

الجمهورية الجزائرية الديمقراطية الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



Université Blida1

Faculté des Sciences
Département de Mathématiques



Mémoire de Fin d'Etude

En vue de l'obtention du diplôme de

Master En Mathématiques

Spécialité : Modélisation Stochastique et Statistique

Thème

L'IMPACT DE DEUX TYPES DE REGRESSION SUR UN PLAN D'EXPERIENCES NUMERIQUES

Mémoire présenté par

El-Airech Meriem et Hamiche Halima

Devant le jury composé de :

S.Kerdjoudj	Université Blida1	Président
M.Raouti	Université Blida1	Examineur
H. Elmoassaoui	Maître de conférences, U. de Blida 1	Rapporteur

ملخص

على الرغم من التقدم في أدوات الكمبيوتر ، تظل الدراسة الرقمية لبعض الظواهر الفيزيائية المعقدة مكلفة للغاية من حيث وقت الحساب . لذلك من الضروري اللجوء الى طرق التخطيط التجريبية. تقدم هذه الأطروحة وتحلل مقارنة بين نهجين الانحدار، الأكثر استخدامًا في منهجية التصميمات العددية للتجارب: الانحدار متعدد الخطوط والانحدار بواسطة Kriging. وأخيرًا ، قمنا بتنفيذ تطبيق MATLAB لمقارنة هذين النهجين .

الكلمات المفتاحية: تصميم التجارب ، سطح الاستجابة ، Kriging ، الانحدار متعدد الخطوط ، تصاميم التجارب الرقمية .

RESUME

Malgré les progrès des outils informatiques, l'étude numérique de certains phénomènes physiques complexes reste très coûteuse en temps de calcul. Il est donc nécessaire de faire appel à des méthodes de planification d'expériences. Ce mémoire présente et analyse une comparaison entre les deux approches de régression, les plus utilisées dans la méthodologie des plans d'expériences numérique: la Régression multilinéaire et la Régression par Krigeage. Enfin, nous réalisons une application en Matlab pour comparer ces deux approches.

Mots clés : Plan d'expériences, Surface de réponse, Krigeage, Régression multilinéaire, Plans d'expériences numériques.

ABSTRACT

Despite the progress in computer tools, the digital study of certain complex physical phenomena remains very costly in terms of computation time. It is therefore necessary to appeal to experimental planning methods. This thesis presents and analyzes a comparison between the two regression approaches, the most used in the methodology of numerical designs of experiments: multilinear Regression and Regression by Kriging . Finally, we carry out an application in MATLAB to compare these two approaches.

Key words : Expérimental designs ,Response surface , Kriging ,Multilinear Regression ,Digital experiment designs.

Remerciements

Nous tenons avant toute chose, à remercier **DIEU** et à lui rendre grâce pour la force et le courage qu'il nous a donné.

Ainsi nous tenons également à exprimer nos vifs remerciements à notre encadreur Monsieur Elmossaoui Hicham pour avoir d'abord proposé ce thème, pour le suivi continué tout le long de la réalisation de ce mémoire et qui n'a pas cessé de nous donner ses conseils.

Nous voudrions aussi remercier nos familles, pour le soutien et l'aide inconditionnel qu'elle nous apporte chaque jour.

Et je remercie mon binôme pour son aide continue et son souci de travailler parfaitement.

Nous adressons nos sincères remerciements et gratitude aux membres de Jury pour nous avoir consacré leur précieux temps.

Nous tenons à remercier vivement toutes les personnes qui nous ont aidés pour réaliser ce mémoire surtout nos amies Maroua et Houria , ainsi qu'à tous ceux qui nous ont aidé de près ou de loin à accomplir ce travail.

Merci à tous

Dédicaces

A Papa et maman, merci pour votre amour et votre soutien.

A tous ceux qui m'ont enseigné et surtout à Monsieur **ELMOSSAOUI**.

A mes chers frères et chères sœurs. .

A mes camarades.

A tous ceux qui ont contribué à mon bonheur.

HAMICHE HALIMA

Dédicaces

A Papa et maman, merci pour votre amour et votre soutien.

A tous ceux qui m'ont enseigné et surtout à. Monsieur **ELMOSSAOUI**.

A mes chers frères et chères sœurs et leur enfants. .

A mes camarades et mes amis et mon amie d'enfance Maroua.

A tous ceux qui ont contribué à mon bonheur.

ELAIRECH MERIEM

TABLE DES MATIERES

RESUME	
REMERCIEMENTS	
TABLE DES MATIERES	
LISTE DES ILLUSTRATIONS, GRAPHIQUES ET TABLEAUX	
INTRODUCTION	7
1. Les plans d'expériences	9
1.1. Les plans d'expériences classiques	9
1.1.1 Intérêt des plans d'expériences et leur utilisation	9
1.1.2 Vocabulaire de base	9
1.1.3. Réalisation des essais	10
1.1.4. Matrice d'expériences	10
1.1.5. Variable centré réduite	11
1.1.6. Différents types de plan d'expériences usuels	11
1.1.7. Pour quoi utilise-t-on encore des plans d'expériences classiques en Simulation ?	15
1.2 Les plans d'expériences numériques	17
2. Notion de modélisation mathématique	28
2.1. Modèle mathématique pour plans d'expérience	28
2.2. Modèle pour plan de mélange	30
2.3. Méthodes Mars et Polymars	37
2.4. Réseaux de Neurones	39
3. Régression Multilinéaire	40
3.1. Le principe de la Régression Multilinéaire	40
3.2. Hypothèses de la Régression Multilinéaire	42
3.3. Estimation des coefficients par la méthode des Moindres Carrés	42
3.4. Propriétés des coefficients	45
3.5. Analyse de la variance	46
3.6. Tests statistiques	49
4. Description du modèle par Krigage	52
4.1. Démarche générale de résolution des équations du Krigage	52
4.2. Krigage universel	54
4.3. Krigage avec dérivé externe	54

4.4. Lien entre le Krigeage avec modèle de tendance et le Krigeage sur les résidus d'une Régression	58
4.5. Problème de l'analyse Variographique en Krigeage avec modèle de tendance	59
4.6. Le Variogramme	60
4.7. Application : Etude comparative	60
CONCLUSION	65
APPENDICE	66
REFERENCES	72

LISTE DES ILLUSTRATIONS, GRAPHIQUES ET TABLEAUX

Figure 1.1	Domaine de variation du facteur	10
Figure 1.2	Plan composite pour l'étude de deux facteurs	14
Figure 1.3	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	24
Figure 1.4	Les 100, 1000, 10000 premiers points de la suite de halton en base 2 et 3	26
Figure 1.5	Les 100, 1000 et 10000 premiers points d'une suite de Sobol	27
Figure 1.6	Les 50, 250 et 250 premiers points d'une suite de Faure	27
Figure 2.1	Modélisation par un polynôme d'ordre un	31
Figure 2.2	Modélisation par un polynôme d'ordre deux	32
Figure 2.3	Modélisation par un polynôme d'ordre trois	33
Figure 2.4	Modélisation par un polynôme synergique d'ordre trois	34
Figure 3.1	Droite des moindres carrés	41
Figure 4.1	Boxplots de critère R^2 calculé sur 40 plans à 30 points en 3 dimensions et à 40 points e 5 dimensions	62
Figure 4.2	Boxplots de critère R^2 ajustés calculé sur 40 plans à 30 points en 3 dimensions et à 40 points 5 dimensions	63
Figure 4.3	Boxplots de critère de Fisher calculé sur 40 plans à 30 points en 3 dimensions et à 40 points 5 dimensions	64
Tableau 1.1	Matrice d'expérience	11
Tableau 1.2	Matrice d'expériences $2^2 \times 3^1$	13
Tableau 3.1	Présentation des résultats d'essais pour une régression simple	40
Tableau 3.2	Présentation des résultats d'expériences Pour une régression multiple	41
Tableau 4.1	Résultats des critères d'évaluation pour 3 dimensions et 5 dimensions	61

INTRODUCTION

Les plans d'expériences permettent d'organiser au mieux les essais qui accompagnent une recherche scientifique ou des études industrielles. Ils sont applicables à de nombreuses disciplines et à toutes les industries à partir du moment où l'on recherche le lien qui existe entre une grandeur d'intérêt y et des variables x_i . Donc il faut penser aux plans d'expériences si l'on s'intéresse à une fonction du type :

$$y = f(x_i)$$

Avec les plans d'expériences on obtient le maximum de renseignements avec le minimum d'expériences. Pour cela, il faut suivre des règles mathématiques et adopter une démarche rigoureuse. Il existe de nombreux plans d'expériences classiques et numériques adaptés à tous les cas rencontrés par un expérimentateur.

La diversité des plans d'expériences proposés dans la littérature vient du fait qu'il n'existe pas un plan parfait où tous les critères d'optimalité sont vérifiés en même temps. Chaque plan présente des avantages pour un certain critère d'optimalité et un inconvénient pour un autre. Il s'agira donc de trouver un compromis pour les besoins de chaque étude.

L'emploi des plans d'expériences, pour l'étude empirique d'une loi de réponse, pose aux statisticiens ou aux chercheurs des problèmes particuliers. Alors qu'ils ont peu d'informations sur cette loi, ils ne peuvent disposer en général que d'un échantillon d'observations très limité au regard du nombre des paramètres des modèles qu'ils peuvent envisager pour leurs analyses. Avant toute observation de la réponse, il doit donc préciser, non seulement quels modèles doivent utiliser, mais encore, comment organiser les expériences. En effet, la qualité de l'analyse statistique dépend étroitement du plan expérimental utilisé pour observer la réponse.

L'objectif principal de notre travail est d'étudier les deux approches de régression les plus utilisées dans la méthode des plans d'expériences numériques : la régression multilinéaire et la régression par krigeage. Nous cherchons à réaliser une comparaison entre ces deux approches selon un certain nombre de paramètres statistiques pour les besoins de chaque étude.

Le mémoire est composé de quatre chapitres organisés comme suit:

Le premier chapitre présente la théorie de la méthodologie des plans d'expériences et un état d'art sur les plans d'expériences couramment utilisés en expérimentation numérique, ainsi que certains critères pour étudier la qualité de ces plans.

Les importantes notions de modélisation mathématique ont été décrites au chapitre deux.

Le troisième chapitre introduit la théorie de la régression multilinéaire. Le chapitre commence par la présentation de la régression linéaire simple, puis sa généralisation pour une régression multilinéaire. Pour valider une telle régression, un rappel sur les outils statistiques a été effectué (Analyse de la variance, Analyse des résidus etc...).

Le quatrième chapitre est consacré à la représentation et à l'étude de la régression par Krigage. Aussi, une comparaison a été effectuée entre les deux régressions pour les plans numériques de Strauss Marqué et cela pour les besoins de chaque étude et chaque utilisateur.

Enfin, une conclusion termine ce travail, et en annexe, nous présentons les programmes développés dans le logiciel MATLAB permettant de réaliser les illustrations numériques.

CHAPITRE I

LES PLANS D'EXPERIENCES

Dans ce chapitre sont synthétisés et résumés les différentes hypothèses entrant en jeu dans l'utilisation de la méthode des plans d'expériences. Une méthode utile à tout expérimentateur qui entreprend des recherches scientifiques ou des études industrielles. Une partie importante dans ce chapitre sera consacrée à l'étude des plans d'expériences numériques.

1.1 Plans d'expériences classiques

1.1.1 Intérêt des plans d'expériences et leur utilisation

Les plans d'expériences sont utilisés dans les études industrielles en recherche de développement. Ils interviennent dans de nombreux domaines industriels. On peut notamment citer :

- Industries chimiques, pétrochimiques et pharmaceutiques.
- Industries mécaniques et automobiles.
- Industries métallurgiques.

Leur utilisation vise les buts suivants :

- Détermination des facteurs clés dans la conception d'un nouveau produit ou d'un nouveau procédé.
- Optimisation des réglages d'un procédé de fabrication ou d'un appareil de mesure.
- Prédiction par modélisation du comportement d'un procédé.

Les plans d'expériences s'inscrivent dans une démarche générale d'amélioration de la qualité.

1.1.2 Vocabulaire de base

Le scientifique est souvent amené à comprendre comment réagit un système en fonction des facteurs susceptibles de le modifier. Pour visualiser cette évolution, il mesure une réponse et va ensuite essayer d'établir des relations de cause à effet entre les réponses et les facteurs. Parmi les facteurs on distinguera:

- Les facteurs contrôlables qui dépendent directement du choix du technicien (pression, température, matériau...etc.).
- Les facteurs non contrôlables qui varient indépendamment du choix du technicien (Conditions climatiques, environnement d'utilisation...).
- Les facteurs d'entrée dont on cherche à analyser une influence (matière première, vitesse d'agitation, température, rendement ...).

Les facteurs étudiés dans un plan d'expériences sont bien entendu les facteurs d'entrée. Un facteur est une grandeur le plus souvent mesurable mais il peut s'agir d'une grandeur qualitative comme les différents lots d'une matière première. Lorsque l'on étudie l'influence d'un facteur, en général, on limite ses variations entre deux bornes (borne inférieure c'est le niveau bas ; borne supérieure c'est le niveau haut).

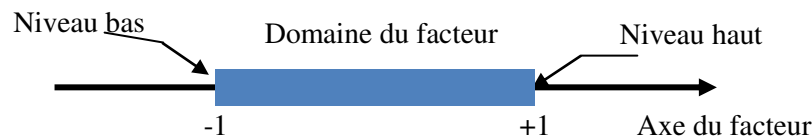


Figure 1.1 : Domaine de variation du facteur.

Une réponse expérimentale d'un système est une manifestation mesurable que l'on observe lorsque l'on fait varier les facteurs étudiés pour connaître l'effet des facteurs sur le système. La réponse peut être de type quantitatif continu, comme un rendement, caractéristique mécanique,...etc., ou de type qualitatif. Les réponses quantitatives sont en général plus faciles à traiter.

1.1.3 Réalisation des essais

Un soin tout particulier doit être apporté à l'exécution des essais. Si on ne réalise pas personnellement les essais, il faut notamment vérifier que les facteurs contrôlables mais non étudiés soient bien fixés à des valeurs précises. De même si un des facteurs étudiés est un composé chimique, il est bien sûr préférable de ne pas avoir à changer de lot de matière première durant l'ensemble de l'expérimentation.

1.1.4 Matrice d'expériences

Une matrice d'expériences est un objet mathématique qui représente, sous forme codée ou normée, l'ensemble des expériences à réaliser. C'est un tableau constitué de n lignes, correspondant aux n expériences, et de k colonnes, correspondant aux k variables

(facteurs) étudiées. La matrice d'expériences (Tableau 1.1) définit des essais. L'élément ij de la matrice ainsi formée correspond à la valeur du niveau que prend la $j^{i\text{ème}}$ variable à la $i^{i\text{ème}}$ expérience. La matrice d'expériences, définit alors, les essais à réaliser. Le terme essai est l'équivalent de point d'expérience.

Tableau 1.1 : Matrice d'expériences

N° essai	Facteur 1	Facteur 2
1(A)	-1	-1
2(B)	+1	-1
3(C)	-1	+1
4(D)	+1	+1

1.1 .5 Variable centré réduite

L'utilisation des variables centrées réduites présente l'intérêt de pouvoir généraliser la théorie des plans d'expériences quels que soient les facteurs ou les domaines d'études retenus. Remplacer les variables naturelles par les variables codées va permettre d'avoir pour chaque facteur le même domaine de variation (entre -1 et +1) et de pouvoir ainsi comparer entre eux l'effet des facteurs. Le niveau bas est ainsi codé - 1 alors que le niveau haut est codé + 1.

Le passage des variables d'origine z aux variables centrées réduites x (variables sans dimensions), et inversement, est donné par la formule suivante :

$$x = \frac{z - z_0}{pas}$$

Où $z_0 = (\text{niveau haut} + \text{niveau bas})/2$, $pas = (\text{niveau haut} - \text{niveau bas})/2$.

1.1.6 Différents types de plan d'expériences classiques

Dans cette section, nous introduisons les principaux types de plans d'expériences. Ces plans peuvent être classés en deux catégories :

- Les plans factoriels.
- Les plans pour surface de réponse.

Ces deux catégories sont liées aux objectifs d'utilisation possibles de la méthode des plans d'expériences.

1.1.6.1 Les plans factoriels

Ces plans permettent de découvrir les facteurs les plus influents sur une réponse donnée. On ne cherche pas vraiment à obtenir une relation précise entre les variations des facteurs et celle de la réponse. Ces plans sont des plans discrets orthogonaux. Les facteurs sont discrétisés pour ne prendre qu'un nombre fini de niveaux. L'orthogonalité d'un plan se définit de la manière suivante :

Définition 1.1 : Un plan d'expériences est orthogonal si :

- A chaque niveau d'un facteur, tous les niveaux de n'importe quel autre facteur lui sont associés le même nombre de fois dans le plan.
- A chaque niveau de chaque facteur apparaît le même nombre de fois.

A Plan factoriel complet

Un plan factoriel complet est un plan pour lequel toutes les combinaisons possibles aux limites du domaine d'étude auront été réalisées : c'est le nombre maximal d'essais pour un plan d'expériences factoriel. Le plan du Tableau 1.2 est un plan complet de trois facteurs respectivement à 2, 2 et 3 niveaux. Lorsqu'on dispose de k facteurs à 2 niveaux, le plan complet est noté 2^k . Pour 3 niveaux, on le note 3^k . Si on a k_1 facteurs à 2 niveaux et k_2 facteurs à 3 niveaux, on le note le plan $2^{k_1}3^{k_2}$.

Tableau 1.2 : Matrice d'expériences $2^2 \times 3^1$

N°	A	B	C
1	1	1	1
2	1	1	2
3	1	1	3
4	1	2	1
5	1	2	2
6	1	2	3
7	2	1	1
8	2	1	2
9	2	1	3
10	2	2	1
11	2	2	2
12	2	2	3

B Plans factoriels fractionnaires

En pratique, les plans complets ne sont pas souvent utilisés à cause de ce grand nombre d'expériences. On utilise plutôt des plans fractionnaires qui représentent une fraction orthogonale du plan complet. Cette fraction consiste à ne garder que certaines expériences du plan complet tout en conservant la propriété d'orthogonalité. La solution proposée par les plans fractionnaires, consiste à utiliser pour l'étude de facteurs, les matrices des effets des plans complets 2^{k-1} , 2^{k-2} ... etc. L'avantage des plans fractionnaires est évident, la charge expérimentale est divisée par 2^2 . Si par exemple on veut étudier 3 facteurs en ne faisant que 4 essais. Nous prenons la précaution de choisir les quatre essais pour que la matrice X soit une matrice orthogonale.

1.1.6.2 Les plans pour surface de réponse

L'objectif de cette catégorie de plans est plus que de hiérarchiser les effets des différents facteurs, de décrire le plus précisément possible le comportement de la réponse en fonction des variations des facteurs. Le but de ce type d'étude est donc de parvenir à une modélisation du phénomène étudié basée sur l'expérimentation. Ces plans permettent de déterminer à quelles valeurs les facteurs d'entrée d'un dispositif doivent être ajustés pour obtenir une ou plusieurs réponses souhaitées, ils s'appuient sur l'usage de modèles de

nature polynomiale. Signalons l'existence de nombreuses références sur les plans de surface de réponse. Il en existe plusieurs types : les plans factoriels complets à trois niveaux, les plans de Box et Behnken, les plans composites, les plans optimaux... etc.

A Plans composites

Les plans composites permettent de calculer un modèle polynomial du second degré. On parle de modèle quadratique. Ils sont composés d'une première partie qui est un plan fractionnaire ou complet, puis d'un ou plusieurs essais au centre du domaine d'étude et enfin d'essais supplémentaires permettant le calcul du modèle quadratique. Pour deux facteurs la figure ci-dessous montre la disposition des points expérimentaux pour un tel plan :

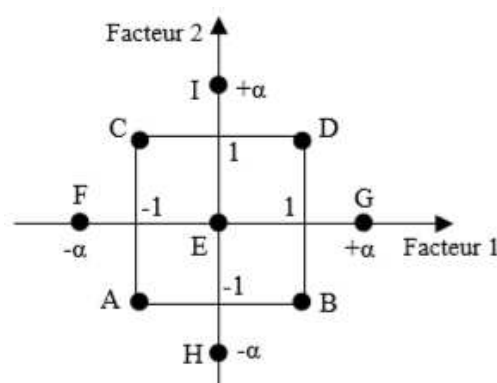


Figure 1.2 : Plan composite pour l'étude de deux facteurs.

B Plans de Box-Behnken

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'hyper cube $[-1, 1]^k$. Ces plans d'expériences sont composés d'une matrice factorielle à deux niveaux (2^k points) et de blocs incomplets équilibrés en suivant un arrangement particulier. Ces matrices sont ensuite complétées d'un point ou plusieurs points au centre du domaine. Ces plans sont une alternative aux plans composites parce qu'ils requièrent seulement 3 niveaux pour chaque facteur et permettent aussi de modéliser une surface de réponse de degré 2.

C Plans D-optimaux

Les plans D-optimaux font partie d'une famille plus grande appelée plans optimaux alphabétiques (du nom des critères d'optimalité remplis par les plans : A, D, E,... etc.). Ils

sont particulièrement bien adaptés à des problématiques avec contraintes (telles que contrainte du domaine de variation : essais impossibles, réutilisation d'essais, certains essais ont été réalisés au préalable mais pas au niveau des points expérimentaux préconisés par la théorie des plans d'expériences, contrainte sur le nombre d'essais maximal à effectuer, dans cette situation, la qualité du plan se dégrade, il y a notamment perte d'orthogonalité. La précision des estimateurs que l'on obtiendra sera beaucoup plus faible. La question qui se pose est donc de trouver d'autres expériences dans ce domaine sous contraintes pour permettre une estimation optimisée de ce modèle : c'est l'objectif du plan D-optimal.

D Plans de type Space Filling

Sont de plans Beaucoup utilisés dans le cas des essais numériques, permettent d'étudier des modèles complexes, non linéaires par exemple. Dans ce type de plans, chaque facteur prend beaucoup de niveaux équi-répartis sur son domaine de variation. Les propriétés qui peuvent être satisfaites par ces plans sont :

- Le remplissage : maximisation de la distance entre les deux points les plus proches du plan.
- L'indépendance : maximisation du déterminant de la matrice de corrélation des paramètres.
- L'uniformité (discrépance) : minimisation de la distance à une répartition uniforme.

1.1.7 Pourquoi utilise-t-on encore des plans d'expériences classiques en simulation ?

Les plans d'expériences classiques ne sont pas remis en cause. Si on les utilise dans le domaine où ils sont applicables, ils sont de très bonne qualité. Mais si les hypothèses initiales ne sont pas vérifiées, comme cela semble être souvent le cas en simulation, leur utilisation n'est pas appropriée. Cependant, malgré la complexité des modèles numériques, un simple modèle polynomial peut être suffisant pour approcher la réponse du simulateur et ce pour diverses raisons. La réponse du simulateur peut-être lisse même si le modèle mathématique décrivant le phénomène simulé est très complexe. Par exemple l'industrie pétrolière utilise des simulateurs d'écoulement pour prédire des productions cumulées d'huile. On constate que ce type de réponse est suffisamment régulier pour être résumé par une surface de réponse polynomiale [1].

Dans de nombreuses applications, le domaine d'étude peut être plus petit que le domaine de variation des variables d'entrée. Sur un domaine restreint il est alors possible de représenter la réponse du simulateur par un simple modèle polynomial même si globalement la réponse a un comportement plus complexe. Dans l'exemple précédent, les variables d'entrée du simulateur sont connues avec une plage d'incertitude due aux erreurs de mesure de ces variables lors de la phase de caractérisation du réservoir sur le terrain. Le domaine expérimental représente alors ces incertitudes sur les variables d'entrée, il est donc suffisamment petit pour justifier l'utilisation d'un modèle polynomial même pour des réponses complexes telles que la production de gaz ou d'eau [2].

Le modèle polynomial peut être très utile pour faire une analyse de sensibilité. Iman et Helton [3] ont montré au travers de nombreux exemples que, même si la surface de réponse n'est pas de qualité, elle permet tout de même de trier par ordre de contribution les variables d'entrée influant sur la réponse.

Enfin une dernière raison pour laquelle la méthode des plans d'expériences classique est souvent utilisée vient du fait qu'elle est bien connue des ingénieurs et l'interprétation physique du modèle polynomial est simple, de plus elle est disponible en logiciel. Cependant « l'automatisation » de cette méthode amène quelquefois à une mauvaise utilisation, le logiciel étant alors utilisé en boîte noire sans s'assurer que l'approche est bien adaptée au problème à traiter.

1.1.7.1 Avantages et Inconvénients

Les deux principaux avantages de cette méthode sont sa simplicité et le faible coût de simulation. En effet, si une surface de réponse polynomiale est adaptée à la réponse étudiée, alors cette approche est celle qui nécessite le moins de simulations. Cependant, il se peut que ce modèle statistique soit trop simple au regard de la complexité de la réponse du simulateur. Dans de nombreux cas il est nécessaire de faire appel à des modèles statistiques plus sophistiqués tels que le Krigeage. De plus, comme nous l'avons signalé précédemment, les points des plans classiques sont positionnés sur les bords du domaine expérimental. Ils ne sont donc pas adaptés pour détecter les irrégularités à l'intérieur du domaine, ni d'ailleurs pour ajuster un modèle statistique autre que polynomial. D'autres types de plans sont alors envisagés.

1.2 Les plans d'expériences numériques

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).

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.

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. Une autre 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, c'est-à-dire de l'erreur de mesure, tandis qu'en numérique, c'est une erreur liée au modèle et non à l'expérimentation. 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.

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).

Pour résumer, un plan d'expériences numérique efficace doit être capable d'approcher une assez grande variété de modèles et de fournir une information sur tout le domaine expérimental avec un nombre de simulations le plus faible possible. Cette non dépendance n'est évidemment pas absolue, mais relative à l'influence des autres variables. Dans le cas, où il n'y a pas eu d'étude de sensibilité au préalable. Nous supposons ici implicitement qu'ils sont relativement lisses.

Pour explorer efficacement, les simulations doivent :

- Approcher l'orthogonalité des variables d'entrée.
- 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 plans de type Space Filling paraît être un avantage pour certains met modèles tels que le Krigeage afin d'éviter les problèmes numériques.

1.2.1 Plans d'expériences numériques usuelles

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 le 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.

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, aussi de la simulation de crues et de calcul de rejet de CO₂, de la simulation d'une soufflerie à Mach, de la simulation d'accident nucléaire, de la simulation de crash automobile, ...etc.

1.2.2 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.

- **Hyper cubes latins**

La méthode d'échantillonnage par hyper cube latin a été introduite pour évaluer numériquement les intégrales multiples. Elle permet d'assurer le non redondance de l'information au travers d'une bonne répartition des projections sur les axes factoriels. En pratique, les hyper cubes 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 : $[0, \frac{1}{n}]$, $[\frac{1}{n}, \frac{2}{n}]$, \dots , $[\frac{n-1}{n}, 1]$.

En faisant le produit cartésien de ces intervalles, on obtient un maillage de n^k cellules de même taille. Ensuite, on choisit n cellules parmi les n^k 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.

Définition 1.2: Un hyper cube latin à n points sur $[0,1]^k$, est défini par l'ensemble de points X^i tel que :

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

Où π_j est une permutation de $1, \dots, n$ et $U_j^{(i)} \sim U[0,1]$ est une valeur aléatoire de distribution uniforme sur $[0,1]$. Ainsi, $(\pi_1(i), \dots, \pi_k(i))$ représente la cellule dans laquelle se trouve le point X^i , et $(U_1^{(i)}, \dots, U_k^{(i)})$ désigne à quel endroit le point se trouve dans la cellule. L'hyper cube latin obtenu correspond donc à la matrice de n lignes et k colonnes à coefficients X_j^i .

Nous pouvons choisir de placer le point au centre des cellules pour ainsi supprimer l'aléa de la planification. Un hyper cube latin, défini par la matrice π , est un plan très facile à construire puisque chaque colonne est une permutation de 1 à n .

Les points d'un hyper cube latin ont la propriété intéressante d'être uniformément distribués sur les axes factoriels. 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 k colonnes, et donc $(n!)^{k-1}$ hyper cubes latins éventuels qui n'assurent pas tous l'uniformité des points dans l'espace. C'est le cas par exemple de l'hyper cube latin 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.

Pour pallier ces problèmes, il existe deux possibilités. La première consiste à construire les $(n!)^{k-1}$ hyper cubes 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'hyper cubes latins à générer, même si la construction d'un hyper cube latin demande peu de temps d'exécution. La deuxième possibilité, plus réaliste, consiste à générer un hyper cube 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 hyper cube latin orthogonal pour les effets principaux, voire plus si possible. Il est aussi possible d'appliquer des critères d'uniformité, ce qui revient à sélectionner un hyper cube latin ayant de bonnes propriétés d'uniformité. De nombreuses références sont disponibles sur ce sujet propose une construction, basée sur un algorithme d'échanges, d'hyper cubes latins optimaux quel que soit le critère.

- **Tableaux orthogonaux**

Définition 1.3 : Un tableau orthogonal de force t à q symboles est une matrice de n lignes et $(k > t)$ 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 $OA_\lambda(n, k, q, t)$.

Géométriquement, cela revient à découper les axes du cube unité en q segments égaux afin d'obtenir k^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.

. Remarquons qu'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.

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

Définition 1.4 : Un échantillonnage par tableau orthogonal (abusivement appelé par la suite tableau orthogonal) à n points sur $[0,1]^k$ 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, $(\pi_1(A_1^{(i)}), \dots, \pi_k(A_k^{(i)}))$ représente la cellule dans laquelle se trouve le point X^i , et $(U_1^{(i)}, \dots, U_k^{(i)})$ 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 hyper cubes latins :

Propriété 1.1 :

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 .

Propriété 1.2 :

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

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 k . A titre d'exemple, le tableau orthogonal $OA_1(6^2, 4, 6, 2)$ ne peut exister. En effet, Tarry [4] a démontré qu'il n'existe pas de carré gréco latin d'ordre 6.

- **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 1.5 : 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 $(\mathbf{Z}/q\mathbf{Z})^k$.

Dans ce cas, on le note $OA_\lambda((\mathbf{Z}/q\mathbf{Z})^k, t)$. $\mathbf{Z}/q\mathbf{Z} = \{0, \dots, q-1\}$ (ou bien encore noté \mathbf{Z}^q) est un corps sur \mathbf{Z} car on suppose q premier.

- **Hyper cubes 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 hyper cubes 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 [5] pour remédier à ces problèmes est de combiner les propriétés d'orthogonalité d'un tableau orthogonal avec les bonnes propriétés de projection des hyper cubes 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.

Proposition 1.1

Soit A un tableau orthogonal de type $OA_1(n, k, q, t)$. 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 hyper cube 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'hyper cube latin $A = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 3 & 1 & 2 \end{pmatrix}$.

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

$$\lfloor qij / q \rfloor, i = 1, \dots, q^2.$$

Avec $\lfloor \cdot \rfloor$ qui désigne la partie entière inférieure.

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 hyper cube latin. A ce stade, nous avons donc un hyper cube latin basé sur un tableau orthogonal. Il nous suffit alors de choisir un point au hasard dans chacune des petites cellules.

1.2.3 Plans minimax

Johnson et Moore et Ylvisaker [6] ont établi non seulement le critère minimax 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.

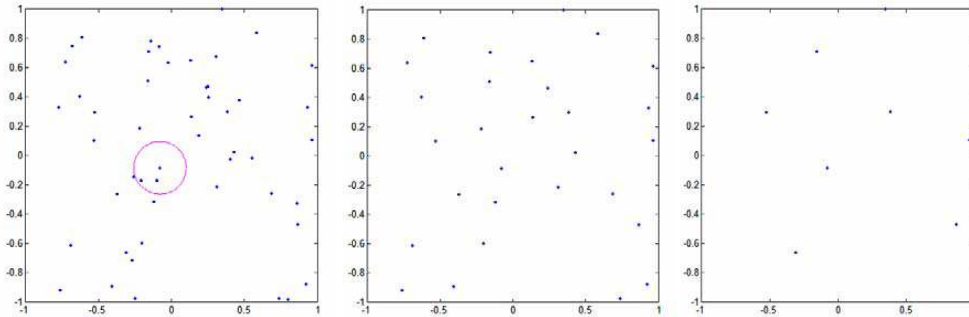


Figure 1.3 : A gauche, un plan initial (aléatoire) de 50 points avec la première boule d'élimination. Au centre, le plan à 28 points obtenu avec la boule d'élimination de la figure de gauche. A droite, le plan à 8 points obtenu avec une boule d'élimination de rayon supérieur

1.2.4 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 [7] appliquent cette définition 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 .

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

$$c(x) = \begin{pmatrix} \delta_1^2 & \text{cov}(x^1, x^2) & \text{cov}(x^1, x^n) \\ & \delta_2^2 & \vdots \\ & & \delta_n^2 \end{pmatrix}$$

Où δ_i est l'écart-type de X^i sont réduites, $c(x)$ s'identifie avec la matrice de corrélation :

$$c(x) = \delta^2 \begin{pmatrix} 1 & \dots & \delta_{ij} \\ \vdots & \ddots & \vdots \\ \delta_{ij} & \dots & 1 \end{pmatrix}$$

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

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

On reconnaît en $\gamma(h_{ij})$ le variogramme où h_{ij} est la distance entre les points i et j et a la portée dudit variogramme. On peut alors calculer pour $(x^1, \dots, x^n)^t$ vecteur de points dans un espace à k dimensions, une matrice de corrélation spatiale par :

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

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. 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.

1.2.5 Plan d'expériences numériques par processus ponctuels markoviens marqués de Strauss

L'idée principale est de considérer chaque expérience x_i comme un point ou particule défini sur $[0,1]^k$ et chaque configuration x comme une matrice d'expériences où chaque point de cette configuration, sera caractérisé par une marque m_i définie sur l'espace de marques M [8, 9]. Le point avec sa marque forment un objet définie par (x_i, m_i) . On assimile ainsi les n objets (plan d'expériences) à la réalisation d'un processus ponctuel marqué X . Les processus de Strauss incorporent des potentiels d'interactions. Ces interactions sont assimilées à des propriétés de voisinage telles que définies par un champ de Markov au sens de Ripley-Kelly. Le potentiel d'interaction le plus utilisé est l'interaction entre paires d'objets. Ces processus objets sont importants pour la modélisation des phénomènes répulsifs. La densité de probabilité d'un processus ponctuel marqué de Strauss pour une configuration x de points est donnée par :

$$\pi(x) = \alpha \beta^{n(x)} \gamma^{s(x)}$$

Avec,

- α est la constante de normalisation.
- $0 < \gamma < 1$ est un coefficient d'interaction.
- $\beta > 0$ est la densité du processus.

- $s(x) = \sum_{i < j} 1_{\{x_i \sim y_j\}}$ le nombre de paires de voisins par rapport à la relation \sim . 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 . Par conséquent $s(x) = \sum_{i < j} 1_{\|x_i - x_j\| \leq r}$.
- $n(x) = \sum_{i=1}^M n(x \times \{i\})$, $(x \times \{i\})$ représente le nombre de points dans la configuration x ayant la marque i .

1.2.6 Les suites à faible discrédance

Les suite à faible discrédance ont l'objectif de bien remplir l'espace expérimental. Ces suites de points sont initialement utilisées pour remplacer les suites aléatoires dans la méthode de Monte Monte-Carlo, d'où la dénomination de méthodes de Quasi Monte-Carlo.

Définition 1.6 : Soit un entier ≥ 2 . Tout entier $i \in N$ peut se décomposer dans la base b de manière unique :

$$i = \sum_{s=0}^m a_s b^s$$

Avec $a_s \in \{0, 1, \dots, b-1\}$. A partir de cette décomposition, on définit la fonction radicale inverse par $\phi_b(i) = \sum_{s=0}^m a_s b^{-s-1}$ pour tout $i \in N$, où, $\phi_b(i) \in [0, 1]$ Pour tout $i \in N$. 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 . Par exemple, en base $b = 4$, les 6 premiers points de la suite de Van Der Corput sont :

$$x^0 = 0, x^1 = \frac{1}{4}, x^2 = \frac{1}{2}, x^3 = \frac{3}{4}, x^4 = \frac{1}{16}, x^5 = \frac{5}{16}.$$

- **Les suites de Halton :** Ces suites se déduisent de la suite de Van Der Corput au cas multidimensionnel, autrement dit les suites de Halton sont une généralisation en dimension $k \geq 1$ des suites de Van Der Corput.

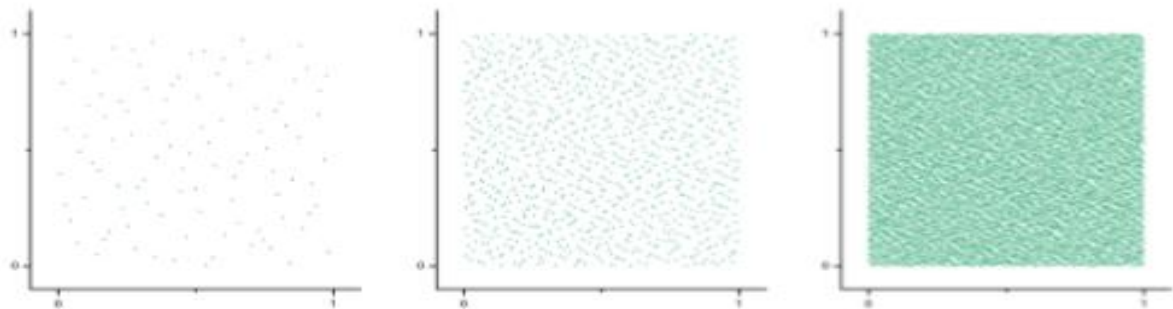


Figure 1.4 : les 100, 1000, 10000 premiers points de la suite de halton en base 2 et 3.

- **Les suites de Sobol :** Ces suites permettent de répartir les différents points dans l'espace en minimisant la distance entre chaque observation. Leur construction est assez compliquée, et elles s'obtiennent par des récurrences linéaires à partir de polynômes primitifs sur le corps $Z_2 = \{0,1\}$. Elles sont nommées quasi-aléatoires car on peut toujours trouver les coordonnées du deuxième point à partir du premier et ainsi de suite.

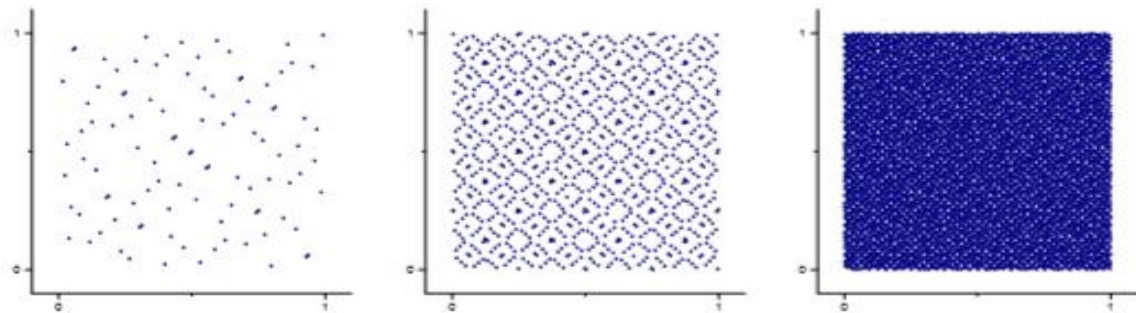


Figure 1.5 : les 100, 1000 et 10000 premiers points d'une suite de Sobol.

- **Suites de Faure :** Ces suites sont considérées comme meilleures que celles de Halton et de Sobol. La figure ci-dessous montre une distribution de 50, 250 et 500 points d'une suite de Faure pour deux dimensions.

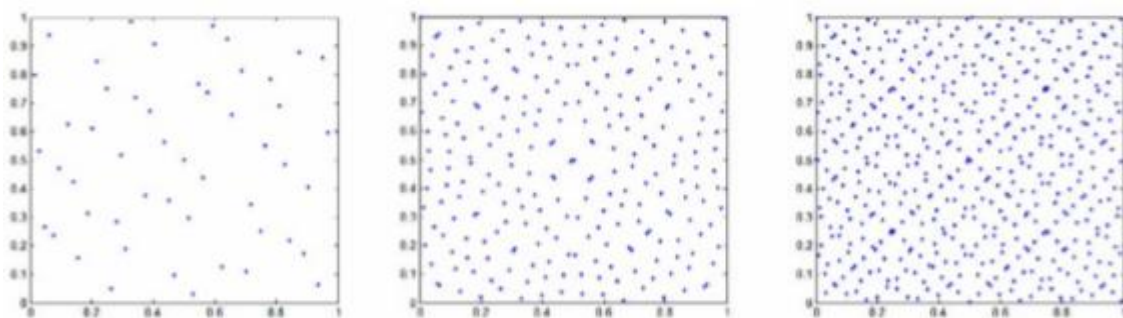


Figure 1.6 : les 50, 250 et 500 premiers points d'une suite de Faure.

CHAPITRE 2

NOTION DE MODELISATION MATHEMATIQUE

Dans ce chapitre nous considérons le cas, où les modèles mathématiques dont les réponses mesurées au cours d'une expérimentation sont des grandeurs aléatoires, qu'il faut traiter d'une manière aléatoire. La réalité doit être prise en compte. Malgré tous les soins que l'on peut prendre, la même expérience ne donnera pas exactement le même résultat à chaque détermination. Chaque mesure est entachée d'une incertitude appelée erreur expérimentale ou erreur de mesure ou erreur pure. Ceci est dû à la variation de certains paramètres qui ne sont pas connus par l'expérimentateur.

2.1 Modèle mathématique pour plans d'expérience

Le modèle est une relation entre les facteurs x_1, x_2, \dots, x_k et la réponse que l'on désire étudier. La fonction $y = f(x_1, \dots, x_k)$ correspondant à une surface de réponse (appelée hyper surface si $k > 3$) est trop générale et il est d'usage d'en prendre un développement limité de Taylor [4]. Si les dérivées du développement de Taylor peuvent être considérées comme constantes, le développement précédent prendra la forme d'un polynôme de degré plus ou moins élevé :

$$y = a_0 + \sum a_i x_i + \sum a_{ij} x_i x_j + \sum a_{ii} x_i^2 + \dots$$

- y est la grandeur à la quelle s'intéresse l'expérimentateur.
- x_i représente le niveau attribué au facteur i . C'est la valeur de la coordonnée centrée réduite du facteur i retenue par l'expérimentateur pour réaliser un essai. Cette valeur est parfaitement connue.
- $a_0, a_i, a_{ij}, a_{ii}, \dots$ sont les coefficients du modèle. Ils ne sont pas connus et doivent être calculés à partir des résultats des expériences.

On peut classer les modèles en trois catégories :

2.1.1 Modèle du mathématicien

On exprime la valeur de la réponse en fonction d'une relation mathématique faisant intervenir des coefficients (les a_i) et les niveaux des facteurs (les x_i). Le modèle

mathématique du mathématicien est caractérisé par l'absence d'erreur expérimentale. Par exemple pour deux facteurs :

$$y = a_0 + a_1x_1 + a_2x_2 + a_{12}x_1x_2 + a_{11}x_1^2 + a_{22}x_2^2.$$

2.1.2 Modèle de l'expérimentateur

Deux compléments doivent être apportés au modèle précédemment écrit. Le premier est le manque d'ajustement ou bien l'erreur d'ajustement. Cette expression traduit le fait que le modèle choisi par l'expérimentateur avant les essais est probablement un peu différent du modèle réel qui régit le phénomène étudié. Il y a un écart entre ces deux modèles. Cet écart est le manque d'ajustement Δ (Lack of Fit). Le second est la prise en compte de la nature aléatoire de la réponse.

En effet si l'on mesure plusieurs fois une réponse en un même point expérimental, on n'obtient pas exactement le même résultat. Il y a une dispersion des résultats. Les dispersions ainsi constatées sont appelées erreurs aléatoires δ_y . Ces deux écarts sont souvent réunis dans un seul écart appelé, le résidu. Le modèle utilisé par l'expérimentateur pour deux facteurs s'écrit alors :

$$y = a_0 + a_1x_1 + a_2x_2 + a_{12}x_1x_2 + a_{11}x_1^2 + a_{22}x_2^2 + \Delta + \delta_y.$$

Ce modèle se compose de trois parties. La première partie est un modèle du mathématicien, la seconde partie est l'écart d'ajustement. L'objectif de l'expérimentateur est de réduire ce terme pour le ramener à une valeur de l'ordre de grandeur de l'erreur expérimentale. La troisième partie est l'erreur aléatoire. Cette erreur est due à la variation des niveaux des facteurs non contrôlés.

2.1.3 Modèle du statisticien

Le modèle précédent est trop compliqué pour faire des calculs. On effectue les simplifications suivantes :

- L'erreur d'ajustement entre le modèle choisie à priori et le modèle réel est supposée négligeable.
- Les erreurs aléatoires sont issues d'une même population, pour tout le domaine d'étude. Il s'agit de valeurs ayant la même moyenne, le même écart type et provenant de la même distribution. Il faut bien voir que les valeurs de l'erreur sont différentes les unes des autres, puisqu'elles résultent de tirages au hasard dans une population donnée.

Pour un modèle polynomial du second degré à deux facteurs, le modèle de statisticien s'écrit :

$$y = a_0 + a_1x_1 + a_2x_2 + a_{12}x_1x_2 + a_{11}x_1^2 + a_{22}x_2^2 + \delta_y.$$

2.2 Modèle pour plans de mélange

A Modèle mathématique d'ordre 1

Tout comme dans le cas classique le modèle polynomial le plus simple à mettre en œuvre est celui de degré égal à un. Là aussi un tel modèle peut être intéressant lorsque, par exemple, le nombre de composants est élevé et qu'une première étude est nécessaire afin d'évaluer quels sont les plus influents sur la réponse étudiée (technique de Criblage). Le modèle statistique classique est toujours de la forme $y(X) = f(X) + \varepsilon(x)$. Supposons tout d'abord que la loi de réponse peut être correctement approchée au voisinage ε du centré du simplexe par :

$$\forall x \in \varepsilon, f(x) = \beta_0 + \sum_{i=1}^m \beta_i x_i.$$

Ce modèle ne tient cependant pas compte de l'hypothèse fondamentale (H_2). On peut l'introduire simplement en remarquant que $B_0 = B_0 \times 1$ et donc :

$$\forall x \in \varepsilon, f(x) = \beta_0(\sum_{i=1}^m x_i) + \sum_{i=1}^m \beta_i x_i = \sum_{i=1}^m (\beta_0 + \beta_i)x_i.$$

On constate donc qu'il est donc inutile de conserver l'effet moyen général β_0 (qui devient impossible à estimer ici) et en posant $b_i = \beta_0 + \beta_i$ ($i = \dots, m$) il vient :

Proposition 2.1 : Le modèle polynomial d'ordre un adapté à l'étude des mélanges, pour m composants, est donné par :

$$\forall x \in \varepsilon, f(x) = \sum_{i=1}^m b_i x_i.$$

Il en résulte que le nombre de paramètres inconnus d'un tel modèle est :

$$P = m.$$

Notons que lorsque ce modèle est mis en œuvre avec un plan d'expérience adapté, la matrice du modèle $x \in M(n, m)$ est donc identique à la matrice du plan D .

Voici un exemple de surface ajustée à l'aide d'un modèle d'ordre un. Il s'agit de la représentation graphique du modèle ajusté au sens des moindres carrés obtenu à partir d'un réseau de Scheffé de type $\{3,3\}$.

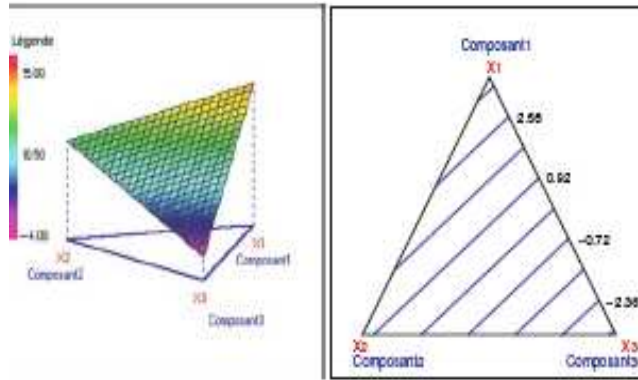


Fig. 2.1 Modélisation par un polynôme d'ordre un.

B Modèle d'ordre deux

Le modèle polynomial d'ordre un présenté précédemment n'est pas assez riche afin de décrire correctement bon nombre de situations pratiques. Il est alors naturel d'utiliser un modèle polynomial de degré supérieur. Partant du modèle classique d'ordre deux on montre (tout comme dans le paragraphe précédent) que l'hypothèse (H_2) entraîne la suppression de l'effet moyen général β_0 mais aussi celle de tous les effets quadratiques β_{ii} . Ceci entraîne que :

Proposition 2.2 Le modèle polynomial d'ordre deux adapté à l'étude des mélanges, pour m composants, est donné par :

$$\forall x \in \varepsilon, f(x) = \sum_{i=1}^m b_i x_i + \sum_{i < j} b_{ij} x_i x_j.$$

Il en résulte qu'un tel modèle est constitué par m paramètres inconnus (les b_i) auxquels il faut rajouter les c_2^m paramètres b_{ij} (il y en a autant que de choix non-ordonnés de 2 éléments parmi m). On a donc :

$$p = m + c_2^m = m + \frac{m(m-1)}{2} = \frac{m(m+1)}{2}.$$

Voici un exemple de surface ajustée au sens des moindres carrés à l'aide d'un modèle d'ordre 2.

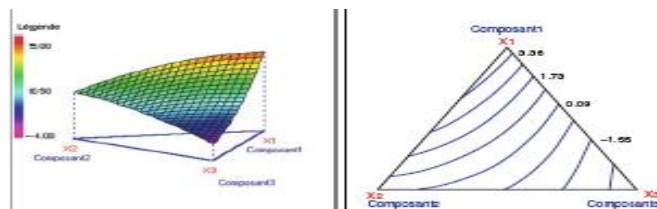


Fig. 2.2 Modélisation par un polynôme d'ordre deux.

Cette modélisation utilise les mêmes données que pour l'exemple du paragraphe 2.2.1. On constate ici une amélioration de l'ajustement par rapport au cas affine car la surface est maintenant légèrement courbée mais la qualité de l'ajustement reste insuffisante car $R^2=0.667$.

C Modèle d'ordre trois complets

Il a été montré précédemment que l'utilisation d'un modèle polynomial d'ordre deux dans le cas des mélanges entraîne la disparition de l'effet moyen général ainsi que des effets quadratiques. Il en résulte que ce modèle est beaucoup moins riche que le modèle d'ordre deux classique et va dans certaines situations s'avérer trop pauvre pour décrire correctement le phénomène étudié. Ceci implique donc que l'utilisation d'un modèle polynomial d'ordre trois est envisageable. La prise en compte de l'hypothèse fondamentale (H_2) conduit alors à considérer la classe de modèles suivante :

Proposition 2.3 Le modèle polynomial d'ordre trois adapté à l'étude des mélanges, pour $m \geq 3$ composants, est donné par :

$$\forall x \in \varepsilon, f(x) = \sum_{i=1}^m b_i x_i + \sum_{i < j} b_{ij} x_i x_j + \sum_{i < j} \delta_{ij} x_i x_j (x_i - x_j) + \sum_{i < j < k} b_{ijk} x_i x_j x_k.$$

Les paramètres inconnus d'un tel modèle sont alors : les b_i au nombre de m , les b_{ij} ainsi que les δ_{ij} au nombre de c_m^2 et enfin les b_{ijk} au nombre de c_m^3 . Le nombre total de paramètres inconnus est donc donné par :

$$p = m + 2c_m^2 + c_m^3 = \frac{m(m+1)(m+2)}{6}.$$

Voici un exemple de surface ajustée au sens des moindres carrés à l'aide d'un modèle d'ordre trois complets.

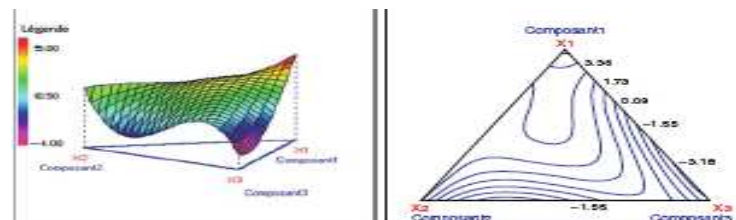


Fig. 2.3 Modélisation par un polynôme d'ordre trois.

Cette surface est toujours obtenue à l'aide des mêmes données que dans les paragraphes précédents. Il est clair ici que l'introduction des termes cubiques dans le

modèle a beaucoup modifié l'allure de la surface ajustée par rapport à ce que l'on avait pour l'ordre un ou deux. On retrouve bien ce résultat quantitativement puisque le coefficient de corrélation linéaire multiple est maintenant $R^2 = 0.917$. Ce résultat est bien entendu lié directement à l'enrichissement du modèle qui a 10 paramètres inconnus au total (contre respectivement 6 et 3 pour les modélisations de degrés 2 et 1).

D Modèle synergique d'ordre trois

Le modèle complet d'ordre trois permet d'effectuer des modélisations relativement complexes (voir la figure 2.3). En contrepartie le nombre de paramètres inconnus de ce modèle est assez élevé et peut constituer un obstacle au niveau du nombre d'expériences à réaliser. C'est pourquoi on s'oriente parfois vers un modèle d'ordre trois plus simple. Une solution classique consiste à supprimer les coefficients δ_{ij} . Ceci conduit alors au modèle d'ordre trois suivant qualifié de synergique (ou aussi de "spécial" dans la littérature anglophone) :

Définition 2.1 : Le modèle polynomial synergique d'ordre trois adapté à l'étude des mélanges, pour $m \geq 3$ composants, est donné par :

$$\forall x \in \varepsilon, f(x) = \sum_{i=1}^m b_i x_i + \sum_{i<j} \sum b_{ij} x_i x_j + \sum_{i<j<k} \sum \sum b_{ijk} x_i x_j x_k.$$

Les coefficients inconnus d'un tel modèle sont alors les b_i (au nombre de m), les b_{ij} (au nombre de c_m^2) ainsi que les b_{ijk} (au nombre de c_m^3). Ceci donne pour total :

$$p = m + c_m^2 + c_m^3 = \frac{m(m^2+5)}{6}.$$

Voici une analyse de la même situation qu'aux paragraphes précédents avec cette fois un ajustement réalisé à l'aide du modèle synergique d'ordre trois.

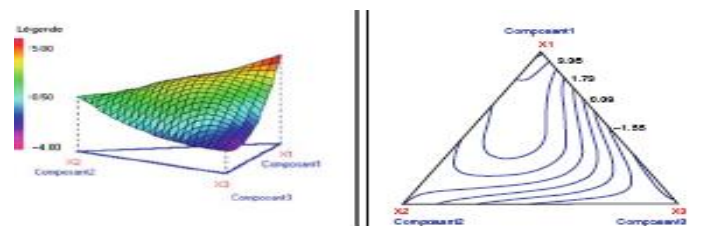


Fig. 2.4 Modélisation par un polynôme synergique d'ordre trois.

Concernant la qualité de l'ajustement on obtient ici $R^2 = 0.791$. Ceci montre donc tout l'intérêt de ce modèle car la qualité de l'ajustement bien supérieure à celle obtenue avec le modèle d'ordre deux alors qu'un seul paramètre inconnu supplémentaire a été rajouté (pour un total de 7 paramètres). L'ajustement est par contre moins bon qu'avec le

modèle complet d'ordre trois mais celui-ci n' nécessite l'utilisation de trois paramètres δ_{ij} supplémentaires

E Modèle synergique d'ordre quelconque

Une fois présente le modèle synergique d'ordre trois il est naturel d'essayer de le généraliser a un ordre quelconque des lors que plus de 3 facteurs sont utilisés. En effet ce modèle a été obtenu en rajoutant au modèle d'ordre deux les monômes de la forme $x_i x_j x_k$ et on peut envisager maintenant de rajouter aussi ceux de la forme $x_i x_j x_k x_l$ (a l'ordre 4) on bien $x_i x_j x_k x_l x_m$ (a l'ordre 5), etc... En d'autres termes il est possible d'utiliser la classe des polynômes affins et ceci entraîne la définition suivante :

Définition 2.2

Soit un entier λ et $m \geq \lambda$ composants. Le modèle polynomial synergique d'ordre λ est donné par ($\forall x \in \mathcal{E}$) :

$$f(x) = \sum_{i_1=1}^m b_{i_1} x_{i_1} + \sum_{i_1 < i_2} b_{i_1 i_2} x_{i_1} x_{i_2} + \sum_{i_1 < i_2 < i_3} b_{i_1 i_2 i_3} x_{i_1} x_{i_2} x_{i_3} + \dots + \sum_{i_1 < i_2 < \dots < i_\lambda} b_{i_1 i_2 \dots i_\lambda} x_{i_1} x_{i_2} \dots x_{i_\lambda}$$

Dans le cas particulier ou $\lambda = m$ le modèle synergique est dit complet.

Les coefficients inconnus d'un tel modèle sont alors les b_{i_1} (au nombre de m), les $b_{i_1 i_2}$ (au nombre de c_m^2), les $b_{i_1 i_2 i_3}$ (au nombre de c_m^3), etc... Le nombre total de paramètres inconnus est donc égal a :

$$P = \sum_{i=1}^{\lambda} c_m^i$$

Dans le cas particulier du modèle synergique complet il vient :

$$P = \sum_{i=1}^{\lambda} c_m^i = 2^m - 1 \quad \text{car} \quad P = \sum_{i=0}^{\lambda} c_m^i = 2^m$$

D'après la formule du binôme de Newton.

2.1.5 Modèle contenant tous les effets d'interactions

Il est parfois souhaitable d'utiliser le modèle à effets d'interactions le plus riche possible. Il s'agit donc, pour m facteurs, de considérer le modèle à effets d'interactions d'ordre m (i.e. Contenant tous les effets d'interactions jusqu'à l'ordre m). Pour tout point x du domaine expérimental $\varepsilon \subset R^m$ un tel modèle est donc donné par :

$$f(x) = \beta_0 + \sum_{i_1=1}^m \beta_{i_1} x_{i_1} + \sum_{i_1 < i_2} \sum \beta_{i_1 i_2} x_{i_1} x_{i_2} + \sum_{i_1 < i_2 < i_3} \sum \beta_{i_1 i_2 i_3} x_{i_1} x_{i_2} x_{i_3} + \dots + \sum_{i_1 < i_2 < \dots < i_m} \dots \sum \beta_{i_1 i_2 \dots i_m} x_{i_1} x_{i_2} \dots x_{i_{m-1}} + \beta_{i_1 \dots i_m} x_{i_1} x_{i_2} \dots x_{i_m}$$

2.1.6 Modèle additif

Considérons ici un modèle très souvent utilisé lorsque m facteurs qualitatifs avec respectivement h_1, h_2, \dots, h_m modalités interviennent. Un tel modèle postule que les effets des différents facteurs s'ajoutent les uns aux autres de manière indépendante. Ce modèle statistique peut toujours être écrit mathématiquement sous la forme générale $y(x) = f(x) + \varepsilon(x)$ avec ici $x = (i_1, i_2, \dots, i_m) \in \varepsilon$ vecteur associé aux différents codages naturels de l'expérience réalisée. On dit que l'on utilise un modèle additif si et seulement si :

$$f(i_1, i_2, \dots, i_m) = \beta_0 + \beta_1^{[i_1]} + \dots + \beta_m^{[i_m]}$$

On peut encore écrire ce modèle sous la forme complète donnée ci-dessous :

$$f(i_1, i_2, \dots, i_m) = \beta_0 + \sum_{j_1=0}^{h_1-1} \beta_1^{[j_1]} \delta_{i_1 j_1} + \sum_{j_2=0}^{h_2-1} \beta_2^{[j_2]} \delta_{i_2 j_2} + \dots + \sum_{j_m=0}^{h_m-1} \beta_m^{[j_m]} \delta_{i_m j_m}$$

Où δ_{ij} est le symbole de Kronecker (i.e. $\delta_{ij} = 1$ si $i = j$, $\delta_{ij} = 0$ sinon). Pour un tel modèle on dit aussi que (avec $j = 1, \dots, m$ et $i = 0, \dots, h_j - 1$) :

$$\begin{cases} \beta_0 \text{ (i.e. La constante) est l'effet moyen général.} \\ \beta_j^{[i]} \text{ est l'effet de la modalité } i \text{ du facteur } j. \end{cases}$$

Le nombre total de paramètres (note p^* pour le distinguer du nombre de paramètres inconnus à estimer $p \leq p^*$) de ce modèle est donné par :

$$p^* = 1 + \sum_{i=1}^m h_i$$

Notons enfin que matriciellement l'écriture du modèle additif est $y = x\beta + \varepsilon$ avec :

$$X = [\mathbb{I}_n \mid D] = [\mathbb{I}_n \mid X_1 \mid \dots \mid X_m].$$

Ou $X_i (i = 1, \dots, m)$ est la matrice des indicatrices des modalités du facteur i . Le vecteur $\beta \in R^{p^*}$ contient tous les paramètres du modèle, il peut naturellement être partitionné en :

$${}^t\beta = ({}^t\beta_0 \mid {}^t\beta_1 \mid \dots \mid {}^t\beta_m).$$

Ou le vecteur $\beta_i \in R^{h_i} (i = 1, \dots, m)$ regroupe tous les effets relatifs au facteur i (attention à ne pas confondre le vecteur des effets des modalités β_i pour facteurs qualitatifs et le réel β_i traduisant l'effet linéaire d'un facteur quantitatif).

2.1.7 Modèle incomplet

Jusqu'à présent les modèles linéaires considérés sont complets, c'est-à-dire qu'ils contiennent tous les effets d'interactions possibles (pour l'ordre choisi). On peut cependant envisager des situations où une connaissance préalable du phénomène étudié peut conduire à un modèle incomplet. Il convient d'être très prudent avec une telle démarche car, en cas de doute, il est préférable d'utiliser un modèle complet (si, bien sûr, le nombre de facteurs n'est pas trop élevé) quitte à avoir la confirmation a posteriori que certains effets d'interactions sont non-significatifs. Négliger un certain nombre d'interactions peut cependant être envisageable si, par exemple, les spécialistes du phénomène étudié sont en mesure d'assurer que de telles interactions ne peuvent pas avoir lieu (un chimiste affirmant clairement que les composants 1 et 2 n'ont aucun effet l'un sur l'autre peut entraîner la suppression de l'interaction 12).

2.3 Méthodes de MARS et PolyMARS

Ces méthodes sont des modèles polynomiaux par morceaux qui sont particulièrement bien adaptés en grande dimension. La méthode MARS (Multivariate Additive Régression Spline) proposée par Friedman [10] 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 et Bose et Stone [11] permet de construire des modèles Polymars (Polychotomous régression based on MARS). Ces modèles peuvent s'appliquer dans le cas des réponses multiples et paraissent être davantage adaptés 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).

2.3.1 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 :

$$(x - t)_+ = \begin{cases} x - t & \text{si } x > t \\ 0 & \text{sinon} \end{cases} \text{ ou } (t - x)_+ = \begin{cases} t - x & \text{si } x < t \\ 0 & \text{sinon} \end{cases}$$

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.

2.3.2 Méthode PolyMARS.

Cette méthode permet de prendre en compte plusieurs réponses simultanément. Par souci 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 prédicateurs 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 $(x_j - t)_+ (x_k - s)_+$.

2.3.3 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 [27]. 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 nœuds 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.

2.4 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 $\theta_j, j = 1, \dots, d + 1$ sont les paramètres de la fonction. La fonction sigmoïde est appelée fonction d'activation. On trouve par exemple, $a(x) = x$ 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.).

2.4.1 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.

CHAPITRE3

REGRESSION MULTILINÉAIRE

Dans ce chapitre, nous allons décrire une méthode de Régression utilisées dans la théorie des plans d'expériences et permettant de calculer les coefficients du modèle mathématique recherché à partir d'un certain nombre d'expériences. Cette méthode est une méthode classique et très connue sous le nom de méthode des moindres carrés. La validation du modèle obtenu à l'aide de cette méthode de Régression est effectuée par une analyse de la variance et une analyse des résidus.

3.1 Le principe de la Régression multilinéaire

La Régression multilinéaire est une généralisation de la Régression linéaire simple. Dans la Régression linéaire simple l'expérimentateur dispose des résultats de N expériences.

Tableau 3.1 : Présentation des résultats d'essais pour une Régression simple.

X	y
x_1	y_1
x_2	y_2
\vdots	\vdots
\vdots	\vdots
x_i	y_i
\vdots	\vdots
x_N	y_N

Au cours de l'essai $n^\circ i$, la variable explicative ou le facteur x à la valeur x_i , la réponse mesurée est y_i . Si les points (x_i, y_i) s'alignent correctement dans le plan (x,y) , l'expérimentateur cherchera l'équation de la droite qui s'ajuste au mieux à ces résultats.

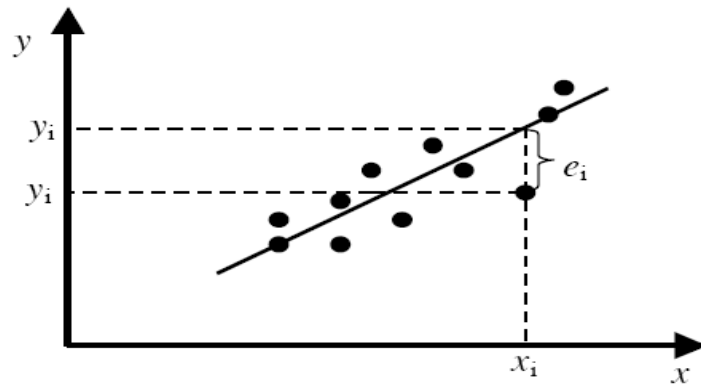


Figure 3.1 : Droite des moindres carrés.

Sur la Figure 3.1, nous voyons qu'à chaque valeur x_i de x correspondent deux valeurs de la réponse : y_i valeur mesurée et \hat{y}_i valeur donnée par la droite, avec $y_i = \hat{y}_i + e_i$. Cette droite recherchée satisfait au critère des moindres carrés $\sum e_i^2$ minimum. La démarche se généralise au cas où l'expérimentateur désire étudier une réponse y en fonction non plus d'un seul facteur x , mais de k facteurs x_1, x_2, \dots, x_k . Le tableau des résultats d'une Régression linéaire multiple est présenté par le Tableau 3.2 :

Tableau 3.2 : Présentation des résultats d'expériences pour une Régression multiple.

x_1	...	x_i	...	x_k	y
x_{11}	...	x_{1i}	...	x_{1k}	y_1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_{i1}	...	x_{ii}	...	x_{ik}	y_i
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_{N1}	...	x_{Ni}	...	x_{Nk}	y_N

Au cours de l'essai i , le facteur x_1 a la valeur x_{i1} , le facteur x_2 a la valeur x_{i2} , ..., le facteur x_j a la valeur x_{ij} , ..., et le facteur x_k a la valeur x_{ik} . La réponse observée est y_i . L'expérimentateur fait l'hypothèse que la réponse peut être approchée par un modèle linéaire de type :

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_jx_j + \dots + a_qx_q.$$

Dans lequel une variable quelconque x_j peut représenter :

- L'un des facteurs x_1, x_2, \dots, x_k .
- Une interaction quelconque entre ces facteurs.

- Une fonction de ces facteurs : $x_i^2, \log(x_i), \dots$ etc.

Les coefficients a_i du modèle sont estimés par la méthode de moindre carré.

3.2 Hypothèses de la Régression multilinéaire

La valeur attendue pour y dans les conditions de l'expérience i , si le modèle ci-dessus est valide est:

$$\eta(y_i) = a_0 + a_1x_{i1} + a_2x_{i2} + \dots + a_qx_{iq}$$

En raison des erreurs aléatoires, la valeur y_i obtenue expérimentalement diffère de $\eta(y_i)$ d'une quantité ε_i

$$y_i = \eta(y_i) + \varepsilon_i$$

La méthode des moindres carrés ne nécessite aucune hypothèse sur la distribution des réponses. Cependant, elle possède certaines propriétés intéressantes sous réserve que soient satisfaites le mieux possible les conditions suivantes:

- les erreurs ε_i doivent être distribuées suivant une loi normale de moyenne 0 et de variance σ^2 , $N(0, \sigma)$, ce qui s'écrit :

$$E(\varepsilon_i) = 0$$

$$var(\varepsilon_i) = \sigma^2$$

3.3 Estimation des coefficients par la méthode des moindres carrés

Lorsque on estime les inconnues a_0, a_1, \dots, a_q par $\hat{a}_0, \hat{a}_1, \dots, \hat{a}_q$ nous pouvons calculer la réponse au point i par :

$$\hat{y}_i = \hat{a}_0 + \hat{a}_1x_{i1} + \hat{a}_2x_{i2} + \dots + \hat{a}_jx_{ij} + \dots + \hat{a}_qx_{iq}$$

La valeur \hat{y}_i diffère du résultat expérimental y_i de la quantité e_i (de même que les \hat{a}_i sont les estimateurs de a_i , les e_i , sont des estimations des ε_i)

$$y_i = \hat{y}_i + e_i.$$

Nous écrivons cette égalité quelque soit i , nous obtenons le système linéaire :

$$\left\{ \begin{array}{l} y_1 = \hat{a}_0 + a_1x_{11} + \hat{a}_2x_{12} + \dots + \hat{a}_jx_{1j} + \dots + \hat{a}_qx_{1q} + e_1 \\ y_2 = \hat{a}_0 + \hat{a}_1x_{21} + \hat{a}_2x_{22} + \dots + \hat{a}_jx_{2j} + \dots + \hat{a}_qx_{2q} + e_2 \\ \vdots \\ y_i = \hat{a}_0 + a_1x_{i1} + \hat{a}_2x_{i2} + \dots + \hat{a}_jx_{ij} + \dots + \hat{a}_qx_{iq} + e_i \\ \vdots \\ y_N = \hat{a}_0 + a_1x_{N1} + \hat{a}_2x_{N2} + \dots + \hat{a}_jx_{Nj} + \dots + \hat{a}_qx_{Nq} + e_N \end{array} \right.$$

Nous cherchons les valeurs des \hat{a}_j qui minimisent la somme des carrés des écarts $\sum e_i^2$. Concrètement nous cherchons le modèle linéaire qui passe au plus près de l'ensemble des points expérimentaux. Pour simplifier l'écriture, nous adoptons la notation matricielle :

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_i \\ \dots \\ y_N \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1j} & \dots & x_{1q} \\ 1 & x_{21} & \dots & x_{2j} & \dots & x_{2q} \\ \vdots & & & \vdots & & \vdots \\ 1 & x_{i1} & \dots & x_{ij} & \dots & x_{iq} \\ \vdots & & & \vdots & & \vdots \\ 1 & x_{N1} & \dots & x_{Nj} & \dots & x_{Nq} \end{pmatrix} \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_j \\ \vdots \\ \hat{a}_q \end{pmatrix} + \begin{pmatrix} e_0 \\ e_1 \\ \vdots \\ e_i \\ \vdots \\ e_N \end{pmatrix}$$

Le système à résoudre s'écrit : $y = X\hat{A} + e$, et le critère des moindres carrés ${}^t e e$ soit minimale on a :

$$\begin{aligned} {}^t e e &= {}^t (y - X\hat{A}) (y - X\hat{A}) \\ &= {}^t (y - \hat{A}'X) (y - X\hat{A}) \\ &= {}^t y y - \hat{A}'Xy - {}^t y X\hat{A} + \hat{A}'XX\hat{A} \\ &= {}^t y y - 2 \hat{A}'Xy + \hat{A}'XX\hat{A} \end{aligned}$$

Calculons la dérivée de ${}^t e e$ par rapport à l'inconnue \hat{A} :

$$\frac{\partial({}^t e e)}{\partial \hat{A}} = \frac{\partial({}^t Y Y)}{\partial \hat{A}} - 2 \frac{\partial({}^t \hat{A}' X Y)}{\partial \hat{A}} + \frac{\partial({}^t \hat{A}' X X \hat{A})}{\partial \hat{A}}$$

Où :

- $\frac{\partial({}^tYY)}{\partial \hat{A}} = 0$ car tYY ne dépend de \hat{A}
- $\frac{\partial({}^t\hat{A}{}^tXY)}{\partial \hat{A}} = {}^tXY$ car ${}^t\hat{A}{}^tXY$ est une forme linéaire en \hat{A}
- $\frac{\partial({}^t\hat{A}{}^tXX\hat{A})}{\partial \hat{A}} = 2 {}^tXX\hat{A}$ car ${}^t\hat{A}{}^tXX\hat{A}$ est une forme quadratique en \hat{A}

Il vient donc :

$$\frac{\partial({}^tee)}{\partial A} = -2{}^tXY + 2{}^tXX\hat{A}$$

La valeur de A qui minimise tee doit vérifier :

$$\frac{\partial({}^tee)}{\partial A} = 0 \Rightarrow -2{}^tXY + 2{}^tXXA = 0 \Rightarrow {}^tXXA = {}^tXY$$

Si la matrice $({}^tXX)^{-1}$ n'est pas singulière on a :

$$\hat{A} = ({}^tXX)^{-1}{}^tXY \dots \dots \dots (3.1)$$

Cette relation est fondamentale et nous l'utiliserons constamment par la suite. Elle est valable pour tous les modèles polynomiaux quelque soit leur degré et quel que soit le nombre de coefficients. Les coefficients ainsi obtenu sont utilisés pour écrire le modèle mathématique. Ce modèle permet de calculer les réponses (qui sont souvent appelées les réponses prédites \hat{y}) dans tout le domaine d'étude.

$$\hat{Y} = X\hat{A}$$

Elles sont très utiles pour faire des prévisions, tracer les diagrammes d'iso réponses, chercher des optimums ou des valeurs bien précises. C'est un précieux outil de prévision.

3.4 Propriétés des coefficients

Les coefficients \hat{a}_i sont fonction des résultats expérimentaux et doivent être considérés comme des variables aléatoires, donc nous pouvons calculer leur espérance mathématique ainsi que leur variance.

3.4.1 Espérance mathématique des coefficients

D'après la formule de (3.1) l'espérance mathématique de \hat{A} a pour expression :

$$\begin{aligned} E(\hat{A}) &= E [({}^tXX)^{-1}{}^tXY] \\ &= ({}^tXX)^{-1}{}^tX E(Y) \end{aligned}$$

Car les éléments de X sont considérés comme fixes. En désignant par A le vecteur des coefficients vrais et ε le vecteur des N écarts entre les résultats expérimentaux et les réponses théoriques alors:

$$Y = XA + \varepsilon$$

Et

$$\begin{aligned} E(Y) &= E(XA + \varepsilon) = E(XA) \\ &= X E(A) \end{aligned}$$

Car $E(\varepsilon) = 0$ par hypothèse. Nous trouvons :

$$E(\hat{A}) = ({}^tXX)^{-1} {}^tXXA = A$$

Le résultat que nous venons d'établir signifie que les distributions des \hat{a}_i sont centrés sur les valeurs vraies a_i

3.4.2 Variance des coefficients

Par définition la variance de \hat{A} est :

$$\text{var}(\hat{A}) = E[(\hat{A} - A) {}^t(\hat{A} - A)]$$

Remplaçons \hat{A} par $({}^tXX)^{-1} {}^tXy$ et y par $XA + \varepsilon$. Nous obtenons :

$$(\hat{A} - A) = ({}^tXX)^{-1} {}^tX(XA + \varepsilon) - A = A + ({}^tXX)^{-1} {}^tX\varepsilon - A = ({}^tXX)^{-1} {}^tX\varepsilon$$

Puisque ${}^t(\hat{A} - A) = {}^t\varepsilon X ({}^tXX)^{-1}$

Donc $\text{var}(\hat{A}) = E[({}^tXX)^{-1} {}^tX \varepsilon {}^t\varepsilon X ({}^tXX)^{-1}] = ({}^tXX)^{-1} {}^tX E(\varepsilon {}^t\varepsilon) X ({}^tXX)^{-1}$

Remplaçons $E(\varepsilon {}^t\varepsilon)$ par $E[(\varepsilon - 0) {}^t(\varepsilon - 0)] = \text{var}(\varepsilon) = \sigma^2$. Nous pouvons écrire :

$$\begin{aligned} \text{var}(\hat{A}) &= ({}^tXX)^{-1} {}^tX \sigma^2 X ({}^tXX)^{-1} \\ &= \sigma^2 ({}^tXX)^{-1} {}^tXX ({}^tXX)^{-1} \\ &\rightarrow \text{var}(\hat{A}) = \sigma^2 ({}^tXX)^{-1} \end{aligned}$$

En multipliant les éléments de la matrice $({}^tXX)^{-1}$ par σ^2 , nous obtenons la matrice de variance covariance :

$$\text{var}(\hat{A}) = \begin{pmatrix} \text{var}(\hat{a}_0) & \cdots & & & \cdots & \text{cov}(\hat{a}_q, \hat{a}_0) \\ \vdots & & & & & \vdots \\ & & \text{var}(\hat{a}_i) & \text{cov}(\hat{a}_j, \hat{a}_i) & & \\ & & \text{cov}(\hat{a}_i, \hat{a}_j) & \text{var}(\hat{a}_j) & & \\ \vdots & & & & & \vdots \\ \text{cov}(\hat{a}_0, \hat{a}_q) & \cdots & & & \cdots & \text{var}(\hat{a}_q) \end{pmatrix}$$

La matrice de variance-covariance ci-dessus est comme $({}^tXX)^{-1}$, une matrice carrée symétrique. Les éléments de la diagonale sont les variances des coefficients, tandis que les termes hors diagonale correspondent aux covariances.

3.5 Analyse de la variance

Une fois le modèle ajusté, le problème de la qualité de l'ajustement obtenu se pose alors. Il est possible d'obtenir des indicateurs numériques permettant de quantifier ceci à l'aide des techniques dites d'analyse de la variance. Ces techniques reposent sur des décompositions judicieuses en sommes de carrés. Désignons dans la suite par \bar{Y} la réponse moyenne observée et par Y^* le vecteur des réponses observées centrées. Remarquons que si 1_N est l'indicatrice d'ordre N (c-à-d le vecteur de \mathbb{R}^n dont toutes les composantes sont égales à 1) alors Walter [12]:

$$\bar{Y} = \frac{1}{N} 1_N^T Y \text{ et } Y^* = Y - \bar{Y} 1_N$$

Nous définissons ensuite les trois sommes classiques suivantes (la notation SS venant de l'anglais Sum of Squares):

- Somme totale des carrés : $\sum_{i=1}^N (Y_i - \bar{Y})^2$ (désignée par SST)
- Somme des carrés due à la Régression : $\sum_{i=1}^N (\hat{Y}_i - \bar{Y})^2$ (désignée par SSR)
- Somme des carrés due à l'erreur : $\sum_{i=1}^N (Y_i - \hat{Y}_i)^2$ (désignée par SSE)

Proposition 3.1

Pour le modèle des moindres carrés, si $P = X (X^T X)^{-1} X^T$ est le projecteur orthogonal de \mathbb{R}^n sur le sous espace $Im(X)$ et si $I_n \subset Im(X)$ alors les sommes carrées ci dessus s'écrivent :

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = Y^T Y - n\bar{Y}^2, \quad \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = Y^T P Y - n\bar{Y}^2 \text{ et } \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = Y^T (I_n - P) Y$$

Il en dérive la relation fondamentale suivante : $\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$

Preuve : Sous une forme matricielle, on peut écrire :

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = (Y - \bar{Y} \mathbf{1}_n)^T (Y - \bar{Y} \mathbf{1}_n) = Y^T Y - \bar{Y} \mathbf{1}_n^T \mathbf{1}_n - \bar{Y} \mathbf{1}_n^T Y + \bar{Y}^2 \mathbf{1}_n^T \mathbf{1}_n, \text{ puisque } \mathbf{1}_n^T Y = Y^T \mathbf{1}_n = n\bar{Y}$$

et $\mathbf{1}_n^T \mathbf{1}_n = n$, donc $\sum_{i=1}^n (Y_i - \bar{Y})^2 = Y^T Y - n\bar{Y}^2$. Pour $\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = (Y - \hat{Y})^T (Y - \hat{Y})$ et comme

$\hat{Y} = X (X^T X)^{-1} X^T Y = P Y$ alors $Y - \hat{Y} = Y - P Y = Y(I_n - P)$, on obtient donc :

$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = (Y - \hat{Y})^T (Y - \hat{Y}) = Y^T (I_n - P)(I_n - P) Y = Y^T (I_n - P) Y$$

Pour $\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = Y^T P Y - n\bar{Y}^2$, on a : $\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = \hat{Y}^T \hat{Y} - 2\bar{Y} \hat{Y}^T \mathbf{1}_n + \bar{Y}^2 \mathbf{1}_n^T \mathbf{1}_n$. Or, Pour le

modèle des moindres carrés : $X^T Y = X^T X \hat{A}$. Et par conséquence :

$X^T Y = X^T X \hat{A} = X^T \hat{Y}$, en multipliant les deux parties par $\frac{1}{n} \mathbf{1}_n$, on obtient :

$$X^T \frac{1}{n} Y \mathbf{1}_n = X^T \frac{1}{n} \hat{Y} \mathbf{1}_n \Rightarrow X^T \bar{Y} = X^T \bar{\hat{Y}}$$

Par conséquence, $\bar{Y} = \bar{\hat{Y}}$. D'où :

$$\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = \hat{Y}^T \hat{Y} - 2\bar{Y} \hat{Y}^T \mathbf{1}_n + \bar{Y}^2 \mathbf{1}_n^T \mathbf{1}_n = \hat{Y}^T \hat{Y} - n\bar{Y}^2 = Y^T P Y - n\bar{Y}^2.$$

Ceci entraîne bien que :

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Car, $\sum_{i=1}^n (Y_i - \bar{Y})^2 = Y^T Y - n\bar{Y}^2 + Y^T (I_n - P) Y = Y^T P Y - n\bar{Y}^2$.

Pour Y vecteur aléatoire de \mathbb{R}^n et $M(n, n)$ une matrice non aléatoire, nous appelons nombre de degrés de liberté de $Y^T M Y$ le rang de la matrice M . Cette notion de degrés de liberté provient de la loi usuelle du khi-Deux. En effet, on montre que si Y est un vecteur aléatoire de loi gaussienne $N(\mu, \delta^2 I_n)$ et si M est la matrice d'un projecteur alors $Y^T M Y$ suit une loi du khi-Deux décentrée, de paramètre de décentrage $\frac{1}{2} \mu^T A \mu$ avec un nombre de degrés de liberté égal au rang de M [13]. Les degrés de liberté associés aux différentes sommes de carrés sont donnés ci-dessous.

Proposition 3.2 Les sommes de carrés SST, SSE et SSR sont associées, respectivement, à $(n - 1)$, $(n - p)$ et $(p - 1)$ degrés de liberté. Ceci permet de définir les sommes moyennes de carré suivantes:

$$MST = \frac{\sum_{i=1}^n Y_i^2 - n\bar{Y}^2}{n-1}, \quad MSR = \frac{\sum_{i=1}^n \hat{Y}_i^2 - n\bar{Y}^2}{p-1} \quad \text{et} \quad MSE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-p}$$

(Avec la notation MS dénote la somme moyenne des carrés).

Preuve : d'après la proposition 1.4. Nous avons:

$\sum_{i=1}^n (Y_i - \bar{Y})^2 = Y^T Y - n\bar{Y}^2 = Y^T (I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T) Y$ et puisque $\frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T = \mathbf{1}_n (\mathbf{1}_n^T \mathbf{1}_n)^{-1} \mathbf{1}_n^T$ est le projecteur orthogonal de \mathbb{R}^n sur $Im X$ alors $(I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T)$ est la projection orthogonale \mathbb{R}^n sur les sous espace $(Im X)^\perp$, donc la matrice $(I_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T)$ a pour rang $n-1$ car la dimension de $(Im X)^\perp$ est égale à $n-1$.

De même que précédemment :

$\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = Y^T P Y - n\bar{Y}^2 = Y^T (P - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T) Y$, si $\mathbf{1}_n \in Im X$ il se retrouve donc aussi dans le noyau de $P - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T$ et le rang de $P - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T$ est égale à $p-1$ [13].

D'où, puisque par définition la variance est obtenue en divisant la somme des carrés par le nombre de degrés de liberté de ces sommes. On obtient ainsi les variances :

$$\left\{ \begin{array}{l}
 \text{Variance totale} = \frac{\sum_{i=1}^n y_i^2 - n\bar{y}^2}{n-1} \\
 \text{Variance d'ajustement} = \frac{\sum_{i=1}^n \hat{y}_i^2 - n\bar{y}^2}{p} \\
 \text{Variance résiduelle} = \frac{\sum_{i=1}^n e_i^2}{n-p-1}
 \end{array} \right.$$

3.6 Tests statistiques

3.6.1 Le coefficient de corrélation multiple R^2

On appelle R^2 le rapport de la partie des résultats expliqués par le modèle à la totalité de ce qu'il devrait expliquer :

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

D'après cette formule, nous voyons que le rapport R^2 varie entre 0 et 1. Le modèle ajusté est d'autant plus "proche" des réponses observées que R^2 est proche de 1. Un seuil classique consiste à valider le modèle dès lors que $R^2 \geq 0.95$.

3.6.2 Le coefficient de $R_{ajusté}^2$

Par définition ce coefficient est donné par la relation :

$$R_{ajuste}^2 = \frac{\frac{\sum_{i=1}^n e_i^2}{n-p-1}}{\frac{\sum_{i=1}^n y_i^2 - n\bar{y}^2}{n-1}}$$

Ces deux rapports varient entre 0 et 1. Ils mesurent en quelque sorte le degré d'ajustement du modèle. Plus ils sont voisins de 1 et plus le modèle représente bien les points expérimentaux.

3.6.3 Comparaison du R^2 et du $R_{ajusté}^2$

Partons d'un modèle ayant p coefficients. Ce modèle a un R^2 et un $R_{ajusté}^2$. Si on retire un coefficient de ce modèle, les réponses calculées sont un peu moins bien expliquées par le nouveau modèle, le R^2 diminue. En ce qui concerne le $R_{ajusté}^2$, il y a une inconnue de moins à calculer pour le modèle, on augmente donc le nombre de degrés de liberté des résidus. Si le coefficient retiré est proche de 0, le $R_{ajusté}^2$ peut augmenter. Si le coefficient retiré est peu élevé, le $R_{ajusté}^2$ peut augmenter légèrement. Le $R_{ajusté}^2$ est donc une statistique qui donne des indications sur la puissance explicative des coefficients. Le $R_{ajusté}^2$ est un moyen de discriminer les modèles possibles. Quand on hésite entre plusieurs modèles, on peut choisir celui qui a le plus fort $R_{ajusté}^2$.

3.6.4 Le F de Fisher

On appelle coefficient de Fisher le réel :

$$F = \frac{\frac{SSR}{p-1}}{\frac{SSE}{n-p}}$$

Si le F de Fisher est élevé, la variance de la Régression (réponses calculées) est plus grande que la variance des résidus. La probabilité que les coefficients du modèle soient tous nuls est alors très faible. Pour avoir des coefficients significatifs, il faut un F de Fisher élevé, c'est-à-dire une probabilité (p-value) faible. Le modèle sera déclaré globalement significatif si $F > F_{1-\alpha, p, n-p-1}$. On peut dire aussi que la qualité du modèle est d'autant

meilleure que la variance d'ajustement est plus faible par rapport à la variance expérimentale [14].

CHAPITRE 4

REGRESSION PAR KRIGEAGE ET ETUDE COMPARATIVE AVEC LA REGRESSION MULTILINEAIRE

Ce chapitre présente la méthode de Krigeage, cette dernière a été développée dans le domaine de la géostatistique par Matheron [23, 24] et a été utilisée en planification d'expériences numériques par Isaaks et Srivastava [28]. Il existe trois types de Krigeage, un Krigeage simple, ordinaire et universel, mais le type le plus utilisé et le plus générale c'est le Krigeage universel, dont nous discuterons ultérieurement. Dans un deuxième temps, nous fournissons une comparaison entre la Régression multilinéaire et la Régression par Krigeage. Cette comparaison a été réalisée en développant un programme dans le logiciel Matlab présenté en annexe de ce manuscrit.

4.1 Démarche générale de résolution des équations du Krigeage

L'objectif du Krigeage est de prévoir la valeur de la variable régionalisée à interpoler $Y(\cdot)$ en un site non échantillonné noté x_0 . La première étape pour atteindre ce but consiste à déterminer le voisinage de Krigeage, ce voisinage se définit par le domaine du champ D contenant x_0 ainsi que les sites $x_{[1]}$ à $x_{[n_0]}$ associés aux observations utilisées dans la prévision de $Y(x_0)$. Ces sites doivent former un sous-ensemble de l'ensemble des sites d'observations, donc $\{x_{[1]}, \dots, x_{[n_0]}\} \subseteq \{x_1, \dots, x_n\}$. Il sera aussi noté $v(x_0)$. Le choix du voisinage de krigeage se base sur une certaine connaissance de la structure de dépendance spatiale entre les observations. La taille n_0 de ce voisinage doit cependant être assez grande pour mener à une prévision précise.

Par la suite, la formule de prévision par Krigeage peut être trouvée. Pour effectuer cette étape, la démarche proposée par Matheron [22] sera suivie dans ce mémoire. Chauvet conseille de procéder par l'écriture des contraintes du Krigeage que voici :

- **Contrainte de linéarité**

La contrainte de base du Krigeage est que la prévision prenne la forme d'une combinaison linéaire des données. Elle doit donc s'écrire ainsi :

$$\hat{Y}(x_0) = a + \sum_{i \in v(x_0)} \lambda_i Y(x_i)$$

Les poids λ_i et la constante a sont les inconnus du problème.

- **Contrainte d'autorisation**

Il faut s'assurer que l'espérance et la variance de l'erreur de prévision $\hat{Y}(x_0) - Y(x_0)$ existent. Cette contrainte n'intervient que dans le cas où la fonction aléatoire $Z(\cdot)$ est supposée stationnaire intrinsèque.

- **Contrainte de non-biais**

La prévision par Krigeage doit posséder la propriété d'absence de biais. Il faut donc que :

$$E[\hat{Y}(x_0) - Y(x_0)] = 0.$$

- **Contrainte d'optimalité**

Les poids λ_i et la constante a sont déterminés de façon à minimiser : $var[\hat{Y}(x_0) - y(x_0)]$ sous les contraintes précédentes.

Cette démarche mènera à la résolution d'équations. Pour faciliter l'écriture de ces équations qui s'avèrent parfois longues, une notation matricielle sera employée. Voici donc tout d'abord quelques remarques sur cette notation :

- Y est le vecteur $n_0 \times 1$ des variables aléatoires $Y(x_{[1]})$ à $Y(x_{[n_0]})$ intervenant dans la prévision de $Y(x_0)$.
- λ est le vecteur $n_0 \times 1$ des poids associés aux variables aléatoires $Y(x_{[1]})$ à $Y(x_{[n_0]})$
- Z est le vecteur $n_0 \times 1$ des erreurs associées aux variables aléatoires $Y(x_{[1]})$ à $Y(x_{[n_0]})$
- Dans le cadre stationnaire de second ordre, nous avons :
 - ❖ Σ est la matrice $n_0 \times n_0$ de variances-covariances de Z , dont la diagonale est composée uniquement de σ^2 , la variance commune à toutes les erreurs $Z(x)$ pour $x \in D$
 - ❖ c_0 est le vecteur $n_0 \times 1$ des covariances entre Z et $Z(x_0)$
- Dans le cadre stationnaire intrinsèque, nous avons:

- ❖ Γ est la matrice $n_0 \times n_0$ dont l'élément (i, j) est $\gamma(x_{[i]} - x_{[j]})$ soit le Semi-Variogramme entre $Z(x_{[i]})$ et $Z(x_{[j]})$, les éléments i et j de Z .
- ❖ γ_0 est le vecteur $n_0 \times 1$ dont l'élément i est $\gamma(x_{[i]} - x_0)$, le Semi-Variogramme entre $Z(x_{[i]})$ et $Z(x_0)$.

Notons que la prévision $\hat{Y}(x_0)$ obtenue est une variable aléatoire. Sa valeur numérique est calculée en remplaçant les variables aléatoires $Y(x_i)$ par les valeurs régionalisées observées $y(x_i)$.

4.2 Krigeage universel

Le modèle de base du Krigeage universel est :

$$Y(x) = \sum_{j=0}^p f_j(x) \beta_j + Z(x) \quad x \in D$$

Avec :

- $f_j(x)$ fonction de base connue.
- $\beta_j = (\beta_0, \dots, \beta_p)^t$ un vecteur de coefficients estimés.
- Z un processus gaussien stationnaire centré et de forme de covariance factorisée gaussienne.

4.3 Krigeage avec dérivé externe

En Krigeage avec dérivé externe, le modèle s'écrit :

$$Y(x) = \sum_{j=0}^p f_j(w) \beta_j + Z(x), \quad x \in D$$

Où : $w = (w_1(x), \dots, w_q(x))$ est le vecteur des valeurs prises par les q variables régionalisées auxiliaire au point x . Si une tendance linéaire est choisie, les $f_j(w)$ seront $f_0(w) = 1, f_1(w) = w_1, \dots, f_q(w) = w_q$.

Dans les paragraphes qui suivent, nous utilisons la notation du Krigeage universel, mais la démarche est exactement la même en Krigeage avec dérivé externe en changeant les $f_j(x)$ par des $f_j(w)$. En fait, le modèle de tendance peut aussi être composé de fonctions des coordonnées spatiales et de variables régionalisées auxiliaires simultanément.

Sous forme matricielle, le modèle utilisé pour prévoir $Y(x_0)$ s'énonce comme suit :

$$Y^* = \begin{pmatrix} s_0 \beta \\ X \beta \end{pmatrix} + Z^* \quad \text{avec} \quad \begin{cases} E[Z^*] = 0 \\ \Gamma, \gamma_0 \text{ connus} \end{cases}$$

Où : $Y^* = (Y(x_0), Y)$, $Z^* = (Z(x_0), Z)$, $\beta = (\beta_0, \dots, \beta_p)$, $s_0 = (f_0(x_0), \dots, f_p(x_0))$
 et X est une matrice $n_0 \times (p + 1)$ dont l'élément (i, j) est $f_j(x_i)$.

Suivons la démarche des contraintes afin de prévoir $Y(x_0)$ par krigeage universel.
 On considère les contraintes suivantes :

- **Contrainte de linéarité**

Encore une fois, $\hat{Y}(x_0)$ doit être une combinaison linéaire des $Y(x_i)$. Il s'écrit donc sous la forme :

$$\hat{Y}(x_0) = a + \sum_{i \in v(x_0)} \lambda_i Y(x_i) = a + \lambda^t Y$$

- **Contrainte d'autorisation**

A cause de l'hypothèse de stationnarité intrinsèque, il faut s'assurer que l'erreur $\hat{Y}(x_0) - Y(x_0)$ soit une combinaison linéaire d'accroissement de $Z(\cdot)$. Pour en être une, on à la somme des poids des termes aléatoires doit valoir zéro. L'erreur de prévision s'écrit :

$$\begin{aligned} \hat{Y}(x_0) - Y(x_0) &= a + \sum_{i \in v(x_0)} \lambda_i Y(x_i) - Y(x_0) \\ &= a + \sum_{i \in v(x_0)} \lambda_i (s_i \beta + Z(x_i)) - s_0 \beta - Z(x_0) \\ &= a + \sum_{i \in v(x_0)} \lambda_i s_i \beta - s_0 \beta + \sum_{i \in v(x_0)} \lambda_i Z(x_i) - Z(x_0) \end{aligned}$$

Avec $a + \sum_{i \in v(x_0)} \lambda_i s_i \beta - s_0 \beta$ termes non aléatoires. Il faut donc encore que $\sum_{i \in v(x_0)} \lambda_i = 1$ pour assurer l'existence de l'espérance et de la variance de $\hat{Y}(x_0) - Y(x_0)$.

- **Contrainte de non biais**

L'espérance de l'erreur de prévision s'écrit :

$$\begin{aligned} E[\hat{Y}(x_0) - Y(x_0)] &= E \left[a + \sum_{i \in v(x_0)} \lambda_i Y(x_i) - Y(x_0) \right] \\ &= a + \sum_{i \in v(x_0)} \lambda_i s_i \beta - s_0 \beta \\ &= a + \sum_{i \in v(x_0)} \lambda_i \sum_{j=0}^p f_j(x_i) \beta_j - \sum_{j=0}^p f_j(x_0) \beta_j \\ &= a + \sum_{j=0}^p \left(\sum_{i \in v(x_0)} \lambda_i f_j(x_i) - f_j(x_0) \right) \beta_j \end{aligned}$$

Afin que cette espérance vaille zéro pour tout $\beta_j (j = 0, \dots, p)$, il faut que $a = 0$ et $\sum_{i \in v(x_0)} \lambda_i f_j(x_i) - f_j(x_0) = 0$ pour $j = 0, \dots, p$, ces contraintes s'écrivent $\lambda^t X = s_0^t$ sous forme matricielle.

Ainsi, sans oublier la contrainte d'autorisation $\sum_{i \in v(x_0)} \lambda_i = 1$, il y a au total $p + 2$ contraintes sur les poids λ en Krigeage universel. toutefois, afin de simplifier les calculs, on suppose que $f_0(\cdot) = 1$. Ainsi $\sum_{i \in v(x_0)} \lambda_i f_j(x_i) = f_j(x_0)$ pour $j = 0$ revient à $\sum_{i \in v(x_0)} \lambda_i = 1$. Il permet d'éliminer une contrainte. L'estimateur devient donc :

$$\hat{Y}(x_0) = \sum_{i \in v(x_0)} \lambda_i Y(x_i) \text{ avec } \sum_{i \in v(x_0)} \lambda_i f_j(x_i) = f_j(x_0) \text{ pour } j = 0, \dots, p$$

- **Contrainte d'optimalité**

Finalisons la démarche en minimisant $var[\hat{Y}(x_0) - Y(x_0)]$. Tel que :

$$\begin{aligned} var(\hat{Y}(x_0) - Y(x_0)) &= E((\hat{Y}(x_0) - Y(x_0))^2) = E((\lambda y(x_i) - y(x_0))^2) \\ &= E\left(\left(\sum_{i \in v(x_0)} \lambda(\beta s_i + z(x_i)) - \beta s_0 - z(x_0)\right)^2\right) = E\left(\left(\sum_{i \in v(x_0)} \lambda \beta s_i + \sum_{i \in v(x_0)} \lambda z(x_i) - \beta s_0 - z(x_0)\right)^2\right) \\ &= E\left(\left(\sum_{i \in v(x_0)} \lambda_i z(x_i) - z(x_0)\right)^2\right) = E\left(\left(\sum_{i \in v(x_0)} \lambda_i z(x_i)\right)^2 + z(x_0)^2 - 2 \sum_{i \in v(x_0)} \lambda_i z(x_i) z(x_0)\right) \\ &= E\left(\left(\sum_{i \in v(x_0)} \lambda_i z(x_i)\right)^2 + z(x_0)^2 - 2z(x_0) \sum_{i \in v(x_0)} \lambda_i z(x_i)\right) = \\ &= E\left(\sum_{i \in v(x_0)} \sum_{j \in v(x_0)} \lambda_i \lambda_j z(x_i) z(x_j) - 2z(x_0) \sum_{i \in v(x_0)} \lambda_i z(x_i) + z(x_0)^2\right) = \\ &= E\left(\sum_{i \in v(x_0)} \sum_{j \in v(x_0)} \lambda_i \lambda_j z(x_i) z(x_j) - \sum_{i \in v(x_0)} \lambda_i z(x_i)^2 + \sum_{i \in v(x_0)} \lambda_i z(x_i)^2 - 2z(x_0) \sum_{i \in v(x_0)} \lambda_i z(x_i) + z(x_0)^2\right). \end{aligned}$$

Pour la première partie du résultat précédent, nous avons :

$$\begin{aligned} &\sum_{i \in v(x_0)} \sum_{j \in v(x_0)} \lambda_i \lambda_j z(x_i) z(x_j) - \sum_{i \in v(x_0)} \lambda_i z(x_i)^2 \\ &= \sum_{i \in v(x_0)} \sum_{j \in v(x_0)} \lambda_i \lambda_j z(x_i) z(x_j) - \sum_{j \in v(x_0)} \lambda_j \sum_{i \in v(x_0)} \lambda_i z(x_i)^2 \\ &= \sum_{i \in v(x_0)} \sum_{j \in v(x_0)} \lambda_i \lambda_j z(x_i) z(x_j) - \frac{1}{2} \sum_{i \in v(x_0)} \sum_{j \in v(x_0)} \lambda_i \lambda_j z(x_i)^2 \\ &\quad - \frac{1}{2} \sum_{i \in v(x_0)} \sum_{j \in v(x_0)} \lambda_i \lambda_j z(x_j)^2 = \\ &= -\frac{1}{2} \sum_{i \in v(x_0)} \sum_{j \in v(x_0)} \lambda_i \lambda_j (-2z(x_i) z(x_j) + z(x_i)^2 + z(x_j)^2) \\ &= -\frac{1}{2} \sum_{i \in v(x_0)} \sum_{j \in v(x_0)} \lambda_i \lambda_j (z(x_i) - z(x_j))^2. \end{aligned}$$

Pour la seconde partie :

$$\begin{aligned}
& \sum_{i \in v(x_0)} \lambda_i z(x_i)^2 - 2z(x_0) \sum_{i \in v(x_0)} \lambda_i z(x_i) + z(x_0)^2 \\
&= \sum_{i \in v(x_0)} \lambda_i z(x_i)^2 - 2z(x_0) \sum_{i \in v(x_0)} \lambda_i z(x_i) + \sum_{i \in v(x_0)} \lambda_i z(x_0)^2 \\
&= \sum_{i \in v(x_0)} \lambda_i (z(x_i)^2 - 2z(x_i)z(x_0) + z(x_0)^2) \\
&= \sum_{i \in v(x_0)} \lambda_i (z(x_0) - z(x_i))^2.
\end{aligned}$$

Donc :

$$\begin{aligned}
\text{var} \left(\hat{Y}(x_0) - Y(x_0) \right) &= E \left(\left(\hat{Y}(x_0) - Y(x_0) \right)^2 \right) \\
&= E \left(-\frac{1}{2} \sum_{i \in v(x_0)} \sum_{j \in v(x_0)} \lambda_i \lambda_j \left(Z(x_i) - Z(x_j) \right)^2 \right) + E \left(\sum_{i \in v(x_0)} \lambda_i \left(Z(x_0) - Z(x_i) \right)^2 \right) \\
&= -\frac{1}{2} \sum_{i \in v(x_0)} \sum_{j \in v(x_0)} \lambda_i \lambda_j E \left(\left(Z(x_i) - Z(x_j) \right)^2 \right) + \sum_{i \in v(x_0)} \lambda_i E \left(\left(Z(x_0) - Z(x_i) \right)^2 \right) \\
&= -\sum_{i \in v(x_0)} \sum_{j \in v(x_0)} \lambda_i \lambda_j \frac{1}{2} \text{var} \left(Z(x_i) - Z(x_j) \right) + 2 \sum_{i \in v(x_0)} \lambda_i \frac{1}{2} \text{var} \left(Z(x_0) - Z(x_i) \right) \\
&= -\sum_{i \in v(x_0)} \sum_{j \in v(x_0)} \lambda_i \lambda_j \gamma(x_i - x_j) + 2 \sum_{i \in v(x_0)} \lambda_i \gamma(x_0 - x_i) \\
&= -\lambda^t \Gamma \lambda + 2\lambda^t \gamma_0
\end{aligned}$$

Cette expression doit être minimisée en respectant les contraintes d'autorisation et de non-biais $\lambda^t X = s_0^t$. Cette étape sera effectuée à l'aide d'un vecteur noté l , de $p + 1$ lagrangiens. La fonction à minimiser est donc : $f(\lambda, l) = -\lambda^t \Gamma \lambda + 2\lambda^t \gamma_0 + 2(\lambda^t X - s_0^t)l$. Son gradient est :

$$\frac{\partial}{\partial \lambda} f(\lambda, l) = -2\Gamma\lambda + 2\gamma_0 + 2Xl$$

Il vaut zéro au point $\hat{\lambda} = \Gamma^{-1}(\gamma_0 + Xl)$. En utilisant la contrainte $X^t \hat{\lambda} = X^t \Gamma^{-1}(\gamma_0 + Xl) = s_0$, le vecteur de Lagrange l est estimé par :

$$\hat{l} = (X^t \Gamma^{-1} X)^{-1} (s_0 - X^t \Gamma^{-1} \gamma_0)$$

Ensuite, l est remplacé par \hat{l} dans $\hat{\lambda}$. L'unique point minimum de la fonction $f(\lambda, l)$ est :

$$\hat{\lambda} = \Gamma^{-1}(\gamma_0 + X(X^t \Gamma^{-1} X)^{-1} (s_0 - X^t \Gamma^{-1} \gamma_0))$$

Ainsi, $Y(x_0)$ est prévu par

$$\hat{Y}(x_0) = (\gamma_0 + X(X^t \Gamma^{-1} X)^{-1} (s_0 - X^t \Gamma^{-1} \gamma_0))^t \Gamma^{-1} Y$$

En outre, la variance de Krigeage devient :

$$\begin{aligned} \delta^2(x_0) &= var [\hat{Y}(x_0) - Y(x_0)] \\ &= \gamma_0^t \Gamma^{-1} \gamma_0 - (s_0 - X^t \Gamma^{-1} \gamma_0)^t (X^t \Gamma^{-1} X)^{-1} (s_0 - X^t \Gamma^{-1} \gamma_0). \end{aligned}$$

4.4 Lien entre le Krigeage avec modèle de tendance et le Krigeage sur les résidus d'une Régression

Dans un cas stationnaire d'ordre deux, une prévision par Krigeage avec modèle de tendance prend la forme :

$$\hat{Y}(x_0) = (c_0 + X(X^t \Sigma^{-1} X)^{-1} (s_0 - X^t \Sigma^{-1} c_0))^t \Sigma^{-1} Y$$

La même prévision peut être obtenue en effectuant un Krigeage simple avec espérance nulle sur les résidus d'une Régression linéaire qui tient compte de la dépendance spatiale des erreurs [17]. Les paramètres du modèle sont donc estimés par une méthode des moindres carrés généralisés ou par une méthode du maximum de vraisemblance avec une matrice de variances-covariances non diagonale. Dans ce cas, une prévision de $Y(x_0)$ est formée en additionnant la prévision de la tendance générale par Régression à la prévision par Krigeage simple dont l'erreurs $e = Y - X\hat{\beta}_{gls}$:

$$\begin{aligned} \hat{Y}(x_0) &= s_0^t \hat{\beta}_{gls} + \hat{e}_{KS}(x_0) \\ &= s_0^t \hat{\beta}_{gls} + c_0^t \Sigma^{-1} e \\ &= s_0^t \hat{\beta}_{gls} + c_0^t \Sigma^{-1} (Y - X\hat{\beta}_{gls}) \\ &= (s_0^t - c_0^t \Sigma^{-1} X) \hat{\beta}_{gls} + c_0^t \Sigma^{-1} Y \\ &= (s_0^t - c_0^t \Sigma^{-1} X) (X^t \Sigma^{-1} X)^{-1} X^t \Sigma^{-1} Y + c_0^t \Sigma^{-1} Y \\ &= (c_0 + X(X^t \Sigma^{-1} X)^{-1} (s_0 - X^t \Sigma^{-1} c_0))^t \Sigma^{-1} Y \end{aligned}$$

Cependant, certains utilisent cette approche en effectuant une Régression qui ne tient pas compte de la dépendance spatiale des données, par exemple une Régression des moindres carrés ordinaires ($\hat{\beta}_{ols} = (X^t X)^{-1} X^t Y$). Dans ce cas, les prévisions ne sont pas tout à fait les mêmes que celles obtenues par Krigeage avec modèle de tendance. Ce type de Krigeage est parfois appelé en anglais Detrended Kriging [18, 19]. Enfin, une autre approche parfois utilisée consiste simplement à effectuer un Krigeage ordinaire sur des erreurs $Y(x_i) - w(x_i)$. Ce type de Krigeage porte parfois le nom de Krigeage résiduel [20]. Ce Krigeage est en fait un Krigeage avec dérivé externe avec une seule variable régionalisée auxiliaire, donc $q = 1$ et $w = w(x)$, en fixant $f_0(w) = 1$, $f_1(w) = w(x)$ et $\beta_1 = 1$. Ainsi, β_0 est le seul paramètre inconnu dans la dérivé. Le modèle s'écrit donc $Y(x) = \beta_0 + w(x) + Z(x)$, ou $Y(x) - w(x) = \beta_0 + Z(x)$ pour $x \in D$. Ce modèle est particulièrement intéressant lorsque la variable régionalisée auxiliaire mesure le même phénomène naturel que la variable régionalisée à interpoler, mais avec moins de précision.

4.5 Problème de l'analyse Variographique en Krigeage avec modèle de tendance

L'analyse Variographique est une étape problématique en Krigeage avec un modèle de tendance. Le Semi-Variogramme est estimé à partir des résidus $e(x_i) = y(x_i) - \hat{\mu}(x_i)$ obtenus suite à une estimation de β . Il s'agit d'un problème de Régression qui peut être résolu par une méthode des moindres carrés. Cependant, le modèle de Régression est particulier : les erreurs ne sont pas indépendantes.

Une estimation correcte de β requiert la connaissance de la structure de dépendance spatiale de la fonction aléatoire résiduelle $Z(\cdot)$. Toutefois, cette structure est inconnue. L'analyse Variographique vise justement à l'estimer. Le problème revient donc à son point de départ.

Afin de sortir de ce cercle vicieux, une solution consiste à d'abord obtenir un estimateur des moindres carrés ordinaires de β , estimer ensuite un Semi-Variogramme sur les résidus, puis calculer un estimateur des moindres carrés généralisés de β , et ainsi de suite [21]. Au fil des itérations, les résidus des moindres carrés ordinaires s'approchent de plus en plus des résidus des moindres carrés généralisés. Nalder et Wein [17] affirment qu'en pratique une seule itération suffit pour obtenir des résultats de Krigeage satisfaisants.

Cependant, Matheron [21] soulève que même en appliquant cette méthode itérative, l'estimation du Semi-Variogramme reste biaisée [26, 27]. Les conséquences de ce biais sont cependant difficiles à évaluer. Ainsi, le Krigeage avec un modèle de tendance reste

une méthode qui peut s'avérer bonne, mais la prudence est de mise lors de son utilisation. Le biais dans l'estimation du Semi-Variogramme en Krigeage avec modèle de tendance est une des motivations au développement de méthodes bayésiennes.

Les trois sections précédentes ont présentés l'essentiel de la théorie du Krigeage formalisée par Matheron [23, 24, 25, 26, 27]. Ces sections devraient donc suffire à une compréhension de la méthode dans le but d'une utilisation de base. Cependant, avant de passer à la mise en œuvre du Krigeage, il est intéressant de s'attarder sur certains aspects théoriques plus poussés ainsi que sur d'autres types de Krigeage développés plus récemment. Les remarques théoriques traitées ici touchent la normalité des données, le travail sur des données transformées, notamment par la fonction logarithmique, et le Cokrigeage, une extension du Krigeage dans le cas multi variable.

4.6 Le Variogramme

Le Variogramme est défini pour toute fonction aléatoire intrinsèque et dépendante uniquement de l'inter distance h , alors que la fonction de covariance ne l'est que pour le cas d'une fonction aléatoire stationnaire d'ordre 2. De plus, l'estimation du Variogramme n'est pas biaisée par la moyenne, au contraire de la covariance.

L'étude de la structure par le Variogramme consiste à suivre l'évolution de variation quadratique moyenne de l'accroissement de la fonction $Z(x)$ en fonction de h d'amplitude croissante, où h est le vecteur reliant deux points dans D . On obtient ainsi le Variogramme dans une direction donnée.

On suppose habituellement que l'espérance des accroissements est stationnaire et nulle. Dans le cas de l'hypothèse intrinsèque, la fonction Semi- Variogramme $\gamma(h)$ est définie par la relation :

$$\gamma(x + h, x) = \gamma(h) = \frac{1}{2} E \left[(Z(x + h) - Z(x))^2 \right], \forall x.$$

$$\text{Et } \text{var}(Z(x + h) - Z(x)) = E \left[(Z(x + h) - Z(x))^2 \right] = 2\gamma(h), \forall x.$$

4.7 Application : Etude comparative

Dans cette partie, nous présentons les résultats de comparaison effectuée, sous forme de Plot Box. Pour plusieurs dimensions et pour trois modèles : Régression multilinéaire pour un modèle de premier degré (MLR1), Régression multilinéaire pour un modèle de deuxième degré (MLR2) et une Régression par Krigeage (Krigeage). Le plan numérique choisi est celui de Strauss Marqué. Cette comparaison a été conduite sur 40

plans afin de donner un sens aux résultats. Nous avons considéré une fonction déterministe de trois et cinq dimensions définie sur $[0,1]^3$ et $[0,1]^5$ respectivement.

$$f(x_1, x_2, \dots, x_k) = \sum_{i=1}^k e^{x_k} \left[a_i - x_k + \log \left(\sum_{j=1}^k x_j \right) \right], k \in \{3, 5\}.$$

Où, $a = (6.5, 4.2, 5, 12, -14)$.

Les résultats concernant les suites à faible discrédance sont présentés dans le tableau 4.1 suivant :

Tableau 4.1 : Résultats des critères d'évaluation pour 3 dimensions et 5 dimensions

Suite \ Critère	R ²			R ² ajusté			Fisher		
	Halton	DIMENSION 3							
MLR1		MLR2	Krigeage	MLR1	MLR2	Krigeage	MLR1	MLR2	Krigeage
0.90		0.98	0.81	0.30	0.3	0.5	28.8	58.88	7.77
DIMENSION 5									
MLR1		MLR2	Krigeage	MLR1	MLR2	Krigeage	MLR1	MLR2	Krigeage
0.90		0.93	0.7	0.19	0.20	0.80	10.18	7.74	11.50
Sobol	DIMENSION 3								
	MLR1	MLR2	Krigeage	MLR1	MLR2	Krigeage	MLR1	MLR2	Krigeage
	0.93	0.91	0.77	0.10	0.11	0.40	37.23	58.92	14.88
	DIMENSION 5								
	MLR1	MLR2	Krigeage	MLR1	MLR2	Krigeage	MLR1	MLR2	Krigeage
	0.94	0.91	0.55	0.27	0.14	0.67	5.82	16.68	16.45

4.7.1 Comparaison selon le coefficient R^2

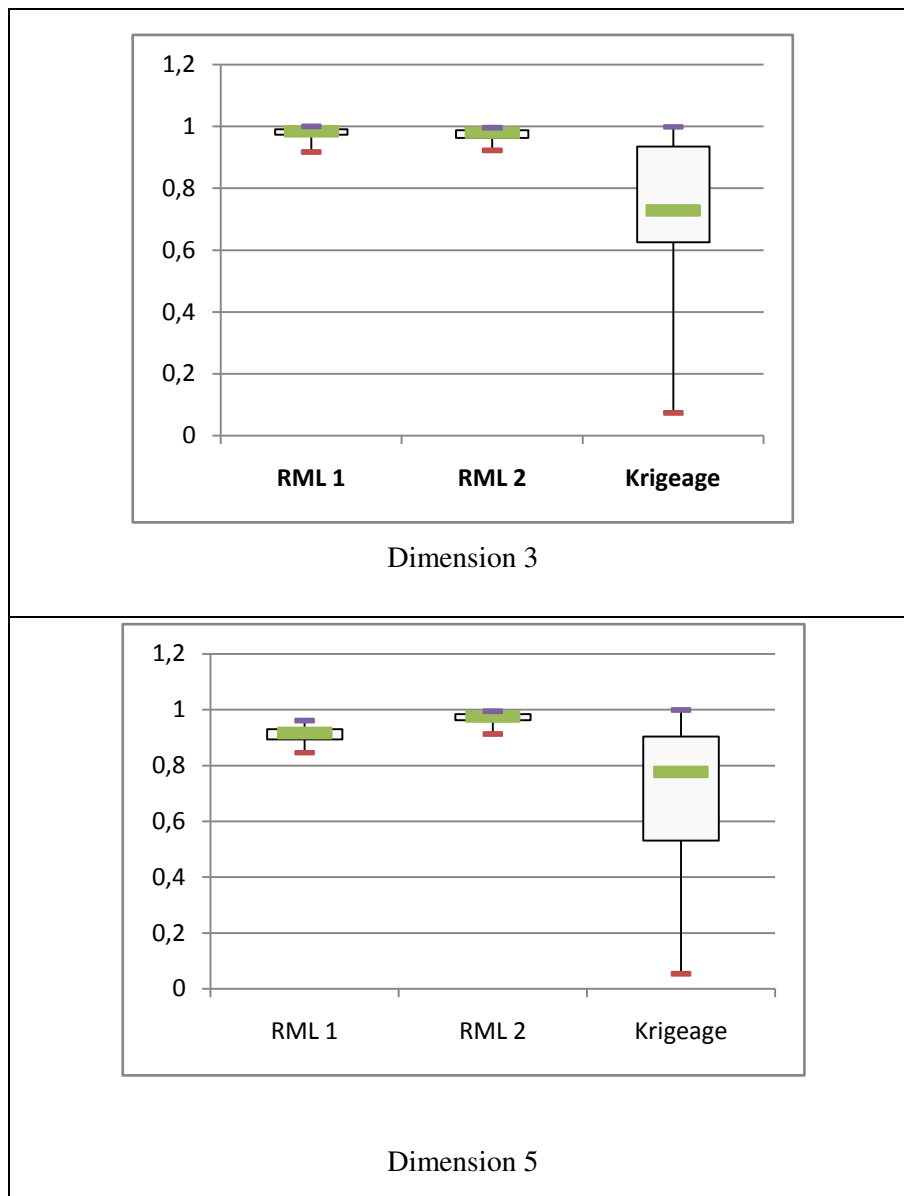


Figure 4.1 : Boxplots de critère R^2 calculé sur 40 plans à 30 points en 3 dimensions et à 40 points e 5 dimensions

Un examen approfondi sur les résultats ci-dessus, donne une idée sur la qualité de modèle à postulé. Il est clair que les deux premiers modèles sont les meilleurs vis-à-vis de ce critère. Pour le modèle par Krigeage les résultats sont moins satisfaisants. Cela, peut être dû au choix de la fonction de Variogramme, qui reste la problématique de la régression par Krigeage.

4.7.2 Comparaison selon le coefficient R^2 ajusté

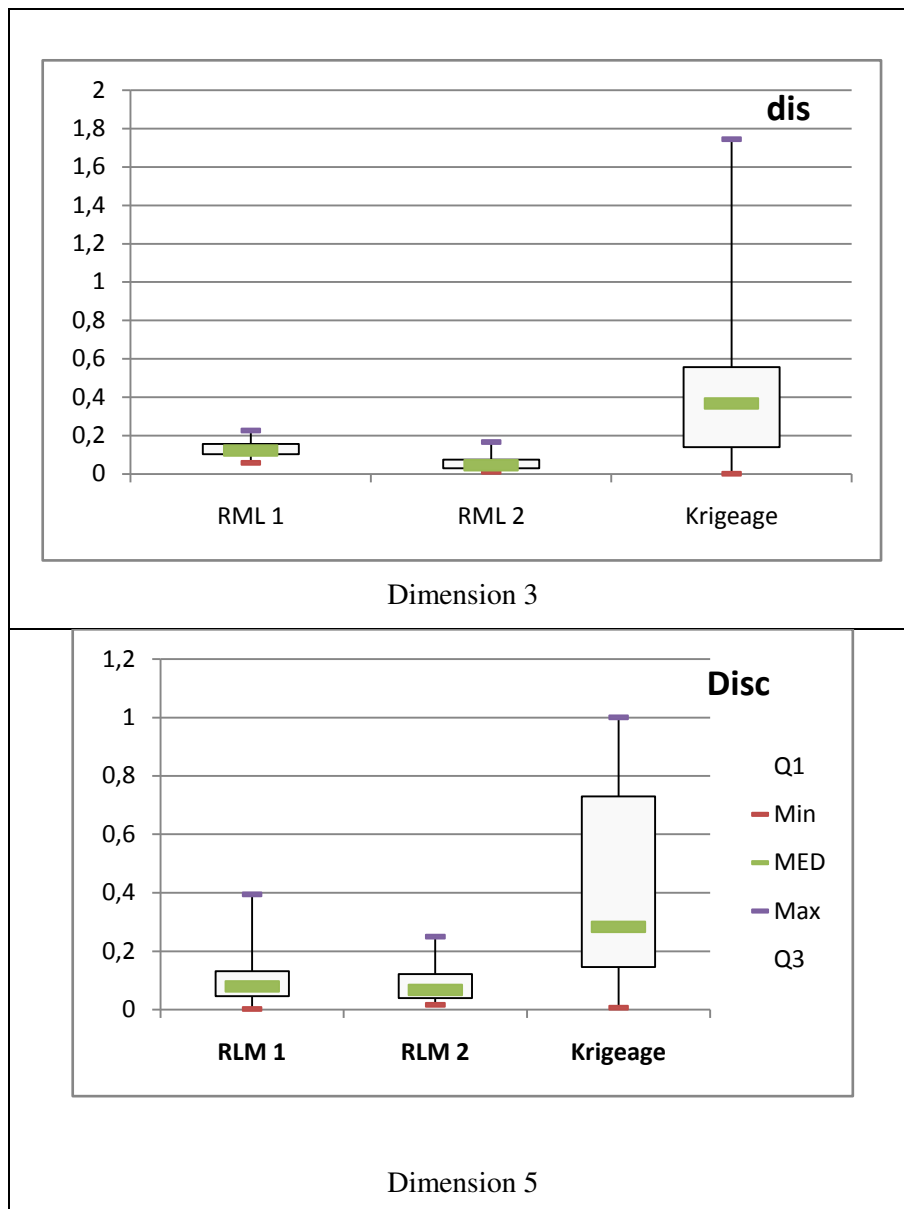


Figure 4.2 : Boxplots de critère R^2 ajustés calculé sur 40 plans à 30 points en 3 dimensions et à 40 points 5 dimensions

Selon les résultats ci-dessus, le modèle par Krigeage est le plus fiable en terme de Régression. C'est exactement ce qu'on doit avoir avec ce type de Régression conçu pour les plans numérique à la base.

4.7.3 Comparaison selon la valeur de Fisher

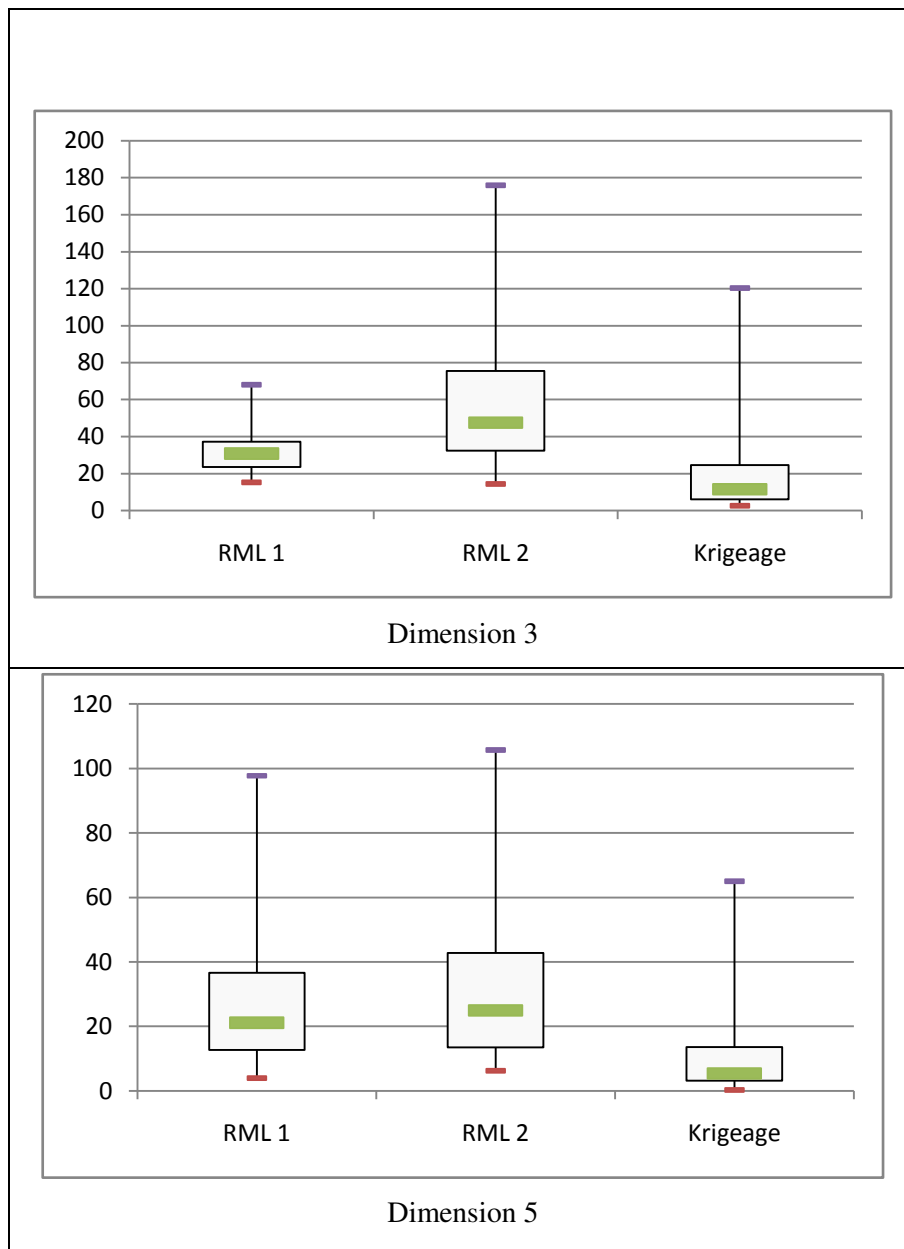


Figure 4.3 : Boxplots de critère de Fisher calculé sur 40 plans à 30 points en 3 dimensions et à 40 points 5 dimensions

Selon les résultats ci-dessus, le teste de Fisher est largement satisfait pour les trois Régression.

CONCLUSION

Dans les plans d'expériences, le modèle adopté par les statisticiens, en plus de sa simplicité puisqu'il se présente toujours sous forme d'un polynôme, permet de rassembler en une seule relation l'ensemble des informations pouvant être apportées par chaque facteur séparément ou conjointement (le nombre de ces facteurs pouvant largement dépasser 2 facteurs). Les mathématiciens et statisticiens peuvent donc contribuer et apporter des informations supplémentaires aux études réalisées par les autres disciplines pour peu qu'ils s'imprègnent quand même du sujet et des résultats de ces dernières.

Dans ce travail, nous nous sommes intéressés en premier lieu à présenter un état d'art sur les plans d'expériences numériques et à l'étude de deux régressions, une régression multilinéaire et une régression par Krigeage qui est une méthode stochastique d'interpolation spatiale qui prévoit la valeur d'un phénomène naturel en des sites non échantillonnés par une combinaison linéaire sans biais et à variance minimale des observations du phénomène en des sites voisins. Cette dernière semble être une régression alternative à la régression multilinéaire pour les plans d'expériences numériques, parce qu'elle propose un modèle statistique plus flexible.

En second lieu, une comparaison a été réalisée entre les régressions étudiées dans ce travail. Et une application en Matlab a été effectuée pour réaliser une telle comparaison.

APPENDICE

CODE MATLAB POUR LES RESULTATS PRESENTES DANS LE CHAPITRE 4

Code Matlab : Algorithme

```
n=; R=; t=; b=; w=3;
eps=; NMC=;
Rap=0;
Cov=0;
Mind=0;
Desc=0;
resultt=[];
for i=1:n
Position(i)=i;
end
for i=1:n
probabilite(i)=Position(i)/n;
end
for nbrplan=1:10
%la creation de X
Initiale=[];
for i=1:n
for j=1:w
Initiale(i,j)=random('uniform',0,1);
end
end
for i=1:n
for j=1:w
X(i,j)=Initiale(i,j);
end
end
```

```

SX0=0;
for i=1:n
for j=i+1:n
for k=1:w
d(k)=X(i,k)-X(j,k);
end
D=norm(d);
if(D<=R)
SX0=SX0+1;
end
end
end
for i=1:n
for j=1:w
Y(i,j)=X(i,j);
end
end
for N=1:NMC
for nbrel=1:n
% le calcul de nombre d'interactions entre les paires de points des
% deux configurations X et Y
hh=rand;
test=0;
j=1;
while((j<=n)&&(test==0))
if (hh < probabilite(j))
k=Position(j);
test=1;
end
j=j+1;
end
for j=1:w

```

```

y(1,j)=rand;
Y(1,j)=y(1,j);
end
Sy=0;
for i=1:n
    if (i~=k)
        for j=1:w
            d(j)=X(i,j)-y(1,j);
        end
        Dy=norm(d);
        if(Dy<=R)
            Sy=Sy+1;
        else
            Sy=Sy;
        end
    end
end
end
Sx=0;
for i=1:n
    if (i~=k)
        for j=1:w
            d(j)=X(i,j)-X(k,j);
        end
        Dx=norm(d);
        if(Dx<=R)
            Sx=Sx+1;
        else
            Sx=Sx;
        end
    end
end
end
%%%%%% le calcul des marques pour la configuration courante

```



```

somme=0;
inverse=inv(X'*X);
for i=1:n
    m1=(X(i,:)*(inverse)*X(i,:));
    if(m1<=eps)
        somme=somme+1;
    end
end
som=0;
inverse=inv(Y'*Y);
for i=1:n
    m2=(Y(i,:)*(inverse)*Y(i,:));
    if(m2<=eps)
        som=som+1;
    end
end
%%% le calcul de probabilité d'acceptation
f1=b^som;
by=(t^Sy)*(f1);
bx=(t^Sx)*(b^somme);
a=min(1,by/bx);
if (a==1)
    for j=1:w
        Xk,j=y(1,j);
        Y=X;
    end
end
%%% %%% %%% %%% %%% %%% %%% %%% %%% %%% calcul de nbr d'interactions entre les pts ds la conf retenue
SX=0;
for i=1:n
    for j=i+1:n
        for k=1:w

```

```

d(k)=X(i,k)-X(j,k);
end
D=norm(d);
if(D<=R)
    SX=SX+1;
else
    SX=SX;
end
end
end
end
end
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% clacul de vaiogramme

for k=1:n
    s0=X(k,:);
    p=0;
    for q=1:n
        if q~=k
            p=p+1;
            gama0(p)=(rep(k)-rep(q))^2/2;
        end
    end
    p=0;
    for i=1:n
        if i~=k
            p=p+1;
            q=0;
            for j=1:n
                if j~=k
                    q=q+1;
                    M(p,q)=(rep(i)-rep(j))^2/2;
                end
            end
        end
    end
    p=0;
    for i=1:
        if i~=k
            p=p+1;
            repk(p)=rep(i);
        end
    end

```

```

F(p,j)=X(i,j);
end
end
end
m=inv(M);
repp(k)=(gama0'+F*inv(F'*m*F)*(s0-F'*m*gama0'))'*m*repk';
end
s1=0;
s2=0;
sse=0;
for i=1:n
e(i)=rep(i)-repp(i);
s1=s1+rep(i)*rep(i);
s2=s2+repp(i)*repp(i);
sse=sse+e(i)*e(i);
end
moyrep=mean(rep);
moyrepp=mean(repp);
sst=s1-n*moyrep^2;
ssaj=s2-n*moyrepp^2;
r2=1-sse/sst;
fish=(ssaj/(w-1))/(sse/(n-w-1));
r2aj=(sse/(n-w-1))/(sst/(n-1));
if (r2 > 0.7) && (r2 < 1)
nbrplan1=nbrplan1+1;
resultt(nbrplan1,1)=r2;
resultt(nbrplan1,2)=r2aj;
resultt(nbrplan1,3)=sse;
resultt(nbrplan1,4)=fish;
end
end
%xlswrite('res.xls',resultt);
%% %% %% %% %% %% %% %% %% %% %% %% %% %% %% %% %% %% %% %% %% %% %% %% %% %% %% %% %%

```

REFERENCES BIBIOGRAPHIQUES

1. Dejean J.-P., Blanc G. Managing Uncertainties on Production Prediction Using Integrated Statistical Methods. Paper SPE 56696 presented at the 1999 SPE Annual Technical Conference and Exhibition, Houston, U.S.A., October 3-6 (1999).
2. Jourdan A., Zambalas-Mezghani I. Response surface designs for scenario management and uncertainty quantification in reservoir production. *Mathematical Geology* 36 (8), 965-985 (2004).
3. Iman R., Helton J.C. .An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models. *Risk Analysis*, 8, 71-90 (1988).
4. Tarry G. Le problème de 36 officiers, *Compte-rendu de l'Assoc. Français Avanc. Sci. Naturel* 2, 170-203 (1901).
5. Tang B. Orthogonal array-based latin hypercubes. *Journal of the American Statistical Association*, 88, 1392-1397(1993).
6. Johnson M.E., Moore L.M., Ylvisaker D. Minimax and maximin distance designs. *J. of Statist. Planning and Inference* **26**, 131-148(1990).
7. Shewry M.C., Wynn H.P. Maximum Entropy Sampling. *Journal of Applied Statistics*, **14**, 165-170 (1987).
8. Elmoosaoui,H.,Oukid,N. and Hananne, F . "Construction of computer experement designs using marked point processes" ,*Afrika Matematika* ,Doi.Org/10.1007/S13370-020-00770-9,Springer (2020).
9. Elmoosaoui, H . « Contribution à la méthodologie de la recherche expérimentale » Thèse d'état, université Blida 1(2020).
10. Friedman J. Multivariate Adaptative Resgression Splines (Invited paper) .*Annals of statistics* , 101-141p (1991).
11. Kooperberg C., Bose S., Stone C.J. Polychotomous Regression, *Journal of the American Statistical Association*, **92**, 117-127 (1997).
12. Walter, T. 'Plans d'expériences :Construction et analyse statistique'', Edition Springer (2010).
13. Searle, S. R., "Linear Models", John Wiley & Sons, New York, (1971).

14. Dodge, Y. et Rousson, V., "Analyse de régression appliquée", Dunod, n° ISBN : 2-10-048659-4, (2004).
15. Chauvet, P. Aide mémoire de la géostatistique linéaire. Cahiers de Géostatistique, Fascicule 2. Ecole Nationale Supérieure des Mines de Paris, Centre de Géostatistique, Fontainebleau (1999).
16. Hengl, T., Geuvelink, G. et Stein, A. Comparison of kriging with external drift and régression-kriging. Technical note, ITC. Disponible en ligne : http://www.itc.nl/library/Papers_2003/misc/hengl_comparison.pdf (Page consultée le 28 avril 2005)(2003).
17. Nalder, I. A. et Wein, R. W. Spatial interpolation of climatic normals : test of a new method in the canadien boreal forest. *Agricultural and Forest Meteorology*, 92:211{225 (1998).
18. Phillips, D. L., Dolph, J. et Marks, D. A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agricultural and Forest Meteorology*, 58:119{141(1992).
19. Hessami, M. Comparison of soft computing systems for the post-calibration of weather radar. Thèse de doctorat, Université Laval, Québec (2002).
20. Cressie, N. A. C. *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons Inc., New York. Revised reprint of the 1991 edition, A Wiley-Interscience Publication(1993).
21. Matheron, G. Le krigeage universel. Les cahiers du Centre de morphologie mathématique de Fontainebleau, Fascicule 1. Ecole de Mines de Paris, Fontainebleau (1969).
22. Matheron, G. *Traite de géostatistique appliquée*, Tome I. Mémoires du Bureau de Recherches Géologiques et Minières, No.14. Editions Technip, Paris (1962).
23. Matheron, G. Principles of geostatistics. *Economic Geology*, 58:1246{1266 (1963a).
24. Matheron, G. *Traite de géostatistique appliquée*, II : Le Krigeage. Mémoires du Bureau de Recherches Géologiques et Minières, No.24. Editions B. R. G. M., Paris (1963b).
25. Matheron, G. *Les variables régionalisées et leur estimation*. Masson, Paris (1965).

26. Matheron, G. La théorie des variables régionalisées, et ses applications. Les cahiers du Centre de morphologie mathématique de Fontainebleau, Fascicule 5. Ecole des Mines de Paris, Fontainebleau (1970).
27. Stone C.J., Hansen M.H., Kooperberg C., Truong Y.K. Polynomial splines and their tensor products in extended linear modeling. *Ann. Statist.* **25**, 1371-1470 (1997).
28. Isaaks, E. et Srivastava, R. *Applied Geostatistics*. Oxford University Press, New York (1989).