



HAL
open science

Planification d'expériences numériques en phase exploratoire pour la simulation des phénomènes complexes

Jessica Franco

► **To cite this version:**

Jessica Franco. Planification d'expériences numériques en phase exploratoire pour la simulation des phénomènes complexes. Analyse fonctionnelle [math.FA]. Ecole Nationale Supérieure des Mines de Saint-Etienne, 2008. Français. NNT : 2008EMSE0024 . tel-00803107

HAL Id: tel-00803107

<https://theses.hal.science/tel-00803107>

Submitted on 21 Mar 2013

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.



N° d'ordre : 489 MA

THESE
présentée par

Jessica FRANCO

Pour obtenir le grade de Docteur
de l'Ecole Nationale Supérieure des Mines de Saint-Etienne

Spécialité : Mathématiques Appliquées

*Planification d'expériences numériques en phase exploratoire
pour la simulation des phénomènes complexes*

Soutenue publiquement à Saint-Etienne le 10 septembre 2008

Membres du jury

Président

Bernard Guy Professeur, Ecole Nationale Supérieure des Mines de Saint-Etienne

Rapporteurs

Luc Pronzato Directeur de Recherche CNRS, Université de Nice Antipolis

Michelle Sergent Professeur, Université Paul Cézanne

Examineurs

Xavier Bay Docteur, Ecole Nationale Supérieure des Mines de Saint-Etienne

Bernard Corre Ingénieur senior, Total

Astrid Jourdan Docteur, Ecole Internationale des Sciences du Traitement de l'Information

Directeurs de thèse

Laurent Carraro Professeur, Ecole Nationale Supérieure des Mines de Saint-Etienne

Georges Oppenheim Professeur, Université d'Orsay

■ Spécialités doctorales :

SCIENCES ET GENIE DES MATERIAUX
MECANIQUE ET INGENIERIE
GENIE DES PROCEDES
SCIENCES DE LA TERRE
SCIENCES ET GENIE DE L'ENVIRONNEMENT
MATHEMATIQUES APPLIQUEES
INFORMATIQUE
IMAGE, VISION, SIGNAL
GENIE INDUSTRIEL
MICROELECTRONIQUE

Responsables :

J. DRIVER Directeur de recherche – Centre SMS
A. VAUTRIN Professeur – Centre SMS
G. THOMAS Professeur – Centre SPIN
B. GUY Maître de recherche – Centre SPIN
J. BOURGOIS Professeur – Centre SITE
E. TOUBOUL Ingénieur – Centre G2I
O. BOISSIER Professeur – Centre G2I
JC. PINOLI Professeur – Centre CIS
P. BURLAT Professeur – Centre G2I
Ph. COLLOT Professeur – Centre CMP

■ Enseignants-chercheurs et chercheurs autorisés à diriger des thèses de doctorat (titulaires d'un doctorat d'Etat ou d'une HDR)

AVRIL	Stéphane	MAMécanique & Ingénierie	CIS
BATTON-HUBERT	Mireille	MASciences & Génie de l'Environnement	SITE
BENABEN	Patrick	PR 2 Sciences & Génie des Matériaux	CMP
BERNACHE-ASSOLANT	Didier	PR 1 Génie des Procédés	CIS
BIGOT	Jean-Pierre	MRGénie des Procédés	SPIN
BILAL	Essaïd	DR Sciences de la Terre	SPIN
BOISSIER	Olivier	PR 2 Informatique	G2I
BOUCHER	Xavier	MAGénie Industriel	G2I
BOUDAREL	Marie-Reine	MASciences de l'inform. & com.	DF
BOURGOIS	Jacques	PR 1 Sciences & Génie de l'Environnement	SITE
BRODHAG	Christian	MRSciences & Génie de l'Environnement	SITE
BURLAT	Patrick	PR 2 Génie industriel	G2I
CARRARO	Laurent	PR 1 Mathématiques Appliquées	G2I
COLLOT	Philippe	PR 1 Microélectronique	CMP
COURNIL	Michel	PR 1 Génie des Procédés	SPIN
DAUZERE-PERES	Stéphane	PR 1 Génie industriel	CMP
DARRIEULAT	Michel	ICM Sciences & Génie des Matériaux	SMS
DECHOMETS	Roland	PR 1 Sciences & Génie de l'Environnement	SITE
DESRAYAUD	Christophe	MAMécanique & Ingénierie	SMS
DELAFOSSE	David	PR 1 Sciences & Génie des Matériaux	SMS
DOLGUI	Alexandre	PR 1 Génie Industriel	G2I
DRAPIER	Sylvain	PR 2 Mécanique & Ingénierie	CIS
DRIVER	Julian	DR Sciences & Génie des Matériaux	SMS
FOREST	Bernard	PR 1 Sciences & Génie des Matériaux	CIS
FORMISYN	Pascal	PR 1 Sciences & Génie de l'Environnement	SITE
FORTUNIER	Roland	PR 1 Sciences & Génie des Matériaux	CMP
FRACZKIEWICZ	Anna	MRSciences & Génie des Matériaux	SMS
GARCIA	Daniel	CR Génie des Procédés	SPIN
GIRARDOT	Jean-Jacques	MRInformatique	G2I
GOEURIOT	Dominique	MRSciences & Génie des Matériaux	SMS
GOEURIOT	Patrice	MRSciences & Génie des Matériaux	SMS
GRAILLOT	Didier	DR Sciences & Génie de l'Environnement	SITE
GROSSEAU	Philippe	MRGénie des Procédés	SPIN
GRUY	Frédéric	MRGénie des Procédés	SPIN
GUILHOT	Bernard	DR Génie des Procédés	CIS
GUY	Bernard	MRSciences de la Terre	SPIN
GUYONNET	René	DR Génie des Procédés	SPIN
HERRI	Jean-Michel	PR 2 Génie des Procédés	SPIN
KLÖCKER	Helmut	MRSciences & Génie des Matériaux	SMS
LAFORST	Valérie	CR Sciences & Génie de l'Environnement	SITE
LI	Jean-Michel	EC (CCI MP) Microélectronique	CMP
LONDICHE	Henry	MRSciences & Génie de l'Environnement	SITE
MOLIMARD	Jérôme	MASciences & Génie des Matériaux	SMS
MONTHEILLET	Frank	DR 1 CNRS Sciences & Génie des Matériaux	SMS
PERIER-CAMBY	Laurent	PR 1 Génie des Procédés	SPIN
PIJOLAT	Christophe	PR 1 Génie des Procédés	SPIN
PIJOLAT	Michèle	PR 1 Génie des Procédés	SPIN
PINOLI	Jean-Charles	PR 1 Image, Vision, Signal	CIS
STOLARZ	Jacques	CR Sciences & Génie des Matériaux	SMS
SZAFNICKI	Konrad	CR Sciences de la Terre	SITE
THOMAS	Gérard	PR 1 Génie des Procédés	SPIN
VALDIVIESO	François	MASciences & Génie des Matériaux	SMS
VAUTRIN	Alain	PR 1 Mécanique & Ingénierie	SMS
VIRICELLE	Jean-Paul	MRGénie des procédés	SPIN
WOLSKI	Krzysztof	CR Sciences & Génie des Matériaux	SMS
XIE	Xiaolan	PR 1 Génie industriel	CIS

Glossaire :

PR 1 Professeur 1^{ère} catégorie
PR 2 Professeur 2^{ème} catégorie
MA(MDC)Maître assistant
DR (DR1) Directeur de recherche
Ing. Ingénieur
MR(DR2) Maître de recherche
CRChargé de recherche
EC Enseignant-chercheur
ICM Ingénieur en chef des mines

Centres :

SMS Sciences des Matériaux et des Structures
SPIN Sciences des Processus Industriels et Naturels
SITE Sciences Information et Technologies pour l'Environnement
G2I Génie Industriel et Informatique
CMP Centre de Microélectronique de Provence
CIS Centre Ingénierie et Santé

A tous un grand merci

Comme le veut la tradition, je vais tenter de satisfaire au difficile exercice de la page des remerciements, peut-être la tâche la plus ardue de ces années de thèse. Non qu'exprimer ma gratitude envers les personnes en qui j'ai trouvé un soutien soit contre ma nature, bien au contraire.

Je tiens à exprimer tout d'abord mes remerciements aux membres du jury, qui ont accepté d'évaluer mon travail de thèse.

Merci à mes directeurs de thèse Laurent Carraro, professeur à l'Ecole des Mines de Saint-Etienne et Georges Oppenheim, professeur à l'université d'Orsay, pour leur bienveillance à l'égard de mon travail.

Je tiens à remercier tout particulièrement Bernard Corre, qui, en tant que tuteur de thèse de Total, a su m'encadrer en gardant un œil critique et avisé ainsi que m'orienter et épanouir mon goût initial pour les applications en ingénierie réservoir. Sa disponibilité, son savoir et sa capacité à résoudre tous (sans exagération) les problèmes (de tout genre) ont quelque chose de fascinant. Plus qu'un encadrant, il a su m'aider aussi bien dans le travail que dans la vie lorsque j'en avais besoin. Il a toujours montré de l'intérêt pour mes travaux et répondu à mes sollicitations lorsque le besoin s'en faisait sentir. J'espère que cette thèse sera un remerciement suffisant au soutien et à la confiance sans cesse renouvelée. Aussi, je suis ravie de poursuivre ma vie professionnelle avec cet homme.

Merci à Xavier Bay, co-directeur et maître assistant à l'Ecole des Mines de Saint-Etienne, pour cette collaboration fructueuse qui m'a permis d'aborder quasi-sereinement les chaînes de Markov et les processus ponctuels !

Merci à Bernard Guy, co-directeur et professeur à l'Ecole des Mines de Saint-Etienne, pour l'intérêt qu'il a porté à mon travail.

Pour ses précieux conseils de tous ordres, sa disponibilité et sa confiance, je remercie tout particulièrement Astrid Jourdan, docteur en poste à l'EISTI. Son expérience et ses grandes compétences ont permis l'accomplissement de ce travail. Elle fût également une relectrice d'exception ! Qu'elle trouve ici les marques de ma reconnaissance, de mon respect et de mon amitié.

Enfin, je remercie Michelle Sergent, professeur à l'Université Paul-Cézanne et Luc Pronzato, directeur de recherche CNRS à l'Université de Nice Antipolis, de m'avoir fait l'honneur d'être les rapporteurs de cette thèse. J'éprouve un profond respect pour leur travail et leur parcours. Le regard critique, juste et avisé qu'ils ont porté sur mes travaux ne peut que m'encourager à être encore plus perspicace dans mon travail.

Merci au groupe de travail DICE pour sa sollicitude et la confiance envers mes travaux et plus particulièrement l'ONERA pour sa collaboration avec une partie de mon travail. Merci à Total d'avoir financé ma thèse en contrat CIFRE et de me laisser poursuivre ma vie professionnelle avec eux !

Je tiens à remercier l'ensemble de l'équipe 3MI et plus particulièrement Eric Touboul pour son incroyable sympathie et sa bonne humeur constante.

Ce n'est pas tout et j'y tiens ! Merci à Delphine Dupuy, collaboratrice et amie exceptionnelle, pour évidemment une liste de choses innombrables : l'aide apportée à mon travail notamment d'avoir été une relectrice impitoyable de mes œuvres diverses et variées, sa sérénité lorsque je faisais des boulettes pour ses applications, les soirées footing (même si celles-ci n'ont pas été aussi nombreuses que souhaitées probablement à cause de notre enthousiasme à cela et de notre difficulté à parler en courant...), les restaurants (pizzerias bien souvent) avec une profonde étude des desserts au chocolat, son accent stéphanois et les longs

moments passés au téléphone à discuter et à me remonter le moral dans les moments les plus durs.

Difficile en quelques mots de lui dire toute ma reconnaissance et mon estime.

Merci « aux gens de G&I » pour cette ambiance ineffable qui vous donne envie d'y manger (donc d'y boire) voire d'y dormir et leur bonne humeur. Ce sont toutes ces personnes, avec des qualités humaines incroyables, qui m'ont donné envie d'incorporer le métier d'ingénieur réservoir. Un merci particulier à quelques-unes d'entre-elles : Jerôme pour être un chef d'exception, Elodie et Céline pour leur grande amitié, Nicolas pour ses bourdes qui nous font rire, Emmanuel F. pour son savoir. Je les remercie tous pour leurs précieux conseils et de m'avoir aidée dans les derniers moments (les plus difficiles) de la thèse. Je m'excuse également auprès d'Emmanuel F. et Elodie pour les avoir distraits, par multiples reprises, de leur travail ! Je n'oublierai certainement pas mon excellentissime co-bureau Emmanuel C. avec lequel j'ai pu avoir de grandes discussions !

J'en profite pour remercier mon ordinateur portable trombinoscope qui m'est resté fidèle malgré les 3 années passées à se faire bien souvent insulter et à se promener inutilement. Je n'oublie pas les potes de l'escalade qui m'ont divertie, distraite et décentrée de mon objectif principal que fût ma rédaction de thèse : Yann pour sa patience et sa capacité à me supporter (une voie de 12 mètres en râlant peut prendre du temps...), Hélène pour son encouragement lors de mes multiples tentatives de jetés et son humeur joyeuse, Mathieu (mon jumeau) pour ses gigantesques jetés qui me faisaient virevolter et les séances de torture, Laurent L. pour son caractère béarnais exceptionnel, Laurent DLV pour être mon anti-moustique et ses « sec sec sec sec », Pierre-Xavier pour ses « dynamique », et les anciens (au sens expérimentés bien sûr !) Stéphane, Bernard, ... Et bien évidemment, je les remercie tous pour leur amitié !

Commentaire : Attention, avant de lire le paragraphe suivant il est important de prendre l'accent de Saint-Etienne !

Fouilla ! Je n'oublierai pas les « potes de Sainté », ceux qui m'ont écoutée dégoïser et maronner de nombreuses fois des histoires dignes des feux de l'amour. J'ai aimé faire la pampille avec eux à la mine ou à la colloc' à pitancher jusqu'à point d'heure ! Mouai, ben merci à Pandi Panda, Bafé et Lucienne rencontrés en développement personnel (ce qui est sûr c'est que l'on a jamais développé quoi que ce soit) pour m'avoir supportée pendant 2 ans, Rém, Tanguy, la morue et puis les autres aussi ! Je les remercie tous pour m'avoir fait passer des moments merveilleux !

Cela va de soi, je remercie évidemment ma mère pour son irremplaçable et inconditionnel soutien.

Je sais que j'oublie des gens. Mais un fait est certain. Bien que je demeure la principale actrice de ce travail de thèse de doctorat, je ne peux ignorer que la réalisation de celle-ci n'aurait jamais été possible sans la contribution de ces dizaines de personnes. Merci à toutes et à tous !

A la mémoire de mon père, Julio Franco.

TABLE DES MATIERES

INTRODUCTION	1
PARTIE I. COMPTE-RENDU DE LA LITTERATURE	16
CHAPITRE 1 PLANS D'EXPERIENCES NUMERIQUES USUELS	18
1.1. PLANS STANDARD	19
1.1.1. Plans factoriels complets	19
1.1.2. Plans factoriels fractionnaires	20
1.1.3. Plans composites	21
1.1.4. Plans de Box-Behnken	21
1.1.5. Plans de Doehlert ou réseaux uniformes	22
1.1.6. Conclusions	24
1.2. PLANS « MARGINAUX »	25
1.2.1. Hypercubes latins	25
1.2.2. Tableaux orthogonaux	28
1.2.3. Hypercubes latins basés sur les tableaux orthogonaux de force 2	32
1.3. SUITES A FAIBLE DISCREPANCE	35
1.3.1. Suites de Halton	37
1.3.2. Suites de Hammersley	40
1.3.3. Suites de Sobol'	41
1.3.4. Suites de Faure	43
1.3.5. Good Lattice Points	44
1.3.6. Suites et réseaux	45
1.4. RECAPITULATIF	51
CHAPITRE 2 CRITÈRES ET PLANS OPTIMAUX	52
2.1. CRITERES D'UNIFORMITE BASES SUR LA DISCREPANCE ET PLANS A FAIBLE DISCREPANCE	53
2.1.1. La discrèpance	53
2.1.2. Approches calculatoires de la discrèpance en norme infinie	55
2.1.3. Discrèpance en norme 2	58
2.1.4. Exemples	62
2.1.5. Plans à faible discrèpance	63
2.2. CRITERES DE DISTANCE ET PLANS OPTIMAUX	67
2.2.1. Mesure de recouvrement	67
2.2.2. Le rapport de distances	68
2.2.3. Distances Maximin et Minimax	68
2.2.4. Exemples	70
2.2.5. Algorithmes d'amélioration de critères de distance	71
2.3. CRITERE D'ENTROPIE ET PLANS A ENTROPIE MAXIMALE	76
2.3.6. Définition de l'entropie	76
2.3.7. Plans à entropie maximale	79
2.3.8. Propriétés des plans à entropie maximale	82
PARTIE II. CONSTRUCTION DE NOUVEAUX PLANS D'EXPERIENCES	86
CHAPITRE 3 PLANS D'EXPERIENCES PAR METHODES MCMC	88
3.1. METHODES MCMC	90
3.2. CHAINES DE MARKOV	92

3.2.1.	Un cas simple dans un espace d'états discret	93
3.2.2.	Cas d'un espace d'états continu	100
3.3.	PROCESSUS PONCTUELS ET PLANIFICATION D'EXPERIENCES	103
3.3.1.	Le processus ponctuel de Strauss	104
3.3.2.	Etude des plans de Strauss	106
3.3.3.	Extensions : processus ponctuel de Gibbs	109
3.3.4.	Ergodicité uniforme	114
3.4.	CONCLUSIONS	117
<u>CHAPITRE 4 VERS DES EXTENSIONS...</u>		118
4.1.	DISTANCE A DENSITE ET PLANS D'EXPERIENCES ASSOCIES	119
4.1.1.	Distance à densité	120
4.1.2.	Processus ponctuels et distance à densité	123
4.1.3.	Distance à densité gaussienne	124
4.2.	PLANS D'EXPERIENCES CONTRAINTS	127
4.2.4.	Approche <i>a posteriori</i>	127
4.2.5.	Approche <i>a priori</i> : Plans Step-by-Step (SbS)	128
4.3.	PLANS ITERATIFS EN PHASE EXPLORATOIRE : ZOOM ET SHIFT	131
4.3.1.	Plans itératifs : zoom	131
4.3.2.	Déplacement de zone de recherche	132
<u>PARTIE III. ANALYSE DES PLANS D'EXPERIENCES</u>		136
<u>CHAPITRE 5 UN NOUVEAU CRITERE STATISTIQUE EN FORME DE RADAR POUR LA MESURE DE L'UNIFORMITE DES PLANS D'EXPERIENCES ET DE LEUR PROJECTION SUR LES SOUS-ESPACES</u>		138
5.1.	LE RADAR D'UNIFORMITE	139
5.2.	APPLICATIONS DU RADAR D'UNIFORMITE	146
5.2.1.	Exemple 1 : détection de lacunes dans les suites de Halton en 15D à l'aide du radar 2D	146
5.2.2.	Exemple 2 : détection de lacunes dans les tableaux orthogonaux à l'aide du radar 3D	147
5.3.	UNE STATISTIQUE GLOBALE POUR LE RADAR 2D	150
5.3.1.	Exemple 3 - Vers une extension du radar d'uniformité	150
5.3.2.	Loi asymptotique de la statistique globale pour $n > 100$	151
5.4.	DISCUSSION	153
<u>CHAPITRE 6 ARBRES DE LONGUEUR MINIMALE : UNE NOUVELLE APPROCHE POUR DETERMINER LA QUALITE DES PLANS D'EXPERIENCES NUMERIQUES</u>		155
6.1.	METHODE D'ETUDE DE REPARTITION SPATIALE D'OBJETS	157
6.1.1.	Géométrie algorithmique	157
6.1.2.	Rappels sur la théorie des graphes	158
6.1.3.	Quelques mots sur les graphes de Gabriel et ceux des voisins relatifs	159
6.1.4.	Propriétés des Arbres de Longueur Minimale	160
6.2.	APPLICATION	161
6.2.1.	Mise en œuvre et interprétation	161
6.2.2.	Exemples d'application	162
6.3.	DISCUSSION	164
<u>CHAPITRE 7 ETUDE INTRINSEQUE DES PLANS</u>		165
7.1.	ETUDE INTRINSEQUE DES PLANS A L'AIDE DES CRITERES USUELS	166
7.1.1.	Plans à 30 points en dimension 3	167
7.1.2.	Plans à 100 points en dimension 10	169
7.2.	ETUDE DES PLANS AVEC LE CRITERE DES ALM	172
7.3.	DISCUSSION	176

PARTIE IV. APPLICATIONS DES PLANS D'EXPERIENCES	178
CHAPITRE 8 CAS EN DIMENSION 3 : COMPARAISON DES PLANS	180
8.1. PRESENTATION DE L'APPLICATION	182
8.1.1. Les données	182
8.1.2. Méthodologie	183
8.2. COMPARAISON DES PLANS D'EXPERIENCES	184
8.2.1. Les plans	184
8.2.2. Résultats de l'étude avec le modèle PolyMARS	185
8.2.3. Résultats de l'étude avec une régression de degré 2	189
8.2.4. Résultats de l'étude avec le krigeage ordinaire	191
8.2.5. Conclusions	193
8.3. QUALITE DE LA SURFACE DE REPONSE OBTENUE A PARTIR DE PLANS D'EXPERIENCES A 30 POINTS	194
8.3.1. Etude avec le modèle PolyMARS	195
8.3.2. Etude avec le modèle de régression quadratique	198
8.4. CONCLUSIONS DE L'ETUDE	199
CHAPITRE 9 CAS D'ECOLE : FONCTION DETERMINISTE EN DIMENSION 10	200
9.1. LES DONNEES	201
9.2. LES RESULTATS	202
9.3. CONCLUSIONS	204
CHAPITRE 10 APPLICATION A L'EXPLORATION PETROLIERE	205
10.1. PRÉSENTATION DU CAS D'ÉTUDE	206
10.1.1. Contexte de l'étude	206
10.1.2. Description des données	206
10.1.3. Schéma d'une étude	207
10.2. ETUDE DETAILLE D'UN JEU DE DONNEES	208
10.2.1. Etude préliminaire des données simulées	208
10.2.2. Construction des différents modèles et interprétation	210
10.2.3. Comparaison des modèles ajustés	218
10.3. ETUDE DE L'IMPACT DU PLAN D'EXPERIENCES	220
10.3.1. Les modèles additifs	221
10.3.2. Les modèles linéaires	222
10.3.3. MARS	224
10.3.4. PolyMARS	225
10.3.5. Krigeage ordinaire	226
10.3.6. Réseaux de neurones	227
10.3.7. Conclusions	228
10.4. ETUDE DE L'IMPACT DU NOMBRE DE POINTS	230
10.5. CONCLUSIONS	232
CONCLUSIONS ET PERSPECTIVES	235
ANNEXES	241
BIBLIOGRAPHIE	264

TABLE DES ILLUSTRATIONS

Figure 1. Schéma représentant les facteurs et les réponses	4
Figure 2. Un plan factoriel complet à 5 niveaux	19
Figure 3. Un plan de Doehlert à 45 points dans le carré unité avec son simplexe initial : le triangle équilatéral en rouge	22
Figure 4. Un échantillonnage par hypercube latin à 5 points en dimension 2	26
Figure 5. Un échantillonnage par hypercube latin à 5 points en dimension 2	27
Figure 6. Un hypercube latin à 20 points en dimension 2	27
Figure 7. Un plan généré par un tableau orthogonal $OA_1(25,5,5,2)$ dont les points sont centrés et projetés sur le sous-espace (X_1, X_2)	29
Figure 8. Un plan généré par un tableau orthogonal $OA_1(25, 5, 5, 2)$ dont les points sont randomisés et projetés sur le sous-espace (X_1, X_2)	30
Figure 9. Une distribution de 49 points issue d'un tableau orthogonal de force 2 linéaire en dimension 3	32
Figure 10. (De gauche à droite). Un hypercube latin, un hypercube latin de Tang et un tableau orthogonal de force 2 à 4 points en dimension 2.....	33
Figure 11. Un hypercube latin généré par un tableau orthogonal $OA_1(25,5,5,2)$ dont les points sont randomisés et projetés sur le sous-espace (X_1, X_2)	34
Figure 12. (De la gauche vers la droite) Les 50, 250 et 500 premiers points d'une suite de Halton en bases 2 et 3	38
Figure 13. Des mauvaises projections en 2D. A gauche : les 80 premiers points d'une suite de Halton en bases 11 et 13. A droite : les 100 premiers points d'une suite de Halton en bases 17 et 19	38
Figure 14. (De la gauche vers la droite) Les 50, 250 et 500 premiers points d'une suite de Hammersley en base 2	40
Figure 15. Les 50, 250 et 500 premiers points d'une suite de Sobol' en dimension 2	42
Figure 16. A gauche : les 80 premiers points d'une suite de Sobol' en dimension 6 en projection sur le sous-espace des dimensions 5 et 6. A droite : les 100 et 150 premiers points d'une suite de Sobol' en dimension 8 sur le sous-espace des dimensions 7 et 8	42
Figure 17. Les 50, 250 et 500 premiers points d'une suite de Faure en dimension 2.....	43
Figure 18. Les 80 premiers points d'une suite de Faure en dimension 6, les 100 et 200 premiers points d'une suite de Faure en dimension 8, projetés sur le carré unité défini par les deux dernières dimensions	44
Figure 19. Les 21 premiers termes d'un good lattice points en dimension 2, avec des générateurs différents	44
Figure 20. Une première caractérisation possible des pavés élémentaires pour 9 points	47
Figure 21. Une seconde caractérisation possible des pavés élémentaires pour 9 points	47
Figure 22. Une troisième caractérisation possible des pavés élémentaires pour 9 points	48
Figure 23. Les 50, 250 et 500 premiers points d'une suite de Niederreiter en dimension 2.....	49
Figure 24. Les 80 premiers points d'une suite de Niederreiter en dimension 6, les 100 et 200 premiers points d'une suite de Niederreiter en dimension 8, projetés sur le carré unité défini par les deux dernières dimensions	49
Figure 25. Les 80 premiers points d'une suite de Hammersley en dimension 2 avec un sous-ensemble J défini par x et x' pour le calcul de la discrédance extrême.....	54
Figure 26. Un sous-ensemble J pour le calcul de la discrédance à l'origine	54
Figure 27. Plan factoriel et visualisation de deux rectangles parallèles aux axes n'étant pas échantillonnés uniformément.....	55
Figure 28. L'écart relatif entre la valeur exacte et approchée de la discrédance en fonction du nombre d'orthants sur un plan pseudo aléatoire à 100 points en dimension 2	57
Figure 29. Le temps d'exécution du calcul approché de la discrédance en fonction de la dimension	57
Figure 30. Un sous-ensemble J pour le calcul de la discrédance centrée	60

Figure 31. Pour $\mathbf{x} = (0.7, 0.75)$, le volume total des 2 sous-ensembles J est 0.6 et la proportion de points totale est $49/80=0.6125$. La différence entre ces deux grandeurs est donc 0.0125	60
Figure 32. Pour $\mathbf{x} = (0.8, 0.35)$ et $\mathbf{x}' = (0.3, 0.8)$, le volume des 2 enveloppes est égal à 0.275, et la proportion de points totale est $21/80=0.26$. La différence obtenue entre ces deux grandeurs est donc 0.015	61
Figure 33. De la gauche vers la droite. Un plan aléatoire, une suite de Halton, et un plan factoriel complet à 49 points en dimension 2	62
Figure 34. A gauche un plan à faible discrédance et à droite une suite de Halton à 20 points en 2D	65
Figure 35. Représentation des projections sur les facteurs (X_7, X_8) d'un plan à faible discrédance sur la figure de gauche et d'une suite de Halton sur la figure de droite à 80 points en 8D	65
Figure 36. Un graphe représentant l'historique des échanges pour un plan	72
Figure 37 Un plan minimax à 25 points (cercle) généré à partir d'une grille régulière à 400 points et d'un premier plan minimax à 15 points (croix)	73
Figure 38. A gauche, un plan initial (aléatoire) de 50 points avec la première boule d'élimination. Au centre, le plan à 28 points obtenu avec la boule d'élimination de la figure de gauche. A droite, le plan à 8 points obtenu avec une boule d'élimination de rayon supérieur	74
Figure 39. A gauche, un plan de 20 points obtenu à partir d'une suite de Halton à 50 points. A droite, une suite de Halton à 20 points en dimension 2	75
Figure 40. Représentation d'un variogramme sphérique avec une portée égale à 2	82
Figure 41. Représentation d'un variogramme exponentiel avec une portée égale à 2	82
Figure 42. Représentation d'un variogramme gaussien avec une portée égale à 2	83
Figure 43. Représentation d'un variogramme puissance ($\alpha = 0.5$) avec une portée égale à 2	83
Figure 44. Un plan à entropie maximale à 40 points en 2D avec une portée de 0.7 obtenu à l'aide d'une fonction de corrélation exponentielle (respectivement sphérique) à gauche (respectivement à droite)	84
Figure 45. A gauche, un plan à entropie maximale à 20 points en 2D obtenu avec une portée de 0.1. A droite, un plan à entropie maximale à 20 points en 2D obtenu avec une portée de 0.9	84
Figure 46. A gauche, un plan initial, aléatoire, à 20 points en 2D. A droite, un plan à 20 points en 2D construit en maximisant l'entropie	85
Figure 47. Un plan anisotrope où l'on considère une portée de 0.2 dans la direction des abscisses	85
Figure 48. Représentation des éléments de E dans $[0,1]^2$	94
Figure 49. Représentation des transitions de E dans E	96
Figure 50. A gauche, un plan initial (aléatoire) de 20 points en dimension 2. A droite, le plan de Strauss obtenu à partir du plan de gauche avec la représentation des sphères de rayon R/2	106
Figure 51. A gauche, un plan de Strauss à 20 points en dimension 2 de paramètres R=0.1 et $\gamma=0.1$. A droite, un plan de Strauss de paramètres R=0.3 et $\gamma=0.1$	107
Figure 52. Un plan de Strauss à 50 points avec R=0.8 et $\gamma=0.1$	107
Figure 53. A gauche, un plan de Strauss de paramètres R=0.19 et $\gamma=0.1$. A droite, un plan de Strauss de paramètres R=0.19 et $\gamma=0.001$	108
Figure 54. Un plan de Strauss à 20 points en dimension 2 obtenu à partir d'une distribution initiale (en bleu) où les points sont concentrés au centre du domaine	109
Figure 55. Un plan de Strauss à 20 points en dimension 2 (R = 0.19, R1d = 0.0375, $\gamma = 0.1$, $\gamma_{ID} = 0.01$)	110
Figure 56. La fonction de potentiel puissance pour différents α et R=0.2	111
Figure 57. A gauche. Un plan de Strauss obtenu à partir d'un potentiel tout ou rien. A droite. Un plan de Strauss obtenu à partir d'un potentiel puissance $\alpha=0.5$	112
Figure 58. Un plan de Strauss-Gibbs combinant interactions globales et sur les marges avec un potentiel puissance de type puissance	113
Figure 59. Configuration la plus défavorable au mélange	116
Figure 60. Exemple de paires de point qui sont à la même distance h	119
Figure 61. Loi normale de moyenne nulle et d'écart-type 1 et distance $\delta(\mathbf{a}, \mathbf{b})$ associée	121
Figure 62. Loi normale centrée-réduite et représentation des boules B(0,1/4) et B(0.5,1/4)	122

Figure 63. Loi normale centrée réduite et représentation de la boule $B(1,1/4)$	122
Figure 64. Un plan de Strauss de distance à densité gaussienne isotrope ($R=0.28, \sigma=2$)	125
Figure 65. Un plan de Strauss de distance à densité gaussienne isotrope avec interactions entre les points.....	125
Figure 66. Un plan de Strauss de distance à densité gaussienne anisotrope.....	126
Figure 67. Un plan à distance à densité gaussienne avec $\sigma=1000$ et $R=0.4$	126
Figure 68. Un plan de Strauss généré dans $[0,1]^2$ puis on conserve les points.....	128
Figure 69. Une distribution de 40 points en dimension 2 obtenue à partir du point bleu et rouge	129
Figure 70. Un plan à entropie maximale à 20 points en 2D où l'on veut affiner la zone définie par le pavé.....	131
Figure 71. Représentation de l'échantillonnage obtenu dans la zone choisie en ayant considéré les simulations déjà effectuées.....	132
Figure 72. Plan initial avec la représentation de la nouvelle zone de recherche	133
Figure 73. Représentation de l'échantillonnage de la nouvelle zone de recherche avec en rouge les expériences initiales et en bleu les expériences itérées une à une.....	133
Figure 74. Représentation de l'échantillonnage dans la nouvelle zone de recherche où l'on a considéré les points de la zone et les points d'une « enveloppe » de cette zone	134
Figure 75. A gauche, les projections des points sur un axe L_a . A droite, l'histogramme des projections.....	139
Figure 76. La loi des projections pour un domaine cubique en dimension 2	140
Figure 77. A gauche : le plan (X_{14}, X_{15}) d'une suite de Halton ; à droite : la courbe obtenue à l'aide du radar ..	147
Figure 78. A gauche. Un tableau orthogonal linéaire à 49 points en dimension 3 de force 2. A droite. Les points du tableau orthogonal projetés sur (X_1, X_2)	148
Figure 79. A gauche, les p-valeurs en $-\log_{10}$ du test de Kolmogorov-Smirnov. A droite, les valeurs du test de Kolmogorov Smirnov avec la représentation en épingle du radar.....	149
Figure 80. De gauche à droite. Les fonctions de répartition (après transformation) des projetés sur L_1, L_2 et L_3	149
Figure 81. Les 100 premiers termes d'une suite de Halton en dimension 8 projetés sur (X_3, X_6)	150
Figure 82. Interprétation de la probabilité $P(A_S \cap A_T)$	152
Figure 83. Radar d'uniformité avec la statistique de Cramér-Von Mises pour les exemples 1 et 3.....	154
Figure 84. Illustration d'un arbre de longueur minimale	159
Figure 85. Représentation de la répartition des distributions dans le plan (μ, σ)	161
Figure 86. Représentation du critère de l'ALM pour les plans à 30 points en 3D où figurent les plans de divergence minimale	162
Figure 87. Représentation des boxplots sur les critères de discrédance en norme L_2 (classique et centrée) et de distance (MinDist et recouvrement) calculés sur les plans à 30 points en dimension 3	167
Figure 88. Représentation des boxplots sur les critères de discrédance en norme L_2 (classique et centrée) et de distance (Mindist et recouvrement) calculés sur les plans à 100 points en dimension 10.....	170
Figure 89. Représentations du critère lié aux ALM pour les plans étudiés en dimension 2, 3,5, 10 et 20. A gauche, les plans étudiés comptent 10 points par dimension. A droite, les plans étudiés comptent 20 points par dimension.....	174
Figure 90. Représentation du critère lié aux ALM sur les plans de Strauss	175
Figure 91. Représentation de la surface approchée de la fonction chameau obtenue à partir d'une grille et d'un plan de Strauss (points en rouge)	181
Figure 92. Représentation de la surface réelle	182
Figure 93. Evolution des critères en fonction du nombre de points (Strauss / plans standard)	185
Figure 94. Représentation des coupes de la surface approchée pour le plan de Doehlert à 67 points.....	186
Figure 95. Représentation des coupes de la surface approchée pour le plan de Doehlert à 67 points permuté... ..	186
Figure 96. Evolution des critères en fonction du nombre de points (Strauss).....	187
Figure 97. Evolution des critères en fonction du nombre de points (hypercubes latins et Strauss 1D).....	187
Figure 98. Evolution des critères en fonction du nombre de points (Strauss / aléatoires /Minimax)	188
Figure 99. Comparaison en coupe de la surface simulée et approchée par PolyMARS.....	189

Figure 100. Critères pour le modèle quadratique (Strauss / plans standard)	189
Figure 101. Critères pour le modèle quadratique (Strauss et autres SFD)	190
Figure 102. Krigeage ordinaire	191
Figure 103. Krigeage ordinaire avec les suites à faible discrédance et les plans de Strauss	192
Figure 104. A gauche, représentation du boxplot pour mesurer la qualité d'ajustement du modèle PolyMARS. A droite, représentation du critère R^2 moyen pour chaque type de plans	195
Figure 105. A gauche, représentation du boxplot pour mesurer la prédiction du modèle PolyMARS. A droite, représentation du critère $RMSE$ moyen pour chaque type de plans	196
Figure 106. Représentation des critères pour comparer quelques SFD avec les plans de Strauss de distance à densité gaussienne ($\sigma=1$, $\rho=0$, $R=0.2$).....	197
Figure 107. Boxplot pour mesurer la qualité d'ajustement (à gauche) et de prédiction (à droite)	198
Figure 108. Coupe de la fonction de Jin en fonction de x_1 et x_3 ($x_i = 3$, $i \neq 1, 3$).....	200
Figure 109. Boxplots des critères de qualité d'ajustement (à gauche) et de validation (à droite) des plans	202
Figure 110. Configuration initiale du réservoir pétrolier et lignes de courant obtenues (chaque couleur de la figure de droite correspond à la zone d'influence d'un puits injecteur).....	206
Figure 111. Domaines de variation pour la position des puits P2, P5, I2 et I3.....	207
Figure 112. Production cumulée d'huile pour les différentes configurations.....	208
Figure 113. Représentation des 20 meilleures configurations parmi les 80 initialement simulées ainsi que du cas de base.....	209
Figure 114. Fonction de répartition empirique (à gauche) et histogramme et lissage par noyau de la réponse (à droite).....	210
Figure 115. Représentation de la réponse en projection.....	210
Figure 116. Représentation des différentes composantes du modèle additif ajusté	211
Figure 117. Etude des valeurs prédites du modèle additif mixte.....	212
Figure 118. Etude de la normalité des résidus du modèle additif mixte.....	212
Figure 119. Illustration du compromis biais-variance.....	214
Figure 120. Représentation des fonctions de base du modèle MARS de degré 1	214
Figure 121. Représentation des différentes composantes du modèle ajusté par les réseaux de neurones	218
Figure 122. Représentation du critère basé sur les ALM pour chacun des plans de l'étude	221
Figure 123. Ajustement des modèles additifs associés à différents plans d'expériences	221
Figure 124. Prédiction des modèles additifs associés à différents plans d'expériences	222
Figure 125. Ajustement des modèles linéaires associés à différents plans d'expériences.....	222
Figure 126. Prédiction des modèles linéaires associés à différents plans d'expériences.....	223
Figure 127. Résultats pour le modèle stepwise	223
Figure 128. Ajustement des modèles MARS associés à différents plans d'expériences.....	224
Figure 129. Prédiction des modèles MARS associés à différents plans d'expériences.....	224
Figure 130. Ecarts absolus maximaux lors de l'utilisation des modèles MARS associés à différents plans d'expériences en prédiction sur le plan test.....	225
Figure 131. Ajustement des modèles PolyMARS associés à différents plans d'expériences	225
Figure 132. Prédiction des modèles PolyMARS associés à différents plans d'expériences	226
Figure 133. Prédiction des modèles de krigeage	226
Figure 134. Représentation des portées obtenues avec le krigeage ordinaire à partir des plans de Strauss	227
Figure 135. Ajustement des réseaux de neurones associés à différents plans d'expériences	228
Figure 136. Prédiction des réseaux de neurones associés à différents plans d'expériences	228
Figure 137. Représentation du coefficient de détermination en fonction du nombre de points	230
Figure 138. Qualité de prédiction sur le plan test des modèles ajustés	231
Figure 139. Représentation des portées selon les directions pour les différents	231
Figure 140. Représentation de la RMA pour le modèle PolyMARS ($a=3$) concernant les plans de Strauss et les plans à entropie maximale.....	233

Figure 141. Représentation des 3 meilleures configurations (orange, rose, turquoise) parmi les 80 initialement simulées, du cas de base (en bleu) ainsi que de la solution trouvée par le problème d'optimisation (en rouge).	234
Figure 142. Représentation du critère lié aux ALM pour des plans à 30 points en dimension 3	240
Figure 143. Boxplot pour mesurer la qualité d'ajustement (à gauche) et de prédiction (à droite) du modèle de régression quadratique des plans de Strauss, des plans à entropie maximale (dmax) et des plans à divergence minimale (KL).....	240
Figure 144. Boxplot du critère de meshratio sur les plans à 30 points en 3D.	258
Figure 145. Boxplot du critère de meshratio sur les plans à 100 points en 10D.	259
Figure 146. Valeurs (moyenne et écart-type) des critères de qualité du modèle pour chaque classe de plans....	260

TABLE DES ILLUSTRATIONS

Tableau 1. Matrice de Doelhart à 13 expériences sur $[0, 1]^3$	23
Tableau 2. Coordonnées du $(d + 1)$ -ième point du simplexe	24
Tableau 3. Récapitulatif des propriétés de certains plans d'expériences	51
Tableau 4. Discrédances en norme infinie pour $n=16, 25, 36$ et 49 pour les 3 types de plans suivants : aléatoires, suite de Halton, plan factoriel en dimension 2	62
Tableau 5. Valeurs des critères de distance pour les 3 types de plans suivants : aléatoires, suite de Halton, plan factoriel où $n=16, 25, 36$ et 49	70
Tableau 6. Critères de discrédance en norme L_2 et critères de distance pour les suites à faible discrédances à 30 points en 3D	168
Tableau 7. Critères de discrédance en norme L_2 et critères de distance pour les suites à faible discrédances à 100 points en 10D	171
Tableau 8. Liste des plans d'expériences utilisés dans les études avec le modèle PolyMARS et la régression quadratique	184
Tableau 9. Liste des plans d'expériences testés	191
Tableau 10. Liste des plans étudiés avec leur caractéristique	194
Tableau 11. Résultats des critères de qualité d'ajustement (R^2) et de validation ($RMSE$) pour les suites à faible discrédance	203
Tableau 12. Critères de validation pour les modèles de régression	213
Tableau 13. Comparaison des modèles PolyMARS ajustés avec différentes valeurs du paramètre de pénalisation α	215
Tableau 14. Etude des qualités de prédiction du modèle de krigeage ordinaire	217
Tableau 15. Comparaison des modèles de réseaux de neurones avec différentes paramétrisations	217
Tableau 16. Comparaison des modèles ajustés avec le plan de Strauss à 80 points	218
Tableau 17. Liste des plans d'expériences utilisés dans cette étude	220
Tableau 18. Variables sélectionnées triées par ordre d'importance et leur p-valeur	262
Tableau 19. Récapitulatif du modèle MARS de degré 1 ajusté	263
Tableau 20. Krigeage ordinaire avec le programme R	263
Tableau 21. Krigeage universel avec tendance linéaire (programme R)	263
Tableau 22. Krigeage universel avec tendance linéaire (programme R)	263

INTRODUCTION

Ce travail est l'objet d'une collaboration entre le département 3MI (Méthodes et Modèles Mathématiques pour l'Industrie) de l'Ecole Nationale Supérieure des Mines de Saint-Etienne (ENSM-SE) et le service Géomodélisation et Incertitudes (G&I) de Total.

Le présent travail s'inscrit dans le contexte de la simulation des phénomènes complexes par des codes de calculs très onéreux en temps de calcul. A l'initiative de Total et de l'ENSM-SE, un consortium réunissant les principaux acteurs de l'industrie a été créé pour réaliser une synergie de compétences. Cette structure, baptisée DICE (*Deep Inside Computer Experiments*), regroupe divers industriels et organismes publics de recherche tels que EDF, IRSN, ONERA, Renault et Total ainsi que des partenaires académiques (les universités Joseph Fourier de Grenoble, Paul Cézanne d'Aix-Marseille, Orsay de Paris VI et les Mines de Saint-Etienne). Son objectif est d'améliorer la compréhension de toute la thématique des codes de calculs lourds comprenant la sensibilité globale, la propagation d'incertitudes, l'estimation de quantiles, l'optimisation ou bien encore la résolution de problèmes inverses ainsi que d'innover avec des techniques de plans d'expériences numériques, de surfaces de réponses, de criblage et de screening.

Le travail présenté dans ce mémoire sur la planification d'expériences numériques en phase exploratoire pour les codes simulant les phénomènes complexes est l'un des axes de recherche financé par le consortium DICE et tout particulièrement par un contrat de thèse industriel (CIFRE) avec Total.

1. CONTEXTE SCIENTIFIQUE

L'objectif est de proposer une formulation des problèmes abordés, suffisamment générale pour s'adapter aux différentes applications. On s'inspire dans cette partie assez largement de Santner (2003), qui donne un formalisme général pour les *computer experiments*, sans pour autant l'adopter totalement. A noter que les travaux fondateurs, dans ce domaine, sont principalement ceux de Sacks, Welch, Mitchell et Wynn (1989).

Afin de concrétiser démarche et notations, on illustre le plus possible nos propos à l'aide de problèmes de l'exploration - production pétrolière. La simulation numérique de réservoir consiste à discrétiser un réservoir à la fois dans l'espace et dans le temps pour représenter les écoulements de celui-ci. Ainsi, elle est le seul moyen de résoudre un problème complexe comme par exemple décrire de façon quantitative les écoulements de plusieurs phases dans un réservoir hétérogène. La simulation peut par ailleurs résoudre des problèmes simples d'une façon plus satisfaisante qu'en utilisant les méthodes analytiques classiques. Il est par exemple

tout à fait possible de construire en quelques jours un modèle de réservoir très simplifié pour évaluer une nouvelle découverte. L'utilisation des simulations se justifie par des choix économiques, la crédibilité des résultats, l'aide à la décision et la surveillance du champ.

De nombreux phénomènes complexes de l'industrie sont modélisés par des codes de simulation numérique sophistiqués comme outil de prévision, d'optimisation et de décision. C'est le cas de la simulation numérique de production d'un réservoir pétrolier comme nous venons de le voir mais aussi de la simulation de crues et de calcul de rejet de CO₂, de la simulation d'une soufflerie à Mach 3, de la simulation d'accident nucléaire, de la simulation de crash automobile, etc.

Avant de discuter plus précisément des simulateurs, nous présentons dans un premier temps la caractérisation des phénomènes réels.

Le phénomène réel. Un phénomène est supposé être caractérisé par :

- des **variables d'intérêt** regroupées dans un vecteur $y_{réel}$ de dimension l^1 . Par analogie à un contexte de régression, la variable $y_{réel}$ sera appelée réponse. Par exemple, les variables d'intérêt peuvent être la production d'huile au cours du temps, l'instant où le puits commence à produire de l'eau (breakthrough), la production d'eau au cours du temps, etc.
- des **variables d'environnement** le plus souvent mal connues, regroupées dans une quantité v_{env} . La variable v_{env} est en général de grande dimension, voire souvent de dimension infinie (champs définis dans un volume...). Les variables d'environnement sont essentiellement les variables descriptives du réservoir, par exemple nature des faciès, champ de porosité, champ de perméabilité, perméabilités relatives, viscosités, etc.
- des **variables de contrôle** qui peuvent être fixées par l'utilisateur pour atteindre divers objectifs ; elles sont regroupées dans la quantité u_{cont} que l'on suppose, pour simplifier, de dimension finie. Par exemple, les variables de contrôle fixent le schéma de développement du champ : position des puits producteurs et injecteurs, scénarios de production... Elles sont déterminées en phase d'appréciation / développement, puis modifiées au cours de la production (forage de nouveaux puits par exemple).

¹ Ce vecteur peut comporter une structure particulière, par exemple si la réponse est constituée de variables observées au cours du temps.

Ces variables sont liées selon le schéma suivant, de type boîte noire.

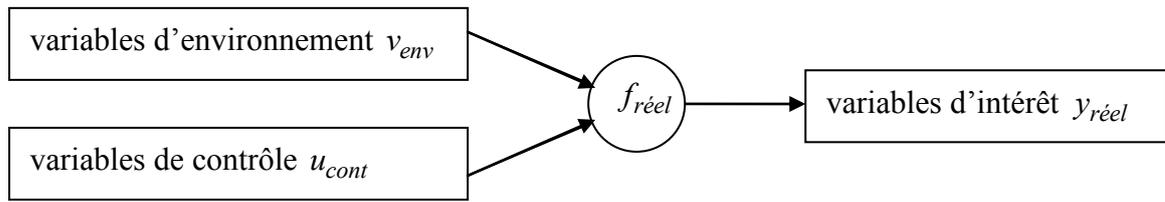


Figure 1. Schéma représentant les facteurs et les réponses

Sa traduction fonctionnelle est la suivante :

$$y = f_{réel}(v_{env}, u_{cont}).$$

Remarque. La variable temps, n'est pas explicitée dans les relations ci-dessus car elle n'intervient pas obligatoirement. Néanmoins, de nombreux problèmes possèdent une dimension temporelle. Il peut alors être intéressant de faire apparaître le temps dans la description du phénomène.

Le simulateur. A ce stade, de nombreuses questions restent sans réponse, comme par exemple l'effet de la méconnaissance de v_{env} sur $y_{réel}$. C'est une des raisons pour lesquelles la fonction $f_{réel}$ est approchée par un simulateur. Il est d'ailleurs bon de noter ici que la description du phénomène par les variables d'environnement est déjà liée à une modélisation de celui-ci, et contient donc en germe la construction du simulateur. Il est aisé de s'en convaincre, par exemple, avec l'intervention d'une variable d'environnement comme la perméabilité. Celle-ci est en effet liée à un modèle d'écoulement *via* la loi de Darcy.

Dans ce contexte s'ajoute une nouvelle classe de variables : les variables de simulation qui permettent de paramétrer, et éventuellement d'ajuster, le simulateur. Evidemment, pour leur intégration au simulateur, toutes ces variables (discrétisation en espace, orientation de la grille, gestion de la discrétisation en temps, critère de convergence, ...) sont de dimension finie et sont en général fixées avant toute simulation.

Problèmes à résoudre. Ils sont globalement de deux types : problèmes directs et problèmes inverses. Le développement des techniques de modélisation, renforcée par l'augmentation de puissance de calcul des ordinateurs, a mené au développement de simulateurs d'une très grande complexité. Ce gain de réalisme se paye par un temps de calcul souvent prohibitif. A titre d'exemple, dans le domaine pétrolier, pour une configuration de puits donnée, une seule simulation de réservoir peut prendre jusqu'à 10 jours CPU pour fournir les profils de

production attendus. Cela implique aussi une difficulté à cerner globalement le simulateur. En effet, 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. En ingénierie réservoir, on pense par exemple aux paramètres géométriques (taille du réservoir, emplacement des failles, etc.), aux paramètres d'exploitation (type de récupération, position et nombre de puits, débits de production, injection d'eau ou de gaz, etc.), aux paramètres pétrophysiques (perméabilités, porosités, etc.), etc. La solution de ce modèle numérique, comme par exemple l'estimation de la production d'hydrocarbure en fonction des différents paramètres d'entrée, est donc déduite de données incertaines.

Deux types de problèmes, parfois complémentaires, se posent :

- Analyse de sensibilité : Quels sont les paramètres ou les combinaisons de paramètres influents sur la réponse du simulateur ? En effet, afin de simplifier le problème étudié, il est d'usage de définir les paramètres susceptibles d'influencer la réponse (Saltelli *et al.*, 2000) et leur domaine de variation (domaine expérimental).
- Propagation d'incertitudes : Comment propager les incertitudes affectant certaines des variables d'entrée sur la réponse du simulateur ? En effet, une faible variation de l'un des paramètres du modèle peut entraîner des prévisions de production très différentes. La gestion des incertitudes devient donc essentielle en ingénierie réservoir, elle permet non seulement d'accroître la production (et par ce biais de traiter un problème d'optimisation), mais précise surtout les enjeux économiques. Ce problème ne sera pas traité en tant que tel dans ce mémoire, le lecteur souhaitant en savoir plus peut se référer à la thèse de Scheidt (2006).

Les difficultés à surmonter. Les difficultés essentielles proviennent du coût de calcul du simulateur et de la dimension du problème à traiter. Même après une diminution drastique de la dimension, il devient impossible pour certaines applications (par exemple, le calcul de distribution de probabilité par des techniques de Monte-Carlo) d'envisager d'utiliser directement le simulateur. Ainsi, l'exploration de ces codes est difficile et nécessite la mise en œuvre de méthodes spécifiques.

La solution. L'idée est de remplacer le simulateur par une ou plusieurs fonctions approchées. Ce sont en général des fonctions relativement simples qui sont construites à l'aide de méthodes d'approximation ou d'interpolation et à partir d'*expériences numériques*, i.e. de divers *runs* du simulateur pour des valeurs différentes des variables d'environnement et de

contrôle. Ce jeu de variables d'environnement est appelé *jeu d'apprentissage* ou *plan d'expériences numériques*.

Evidemment, la possibilité dans ce contexte de pouvoir fixer le plan d'expériences représente un degré de liberté déterminant, et une bonne interaction entre le choix de la surface de réponse et la construction du plan d'expériences est sans aucun doute un facteur prépondérant de succès.

On s'intéresse ici tout particulièrement aux *plans d'expériences numériques* sur lesquels on évalue le simulateur et qui servent de base pour construire une approximation du code de calcul.

Les difficultés sont doubles. Les expériences, pour lesquelles les simulations seront réalisées, doivent être échantillonnées de façon optimale afin d'estimer les paramètres du modèle. Ce travail se faisant en phase exploratoire, à savoir sans aucune simulation n'ait encore été effectuée, le plan d'expériences doit satisfaire à tous les types de surfaces de réponse. Ainsi le modèle choisi, *a priori* ou de manière itérative, pour décrire correctement le comportement de la réponse doit être estimé, dans un premier temps, à partir d'un petit nombre d'expériences. Nous expliquons ici comment choisir les expériences afin d'assurer la meilleure estimation possible avec un nombre de simulations assez faible lorsque nous sommes sans aucune connaissance sur le phénomène. Avant cela, nous proposons une revue des différences essentielles entre les planifications d'expériences classiques et numériques afin de mieux comprendre le cadre de ce travail.

2. DES PLANS D'EXPERIENCES CLASSIQUES AUX PLANS D'EXPERIENCES NUMERIQUES.

La différence fondamentale entre les plans d'expériences numériques et les plans classiques développés en laboratoire (expériences physiques) réside dans le fait qu'en laboratoire, on suppose la présence d'erreurs aléatoires dues aux mesures, aux conditions expérimentales, etc. Ce qui n'est en principe pas le cas avec une expérience numérique (simulation déterministe). Nous n'envisageons pas ici le cas des simulateurs bruités, à savoir des simulations entachées de bruit de simulation (Monte-Carlo ou autre). La présence d'une racine aléatoire dans le simulateur peut aussi induire des différences au niveau des réponses. Nous n'envisageons pas ce cas dans ce travail.

Remarque. Il convient de prendre conscience que la répétition d'une simulation, à variables de contrôle et d'environnement fixées, pour lesquelles on ferait varier les variables de simulation

donnerait des résultats différents. Cette pratique se rapprocherait donc de la répétition d'expériences et générerait un bruit de nature numérique.

A variables fixées, on suppose alors que le simulateur reste inchangé au fil du temps. Aussi, lorsqu'un plan est lancé séquentiellement et que le code est écrit et exécuté par une tierce personne, il peut être intéressant de dupliquer un des points dans le but de vérifier que celui-ci est resté inchangé.

Dans le cas général, les techniques des plans classiques telles que la randomisation et la répétabilité ne sont plus forcément de mise. En effet, les plans classiques ont tendance à positionner les points aux bords du domaine pour prendre en compte la variation aléatoire et offrir une tendance plus fiable en présence d'erreurs de mesures; on pense par exemple aux plans Box Behnken, plans composites, etc. (voir paragraphe 1.1).

La différence entre les approches classiques et numériques est dans la nature de l'erreur du modèle associé aux paramètres du plan. Avec l'approche classique, l'erreur dépend d'un bruit, *i.e.* de l'erreur de mesure, tandis qu'en numérique, c'est une erreur liée au modèle et non à l'expérimentation. Une autre différence réside entre les modèles paramétriques et non-paramétriques.

Il existe cependant des attributs communs entre ces deux types de plans. Dans les deux cas, l'objectif principal est d'extraire le plus d'information possible avec un nombre d'expériences limité. Une autre caractéristique est que l'échantillonnage des points se fait indépendamment et donc se prête à une évaluation désordonnée. En numérique, il sera donc possible de faire des calculs parallèles ou bien d'utiliser des multiprocesseurs tout comme en expérimentation classique où il est possible de faire plusieurs manipulations en parallèle.

Pour résumer, les méthodes de planification d'expériences basées sur les travaux de Fischer (1926) ne suffisent plus pour deux principales raisons. Tout d'abord, les grands codes de calcul décrivent des phénomènes d'une complexité de plus en plus importante, ce qui implique que les modèles correspondants sont souvent non linéaires et/ou non paramétriques. La non-linéarité, par rapport aux variables d'entrées, est d'autant plus probable que le domaine expérimental est large car, généralement, aucune contrainte expérimentale en simulation ne permet de réduire la plage de variation des paramètres d'entrées.

Ensuite, la nature de l'expérience est différente. Les expériences numériques étant des simulations et, hormis le cas des codes stochastiques mettant en œuvre une méthode de Monte-Carlo, elles fournissent la même réponse pour des conditions expérimentales identiques (incluant les paramètres algorithmiques et informatiques). La répétition d'une

expérience dans les mêmes conditions est donc sans intérêt puisqu'elle ne fait pas gagner d'information. Alors que l'on sait bien que la répétition d'expériences classiques est d'un grand intérêt, en particulier pour évaluer et/ou contrôler l'erreur dite de répétabilité.

Notre objectif étant de modéliser le code en phase exploratoire (quand aucune simulation n'a encore été réalisée), on recherchera souvent à satisfaire les deux contraintes suivantes :

1. répartir les points dans l'espace le plus uniformément possible de façon à capter les non-linéarités (par rapport aux variables d'entrées) mais aussi à éviter les répétitions.
2. faire en sorte que ce remplissage de l'espace subsiste par réduction de la dimension.

La première contrainte s'explique également si les relations supposées entre les entrées et les sorties sont incorrectes. Par exemple, si la relation est supposée essentiellement linéaire alors qu'en réalité elle est hautement non-linéaire, il est important de pouvoir détecter les fortes non-linéarités. L'erreur, qui augmente parce que le modèle supposé est incorrect, est souvent référée comme le biais du modèle qui peut se détecter à l'aide de différents outils (quantiles, scatter-plot, ...). La possibilité de le déceler est améliorée avec un choix judicieux du plan d'expériences, en observant la réponse pour des valeurs d'entrées différentes. Cette première contrainte est à l'origine des recherches sur les *Space-Filling Designs* (SFD) ; nous verrons qu'elle peut être commune avec la planification d'expériences au sens classique du terme, même si ceux-ci (SFD) restent très peu utilisés dans ce contexte.

La deuxième contrainte vient de l'observation fréquente (en grande dimension) que le code ne dépend² réellement que de quelques variables influentes, facteurs³ ou « composantes principales », ou plutôt directions discriminantes, constituées de combinaisons linéaires de ces variables. Il est donc souhaitable que la propriété de remplissage de l'espace soit encore satisfaite en projection sur des sous-espaces.

Une dernière contrainte de taille résulte d'un compromis : il faut d'une part que le nombre de simulations soit suffisamment important pour comprendre la dépendance du simulateur en ses variables d'entrée (et donc estimer tous les coefficients du modèle approché) et d'autre part, il doit être limité pour réduire le coût de simulation.

Pour résumer, un plan d'expériences efficace doit être capable d'approcher une assez grande variété de modèles⁴ et de fournir une information sur tout le domaine expérimental avec un nombre de simulations le plus faible possible.

² Cette non dépendance n'est évidemment pas absolue, mais relative à l'influence des autres variables.

³ Dans le cas, où il n'y a pas eu d'étude de sensibilité au préalable.

⁴ Nous supposons ici implicitement qu'ils sont relativement lisses.

Pour explorer efficacement, les simulations doivent :

- remplir l'espace, l'ensemble des cas expérimentaux doivent être un sous-ensemble représentatif des points dans le domaine expérimental,
- permettre d'examiner un grand nombre de variables efficacement,
- permettre d'analyser et d'estimer autant d'effets d'interactions que possible,
- et enfin le plan d'expériences doit être facile à générer.

A noter que le plan doit convenir à l'analyse souhaitée et que celle-ci devra délivrer autant d'information que possible. Les deux ne peuvent pas être considérés indépendamment. Par exemple, l'utilisation des SFD paraît être un avantage pour certains métamodèles tels que le krigeage afin d'éviter les problèmes numériques.

3. MODELISATION DES SIMULATEURS NUMERIQUES

3.1. Formalisation

Afin de simplifier les notations précédentes, nous proposons la formalisation suivante. Tout au long de ce mémoire, on étudiera des simulateurs numériques, notés par f , qui calculent une réponse y . On a donc :

$$y = f(\mathbf{x})$$

où $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ est le vecteur des d **variables d'entrée** ou **facteurs** (paramètres d'entrée pouvant varier pendant les expériences).

Tous les **facteurs** susceptibles d'avoir de l'influence sur le phénomène doivent être pris en compte, faute de quoi les résultats risquent d'être faussés et donc inutilisables. Un facteur peut être qualitatif ou quantitatif, continu ou discret. Nous considérons ici uniquement des facteurs quantitatifs continus. Le nombre de variables d'entrée d du simulateur correspond à la dimension du domaine expérimental.

Le **domaine expérimental** est donc un espace d -dimensionnel défini par une borne inférieure et une borne supérieure pour chaque variable. Les influences des facteurs peuvent être déterminées en imposant à ceux-ci des variations pouvant entraîner des modifications du comportement du phénomène étudié. Un domaine de variation est défini pour chaque facteur à l'intérieur duquel il peut prendre différentes valeurs (ou niveaux). Désormais, le domaine

expérimental est ramené, à l'aide d'une transformation linéaire, à l'hypercube unité $[0,1]^d$ (ou parfois à $[-1,1]^d$). Cette mise à l'échelle est une convenance pour représenter les figures mais aussi une technique mathématique, par exemple pour permettre le bon conditionnement de certaines matrices. Nous supposons en général que chaque composante x_i prend ses valeurs continûment dans un intervalle borné. Cependant, le domaine expérimental peut être plus complexe en pratique.

Une **expérience** (ou point) est défini par un ensemble de conditions opératoires correspondant mathématiquement à un vecteur \mathbf{x} de longueur d , de la forme (x_1, \dots, x_d) .

Le **plan d'expériences** (ou plan d'apprentissage) est un objet mathématique qui représente, sous forme codée ou normée, l'ensemble des n expériences à réaliser. En expérimentation numérique, cela correspond aux n vecteurs sur lesquels on évalue le simulateur. Un plan d'expériences est représenté par une matrice $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ constituée de n lignes (correspondant aux n expériences) et d colonnes (correspondant aux d facteurs) avec

$$\mathbf{x}_j = (x_j^1, x_j^2, \dots, x_j^d)^t$$

où l'élément x_j^i de la matrice ainsi formée est la valeur que prend la j -ième variable à la i -ième expérience.

La **réponse** est la quantité y obtenue pour un plan spécifié, *i.e.* lorsque l'on fait varier les facteurs étudiés. Mathématiquement, cette réponse est représentée par $f(\mathbf{x})$. Les variables de sortie d'un simulateur sont la plupart du temps vectorielles. Pour simplifier, nous supposons que $y \in \mathbb{R}$. La valeur de la réponse simulée au i -ième point d'apprentissage $\mathbf{x}^i = (x_1^i, \dots, x_d^i) \in \mathbb{R}^d$ sera notée par $y_i = f(\mathbf{x}^i)$. Aussi, le vecteur des réponses sur le plan d'expériences est de la forme :

$$\mathbf{Y} = (y_1, \dots, y_n)^t = (f(\mathbf{x}^1), \dots, f(\mathbf{x}^n))^t.$$

La donnée du plan \mathbf{X} constitué de n expériences et du vecteur \mathbf{Y} permet de construire une réponse approchée, notée

$$\hat{\mathbf{Y}} = (\hat{y}_1, \dots, \hat{y}_n)^t.$$

On notera alors $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1 \dots \hat{\varepsilon}_n)^t$ le vecteur des résidus, *i.e.* les écarts entre les observations \mathbf{Y} et les valeurs $\hat{\mathbf{Y}}$ prédites par le modèle.

3.2. Protocole expérimental

Toutes les applications de simulations numériques suivent la même démarche générale, cette section donne la méthodologie de résolution adoptée pour étudier un Space-Filling Design et un modèle. Les étapes principales constituant la démarche méthodologique sont les suivantes :

Choix des facteurs et du domaine expérimental (screening). Pour observer une éventuelle variation de la réponse, il faut obligatoirement faire varier un ou plusieurs facteurs. Dans le cas d'un grand nombre de facteurs, un *screening* doit permettre de détecter les facteurs influents et leur domaine de variation. Cette étape revient à identifier le domaine expérimental. Une analyse de sensibilité permet de réduire le nombre de variables et le domaine expérimental.

Proposition d'un modèle. Ces modèles (ou surfaces de réponse ou métamodèles) peuvent être infiniment variés et dépendent du type de problème étudié : modèles linéaires ou non linéaires, paramétriques ou non-paramétriques, etc. Si nous connaissions ces modèles, il serait facile de les utiliser pour faire de la prévision. Cette étape inclut la sélection d'une classe de modèles pour approcher la (ou les) sortie(s) du simulateur. Nous ne demandons pas à ces modèles d'expliquer le phénomène étudié en détail car soit cela se révélerait impossible, soit cela nécessiterait de très longues recherches. Ils doivent simplement permettre de bien représenter la réponse étudiée dans le domaine expérimental et, ainsi, d'estimer la réponse de manière acceptable. En effet, une prévision n'a guère d'intérêt si son incertitude est trop grande. L'estimation de la réponse étudiée doit être connue en chaque point du domaine expérimental avec une incertitude non pas minimale mais acceptable au sens où si l'on réalise l'expérience à l'aide du simulateur la réponse doit être de qualité semblable. Différentes classes de métamodèles sont données en annexe A.

Plan d'expériences. Le choix d'un plan d'expériences dépend des objectifs et des qualités souhaitées. Certaines de ces qualités peuvent parfois être elles-mêmes liées au métamodèle considéré à l'étape précédente. Généralement, nous ne faisons pas d'hypothèse sur celui-ci, et le plan doit alors être satisfaisant pour toutes les classes de métamodèles. Ainsi, cette étape consiste à choisir un plan initial \mathbf{X} avec un nombre approprié d'expériences et de niveaux pour chaque variable afin d'assurer en particulier le recouvrement de l'espace. Nous parlerons

dans ce mémoire des différents Space-Filling Designs envisageables. Le nombre de simulations dépend de la complexité du phénomène mais aussi de la complexité du modèle choisi à l'étape suivante. On évaluera le simulateur sur les vecteurs d'entrée de \mathbf{X} pour construire la réponse \mathbf{Y} du système.

Construction du modèle. A partir des valeurs du plan d'expériences \mathbf{X} et de la réponse \mathbf{Y} du système, on ajuste un modèle (ou métamodèle) que l'on souhaite prédictif. Ce modèle doit être rapide à ajuster, notamment en grande dimension, et facile à interpréter.

Validation du métamodèle : A la fin de l'étape précédente, nous avons obtenu un métamodèle mais nous ne pouvons pas directement utiliser celui-ci sans vérifier qu'il représente correctement la réponse dans le domaine expérimental.

Un premier diagnostic consiste à étudier les valeurs du modèle aux points du plan d'expériences : nous parlerons dans ce cas d'ajustement. Pour ce faire, il existe des critères permettant de tester la bonne adéquation entre les observations et les prévisions sur les données utilisées lors de la construction du modèle. C'est par exemple, le rôle du coefficient de détermination R^2 .

Une fois le modèle accepté, il est nécessaire de vérifier la capacité du modèle à prédire en n'importe quel point du domaine expérimental la valeur de la réponse étudiée. Il s'agit ici de comparer les prévisions du modèle en des points différents de ceux du plan d'expériences.

Pour vérifier que le modèle représente bien le phénomène étudié dans le domaine expérimental, le critère usuel est l'erreur quadratique moyenne ou *MSE* (Mean Squarred Error). Plus la valeur de la MSE est petite et plus le métamodèle est bon.

Le nombre de simulation étant souvent limité, il devient difficile de faire cette comparaison sur un nombre de points représentatifs. Une possibilité est de mettre en œuvre des techniques de validation croisée. Différents critères pour mesurer la qualité d'un modèle en ajustement ou en prédiction sont donnés à l'annexe B.

Interprétation du modèle et analyse de sensibilité. L'interprétation du modèle est importante, en particulier pour les modèles complexes tels que le krigeage ou les réseaux de neurones. Interpréter le modèle permet de comprendre la relation entre les entrées et les sorties du code.

Complémentarité ou séquentialité. La phase d'interprétation des résultats peut mettre en évidence des déficiences dans la démarche préalablement choisie. Ceci nécessite donc de proposer un nouveau modèle mais aussi de nouvelles expériences et parfois même dans un nouveau domaine expérimental. Pour cela, il y a un intérêt réel à utiliser les informations obtenues auparavant que ce soit celles apportées par le métamodèle initial ou par les simulations déjà effectuées.

Cette méthodologie a été utilisée lors des applications de la partie IV.

4. OBJECTIFS ET ORGANISATION DU MEMOIRE DE THESE

Ce travail a été motivé par le fait qu'il existe peu de description précise et critique des plans d'expériences numériques dans la littérature. Bien souvent on trouve des descriptions à travers différents types de littérature. Signalons deux ouvrages qui discutent des plans d'expériences dans le contexte de phénomènes complexes : Santner *et al.* (2003) et Fang *et al.* (2006).

L'objectif de ce manuscrit est d'étudier les avantages et les inconvénients des plans d'expériences numériques et aussi de proposer des nouvelles méthodes. Pour cela, il est nécessaire de développer des outils spécifiques.

Nous tenterons également de répondre à d'autres types de questions. Par exemple, bien que la discussion autour de l'uniformité reste essentielle nous développerons des techniques pour générer des plans selon des distributions hétérogènes dans le cas où nous sommes en possession d'information *a priori* sur le phénomène. Ce mémoire doit aussi permettre des ouvertures vers des questions essentielles :

Quel plan d'expériences pour quelle surface de réponse ? Quel plan d'expériences choisir selon l'objectif de l'étude (propagation d'incertitudes, optimisation, problème inverse) ?

Le présent document se décompose en quatre parties.

Une première partie fait un état de l'art des plans d'expériences couramment utilisés en expérimentation numérique ainsi que de certains critères pour étudier la qualité intrinsèque de ces plans au niveau du remplissage de l'espace. Ceci nous amène également à traiter des plans basés sur un (ou des) critère(s) spécifique(s).

Une deuxième partie présente une nouvelle méthode de planification d'expériences basée sur des méthodes de Monte-Carlo par chaînes de Markov. Nous proposerons également dans cette

partie des extensions qui permettent de traiter des problèmes particuliers tels que la génération de simulation dans un domaine contraint ou bien encore la génération de points de manière préférentielle (hétérogène) qui peut se révéler particulièrement intéressante lorsqu'il y a une quelconque connaissance métier.

La troisième partie présente de nouveaux critères intrinsèques aux plans pour étudier et qualifier la distribution des points. Nous appliquerons tous ces critères sur les plans d'expériences utilisés lors des applications afin d'en déduire des premières informations.

La dernière partie concerne l'application des outils innovants en les comparant avec ceux habituellement utilisés. Nous pourrions voir deux types d'application : d'une part théorique sur un cas d'école (10D) et d'autre part plus concrète en ingénierie réservoir (3D et 8D). Ces applications permettront d'ouvrir la discussion sur les plans à utiliser selon nos objectifs.

PARTIE I.

COMPTE-RENDU DE LA LITTERATURE

Cette partie est consacrée à présenter un historique sur un certains nombre de techniques (plus ou moins récentes) en planification d'expériences.

Le premier chapitre discute des plans d'expériences usuels tels que les plans standard jusqu'alors utilisés en expérimentation classique, les hypercubes latins, les suites à faible discrédance, etc. Le second s'attache à présenter les critères d'optimalité spécifiques sur lesquels des techniques de planification sont basées.

Le but est de donner au lecteur un bref aperçu des connaissances actuelles sur les plans d'expériences numériques en phase exploratoire et de lui permettre de situer les travaux de la partie II et III dans cette thématique.

Chapitre 1

Plans d'expériences numériques usuels

Comme nous avons pu l'indiquer en introduction, les objectifs attendus des plans d'expériences pour l'exploitation des codes de calcul en phase initiale sont principalement le remplissage de l'espace et la non redondance des points projetés sur les sous-espaces. Nous garderons à l'esprit également que les méthodes de génération des plans doivent être peu coûteuses en temps d'exécution et doivent permettre d'obtenir des plans efficaces avec un faible nombre d'expériences.

Ce chapitre a pour but de fournir un état de l'art critique et le plus concis qui soit des plans d'expériences paraissant les plus adaptés en phase d'exploration vis-à-vis essentiellement des propriétés énoncées ci-dessus.

Dans une première partie, nous évoquons les plans historiques dédiés à l'expérimentation classique. Cette première classe de plans est référée sous le nom des plans standard. Ensuite, nous étudions les plans « marginaux » qui sont construits de manière à ce que la répartition des points en projection sur les arêtes de l'hypercube unité (encore appelées les marges de dimension 1) soit uniforme. Enfin, nous discutons des suites à faible discrédance qui sont générées de façon à répondre, d'un certain point de vue, à la question du remplissage de l'espace.

1.1. Plans standard

Nous avons choisi de discuter des plans standard dans ce mémoire car ils ont été développés à l'origine pour des applications de surface de réponse. Parmi ces plans, les plus courants sont les plans factoriels, les plans de Box-Behnken et composites centrés qui sont faciles à générer. Nous avons choisi de discuter aussi des plans de Doehlert qui répondent particulièrement bien à la question du remplissage de l'espace.

Dans cette partie, nous faisons une brève description de ces plans puis on s'attache à étudier les différentes propriétés recherchées ici : remplissage de l'espace, non redondance, coût, etc. A noter que ces différents plans seront souvent cités tout au long de ce mémoire à titre de comparaison pour étudier l'intérêt des Space-Filling Designs.

1.1.1. Plans factoriels complets

La méthode la plus simple pour répondre à la problématique du remplissage de l'espace consiste à choisir les points sur une grille régulière dans le domaine expérimental.

Description. Pour obtenir une grille régulière à k niveaux, il suffit de choisir k valeurs réparties à différents intervalles réguliers pour chaque facteur. Par exemple, dans le carré unité $[0, 1] \times [0, 1]$ et en choisissant 5 niveaux, on peut considérer la grille de points (voir Figure 2) suivante $\{0, 0.25, 0.5, 0.75, 1\} \times \{0, 0.25, 0.5, 0.75, 1\}$.

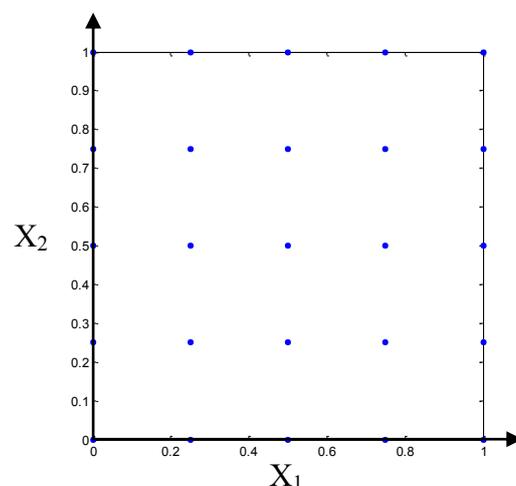


Figure 2. Un plan factoriel complet à 5 niveaux.

Il est évident que plus le nombre de niveaux sera important et plus le remplissage de l'espace sera satisfaisant mais cela entraîne une « explosion » du nombre de simulations. Il est donc essentiel de trouver un bon compromis et de choisir les niveaux paraissant les plus adaptés au problème.

Discussion. Cette méthode reste intéressante si la dimension du problème est petite, c'est-à-dire 2, voire 3. Dès que la dimension est plus élevée, le nombre de simulations $n = k^d$ augmente considérablement, et l'idée d'utiliser une grille devient alors inenvisageable. De plus, seulement k valeurs différentes apparaissent pour chaque dimension. Dans le cas où la réponse dépend très fortement de seulement quelques variables (par exemple, une ou deux en dimension 5), le plan factoriel conduit à beaucoup de répliques inutiles. Ce type de plan devient donc inutilisable lorsque la dimension augmente car on perd la plupart des points en projection sur les axes factoriels. Par exemple, si la réponse s'écrit tel que $f(X_1, X_2) = f_1(X_1)$ ou $f(X_1, X_2) = f_2(X_2)$ alors le plan factoriel de la Figure 2 est mal adapté puisque l'information donnée par celui-ci se réduit à 5 points au lieu de 25.

Remarque. Si le nombre de coefficients du modèle à estimer est proche du nombre d'expériences, il est conseillé d'ajouter au plan factoriel quelques points répartis uniformément dans le domaine.

1.1.2. Plans factoriels fractionnaires

Le contexte dans lequel nous sommes ne nous autorisant pas à effectuer un nombre important de simulations, les plans factoriels complets sont inadapés. Pourtant le principe sur lequel ils reposent reste intéressant. C'est pourquoi, les plans factoriels fractionnaires peuvent constituer une bonne alternative. Le principe consistant à prendre des sous-ensembles de plans factoriels complets permet de réduire considérablement le coût en nombre de simulations (pour plus de détails voir Myers et Montgomery, 1995). Cependant, les problèmes évoqués en relation avec les plans factoriels au sujet des projections sur les axes factoriels demeurent toujours présents. De nouveaux problèmes d'alignements dus aux alias de la fraction apparaissent ; de la même façon que c'est le cas pour les tableaux orthogonaux linéaires (voir Figure 9).

1.1.3. Plans composites

Description. Une matrice d'expériences composite est une combinaison

- d'une matrice d'expériences factorielle à 2 niveaux, complète (2^d) ou fractionnaire (2^{d-r}), dont les points sont pris parmi les sommets de l'hypercube $[-1, 1]$ par exemple)
- d'une matrice d'expériences axiale composée de points disposés symétriquement sur chaque axe, à une distance⁵ α du centre du domaine
- et d'un point central⁶ pour d facteurs qui permet d'obtenir une information sur la variabilité du phénomène ou tester la validité du modèle (par exemple, dans le cas d'un modèle linéaire de degré 1, tester la présence d'une courbure).

Un composite à faces centrées sur le domaine cubique $[-1, 1]^2$ correspond donc à un plan factoriel à 3 niveaux (-1, 0, 1). A noter que ces plans sont adaptés à la démarche « one-at-a-time » qui va imposer des points sur les axes et les plans factoriels 2^d .

Discussion. Les plans composites sont souvent utilisés en expérimentation classique pour approcher des surfaces de réponse de degré 2 (Myers et Montgomery, 1995). Il est possible de générer différents plans de ce type en faisant varier la distance entre le point central et les points aux bords du domaine. On peut citer par exemple, les plans composites centrés ordinaires, les plans composites à faces centrées, les plans composites centrés inscrits, etc.

Le nombre d'expériences des plans composites augmente rapidement avec le nombre de variables étudiées et cette augmentation de points est surtout due à la matrice factorielle. Ces plans ne répondent pas convenablement aux objectifs de remplissage de l'espace et notamment aux critères de bonne répartition de points en projection. En effet, ces plans ne testent que 3 ou 5 niveaux (selon la valeur de α) par paramètres et cela avec n'importe quelle taille du plan.

1.1.4. Plans de Box-Behnken

Description. Un plan de Box-Behnken est un plan dont les variables ne prennent que 3 niveaux $(-a, 0, +a)$ si l'on considère pour le domaine expérimental l'hypercube $[-1, 1]^d$. Ces plans d'expériences sont composés d'une matrice factorielle à 2 niveaux (2^d points) et de

⁵ Pour un domaine cubique, on considère donc $\alpha = 1$

⁶ En expérimentation classique, il est conseillé de considérer plusieurs points au centre. Ce qui n'a plus de sens lorsque le simulateur est purement déterministe.

blocs incomplets équilibrés en suivant un arrangement particulier. Ces matrices sont ensuite complétées d'un point au centre du domaine. Ces plans sont une alternative aux plans composites parce qu'ils requièrent seulement 3 niveaux pour chaque facteur (Box et Behnken, 1960) et permettent aussi de modéliser une surface de réponse de degré 2. On peut trouver dans Myers et Montgomery (1995) ou bien aussi dans Dreesbeske, Fine et Saporta (1994) la méthode de construction de ces plans et plus précisément la manière de former les blocs.

Discussion. En termes de nombre de points, un plan de Box-Behnken est comparable à un plan composite en dimension 3 et 4. A noter qu'il n'existe pas de Box-Behnken en dimension 2. Ces plans ayant leurs points sur les axes factoriels du domaine, ils ne remplissent pas correctement l'espace.

1.1.5. Plans de Doehlert ou réseaux uniformes

Description. Les plans de Doehlert (1970) sont une partie intégrante de la famille des réseaux. La méthode de génération est itérative et consiste à considérer un simplexe initial dans le domaine d'exploration et à lui appliquer des isométries (translations et rotations) à partir d'un de ses sommets (en général, on considère des translations). En itérant ce procédé, on obtient alors une distribution de points particulière (Figure 3). Pratiquement, cela revient, pour chaque variable sur $[-1, 1]$, à soustraire successivement les coordonnées de chaque point du simplexe initial aux autres.

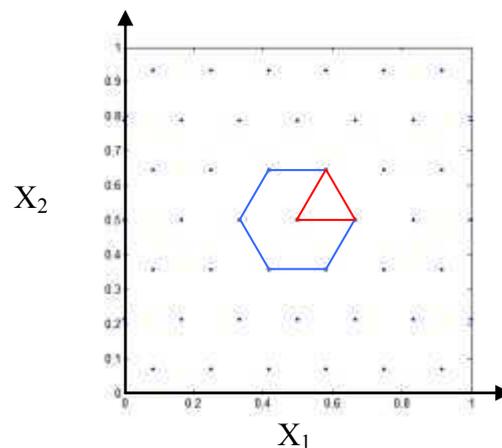


Figure 3. Un plan de Doehlert à 45 points dans le carré unité avec son simplexe initial : le triangle équilatéral en rouge

Par exemple pour $d = 3$, les coordonnées du simplexe initial sont, sur $[0, 1]$, données par les 4 premières expériences du tableau ci-dessous. La matrice à 13 expériences de Doehlert est présentée dans le Tableau 1.

N°	X_1	X_2	X_3
1	0.5	0.5	0.5
2	1	0.5	0.5
3	0.75	0.933	0.5
4	0.75	0.6445	0.9085
5	0	0.5	0.5
6	0.25	0.067	0.5
7	0.75	0.067	0.5
8	0.25	0.9330	0.5
9	0.25	0.3555	0.0915
10	0.75	0.3555	0.0915
11	0.5	0.7885	0.0915
12	0.25	0.6445	0.9085
13	0.5	0.2115	0.9085

Tableau 1. Matrice de Doehlert à 13 expériences sur $[0, 1]^3$

Discussion. Ces plans permettent avant tout de générer des points uniformément dans l'espace ; ce qui est particulièrement utile lorsque l'on veut explorer la totalité du domaine (bornes et intérieur) sans proposer de modèle *a priori* représentant la réponse. La densité de points sera d'ailleurs déterminée par la taille que l'on accorde au simplexe initial. Ces plans ont aussi de nombreux autres avantages.

Une de leurs caractéristiques importantes est de permettre une démarche séquentielle dans l'étude d'une surface de réponse. Par exemple, si la zone expérimentale la plus intéressante ne se trouve pas dans le domaine préalablement choisi mais plutôt située au voisinage de celui-ci ; il est possible de construire autour de l'un des points du plan initial, un nouvel ensemble d'expériences qui comprend une partie des simulations déjà effectuées sans pour autant détruire l'uniformité préexistante. A noter que cette propriété est par exemple non valable concernant les plans factoriels.

Chaque simplexe étant déductible du simplexe de dimension immédiatement inférieure, il est facile d'augmenter le nombre de facteurs progressivement sans pour autant recommencer toutes les expériences lorsque l'on s'aperçoit que le domaine d'exploration ne convient pas. Pour cela, il suffit d'ajouter une d -ième colonne de zéros pour les d premiers points du simplexe de dimension $d - 1$ et d'ajouter un nouveau point dont les coordonnées sont données dans le tableau ci-dessous.

<i>Variable</i>	$d - q (> 0)$...	$d - 3$	$d - 2$	$d - 1$
<i>Point</i> $d + 1$	$\frac{1}{\sqrt{(2(d+1)-d)(d-q)}}$...	$\frac{1}{\sqrt{(2(d-1)(d-2)}}$	$\frac{1}{\sqrt{(2d(d-1)}}$	$\sqrt{(d+1)/2d}$

Tableau 2. Coordonnées du $(d + 1)$ -ième point du simplexe

Ces plans permettent aussi de simplifier l'étude à une zone restreinte en changeant les dimensions du simplexe initial.

Le nombre minimal de points distincts, n , d'un plan de Doehlert pour un nombre de facteurs d donné est $n = d^2 + d + 1$, ce qui en grande dimension paraît être un avantage. Par exemple si $d = 20$, n sera égal à 421. Le cas $d = 2$ est représenté (à un facteur d'échelle près) sur la Figure 3, on peut vérifier que le simplexe de base est composé des 6 points de l'hexagone et du point central

Cependant, l'uniformité de ces plans induit une perte d'information due aux alignements dans l'espace comme on peut le constater sur la Figure 3.

1.1.6. Conclusions

Les plans étudiés auparavant sont généralement de bonne qualité concernant le remplissage de l'espace. Cependant, la plupart de ces méthodes sont soit coûteuses en nombre de simulations ou/et soit une cause de perte d'information dans le contexte des plans d'expériences numériques. C'est essentiellement pour ces raisons que nous avons choisi de n'étudier qu'une infime partie de tous les plans d'expériences classiques connus. Cela ne signifie pas qu'il faut les écarter pour autant des plans d'expériences numériques, le lecteur souhaitant en savoir davantage sur ces plans peut se référer à de nombreux ouvrages, par exemple Myers et Montgomery (1995), Lewis, Mathieu et Phan-Tan-Luu (1999).

1.2. Plans « marginaux »

Nous introduisons dans cette section des plans qui, par construction, ont des bonnes propriétés en termes de non redondance, et de non alignement sur certains sous-espaces. Cependant, rien ne garantit qu'ils remplissent bien l'espace. Nous regarderons ici ce qu'il en est.

On définit par ailleurs ce que l'on entend par marges. La notion de marges se réfère à des sous-espaces factoriels. Par exemple, les marges de dimension 1 correspondent aux axes factoriels.

1.2.1. Hypercubes latins

La méthode d'échantillonnage par hypercube latin (Mac Kay, Conover, Beckman, 1979) a été introduite pour évaluer numériquement les intégrales multiples. Elle permet d'assurer la non redondance de l'information au travers d'une bonne répartition des projections sur les axes factoriels. En pratique, les hypercubes latins sont très utilisés en planification d'expériences numériques notamment pour leur simplicité d'usage et de construction.

Description. Chaque axe $[0, 1]$ du cube unité est découpé en n segments de même longueur

de la façon suivante $\left\{ \left[0, \frac{1}{n} \right], \left[\frac{1}{n}, \frac{2}{n} \right], \dots, \left[\frac{n-1}{n}, 1 \right] \right\}$. En faisant le produit cartésien de ces

intervalles, on obtient un maillage de n^d cellules de même taille. Ensuite, on choisit n cellules parmi les n^d possibles de telle manière que les cellules des marges de dimension 1 soient représentées une fois chacune. Enfin, on tire un point au hasard dans chaque cellule présélectionnée.

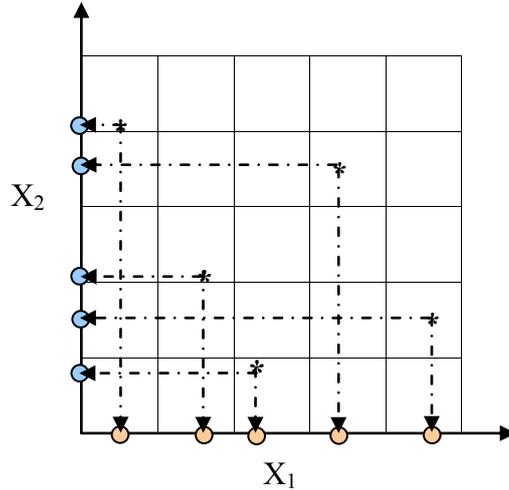


Figure 4. Un échantillonnage par hypercube latin à 5 points en dimension 2

Définition

Un hypercube latin à n points sur $[0, 1]^d$, est défini par l'ensemble de points X^i tel que

$$X_j^i = \frac{\pi_j(i) + U_j^{(i)}}{n}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq d,$$

où π_j est une permutation de $1, \dots$ et $U_j^{(i)} \sim U[0,1]$ est une valeur aléatoire de distribution uniforme sur $[0, 1]$.

Ainsi, $(\pi_1(i), \dots, \pi_d(i))$ représente la cellule dans laquelle se trouve le point X^i , et $(U_1^{(i)}, \dots, U_d^{(i)})$ désigne à quel endroit le point se trouve dans la cellule.

L'hypercube latin obtenu correspond donc à la matrice de n lignes et d colonnes à coefficients X_j^i .

Remarque 1. Nous pouvons choisir de placer le point au centre des cellules pour ainsi supprimer l'aléa de la planification.

Remarque 2. Un hypercube latin, défini par la matrice π , est un plan très facile à construire puisque chaque colonne est une permutation de 1 à n .

Discussion. Les points d'un hypercube latin ont la propriété intéressante d'être uniformément distribués sur les axes factoriels (voir Figure 4). Cependant, cette propriété n'assure pas l'uniformité des points sur tout le domaine expérimental. Pour n fixé, il existe $n!$ permutations possibles pour chacune des d colonnes, et donc $(n!)^{d-1}$ hypercubes latins

éventuels qui n'assurent pas tous l'uniformité des points dans l'espace. C'est le cas par exemple de l'hypercube latin de la Figure 5 où les points sont placés sur l'une des diagonales du domaine expérimental. Si le processus réel est une fonction de $(X_2 - X_1)$, alors l'information donnée par ce plan d'expériences se réduit à 1 point au lieu de 5.

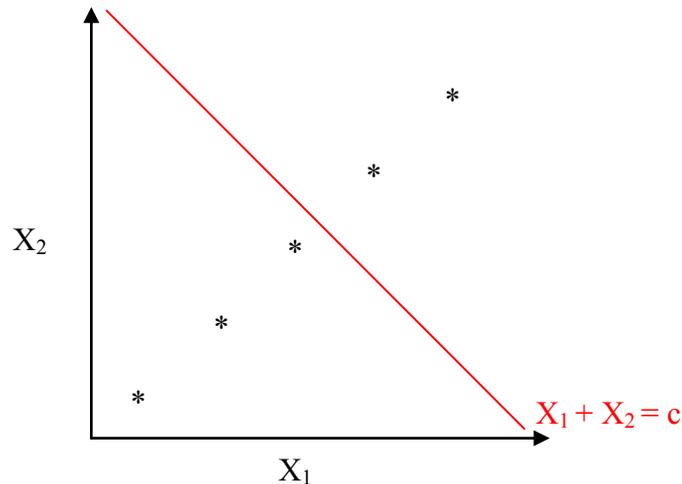


Figure 5. Un échantillonnage par hypercube latin à 5 points en dimension 2

De façon générale, le contexte dans lequel nous sommes ne nous autorisant pas à prendre un grand nombre de points, nous aurons un nombre de mailles réduit. Il sera donc fréquent d'obtenir des plans où l'on peut observer des zones lacunaires et des agglomérats de points comme on peut l'observer sur la Figure 6.

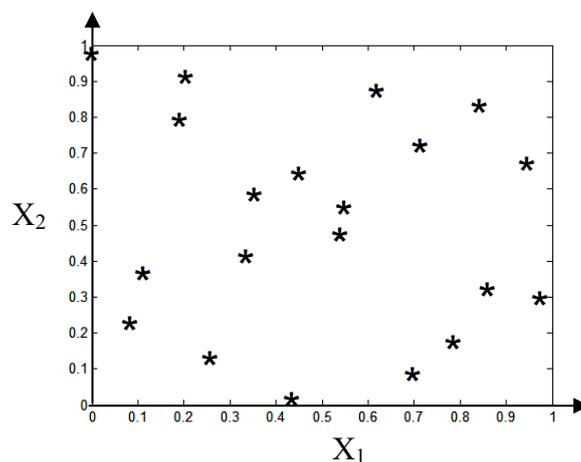


Figure 6. Un hypercube latin à 20 points en dimension 2

Amélioration. Pour pallier ces problèmes, il existe deux possibilités. La première consiste à construire les $(n!)^{d-1}$ hypercubes latins possibles et de choisir le meilleur à l'aide d'un critère de sélection. Cependant, cette méthode n'est pas envisageable du fait du grand nombre

d'hypercubes latins à générer, même si construire un hypercube latin demande peu de temps d'exécution. La deuxième possibilité, plus réaliste, consiste à générer un hypercube latin et à lui appliquer un critère afin de le rendre meilleur au sens des exigences souhaitées.

On peut chercher par exemple à améliorer le critère de corrélation. L'objectif est alors de construire un hypercube latin pour les effets principaux, voire plus si possible. On peut citer Kenny (1998), Tang (1998) et Owen (1994).

Il est aussi possible d'appliquer des critères d'uniformité, ce qui revient à sélectionner un hypercube latin ayant de bonnes propriétés d'uniformité. De nombreuses références sont disponibles sur ce sujet : Johnson *et al.* (1990), Fang *et al.* (1994), Fang et Zhu (1993). Park (1994) propose une construction, basée sur un algorithme d'échanges, d'hypercubes latins optimaux quel que soit le critère. Quelques uns de ces hypercubes latins améliorés seront comparés lors des applications (voir la partie IV). Pour donner un exemple, un hypercube latin maximin résulte d'un bon compromis entre le critère de maximin (qui maximise la distance minimale entre une paire de points, voir section 2.2.3) et les bonnes propriétés de projection des hypercubes latins.

1.2.2. Tableaux orthogonaux

1.2.2.1. Cas général

Conceptuellement, les tableaux orthogonaux (Owen, 1992) sont très proches des hypercubes latins. En effet, ils ont l'avantage d'avoir les bonnes propriétés de projection en grande dimension qu'ont les hypercubes latins en dimension 1.

Définition

Un tableau orthogonal de force t à q symboles est une matrice de n lignes et $d(>t)$ colonnes avec ses éléments prenant q valeurs possibles, telle que toutes les sous matrices de taille $n \times t$ aient q^t lignes apparaissant le même nombre λ fois.

Il est évident que $n = \lambda q^t$. On note un tel tableau orthogonal $OA_\lambda(n, d, q, t)$.

Description. Géométriquement, cela revient à découper les axes du cube unité en q segments égaux afin d'obtenir q^d cellules de même taille. On choisit n cellules de telle façon qu'elles forment un tableau orthogonal défini comme ci-dessus. Cette définition assure que chaque t colonnes de la matrice du plan, soit chaque t -uplet, apparaît exactement λ -fois.

Remarque. Un tableau orthogonal de force 1 est un hypercube latin.

Comme pour les hypercubes latins, on peut choisir le point aléatoirement dans la cellule mais aussi au centre de la cellule. Dans ce dernier cas, toutes les projections sur les sous-espaces de dimensions t conduisent alors à une grille régulière comme on peut le constater sur la Figure 7.

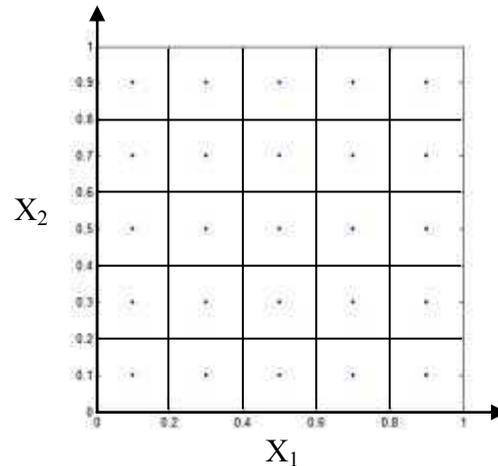


Figure 7. Un plan généré par un tableau orthogonal $OA_1(25,5,5,2)$ dont les points sont centrés et projetés sur le sous-espace (X_1, X_2)

Définition

Un échantillonnage par tableau orthogonal (abusivement appelé par la suite tableau orthogonal) à n points sur $[0,1]^d$ est un ensemble de points X^i tel que

$$X_j^i = \frac{\pi_j(A_j^{(i)}) + U_j^{(i)}}{q}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq d,$$

où π_j est une permutation de $0, \dots, q-1$, $A_j^{(i)}$ sont les éléments du tableau orthogonal et

$U_j^{(i)} \sim U[0,1]$ est une valeur aléatoire de distribution uniforme sur $[0,1]$.

Ainsi, $(\pi_1(A_1^{(i)}), \dots, \pi_d(A_d^{(i)}))$ représente la cellule dans laquelle se trouve le point X^i , et

$(U_1^{(i)}, \dots, U_d^{(i)})$ désigne à quel endroit le point se trouve dans la cellule.

Le tableau orthogonal correspond à la matrice de n lignes et d colonnes à coefficients X_j^i .

Concernant la génération des plans, on a une propriété analogue à celle des hypercubes latins :

Propriété

Si on effectue une permutation des symboles sur chaque colonne d'un tableau orthogonal de force t , alors on obtient toujours un tableau orthogonal de force t .

Discussion. La Figure 8 montre les points générés par un tableau orthogonal de type $OA_1(25,5,5,2)$ projetés sur le sous-espace (X_1, X_2) . Nous pouvons observer qu'il y a bien un unique point par cellule et donc que les points projetés sur le sous-espace (X_1, X_2) sont répartis uniformément.

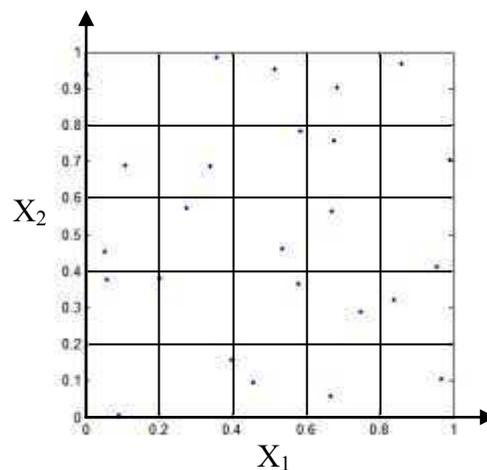


Figure 8. Un plan généré par un tableau orthogonal $OA_1(25, 5, 5, 2)$ dont les points sont randomisés et projetés sur le sous-espace (X_1, X_2)

Propriété

Un tableau orthogonal de force t et d'index λ est aussi un tableau orthogonal de force $s \leq t$ et d'index $\lambda' = \frac{n}{q^s} \geq \lambda$.

Si l'on reprend les exemples des Figure 7 et Figure 8, ces tableaux orthogonaux de force 2 et d'index $\lambda = 1$ (1 point par cellule) sont également de force 1 et d'index $\lambda' = 5$ (5 points par strate). On constate le même problème que pour les plans factoriels en projection sur les marges de dimension 1 concernant les répétitions sur les axes.

Les tableaux orthogonaux de force t permettent de généraliser aux dimensions t les bonnes propriétés de projections en dimension 1 des hypercubes latins. Ainsi, les points sont globalement répartis plus uniformément que pour les hypercubes latins. Cependant, la

propriété ci-dessus implique que les projections sur les sous-espaces de dimension strictement inférieure à t conduisent à la réplication des niveaux. Aussi cela devient un inconvénient majeur lorsque l'on ne connaît pas à l'avance le nombre de facteurs réellement influents sur le phénomène.

Ces plans ont aussi des inconvénients concernant le coût, n , de la simulation qui est élevé car n est égal à λq^t . Il est donc préférable d'utiliser des tableaux orthogonaux d'index $\lambda = 1$ et même avec ce paramétrage, le coût reste très élevé. On remarque cependant que la taille du plan ne dépend pas du nombre de facteurs mais uniquement du nombre de niveaux et de la force. Ces plans peuvent donc devenir intéressants à partir d'un certain nombre de facteurs.

A noter également que les tableaux orthogonaux n'existent que pour certaines valeurs de n , en particulier pour $n = \lambda q^t$, et pour certaines valeurs de d . A titre d'exemple, le tableau orthogonal $OA_4(6^2, 4, 6, 2)$ ne peut exister. En effet, Tarry (1901) a démontré qu'il n'existe pas de carré gréco latin⁷ d'ordre 6.

1.2.2.2. Cas particulier des tableaux orthogonaux linéaires

Les tableaux orthogonaux linéaires ont été considérés car ils sont plus faciles à mettre en œuvre que dans le cas général.

Définition

Un tableau orthogonal linéaire est un tableau orthogonal tel que :

- le nombre de symboles q est un nombre premier

- les lignes sont toutes distinctes et constituent un sous-espace vectoriel de $(\mathbf{Z} / q\mathbf{Z})^d$

Dans ce cas, on le note $OA_\lambda((\mathbf{Z} / q\mathbf{Z})^d, t)$.

$\mathbf{Z} / q\mathbf{Z} = \{0, \dots, q-1\}$, (ou bien encore noté \mathbf{Z}^q) est un corps sur \mathbf{Z} car on suppose q premier.

⁷ Tableau de n lignes et n colonnes remplies avec n^2 paires différentes tel que chaque ligne et chaque colonne ne contienne qu'un seul exemplaire.

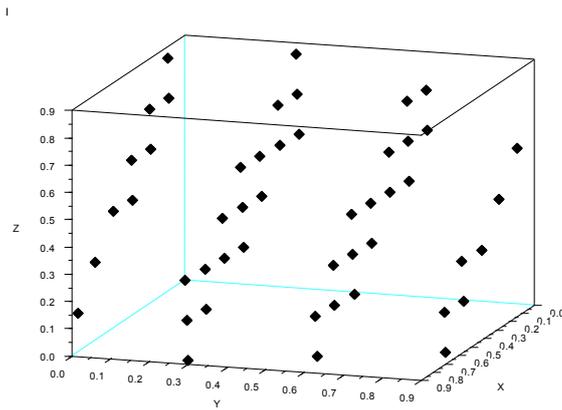


Figure 9. Une distribution de 49 points issue d'un tableau orthogonal de force 2 linéaire en dimension 3

Remarque. Un tableau orthogonal linéaire de force t , n'est autre qu'un tableau orthogonal de force t avec une structure de sous-espace vectoriel. Notamment, un tableau orthogonal linéaire de force 1 est toujours un hypercube latin.

Concernant la méthode de construction des tableaux orthogonaux linéaires, le lecteur pourra se référer au travail de thèse de Jourdan (2000). En observant les plans générés par cette méthode, nous avons pu remarquer qu'ils comportaient des problèmes d'alignements en plus de ceux que peuvent avoir les tableaux orthogonaux déjà précédemment cités. Les points sont situés sur des plans parallèles (voir Figure 9). Ce phénomène s'explique par le fait que nous travaillons dans \mathbf{Z}^q . Aussi, il est possible de trouver l'équation du plan sur lesquels les points sont situés. Dans l'exemple ci-dessus, le tableau orthogonal est linéaire de force 2 sur \mathbf{Z}^7 , son procédé de construction implique que les points vérifient la relation $x + 3y + z = 0 \pmod{7}$. Les points se trouvent alors sur 5 plans parallèles (4 d'entre-eux se distinguent bien sur la Figure 9, le cinquième est composé uniquement du point d'origine). Il est clair que, dans ce cas, la distribution des projections sur l'axe parallèle à ces plans ne sera pas satisfaisante.

1.2.3. Hypercubes latins basés sur les tableaux orthogonaux de force 2

Les tableaux orthogonaux sont couramment employés pour la planification d'expériences et leur succès est dû à leurs bonnes propriétés d'uniformité. Concernant les tableaux orthogonaux de force $t > 1$, cette uniformité n'est garantie que sur les sous-espaces de dimension t . Ainsi, les tableaux orthogonaux présentent des répétitions sur les axes factoriels.

Il semble alors préférable d'utiliser des hypercubes latins afin de garantir une bonne représentation des axes factoriels. Cependant, ces plans ne garantissent pas, en général, une distribution uniforme sur les sous-espaces de dimension $t > 1$. Pour résumer, aucune des 2 méthodes n'est complètement satisfaisante.

L'idée proposée par Tang (1993) pour remédier à ces problèmes est de combiner les propriétés d'orthogonalité d'un tableau orthogonal avec les bonnes propriétés de projection des hypercubes latins. Tang fournit un algorithme pour générer ces plans à partir de tableaux orthogonaux de force 2, ce qui permet de les faire hériter des bonnes propriétés d'uniformité sur les marges de dimension 1.

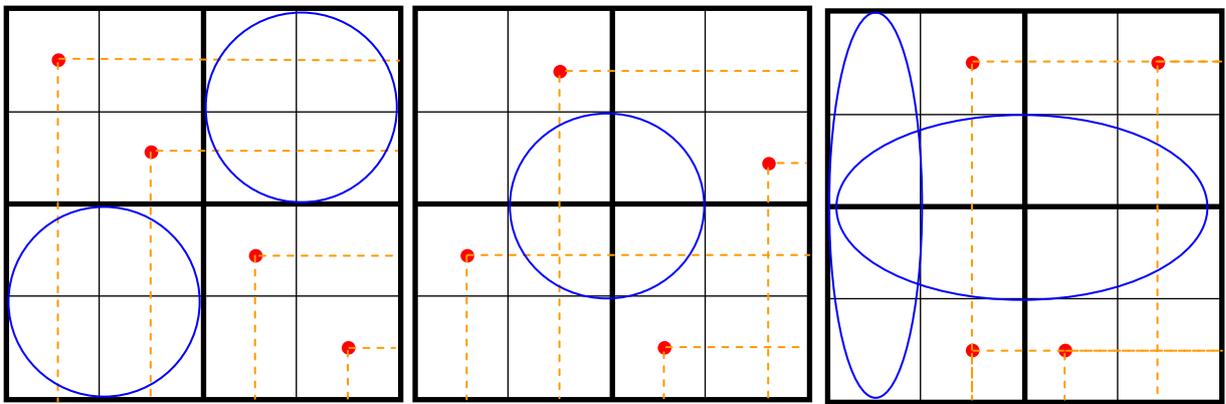


Figure 10. (De gauche à droite). Un hypercube latin, un hypercube latin de Tang et un tableau orthogonal de force 2 à 4 points en dimension 2

Proposition - Définition

Soit A un tableau orthogonal de type $OA_1(n, d, q, 2)$. Pour chaque colonne de A , on remplace chaque élément par une permutation de l'ensemble à q éléments suivant

$$[kq+1, kq+2, \dots, (k+1)q], \forall k=0, \dots, q-1.$$

pour obtenir un hypercube latin.

Par exemple, si $A = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}$ et que l'on considère les permutations suivantes

$0 \rightarrow 0$ puis $0 \rightarrow 1$ et $1 \rightarrow 3$ puis $1 \rightarrow 2$ pour chaque dimension (colonne de la matrice A),

alors on obtient l'hypercube latin $\begin{pmatrix} 0 & 1 & 3 & 2 \\ 0 & 3 & 1 & 2 \end{pmatrix}$.

Réciproquement, à partir de l'hypercube latin, il est possible de générer le tableau orthogonal de type $OA_1(n, d, q, 2)$ où les coefficients sont définis par

$$\lfloor q_{ij} / q \rfloor, i=1, \dots, n$$

avec $\lfloor . \rfloor$ qui désigne la partie entière inférieure.

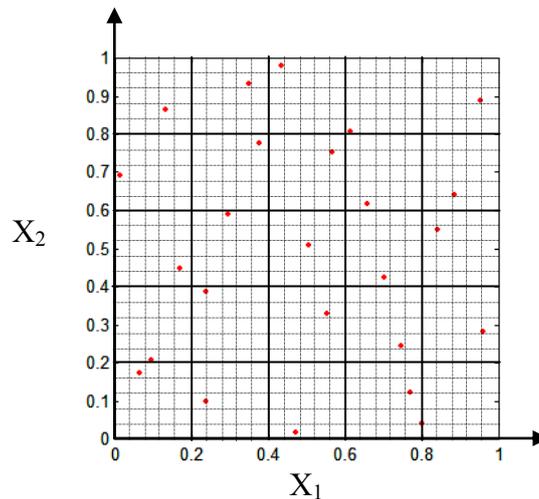


Figure 11. Un hypercube latin généré par un tableau orthogonal $OA_1(25, 5, 5, 2)$ dont les points sont randomisés et projetés sur le sous-espace (X_1, X_2)

Description. Cette technique est une méthode d'échantillonnage à trois degrés. Géométriquement, on effectue un premier échantillonnage du cube unité en q^d cellules. Parmi lesquelles, on en choisit $n = q^2$ de telle manière qu'elles forment un tableau orthogonal de force 2. On choisit une petite cellule à l'intérieur des n cellules présélectionnées de façon à ce qu'elles constituent un hypercube latin. A ce stade, nous avons donc un hypercube latin basé sur un tableau orthogonal. Il nous suffit alors de choisir un point au hasard dans chacune des petites cellules (voir Figure 11).

Discussion. Il est nécessaire de prendre des permutations distinctes pour les d colonnes afin d'éviter des alignements si les points sont placés au centre des cellules. Si les hypercubes latins sont générés à partir de tableaux orthogonaux linéaires, les alignements des tableaux orthogonaux seront conservés.

1.3. Suites à faible discrédance

Dans la partie précédente, nous avons discuté de plans dont les points sont bien répartis en projection mais pas forcément dans l'espace. Nous présentons ici des plans dont l'objectif est de bien remplir l'espace. Aussi, nous nous attachons à regarder leurs propriétés en projection.

Ces suites de points sont initialement utilisées pour remplacer les suites aléatoires dans la méthode de Monte-Carlo, d'où la dénomination de méthodes de quasi Monte-Carlo. La plupart de ces suites à faible discrédance (Thiérmard, 2000) sont générées par des algorithmes déterministes afin d'obtenir des points répartis le plus uniformément possible dans le domaine expérimental.

On donne ici au lecteur une définition simple et basique de la discrédance pour comprendre en quoi elle répond à nos attentes de remplissage de l'espace. La définition de Niederrieter (1987), donnée à la section 2.1, permet quant à elle de comprendre le principe sur lequel repose la discrédance. Le lecteur pourra également se reporter à cette section pour les différentes méthodes de calcul de la discrédance.

La discrédance mesure l'écart entre une distribution de points uniforme et une distribution de points donnée ; en d'autres termes elle permet de mesurer l'irrégularité de la distribution.

Si (dans le cas de la dimension 1) l'on considère la fonction de répartition empirique \hat{F}_n des points x^0, \dots, x^{n-1} , on peut définir la discrédance par

$$D_n(X) = \sup_{x \in [0,1]} |\hat{F}_n(x) - F_U(x)|$$

avec F_U la fonction de répartition de la loi uniforme sur $[0,1]$.

Remarque. $D_n(X)$ est la statistique de Kolmogorov-Smirnov permettant de tester l'adéquation à la loi uniforme.

Définition

Soit X un espace compact et μ une probabilité régulière sur la tribu borélienne de X . Une suite $(x_n)_{n \in \mathbb{N}}$ de points de X est uniformément distribuée si pour tout $f \in \mathcal{C}(X)$,

$$\frac{1}{n} \sum_{k=1}^n f(x_k) \xrightarrow{n \rightarrow \infty} \int_X f d\mu$$

Une propriété importante, ici, est la suivante.

Propriété

Une suite (x_n) est uniformément distribuée si $\lim_{n \rightarrow \infty} D_n(x) = 0$.

Il existe de nombreux résultats de majoration de la discrédance. Commençons par énoncer le résultat le plus fameux, connu sous le nom d'inégalité de Koksma-Hlawka.

Théorème (Hlawka, 1961)

Si f est une fonction à variation $V(f)$ bornée au sens de Hardy et Krause, alors pour toute suite de points x^1, \dots, x^n sur $[0, 1]^d$, on a

$$\left| \frac{1}{n} \sum_{i=1}^n f(x^i) - \int_{[0,1]^d} f(t) dt \right| \leq V(f) D_n(X).$$

Ainsi, l'erreur d'approximation précédente est, dans le pire des cas, égale au produit de la variation $V(f)$ (une grandeur qui ne reflète que l'irrégularité de la fonction f) et de la discrédance $D_n(X)$ (qui mesure uniquement la qualité de la répartition de la suite).

Remarque. Les suites discutées ici seront finies et indicées par convention (pour contenir l'origine du domaine) de 0 à $n-1$.

Il est possible de construire des suites pour lesquelles la discrédance est plus faible que celle d'une suite aléatoire égale à $1/\sqrt{n}$. On les appelle des **suites à faible discrédance**. Ces suites ont la particularité de remplir le cube unité uniformément et de manière extrêmement régulière.

L'idée naturelle venant à l'esprit pour avoir les points les plus uniformément répartis est de considérer une grille régulière. Cependant, il est possible de montrer que la discrédance d'une telle distribution est de l'ordre de $\frac{1}{n}$ ce qui est en fait un bien médiocre résultat. Afin d'en comprendre l'origine, nous y reviendrons au chapitre 2.1 consacré en particulier au calcul de la discrédance.

Aussi, nous verrons que la complexité du calcul de la discrédance est dépendant de la dimension et devient donc vite inenvisageable en grande dimension. C'est pourquoi, les méthodes de suites à faible discrédance proposées ici semblent intéressantes car elles sont simples à mettre en œuvre et garantissent une faible discrédance.

Des exemples de suites à faible discrédance sont données par Halton (1960), Hammersley (1960), Sobol' (1967), Faure (1982) et Niederreiter (1987). Dans les sections suivantes, nous étudierons la construction et les propriétés de ces différentes suites utilisées en planification d'expériences.

Une notion importante sur laquelle repose la construction de la plupart de ces suites est la définition de la fonction radicale inverse en base b donnée ci-dessous.

Toutes les suites données ci-après sont définies pour tout n . Nous verrons que la plupart de ces suites sont intéressantes vis-à-vis de l'aspect itératif. Par exemple, lorsque l'on souhaite rajouter q points à un plan de taille n . C'est notamment cet aspect qui va influencer le choix de la suite à utiliser.

Définition

Soit b un entier ≥ 2 . La fonction radicale inverse en base b est donnée par

$$\phi_b(i) = \frac{p_0}{b} + \frac{p_1}{b^2} + \dots + \frac{p_m}{b^{m+1}}$$

avec $i = p_0 + p_1b + \dots + p_mb^m$ et $m = \begin{cases} 1 + \lfloor \log_b(i) \rfloor & \text{si } i \neq 0 \\ 1 & \text{sinon} \end{cases}$.

La suite $C_b = \{x^0, x^1, \dots, x^{n-1}\}$ avec $x^i = \phi_b(i)$ est appelée suite de Van Der Corput en base b .

1.3.1. Suites de Halton

Les suites de Halton sont la version en dimension $d \geq 1$ des suites de Van Der Corput qui quant à elles constituent la version unidimensionnelle de ces suites. L'idée pour générer les suites de Halton consiste à utiliser une base différente pour chaque dimension.

Définition

Une suite de Halton $H_{b_1, \dots, b_d} = \{x^0, x^1, \dots, x^{n-1}\}$ en bases b_1, \dots, b_d est définie par

$$x^j = (\phi_{b_1}(i), \dots, \phi_{b_d}(i)) \in [0, 1]^d,$$

où b_1, \dots, b_d sont des entiers positifs premiers entre eux deux à deux.

Remarque. Afin de minimiser au mieux la discrédance, il convient de prendre pour les bases les d premiers nombres premiers. En effet, cela permet de minimiser la constante intervenant

dans le terme dominant de la majoration de la discrédance d'une telle suite (voir Faure, 1980 repris dans Niederreiter, 1992).

Ces suites ont l'avantage d'être faciles à implémenter et d'avoir un temps d'exécution très faible. Pour passer de $x^i = \phi_b(i)$ à $x^{i+1} = \phi_b(i+1)$ il suffit d'effectuer une addition, en base b , à droite de $\frac{1}{b}$.

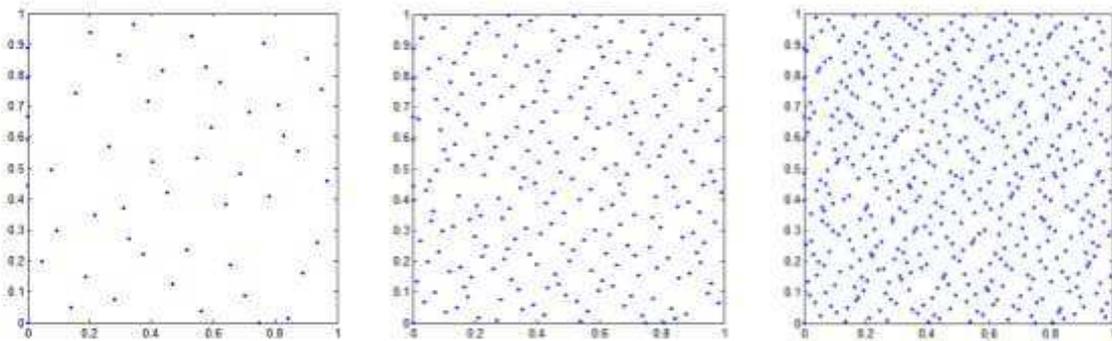


Figure 12. (De la gauche vers la droite) Les 50, 250 et 500 premiers points d'une suite de Halton en bases 2 et 3

Discussion. Malgré l'uniformité de la distribution dans le cube unité (Figure 12), ces suites présentent des pathologies en dimension élevée. En effet, lorsque l'on projette les points d'une telle suite dans le carré unité, l'allure de la structure est parfois décevante (Figure 13). Nous pouvons observer des diagonales décalées à chaque itération, qui laissent place à de nombreuses zones lacunaires dans le carré unité. Ce phénomène résulte de la régularité des sous-suites de longueur b des suites de Van der Corput. A noter cependant que cette pathologie ne se manifeste pas pour n'importe quelle paire de bases. En général, il est plus fréquent d'observer ce phénomène en dimension élevée et lorsque la différence entre les deux bases est très petite par rapport aux bases utilisées.

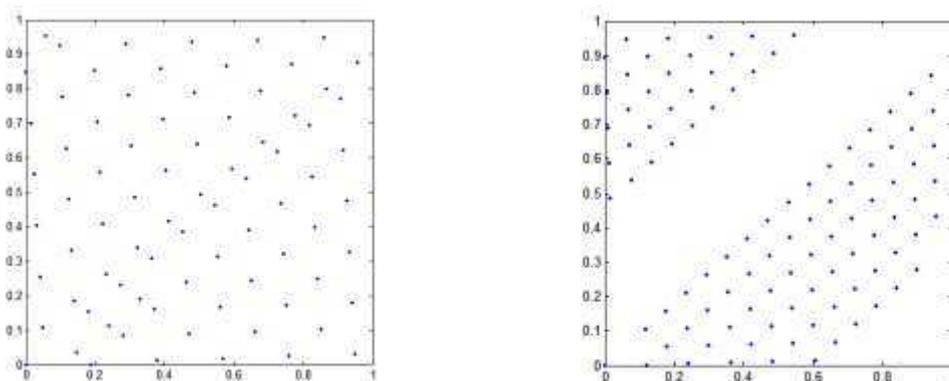


Figure 13. Des mauvaises projections en 2D. A gauche : les 80 premiers points d'une suite de Halton en bases 11 et 13. A droite : les 100 premiers points d'une suite de Halton en bases 17 et 19

Amélioration. Une solution pour corriger cette pathologie est de casser les corrélations entre les suites de Van Der Corput des différentes dimensions. Braaten et Weller (1979) ont amélioré l'uniformité des points d'une suite de Halton en appliquant sur chaque nombre premier b_j une permutation π_{b_j} sur $\{0, \dots, b_j - 1\}$ tel que $\pi_{b_j}(0) = 0$. Ainsi, à partir du nouveau développement de i en base b_j :

$$\varphi_{b_j}(i) = \frac{\pi_{b_j}(p_0)}{b_j} + \frac{\pi_{b_j}(p_1)}{b_j^2} + \dots + \frac{\pi_{b_j}(p_m)}{b_j^{m+1}},$$

on obtient une suite de Halton $Hg_{b_1, \dots, b_d} = \{x^0, x^1, \dots, x^{n-1}\}$ avec $x^i = (\varphi_{b_1}(i), \dots, \varphi_{b_d}(i))$ dite généralisée. La démarche est aussi dénommée comme une méthode de *scrambling*.

Ces suites permettent d'améliorer l'uniformité de la répartition mais ont des caractéristiques encore mal connues. Il existe à cet effet plusieurs méthodes de permutations sur les éléments des bases. Braaten et Weller (1979) ont amélioré les suites de Halton en prenant des permutations qui minimisent la discrédance de chaque suite de Van Der Corput en base b . Cette méthode ne permet pas d'avoir une seule permutation qui implique une meilleure discrédance. Ainsi une table de permutations est fournie jusqu'à la dimension 16. Tuffin (1998) s'est également intéressé au choix des permutations en particulier pour les dimensions plus élevées. Kocis et Whiten (1997) proposent aussi deux méthodes pour améliorer les suites de Halton.

1.3.2. Suites de Hammersley

Une suite de Hammersley en dimension d est construite à partir d'un terme dépendant du nombre de points et d'une suite de Halton en dimension $d-1$.

Définition

Une suite de Hammersley $H_{b_1, \dots, b_{d-1}}^n = \{x^0, \dots, x^{n-1}\}$ en bases b_1, \dots, b_{d-1} est définie par

$$x^i = \left(\frac{i}{n}, \phi_{b_1}(i), \dots, \phi_{b_{d-1}}(i) \right) \in I^d$$

où b_1, \dots, b_{d-1} sont des entiers positifs premiers entre eux deux à deux.

Remarque. Nous prendrons également le soin de prendre les $d-1$ premiers nombre premiers afin de réduire au mieux la discrédance de $H_{b_1, \dots, b_{d-1}}^n$.

Discussion. Ces suites étant construites à partir de suites de Halton, elles présentent le même phénomène de diagonales successives. De plus, il est impossible de rajouter des points supplémentaires à ces suites sans pour autant perturber la discrédance. Dans le cas où le nombre de points à générer n 'est pas connu à l'avance, il est déconseillé d'utiliser une suite de Hammersley. Aussi, ces suites ont perdu l'aspect itératif des suites de Halton qui permettait de rajouter des points facilement.

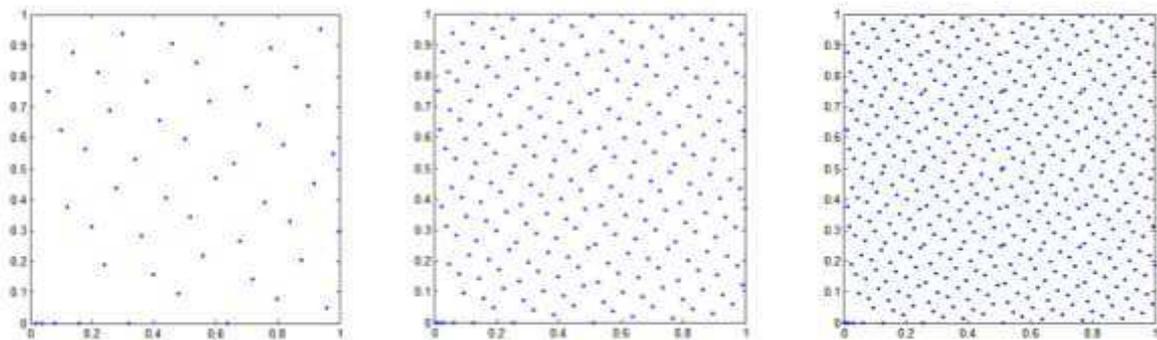


Figure 14. (De la gauche vers la droite) Les 50, 250 et 500 premiers points d'une suite de Hammersley en base 2

1.3.3. Suites de Sobol'

Elles sont définies à partir de polynômes primitifs sur $\mathbf{Z}_2 = \{0, 1\}$. On rappelle à cet effet la définition d'un polynôme primitif.

Définition

Un polynôme $p(t)$ de degré s de la forme $t^s + u_1 t^{s-1} + \dots + u_s$ est primitif sur le corps \mathbf{Z}_2 s'il est irréductible sur \mathbf{Z}_2 et si le plus petit entier i pour lequel il divise $t^i - 1$ (ou $t^i + 1$) est égal à $2^s - 1$.

La notion d'irréductibilité consiste à vérifier si le polynôme $p(t)$ n'est divisible par aucun autre polynôme de degré inférieur. Le plus petit entier positif i tel que $p(t)$ divise $t^i - 1$ est aussi appelé l'ordre du polynôme.

Un polynôme primitif de degré s doit posséder les monômes 1 et t^s , ainsi qu'un nombre impair de termes.

Définition

Une suite de Sobol' $S = \{x^0, x^1, \dots\}$, en dimension 1 est définie comme suit

$$x^i = \frac{1}{2^m} \left(\bigoplus_{k=1}^m p_k l_k \right),$$

où (p_1, \dots) est la représentation binaire de i , $m = \begin{cases} 1 & \text{si } i = 0 \\ 1 + \lfloor \log_2 i \rfloor & \text{sinon} \end{cases}$ et où le

symbole \oplus désigne simplement l'addition dans \mathbf{Z}_2 ⁸.

Les l_k avec $k > s$ sont obtenus en utilisant la relation de récurrence suivante

$$l_k = 2u_1 l_{k-1} \oplus 2^2 u_2 l_{k-2} \oplus \dots \oplus l_{s-1} l_{i-s+1} \oplus (2^s l_{i-s} \oplus l_{i-s})$$

où les u_k sont les coefficients d'un polynôme $t^s + u_1 t^{s-1} + \dots + u_s$ primitif de degré s sur \mathbf{Z}_2 et les nombres l_1, \dots, l_s doivent être des entiers impairs tels que $1 \leq l_k \leq 2^k$ pour $k = 1, \dots, s$.

⁸ L'addition modulo 2 se réalise par un « ou exclusif ».

Afin de construire une suite de Sobol' en dimension d , il suffit de choisir d polynômes primitifs distincts.

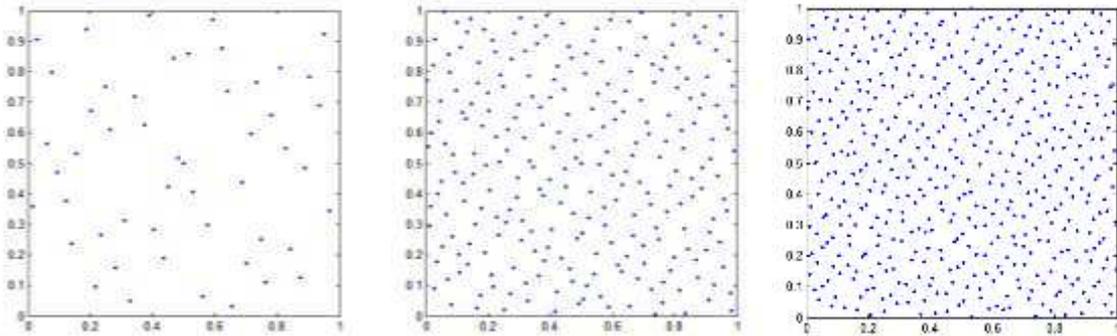


Figure 15. Les 50, 250 et 500 premiers points d'une suite de Sobol' en dimension 2

Discussion. Ces suites ont l'avantage d'être rapides à construire car la nature binaire des ordinateurs permet de réduire le temps de calcul. Elles ont aussi l'avantage de préserver, généralement, l'uniformité de la distribution lorsque la dimension augmente.

Cependant, en raison de la nature arbitraire des états initiaux l_1, \dots, l_s de ces récurrences, les points sont souvent mal répartis dans l'hypercube unité (voir Figure 16). Seulement, on ne peut difficilement prédire les états initiaux qui donnent des distributions non uniformes. Un autre inconvénient apparaît si l'on souhaite rajouter de nouveaux points à une distribution. Dans ce cas, on observe qu'ils tombent dans le voisinage de ceux produits précédemment. Augmenter le nombre de points n'est donc pas une solution envisageable.

Amélioration. Sobol' et Levitan (1976) proposent une table de valeurs jusqu'en dimension 16 de manière à avoir des distributions de points plus uniformes. Une solution pour conjurer ce problème est d'utiliser des méthodes de scrambling.

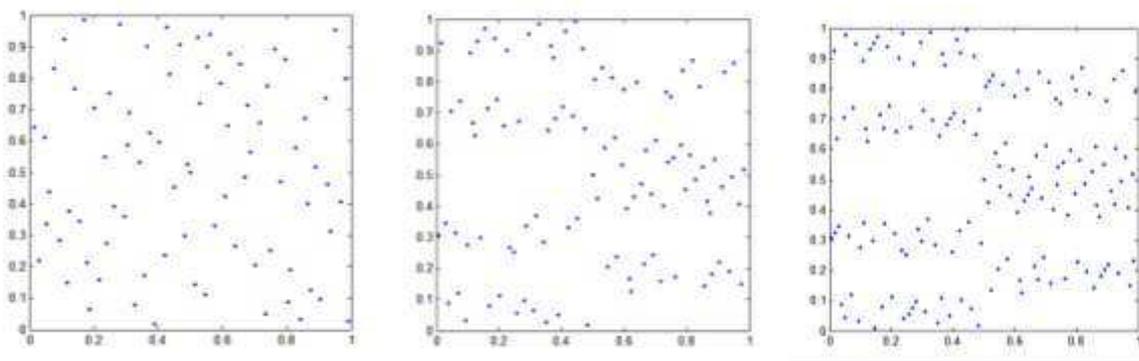


Figure 16. A gauche : les 80 premiers points d'une suite de Sobol' en dimension 6 en projection sur le sous-espace des dimensions 5 et 6. A droite : les 100 et 150 premiers points d'une suite de Sobol' en dimension 8 sur le sous-espace des dimensions 7 et 8

1.3.4. Suites de Faure

Les suites de Faure sont définies à partir de la fonction radicale inverse, ϕ_b , et une matrice génératrice de Pascal $C = (c_{kl})$ donnée par

$$c_{kl} = \binom{l-1}{k-1} = \begin{cases} \frac{(l-1)!}{(k-1)!(l-k)!} & \text{si } k \leq l, \\ 0 & \text{sinon,} \end{cases} \quad \forall l, k \in \mathbb{N}^*.$$

Définition

Soit $b \geq d$ un nombre premier. La suite de Faure $F = \{x^0, \dots, x^{n-1}\}$ en dimension d est définie comme suit

$$x_j^i = \phi_b(C^{j-1}i)$$

où $C^{j-1} = \binom{l-1}{k-1} (j-1)^{l-k} \text{ mod } b$, est la matrice génératrice de la j -ième dimension d'une suite de Faure en dimension d .

Remarque. Il est recommandé afin d'obtenir une distribution uniforme de choisir comme base b le plus petit nombre premier supérieur ou égal à d .

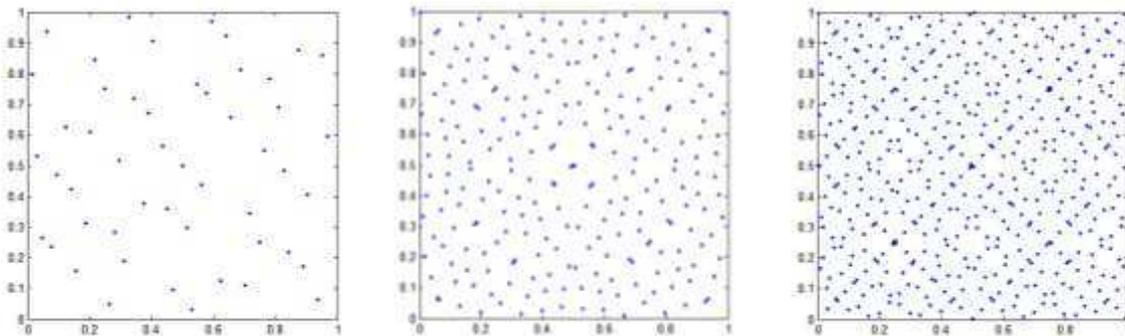


Figure 17. Les 50, 250 et 500 premiers points d'une suite de Faure en dimension 2

Discussion. Les suites de Faure permettent de générer une distribution uniforme localement (Figure 17). Cependant, les projections sur le carré unitaire des points de l'hypercube sont échantillonnées en bandes (Figure 18). Similairement aux suites de Sobol', nous pouvons remarquer que les nouveaux points sont placés dans le voisinage de ceux produits précédemment. Les techniques de scrambling permettent également d'amoindrir ces phénomènes.

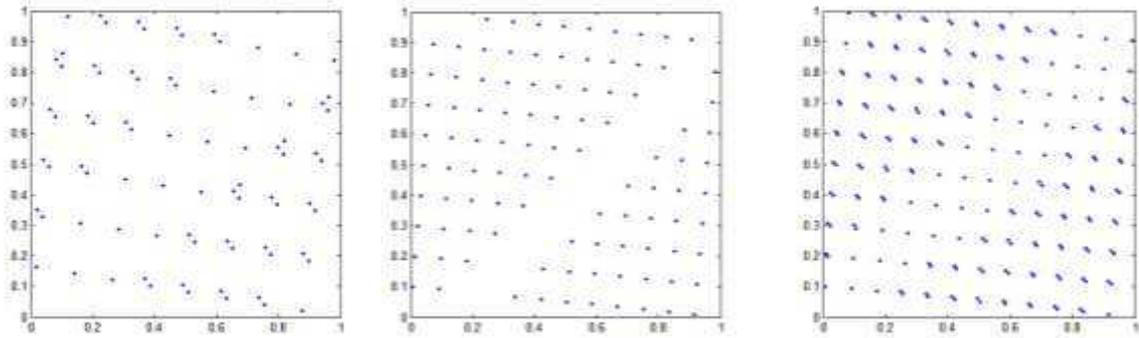


Figure 18. Les 80 premiers points d'une suite de Faure en dimension 6, les 100 et 200 premiers points d'une suite de Faure en dimension 8, projetés sur le carré unité défini par les deux dernières dimensions

1.3.5. Good Lattice Points

Les *good lattice points* sont la forme la plus simple des méthodes de réseaux (voir Sloan et Joe, 1994). Les *good lattice points* peuvent être considérés comme des suites à faible discrédance qui présentent des propriétés analogues à celles des grilles régulières telles que l'espacement régulier entre les points.

Définition

P est un « *good lattice points* » s'il est défini par

$$P = \left(\left\{ \frac{i-0.5}{n} \right\}, \left\{ \frac{g_2 i - 0.5}{n} \right\}, \dots, \left\{ \frac{g_d i - 0.5}{n} \right\} \right), j = 1, \dots, d \text{ et } i = 1, \dots, n$$

où $\{x\} = x - \lfloor x \rfloor$ est la partie fractionnaire de x , $(n; g_1, \dots, g_d)$ un vecteur d'entiers satisfaisant $1 \leq g_j < n$, $g_k \neq g_j$ pour $k \neq j$, et $d < n$ avec $g_1 = 1$.

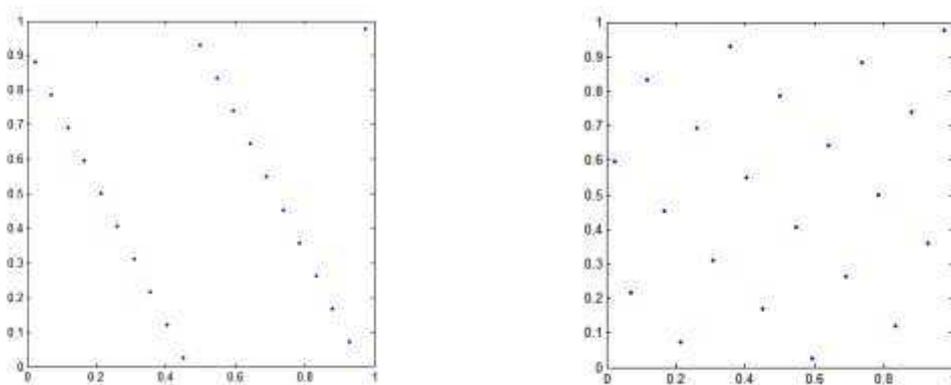


Figure 19. Les 21 premiers termes d'un good lattice points en dimension 2, avec des générateurs différents

Le terme « good » se réfère à un choix judicieux de n et des g_j . En effet, les générateurs sont déterminés de façon à minimiser la discrédance comme le montre les figures ci-dessus. Les tables de générateurs sont disponibles dans les ouvrages de Wang et Fang (1981) ou Sloan et Joe (1994). Les figures ci-dessus représentent un *good lattice points* en dimension 2 avec $n = 21$ points. Sur la figure de gauche, nous avons pris des générateurs quelconques $g_1 = 1$ et $g_2 = 19$. Tandis que sur la figure de droite, les générateurs ont été choisis à l'aide de la table trouvée dans Fang *et al.* (1994) et valent $g_1 = 1$ et $g_2 = 13$.

Un *good lattice points* est en réalité un hypercube latin qui présente une faible discrédance. Ceci a pour avantage d'exclure les hypercubes latins dont les points sont mal répartis comme par exemple ceux dont les points se situent sur la diagonale. Cependant, cela n'évite pas les alignements comme l'illustre la Figure 19 où les points sont positionnés sur 6 droites parallèles.

L'avantage des *good lattice points* est donc leur régularité qui constitue également un inconvénient à cause des alignements obtenus. Une autre difficulté est celle d'ajouter des points à la distribution préexistante.

1.3.6. Suites et réseaux

Les réseaux ont l'avantage de pouvoir combiner une faible discrédance avec des bonnes propriétés en projection sur certains sous-espaces. Nous verrons, après une brève introduction de ces concepts, que beaucoup de suites à faible discrédance étudiées ci-dessus peuvent être considérées comme des (t,m,d)-réseaux et des (t,d)-suites.

Définition

Un pavé élémentaire en base $b \geq 2$ est de la forme

$$P = \prod_{j=1}^d \left[\frac{a_j}{b^{q_j}}, \frac{a_j + 1}{b^{q_j}} \right]$$

où $q_j \geq 0$, $0 \leq a_j < b^{q_j}$ et a_j, q_j sont entiers.

Un pavé élémentaire E est un pavé du cube unité $[0,1]^d$ dont le j -ième axe a pour longueur $1/b^{q_j}$. Si on divise le j -ième axe en b^{q_j} segments égaux et qu'on répète cette

division sur les autres axes, le pavé obtenu est un pavé élémentaire ayant pour volume

$$b^{-\sum_{j=1}^d q_j}.$$

Définition

Soit $0 \leq t \leq m$ des entiers. Un (t, m, d) -réseau en base b est une suite de b^m points de $[0, 1]^d$ tel que chaque pavé élémentaire en base b de volume b^{t-m} contient exactement b^t points de la suite.

Remarque. Si $t = m$, alors tous les points de la suite sont dans $[0, 1]^d$.

Définition

Une suite de points $\{x^0, x^1, \dots, x^{n-1}\} \in [0, 1]^d$ est une (t, d) -suite en base b si pour tout $k \geq 0$ et $m \geq t$, la suite $\{x^{kb^m}, \dots, x^{(k+1)b^m-1}\}$ est un (t, m, d) -réseau en base b .

Théorème (Niederreiter, 1987)

Il n'existe aucune suite $(0, d)$ -suite en base b avec $b < d$.

Exemple. Etudions une $(0, 2)$ -suite résultante d'un $(0, 2, 2)$ -réseau en base 3 avec $n = 3^2 = 9$ points. La théorie des réseaux garantit que chaque pavé élémentaire en base 3 de volume 3^{-2} contient exactement un point de la suite. Les pavés sont de la forme

$$\left[\frac{a_1}{3^{q_1}}, \frac{a_1+1}{3^{q_1}} \right] \times \left[\frac{a_2}{3^{q_2}}, \frac{a_2+1}{3^{q_2}} \right] \text{ dans } [0, 1]^2 \text{ avec les entiers } a_j, q_j \text{ tel que } q_j \geq 0 \text{ et}$$

$0 \leq a_j < 3^{q_j}$ pour $j = 1, 2$. Le volume d'un pavé élémentaire étant $b^{-\sum_{j=1}^d q_j}$, les pavés élémentaires de volume b^{-m} doivent donc vérifier $q_1 + q_2 = 2$. Comme q_1 et q_2 sont des entiers positifs, on a pour $q = (q_1, q_2)$ les 3 possibilités suivantes $q = (0, 2)$, $q = (2, 0)$ et $q = (1, 1)$ où $q = (q_1, q_2)$ qui sont représentées sur les figures ci-dessous.

Si $q = (0, 2)$ alors $\begin{cases} b^{q_1} = 1 \\ b^{q_2} = 9 \end{cases} \Rightarrow \begin{cases} a_1 = 0 \\ a_2 \in \{0, \dots, 8\} \end{cases}$ car les a_j doivent vérifier $0 \leq a_j < b^{q_j}$.

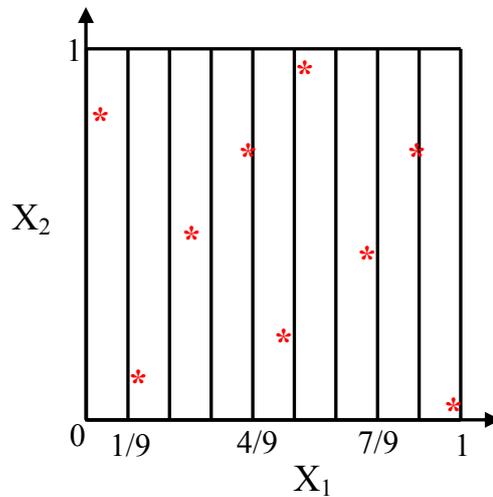


Figure 20. Une première caractérisation possible des pavés élémentaires pour 9 points

On s'aperçoit alors que la distribution a de bonnes propriétés d'uniformité en projection sur le facteur X_1 .

Si $q = (2, 0)$ alors $\begin{cases} b^{q_1} = 9 \\ b^{q_2} = 1 \end{cases} \Rightarrow \begin{cases} a_1 \in \{0, \dots, 8\} \\ a_2 = 0 \end{cases}$ car les a_j doivent vérifier $0 \leq a_j < b^{q_j}$.

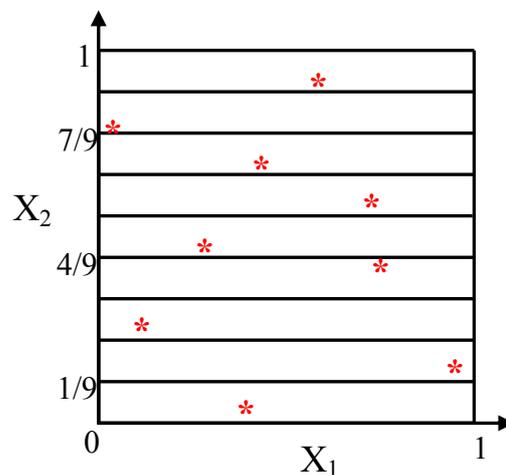


Figure 21. Une seconde caractérisation possible des pavés élémentaires pour 9 points

Dans ce cas, la distribution a de bonnes propriétés d'uniformité en projection sur le facteur X_2 .

Si $q = (1, 1)$ alors $\begin{cases} b^{q_1} = 3 \\ b^{q_2} = 3 \end{cases} \Rightarrow \begin{cases} a_1 \in \{0, 1, 2\} \\ a_2 \in \{0, 1, 2\} \end{cases}$ car les a_j doivent vérifier $0 \leq a_j < b^{q_j}$.

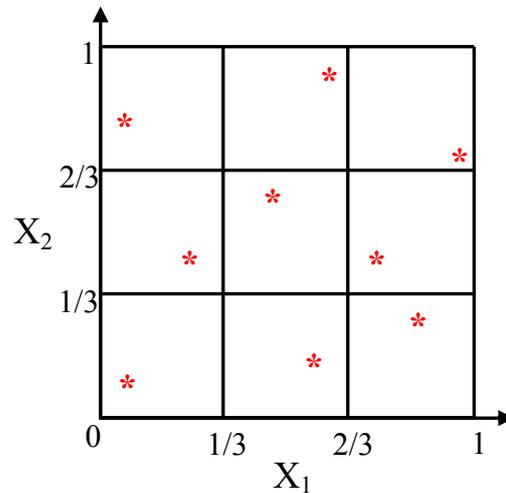


Figure 22. Une troisième caractérisation possible des pavés élémentaires pour 9 points

On retrouve alors un tableau orthogonal de force 2 à 3 niveaux, ce qui implique l'uniformité de la distribution dans le carré unité mais n'apporte pas d'intérêt vis-à-vis des projections sur les sous-espaces, ici de dimension 1.

Un réseau permet donc d'obtenir une distribution uniforme en projection et ce quelle que soit la dimension de la projection. Cette propriété semble indiquer une bonne répartition des points dans le cube unité. On remarque de plus, que pour m et b fixés, plus la valeur de t est petite, meilleures seront les propriétés d'uniformité d'un (t,m,d) -réseau en base b . Cette intuition est confirmée par la proposition suivante concernant la discrédance des réseaux.

Proposition (Niederreiter, 1987)

Pour $b \geq d$ un (t,m,d) -réseau a une discrédance de l'ordre de $O\left(n^{-1}(\log n)^{d-1}\right)$ et une (t,d) -suite a une discrédance de l'ordre de $O\left(n^{-1}(\log n)^d\right)$.

Avec les concepts énoncés plus haut, il est possible de faire le lien avec des méthodes d'échantillonnage étudiées précédemment. Un hypercube latin de taille b est un $(0,1,d)$ -réseau en base b . En effet, chaque pavé élémentaire de volume $1/b$ contient exactement 1 point de la suite. Un hypercube latin basé sur un tableau orthogonal $OA(b^2, d, b, d)$ de taille $n = b^2$ est un $(0,2,d)$ -réseau en base b .

On distingue également un lien avec les suites étudiées précédemment. Les suites de Van Der Corput en base b sont des $(0,1)$ -suites en base b . Les suites de Sobol' sont des (t,d) -suites

en base 2, Sobol' (1967) fournit les valeurs de t qu'il faut utiliser pour une dimension donnée. Les suites de Faure en base b sont des $(0,d)$ -suites en base b . A noter que les suites de Halton et Hammersley, étant construites à partir de plusieurs bases, ne peuvent être des (t,d) -suites.

Il existe d'autres (t,d) -suites. Nous pouvons citer par exemple, les suites de Niederreiter (1987) qui sont des $(0,d)$ -suites définies pour toute base $b \geq d$, où b est une puissance d'un nombre premier. Des tables contenant les meilleurs paramètres connus pour les (t,d) -suites et les (t,m,d) -réseaux en base $b = 2, 3, 5$ et 10 ont été construites par Mullen, Mahalanabis et Niederreiter (1995).

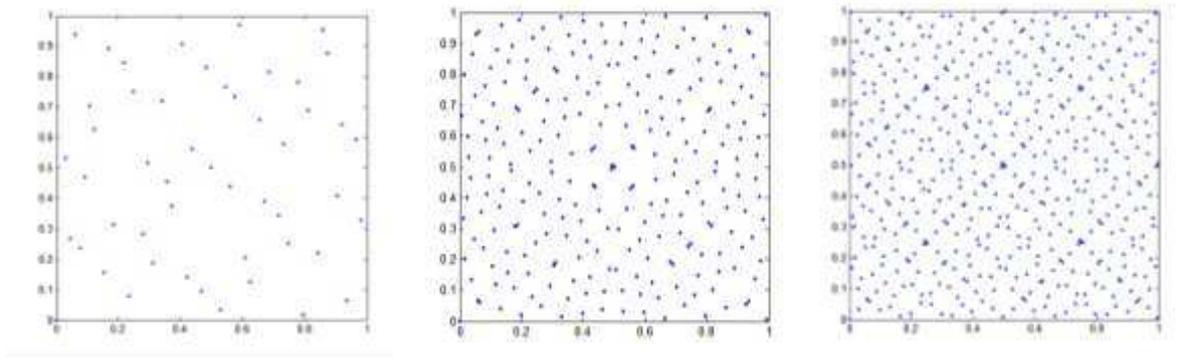


Figure 23. Les 50, 250 et 500 premiers points d'une suite de Niederreiter en dimension 2

On peut remarquer que les suites de Niederreiter sont similaires aux suites de Faure en dimension 2. Cependant, les suites de Niederreiter sont bien meilleures lorsque l'on projette les points dans le carré unité. En effet, les points sont répartis plus uniformément et lorsqu'il y a ajout de nouveaux points, ceux-ci sont placés de façon à remplir l'espace le plus uniformément possible et non pas dans le voisinage des points placés précédemment comme c'est le cas pour les suites de Faure.

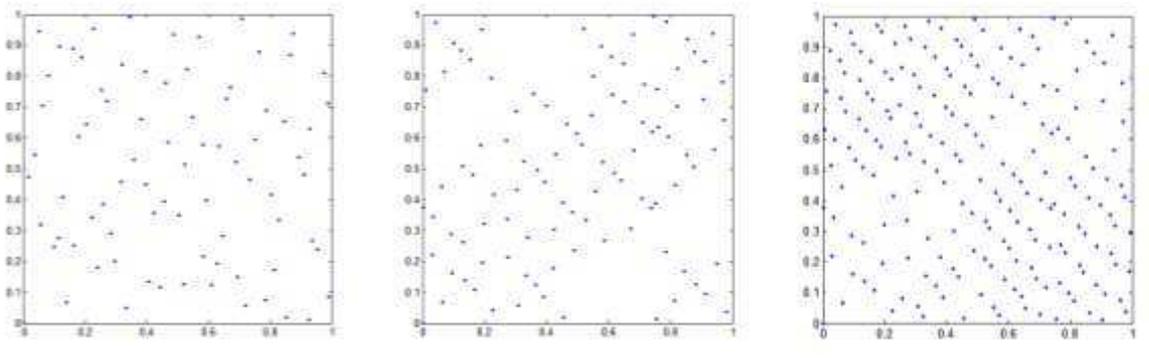


Figure 24. Les 80 premiers points d'une suite de Niederreiter en dimension 6, les 100 et 200 premiers points d'une suite de Niederreiter en dimension 8, projetés sur le carré unité défini par les deux dernières dimensions

Un avantage des (t,m,d) -réseaux est qu'ils permettent d'obtenir une répartition uniforme sur les sous-espaces de dimension s pour $1 \leq s \leq m$. Cette propriété devient plus forte quand le

nombre de points de la suite croît c'est-à-dire lorsque m augmente. Ceci implique de pouvoir rajouter des points sans pour autant détruire l'uniformité de la suite contrairement à d'autres techniques d'échantillonnage. Afin de rajouter des points en conservant l'uniformité, il suffit d'augmenter le paramètre m . Par ce biais, ce sont les contraintes d'uniformité en projection qui entraînent une répartition uniforme des points dans le domaine expérimental de dimension d . Cependant, selon les bases choisies, on peut observer des corrélations entre les points des pavés élémentaires successifs qui forment une distribution régulière et ne sont pas uniformément distribués dans ces intervalles.

Pour plus de détails sur les (t,d) -suites et (t,m,d) -réseaux, le lecteur peut se référer à Niederreiter (1992). Owen (1997) explique les limites des (t,d) -suites en très grande dimension.

1.4. Récapitulatif

Le tableau ci-dessous permet de récapituler les propriétés des plans étudiés au sein de ce chapitre. Chaque type de plan est critiqué en regardant si :

- les distributions obtenues permettent ou non de remplir l'hypercube uniformément,
- les distributions, notamment en projection dans le carré unité, suivent des alignements,
- les distributions sont séquentielles c'est-à-dire s'il est possible de rajouter des points à cette distribution sans pour autant détruire l'uniformité initiale,
- à partir des distributions obtenues, les distributions en dimension inférieure ont une meilleure uniformité.

	Remplissage de l'espace	Motif	Séquentialité	Réduction de dimension
Aléatoires	Non : lacunes.	Non	Oui	Oui
Doehlert	Oui	Oui	Oui	Non
Grilles régulières	Oui	Oui	Non	Non
Hypercubes latins	Oui, en version centrée. Sinon, lacunes.	Oui, en version centrée	Non	Oui, en dimension 1
Hypercubes latins de Tang	Oui	Non	Non	Oui
Tableaux orthogonaux	Oui, en version centrée. Sinon, lacunes.	Oui	Non	Oui, en dimension égale à la force du tableau.
Good lattice points	Oui, si les paramètres sont bien choisis.	Oui	Oui	Non
Halton	Oui	Oui, en dimension élevée.	Oui	Non
Hammersley	Oui	Oui, en dimension élevée.	Oui	Non
Faure	Oui	Oui	Oui	Non
Niederreiter	Oui	Non	Oui	Non
Sobol'	Oui	Non, si les états initiaux sont bien choisis.	Oui	Non

Tableau 3. Récapitulatif des propriétés de certains plans d'expériences

Chapitre 2

Critères et plans optimaux

Etudier l'uniformité d'une distribution de points est un travail fastidieux, voire même impossible en grande dimension. Il est donc utile de se référer à des critères afin de pouvoir décider si une distribution est uniforme ou non ou/et si elle vérifie également le bon remplissage de l'espace. Il existe pour cela des critères *intrinsèques* aux plans d'expériences à savoir des critères qui ne préjugent en rien de la qualité de la surface de réponse déterminée par la suite à partir de ces plans.

Aussi, nous discutons dans un premier temps les critères de discrédance qui permettent de mesurer si la distribution donnée est proche d'une distribution uniforme.

Ensuite, nous introduisons la classe des critères de distance qui comparent une distribution donnée à une grille régulière.

Enfin, nous présentons le critère d'entropie qui mesure la quantité d'information contenue dans le plan. Contrairement aux deux précédents, ce critère n'est pas intrinsèque car il dépend d'un modèle statistique. Nous verrons cependant qu'il est lié à la distance entre les points et assure ainsi une bonne distribution.

Dans chacune des parties, nous discutons de plans optimaux au sens de ces critères, choisis pour leur bonne qualité d'uniformité ou/et de remplissage. Aussi, on y retrouve les plans à faible discrédance, des plans améliorés à l'aide de critères de distance, et enfin des plans où l'on maximise l'entropie associée.

2.1. Critères d'uniformité basés sur la discrédance et plans à faible discrédance

2.1.1. La discrédance

La définition de base a été donnée à la section 1.3. On rappelle que la discrédance mesure l'écart entre une distribution de points donnée et une distribution de points uniforme ; en d'autres termes, elle mesure l'irrégularité de la distribution. On donne ci-dessous la définition permettant de comprendre le principe sur lequel elle est basée.

Remarque. Si on reparamètre le domaine, par exemple si on veut étudier $(X_1)^2$ au lieu de X_1 , alors l'objectif ne sera pas exactement le même. En effet, assurer l'uniformité du facteur X_1 n'implique pas l'uniformité pour le facteur $(X_1)^2$. Le même problème apparaît lorsque l'on veut définir une loi *a priori* informative en bayésien, où dans ce cas l'uniformité de la distribution ne sera pas l'objectif recherché.

Définition (Niederreiter, 1987)

Soit X la suite de n points x^1, \dots, x^n dans $[0, 1]^d$ et soit J un sous-ensemble de $[0, 1]^d$,

avec les notations précédentes on pose : $D_n(J, X) = \frac{A(J, X)}{n} - \lambda_d(J)$

où $A(J, X)$ est le nombre d'indices i , $1 \leq i \leq n$, tel que $x^i \in J$ et $\lambda_d(J)$ est la mesure de Lebesgue (ou volume) de J .

La **discrédance extrême** de X , notée $D_n(X)$ est définie par $D_n(X) = \sup_{J \in \mathfrak{S}} |D_n(J, X)|$

où \mathfrak{S} est l'ensemble des parties de $[0, 1]^d$ de la forme $\prod_{i=1}^d [a_i, b_i]$.

La **discrédance à l'origine** de X , notée $D_n^*(X)$ est définie par $D_n^*(X) = \sup_{J \in \mathfrak{S}^*} |D_n(J, X)|$

où \mathfrak{S}^* est l'ensemble des parties de $[0, 1]^d$ de la forme $\prod_{i=1}^d [0, b_i]$.

Remarque. Soit $\mu_X = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ la probabilité uniforme sur X , alors $D_n(J, X)$ s'écrit comme

la distance entre celle-ci et la mesure de Lebesgue soit $D_n(J, X) = \mu_X(J) - \lambda_d(J)$.

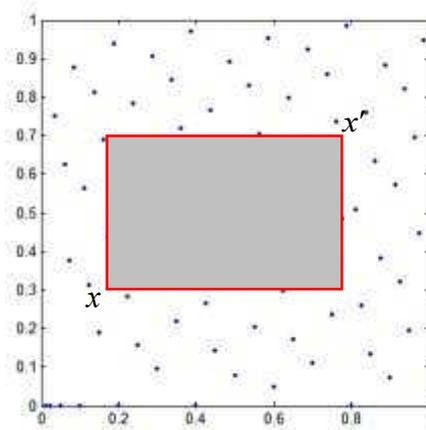


Figure 25. Les 80 premiers points d'une suite de Hammersley en dimension 2 avec un sous-ensemble J défini par x et x' pour le calcul de la discrédance extrême

Pour $x = (0.2, 0.3)$ et $x' = (0.8, 0.7)$, le volume du rectangle J est égal à 0.24, et la proportion de points dans J est $18/80 = 0.225$. On a donc $D_n(J, X) = 0.015$. En effectuant le même procédé sur plusieurs sous-ensembles J et en prenant la valeur maximale acquise, on obtient la discrédance extrême de la suite X .

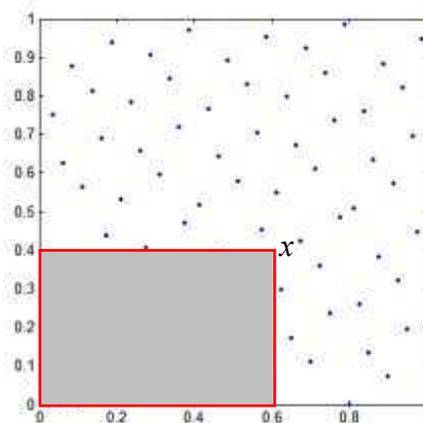


Figure 26. Un sous-ensemble J pour le calcul de la discrédance à l'origine

$D_n(X)$ et $D_n^*(X)$ sont des discrédances en norme L_∞ . Les discrédances en norme L_2 définissent une nouvelle mesure de la non uniformité. Elles seront données dans la section 2.1.3.

Maintenant que nous venons de voir la définition exacte de la discrédance, il devient possible d'expliquer les mauvais résultats de discrédance d'une grille. Pour en comprendre l'origine, il faut se rendre compte que pour avoir une faible discrédance, il faut échantillonner le plus uniformément possible tous les rectangles parallèles aux axes. Ainsi pour une grille, il y a des rectangles dont l'échantillonnage n'est pas bon. Par exemple, il peut y avoir des problèmes selon le placement des bords par rapport à la grille.

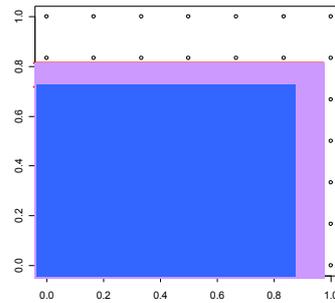


Figure 27. Plan factoriel et visualisation de deux rectangles parallèles aux axes n'étant pas échantillonnés uniformément

2.1.2. Approches calculatoires de la discrédance en norme infinie

Les valeurs de la discrédance sont faciles à calculer en dimension 1 et 2. Le calcul de la discrédance devient très difficile en grande dimension.

2.1.2.1. En dimension 1

La discrédance d'une suite de n points ne dépend pas de l'ordre des points.

Théorème (Niederreiter, 1972)

Soit X une suite de n points x^1, \dots, x^n avec $0 \leq x^1 \leq \dots \leq x^n \leq 1$, alors

$$D_n^*(X) = \frac{1}{2n} + \max_{1 \leq i \leq n} \left| x_i - \frac{2i+1}{2n} \right|.$$

Remarque. Ce théorème est facile à prouver et même plus ancien si l'on se réfère au test de Kolmogorv-Smirnov.

Ce théorème montre que la meilleure suite uniformément répartie en dimension 1 est la suite de n points x^1, \dots, x^n des points milieux avec $x^i = \frac{2i+1}{2n}$. Ainsi, pour toute suite X de n points dans l'intervalle $[0, 1]$, on a $D_n^*(X) \geq \frac{1}{2n}$.

2.1.2.2. En dimension 2

Bundschuh et Zhu (1993) ont généralisé une formule établie par De Clerck (1986) concernant la discrédance des suites de n points x^1, \dots, x^n sur $[0, 1]^2$ dans le cas où les composantes ne sont pas nécessairement distinctes. Voici le résultat qu'ils ont établi.

Théorème

Soit une suite x^1, \dots, x^n dans $[0, 1]^2$ dont les n points ont été préalablement triés dans l'ordre croissant de leur première composante : $0 \leq x_1^1 \leq \dots \leq x_1^n \leq 1$.

On se donne la paire de points auxiliaires $x^0 = (0, 0)$ et $x^{n+1} = (1, 1)$. Pour chaque indice $i \in \{0, \dots, n\}$, on note $\{\xi_0^i, \dots, \xi_{i+1}^i\}$ la séquence obtenue en réordonnant le sous-ensemble de secondes composantes $\{x_2^0, \dots, x_2^i, x_2^{n+1}\}$ de manière à satisfaire

$$0 = \xi_0^i \leq \xi_1^i \leq \dots \leq \xi_i^i \leq \xi_{i+1}^i = 1.$$

Avec ces notations, on a

$$D_n^*(X) = \max_{0 \leq i \leq n} \max_{0 \leq k \leq i} \max \left\{ \left| \frac{k}{n} - x_1^i \xi_k^i \right|, \left| \frac{k}{n} - x_1^{i+1} \xi_{k+1}^i \right| \right\}.$$

2.1.2.3. En dimension quelconque

Le calcul de la discrédance est très difficile. La complexité des algorithmes permettant de la calculer croît exponentiellement avec la dimension. Ainsi, la formule de Bundschuh et Zhu (1993) vue pour le cas de la dimension 2 reste exploitable uniquement pour les petites dimensions. Dobkin et Eppstein (1993) présentent une première théorie permettant de calculer la discrédance exacte à partir d'un échantillonnage de l'espace.

En norme L_∞ , il est d'usage de considérer une méthode de calcul approché de la discrédance. La plus simple et la plus fréquente d'entre elles consiste à considérer plusieurs

orthants aléatoirement dans l'hypercube unité et de leur appliquer la définition de Niederreiter (1987) vue auparavant. Par souci d'exactitude, il semble intéressant d'étudier l'écart relatif entre la valeur de la discrédance obtenue avec la méthode de Bundshu et Zhu et la valeur de la discrédance approchée en fonction du nombre d'orthants que l'on choisit.

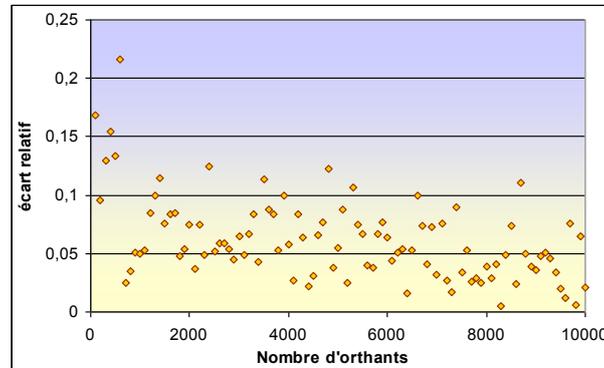


Figure 28. L'écart relatif entre la valeur exacte et approchée de la discrédance en fonction du nombre d'orthants sur un plan pseudo aléatoire à 100 points en dimension 2

On en déduit qu'afin d'avoir la meilleure précision possible, il est préférable de choisir un nombre d'orthants assez grand. Il devient normal de se demander si le temps de calcul va rester compétitif même en grande dimension. C'est pourquoi, il nous a semblé intéressant d'étudier le temps d'exécution du calcul de la discrédance en fonction de la dimension. Pour ce faire on a pris des plans pseudo aléatoires à 10 points par dimension et un nombre d'orthants égal à 10000 pour avoir la meilleure précision possible.

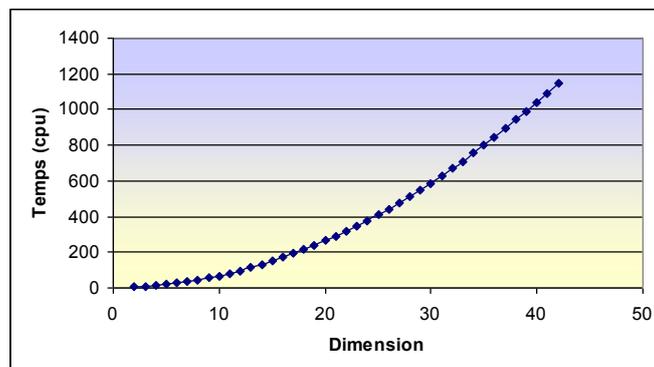


Figure 29. Le temps d'exécution du calcul approché de la discrédance en fonction de la dimension

Une bonne précision de la discrédance en grande dimension ne peut être obtenue en un temps de calcul raisonnable avec les méthodes approximatives ; même si la Figure 29 ne nous montre pas que la complexité de l'algorithme est exponentielle en la dimension.

2.1.3. Discrédance en norme 2

Seule la discrédance en norme L_2 est facilement calculable quelle que soit la dimension.

En voici plusieurs. On note X la suite de n points x^1, \dots, x^n sur $[0, 1]^d$. Cette section présente dans un premier temps la définition des différentes discrédances en norme 2 ainsi que leurs approches calculatoires reprises d'Hickernell (1998).

2.1.3.1. Discrédance extrême et à l'origine

Définition

La discrédance L_2 d'une suite de n points x^1, \dots, x^n sur $[0, 1]^d$ est définie par

$$DL2_n(X)^2 = \int_{[0, 1]^{2d}} (D_n(J, X))^2 da db$$

où J est l'ensemble des parties de $[0, 1]^d$ de la forme $\prod_{i=1}^d [a_i, b_i]$.

Définition

La discrédance L_2 à l'origine d'une suite de n points x^1, \dots, x^n sur $[0, 1]^d$ est définie par

$$DL2_n^*(X)^2 = \int_{[0, 1]^d} (D_n(J, X))^2 db$$

où J est l'ensemble des parties de $[0, 1]^d$ de la forme $\prod_{i=1}^d [0, b_i]$.

En dimension d , les calculs de $DL2_n$ et $DL2_n^*$ peuvent se faire par les formules explicites suivantes :

$$DL2_n(X)^2 = \left(\frac{1}{12}\right)^d - \frac{2^{1-d}}{n} \sum_{i=1}^n \prod_{j=1}^d x_j^i (1 - x_j^i) + \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \prod_{j=1}^d (1 - \max(x_j^i, x_j^k)) \min(x_j^i, x_j^k)$$

et

$$DL2_n^*(X)^2 = \left(\frac{1}{3}\right)^d - \frac{2^{1-d}}{n} \sum_{i=1}^n \prod_{j=1}^d (1 - (x_j^i)^2) + \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \prod_{j=1}^d (1 - \max(x_j^i, x_j^k)).$$

2.1.3.2. Discr ance modifi e

D finition

La discr ance L_2 modifi e d'une suite de n points x^1, \dots, x^n sur $[0, 1]^d$ est d finie par

$$DL2_n^M(X)^2 = \sum_{u \neq \emptyset} \int_{[0, 1]_u} (D_p(J_u, X))^2 db_u$$

avec $[0, 1]_u$ la projection de l'hypercube unit  sur les composantes u  tant un sous-ensemble de $\{1, \dots, d\}$, $p = \text{Card}(u)$ et J_u est la projection du sous-ensemble J de la forme

$$\prod_{i=1}^d [0, b_i].$$

La discr ance modifi e consid re les projections sur tous les sous-espaces et est d finie par

$$DL2_n^M(X)^2 = \left(\frac{4}{3}\right)^d - \frac{2^{1-d}}{n} \sum_{i=1}^n \prod_{j=1}^d \left(3 - (x_j^i)^2\right) + \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \prod_{j=1}^d \left(2 - \max(x_j^i, x_j^k)\right).$$

2.1.3.3. Discr ance centr e

D finition

La discr ance L_2 centr e d'une suite de n points x^1, \dots, x^n sur $[0, 1]^d$ est d finie par

$$DL2_n^C(X)^2 = \sum_{u \neq \emptyset} \int_{[0, 1]_u} (D_p(J_u, X))^2 dx_u$$

avec $[0, 1]_u$ la projection de l'hypercube unit  sur les composantes u  tant un sous-ensemble de $\{1, \dots, d\}$, $p = \text{Card}(u)$ et o  J_u est la projection du sous-ensemble construit   partir du point consid r  x et de son sommet le plus proche.

Remarque. En dimension 2, l'ensemble J_u peut  tre de 4 formes diff rentes dont un est repr sent  sur la Figure 30. Plus g n ralement, en dimension d , il existe 2^d cas.

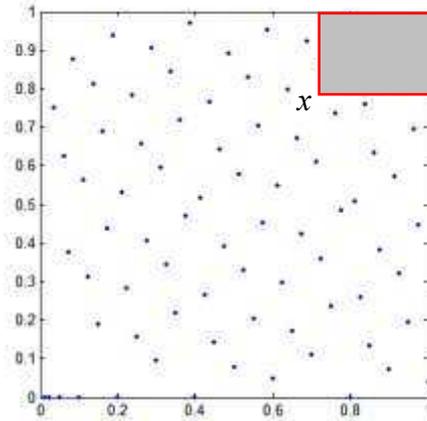


Figure 30. Un sous-ensemble J pour le calcul de la discrédance centrée

Hickernell (1998) donne une expression analytique de la discrédance centrée :

$$DL2_n^C(X)^2 = \left(\frac{13}{12}\right)^2 - \frac{2}{n} \sum_{i=1}^n \prod_{j=1}^d \left(1 + \frac{1}{2}|x_j^i - 0.5| - \frac{1}{2}|x_j^i - 0.5|^2\right) + \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \prod_{j=1}^d \left(1 + \frac{1}{2}|x_j^i - 0.5| + \frac{1}{2}|x_j^k - 0.5| - \frac{1}{2}|x_j^i - x_j^k|\right).$$

2.1.3.4. Discrédance symétrique

Définition

La discrédance L_2 symétrique d'une suite de n points x^1, \dots, x^n sur $[0, 1]^d$ est définie par

$$DL2_n^S(X)^2 = \sum_{u \neq 0} \int_{[0, 1]_u} (D_p(J_u, X))^2 dx_u$$

avec J_u la projection de l'intervalle J sur le sous-espace défini par les composantes u

où J est la réunion des sous-ensembles symétriques i.e. dont la somme des coordonnées des sommets sont égales.

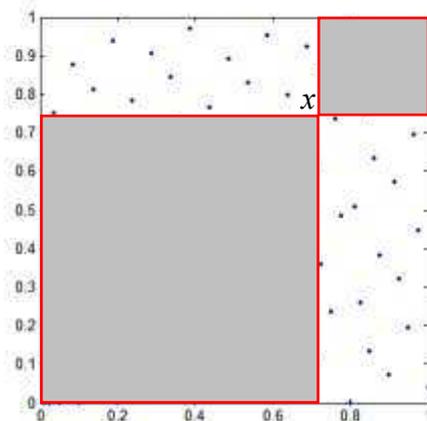


Figure 31. Pour $x = (0.7, 0.75)$, le volume total des 2 sous-ensembles J est 0.6 et la proportion de points totale est $49/80=0.6125$. La différence entre ces deux grandeurs est donc 0.0125

On dispose également d'une formule analytique pour calculer cette discrédance :

$$DL2_n^S(X)^2 = \left(\frac{4}{3}\right)^d - \frac{2}{n} \sum_{i=1}^n \prod_{j=1}^d \left(1 + 2x_j^i - 2(x_j^i)^2\right) + \frac{2^d}{n^2} \sum_{i=1}^n \sum_{k=1}^n \prod_{j=1}^d \left(1 - |x_j^i - x_j^k|\right).$$

2.1.3.5. Discrédance « enveloppée »

Définition

La discrédance L_2 « enveloppée » d'une suite de n points x^1, \dots, x^n sur $[0, 1]^d$ est définie par

$$DL2_n^E(X)^2 = \sum_{u \neq 0} \int_{[0, 1]^u} (D_n(J_u, X))^2 dx_u$$

avec J_u la projection de J sur le sous-espace défini par les composantes u

$$\text{où } J = \bigotimes_{j=1}^d K(x_j, x'_j) \text{ avec } K(x_j, x'_j) = \begin{cases} [x_j, x'_j] \text{ si } x_j \leq x'_j \\ [0, x'_j] \cup [x_j, 1] \text{ sinon} \end{cases}$$

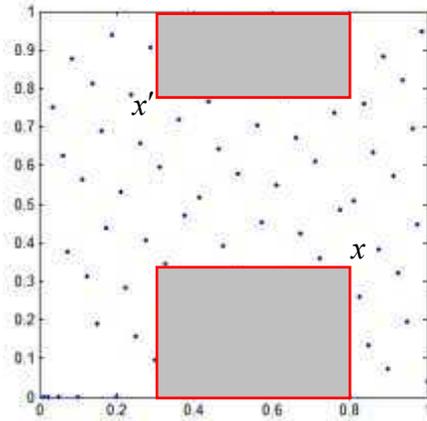


Figure 32. Pour $x = (0.8, 0.35)$ et $x' = (0.3, 0.8)$, le volume des 2 enveloppes est égal à 0.275, et la proportion de points totale est $21/80=0.26$. La différence obtenue entre ces deux grandeurs est donc 0.015

Pleming et Manteufel (2005) introduisent la discrédance « enveloppée » comme étant la plus sensible pour détecter les zones lacunaires en projection sur les marges de dimension 1. Ils l'utilisent pour mesurer l'uniformité des hypercubes latins réitérés. Il existe également une formule analytique pour la calculer :

$$DL2_n^E(X)^2 = \left[-\left(\frac{4}{3}\right)^d + \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \prod_{j=1}^d \left(\frac{3}{2} - |x_j^i - x_j^k| \left(1 - |x_j^i - x_j^k|\right)\right) \right] \times n.$$

2.1.4. Exemples

Quelle que soit la discrédance, plus la valeur calculée pour un plan donné est faible et plus les points du plan seront distribués uniformément.

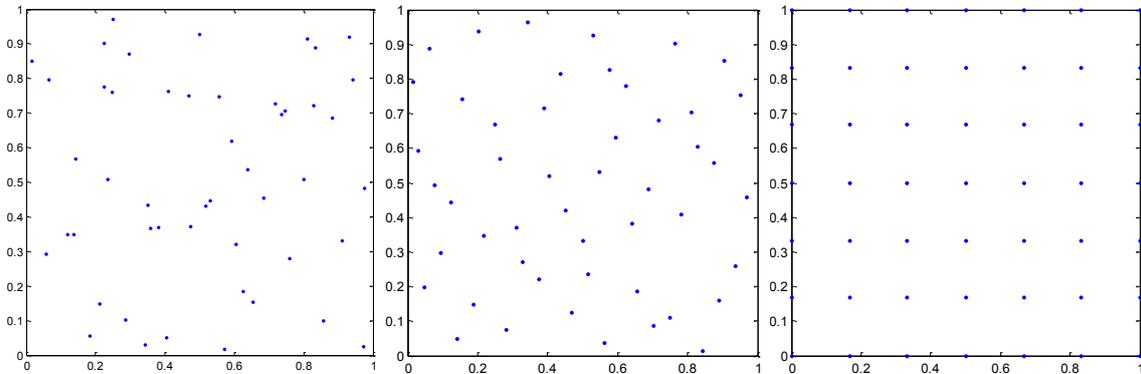


Figure 33. De la gauche vers la droite. Un plan aléatoire, une suite de Halton, et un plan factoriel complet à 49 points en dimension 2

	Aléatoire	Halton	Factoriel
$n=16$	0.26	0.17	0.44
$n=25$	0.21	0.14	0.36
$n=36$	0.26	0.10	0.31
$n=49$	0.12	0.08	0.27

Tableau 4. Discrédances en norme infinie pour $n=16, 25, 36$ et 49 pour les 3 types de plans suivants : aléatoires, suite de Halton, plan factoriel en dimension 2

Le tableau ci-dessus montre que pour le plan factoriel, on obtient les discrédances les plus mauvaises. Ceci implique que la discrédance seule ne suffit pas à mesurer la qualité d'un plan car nous ne souhaitons pas uniquement mesurer l'uniformité des points mais aussi le remplissage de l'espace. A noter que cette remarque est valable pour n'importe quel nombre de points.

Ainsi, il est nécessaire de choisir plusieurs critères afin de mesurer la bonne répartition des points dans l'hypercube unité. Ce que permettent entre autre les mesures basées sur les critères de distances présentés dans la section suivante.

2.1.5. Plans à faible discrédance

La notion de plans à faible discrédance décrits ici est différente de celle des suites détaillées dans la section 1.3 pour les raisons mentionnées ci-dessous.

Pour commencer, précisons la définition que l'on donne aux plans à faible discrédance. Nous appellerons plans à faible discrédance des plans **stochastiques** obtenus à l'aide d'un simple algorithme d'échange qui privilégie les configurations minimisant la discrédance.

Il existe de nombreux résultats de majoration de la discrédance. A la section 1.3, nous avons énoncés l'inégalité de Koksma-Hlawka. Il existe aussi des majorations pour chacune des suites discutées précédemment. Par exemple, Faure (1980) montre que la discrédance à l'origine d'une suite de Halton à n points en dimension d générée à partir des bases b_1, \dots, b_d est majorée par

$$\frac{d}{n} + \frac{1}{n} \prod_{j=1}^d \frac{b_j - 1}{2 \log b_j} \log n + \frac{b_j + 1}{2}.$$

Le théorème de Faure (1981) ci-dessous montre qu'il est possible de diminuer la discrédance ne serait ce qu'en généralisant les suites c'est-à-dire en affectant des permutations aux éléments de la suite. L'idée d'obtenir des plans à faible discrédance à l'aide d'un simple algorithme d'échange ne paraît pas être alors complètement irréaliste.

Algorithme 1. Algorithme pour générer un plan à faible discrédance.

Initialisation : choisir un vecteur de n points $X_{(0)}$ au hasard et fixer une discrédance en norme L_2

Calcul de la discrédance du plan associé : $disc_{L_2}(X_{(0)})$

POUR $k = 1$ à N_{Max}

Choisir une expérience i au hasard dans $\{1, \dots, n\}$

Simuler z^i uniformément dans $[0, 1]^d$

Accepter de remplacer x^i par z^i si $disc_{L_2}(X_{(k)}) > disc_{L_2}(X_{(k-1)})$

Fin pour

Théorème. (Faure, 1981)

Si X est la suite de Van Der Corput généralisée en base $b=12$ engendrée par la permutation

$$\sigma_{12} = (0\ 5\ 9\ 3\ 7\ 1\ 10\ 4\ 8\ 2\ 6\ 11),$$

on obtient

$$\limsup \frac{n D_n^*(X)}{\log n} \approx 0.224.$$

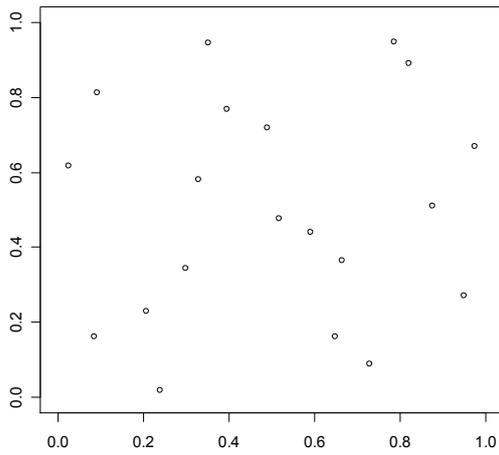
L'intérêt de discuter de ces plans à faible discrédance est double. D'une part, nous avons le souhait de montrer qu'il est possible d'obtenir des plans à plus faible discrédance que celle obtenue pour la majorité des suites présentées dans la section 1.3. D'autre part, ces plans non déterministes, par nature, ont le mérite de ne pas hériter des mauvaises propriétés (en grande dimension) de projections des suites causées par leur régularité.

La génération de ces plans nécessitant le calcul de la discrédance, ils sont construits uniquement à partir de la discrédance en norme L_2 , beaucoup plus simple à calculer que la discrédance en norme L_∞ . Il devient alors possible d'envisager des plans basés sur n'importe laquelle des discrédances définies au paragraphe 2.1.3. Ce qui signifie également la possibilité de construire des plans où l'on considère les points en projection sur les marges en envisageant la discrédance centrée ou modifiée. Aussi, les plans discutés par la suite seront générés à l'aide de la discrédance centrée.

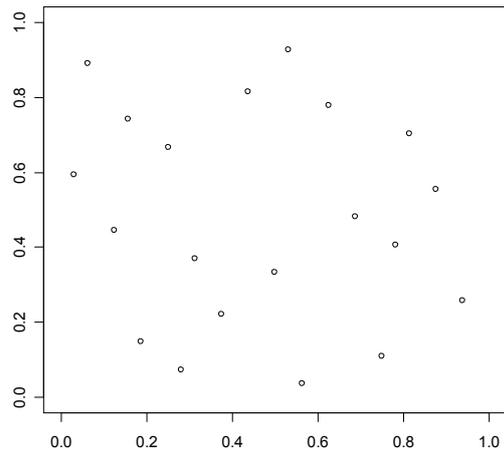
L'inconvénient de ces plans, outre le fait qu'ils ne peuvent être basés uniquement sur des discrédances en norme L_2 , est leur temps d'exécution. En effet, même si le calcul des discrédances en norme L_2 est obtenu à l'aide de formules analytiques simples, les suites à faible discrédance sont très rapides à mettre en œuvre. Par exemple, générer une suite de Halton à 600 points en dimension 60 est immédiat alors qu'un plan à faible discrédance de cet ordre de grandeur est inenvisageable.

Maintenant, vérifions sur quelques exemples, que ces plans permettent de réduire considérablement la discrédance annoncée des suites à faible discrédance.

Sur l'exemple ci-dessous, nous avons généré un plan à 20 points en dimension 2 à l'aide de l'algorithme d'échange qui nous donne une discrédance de 0.041 alors que la discrédance pour la suite de Halton de même caractéristique est égale à 0.06.



Discrédance centrée $L_2 = 0.041$

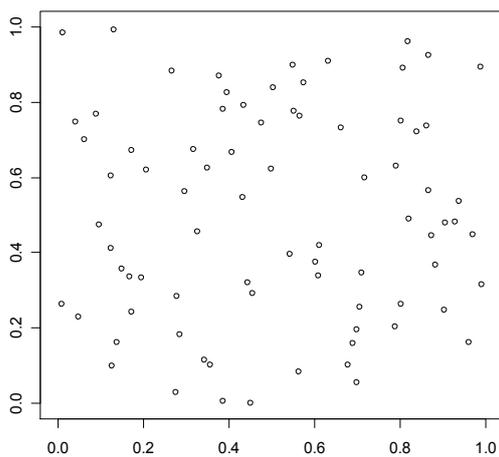


Discrédance centrée $L_2 = 0.06$

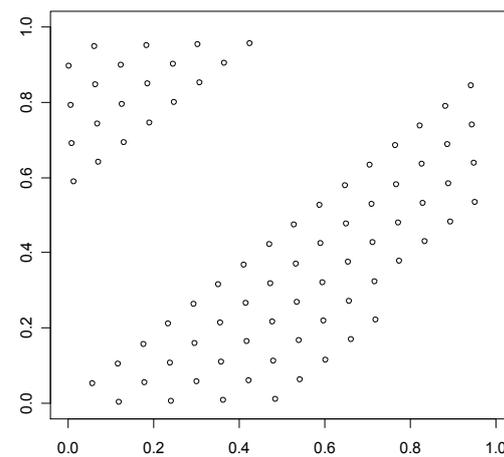
Figure 34. A gauche un plan à faible discrédance et à droite une suite de Halton à 20 points en 2D

Une autre possibilité est d'initialiser l'algorithme d'échange avec une suite à faible discrédance, au lieu de choisir une distribution aléatoire, afin de diminuer plus rapidement la discrédance. Par exemple, si l'initialisation est faite à partir d'une suite de Halton, il est possible de diviser par 2 la discrédance en moitié moins d'itérations.

Regardons ce qu'il en est maintenant en grande dimension et plus particulièrement au regard des projections. L'exemple ci-dessous montre les projections d'un plan à faible discrédance et d'une suite de Halton sur les facteurs (X_7, X_8) .



Discrédance centrée $L_2 = 0.12$



Discrédance centrée $L_2 = 0.15$

Figure 35. Représentation des projections sur les facteurs (X_7, X_8) d'un plan à faible discrédance sur la figure de gauche et d'une suite de Halton sur la figure de droite à 80 points en 8D

A titre de comparaison, la discr pance centr e L_2 d'une distribution al atoire est en moyenne (sur 1000 plans   80 points en dimension 8)  gale   0.22. Nous avons donc l'assurance ici d'obtenir des distributions de meilleure qualit  qu'un plan al atoire.

Le caract re stochastique de ces plans est le d faut de leur qualit . En effet, ces plans permettent d' viter les  ventuels probl mes num riques que l'on peut avoir avec des plans   structure r guli re. Cependant, g n rer un plan en tr s grande dimension (par exemple en dimension 60) devient vite inenvisageable   cause du temps de calcul.

2.2. Critères de distance et plans optimaux

L'objectif des critères basés sur la distance entre les points est de quantifier la proximité entre une distribution de points donnée et celle d'une grille régulière. Nous avons choisi de discuter uniquement des critères d'uniformité les plus courants basés sur la distance entre les points voisins. Pour des compléments sur d'éventuelles mesures d'uniformité le lecteur peut se référer à Gunzburger (2004).

L'idée ici est de générer des plans dont les points sont proches d'une grille régulière mais pas exactement ceux d'une grille à cause des mauvais alignements que cela peut engendrer. On rappelle que nous recherchons des plans avec des distributions quasi-périodiques c'est-à-dire qui sont un bon compromis entre une grille régulière et une bonne uniformité (dictée par exemple avec le critère de discrédance).

On note $\text{dist}(x^i, x^k)$ la distance euclidienne entre deux points x^i et x^k telle que :

$$\text{dist}(x^i, x^k) = \left[\sum_{j=1}^d (x_j^i - x_j^k)^2 \right]^{1/2}.$$

2.2.1. Mesure de recouvrement

Définition

Soit $X = \{x^1, \dots, x^n\} \subset [0, 1]^d$ une suite de n points en d dimensions.

On définit la mesure de recouvrement λ par

$$\lambda = \frac{1}{\bar{\gamma}} \left(\frac{1}{n} \sum_{i=1}^n (\gamma_i - \bar{\gamma})^2 \right)^{1/2}$$

avec $\gamma_i = \min_{k \neq i} \text{dist}(x^i, x^k)$ désignant la distance minimale entre le point x^i et les autres

points de la suite et $\bar{\gamma} = \frac{1}{n} \sum_{i=1}^n \gamma_i$ la moyenne des γ_i .

Si les points sont sur une grille régulière, on a $\gamma_1 = \dots = \bar{\gamma}$ et donc $\lambda = 0$.

Ainsi, plus λ est faible, et plus les points sont proches de ceux d'une grille régulière. Cette expression fait apparaître explicitement le coefficient de variation de l'échantillon des γ_i , à savoir le rapport de l'écart-type avec la moyenne.

2.2.2. Le rapport de distances

Définition

Soit $X = \{x^1, \dots, x^n\} \subset [0, 1]^d$ une distribution de n points en d dimensions.

On définit le rapport de distance par

$$R = \frac{\max_{i=1, \dots, n} \gamma_i}{\min_{i=1, \dots, n} \gamma_i}$$

avec $\gamma_i = \min_{k \neq i} \text{dist}(x^i, x^k)$.

Dans le cas où les points sont sur une grille régulière, on a $\gamma_1 = \dots$ et donc $R = 1$. Ainsi, plus R est proche de 1, et plus les points sont proches de ceux d'une grille régulière.

2.2.3. Distances Maximin et Minimax

Johnson *et al.* (1990) ont introduit les distances maximin et minimax afin de construire des plans répondant à la question de remplissage de l'espace.

Définition (Chen, 2003)

On définit les critères suivants au sens de la distance euclidienne

$$\text{MinDist} = \min_{x_i \in X} \min_{\substack{x_k \in X \\ k \neq i}} \text{dist}(x^i, x^k)$$

et

$$\text{AvgDist} = \frac{1}{n} \sum_{x_i \in X} \min_{\substack{x_k \in X \\ k \neq i}} \text{dist}(x^i, x^k)$$

où $\mathbf{X} = \{\mathbf{x}^1, \dots, \mathbf{x}^n\}$ est le plan d'expériences à n points en d dimensions.

Maximiser le critère *MinDist* revient à considérer le critère de distance **maximin**. Si la valeur de *MinDist* est grande, alors il n'y a aucun point qui soit proche au sens de cette grandeur. Ainsi, plus les critères *MinDist* et *AvgDist* seront grands et plus cela traduira une distribution uniforme.

Définition

On définit les critères suivants au sens de la distance euclidienne :

$$MaxDist(D, \mathbf{X}) = \sup_{x_i \in D} \min_{x_k \in \mathbf{X}} dist(x^i, x^k)$$

et

$$EDist(D, \mathbf{X}) = \int_D \min_{x_k \in \mathbf{X}} dist(x^i, x^k) dx^i$$

où D est un ensemble de points du domaine expérimental et $\mathbf{X} = \{\mathbf{x}^1, \dots, \mathbf{x}^n\}$ le plan d'expériences à n points en dimension d .

Remarque. L'ensemble D considéré dans la définition est généralement $D = [0, 1]^d$. Ainsi, le supremum et l'intégrale ci-avant sont définies sur des ensembles infinis de points. Cependant, afin de calculer ces 2 expressions simplement, nous considérons un échantillonnage fini de points dans D , à savoir un plan factoriel ou un échantillonnage de ce type.

Minimiser le critère *MaxDist* revient à considérer le critère de distance **minimax**. Le principe du critère minimax est de mesurer la distance maximale entre un point du domaine expérimental n'appartenant pas au plan, qui varie, et les points du plan d'expériences. Afin de calculer les critères, on prend généralement pour les points de D une grille régulière. Aussi, en grande dimension il n'est plus envisageable de considérer ce critère sans prendre pour l'ensemble D un plan aléatoire.

Le critère *EDist* correspond au critère de dispersion introduit par Niederreiter (1992).

2.2.4. Exemples

Considérons à présent la valeur des critères basés sur les mesures de distance pour les différents plans de l'exemple précédent (voir Tableau 4 et Figure 33)

<i>n</i>	Aléatoire				Halton				Factoriel			
	λ	R	MinDist	MaxDist	λ	R	MinDist	MaxDist	λ	R	MinDist	MaxDist
16	0.5	7.56	0.049	0.43	0.17	1.81	0.16	0.35	3.10^{-8}	1	0.33	0.23
25	0.43	4.03	0.049	0.32	0.17	1.80	0.12	0.27	0	1	0.25	0.17
36	0.68	19.8	0.012	0.29	0.19	2.06	0.097	0.18	3.10^{-8}	1	0.2	0.14
49	0.58	11.14	0.016	0.19	0.22	2.81	0.053	0.17	2.10^{-6}	1	0.17	0.11

Tableau 5. Valeurs des critères de distance pour les 3 types de plans suivants : aléatoires, suite de Halton, plan factoriel où $n=16, 25, 36$ et 49

L'ensemble D considéré pour calculer le critère $MaxDist$ est un plan factoriel à 30 niveaux ; ce qui donne dans ce cas une bonne interprétabilité à ce critère. A noter que l'ensemble D doit être formé de beaucoup plus de points que l'ensemble de points testé pour rendre compte du positionnement des points et que ces points ne peuvent pas être simplement choisis de manière aléatoire. Ce qui s'avère très difficile à assurer en grande dimension. C'est une des principales raisons qui explique que dans la suite ce critère ne sera pas étudié.

Les résultats du tableau ci-dessus sont bien ceux attendus, à savoir que les plans factoriels ressortent comme étant de meilleure qualité au regard de ces critères.

Conclusions. Il est nécessaire de considérer les deux catégories de critères – discrédance et distance – pour déterminer si la distribution d'un plan remplit bien l'espace ou non. Cependant, ce sont surtout des critères de comparaison et non de qualification de plans. C'est pourquoi, nous introduisons dans le Chapitre 6 un critère permettant de qualifier un plan. Le Chapitre 5, quant à lui, présentera un outil de visualisation de l'uniformité des plans d'expériences en projection sur les sous-espaces.

2.2.5. Algorithmes d'amélioration de critères de distance

En définissant une métrique sur les ensembles de points, il est possible d'utiliser les distances pour calculer un critère de remplissage de l'espace permettant de choisir un sous-ensemble de points.

Dans cette partie, nous présentons deux algorithmes. L'un, d'échange, qui est basé sur le critère minimax et l'autre, basé sur un processus de mort. Ce dernier, comparativement au premier, reste peu cité dans la littérature et semble pourtant être plus robuste au niveau du remplissage de l'espace en grande dimension.

Le principe de base de ces algorithmes consiste à déterminer un sous-ensemble de points, à partir d'un grand ensemble de candidats, qui améliore un critère géométrique de remplissage de l'espace. L'avantage de ces méthodes réside dans le fait qu'elles permettent de générer des plans avec un critère spécifique supérieur à ce que l'on obtient généralement.

2.2.5.1. Plans minimax

Johnson *et al.* (1990) ont établi non seulement le critère minimax vu ci-avant mais aussi les plans construits sur ce critère. Ces plans sont étudiés ici car leur principe de construction est par nature moins simple que celui des plans maximin non discutés dans ce mémoire. Ces plans sont aussi référencés dans le logiciel R sous le nom de « cover designs » en tant que Space-Filling Designs.

Dans ce type de procédé, les points de l'ensemble candidat ne servent pas seulement pour les points du plan mais permettent aussi de déterminer le critère de remplissage de l'espace afin d'évaluer la qualité du plan. Le critère utilisé ici cherche à minimiser la distance maximale entre les points de l'ensemble candidat et les points du plan, c'est-à-dire celui défini ci-avant comme étant le critère minimax. A ce titre, le type de plan obtenu par cet algorithme peut être classé dans la série des plans minimax (critère *MaxDist*).

Afin de trouver le plan optimal au sens de ce critère, l'algorithme utilise une distribution initiale aléatoire et fait diminuer le critère à l'aide d'un simple algorithme d'échange du même type que celui utilisé pour construire les plans à faible discrédance. Cependant, ici, les échanges se font entre les points du plan et un ensemble de points candidats fini qui sert à calculer le critère que l'on cherche à optimiser.

L'utilisation des algorithmes d'échange n'est pas récente : Kennard et Stone (1969), Fedorov (1972) et Mitchell (1974) proposaient déjà de ce type d'algorithmes pour générer des plans uniformes.

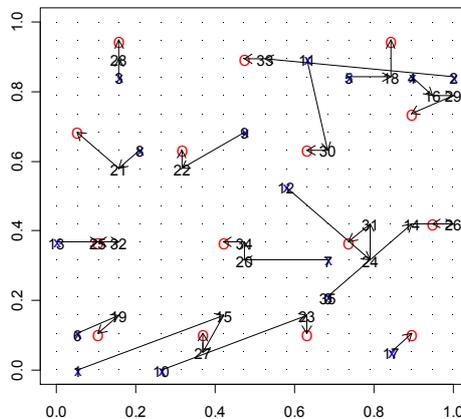


Figure 36. Un graphe représentant l'historique des échanges pour un plan

Le principe de l'algorithme est simple : pour un point donné du plan, on remplace ce point avec un de l'ensemble candidat et on examine si cet échange permet d'améliorer le critère. Si l'amélioration est avérée, le nouveau point est adopté dans le plan et l'ancien point est ajouté à l'ensemble des candidats. Ce processus continue jusqu'à ce qu'il n'y ait plus d'échange possible ou que le nombre d'itérations préalablement fixé soit atteint. Ainsi la convergence de l'algorithme est assurée, mais aucune garantie n'est donnée d'avoir généré le meilleur plan possible.

Royle et Nychka (1998) proposent un calcul efficace du critère de distance lorsque deux points du plan sont échangés ; qui est aussi un des plus grand succès de l'approche Fedorov.

Extensions. Une première extension consiste à fixer des points du plan, c'est-à-dire des points qui ne peuvent être échangés (voir Figure 37). Ceci est nécessaire dans une approche séquentielle *i.e.* si l'on veut accumuler une suite de plans en augmentant la taille et en conservant les points déjà générés.

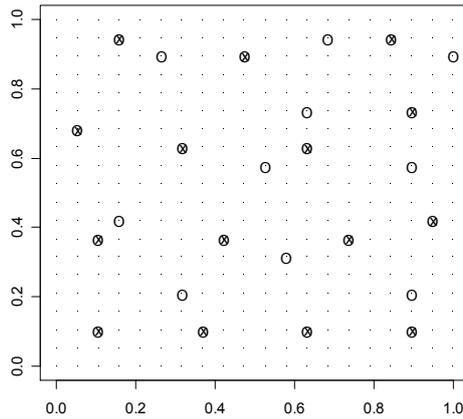


Figure 37 Un plan minimax à 25 points (cercle) généré à partir d'une grille régulière à 400 points et d'un premier plan minimax à 15 points (croix)

Une autre extension considère des métriques différentes de la distance euclidienne généralement utilisée. Pour terminer, l'algorithme peut considérer uniquement les échanges entre les points candidats et les plus proches voisins, ce qui permet de diminuer les temps de calcul tout particulièrement quand l'ensemble candidat est grand.

Il paraît clair qu'en grande dimension ces plans ne pourront être de bonne qualité vis-à-vis du remplissage de l'espace. En effet, le procédé calculatoire qui utilise des points aléatoires à la place d'une grille régulière ne permet pas d'écarter au mieux les points.

Nous n'avons pas souhaité discuter ici des plans maximin car ils restent très simples à générer (également à l'aide d'un algorithme d'échange). Cependant, ils paraissent être de meilleure qualité que les plans minimax et équivalents aux plans à entropie maximale présentés dans la section 2.3.7.

2.2.5.2. Algorithme de Wooton Sergent et Phan-Tan-Luu

L'algorithme de WSP (Sergent 1989, Sergent *et al.* 1997) repose sur l'algorithme de Kennard et Stone (1969) et Wooton (1975).

Il se déroule de la manière suivante. Dans un premier temps, on choisit le point de la distribution initiale le plus proche du centre de l'hypercube unité (choix arbitraire de l'algorithme). Ensuite, on élimine les points, autres que le centre, compris dans la sphère d'élimination de rayon D_{min} fixé à l'initialisation de l'algorithme. On réitère ce procédé en considérant une nouvelle sphère d'élimination centrée au point le plus proche du centre de la sphère de l'itération précédente.

Une fois tous les points de la distribution initiale parcourus (éliminés ou conservés), le processus est terminé. On obtient alors une distribution de n points répartis uniformément

dans l'hypercube où n est inférieur ou égal au nombre de points de la distribution initiale. Il est évident que plus la distance D_{min} est faible et plus le plan compte de points.

En résumé, l'algorithme de WSP assure que les points de l'espace sont choisis de telle façon qu'ils sont à la fois au moins à une distance minimale (D_{min}) de chaque point déjà inclus dans le plan et également, aussi près que possible du centre de l'hypercube unité.

Une extension possible est de rechercher toutes les solutions en faisant varier la distance minimale et de choisir celle présentant un bon compromis entre la qualité du critère et le nombre de points. Le critère généralement utilisé est le rapport entre la distance minimale et la distance maximale des plus proches voisins.

L'inconvénient principal de cet algorithme est la difficulté à obtenir un nombre de points souhaité et de fixer la distance minimale pour s'en approcher. De plus, les défauts de la distribution initiale persisteront dans le plan final. Il est clair que si le plan initial a des zones de vide, le plan ne pourra avoir d'expériences dans cette zone. Pour amoindrir ces phénomènes, il est préférable de considérer un plan initial avec un grand nombre de points ou de points déjà bien répartis.

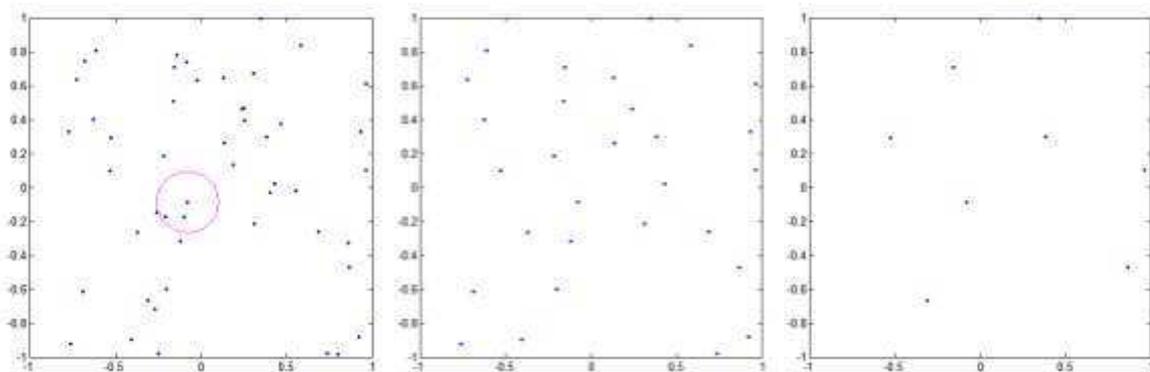


Figure 38. A gauche, un plan initial (aléatoire) de 50 points avec la première boule d'élimination. Au centre, le plan à 28 points obtenu avec la boule d'élimination de la figure de gauche. A droite, le plan à 8 points obtenu avec une boule d'élimination de rayon supérieur

Il est également possible de choisir une suite à faible discrépanance comme plan initial. Le plan ainsi obtenu a généralement une discrépanance plus faible que lorsque le plan initial est aléatoire. L'exemple suivant (Figure 39) montre qu'en choisissant une suite de Halton comme distribution initiale, on obtient une discrépanance plus faible pour le plan final que pour une suite de Halton de mêmes paramètres. Ainsi, le plan de gauche de la Figure 39 est de meilleure qualité par rapport à la suite de Halton (à droite) que ce soit au point de vue de la discrépanance ou même du critère de distance associé à l'algorithme WSP.

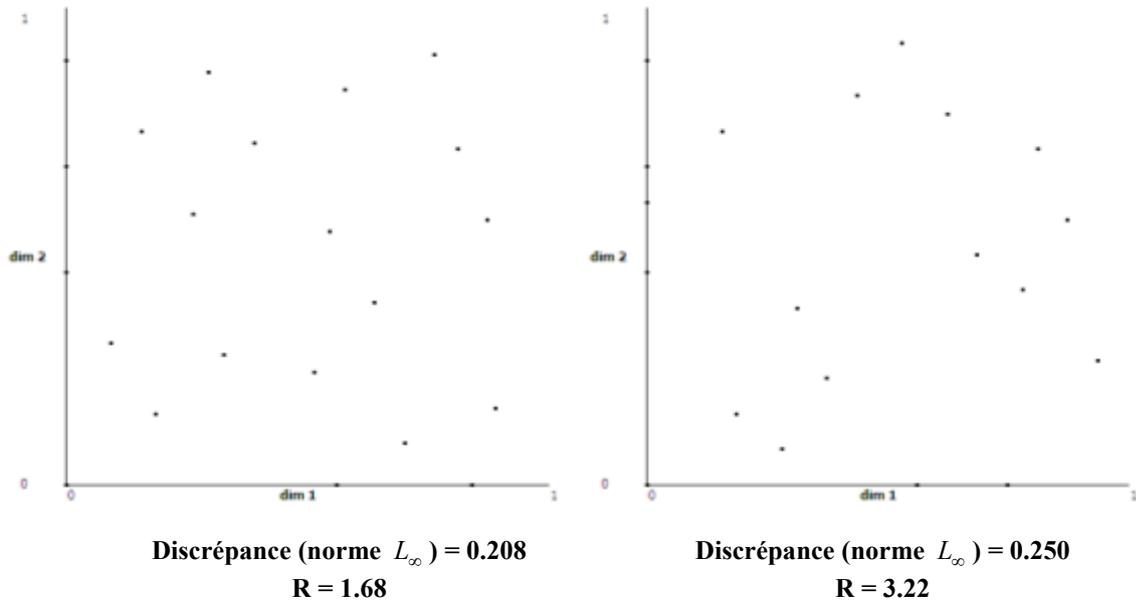


Figure 39. A gauche, un plan de 20 points obtenu   partir d'une suite de Halton   50 points. A droite, une suite de Halton   20 points en dimension 2

2.3. Critère d'entropie et plans à entropie maximale

Ce critère est différent de ceux présentés précédemment dans le sens où il ne permet pas à proprement parler d'étudier l'uniformité et/ou le remplissage de l'espace d'un plan en phase exploratoire. En effet, le calcul de l'entropie n'est bien souvent envisageable que si la distribution sous-jacente est connue ; hypothèse que nous ne possédons généralement pas en phase exploratoire.

Le souhait de définir la notion d'entropie, ici, est motivé par la présentation d'une méthode de plans optimaux au sens de ce critère dans la section 2.3.7. En effet, même si le critère n'est pas intrinsèque, les plans générés présentent de bonnes propriétés au regard du remplissage de l'espace. De plus, ce critère permet de prendre en compte le caractère anisotrope des variables qui lui peut être pressenti dès la phase exploratoire grâce à la connaissance du phénomène physique.

2.3.6. Définition de l'entropie

Un peu d'histoire. La notion d'entropie a été introduite officiellement en thermodynamique au milieu du *XIX^{ème}* siècle par Clausius pour compléter le principe de conservation de l'énergie. Boltzmann a été le premier à formaliser de manière probabiliste l'entropie qui fut utilisée par Shannon en 1948 dans le contexte de la théorie de l'information. Celui-là même a défini, dans un cadre probabiliste, une notion d'entropie, qui a des interprétations diverses en informatique, en thermodynamique et en probabilités, pourtant toutes liées par des intuitions communes. Shewry et Wynn (1987) l'ont décrit comme étant « la quantité d'information dans une expérience ». Plus généralement, l'entropie mesure la quantité d'information contenue dans une distribution. Voilà dans le cadre probabiliste, la définition qui a été retenue.

Définition

L'entropie d'une variable aléatoire continue X à densité de probabilité f est définie par :

$$H(X) = - \int_{\mathbf{x} \in \mathcal{R}} f(x) \log(f(x)) dx = -E_X(\log f(X))$$

avec la convention $0 \ln(0) = 0$.

Pour une étude mathématique simplifiée, on choisit le logarithme népérien, ceci n'ayant aucune influence sur les résultats énoncés car l'entropie est alors translatée d'une constante.

Par ailleurs, l'entropie d'un vecteur aléatoire $X = (X_1, \dots, X_d)$ à valeurs dans \mathbb{R}^d et de même densité f se définit de façon analogue à la dimension un :

$$H(X) = - \int_{x \in \mathbb{R}^d} f(x) \ln(f(x)) d\mu(x)$$

où f est la densité de probabilité et μ la mesure de Lebesgue.

Remarque. La valeur de l'entropie ne dépend pas des valeurs prises par X , mais uniquement de la densité de probabilité d'où l'impossibilité de la calculer directement à partir d'un plan d'expériences.

Entropie et quantité d'information. Shannon (1948) décrit l'entropie comme une quantité d'information. Afin de comprendre cette notion, nous allons considérer une variable aléatoire discrète X dont la distribution est donnée par $P(X = x_i) = p_i$. L'entropie de X est alors définie par :

$$H(X) = - \sum_{i=1}^n p_i \ln(p_i)$$

D'une part, on remarque que si toutes les valeurs de X sont distribuées de manière régulière (c'est-à-dire où $p_1 = p_2 = \dots = 1/n$), alors X est la plus imprévisible possible et $H(X) = \ln(n)$ atteint sa valeur maximale. Cette propriété est connue comme le principe d'entropie maximale. Si au contraire X prend une seule valeur avec la probabilité 1 alors X est complètement prévisible et l'entropie associée sera minimale avec $H(X) = 0$.

D'autre part, on note que l'entropie doit augmenter avec le nombre d'états possibles ce qui signifie que plus il y a de choix et plus l'incertitude est grande.

Cet exemple montre que la quantité d'information disponible sur la distribution peut se mesurer par $I = -H$. Shannon (1948) montre que cette quantité d'information I apportée par un paramètre θ est la seule fonction ayant la propriété d'additivité à une constante près et la propriété de continuité.

Gain d'une expérience. Lindley (1956) utilise la notion de quantité d'information au sens de Shannon afin de mesurer l'information apportée par une expérience. Notons e une expérience d'un système aléatoire X de densité $p(x|\theta)$ et θ un paramètre. Généralement θ est le paramètre du modèle statistique que nous souhaitons estimer. Selon le modèle choisi, θ est défini différemment. Par exemple, en régression, θ peut être le vecteur des paramètres de la régression ; alors que si on considère un processus gaussien, θ correspond au vecteur des différents paramètres (espérance, variance, corrélations).

Si la distribution *a priori* de θ est notée $\pi(\theta)$ et si x est la donnée obtenue après l'expérience e alors la densité *a posteriori* de θ est proportionnelle à

$$\pi(\theta|x) \propto p(x|\theta)\pi(\theta).$$

L'information globale sur θ , avant e est

$$I(\theta) = \int \pi(\theta) \log(\pi(\theta)) d\theta$$

et après e ,

$$I(\theta|x) = \int \pi(\theta|x) \log(\pi(\theta|x)) d\theta.$$

L'accroissement d'information est donc donné par

$$I(\theta|x) - I(\theta).$$

Le gain d'information est défini en tenant compte de la loi conjointe de X et de θ ,

$$g(e) = E[I(\theta|x) - I(\theta)] = E_x E_\theta \log \left(\frac{\pi(\theta|x)}{\pi(\theta)} \right).$$

Maximisation de l'entropie. Avec la définition ci-dessus, l'idée est de sélectionner une expérience e dans un ensemble E qui maximise le gain. Shewry et Wynn (1987) soulignent plusieurs problèmes liés à cette stratégie et notamment sur la définition de l'ensemble E . Ils proposent de considérer E comme un ensemble fini d'expériences possibles et établissent ainsi le lien entre gain et l'entropie.

Considérons le domaine expérimental E comme discret et composé de N points auxquels on associe une réponse Y_i , $i=1, \dots, N$. Partitionnons le domaine expérimental en deux

ensembles, D le plan d'expériences et \bar{D} les points restants. Alors, la décomposition classique de l'entropie s'écrit

$$H(Y_E) = H(Y_D) + E_{Y_D} [H(Y_{\bar{D}}|Y_D)].$$

L'espérance dans le membre de droite correspond à $-g(e)$ dans le cas où $Y_{\bar{D}}$ joue le rôle de θ et e au plan D . On cherche donc à minimiser $H(Y_E) - H(Y_D)$. L'entropie totale du système, $H(Y_E)$, étant fixe, cela revient à maximiser l'entropie aux points du plan $H(Y_D)$.

Construire un plan à entropie maximale a donc pour but de maximiser la quantité d'information apportée par les expérimentations relativement à un paramètre θ . On trouve au sujet des plans à entropie maximale de nombreuses discussions, nous retiendrons principalement celles de Koelher et Owen (1996) et Santner *et al.* (2003). Le critère d'entropie a été, par la suite, appliqué aux plans d'expériences pour l'approximation d'un modèle numérique complexe déterministe par Mitchell et Scott (1987) puis aussi par Currin *et al.* (1991) et dernièrement par Sebastiani et Wynn (2000). La section suivante propose une méthode simple, basée sur un algorithme d'échanges, pour générer des plans à entropie maximale où l'on présuppose un modèle sous-jacent.

Remarque. L'approche combine une information *a priori* et des données expérimentales avec un modèle sous-jacent pour produire une distribution *a posteriori*. Elle nous place donc entièrement dans le contexte bayésien. On pourra se référer à l'article de Chaloner et Verdinelli (1995) pour un état des lieux complet concernant les plans bayésiens (Walter et Pronzato, 1997).

2.3.7. Plans à entropie maximale

La définition générale des plans à entropie maximale nécessite d'avoir la réponse aux points du plan d'expériences. Ce qui ne fait pas *a priori* de l'entropie un critère intrinsèque.

Shewry et Wynn (1987) appliquent le calcul de l'entropie de la loi d'un processus gaussien de façon à ne pas avoir à utiliser les réponses aux points du plan d'expériences pour construire celui-ci.

La méthode qui suit propose un remplissage de l'espace, en distribuant des points selon une matrice de corrélation spatiale. Dans le cas particulier d'un processus gaussien centré f ,

Shewry et Wynn (1987) démontrent⁹ que $H(Y_X)$ dépend simplement de $\ln(\det(C(X)))$ où $C(X)$ dénote la matrice de covariance. Ainsi, générer un plan à entropie maximale revient à, sous l'hypothèse de stationnarité, maximiser le déterminant de la matrice de corrélation.

Remarque. Si le modèle est linéaire, alors le déterminant s'exprime en fonction de la matrice du plan. Aussi, le plan obtenu à l'aide d'une méthode classique telle qu'un algorithme d'échange est, dans ce cas, D-optimal.

Cette équivalence existe uniquement si les réponses aux points suivent une loi normale multidimensionnelle, mais sans aucune condition sur la structure de covariance.

Soit $X = (X^1, \dots, X^n)^t$ un vecteur de variables aléatoires, la matrice de variance – covariance de X s'écrit :

$$C(X) = \begin{pmatrix} \sigma_1^2 & \text{cov}(X^1, X^2) & \text{cov}(X^1, X^n) \\ & \sigma_2^2 & \\ & & \dots \\ & & & \sigma_n^2 \end{pmatrix}$$

où σ_i est l'écart-type de X^i et $\text{cov}(X^i, X^j) = \sigma_i \sigma_j \rho_{ij}$ est la covariance de X^i et X^j .

Si les variables X^i sont réduites, $C(X)$ s'identifie avec la matrice de corrélation :

$$C(X) = \sigma^2 \begin{pmatrix} 1 & \rho_{ij} \\ & \dots \\ \rho_{ij} & 1 \end{pmatrix}.$$

Définissons à présent une matrice de corrélation spatiale $C = [\rho_{ij}]$ avec :

$$\begin{cases} \rho_{ii} = 1 \\ \rho_{ij} = 1 - \gamma(h_{ij}) \\ \rho_{ij} = 0 \text{ si } h_{ij} > a \end{cases} \quad (1)$$

On reconnaît en $\gamma(h)$ le variogramme où h_{ij} est la distance entre les points i et j et a la portée dudit variogramme (voir section 2.3.3. et annexe A).

⁹ Cette démonstration a été excellemment reprise dans Koelher et Owen(1996).

On peut alors calculer pour $X = (\mathbf{x}^1, \dots, \mathbf{x}^n)^t$, vecteur de points dans un espace à d dimensions, une matrice de corrélation spatiale par :

$$C(X) = \begin{pmatrix} 1 & & \rho_{ij} \\ & \dots & \\ \rho_{ij} & & 1 \end{pmatrix}$$

où ρ_{ij} est une fonction de la distance entre les points i et j , calculée selon un modèle de corrélation spatiale que l'on attribue *a priori* à l'espace étudié (voir (1) ci-dessus).

Le déterminant de $C(X)$ est maximum pour $\rho_{ij} = 0$, c'est-à-dire lorsque les points pris deux à deux sont éloignés l'un de l'autre d'une distance supérieure à la portée a de la fonction de corrélation spatiale.

On cherche donc à maximiser le déterminant de $C(X)$ par un algorithme d'échange du type de Fedorov ou de Mitchell comme ci-après :

Algorithme 2. Algorithme pour générer un plan à entropie maximale.

Initialisation : choisir un vecteur de n points $X_{(0)}$ au hasard et fixer un variogramme

Calcul du déterminant de la matrice de corrélation spatiale associée : $\det(C(X_{(0)}))$

POUR k = 1 à NMax

Choisir une expérience i au hasard dans $\{1, \dots, n\}$

Simuler z^i uniformément dans $[0, 1]^d$

Accepter de remplacer x^i par z^i si $\det(C(X_{(k)})) > \det(C(X_{(k-1)}))$

Fin pour

Remarque. L'algorithme classique DETMAX teste tous les points d'une grille régulière.

2.3.8. Propriétés des plans à entropie maximale

Choix du variogramme. Comme l'illustre la Figure 44, nous avons vu précédemment que l'équivalence entre la maximisation de l'entropie et du déterminant de la matrice de corrélation ne dépend pas de la fonction de corrélation choisie. On rappelle ci-après quelques-uns des principaux modèles de variogrammes.

Modèle sphérique

$$\gamma(h) = \frac{3}{2} \frac{h}{a} - \frac{1}{2} \left[\frac{h}{a} \right]^3 \quad \text{si } h \leq a$$

$$\gamma(h) = 1 \quad \text{si } h > a$$

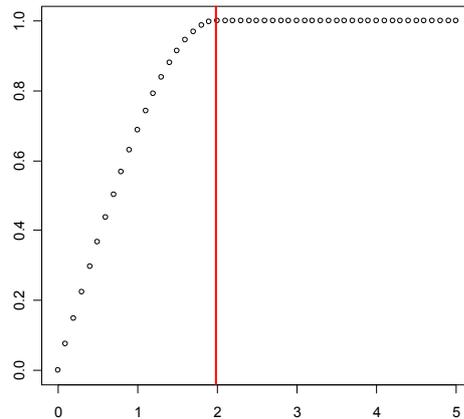


Figure 40. Représentation d'un variogramme sphérique avec une portée égale à 2

Modèle exponentiel

$$\gamma(h) = 1 - \exp\left[-\frac{3h}{a}\right]$$

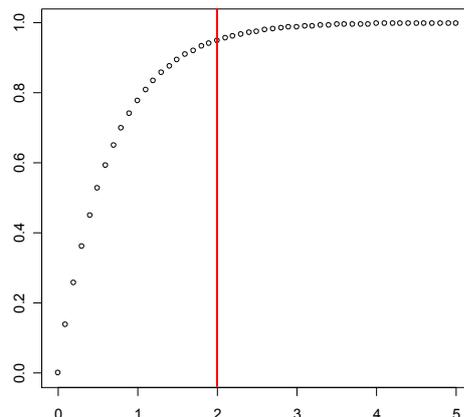


Figure 41. Représentation d'un variogramme exponentiel avec une portée égale à 2

Modèle gaussien

$$\gamma(h) = 1 - \exp\left[-\frac{3h^2}{a^2}\right]$$

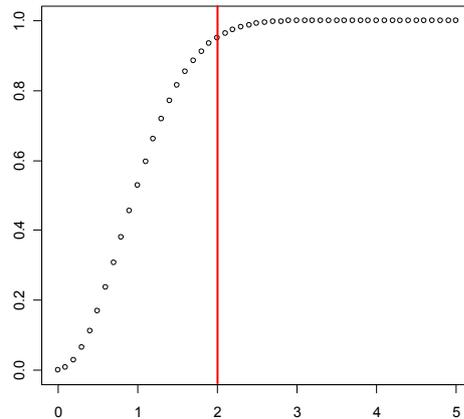


Figure 42. Représentation d'un variogramme gaussien avec une portée égale à 2

Modèle en puissance

$$\gamma(h) = \left[\frac{h}{a}\right]^\alpha \quad \text{avec } 0 < \alpha < 2$$

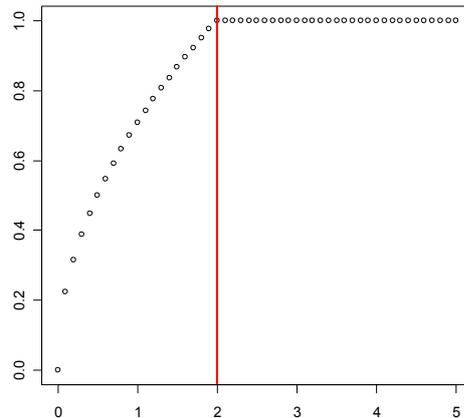


Figure 43. Représentation d'un variogramme puissance ($\alpha = 0.5$) avec une portée égale à 2

A noter que les portées théoriques (représentées en rouge) sont, excepté pour le variogramme sphérique, différentes des portées pratiques.

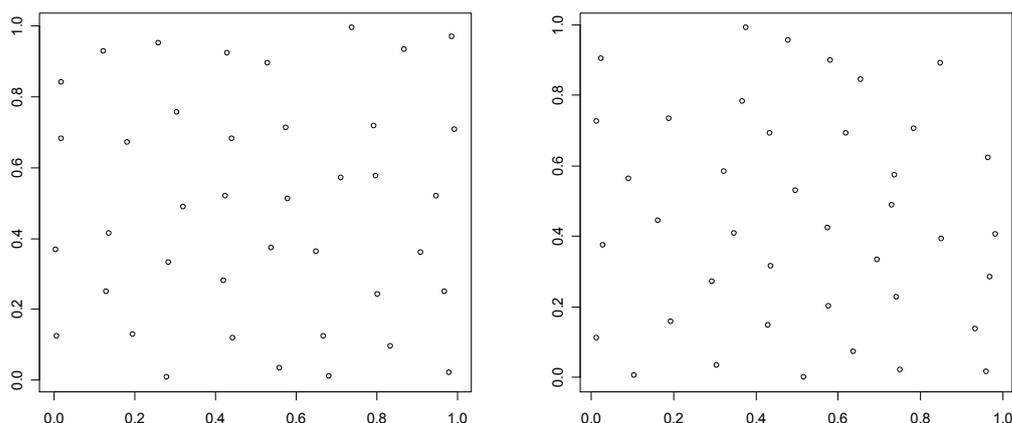


Figure 44. Un plan à entropie maximale à 40 points en 2D avec une portée de 0.7 obtenu à l'aide d'une fonction de corrélation exponentielle (respectivement sphérique) à gauche (respectivement à droite)

La figure ci-dessus conforte le fait que le choix du variogramme influe peu sur la distribution finale des points car nous cherchons par le procédé de construction à annuler son effet en donnant une prépondérance à la portée et non au comportement à l'origine.

Par la suite, nous considérons uniquement des plans générés avec un variogramme sphérique car dans la mesure où la portée théorique est la même que la portée réelle cela semble préférable ; notamment au niveau du contrôle de la convergence de l'algorithme. En réalité, ce sont bien les paramètres de corrélation c'est-à-dire les portées qui vont influencer sur le résultat obtenu (voir Figure 45) et non le choix de la fonction de corrélation.

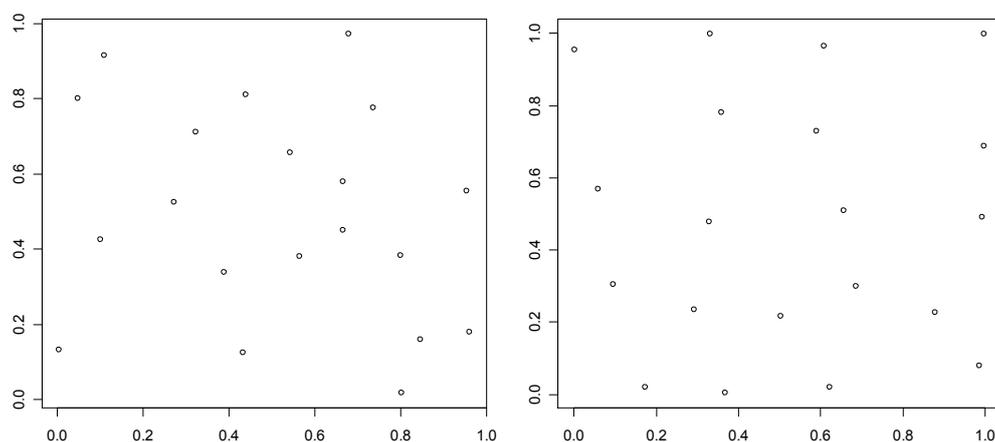


Figure 45. A gauche, un plan à entropie maximale à 20 points en 2D obtenu avec une portée de 0.1. A droite, un plan à entropie maximale à 20 points en 2D obtenu avec une portée de 0.9

Pratiquement, la portée sera choisie assez grande relativement à la dimension de l'espace afin d'optimiser au mieux le remplissage de ce dernier. La racine de la dimension semble être une bonne alternative.

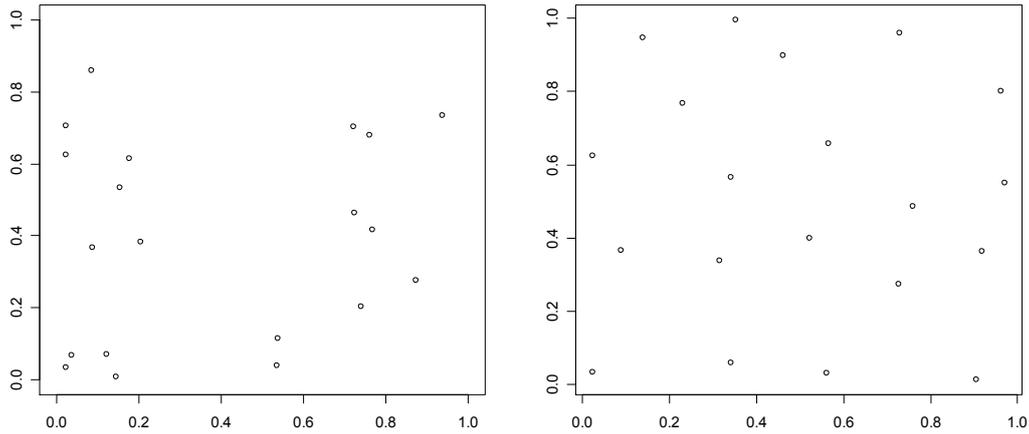


Figure 46. A gauche, un plan initial, aléatoire, à 20 points en 2D. A droite, un plan à 20 points en 2D construit en maximisant l'entropie

Space-Filling Designs. L'interprétation de la maximisation du déterminant de la matrice de covariance a été donnée par Johnson *et al.* (1990). Ils ont montré que, si la corrélation entre les points est extrêmement faible et que c'est une fonction décroissante de la distance entre les points, alors le critère d'entropie maximise la distance minimale des points du plan et favorise ainsi les distributions avec peu de points proches au sens de cette grandeur. Aussi, ils affirment que ces plans ont tendance à imposer des points sur les bords du domaine.

Anisotropie. A noter que l'hypothèse d'un modèle sous-jacent peut faire toute la richesse de cette méthode car elle permet ainsi de prendre en compte une information *a priori* comme par exemple l'anisotropie d'un phénomène (voir Figure 47). Sur la figure ci-dessous, nous avons considéré des portées directionnelles choisies de manière à ce qu'il y ait une portée plus faible pour le facteur X_1 que pour le facteur X_2 . Ainsi, le plan obtenu a des points plus rapprochés dans la direction X_1 que dans la direction X_2 .

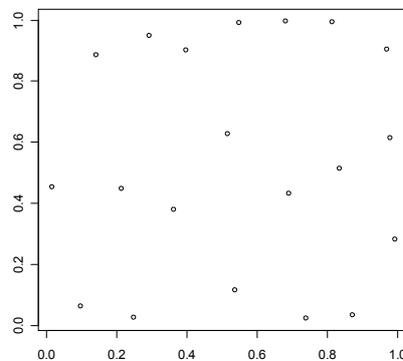


Figure 47. Un plan anisotrope où l'on considère une portée de 0.2 dans la direction des abscisses et une portée de 0.6 dans la direction des ordonnées

PARTIE II.

CONSTRUCTION DE NOUVEAUX PLANS

D'EXPERIENCES

Cette partie est consacrée à des techniques innovantes en expérimentation numérique pendant la phase initiale.

Le premier chapitre exploite les méthodes de Monte-Carlo par chaînes de Markov pour générer des plans de type Space-Filling.

Le second chapitre présente des extensions : des plans hétérogènes basés sur une distance à densité et des plans d'expériences générés dans un espace défini par des contraintes.

Chapitre 3

Plans d'expériences par méthodes McMC

On commence par rappeler brièvement quelques unes des raisons de la non-performance des plans usuels : le nombre de simulations élevé, le non-remplissage de l'espace ou bien encore les mauvaises propriétés en projection sur les sous-espaces.

Les chapitres précédents ont fait ressortir que principalement les plans de WSP et à entropie maximale permettent de bien remplir l'espace et ne possèdent pas de phénomènes particuliers d'alignements en projections. On se souvient cependant que les plans de WSP, utilisant un processus de mort, ne permettent pas de fixer au préalable le nombre de points souhaités. De plus, leur qualité au regard du remplissage de l'espace est fortement liée au choix de la distribution initiale. Les plans à entropie maximale quant à eux présupposent un modèle sous-jacent, et nous préférons nous affranchir d'hypothèses supplémentaires même si celles-ci sont faibles, peuvent devenir des avantages dans certains cas, et n'induisent pas d'erreurs. Nous retiendrons essentiellement que ces 2 types de plans n'offrent pas la possibilité d'imposer aux points d'avoir une bonne répartition en projection sur les marges à l'image des hypercubes latins.

Partant du constat que les méthodes usuelles de construction de plans d'expériences ne sont pas adaptées et sont généralement basées sur des heuristiques et non sur des fondements théoriques, nous introduisons un nouveau type de Space-Filling Designs spécifié à partir du processus ponctuel de Strauss. Ces plans construits à partir de méthodes Monte-Carlo par chaînes de Markov (McMC) sont facilement modulables et peuvent ainsi répondre à divers objectifs. Signalons en particulier les bonnes propriétés de projection des expériences sur les

axes factoriels. Nous verrons encore qu'il est possible de générer des points de manière hétérogène (préférentielle) si l'on dispose d'informations *a priori*.

La première partie de ce chapitre argumentera l'introduction des méthodes McMC en planification d'expériences numériques et présentera l'algorithme général de Metropolis-Hastings (MH). La deuxième partie sera consacrée aux chaînes de Markov afin d'étudier la convergence de l'algorithme de MH. Enfin, on présentera la nouvelle famille de plans en question en s'appuyant sur la méthodologie McMC décrite auparavant.

3.1. Méthodes McMC

On introduit dans ce chapitre des plans qui sont par nature aléatoires ou randomisés. Ils sont définis par une loi de probabilité sur le domaine cubique $[0,1]^d \times \dots \times [0,1]^d$:

$$\pi(dx) = k \times \pi(x)dx ; x = (x^1, \dots, x^n) \quad (1)$$

Le choix de la densité non normalisée $\pi(x)$ sera lié aux contraintes imposées aux plans d'expériences. Nous verrons ultérieurement les raisons pour lesquelles on considère un processus de Strauss ou plus généralement un processus ponctuel de Gibbs.

A ce stade, disons simplement qu'il est impossible en pratique de simuler directement de tels plans à cause de la relative complexité de la loi $\pi(dx)$. En particulier, il n'existe pas d'expression analytique simple de la constante de normalisation k intervenant dans la densité du processus (1). Comme l'espace est de grande dimension ($n \times d$), l'estimation de cette constante n'est pas envisageable.

L'alternative classique est de simuler une chaîne de Markov de loi asymptotique (ou stationnaire) $\pi(\cdot)$ (Meyn, Tweedie, 1993). C'est le principe des méthodes dites de Monte-Carlo par chaînes de Markov (McMC). La souplesse d'une telle approche réside dans le fait que l'on approche la loi π par itérations successives. Le choix du mécanisme de transition de la chaîne de Markov sous-jacente rend une transition d'état à état facile à simuler. Les processus que l'on souhaite simuler ne sont connus que par leur densité non normalisée, ainsi la construction de la chaîne de Markov ne doit pas faire intervenir la constante de normalisation. C'est le cas de l'algorithme de Metropolis-Hastings (MH, Metropolis *et al.* 1953, Hastings 1970) qui consiste à proposer à chaque étape des transitions de l'état courant x vers un état y *a priori* quelconque. La transition proposée est ensuite acceptée ou non selon une certaine probabilité. Le taux d'acceptation fait intervenir la densité de manière à assurer la convergence de l'algorithme vers la loi π .

L'algorithme de MH se décrit donc en 2 étapes :

- on fait une proposition de changement d'états $x \mapsto y$ selon une loi de probabilité $Q(x, dy)$
- on accepte y avec la probabilité $a(x, y)$; sinon, on reste dans l'état x .

La donnée simultanée de Q et a caractérise entièrement l'algorithme de MH ou la chaîne de Markov sous-jacente. En général, le choix de la loi Q est un compromis entre facilité de simulation et capacité à bien explorer l'espace des configurations. Le choix de la probabilité d'acceptation a est plus limité : il est dicté essentiellement par l'objectif de simuler (asymptotiquement) une loi de probabilité π donnée. C'est le cas du choix usuel (supposant Q symétrique)

$$a(x, y) = \min\left(1, \frac{\pi(y)}{\pi(x)}\right)$$

qui favorise les configurations les plus probables pour la loi π . On remarque que le calcul de la probabilité $a(x, y)$ ne fait pas intervenir la constante de normalisation k .

Remarque. Le choix de l'état de destination y ne dépend pas forcément de la densité à simuler.

Pour aller plus loin dans la compréhension d'une telle méthode, il nous faut d'abord étudier l'équivalent probabiliste d'une suite de points générés par itérations successives d'une fonction $f(\cdot)$ déterministe :

$$z_0 \text{ quelconque ; } z_{n+1} = f(z_n) \text{ pour } n \geq 0 .$$

3.2. Chaînes de Markov

On appellera chaîne de Markov toute suite $(X_p)_{p \geq 0}$ de variables aléatoires à valeurs dans un même espace E (désigné espace des états ou configurations) et obéissant à une dynamique de la forme :

$$X_0 \text{ de loi initiale } \nu \text{ quelconque ; } X_{p+1} = f(X_p, U_p) \text{ pour } p \geq 0$$

où les U_p sont des vecteurs aléatoires indépendants et identiquement distribués. La fonction f , qui est à valeurs dans E , représente typiquement une itération ou boucle d'un algorithme de simulation MCMC. La taille des vecteurs U_p est le nombre d'appels à la fonction *RAND* dans cette boucle (cette taille peut être finie ou non).

La traduction probabiliste de la dynamique ci-dessus est qu'à tout instant $p \geq 0$, la loi conditionnelle de X_{p+1} sachant X_p, X_{p-1}, \dots, X_0 (présent + passé) est égale à la loi conditionnelle de X_{p+1} sachant X_p . Dit autrement, pour prévoir l'évolution future de la suite à un instant p , connaissant la dernière valeur X_p , le passé X_{p-1}, \dots, X_0 n'apporte aucune information supplémentaire. Cela ne signifie pas que l'évolution future soit indépendante du passé mais qu'elle n'en dépend qu'à travers le présent X_p : le futur ne dépend du passé qu'à travers le présent! On reconnaît ici précisément la définition usuelle d'une chaîne de Markov à partir de cette propriété dite de Markov (simple).

La question centrale qui va se poser est de savoir si partant d'une loi initiale ν arbitraire (celle de X_0), la loi obtenue au bout de p itérations (donc la loi de X_p) approche la loi stationnaire π de la chaîne (à savoir la loi que l'on cherche à simuler). Pour aborder cette question, on se placera dans un premier temps dans le cadre classique des chaînes de Markov à espace d'états discret. C'est l'objet du paragraphe suivant qui considère un cas d'école dans un but essentiellement pédagogique.

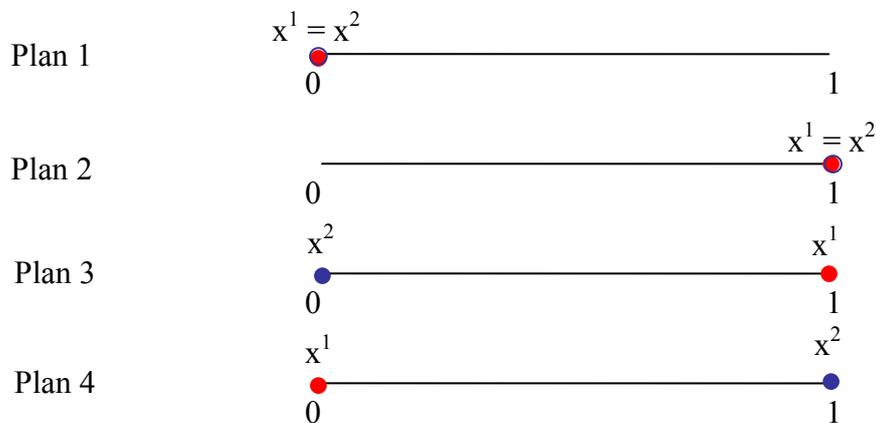
3.2.1. Un cas simple dans un espace d'états discret

Les chaînes de Markov les plus simples sont celles dont les états ne prennent qu'un nombre fini ou dénombrable de valeurs. Cette section est consacrée à une présentation de ces chaînes élémentaires dans le cadre de la planification d'expériences. En effet, nous allons mettre en œuvre la méthodologie MCMC sur un exemple simplifié de planification d'expériences de manière à se familiariser avec les principaux concepts relatifs aux chaînes de Markov.

Pour cela, les plans d'expériences qui nous intéressent *a priori* sont composés seulement de $n = 2$ expériences distribuées dans l'intervalle $[0, 1]$ (cas de la dimension $d = 1$).

Ainsi, un plan d'expériences x est un couple de points (x^1, x^2) dans $[0, 1]^2$. L'espace E est celui des états de l'algorithme de construction itérative d'un plan ou de la chaîne de Markov sous-jacente.

Pour éliminer la difficulté essentiellement technique liée au fait que E est un espace d'états continu, on va encore radicalement le simplifier en le réduisant aux 4 plans d'expériences possibles qui suivent :



Avec la numérotation ci-dessus, on convient que $E = \{1, 2, 3, 4\}$, ce qui fait de E un espace discret fini. On confond tout état i de E avec le plan (x^1, x^2) qui lui correspond.

Une représentation des éléments de E dans l'espace euclidien $[0,1]^2$ des plans d'expériences est la suivante :

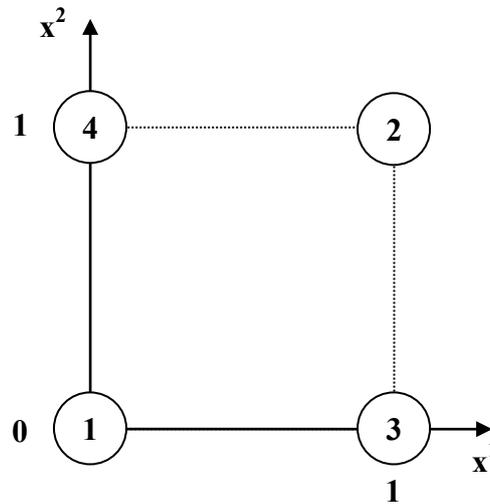


Figure 48. Représentation des éléments de E dans $[0,1]^2$

Soit maintenant π la distribution sur l'espace E des 4 plans d'expériences définie par

$$\pi = (\pi_1 \ \pi_2 \ \pi_3 \ \pi_4) = k \times (\gamma \ \gamma \ 1 \ 1)$$

avec π_i la probabilité de l'état i et où $0 < \gamma \leq 1$ est assimilé à un coefficient d'interaction entre les 2 expériences. Enfin, $k = \frac{1}{2 \times (1 + \gamma)}$ est la constante de normalisation.

Ainsi, on donne une probabilité plus forte aux états 3 et 4, c'est-à-dire aux plans dont les points sont répartis aux extrémités de l'intervalle $[0,1]$ (aucune interaction) et plus faible aux plans qui consistent à répéter deux fois la même expérience (cas d'une interaction). On verra au paragraphe 3.3.1 que cette loi est l'analogue en discret de la loi du processus ponctuel de Strauss dans le cas simple de 2 points distribués dans un espace de dimension 1.

Pour simuler π , on propose l'algorithme de Metropolis-Hastings suivant :

Algorithme 3. Algorithme de Metropolis-Hastings

Initialisation : Soit $n = 2$. Choisir $x = (x^1, x^2)$, par exemple selon la loi uniforme $\nu = \left(\frac{1}{4} \frac{1}{4} \frac{1}{4} \frac{1}{4}\right)$

POUR N = 1 à NMC

POUR k = 1 à n

Choisir i uniformément dans $\{1, 2\}$ ($i = 1$ ou 2 avec probabilité $1/2$)

Simuler u uniformément dans $\{0, 1\}$ ($u = 0$ ou 1 avec probabilité $1/2$)

Accepter le changement $x^i = u$ avec la probabilité

$$\min\left(1, \frac{\pi(y)}{\pi(x)}\right)$$

où $y = (u, x^2)$ si $i = 1$ et $y = (x^1, u)$ si $i = 2$

Fin pour k

Fin pour N

On notera qu'il s'agit d'un algorithme itératif, NMC correspondant au nombre d'itérations McMC effectuées.

Notons P la matrice de transition correspondant à chaque itération élémentaire dans la boucle « Pour $k = 1$ à n » et $K = P^n$ la matrice de transition correspondant à l'ensemble du balayage. Ainsi, la suite $\mathbf{x}_{(0)}, \mathbf{x}_{(1)}, \dots, \mathbf{x}_{(n)}$ des plans générés par cet algorithme est la réalisation d'une chaîne de Markov $(X_p)_{p \geq 0}$ de probabilité (ou noyau de transition) la matrice K . Le calcul conduit à

$$P = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{\gamma}{4} & \frac{\gamma}{4} & 1 - \frac{\gamma}{2} & 0 \\ \frac{\gamma}{4} & \frac{\gamma}{4} & 0 & 1 - \frac{\gamma}{2} \end{pmatrix}$$

où $P(i, j)$ s'interprète comme la probabilité pour passer du plan d'expériences i au plan j .

Donnons quelques exemples de calcul.

$P(3,1)$, correspondant à la probabilité de passer de l'état 3 à l'état 1, est égale au produit des probabilités

- de choisir le point x^1 (parmi 2 possibles) du plan d'expériences à savoir $i = 1$
- de choisir la direction (parmi 2 possibles) pour se retrouver dans l'état 1 à savoir $u = 0$
- d'accepter l'état 1 venant de l'état 3

On a donc,

$$P(3,1) = \frac{1}{2} (\text{choix du point}) * \frac{1}{2} (\text{choix de la direction}) * \gamma (\text{probabilité d'accepter})$$

On a $P(3,1) = P(3,2)$ par symétrie. La transition de l'état 3 vers l'état 4 étant impossible en un seul mouvement $P(3,4) = 0$. On en déduit ainsi que $P(3,3) = 1 - \frac{\gamma}{2}$.

Une représentation classique des transitions de E dans E est la suivante :

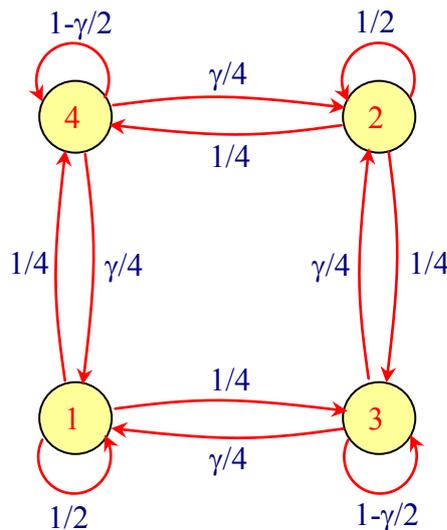


Figure 49. Représentation des transitions de E dans E

Comme $n = 2$, on en déduit

$$K = P^2 = \begin{pmatrix} \frac{2+\gamma}{8} & \frac{\gamma}{8} & \frac{3-\gamma}{8} & \frac{3-\gamma}{8} \\ \frac{\gamma}{8} & \frac{2+\gamma}{8} & \frac{3-\gamma}{8} & \frac{3-\gamma}{8} \\ \frac{\gamma(3-\gamma)}{8} & \frac{\gamma(3-\gamma)}{8} & \frac{\gamma}{8} + \left(1 - \frac{\gamma}{2}\right)^2 & \frac{\gamma}{8} \\ \frac{\gamma(3-\gamma)}{8} & \frac{\gamma(3-\gamma)}{8} & \frac{\gamma}{8} & \frac{\gamma}{8} + \left(1 - \frac{\gamma}{2}\right)^2 \end{pmatrix},$$

soit les valeurs explicites des probabilités de transition de la chaîne X :

$$K(i, j) = P(X_{p+1} = j | X_p = i).$$

3.2.1.1. Réversibilité et mesure invariante

Les transitions π -réversibles vont jouer un rôle important dans la construction des algorithmes de simulation. On dit que le mécanisme de transition P sur E est π -réversible s'il vérifie la condition dite de réversibilité :

$$\pi_i P(i, j) = \pi_j P(j, i).$$

On vérifie facilement que c'est bien le cas ici et cela signifie que, sous la distribution π , la probabilité de passer de x à y est la même que de passer de y à x .

La première conséquence pour π est qu'elle est une loi invariante (ou stationnaire) pour P , c'est-à-dire que l'on a l'égalité suivante

$$\pi P = \pi$$

En effet, on a :

$$(\pi P)_j = \sum_{i=1}^n \pi_i P(i, j) = \pi_j \sum_{i=1}^n P(j, i) = \pi_j$$

Remarque. On peut aussi le vérifier directement par le calcul de πP mais c'est une propriété générale.

A fortiori, on a $\pi K = \pi$ qui traduit que π est une loi invariante pour la chaîne de Markov $(X_p)_{p \geq 0}$. Ainsi, la condition de π -réversibilité implique l'invariance pour π . La plupart des algorithmes de simulation sont en réalité construits pour produire des chaînes de Markov réversibles.

Nous allons voir que la propriété de π -invariance est nécessaire pour obtenir la convergence de la chaîne vers π .

Remarquons simplement que, si π est la loi initiale, la chaîne de Markov associée est (strictement) stationnaire.

3.2.1.2. Ergodicité de la chaîne

On observe ici que P est une matrice *primitive* (ou *régulière*) au sens où il existe $k \geq 1$ tel que P^k ait tous ses termes strictement positifs. Cela revient à dire qu'il est possible en k transitions de visiter n'importe quel état quelque soit le point de départ. De manière plus précise ici, $k=2$ convient et correspond à un balayage (aléatoire) de l'algorithme. On a la minoration suivante

$$K(i, j) \geq c = \gamma/8 > 0.$$

Cette dernière propriété couplée avec le fait que π est invariante pour K conduit au résultat suivant :

Proposition.

La chaîne de Markov $(X_p)_{p \geq 0}$ est uniformément ergodique au sens où, pour toute loi initiale ν , on a

$$\|\nu K^p - \pi\|_1 \leq C^p \|\nu - \pi\|_1 \leq 2C^p \text{ avec } C = 1 - 4c < 1$$

où $\|\mu\|_1 = \sum_{i=1}^4 |\mu(i)|$ désigne la norme L^1 ou norme de la variation totale sur l'espace \mathbf{R}^4 identifié à l'espace des mesures signées sur $E = \{1, 2, 3, 4\}$.

Preuve. Soient ν_1 et ν_2 deux lois initiales quelconques. Notons ν la mesure signée $\nu_1 - \nu_2$ sur $E = \{1, 2, 3, 4\}$ de sorte que

$$\nu(E) = \sum_{i=1}^4 \nu(i) = \nu_1(E) - \nu_2(E) = 0.$$

Puis, pour toute constante c , on a

$$\|\nu_1 K - \nu_2 K\|_1 = \sum_{j=1}^4 |(\nu K)_j| = \sum_{j=1}^4 \left| \sum_{i=1}^4 \nu(i) K(i, j) \right| = \sum_{j=1}^4 \left| \sum_{i=1}^4 \nu(i) (K(i, j) - c) \right|$$

en utilisant le fait que $\sum_{i=1}^4 \nu(i) = 0$.

On en déduit, avec le choix particulier $c = \gamma/8$ qui minore les probabilités $K(i, j)$,

$$\|v_1 K - v_2 K\|_1 \leq \sum_{j=1}^4 \sum_{i=1}^4 |v(i)| (K(i, j) - c) = \sum_{i=1}^4 |v(i)| \sum_{j=1}^4 (K(i, j) - c)$$

soit

$$\|v_1 K - v_2 K\|_1 \leq C \|v_1 - v_2\|_1.$$

Ce qui traduit que l'« opérateur » $v \longrightarrow vK$ est strictement contractant de l'espace des probabilités sur E dans lui-même. Il suffit pour conclure d'utiliser la relation $\pi = \pi K$ qui signifie que la probabilité π est un point fixe de K .

□

3.2.1.3. Conséquence et interprétation

Soit ν une loi initiale arbitraire pour la chaîne de Markov (la loi d'initialisation de l'algorithme). Notons $\nu_p(i) = P_\nu(X_p = i)$ pour $i = 1$ à 4 et $p \geq 1$. La notation standard P_ν rappelle que P dépend de la loi initiale.

La probabilité qu'au bout de p itérations Monte-Carlo le plan d'expériences construit par l'algorithme soit le plan n° i est précisément la probabilité $\nu_p(i)$ lorsqu'on initialise avec la loi ν . En notant $\nu_p = (\nu_p(1) \nu_p(2) \nu_p(3) \nu_p(4)) = \nu K^p$ la loi de X_p à l'étape p de l'algorithme, on vient donc de voir que

$$\sum_{i=1}^4 |\nu_p(i) - \pi_i| \leq 2C^p \text{ avec } 0 < C < 1.$$

A fortiori,

$$\sup_{\nu, A \subset E} |P_\nu(X_p \in A) - \pi(A)| \leq 2C^p$$

ce qui garantit en particulier que, lorsque p augmente, la distribution de X_p s'approche (uniformément en ν) de la loi de probabilité π recherchée.

3.2.2. Cas d'un espace d'états continu

On donne ici une généralisation des résultats obtenus dans le paragraphe précédent qui englobe le cas d'un domaine expérimental continu de la forme $E = ([0,1]^d)^n$.

Pour cela, on se donne un espace mesurable E quelconque et π une loi de probabilité sur E de la forme

$$\pi(dx) = \pi(x)d\mu(x)$$

où μ est une mesure de référence (supposée positive). Dans le cas qui nous intéresse, ce sera la mesure de Lebesgue λ ou une mesure à densité par rapport à λ .

On se donne une chaîne de Markov $(X_p)_{p \geq 0}$ de probabilité ou noyau de transition K (c'est par exemple la chaîne sous-jacente à un algorithme de simulation MCMC). On suppose que π est invariante pour K . En pratique, cela résultera du fait que $K = P^n$ et de la propriété de π -réversibilité pour P :

$$\forall x, y \in E, \quad \pi(dx)P(x, dy) = \pi(dy)P(y, dx).$$

On suppose l'existence d'une constante $c > 0$ telle que

$$K(x, dy) \geq c\mu(dy).$$

Cette dernière propriété couplée avec le fait que π est invariante pour K conduit au résultat général suivant :

Théorème.

Sous les hypothèses $\pi K = \pi$ et $K(x, dy) \geq c\mu(dy)$, avec $c > 0$, la chaîne de Markov $(X_p)_{p \geq 0}$ est uniformément ergodique, au sens où, pour toute loi initiale ν , on a

$$\|\nu K^p - \pi\| \leq 2C^p \text{ où } 0 < C < 1$$

où $\|\mu\| = \sup_{\|f\|_\infty \leq 1} \left| \int f(x)d\mu(x) \right|$ désigne la norme de la variation totale sur l'espace des mesures signées sur E .

Preuve. C'est une généralisation de la preuve donnée à la suite de la proposition du paragraphe précédent. Soient ν_1 et ν_2 deux lois initiales quelconques. Notons ν la mesure signée $\nu_1 - \nu_2$ sur E de sorte que

$$\nu(E) = \nu_1(E) - \nu_2(E) = 0.$$

Soit $\mu = \nu_1 K - \nu_2 K$ la mesure signée

$$\mu(dy) = \int \nu_1(dx) K(x, dy) - \int \nu_2(dx) K(x, dy).$$

Pour toute constante c , $\nu(E) = 0$ entraîne pour f mesurable bornée ($f \in L^\infty$)

$$\int f(y) \mu(dy) = \int f(y) (\nu_1(dx) - \nu_2(dx)) (K(x, dy) - c dy).$$

On pose $g(x) = \int_E f(y) (K(x, dy) - c dy)$. On a $g \in L^\infty$ et

$$\int f(y) \mu(dy) = \int g(x) \nu(dx) = \nu(g).$$

On en déduit, avec le choix particulier c de la condition de minoration sur K ,

$$|g(x)| \leq \int |f(y)| (K(x, dy) - c dy) \leq \|f\|_\infty \times C$$

où $C = 1 - c < 1$.

De là,

$$\|g\|_\infty \leq C \|f\|_\infty.$$

En supposant $\|f\|_\infty \leq 1$, on obtient

$$\left| \int f(y) \mu(dy) \right| = |\nu(g)| \leq \|\nu\|_1 \|g\|_\infty \leq C \|f\|_\infty \|\nu\|_1 \leq C \|\nu\|_1$$

donc

$$\|\nu_1 K - \nu_2 K\|_1 \leq C \|\nu_1 - \nu_2\|_1.$$

A nouveau, l'opérateur $\mu \longrightarrow \mu K$ est strictement contractant de l'espace des probabilités sur E dans lui-même et il suffit d'utiliser la relation $\pi = \pi K$ pour conclure.

□

Conséquence 1. Soit ν une loi initiale arbitraire pour la chaîne de Markov. En notant νK^p la loi de X_p à l'étape p de l'algorithme, on vient donc d'établir la majoration

$$\sup_{\nu, A} \left| P_\nu (X_p \in A) - \pi(A) \right| \leq 2C^p .$$

Cette dernière garantit en particulier que la distribution de X_p s'approche (uniformément en ν) de la loi de probabilité π recherchée lorsque le nombre d'itérations p augmente.

Conséquence 2. Des résultats classiques de la théorie des chaînes de Markov dans des espaces d'états généraux (Doob, 1953) permettent alors d'envisager la méthode de Monte-Carlo pour estimer la moyenne $\pi(h)$ de toute fonctionnelle h sur E (loi forte des grands nombres ou théorème ergodique). De plus, on dispose d'un théorème central limite pour contrôler la convergence.

3.3. Processus ponctuels et planification d'expériences

L'idée principale est de voir chaque expérience comme un point ou particule chargée de l'espace $\Omega = [0,1]^d$. On assimile ainsi les n expériences à la réalisation (conditionnelle au nombre n de points) d'un processus ponctuel (Daley et Vere-Jones, 2003). De manière à bien occuper l'espace, il a été naturel de considérer le processus de Strauss (1975) qui permet de rendre compte de répulsion entre particules. Ainsi, les n expériences sont vues comme la photographie à un instant donné de n particules de même charge électrique en mouvement dans une boîte d-dimensionnelle (on néglige la gravité).

Pour générer un tel processus, nous utiliserons les techniques usuelles de simulation par chaînes de Markov (Meyn, Tweedie, 1993) et plus précisément l'algorithme de Metropolis-Hastings (Metropolis *et al.* 1953 ; Hastings, 1970). C'est essentiellement la propriété de Markov spatiale du processus ponctuel de Strauss qui en pratique permet de simuler de manière efficace ces nouveaux plans d'expériences, notamment en grande dimension.

Pour l'étude de tels plans (aléatoires par nature), le calcul de la moyenne des critères d'uniformité usuels (voir Chapitre 2) est envisagé. C'est la dynamique markovienne de l'algorithme de MH et ses bonnes propriétés de mélange ou d'ergodicité qui vont permettre ce calcul. Ainsi, on peut parler de planification d'expériences par méthodes de Monte-Carlo par chaînes de Markov (McMC).

3.3.1. Le processus ponctuel de Strauss

On considère le processus de Strauss avec interaction de type répulsion dans le domaine $\Omega = [0, 1]^d$ (Ripley, Kelly 1977). La loi $\pi(\cdot)$ de ce processus conditionnellement au nombre n de points est donnée par :

$$\pi(x) = k \times \gamma^{s(x)} \quad (1)$$

où $0 < \gamma < 1$ est un coefficient de répulsion, k la constante de normalisation, et où la fonction $s(x)$ assimilée à un potentiel global d'énergie est définie par :

$$s(x) = s(x^1, \dots, x^n) = \sum_{i < j} 1_{\|x^i - x^j\| \leq R}$$

Il s'agit simplement du nombre de paires d'expériences $\{x^i, x^j\}$ telles que la distance entre x^i et x^j est inférieure ou égale à R .

En termes de particules chargées, c'est aussi le nombre global d'interactions limité uniquement aux interactions de paires (interactions dites d'ordre 2). Chaque particule a une sphère d'influence de rayon $R/2$, l'interaction se produisant exactement lorsque 2 sphères quelconques se rencontrent.

Le choix $0 < \gamma < 1$ permet de donner une probabilité plus ou moins forte aux plans pour lesquels les expériences sont réparties de manière à interagir plus ou moins les unes par rapport aux autres. Le plan est d'autant plus vraisemblable que le nombre d'interactions $s(x)$ est faible. Nous ne considérons pas ici le cas limite où $\gamma = 0$, qui interdirait toute paire de points de distance inférieure ou égale à R . Ce modèle, dit de sphères dures, conduit à des points régulièrement espacés et donc à des alignements. Le cas $\gamma = 1$ correspond exactement à l'indépendance (distribution conjointe uniforme), et dans la version non conditionnelle, au processus de Poisson homogène (Daley et Vere-Jones, 2003).

Dans toute la suite, on se limitera à des réalisations de processus ponctuels conditionnellement à un nombre n de points spécifié à l'avance. Ainsi, dans la dynamique de simulation, ce nombre n n'aura pas la possibilité d'évoluer.

La non-existence d'expression analytique explicite pour la constante de normalisation k intervenant dans la densité du processus (1), même pour des tailles modestes de n et d , motive l'utilisation des méthodes MCMC (voir chapitre 3.1). Dans le contexte de la planification d'expériences, un état de la chaîne est un plan de n expériences.

L'algorithme de MH va toujours procéder en 2 étapes :

- une phase de proposition de changement d'état où le nouvel état ne diffère que par une seule expérience
- et une autre d'acceptation ou de rejet de ce changement

Algorithme 4. Algorithme de Metropolis-Hastings pour générer des plans d'expériences dits de Strauss

Initialisation : choisir un plan d'expériences x^1, \dots, x^n selon une loi de probabilité donnée (loi uniforme par exemple)

POUR N = 1 à NMC

POUR k = 1 à n

Choisir une expérience x^i au hasard

Simuler y^i uniformément dans $[0, 1]^d$

Accepter le changement $x^i = y^i$ avec la probabilité

$$a(x, y) = \min\left(1, \frac{\pi(y)}{\pi(x)}\right)$$

où $y = (x^1, \dots, x^{i-1}, y^i, x^{i+1}, \dots, x^n)$

Fin pour k

Fin pour N

La force de cet algorithme tient au fait que le calcul du rapport $\frac{\pi(y)}{\pi(x)}$ se limite à chaque fois à calculer

$$s(x^i) = \sum_{j \neq i} 1_{\|x^i - x^j\| \leq R} \quad \text{et} \quad s(y^i) = \sum_{j \neq i} 1_{\|y^i - x^j\| \leq R} .$$

En effet, on vérifie que

$$\frac{\pi(y)}{\pi(x)} = \gamma^{s(y^i) - s(x^i)} .$$

L'entier $s(x^i)$ est le nombre de voisins de l'expérience x^i , *i.e.* les points contenus dans la sphère centrée en x^i et de rayon R . Quant à $s(y^i)$, c'est le nombre de voisins de l'expérience i après la transition vers y^i .

On notera que ces plans sont bien de nature aléatoire comme par exemple les plans uniformes, les plans minimax (Johnson *et al.*, 1990) ou bien encore les plans WSP (Sergent

1989, Sergent *et al.* 1997). Cependant, ils sont construits à partir d'une loi de probabilité π spécifiée au préalable. C'est à travers cette loi et ses extensions (processus de Gibbs de potentiel plus général) qu'il sera possible de traduire les propriétés souhaitées. L'utilisation de l'algorithme de MH pour la simulation de la loi π sera justifiée en vérifiant que les hypothèses du théorème énoncé au paragraphe 3.2.2 sont satisfaites pour le mécanisme de transition associé.

Désormais, nous parlerons de *plans de Strauss*.

3.3.2. Etude des plans de Strauss

La figure de droite montre une distribution spatiale de $n=20$ points caractéristique d'un processus de Strauss obtenu après 1000 itérations MCMC. Dans ce cas là, nous avons choisi une distribution initiale uniforme et les paramètres R et γ sont fixés respectivement à 0.19 et 0.1.

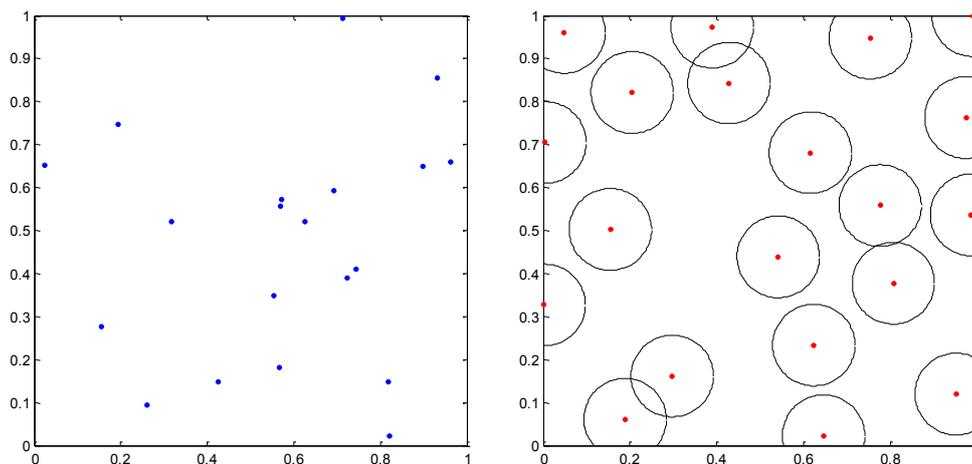


Figure 50. A gauche, un plan initial (aléatoire) de 20 points en dimension 2. A droite, le plan de Strauss obtenu à partir du plan de gauche avec la représentation des sphères de rayon $R/2$

Remarque. Sur les figures, on matérialisera les interactions entre expériences en traçant les cercles de rayon $R/2$, l'intersection de 2 cercles correspondant précisément à une interaction.

Influence des paramètres.

Les figures ci-dessous montrent l'influence du paramètre R sur la distribution finale. Le choix du rayon s'avère être important :

- un rayon trop petit engendre une distribution sans interaction mais avec de nombreuses lacunes (distribution uniforme),
- un rayon trop grand conduit à une distribution avec des agglomérats (clusters).

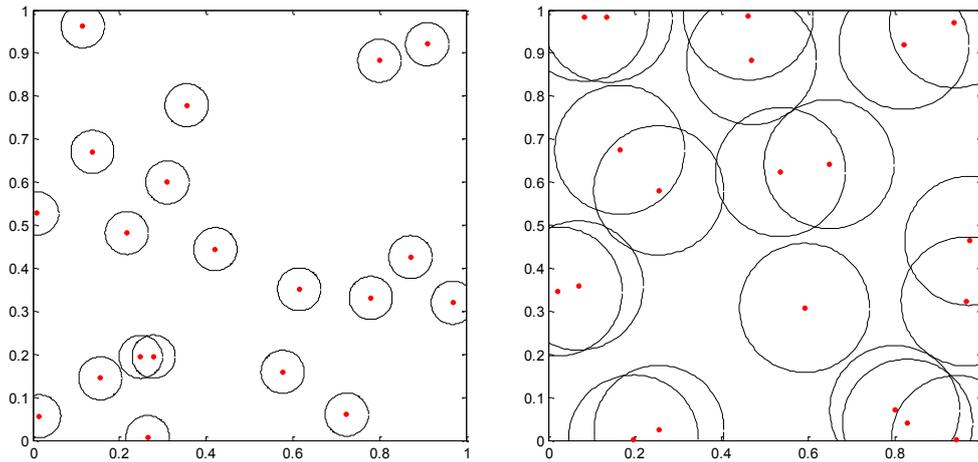


Figure 51. A gauche, un plan de Strauss à 20 points en dimension 2 de paramètres $R=0.1$ et $\gamma=0.1$.
A droite, un plan de Strauss de paramètres $R=0.3$ et $\gamma=0.1$

La figure ci dessous permet d'illustrer et de comprendre, lorsque le rayon choisi est grand, que la loi π rend plus vraisemblable les agglomérats qui correspondent à un nombre global d'interactions compatible avec la loi π .

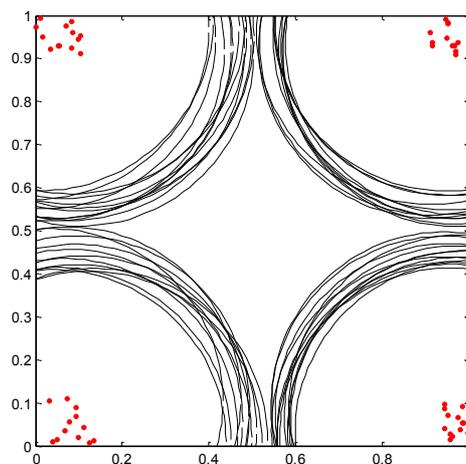


Figure 52. Un plan de Strauss à 50 points avec $R=0.8$ et $\gamma=0.1$

De même que pour le rayon d'interaction, il est important de bien fixer le paramètre de répulsion γ .

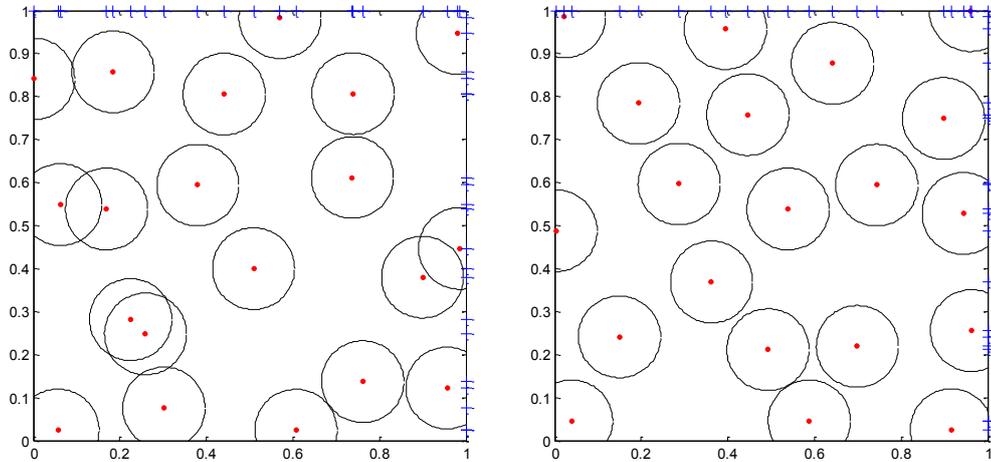


Figure 53. A gauche, un plan de Strauss de paramètres $R=0.19$ et $\gamma=0.1$. A droite, un plan de Strauss de paramètres $R=0.19$ et $\gamma=0.001$

La figure ci-dessus montre qu'il est plus aisé de générer une distribution répondant au critère de remplissage de l'espace avec un paramètre de répulsion faible. Cependant, un paramètre de répulsion trop faible contraint généralement les points à être alignés. Ce qui dégrade la répartition des points en projection, notamment sur les frontières du domaine expérimental comme le montre la Figure 53 pour le facteur X_1 au niveau de la valeur 1.

Ainsi, il est délicat de bien fixer le paramètre de répulsion afin de trouver un compromis entre un remplissage de l'espace optimal et une répartition de points uniforme sur les marges de dimension 1. C'est pourquoi, nous proposons dans la suite une variante qui permet de corriger ce problème en considérant également les interactions de paires sur les d projections.

Le rayon d'interaction est le paramètre le plus sensible à régler. Pour un critère donné, la meilleure solution serait sans doute de tabuler cette valeur selon le nombre de points et la dimension du problème.

Une fois les problèmes de paramétrage résolus, les plans de Strauss permettent de bien répondre à la question du remplissage de l'espace en grande dimension. Pour s'en convaincre, il suffit de calculer les critères intrinsèques tels, par exemple, le critère d'Arbre de Longueur Minimale (voir Chapitre 6).

Remarque. Il est facile de vérifier que pour toute loi initiale, la distribution finale de points se rapproche toujours de celle d'un processus de Strauss. Le choix de la loi initiale impacte uniquement sur la vitesse de convergence.

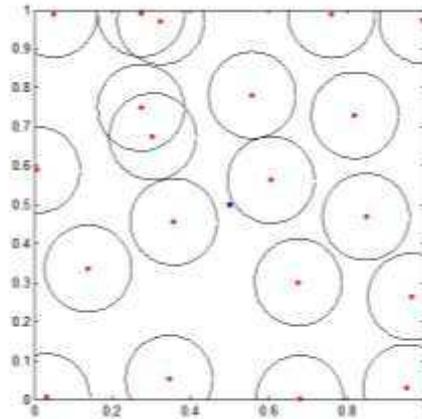


Figure 54. Un plan de Strauss à 20 points en dimension 2 obtenu à partir d'une distribution initiale (en bleu) où les points sont concentrés au centre du domaine

La figure ci-dessus montre un exemple où la distribution initiale (en bleu) se résume à être concentrée au centre de l'hypercube unité (masse de Dirac). La distribution finale observée (en rouge) se rapproche bien d'un processus de Strauss.

3.3.3. Extensions : processus ponctuel de Gibbs

Plusieurs généralisations du processus ponctuel de Strauss sont envisageables. Par exemple, on peut faire intervenir les triplets de points distants deux à deux de moins de R . Nous nous contenterons ici seulement des interactions de paires qui nous paraissent suffisantes pour atteindre nos objectifs.

3.3.3.1. Contraintes sur les marges

Il a été observé que le processus de Strauss ne permet pas toujours de satisfaire à une bonne répartition des points en projection sur les marges de dimension 1.

Une première extension naturelle consiste à considérer des interactions globales, *i.e.* en dimension d , ainsi que des interactions sur chaque projection afin de retrouver les bonnes propriétés des hypercubes latins. Cela fait intervenir deux nouveaux paramètres : γ_{1D} pour l'interaction et R_{1D} pour la longueur de l'intervalle d'influence sur les marges ou projections

unidimensionnelles (voir le paragraphe 3.3.3.2. pour l'expression générale de la densité du processus ponctuel associé).

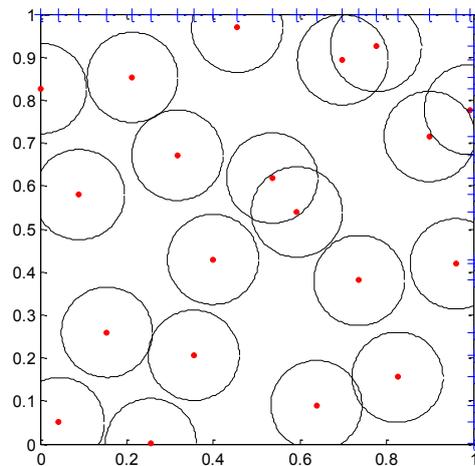


Figure 55. Un plan de Strauss à 20 points en dimension 2 ($R = 0.19$, $R1d = 0.0375$, $\gamma = 0.1$, $\gamma_{1D} = 0.01$)

L'exigence d'une bonne répartition des points en projection sur les axes factoriels permet de corriger l'effet de la grande dimension qui accorde moins de volume au centre du domaine. En effet, Morris et Mitchell (1992) disent à cet effet qu'en grande dimension les plans maximins (similaires aux plans de Strauss) de taille modérée (avec peu d'expériences comparativement à la dimension du problème) sont souvent concentrés sur les bords du domaine expérimental. En considérant des plans maximin avec des arrangements identiques aux hypercubes latins, on propose des plans avec un compromis entre un bon critère de maximin et des bonnes propriétés de projection dans chaque dimension. Ici, il en est de même.

Il est tout à fait possible d'exiger également de bonnes propriétés de projection sur des sous-espaces de dimension 2 ou plus. Même si l'obtention d'une bonne répartition sur les plans factoriels apporte un intérêt évident pour estimer les interactions d'ordre 2 entre les différents facteurs, seules les contraintes sur les marges de dimension un sont présentées ici.

De plus, nous verrons au Chapitre 7 que de vouloir une bonne répartition en grande dimension d et en projection sur les sous-espaces implique beaucoup trop de contraintes et affaiblit au moins un des deux objectifs selon les paramètres imposés.

3.3.3.2. Plans de Strauss-Gibbs

Les plans de Strauss sont issus d'une loi de probabilité qui rend plus vraisemblable un nombre relativement faible d'interactions globales (ainsi que sur les marges, voir paragraphe 3.3.3.1.). Pour pallier le phénomène d'agglomérats de points (voir paragraphe 3.3.2), notamment lorsque le rayon est inadapté, nous allons prendre en compte la distance réelle entre les points pour mesurer la force de l'interaction et non plus se contenter de savoir s'il y a interaction ou pas (sphères de rayon $R/2$ qui se rencontrent).

Cela nous amène à considérer la famille plus large des processus ponctuels dont la distribution π est de la forme :

$$\pi(x) \propto e^{-U(x)}$$

où

$$U(x) = \beta \sum_{i < j} \varphi(\|x^i - x^j\|) + \sum_{k=1}^d \beta_k \sum_{i < j} \varphi_k(\|x_k^i - x_k^j\|)$$

$$\beta = -\ln \gamma > 0 ; \quad \beta_k = -\ln \gamma_k > 0, \quad k = 1, \dots, d$$

- φ de $[0, +\infty[$ dans \mathbf{R} est une fonction continue décroissante telle que $\varphi(0) = 1$ et φ nulle à l'infini. On utilisera la famille ($\alpha > 0, R > 0$) de fonctions suivantes

$$\varphi(r) = \left(1 - \frac{r}{R}\right)^\alpha \quad \text{si } 0 \leq r \leq R \quad (\varphi = 0 \text{ pour } r \geq R)$$

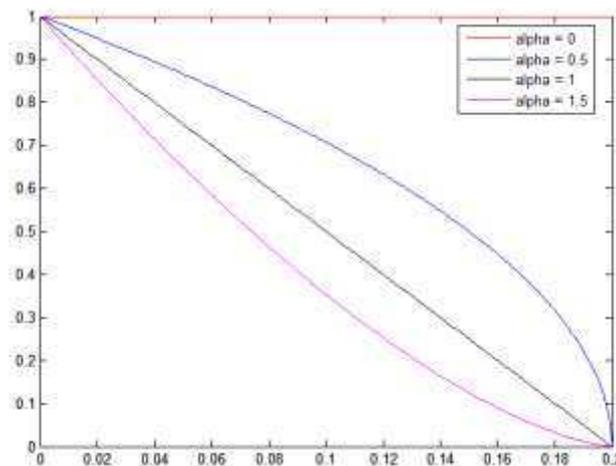


Figure 56. La fonction de potentiel puissance pour différents α et $R=0.2$

Il s'agit de processus ponctuels de Gibbs de potentiel U , de potentiels locaux les fonctions φ et φ_k qui correspondent uniquement à des interactions de paires (Ripley, Kelly 1977). Le cas $\alpha = 0$ (potentiel 0-1 ou dit du « tout ou rien ») correspond exactement aux processus de Strauss discutés jusqu'ici.

Cette extension réduit le risque de points agglomérés et améliore par exemple un critère de type maximin.

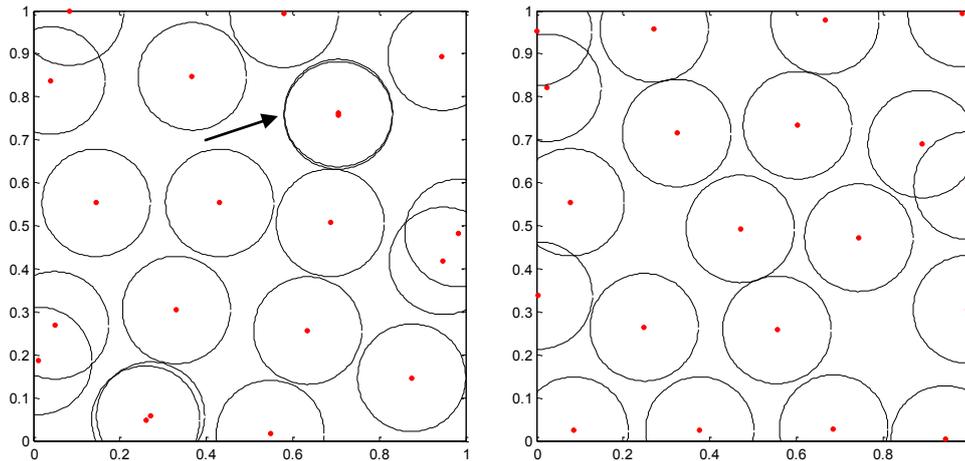


Figure 57. A gauche. Un plan de Strauss obtenu à partir d'un potentiel tout ou rien. A droite. Un plan de Strauss obtenu à partir d'un potentiel puissance $\alpha = 0.5$

La Figure 57 permet de bien comprendre les interactions que l'on privilégie. En effet, sur le plan de gauche généré avec un potentiel de type tout ou rien nous observons qu'il y a deux points de mêmes coordonnées (cercle repéré par la flèche). Généralement, ce phénomène n'apparaît pas quand on utilise un potentiel puissance qui favorise des interactions faibles comme le montre la figure de droite. Ainsi l'utilisation d'un potentiel puissance permet de mieux répondre à la question du remplissage de l'espace.

Cependant, des alignements persistent comme on peut l'observer sur la figure de droite pour le facteur X_2 au niveau de la valeur 0.02. Une solution est donc de combiner les contraintes sur les marges de dimension 1 avec le potentiel puissance afin d'obtenir une répartition uniforme sur les marges de dimension 1 et une répartition optimale au sens du remplissage de l'espace.

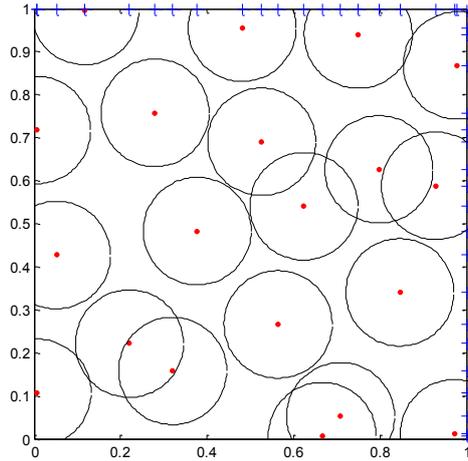


Figure 58. Un plan de Strauss-Gibbs combinant interactions globales et sur les marges avec un potentiel puissance de type puissance

3.3.3.3. Cas de l'hétérogénéité

En phase exploratoire, l'objectif recherché est de distribuer les points uniformément dans l'hypercube unité. Cependant, si nous avons une connaissance métier sur le phénomène, il peut être souhaitable de distribuer les points plus densément dans une certaine partie de l'hypercube unité. Aussi, il doit être possible de générer des points selon une distribution non plus uniforme mais hétérogène ou bien encore anisotrope.

On définit alors la distribution π sur l'espace E des plans d'expériences par

$$\pi(x) \propto \left(\prod_{i=1}^n \lambda(x^i) \right) e^{-U(x)}$$

où λ est une densité de probabilité facile à simuler, par exemple une densité gaussienne tronquée.

C'est la dernière extension envisagée dans ce chapitre. L'algorithme de MH qui a été retenu est le suivant :

Algorithme 5. Algorithme de Metropolis-Hastings pour générer des plans d'expériences dits de Strauss-Gibbs hétérogènes

Initialisation : choisir un plan d'expériences x^1, \dots, x^n selon une loi de probabilité donnée (loi uniforme par exemple)

POUR N = 1 à NMC

POUR k = 1 à n

Choisir une expérience x^i au hasard

Simuler y^i selon la densité $\lambda(\cdot)$ dans $[0,1]^d$

Accepter le changement $x^i = y^i$ avec la probabilité

$$a(x, y) = \min\left(1, e^{U(x)-U(y)}\right) \text{ où } y = \left(x^1, \dots, x^{i-1}, y^i, x^{i+1}, \dots, x^n\right)$$

Fin pour k

Fin pour N

Remarque. Dans la section 3.3.4 nous proposons une méthode originale et différente pour générer des plans selon une distribution donnée.

3.3.4. Ergodicité uniforme

Deux caractéristiques sont à la base de l'algorithme 4 (le plus général, plans de Strauss-Gibbs hétérogènes) :

- la proposition de changement d'une expérience prise au hasard, notée $Q(x, dy)$
- la probabilité $a : E \times E \rightarrow [0,1]$ d'accepter le changement $x \mapsto$ proposé

De manière précise, on a ici

(i) $Q(x, dy) = \sum_{i=1}^n \frac{1}{n} \lambda(y^i) 1_{y^{(-i)}=x^{(-i)}} dy^i$ où $y^{(-i)} = x^{(-i)}$ signifie que $y_j = x_j$ pour $j \neq i$

(ii) $a(x, y) = \min\left(1, e^{U(x)-U(y)}\right)$

Si δ_x désigne la mesure de Dirac en x , chaque transition élémentaire $P: E \times E \rightarrow [0,1]$ de cet algorithme s'écrit :

$$P(x, dy) = a(x, y)Q(x, dy) + \underbrace{\left[\int (1-a(x, z))Q(x, dz) \right]}_{p_x} \delta_x(dy)$$

On remarquera que p_x est la probabilité de rester en x partant de x puisqu'ici $Q(x, \{x\}) = 0$.

Notons

$$K(x, dy) = P^n(x, dy) = P(X_{N+1} \in dy^1 \dots dy^n | X_N = x)$$

le noyau de transition associé au balayage complet des n expériences à chaque itération N de l'algorithme. Par construction, la suite des plans d'expériences $\mathbf{x}_{(0)}, \mathbf{x}_{(1)}, \dots, \mathbf{x}_{(NMC)}$ ainsi générés est la réalisation d'une chaîne de Markov $(X_N)_{N \geq 0}$ de probabilité ou noyau de transition K .

Vérifions que le mécanisme de transition P satisfait la condition de π -réversibilité

$$\forall x, y \in E, \quad \pi(dx)P(x, dy) = \pi(dy)P(y, dx)$$

Soit B un borélien de $E \times E$. On a

$$\int 1_B(x, y) \pi(dx) P(x, dy) = \int 1_B(x, y) \pi(x) a(x, y) Q(x, dy) dx + \int_E \pi(x) p_x \left(\int_E 1_B(x, y) \delta_x(dy) \right) dx$$

et donc,

$$\int 1_B(x, y) \pi(dx) P(x, dy) = \int 1_B(x, y) \pi(x) a(x, y) Q(x, dy) dx + \int_E 1_B(x, x) \pi(x) p_x dx.$$

Par construction, Q et a vérifient

$$Q(x, dy) \lambda(x) dx = Q(y, dx) \lambda(y) dy \text{ et } e^{-U(x)} a(x, y) = e^{-U(y)} a(y, x).$$

Ce qui implique

$$\int 1_B(x, y) \pi(dx) P(x, dy) = \int 1_B(x, y) \pi(y) a(y, x) Q(y, dx) dy + \int 1_B(y, y) \pi(y) p_y dy.$$

Finalement,

$$\int 1_B(x, y) \pi(dx) P(x, dy) = \int 1_B(x, y) \pi(dy) P(y, dx).$$

C'est la condition de π -réversibilité du mécanisme de transition P .

Il est immédiat de constater que la propriété de π -réversibilité sur P définie ci-dessus implique que π est une loi invariante pour P (et donc aussi pour K).

La minoration de K par une constante strictement positive permet d'assurer la propriété de mélange de la chaîne de Markov. Pour cela, on vérifie que

$$K(x, dy) \geq n! \prod_{k=1}^n \left(\frac{1}{n} \times \min \left(1, \frac{\pi(u_k)}{\pi(u_{k-1})} \right) \right) \times dy^1 \dots ,$$

où u_k est la suite des plans itérés avec $u_0 = x$ et $u_n = y$.

Dans le cas du processus ponctuel de Strauss usuel (potentiel tout ou rien), la minoration de $\frac{\pi(u_k)}{\pi(u_{k-1})} = \gamma^{n(u_k) - n(u_{k-1})}$ par γ^{n-1} correspond à la configuration la plus défavorable au mélange qui est représentée sur la figure ci-dessous.

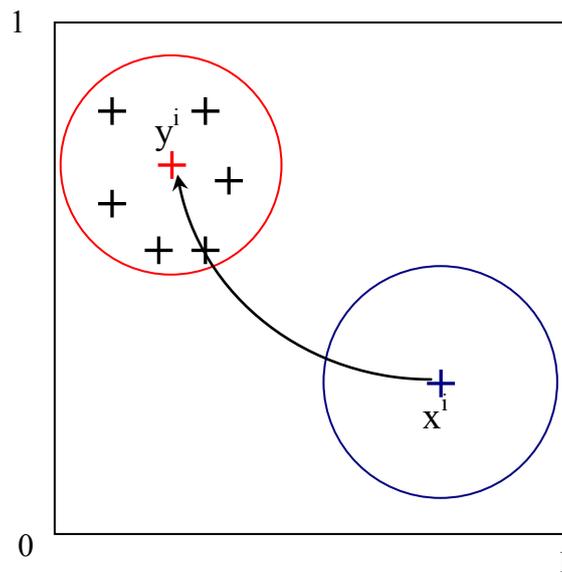


Figure 59. Configuration la plus défavorable au mélange

On obtient donc $K(x, dy) \geq c dy^1 \dots$ avec $c = \frac{n!}{n^n} (\gamma^{n-1})^n > 0$.

On peut vérifier également qu'il existe une minoration analogue avec un potentiel généralisé mais aussi en considérant le processus ponctuel défini au paragraphe 3.3.3.1. .

Nous venons de voir que les conditions d'application du théorème énoncé au paragraphe 3.2.2 sont bien toutes réunies.

3.4. Conclusions

L'utilisation des méthodes McMC dans le contexte de la planification d'expériences permet de construire de nouveaux plans spécifiés à partir d'une loi de Gibbs. Cette approche offre une grande souplesse dans la mesure où l'on peut facilement jouer sur la loi de Gibbs et sa représentation en termes de potentiels locaux d'énergie pour imposer des propriétés, par exemple de remplissage.

Par ailleurs la méthodologie McMC constitue une alternative intéressante à l'approche statistique classique qui consiste à travailler sur des réalisations indépendantes d'une même loi. Ainsi, si l'on souhaite étudier un critère d'optimalité pour une famille de plans particulière, il est possible de calculer l'intervalle de confiance sur la valeur du critère à partir d'une seule simulation McMC (théorème central limite).

Enfin de nombreuses extensions pour la planification d'expériences numériques sont envisageables. Quelques unes d'entre elles sont présentées dans le Chapitre 4. Nous introduirons la notion de distance à densité et de processus ponctuels associés pour générer des points de manière préférentielle (hétérogène). Cette méthode peut s'avérer particulièrement intéressante en grande dimension où les plans uniformes concentrent leurs points naturellement sur les bords du domaine.

Nous présenterons également des plans qui prennent en compte des contraintes spatiales.

Chapitre 4

Vers des extensions...

L'intérêt de ce chapitre est d'apporter divers compléments aux plans d'expériences permettant de mettre en évidence la capacité de certains plans à s'adapter aux problèmes éventuels.

Jusqu'à présent nous sommes restés dans le contexte bien général de la planification d'expériences numériques, notamment celui de l'uniformité de la distribution dans un hypercube. Il ne faut pas écarter cependant le fait que nous puissions être confrontés à des objectifs tels que :

- la distribution de points dans un domaine défini à l'aide de contraintes, particulièrement utile dans les problèmes d'optimisation,
- distribuer des points de manière hétérogène (préférentielle),
- générer des points de manière itérative : par exemple effectuer un zoom sur une certaine partie de l'espace ou bien encore déplacer la zone de recherche tout en conservant les points déjà générés précédemment,
- etc.

Dans ce chapitre, nous commençons par proposer une méthode pour générer des points de manière préférentielle. Cette méthode est basée sur le processus de Strauss discuté dans le chapitre précédent. Ensuite, nous exposons deux types d'approches pour distribuer les points dans des domaines contraints. Ce qui nous amène à présenter une nouvelle méthode, elle aussi basée sur le processus de Strauss. Enfin, nous présentons quelques exemples de plans itératifs.

4.1. Distance à densité et plans d'expériences associés

Dans les chapitres précédents, nous avons exposé différentes méthodes pour générer des points uniformément dans le domaine expérimental. Cet objectif reste essentiel lorsque nous n'avons aucune connaissance *a priori* du code afin de mieux capter les éventuelles non-linéarités.

Cependant, le phénomène complexe étudié peut apprendre des premières informations. Par exemple, on peut savoir que le phénomène varie de façon non-linéaire au centre du domaine. Il doit être particulièrement facile d'intégrer à la démarche d'analyse certaines connaissances métiers.

Ainsi, on cherche à générer une distribution hétérogène et/ou anisotrope de points dans l'espace. Nous avons déjà proposé au paragraphe 3.3.3.3. un moyen d'y parvenir en jouant sur l'intensité d'un processus ponctuel de Gibbs.

Si le modèle de simulation retenu est un processus de Strauss, basé sur un mécanisme de répulsion lié à une distance d'interaction, une technique consiste à doter l'espace d'une métrique¹⁰ hétérogène. L'idée revient alors à considérer une fonction distance variable dans l'espace (voir Figure 60).

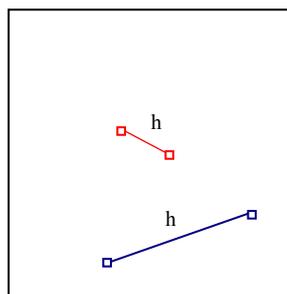


Figure 60. Exemple de paires de point qui sont à la même distance h

Une autre motivation de cette méthode est de corriger le fléau de la grande dimension en densifiant les points au centre du domaine expérimental. Par exemple, l'utilisation d'une distance à densité gaussienne (voir paragraphe 4.1.3) accordera beaucoup plus de points au centre du domaine que l'utilisation d'un plan de Strauss usuel en grande dimension.

¹⁰ Nous verrons ultérieurement qu'en fait ce n'est pas une métrique au sens strict car elle ne vérifie pas les axiomes usuels.

4.1.1. Distance à densité

On considère une fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ positive suffisamment régulière de sorte que f soit intégrable sur tout segment.

On introduit la notion de distance à densité.

Définition.

Pour 2 points A et B , on pose :

$$\delta(A, B) = \int_A^B f(M) dl$$

intégrale curviligne de f sur le segment AB où M décrit le segment AB et où dl désigne un élément de longueur.

L'application δ ainsi définie sera appelée **distance de densité la fonction f** .

En notant a et b les coordonnées respectives des points A et B , on a

$$\delta(A, B) = \int_0^l f(x_s) ds$$

où s est l'abscisse curviligne et $x_s = a + su$ avec $u = \frac{b-a}{\|b-a\|}$ vecteur unitaire ($a \neq b$) et où

$l = \|b-a\|$ est la distance usuelle entre A et B .

On note encore $\delta(A, B) = \delta(a, b)$.

Remarque. La distance euclidienne usuelle correspond au cas f constante égale à 1.

Considérons un exemple en dimension $d=1$. Pour cela, on choisit pour f la densité gaussienne de moyenne $\mu=0$ et d'écart-type $\sigma=1$ (cas d'une densité de probabilité).

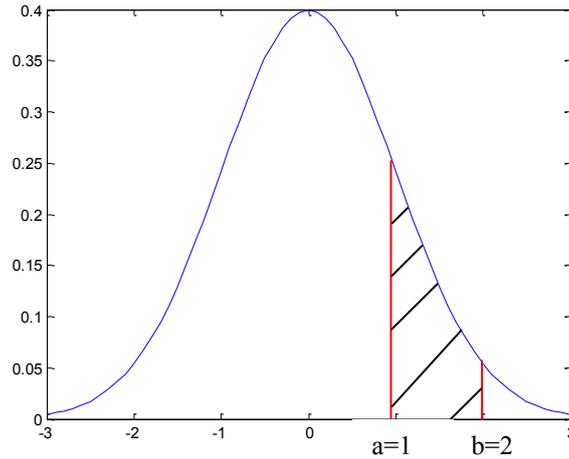


Figure 61. Loi normale de moyenne nulle et d'écart-type 1 et distance $\delta(a, b)$ associée

On a simplement

$$\delta(a, b) = |\phi(b) - \phi(a)|$$

où $\phi(x) = \int_{-\infty}^x \frac{e^{-u^2/2}}{\sqrt{2\pi}} dx$ est la fonction de répartition de la loi normale standard.

Dans ce cas simple ($d = 1$), δ vérifie les axiomes usuels d'une distance :

$$\delta(a, b) \geq 0$$

$$\delta(a, b) = 0 \Leftrightarrow a = b$$

$$\delta(a, b) = \delta(b, a)$$

$$\delta(a, b) \leq \delta(a, c) + \delta(c, b)$$

Mais contrairement à la distance usuelle, on n'a pas la propriété d'invariance suivante :

$$\delta(a, b) \neq \delta(a + c, b + c).$$

Sur l'exemple ci-dessous, la « boule » de centre 0 et de rayon $\frac{1}{4}$ est l'intervalle interquartile $[a_0, b_0] = [-0.67, 0.67]$ et la boule de centre $\frac{1}{2}$ et de même rayon est $[a_1, b_1] = [-0.15, 1.57]$ différente de $1/2 + [a_0, b_0]$ (voir Figure 62).

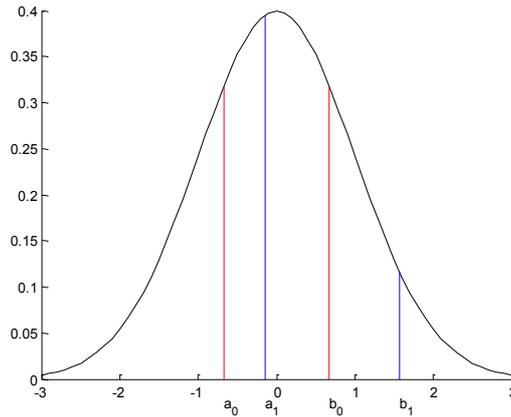


Figure 62. Loi normale centrée-réduite et représentation des boules $B(0,1/4)$ et $B(0.5,1/4)$

Si l'on considère maintenant la boule de centre 1 et de rayon $1/4$, on obtient un intervalle illimité à droite $[0.23, +\infty[$:

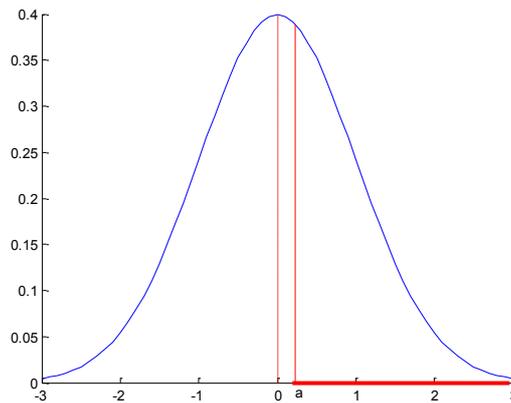


Figure 63. Loi normale centrée réduite et représentation de la boule $B(1,1/4)$

Ainsi, on voit que la distance entre deux points dépend fortement de leurs positions dans l'espace. A l'infini, les points ont tendance à être éloignés relativement à la distance usuelle alors qu'au centre de la distribution les points sont plus proches. A noter que pour σ grand, on se rapproche de la distance usuelle.

Examinons maintenant les propriétés de la distance à densité introduite ici. Elle vérifie clairement les axiomes de positivité et de symétrie. La propriété de séparation peut être en défaut. Quant à l'inégalité triangulaire elle n'est plus valable dans le cas général dès que la dimension est supérieure ou égale à 2.

Nous verrons que seule la propriété de symétrie est nécessaire pour la suite.

4.1.2. Processus ponctuels et distance à densité

On cherche à généraliser le processus de Strauss défini par

$$\left\{ \begin{array}{l} \pi(x) \propto \gamma^{s(x)} \\ s(x) = \sum_{i < j} 1_{\|x^i - x^j\| \leq R} \end{array} \right.$$

où les paramètres du processus sont γ et R comme dans la section 3.3.1.

On rappelle que $s(x)$ compte le nombre d'interactions et que les points x^i et x^j interagissent si et seulement si la distance usuelle qui les sépare est inférieure ou égale à R . On notera que chaque interaction n'est comptée qu'une seule fois par symétrie.

L'idée naturelle pour générer des plans non forcément uniformes est donc de remplacer la distance usuelle par une distance à densité. On définit alors le processus de Strauss par :

$$\left\{ \begin{array}{l} \pi(x) \propto \gamma^{s(x)} \\ s(x) = \sum_{i < j} 1_{\delta(x^i, x^j) \leq R} \end{array} \right.$$

où δ désigne une distance à densité f .

A noter que c'est la propriété de symétrie de δ qui rend correcte cette définition.

Concernant la simulation, on a toujours une expression simple pour le calcul du rapport de vraisemblance dans l'algorithme de Metropolis-Hastings :

$$\frac{\pi(y)}{\pi(x)} = \gamma^{s(y) - s(x)}.$$

Dans le cas général, on considérera le processus ponctuel de Gibbs de potentiel (voir paragraphe 3.3.3.2.)

$$U(x) = \beta \sum_{i < j} \varphi(\delta(x^i, x^j)).$$

L'algorithme de Metropolis-Hastings reste inchangé à celui décrit au paragraphe 3.3.1. La seule modification est au niveau du calcul de la distance entre 2 points. Aussi, au niveau de la convergence de l'algorithme, les hypothèses (π -invariance et condition de mélange) du théorème de la section 3.2.2 sont encore vérifiées.

Ce procédé permet donc d'obtenir de nouveaux plans d'expériences construits à partir de distance à densité. La section suivante décrit concrètement la génération de plans d'expériences dans le cas particulier d'une distance à densité gaussienne.

4.1.3. Distance à densité gaussienne

On considère le cas où f est une densité (de probabilité) gaussienne sur \mathbb{R}^d

$$f(x) = \frac{1}{\sqrt{2\pi}^d} \frac{1}{\sqrt{\det \Gamma}} e^{-\langle x-\mu, \Gamma^{-1}(x-\mu) \rangle / 2}$$

avec Γ une matrice de variances-covariances définie positive.

On montre par le calcul (voir annexe D) que

$$\delta(a, b) = k \times e^{-\rho^2/2} \times \sigma_u \times \left[\phi\left(\sigma_u \langle a - \mu, \Gamma^{-1}u \rangle + l/\sigma_u\right) - \phi\left(\sigma_u \langle a - \mu, \Gamma^{-1}u \rangle\right) \right]$$

avec

$$l = \|b - a\|, \quad u = \frac{b - a}{\|b - a\|}, \quad \sigma_u^2 = \frac{1}{\langle u, \Gamma^{-1}u \rangle}$$

$$\rho^2 = \langle a - \mu, \Gamma^{-1}(a - \mu) \rangle - \frac{\langle a - \mu, \Gamma^{-1}\mu \rangle^2}{\langle \mu, \Gamma^{-1}\mu \rangle}$$

et où $\phi(z) = \int_{-\infty}^z \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt$ désigne la fonction de répartition de la loi normale centrée réduite.

Enfin, k est une constante qui ne dépend que des caractéristiques de la densité f choisie (voir annexe D). En pratique, la métrique est définie à un changement d'échelle près en fixant la valeur de k à 1 (métrique équivalente).

On notera que σ_u est l'écart-type de la loi normale unidimensionnelle induite sur la droite AB et que ρ est la distance de la droite AB au centre μ de la loi pour la métrique euclidienne induite par le produit scalaire $\langle x, \Gamma^{-1}y \rangle$.

L'exemple ci-dessous représente un plan de Strauss de distance à densité gaussienne avec un faible écart-type afin de favoriser la présence de points au centre du domaine.

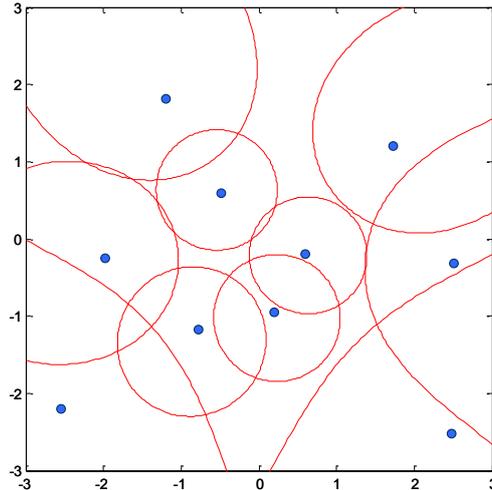


Figure 64. Un plan de Strauss de distance à densité gaussienne isotrope ($R=0.28$, $\sigma=2$)

Attention, contrairement à ce qui a été vu jusqu'ici, on représente les boules de rayon R et non $R/2$. En effet, l'inégalité triangulaire n'étant pas vérifiée dans tous les cas il devient impossible d'observer les interactions à l'aide des sphères de rayon $R/2$. Ainsi, une interaction entre deux points se produit exactement lorsque chacun d'eux est dans le voisinage de l'autre. Il ne suffit pas que les boules s'intersectent. Par exemple, sur la figure ci-dessus on ne compte aucune interaction.

La figure suivante quant à elle représente une distribution où l'on peut voir des interactions entre les points.

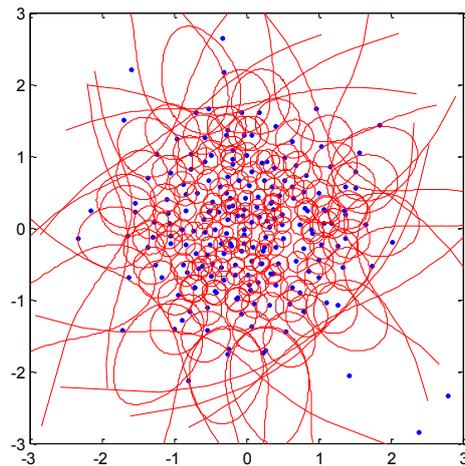


Figure 65. Un plan de Strauss de distance à densité gaussienne isotrope avec interactions entre les points

On observe différentes formes de voisinages¹¹, quasi-sphériques au centre, ovales loin du bord et non bornés. A noter aussi que ces voisinages ne sont pas forcément convexes.

Il est également possible de construire des plans anisotropes en fixant par exemple des écarts-types différents selon les directions (corrélation nulle sur la figure suivante).

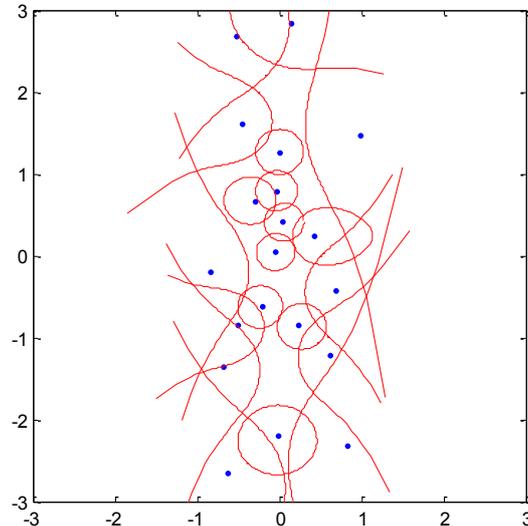


Figure 66. Un plan de Strauss de distance à densité gaussienne anisotrope

Il est immédiat de constater qu'avec un écart-type large, on retrouve un plan de Strauss usuel :

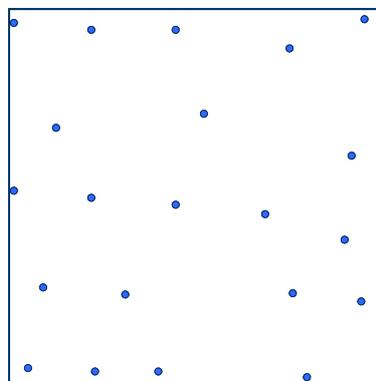


Figure 67. Un plan à distance à densité gaussienne avec $\sigma=1000$ et $R=0.4$

Ces plans ont été utilisés au cas d'étude de dimension 3 (voir paragraphe 8.3.1.). Les résultats étant de qualité moyenne, il serait probablement plus intéressant de les utiliser dans un problème de grande dimension pour corriger le fléau de la grande dimension.

¹¹ Le terme voisinage est impropre dans la mesure où l'on ne peut pas définir une vraie topologie (axiomes des voisinages ou des ouverts non vérifiés en général)

4.2. Plans d'expériences contraints

L'objectif principal de cette section est de répondre à la question suivante : « Est-il possible de construire des SFD dans un domaine quelconque, pas forcément convexe, sans rapport avec les hypercubes traditionnels ? ».

En effet, dès que le domaine de variation des paramètres n'est plus un hypercube, les plans étudiés précédemment ne conviennent pas nécessairement. Par exemple, l'intérêt d'utiliser des hypercubes latins et des tableaux orthogonaux n'est plus évident.

Cette question est en fait à l'origine de nombreux problèmes d'optimisation sous contraintes, qui induisent des portions d'espace interdites. Afin d'y répondre, nous proposons ici deux approches :

- une approche *a posteriori* qui consiste à générer des plans adaptés à cette approche dans l'hypercube et à retenir uniquement la région qui se limite au domaine d'intérêt
- une approche *a priori* qui consiste à travailler directement sur le domaine de variations possibles de paramètres.

4.2.4. Approche *a posteriori*

Si le domaine d'intérêt peut être inscrit dans un hypercube d'entourage, l'idée est de générer un SFD approprié dans cet hypercube et de conserver uniquement les points qui se situent dans la région d'intérêt. Il faut être bien conscient que cette approche peut altérer les qualités du plan résultant, ne serait-ce que, par exemple et de façon évidente, sa discrèpance. Ainsi, cette approche nécessite bien souvent d'améliorer les plans à l'aide de critères. Trosset (1999) introduit à cet effet une variante des plans maximin. Les plans de Strauss sont également une bonne alternative à ce problème.

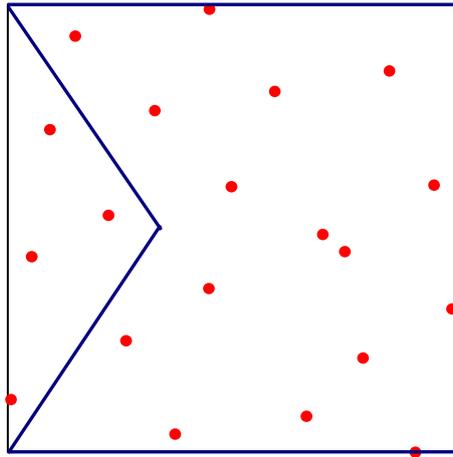


Figure 68. Un plan de Strauss généré dans $[0,1]^2$ puis on conserve les points dans le domaine d'intérêt en bleu

Cette approche de filtrage *a posteriori* peut aussi devenir problématique lorsque les contraintes amènent à réduire considérablement la région d'intérêt et qu'il devient difficile alors de conserver des points dans celle-ci. L'approche décrite dans la section suivante permet de pallier cet inconvénient.

4.2.5. Approche *a priori* : Plans Step-by-Step (SbS)

On travaille directement sur le domaine des variations possibles des paramètres. Dans le cas où le domaine peut être défini précisément et explicitement, cette méthode est facilement envisageable et n'impose pas de grandes modifications aux méthodes détaillées précédemment.

Dans le cas contraire, il nous faut trouver une solution et éliminer la phase « tirer un point au hasard dans le domaine » des algorithmes de génération de SFD qui devient bien souvent inenvisageable. C'est pour répondre à cette question que nous avons introduit les plans Step-by-Step (SbS).

Principe. L'idée est de partir d'un (ou de quelques) point(s) puis d'ajouter des points selon une dynamique globale qui tend à les éloigner les uns des autres d'un rayon R avec une intensité plus ou moins forte γ .

Remarque. γ et R sont les mêmes paramètres que dans un processus de Strauss.

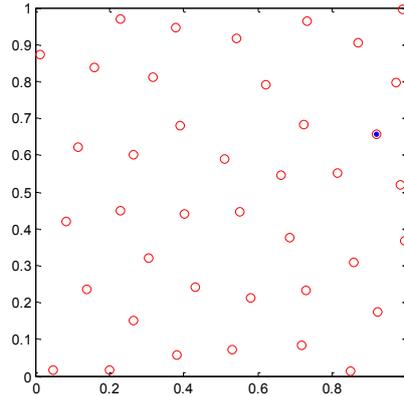


Figure 69. Une distribution de 40 points en dimension 2 obtenue à partir du point bleu et rouge

Algorithme 4. Génération de plans Step-by-Step

Fixer n le nombre d'expériences souhaitées, γ et R . D_c est le domaine défini par des contraintes.

Choisir une (ou quelques : \tilde{i}) expériences $x^1, \dots, x^{\tilde{i}} \in \mathcal{U}_c$ (par une méthode d'optimisation par exemple)

$rejet \leftarrow 0$

TANT QUE ($\tilde{i} \leq n$ **ET** $rejet < 10^5$)

Choisir une expérience x^i au hasard avec $i \in \{1, \dots, \tilde{i}\}$

Simuler y^i sur la sphère de rayon R et de centre x^i

TANT QUE ($y^i \notin D_c$)

$rejet \leftarrow rejet + 1$

Simuler y^i sur la sphère de rayon R et de centre x^i ¹²

Fin tant que

Accepter de remplacer x^i par y^i avec la probabilité

$$a(x, y) = \min\left(1, \frac{\pi(y)}{\pi(x)}\right) \text{ où } y = (x^1, \dots, x^{i-1}, y^i, x^{i+1}, \dots, x^{\tilde{i}})$$

SI y^i est accepté **ALORS** $\tilde{i} \leftarrow \tilde{i} + 1$ **Fin si**

Fin tant que

¹² Il est également possible de simuler un nouveau point x^i au lieu de resimuler un point y^i .

Remarque. Cette approche peut être considérée comme un processus de Strauss local.

Si le nombre de rejets explose, alors l'hypothèse faite sur la valeur du rayon lors de l'initialisation n'est vraisemblablement pas la bonne. Dans ce cas on procède, par exemple, par dichotomie pour trouver la valeur de R la mieux adaptée à notre problème.

Généralement, on opère toujours par dichotomie sur la valeur du rayon pour s'assurer de ne pas considérer un rayon trop petit.

Quelques extensions peuvent également être apportées à cette approche. Par exemple, on pourrait avoir une stratégie de pénalisation avec une probabilité d'acceptation sur le respect des contraintes. Ce qui permettrait parfois de prendre des raccourcis par le domaine non réalisable et d'éviter une stagnation de l'algorithme. Cette extension reste envisageable s'il est possible de calculer la distance au respect des contraintes.

4.3. Plans itératifs en phase exploratoire : zoom et shift

Une fois le plan initial construit, nous pouvons être amenés à considérer seulement une portion du domaine d'expérimentation. Par exemple, dans le cas où l'on veut résoudre un problème d'optimisation, on se limite à la portion du domaine où il est fortement probable de trouver un minimum ou un maximum. On itère alors ce procédé de génération de plans sur chaque sous-région issue chacune d'une partie du domaine, cela consiste à zoomer sur une partie du domaine expérimental. Dans ce type de procédé, il est important pour chaque région concernée de conserver les points générés auparavant afin de ne pas perdre d'information.

Cette section n'a pas pour objectif de présenter des grands fondements mathématiques mais uniquement d'illustrer les possibilités de certains plans d'expériences discutés auparavant.

4.3.1. Plans itératifs : zoom

Le plan initial est un plan à entropie maximale à 20 points en dimension 2 généré à l'aide d'un variogramme sphérique de portée égale à 0.85.

Nous souhaitons faire une première itération afin d'affiner la zone définie par $[0.2, 0.8] \times [0.3, 0.6]$ en y ajoutant 30 points.

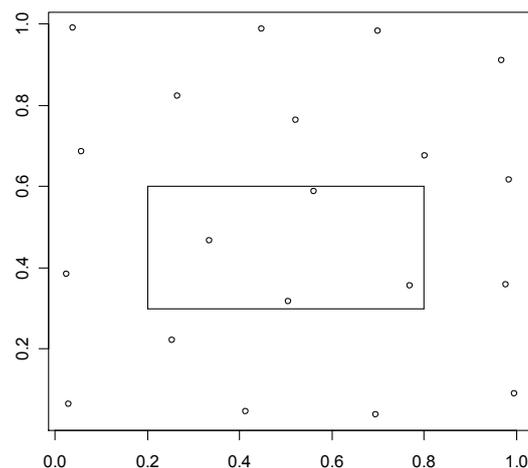


Figure 70. Un plan à entropie maximale à 20 points en 2D où l'on veut affiner la zone définie par le pavé

On génère un plan à entropie maximale tout en considérant et en conservant les 4 points préexistants du nouveau domaine d'étude. C'est-à-dire que si l'on est amené à tirer un point simulé lors de la phase précédente, on ne le considère pas et on va donc ainsi retirer un point jusqu'à ce que l'on choisisse un point « déplaçable ». On obtient :

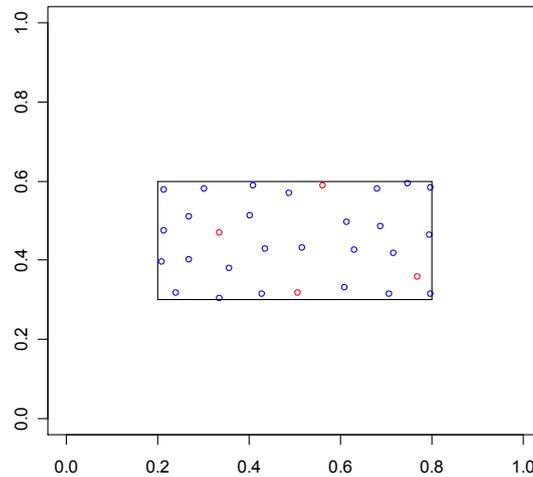


Figure 71. Représentation de l'échantillonnage obtenu dans la zone choisie en ayant considéré les simulations déjà effectuées

Cette méthodologie s'adapte également très bien aux plans de Strauss ; plus généralement à toutes les méthodes itératives de planification telles que les algorithmes d'échange par exemple.

4.3.2. Déplacement de zone de recherche

Lors d'une recherche d'optimum, on peut être amené au cours de la première itération à déplacer la zone de recherche. Chaque zoom ou chaque déplacement se fait en conservant à la fois les points déjà simulés et les propriétés initiales du plan (déterminant maximal ou respect du processus de Strauss à l'état stationnaire).

Voici un exemple de déplacement de zone de recherche.

On part de la zone de recherche initiale $[0.1, 0.5] \times [0.7, 0.95]$. Admettons que l'optimum que nous trouvons dans cette zone de recherche se trouve en $(0.47, 0.7)$ c'est-à-dire au bord du domaine. On aimerait savoir s'il n'y peut pas y avoir un meilleur optimum dans le voisinage de cette zone d'intérêt. On s'avise à faire une recherche d'optimum dans une autre

zone de recherche, par exemple $[0.45, 0.8] \times [0.5, 0.8]$, tout en conservant les points déjà générés auparavant.

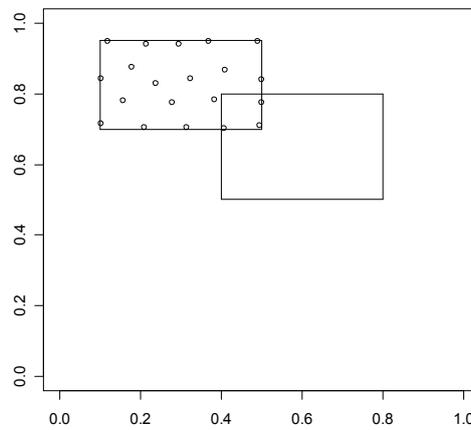


Figure 72. Plan initial avec la représentation de la nouvelle zone de recherche

On génère des points avec la méthode, on obtient :

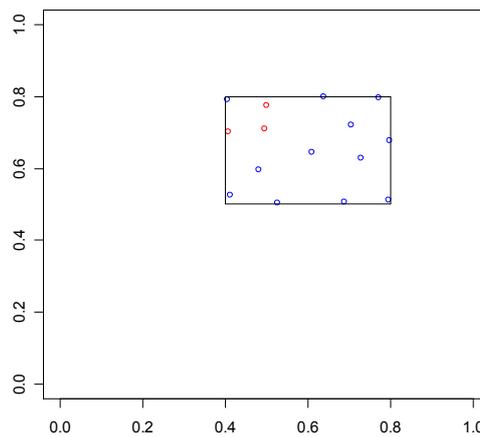


Figure 73. Représentation de l'échantillonnage de la nouvelle zone de recherche avec en rouge les expériences initiales et en bleu les expériences itérées une à une

Il est intéressant de considérer les points proches des bords du nouveau domaine afin d'éviter de faire des simulations trop proches de celles déjà effectuées auparavant. L'idée consiste à rajouter une tolérance sur les coordonnées des points.

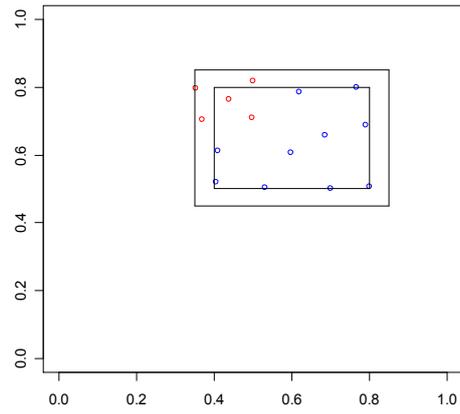


Figure 74. Représentation de l'échantillonnage dans la nouvelle zone de recherche où l'on a considéré les points de la zone et les points d'une « enveloppe » de cette zone

PARTIE III.

ANALYSE DES PLANS D'EXPERIENCES

Cette partie est consacrée à l'analyse intrinsèque des plans d'expériences ainsi qu'à de nouveaux outils pour étudier les propriétés des plans.

Le premier chapitre présente un outil graphique pour vérifier l'hypothèse d'uniformité du plan, par balayage dans toutes les directions.

Le second chapitre introduit un outil basé sur une technique de théorie des graphes afin de savoir vers quel type de distribution (aléatoire, amas, gradient, grille, ...) s'oriente le plan étudié.

Chapitre 5

Un nouveau critère statistique en forme de radar pour la mesure de l'uniformité des plans d'expériences et de leur projection sur les sous-espaces

Dans ce chapitre, nous introduisons une statistique construite sur les projections de dimension 1 pour tester l'uniformité d'un plan d'expériences. Le choix de la dimension 1 est lié à la difficulté d'obtenir la loi théorique des projections sur un espace de dimension supérieure. La restriction à la dimension 1 présente cependant l'avantage d'offrir un outil de visualisation simple basé sur le principe d'un radar. En représentant la valeur de la statistique selon toutes les directions, on obtient une courbe (ou une surface) paramétrée permettant d'identifier les défauts d'uniformité d'un plan au regard de ses projections sur les droites.

Le chapitre est structuré de la façon suivante. Dans un premier temps, on définit la nouvelle statistique et l'outil de visualisation associé, appelé radar d'uniformité, et en donnons quelques propriétés. Puis, on montre des applications du radar à des plans SFD. Une amélioration est alors apportée dans la section 5.3, où l'on définit une statistique globale qui ne dépend pas d'un axe de rotation particulier. Enfin, nous terminons par une discussion permettant de préciser le cadre d'application du radar d'uniformité.

5.1. Le radar d'uniformité

Dans le cadre de l'exploration d'un code de calcul, considérons un plan d'expériences uniforme sur un domaine cubique $\Omega = [-1, 1]^d$. Notons x^1, \dots, x^n les points expérimentaux, et (H_0) l'hypothèse : « x^1, \dots, x^n ont été générés par des tirages aléatoires indépendants selon la loi uniforme dans Ω ». Si le code de calcul ne dépend que d'une seule direction principale, il devient important que les projections sur cet axe soient le mieux réparties possible. Plus généralement, notons L_a la droite engendrée par un vecteur unitaire $a = (a_1, \dots, a_d)$ de Ω , et μ_a la loi de probabilité des projections de x^1, \dots, x^n sur L_a . Idéalement, on pourrait donc vouloir que pour toutes les directions a la loi μ_a soit uniforme. Cependant ceci n'est pas réaliste lorsque L_a n'est pas un axe de coordonnées. Par exemple, dans le cas du domaine cubique $[-1, 1]^2$, les points projetés seront en plus grande quantité sur la partie centrale de l'axe de projection, comme on peut l'observer sur la Figure 75.

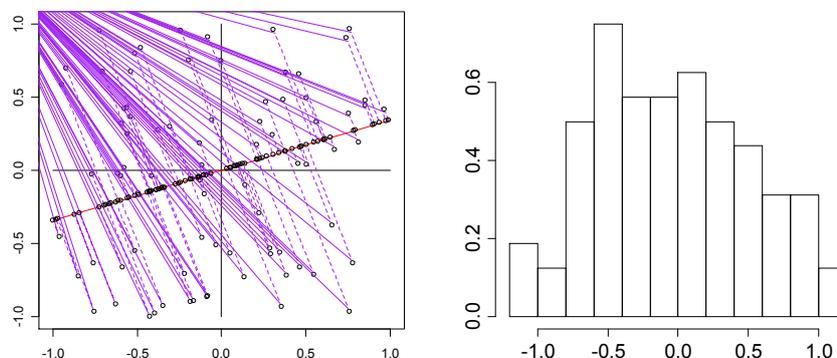
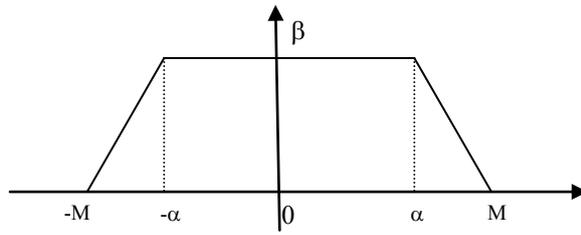


Figure 75. A gauche, les projections des points sur un axe L_a . A droite, l'histogramme des projections

Plus précisément, la répartition des projetés peut se diviser en 3 zones, définies par la projection des sommets du domaine. La loi μ_a admet la densité en forme de trapèze représentée ci-dessous. Les sommets du trapèze correspondent aux sommets du carré Ω projetés sur l'axe L_a avec $a = (\cos \theta, \sin \theta)$.



avec :

$$M = |\cos\theta| + |\sin\theta|$$

$$\alpha = ||\cos\theta| - |\sin\theta||$$

$$\beta = \frac{1}{M + \alpha}$$

Figure 76. La loi des projections pour un domaine cubique en dimension 2

Dans le cas général, la projection sur L_a est une combinaison linéaire de variables aléatoires indépendantes de loi uniforme, ce qui ramène le calcul de sa loi à un problème classique de probabilités abordé dès le XVIII^{ème} siècle par Lagrange (voir la discussion dans Elias et Shiu (1987) à ce sujet). Si on note $X = (X_1, \dots, X_d)$ avec X_1, \dots, X_d des variables aléatoires indépendantes et identiquement distribuées de loi uniforme sur $[-1, 1]$, et Z la projection de X sur L_a avec $a_j \neq 0 \forall j \in \{1, \dots, d\}$, alors la fonction de répartition de Z est donnée par

$$F_a(z) = \left(\prod_{j=1}^d \frac{1}{2a_j} \right) \times \sum_{s \in \{-1, 1\}^d} \varepsilon(s) \frac{(z + s.a)_+^d}{d!}$$

où $s = (s_1, \dots, s_d) \in \{-1, 1\}^d$ sont les sommets de l'hypercube $\Omega = [-1, 1]^d$, $\varepsilon(s) = \prod_{j=1}^d s_j$,

$s.a$ est le produit scalaire des vecteurs s et a , et $(y)_+ = \max(y, 0)$ la partie positive de y .

Il en résulte que pour un axe donné, Z admet une densité continue par morceaux dont les nœuds correspondent aux projections sur l'axe des sommets du domaine.

Preuve. Loi d'une combinaison linéaire de variables aléatoires indépendantes de loi uniforme.

Le résultat final démontré par Ostrowski (1952) utilise des méthodes souvent très techniques. Cependant Elias *et al.* (1987) proposent une méthode de calcul opérationnel plus simple que nous appliquons à notre cas particulier.

Soit $Z = X.a = X_1 a_1 + \dots + X_d a_d$. C'est donc une somme de variables aléatoires indépendantes de loi uniforme sur $[-a_i, a_i]$, $i = 1, \dots, d$. Par conséquent, elle admet une densité

qui s'obtient simplement par $f_a = \left(\prod_{j=1}^d \frac{1}{2a_j} \right) \times \left(1_{[-a_1, a_1]} * \dots * 1_{[-a_d, a_d]} \right)$, où $*$ désigne le produit

de convolution. Pour calculer ce produit des convolutions, Elias et Shiu (1987) proposent d'utiliser les translations. On pose $T_a(f)(x) = f(x+a)$ la translation de la fonction f de $-a$.

On peut alors écrire (avec la convention $0^\alpha = 0, \forall \alpha \geq 0$):

$$\begin{aligned} 1_{[-a_j, a_j]}(x) &= (x+a_j)_+^0 - (x-a_j)_+^0 \\ &= T_{a_j}(x_+^0)(x) - T_{-a_j}(x_+^0)(x) \\ &= (T_{a_j} - T_{-a_j})(x_+^0)(x) \end{aligned}$$

Et donc, $f_a(x) = \left(\prod_{j=1}^d \frac{1}{2a_j} \right) \times \left((T_{a_1} - T_{-a_1})(x_+^0) * \dots * (T_{-a_d} - T_{a_d})(x_+^0) \right)$.

La preuve est alors la conséquence des deux lemmes suivants :

Lemme 1. La translation commute avec la convolution.

$$(T_a(f_{x_1})) * f_{x_2} = f_{x_1} * (T_a(f_{x_2})) = T_a(f_{x_1} * f_{x_2}) \forall x_1, x_2 \in \mathbb{R}$$

Lemme 2.

$$\frac{x_+^m}{m!} * \frac{x_+^n}{n!} = \frac{x_+^{m+n+1}}{(m+n+1)!} \quad \forall n, m \in \mathbb{N}, \quad \forall x \in \mathbb{R}$$

En appliquant le lemme 1, on a

$$f_a(x) = \left(\prod_{j=1}^d \frac{1}{2a_j} \right) \times \left((T_{a_1} - T_{-a_1}) \dots (T_{-a_d} - T_{a_d}) \underbrace{\dots}_{d \text{ fois}} \right)$$

En appliquant le lemme 2, on obtient

$$f_a(x) = \left(\prod_{j=1}^d \frac{1}{2a_j} \right) \times \left((T_{a_1} - T_{-a_1})(T_{a_2} - T_{-a_2}) \dots (T_{a_d} - T_{-a_d}) \frac{x_+^{d-1}}{(d-1)!} \right)$$

D'où,

$$f_a(x) = \left(\prod_{j=1}^d \frac{1}{2a_j} \right) \times \left(\sum_{s \in \{-1,1\}^d} s_1 \dots s_d \frac{x_+^{d-1}}{(d-1)!} \right)$$

Soit enfin

$$f_a(x) = \left(\prod_{j=1}^d \frac{1}{2a_j} \right) \times \left(\sum_{s \in \{-1,1\}^d} \varepsilon(s) \frac{(x+s.a)_+^{d-1}}{(d-1)!} \right).$$

Le résultat annoncé s'obtient en intégrant cette relation. \square

A noter également que la loi des projections est connue dans d'autres situations, comme dans le cas d'un domaine sphérique : si Ω est la boule unité de \mathbb{R}^d , un calcul direct montre que μ_a admet la densité $f_a(x) = \frac{2}{\pi} \sqrt{1-x^2} 1_{[-1,1]}(x)$, la fonction de répartition étant égale à $F_a(x) = \frac{1}{2} + \frac{1}{\pi} (\text{Arc sin } x + x\sqrt{1-x^2})$ pour $x \in [-1, 1]$.

En résumé, pour qu'un plan d'expériences uniforme possède de bonnes propriétés de répartition sur les projections de dimension 1, il faudrait que pour toutes les directions a , la distribution empirique des projections sur L_a soit proche de leur distribution théorique sous l'hypothèse (H_0). Les statistiques d'adéquation de loi sont nombreuses (voir D'agostino et Stephens, 1986) ce qui laisse un grand nombre de choix pour définir un critère adapté à cet objectif. Des contraintes spécifiques permettent de limiter les possibilités. Déjà, il est préférable que la loi de la statistique soit connue afin d'en éviter son calcul approché. En outre, on aimerait que la statistique soit libre, c'est-à-dire que sa loi ne dépende pas de la direction de projection afin de disposer d'un seuil unique de rejet pour tous les angles. Enfin, par souci de cohérence, il serait souhaitable que la statistique retenue puisse s'interpréter en terme de discrédance lorsqu'on projette sur un axe de coordonnées.

Finalement, deux statistiques (au moins) correspondent à ces contraintes : la statistique de Kolmogorov-Smirnov

$$D_n(a) = \sup | \hat{F}_{n,a}(z) - F_a(z) | \quad (1)$$

et la statistique de Cramér-Von Mises

$$N\omega_n^2(a) = \int (\hat{F}_{n,a}(z) - F_a(z))^2 dz \quad (2)$$

où $\hat{F}_{n,a}$ est la fonction de répartition empirique des projections de x^1, \dots, x^n sur L_a , et F_a la fonction de répartition de μ_a . Lorsque L_a est un axe de coordonnées, μ_a est la loi uniforme sur $[0,1]$ et ces statistiques correspondent respectivement aux discrédances L_∞ et L_2

(Niederreiter, 1987). Dans la suite, nous avons choisi de travailler avec la première car les conclusions semblent équivalentes mais avec des graphiques correspondants plus lisibles (voir la section 5.4). Par analogie avec le cas des axes de coordonnées, nous parlerons de *discrédance des projetés* pour désigner la statistique de Kolmogorov-Smirnov de la formule (1).

Continuité et dérivabilité du radar. Nous donnons les démonstrations pour la dimension 2 afin de simplifier les écritures. Le résultat reste valable en dimension 3. Notons

$D^U(z_1, \dots, z_n) = \sup_{z \in [-1, 1]} |\hat{F}_{z_1, \dots, z_n}(z) - z|$, la statistique de Kolmogorov-Smirnov pour la loi uniforme sur $[-1, 1]$.

Proposition.

- (i) D^U est continue
- (ii) D^U admet des dérivées partielles par rapport à tous les x^i , $i = 1, \dots, n$ en (z_1, \dots, z_n) si et seulement si le supremum de $|\hat{F}_{z_1, \dots, z_n}(z) - z|$ est atteint en un seul z_j , $j \in \{1, \dots, n\}$.

Démonstration. On peut écrire

$$D_n(\theta) = D^U(F_\theta(x_1 \cos \theta + y_1 \sin \theta), \dots, F_\theta(x_n \cos \theta + y_n \sin \theta))$$

où $D^U(z_1, \dots, z_n) = \sup_{z \in [0, 1]} |\hat{F}_{z_1, \dots, z_n}(z) - z|$ est la statistique de Kolmogorov-Smirnov pour la

loi uniforme sur $[0, 1]$. On vérifie avec l'expression de F_θ que $(\theta, p) \mapsto F_\theta(x_i \cos \theta + y_i \sin \theta)$ est continue. La question de la continuité du radar d'uniformité se ramène donc à la continuité du

radar par rapport à toutes les variables $(t_1, \dots, t_n) \mapsto \left| \frac{1}{n} \sum_{i=1}^n 1_{\{t > t_i\}} - t \right|$. Si on perturbe les t_i

d'une valeur η_i , alors le supremum varie au plus de la somme des η_i , ce qui permet de garantir la continuité.

Pour la dérivabilité, on vérifie avec l'expression de F_θ que $(\theta, p) \mapsto \dots$ admet des dérivées partielles. La question de dérivabilité du radar d'uniformité se ramène donc à l'existence de dérivées partielles de D^U .

- Supposons que le supremum soit atteint pour un unique z_j où $j \in \{1, \dots, n\}$.

Dans ce cas, les dérivées partielles relatives à z_i , $i \neq j$, existent et valent 0 car une perturbation locale en z_i ne change pas le supremum. Etudions maintenant l'existence de

$\frac{\partial D^U}{\partial z_j}$. Supposons par exemple que : $\sup_{z \in [-1, 1]} |\hat{F}_n(z) - z| = \hat{F}_n(z_j) - z_j > 0$. On choisit ε tel

que $\forall i \neq j, |\hat{F}_n(z_j) - z_j| > |\hat{F}_n(z_i) - z_i| + \varepsilon$ et $\hat{F}_n(z_j) - z_j > \varepsilon$. On a alors :

$$D^U(z_1, \dots, z_j + \varepsilon, \dots, z_n) = D^U(z_1, \dots, z_j) - \varepsilon \quad \text{et}$$

$$D^U(z_1, \dots, z_j - \varepsilon, \dots, z_n) = D^U(z_1, \dots, z_j) + \varepsilon$$

Donc $\frac{\partial D^U}{\partial z_j}(z_1, \dots, z_n)$ existe en (z_1, \dots, z_n) et vaut -1.

De la même façon, si $\sup_{z \in [0, 1]} |\hat{F}_n(z) - z| = z_j - \hat{F}_n(z_j)$, on montre que $\frac{\partial D^U}{\partial z_j}$ existe et vaut 1.

- Supposons que le supremum soit atteint en z_j et en au moins un autre z_k pour $j, k \in \{1, \dots, n\}$. Nous allons montrer que la dérivée partielle en z_j n'existe pas.

Supposons par exemple que $\hat{F}_n(z_j) - z_j > 0$ et soit $\varepsilon > 0$, $\hat{F}_n(z_j) - z_j > \varepsilon$. Distinguons le cas suivant.

- Le supremum est atteint en z_j et z_k , $j \neq k$ avec $z_j = z_k$.

$$\text{Alors } D^U(z_1, \dots, z_j + \varepsilon, \dots, z_n) = D^U(z_1, \dots, z_n) - \varepsilon,$$

mais $D^U(z_1, \dots, z_j - \varepsilon, \dots, z_n) = D^U(z_1, \dots, z_n)$. Donc la dérivée à droite est différente de la dérivée à gauche.

- Sinon, pour tous les z_k , avec $k \neq j$ pour lesquels l'optimum est atteint, $z_j \neq z_k$. Alors

$$D^U(z_1, \dots, z_j - \varepsilon, \dots, z_n) = D^U(z_1, \dots, z_n) + \varepsilon$$

mais $D^U(z_1, \dots, z_j + \varepsilon, \dots, z_n) = D^U(z_1, \dots, z_n)$ car la perturbation locale en z_j ne fait pas changer la valeur du supremum qui est toujours atteint en z_k . La dérivée à droite est donc aussi différente de la dérivée à gauche. \square

La discrédance des projetés fournit un outil de visualisation des défauts d'uniformité selon les projections. Cet outil est basé sur le principe d'un radar, et pour cette raison nous proposons de l'appeler *radar d'uniformité*. Son utilisation diffère selon la dimension de l'espace Ω sur lequel on vise à vérifier l'uniformité des points.

En dimension 2, on calcule la discrédance des projetés selon toutes les directions en effectuant un balayage angulaire dans Ω . Ce qui donne une courbe paramétrée, appelée *radar 2D*, d'équation en coordonnées polaires

$$\theta \mapsto \dots$$

définie sur $[0, 2\pi]$, permettant de visualiser la bonne répartition des points selon toutes les directions. Et ainsi de décider si le plan convient ou non.

En dimension 3, on calcule la discrédance des projetés sur un axe $L_{\theta, \varphi}$ pivotant autour du centre du domaine et défini en coordonnées sphériques par un angle θ pour la longitude et φ pour la latitude. On obtient dans ce cas une surface paramétrée, appelée *radar 3D*,

$$(\theta, \varphi) \mapsto \dots (\varphi)$$

définie sur $[0, 2\pi] \times \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$.

En dimension quelconque, il n'est plus envisageable d'effectuer un balayage angulaire de l'espace Ω et surtout, représenter graphiquement le résultat devient difficile (même si le calcul reste possible). Cependant, l'hypothèse (H_θ) reste valable sur les espaces de coordonnées de dimension 2 et 3. On peut donc appliquer le radar d'uniformité sur tous les couples et/ou triplets de dimensions possibles.

En pratique, la qualité de la représentation peut être dégradée par la discrétisation. Les radars 2D et 3D sont bien des applications continues. En revanche D_n n'est pas dérivable pour tous les axes L_a tels qu'au moins deux points de Ω se projettent en un même point, ce qui permet de comprendre l'origine des nombreuses cassures visibles dans les graphiques de la section suivante.

5.2. Applications du radar d'uniformité

Dans cette section, nous présentons quelques exemples afin de vérifier l'intérêt du radar d'uniformité pour tester la bonne répartition des points expérimentaux en projection. Conformément au cadre d'application, nous nous plaçons dans des cas où l'hypothèse (H_0) de répartition uniforme dans le domaine expérimental est plausible. Pour chaque représentation du radar d'uniformité nous avons ajouté le cercle (ou la sphère) de rayon ks égal à la statistique du test de Kolmogorov-Smirnov associé à un niveau de confiance à 95%. Rappelons que la statistique étant libre, ks ne dépend pas de a . Cela donne un élément de décision, ou *a minima*, de comparaison, avec les plans aléatoires obtenus par tirage uniforme. Si le plan étudié est stochastique (pseudo aléatoire, hypercubes latins et tableaux orthogonaux randomisés) la visualisation du graphe permet de donner les directions a selon lesquelles l'hypothèse (H_0) est rejetée. Dans le cas contraire, le plan étudié est déterministe (suites à faible discrédance), nous ne sommes plus dans le cadre classique d'application du test. Si une des valeurs de $D_n(a)$ est supérieure à ks , alors on peut seulement dire que le plan étudié est moins bon qu'un plan aléatoire au sens où la probabilité qu'un plan aléatoire ait une meilleure discrédance dépasse 95%. Les exemples qui suivent concernent essentiellement ces cas là, car nous avons préféré prendre des SFD connus sans les transformer. Pour autant, en pratique, il suffirait d'appliquer une randomisation ou un scrambling (voir Fang, Li, Sudjianto, 2006) pour se retrouver dans les conditions d'applications usuelles des tests statistiques.

5.2.1. Exemple 1 : détection de lacunes dans les suites de Halton en 15D à l'aide du radar 2D

On considère les 250 premiers points d'une suite à faible discrédance de Halton (1960) (voir section 1.3.1) en dimension 15. En grande dimension, il est intéressant d'appliquer le radar sur toutes les paires de dimensions possibles. Parmi les cas de rejet, on a par exemple la paire (X_{14}, X_{15}) , qu'on peut visualiser sur la Figure 77.

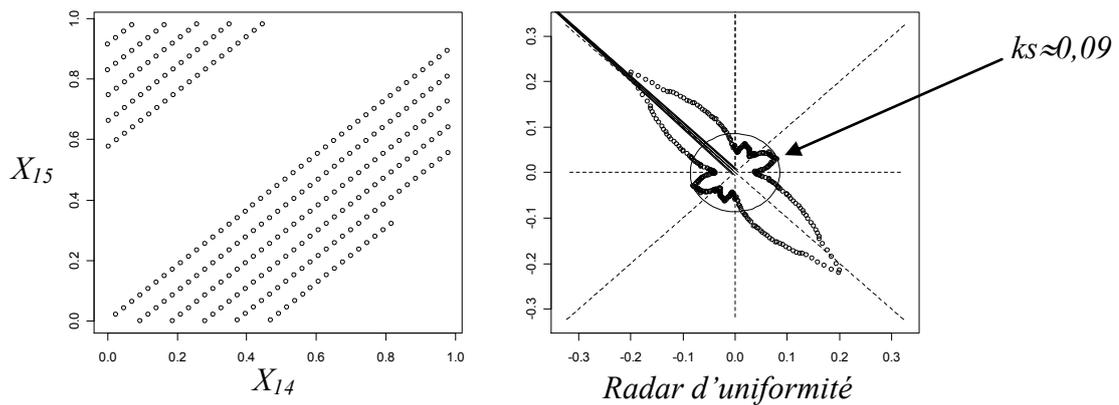


Figure 77. A gauche : le plan (X_{14}, X_{15}) d'une suite de Halton ; à droite : la courbe obtenue à l'aide du radar

Dans cet exemple, comme il y a des valeurs de $D_n(a)$ à l'extérieur du cercle de rayon ks , le radar d'uniformité détecte la mauvaise répartition uniforme dans le plan (X_{14}, X_{15}) . Le plus grand écart à l'uniformité s'observe pour la direction a associée à un angle d'environ 135° , correspondant à la direction orthogonale aux alignements visibles sur la figure de gauche. On retrouve ici le défaut bien connu des suites de Halton en grande dimension de ne pas conserver une faible discrèpance en projection (Morokoff, Caflisch, 1994). A noter cependant que le radar n'est pas fait pour détecter systématiquement les directions d'alignement comme le montre l'exemple suivant.

5.2.2. Exemple 2 : détection de lacunes dans les tableaux orthogonaux à l'aide du radar 3D

On considère un tableau orthogonal linéaire de force 2 centré à 49 points à 3 dimensions (voir section 1.2.2.2.). Comme les hypercubes latins, les tableaux orthogonaux sont souvent recommandés pour les plans d'expériences numériques en raison de leurs bonnes propriétés en projection. En effet, projeté sur une face, un tableau orthogonal de force 2 donne une grille régulière de points. Cependant, la non redondance des projections en dimension 2 n'implique pas une bonne répartition des points, ni sur les axes du domaine (ici 7 paquets de 7 points), ni dans l'espace comme nous pouvons l'observer.

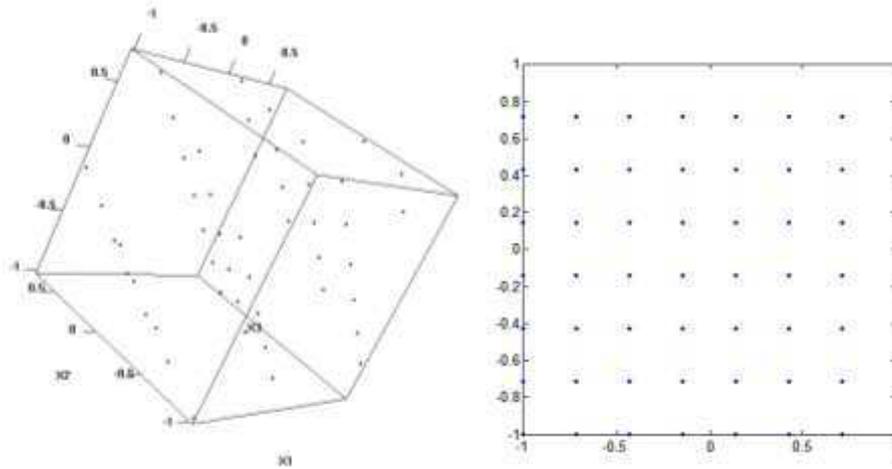


Figure 78. A gauche. Un tableau orthogonal linéaire à 49 points en dimension 3 de force 2. A droite. Les points du tableau orthogonal projetés sur (X_1, X_2)

Le tableau orthogonal étudié ici est linéaire de force 2 et son procédé de construction implique que les points vérifient la relation $x_1 + 3x_2 + x_3 = 0 \pmod{7}$. Les points se trouvent donc sur 5 plans parallèles. En conséquence, la distribution des projections sur l'axe parallèle à ces plans ne sera pas satisfaisante. Ce problème n'est jamais signalé dans le contexte des plans d'expériences numériques, il est pourtant connu dans la littérature des plans d'expériences classiques.

Nous avons appliqué le radar d'uniformité à ce plan et représenté la sphère de rayon ks , les points $D_n(\theta, \varphi)$ et les distances de ces points au centre de la sphère à l'aide d'épingles pour mieux estimer les différences d'uniformité selon les directions. Nous avons également représenté le logarithme des p-valeurs du test en fonction de θ et φ . On constate que le radar détecte bien un problème dans la direction perpendiculaire aux 5 plans, $\theta = 72^\circ, \varphi = 18^\circ$. Il révèle de plus une mauvaise distribution des projections dans la direction $\theta = 35^\circ, \varphi = 40^\circ$, problème que les caractéristiques du plan ne permettaient pas d'anticiper. Notons cependant que le radar ne détecte pas de problème sur les axes de coordonnées pour lesquels les projections sont empilées en 7 paquets de 7 points.

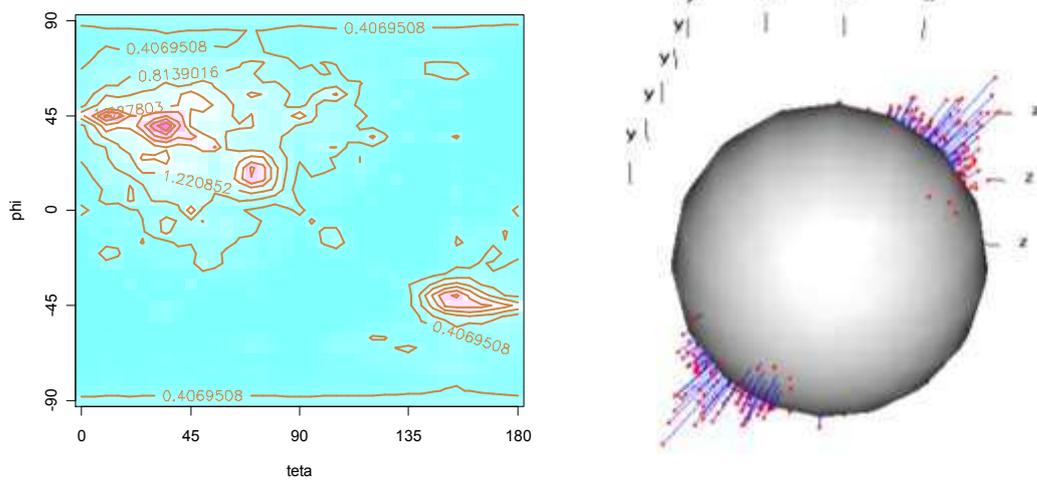


Figure 79. A gauche, les p-valeurs en $-\log_{10}$ du test de Kolmogorov-Smirnov. A droite, les valeurs du test de Kolmogorov Smirnov avec la représentation en épingles du radar

Ceci s'explique quand on regarde la fonction de répartition empirique. En effet, la distance entre la fonction de répartition des points projetés sur un des axes de coordonnées avec celle de la loi uniforme n'est pas très élevée comme le montre la Figure 80 (c). En fait, les alignements peuvent être détectés au moyen de la statistique de Kolmogorov-Smirnov surtout lorsqu'ils ne sont pas répartis régulièrement dans l'espace (comme dans l'exemple 1). Pour illustrer ce point on a représenté les fonctions de répartitions empiriques des projetés (après transformation) sur les 3 axes L1, L2, L3 correspondant respectivement à l'axe défini par $\theta = 35^\circ, \varphi = 40^\circ$, l'axe perpendiculaire aux 5 plans parallèles, et l'axe (0z). L'écart à l'uniformité est clair sur les figures (a), et (b) mais pas sur (c).

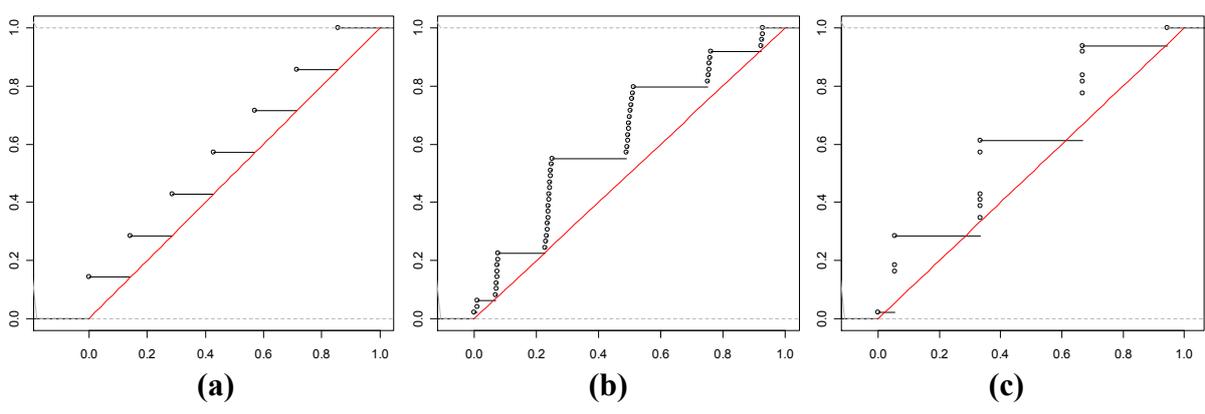


Figure 80. De gauche à droite. Les fonctions de répartition (après transformation) des projetés sur L1, L2 et L3

5.3. Une statistique globale pour le radar 2D

5.3.1. Exemple 3 - Vers une extension du radar d'uniformité

Considérons les 100 premiers points d'une suite de Halton en dimension 8 projetés sur le sous-espace formé par (X_3, X_6) .

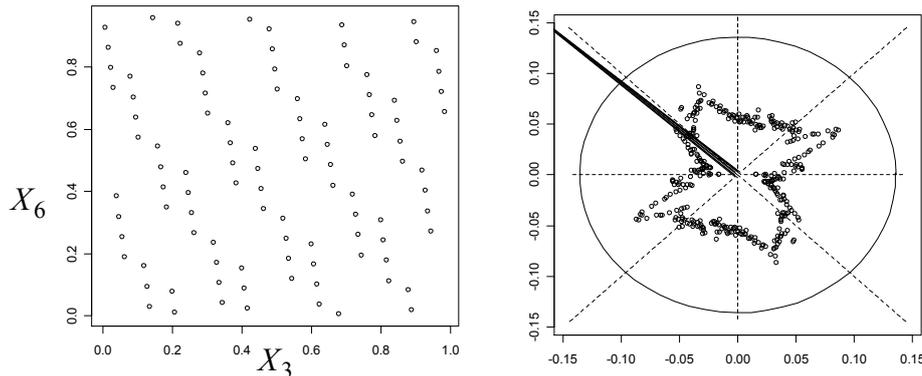


Figure 81. Les 100 premiers termes d'une suite de Halton en dimension 8 projetés sur (X_3, X_6)

Tous les points du radar d'uniformité sont à l'intérieur du cercle de rayon ks et, par ce fait, le radar accepte le plan même si nous pouvons observer que les points du plan (X_3, X_6) ne sont pas répartis uniformément. Cependant, les valeurs de la discrédance sont assez dispersées avec une faible valeur pour l'angle $\theta = 0^\circ$, et des valeurs assez fortes, par exemple pour l'angle $\theta = 29^\circ$ qui semble correspondre à la direction orthogonale d'alignements. L'idée pour rejeter ce type de plan est de définir une nouvelle statistique qui fasse intervenir les discrédances maximale et minimale. De façon à éviter les problèmes d'échelle, nous suggérons de prendre le rapport de ces quantités. On pose donc :

$$G_n = \frac{\sup_{\theta \in [0, 2\pi]} D_n(\theta)}{\inf_{\theta \in [0, 2\pi]} D_n(\theta)} \quad (3)$$

Cette statistique devrait permettre de rejeter les plans qui ont une mauvaise répartition dans une direction comparativement à une autre où les points seront répartis uniformément. Elle a l'avantage d'être globale, c'est-à-dire que seule sa valeur permet d'accepter ou de rejeter un plan. Jusqu'à présent on disposait d'un test statistique pour chaque valeur de θ , ce qui peut être critiquable au niveau décisionnel. Pour une valeur de n fixé, la loi de G_n semble difficile à obtenir autrement que par simulation. Nous avons indiqué en annexe C la table

correspondant à $n = 1, \dots$. D'après cette table, le seuil de rejet à un niveau de 95% pour un plan à 100 points est égal à 4.23. Dans le cas de l'exemple 3, la valeur observée de la statistique G_n est égale à 6.07, ce qui rejette très largement ce plan.

Sur la table statistique, on constate que les valeurs se stabilisent lorsque n augmente, ce qui laisse espérer qu'il est possible de préciser son comportement asymptotique. C'est en effet le cas. Nous décrivons ci-dessous un procédé qui permet de l'obtenir. Celui-ci est semi-analytique car des simulations sont nécessaires pour achever les calculs. En théorie la démarche peut être étendue au cas 3D. La table statistique peut se calculer, similairement au cas 2D, par simulation. Cependant, il est beaucoup plus difficile d'obtenir la loi asymptotique.

5.3.2. Loi asymptotique de la statistique globale pour $n > 100$

Soit $X = (X_1, \dots, X_d)$ avec X_1, \dots, X_d les variables aléatoires indépendantes de loi uniforme sur le carré $[-1, 1]^2$. On note $R_\theta(X)$ la projection de X sur l'axe L_θ . Soit

$(Y_{A_t}^{n,\theta})_{A_t}$ le processus suivant :

$$Y_{A_t}^{n,\theta} = \frac{1}{n} \sum_{i=1}^n 1_{]-\infty, t]}(R_\theta(X_i)) - P(R_\theta(X) < t) = \frac{1}{n} \sum_{i=1}^n 1_{A_t}(X_i) - P(X \in A_t)$$

avec $A_t = \{x \in R^d, R_\theta(x) < t\}$.

Avec ces notations, la discrédance des projetés s'écrit $D_n(\theta) = \sup_{t \in \mathbb{R}} |Y_{A_t}^{n,\theta}|$. D'où,

$$G_n = \frac{\sup_{\theta} D_n(\theta)}{\inf_{\theta} D_n(\theta)} = \frac{\sup_{\theta} \sup_t |Y_{A_t}^{n,\theta}|}{\inf_{\theta} \sup_t |Y_{A_t}^{n,\theta}|}.$$

On cherche alors vers quelle loi tend G_n quand $n \rightarrow \infty$. La théorie des processus empiriques et de leur convergence permet d'identifier la loi limite du processus $(Y_{A_t}^{n,\theta})_{A_t}$. La convergence des lois fini-dimensionnelles vers le processus gaussien de même espérance et covariances est assurée par le théorème de la limite centrée. Par ailleurs, l'ensemble des parties $(A_t)_t$ est une VC-classe. On en déduit que

$$(Y_{A_t}^{n,\theta})_{A_t} \xrightarrow{\text{loi}} (Y_{A_t}^\theta)_{A_t}$$

avec $\left(Y_{A_t}^\theta\right)_{A_t}$ un processus gaussien centré et de structure de covariance :

$$\begin{aligned}
 \text{cov}\left(Y_{A_s}^\mu, Y_{A_t}^\theta\right) &= \frac{1}{n^2} \sum_{i,j}^n P\left(R_\mu(X) < s, R_\theta(X) < t\right) \\
 &\quad - \frac{1}{n} \sum_i^n P\left(R_\mu(X) < s\right) \cdot P\left(R_\theta(X) < t\right) \\
 &\quad - \frac{1}{n} \sum_j^n P\left(R_\theta(X) < t\right) \cdot P\left(R_\mu(X) < s\right) \\
 &\quad + P\left(R_\mu(X) < s\right) \cdot P\left(R_\theta(X) < t\right) \\
 \text{cov}\left(Y_{A_s}^\mu, Y_{A_t}^\theta\right) &= P\left(R_\mu(X) < s, R_\theta(X) < t\right) - P\left(R_\mu(X) < s\right) \cdot P\left(R_\theta(X) < t\right) \\
 \text{cov}\left(Y_{A_s}^\mu, Y_{A_t}^\theta\right) &= P\left(A_s \cap A_t\right) - P\left(A_s\right) \cdot P\left(A_t\right)
 \end{aligned} \tag{4}$$

Par suite,

$$P\left(G_n < y\right) \xrightarrow{n \rightarrow +\infty} P\left(\frac{\sup \sup_t \left|Y_{A_t}^\theta\right|}{\inf \sup_t \left|Y_{A_t}^\theta\right|} < y\right)$$

La probabilité $P\left(A_s \cap A_t\right)$ s'interprète comme la surface d'un polygone délimité par le domaine et les droites A_s et A_t (voir Figure 82). Cette surface peut se calculer analytiquement, par exemple en remarquant que le polygone est une partition des triangles avec un sommet commun, à choisir parmi les sommets du polygone. La loi asymptotique de G_n peut alors s'obtenir par simulation d'un champ gaussien centré et de matrice de covariances définie avec (4).

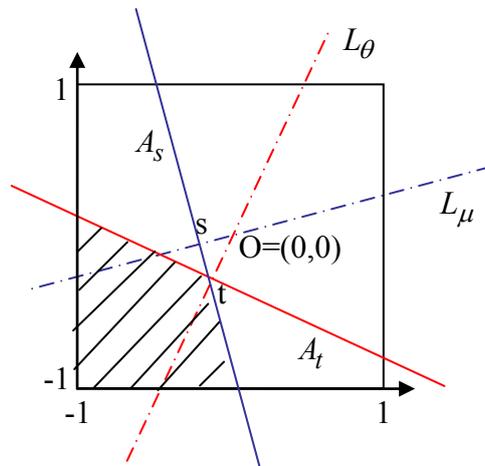


Figure 82. Interprétation de la probabilité $P\left(A_s \cap A_t\right)$

5.4. Discussion

La complexité des phénomènes décrits par des grands codes de calcul incite à répartir les expériences numériques le plus uniformément possible dans le domaine. En outre il est souhaitable que la répartition soit également satisfaisante en projection sur des sous-espaces, au cas où le code ne dépendrait que d'un petit nombre de facteurs ou de composantes principales. Dans ce chapitre, nous avons introduit un critère statistique permettant de généraliser l'utilisation de la discrédance L_∞ aux projections sur tous les sous-espaces de dimension 1. Ce critère est appelé radar d'uniformité car il permet en dimension 2 et 3 de tester graphiquement l'hypothèse d'uniformité du plan, par balayage selon toutes les directions. Nous avons également introduit une statistique globale en dimension 2, qui permet de mieux rejeter des plans décevants mais acceptés par le radar d'uniformité.

L'intérêt de ces critères a été étudié sur des SFD usuels. Sur ces exemples, le radar d'uniformité a réussi à détecter les principaux défauts de ces plans, dont certains se comportent très mal en projection, comme la suite à faible discrédance en dimension 15 de l'exemple 1 ou le tableau orthogonal en dimension 3 de l'exemple 2. Il n'est évidemment pas assuré que le radar détecte des défauts en grande dimension en regardant seulement les projections de dimension 1. Il peut y parvenir lorsqu'il existe une zone vide de forme rectangulaire comme dans les exemples précités. Dans ce cas, la répartition des points est mauvaise lorsqu'on les projette sur la largeur du rectangle. Il peut également détecter des alignements de points, mais échoue si les directions d'alignement sont bien réparties, comme avec le plan factoriel de l'exemple 2. Cela met en évidence un manque de puissance du test de Kolmogorov-Smirnov lorsque l'échantillon est issu en réalité d'une loi continue dont le support est une réunion de petits intervalles répartis régulièrement. En pratique, cette situation n'est pas très préjudiciable car les SFD obtenus par un procédé déterministe sont souvent randomisés ou « scamblés » (voir par exemple Fang, Li et Sudjianto, 2006).

Le radar d'uniformité peut s'adapter à d'autres statistiques d'adéquation de loi, comme la statistique de Cramér-Von Mises (voir section 5.1), qui correspond à la discrédance L_2 lorsqu'on projette sur un axe de coordonnées. A titre d'illustration, nous avons refait les exemples 1 à 3 avec le radar correspondant. Sans surprise, les conclusions sont identiques, car les tests de Kolmogorov-Smirnov et Cramér-Von Mises ne présentent pas de différence nette en termes de puissance. La différence principale est d'ordre graphique : l'allure du radar

défini avec la statistique de Cramér-Von Mises est plus lisse, ce qui vient de la norme L_2 , et présente des variations d'échelle parfois importantes d'un plan à un autre, alors que ces différences sont atténuées par la norme L_∞ dans les exemples montrés ici.

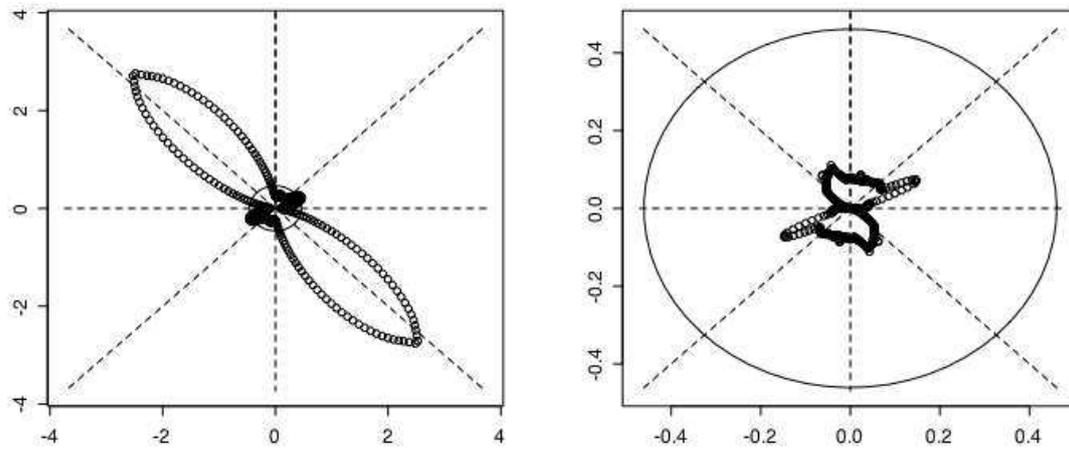


Figure 83. Radar d'uniformité avec la statistique de Cramér-Von Mises pour les exemples 1 et 3

Chapitre 6

Arbres de longueur minimale : Une nouvelle approche pour déterminer la qualité des plans d'expériences numériques

Comme on a pu le voir dans les chapitres précédents, étudier l'uniformité d'une distribution de points est un travail fastidieux, voire même impossible en grande dimension. Il est donc utile de se référer à des critères afin de pouvoir décider si une distribution est uniforme ou non. Le Chapitre 2 présente deux classes de critères mesurant la qualité de la répartition spatiale des points d'un plan d'expériences : la discrédance et les critères basés sur des calculs de distances entre les points. Seulement ces critères permettent, souvent difficilement, de conclure sur l'uniformité des points.

C'est pourquoi, nous proposons ici un nouveau critère qui s'appuie sur la théorie des graphes et des arbres de longueur minimale (ALM) en particulier, pour qualifier la répartition des points d'un plan dans un espace multidimensionnel. Et, par ce biais, de classer les plans selon leur structure : grille, aléatoire, en amas, ... ce que ne permettent pas les autres critères couramment utilisés. Grâce à cette méthode, il sera ainsi possible de vérifier par exemple si le plan étudié est un bon compromis entre un plan aléatoire et un plan factoriel.

A noter que les méthodes de répartition spatiale de points sont bien connues pas les théoriciens des graphes et utilisés ou réinventés dans différents domaines tels que la reconstruction de données géographiques pour un simulateur de vol , en imagerie pour la

retouche d'image, ou bien même pour le diagnostic de cellules cancéreuses, etc. Cependant ces méthodes, et en particulier les ALM, semblent non exploitées dans le contexte de la planification d'expériences numériques.

Ce chapitre vise à introduire les méthodes d'études de répartition spatiale de points, en insistant plus particulièrement sur la méthode des arbres de longueur minimale. Nous présenterons également quelques exemples d'application de ce critère lié aux ALM. Enfin, nous concluons et nous proposerons des perspectives.

6.1. Méthode d'étude de répartition spatiale d'objets

On cherche, au-delà d'une analyse qualitative qui se résume en général à la valeur d'un critère, à quantifier, même sommairement, une distribution de points dans un espace multidimensionnel par un couple (ou un n-uplet) de paramètres qui permet de différencier divers types de plans (au sens de « distribution de points dans un espace »).

Plusieurs méthodes permettent d'étudier une répartition spatiale d'objets. On cite les plus courantes :

- les méthodes statistiques,
- les fonctions de distance,
- les méthodes géométriques.

Nous ne nous appesantirons pas sur les deux premières séries, qui reprennent, entre autres, les critères de distance et de discrédance décrits précédemment, mais aussi des notions de probabilité, de densité de points par unité de volume, des fonctions de distribution de distances entre un point et ses plus proches voisins, voire des estimations de fonctions de corrélation et/ou d'anisotropie. On pense, pour ce qui concerne ce dernier point, aux techniques de variographie utilisées par les géostatisticiens.

On s'attarde par contre d'avantage sur l'algorithmique géométrique, et la théorie des graphes en particulier, dont découle le critère qui fait l'objet de ces travaux.

6.1.1. Géométrie algorithmique

La géométrie algorithmique (Cormen *et al.*, 2002) recouvre un certain nombre de méthodes qui sont toutes basées sur la partition de l'espace à partir de la position des points étudiés. Ceci revient à caractériser pour chaque point sa zone d'influence et les relations avec ses voisins.

Une des techniques les plus connues est sans doute le **diagramme de Voronoï** qui permet, une fois construit, d'évaluer des grandeurs caractéristiques telles que :

- des mesures (surface, volume, périmètre, etc.) de chaque zone d'influence,
- des facteurs de formes,
- les premiers moments des statistiques des grandeurs ci-dessus,
- etc.

Il existe bien entendu d'autres descripteurs de la répartition des points dans l'espace tels que les **graphes de Delaunay**. Il s'agit de construire une triangulation qui relie les points voisins. De la même façon, cette triangulation peut se caractériser par la longueur de chaque arête et le nombre de voisins de chaque point, mais aussi les moments de leurs distributions, etc.

On peut également évoquer :

- les graphes de Gabriel
- les graphes des voisins relatifs
- ...
- et les arbres de longueur minimale, qui nous intéressent ici.

Faisons à présent un bref rappel sur la définition des graphes.

6.1.2. Rappels sur la théorie des graphes

Quelques définitions.

Un **graphe** $G(X,U)$ est un couple constitué :

- d'un ensemble $X=\{x_1, x_2, \dots, x_n\}$
- d'une famille $U=\{u_1, u_2, \dots, u_m\}$ d'éléments $X.X=\{(x,y)/x \in X, y \in X\}$, appelés les arcs.

Un **graphe** est un 1- graphe si tous les éléments (x,y) de $X.X$ n'apparaissent au plus qu'une seule fois dans U .

Un **graphe simple** est un 1-graphe qui ne contient pas de boucles : les arcs sont alors appelés des arêtes.

Un **chemin** est une séquence $\mu=(u_1, u_2, \dots, u_q)$ d'arêtes contiguës et un graphe est dit connexe s'il existe un chemin reliant toute paire de points (x,y) .

Un **arbre** est un graphe simple, connexe, ne contenant pas de circuits.

Remarque. Le métro de Paris est, par exemple, assimilé à un arbre.

Un **arbre** est dit **maximal** s'il contient tous les points.

Un **Arbre de Longueur Minimale (ALM)**, ou *Minimal Spanning Tree (MST)*, est un arbre maximal pondéré dont la longueur totale (somme de toutes ses arêtes) est la plus petite de tous les arbres maximaux. On peut en effet pondérer un arbre en affectant un poids à chaque arête. Néanmoins, pour un ensemble de points donnés, s'il existe plusieurs arbres maximaux, il n'existe qu'un ALM, par définition.

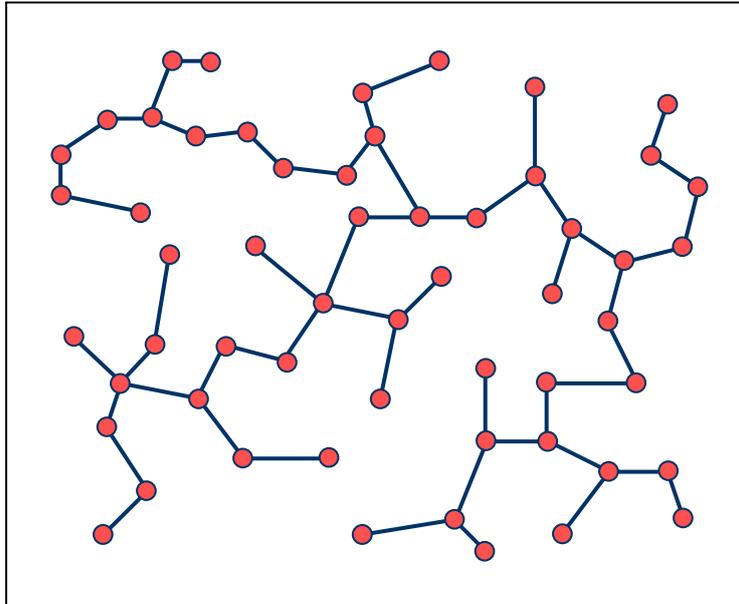


Figure 84. Illustration d'un arbre de longueur minimale

6.1.3. Quelques mots sur les graphes de Gabriel et ceux des voisins relatifs

Rappelons tout d'abord ce qu'est un **polyèdre de Delaunay** (en deux dimensions, on parle de triangle de Delaunay).

Deux points sont voisins s'ils se trouvent sur le même polyèdre de Delaunay. Un polyèdre est dit de Delaunay si l'ensemble de ses sommets se trouvent sur une même hyper sphère vide, c'est-à-dire qui n'englobe aucun autre point. Dans le plan, les sommets d'un triangle de Delaunay se trouvent sur un cercle qui ne contient aucun point de l'ensemble.

Le **Graphe des Voisins Relatifs** est un graphe connexe où deux points α et β sont voisins, *i.e.* sont reliés par une arête, s'ils vérifient l'équation suivante :

$$d(\alpha, \beta) \leq \max(d(\alpha, \gamma), d(\beta, \gamma)) \quad \forall \gamma \neq \alpha, \beta.$$

Le **graphe de Gabriel** est un graphe connexe où deux points α et β ont une arête commune si et seulement si l'hyper sphère de diamètre $[\alpha, \beta]$ est vide, *i.e.* si elle ne contient aucun point.

6.1.4. Propriétés des Arbres de Longueur Minimale

Les arbres de longueur minimale ont des propriétés intéressantes. Ainsi, Hammersley *et al.* (1959) ont montré que la longueur d'un ALM est exprimable analytiquement de façon asymptotique.

Soit une distribution aléatoire uniforme de n points, dans une région de mesure V , dans un espace de dimension d . La longueur du graphe est donné asymptotiquement par :

$$L_{G_\infty} = \alpha_d (Vn^{d-1})^{1/d}$$

En dimension 2, nous obtenons :

$$L_{G_\infty} = \alpha_2 (Vn)^{1/2}.$$

Et la longueur moyenne des arêtes est donnée par :

$$M_A = \alpha_2 \frac{(Vn)^{1/2}}{n-1}.$$

On peut de la même façon calculer :

- l'écart-type de la distribution de la longueur des arêtes : l'interprétation classique des ALM se limite en général aux deux premiers moments de la distribution,
- mais aussi l'histogramme des longueurs des arêtes, les moments d'ordre supérieur et les quantiles,
- la distribution des angles des arêtes par rapport à un axe donné, ce qui est une mesure de l'anisotropie de la distribution des points dans l'espace,
- etc.

6.2. Application

6.2.1. Mise en œuvre et interprétation

Des algorithmes, comme celui de Kruskal (1956) ou Prim (1957), permettent de construire à partir d'un nuage de points l'arbre de longueur minimale. Il utilise pour ce faire la triangulation de Delaunay. Pour une même distribution de points, il peut y avoir différentes solutions possibles. Il est néanmoins démontré que l'histogramme des longueurs d'arêtes est unique et qu'il peut donc être caractérisé, par exemple, par ses deux premiers moments : la moyenne et l'écart-type.

Dans cet espace (μ, σ) , toute population de points occupe une position déterminée qui en caractérise la répartition comme le montre la figure suivante :

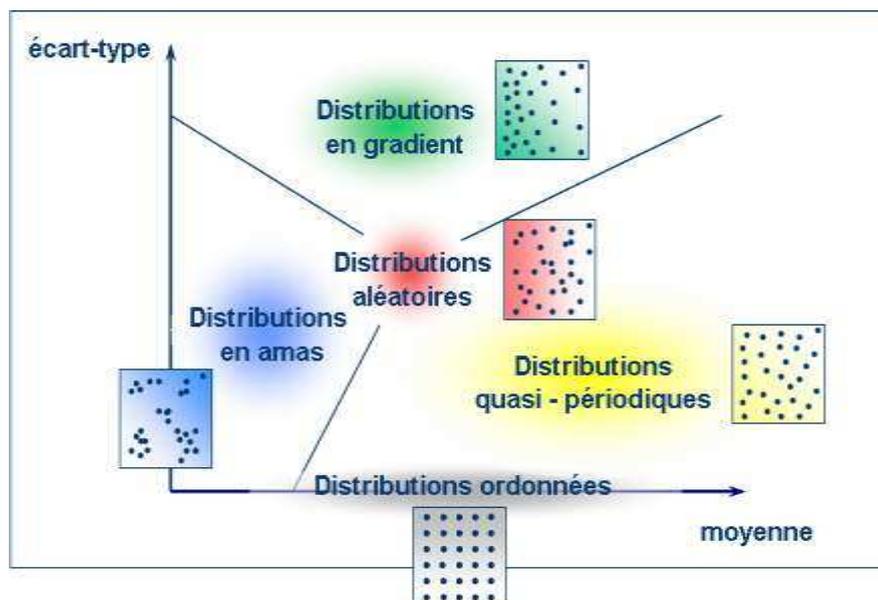


Figure 85. Représentation de la répartition des distributions dans le plan (μ, σ)

Remarque. Cette répartition des distributions ne dépend pas de la dimension.

Wallet et Dussert (1998) ont comparé différentes méthodes d'analyse topographique (distribution des plus proches voisins, distribution radiale, graphe de Voronoï, dimension fractale de la dispersion relative et ALM) sur des simulations de répartitions de points et mis en évidence que celle utilisant l'ALM (moyenne μ et écart-type σ des longueurs de branches) présentait les erreurs standards les plus faibles pour chaque distribution de points étudiée

(distribution aléatoire, distribution hardcore, petits amas et gros amas) et simultanément le meilleur pouvoir de discrimination et la meilleure stabilité.

Les qualités offertes par l'utilisation des paramètres μ et σ de l'ALM et la possibilité d'utiliser facilement des paramètres supplémentaires (histogramme, distribution des angles, etc.) nous ont conduit à retenir ce critère pour étudier les Space-Filling Designs.

6.2.2. Exemples d'application

L'objectif de cette partie est uniquement de présenter un exemple d'application du critère. Une étude plus approfondie est effectuée au paragraphe 7.2.

En dimension 2, les valeurs de la moyenne $\mu = 0.662$ et de l'écart-type $\sigma = 0.311$ correspondantes à une distribution uniforme (Dussert *et al.*, 1986) ont été corroborées par des simulations Monte-Carlo de distributions aléatoires générées avec une méthode de congruence linéaire (Knuth, 1969).

En dimension 3, les valeurs expérimentales retenues pour la caractérisation des distributions aléatoires sont $\mu = 0.708$ et $\sigma = 0.26$. Le critère basé sur les ALM a été mis en œuvre afin de confronter plusieurs types de plans à 30 points en dimension 3. On y retrouve les différents plans de Strauss, des plans à entropie maximale, des hypercubes latins (lhs), les suites à faibles discrédance (sfd) ou bien encore des plans minimax (covd). Afin de conclure sur la qualité de ces plans, nous avons aussi considéré des plans aléatoires et des plans en amas construits à partir du processus de Neyman et Scott (1958).

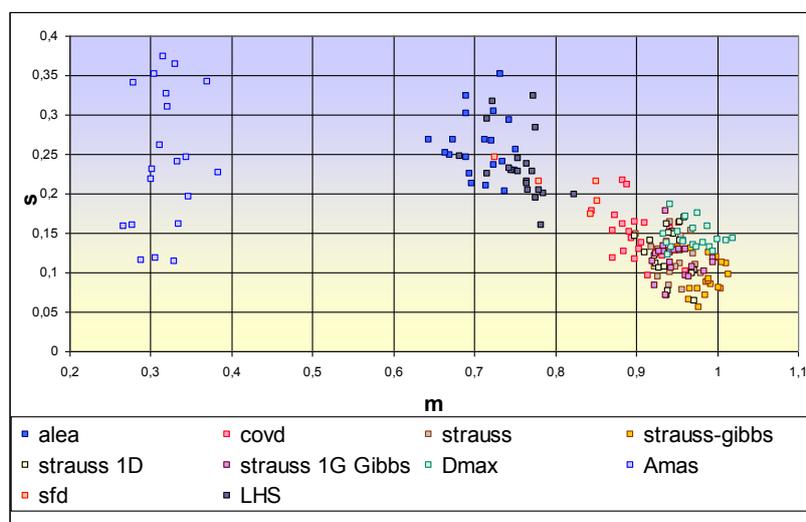


Figure 86. Représentation du critère de l'ALM pour les plans à 30 points en 3D où figurent les plans de divergence minimale

La figure ci-dessus montre par exemple que les hypercubes latins ont des distributions les plus proches de celles des plans aléatoires. Les plans qui paraissent comme les plus intéressants au regard de nos objectifs, c'est-à-dire se situant dans la zone de quasi-périodicité, sont les plans de Strauss sans les contraintes 1D.

6.3. Discussion

Un nouveau critère de qualification intrinsèque des Space-Filling Designs, basé sur les propriétés des Arbres de Longueur Minimale (ALM), a été étudié. Ce critère permet de classer et de qualifier toute distribution de points dans l'espace selon une cartographie, en moyenne et écart-type, des longueurs des arêtes des ALM associés aux plans étudiés.

Le domaine en μ et σ est partitionné en trois zones, permettant de séparer les distributions :

- en amas,
- en gradients,
- aléatoires,
- quasi-périodiques
- ordonnées.

Un critère probablement équivalent à celui étudié ici serait d'utiliser la percolation ; méthode permettant d'étudier la propagation d'une information entre différents sites, dont certains sont liés entre eux, et d'autres non. Cela peut par exemple représenter la propagation d'un feu de forêt, d'une infiltration d'eau, etc.

Chapitre 7

Etude intrinsèque des plans

L'objectif de ce chapitre est de comparer la qualité des distributions de points au sens du remplissage de l'espace à l'aide des critères étudiés jusqu'ici.

Deux études différentes sont accomplies.

Une première qui consiste à appliquer les critères intrinsèques des chapitres 2.1 et 2.2. à différents plans en dimension 3, puis en dimension 10 afin d'être dans le contexte des plans d'expériences numériques. Le choix de ces dimensions est lié aux applications présentées ultérieurement dans la partie IV.

Enfin, la deuxième étude mettra en œuvre le critère des Arbres de Longueur Minimale (ALM) pour étudier la robustesse des plans en grande dimension.

7.1. Etude intrinsèque des plans à l'aide des critères usuels

L'objectif de cette section est de calculer les critères intrinsèques présentés dans le chapitre 2 (distance et discrédance) sur les plans de même caractéristique que ceux utilisés lors des applications 3D et 10D. Ces études permettront ainsi de confirmer ou non le lien entre les propriétés intrinsèques des plans et leur pouvoir prédictif dans les applications.

Pour ce faire, nous étudions deux types de critère : discrédance et distance.

Les plans confrontés dans cette partie sont généralement¹³ les suivants :

- plans aléatoires (*Aleatoire*)
- plans de type minimax¹⁴ (*Minimax*)
- plans à entropie maximale (*Dmax*)
- hypercubes latins de différents types (*rLHs*, *mLHs*, *iLHs*)
- suites à faible discrédance (Faure, Halton, Hammerlsey, Niederreiter, Sobol')
- plans à faible discrédance (*low-disc*)
- plans de Strauss (*Strauss*)
- plans de Strauss-Gibbs (avec un potentiel puissance) (*Strauss_G*)
- plans de Strauss avec des contraintes en 1D (*Strauss1D*)
- plans de Strauss-Gibbs avec des contraintes en 1D (*Strauss1D_G*).

Concernant les plans stochastiques (non-déterministes), c'est-à-dire tous les types de plans excepté les suites à faible discrédance, les critères ont été calculés sur 100 plans afin de donner un sens aux résultats. Ce nombre peut être discutable mais les tests en grande dimension n'ont pas permis d'en tester plus.

¹³ Quelques plans n'ont pas été testés en dimension 10 à cause du temps de calcul

¹⁴ Ces plans sont obtenus à l'aide de la routine `cover.design` du package `fields` de R en considérant une grille en dimension 3 comme distribution initiale et une distribution aléatoire en dimension 10.

7.1.1. Plans à 30 points en dimension 3

Nous avons représenté sur les figures ci-dessous les critères les plus significatifs ; on y trouve : la discrétance et la discrétance centrée en norme L_2 ainsi que les critères MinDist et de recouvrement. D'autres critères ont été envisagés et portés en annexe B.

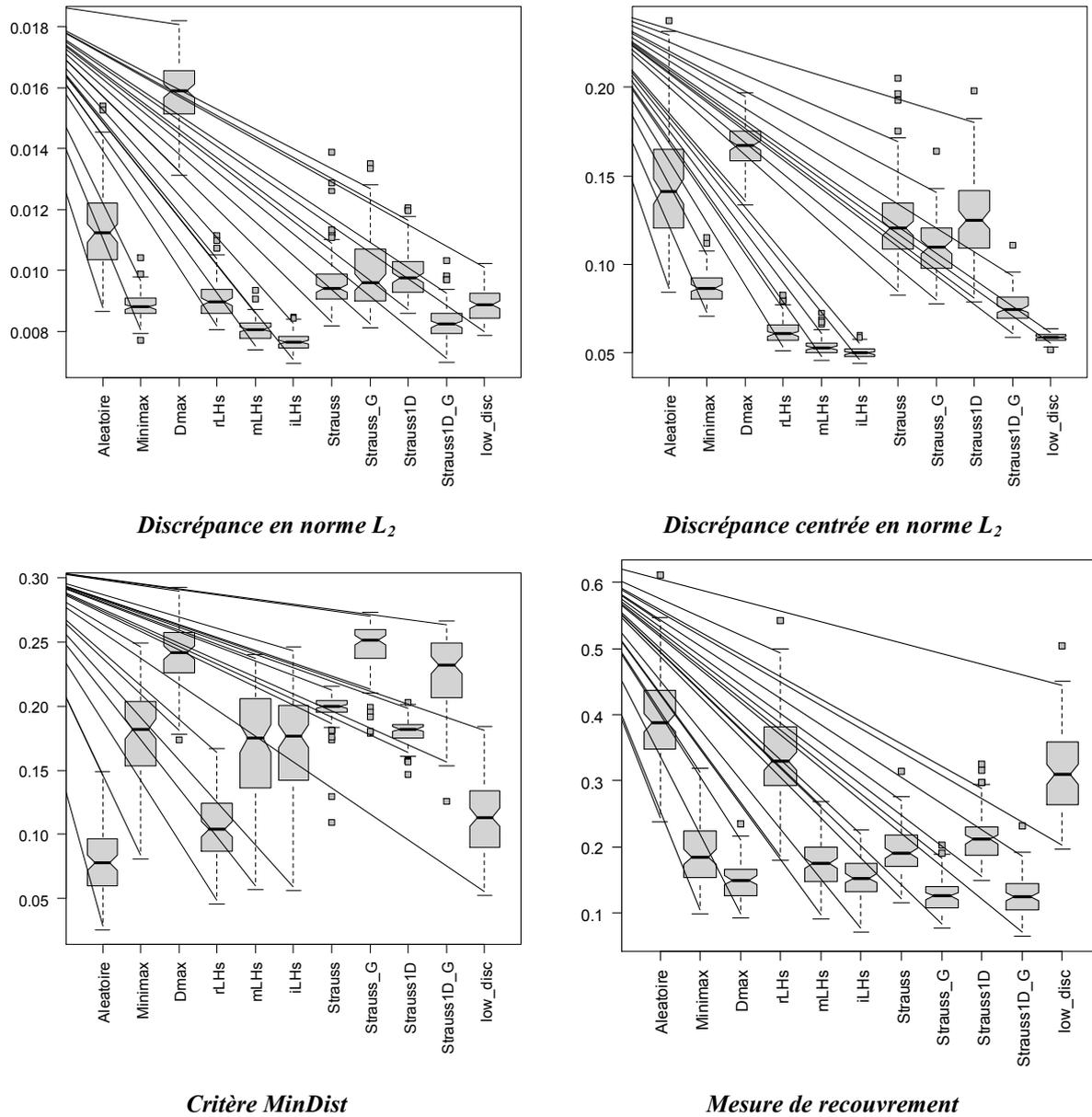


Figure 87. Représentation des boxplots sur les critères de discrétance en norme L_2 (classique et centrée) et de distance (MinDist et recouvrement) calculés sur les plans à 30 points en dimension 3

On rappelle avant l'interprétation des boxplots que :

- plus la discrédance est faible et plus les points sont répartis uniformément,
- plus la mesure de recouvrement est proche de 0 et plus les points se rapprochent d'une grille régulière,
- plus le critère MinDist est grand et plus les points seront éloignés les uns des autres.

Quelques remarques méritent d'être notifiées concernant les figures ci-dessus.

Les hypercubes latins, les Strauss-Gibbs avec considération des contraintes 1D et les plans minimax ont de faibles discrédances à l'image des plans à faible discrédance construits dans ce sens. La discrédance centrée prenant en compte la répartition des points sur les marges de dimension 1, ce résultat est attendu.

Il paraît étonnant que les Strauss avec des contraintes 1D ne donnent pas de bons résultats surtout comparativement à ceux obtenus avec les Strauss-Gibbs avec contraintes 1D. Cela peut s'expliquer par un mauvais choix des rayons d'interactions qui est compensé par le potentiel puissance évitant ainsi les points proches au sens de la distance usuelle.

Les plans à entropie maximale ont généralement la plus mauvaise discrédance, mais ils ressortent comme étant les meilleurs au regard des critères de distance.

Les plans de Strauss, Strauss-Gibbs et Strauss-Gibbs avec des contraintes 1D ont de très bons critères de distance.

Les hypercubes latins améliorés (ilhs et mlhs) sont meilleurs que les hypercubes latins usuels concernant les critères de distance. En 3D, ils permettent donc visiblement de bien remplir l'espace. Les plans minimax semblent aussi de bonne qualité.

Les plans ayant de mauvais résultats dans les 2 catégories de critères sont sans étonnement les plans aléatoires.

	<i>Discrédance</i>	<i>Discrédance centrée</i>	<i>Mindist</i>	<i>Recouvrement</i>
<i>Faure</i>	0.0082	0.034	0.170	0.286
<i>Halton</i>	0.0076	0.031	0.155	<u>0.221</u>
<i>Hammerlsey</i>	<u>0.0075</u>	0.042	<u>0.176</u>	0.235
<i>Niederreiter</i>	0.0080	0.026	0.162	0.228
<i>Sobol'</i>	0.0083	<u>0.025</u>	0.136	0.359

Tableau 6. Critères de discrédance en norme L_2 et critères de distance pour les suites à faible discrédances à 30 points en 3D

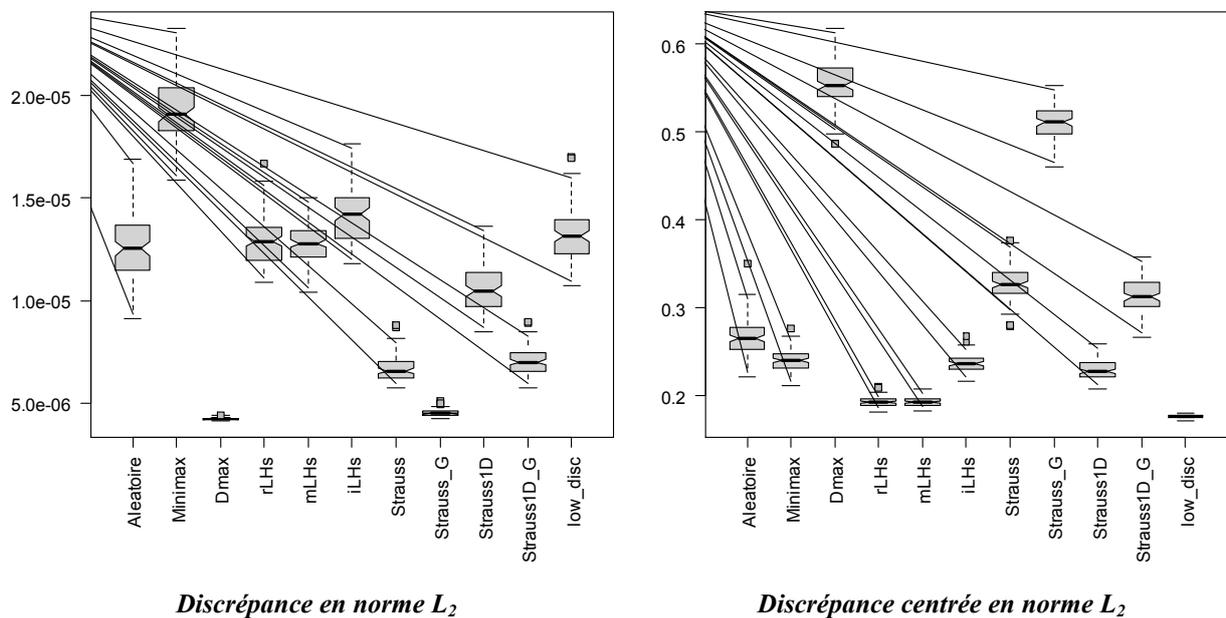
La suite de Hammersley apparaît comme étant celle de meilleure qualité. En effet, le critère MinDist est le plus élevé et la discrédance est la plus faible. Cependant, si l'on considère la discrédance centrée qui prend en compte les points en projection sur les marges, elle ressort comme étant la plus mauvaise.

En comparaison avec les résultats des boxplots, il apparaît que les suites à faible discrédance ont une meilleure discrédance que les plans à faible discrédance. Ces derniers étant construits à partir d'un algorithme d'échange de minimisation de la discrédance centrée, ils restent meilleurs vis-à-vis de ce critère.

Il est intéressant d'observer que les suites sont meilleures vis-à-vis des critères de distance comparés aux plans à faible discrédance et de qualité semblable aux hypercubes latins améliorés.

7.1.2. Plans à 100 points en dimension 10

Nous présentons les résultats concernant les critères intrinsèques de distance et de discrédance. Certains des résultats sont mis en annexe E afin de faciliter l'analyse.



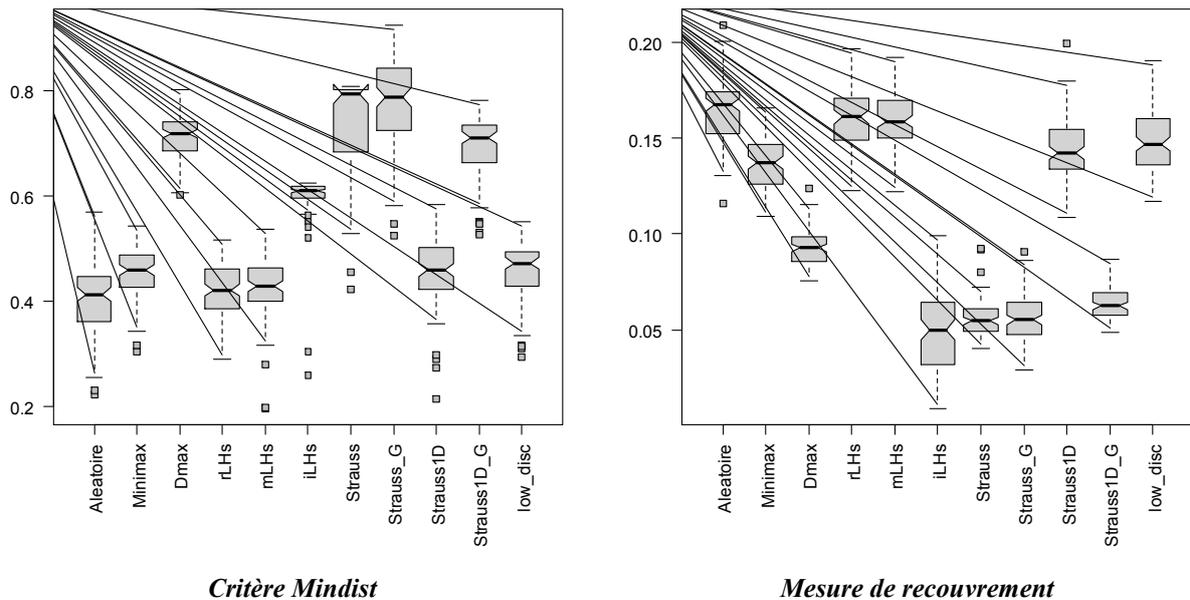


Figure 88. Représentation des boxplots sur les critères de discrédance en norme L_2 (classique et centrée) et de distance (Mindist et recouvrement) calculés sur les plans à 100 points en dimension 10

Les résultats sont quelque peu différents de ceux obtenus en dimension 3.

Par exemple, on remarque que les hypercubes latins usuels et améliorés au regard du critère maximin (mLHs) sont de qualité semblable à celle obtenue avec des plans aléatoires que ce soit au niveau des critères de distance ou de discrédance (excepté pour la discrédance centrée).

Les plans à entropie maximale et les plans de Strauss-Gibbs ont de très faibles dicrédances semblables à celles des suites à faible discrédance.

Les plans à faible discrédance ne semblent pas être de bonne qualité en grande dimension sauf excepté bien évidemment vis-à-vis de la discrédance centrée. Ceci s'explique par le fait que l'algorithme d'échange privilégie beaucoup trop le critère que l'on souhaite optimiser. Une solution qui pourrait être de coupler ce critère à un critère de distance, ce qui amènerait à des plans de qualités médiocres au regard des 2 critères choisis.

On remarque que les plans ayant les plus mauvaises discrédances centrées sont les plans à entropie maximale et les plans de Strauss-Gibbs. Ces plans étant très bons au regard des critères de distance, il apparaît que si les points sont bien répartis dans l'espace alors ils ne sont pas obligatoirement en projection sur les marges. La réciproque est vraie lorsque l'on regarde les résultats obtenus avec les hypercubes latins. A noter que les plans de Strauss-Gibbs avec contraintes 1D ont privilégié la répartition des points dans l'espace au regard de celle en projection sur les marges de dimension un. Tous ces phénomènes nous laissent penser

qu'il est impossible en grande dimension de bien répartir les points dans l'espace et en projection sur les marges 1D.

A noter que les Strauss avec des contraintes 1D ne donnent pas (similairement au cas 3D) de bons résultats surtout comparativement à ceux obtenus avec les Strauss-Gibbs avec contraintes 1D. Il serait intéressant d'effectuer d'autres tests avec plusieurs rayons d'interaction afin de confirmer que ces résultats sont la conséquence d'une paramétrisation non optimale.

Les hypercubes latins améliorés (ilhs) sont meilleurs que les hypercubes latins usuels concernant les critères de distance. En 10D, ils semblent être un bon compromis entre la bonne répartition des points dans l'espace et en projection sur les marges. Les plans de Strauss-Gibbs avec contraintes 1D sont de meilleure qualité que ces derniers notamment au regard des critères de distance.

Les plans à entropie maximale, par leur très bonne discrédance, la plus mauvaise discrédance centrée et une mesure de recouvrement moyenne, semblent différents des plans de Strauss. Les critères permettent de dire que les points de ces plans sont tous éloignés les uns des autres mais qu'il n'y a pas un recouvrement de l'espace optimal. Ce qui conforte la remarque de Jonhson *et al.* sur le fait que ces plans auraient tendance à mettre les points sur les bords du domaine.

Les plans minimax ne semblent pas robustes lorsque l'on augmente la dimension ; c'est ce que traduit la mauvaise qualité des critères de discrédance et de distance associés.

Les suites à faible discrédance ont bien évidemment une faible discrédance. A noter que la suite de Sobol' a une discrédance beaucoup plus faible (rapport de 10) que la moyenne des plans à faible discrédance. Ceci laisse supposer que le nombre d'itérations est trop réduit même si celui-ci ne peut pas être augmenté à cause de la grande dimension et ainsi du temps de calcul engendré.

	<i>Discrédance</i>	<i>Discrédance centrée</i>	<i>Mindist</i>	<i>Recouvrement</i>
<i>Faure</i>	1.91e-05	0.009	0.29	0.35
<i>Halton</i>	9.42e-06	0.011	0.51	0.14
<i>Hammerlsey</i>	1.13e-05	0.011	0.51	0.15
<i>Niederreiter</i>	1.13e-05	0.011	0.36	0.22
<i>Sobol'</i>	1.52e-05	0.002	0.45	0.12

Tableau 7. Critères de discrédance en norme L_2 et critères de distance pour les suites à faible discrédances à 100 points en 10D

7.2. Etude des plans avec le critère des ALM

Cette étude vise à quantifier le critère lié aux ALM sur différents plans couramment utilisés en expérimentation numérique. Elle permet d'autre part de les situer dans le plan (μ, σ) en les comparant aux résultats de plans aléatoires et de plans en amas. Il s'agit aussi de vérifier la robustesse de certains plans dans les grandes dimensions en vérifiant qu'ils se situent dans la zone de quasi-périodicité (voir Figure 85).

Nous avons étudiés les plans suivants :

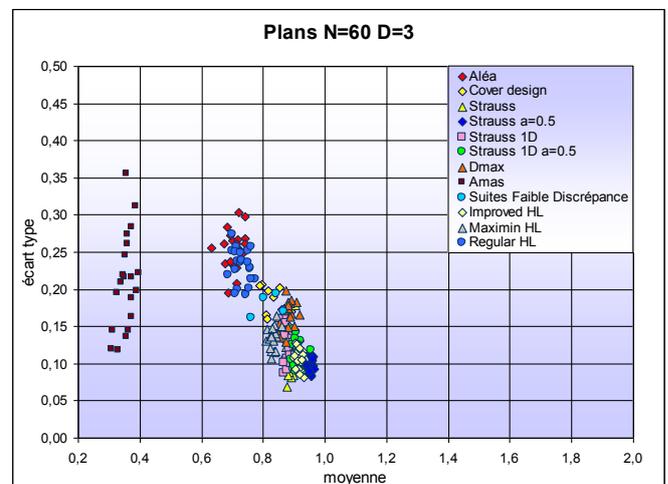
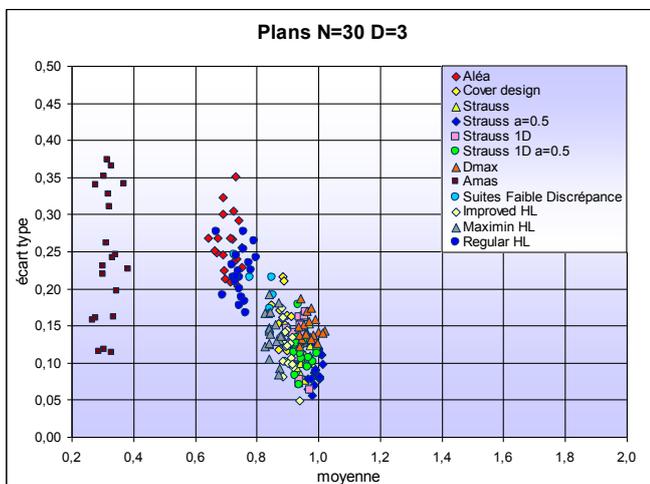
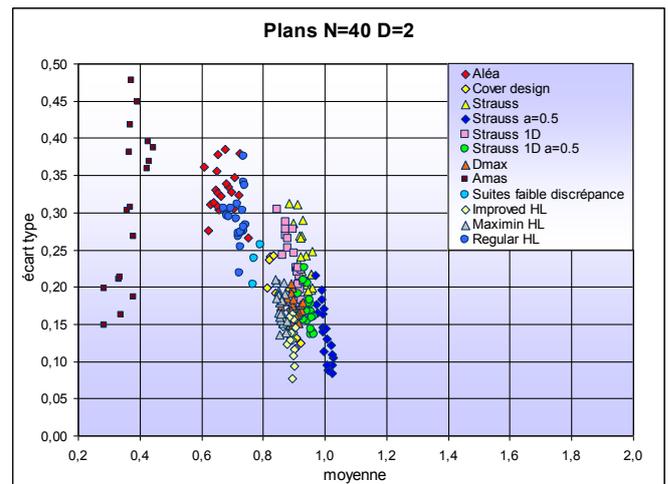
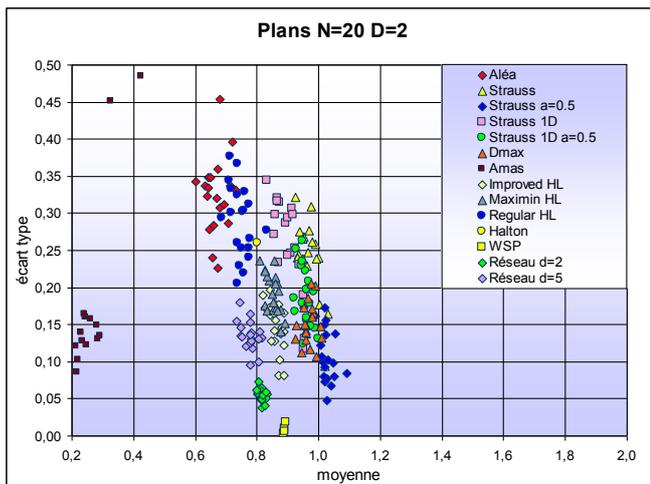
- suites à faible discrèpances : Halton, Hammersley, Faure, Niederreiter, Sobol
- hypercubes latins
- hypercubes latins améliorés au regard du critère de maximin
- hypercubes latins améliorés en optimisant la distance euclidienne entre les points du plan
- plans minimax (*Cover design*)
- plans à entropie maximale
- plans de Strauss
- plans de WSP (Sergent *et al.* 1997)
- tableau orthogonal linéaire (uniquement pour le cas $N = 49$ en dimension 3) (voir Owen, 1992)

Pour les plans introduisant un aléa stochastique (c'est-à-dire tous, à l'exception des suites à faible discrèpance), l'étude a été effectuée avec 20 plans de chaque type afin de confirmer les résultats obtenus. Les résultats étant de bonne qualité, il ne nous a pas semblé judicieux de considérer un plus grand nombre de plans.

Pour chacune des dimensions testées (2, 3, 5, 10 et 20) une première application est effectuée avec des plans comptant 10 points par dimension afin de rester dans le contexte de la planification numérique. Cependant les résultats théoriques énoncés précédemment étant asymptotiques, nous avons également conduit une seconde application sur des plans à 20 points par dimension. Même si le nombre de points choisi pourrait encore être contesté vis-à-vis des résultats asymptotiques.

Les plans en amas ont été générés par des répartitions de Neyman et Scott (1958) c'est-à-dire par l'utilisation des règles suivantes : le centre des amas est uniformément réparti ; chaque amas contient un nombre de point N_i suivant une distribution de moyenne μ ; les points dans un amas sont distribués selon une répartition gaussienne centrée à l'origine des amas et avec un écart-type quelconque.

Les figures ci-dessous présentent les résultats obtenus pour les plans listés précédemment.



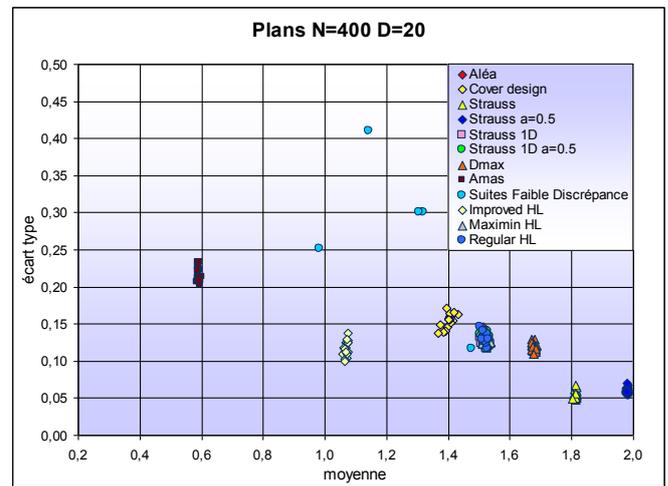
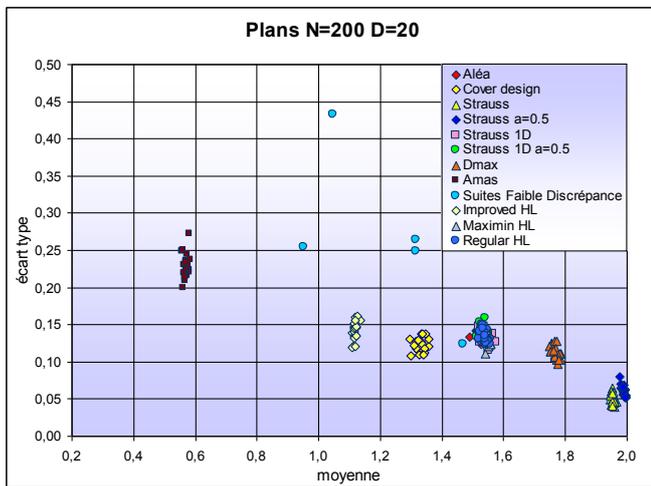
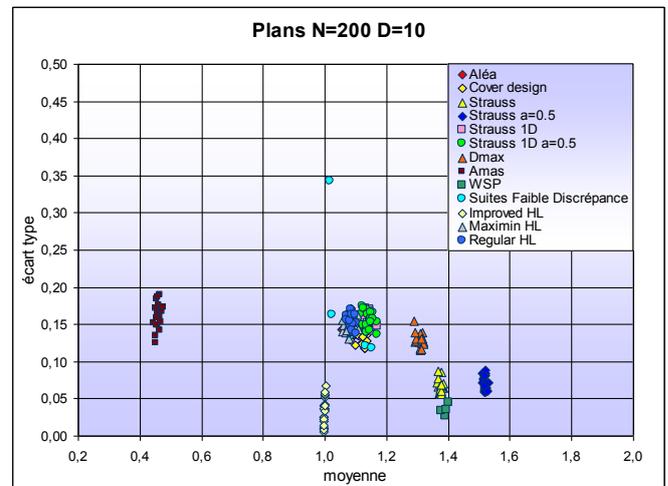
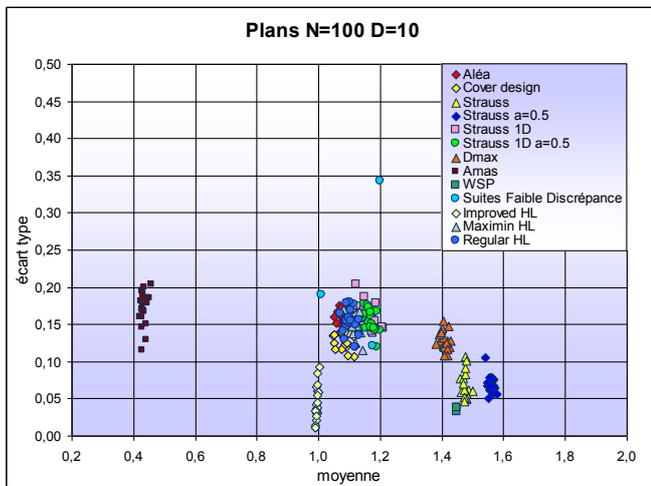
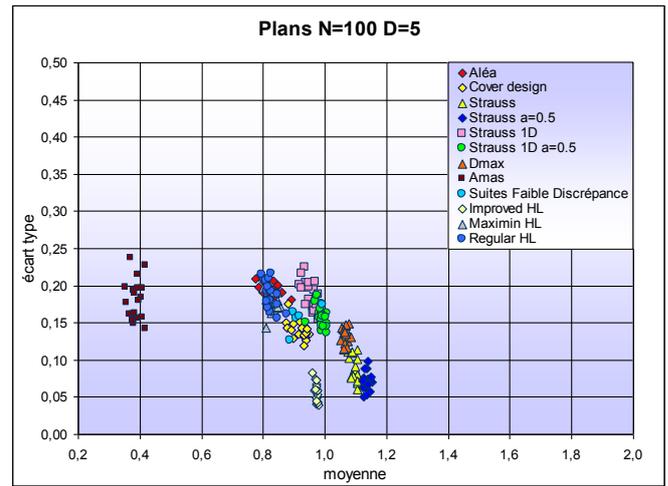
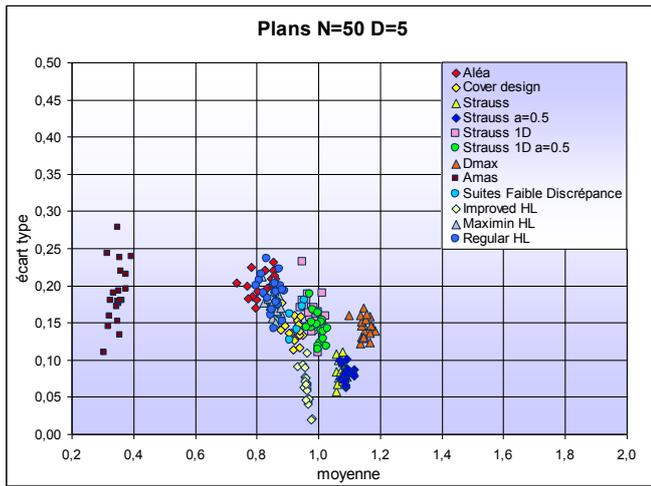


Figure 89. Représentations du critère lié aux ALM pour les plans étudiés en dimension 2, 3,5, 10 et 20. A gauche, les plans étudiés comptent 10 points par dimension. A droite, les plans étudiés comptent 20 points par dimension

Les seuls plans qui conservent de bonnes propriétés en grande dimension, au sens du critère étudié, sont les plans de Strauss, les plans à entropie maximale et les plans WSP¹⁵ (même si ces derniers n'ont pu être étudiés avec autant de détail que les autres). On remarque néanmoins que les plans de Strauss avec contraintes sur les marges se comportent de façon médiocre, l'excès de contraintes faisant, semblerait-il, qu'ils se rapprochent des distributions aléatoires en grande dimension. Les hypercubes latins, quels qu'ils soient, sont inexploitable au-delà de la dimension 5 car ils ont la même pathologie que les plans de Strauss avec des contraintes 1D. Les suites à faible discrédance confirment, *a fortiori*, leur médiocrité en grande dimension (voir le paragraphe 1.3.1).

Au vu de ces résultats, il est intéressant de regarder l'effet de la dimension notamment sur les plans de Strauss et les plans minimax (cover designs).

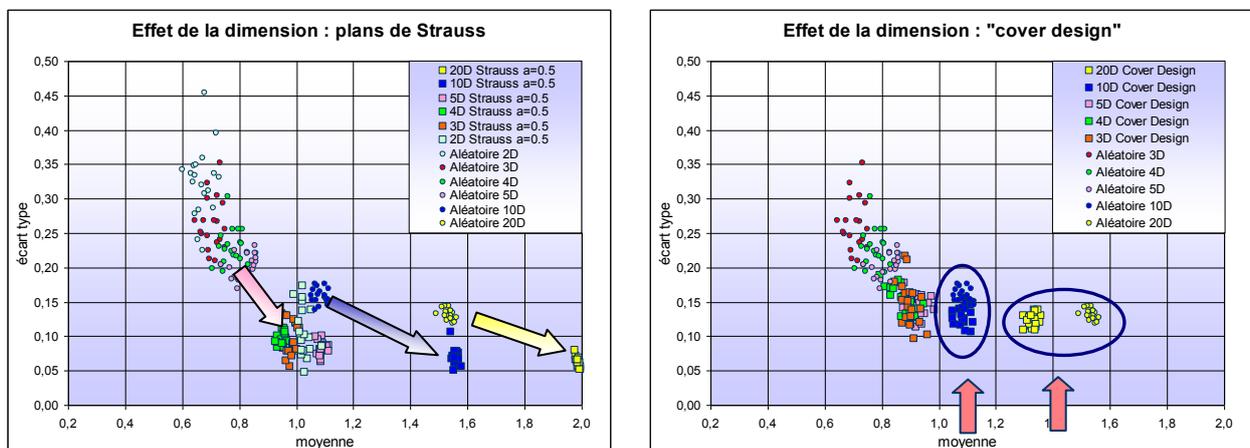


Figure 90. Représentation du critère lié aux ALM sur les plans de Strauss et les plans minimax (cover designs) pour l'étude de l'effet de la dimension

Les figures ci-dessus montrent que les plans minimax (générés par la fonction cover design du logiciel R) ne fournissent pas les résultats attendus au niveau du bon remplissage de l'espace en grande dimension (10 et 20 ici). En effet, ces plans sont équivalents à des plans aléatoires.

¹⁵ Ces plans étant difficiles à générer lorsque l'on veut un nombre de points précis, il nous a été impossible d'en construire dans certains cas ; notamment en grande dimension.

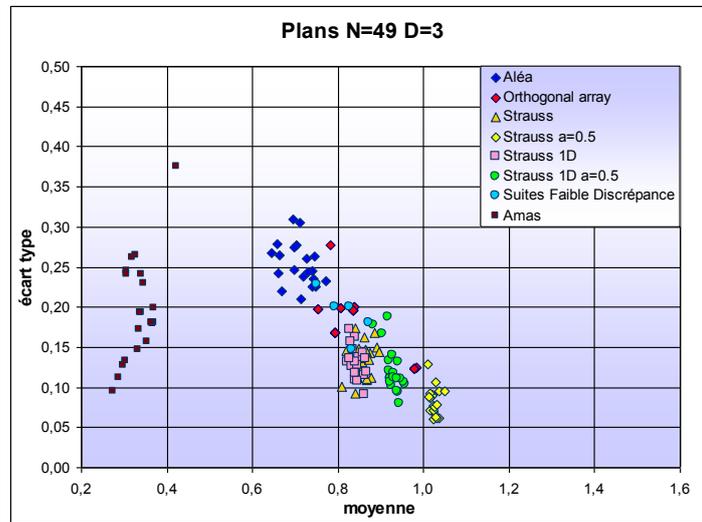
7.3. Discussion

Les plans à entropie maximale, les plans de Strauss, Strauss-Gibbs et les plans de Strauss-Gibbs avec des contraintes 1D ressortent comme étant les meilleurs au regard des critères de distance. Il est intéressant de noter que les plans de Strauss sont ceux parmi les plans précédemment cités qui ont la moins bonne discrédance. Ceci n'a rien d'étonnant notamment quand le rayon d'interaction est mal fixé car ces plans ont tendance à agglutiner les points entre eux pour minimiser le nombre d'interaction global.

L'application du critère lié aux ALM a permis de montrer que les plans de Strauss, les plans d'entropie maximale et les plans de WSP donnent de bons résultats au regard de ce critère. Elle a aussi montré que les plans considérant des contraintes sur les marges de dimension 1 tels que les hypercubes latins ou les plans de Strauss avec contraintes 1D se situent dans la zone des plans aléatoires en grande dimension. Il est également possible d'observer le phénomène inverse car les plans de Strauss avec un potentiel puissance n'ont pas une bonne discrédance centrée. Ces applications ont ainsi confirmé l'impossibilité, en grande dimension, de générer des plans avec des points bien répartis à la fois sur les marges de dimension 1 et dans l'espace. Il faut choisir entre un bon remplissage de l'espace pour capter les non-linéarités et une bonne répartition en projection dans le cas où certains facteurs ne seraient pas influents.

On remarquera que les tableaux orthogonaux linéaires en dimension 3 (voir Figure 78) ont des très bons critères de distance. Cependant, nous avons observé dans la section 1.2.2 que ces tableaux orthogonaux possèdent leurs points sur des plans parallèles. Les critères de distance ne permettent pas de remettre en cause la distribution des points car ils sont calculés en prenant en compte uniquement les distances des points les plus proches (points voisins). Les points étant sur des plans parallèles distincts, ils ne permettent donc pas de faire exploser l'un ou l'autre de ces critères. Seul le critère MaxDist est presque aussi mauvais que pour les plans aléatoires, ce qui induit qu'il ne faut pas uniquement étudier les critères de discrédance (notamment les discrédances modifiées) et de distance pour qualifier un plan de bon ou de mauvais.

C'est, par exemple, dans ce cas qu'il peut être utile de faire appel au radar d'uniformité qui permet de repérer ces problèmes (voir paragraphe 5.2) ou bien aussi au critère lié aux ALM :



PARTIE IV.

APPLICATIONS DES PLANS

D'EXPERIENCES

La dernière partie de ce mémoire a pour objectif d'étudier et de comparer les plans d'expériences discutés jusqu'ici à travers différentes applications.

Les résultats dépendant fortement du cas traité, les conclusions sont à chaque fois relatives aux données. C'est une des raisons pour laquelle, nous avons envisagé 3 applications.

Une première application en faible dimension (3D) confronte quelques plans d'expériences classiques avec des Space-Filling Designs à travers trois types de surfaces de réponses. Cette application a pour objectif principal de conforter l'intérêt réel pour les Space-Filling Designs en planification d'expériences numériques.

Ensuite, un cas d'école en dimension plus élevée (10D) a pour but d'étudier l'influence du plan d'expériences dans le contexte de la haute linéarité et de la grande dimension. Les plans utilisés lors de cette application ayant fait l'objet d'une étude intrinsèque lors de la partie III, nous discuterons du lien entre la distribution des expériences et les résultats obtenus à l'aide de celles-ci.

Enfin, nous traitons une application à l'exploration pétrolière (8D) où l'on confronte différents Space-Filling Designs avec différents types de surfaces de réponse.

Aussi, nous tenterons de dégager des conclusions similaires à ces 3 applications et d'apporter quelques éléments de réponse aux questions du type : Combien d'expériences considérer selon la surface de réponse envisagée ? Quel plan utiliser selon le type de surface de réponse ?, etc.

Chapitre 8

Cas en dimension 3 : comparaison des plans

Le but de cette application est de réaliser une étude comparative entre les plans de Strauss et les plans courants en expérimentation numérique. Le cadre de l'étude est l'approximation d'un code de calcul numérique par une surface de réponse. Suivant l'objectif fixé, différentes méthodes de construction de la surface approchée peuvent être mises en place. On se limite ici à étudier des surfaces ajustées par

- la méthode PolyMARS,
- la régression quadratique avec interactions de degré 2,
- et par krigeage ordinaire.

Le modèle ajusté dépend (plus ou moins selon les modèles) des points d'expérimentation. Par exemple, les figures ci-dessous représentent la surface approchée avec un modèle polynômial de la fonction chameau (voir Scheidt, 2006) construite à partir de différents plans d'expériences.

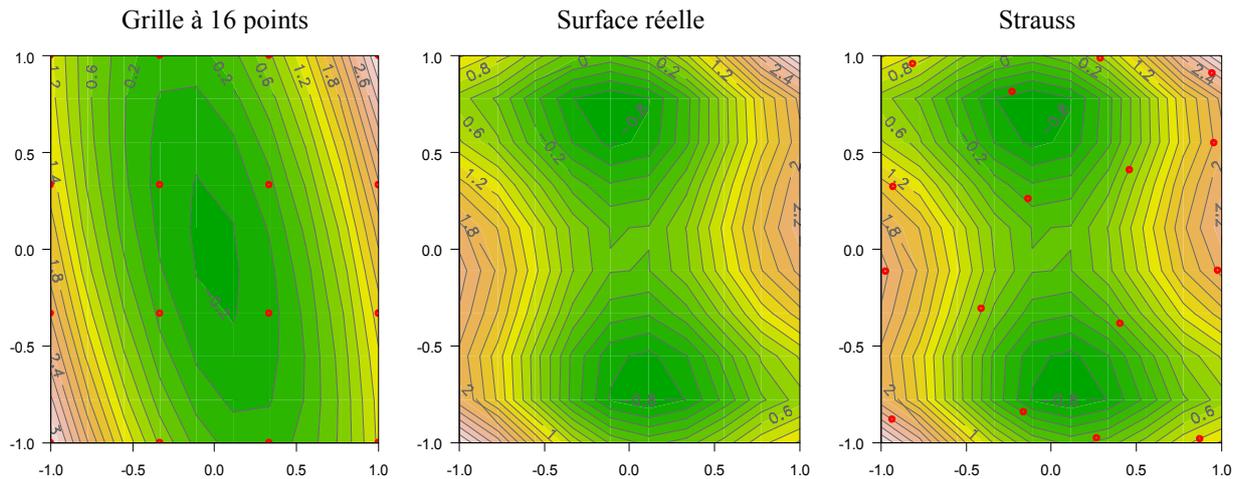


Figure 91. Représentation de la surface approchée de la fonction chameau obtenue à partir d'une grille et d'un plan de Strauss (points en rouge)

Cet exemple montre qu'une grille dans ce cas ne permet pas d'approcher correctement la surface réelle. En effet, la construction trop régulière donne une matrice d'information singulière.

Le principal objectif de ce chapitre n'est pas de comparer les trois types de surfaces de réponses envisagés ici mais d'étudier l'impact du choix du plan selon celui du modèle ainsi que celui du nombre de points à considérer.

8.1. Présentation de l'application

8.1.1. Les données

Le phénomène étudié correspond au simulateur d'écoulement *Eclipse*. Les données ont été initialement utilisées pour construire une surface de réponse permettant d'avoir une représentation simplifiée du code de calcul. La fonction d'intérêt Y représente une production cumulée d'huile totale (FOPT) au temps $T=5114$ jours. Cette réponse dépend de trois paramètres incertains : une porosité, $poro$, un coefficient d'anisotropie verticale de perméabilité, $mk3$, et un coefficient lié aux perméabilités relatives, kr . Dans ce cas d'étude, on dispose d'une connaissance approfondie du phénomène à l'aide d'un plan factoriel complet à 11 niveaux dans $[-1, 1]^3$.

En pratique, le nombre de simulations doit être suffisamment important pour estimer tous les coefficients du modèle approché mais également limité pour réduire le temps de calcul. Ainsi, le choix des points à évaluer est important, d'où l'intérêt de cette étude.

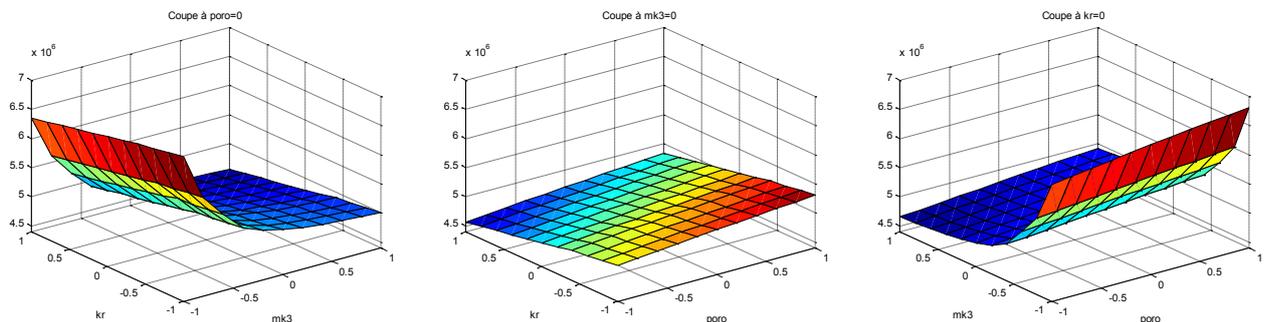


Figure 92. Représentation de la surface réelle

On remarque que la surface réelle est assez lisse et linéaire sauf dans la direction $mk3$.

Cette étude vise à comparer les plans de Strauss avec des plans connus tels que les suites à faible discrédance, des hypercubes latins, des plans D-optimaux, aléatoires, etc.

Dans un premier temps, on effectue une étude globale sur les plans ayant un nombre de points variant de 20 à 100 afin d'analyser l'influence du nombre de points et des méthodes d'expérimentation. Par la suite, une étude plus approfondie pour pallier la nature aléatoire de certains plans (hypercubes latins, Strauss, etc.) sera effectuée sur 100 plans à 30 points ayant les mêmes caractéristiques de construction. Les points du plan n'appartenant pas forcément à la grille 11^3 , une régression linéaire est effectuée afin d'avoir une valeur approchée de la réponse en chacun des points.

Remarque. Pour tirer des conclusions plus pertinentes, il aurait été sans doute préférable de conduire cette étude sur un plus grand nombre de plans. Cependant, des tests effectués sur 1000 plans de Strauss ont donné des résultats sensiblement identiques à ceux obtenus avec 100 plans.

8.1.2. Méthodologie

Les différents plans sont comparés en testant le pouvoir prédictif du modèle ajusté après avoir vérifié la qualité d'ajustement du modèle. En pratique, un modèle de surface de réponse est ajusté pour chaque plan puis la réponse est prédite sur un plan test (dans notre cas, la grille complète 11^3). Plusieurs critères – définis en annexe B – sont calculés et permettent de comparer la surface réelle obtenue par le simulateur sur les 1331 points et la surface approchée.

Dans un premier temps, la méthode PolyMARS¹⁶ est utilisée afin de construire une surface approchée pour chaque plan proposé. Ensuite, une régression quadratique avec interaction a été envisagée. Et pour conclure, les résultats par krigeage ordinaire sont donnés.

¹⁶ Les résultats présentés ont été obtenus avec la fonction `polymars` du package `polyspline` du logiciel R

8.2. Comparaison des plans d'expériences

8.2.1. Les plans

Les plans utilisés dans cette étude peuvent être classés en deux catégories : les plans classiques et les Space-Filling Designs. L'intérêt est de voir si les SFD donnent globalement de meilleurs résultats et même plus particulièrement d'étudier les performances des plans de Strauss.

Les plans testés sont les suivants :

	Plans	Nombre de points	Nombre de plans
Classiques	<i>CFC (+ variante)</i>	15	2
	<i>Box Behnken</i>	13	1
	<i>Doehlert</i>	13 + 67	1+2
	<i>D-optimaux</i>	30 à 50	13
	<i>Plans factoriels à 3, 4 et 5 niveaux</i>	27, 64 et 125	1+1+2
SFD	<i>Hypercubes latins (HL)</i>	(20, 30, ...,100)	10x9
	<i>HL correlation</i> ¹⁷	(20, 30, ...,100)	10x9
	<i>HL maximin</i>	(20, 30, ...,100)	10x9
	<i>Faure</i>	(20, 30, ...,100)	1x9
	<i>Halton</i>	(20, 30, ...,100)	1x9
	<i>Hammersley</i>	(20, 30, ...,100)	1x9
	<i>Sobol</i>	(20, 30, ...,100)	1x9
	<i>Niederreiter</i>	(20, 30, ...,100)	1x9
	<i>Strauss</i> ¹⁸	(20, 30, ...,100)	10x9
	<i>Strauss a</i> ¹⁹	(20, 30, ...,100)	10x9
	<i>Strauss 1D</i> ²⁰	(20, 30, ...,100)	10x9
	<i>Strauss 1D a</i> ²¹	(20, 30, ...,100)	10x9
	<i>Cover.Design</i>	(20, 30, ...,100)	10x9
<i>Tableau orthogonal</i>	49	1	
Aléatoires	(20, 30, ...,100)	10x9	

Tableau 8. Liste des plans d'expériences utilisés dans les études avec le modèle PolyMARS et la régression quadratique

¹⁷ Tous les hypercubes latins utilisés dans cette étude ont été générés avec la routine de Matlab.

¹⁸ Strauss usuel (avec un potentiel tout ou rien)

¹⁹ Strauss-Gibbs avec un potentiel de puissance 0.5

²⁰ Strauss en 3D et avec des contraintes de bonne répartition sur les marges de dimension un.

²¹ Strauss-Gibbs avec un potentiel de puissance 0.5 en 3D et avec des contraintes de bonne répartition sur les marges de dimension un.

8.2.2. Résultats de l'étude avec le modèle PolyMARS

On présente ici les résultats de l'étude menée avec une surface de réponse ajustée par PolyMARS (voir annexe A).

Les figures suivantes représentent le critère R^2 pour mesurer la qualité de l'ajustement du modèle (à gauche) ainsi que la $RMSE$ permettant de mesurer le pouvoir prédictif du modèle (à droite).

Le nombre de plans testés étant important, nous proposons en premier lieu une comparaison entre les plans de Strauss et les plans standard.

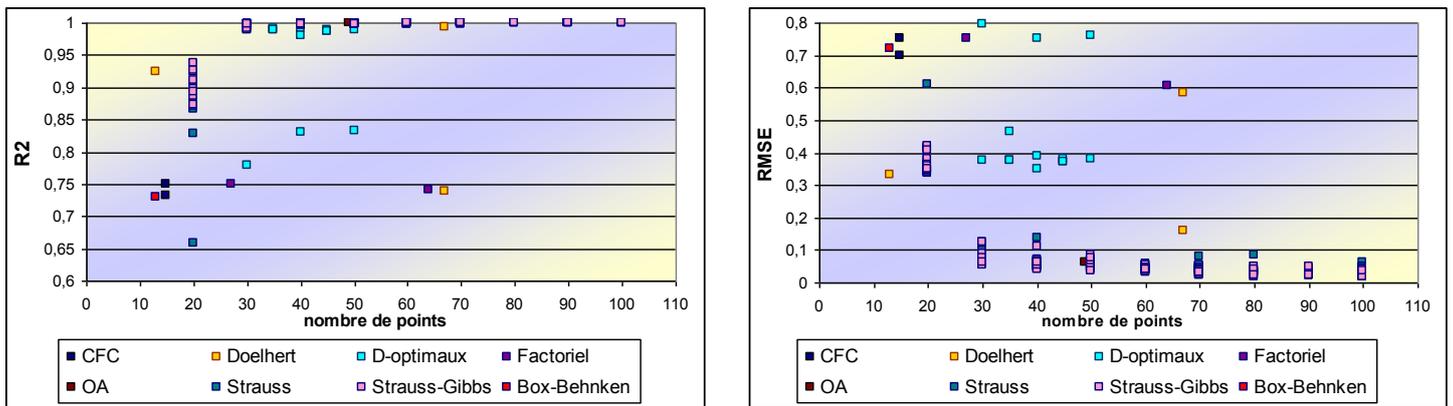


Figure 93. Evolution des critères en fonction du nombre de points (Strauss / plans standard)

La figure précédente montre que les plans ayant moins de 20 points donnent des surfaces de réponse de qualité assez médiocre. Une grande variabilité du critère R^2 est observée même pour les plans de Strauss.

Une remarque intéressante est que le plan factoriel à 27 points donne les résultats similaires à ceux obtenus avec le plan composite à faces centrées à 15 points et ceux du plan Box-Behnken à 13 points. Ce résultat s'explique simplement par le fait que ces 3 plans ont des projections identiques sur les sous-espaces de dimension 2. Ce qui tend à confirmer l'intérêt des Space-Filling Designs.

A noter que le plan de Doehlert à 13 points donne de bons résultats malgré un nombre de points d'apprentissage assez faible. Nous avons aussi testé deux plans de Doehlert à 67 points et on constate une différence significative entre les deux surfaces ajustées car un des plans (voir Figure 94 et Figure 95) a de bonnes projections dans la direction $mk3$ (Predictor 2 ci-dessous). On obtient ainsi une bonne modélisation dans cette direction car la surface est très régulière et présente une courbure uniquement dans la direction $mk3$ (voir Figure 92).

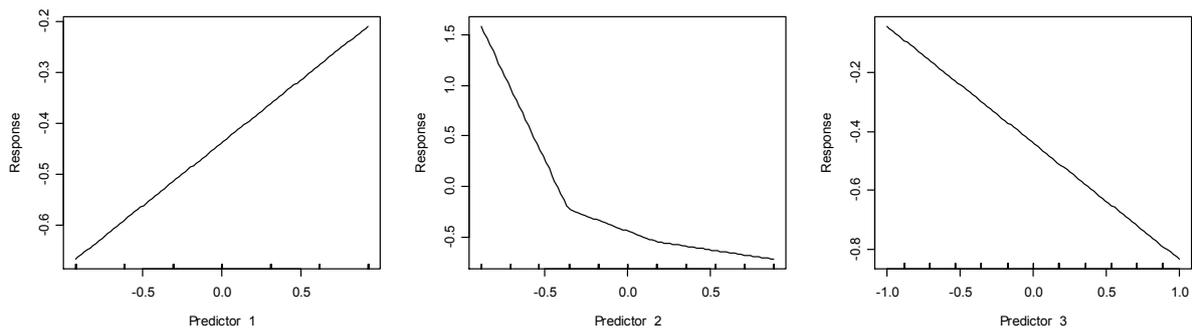


Figure 94. Représentation des coupes de la surface approchée pour le plan de Doehlert à 67 points.

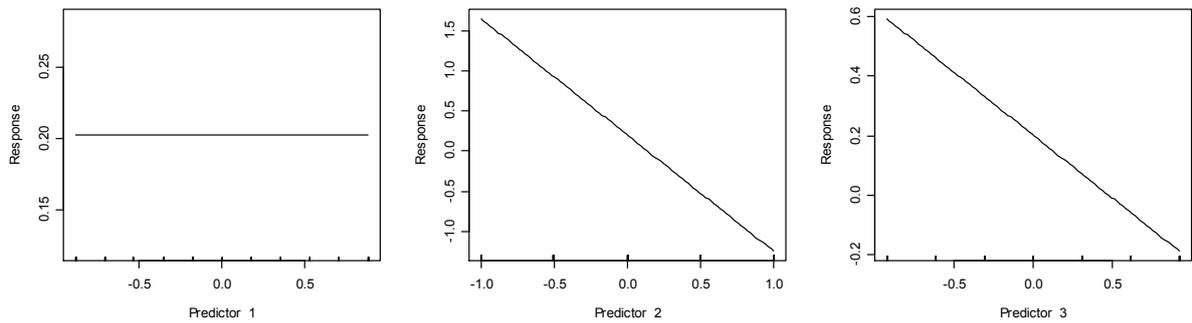
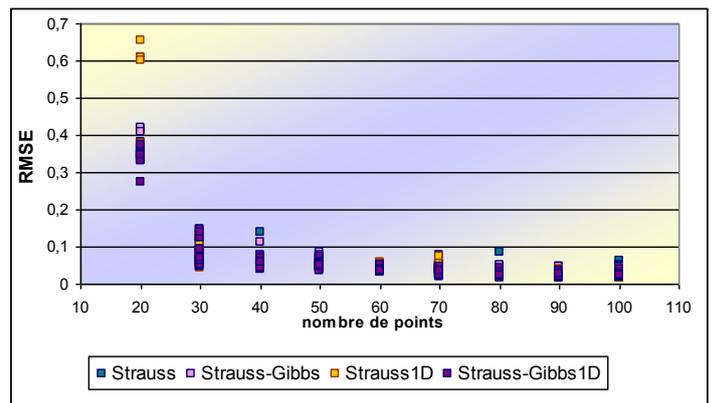
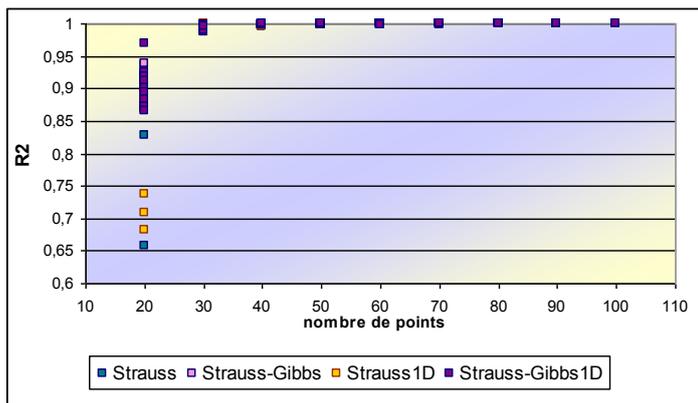


Figure 95. Représentation des coupes de la surface approchée pour le plan de Doehlert à 67 points permuté

Nous nous sommes intéressés plus particulièrement à comparer les critères sur les différents plans de Strauss. Les figures du bas correspondent à un zoom afin de mieux observer les phénomènes lorsque le nombre de points du plan d'expériences est supérieur à 30.



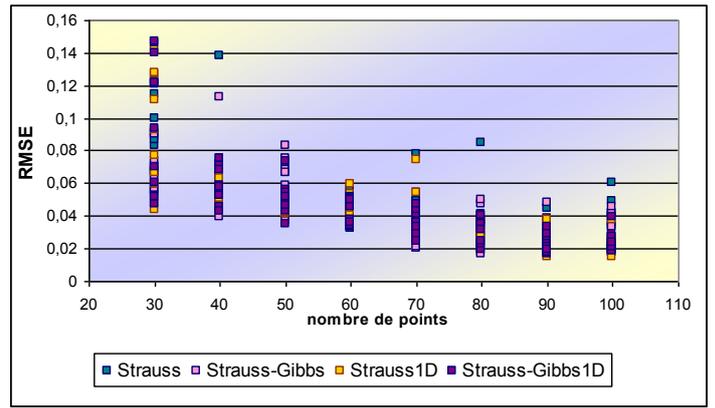
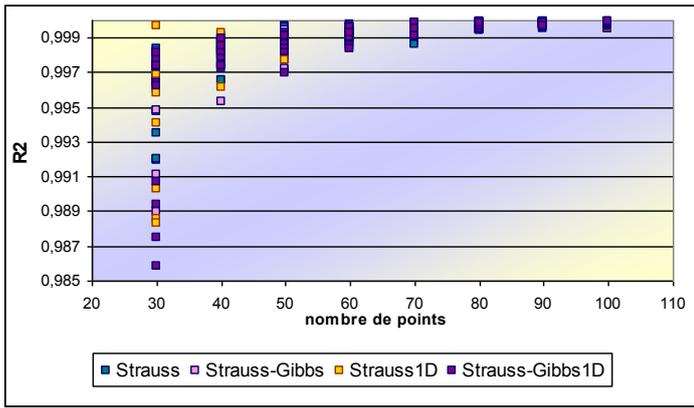


Figure 96. Evolution des critères en fonction du nombre de points (Strauss)

Ces figures montrent une stabilisation et une dispersion plus faible de la valeur des critères quand le nombre de points augmente. D'autre part, on observe que les plans de Strauss-Gibbs à 20 points générés avec un potentiel puissance (Strauss-Gibbs et Strauss-Gibbs 1D) donnent des critères beaucoup moins dispersés que ceux générés avec un potentiel de type tout ou rien. Cela peut s'expliquer par le fait qu'avec le potentiel tout ou rien il est possible d'avoir parfois des points très proches si le rayon d'interaction est mal adapté.

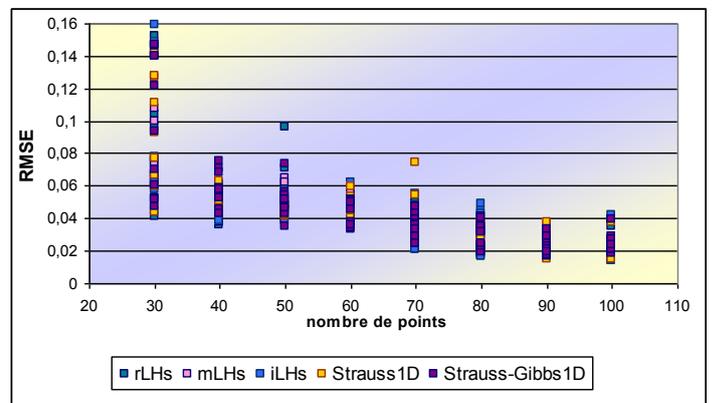
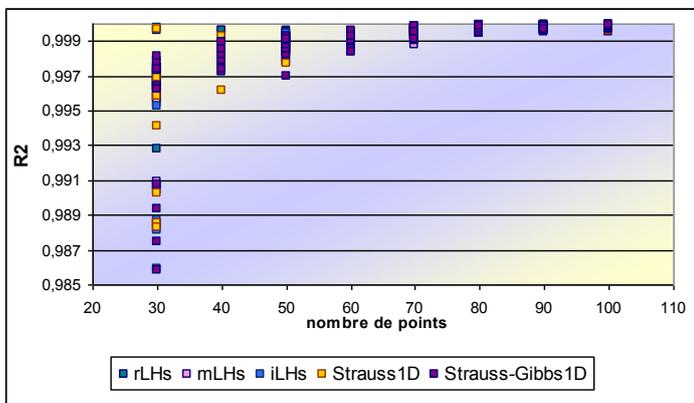
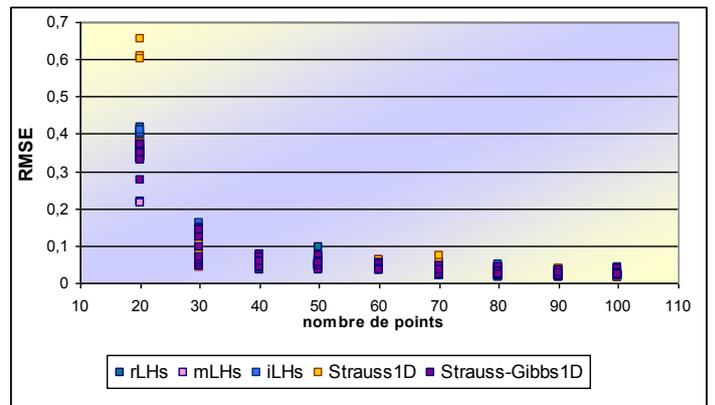
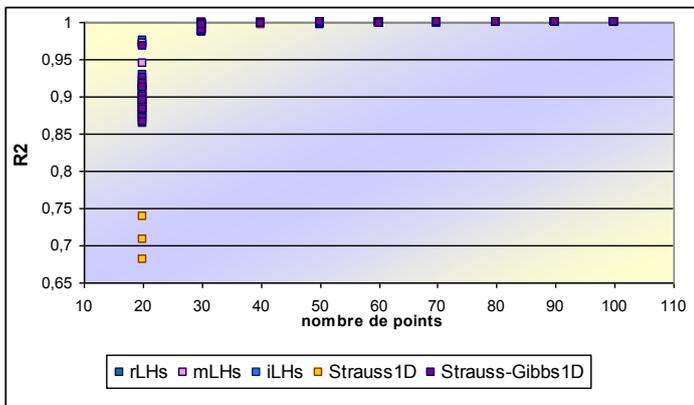


Figure 97. Evolution des critères en fonction du nombre de points (hypercubes latins et Strauss 1D)

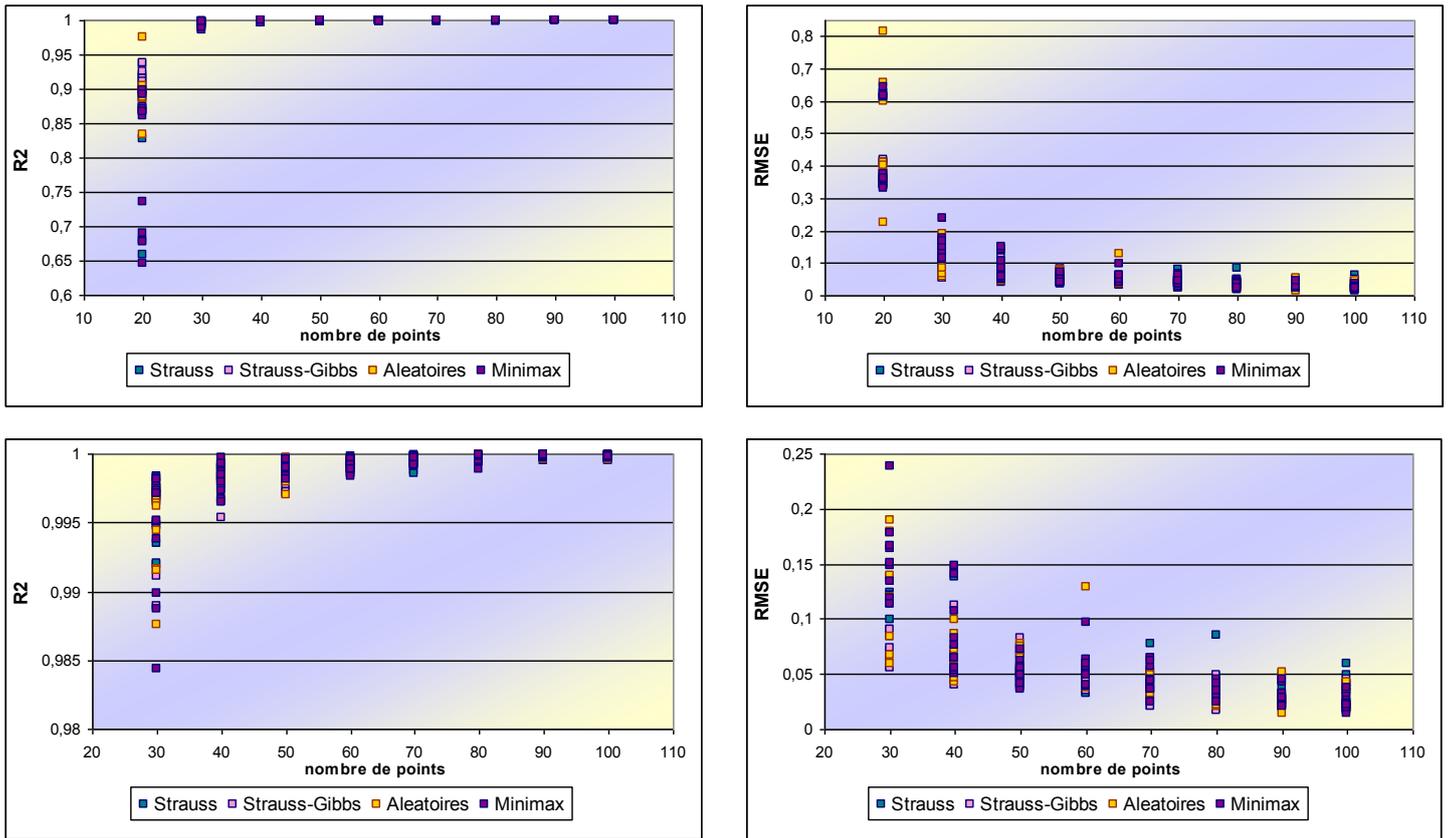


Figure 98. Evolution des critères en fonction du nombre de points (Strauss / aléatoires /Minimax)

On peut en conclure, pour ce cas d'étude, que 10 points par dimension semble être un bon compromis au regard de la contrainte de taille. Nous détaillons à la section suivante une étude comparative pour des plans à 30 points.

On remarque également que plus le nombre de points est important, plus les méthodes utilisées pour générer les SFD n'ont pas d'influence sur le résultat. En effet, la méthode PolyMARS ajuste et prédit correctement la réponse (R^2 tend vers 1 et $RMSE$ tend vers 0) comme on peut l'observer sur la figure ci-dessous obtenue pour un plan de Strauss à 70 points.

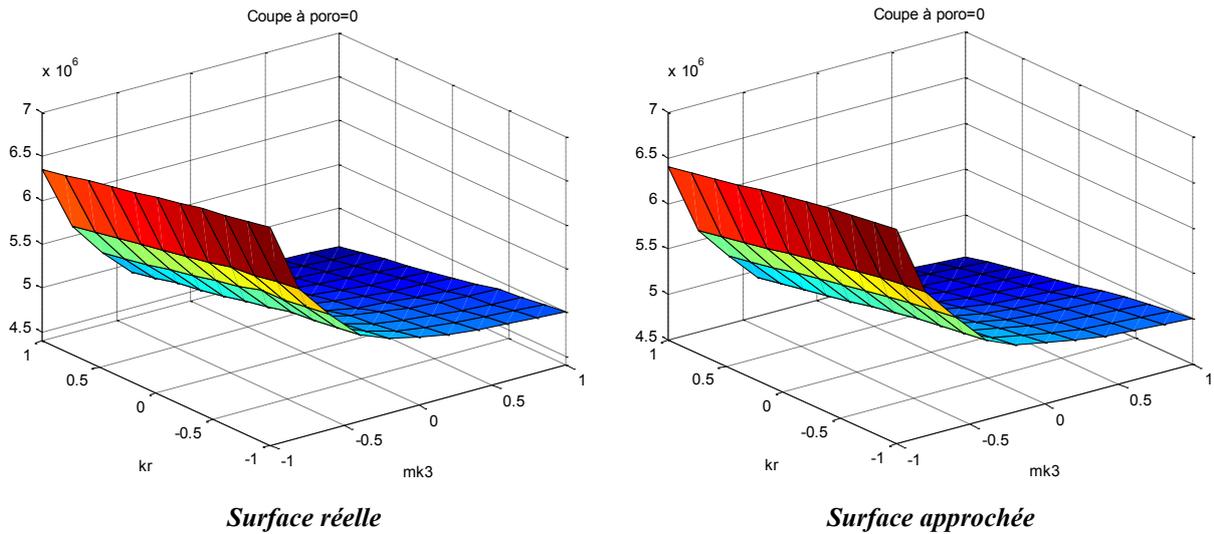


Figure 99. Comparaison en coupe de la surface simulée et approchée par PolyMARS pour un plan de Strauss à 70 points

8.2.3. Résultats de l'étude avec une régression de degré 2

Une étude similaire a été faite avec une surface de réponse ajustée par une régression polynomiale de degré 2 avec interactions d'ordre 2.

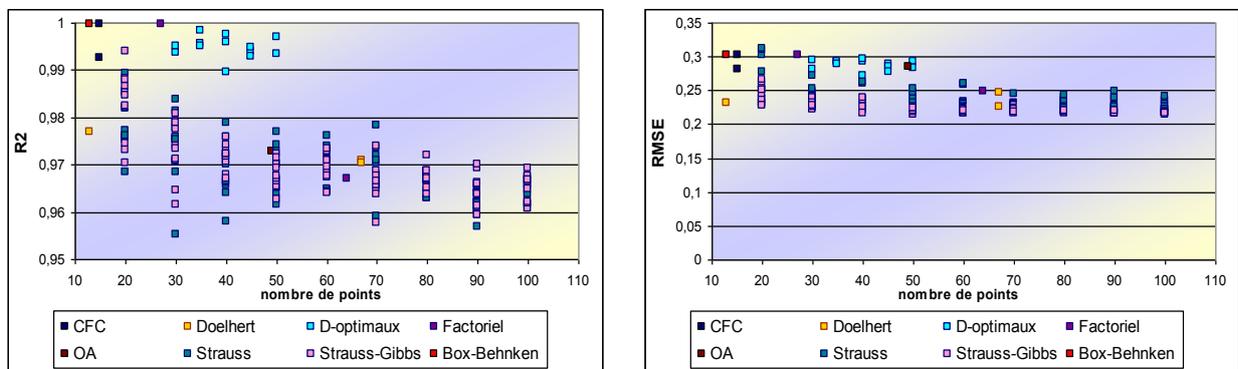


Figure 100. Critères pour le modèle quadratique (Strauss / plans standard)

Les figures précédentes montrent que les plans D-optimaux ajustent mieux les données. Cependant, les valeurs du critère *RMSE* sont relativement élevées : cela s'explique par le compromis biais-variance (illustré au paragraphe 10.2.2.2.).

Nous remarquons que contrairement au modèle PolyMARS, le nombre de points ne semble pas influencer sur les résultats obtenus pour ces plans. En effet, l'asymptote des critères en leur valeur cible observée n'est plus visible avec le modèle quadratique. Cela s'explique par la rigidité du modèle polynomial.

Les figures suivantes montrent que le nombre de points influe principalement sur la dispersion des valeurs calculées.

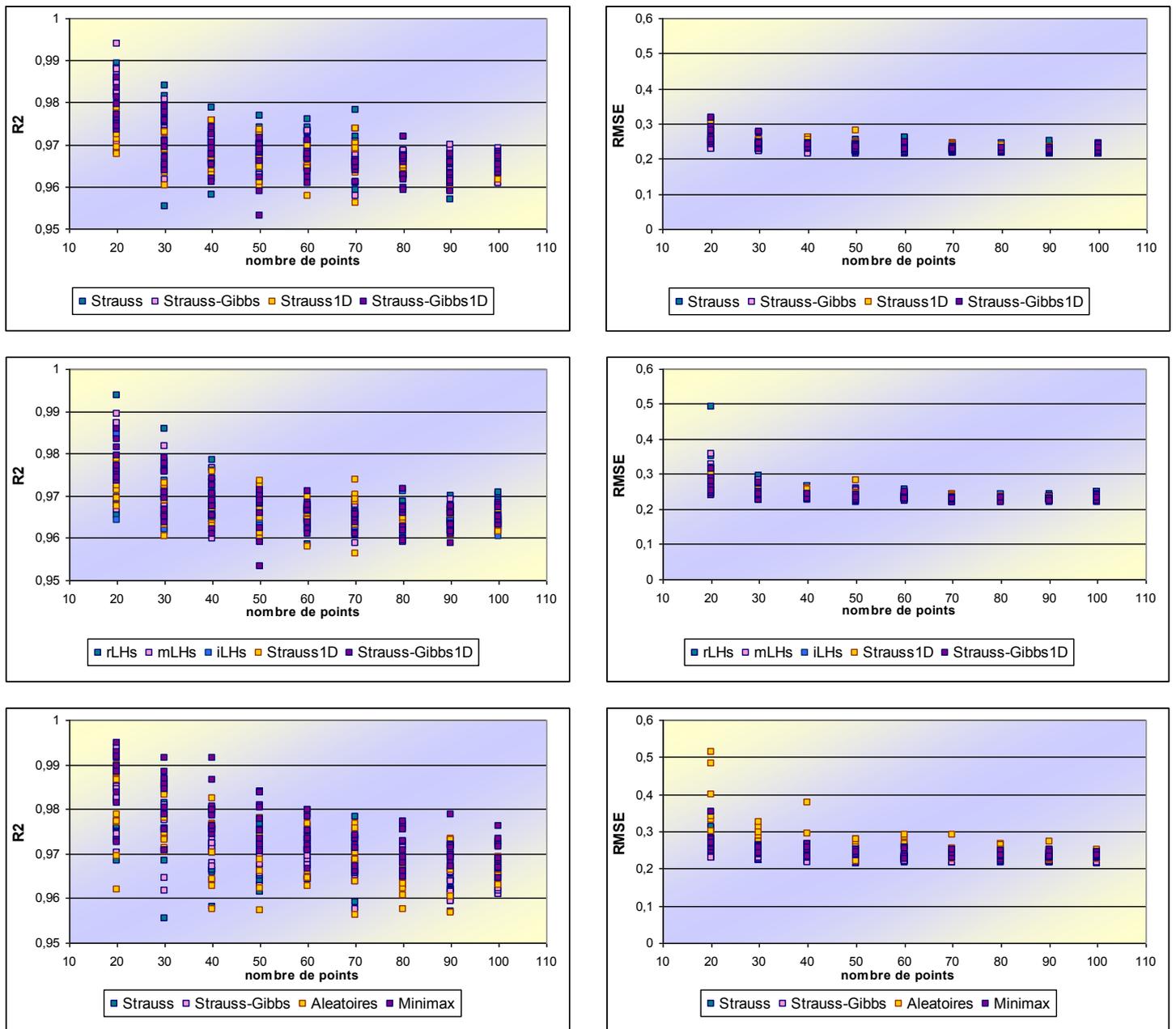


Figure 101. Critères pour le modèle quadratique (Strauss et autres SFD)

On peut remarquer que les critères sont beaucoup plus dispersés pour les plans aléatoires, ce qui conforte l'intérêt des plans SFD.

8.2.4. Résultats de l'étude avec le krigeage ordinaire

Les qualités prédictives du modèle dépendent aussi de la méthode utilisée pour construire la surface de réponse. Pour compléter cette étude comparative des plans d'expériences, des résultats similaires ont été obtenus pour le krigeage.

Le krigeage est une méthode stochastique d'interpolation spatiale qui prévoit la valeur d'une réponse par une combinaison linéaire sans biais et à variance minimale des données (voir annexe A).

La liste des plans que nous avons utilisés pour cette étude, est détaillée dans le tableau ci-dessous.

Plans		Nombre de points	Nombre de plans	Plans		Nombre de points	Nombre de plans
Classiques	<i>CFC (+ variante)</i>	15	2	SFD	<i>HL correlation</i>	20 à 100	9
	<i>Box Behnken</i>	13	1		<i>HL maximin</i>	20 à 100	9
	<i>Doehlert</i>	13 + 67	1+2		<i>Faure</i>	20 à 100	9
	<i>D-optimaux</i>	30 à 50	13		<i>Halton</i>	20 à 100	9
	<i>Plan factoriel à 3 niveaux</i>	27	1		<i>Hammersley</i>	20 à 100	9
	<i>Plans factoriels à 4 et 5 niveaux</i>	64 et 125	1+2		<i>Sobol</i>	20 à 100	9
					<i>Strauss usuel</i>	20 à 100	9
					<i>Tableau orthogonal</i>	49	1

Tableau 9. Liste des plans d'expériences testés

La figure ci-dessous représente le critère $RMSE$ en fonction du nombre de points. A noter que le critère considéré ici n'est pas exactement le même que précédemment :

$$RMSE_{relative} = \sqrt{\frac{\sum_{i=1}^{1331} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{1331} y_i^2}}$$

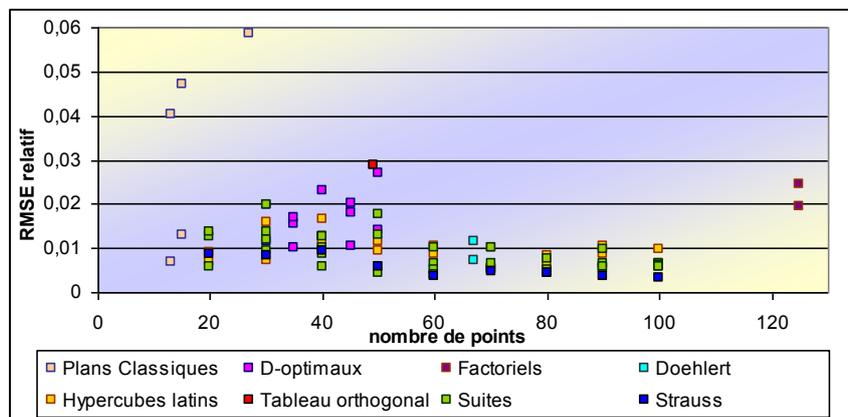


Figure 102. Krigeage ordinaire

Nous obtenons des valeurs très faibles du critère ($<0,06$) dont l'ordre de grandeur semble cohérent avec les données (réponse réelle de l'ordre de 10^7 et résidus en 10^5). Néanmoins, nous ne connaissons pas de valeur cible pour ce critère, il faudrait étudier plus en détail chaque surface approchée pour vérifier si les écarts entre les différents modèles sont réellement significatifs.

Sur la Figure 102, on constate que les plans de Strauss avec certains hypercubes latins et quelques suites à faible discrédance estiment généralement de meilleurs modèles que les autres plans. Cependant, pour des plans ayant un nombre de points supérieur à 60, on dispose de peu de plans classiques de comparaison. On observe malgré tout que les plans factoriels 4^3 et 5^3 donnent des résultats assez médiocres car ils sont équivalents aux résultats obtenus avec des plans de 40 points.

En étudiant uniquement en détail le critère de prédiction $RMSE_{relative}$ obtenu pour les suites à faible discrédance et les plans de Strauss (voir Figure 103) on constate que la suite de Hammersley donne de bons résultats. A noter qu'à cause de la régularité des suites à faible discrédance, il peut y avoir des problèmes numériques récurrents lors de l'optimisation des paramètres du krigeage. Par exemple, la suite de Halton à 20 points ne permet pas d'obtenir un modèle de krigeage. Ces problèmes numériques semblent très peu présents avec les plans de Strauss.

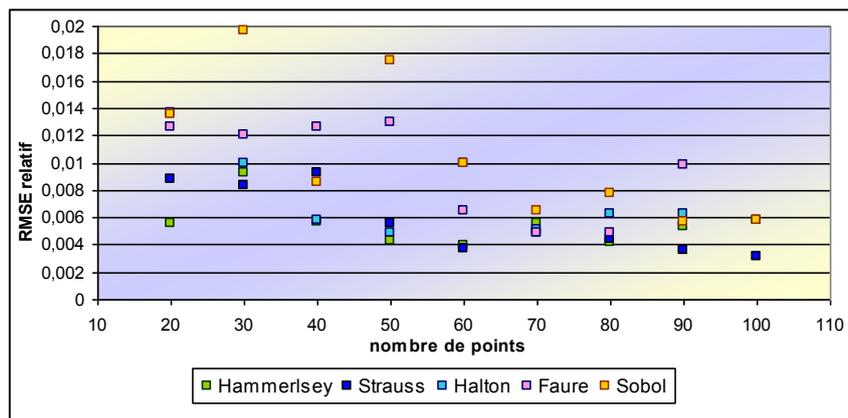


Figure 103. Krigeage ordinaire avec les suites à faible discrédance et les plans de Strauss

8.2.5. Conclusions

Ces 3 études préliminaires montrent que les SFD permettent généralement d'avoir de meilleurs résultats que ce soit avec un modèle PolyMARS, de krigeage ou bien même de régression quadratique.

Nous avons vu également qu'ici la suite de Halton peut induire des problèmes au niveau de l'optimisation des paramètres du krigeage. Il est alors préférable d'envisager des plans dont les points sont bien répartis dans l'espace et de manière « aléatoire » tels que les plans de Strauss ou les plans à entropie maximale.

A noter qu'avec le krigeage, il faut utiliser les SFD avec précaution car la tendance à éloigner les points les uns des autres peut amener trop peu d'information pour estimer les portées petites.

Concernant l'influence du nombre de points, nous avons pu vérifier que celui-ci est assez dépendant du modèle envisagé. Par exemple, pour la régression il n'y a pas beaucoup d'importance à y attacher alors que pour le krigeage on préférera choisir des plans à 50 points.

La section suivante a pour but d'étudier la qualité de la surface de réponse obtenue avec les modèles PolyMARS et de régression quadratique à partir de plans d'expériences à 30 points.

8.3. Qualité de la surface de réponse obtenue à partir de plans d'expériences à 30 points

Au vu des résultats précédents et présentant uniquement dans cette section les résultats pour PolyMARS et pour la régression, cette étude se limite à des plans d'expériences contenant 30 points.

Les critères utilisés pour mesurer la qualité d'ajustement (R^2) et prédictive ($RMSE$) du modèle ont été évalués sur 100 plans afin de mieux rendre compte de leur influence du fait qu'ils introduisent un aléa stochastique.

Les plans étudiés sont les suivants :

Plans		Nombre de plans
Aléatoires		100
SFD	<i>Minimax</i> ²²	100
	<i>Hypercubes Latins (HL)</i>	100
	<i>HL correlation</i> ²³	100
	<i>HL maximin</i> ²⁴	100
	<i>Strauss</i>	100
	<i>Strauss G</i> ²⁵	100
	<i>Strauss 1D</i> ²⁶	100
	<i>Strauss 1D G</i>	100

Tableau 10. Liste des plans étudiés avec leur caractéristique

²² plans de type minimax obtenus à l'aide de la routine *cover.design* du package *Fields* de R

²³ hypercubes latins avec le meilleur critère de corrélation

²⁴ hypercubes latins avec le meilleur critère maximin

²⁵ plans de Strauss avec un potentiel puissance

²⁶ plans de Strauss avec des contraintes sur les marges de dimension 1

8.3.1. Etude avec le modèle PolyMARS

Les figures ci-dessous représentent les boxplots des critères de qualité (ajustement et prédiction) calculés sur tous les plans introduisant de l'aléa stochastique.

Qualité d'ajustement du modèle

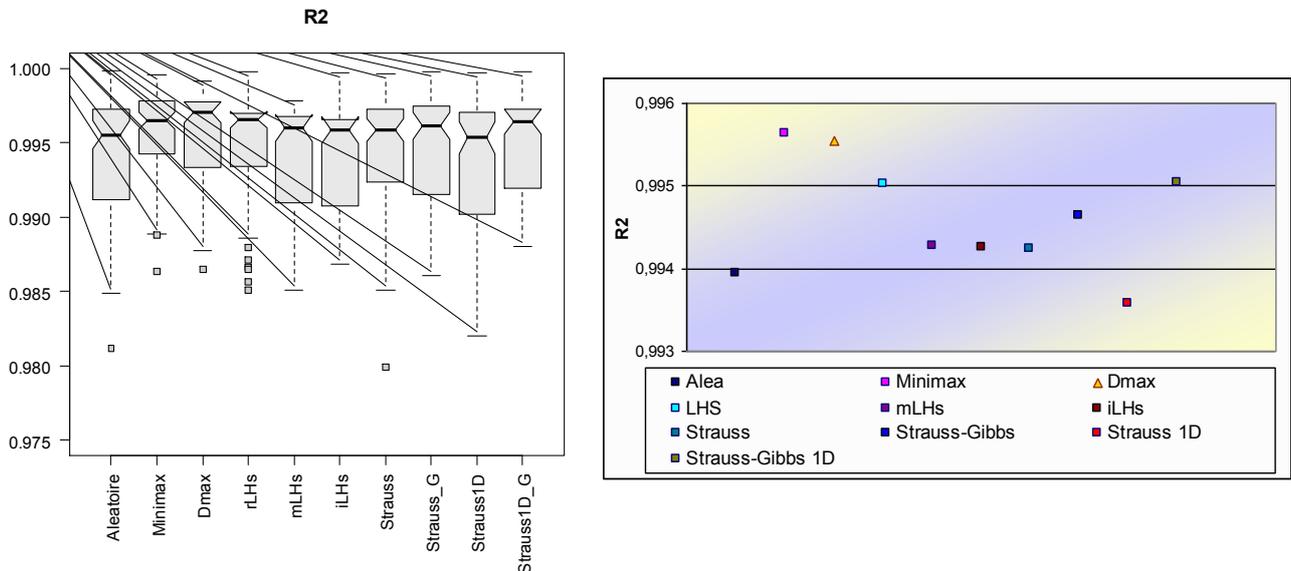


Figure 104. A gauche, représentation du boxplot pour mesurer la qualité d'ajustement du modèle PolyMARS. A droite, représentation du critère R^2 moyen pour chaque type de plans

Généralement les modèles ajustés sont de très bonne qualité quel que soit le plan d'expériences (>0.98). On observe cependant quelques valeurs aberrantes notamment pour les hypercubes latins (rLHs).

Le meilleur ajustement est obtenu à l'aide des plans minimax et des plans à entropie maximale. Il semble étonnant qu'en moyenne les plans de Strauss en 1D ajustent des modèles de moins bonne qualité que les plans aléatoires même si ceux-ci restent de qualité relativement proche. Afin de vérifier si ces remarques sont dues au compromis biais-variance, nous regardons ce qu'il en est au niveau de la prédiction.

Qualité de prédiction du modèle

Tous les modèles ajustés étant pertinents, on s'intéresse donc à la *RMSE* afin de mesurer la qualité de prédiction des modèles.

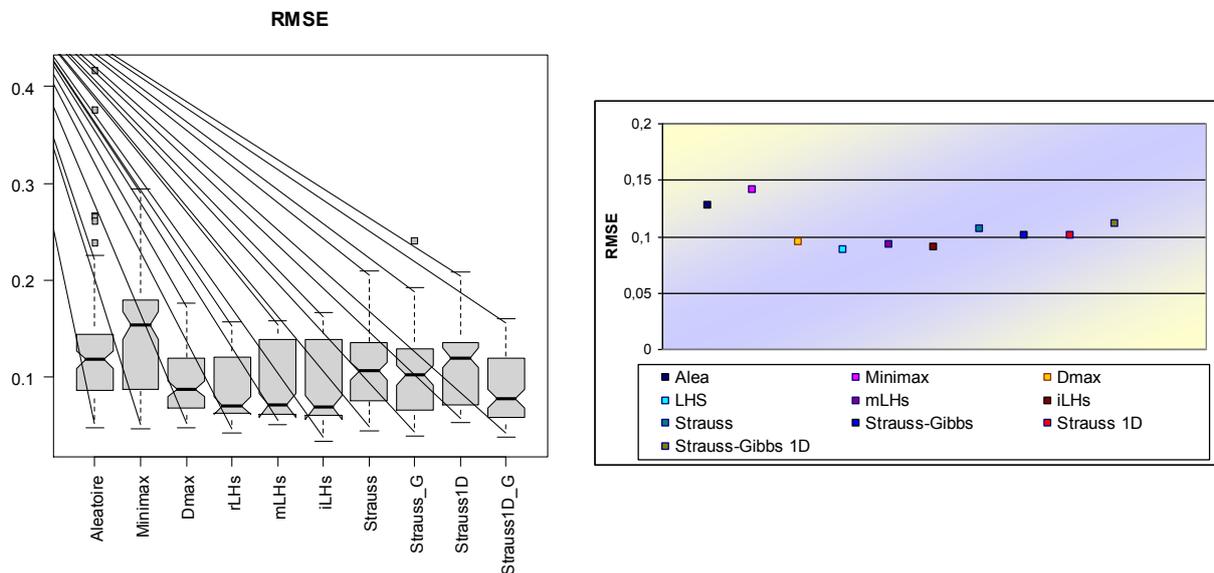


Figure 105. A gauche, représentation du boxplot pour mesurer la prédiction du modèle PolyMARS. A droite, représentation du critère *RMSE* moyen pour chaque type de plans

Comme pour les critères de qualité d'ajustement du modèle, nous avons des critères de qualité de prédiction très bons pour les plans d'expériences numériques.

La figure ci-dessus confirme qu'au niveau prédictif les plans de Strauss 1D ne sont pas plus mauvais que les plans aléatoires.

Remarque. On peut redécouvrir le compromis biais-variance notamment pour les plans minimax.

Cette figure montre bien la différence entre les plans aléatoires et les Space-Filling Designs (n'incluant pas les plans minimax qui eux semblent être de mauvaise qualité). On se rappellera également que dans la partie précédente nous avons vu que la grille (à 27 points) et le plan D-optimal (à 30 points) ne donnent pas de bons résultats que ce soit pour la qualité d'ajustement ou de prédiction du modèle, ce qui conforte l'idée d'utiliser des Space-Filling Designs en phase exploratoire.

Nous nous sommes intéressés à regarder également la potentialité des plans de Strauss à distance de densité gaussienne pour voir s'ils pouvaient nous apporter plus de renseignements au centre du domaine et donc permettre de retrouver la courbure dans la direction *mk3* plus facilement.

Voici les résultats obtenus :

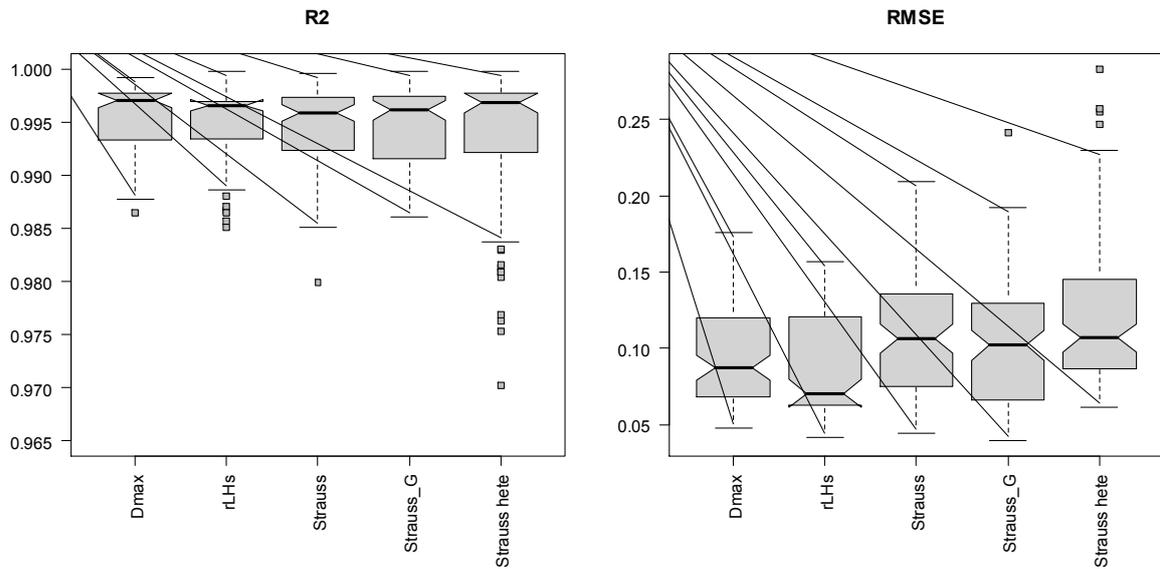


Figure 106. Représentation des critères pour comparer quelques SFD avec les plans de Strauss de distance à densité gaussienne ($\sigma=1$, $\rho=0$, $R=0.2$)

Les figures ci-dessus montrent qu'au niveau de l'ajustement (figure de gauche) il y a un peu plus de dispersion mais que ces plans restent malgré cela de bonne qualité. Au niveau de la prédiction (figure de droite), on voit que ces plans sont d'une moins bonne qualité que les SFD mais ne donnent pas non plus de très mauvais résultats. En effet, si on les compare avec les plans aléatoires et les plans minimax de la Figure 105, on voit que les plans hétérogènes permettent de donner de meilleurs résultats que ces derniers.

Remarque. La courbure se situant vers la valeur -0.5 sur le domaine $[-1,1]^3$, il serait plus pertinent de densifier les points dans cette zone.

La faible dimension de ce problème n'offre apparemment pas la possibilité à ces plans de se démarquer. Il pourrait réellement être intéressant de les tester dans un contexte de dimension plus élevée afin de voir s'ils permettent de corriger les effets de bords de la grande dimension.

8.3.2. Etude avec le modèle de régression quadratique

Les plans utilisés dans cette étude sont les mêmes que ceux mis en comparaison dans le paragraphe précédent.

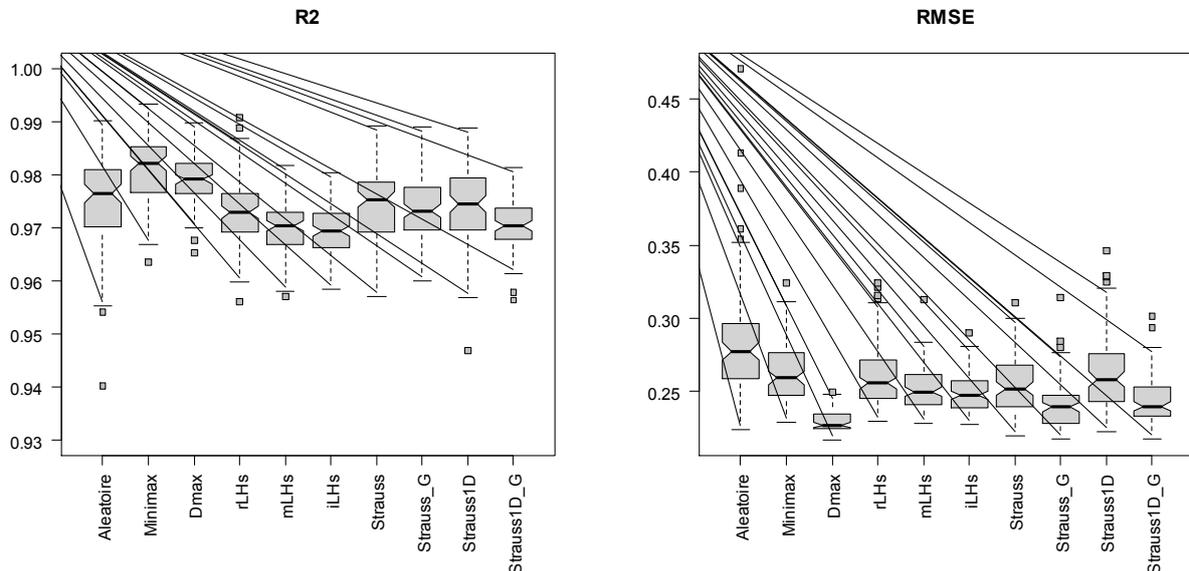


Figure 107. Boxplot pour mesurer la qualité d'ajustement (à gauche) et de prédiction (à droite) du modèle de régression quadratique des différents plans

Similairement à ce que l'on a pu observer avec le modèle PolyMARS nous avons de très bons résultats en ajustement quel que soit le critère et quel que soit le plan. Seuls les plans de types hypercubes latins sont légèrement moins bons.

Au niveau de la prédiction, on voit sur la figure de droite que les résultats relativement médiocres des plans de Strauss avec des contraintes 1D, des hypercubes latins usuels et des plans minimax sont sensiblement identiques. La dégradation de la qualité des plans de Strauss comparativement aux plans de Strauss-Gibbs s'explique encore une fois par le fait que le rayon d'interaction, s'il est mal adapté, peut engendrer de mauvaises configurations de plans. Enfin, les plans à entropie maximale et les plans de Strauss-Gibbs semblent remarquablement bien appropriés pour ce modèle.

8.4. Conclusions de l'étude

Cette étude a permis de montrer l'intérêt des Space-Filling Designs en phase exploratoire. Nous avons vu que globalement les plans à entropie maximale et les plans de Strauss sont ceux qui ont donné les résultats les moins dispersés et les meilleurs. Cependant, cette application n'a pas permis de différencier significativement les plans de Strauss au regard des autres SFD. Cela peut provenir non seulement de la faible dimension du problème proposé mais aussi de la linéarité de la réponse.

C'est pourquoi nous considérons, au chapitre suivant, une application de comparaison des plans en dimension 10 afin d'étudier leur robustesse et leurs qualités en grande dimension.

Chapitre 9

Cas d'école : fonction déterministe en dimension 10

L'objectif de cette étude est de comparer divers plans d'expériences usuels sur une fonction déterministe de 10 variables, considérée « hautement » non linéaire et définie sur $[-10,10]^{10}$ par (Jin *et al.*, 2001) :

$$f(x_1, \dots, x_{10}) = \sum_{j=1}^{10} \exp(x_j) \left\{ c_j + x_j - \ln \left[\sum_{k=1}^{10} \exp(x_k) \right] \right\}$$

avec $(c_1, \dots, c_{10}) = (-6.089; -17.164; -34.054; -5.914; -24.721; -14.986; -24.100; -10.708; -26.662; -22.179)$

La figure suivante représente la coupe²⁷ à $x_i = 3$ pour $i \neq 1, 3$.

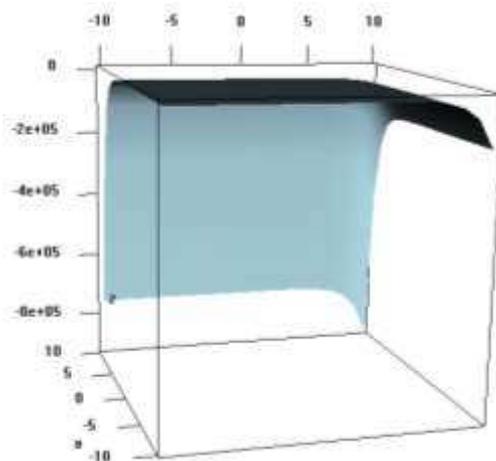


Figure 108. Coupe de la fonction de Jin en fonction de x_1 et x_3 ($x_i = 3$, $i \neq 1, 3$)

²⁷ Cette coupe a été choisie au regard de celle représentée dans l'article de Jin *et al.* 2000

9.1. Les données

Le nombre de points des plans a été fixé volontairement à 100 afin de se trouver dans les conditions les plus probables qui soient des expérimentations numériques.

Voici la liste des plans qui ont été confrontés :

- Plans aléatoires (*Aléatoire*)
- Plans minimax (*Minimax*)
- Plans à entropie maximale (*Dmax*)
- Hypercubes latins usuels (*rLHs*)
- Hypercubes latins maximin (*mLHs*)
- Hypercubes latins améliorés à l'aide d'un critère de distance (*iLHs*)
- Plans de Strauss (*Strauss*)
- Plans de Strauss généralisés (potentiel puissance à $\frac{1}{2}$) (*Strauss_G*)
- Plans de Strauss avec des contraintes 1D (*Strauss1D*)
- Plans de Strauss généralisés avec des contraintes 1D (*Strauss1D_G*)

Les plans testés ici introduisant tous de l'aléa stochastique, l'étude a été conduite sur 100 plans de chaque type afin de donner un sens aux résultats.

Les différents plans sont comparés à l'aide de critères – dont le détail est donné en annexe B – permettant de valider le modèle en ajustement et en prédiction. En pratique, un modèle de surface de réponse de degré 2 calculé à l'aide de la méthode MARS²⁸ (voir annexe A) est ajusté pour chaque plan. Afin de calculer les critères de prédiction du modèle, la réponse est prédite sur un plan test aléatoire de 10000 points.

Le choix d'un modèle MARS de degré 2 pour ce cas d'étude réside dans le fait que ce modèle semble particulièrement adapté à la forme de la réponse et qu'il est rapide à mettre en œuvre en grande dimension. La méthode MARS dépendant plus fortement du plan d'expériences comparativement aux autres méthodes plus classiques telles que les régressions quadratiques ou PolyMARS, nous pourrions probablement tirer plus de conclusions sur la comparaison des plans.

²⁸ MARS de degré 2 revient à faire une régression linéaire par morceaux en considérant les éventuelles interactions.

9.2. Les résultats

Dans cette partie, nous comparons les plans d'expériences entre eux à l'aide du critère R^2 pour mesurer la qualité de l'ajustement et du critère de $RMSE$ pour mesurer la qualité de prédiction. Les résultats concernant les plans introduisant de l'alea stochastique seront présentés sous forme de boxplot, et ceux des plans déterministes sous forme de tableau.

A titre indicatif, les moyennes et écarts-types des différents critères (R^2 , MAE , RMA ,...) sont disponibles en annexe F.

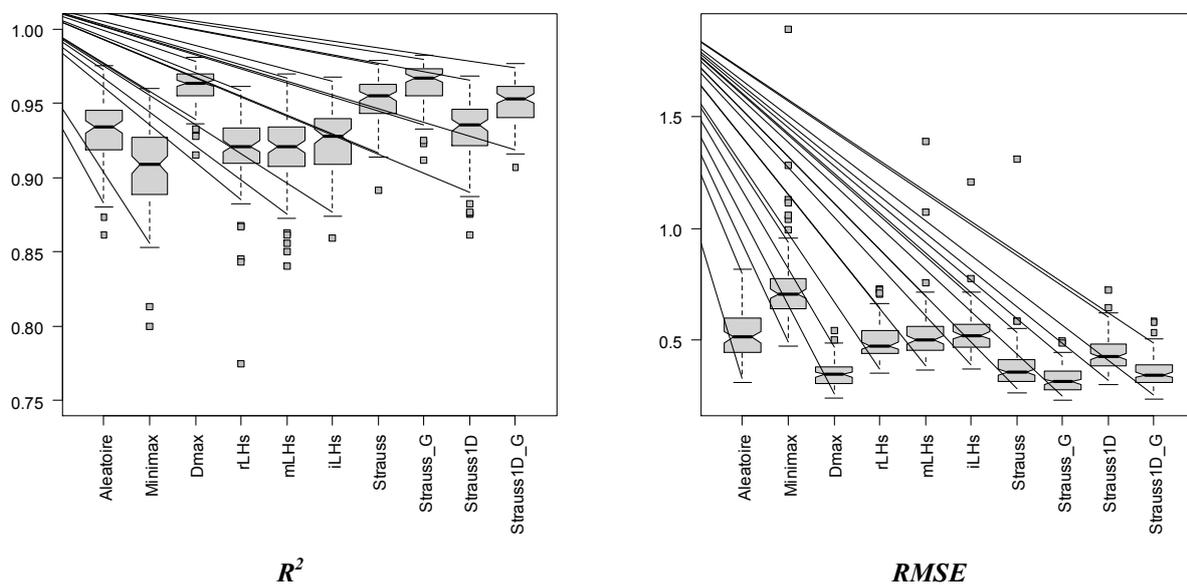


Figure 109. Boxplots des critères de qualité d'ajustement (à gauche) et de validation (à droite) des plans

On voit deux groupes de plans qui se détachent :

- un premier modélisant bien la réponse qui est composé des plans de Strauss sans les contraintes 1D et des plans à entropie maximale,
- un second formé de tous les autres plans fournissant des résultats médiocres.

A noter que les plans minimax ressortent comme étant les plus mauvais que ce soit en prédiction ou en ajustement. Leur construction minimisant la distance maximale entre les points du plan et les points d'une distribution aléatoire, en remplacement d'une grille régulière (inimaginable à générer avec beaucoup de niveaux en dimension 10), en est probablement la raison principale.

Les résultats concernant les suites à faible discrédance sont donnés dans le tableau ci-dessous :

Plans \ Critères	R^2	$RMSE$
Halton	0,93	0,49
Hammersley	0,94	0,52
Faure	0,88	0,63
Sobol	0,88	0,56
Niederreiter	0,92	0,40

Tableau 11. Résultats des critères de qualité d'ajustement (R^2) et de validation ($RMSE$) pour les suites à faible discrédance

La suite de Faure paraît être ici de mauvaise qualité, alors qu'au contraire la suite de Niederreiter permet de relativement bien approcher la réponse.

9.3. Conclusions

Même si ces résultats sont très dépendants du cas étudié, ils montrent que les plans de Strauss (potentiel tout ou rien et potentiel puissance) offrent une meilleure qualité d'ajustement et de validation du modèle que celle obtenue avec les autres plans.

D'autre part, les plans minimax²⁹ ne permettent pas, en grande dimension, de générer de bons SFD.

Nous pouvons également remarquer que les hypercubes latins tant utilisés en expérimentation numérique sont de qualité semblable entre eux et similaire à celle d'un plan aléatoire. Ce qui pourrait confirmer que le fait de considérer les projections en dimension 1 est inenvisageable en grande dimension. Si l'on doit réellement considérer des bonnes projections en 1D, les plans de Strauss avec des contraintes 1D sont légèrement meilleurs que les hypercubes latins améliorés à l'aide du critère de maximin.

Les résultats liés aux plans de Strauss considérant un potentiel généralisé confirment que celui-ci permet de mieux remplir l'espace du fait qu'il prend en compte la force d'interaction entre les points.

Il paraît intéressant de noter que toutes ces conclusions sont similaires avec celles obtenues lors des applications intrinsèques du Chapitre 7.

²⁹ Ces plans ont été générés à l'aide de la routine *cover.design* du package *Fields* de R

Chapitre 10

Application à l'exploration pétrolière

Par le truchement de la résolution d'un problème d'optimisation en 8 dimensions, un certain nombre de techniques étudiées dans le cadre des *Computer Experiments* et du consortium DICE sont mises en œuvre. Ce sont :

- des Space-Filling Designs : notamment les plans de Strauss et les plans à entropie maximale mis en avant dans l'application du cas 10D.
- des surfaces de réponse multiples : modèles additifs, polynômes, réseaux de neurones, et krigeages.

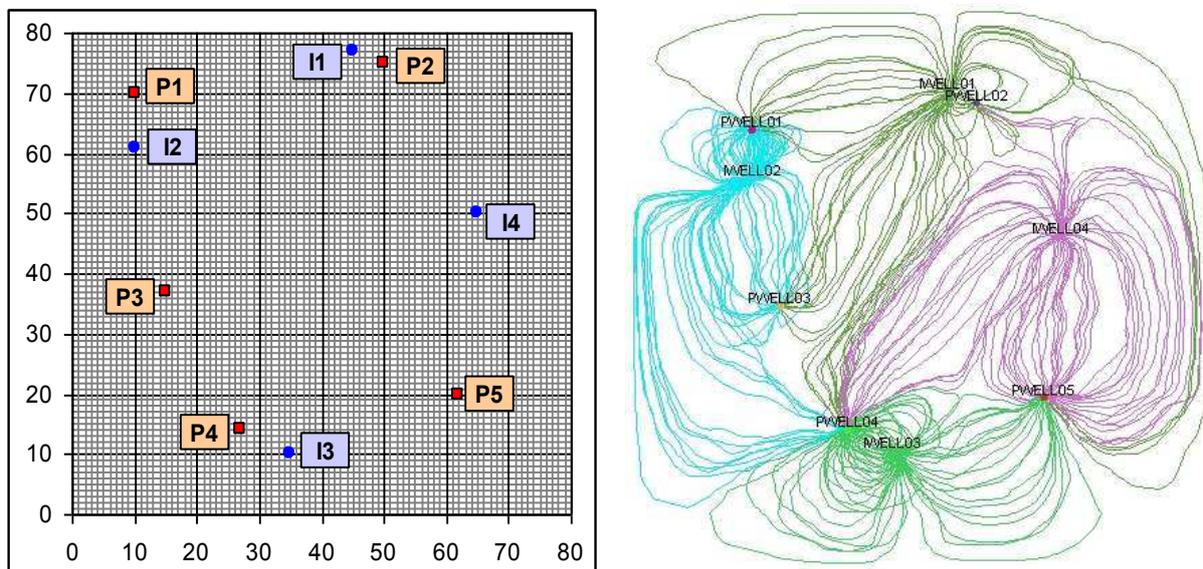
Il s'agit de montrer qu'il est possible de résoudre un problème d'optimisation en l'état actuel des techniques et, par ce biais, de confronter les divers Space-Filling Designs avec différents types de surfaces de réponse.

Ce chapitre se décompose en 5 parties. La première partie présente l'application en décrivant son objectif, les paramètres d'entrée sélectionnés, etc. Nous détaillons ensuite une étude sur un plan de Strauss à 80 points avec différents modèles de surfaces de réponse. La section 3 étudie l'impact du type du plan d'apprentissage sur le modèle obtenu. La section 4 présente les résultats obtenus en faisant varier le nombre de points du plan d'expériences. Enfin, la dernière section résume les résultats obtenus et apporte une réponse au problème d'optimisation.

10.1. Présentation du cas d'étude

10.1.1. Contexte de l'étude

Cette application s'intéresse à un réservoir pétrolier décrit spatialement en 3 dimensions et discrétisé en $80 \times 80 \times 6$ mailles. Le réservoir étudié est constitué de 9 puits : 5 producteurs et 4 injecteurs. Les données sont obtenues à l'aide du simulateur d'écoulement en milieu poreux 3DSL[®] de Streamsim Co (basé sur la technique des lignes de courant). Les réponses fournies par celui-ci sont multiples et temporelles. La variable d'intérêt Y considérée ici est la production cumulée d'huile du réservoir à l'horizon temporel de 7 000 jours.



10.1.2. Description des données

L'objectif final de cette étude consiste à optimiser la production d'huile en fonction de variables de contrôle représentant la position de certains puits : 2 puits producteurs (notés P2 et P5) et 2 puits injecteurs (notés I2 et I3). Les puits étant renseignés uniquement par leurs abscisses (*abs*) et ordonnées (*ord*), l'espace à explorer est de dimension $d = 8$. Les coordonnées des puits sont définies sur une grille et par conséquent sont des nombres entiers.

Les domaines de variation définis autour de la position initiale des puits (encore appelée configuration de base) sont représentés sur la figure suivante.

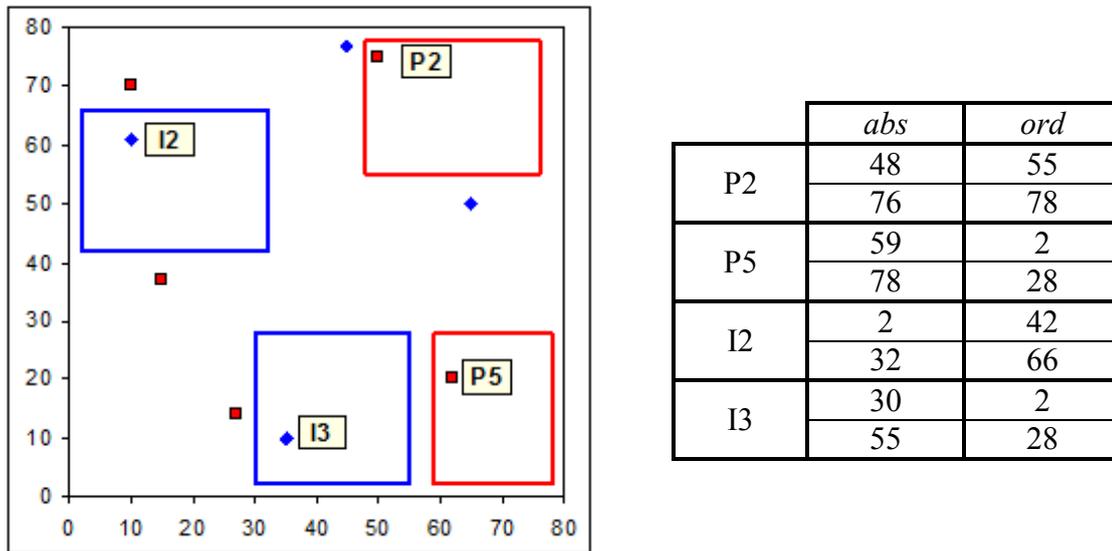


Figure 111. Domaines de variation pour la position des puits P2, P5, I2 et I3

10.1.3. Schéma d'une étude

La méthodologie utilisée pour traiter ce cas d'application est similaire à celle employée dans les applications précédentes et décrite en introduction. On rappelle cependant brièvement les différentes étapes réalisées :

1. Génération du plan d'expériences
2. Lancement des simulations sur 3DSL[®]
3. Etude préliminaire des données
4. Ajustement des surfaces de réponse et analyse des résultats
5. Etude comparative des différents modèles obtenus
6. Résolution du problème.

Les étapes 1 et 2 permettent d'obtenir les données à partir desquelles l'étude est construite. Les résultats présentés ici concernent les étapes 3 à 5. La qualité de prédiction est étudiée avec des points tests issus d'un plan de Strauss à 1000 points.

10.2. Etude détaillée d'un jeu de données

Dans cette section, nous détaillons les résultats obtenus avec un plan de Strauss à 80 expériences. Chaque surface de réponse est analysée en regardant ses propriétés d'ajustement aux données du plan ainsi que ses qualités prédictives à l'aide des critères détaillés en annexe B et calculés sur le plan test.

10.2.1. Etude préliminaire des données simulées

Avant de construire une surface de réponse, il est important d'étudier brièvement le jeu de données afin de détecter d'éventuelles anomalies ou un phénomène inattendu.

On rappelle que l'objectif final de l'application est de maximiser la réponse au temps $t = 7000$ jours en fonction de l'emplacement des 4 puits. L'évolution temporelle de la production d'huile est représentée ci-dessous pour différentes configurations du réservoir. Les courbes correspondent à celles des points fournissant les 3 meilleures productions, à la configuration de base, ainsi qu'aux trois simulations pour lesquelles la production simulée a été la plus faible.

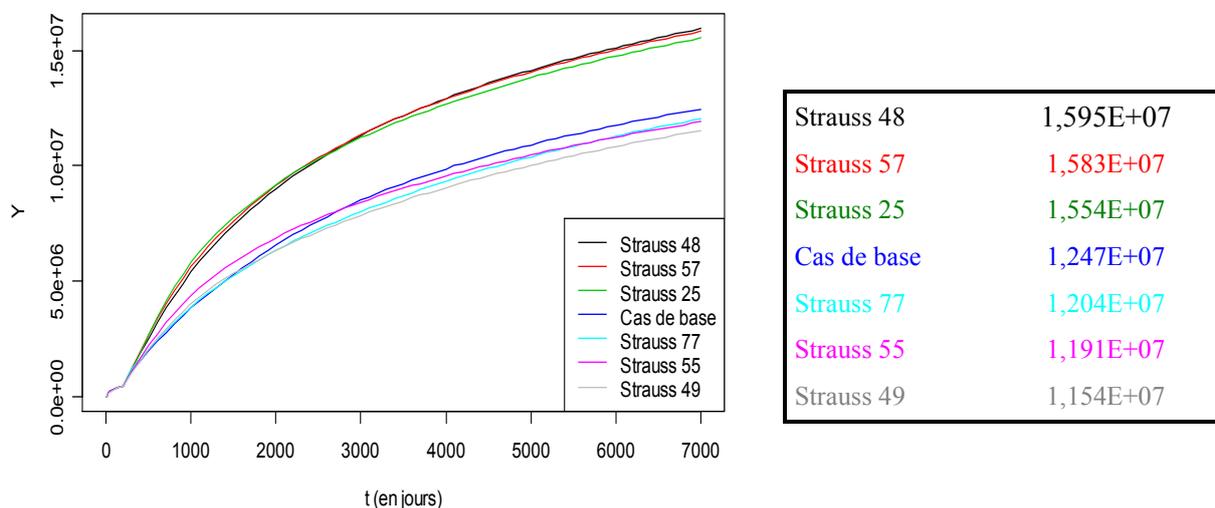


Figure 112. Production cumulée d'huile pour les différentes configurations

La Figure 112 montre que le cas de base ne correspond pas à une configuration idéale du point de vue de l'objectif poursuivi. En effet, le cas de base est proche des simulations les plus faibles. L'ordre de grandeur de la réponse étant élevé ($\sim 10^7$), il est usuel de centrer et réduire les données.

La Figure 113 représente les positions des puits correspondant aux 20 meilleures configurations parmi les 80 proposées par le plan de Strauss.

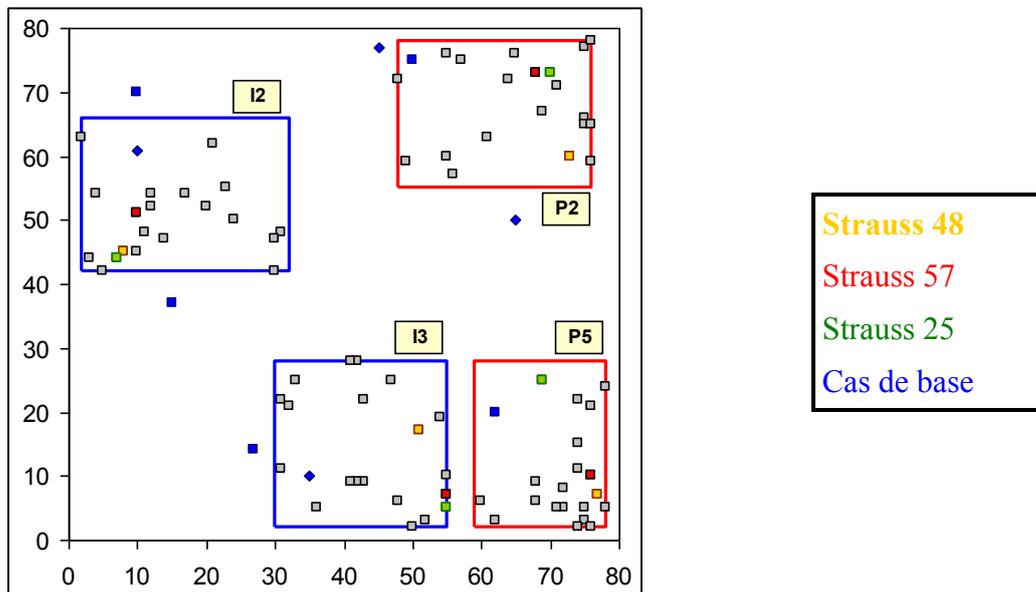


Figure 113. Représentation des 20 meilleures configurations parmi les 80 initialement simulées ainsi que du cas de base

Les nuages de points observés sont difficiles à interpréter car ils sont représentés en projection dans chaque zone de variation. Cependant, la répartition des points à l'intérieur de la zone de variation de certains puits montre des régions semblant favorables à une forte production. Par exemple, on note une zone d'intérêt se situant au quart Sud-Est pour le puits P5. Venant de nous assurer que l'objectif de l'étude a un sens, on complète la phase d'exploration des données en détaillant les valeurs simulées servant de référence à notre étude.

Les figures suivantes représentent la distribution de la réponse. On observe que le jeu de données est assez bien réparti avec une faible queue de distribution pour les productions les plus faibles.

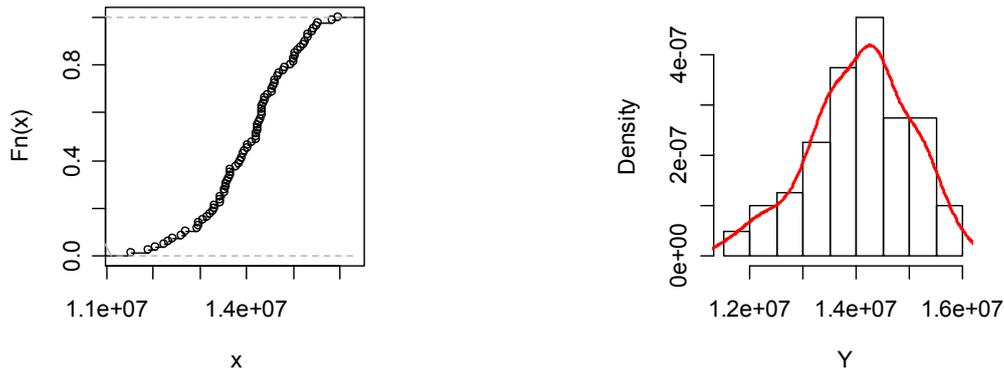


Figure 114. Fonction de répartition empirique (à gauche) et histogramme et lissage par noyau de la réponse (à droite).

La réponse en fonction de chaque variable est représentée en projection sur la figure suivante qui confirme la zone d'intérêt déjà observée du puits P5.

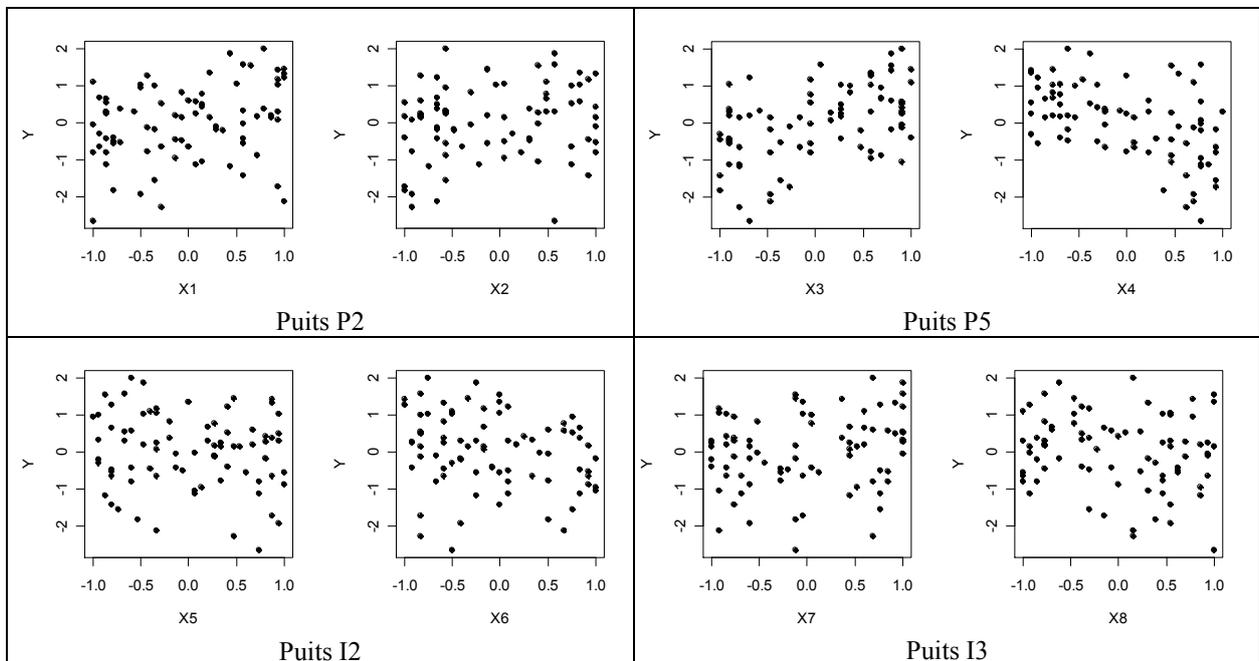


Figure 115. Représentation de la réponse en projection

10.2.2. Construction des différents modèles et interprétation

Plusieurs types de modèles (voir annexe A) ont été ajustés à ces données, cependant seuls les modèles les plus pertinents sont analysés. Par exemple, les modèles obtenus en appliquant une procédure stepwise³⁰ sont proposés bien que le modèle complet ait été étudié auparavant.

³⁰ backward

10.2.2.1. Les modèles additifs

Appliquer un modèle additif permet bien souvent de se faire une première idée sur les variations de la réponse dans chaque direction car le modèle ajusté sera non paramétrique et facile à interpréter grâce à ses propriétés d'additivité. Même si l'hypothèse d'additivité est difficilement vérifiable en dimension 8, les résultats sont intéressants. Une procédure stepwise est appliquée en vue d'aboutir à un modèle moins complexe.

Remarque. Si la relation entre les variables comporte des interactions importantes qui ne sont pas prises en compte dans le modèle, alors celui-ci conduit à des résultats erronés.

Le modèle additif complet est de la forme :

$$Y \approx \alpha + \beta_1 s(X_1) + \dots + \beta_8 s(X_8)$$

où le détail des composantes est donné dans la Table 1 de l'annexe G.

La table d'analyse de la déviance relative au modèle construit en estimant des splines dans chaque direction montre que seules les fonctions $s(X_7)$ et $s(X_8)$ ont une influence sur la réponse (p-valeur < 0.10). Les 8 composantes monodimensionnelles estimées et les bandes de confiance associées aux courbes prédites sont représentées ci-dessous.

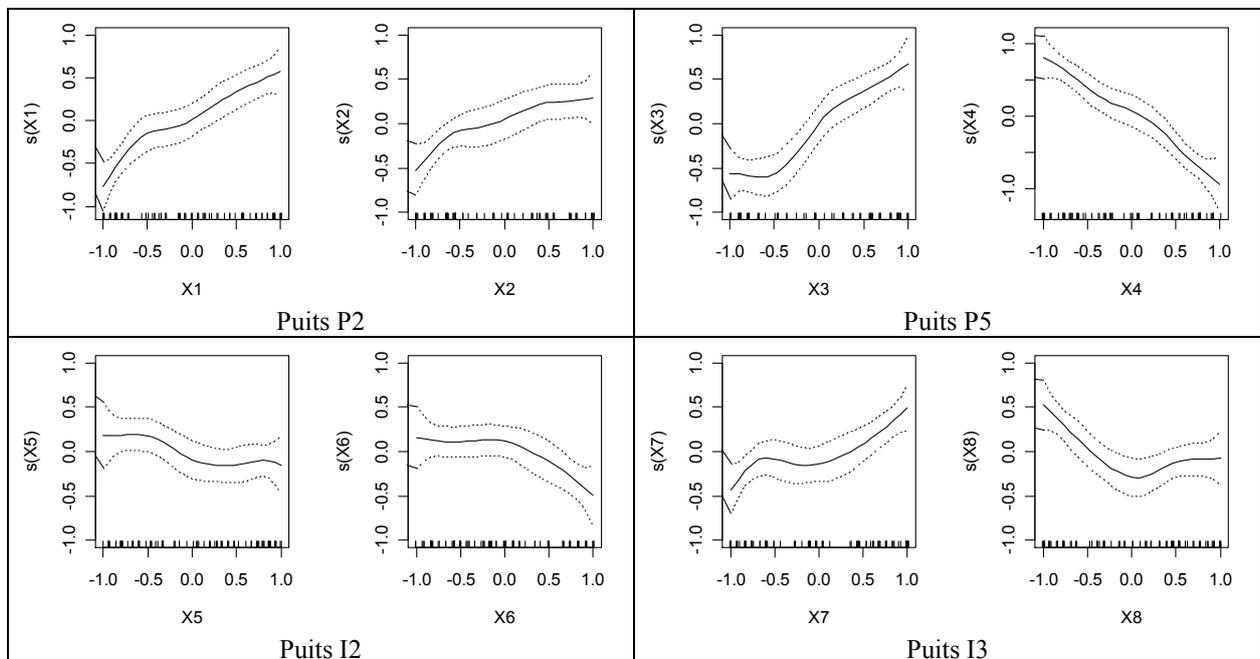


Figure 116. Représentation des différentes composantes du modèle additif ajusté

Le modèle fournissant les résultats présentés à la Figure 116 est assez complexe car il compte 32 degrés de liberté. Le but de la procédure stepwise est de proposer un modèle

simplifié en choisissant de mettre chaque variable ou non. Avec ce type de procédure, on obtient un modèle mixte (mélangeant des splines et des fonctions linéaires) à 17 degrés de liberté avec des composantes majoritairement linéaires à l'exception des directions X3, X7 et X8 (voir Table 2 en annexe G).

La Figure 117 représente les valeurs ajustées en fonction des valeurs réelles, sur le plan d'apprentissage et sur le plan test. Les nuages de points indiquent une forte corrélation entre la prévision et la valeur réelle.

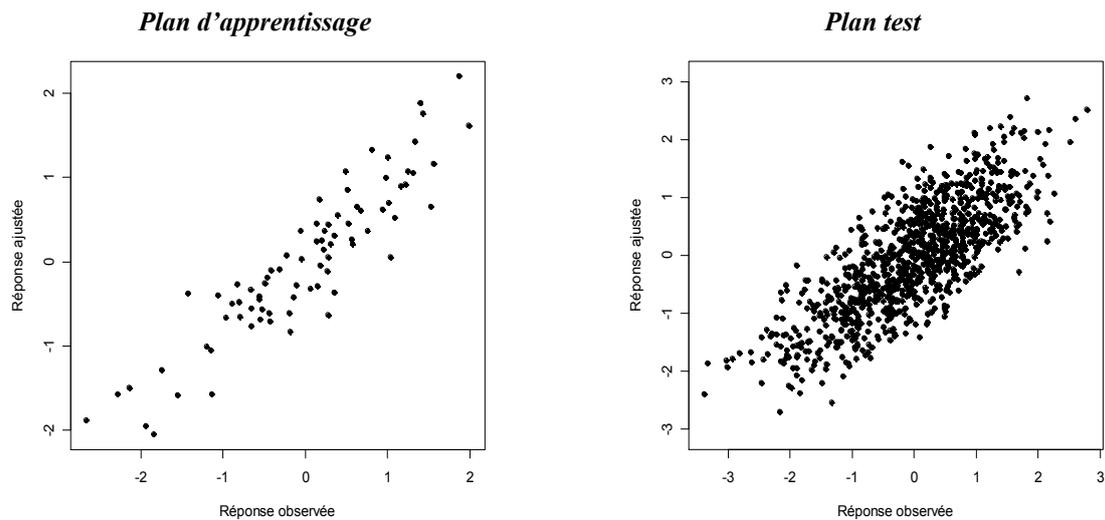


Figure 117. Étude des valeurs prédites du modèle additif mixte

Les résultats semblent valider l'hypothèse d'additivité. Une étude plus approfondie des résidus confirme l'hypothèse d'homoscédasticité ainsi que l'absence de lien entre les variables explicatives et les résidus.

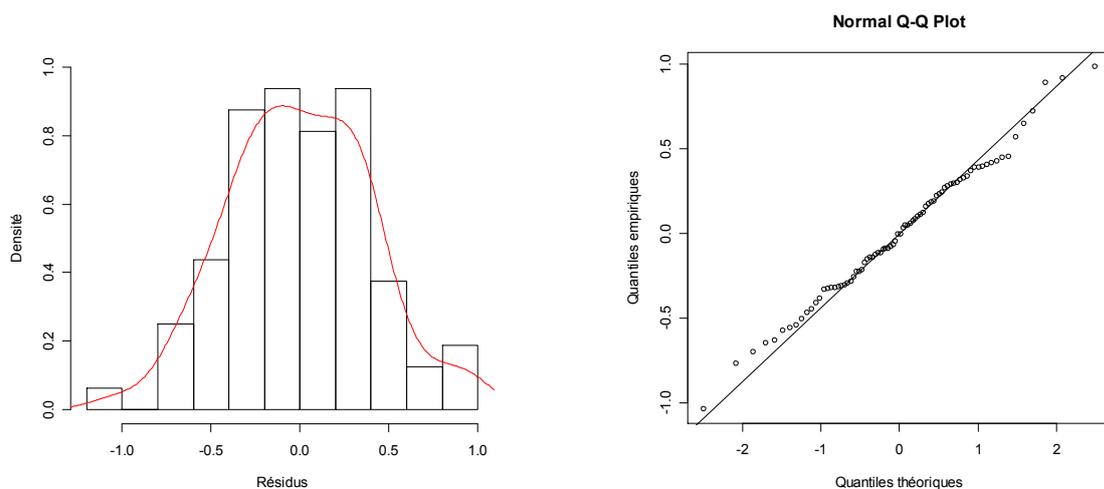


Figure 118. Étude de la normalité des résidus du modèle additif mixte

10.2.2.2. Les modèles linéaires

Nous cherchons à construire le modèle expliquant au mieux les données sans pour autant le surcharger. Le modèle linéaire à effets simples, bien que très minimaliste, permet cependant d'expliquer près de 76% de la variabilité de la réponse. La table d'analyse de la variance correspondante est détaillée en annexe G (Table 3).

On constate que toutes les variables sont significatives et que les plus influentes sont les facteurs correspondant aux coordonnées des puits producteurs P2 et P5.

Les étapes suivantes consistent à complexifier le modèle pour expliquer au mieux les données. Dans un premier temps, on ajoute des interactions entre les termes puis on augmente le degré du polynôme. Afin de ne pas construire un modèle trop proche des données (et qui sera sans doute instable en prédiction), on applique une procédure stepwise pour conserver seulement les termes influents. Le Tableau 12 récapitule les caractéristiques des différents modèles étudiés.

	Linéaire	Linéaire avec interactions	Linéaire avec interactions (stepwise)	Degré 2 (stepwise)	Degré 3 (stepwise)
Ajustement					
R^2	0,76	0,90	0,89	0,90	<u>0,94</u>
R^2_{adj}	0,74	0,82	0,85	0,86	<u>0,89</u>
ddl pour le modèle	8	36	21	24	34
Qualité de prédiction					
$RMSE$	0,63	0,63	<u>0,62</u>	0,64	0,71
RMA	<u>1,98</u>	2,19	2,27	2,37	2,68

Tableau 12. Critères de validation pour les modèles de régression

Le modèle stepwise construit à partir des effets simples et des interactions entre deux facteurs contient 21 termes dont les 8 termes linéaires (voir Tableau 18 en annexe G).

Ce modèle ajuste mieux les données du plan d'expériences ($R^2 = 89\%$) mais les critères évalués sur le plan test montrent que ce modèle a des qualités similaires au modèle linéaire en prédiction. Le modèle retenu pour ce cas d'étude est donc le modèle linéaire à effets simples qui présente l'avantage d'avoir un nombre de degrés de liberté plus faible ainsi que la RMA la meilleure.

Remarque. Dans le Tableau 12, les résultats sont donnés jusqu'à la régression de degré 3. En pratique, on se limite aux effets simples et aux interactions mais ici nous avons souhaité illustrer le compromis biais-variance.

La Figure 119 présente l'évolution du R^2 sur le plan d'apprentissage (noté R_{app}^2), et du R^2 sur le plan test (noté R_{pred}^2) en fonction du nombre de termes dans le modèle. On remarque que l'ajustement s'améliore avec la complexité (croissance de R_{app}^2) tandis que la prévision passe par une zone optimale quand le modèle est de complexité réduite (croissance puis décroissance de R_{pred}^2).

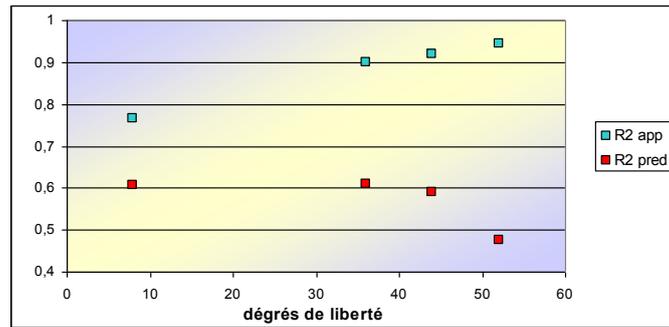


Figure 119. Illustration du compromis biais-variance (Modèle de régression linéaire en augmentant le degré)

10.2.2.3. MARS

Le modèle MARS de degré 1 est composé de 9 fonctions de base linéaires par morceaux. Par construction, ce modèle ne contient pas de termes d'interactions (voir Tableau 19 en annexe G). La figure suivante représente les courbes estimées en faisant varier un facteur à la fois et en fixant les autres à 0.

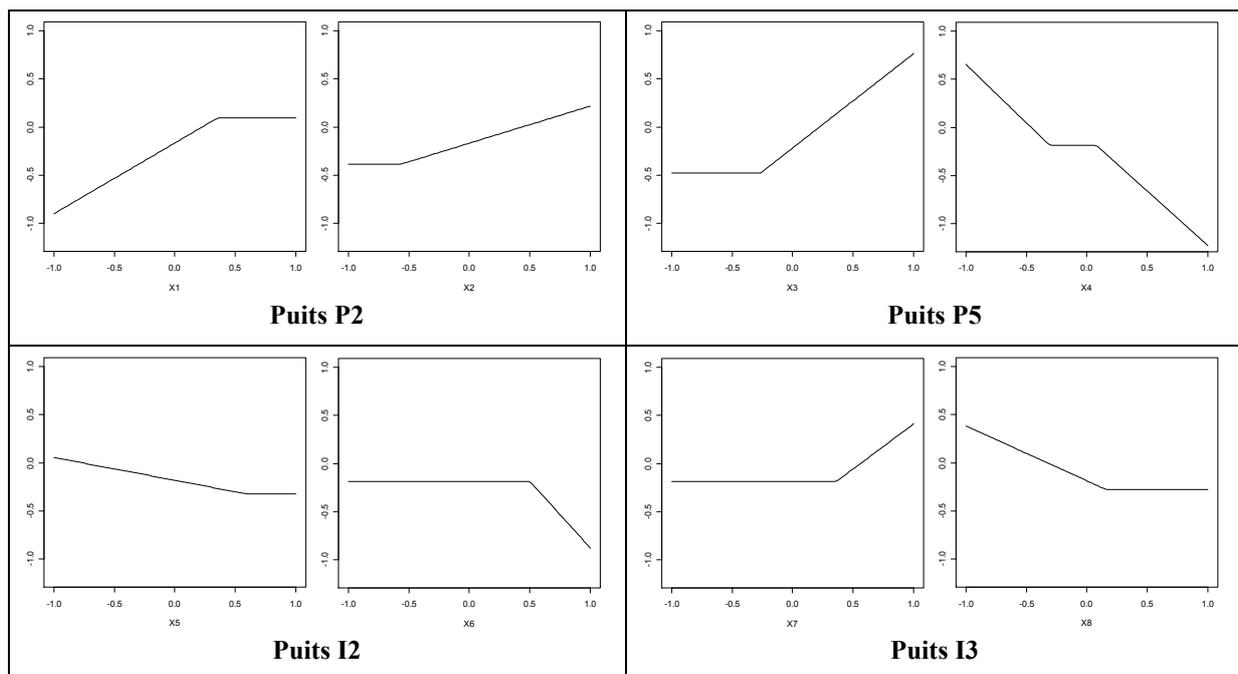


Figure 120. Représentation des fonctions de base du modèle MARS de degré 1

En comparaison avec le modèle additif (voir Figure 116), les composantes de MARS semblent cohérentes et les sens de variation sont respectés.

Le modèle de degré 2 n'est pas retenu car il est composé de 8 fonctions dont 4 interactions. Les critères montrent que le modèle est légèrement meilleur sur les données d'apprentissage mais que le modèle de degré 1 se comporte mieux en prédiction. Les termes de ce modèle ne sont pas précisés ici mais on constate des interactions entre la direction verticale du puits P5 (variable X_4) et les deux coordonnées du puits I3. Or, on peut remarquer sur la Figure 111 que ces deux puits sont situés à proximité.

10.2.2.4. PolyMARS

Les résultats du Tableau 13 permettent d'étudier l'influence du paramètre de pénalisation et les qualités des différents modèles. Le modèle sélectionné est celui ajusté avec $a=3$ car il correspond à un bon compromis entre sa qualité en prédiction et le nombre de fonctions de base égal 7.

	PolyMARS					
	a=0-1	a=2	a=2,7	a=3	a=4	a=5
R^2	0,90	0,85	0,80	0,74	0,68	0,56
Nombre de fonctions dans le modèle	19	12	9	7	5	3
Nombre d'interactions	5	1	1	--	--	--
MAE	0,51	0,50	0,47	0,51	0,54	0,57
$RMSE$	0,64	0,63	0,60	0,64	0,68	0,71
RMA	2,16	2,05	2,18	1,99	2,32	2,19

Tableau 13. Comparaison des modèles PolyMARS ajustés avec différentes valeurs du paramètre de pénalisation a .

Ce tableau montre aussi que le modèle le plus simple contient seulement 3 termes – les fonctions linéaires X_4 , X_3 et X_1 – et permet déjà d'expliquer plus de 50% des variations de la réponse sur le plan d'apprentissage ainsi que sur le plan test. La méthode PolyMARS semble donc particulièrement adaptée à ce jeu de données. Par construction, il faut que les termes linéaires soient considérés avant de pouvoir intégrer des termes linéaires par morceaux et/ou des interactions. Au vu des résultats des modèles précédents, on sait que les termes linéaires jouent un rôle important.

10.2.2.5. Krigeage

Le modèle d'interpolation dans sa forme générale est

$$Y(x) = \sum_{j=0}^d \beta_j f_j(x) + Z(x)$$

où $f_1(\cdot), \dots$ sont des fonctions de base connues, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^T$ un vecteur de coefficients estimés et Z un processus gaussien stationnaire centré et de forme de covariance factorisée gaussienne.

Ce type de modèle est référé sous le nom de krigeage universel. Il est aussi possible de considérer un krigeage ordinaire de la forme $Y(x) = \mu + Z(x)$ où μ est une constante à estimer. Cette forme de krigeage est généralement la plus utilisée en pratique.

Le krigeage ordinaire. Il comporte 10 paramètres: la constante de krigeage, la variance et les portées dans les 8 directions principales qui sont estimés par maximum de vraisemblance³¹. Le modèle de krigeage ordinaire obtenu est donné Tableau 20 en annexe G. Les valeurs estimées de la moyenne (-0,05) et de la variance (1.40) semblent cohérentes compte tenu que les données sont centrées et réduites. Les portées estimées sont comprises entre 1,39 et 4,39 ce qui confirme que la surface approchée est assez lisse. La corrélation entre deux points distants de 1 est forte (0,60) dans la direction où la portée estimée vaut 1,39 et très forte (0,95) dans la direction où $\theta_{est} = 4,39$. Ceci indique une tendance très plate dans la direction X_2 .

Le krigeage universel avec tendance linéaire. La tendance imposée va permettre d'expliquer les grandes variations de la réponse, le processus gaussien quant à lui expliquera la partie résiduelle c'est-à-dire les variations à plus courte distance. Après plusieurs essais, des problèmes de convergence sont apparus lors de l'optimisation de la vraisemblance (portées différentes). Il semblerait que le profil de vraisemblance soit assez plat. Cependant, les modèles obtenus ont des qualités prédictives similaires.

Deux exemples de modèles différents sont donnés en annexe G (Tableau 21 et Tableau 22).

Une portée égale à 10 signifie que la tendance apprend toutes les variations dans cette direction (car les données ont été centrées-réduites) : il ne reste plus assez d'information pour

³¹ Les bornes imposées pour les paramètres d'échelle sont 10^{-3} et 10. La procédure d'estimation des paramètres de covariance est itérée plusieurs fois afin de s'assurer de la bonne convergence.

le bruit. On devrait donc soit enlever la tendance dans cette direction soit modéliser le reste par un bruit blanc.

		KRIGEAGE ORDINAIRE	KRIGEAGE UNIVERSEL
Qualité de prédiction sur le plan test			
RMSE	↘	0,65	0,66
RMA	↘	2,09	2,49

Tableau 14. Etude des qualités de prédiction du modèle de krigeage ordinaire

Conclusions. Les résultats précédents montrent que le krigeage ordinaire *i.e.* sans tendance estime un modèle avec des variations à grandes échelles. L'ajout d'une tendance n'apporte rien au modèle, on privilégie alors le krigeage ordinaire qui est plus simple et plus stable (*RMA* la plus faible).

10.2.2.6. Réseaux de neurones

On obtient différents modèles selon la paramétrisation (nombre de neurones, decay, modèle sous-jacent) de la méthode. Quelques exemples sont donnés dans le Tableau 14 et nous indiquent qu'il est préférable d'utiliser un réseau de neurones basé sur un modèle sous-jacent linéaire simple avec 3 neurones et le paramètre de régularisation decay égal à 1. Dans ce cas, le modèle obtenu contient 31 coefficients et est très bon en prédiction (voir Tableau 16).

Nombre de neurones	2	3	3	3	3	4, 5, 6 et 7	2
decay	1	1	0.1	0.01	2	1	1
Hessien	T	T	T	T	F	T	T
Modèle sous-jacent	Lin.	Lin.	Lin.	Lin.	Lin.	Lin.	Lin.Int
R^2	0,76	0,76	0,88	0,90	0,75	0,76	0,88
Nombre de poids	21	31	31	31	31	41	77
<i>MAE</i>	0,49	0,49	0,49	0,56	0,49	0,49	0,48
<i>RMSE</i>	0,62	0,62	0,62	0,70	0,62	0,62	0,61
<i>RMA</i>	2,02	2,01	2,02	2,41	2,01	2,00	2,21

Tableau 15. Comparaison des modèles de réseaux de neurones avec différentes paramétrisations.

Les 8 composantes mono-dimensionnelles estimées sont représentées ci-dessous

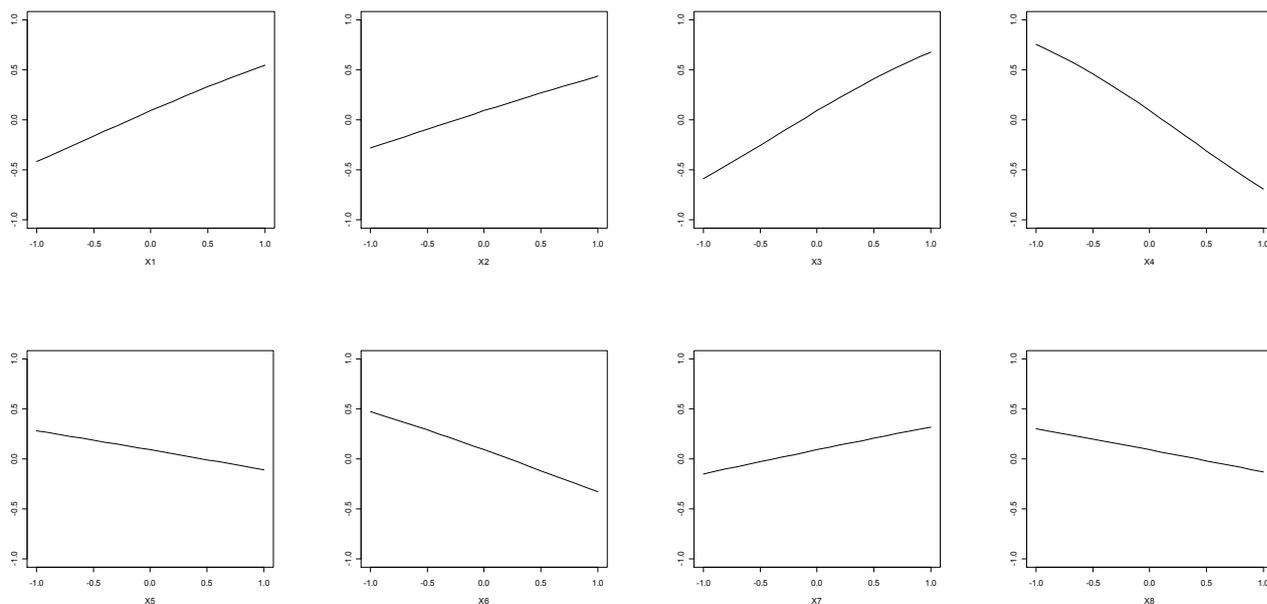


Figure 121. Représentation des différentes composantes du modèle ajusté par les réseaux de neurones

Les figures ci-dessus confirment les tendances observées avec le modèle additif et le modèle MARS.

10.2.3. Comparaison des modèles ajustés

L'étude des différents modèles a montré que les variations de la réponse en fonction des différents facteurs sont majoritairement linéaires. La plupart des modèles retenus contiennent effectivement les effets principaux. Le tableau suivant récapitule les critères évalués pour chaque modèle.

		Régression	Modèle additif	MARS	PolyMARS	Krigeage	Réseau de neurones
		linéaire	mixte	degré 1	a = 3	ordinaire	linéaire
Ajustement							
R^2	~1	0,76	0,84	0,78	0,74	--	0,75
R^2_{adj}	~1	0,74	0,80	0,67	--	--	--
DDL modèle ³²	--	8	17	9	7	--	31
Prédiction							
RMSE	↘	0,63	0,64	0,65	0,64	0,65	0,62
RMA	↘	1,98	1,98	1,94	1,99	2,09	2,01

Tableau 16. Comparaison des modèles ajustés avec le plan de Strauss à 80 points

³² Pour les méthodes MARS et PolyMARS, le nombre de degrés de liberté indiqué correspond au nombre de fonctions de base du modèle ajusté.

Critères en apprentissage. Le modèle additif mixte permet d'obtenir des valeurs ajustées proches de la réalité sur le plan d'expériences. Cependant, ce modèle figure parmi les modèles les plus complexes : 17 degrés de liberté contre moins de 10 pour la plupart des autres modèles. Les modèles proposés sont en moyenne assez bons, même les plus simples tels que la régression linéaire, car ils expliquent au moins 74% des variations totales de la réponse.

Critères en prédiction sur l'ensemble test. Le Tableau 16 montre que les modèles retenus sont souvent les plus simples tels la régression linéaire ou PolyMARS. Le coefficient *RMSE* varie très peu d'un modèle à l'autre ($\sim 0,63$).

Conclusions. Pour ce cas d'étude, le modèle linéaire et PolyMARS semblent les plus adaptés. En effet, les réseaux de neurones, offrant ici les meilleures qualités en prédiction, fournissent des modèles trop complexes par rapport à ces 2 méthodes qui auront donc tendance à être privilégiées.

10.3. Etude de l'impact du plan d'expériences

Cette section est consacrée à une étude comparative entre les plans d'expériences à 80 points et les métamodèles.

Les méthodes de surfaces de réponse étudiées sont les suivantes :

- régression linéaire à effets simples avec ou sans interactions,
- modèle additif avec des splines,
- MARS de degré 1,
- PolyMARS (avec un paramètre de pénalisation $\alpha=3$),
- réseaux de neurones (nombre de neurones=3 et decay=1),
- krigeage ordinaire.

Les résultats de chaque méthode seront détaillés afin de proposer une étude comparative des plans ainsi que des métamodèles pour confronter les conclusions à celles de la section précédente.

Les plans d'expériences à partir desquels les différents modèles ont été ajustés sont détaillés dans le tableau ci-dessous :

Plans	Nombre de plans
Suite de Halton	1
Suite de Faure	1
Suite de Sobol'	1
Plans aléatoires	10
Plans à entropie maximale (Dmax)	10
Plans minimax	10
Plans de Strauss	10

Tableau 17. Liste des plans d'expériences utilisés dans cette étude

Qualité intrinsèque des plans. La figure ci-dessous représente la moyenne et l'écart-type des longueurs des arêtes de l'Arbre de Longueur Minimum (voir Chapitre 6) associé à chacun des plans.

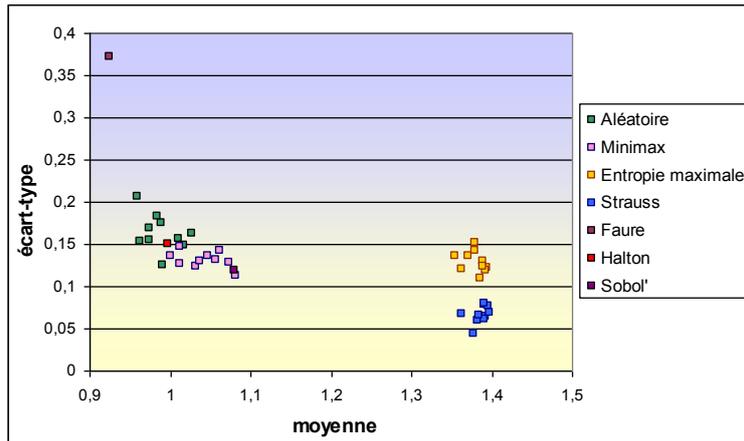


Figure 122. Représentation du critère basé sur les ALM pour chacun des plans de l'étude

On remarque que les plans minimax, Halton et Sobol' sont de qualité équivalente à ceux des plans aléatoires. Au vu de la position de la suite de Faure dans le plan (μ, σ) , sa distribution s'apparente à une de type gradient. Les plans de Strauss et les plans à entropie maximale sont de qualité similaire et semblent bien remplir l'espace (moyenne élevée et écart-type faible).

Nous verrons par la suite si les propriétés intrinsèques des plans influent particulièrement sur la qualité des métamodèles.

10.3.1. Les modèles additifs

Les modèles considérés sont composés de splines dans les 8 directions principales.

Apprentissage.

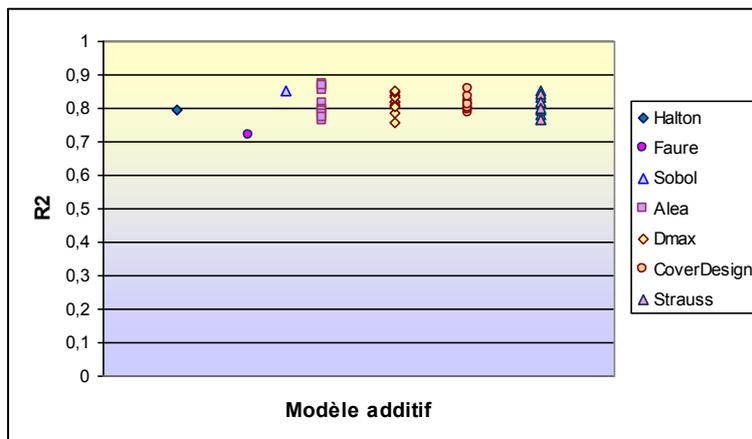


Figure 123. Ajustement des modèles additifs associés à différents plans d'expériences

Les qualités d'ajustement du modèle sur les données d'apprentissage varient peu d'un plan à l'autre. Pour les plans non déterministes, on constate que la dispersion du critère est assez faible.

Le résultat médiocre qu'offre la suite de Faure peut probablement s'expliquer par le fait que le plan a de mauvaises projections dans chacune des 8 directions.

Prédiction.

En termes de prédiction, le critère *RMSE* varie entre 0,6 et 0,8. On observe également peu de différences entre les plans.

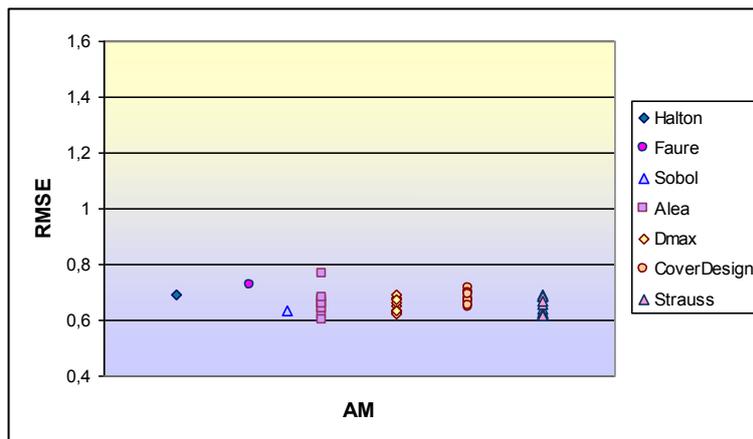


Figure 124. Prédiction des modèles additifs associés à différents plans d'expériences

10.3.2. Les modèles linéaires

Les modèles présentés ici correspondent à une régression à effets simples et à un modèle avec des termes d'interactions deux à deux. Les termes au carré n'ont pas été considérés car ils n'apportent rien de plus au modèle (voir Tableau 12).

Apprentissage.

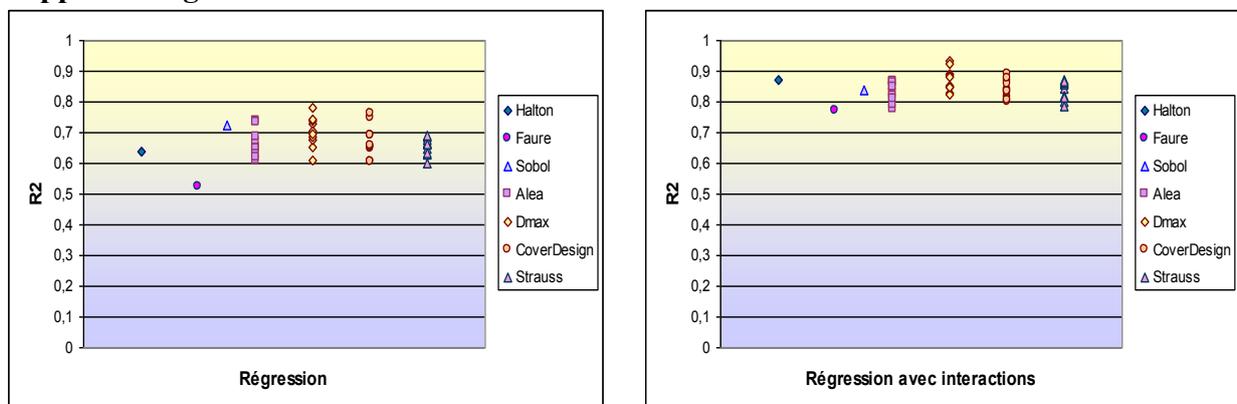


Figure 125. Ajustement des modèles linéaires associés à différents plans d'expériences

Concernant le modèle à effets simples, on observe sur la figure de gauche, une dispersion plus importante des valeurs du critère R^2 pour les plans à entropie maximale (Dmax) et les plans minimax (CoverDesign). A noter que les plans de Strauss offrent une certaine robustesse (faible dispersion) aux critères d'ajustement comparativement à celle des autres méthodes de planification d'expériences. Le modèle avec interactions est plus proche des données d'apprentissage. Ce phénomène a déjà été observé à la section précédente.

Prédiction.

Comme le montre les figures ci-dessous, le pouvoir prédictif du modèle avec interactions est généralement de moins bonne qualité que celui du modèle à effets simples.

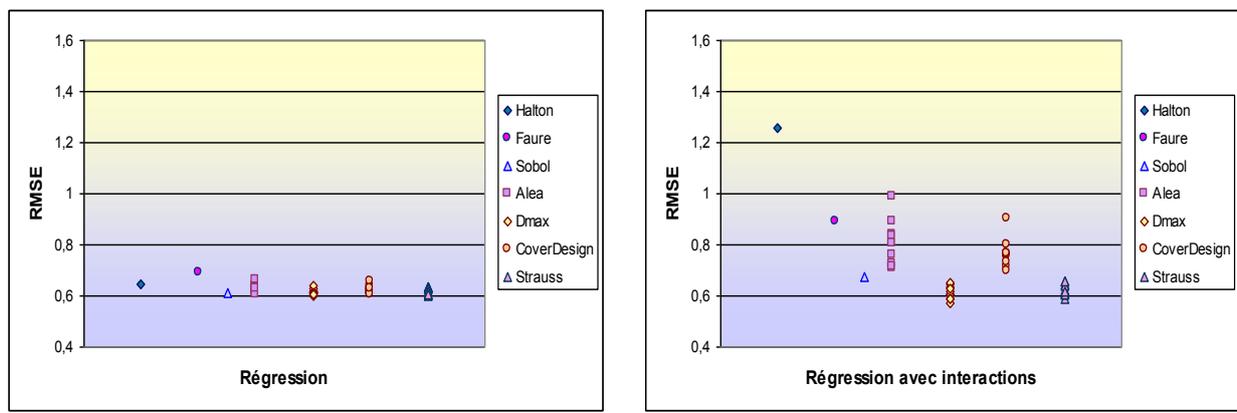


Figure 126. Prédiction des modèles linéaires associés à différents plans d'expériences

Les plans de Strauss et les plans à entropie maximale fournissent des modèles qui sont meilleurs de plusieurs points de vue : d'une part, les valeurs de la $RMSE$ sont les plus faibles et d'autre part les valeurs calculées varient peu d'un plan à l'autre.

Si on applique une méthode stepwise sur le modèle linéaire avec interactions, les résultats sont similaires à ceux de la régression avec interactions :

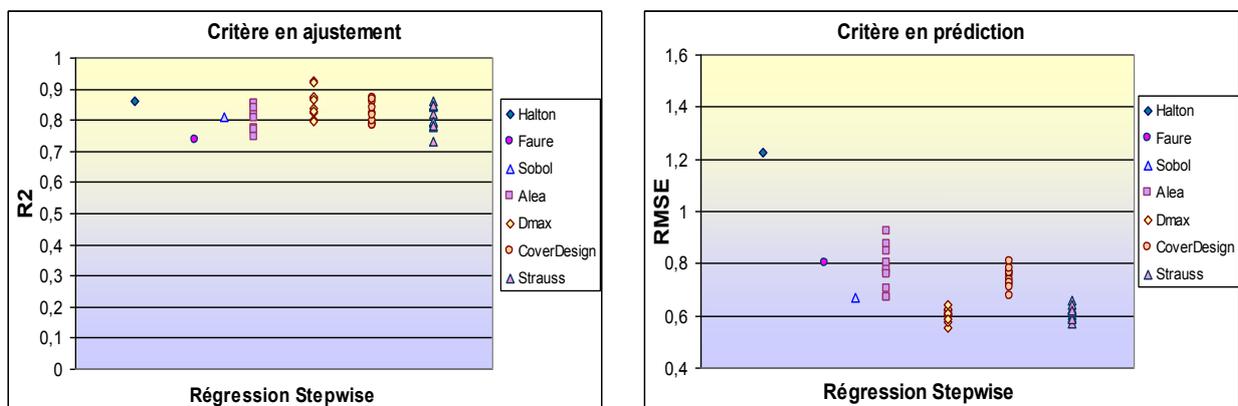


Figure 127. Résultats pour le modèle stepwise

Au vu de ces résultats, on privilégiera la régression linéaire simple qui donne de meilleurs résultats avec une dispersion plus faible.

10.3.3. MARS

On rappelle que les points du plan d'expériences interviennent au cours de la construction d'un modèle MARS dans la mesure où les projections des points sont utilisées en tant que nœuds potentiels pour les fonctions de base.

Apprentissage.

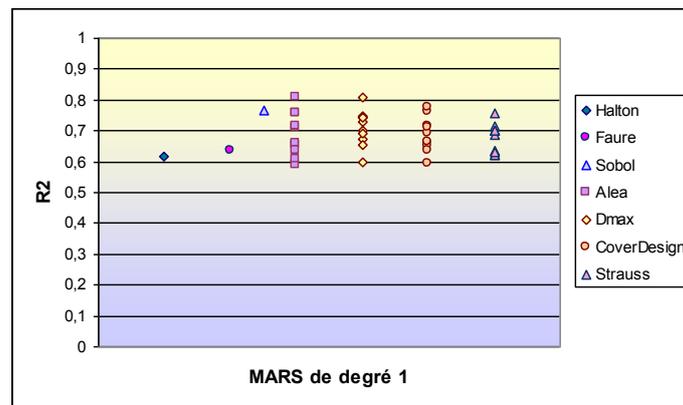


Figure 128. Ajustement des modèles MARS associés à différents plans d'expériences

En termes d'ajustement, on constate que les modèles expliquent entre 60 et 80% de la variabilité totale de la réponse. La dispersion observée pour les différents plans est légèrement plus importante que celle du modèle linéaire à effets simples (voir Figure 125 de gauche).

Prédiction.

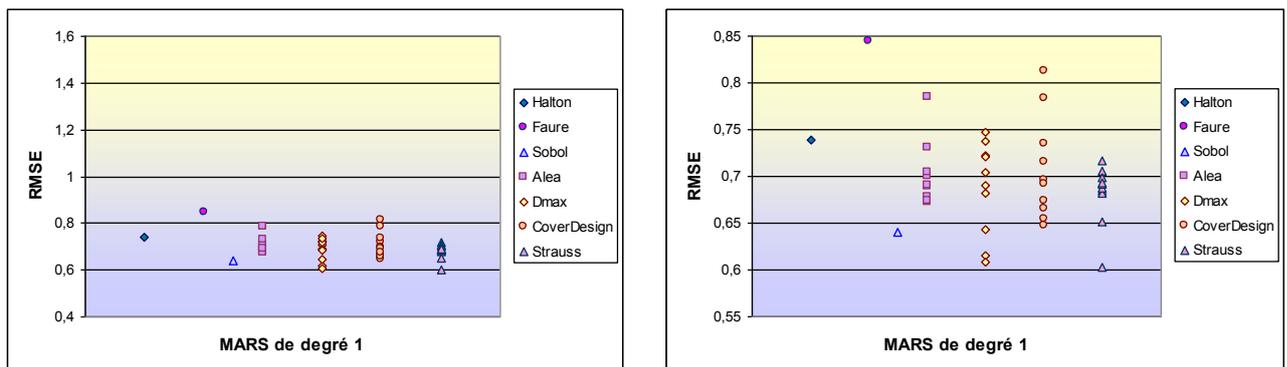


Figure 129. Prédiction des modèles MARS associés à différents plans d'expériences

La Figure 129 de droite représente les valeurs de la *RMSE* avec un effet de zoom afin de mieux observer les sensibilités des résultats aux plans. Le pouvoir prédictif des modèles MARS est assez satisfaisant. Pour compléter cette étude la Figure 130 représente la valeur du

critère *RMA* pour chacun des modèles MARS de degré 1. Ce critère permet de déceler si une région du domaine d'intérêt est mal approchée par le modèle proposé.

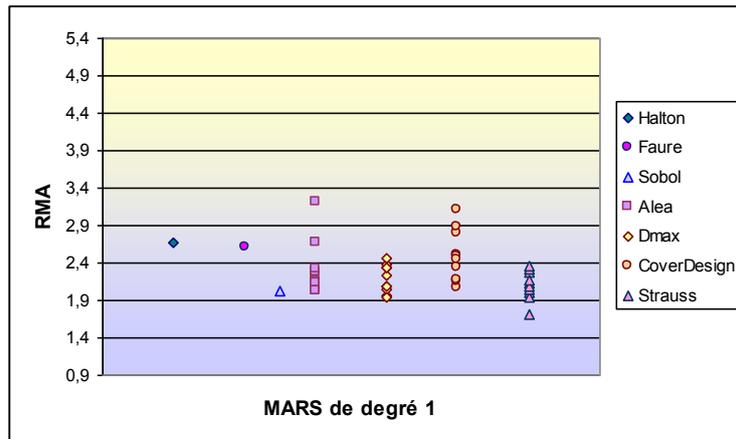


Figure 130. Ecarts absolus maximaux lors de l'utilisation des modèles MARS associés à différents plans d'expériences en prédiction sur le plan test

Pour les modèles MARS de degré 1, on observe que les plans de Strauss et les plans à entropie maximale donnent des critères plus faibles et moins dispersés. Couplés aux critères *RMSE*, cela montre que la réponse prédite est généralement plus proche de la vraie réponse aux points du plan test et que l'erreur maximale commise est plus faible.

10.3.4. PolyMARS

En prenant en compte les résultats de l'étude précédente, les modèles avec le paramètre de pénalisation $\alpha=3$ sont analysés ci-dessous.

Apprentissage.

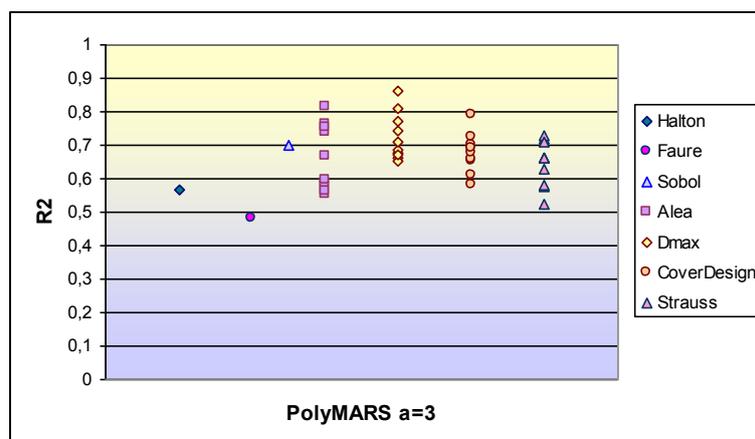


Figure 131. Ajustement des modèles PolyMARS associés à différents plans d'expériences

Les valeurs du critère R^2 pour la méthode PolyMARS sont beaucoup plus dispersées que pour les méthodes présentées avant. Il semblerait que les plans à entropie maximale permettent d'obtenir un modèle plus proche des données d'apprentissage.

Prédiction.

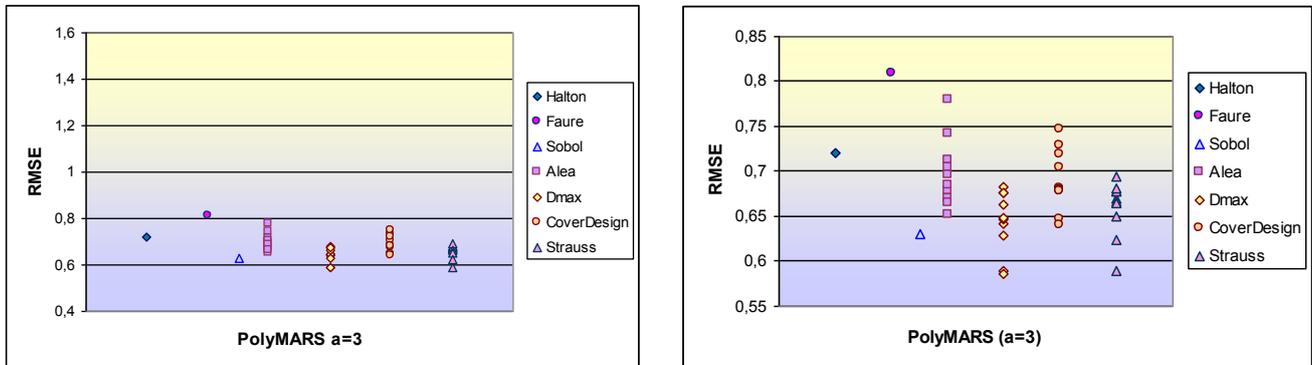


Figure 132. Prédiction des modèles PolyMARS associés à différents plans d'expériences

La Figure 132 montre que les modèles ajustés ont une $RMSE$ entre 0,6 et 0,8. Les plans à entropie maximale et Strauss ont des propriétés légèrement meilleures que les plans aléatoires ou les plans minimax. On remarque cependant que bien que les plans de Strauss et les plans à entropie maximale fournissent des critères en prédiction sensiblement identiques, il en est différemment au niveau de l'ajustement. Ainsi, on préférera choisir les plans à entropie maximale qui sont meilleurs en ajustement.

10.3.5. Krigeage ordinaire

Au vu des résultats précédents, nous présentons uniquement ceux concernant le krigeage ordinaire³³.

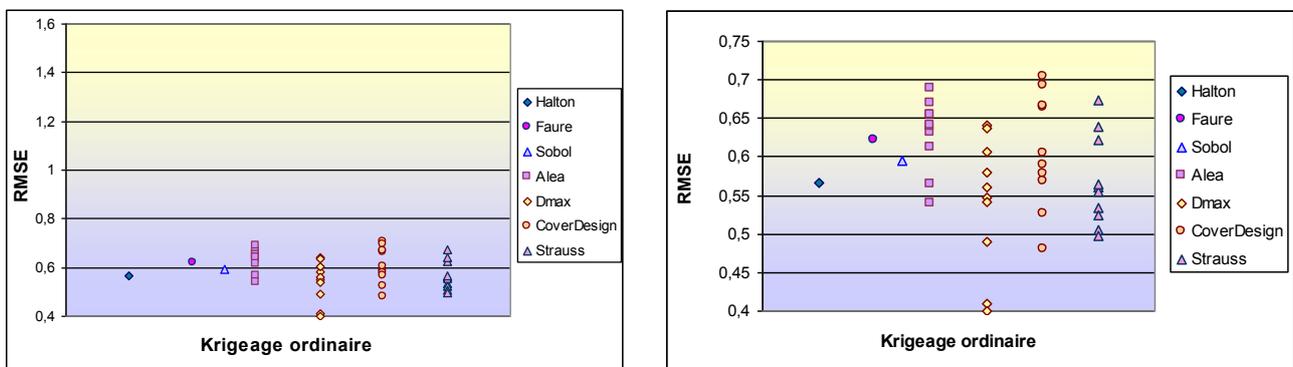


Figure 133. Prédiction des modèles de krigeage

³³ Les résultats présentés ici ont été obtenus avec un code de krigeage en R développé au sein du consortium DICE. Les paramètres sont estimés par maximum de vraisemblance en utilisant un algorithme génétique pour les paramètres de portée.

Interprétation. Les résultats sont de qualité semblable à un modèle linéaire. L'avantage du krigeage est la souplesse de la méthode qui explique en partie la légère disparité des critères comparé au modèle linéaire. La régression linéaire est néanmoins beaucoup plus simple à mettre en œuvre.

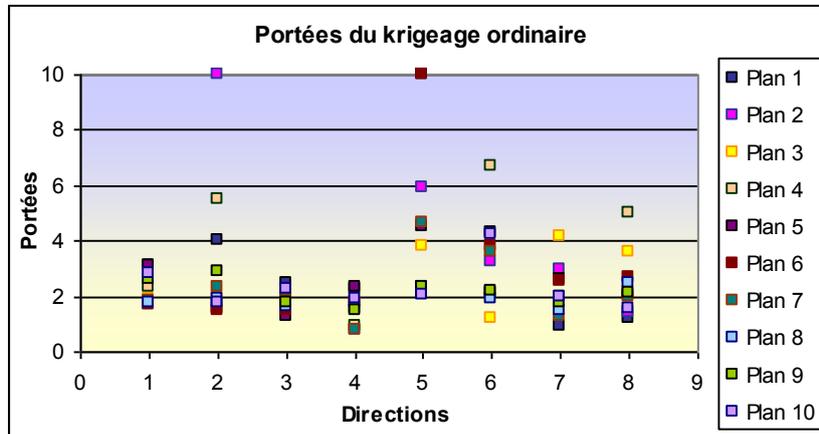


Figure 134. Représentation des portées obtenues avec le krigeage ordinaire à partir des plans de Strauss

On remarque que les portées sont généralement grandes même si les estimations de celles-ci varient d'un plan à l'autre. Ce qui confirme le fait que la surface étudiée est assez lisse. Certaines directions semblent se démarquer, les directions X_1 (abscisse du puits P2), X_3 et X_4 (puits P5) où il y a les portées les plus faibles. A noter que ces facteurs sont ressortis comme les plus influents avec le modèle de régression linéaire (voir Table 3 en annexe G).

10.3.6. Réseaux de neurones

Les résultats présentés ont été obtenus avec les paramètres suivants :

- formule de base : linéaire à effets simples
- decay = 1
- nombre de neurones = 3
- maxit = 5000

Apprentissage

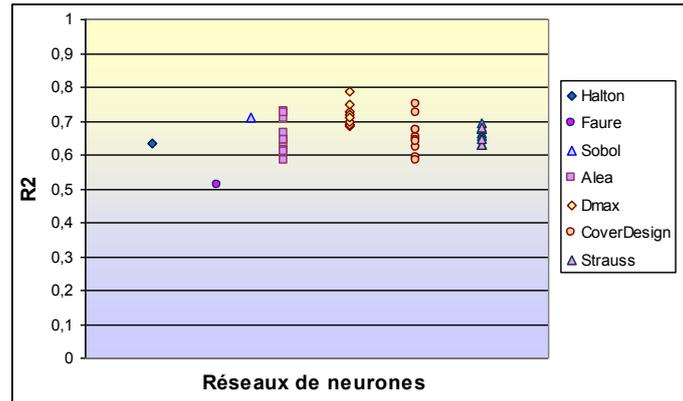


Figure 135. Ajustement des réseaux de neurones associés à différents plans d'expériences

Pour les réseaux de neurones, on constate que les plans à entropie maximale expliquent en moyenne 75% de la variabilité de la réponse.

Prédiction

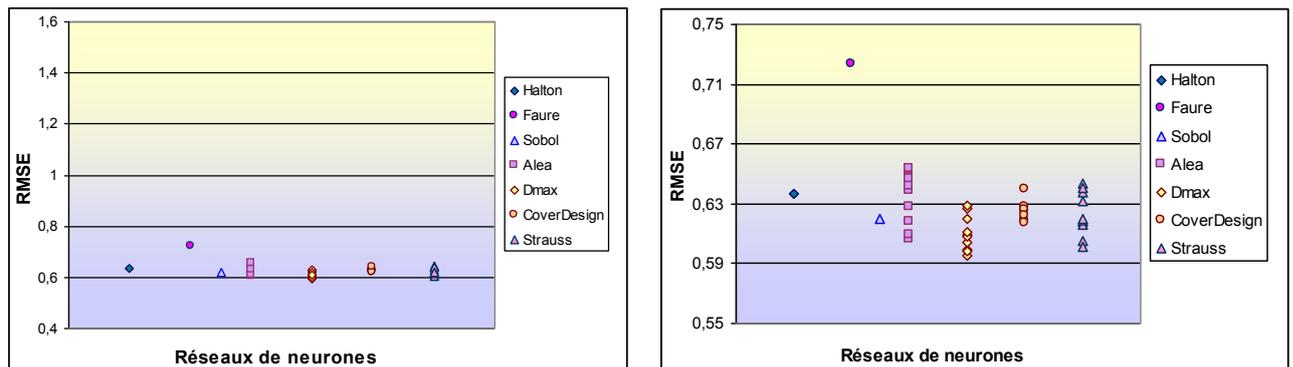


Figure 136. Prédiction des réseaux de neurones associés à différents plans d'expériences

En termes de prédiction, les réseaux de neurones donnent des résultats très satisfaisants et similaires à ceux de la régression simple. La Figure 136 montre une certaine robustesse pour toutes les catégories de plans.

10.3.7. Conclusions

D'après les résultats précédents, il ressort que le krigeage ordinaire est le meilleur modèle en prédiction. La régression linéaire à effets simples donne aussi de bons résultats. Il faut alors se demander quel modèle considérer : soit un modèle simple et d'un peu moins bonne qualité, ou un modèle plus complexe de meilleure qualité.

Concernant la nature du plan d'expériences, nous observons, que quel que soit le modèle approché les plans de Strauss et les plans à entropie maximale sont robustes et donnent généralement les meilleurs résultats en ajustement.

A noter également que selon les surfaces de réponse utilisées, le plan d'expériences paraît avoir plus ou moins d'influence. Par exemple, PolyMARS privilégie les plans de Strauss et les plans à entropie maximale. Au contraire, le modèle linéaire à effets simples ne semble pas être influencé par le choix du plan d'expériences. En effet, la rigidité de cette méthode et le nombre de degrés de liberté assez important (80 données pour estimer 9 paramètres) impliquent la proximité des modèles.

10.4. Etude de l'impact du nombre de points

L'objectif de cette partie est d'étudier l'impact du nombre de points du plan d'apprentissage sur le modèle ajusté. Pour cela, à partir d'un plan initial à 80 points (utilisé dans l'étude précédente), nous avons rajouté 10 points de manière à ce que les points suivent toujours un processus de Strauss. En effet, les plans de Strauss offrent la possibilité de zoomer sur une zone de recherche c'est-à-dire d'affiner la distribution en rajoutant des points de manière à remplir l'espace tout en considérant les points déjà existants.

Nous avons opéré de cette manière de façon à obtenir des plans à 80, 90, 100, 110, 120, 130, 140, 150 et 160 points, imbriqués les uns dans les autres. Cette méthode nous permet non seulement d'économiser un très grand nombre de simulations mais aussi d'étudier l'impact direct du nombre d'expériences sur la surface de réponse obtenue.

Les métamodèles considérés ici sont ceux de la section précédente.

Apprentissage.

La figure ci-dessous permet de visualiser l'évolution du coefficient de détermination en fonction du nombre de points du plan d'expériences.

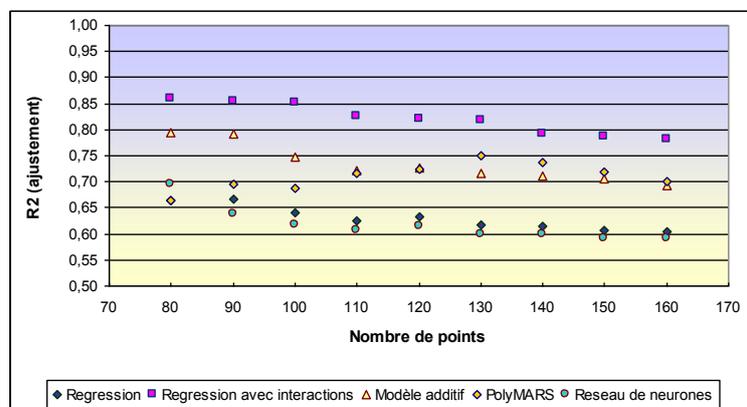


Figure 137. Représentation du coefficient de détermination en fonction du nombre de points

Les coefficients de détermination sont compris entre 0,6 et 0,9. On observe globalement que l'augmentation du nombre de points diminue légèrement les propriétés d'ajustement du modèle excepté pour PolyMARS.

Prédiction sur l'ensemble test.

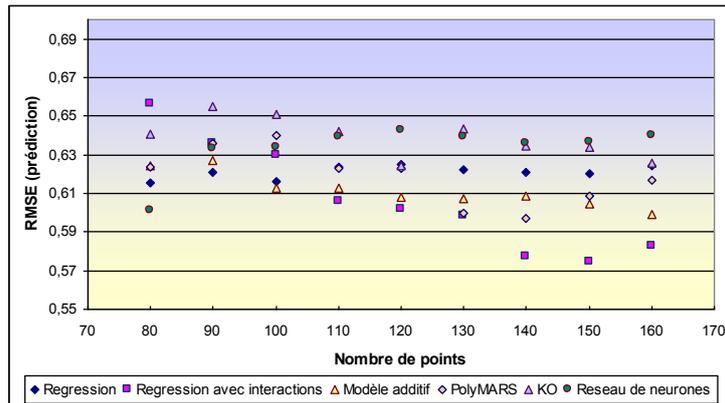


Figure 138. Qualité de prédiction sur le plan test des modèles ajustés

Plusieurs phénomènes méritent d'être remarqués :

- le modèle de régression avec interactions est plus précis si le plan compte plus de points,
- les réseaux de neurones ne donnent pas de meilleurs résultats avec des plans comptant plus de points,
- les résultats du krigeage étant difficilement interprétables (cassure au niveau du plan à 120 points), nous avons représenté les portées sur la figure ci-dessous.

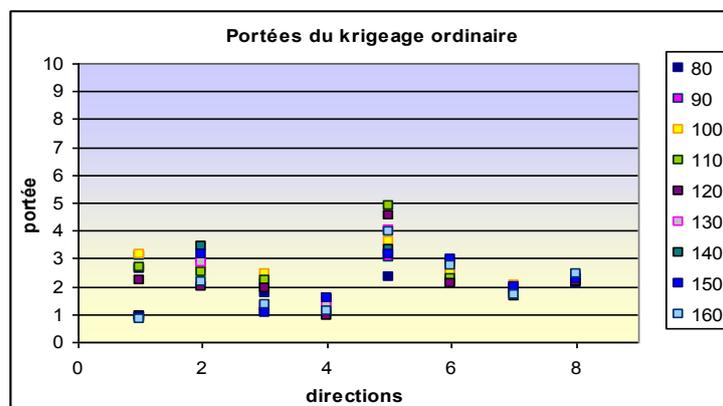


Figure 139. Représentation des portées selon les directions pour les différents plans de Strauss imbriqués

On remarque que le puits I3 (directions 7 et 8) ne semble pas être influencé par le plan utilisé même si ce dernier apparaît comme influent du fait des faibles portées estimées pour ses deux coordonnées.

Au vu de ces résultats, on peut se demander s'il est plus intéressant de considérer des plans avec un nombre de points plus élevé (par exemple 250).

10.5. Conclusions

Certaines configurations de puits permettant de trouver une production plus élevée que celle obtenue avec le cas de base, l'objectif initial a un intérêt réel. Ainsi, après avoir rappelé les conclusions, nous vérifierons qu'il est possible de répondre au problème initial.

Métamodèle. L'hypothèse d'additivité est confirmée avec le modèle additif. La régression linéaire simple expliquant 76% de la réponse semble dire que le modèle est essentiellement linéaire. L'ajout d'interactions dans ce modèle n'influence pas son pouvoir prédictif. Cette conclusion semble être confortée au regard des bons résultats des modèles MARS et PolyMARS. Les valeurs estimées des portées à l'aide du krigeage ordinaire confirment que la surface est lisse.

Le phénomène étant essentiellement linéaire, tous les modèles sont de qualité équivalente. On privilégie alors le modèle PolyMARS qui semble être généralement le moins complexe et aussi un des plus simple à mettre en oeuvre.

Physique du phénomène. Le modèle avec interactions indique une interaction entre la coordonnée verticale du puits P5 et les 2 coordonnées du puits I3. Le krigeage ordinaire avec des portées relativement faibles pour la coordonnée horizontale du puits P2 et le puits P5, montre un impact plus important de ces facteurs sur le phénomène.

Plans d'expériences. L'étude intrinsèque à l'aide des ALM a fait ressortir les plans de Strauss et les plans à entropie maximale dans la zone de quasi-périodicité (voir Figure 122). Ces plans ressortent également comme étant les meilleurs et cela avec n'importe quel métamodèle. Ainsi, ce résultat confirme l'intérêt de considérer des Space-Filling Designs en phase exploratoire. Ces plans sont aussi les plus robustes car on observe beaucoup moins de dispersion au niveau des critères que ce soit en ajustement ou en prédiction. La suite de Sobol' donne dans ce cas des critères de bonne qualité.

Au vu des résultats obtenus avec les ALM et connaissant les défauts de ces suites en projection, nous privilégierons soit un plan de Strauss-Gibbs soit un plan à entropie maximale. Ces deux derniers types de plans étant très proches, et afin de décider lequel choisir nous avons tracé la *RMA* (plus grande erreur absolue) pour le modèle PolyMARS sur la figure suivante.

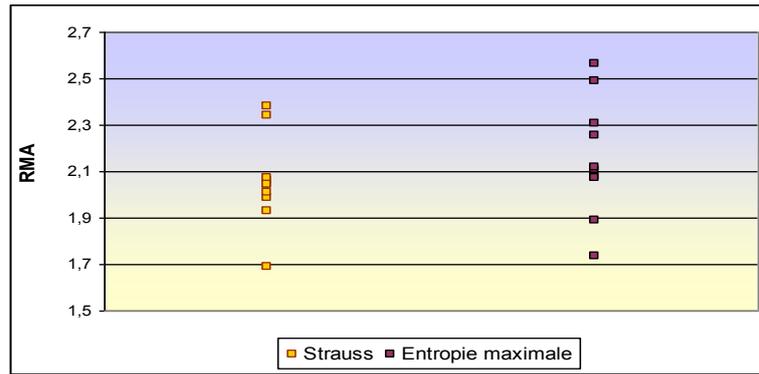


Figure 140. Représentation de la *RMA* pour le modèle PolyMARS ($a=3$) concernant les plans de Strauss et les plans à entropie maximale

La moyenne et l'écart-type de la *RMA* étant plus faibles pour les plans de Strauss, notre choix, pour résoudre le problème d'optimisation, s'est naturellement porté sur le plan de Strauss qui minimise la *RMSE* obtenue avec le modèle PolyMARS.

Résumé. Les comparaisons de plans d'expériences et de métamodèles précédentes nous ont permis de retenir un plan d'expériences et un métamodèle afin de résoudre le problème initial d'optimisation. Nous avons donc retenu un plan de Strauss et la méthode PolyMARS ($a=3$). Ce couple plan/métamodèle fournit un modèle avec 8 fonctions de base. Le modèle est composé de tous les effets simples sauf X_5 et de l'interaction entre X_1 et X_2 . On remarque dans ce cas l'importance du puits P2 dans ce problème et la non influence de l'abscisse de I2.

Optimisation. La méthode choisie pour résoudre le problème d'optimisation est celle de Broyden, Fletcher, Goldfarb and Shanno (BFGS, 1970). Cette méthode permet à la base de résoudre un problème d'optimisation non-linéaire sans contrainte. Aussi, la version considérée ici permet d'introduire des contraintes (Byrd *et al.*, 1995). On rappelle brièvement que l'idée principale de cette méthode est d'éviter de construire explicitement la Hessienne et de construire à la place une approximation de la dérivée³⁴ seconde de la fonction à minimiser, en analysant les gradients successifs.

La convenance choisie ici pour l'initialisation de l'optimisation est de considérer le cas de base (qui vérifie bien les contraintes) qui est déterminé par les coordonnées de puits suivantes

$$(P5; P2; I2; I3) = ((50, 75); (62, 20); (10, 61); (35, 10))$$

celui-ci fournit une production égale à $1,247E+07$ avec le simulateur et à $1,265E+07$ avec le modèle de PolyMARS retenu.

³⁴ Qui permet l'application de la méthode de Quasi-Newton

La solution obtenue avec la méthode L-BFGS-B³⁵ correspond aux coordonnées suivantes (représentées en rouge sur la figure ci-dessous)

$$(P5; P2; I2; I3) = ((65, 78); (74, 6); (10, 47); (52, 2))$$

qui fournit une production égale à 1,590E+07 sur le modèle.

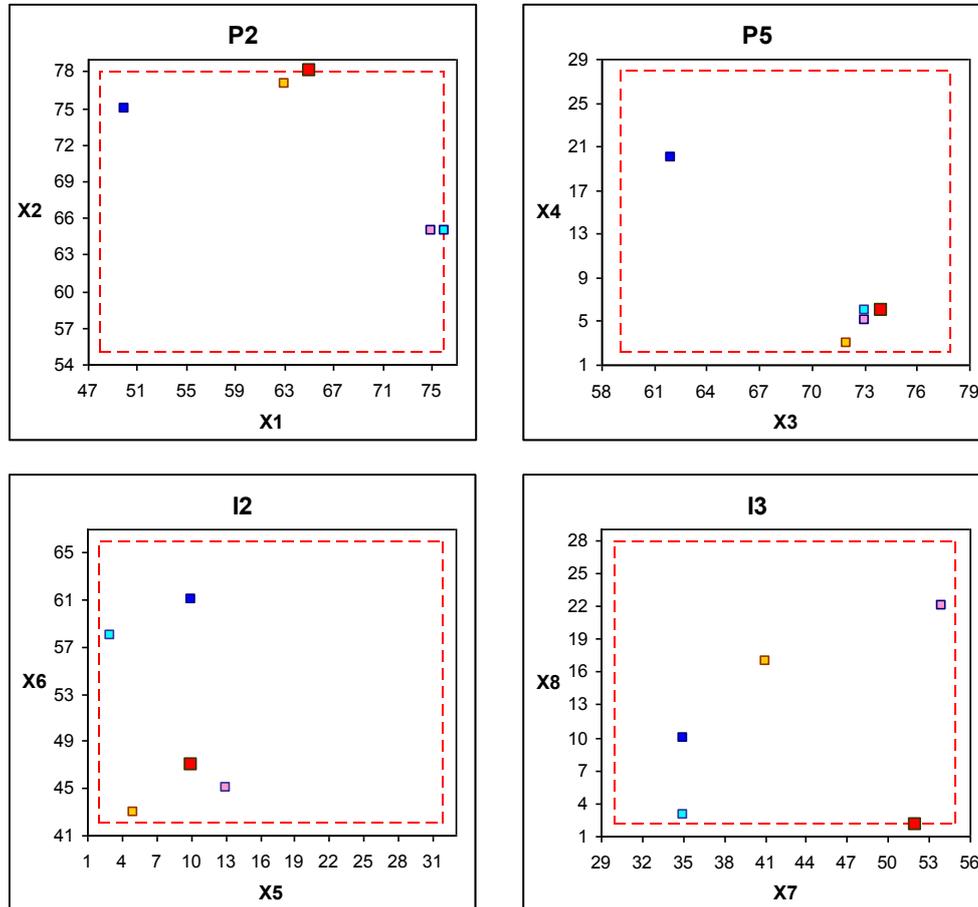


Figure 141. Représentation des 3 meilleures configurations (orange, rose, turquoise) parmi les 80 initialement simulées, du cas de base (en bleu) ainsi que de la solution trouvée par le problème d'optimisation (en rouge)

La solution trouvée par le problème d'optimisation est cohérente notamment au regard de la position de P5.

³⁵ Utilisation de la fonction optim du logiciel R

CONCLUSIONS ET PERSPECTIVES

Cette thèse est consacrée à la planification d'expériences numériques en phase exploratoire sur les gros codes de calcul. En simulation numérique, les difficultés essentielles proviennent du coût du calcul du simulateur et de la grande dimension du problème à traiter. L'approche, actuellement privilégiée, consiste à construire un modèle d'approximation (régression, etc.) ou d'interpolation (krigeage, etc.) à partir d'un nombre réduit de simulations. Ces simulations sont effectuées dans des conditions fixées *a priori* et organisées selon un plan d'expériences numériques. Cette approche de modélisation de la réponse nécessite une connaissance *a priori* du type de dépendance entre la réponse et les paramètres et atteint ses limites lorsque la réponse étudiée est complexe. Ce travail se place dans le contexte de la phase exploratoire, à savoir lorsque cette dépendance n'est pas connue.

Cette situation intervient fréquemment en industrie pétrolière, et même plus généralement dans les codes simulant des phénomènes complexes. Il faut ainsi prendre en compte cette dépendance complexe, souvent fortement non-régulière, de ces réponses par rapport aux nombreux paramètres physiques qui décrivent le phénomène. En ingénierie réservoir, ces paramètres peuvent être de différents types : géométriques (taille du réservoir, emplacement des failles), d'exploitation (type de récupération, position et nombre de puits) ou bien encore pétrophysiques (perméabilités, porosités).

Le problème consiste à construire un modèle approché du modèle numérique (d'écoulement par exemple en industrie pétrolière) qui soit le plus prédictif possible. La construction de ce modèle se faisant à partir d'un jeu de données connu, il est nécessaire de choisir les expériences de façon optimale tout en contrôlant le nombre de simulations numériques requises pour calibrer le modèle approché afin d'amoindrir le compromis biais/variance.

Les plans d'expériences doivent donc posséder de bonnes propriétés telles que

- le remplissage de l'espace afin de capter les éventuelles non-linéarités mais aussi pour éviter les répétitions qui ne sont plus forcément de mise
- le remplissage de l'espace doit également subsister par réduction de la dimension car le code ne dépend généralement que de quelques variables influentes, facteurs ou composantes principales.

CONTRIBUTIONS.

Cette thèse propose une nouvelle méthode de planification d'expériences qui s'adapte facilement aux exigences du contexte dans lequel nous nous plaçons. Cette technique basée sur des cas particuliers du modèle de Strauss et de Gibbs est une façon de remplir l'espace en

interdisant que des points soient trop proches. Ce qui est bien adapté chaque fois qu'un individu nécessite un espace propre : modèles de sphères impénétrables en physique, répartition d'arbres dans une forêt, répartition d'animaux sur un territoire, centres de cellules d'un tissu cellulaire, répartition d'un type de commerce dans une ville, etc. Les plans de Strauss ainsi construits ont des points bien répartis dans l'espace, mais à condition d'avoir paramétré correctement la méthode. En effet, la paramétrisation du rayon d'interaction est délicate et influe fortement sur la distribution finale : un rayon trop petit engendre une distribution sans interaction mais avec de nombreuses lacunes et un rayon trop grand conduit à une distribution avec des agglomérats. Cette paramétrisation est d'autant plus difficile que la dimension est grande. L'utilisation du potentiel de Gibbs, prenant en compte la force des interactions entre les points, permet d'avoir une marche de manœuvre plus importante pour fixer ce paramètre qu'on pourra ainsi surestimer. Une autre solution, plus drastique, est de procéder par dichotomie sur la valeur du rayon et de valider la qualité des plans à l'aide de critères intrinsèques.

La complexité des phénomènes décrits par des codes de calculs lourds incite à répartir les expériences numériques le plus uniformément possible dans le domaine. En outre il est souhaitable que la répartition soit également satisfaisante en projection sur des sous-espaces, au cas où le code ne dépendrait que d'un petit nombre de facteurs ou de composantes principales. Aussi, l'approche des plans spécifiés à partir des méthodes MCMC offrent une grande souplesse dans la mesure où l'on peut jouer facilement sur la loi de Gibbs et sa représentation en termes de potentiels locaux d'énergie pour imposer des propriétés telles que le remplissage de l'espace ou la bonne projection des points sur les marges de dimension 1.

Le travail de cette thèse a été par ailleurs de proposer et d'utiliser différents outils d'étude de la qualité des plans d'expériences pour s'assurer de leurs propriétés. Nous avons basé essentiellement nos recherches sur des outils de visualisation qui permettent de conclure rapidement sur la qualité des plans.

Un premier critère statistique permet de généraliser l'utilisation de la discrédance L^∞ aux projections sur tous les sous-espaces de dimension 1. Ce critère est appelé radar d'uniformité car il permet en dimension 2 et 3 de tester graphiquement l'hypothèse d'uniformité du plan, par balayage dans toutes les directions. Nous avons également introduit une statistique globale en dimension 2, qui permet de mieux rejeter des plans décevants mais néanmoins acceptés par le radar d'uniformité.

Enfin, un nouveau critère de qualification intrinsèque des plans, basé sur les propriétés des Arbres de Longueur Minimale (ALM), est proposé. Ce critère permet de classer et de qualifier toute distribution de points dans l'espace selon une cartographie, en moyenne et écart-type, des longueurs des arêtes des ALM associés aux plans étudiés.

Le contexte de la planification numérique implique généralement cette quête d'uniformité au sein des plans d'expériences. Cependant, il s'avère qu'elle n'est pas toujours recherchée. La connaissance métier du phénomène peut donner des intuitions sur certaines zones du domaine expérimental. Ainsi, il doit être possible de distribuer des points de manière préférentielle en privilégiant les zones les moins connues. Les plans de Strauss s'adaptent encore une fois à cette problématique. L'idée est de doter l'espace d'une « métrique » hétérogène et ainsi de considérer une fonction distance variable dans l'espace. Une autre motivation de cette méthode est de corriger le fléau de la grande dimension en densifiant les points au centre du domaine. Par exemple, l'utilisation d'une distance à densité gaussienne accordera beaucoup plus de points au centre du domaine que l'utilisation d'un plan de Strauss usuel en grande dimension.

Un autre problème pouvant intervenir lors de la planification d'expériences est celui de générer un certain nombre de points dans un domaine défini à l'aide de contraintes, particulièrement utile dans les problèmes d'optimisation. Ainsi, il est bien souvent difficile de générer un nombre de points fixé dans cet espace, parfois très réduit, et avec des expériences les mieux réparties qui soient. La méthode *a priori*, Step-by-Step, propose d'ajouter, à partir de quelques expériences réalisables, des points selon une dynamique globale qui tend à les éloigner les uns des autres d'un rayon fixé avec une intensité plus ou moins forte (similairement au processus de Strauss).

Les plans séquentiels (évolutifs) liés au modèle n'ont pas été discutés dans ce travail qui se limitait essentiellement à la phase exploratoire. Il est bien évident que le rajout d'expériences de manière optimale selon le type de surface de réponse est un sujet intéressant. On peut citer par exemple les travaux de C. Scheidt à ce sujet. Les seuls plans évolutifs que nous proposons ici restent dans le contexte de la phase exploratoire soit parce que nous n'avons pas assez appris sur le phénomène, soit parce que le modèle ne convient pas. Les plans de Strauss permettent d'ajouter, judicieusement, et de contrôler le nombre de simulations à chaque itération jusqu'à ce que la qualité de prédiction de l'approximation soit jugée satisfaisante.

Nous avons montré la robustesse des plans de Strauss et des plans à entropie maximale en traitant des cas d'écoles et des exemples rencontrés en ingénierie réservoir.

A noter que ces plans étant à la base développés pour la phase exploratoire et les codes de calculs, il ne faut pas pour autant les restreindre à ce contexte. Il est tout à fait envisageable d'utiliser des plans de Strauss lors d'expérimentation classique, ou alors en faible dimension, etc.

PERSPECTIVES.

L'utilisation des plans de Strauss ainsi que les plans à entropie maximale semble être prometteuse en expérimentation numérique. Aussi, nous pensons qu'il peut être intéressant d'apporter des améliorations à ces deux méthodes.

Le problème de paramétrage du rayon d'interaction reste primordial concernant les plans de Strauss. Il pourrait être intéressant de tabuler ces plans, plus précisément de tabuler le rayon d'interaction selon le nombre de points à générer et la dimension du problème. Un point essentiel concernant les validations de modèles qui sont souvent basées sur la validation croisée n'a pas été traité lors de ce travail. Il serait donc particulièrement intéressant d'étudier la robustesse des plans dans ce cas. Une perspective pourrait être de proposer une méthode permettant de repositionner les points selon les expériences choisies pour la validation croisée ; ce qui revient à choisir au préalable les expériences qui vont servir à la validation du modèle.

Il pourrait être intéressant d'étudier le comportement de tous les plans d'expériences au regard de la validation croisée. La validation croisée très utilisée en expérimentation classique même si cela ne semble pas très pertinent car retirer des points à un plan factoriel fractionnaire, par exemple, détruit sa structure. Il est légitime de se demander ce que deviennent les propriétés initiales du plan. Il paraît évident que lors de la validation croisée, il faut écarter tous les plans ayant une structure rigide tels que les plans standards, les hypercubes latins, etc.

Il serait pertinent concernant les plans à entropie maximale de ne plus faire d'hypothèses sur un modèle sous-jacent. Similairement à la discrédance, la divergence de Kullback-Leibler (1951) permet de mesurer l'écart entre la distribution empirique et la loi uniforme. Dans le cas de la loi uniforme, minimiser la divergence de K-L revient à maximiser l'entropie. La maximisation de l'entropie n'aurait pas ici pour objectif d'augmenter la quantité d'information (au sens de Shannon) contenue dans l'échantillon relativement à des paramètres du modèle mais de s'approcher d'une distribution uniforme. Afin d'estimer l'entropie, nous avons choisi la méthode des noyaux d'Epanechnikov multidimensionnelle avec une taille de

fenêtre déterminée par la distance aux plus proches voisins (méthode NPPV). A ce sujet, Beirlant *et al.* (2001) proposent un état de l'art sur les différentes techniques d'estimation de l'entropie. Une idée est donc de construire de façon empirique des plans de divergence K-L minimale à l'aide d'un simple algorithme d'échange en estimant à chaque étape l'entropie. Une étude a été lancée sur ce sujet et les résultats semblent prometteurs comme on peut le voir sur les figures ci-contre. La figure ci-dessous reprend l'étude présentée au paragraphe 6.2.2 en y incluant les plans à divergence K-L minimale que l'on dénommera plans K-L.

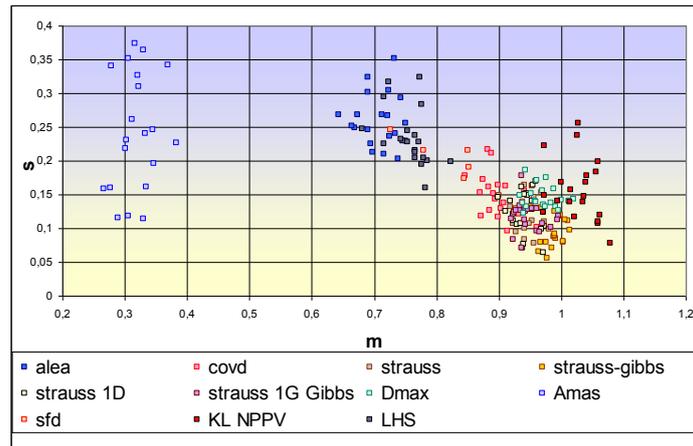


Figure 142. Représentation du critère lié aux ALM pour des plans à 30 points en dimension 3

On remarque que ces plans se situent dans la zone proche des plans à entropie maximale. Afin de confirmer leur qualité, nous les avons testés à l'aide de l'application en dimension 3.

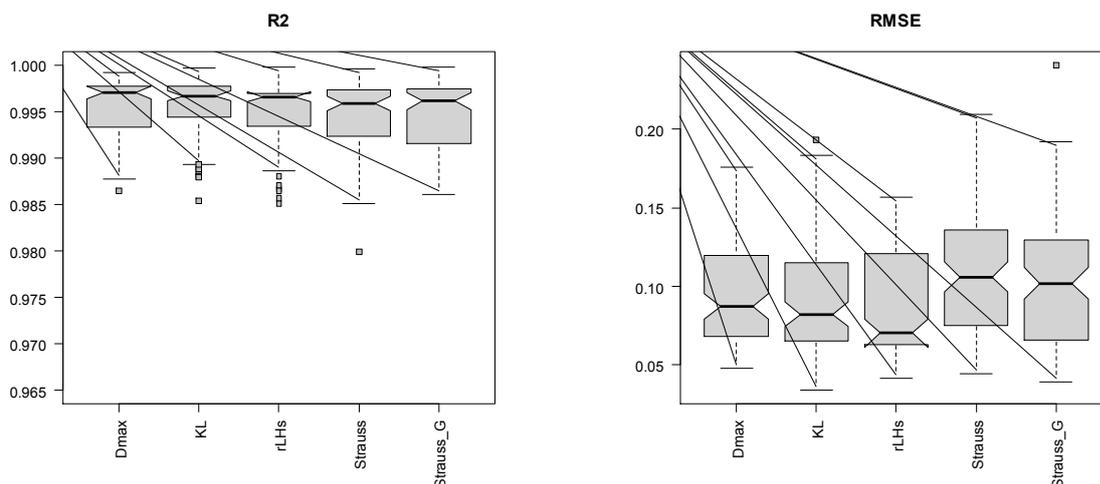


Figure 143. Boxplot pour mesurer la qualité d'ajustement (à gauche) et de prédiction (à droite) du modèle de régression quadratique des plans de Strauss, des plans à entropie maximale (dmax) et des plans à divergence minimale (KL)

Il ressort, à travers cette application, que les plans à divergence minimale semblent prometteurs. Cependant, il faudrait les tester en grande dimension pour s'en assurer.

ANNEXES

Annexe A.

Rappels sur les surfaces de réponse

Le lecteur ne doit pas attendre de cette partie la description détaillée de chacune des surfaces de réponse discutées ici. Seule l'écriture de la relation entre les entrées et les sorties sera fournie, l'intérêt n'étant pas de redéfinir chaque métamodèle en tant que tel. L'objectif de cette partie est de discuter des métamodèles vis-à-vis des objectifs de l'expérimentateur, de fournir les références qui s'y prêtent et d'exposer les points essentiels retenus lors de leur utilisation dans les applications.

A.1. Régression linéaire

Très utilisée en expérimentation classique il existe de nombreux ouvrages qui expliquent et fournissent les hypothèses de la méthode (Myers et Montgomery, 1995).

Le modèle. On note $f(X) = (f_0(X), \dots, f_M(X))$ un vecteur de fonctions connues. On suppose $Y = f(X)\beta + \varepsilon$ où $\beta = (\beta_0, \dots, \beta_M)^t$ est le vecteur constitué des paramètres du modèle et où ε est de loi normale centrée de variance constante égale à σ^2 .

L'analyse du modèle se fait à l'aide de la table d'ANOVA.

Discussion. Malgré la rigidité de ce modèle, il reste souvent utilisé. En effet, c'est un des modèles le plus simple à construire qui donne des informations initiales nécessaires en grande dimension quand le nombre d'expériences est moindre et que les autres méthodes sont trop gourmandes. La régression est tout particulièrement utilisée dans les techniques de criblage pour sélectionner les variables influentes. Ce qui se justifie par le fait que cette méthode offre la possibilité d'utiliser les tests statistiques comme outil d'aide à la décision.

De plus, une relation de type polynômiale entre les entrées et la réponse suffit parfois à expliquer le phénomène. A noter également que cette remarque devient invraisemblable lorsque les phénomènes modélisés sont très complexes, par exemple si les variables interagissent beaucoup entre elles.

A.2. Modèles additifs

Les modèles additifs sont un cas particulier des modèles additifs généralisés (GAM) développés par Hastie et Tibshirani (2001). Cette méthode présente l'avantage d'accorder à la surface plus de flexibilité qu'un modèle linéaire. La structure additive offre un réel outil de visualisation qui permet de représenter l'effet de chaque variable sur la réponse. Cette méthode est aussi particulièrement adaptée aux cas de grande dimension.

Le modèle. L'hypothèse d'additivité sur laquelle est basée ce modèle suppose que la réponse s'écrit comme une somme de fonctions en chacune des variables explicatives. En comparaison avec la régression linéaire, chaque terme linéaire est remplacé par une fonction plus générale et la réponse Y se décompose sous la forme suivante :

$$Y = \alpha + f_1(x_1) + \dots + f_d(x_d) + \varepsilon$$

où les $f_j (j=1, \dots, d)$ sont des fonctions de la variable associée x_j à estimer et ε un bruit blanc indépendant des variables. L'unicité de la décomposition est assurée en supposant que $E(f_j) = 0$. La constante α représente alors le niveau moyen des données.

Discussion. Cette décomposition présente l'avantage d'ajuster des modèles non paramétriques simples et faciles à interpréter. Si le problème contient des interactions importantes, le modèle additif peut conduire à des résultats erronés. La généralisation du modèle ci-dessus peut se faire en prenant des termes dépendant de plusieurs prédicteurs, souvent limité au nombre de 2.

Les modèles additifs ont l'avantage d'être un outil de visualisation des variations de la réponse dans chaque direction principale, ce qui peut amener à simplifier le modèle. L'étude des données par cette méthode fournit une information *a priori* qui pourra par la suite être utilisée pour régler les paramètres lors de l'ajustement d'un autre modèle. Citons par exemple l'étude de la nuée variographique empirique pour le krigeage.

A.3. MARS et PolyMARS

Ces méthodes sont des modèles polynomiaux par morceaux qui sont particulièrement bien adaptés en grande dimension.

La méthode MARS (Multivariate Additive Regression Spline) proposée par Friedman (1991) utilise des polynômes par morceaux de degré 1 *i.e.* des splines linéaires pour estimer les fonctions de régression. Une extension de MARS proposée par Kooperberg (1997) permet de construire des modèles PolyMARS (Polychotomous regression based on MARS). Ces modèles peuvent s'appliquer dans le cas des réponses multiples et paraissent être davantage adaptée à la grande dimension que la méthode MARS.

La principale similitude entre ces 2 méthodes est la construction du modèle ajusté qui se déroule en trois phases :

- phase d'addition de fonctions de base dans le modèle,
- phase de suppression,
- et sélection du modèle final selon un critère de validation croisée généralisée (GCV).

Méthode MARS. C'est une généralisation de la régression linéaire stepwise où les fonctions de base sont des splines linéaires. Le modèle s'écrit sous la forme suivante :

$$Y = \beta_0 + \sum_{j=1}^J \beta_j h_j(X)$$

où les paramètres β_j sont estimés par moindres carrés et où les fonctions de base h_j sont des splines linéaires avec un nœud à la valeur t *i.e.* des fonctions de la forme :

$$(x-t)_+ = \begin{cases} x-t & \text{si } x > t \\ 0 & \text{sinon} \end{cases} \quad \text{ou} \quad (t-x)_+ = \begin{cases} t-x & \text{si } x < t \\ 0 & \text{sinon} \end{cases}$$

Les fonctions de base sont construites en définissant un nœud pour chaque valeur x_j^i du plan d'expériences. Notons que seuls les nœuds « lointains » du nœud déjà choisi peuvent être candidats pour un nouveau nœud en cette variable.

La méthode MARS peut donc être vue comme une régression linéaire avec des fonctions de base construites à partir des observations au lieu d'être imposées.

Méthode PolyMARS. Cette méthode permet de prendre en compte plusieurs réponses simultanément. Par soucis de simplicité, nous considérons ici uniquement le cas d'une seule réponse. La réponse Y est approchée par des fonctions qui dépendent au plus de deux

prédicteurs x_j . Similairement à MARS, les fonctions de base utilisées sont des splines linéaires et peuvent être ainsi de la forme 1 , x_j , $(x_j - t)_+$, $x_j x_k$, $(x_j - t)_+ x_k$ ou $(x_j - t)_+ (x_k - s)_+$.

Différences entre MARS et PolyMARS. La principale différence se situe au niveau des possibilités pour les fonctions de base. Les nombreux choix concernant les fonctions de base pour la méthode PolyMARS peuvent parfois induire un temps de calcul plus important.

MARS explore un espace de fonctions alors que PolyMARS utilise la notion d'espace autorisé c'est-à-dire que la construction de l'espace satisfait les contraintes suivantes :

- ajout des effets simples x_j
- ajout d'une spline linéaire $(x_j - t)_+$ si le prédicteur x_j est inclus dans le modèle
- ajout d'un terme d'interaction $x_j x_k$ si les deux prédicteurs sont inclus dans le modèle individuellement
- ajout de $(x_j - t)_+ x_k$ si les deux termes sont inclus dans le modèle
- ajout de $(x_j - t)_+ (x_k - s)_+$ si $(x_j - t)_+ x_k$ et $(x_k - s)_+ x_j$ sont dans le modèle

Le respect de cet ordre permet d'obtenir des modèles plus simples et donc plus faciles à interpréter. D'autres raisons telles que la réduction de la variance résiduelle et des arguments liés au taux de convergence permettent de justifier cette approche (Stone *et al*, 1997). Cependant, les contraintes de construction peuvent parfois faire obstacle à la bonne représentation de la réponse.

Lors de la phase de suppression, cet ordre doit aussi être respecté. Par exemple, il faut que le terme d'interaction ait été préalablement supprimé avant que les effets principaux le soit.

La méthode PolyMARS permet de diminuer le nombre de nœuds³⁶ et donc de gagner en efficacité par rapport à la méthode MARS qui, quand à elle, teste toutes les valeurs observées. D'autre part, les procédures d'addition et de suppression dans PolyMARS sont basées sur des tests statistiques et non sur un critère GCV.

³⁶ L'utilisateur peut contrôler le nombre de nœuds explorés et ceci avec les dangers que cela comporte.

A.4. Réseaux de neurones

Un neurone est une fonction non linéaire, paramétrée, bornée, en général, une fonction sigmoïde :

$$a\left(\sum_{j=1}^{d+1}\theta_j x_j\right)$$

où x_j , $j=1, \dots, d$ sont les variables, $x_{d+1}=1$, et θ_j , $j=1, \dots, d+1$ sont les paramètres de la fonction. La fonction sigmoïde est appelée fonction d'activation. On trouve par exemple,

$$a(x) = x \text{ ou bien encore } a(x) = \frac{1}{1 + e^{-x}}.$$

Un réseau de neurones est une combinaison de neurones :

$$f(x, \Theta) = \theta + \sum_{k=1}^{N_c} \theta_k a\left(\sum_{j=1}^{d+1} \theta_{kj} x_j\right)$$

où N_c est le nombre de neurones dans la couche cachée et $\Theta = \{\theta, \theta_k, \theta_{jk}\}$ $k=1, \dots, N_c$ et $j=1, \dots, d+1$ sont les paramètres inconnus du réseau, encore appelés poids.

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, 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.).

Une description détaillée du fonctionnement d'un réseau de neurones se trouve dans (Thiria, 1997)

Discussion. Les réseaux de neurones donnent une approximation universelle : toute fonction bornée suffisamment régulière, peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones à une couche cachée avec un nombre fini de neurones ayant même fonction d'activation et un neurone de sortie linéaire.

La propriété importante des réseaux de neurones est celle de la parcimonie. Pour une précision donnée, donc un nombre de neurones cachés donné, le nombre de paramètres à estimer du réseau est proportionnel au nombre de variables d'entrées. Ainsi, quand le nombre

de variables est grand, il est plus avantageux d'utiliser un réseau de neurones qu'un modèle polynômial.

Enfin, l'apport des réseaux de neurones réside dans leur capacité à modéliser des processus non linéaires grâce aux fonctions d'activation utilisées.

A.5. Krigeage

Cette méthode a été développée dans le domaine de la géostatistique par Matheron dans les années 60 (1963) et a été utilisée en planification d'expériences numériques par Sacks (1989). Le krigeage présente l'avantage d'être une méthode interpolante aboutissant à la construction d'un modèle probabiliste. Cette méthode va donc dépendre fortement des points d'expériences proposés. Le krigeage consiste à construire un interpolateur qui est une fonction linéaire des observations tenant compte de la structure spatiale des données.

Le modèle d'interpolation dans sa forme générale est

$$Y(x) = \sum_{j=0}^d \beta_j f_j(x) + Z(x)$$

où $f_1(\cdot), \dots, f_d(\cdot)$ sont des fonctions de base connues, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^T$ un vecteur de coefficients estimés et Z un processus gaussien stationnaire centré et de forme de covariance factorisée gaussienne.

Ce type de modèle est référé sous le nom de krigeage universel. Il est aussi possible de considérer un krigeage ordinaire de la forme $Y(x) = \mu + Z(x)$ où μ est une constante à estimer. Cette forme de krigeage est généralement la plus utilisée en pratique.

Le krigeage est longuement détaillé par Santner *et al.* (2003) et Koelher et Owen (1996). Nous nous contenterons donc ici de discuter des points forts et des points faibles de cette méthode.

Discussion. Le krigeage est une méthode intermédiaire entre des méthodes rigides telles la régression où les fonctions utilisées sont imposées en amont et des méthodes souples telles les modèles additifs ou PolyMARS où l'estimation repose sur la construction de splines dans chaque direction. Concernant le krigeage, la régularité de la fonction est choisie au travers une fonction de corrélation et de paramètres déterminés par les observations.

Un autre avantage du krigeage est de pouvoir associer une variance de prévision à la surface. Ainsi, il est intéressant de tenir compte de cette information en jouant différemment sur les zones connues avec précision et celles où on ne dispose que de très peu d'observations.

Les principaux inconvénients du krigeage se situent au niveau numérique. En effet, l'inversion de la matrice de corrélation est difficile à cause du mauvais conditionnement. Ce qui se justifie, en grande dimension, par le nombre d'observations souvent insuffisant pour permettre l'identification des paramètres de portée. Ainsi, la fonction de vraisemblance est souvent plate au voisinage de l'optimum. Pire encore, le profil de vraisemblance peut présenter des optima locaux. Il convient alors de choisir des algorithmes d'optimisation adaptés.

Contrairement aux autres méthodes présentées ci-dessus, l'interprétation de l'estimateur du krigeage est difficile.

A.6. Variogramme

Nous terminons ce paragraphe par la définition du variogramme nécessaire à la bonne compréhension des plans d'entropie maximale.

Définition du variogramme. En géostatistique, la fonction la plus utilisée pour décrire la continuité d'un processus est le variogramme. Et cela surtout parce qu'elle est plus simple à estimer que la covariance (qui demande l'estimation préalable de l'espérance mathématique), mais aussi parce qu'elle permet d'accommoder des situations où la variance du processus Z à modéliser est inconnue.

Le variogramme théorique est défini par :

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

sous l'hypothèse d'un processus stationnaire.

Cette fonction, croissante en fonction de h , synthétise beaucoup d'information concernant le comportement conjoint des variables aléatoires. Si la valeur du variogramme converge vers une valeur limite, appelée palier, cela signifie qu'il y a une distance h à partir de laquelle $Z(x)$ et $Z(x+h)$ sont non corrélés. On définit alors :

- portée a : il s'agit de la distance à partir de laquelle deux observations ne sont plus corrélées
- palier : c'est la variance de la variable aléatoire $Z(x)$
- effet de pépite : il s'agit de variations à courte échelle, erreur de localisation, d'analyse et de précision analytique. La pépite traduit la discontinuité du processus à très courtes distances.

Annexe B.

Rappels sur les critères de qualité des surfaces de réponse

Cette partie traite des différents critères permettant de mesurer la qualité de la surface ajustée tant en ajustement du modèle qu'en prédiction.

Après avoir construit un modèle, la question de la validité du modèle obtenu se pose. Un premier diagnostic consiste à étudier les valeurs du modèle aux points du plan d'expériences : nous parlerons dans ce cas **d'ajustement**. Pour se faire il existe des critères permettant de tester la bonne adéquation entre les observations et les prévisions sur les données utilisées pour la construction du modèle. C'est notamment le rôle des coefficients de détermination R^2 et R^2 ajusté, de l'étude des résidus, etc.

Une fois le modèle accepté, il est nécessaire de quantifier la qualité de la surface ajustée en termes de **prédiction**. Il s'agit ici de comparer les prévisions du modèle en des points différents de ceux du plan d'expériences. Dans le cas où le nombre de simulations réalisables n'est pas contraint, il est alors possible de définir un ensemble de points tests sur lequel évaluer des critères d'erreur en prédiction (MAE , $RMSE$, etc.). Dans le cas contraire, il est possible d'utiliser les techniques de validation croisée.

AJUSTEMENT

Le but ici est de proposer des indicateurs permettant de vérifier la bonne adéquation entre les données simulées et leur approximation calculée à partir du métamodèle. A noter que ce paragraphe ne concerne pas les méthodes d'interpolation où le métamodèle passe par les points du plan d'expériences.

Coefficients de détermination. Le coefficient de détermination est le critère généralement utilisé en régression linéaire pour tester l'ajustement du modèle. Ce coefficient est défini par

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

où \bar{y} est une estimation de la moyenne de la réponse et n est le nombre de points du plan d'expériences.

Ce critère permet de mesurer le pourcentage de la variabilité totale de la réponse expliquée par le modèle. Le coefficient ne doit pas être utilisé pour comparer des modèles entre eux dans la mesure où il dépend fortement du modèle. En effet, le R^2 augmente lorsque le nombre de termes grandit même si tous les prédicteurs ne sont pas significatifs. Pour pallier ce problème, on utilise plutôt le coefficient de détermination ajusté défini par :

$$R_{adj}^2 = 1 - \frac{\frac{1}{n-(p+1)} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}$$

où p désigne le nombre de termes dans le modèle (constante non incluse). Si on exprime R_{adj}^2 en fonction de R^2 , on constate que le R_{adj}^2 est toujours plus faible que le R^2 et que la différence entre ces deux coefficients est d'autant plus grande que le nombre de prédicteurs est élevé. Le R_{adj}^2 est donc un compromis entre un modèle qui représente fidèlement la variabilité de la réponse et un modèle qui n'est pas trop riche en prédicteurs.

Résidus. Certaines méthodes ont des hypothèses sous-jacentes sur les *résidus* $\hat{\varepsilon}_1, \dots$. Par exemple, en régression on suppose que les résidus sont centrés. Une étude des résidus peut donc s'avérer utile pour comparer les informations fournies par les *coefficients de détermination*.

PREDICTION

Le plan d'expériences représente seulement un petit échantillon des valeurs possibles pour les variables explicatives alors que nous espérons pouvoir utiliser le modèle ajusté pour construire une valeur approchée de la réponse en n'importe quel point du domaine d'étude. Cette deuxième étape permet donc d'étudier les qualités de généralisation du modèle proposé c'est-à-dire son comportement en prédiction.

Nous supposons que nous disposons d'un nombre raisonnable de points tests pour étudier les qualités prédictives du modèle. Ce sera notamment le cas pour l'application 3D. Les indicateurs proposés ci-dessous mesurent généralement l'écart entre la prédiction calculée via le modèle ajusté et la valeur réelle de la réponse donnée par le simulateur.

Un coefficient de détermination R^2 évalué sur un ensemble test (aussi appelé R^2 externe) donne une indication sur les qualités d'un métamodèle. La valeur du R^2 en prédiction peut être négative ce qui signifie que le modèle crée de la variabilité par rapport à un modèle constant.

Erreur quadratique moyenne : MSE . La MSE (Mean Square Error) correspond à la moyenne des erreurs de prédiction au carré (critère de type L_2) :

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

où m représente le nombre de données de l'ensemble test. Ce critère mesure l'erreur quadratique moyenne commise sur les données tests. Concrètement, une faible valeur de la MSE signifie que les valeurs prédites sont proches des valeurs réelles. On raisonnera parfois sur la $RMSE$ (Root Mean Squared Error) définie par : $RMSE = \sqrt{MSE}$.

Remarque. Le critère $RMSE$ dépend de l'ordre de grandeur des valeurs observées. Ainsi ce critère peut varier beaucoup d'une application à l'autre. Il existe deux possibilités pour remédier à ce problème. La première est de centrer réduire la réponse (ce qui est généralement envisagé). La deuxième est de considérer des critères relatifs en moyenne (RRM) et en écart-type (RRE) définis comme suit :

$$RRM = \frac{1}{m} \sum_{i=1}^m \frac{y_i - \hat{y}_i}{y_i} \text{ et } RRE = \sqrt{\text{Var} \left(\frac{y_i - \hat{y}_i}{y_i} \right)}$$

Erreur absolue moyenne : MAE . Le critère MAE (Mean Absolute Error) (correspondant à la norme L_1) est défini par :

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|.$$

Ce critère est une mesure similaire à la $RMSE$. Il est néanmoins plus robuste car moins sensible aux valeurs extrêmes que la MSE .

Toutes les mesures de distance (MSE , $RMSE$ et MAE) sont équivalentes et permettent de quantifier la proximité de la solution approchée par rapport aux données simulées. Une petite valeur de ces critères signifie que le modèle estimé est proche de la réalité.

Résidu maximal : RM . Le résidu maximal (correspondant à la norme L_∞) permet de vérifier si toutes les observations ont été correctement prédites. Il est défini par :

$$RM = \max_{1 \leq i \leq m} |y_i - \hat{y}_i|.$$

On peut également introduire le critère RMA (Relative Maximum Absolute Error) comme étant le résidu maximal standardisé :

$$RMA = \frac{\max_{1 \leq i \leq m} |y_i - \hat{y}_i|}{\sigma_y}.$$

Une grande valeur de ces critères indique une erreur importante d'approximation en un point du plan d'expériences même si la précision indiquée par le coefficient R^2 est satisfaisante. A noter que dans le cas de données normalisées, ces deux critères sont égaux.

Annexe C.

Table statistique de la statistique globale du radar d'uniformité

TABLE STATISTIQUE DE G_N : VALEURS DE g_N TELLES QUE $P = P(G_N < g_N)$.

N	P=0,80	P=0,85	P=0,90	P=0,95	P=0,99
1	1,94	1,97	1,98	1,99	1,99
2	2,52	2,82	3,08	3,33	3,47
3	2,68	3,00	3,30	3,63	3,84
4	2,80	3,12	3,42	3,80	4,07
5	2,86	3,19	3,50	3,88	4,12
6	2,89	3,25	3,58	3,97	4,26
7	2,93	3,28	3,59	4,01	4,28
8	2,94	3,29	3,64	4,03	4,31
9	2,95	3,30	3,61	4,00	4,35
10	2,97	3,34	3,66	4,06	4,33
11	2,98	3,34	3,68	4,13	4,46
12	2,99	3,33	3,67	4,08	4,34
13	3,00	3,36	3,70	4,11	4,40
14	3,00	3,36	3,72	4,11	4,41
15	3,00	3,34	3,68	4,12	4,43
16	3,00	3,36	3,70	4,10	4,43
17	3,03	3,38	3,70	4,13	4,49
18	3,01	3,35	3,68	4,07	4,39
19	3,04	3,41	3,75	4,15	4,50
20	3,04	3,41	3,74	4,15	4,47
21	3,04	3,40	3,73	4,17	4,48
22	3,05	3,41	3,74	4,12	4,42
23	3,05	3,39	3,73	4,16	4,47
24	3,06	3,42	3,73	4,16	4,42
25	3,05	3,40	3,72	4,15	4,48
26	3,06	3,43	3,74	4,11	4,46
27	3,05	3,40	3,72	4,11	4,44
28	3,06	3,41	3,76	4,18	4,49
29	3,06	3,43	3,75	4,17	4,47

30	3,05	3,42	3,76	4,18	4,49
31	3,06	3,40	3,72	4,11	4,41
32	3,07	3,44	3,79	4,27	4,58
33	3,06	3,40	3,71	4,12	4,43
34	3,05	3,40	3,71	4,13	4,41
35	3,07	3,42	3,76	4,17	4,42
36	3,08	3,43	3,76	4,16	4,47
37	3,09	3,46	3,78	4,21	4,53
38	3,08	3,45	3,78	4,16	4,47
39	3,07	3,44	3,77	4,23	4,53
40	3,06	3,42	3,75	4,17	4,45
41	3,08	3,43	3,76	4,16	4,53
42	3,09	3,43	3,76	4,18	4,47
43	3,08	3,44	3,75	4,21	4,53
44	3,10	3,45	3,76	4,15	4,51
45	3,09	3,44	3,77	4,17	4,44
46	3,07	3,43	3,75	4,15	4,52
47	3,07	3,43	3,78	4,22	4,51
48	3,07	3,42	3,75	4,15	4,41
49	3,10	3,44	3,78	4,21	4,50
50	3,08	3,44	3,77	4,21	4,53
51	3,09	3,43	3,77	4,18	4,47
52	3,11	3,45	3,79	4,22	4,54
53	3,09	3,46	3,78	4,15	4,48
54	3,10	3,45	3,81	4,24	4,54
55	3,09	3,44	3,76	4,20	4,50
56	3,10	3,44	3,79	4,19	4,52
57	3,09	3,46	3,79	4,19	4,48
58	3,09	3,45	3,78	4,20	4,49
59	3,11	3,44	3,78	4,18	4,52
60	3,08	3,44	3,76	4,17	4,45
61	3,11	3,47	3,82	4,20	4,54
62	3,10	3,47	3,80	4,19	4,48
63	3,09	3,45	3,78	4,19	4,46
64	3,09	3,47	3,79	4,23	4,53
65	3,10	3,47	3,80	4,22	4,54
66	3,11	3,46	3,78	4,16	4,45
67	3,10	3,45	3,79	4,23	4,52
68	3,09	3,43	3,77	4,15	4,44
69	3,09	3,44	3,77	4,18	4,44
70	3,10	3,46	3,78	4,21	4,50
71	3,11	3,46	3,77	4,17	4,49
72	3,10	3,44	3,78	4,18	4,42
73	3,09	3,45	3,79	4,18	4,54

74	3,11	3,47	3,80	4,17	4,46
75	3,10	3,44	3,77	4,21	4,47
76	3,11	3,48	3,80	4,22	4,51
77	3,09	3,43	3,78	4,19	4,44
78	3,11	3,45	3,78	4,21	4,48
79	3,08	3,44	3,78	4,20	4,46
80	3,10	3,46	3,80	4,21	4,52
81	3,11	3,48	3,79	4,19	4,52
82	3,11	3,49	3,83	4,25	4,51
83	3,11	3,46	3,78	4,17	4,44
84	3,11	3,45	3,77	4,18	4,51
85	3,09	3,46	3,77	4,23	4,55
86	3,10	3,46	3,78	4,18	4,49
87	3,08	3,44	3,76	4,15	4,44
88	3,11	3,47	3,77	4,21	4,48
89	3,10	3,47	3,79	4,17	4,46
90	3,08	3,43	3,74	4,15	4,44
91	3,07	3,43	3,76	4,14	4,43
92	3,09	3,46	3,78	4,19	4,52
93	3,09	3,46	3,81	4,23	4,53
94	3,08	3,43	3,74	4,18	4,52
95	3,09	3,44	3,77	4,21	4,54
96	3,11	3,45	3,76	4,19	4,52
97	3,12	3,45	3,78	4,20	4,47
98	3,08	3,44	3,79	4,21	4,46
99	3,09	3,45	3,77	4,21	4,50
100	3,10	3,44	3,79	4,23	4,51

Annexe D.

Distance à densité et plans d'expériences associés

Pour le calcul, on utilise le fait que $s \rightarrow \phi(x_s)$ est une densité gaussienne 1D à une constante près.

Précisément, il faut écrire que

$$\begin{aligned} \langle x_s, \Gamma^{-1}x_s \rangle &= \langle a - \mu, \Gamma^{-1}(a - \mu) \rangle + 2s \langle a - \mu, \Gamma^{-1}u \rangle + s^2 \langle u, \Gamma^{-1}u \rangle \\ &= \frac{\left(s + \frac{\langle a - \mu, \Gamma^{-1}u \rangle}{\langle u, \Gamma^{-1}u \rangle} \right)^2}{\sigma_u^2} + \rho^2 \end{aligned}$$

$$\text{où } \left\{ \begin{array}{l} \rho^2 = \langle a - \mu, \Gamma^{-1}(a - \mu) \rangle - \frac{\langle a - \mu, \Gamma^{-1}u \rangle^2}{\langle u, \Gamma^{-1}u \rangle} \geq 0 \\ \sigma_u^2 = \frac{1}{\langle u, \Gamma^{-1}u \rangle} > 0 \end{array} \right.$$

On en déduit

$$\delta(A, B) = \frac{\sqrt{2\pi}}{\sqrt{2\pi}^d} \frac{1}{\sqrt{\det \Gamma}} e^{-\frac{\rho^2}{2}} \int_0^l \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma_u^2} \left(s + \frac{\langle a - \mu, \Gamma^{-1}u \rangle}{\langle u, \Gamma^{-1}u \rangle} \right)^2} ds$$

Avec le changement de variable suivant $t = s + \frac{\langle a - \mu, \Gamma^{-1}u \rangle}{\langle u, \Gamma^{-1}u \rangle}$ où $dt = \frac{ds}{\sigma_u}$, on trouve

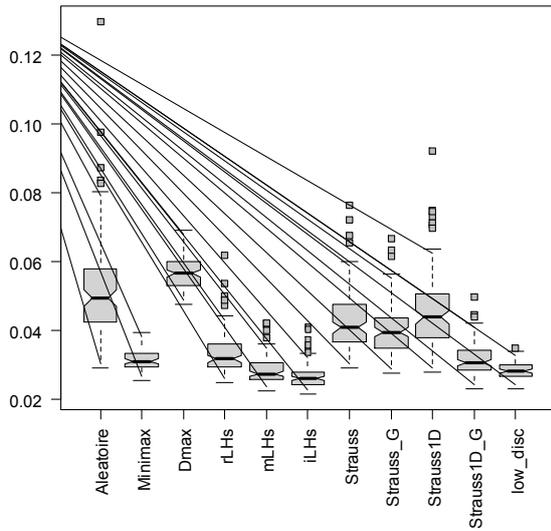
$$\delta(a, b) = k \times e^{-\rho^2/2} \times \sigma_u \times \left[\phi \left(\sigma_u \langle a - \mu, \Gamma^{-1}u \rangle + \frac{l}{\sigma_u} \right) - \phi \left(\sigma_u \langle a - \mu, \Gamma^{-1}u \rangle \right) \right]$$

avec $\phi(z) = \int_{-\infty}^z \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt$.

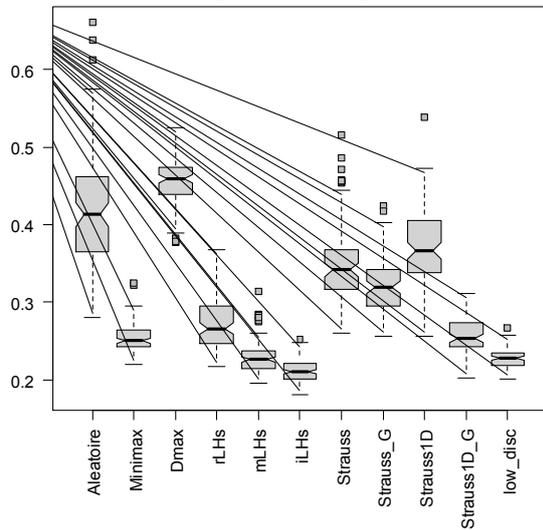
Annexe E.

Résultats des critères intrinsèques pour les plans en dimensions 3 & 10

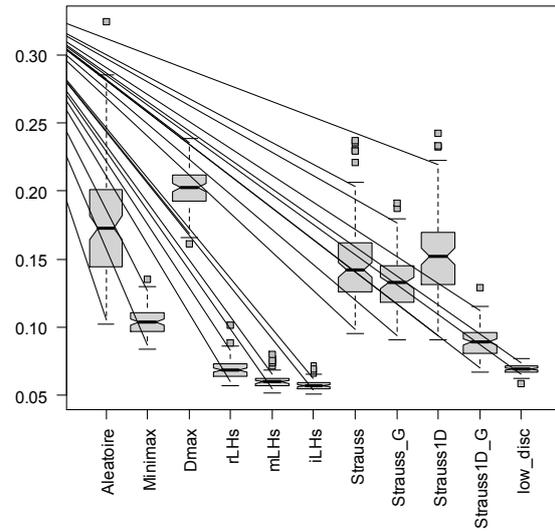
Résultats des plans en dimension 3.



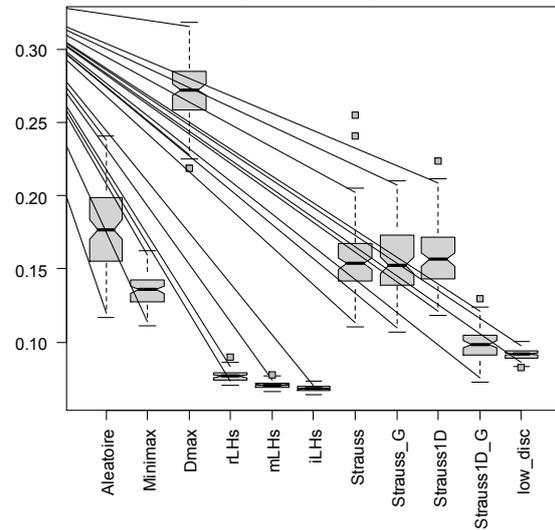
Discrépance à l'origine



Discrépance symétrique



Discrépance modifiée



Discrépance « enveloppée »

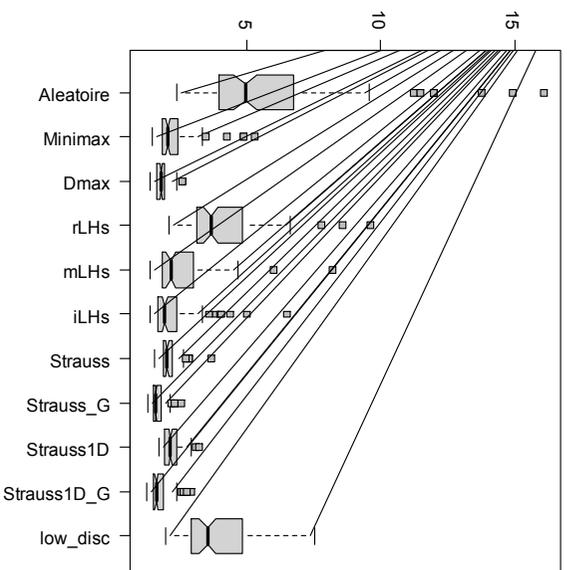
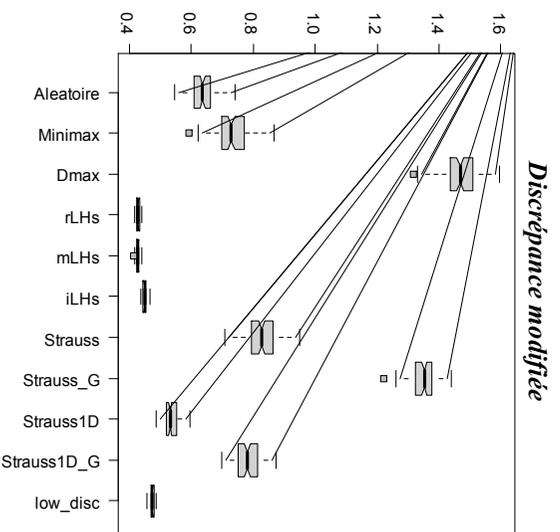
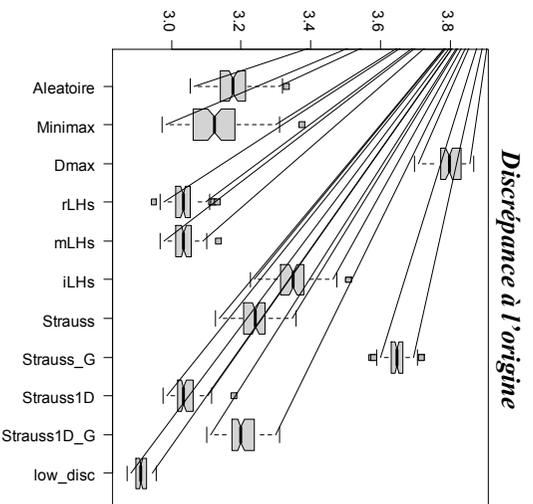
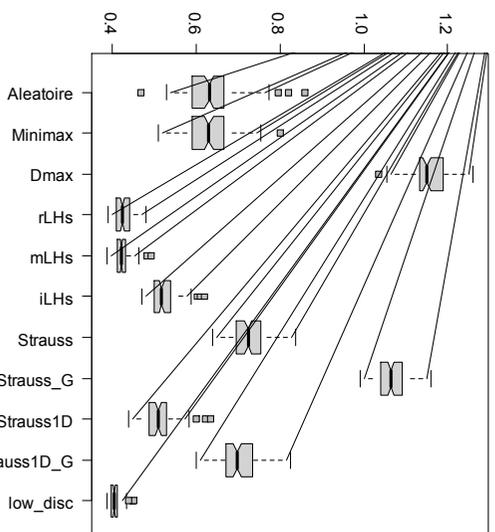
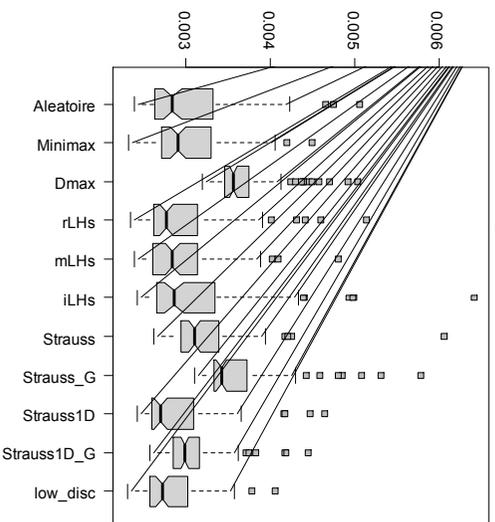


Figure 144. Boxplot du critère de meshratio sur les plans à 30 points en 3D.

Résultats des plans en dimension 10.



Discrepance symétrique

Discrepance « enveloppée »

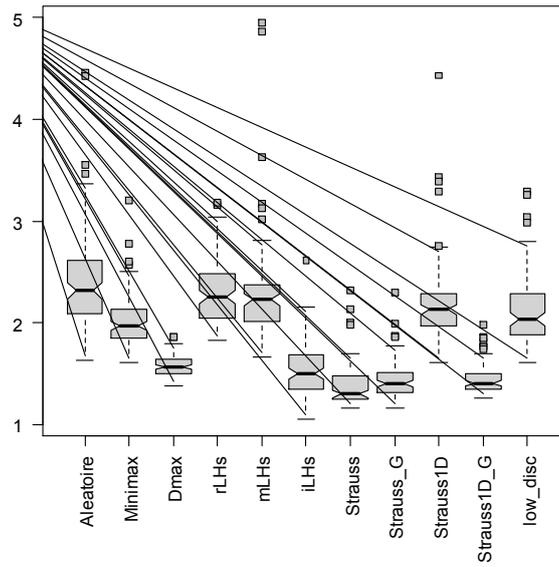


Figure 145. Boxplot du critère de meshratio sur les plans à 100 points en 10D.

Annexe F.

Critères extrinsèques sur le cas 10D

Plans	Critères	R^2		$RMSE$		MAE		RMA	
		\nearrow	\searrow	\nearrow	\searrow	\searrow	\searrow	\searrow	\searrow
Aleatoires	moyenne	0,932	0,253	0,716	0,523	0,355	3,691		
	écart-type	0,021	0,032	0,108	0,101	0,063	1,096		
Dmax	moyenne	0,961	0,261	0,873	0,351	0,288	1,669		
	écart-type	0,013	0,038	0,047	0,062	0,055	0,436		
Minimax	moyenne	0,907	0,220	0,427	0,735	0,465	5,623		
	écart-type	0,028	0,031	0,379	0,181	0,072	2,439		
rLHs	moyenne	0,919	0,278	0,750	0,493	0,336	3,641		
	écart-type	0,026	0,032	0,088	0,083	0,050	1,125		
mLHs	moyenne	0,918	0,278	0,705	0,527	0,353	3,923		
	écart-type	0,024	0,031	0,207	0,132	0,055	2,007		
iLHs	moyenne	0,925	0,275	0,482	0,574	0,369	4,449		
	écart-type	0,024	0,030	2,243	0,436	0,093	5,828		
Strauss	moyenne	0,951	0,254	0,844	0,377	0,264	2,758		
	écart-type	0,016	0,035	0,168	0,118	0,049	2,103		
Strauss-Gibbs	moyenne	0,963	0,251	0,892	0,324	0,257	1,786		
	écart-type	0,015	0,039	0,041	0,058	0,047	0,595		
Strauss1D	moyenne	0,931	0,273	0,802	0,439	0,305	3,232		
	écart-type	0,021	0,032	0,072	0,075	0,048	1,276		
Strauss-Gibbs 1D	moyenne	0,937	0,265	0,533	0,488	0,302	4,238		
	écart-type	0,017	0,028	2,498	0,481	0,091	7,744		

Figure 146. Valeurs (moyenne et écart-type) des critères de qualité du modèle pour chaque classe de plans

Annexe G.

Résultats numériques du cas 8D

DF for Terms and F-values for Nonparametric Effects

	Df Npar	Df Npar	F	Pr(F)	
(Intercept)	1				
s(X1)	1	3	1.2286	0.30981	
s(X2)	1	3	1.2054	0.31813	
s(X3)	1	3	1.7594	0.16784	
s(X4)	1	3	0.7266	0.54122	
s(X5)	1	3	0.9592	0.41983	
s(X6)	1	3	1.2549	0.30065	
s(X7)	1	3	2.2667	0.09299	.
s(X8)	1	3	3.5168	0.02212	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 1. Table d'analyse de la déviance pour le modèle additif complet

DF for Terms and F-values for Nonparametric Effects

	Df Npar	Df Npar	F	Pr(F)	
(Intercept)	1				
X1	1				
X2	1				
s(X3)	1	3	2,1500	0,102954	
X4	1				
X5	1				
X6	1				
s(X7)	1	3	3,5416	0,019577	*
s(X8)	1	3	4,4603	0,006694	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 2. Table d'analyse de la déviance pour le modèle additif mixte

Coefficients:					
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-0.011	0.058	-0.196	0.845291	
X1	0.507	0.088	5.781	1.83e-07	***
X2	0.393	0.088	4.462	2.98e-05	***
X3	0.668	0.086	7.732	5.24e-11	***
X4	-0.814	0.089	-9.099	1.54e-13	***
X5	-0.229	0.092	-2.490	0.015116	*
X6	-0.357	0.095	-3.754	0.000353	***
X7	0.237	0.085	2.783	0.006894	**
X8	-0.310	0.089	-3.499	0.000811	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 0.5118 on 71 degrees of freedom					
Multiple R-Squared: 0.7646, Adjusted R-squared: 0.738					
F-statistic: 28.82 on 8 and 71 DF, p-value : <2.2e-16					

Table 3. Table d'analyse de la variance pour le modèle linéaire à effets simples

	Variables	p-valeur		Variables	p-valeur
1	X4	< 2e-16	12	X3:X7	0,01929
2	X3	1,29E-13	13	X4:X6	0,028459
3	X1	3,35E-11	14	X5	0,033088
4	X2	0,00000368	15	X2:X6	0,037653
5	X6	0,0000106	16	X3:X6	0,060883
6	X3:X4	0,0000667	17	X5:X7	0,066969
7	X7	0,000193	18	X2:X5	0,072736
8	X3:X5	0,001367	19	X1:X8	0,097857
9	X8	0,002047	20	X6:X7	0,124237
10	X1:X2	0,005325	21	X4:X8	0,124613
11	X5:X8	0,016756	22	(Intercept)	0,544704

Tableau 18. Variables sélectionnées triées par ordre d'importance et leur p-valeur

Fonctions de base	Coefficients
$(X_6 - 0,500)^+$	-1,380
$(-0,307 - X_4)^+$	1,213
$(X_4 - 0,077)^+$	-1,129
$(X_3 + 0,263)^+$	0,979
$(X_7 - 0,360)^+$	0,936
$(0,357 - X_1)^+$	-0,735
(Intercept)	-0,617
$(0,154 - X_8)^+$	0,570
$(X_2 + 0,565)^+$	0,385
$(0,600 - X_5)^+$	0,237

Tableau 19. Récapitulatif du modèle MARS de degré 1 ajusté

	θ_{est}		μ	σ^2	-2Log(L)
P2	3,64	4,39	-0,05	1,40	145,81
P5	1,96	2,47			
I2	2,79	2,17			
I3	1,39	2,51			

Tableau 20. Krigeage ordinaire avec le programme R

Deux exemples de modèles de KU :

	θ_{est}		Tendance β		σ^2	-2Log(L)	Critères
-0.003							
P2	10	10	0.48	97.01	0.25	97.01	
P5	0.01	10	0.74	-0.78			RMSE 0.66
I2	1.85	0.71	-0.29	-0.34			MAE 0.52
I3	0.44	10	0.31	-0.22			RMA 2.49

Tableau 21. Krigeage universel avec tendance linéaire (programme R)

	θ_{est}		Tendance β		σ^2	-2Log(L)	Critères
0.02							
P2	1.08	10	0.53	0.39	0.27	102.60	
P5	10	10	0.69	-0.84			RMSE 0.65
I2	0.25	10	-0.26	-0.34			MAE 0.52
I3	10	0.06	0.30	-0.30			RMA 1.99

Tableau 22. Krigeage universel avec tendance linéaire (programme R)

BIBLIOGRAPHIE

- Beirlant J., Dudewicz E.J., Györfi L., Van Der Meulen E.C. (1997). Nonparametric entropy estimation : an overview. *Int. J. Math. Stat. Sci.*, **6**(1) 17-39.
- Box G.E.P., Behnken D.W. (1960). Some New Three-Level Designs for the Study of Quantitative Variables, *Technometrics*, **2**, 455-475.
- Braaten E. et Weller G. (1979). An improved low-discrepancy sequence for multidimensional quasi-Monte-Carlo integration, *J. Comput. Phys.* **33**, 249–258.
- Broyden, C.G. (1970). The Convergence of a Class of Double-rank Minimization Algorithms. *Journal of Institute of Mathematics and Its Applications*, **6**, 76-90.
- Bundschuh P., Zhu Y. (1993). A method for exact calculation of the discrepancy of low-dimensional finite points sets I, *Abh. Math. Sem. Univ. Hamburg* **63**, 115-133.
- Byrd, R. H., Lu, P., Nocedal, J. and Zhu, C. (1995) A limited memory algorithm for bound constrained optimization. *SIAM J. Scientific Computing*, **16**, 1190–1208.
- Chaloner K., Verdinelli I. (1995). Bayesian Experimental Design: A review. *Statistical Science* **10**, 237-304.
- Chen V.C.P., Tsui K.L., Barton R.R, Allen J.K. (2003). A review of design and modeling in computer experiments. *Handbook of Statistics* **22**, 231-261.
- Cormen T.H., Leiserson C.E., Rivest D.L., Stein C. (2002). *Introduction à l'algorithmique*. Edition Sciences Sup., Dunod.
- Corre B. (1987). Cantidad de vapor optima economica a inyectar en forma alternada, *Simposio internacional SRM 2113*, **2**, 763-788.
- Currin C., Mitchell T., Morris M. Ylvisaker D. (1991). Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments. *Journal of the American Statistical Association*, **86**, 953-963.
- D'agostino R.B., Stephens M.A. (1986). *Goodness-of-fit Techniques*. Marcel Dekker, New-York.
- Daley D.J., Vere-Jones D. (2003). *An Introduction to The Theory of Point processes. Elementary Theory and Methods*, 2nd edn, New York, Springer, I 4.
- De Clerck L. (1986). A method for exact calculation of the star discrepancy of plane sets applied to the sequences of Hammersley. *Monatsh. Math.* **101**, 261-278.
- Dobkin D.P., Eppstein D. (1993). Computing the discrepancy. *Proceedings of the ninth annual symposium on computational geometry*, 47-52.
- Doehlert D.H. (1970). Uniform shell designs, *Applied Statistics*, **19**, 231-239.
- Doob J.L. (1953). *Stochastic Processes*, John Wiley & Sons.
- Droesbeke J.J., Fine J., Saporta G. (1994). *Plans d'expériences, Applications à l'entreprise*. Ed. TechnipTechnip.
- Dussert C., Rassigni G., Rassigni M., Palmari J. (1986). Minimal spanning tree: A new approach for studying order and disorder. *Physical Review B.*, **34**, 5, 3528-3531.
- Elias S.W., Shiu W. (1987). Convolution of Uniform Distributions and Ruin Probability. *Scandinavian Actuarial*, 191-197.
- Fang K.-T., Li R., Sudjianto A. (2006). *Design and Modeling for Computer Experiments*. Chapman & Hall.

- Fang K.T., Wang Y., Bentler P. (1994). Some applications of number of theoretic methods in statistics. *Statist. Science*, **9**, 419-428.
- Fang K.T., Zhu.L.X. (1993). On uniformly random design. Technical report. **math-9**, Hong Kong Baptist college.
- Faure H. (1980). Suite à faible discrédance dans T^s , *Publ. Dép. Math.*, Université de Limoges, Limoges, France.
- Faure H. (1982). Discrédance de suites associées à un système de numération (en dimension s), *Acta Arith.* **41**, 337–351.
- Faure H. (1981). Discrédance de suites associées à un système de numération (en dimension un), *Bull. soc. Math. France*, **109**, 143-182.
- Fedorov V.V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.
- Fisher R.A. (1926). The arrangement of field experiments. *J. Ministry Agric.* **33**, 503-513.
- Fletcher, R. (1970). A New Approach to Variable Metric Algorithms, *Computer Journal*, **13**, 317-322.
- Friedman J. Multivariate Adaptative Regression Splines (invited paper). (1991) *Annals of Statistics*, 10 1-141.
- Goldfarb, D. (1970). Family of Variable Metric Updates Derived by Variational Means, *Mathematics of Computation*, **24**, 23-26.
- Gunzburger M., Burkardt J. (2004). Uniformity measures for point samples in hypercubes. www.csit.fsu.edu/~burkardt/pdf/ptmeas.pdf.
- Halton J. H. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, *Numer. Math.* **2**, 84–90.
- Hammersley J.M. (1960). Monte-Carlo methods for solving multivariate problems, *Annals of the New York Academy of Sciences* **86**, 844–874.
- Hammersley J.M., Halton J.H., Bearwood J. (1959). The shortest path through many points, *Camb. Philos. Soc. Proc.*, **55**, 299-327.
- Hastie T., Tibshirani R., Friedman J. (2001). *The Elements of Statistical Learning*, Springer.
- Hastings W.K. (1970). Monte-Carlo Sampling methods using Markov chains and their applications, *Biometrika*, **57**, 97-109.
- Hickernell F.J. (1998). A generalized discrepancy and quadrature error bound. *Mathematics of Computation*, **67**, 299-322.
- Hlawka E. (1961). Funktionen von Beschränkter Variation in der Theorie der Gleichverteilung. *Annali di Matematica Pura Ed Applicata*, **54**, 325-333.
- Jin R., Chen W., Simpson T.W. (2001). Comparative studies of metamodelling techniques under multiple modelling criteria. *Struct Multidisc Optim*, Springer-Verlag, **23** 1-13.
- Johnson M.E., Moore L.M., Ylvisaker D. (1990). *Minimax and maximin distance designs*. *J. of Statis. Planning and Inference* **26**, 131-148.
- Jourdan A. (2000). Analyse statistique et échantillonnage d'expériences simulées, Thèse de l'Université de Pau et des Pays de l'Adour.
- Kennard R.W., Stone L.A. (1969). Computer aided design of experiment. *Technometrics*, **11** (1), 137-148.

- Kenny Q.Y. (1998). Orthogonal Column Latin Hypercube and their application in computer experiments. *J. Amer. Statistic. Assoc.*, **93**, 1430-1439.
- Knuth D.E. (1969). *The Art of Computer Programming*. Addison-Wesley, New-York, Vol. 2.
- Koehler J.R. et Owen A.B. (1996). Computer Experiments, *Handbook of Statistics*, **13**, 261-308.
- Kocis L. and. Whiten W. J (1997). Computational investigations of low-discrepancy sequences, *ACM Transactions on Mathematical Software* **23**, No. 2, 266–294.
- Kooperberg C., Bose S., Stone C.J. (1997). Polychotomous Regression, *Journal of the American Statistical Association*, **92**, 117-127.
- Kruskal J.B. (1956). On the shortest spanning subtree of a graph and travelling salesman problem, *Proc. Amer. Math. Soc.*, **7**, 48-50.
- Kullback S., Leibler R.A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22** 79-86.
- Lindley D.V. (1956). On a Measure of the Information Provided by an Experiment. *The Annals of Mathematical Statistics*, **27**, 986-1005.
- Lewis G.A., Mathieu D., Phan-Tan-Luu R. (1999). *Pharmaceutical Experimental Design*. Marcel Dekker.
- Matheron G. (1963). Principles of geostatistics, *Econm. Geol.* **58**, 1246-1266.
- Mc Kay M.D., Conover W.J. Beckman R.J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, **21**, 239-245.
- Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H. and Teller E. (1953). Equation of state calculations by fast computing machines, *Journal of Chemical Physics*, **21**, 1087-1092.
- Meyn S.P., Tweedie R.L. (1993). *Markov Chain and Stochastic Stability*, New York, Springer
- Mitchell T.J. (1974). An algorithm for the construction of «D-optimal» experimental designs. *Technometrics*, **16** (2), 203-210.
- Mitchell T.J., Scott D.S. (1987). A computer program for the design of group testing experiments. *Communications in Statistics – Theory and methods*, **16**, 2943-2955.
- Morokoff W., Caflish R.E. (1994). Quasi-random sequences and their discrepancies. *SIAM J. Sci. Stat. Computing*, **15**, 1251-1279.
- Morris M.D., Mitchell T.J. (1992). Exploratory designs for computational experiments, ORNTL/TM-12045, Oak Ridge National Laboratory, Oak Ridge, TN.
- Mullen G.L., Mahalanabis A., Niederreiter H. (1995). Tables of (t,m,s)-net and (t,s)-sequence parameters. In Niederreiter H. and Shiue, P.J.-S., editors, *Monte-Carlo and Quasi-Monte-Carlo Methods in Scientific Computing*, 229-317, New-York. Springer Verlag.
- Myers R.H., Montgomery D.C. (1995). *Response Surface Methodology*, Wiley New-York.
- Neyman J., Scott E.L. (1958). Statistical approach to problems of cosmology, *J.R. Statist. Soc. B*, **20**, 1-43.
- Niederreiter H. (1972). Discrepancy and convex programming. *Ann. Math. Pur. Appl.* **93**, 89-97.

- Niederreiter H. (1987). Point sets and sequences with small discrepancy. *Monasth. Math.* **104**, 273-337.
- Niederreiter H. (1987). Low-Discrepancy and Low-Dispersion Sequences, *Journal of number theory*, **30**, 51-70.
- Niederreiter H. (1992). *Random number generation and quasi-Monte-Carlo methods*. CBMS-NFS, SIAM, Philadelphia.
- Ostrowski A.M. (1952). Two Explicit Formulae for the Distribution Function of the Sums of n Uniformly Distributed Independent Variables, *Arch. Math*, **3**, 451-459.
- Owen A.B. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statistica Sinica* **2**, 439-452.
- Owen A.B. (1994). Controlling correlations in Latin Hypercube samples. *J. Amer. Statist. Assoc.*, **89**, 1517-1522.
- Owen A.B. (1997). *Latin Supercube Sampling for Very High Dimensional Simulations*. Stanford University.
- Park J.S. (1994). Optimal Latin-hypercube designs for computer experiments. *Journal Statistical Planning and Inference*, **39**, 95-111.
- Pleming J.B., Manteufel R.D. (2005). Replicated Latin Hypercube Sampling. *AIAA 2005-1819*.
- Prim R.C. (1957). Shortest Connection networks and Some Generalizations. *Bell system Technical Journal*, **36**, 1389-1401.
- Ripley B.D., Kelly F.P. (1977). Markov point processes, *Journal of the London Mathematical Society*, **15**, 188-192.
- Royle J.A., Nychka D. (1998). An algorithm for the construction of spatial coverage designs with implementation in Splus. *Computer & Geosciences*, **24**, 5, 479-488.
- Sacks J., Schiller S.B., Welch W.J. (1989). Designs for Computer Experiments, *Technometrics* **31**, 41-47.
- Sacks J., Welch W.J, Mitchell T.J., Wynn H.P (1989). Design and Analysis of Computer Experiments, *Statistical Science*, **4**, 409-435.
- Saltelli A., Chan K., Scott E.M. (2000). *Sensitivity Analysis*, Wiley, New-York.
- Santner T.J., Williams B.J., Notz W.I. (2003). *The Design and Analysis of Computer Experiments*, Springer, 121-161.
- Scheidt C. (2006). *Analyse statistique d'expériences simulées : Modélisation adaptative de réponse non-régulières par krigeage et plans d'expériences*. Université Louis Pasteur Strasbourg.
- Sebastiani P., Wynn H.P. (2000) Maximum entropy sampling and optimal Bayesian experimental design. *J. of Royal Statist. Soc.*, **62**, 145-157.
- Sergent M. (1989). *Contribution de la Méthodologie de la Recherche Expérimentale à l'élaboration de matrices uniformes : Application aux effets de solvants et de substituants*, Ph D Thesis, Université Aix Marseille III
- Sergent M, Phan-Tan-Luu R., Elguero J. (1997) Statistical Analysis of Solvent Scales. Part 1, *Anales de Quimica Int. Ed.*, **93**, 3-6.

- Shanno D.F. (1970). Conditioning of Quasi-Newton Methods for Function Minimization, *Mathematics of Computation*, **24**, 647-656.
- Shannon C.E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, **27**, 379-423, 623-656.
- Shewry M.C., Wynn H.P. (1987). Maximum Entropy Sampling. *Journal of Applied Statistics*, **14**, 165-170.
- Simpson T.W., Lin D.K.J., Chen W. (2001). Sampling strategies for computer experiments: design and analysis, *International Journal of Reliability and Applications*, 2(3),209-240.
- Sloan I.H., Joe S. (1994). *Lattice Methods for Multiple Integration*. Oxford Science Publications, Oxford.
- Sobol' I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals, *USSR Comput. Math. Math. Phys.* **7**, No. 4, 86–112.
- Sobol' I.M. and Levitan Y.L. (1976). The production of points uniformly distributed in a multidimensional cube. pp 40, Institute of Applied Mathematics, USSR Academy of Sciences.
- Stone C.J., Hansen M.H., Kooperberg C., Truong Y.K. (1997). Polynomial splines and their tensor products in extended linear modeling. *Ann. Statist.* **25**, 1371-1470.
- Strauss D.J. (1975). A model for clustering, *Biometrika*, **62**, 467-475
- Tan K.S. and Boyle P.P. (2000). Applications of randomized low discrepancy sequences to the valuation of complex securities, *Journal of Economic Dynamics & Control*.**24**, 1747-1782.
- Tang B. (1993). Orthogonal array-based latin hypercubes. *Journal of the American Statistical Association*, **88**, 1392-1397.
- Tang B. (1998). Selecting latin hypercubes using correlation criteria. *Statistica Sinica*, **8**, 965-977.
- Tarry G. (1901). Le problème de 36 officiers, *Compte-rendu de l'Assoc. Français Avanc. Sci. Naturel* **2**, 170-203.
- Thiérmard E. (2000). Sur le calcul et la majoration de la discrédance à l'origine. Thèse No 2259, Département de mathématiques, école polytechnique fédérale de Lausanne.
- Thiria S. Lechevallier Y., Gascuel O., Canu S. (1997). *Statistique et méthodes neuronales*, éditions Dunod : Discrimination, Prévision, Représentation structurée, Validation, Validité.
- Trosset M.W. (1999). Approximate maximin distance designs. In *ASA Proceedings of the Section on Physical and Engineering Sciences*, American Statistical Association, 223-227.
- Tuffin B. (1998). A new permutation choice in Halton sequences, in *Monte-Carlo and Quasi-Monte-Carlo Methods 1996*, Lecture Notes in Statistics 127 (H. Niederreiter, P. Hellekalek, G. Larcher, and P. Zinterhof, eds.), Springer, 427–435.
- Wallet F., Dussert C. (1998). Comparison of spatial point patterns and processes characterization methods, *Europhysics Lett.*, **42**, 493-498.
- Walter E., Pronzato L. (1997). *Identification of Parametric Models from Experimental Data*. Springer-Verlag.
- Wang Y. Fand K.T. (1981). A note on uniform distribution and experimental deign. *Kexue Tongba*, **26**, 485-489.

Wooton R., Cranfield R., Sheppey G.C., Goodford P.J. (1975). Physicochemical activity relationships in practice. 2. Rational selection of benzenoid substituents. *Journal Medical Chem* 18, 607-61.

3DSL® <http://www.streamsim.com> (juin 2008)

**Ecole Nationale Supérieure des Mines
de Saint-Etienne**

N° d'ordre : 489 MA

Jessica Franco

Title of thesis

Exploratory Designs for Computer Experiments of Complex Physical Systems Simulation

Speciality

Applied Mathematics

Keywords

Computer Experiments; Experimental Designs; Space-Filling Designs; Uniformity Criteria

Summary

Computer experiments are always more used for complex physical systems simulation. Such problems, generally very large dimension, require sophisticated and time-consuming computer codes. In this context, it becomes difficult to use the simulator for each study. A possible approach consists in planning a limited number of simulations according to an experimental design. From this a metamodel may be fitted, which provides a simplified version of the simulator.

In this work, we focus on exploratory designs, i.e. we assume that the relationship between the response and inputs is unknown, generated with deterministic simulators. Therefore, designs should allow one to fit a variety of models and should provide information about all portions of the experimental region. If we expect the response to depend on only a few of the inputs, then it is preferable that points are evenly spread across the projection onto these factors.

The objective of this work is to suggest on one hand criteria quantifying the designs intrinsic quality and on other hand mechanics which allow to meet the objectives.

Two new criteria were developed in this work. A graphic tool called “uniformity radar” is proposed in order to assess the points uniformity across the projection onto 2D and 3D sub-spaces. The second tool, based on the minimal spanning tree and in shape of cartography, allows to decide on which typology the distribution is similar.

The key point of this work is the use of Markov chains Monte Carlo methods to build experimental designs. We take into account the Strauss process and more generally Gibbs process (with repulsive interactions) to meet the requirements (space-filling and good projections). These new experimental designs are readily adaptable, several extensions are considered such as the heterogeneity, the constrained designs, etc.)

Finally, a comparison of designs is made in industrial context to show the interest of this work.

**Ecole Nationale Supérieure des Mines
de Saint-Etienne**

N° d'ordre : 489 MA

Jessica Franco

Titre de la thèse :

Planification d'expériences numériques en phase exploratoire pour la simulation des phénomènes complexes

Spécialité

Mathématiques Appliquées

Mots clefs

Plans d'expériences numériques, Space-Filling Designs, Critères d'uniformité

Résumé

Depuis quelques années, la simulation numérique modélise des phénomènes toujours plus complexes. De tels problèmes, généralement de très grande dimension, exigent des codes de simulation sophistiqués et très coûteux en temps de calcul (parfois plusieurs jours).

Dans ce contexte, le recours systématique au simulateur devient illusoire. L'approche privilégiée consiste à définir un nombre réduit de simulations organisées selon un plan d'expériences numériques et d'ajuster, à partir de celui-ci, une surface de réponse pour approcher le simulateur.

Dans le cadre de ces travaux, nous nous sommes intéressés aux plans d'expériences générés à partir de simulateurs déterministes et en phase exploratoire *i.e.* lorsque la dépendance entre les entrées et les sorties est a priori inconnue. Ainsi, les plans élaborés requièrent certaines propriétés comme le remplissage de l'espace afin de capter les non linéarités, ou bien encore la bonne répartition des points en projection dans le cas où le code ne dépendrait que de quelques variables influentes.

Le but de cette étude est de proposer d'une part des indicateurs quantifiant la qualité intrinsèque des plans et d'autre part des méthodes de planification d'expériences qui permettent de répondre aux objectifs souhaités dans notre contexte.

Concernant le premier point, deux nouveaux critères ont été développés dans ce travail. Un outil graphique appelé « radar d'uniformité » est proposé dans le but d'étudier l'uniformité des points en projection sur les sous-espaces de dimension 2 et 3. Un deuxième outil, basé sur les arbres de longueur minimale et s'apparentant à une cartographie permet de décider de quelle typologie se rapproche le plan étudié.

Le point essentiel de ce travail concerne une nouvelle méthode de planification d'expériences basée sur la simulation des échantillons selon des lois de probabilité par une méthode de Monte Carlo par chaînes de Markov (McMC). Les processus ponctuels de Strauss et, plus généralement, de Gibbs (avec interactions de type répulsion) ont été envisagés afin de répondre aux exigences souhaitées (remplissage de l'espace et en projection). Ces nouveaux plans étant facilement modulables, plusieurs extensions sont considérées telles que l'hétérogénéité, les plans contraints, etc.

Enfin, une comparaison des plans est effectuée dans un contexte applicatif afin de montrer l'intérêt de ce travail.