que nous donnons ici sont données de manière isotrope, c'est-à-dire que $Cov(h) = Cov(\|\mathbf{h}\|)$.

Covariance sphérique généralisée :

La covariance sphérique généralisée dans \mathbb{R}^n s'écrit de la forme :

$$Cov_n(h) = \begin{cases} \theta^n v_{n-1} \int_{h/\theta}^1 (1-u^2)^{(n-1)/2} du & \text{si } h \leq \theta \\ 0 & \text{sinon} \end{cases}$$

Dans cette expression, v_n est le volume d'une boule de diamètre unité dans \mathbb{R}^n , cette valeur peut être déduite du volume V_n d'une boule de rayon unité dont l'expression est :

$$v_n = \frac{V_n}{2^n} = \frac{\pi^{n/2}}{2^{n-1} n \Gamma(n/2)}$$

Ce modèle est très utilisé en géologie pour $n \leq 3$, nous retrouvons le modèle sphérique très connu valide en \mathbb{R}^3 :

$$Cov_3(h) = \begin{cases} \sigma^2 & \text{si } h = 0 \\ \sigma^2 \left[1 - 1.5 \frac{h}{\theta} + 0.5 \left(\frac{h}{\theta} \right)^3 \right] & \text{si } h \leq \theta \\ 0 & \text{sinon} \end{cases}$$

Ce variogramme a un comportement linéaire à l'origine et atteint son pallier en $h = \theta$. Le paramètre d'échelle θ du variogramme sphérique coïncide avec la portée a , c'est-à-dire que nous avons $\theta = a$.

Covariance exponentielle :

La fonction de covariance exponentielle associée au paramètre $\theta > 0$ est définie par :

$$Cov(h) = \sigma^2 \exp \left\{ -\frac{h}{\theta} \right\}$$

Ce variogramme atteint son pallier c asymptotiquement, quand $h \rightarrow \infty$. Sa portée est environ $a = 3\theta$.

Covariance gaussienne :

La fonction de covariance préconisée la plupart du temps dans le contexte des expériences simulées, est la fonction de covariance gaussienne.

$$Cov(h) = \sigma^2 \exp \left\{ -\left(\frac{h}{\theta} \right)^2 \right\}$$

La fonction de covariance gaussienne est continue et différentiable à l'origine. Ce variogramme est utilisé dans la plupart des cas, de façon à avoir un processus ayant de bonnes qualités (continu et infiniment différentiable). De plus, la fonction de covariance gaussienne donne des réalisations plus lisses qu'une covariance exponentielle. La portée est environ $a = \theta\sqrt{3}$.

Variogramme puissance :

Enfin, nous présentons le variogramme puissance. Ce variogramme est non-stationnaire, c'est un modèle sans pallier. Nous avons $\lim_{x \rightarrow \infty} \gamma(h) = +\infty$. De plus, la variance, la covariance et la moyenne ne sont pas définies, c'est pour cette raison que nous présentons le variogramme. Il a pour expression :

$$\gamma(h) = h^a \text{ avec } 0 < a < 2$$

L'avantage de ce type de variogramme est qu'il est invariant par changement d'échelle. En effet, seule l'estimation du rapport d'anisotropie entre les pentes est nécessaire [17]. Une fois ce rapport défini, il est possible de multiplier les pentes par une constante.

Toutefois, ce variogramme n'est pas utilisable lors d'un krigeage simple, car il est non-stationnaire et n'a pas de portée. Il est généralement utilisé lors du krigeage ordinaire.

Par convention, une illustration des différents types de variogramme est représentée sur la Figure B.8. Nous présentons les variogrammes et pas les covariances.

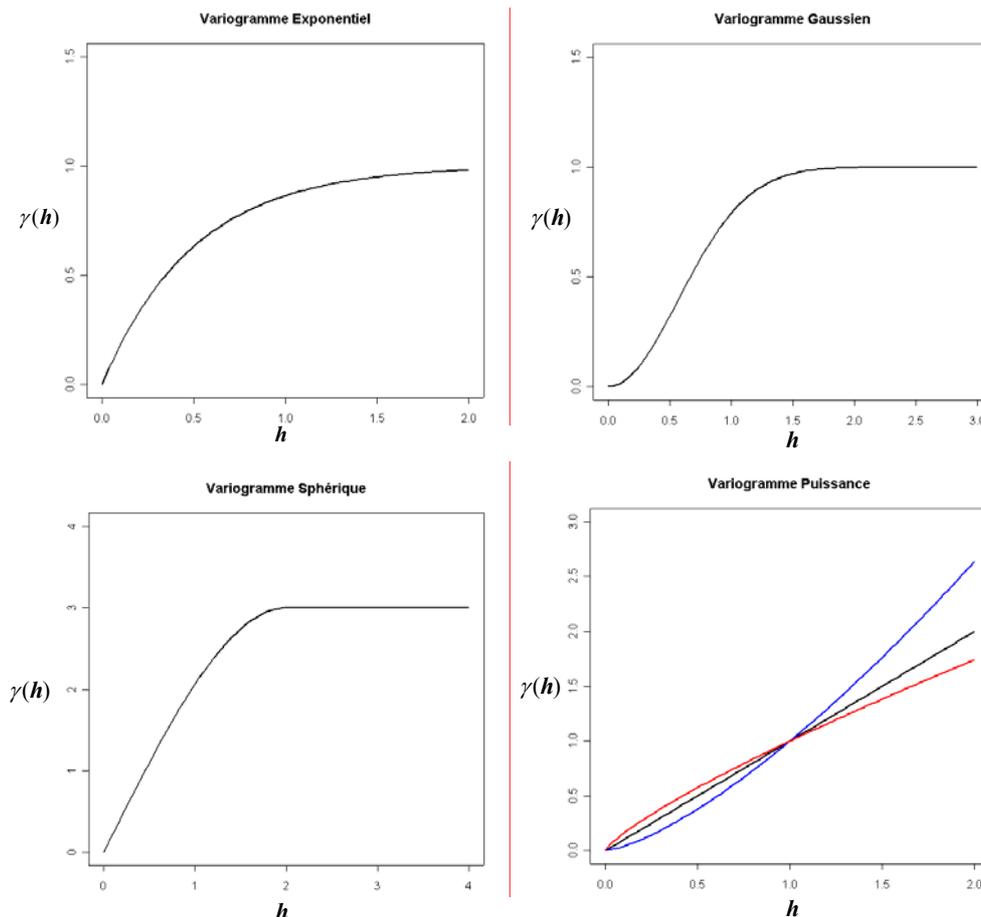


Figure B.8 : Représentation graphique des différents variogrammes

Généralement, si le nombre de données est important, les variogrammes, ainsi que leurs paramètres sont directement estimés à partir du variogramme empirique. Toutefois, il est possible, particulièrement en ingénierie de réservoir où l'acquisition des données est

extrêmement coûteuse, de ne pas pouvoir disposer d'un variogramme empirique fiable. Dans ce cas, il faut définir une structure de variogramme (exponentiel, gaussien, etc.) et estimer ensuite les paramètres. Nous présentons ici différents moyens d'estimer la longueur de corrélation et la variance σ^2 .

B.4.4 Estimation de la longueur de corrélation et σ^2 dans le cas d'un variogramme gaussien

Tous les estimateurs de la fonction de corrélation apparus dans la bibliographie supposent que $R(\cdot) = R(\cdot/\psi)$ où ψ est un vecteur de paramètres, précisant la structure de la fonction de covariance et la nature de la portée [57].

Par exemple, considérons la fonction de covariance exponentielle :

$$\text{Cov}(\mathbf{h}/\psi) = \sigma^2 \exp \left\{ -\sum_{j=1}^d |h_j / \theta_j|^p \right\}$$

où $\psi = (\theta_1, \dots, \theta_d, p)$ est un vecteur à $d+1$ composantes représentant d'une part, les paramètres d'échelle (longueurs de corrélation) et d'autre part, la puissance associée. Dans ce cas, la matrice de corrélation $\mathbf{R} = \mathbf{R}(\psi)$ ne dépend que de ψ tout comme le vecteur des corrélations $\mathbf{r}_0 = \mathbf{r}_0(\psi)$.

Nous allons proposer ici trois méthodes possibles pour l'estimation du paramètre ψ qui amènent chacune à un estimateur différent. Toutes les méthodes, excepté la dernière, supposent que les données suivent une distribution gaussienne, c'est-à-dire que nous avons :

$$[\mathbf{Y} | \boldsymbol{\beta}, \sigma^2, \psi] \sim N_{\eta}[\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{R}]$$

Dans ce paragraphe, nous allons principalement nous concentrer sur l'estimation des paramètres de corrélation ψ plutôt que sur celle de la variance du processus σ^2 . Ceci est justifié par le fait que le prédicteur \hat{Y}_0 ne dépend que de ψ et pas de σ^2 . En effet, l'estimateur de krigeage peut s'utiliser indépendamment avec la fonction de covariance ou de corrélation du processus, car le paramètre σ^2 disparaît quand on forme le produit matriciel ${}^t \mathbf{r}_0 \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$.

- **Estimation par Maximum de Vraisemblance (Maximum Likelihood)**

Le choix le plus populaire pour un estimateur de ψ est certainement celui du maximum de vraisemblance (MLE) utilisant l'hypothèse d'une loi normale multivariée [44].

Dans ce cas, la fonction de vraisemblance est définie par :

$$L(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\psi}) = (2\pi)^{-n/2} \sigma^{-n} (\det \mathbf{R})^{1/2} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right] \quad (\text{B.18})$$

On utilise le logarithme de cette fonction, la log-vraisemblance, qui est à une constante près :

$$\ell(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\psi}) = -\frac{1}{2} \left[n \log(\sigma^2) + \log(\det(\mathbf{R})) + {}^t (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) / \sigma^2 \right] \quad (\text{B.19})$$

où $\det(\mathbf{R})$ représente le déterminant de la matrice \mathbf{R} .

Étant donné un $\boldsymbol{\psi}$ fixé, l'estimateur de $\boldsymbol{\beta}$ par maximum de vraisemblance est son estimateur par moindres carrés généralisés (estimateur de Gauss-Markov), c'est-à-dire :

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\boldsymbol{\psi}) = ({}^t \mathbf{X} \mathbf{R}^{-1} \mathbf{X})^{-1} {}^t \mathbf{X} \mathbf{R}^{-1} \mathbf{Y} \quad (\text{B.20})$$

De plus, l'estimateur de σ^2 par maximum de vraisemblance est donné par :

$$\hat{\sigma}^2 = \hat{\sigma}^2(\boldsymbol{\psi}) = \frac{1}{n} {}^t (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (\text{B.21})$$

En substituant ces valeurs dans l'équation (B.19), nous obtenons que le maximum de l'équation (B.19) est le même que celui de la fonction :

$$\ell(\boldsymbol{\psi}) = \ell(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2, \boldsymbol{\psi}) = -\frac{1}{2} \left[n \log(\hat{\sigma}^2(\boldsymbol{\psi})) + \log(\det(\mathbf{R}(\boldsymbol{\psi}))) + n \right] \quad (\text{B.22})$$

qui ne dépend plus que du paramètre $\boldsymbol{\psi}$. Ainsi, l'estimateur $\hat{\boldsymbol{\psi}}$ est déterminé de manière à minimiser la fonction :

$$n \log(\hat{\sigma}^2(\boldsymbol{\psi})) + \log(\det(\mathbf{R}(\boldsymbol{\psi}))) \quad (\text{B.23})$$

où $\hat{\sigma}^2$ est défini par l'équation (B.21).

Le prédicteur correspondant à $\hat{\boldsymbol{\psi}}$ est appelé un MLE EBLUP estimateur de $Y(\mathbf{x}_0)$.

Existence et unicité des estimateurs $\boldsymbol{\beta}$ et σ^2 :

Comme nous l'avons montré précédemment, nous procédons par élimination substitution pour déterminer les paramètres $\boldsymbol{\beta}$ et σ^2 . L'objectif ici est de montrer l'existence et l'unicité des estimateurs par maximum de vraisemblance de $\boldsymbol{\beta}$ et σ^2 .

La première phase de cette procédure consiste à exprimer les paramètres $\boldsymbol{\beta}$ et σ^2 en fonction du seul paramètre $\boldsymbol{\psi}$, celui-ci étant fixé. Pour ce faire, nous considérons le logarithme du maximum de vraisemblance :

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2} \log(\det(\mathbf{R})) - \frac{1}{2\sigma^2} {}^t (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

La fonction ℓ possède un point stationnaire $(\hat{\beta}_\varphi, \hat{\sigma}_\varphi^2)$, caractérisé par le système d'équations :

$$\begin{cases} \frac{\partial \ell(\hat{\beta}_\varphi, \hat{\sigma}_\varphi^2)}{\partial \beta} = 0 \\ \frac{\partial \ell(\hat{\beta}_\varphi, \hat{\sigma}_\varphi^2)}{\partial \sigma^2} = 0 \end{cases} \quad (\text{B.24})$$

Le système (B.24) est équivalent à :

$$\begin{cases} -\frac{1}{\sigma^2} ({}^t \mathbf{X} \mathbf{R}^{-1} \mathbf{Y} - {}^t \mathbf{X} \mathbf{R}^{-1} \mathbf{X} \beta) = 0 \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} {}^t (\mathbf{Y} - \mathbf{X} \beta) \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X} \beta) = 0 \end{cases} \quad (\text{B.25})$$

soit encore :

$$\begin{cases} \hat{\beta}_\varphi = ({}^t \mathbf{X} \mathbf{R}^{-1} \mathbf{X})^{-1} {}^t \mathbf{X} \mathbf{R}^{-1} \mathbf{Y} \\ \hat{\sigma}_\varphi^2 = \frac{1}{n} {}^t (\mathbf{Y} - \mathbf{X} \hat{\beta}_\varphi) \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\beta}_\varphi) \end{cases} \quad (\text{B.26})$$

Par ailleurs, la matrice Hessienne de la fonction ℓ est semi-définie négative en $(\hat{\beta}_\varphi, \hat{\sigma}_\varphi^2)$:

$$\begin{cases} \frac{\partial^2 \ell}{\partial^2 \beta} (\hat{\beta}_\varphi, \hat{\sigma}_\varphi^2) = -\frac{n {}^t \mathbf{X} \mathbf{R}^{-1} \mathbf{X}}{{}^t (\mathbf{Y} - \mathbf{X} \hat{\beta}_\varphi) \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\beta}_\varphi)} \leq 0 \\ \frac{\partial^2 \ell}{\partial \beta \partial \sigma^2} (\hat{\beta}_\varphi, \hat{\sigma}_\varphi^2) = \frac{-{}^t \mathbf{X} \mathbf{R}^{-1} \mathbf{Y} + \hat{\beta}_\varphi {}^t \mathbf{X} \mathbf{R}^{-1} \mathbf{X}}{(\sigma_\varphi^2)^2} = 0 \\ \frac{\partial^2 \ell}{\partial \sigma^2 \partial \beta} (\hat{\beta}_\varphi, \hat{\sigma}_\varphi^2) = \frac{-{}^t \mathbf{X} \mathbf{R}^{-1} \mathbf{Y} + \hat{\beta}_\varphi {}^t \mathbf{X} \mathbf{R}^{-1} \mathbf{X}}{(\sigma_\varphi^2)^2} = 0 \\ \frac{\partial^2 \ell}{\partial (\sigma^2)^2} (\hat{\beta}_\varphi, \hat{\sigma}_\varphi^2) = -\frac{1}{2} \frac{{}^t (\mathbf{Y} - \beta_\varphi \mathbf{X}) \mathbf{R}^{-1} (\mathbf{Y} - \beta_\varphi \mathbf{X})}{(\sigma_\varphi^2)^3} \leq 0 \end{cases} \quad (\text{B.27})$$

Le point stationnaire $(\hat{\beta}_\varphi, \hat{\sigma}_\varphi^2)$ est donc bien un maximum de la fonction ℓ (ℓ est concave) et il est unique. Les paramètres $\hat{\beta}_\varphi$ et $\hat{\sigma}_\varphi^2$ sont donc les estimateurs du maximum de vraisemblance de β et σ^2 .

- **Estimation par Maximum de Vraisemblance Restreint**

Comme précédemment, supposons que $\mathbf{R}(\cdot)$ (et donc \mathbf{R} et r_θ) dépend d'un vecteur fini de paramètres. La méthode du maximum de vraisemblance restreinte (REML) des paramètres de la variance et covariance a été introduite par Patterson et Thompson en 1971 pour réduire le

biais des estimateurs du type maximum de vraisemblance. Certains auteurs utilisent le terme « Estimateurs du maximum de vraisemblance marginaux » pour le même concept. L'estimateur REML de $\boldsymbol{\psi}$ maximise la fonction de vraisemblance de combinaisons linéaires indépendantes (i.e. $\mathbf{C}\mathbf{Y}$) de \mathbf{Y} , ce qui implique que chaque combinaison linéaire est orthogonale à $\mathbf{X}\boldsymbol{\beta}$, le vecteur espérance de \mathbf{Y} . Supposons que \mathbf{X} est de rang plein p , cette méthode correspond alors à choisir une matrice \mathbf{C} de dimension $(n-p) \times n$ et de rang plein qui satisfasse l'équation $\mathbf{C}\mathbf{X} = \mathbf{0}$. L'estimateur REML de $\boldsymbol{\psi}$ est alors le maximum de la fonction de vraisemblance des « données transformées » :

$$\mathbf{W} \equiv \mathbf{C}\mathbf{Y} \sim \mathcal{N} \left[\mathbf{C}\mathbf{X}\boldsymbol{\beta} = \mathbf{0}, \sigma^2 \mathbf{C}\mathbf{R}(\boldsymbol{\psi})' \mathbf{C} \right] \quad (\text{B.28})$$

Notons que \mathbf{W} contient p observations de moins que \mathbf{Y} , mais que \mathbf{W} a l'avantage de ne contenir aucun des paramètres inconnus de $\boldsymbol{\beta}$. On peut montrer que l'estimateur REML de $\boldsymbol{\psi}$ est indépendant du choix des combinaisons linéaires utilisées pour construire \mathbf{W} , mais le nombre de colonnes de \mathbf{C} doit être maximal dans le sens où \mathbf{C} doit être de rang $n - p$. Avec quelques notions d'algèbre, on démontre facilement que l'estimateur $\tilde{\sigma}_{MVR}^2$ par maximum de vraisemblance restreint de σ^2 est :

$$\tilde{\sigma}_{MVR}^2 = \frac{n}{n-p} \hat{\sigma}^2 = \frac{1}{n-p} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})' \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (\text{B.29})$$

où $\hat{\sigma}^2$ est le MLE de σ^2 (cf. (B.21)) et finalement, l'estimateur REML de $\boldsymbol{\psi}$ est le minimum de la fonction :

$$(n-p) \log(\tilde{\sigma}_{MVR}^2(\boldsymbol{\psi})) + \log(\det(\mathbf{R}(\boldsymbol{\psi}))) \quad (\text{B.30})$$

Cet estimateur a l'avantage d'être un estimateur sans biais par rapport à l'estimateur par maximum de vraisemblance classique.

- **Validation Croisée**

La méthode de validation croisée est très populaire dans le choix des paramètres de modèles paramétriques.

Nous supposons une fois de plus que la fonction de corrélation est une fonction paramétrique avec $R(\cdot) = R(\cdot/\boldsymbol{\psi})$, et donc $\mathbf{R} = \mathbf{R}(\boldsymbol{\psi})$ et $\mathbf{r}_0 = \mathbf{r}_0(\boldsymbol{\psi})$. Pour $1 \leq i \leq n$, soit $\hat{Y}_{-i}(\boldsymbol{\psi})$, noté aussi classiquement $\hat{Y}_{(i)}(\boldsymbol{\psi})$, le prédicteur de $Y(\mathbf{x}_i)$ lorsque $\boldsymbol{\psi}$ est le vrai paramètre de corrélation basé sur toutes les données excepté le point $(\mathbf{x}_i, Y(\mathbf{x}_i))$. L'estimateur de validation croisée de $\boldsymbol{\psi}$ minimise l'erreur de prédiction moyenne au carré :

$$XV - PE(\boldsymbol{\psi}) = \sum_{i=1}^n \left(\hat{Y}_{-i}(\boldsymbol{\psi}) - y(\mathbf{x}_i) \right)^2$$

Des formes plus générales de la validation croisée sont proposées dans [57].

Références

- [1] Aziz, K. and Settari, A., *Petroleum Reservoir Simulation*, Applied Science Publishers, London, 1979.
- [2] Bishop, C. M., *Neural Networks for Pattern Recognition*, Oxford, 1995.
- [3] Brua J. Y., *Modélisation de réponses non-linéaires par plans d'expériences : étude bibliographique*, IFP, 2003.
- [4] Butler, N. A., Optimal and Orthogonal Latin hypercube designs for computer experiments, *Biometrika*, 88, 3, p 847-857, 2001.
- [5] Castellini, A., Landa, J. L., Kikani, J., *Practical Methods for Uncertainty Assessment of Flow Predictions for Reservoir with Significant History – A field case study*, ECMOR IX, Cannes, 30 August - 2 September, France, 2004.
- [6] Chauvet, P., *Aide-mémoire de Géostatistique linéaire (Fascicule 2)*, Ecole des Mines de Paris, 1991.
- [7] Chilès J.-P., Delfiner, P., *Geostatistics, Modeling Spatial Uncertainty*, Wiley Series in probability and statistics, 1999.
- [8] Christensen, R., *Plane Answers to Complex Questions, the Theory of Linear Models*, Editions Springer, 1996.
- [9] Chung, H.-S. and Alonso, J. J., *Using Gradients to Construct Cokriging Approximation Model for High-Dimensional Design Optimization Problems*, 40th AIAA Aerospace Science Meeting and exhibit, AIAA 2002-0317, Reno, January 2002.
- [10] Collombier D., *Plans d'expérience factoriels*, Springer, 1996.
- [11] Cook, R. D. et Weisberg, S., *Residuals and Influence in Regression*, Chapman and Hall, London 1982.
- [12] Corre, B., Thore, P., de Feraudy, V. and Vincent G., *Integrated Uncertainty Assessment For Project Evaluation and Risk Analysis*, SPE 65205, SPE European Petroleum Conference, Paris, France, oct. 24-25, 2000.
- [13] Cressie, N., *Statistics for Spatial Data*, Wiley, New-York 1991.
- [14] Currin, C., Mitchell, T. J., Morris, M. D. and Ylvisaker, D., Bayesian prediction of deterministic functions, with an application to the design and analysis of computer experiments, *Journal of the American Statistical Association* 86, 953-963, 1991.
- [15] Damslet E., Hage, A. and Volden, R., *Maximum Information at Minimum Cost! A North Field Development Study With an Experimental Design*, JPT (December 1992), p.1350

- [16] Dejean, J.-P. and Blanc, G., *Managing Uncertainties on Production Predictions Using Integrated Statistical Methods*, SPE 56696, SPE Annual Technical Conference and Exhibition, Houston, USA, oct. 3-6, 1999.
- [17] Deutsch, C. V., Journel, A. G., *GSLIB : Geostatistical Software Library and User's Guide*, Oxford University Press, 1992.
- [18] Droesbeke J.J, Fine J., Saporta G., *Plans d'expériences. Application à l'entreprise*, Technip, 1997.
- [19] Draper, N. R., Smith, M., *Applied Regression Analysis*, Wiley, N-Y, 3rd Edition, 1998.
- [20] Egeland T., Hatlebakk, E., Holden, L. and Larsen, E.A., *Designing Better Decisions*, SPE 24275, SPE European Petroleum Conference, Stavanger, Norway, may.25-27 , 1992.
- [21] Fang, K. T., Lin, D. K. J., Winker, P. and Zhang, Y., Uniform Design : Theory and Application, *Technometrics*, Vol. 42, No. 3, 2000.
- [22] Fang, K. T., Li, R, Sudjiant, A., *Design and Modeling for Computer Experiments*, Chapman & Hall/CRC; London, 2005.
- [23] Friedmann, F., Chawathé, A. and Larue, D. K., *Assessing Uncertainty in Channelized Reservoirs Using Experimental Designs*, SPE 85117, SPE Annual Technical Conference and Exhibition, New Orleans, USA, 30 sept.-3 oct. , 2001.
- [24] Johnson, M. E., Moore, L. M., Ylvisaker, D., Minimax and Maximin distance designs, *Journal of Statistical Planning and Inference* 26, p 131-148, 1990.
- [25] Jourdan A., Zabalza-Mezghani I., Response surface designs for scenario management and uncertainty in reservoir production, *Mathematical Geology*, Vol.36 n°8, November 2004
- [26] Journel, A. G., Huijbregts, Ch. J., *Mining Geostatistics*, Academic Press, 1978.
- [27] Journel, A. G., *Geostatistics for Reservoir Characterization*, SPE Annual Conference and Exhibition, New Orleans, USA, sept. 23-26, 1990.
- [28] Kenny, Q. Y., Orthogonal Column Latin Hypercubes and Their Application in Computer Experiments, *Journal of the American Statistical Association* 93, No. 444, 1430-1439, 1998.
- [29] Koehler, J. R. and Owen, A., B., *Computer Experiments*, Handbook of Statistics, Vol. 13, p 261-308, 1996.
- [30] Landa, L. L. and Güyagüler, B., *A Methodology for History Matching and the Assessment of Uncertainties Associated with Flow Prediction*, SPE 84465, 2003.

- [31] Lechner J.P. and Zangl, G., *Treating Uncertainties in Reservoir Performance Prediction with Neural Networks*, SPE 94357, SPE Europec/EAGE Annual Conference, Madrid, Spain, June 13-16, 2005.
- [32] Manceau, E., Zabalza-Mezghani, I., and Feraille M., *Cougar_Opt software user Manual*, IFP, Version 2.0, 2002.
- [33] Manceau, E., Zabalza-Mezghani, I., and Roggero, F., *Use of Experimental Design to Make Decisions in an Uncertain Reservoir Environment - From Reservoir Uncertainties to Economic Risk Analysis*, OAPEC seminar, 26-28th June 2001, Rueil-Malmaison, France.
- [34] Matheron, G., Principles of geostatistics, *Economic Geology* 58, 1246-1266, 1963
- [35] McKay, M. D., Beckman, R. J. and Conover, W. J., A comparison of three methods for selecting values of input variables in the analysis of output from computer code, *Technometrics* 21, 239-245, 1979.
- [36] Mitchell, T. J. and Scott, D. S., A computer program for the design of group testing experiments. *Communications in Statistics - Theory and Methods* 16, 2943-2955, 1987
- [37] Montgomery, D. C., Peck, E. A., *Introduction to linear regression analysis, second edition*, J. Wiley, 1991
- [38] Morris, M. D. and Mitchell, T. J., Exploratory designs for computational experiments, *Journal of Statistical Planning and Inference* 43, 381-402. 1995.
- [39] Müller, W. G., *Collecting Spatial Data : Optimum Design of Experiments for Random Fields*, 2nd edition, Physica-Verlag 1998.
- [40] Nelder, J.A., Mead, R., A simplex Method for Optimization, *The computer Journal*, n°7, pp. 308, 1965.
- [41] Neter, J., Kutner M. H., Nachtsheim, C.J., Wasserman, W., *Applied Linear Models*, Irwin, Chicago, 4th edition (1996)
- [42] Owen, A. B; Orthogonal arrays for computer experiments, integration and visualization, *Statistica Sinica* 2, 439-452, 1992.
- [43] Park, J. S., Optimal Latin-Hypercube designs for computer experiments, *Journal of Statistical Planning and Inference* 39, 95-111, 1994.
- [44] Pawitan, Y., *In all likelihood : statistical modelling and inference using likelihood*, Oxford University Press, 2001
- [45] Peaceman, D.-W., *Fundamentals of Numerical Reservoir Simulation*, Elsevier Science Publishers, New York, 1977.

- [46] Portella, R.C.M., Salomao, M.C., Blauth, M., and Duarte, R.L.B., *Uncertainty Quantification to Evaluate the Value of Information in a Deepwater Reservoir*, SPE 79707, SPE Reservoir Simulation Symposium, Houston, USA, feb. 3-5, 2003.
- [47] Rahon, D., et Roggero, F., *Utilisation des Réseaux de Neurones pour la prévision des incertitudes*, Rapport IFP n,° 44 880, Septembre 1998.
- [48] RamaRao, B.-S., La Venue, A.-M., De Marsily, G. and Marietta, M.-G., Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields : 1. Theory and computational experiments. *Water Resources Research*, 31(3) : 475-493, March 1995.
- [49] Rao, C. R., Mitra, S.K., *Generalized Inverse of Matrices and its Applications*, Wiley, N-Y, 1971.
- [50] Ripley, B., *Spatial statistics*, Wiley, New-York, 1981.
- [51] Robinson, G. K., That BLUP is a good thing: the estimation of random effects. With comments and a rejoinder by the author. *Statist. Sci.* **6** , 15-51, 1991.
- [52] Robinson, R., Neural Network Offer an Alternative to Traditional Regression, *Geobyte*, February 1991.
- [53] Sacks, J. and Schiller, S., Spatial designs, *Statistical Decision Theory and Related Topics IV* 2, 385-399 Springer, New-York, 1988.
- [54] Sacks, J., Schiller, S. and Welch, W., Design for computer Experiments, *Technometrics*, 31, 41-47, 1989.
- [55] Sacks, J., Welch, W., Mitchell, T., and Wynn, P., Design and Analysis of Computer Experiments (with discussion), *Statistical Science* 4, 409-435, 1989.
- [56] Saltelli, A., Chan, K., and Scott, E. M., *Sensitivity Analysis*, Wiley Series in Probability and Statistics, 2000.
- [57] Santer, T. J., Williams, B. J. and Notz, W. I., *The Design and Analysis of Computer Experiments*, Springer Series in Statistics, 2003.
- [58] Saporta G., *Probabilités. Analyse des données et statistiques*, Technip, 1990.
- [59] Scheidt C., Zabalza-Mezghani I., *Assessing Uncertainty and Optimizing Production schemes - Experimental Designs for non-linear production response modeling - An application to early water breakthrough prevention*, ECMOR IX, Cannes, 30 August - 2 September, France, 2004.
- [60] Scheidt C., Zabalza-Mezghani I., Collombier D., *Méthode de modélisation de la production d'un gisement pétrolier*, Brevet : registration n° 04/09.177, September 2004.

- [61] Scheidt C., Zabalza-Mezghani I., Feraille M., Collombier D., *Adaptive Evolutive Experimental Designs for Uncertainty Assessment*, IAMG, Toronto, Canada, 21-26 août 2005
- [62] Scheidt C., *Construction d'un modèle approché pour la quantification d'incertitudes : Krigeage ordinaire / Régression et Krigeage résiduel*, Rapport IFP n° ??, Octobre 2005.
- [63] Shewry, M. C. and Wynn, H. P., Maximum entropy sampling, *Journal of Applied Statistics* 14, 165-170. 1991.
- [64] Silvey S.D., *Optimal Design, An Introduction to the Theory for Parameter Estimation*, Chapman and Hall, 1980.
- [65] Simpson, T. W., Lin, D. K. J. and Chen W., Sampling Strategies for Computer Experiments : Design and Analysis, *International Journal of Reliability and Applications*, 2001.
- [66] Spendlay, W., Hext G. R. and Himsforth F.R., Sequential Application of Simplex Designs in Optimization and Evolutionary Operation, *Technometrics*, n°4, pp. 441, 1962.
- [67] Tang, B., Orthogonal Array-Based Latin Hypercubes, *Journal of the American Statistical Association*, 88, 1392-1397, 1993.
- [68] Thiria, S. , Lechevallier, Y., Gascuel, O., and Canu, S., *Statistique et Méthodes Neuronales*. Sciences SUP. Dunod, 1997.
- [69] Van Elk J. F., Guerrero, L., Vijayan, K. and Gupta, R., *Improved Uncertainty Management in Field Development Studies through the Application of the Experimental Design Method to the Multiple Realizations Approach*, SPE 64462, SPE Annual Technical Conference and Exhibition, Dallas, USA, oct. 1-4, 2000.
- [70] Venkataraman, R., *Application of the Method of Experimental Design to Quantify Uncertainty in Production Profiles*, SPE 59422, SPE Asia-Pacific Conference, Yokohama, Japan, apr. 25-26, 2000.
- [71] Walstrom, J. E., Mueller, T. D., McFarlane, *Evaluating Uncertainty in Engineering Calculations*, SPE 1928, SPE Annual Fall Meeting, Houston, USA, oct. 1-4, 1967.
- [72] White, C. D., Willis, B. J., Narayanan, K. and Dutton, S. P., *Identifying and Estimating Significant Geologic Parameters With Experimental Design*, SPE Journal, September 2001
- [73] Wu, C. F. J. and Hamada, M., *Experiments: Planning, Analysis, and Parameter Design Optimization*, J. Wiley, New-York, 2000.

- [74] Yeten, B., Castellini, A. , Guyaguler, B., Chen W. H., *A Comparison Study on Experimental Design an Response Surface Methodologies*, SPE 93347, SPE Reservoir Simulation Symposium, Houston, USA, 31 jan.- 2 feb., 2005.
- [75] Zabalza, I., Dejean, J.P., Collombier, D., Blanc, G., *Prediction and Density Estimation of a Horizontal Well Productivity Index using Generalized Linear Models*, paper B31 presented at the 1998 ECMOR VI, Peebles, September 8-11
- [76] Zabalza-Mezghani, I., *Analyse Statistique et Planification d'expérience en ingénierie de réservoir*, Thèse IFP& Université de Pau et des Pays de l'Adour, 24 Mai 2000.
- [77] Zabalza-Mezghani I., Scheidt C., Manceau E., *Quantification of Uncertainty for Improved Recovery Prediction*, WPC 1st youth forum, Pekin, 17-23 October, 2004.
- [78] Zabalza-Mezghani I. Scheidt C., Feraille M., Manceau E., *Quantification of Uncertainty for Improved Recovery Prediction*, Poster at the WPC 1st youth forum, Pekin, 17-23 October, 2004.
- [79] Zabalza-Mezghani I., Manceau E., Feraille M., Jourdan A., *Uncertainty Management: From Geological Scenarios to Production Scheme Optimization*, *Journal of Petroleum Science and Engineering special issue: Risk Analysis Applied to Petroleum Exploration and Production*, 2004.