

الجمهورية الجزائرية الديمقراطية الشعبية

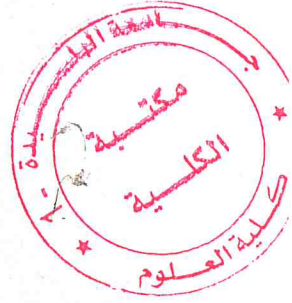
République Algérienne Démocratique et Populaire  
Ministère de L'enseignement Supérieur et de la Recherche Scientifique

Université Saad Dahlab 1  
Faculté des sciences  
Département de mathématiques

Mémoire de fin d'étude en vue de l'obtention d'un Master en  
Mathématiques Appliqués

Option :

MODELISATION STATISTIQUE ET STOCHASTIQUE



*Méthodologie de la recherche expérimentale.*

*PLANS EN PARCELLES DIVISEES*

Réalisé par :

Proposé et encadré par :

Abdel Aziz Sid Ahmed  
Guerfaoui Houssam eddine

M. El Mossaoui.H

Promotion 2017-2018.

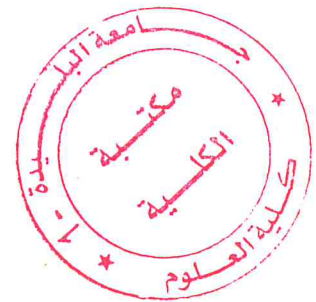
## Remerciement

Ce mémoire n'aurait pas été réalisé sans l'intervention, d'un grand nombre de personnes, nous souhaitons ici les en remercier.

Nous tenons d'abord à remercier très chaleureusement Mr ELMOSSAOUI HICHEM, notre promoteur, pour son sujet de projet de fin d'étude qui nous a passionné, et également pour sa précieuse aide et ses conseils, la patience, la confiance qu'il nous a témoignés ont été déterminants dans la réalisation de notre travail.

Nous voudrions également remercier les jurés pour avoir accepté d'examiner et de juger notre travail. Nos remerciements s'étendent également à tous nos enseignants durant les années des études. Enfin, nous adressons nos plus sincères remerciements à tous nos proches et amis, qui nous ont toujours encouragées au cours de la réalisation de ce mémoire.

Merci à tous et à toutes.



# Table des matières

1.	CHAPITRE I .....	3
1.1.	Intérêts des plans d'expériences .....	3
1.2.	Vocabulaire de base des plans d'expérience .....	4
1.2.1.	Réponses, facteurs .....	5
1.2.2.	Niveaux .....	7
1.2.3.	Domaine de variation du facteur .....	8
1.2.4.	Espace expérimental.....	8
1.2.5.	Domaine d'étude .....	9
1.2.6.	Variables centrées réduites .....	10
1.3.	Les domaines d'applications des plans d'expériences .....	11
1.4.	Différents type de plans d'expériences .....	11
1.4.1.	Plans factoriels .....	12
1.4.2.	Plans pour surface de réponse .....	15
1.5.	Modèles statistiques.....	16
1.5.1.	Modèle linéaire.....	17
1.5.2.	Le modèle linéaire mixte.....	20
1.6.	Critère des plans optimaux .....	22
1.6.1.	Critère D : optimisation de la qualité de l'information .....	22
1.6.2.	Critère A : optimisation de la qualité des coefficients du modèle .....	23
1.6.3.	Critère G : optimisation de la qualité de prévision du modèle.....	23

<b>CHAPITRE 02</b> .....	25
Les plans en parcelles divisées .....	25
2.1. La structure du modèle et de la variance .....	28
2.2. Maximum de vraisemblance : .....	29
2.3. La matrice d'information de MLE (ou bien EVM) : .....	31
2.4. Les Structures des plans .....	34
2.4.1. Facteurs .....	34
2.4.2. Unité expérimentale .....	34
2.4.3. Unités d'observation : .....	35
2.4.4. Unités structurelles : .....	35
2.4.5. Traitements observationnels .....	35
2.4.6. Traitements structurels .....	35
3. <b>CHAPITRE 3</b> .....	37
3.1. Modèle et structure de la variance : .....	37
3.2. Optimalité de conception et génération .....	42
4. <b>CHAPITRE 4</b> .....	44
4.1. Génération de conception : .....	44
4.2. Liste des conceptions possibles : .....	45
4.3. un algorithme : .....	46
4.4. Quelques conceptions D- optimales .....	47
CONCLUSION.....	53

## Table des figures

Figure 1.1: La représentation générale d'un système .....	4
Figure 1.2 : Domaine de variation du facteur.....	7
Figure 1.3 : Définition de l'espace expérimental. ....	8
Figure 1.4: niveau des facteurs définissant des points expérimentaux dans l'espace expérimental centrée. ....	9
Figure 1.5 Deux facteurs à deux niveaux chacun. Figure 1.6 Deux facteurs à trois niveaux chacun.....	9
Figure 1.7 : Région avec restriction .....	10
Figure 4.1 la région D-optimal de la première conception.....	47
Figure 4.2 la région D-optimal de cas de 6 parcelles entières de taille 2 .....	48
Figure 4.3 la région D-optimal de cas de 6 parcelles entières de taille 5 .....	49
Figure 4.4 la région D-optimal de cas de 12 parcelles entières de taille 2.....	50
Figure 4.5 la région D-optimal de cas de 12 parcelles entières de taille 5 .....	51

## Liste des tableaux

Tableau 1.1 Exemple de plan complet pour deux facteurs.....	12
Tableau 1.2 : Nombre d'essai des plans complets suivant le nombre de facteurs à 2 niveaux .....	13
Tableau 2.1 Sous-bloc (whole plot) 1 : Cuisson à température 1. Le sous-bloc (whole plot) 1 représente tous les plateaux de brownies cuits à la première température. Les sous-blocs (whole plot) correspondent à chaque plateau de brownies.....	26
Tableau 2.2 Sous-bloc (whole plot) 2 : Cuisson à température 1. Il s'agit d'une réplique du premier sous-bloc (whole plot). .....	27
Tableau 2.3 Sous-bloc (whole plot) 3 : Cuisson à température 2. Le sous-bloc (whole plot) représente tous les plateaux de brownies cuits à la deuxième température. Les sous-blocs (whole plot) correspondent à chaque plateau de brownies.....	27
Tableau 2.4 Sous-bloc (whole plot) 4 : Cuisson à température 2. Il s'agit d'une réplique du troisième sous-bloc (whole plot).....	27
Tableau 4.1 : représentation de l'espace des conceptions.....	45
Tableau 4.2 : les régions d'optimalité .....	47
Tableau 4.3 : cas de 6 parcelles entières de taille 2.....	48
Tableau 4.4 : cas de 6 parcelles entières de taille 5.....	49
Tableau 4.5 : cas de 12 parcelles entières de taille 2.....	50
Tableau 4.6 : cas de 12 parcelles entières de taille 5.....	51

## Liste des indices :

<i>L'indice</i>	<i>Définition</i>
$e_i$	L'erreur aléatoire
$\tau_i$	Les effets aléatoires
$y_i$	Les réponses
$x_i$	Les facteurs de nuisance
$Z$	L'espace paramétrique
$A$	Un variable d'origine
$X$	Un variable centrée réduite
$A_0$	La valeur centrale en unité courante
$k_i$	Le nombre des niveaux
$l_{x_i}$	Le niveau de facteur $x_i$
$N$	Nombre totale d'expériences
$f^t(x \setminus e_i)$	le vecteur associé au modèle linéaire choisi par rapport à la $i^{ième}$ expérience
$\beta$	Vecteur des coefficients

$\epsilon_i$	L'erreur aléatoire pour la $i^{\text{ème}}$ réponse
$var(..) = \sigma^2$	La variance
$E(...)$	L'espérance mathématique
$V$	la matrice de variance des réponses
$\sigma_i^2$	les composantes de la variance
$I_n$	La matrice d'identité
$\zeta$	Le vecteur des facteurs aléatoires
$g(\zeta \setminus e_j)$	Les vecteurs associant aux effets aléatoires
$Det(...)$	Le déterminant d'une matrice
$trace(...)$	La trace d'une matrice
$({}^tXX)$	La matrice d'information
$({}^tXX)^{-1}$	La matrice de dispersion
$C_{ii}$	Les coefficients de la variance
$Z_j$	La matrice indicatrice associée à la $j^{\text{ème}}$ composante de la variance



$L(\theta, v y)$	La fonction de vraisemblance
$\theta$	Le vecteur des paramètres fixées
$M   \frac{\theta}{\sigma^2}  $	La matrice d'information
$t_i$	Les traitements
$r_i$	Le nombre de répétition du $i^{\text{ème}}$ niveau du facteur WP
WP	Le facteur difficile à changer
SP	Le facteur facile à changer
$y_{ijk}$	La réponse de la $k^{\text{ème}}$ répétition du $i^{\text{ème}}$ niveau du facteur WP et le $j^{\text{ème}}$ niveau du facteur SP
$\alpha_i$	Est l'effet du $i^{\text{ème}}$ niveau du facteur WP
$\beta_j$	Est l'effet du $j^{\text{ème}}$ niveau du facteur SP
$e_{ijk}$	Le terme d'erreur correspondant à l'individu $y_{ijk}$
$\gamma_{ik}$	Le terme d'erreur de la $k^{\text{ème}}$ répétition du $i^{\text{ème}}$ niveau du facteur WP

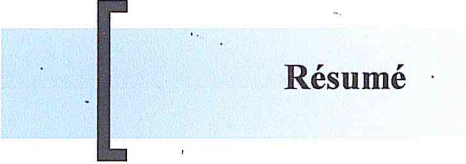
a	Le niveau de facteur WP
b	Le niveau de facteur SP
$J_n$	Matrice des 1
$1_n$	Vecteur des 1
sesq (.)	la somme des éléments carrés de la matrice entre parenthèses
$\Sigma^T$	Vecteur des composants de la variance
MIVC	Les composantes principales des effets et de la variance d'interaction
WSEVC	les composantes de variance d'erreur de parcelle et de sous-parcelles

## ملخص

نهدف من خلال هذه الدراسة إلى نمذجة ومحاكاة تجارب وعلى أرض الواقع على شكل معادلات رياضية من أجل توقع النتائج مسبقا وأيضا من أجل تخفيض التكاليف التي تكون باهظة أحيانا عند إجراء عدد كبير من التجارب.

لإجراء أي تجربة عموما توجد مجموعة من العوامل و المتغيرات بمستويات مختلفة , لكن في بعض الأحيان نصطدم ببعض العراقيل أثناء إجراء التجارب أين يكون تغيير مستويات أحد العوامل أو بعضها أصعب من الأخرى و هنا يبرز مفهوم "plan en parcelles divisées" الذي هو محور دراستنا. بحيث تنقسم العوامل إلى قسمين : أولها العوامل الصعبة التغيير و ثانيها العوامل السهلة التغيير .

خلال هذه الدراسة اهتمنا بشكل أكبر بالحالة التي يكون فيها عامل صعب التغيير و اخر سهل التغيير لأنها الحالة الأكثر وجودا على أرض الواقع هذا من جهة و لأن الحالات التي يكون فيها أكثر من عامل صعب التغيير تصبح أكثر تعقيدا و نمذجتها تزداد صعوبة.



## Résumé

L'objectif de cette étude est de modéliser et de transformer des expériences réelles à des équations mathématiques afin de prédire les résultats, ainsi que pour réduire les coûts qui peuvent être parfois élevés surtout lorsqu'on réalise un grand nombre d'expériences.

Généralement pour réaliser une expérience, on possède un ensemble de facteurs avec des niveaux différents, mais parfois, nous rencontrons des obstacles en faisant des expériences où le changement d'un niveau de certains facteurs est plus difficile que d'autres, ce qui met en évidence le concept de plan en parcelles divisées, qui est au centre de notre étude. Ces facteurs sont divisés en deux catégories : les facteurs difficiles à changer et les facteurs faciles à changer.

Au cours de cette étude, nous sommes, plus intéressés par la situation dans laquelle nous avons un seul facteur difficile à changer et un seul facteur facile à changer, car il s'agit d'une situation très présente dans les cas réels et aussi parce que les situations dans lesquelles plusieurs facteurs sont difficiles à changer deviennent plus complexes et la modélisation devient plus difficile.

## INTRODUCTION GÉNÉRALE

Le terme plan d'expériences vient de l'anglais (Design of experiments) qui se traduit par (Conception des expériences). La méthode des plans d'expériences englobe aussi bien la séquence d'essais à réaliser que l'analyse algébrique et statistique des résultats. L'efficacité d'un plan d'expériences en terme d'informations acquises et de la fiabilité de ces informations, est supérieure à toute autre séquence d'essais non structurée de même volume.

En effet, les plans d'expériences servent à optimiser l'organisation des essais. Cette organisation permet de maximiser le ratio (informations recueillies / nombre d'essais réalisés). Elle permet aussi d'obtenir la meilleure précision possible sur la modélisation des résultats.

Cette méthode est basée sur des règles mathématiques strictes, et exige une démarche rigoureuse de la part de l'expérimentateur. Elle encourage le travail de groupe, et la mise en commun et la synergie du savoir et du savoir-faire, et la réflexion préalable à l'action.

Les plans d'expériences peuvent être classés en plusieurs catégories : plans de criblage, plans d'optimisation, plans de Mélanges. ..etc. Dans notre travail, nous nous intéressons plus particulièrement aux plans en parcelles divisées. Le plan standard en parcelles divisées se caractérise par deux tailles d'unités expérimentales. Les plans en parcelles divisées sont apparus dans l'agriculture où il s'agissait généralement d'appliquer un facteur à une grande parcelle de terre (par exemple, un traitement de fertilisation et de paillis). Ce facteur était appelé facteur de sous-blocs (ou facteur des grandes parcelles). Un autre facteur était ensuite appliqué à l'intérieur de ce sous-bloc (par exemple, diverses variétés de semences). Ce facteur était appelé facteur de sous-parcelles (ou facteur des petites parcelles). Nos deux unités expérimentales dans ce cas sont le sous-bloc et la sous-parcelles. Dans la mesure où nous avons deux tailles d'unités expérimentales, nous avons deux sources d'erreur expérimentale. Cette source d'erreur supplémentaire va affecter les tests d'hypothèses ultérieurs que nous serons amenés à réaliser.

Mais les plans en parcelles divisées sont utilisés bien au-delà du simple domaine agricole dans lequel ils ont été conçus initialement. Il n'est pas rare de rencontrer ce type de

plan dans le domaine industriel où l'expérimentateur dispose de deux facteurs, dont l'un est considéré difficile à changer tandis que l'autre est facile à changer. Le facteur difficile à changer n'est pas modifié à chaque essai en raison des difficultés, alors que le facteur facile à changer est au contraire modifié à chaque essai.

Dans le premier chapitre, nous donnerons les principes de base nécessaires à la compréhension de la théorie des plans d'expériences, ces principes sont présentés à partir de la notion d'espace expérimental. La représentation géométrique des points expérimentaux est très parlante mais elle est vite limitée lorsque la dimension de l'espace est supérieure à trois. C'est pourquoi l'on utilise la représentation matricielle.

Dans le chapitre deux, les plans en parcelles divisées seront abordés. Dans ce même chapitre nous détaillons l'écriture du modèle général ainsi la méthode d'estimation de vraisemblance.

Le troisième chapitre sera consacré à l'étude d'un cas particulier, il s'agit le cas où nous avons un seul facteur difficile à changer et un seul facteur facile à changer.

Enfin dans le dernier chapitre, nous présentons les résultats de notre application sur l'étude de quelques cas pratiques avec des résultats graphiques indiquant la région D-optimale.



## CHAPITRE I

### INTRODUCTION A LA METHODE DES PLANS D'EXPERIENCES

La méthodologie de la recherche expérimentale est l'étude de l'ensemble des stratégies proposées pour résoudre un problème posé par la recherche expérimentale afin d'en déduire un sous ensemble permettant effectivement de résoudre ce problème. Le chercheur choisit dans ce sous ensemble, la stratégie la plus pertinente, compte tenu de ses contraintes économiques et autres.

#### 1.1. Intérêts des plans d'expériences

Les plans d'expériences sont utilisés dans les études industrielles en recherche 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 aux buts suivants :

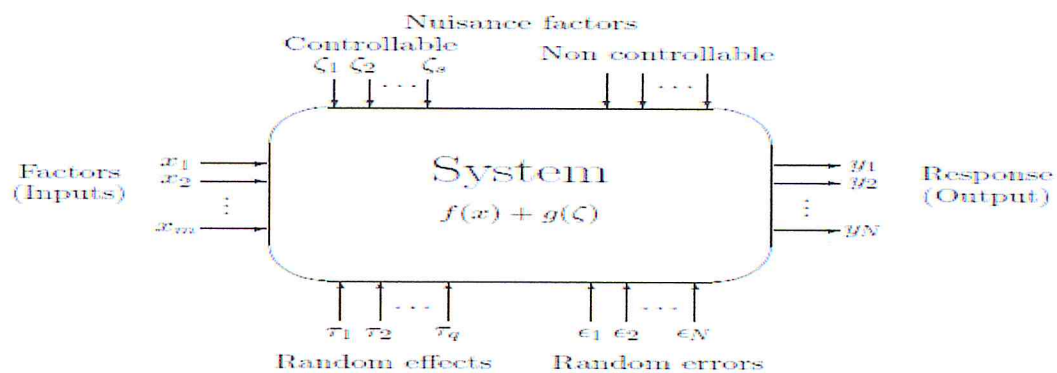
- Détermination des facteurs clés dans la conception d'un nouveau produit ou d'un nouveau procédé
- Optimisation des réglages d'un procédé de fabrication ou d'un d'appareil de mesure
- Prédiction par modélisation du comportement d'un procédé

Les plans d'expériences s'inscrivent dans une démarche générale d'amélioration de la qualité. Le succès de la démarche originale des plans d'expériences réside dans la possibilité d'interprétation de résultats expérimentaux avec un effort minimal sur le plan expérimental : la

minimisation du nombre nécessaire d'expériences permet un gain en temps et en coût financier.[1].

## 1.2. Vocabulaire de base des plans d'expérience

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. [1]



**Figure 1.1:** La représentation générale d'un système

Parmi les facteurs on distinguera :

- Les facteurs contrôlables qui dépendent directement du choix du technicien (choix d'un matériau, d'une pression,...).
- Les facteurs non contrôlables qui varient indépendamment des choix du technicien (conditions climatiques, environnement d'utilisation,...)
- Les facteurs d'entrées telles que la matière première dans un processus de fabrication.



Il n'est pas simple de réaliser de telles relations entre réponses et facteurs, surtout s'il existe des interactions entre les facteurs, les plans d'expériences permettent de réaliser ce type de relation, en minimisant le nombre d'expériences tout en maximisant la précision du résultat.

La science statistique est impliquée dans le processus d'expérimentation parce que tous les systèmes ont un élément très important: La variation, La variation totale est causée par différentes sources, chacune contribue une quantité différente et leurs effets ne peuvent pas être négligés. Quand le système est contrôlé et stable, on s'attend à ce que la variabilité soit aléatoire.

Une variation complètement aléatoire est appelée erreur aléatoire. Et elle est associée aux variables qui ne sont pas sous le contrôle de l'expérimentateur. Dans la Figure 1.1, les flèches entrants étiquetés  $e_1, e_2, \dots, e_n$  représentent les erreurs aléatoires. Le concept technique est que l'impact de ses sources sur les sorties du système est inconnu. Mais il est connu que les résultats expérimentaux sont influencés par eux. Par conséquent, les sorties sont considérées comme variables aléatoires.

Une expérience consiste à appliquer des apports spécifiques au système à étudier, et obtenir des données de sortie en mesurant une caractéristique choisie. Chaque mesure forme un point de données où échantillon appelé observation. Le résultat général est connu sous le nom réponse. Les réponses comme l'indique la Figure 1.1 sont désignées par  $y_i, i = 1, \dots, n$ . Notez que chaque  $y_i$  a une erreur aléatoire incorporée et elle peut avoir des effets aléatoires supplémentaires  $\tau_i, i = 1, \dots, q_i$ . Dans ce cas, plus qu'une réponse peut être mesurée. Par l'utilisation de méthodes statistiques appropriées, l'information contenue dans la réponse mesurée peut être extraite et des connaissances sur le système sont acquises. Et des décisions éclairées peuvent alors être faites basées sur cette analyse.

### 1.2.1. Réponses, facteurs

La grandeur d'intérêt mesurée à chaque essai est appelée la réponse. La valeur de cette réponse, dépend de plusieurs variables. Au lieu de terme variable nous emploierons le mot facteur

Il existe deux types de facteurs, facteur qualitatifs et quantitatifs :

- **Un facteur quantitatif** : se traduit par une grandeur numérique mesurable, ce sera une vitesse, une température, une pression, une intensité, une résistance....
- **Un facteur qualitatif** : ne peut pas donner lieu à quantification directe, on peut seulement identifier ses différents niveaux, ce sera une marque, un procédé, une méthode, un fournisseur, un niveau de qualité. Les facteurs qualitatifs (ou quantitatifs) peuvent être soit des facteurs contrôlés ou non contrôlés. Un facteur contrôlé est un facteur que l'on peut maîtriser, gérer, modifier. On peut donc en fixant volontairement les valeurs.

Exemple des facteurs contrôlés :

- la température d'un four.
- Le type de fournisseur.

Exemple des facteurs non contrôlés :

- La température ambiante.
- Le taux d'hydrométrie
- L'hétérogénéité d'une éprouvette.

Les facteurs non contrôlés seront, soit des facteurs considérés comme non retenus pour l'étude car non influents et laissés à leurs niveaux habituels, soit des facteurs inconnus ou non discernables que l'on subit au cours de l'expérimentation.

Généralement, des multiples facteurs sont impliqués et une expérience correctement conçue doit permettre d'étudier leurs effets sur la réponse simultanément. Les facteurs et leurs effets peuvent être distingués comme suit :

- **Fixe**: Quand il est clair quels niveaux doivent être utilisés en raison des intérêts spécifiques de l'étude, Et les valeurs assignées à ces niveaux ne changeront pas pendant l'expérience, alors ce facteur est défini comme fixé (réparé) et il contribue des effets fixes à la réponse.

- **Aléatoire** : Si on peut considérer que les niveaux de facteur inclus dans l'expérience sont seulement un échantillon aléatoire d'une population de niveaux, donc ce facteur et ses effets sont définis comme aléatoire.

On distingue aussi, une distinction supplémentaire est entre des facteurs de nuisance et les facteurs d'entrés qui sont explicitement étudiés et leurs niveaux doivent être ajustés durant l'expérience. Dans la Figure 1.1 sont désignés par  $x_i (i = 1, \dots, k)$ . Les facteurs de nuisance entrent aussi dans le système, mais ne peuvent pas être changés constamment, ils peuvent être subdivisés en deux type contrôlable c'est le cas où le facteur de nuisance ne contribue pas à l'erreur aléatoire et incontrôlable, c'est le cas où leurs effets s'élèvent à l'erreur aléatoire.

On peut aussi défini un facteur comme étant un facteur **difficile à changer** c'est les cas où son changement est expansif, ou bien **facile à changer** dans le cas contraire. La présence d'un facteur difficile à changer nécessite l'utilisation d'une technique de blocks appelée **parcelles divisées** en anglais on dit **split plot**(voir le chapitre suivant )

### 1.2.2. Niveaux

C'est la valeur donnée à un facteur pour réaliser un essai lorsque l'on étudié l'influence d'un facteur, en général, on limite ses variations entre deux bornes (borne inférieure c'est le niveau bas, borne supérieur c'est le niveau haut). [2]

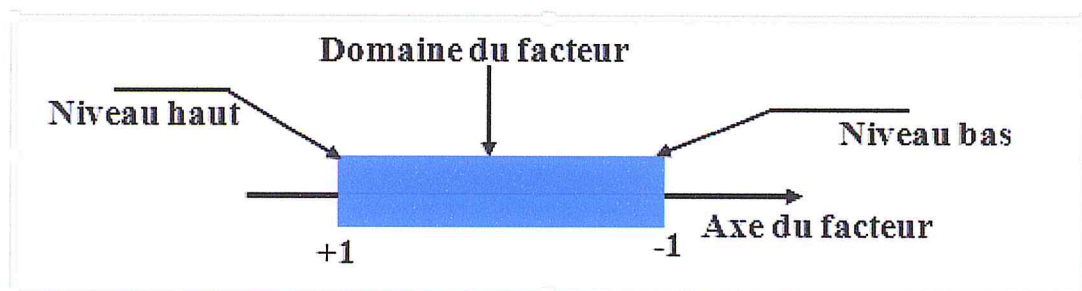


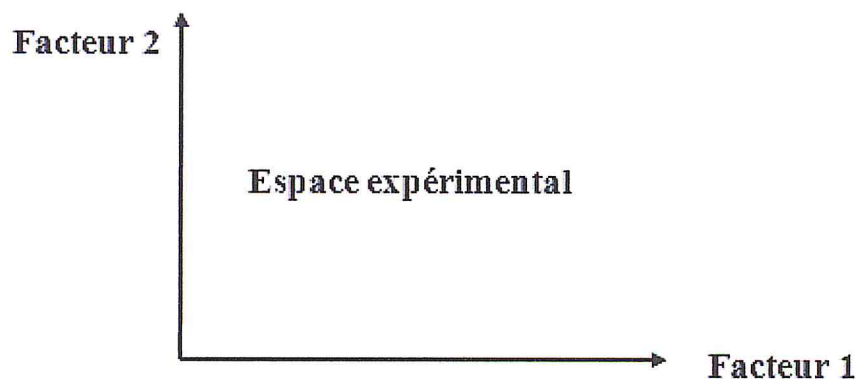
Figure 1.2 : Domaine de variation du facteur.

### 1.2.3. Domaine de variation du facteur

Est l'ensemble de toutes les valeurs qui peut prendre le facteur entre le niveau bas et le niveau haut.

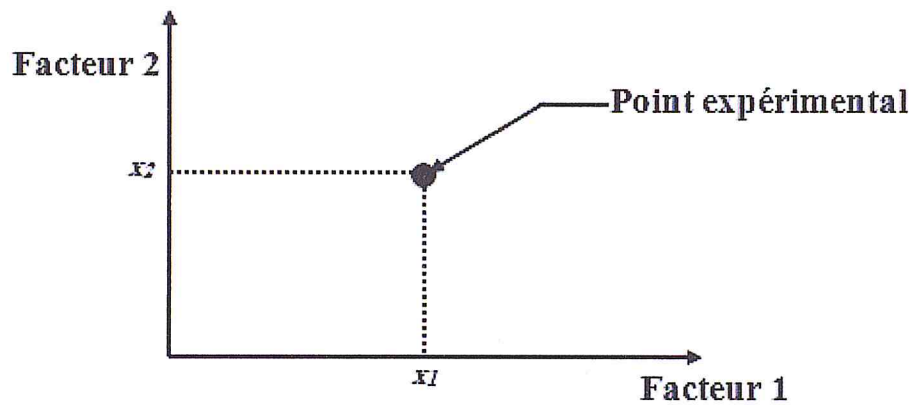
### 1.2.4. Espace expérimental

Dans le cas où en a un second facteur, il est représenté lui aussi par un axe et on définit son niveau haut, son niveau bas et son domaine de variation, ce second axe est disposé orthogonalement au premier. On obtient ainsi un repère cartésien qui définit un espace euclidien à deux dimensions, c'est l'espace expérimental.



**Figure 1.3 :** Définition de l'espace expérimental.

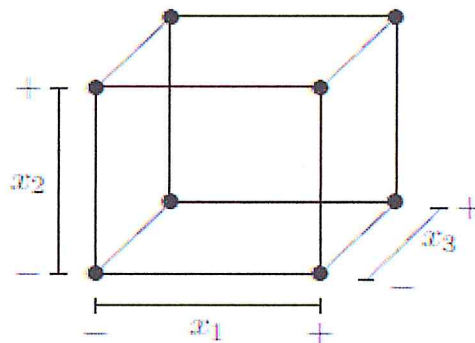
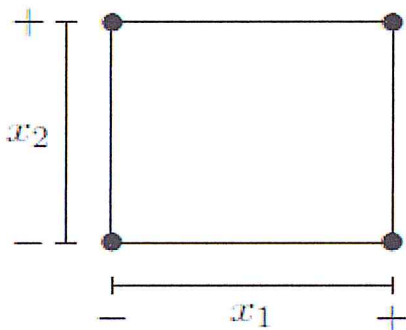
Le niveau  $x_1$  du facteur 1 et le niveau  $x_2$  du facteur 2 et peuvent être considérés comme les coordonnées d'un point de l'espace expérimental.



**Figure 1.4:** niveau des facteurs définissant des points expérimentaux dans l'espace expérimental centrée.

### 1.2.5. Domaine d'étude

Le domaine d'étude est la réunion des domaines de chaque facteur, ce domaine d'étude est la partie de l'espace expérimental retenu par l'expérimentateur pour faire ses essais. Dans le cas des facteurs fixés la région du domaine d'étude est notée par, cette région est limitée par le niveau bas et haut de chaque facteur, par exemple dans le cas de deux facteurs, la région est rectangulaire et sera cubique pour trois facteurs (voir Figure 1.2):



**Figure 1.5** Deux facteurs à deux niveaux chacun. **Figure 1.6** Deux facteurs à trois niveaux chacun.

Pour des facteurs aléatoires le concept analogue est appelé espace paramétrique. Il sera noté  $Z$ . Il se réfère aux valeurs des composants de variances. Il est défini par l'ensemble des valeurs qui vont être considérées acceptables pour l'expérience selon la quantité (le montant) de connaissance du système. Souvent, il y a des restrictions des facteurs ou de certains traitements, si les niveaux extrêmes des facteurs ne peuvent pas être utilisés en même temps dans ce cas le domaine d'étude sera irrégulier, considérons par exemple le cas de deux facteurs où l'un doit être plus grand que l'autre. Dans ce cas la région a une forme triangulaire. Pour le cas où  $x_1 > x_2$  la région de conception limitée est montrée par un triangle gris clair dans la Figure 1.3. La seconde possibilité est lorsque  $x_2 > x_1$  et la région apparaît par le triangle gris foncé sur la même figure. L'approche de conception optimale est un outil puissant pour concevoir des expériences avec des régions de conception limitées;

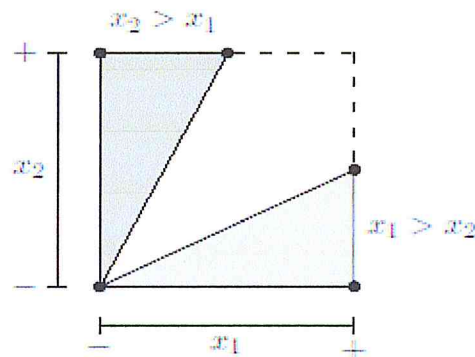


Figure 1.7 : Région avec restriction

### 1.2.6. Variables centrées réduites

Lorsque l'on attribue la valeur -1 au niveau bas d'un facteur et la valeur +1 au niveau haut, on effectue deux modifications :

- **On change l'unité de mesure** : par exemple, si le niveau bas d'un facteur est 10°C et le niveau haut 30°C, il y a 20°C entre ces deux valeurs, soit 20 fois l'unité courante

de température. Entre -1 et +1 il y a deux unités nouvelles, la nouvelle unité vaut donc 10°C. On le donne le nom de *pas*.

• **On déplace l'origine des mesures** : dans l'exemple choisi le milieu de l'intervalle [-1, +1] correspond à une température de 20°C. La nouvelle origine, notée zéro, diffère donc de l'origine exprimée en unités courantes [7].

Ces deux modifications entraînent l'introduction de nouvelles variables que l'on appelle **variables centrées réduites** : Centrées pour indiquer le changement d'origine et réduites pour signaler la nouvelle unité.

Le passage des variables d'origine *A* aux variables centrées réduites *X*, et Inversement, est donné par la formule suivante :

$$X = \frac{A - A_0}{pas}$$

*A*<sub>0</sub> : est la valeur centrale en unité courante.

L'intérêt des variables centrées réduites est de pouvoir présenter les plans d'expériences de la même manière quels que soient les domaines d'étude retenus et quels que soient les facteurs [2].

### 1.3. Les domaines d'applications des plans d'expériences

Les plans d'expériences sont habituellement appliqués dans la recherche et le développement de nouveaux produits ou procédés. Les secteurs d'activité concernés peuvent être aussi différents que l'agronomie, la chimie, l'électronique, la mécanique ou le marketing, ce qui est nouveau, c'est leur application à la qualité.

### 1.4. Différents type de plans d'expériences

Il existe deux types de plans d'expériences :

- les plans factoriels,
- les plans pour surface de réponse.

#### 1.4.1. Plans factoriels

Les plans factoriels sont des plans discrets orthogonaux. Les facteurs sont discrétisés pour ne prendre qu'un nombre fini de valeurs, les niveaux. L'orthogonalité d'un plan se définit de la manière suivante :

**Définition 1** (Orthogonalité) : Un plan d'expériences est orthogonal si :

- à 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,
- chaque niveau de chaque facteur apparaît le même nombre de fois. [2]

**Exemple1** propose un plan factoriel à deux facteurs  $X_1$  prend 3 niveaux et  $X_2$  prend 2 niveaux.

En pratique, dans ces plans, les facteurs prennent 2 à 5 niveaux. Il existe deux types de plans factoriels :

- les plans complets,
- les plans fractionnaires.

**Tableau 1.1**Exemple de plan complet pour deux facteurs

#	$X_1$	$X_2$
1	-1	-1
2	-1	1
3	0	-1
4	0	1



5	1	-1
6	1	1

**Tableau 1.2 :** Nombre d'essai des plans complets suivant le nombre de facteurs à 2 niveaux

Nombre de facteurs	Nombre d'essai
2	4
3	8
4	16
5	32
10	1024
20	1048576

Les plans factoriels complets sont des plans orthogonaux dont les facteurs prennent 2 ou 3 niveaux et où toutes les combinaisons possibles entre tous les niveaux des facteurs sont évaluées. Le plan du Tableau 1.1 est un plan complet de deux facteurs respectivement à 3 et 2 niveaux. Lorsqu'on dispose de  $d$  facteurs à 2 niveaux, le plan complet est noté  $2^d$ . Pour 3 niveaux, on le note  $3^d$ . Si on a  $d_1$  facteurs à 2 niveaux et  $d_2$  facteurs à 3 niveaux, on note le plan  $2^{d_1} 3^{d_2}$ . Ces plans permettent d'estimer les effets des facteurs et toutes les interactions mais conduisent rapidement à un nombre important d'expériences, ce qui peut rendre la méthode coûteuse. Le Tableau 1.2 donne le nombre d'essais nécessaires pour effectuer un plan complet avec des facteurs à 2 niveaux. En pratique, ces plans 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 (cf. Tableau 1.1)

Parmi les plans fractionnaires, on trouve

- les plans de Plackett et Burman,
- les plans de Box,
- les plans de Taguchi.

Les plans de Plackett et Burman sont des plans restrictifs puisqu'ils ne permettent pas de prendre en compte les interactions. Ce sont donc des plans de résolution III où les effets des facteurs sont confondus avec les interactions. Le lecteur intéressé par ce type de plan pourra se référer à (Plackett and Burman, 1946).

Les plans de Box sont des plans plus récents. Notés  $2^{d-k}$ ,  $k < d$ , correspondant au nombre d'expériences pour  $d$  facteurs à 2 niveaux, ces plans peuvent être de la résolution III pour les plus grands  $k$  à la résolution VII pour les plus petits  $k$ . (Box et al., 1978) ont établi un tableau des différents plans possibles pour des facteurs à 2 niveaux (jusqu'à 10 facteurs) avec les alias permettant de construire ces plans. Enfin, les plans de Taguchi sont des plans fractionnaires orthogonaux tabulés. De nombreuses tables sont disponibles dans la littérature mais leur utilisation nécessite un minimum de connaissance sur leur fonctionnement. Le lecteur intéressé pourra se référer à (Sabre, 2007). Beaucoup utilisées en industrie, les tables de Taguchi ont été conçues par le statisticien Genichi Taguchi dans le but de minimiser l'effet des alias et des erreurs de mesure. Dans un plan de Taguchi, les facteurs peuvent prendre un nombre de niveaux constant (tous les facteurs ont le même nombre de niveaux : 2, 3, 4, 5) ou mixte (2 et 3, 2 et 4). Ces plans sont notés  $L_g(m^n)$  ou  $L_g(m^n m^m)$  selon le cas constant ou le cas mixte,  $g$  étant le nombre d'expériences,  $m$  le nombre de niveaux par facteurs et  $n$  le nombre maximum de facteurs que le plan peut prendre en compte. Il existe deux types de tables de Taguchi :

- les tables de type 1 permettant l'identification des interactions,
- les tables de type 2 ne permettant pas l'identification des interactions.

Plus généraux que les plans de Box, les plans de Taguchi sont également plus faciles d'utilisation puisque les plans sont déjà construits. Le choix se fait par rapport au nombre de facteurs, au nombre de niveaux par facteur et à la résolution souhaitée.

#### 1.4.2. Plans pour surface de réponse

Les plans étudiés dans la partie précédente permettent d'établir des modèles linéaires avec interactions, ce qui ne rend pas toujours compte correctement du phénomène étudié. Il existe en effet des situations où de tels modèles vont s'avérer trop pauvres, car la relation entre la sortie et les entrées présentent un comportement quadratique voire non-linéaire. Afin de pallier ce type de problème, on utilise plutôt des plans pour surface de réponse. Il en existe plusieurs types

- plan factoriel complet  $3^d$ .
- plan de Box et Behnken,
- plan composite centré,
- plan D-optimal,
- plan « space-filling » (plans latins, plans maximin, etc).

Le but de ces plans est de modéliser avec précision la relation entre la réponse et les facteurs d'entrée. Ceux-ci sont quantitatifs, discrétisés de manière plus ou moins fine sur leur domaine de variation.

Le plan factoriel complet  $3^d$  est le plan pour surface de réponse le plus simple et le plus naturel à imaginer. On a vu dans la partie sur les plans factoriels que ce plan pouvait être utilisé pour des facteurs discrets. Il peut également être utilisé pour des facteurs continus, à condition de coder les niveaux des facteurs par une base orthogonale de polynômes afin d'assurer l'orthogonalité du plan. Ce type de plans permet d'établir des modèles polynomiaux jusqu'à l'ordre 2 (termes quadratiques et interactions).

Les plans de Box et Behnken sont des fractions du plan complet  $3^d$ . Ils permettent également d'estimer les coefficients de modèles quadratiques, mais avec moins d'expériences.

Les plans composites centrés sont constitués d'un plan factoriel complet ou fractionnaire à 2 niveaux, de points situés sur les axes des facteurs à une certaine distance du centre du domaine et d'expériences au centre du domaine.

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, I, voir le **paragraphe 1.7**). 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. Ces trois types de plans sont étudiés dans (Myers and Montgomery, 2009) et (Pukelsheim, 1993).

Beaucoup utilisés dans le cas des essais numériques, les plans latins hypercubes permettent d'étudier des modèles complexes, non linéaires par exemple. Dans ce type de plan, chaque facteur prend beaucoup de niveaux équi-répartis sur son domaine de variation.

### 1.5. Modèles statistiques

Le modèle est une relation entre les facteurs et la réponse que l'on désire étudier. La détermination d'un modèle approprié, est une partie intégrante dans la conception statistique des plans d'expériences. C'est un effort d'équipe entre l'expérimentateur et le statisticien pour

choisir le meilleur modèle possible. Cette section présente les modèles statistiques les plus utilisés. Des méthodes d'évaluation pour les paramètres impliqués sont décrites dans le chapitre suivant

### 1.5.1. Modèle linéaire

Le modèle statistique linéaire avec des facteurs fixes et des effets fixes a l'utilisation la plus répandue. Il est largement documenté dans la littérature et utilisé dans le cas des plans de criblage et ceux de surface de réponse. [3].

Considérons les cas de  $N$  expériences et  $m$  facteurs à effet fixe, le vecteur  $x = (x_1, \dots, x_m)^t$  vecteur des facteurs. Chaque facteur  $x_i$  possède  $k_i$  niveaux regroupés dans un vecteur  $l_{x_i} = (l_{x_i}^1, \dots, l_{x_i}^{k_i})^t$ . Pour chaque facteur, uniquement un seul niveau est utilisé au cours d'expérience donnée. Donc chaque expérience est un vecteur de  $N$  composantes désigné par  $e_j = (l_1, \dots, l_m)^t$  dont  $j = 1, \dots, N$  et  $l_i \in l_{x_i}$ . Le modèle linéaire pour une expérience  $i$  peut être écrit comme :

$$y_i = f^t(x \setminus e_i) \beta + \epsilon_i \quad i = 1, \dots, N \quad (1.1)$$

Où

$y_i$  : est la réponse mesurée au cours de l'expérience  $e_i$

$f^t(x \setminus e_i)$  : le vecteur associé au modèle linéaire choisi par rapport à la  $i^{\text{ème}}$  expérience.

$\beta$  : vecteur des coefficients

$\epsilon_i$  : l'erreur aléatoire pour la  $i^{\text{ème}}$  réponse.

Le modèle (1.1) s'écrit d'une manière simple en notation matricielle:

$$y = X\beta + \epsilon, \quad (1.2)$$

Avec  $y$  : vecteur des réponses.

$X$  : matrice de calcul, qui dépend des points expérimentaux choisis pour exécuter le plan du modèle postulé. On a  $X = [f(e_1), \dots, f(e_N)]^T$

$\beta$  : vecteur des coefficients.

$e$  : Vecteur des résidus.

A titre d'exemple considérons les cas de deux facteurs chacun possède trois niveaux codés par (-1,0,1) et considérons un modèle quadratique avec interactions:

$$y_i = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_{11} x_1^2 + \theta_{22} x_2^2 + \theta_{12} x_1 x_2 + \varepsilon_i$$

Dans ce cas, nous avons :

$$f(x) = (1 \ x_1 \ x_2 \ x_1^2 \ x_2^2 \ x_1 x_2) \theta = (\theta_0 \ \theta_1 \ \theta_2 \ \theta_{11} \ \theta_{22} \ \theta_{12})^T$$

En considérons toutes les combinaisons possibles entre les niveaux des deux facteurs, alors le nombre d'expérience est 9, la matrice d'expérience sera donnée comme:

$$\begin{array}{l}
 f(x|_{t_1}) = (1 \ -1 \ 1) \\
 f(x|_{t_2}) = (1 \ 0 \ -1) \\
 f(x|_{t_3}) = (1 \ 1 \ -1) \\
 f(x|_{t_4}) = (1 \ -1 \ 0) \\
 f(x|_{t_5}) = (1 \ 0 \ 1) \\
 f(x|_{t_6}) = (1 \ 1 \ 1) \\
 f(x|_{t_7}) = (1 \ 0 \ 0) \\
 f(x|_{t_8}) = (1 \ -1 \ -1) \\
 f(x|_{t_9}) = (1 \ 1 \ 0)
 \end{array}
 \quad
 X =
 \begin{bmatrix}
 1 & -1 & -1 & 1 & 1 & 1 \\
 1 & -1 & 0 & 1 & 0 & 0 \\
 1 & -1 & 1 & 1 & 1 & -1 \\
 1 & 0 & -1 & 0 & 1 & 0 \\
 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 1 & 0 & 1 & 0 \\
 1 & 1 & -1 & 1 & 1 & -1 \\
 1 & 1 & 0 & 1 & 0 & 0 \\
 1 & 1 & 1 & 1 & 1 & 1
 \end{bmatrix}
 \quad
 Y =
 \begin{bmatrix}
 y_8 \\
 y_4 \\
 y_1 \\
 y_2 \\
 y_7 \\
 y_\varepsilon \\
 y_3 \\
 y_9 \\
 y_\epsilon
 \end{bmatrix}$$

Remarquons que ces modèles sont appelés linéaires parce que les paramètres  $\theta$  apparaissent linéaires dans l'équation. Aussi, il y a seulement une partie stochastique qui est

l'erreur aléatoire. Cependant, elle possède certaines propriétés intéressantes sous réserve qu'elles soient satisfaites le mieux possible les conditions suivantes:

Les erreurs  $\varepsilon_i$  doivent être distribuées suivant une loi Normale de moyenne 0 et de variance  $\sigma_\varepsilon^2$ ,  $N(0, \sigma)$ , ce qui s'écrit :

$$\checkmark \quad E(\varepsilon_i) = 0$$

$$\checkmark \quad \text{var}(\varepsilon_i) = \sigma_\varepsilon^2$$

généralement, on peut écrire:

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}_N, \sigma \mathbf{I}_N), \quad (1.3)$$

Sous ces hypothèses, la matrice de variance des réponses s'écrit:

$$\text{var}(\mathbf{y}) = \mathbf{V} = \sigma_\varepsilon^2 \mathbf{I}_N \quad (1.4)$$

Dans le cas où les observations (réponses) n'ont pas la même variance alors de même que précédemment on aura:

$$E(\varepsilon_i) = 0, \quad (1.5.a)$$

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, \quad (1.5.b)$$

$$\text{var}(\varepsilon_i) = \sigma_i^2, \quad (1.5.c)$$

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}_N, \mathbf{V}), \quad (1.5.d)$$

Où, les paramètres  $\sigma_i^2$  sont appelés les composantes de variance, sont généralement inconnus et doivent être estimés. Dans ce cas la matrice de variance covariance s'écrit:

$$V = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_N^2 \end{bmatrix} \quad (1.6)$$

Dans la cas où  $cov(\varepsilon_i, \varepsilon_j) \neq 0$ , alors la matrice de variance covariance peut s'écrire comme :

$$V = \begin{bmatrix} V_1 & 0 & \dots & 0 \\ 0 & V_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & V_r \end{bmatrix} \sigma^2 + \sigma_\varepsilon^2 I_n \quad (1.7)$$

Où, les matrices  $V_1, \dots, V_r$  regroupent les corrélations des réponses associées avec  $\sigma_1, \dots, \sigma_r$  respectivement. et  $\sigma^2 = (\sigma_1^2, \dots, \sigma_r^2)$ .

### 1.5.2. Le modèle linéaire mixte

La forme générale du modèle linéaire mixte est donnée par:

$$y_{ij} = f^T(X|t_i)\theta + g(\zeta|t_i)\tau + \varepsilon_{ij} \quad (1.8)$$

Qui peut être écrit en forme (formulaire) matricielle comme:

$$y = X\theta + Z\tau + \epsilon \quad (1.9)$$

Où,  $i = 1, \dots, N$  est l'indice des observations et  $j = 1, \dots, \nu$  reflète le groupement des observations. Pour la partie stochastique le vecteur des facteurs aléatoires est  $\zeta = (\zeta_1, \dots, \zeta_s)^t$  et  $l_{\zeta_j}$  ( $j = 1, \dots, s$ ) le nombre de niveaux pour le facteur aléatoire  $\zeta_j$ .  $g(\zeta|e_j)$  sont les vecteurs associés aux effets aléatoires, généralement connus comme des vecteurs indicateurs. Le



vecteur  $\tau = (\tau_1, \dots, \tau_r)$  contient les paramètres pour les effets aléatoires qui incluent des effets principaux plus des interactions entre eux.

La matrice  $Z$  est la matrice d'expériences pour la partie stochastique qui est analogue avec la matrice  $X$  dans le cas déterministe. Et il peut être partitionné en matrices  $Z_j$  chacune est associée à un facteur difficile à changer.

Notons  $s$  le nombre de facteurs aléatoires principaux et  $r$  être le nombre total de coefficients aléatoires dans le modèle. Pour chacun on associe une variance  $\sigma_j, j = 1, \dots, s, \dots, r$ . Qui peuvent être regroupés dans un vecteur  $\sigma = (\sigma_1, \dots, \sigma_r)^t$ . Et supposons que  $\tau$  est partitionné en  $\tau_j$  vecteurs où chaque effet dans  $\tau_j$  correspond à un seul niveau des  $l_{\zeta_j}$  niveaux de  $j^{\text{ième}}$  le effet aléatoire. Notons :

$$\tau_0 = \epsilon, l_{s0} = N, \text{ et } Z_0 = I_N$$

Donc, la partie aléatoire du modèle peut être écrite plus particulièrement comme:

$$Z_{\tau} = \sum_{j=0}^r Z_j \tau_j \quad (1.10)$$

Où,

$$\tau_0 = \epsilon, l_{s0} = N, \text{ and } Z_0 = I_N$$

Par conséquent, l'espace des paramètres pour la partie stochastique est représentée par les vecteurs  $\tau$  et  $\epsilon$ .

$$E(\tau_i) = 0 ; \text{ var}(\tau_i) = \sigma_i^2 I_{l_{\zeta_i}} ; \text{ cov}(\tau_i, \tau_j^T) = 0 \text{ pour } i \neq j$$

$$\text{var}(\tau) = \text{diag}(\sigma_0^2 I_{l_{\zeta_0}}, \sigma_1^2 I_{l_{\zeta_1}}, \dots, \sigma_r^2 I_{l_{\zeta_r}})$$

L'espérance et la variance de la réponse peuvent être écrites:

$$E(y) = X_{\theta} \quad (1.11)$$

$$\text{var}(y) = V = \sum_{i=0}^r \sigma_i^2 Z_i Z_i^T \quad (1.12)$$

## 1.6. Critère des plans optimaux

Pour obtenir un plan optimal capable de donner le maximum d'informations, nous allons définir les critères sur lesquels nous nous sommes appuyés pour obtenir un nombre réaliste d'expériences, en tenant compte des contraintes qui peuvent exister entre les facteurs d'entrée. Parmi plusieurs critères de qualité, nous avons testé les 3 principaux les critères D, A et G. [4].

### 1.6.1. Critère D : optimisation de la qualité de l'information

Le critère D est celui qui permet d'optimiser le nombre d'expériences à réaliser. Une matrice d'expériences est dite optimale D si le déterminant de la matrice de dispersion  $({}^tXX)^{-1}$  avec (X) la matrice du modèle, est minimal ce qui veut dire que le déterminant de la matrice d'information  $({}^tXX)$  est maximal puisque

$$\text{Det}(t_{XX}) = \frac{1}{\text{Det}({}^tXX)^{-1}} \quad (1.13)$$

On dira qu'un plan d'expériences est D-optimal si la position des points d'expériences maximise la valeur du déterminant de la matrice d'information. Le D de D-optimal vient donc de déterminant et on parle de D-optimalité.

Plus le nombre d'expériences est élevé plus la matrice d'expériences comporte de points, et plus le déterminant de la matrice d'information est grand. La matrice qui apporte le maximum d'informations selon le critère D est celle qui est constituée de la totalité des points expérimentaux. Il faut donc chercher un nombre réaliste d'expériences (N) qui maximise le déterminant de la matrice d'information, en d'autre termes, il faut chercher le nombre N d'expériences qui couvrent le plus grand volume (ou surface dans le cas de deux facteurs) possible dans la région expérimentale.

Si le déterminant est nul, son inverse n'existe pas et il est impossible de mettre en œuvre la méthode des moindres carrés qui conduit à l'estimation des coefficients du modèle. La valeur du déterminant peut être très grande. La plupart des logiciels indique le logarithme décimal de ce déterminant. Plus le déterminant est grand, plus la qualité d'estimation des coefficients du

modèle est grande car les incertitudes qui affectent les coefficients du modèle et les effets moyens des facteurs sont faibles.

### 1.6.2. Critère A : optimisation de la qualité des coefficients du modèle

La matrice d'expériences est dite optimale si la trace de la matrice de dispersion  $({}^t\mathbf{XX})^{-1}$  est minimale, ce qui veut dire que la somme des variances des coefficients du modèle est minimale.

L'incertitude associée à chacun des coefficients du modèle s'appelle l'erreur-type, on montre que cette erreur est proportionnelle aux termes diagonaux de la matrice de dispersion appelés coefficients de variance et notés  $C_{ii}$ .

La construction d'un plan optimal consiste à extraire les traitements qui minimisent la trace de la matrice de dispersion c'est-à-dire la somme des termes diagonaux :

$$\text{Trace}({}^t\mathbf{XX})^{-1} = \sum_{i=1}^p C_{ii} \quad (1.14)$$

Il existe un lien entre le critère D et le critère A : l'augmentation du premier se traduit par une diminution du second. Pour cette raison, on s'intéresse essentiellement au critère du déterminant

### 1.6.3. Critère G : optimisation de la qualité de prévision du modèle

Les coefficients du modèle étant affectés d'une incertitude, il en est de même pour les prévisions faites à partir du modèle, pour toute combinaison des modalités des facteurs notée symboliquement  $x_0$ .

La variance de prévision de la réponse s'écrit :

$$\text{Var}(\hat{y}(x_0)) = \sigma^2 \{ {}^t f(x_0) ({}^t\mathbf{XX})^{-1} f(x_0) \} \quad (1.15)$$

Avec  $\sigma^2$  la variance expérimentale,  $f(x_0)$  la fonction du modèle pour la combinaison des modalités considérées

La matrice de dispersion étant définie à partir d'une matrice d'expériences donnée, pour une variance expérimentale supposée constante, la variance de prévision de la réponse ne dépend que de la nature de la combinaison des modalités des facteurs. On pourra donc détecter la combinaison pour laquelle la variance de prévision est la plus grande et introduire cette combinaison, comme expérience complémentaire dans la matrice d'expériences. Pour cela, on introduit la fonction de variance standardisée :

$$d(\xi_N, x_0) = \frac{N \cdot \text{Var}(\hat{y}(x_0))}{\sigma^2} \quad (1.16)$$

Cette fonction est calculée systématiquement, pour chacune des combinaisons non retenue, on identifie la combinaison pour laquelle  $d(\xi_N, x_0)$  est maximale. S'il en existe plusieurs, on en retient une au hasard et on l'introduit dans la matrice d'expériences. Le déterminant de la matrice d'information de taille  $N+1$ , s'écrit :

$$\det({}^t X_{N+1} X_{N+1}) = \det({}^t X_N X_N) \left\{ 1 + \frac{d(\xi_N, x_0)}{N} \right\} \quad (1.17)$$

Lorsqu'une matrice d'expériences est optimale, le maximum de la fonction de variance standardisée est égal au nombre d'inconnues  $p$  associées à la forme générale du modèle. On peut donc calculer un nouveau critère :

$$G - \text{efficacité} = \frac{p}{\max\{d(\xi_N, x_0)\}} \quad (1.18)$$

En conclusion, la construction d'un plan optimal nécessite donc :

- un déterminant de la matrice d'information maximal,
- une trace de la matrice de dispersion minimale,
- un coefficient G-efficacité maximal



## CHAPITRE 2

### LES PLANS EN PARCELLES DIVISEES

Un plan en parcelles divisées (split plot) est un plan d'expériences incluant au moins un facteur difficile à changer (hard to change), qu'il n'est pas simple de randomiser complètement en raison de contraintes de temps et de coût. Dans une expérience en parcelles divisées, les niveaux du facteur difficile à changer demeurent constants pour plusieurs essais expérimentaux, qui sont traités comme un sous-bloc. Les facteurs faciles à changer (easy to change) varient sur ces essais, chaque combinaison de ces derniers étant considérée comme une sous-parcelle au sein du sous-bloc (whole plot). Vous devez randomiser l'ordre dans lequel vous exécutez les sous-blocs et les sous-parcelles au sein des sous-blocs. [4]

Historiquement, les plans en parcelles divisées ont été développés à l'origine par Fisher (1925) pour usage dans des expériences agricoles. Sont principalement la forme modifiée de plans en blocs aléatoires. Ces plans sont utilisés dans des situations où la randomisation complète de séries parmi le bloc n'est pas possible.

Ces plans sont largement utilisés dans les expériences industrielles où un ensemble de facteurs peut nécessiter un lot important de matériel expérimental (Whole Plot factors), tandis qu'un autre ensemble de facteurs peut être appliqué dans le cas de lot plus réduit de matériel expérimental (Sub Plot factors). Une autre situation qui conduit à l'usage de plan d'unité divisée (Whole-Plot design) dans l'expérience industrielle est celle où il existe un ou plusieurs facteurs appelés facteurs difficiles à changer, qu'il(s) est (sont) difficile(s) à changer en raison de contraintes de temps et de coût (facteurs Whole-Plot); et les autres facteurs, appelés facteurs faciles à changer (facteurs Split-Plot), ont des modalités faciles à changer.

Généralement, le plan en parcelles divisées (Split-Plot design) peut être utilisé pour tous cas d'expérience faisant intervenir deux types d'unité expérimentale (grande et petite), affectés aléatoirement et indépendamment sur les deux niveaux différents. Le plan optimal pour les expériences en parcelles divisées a reçu une attention de la part de Goos et Vandebroek (2001b, 2003a, 2004). Le travail a été basé sur une structure de plan plus complexe ayant utilisé un

modèle polynomial de premier et de second ordre afin de pouvoir représenter la réaction de celle-ci, mais en général le travail utilisait un modèle linéaire d'effet mixte qui supposait des effets fixes pour les modalités des facteurs des sous-blocs et ceux des parcelles divisées, ainsi que deux composantes de variance associées à l'erreur liée aux sous-blocs et à l'erreur liée aux parcelles divisées. Le but de ce travail de construire des plans D-optimaux pour des estimateurs de maximum de vraisemblance dans une expérience dans le cas de sous blocs (facteur WP), et une autre dans le cas de parcelles divisées (facteur SP), avec l'hypothèse d'effet aléatoire pour les deux facteurs. [5].

### Exemple

Une grande boulangerie élabore une nouvelle recette de brownie. Ils testent deux niveaux de chocolat et de sucre avec deux températures de cuisson différentes. Toutefois, pour gagner du temps, elle décide de faire cuire plusieurs plateaux de brownies à la fois plutôt que de faire cuire chaque plateau individuellement. L'exemple du brownie inclut 2 sous-blocs (whole plot) répétés (4 sous-blocs (whole plot) au total). Chaque sous-bloc (whole plot) comporte 4 sous-parcelles (sub-plot). Le sous-bloc (whole plot) correspond à tous les plateaux de brownies cuits à la température. Les sous-blocs (whole plot) correspondent à chaque plateau de brownies.

**Tableau 2.1** Sous-bloc (whole plot) 1 : Cuisson à température 1. Le sous-bloc (whole plot) 1 représente tous les plateaux de brownies cuits à la première température. Les sous-blocs (whole plot) correspondent à chaque plateau de brownies.

Plateau 1	Plateau 2	Plateau 3	Plateau 4
(Chocolat 1, Sucre 1)	(Chocolat 1, Sucre 2)	(Chocolat 2, Sucre 1)	(Chocolat 2, Sucre 2)

**Tableau 2.2** Sous-bloc (whole plot) 2 : Cuisson à température 1. Il s'agit d'une réplique du premier sous-bloc (whole plot).

Plateau 1	Plateau 2	Plateau 3	Plateau 4
(Chocolat 1, Sucre 1)	(Chocolat 1, Sucre 2)	(Chocolat 2, Sucre 1)	(Chocolat 2, Sucre 2)

**Tableau 2.3** Sous-bloc (whole plot) 3 : Cuisson à température 2. Le sous-bloc (whole plot) représente tous les plateaux de brownies cuits à la deuxième température. Les sous-blocs (whole plot) correspondent à chaque plateau de brownies.

Plateau 1	Plateau 2	Plateau 3	Plateau 4
(Chocolat 1, Sucre 1)	(Chocolat 1, Sucre 2)	(Chocolat 2, Sucre 1)	(Chocolat 2, Sucre 2)

**Tableau 2.4** Sous-bloc (whole plot) 4 : Cuisson à température 2. Il s'agit d'une réplique du troisième sous-bloc (whole plot).

Plateau 1	Plateau 2	Plateau 3	Plateau 4
(Chocolat 1, Sucre 1)	(Chocolat 1, Sucre 2)	(Chocolat 2, Sucre 1)	(Chocolat 2, Sucre 2)

### 2.1. La structure du modèle et de la variance

Le modèle générale (avec des effets mixtes) peut s'écrire comme suit :

$$y_{ij} = f^T(X|_{t_i})\theta + g(\zeta|_{t_j})\tau + \epsilon_{ij} \quad (2.1)$$

$i = 1 \dots N$ : observations,  $j = 1 \dots v$ : groupement (des bloc ou des étapes)

$y_{ij}$ : les réponses,  $f(X)$ : est la relation fonctionnelle pour les facteurs fixées,  $\theta$ : est le vecteur des paramètres fixées (de taille  $d \times 1$ ). Pour la partie aléatoire on a  $\zeta = (\zeta_1 \dots \dots \zeta_s)^T$ : est le vecteurs des facteurs aléatoires et  $l_{\zeta_i}$  représente le niveau de  $\zeta_i$  tel que  $i = 1, \dots, s$ ,  $g(\zeta|_{t_j})$ : est la matrice associé aux effets aléatoires (matrice indicatrice),  $\tau = (\tau_1 \dots \dots \tau_p)^T$ : est le vecteur qui contient les paramètres des effets aléatoire (voir la **figure 1.1**),  $\epsilon_{ij}$ : l'erreur de l'observation  $ij$ .

L'écriture matricielle du modèle

$$Y = X\theta + Z\tau + \epsilon \quad (2.2)$$

$Y$ : vecteur des réponses de taille  $N \times 1$ .

$X$ : est une matrice de taille  $N \times d$  écrit comme suit :

$$X = \begin{bmatrix} f^T(X|_{t_1}) \\ f^T(X|_{t_2}) \\ \vdots \\ f^T(X|_{t_N}) \end{bmatrix}, X \text{ est appelé la matrice d'expérience.}$$

$\epsilon$ : un vecteur contient les erreur aléatoires.

Autre écriture matricielle de modèle

$$Y = u + Z_0\tau_0 + Z_1\tau_1 \dots \dots + Z_p\tau_p \quad (2.3.a)$$

$u$  est la moyenne générale,  $Z_j$  est une matrice indicatrice associée à la  $j^{\text{ème}}$  composante de la variance,  $\tau_j$  est un vecteur d'effets aléatoires suivant la distribution normale associé à la  $j^{\text{ème}}$



composante de la variance telle que  $c_j \sim N(0, \sigma_j^2)$ . La variance de la matrice des observations peut s'écrire comme :

$$\text{Var}(Y) = V = \sum_{j=0}^p \sigma_j^2 Z_j Z_j^t \quad (2.2.b)$$

## 2.2. Maximum de vraisemblance :

La méthode d'estimation de vraisemblance (MLE) consiste à maximiser la fonction de vraisemblance du modèle compte tenu des données observées  $\mathbf{y}$ , et en fonction des paramètres  $(\theta, \sigma^2)$ . La maximisation se fait, alors, en calculant le différentiel de la fonction de vraisemblance, ou encore d'une manière plus pratique, le log-vraisemblance par rapport aux paramètres clés. Les dérivés résultantes sont fixées à zéro, et ensuite résolues soit pour  $\theta$  ou bien pour  $\sigma_i^2$ ,  $i=0 \dots \dots r$ . Cette méthode est choisie en raison de ses bonnes (asymptotiques) propriétés statistiques, une conséquence utile qui en découle est que, asymptotiquement, il est toujours possible d'avoir la matrice d'informations des estimateurs. En plus, la formulation est très générale, donc des types différents de modèles peuvent être évalués sans recourir à des ajustements importants dans la méthodologie. [3].

Considérons le modèle décrit par (1.8.b) et (1.8.a) mais prévoyons une matrice générale de variance-covariance  $V$ . Admettons que la variable-réponse est normalement distribuée, le modèle peut être énoncé comme  $\mathbf{y} \sim N_N(X\theta, V)$  qui donne la fonction de vraisemblance suivante:

$$L = L(\theta, v | \mathbf{y}) = \frac{\exp(-\frac{1}{2}(\mathbf{y} - X\theta)^T \mathbf{v}^{-1} (\mathbf{y} - X\theta))}{(2\pi)^{\frac{1}{2}N} |v|^{\frac{1}{2}}} \quad (2.3)$$

Et la fonction LOG-vraisemblance est :

$$l = \log(L) = \frac{1}{2} N \log(2\pi) - \frac{1}{2} \log(|v|) - \frac{1}{2} (\mathbf{y} - X\theta)^T \mathbf{v}^{-1} (\mathbf{y} - X\theta). \quad (2.4)$$

Les dérivées peuvent être obtenues en utilisant les résultats de l'algèbre matricielle. La première dérivée par rapport à  $\theta$  est :

$$\begin{aligned}\frac{\partial l}{\partial \theta} &= \frac{\partial}{\partial \theta} \left| -\frac{1}{2} (-y^T V^{-1} X \theta - \theta^T X^T V^{-1} y + \theta^T X^T V^{-1} X \theta) \right| \\ &= X^T V^{-1} y - X^T V^{-1} X \theta.\end{aligned}\quad (2.5)$$

Et celle par rapport à  $\sigma^2$  est:

$$\begin{aligned}\frac{\partial l}{\partial \sigma_i^2} &= \frac{\partial}{\partial \sigma_i^2} \left| -\frac{1}{2} \log(|V|) - \frac{1}{2} (y - X\theta)^T V^{-1} (y - X\theta) \right| \\ &= -\frac{1}{2} \left( \text{tr} \left( V^{-1} \frac{\partial V}{\partial \sigma_i^2} \right) - (y - X\theta)^T V^{-1} \frac{\partial V}{\partial \sigma_i^2} V^{-1} (y - X\theta) \right).\end{aligned}\quad (2.6)$$

Les équations de maximum de vraisemblance (EMV) sont obtenues lorsque (2.6) et (2.7) sont égales à zéro. Quand on les résout pour les paramètres, les valeurs ainsi obtenues maximisent la fonction log-vraisemblance, mais à proprement parler, elles ne sont pas toujours les estimations de vraisemblances maximales. Afin que cela puisse se produire, la dérivée seconde doit être vérifiée et la solution ne doit pas se trouver à la frontière de l'espace des paramètres. Cependant, avec les hypothèses déjà faites, ces solutions satisferont les EMVs. Les EMVs des paramètres fixes du modèle (2.2) peuvent s'écrire explicitement comme :

$$\hat{\theta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y. \quad (2.7)$$

Qui démontrent, en comparant les équations (1.7) et (2.8), que sous les hypothèses de normalité et de v.i.i la EMV est équivalente à l'estimateur GLS pour  $\theta$ . Pour le vecteur  $\sigma^2$  des composantes de la variance, l'équation

$$\text{tr} \left( V^{-1} \frac{\partial V}{\partial \sigma_i^2} \right) = (y - X\theta)^T V^{-1} \frac{\partial V}{\partial \sigma_i^2} V^{-1} (y - X\theta) \quad (2.8)$$

est non linéaire par rapport aux éléments  $\sigma_i^2$ , comme ils apparaissent dans les deux membres de l'équation à travers  $V^{-1}$  et éventuellement dans  $\frac{\partial V}{\partial \sigma_i^2}$ , par conséquent elle doit être résolue numériquement. A cet effet, en général, il n'y a pas d'expressions explicites pour les estimations de  $\hat{\sigma}^2$ . Naturellement, la variance-covariance  $\hat{V}$  estimée s'obtient quand  $\hat{\sigma}^2$  est employée dans  $V$ . Des procédures itératives fréquemment utilisées pour l'estimation non-linéaire peuvent être employées dans ce contexte en utilisant les équations (2.8) et (2.9).

### 2.3. La matrice d'information de MLE (ou bien EVM) :

La matrice d'information pour un modèle fixe est constituée aussi de deux parties principales, une pour  $\theta$  et l'autre pour  $\sigma^2$ . Pour la méthode EVM, ce que l'on obtient est la matrice d'information de Fisher définie comme:

$$M \begin{vmatrix} \theta \\ \sigma^2 \end{vmatrix} = -E \begin{vmatrix} \frac{\partial^2 l}{\partial \theta \partial \theta^T} & \frac{\partial^2 l}{\partial \theta \partial \sigma^2} \\ \frac{\partial^2 l}{\partial \sigma^2 \partial \theta} & \frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^2} \end{vmatrix} \quad (2.9)$$

Celle-ci est l'espérance négative de la seconde dérivée de la fonction log-vraisemblance par rapport aux deux vecteurs de paramètres. Les premières dérivées de  $l$  par rapport à  $\theta$  et  $\sigma^2$  ont déjà été calculées dans les équations (2.6) et (2.7), respectivement. Afin d'obtenir les dérivées secondes, on pourra de nouveau utiliser la formule. En commençant par l'équation (2.6), la dérivée seconde de log-vraisemblance par rapport à des paramètres fixes peut être calculée facilement.

$$\begin{aligned} \frac{\partial^2 l}{\partial \theta \partial \theta^T} &= \frac{\partial l}{\partial \theta} [X^T V^{-1} y - X^T V^{-1} X \theta] \\ &= -X^T V^{-1} X. \end{aligned} \quad (2.10)$$

Les deux dérivées partielles croisées (les éléments extra-diagonaux) dans la partie droite de (2.10) sont les mêmes par symétrie, autrement dit:

$$\frac{\partial^2 l}{\partial \sigma_i^2 \partial \theta^T} = \frac{\partial^2 l}{\partial \theta \partial \sigma_i^2}$$

D'après le second membre de l'expression ci-dessus et en utilisant la dérivée seconde correspondante est:

$$\begin{aligned} \frac{\partial^2 l}{\partial \theta \partial \sigma_i^2} &= \frac{\partial^2 l}{\partial \sigma_i^2} [X^T V^{-1} y - X^T V^{-1} X \theta] = \frac{\partial^2 l}{\partial \sigma_i^2} [X^T V^{-1} (y - X \theta)] \\ &= -|X^T V^{-1} \frac{\partial v}{\partial \sigma_i^2} V^{-1} (y - X \theta)|. \end{aligned} \quad (2.11)$$

En utilisant (2.7), la dérivée seconde du log-vraisemblance par rapport aux composantes de la variance est:

$$\begin{aligned}
\frac{\partial^2 l}{\partial \sigma_i^2 \partial \sigma_j^2} &= \frac{\partial^2 l}{\partial \sigma_j^2} \left| -\frac{1}{2} \left( \text{tr} \left( \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \right) - (\mathbf{y} - \mathbf{X}\theta)^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\theta) \right) \right| \\
&= -\frac{1}{2} \text{tr} \left( -\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \right) + \frac{1}{2} \left[ -2(\mathbf{y} - \mathbf{X}\theta)^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\theta) \right] \\
&= -\frac{1}{2} \text{tr} \left( -\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \right) - |(\mathbf{y} - \mathbf{X}\theta)^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\theta)| \quad .(2.12)
\end{aligned}$$

L'évaluation de l'espérance de (2.11) et le multiplier par (-1) donne:

$$-E(-\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}) = \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}.$$

Pour la dérivée croisée de (2.11), le résultat est le suivant:

$$\begin{aligned}
-E \left| \frac{\partial^2 l}{\partial \theta \partial \sigma_i^2} \right| &= -E \left( -|\mathbf{X}^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\theta)| \right) \\
&= \mathbf{X}^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{V}^{-1} E(\mathbf{y} - \mathbf{X}\theta) = 0 \quad .(2.13)
\end{aligned}$$

Qui découle de  $E(\mathbf{y}) = \mathbf{X}\theta$  comme ça donne  $E(\mathbf{y} - \mathbf{X}\theta) = 0$ .

$$-E \left| \frac{\partial^2 l}{\partial \sigma_i^2 \partial \theta^T} \right| = 0.$$

Le résultat correspondant à la dérivée seconde par rapport aux composantes de la variance seulement, en utilisant (2.13) est:

$$\begin{aligned}
-E \left| \frac{\partial^2 l}{\partial \sigma_i^2 \partial \sigma_j^2} \right| &= -E \left( -\frac{1}{2} \text{tr} \left( -\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \right) \right. \\
&\quad \left. - |(\mathbf{y} - \mathbf{X}\theta)^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\theta)| \right)
\end{aligned}$$

$$\begin{aligned}
&= -\frac{1}{2} \text{tr} \left( -\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \right) + \text{tr} \left( \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \mathbf{V}^{-1} \mathbf{V} \right) \\
&= \frac{1}{2} \text{tr} \left( \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \right). \tag{2.14}
\end{aligned}$$

Ceci complète les éléments du deuxième membre de (2.16). D'abord, il y a lieu de noter que la matrice d'information est diagonale par blocs d'après (2.14). Cela implique que les entrées dans la diagonale principale sont indépendantes. Il est ainsi possible d'écrire les matrices d'information pour la partie fixe séparément de celles de la partie aléatoire. L'élément (1;1) correspond à la matrice d'information des paramètres fixes  $\theta$  qui s'écrit comme suit:

$$M(\theta) = \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}. \tag{2.15}$$

De même, l'élément (2;2) dénote la matrice d'information des paramètres de la variance  $\sigma^2$  dont l'expression correspondante est:

$$M(\sigma^2) = \frac{1}{2} \text{tr} \left( \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \right). \tag{2.16}$$

La matrice d'information EVM complète du modèle mixte général (2.2) est:

$$M \begin{vmatrix} \theta \\ \sigma^2 \end{vmatrix} = \begin{matrix} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} & 0 \\ 0 & \frac{1}{2} \text{tr} \left( \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_i^2} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \sigma_j^2} \right)_{i=j=0, \dots, r} \end{matrix}. \tag{2.17}$$

Mais il est également vrai que :

$$\text{var} \begin{vmatrix} \hat{\theta} \\ \hat{\sigma}^2 \end{vmatrix} = \mathbf{M}^{-1} \begin{vmatrix} \theta \\ \sigma^2 \end{vmatrix}. \tag{2.18}$$

Finalement, il y a lieu de noter que l'expression dans (2.16) est la même que la matrice d'information obtenue par la méthode des moindres carrés généralisés dans la section précédente est donnée par (2.16). Ceci est attendu du fait de l'équivalence de EVM et MCG mentionnée dans ce chapitre (Maximum de vraisemblance Paragraphe 1). Une autre forme de  $M(\sigma^2)$  pourrait être obtenue par le biais de la méthode ANOVA. Cependant, elle ne sera pas équivalente à (2.17), il n'y a pas d'expression générale pour elle, et elle pourrait s'obtenir comme l'inverse

de la matrice variance-covariance des estimateurs. Cette approche n'est explorée ni dans la littérature des plans optimaux ni dans ce document.

## 2.4. Les Structures des plans

### 2.4.1. Facteurs

Il est bien important d'avoir des définitions claires sur le type de facteurs à utiliser. Ces définitions sont, en premier lieu, complètement en rapport avec le sujet. Par la suite, la classification des facteurs se fait, aussi, en fonction de considérations statistiques. Le nombre d'étapes est lié aux conditions de l'expérience, et au modèle considéré également. Des précautions similaires sont nécessaires pour décider du rôle ainsi que l'importance relative de chaque facteur. Une fois les facteurs définis, ce qui suit est supposé connu :

- le nombre total de facteurs impliqués,
- l'ensemble de facteurs fixes et aléatoires,
- les facteurs fixes quantitatifs et qualitatifs,
- l'étendue des niveaux des facteurs fixes et le nombre de niveaux des facteurs aléatoires,
- toutes les restrictions sur les facteurs, y compris les facteurs difficiles de changer.

Il est important de tenir compte d'autres considérations avant d'injecter ces informations dans un modèle statistique. La définition de facteurs emboîtés et de facteurs croisés est particulièrement importante. Néanmoins, ces caractéristiques peuvent ne pas être connues au moment de la sélection des facteurs. Leur reconnaissance peut être complétée par la définition d'autres éléments clés de la structure complète du plan, exemple le choix des parcelles (Whole-Plots). Cette dernière, de type et de taille particuliers, sera expliquée ultérieurement. Les facteurs croisés et les facteurs emboîtés seront examinés plus tard.

### 2.4.2. Unité expérimentale

Il est généralement admis que l'identification adéquate des unités dans une expérience est cruciale, voir e.g. Hinkelmann et Kempthorne (2005) et Casella (2008). Faute de procéder à une identification correcte peut mener à un plan qui ne répond pas aux objectifs de la phase expérimentale et qui, par conséquent, produit des conclusions erronées. En réalité,

cet effort peut s'avérer le plus obscure dans la phase de la conception. L'exemple le plus simple est que les unités expérimentales peuvent être différentes si le même système est étudié dans une expérience séparée et avec des objectifs différents. La littérature des expériences est riche en matière de définitions données les unités. Des présentations ainsi que des discussions utiles sont données par Bailey (2008) et Milliken et Johnson (2009). La détermination de la taille de l'unité dépend du contexte, donc elle est subjective. Cela pourrait être la plus petite, petite, grande, la plus grande, etc. Pour les expériences à plusieurs étapes, il est nécessaire de déterminer le nombre de différentes tailles d'unités impliquées, et quelles unités appartiennent à quelle étape de la structure du plan. Il existe une interaction entre ces concepts. Parfois il est possible de déterminer la hiérarchie en premier lieu, et ensuite la taille des unités sera définie en conséquence. Dans d'autres occasions, la taille de l'unité peut afficher sa position dans la hiérarchie. On peut distinguer deux types différents d'unités

#### **2.4.3. Unités d'observation**

Celles-ci sont les unités de plus petite taille affectées au niveau le plus bas dans la hiérarchie du plan. C'est à ce niveau que les observations sont obtenues, d'où le nom les unités d'observation.

#### **2.4.4. Unités structurelles**

Ce sont des unités plus grandes, intermédiaires, voire conceptuelles (non physiques) seront associées à tous les autres niveaux de la hiérarchie. Ces unités seront indépendamment appelées les unités structurelles.

#### **2.4.5. Traitements observationnels**

Ceux-ci représentent l'ensemble des traitements  $t_i$ ,  $i=1, \dots, N$ , appliqués aux unités d'observation.

#### **2.4.6. Traitements structurels**

Ces derniers constituent l'ensemble des traitements  $t_j$ ,  $j=1, \dots, s$ , appliqués aux unités structurelles; ils sont obtenus par l'arrangement des facteurs utilisé dans le niveau correspondant. Quand il s'avère nécessaire de faire la distinction entre les deux types de traitements,  $t_i$  désigne l'ensemble des traitements observationnels,  $t_j$  désigne celui des traitements structurels. En général, il est admis que la plus grande unité se trouve au niveau qui

---

est au sommet de la hiérarchie, la deuxième plus grande au niveau juste en bas, ainsi de suite jusqu'à ce que les unités observationnelles sont trouvées au niveau le plus bas.

L'emploi de différentes tailles d'unités est synonyme d'un arrangement emboîté. Cela signifie aussi qu'il faut spécifier une hiérarchie, mais pas toujours. C'est dire que l'emboîtement en soi ne nécessite pas un modèle pour une variabilité multi-niveaux, mais une hiérarchie oui. En mettant ces idées ensemble, il devient clair que la définition précise d'unités expérimentales est essentielle vue qu'elle peut déterminer la structure globale du plan. Par contre, si la structure d'un plan est définie en premier lieu, les unités expérimentales doivent être affectées au niveau hiérarchique qui convient





## CHAPITRE 3

### PLANS A PARCELLES DIVISEES POUR UN SEUL FACTEUR DIFFICILE A CHANGER ET UN SEUL FACTEUR FACILE A CHANGER (1WP ET 1 SP)

#### 1.1. Modèle et structure de la variance

L'équation du modèle pour la conception de la parcelle divisée avec un facteur WP (facteur A) et un facteur SP (facteur B) peuvent être écrits ainsi [4]:

$$y_{ijk} = u + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_{ik} + e_{ijk} \quad i = 1 \dots a, j = 1 \dots b, k = 1 \dots r_i,$$

Où

$r_i$  est le nombre de répétition du  $i^{\text{ème}}$  niveau du facteur A, mais dans notre cas tous les  $r_i$  sont égaux donc le nombre total des parcelles est égal à  $ar$  (nombre des niveaux de A \* nombre de répétition).

$y_{ijk}$  est la réponse de la  $k^{\text{ème}}$  répétition du  $i^{\text{ème}}$  niveau du facteur A et le  $j^{\text{ème}}$  niveau du facteur B.

$u$  est la moyenne générale,  $\alpha_i$  est l'effet du  $i^{\text{ème}}$  niveau du facteur A,  $\beta_j$  est l'effet du  $j^{\text{ème}}$  niveau du facteur B,  $(\alpha\beta)_{ij}$  est l'effet d'interaction du  $i^{\text{ème}}$  niveau du facteur A et du  $j^{\text{ème}}$  niveau du facteur B,  $\gamma_{ik}$  est le terme d'erreur de la  $k^{\text{ème}}$  répétition du  $i^{\text{ème}}$  niveau du facteur A,  $e_{ijk}$  est le terme d'erreur correspondant à l'individu  $y_{ijk}$ .

les variables aléatoires  $\alpha_i$ ,  $\beta_j$ ,  $(\alpha\beta)_{ij}$ ,  $\gamma_{ik}$ ,  $e_{ijk}$  suivent la loi normale avec une moyenne nulle et de variance  $\sigma_\alpha^2$ ,  $\sigma_\beta^2$ ,  $\sigma_{\alpha\beta}^2$ ,  $\sigma_\gamma^2$ ,  $\sigma_e^2$  respectivement.

Le modèle peut être écrit sous forme matricielle comme suit:

$$Y = u1 + Z_1\gamma_1 + Z_2\gamma_2 + Z_3\gamma_3 + Z_4\gamma_4 + Z_5\gamma_5 \quad (3.1)$$

Y est un vecteur de  $abk$  observations (nombre de niveau du facteur A \* nombre de niveau du facteur B, nombre de répétition),  $u$  est la moyenne générale,  $Z_i$  est une matrice d'indicateurs associé à la  $i^{\text{ème}}$  composante de variance,  $\gamma_i$  est un vecteur de distribution normale des effets aléatoires associé à la  $i^{\text{ème}}$  composante de variance telle que  $\gamma_i \approx (0, \sigma_i^2 I)$ .

La matrice de variance des observations peut être écrite comme :

$$\text{Var}(Y) = V = \sum_{i=1}^5 \sigma_i^2 Z_i Z_i^t \quad (3.2)$$

Les matrices  $Z$  sont définies comme suit :

$$Z_0 = I_a \otimes I_b \otimes I_r \otimes I_n Z_0 Z_0' = I_a \otimes I_b \otimes I_r \otimes I_n \quad (3.3.a)$$

$$Z_1 = I_a \otimes 1_b \otimes 1_r \otimes I_n Z_1 Z_1' = I_a \otimes J_b \otimes J_r \otimes J_n \quad (3.3.b)$$

$$Z_2 = 1_a \otimes I_b \otimes 1_r \otimes 1_n Z_2 Z_2' = J_a \otimes I_b \otimes J_r \otimes J_n \quad (3.3.c)$$

$$Z_3 = I_a \otimes I_b \otimes 1_r \otimes 1_n Z_3 Z_3' = I_a \otimes I_b \otimes J_r \otimes J_n \quad (3.3.d)$$

$$Z_4 = I_a \otimes 1_b \otimes I_r \otimes 1_n Z_4 Z_4' = I_a \otimes J_b \otimes I_r \otimes J_n \quad (3.3.e)$$

**Remarque :**

- 1-  $(\otimes)$  est le produit de Kroneker
- 2-  $1_a$  est un vecteur de (a) colonnes et ses composants sont à 1
- 3-  $J$  est une matrice dont toutes ses composantes sont à 1

Donc d'après la formule (...) on aura :

$$V = (I_a \otimes J_b \otimes J_r \otimes J_n) \sigma_\alpha^2 + (J_a \otimes I_b \otimes J_r \otimes J_n) \sigma_\beta^2 + (I_a \otimes I_b \otimes J_r \otimes J_n) \sigma_{\alpha\beta}^2 + (I_a \otimes J_b \otimes I_r \otimes J_n) \sigma_\gamma^2 + (I_a \otimes I_b \otimes I_r \otimes I_n) \sigma_e^2 \quad (3.4)$$

$$L = L(\beta, v|y) = \frac{\exp(-\frac{1}{2}(y - X\beta)^T v^{-1} (y - X\beta))}{(2\pi)^{\frac{1}{2}N} |v|^{\frac{1}{2}}} \quad (3.5)$$

Et la fonction LOG-vraisemblance est :

$$l = \log(L) = -\frac{1}{2}N \log(2\pi) - \frac{1}{2} \log(|V|) - \frac{1}{2} (y - X\beta)^T V^{-1} (y - X\beta). \quad (3.6)$$

Cette matrice est donnée comme suit :

$$\begin{bmatrix} \frac{\partial^2 l}{\partial \beta \partial \beta^t} & \frac{\partial^2 l}{\partial \beta \partial \sigma^{2t}} \\ \frac{\partial^2 l}{\partial \beta \partial \sigma^{2t}} & \frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^{2t}} \end{bmatrix}$$

Ce qui donne ensuite :

$$\ln \begin{vmatrix} \beta \\ \sigma^2 \end{vmatrix} = \begin{vmatrix} X^T v^{-1} X & 0 \\ 0 & \frac{1}{2} \text{tr}(v^{-1} Z_i Z_i^t v^{-1} Z_j Z_j^t) \end{vmatrix} \quad i, j = 1 \dots \dots \dots 5 \quad (3.7)$$

Le modèle de ce travail ne concerne que les composantes de la variance et donc la matrice d'information devient[7] :

$$\ln[\sigma^2] = \frac{1}{2} \{ \text{tr}(v^{-1} Z_i Z_i^t v^{-1} Z_j Z_j^t) \} \quad i, j = 1 \dots \dots \dots 5 \quad (3.8)$$

Cela se réduit à

$$\frac{1}{2} \{ \text{tr}(v^{-1} Z_i Z_i^t v^{-1} Z_j Z_j^t) \} = \frac{1}{2} \{ \text{sesq}(Z_i^t v^{-1} Z_j) \} \quad i, j = 1 \dots \dots 5 \quad (3.4)$$

Où sesq (.) est la somme des éléments carrés de la matrice entre parenthèses. La procédure commence par le calcul de  $v^{-1}$ , [3]



$$\varphi_4 = \sigma_e^2 + nr\sigma_{\alpha\beta}^2 + bn\sigma_\gamma^2 + bnr\sigma_\alpha^2$$

$$\varphi_5 = \sigma_e^2 + nr\sigma_{\alpha\beta}^2 + bn\sigma_\gamma^2 + bnr\sigma_\alpha^2 + anr\sigma_\beta^2$$

L'inverse de  $v$  est également obtenu en utilisant les résultats de Henderson et Searle (1979),

$$v^{-1} = \varphi_0^{-1}(I_a \otimes C_b \otimes C_r \otimes C_n) + \varphi_1^{-1}(C_a \otimes C_b \otimes \bar{J}_r \otimes \bar{J}_n) + \varphi_2^{-1}(I_a \otimes \bar{J}_b \otimes C_r \otimes \bar{J}_n) + \varphi_3^{-1}(C_a \otimes \bar{J}_b \otimes \bar{J}_r \otimes \bar{J}_n) + \varphi_4^{-1}(\bar{J}_a \otimes C_b \otimes \bar{J}_r \otimes \bar{J}_n) + \varphi_5^{-1}(\bar{J}_a \otimes \bar{J}_b \otimes \bar{J}_r \otimes \bar{J}_n) \quad (3.11)$$

Une matrice d'informations peut ensuite être obtenue de Searle et al (2006) p. 247.

$$M \begin{bmatrix} \sigma_e^2 \\ \sigma_\gamma^2 \\ \sigma_{\alpha\beta}^2 \\ \sigma_\beta^2 \\ \sigma_\alpha^2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} t_{ee} & t_{\gamma\gamma}/bn & t_{\alpha\beta}/rn & t_{\beta\beta}/arn & t_{\alpha\alpha}/brn \\ t_{\gamma\gamma}/bn & t_{\gamma\gamma} & t_{\alpha\alpha}/r & abrn^2/\varphi_5 & t_{\alpha\alpha}/r \\ t_{\alpha\beta}/rn & t_{\alpha\alpha}/r & t_{\alpha\beta} & t_{\beta\beta}/a & t_{\alpha\alpha}/b \\ t_{\beta\beta}/arn & abrn^2/\varphi_5 & t_{\beta\beta}/a & t_{\beta\beta} & ab(rn)^2/\varphi_5 \\ t_{\alpha\alpha}/brn & t_{\alpha\alpha}/r & t_{\alpha\alpha}/b & ab(rn)^2/\varphi_5 & t_{\alpha\alpha} \end{bmatrix} \quad (3.12)$$

$$t_{ee} = \frac{v_e}{\varphi_0^2} + \frac{v_{\gamma\gamma}}{\varphi_1^2} + \frac{t_{\alpha\beta}}{(rn)^2}, \quad t_{\alpha\alpha} = (brn)^2 \left( \frac{v_\alpha}{\varphi_4^2} + \frac{1}{\varphi_5^2} \right)$$

$$t_{\beta\beta} = (arn)^2 \left( \frac{v_\beta}{\varphi_3^2} + \frac{1}{\varphi_5^2} \right), \quad t_{\alpha\beta} = (rn)^2 \left( \frac{v_{\alpha\beta}}{\varphi_2^2} + \frac{v_\beta}{\varphi_3^2} + \frac{v_\alpha}{\varphi_4^2} + \frac{1}{\varphi_5^2} \right)$$

$$t_{\gamma\gamma} = (bn)^2 \left( \frac{v_\gamma}{\varphi_1^2} + \frac{v_\alpha}{\varphi_4^2} + \frac{1}{\varphi_5^2} \right)$$

Tel que :

$$v_e = a(b-1)(r-1), \quad v_\gamma = a(r-1), \quad v_{\alpha\beta} = (a-1)(b-1)$$

$$v_\alpha = a-1$$

## 1.2. Optimalité de conception et génération

Aucune expression analytique sous forme fermée n'est disponible pour la matrice de variance-covariance dans ce modèle à effets aléatoires linéaires.

Nous examinons l'optimalité en utilisant la matrice variance-covariance asymptotique.

Une conception parmi un groupe de conceptions avec le même nombre et la même taille de parcelles (whole plots) est considérée comme optimale si elle minimise un critère d'optimalité lié à la matrice variance-covariance des estimations de paramètres.

De manière équivalente, nous recherchons la conception qui maximise un critère d'optimalité lié à la matrice d'informations de cinq composantes de variance. La conception optimale dans un modèle à effets aléatoires linéaires dépend de la taille relative des valeurs vraies des composantes de variance est nous ne pourrons pas étudier l'optimalité à moins qu'une hypothèse soit faite sur ses valeurs. Puisque l'optimalité pour des tels modèles est similaire à celle des modèles non linéaires.

Nous empruntons une idée d'optimisation sur la théorie des modèles non linéaires et nous utilisons l'optimalité locale. Les cinq composantes de la variance ont été classées en deux ensembles, le premier ensemble étant comprend des composantes principales des effets et de la variance d'interaction (MIVC) qui consistent de  $\sigma_\alpha^2$ ,  $\sigma_\beta^2$  et  $\sigma_{\alpha\beta}^2$ . Le second ensemble comprend les composantes de variance d'erreur de parcelle et de sous-parcelles (WSEVC)  $\sigma_\gamma^2$  et  $\sigma_e^2$ .

Le travail assignera initialement la valeur proportionnelle de la variance composantes aux deux ensembles de telle sorte que la somme est égale à un, et ensuite, on distribue la valeur proportionnelle à chaque ensemble sur la base initiale allocation.

**Par exemple :**

$$\text{MIVC} = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 = 0.5 \text{WSEVC} = \sigma_\gamma^2 + \sigma_e^2 = 0.5$$

la sous attribution de MIVC et WSEVC est :

$$\sigma_{\alpha}^2 = 0.05, \sigma_{\beta}^2 = 0.40, \sigma_{\alpha\beta}^2 = 0.05 \text{ et } \sigma_{\gamma}^2 = 0.42, \sigma_e^2 = 0.08.$$



## CHAPITRE 4

### QUELQUES EXEMPLES SUR LES CAS DE 1WP 1SP

Dans ce chapitre, nous essayons d'examiner quelques cas particuliers du chapitre précédent. Comme étant indiqué dans le chapitre 3, l'optimalité de plan d'expériences dépend de la valeur réelle des composantes de la variance. La valeur initiale des composantes de la variance sera affectée à chacune des cinq composantes de la variance. Nous avons utilisé cette approche pour rendre la comparaison facile.

Un code MATLAB a été écrit dans le contexte de la matrice d'information de la section (2.1) de telle sorte que les plans énumérés pour un nombre particulier et une taille spécifique de la parcelle peuvent être comparés en fonction de n'importe quelle configuration des vraies valeurs des composantes de la variance. [5].

#### 4.1. Génération de plans

Le travail consiste à assigner  $ar=R$  parcelles entières de taille égale, d'une façon au hasard aux niveaux du facteur difficile à changer. Chaque niveau du facteur de sous-parcelle (facteur facile à changer) est appliqué une fois dans chaque parcelle et une observation est mesurée dans chaque sous-parcelle. La structure de conception résultante est équilibrée. Le nombre de plans possibles est égal au nombre total de façons de partitionner  $ar = \sum_1^a r_i$  avec  $r_1 = r_2 = r_3 = \dots = r_a$  et  $2 \leq a \leq ar$ . A titre exemple, considérons six parcelles entières de taille deux, la taille de la parcelle est égal au nombre de niveaux disponibles pour le facteur B de la sous-parcelle (en ce cas  $b = 2$ ). Le nombre de conceptions possibles est égal au 6, le nombre de partitionnement possible :  $a=2, r=3$  [2, 3] et  $a=3, r=2$  [3, 2]. Nous listons les possibles conceptions pour divers nombre de parcelles entre 6 et 30 qui satisfont la condition d'équilibre.



## Liste des plans possibles

Nombre de parcelles	Plans possibles
6	[a=3, r=2][a=2, r=3]
8	[a=4, r=2][a=2, r=4]
10	[a=5, r=2][a=2, r=5]
12	[a=6, r=2][a=2, r=6] [a=3, r=4][a=4, r=3]
14	[a=7, r=2][a=2, r=7]
15	[a=3, r=5][a=5, r=3]
16	[a=2, r=8][a=8, r=2][a=4, r=4]
18	[a=2, r=9][a=9, r=2][a=3, r=6][a=6, r=3]
20	[a=2, r=10][a=10, r=2][a=4, r=5][a=5, r=4]
24	[a=2, r=12] [a=12, r=2] [a=4, r=6] [a=6, r=4] [a=8, r=3] [a=3, r=8]
28	[a=14, r=2] [a=2, r=14] [a=7, r=4] [a=4, r=7]
30	[a=2, r=15] [a=15, r=2] [a=5, r=6] [a=6, r=5] [a=3, r=10] [a=10, r=3]

## 4.2. Algorithme

- (1) Déterminer toutes les plans possibles pour un nombre de parcelles et une taille fixes.
- (2) Spécifiez les informations disponibles sur les composants individuels de la variance. C'est-à-dire la valeur disponible pour MIVC et WSEVC. Tels que:

$$\text{MIVC} + \text{WSEVC} = 1$$

- (3) Redistribuer la valeur ci-dessus aux composantes de la variance dans chaque ensemble.
- (4) Calculer le critère d'optimalité (D-optimal) pour toute les plans possibles en utilisant le code de programme MATLAB et identifier le plan D-optimale.
- (5) En n'apportant aucune modification à MIVC et basé sur la conception D-optimale obtenue en 4, on peut obtenir les régions où d'autres plans optimales de l'espace de conception. Quand  $a \leq r$  à partir de plan optimale identifié en (4) augmenter la valeur actuelle de  $\sigma_\gamma^2$  de 0,01 et en diminuer la valeur de  $\sigma_e^2$  par 0,01 dans l'ordre jusqu'à arriver à un autre plan optimale dans l'espace des conceptions. Quand  $a \geq r$  à partir de plan optimale obtenu en (4) en diminuant la valeur de  $\sigma_\gamma^2$  de 0,01 et en augmentant la valeur de  $\sigma_e^2$  par 0,01 en séquence jusqu'à ce qu'un autre plan dans l'espace de conception soit optimale.

## 4.3. Exemple

Considérons la situation où nous avons 6 parcelles de taille 2, en suivant les étapes de l'algorithme précédent, la liste des plans dans l'espace de conceptions est :  $\{a=2, b=2, r=3\}$  et  $\{a=3, b=2, r=3\}$ . En prenant  $\text{MIVC}=0.5$  et  $\text{WSEVC}=0.5$  par (2) et appliquer (3) les composants individuels de variance sont tels que :

$$= 0.05 \quad \sigma_\beta^2 = 0.4 \quad \sigma_{\alpha\beta}^2 = 0.05 \quad \sigma_\gamma^2 = 0.30 \quad \sigma_e^2 = 0.20$$

En utilisant le code MATLAB, le D-optimal est  $\{a = 2, b = 2, r = 3\}$ . Pour ce plan, nous augmentons la valeur de l'erreur de la parcelle de 0,30 et en réduisant la valeur de l'erreur de sous-parcelle de 0,20 en séquences de 0,01, il y a un changement dans le plan D-optimale à une certaine configuration. En général, pour un MIVC fixe, la région de valeur proportionnelle pour laquelle les deux conceptions sont optimales est donnée ci-dessous :

Tableau 4.1 : les régions d'optimalité

Conceptions	Parcelles entières	Sous-parcelles
$\{a=2, b=2, r=3\}$ .	$0.01 \leq \sigma_\gamma^2 \leq 0.42$	$0.08 \leq \sigma_e^2 \leq 0.49$
$\{a=3, b=2, r=2\}$ .	$0.43 \leq \sigma_\gamma^2 \leq 0.49$	$0.01 \leq \sigma_e^2 \leq 0.07$

Dans ce cas il est claire que la région d'optimalité de premier plan candidat est grande que le deuxième plan candidat donc le plan D-optimale est le premier.

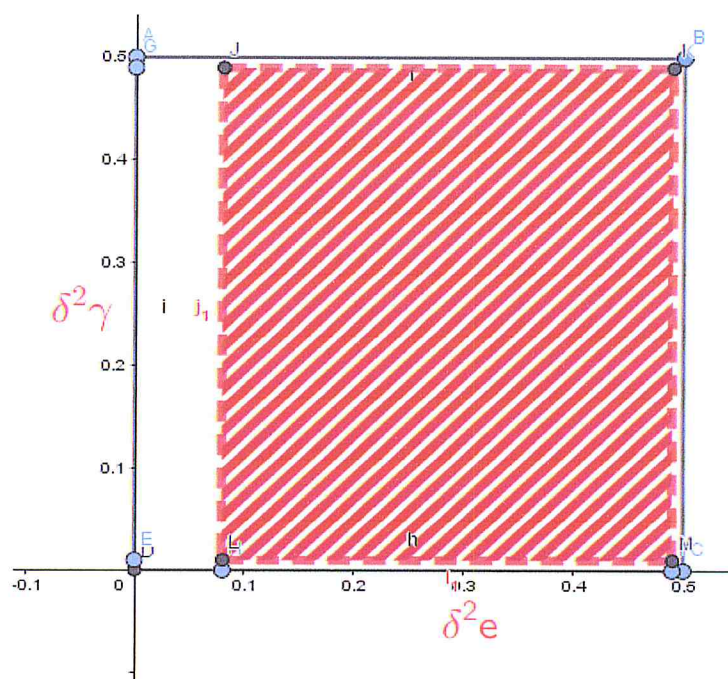


Figure 4.1 la région D-optimale de la première conception

#### 4.4. Quelques plans D- optimales:

Pour un nombre fixé de parcelles et une taille spécifique , nous présentons dans les tableaux ci dessus les régions d'optimalités pour le plan optimal-D.

Tableau 4.2 :cas de 6 parcelles entières de taille 2

Nombre De parcelles	Taille	Proportion initiale	Proportionnelle de composants de variance	D-Optimale conception	Région d'optimalité pour la conception D-optimale
6	2	MIVC=0.5 WSEVC=0.5	$\sigma_{\alpha}^2 = 0.05$ $\sigma_{\beta}^2 = 0.40$ $\sigma_{\alpha\beta}^2 = 0.05$ $\sigma_{\gamma}^2 = 0.30$ $\sigma_e^2 = 0.20$	{a=2,b=2, r=3}.	$0.08 \leq \sigma_e^2 \leq 0.49$ $0.01 \leq \sigma_{\gamma}^2 \leq 0.42$

Dans le cas de 6 parcelles de taille 2, le plan D-optimale est  $\{a=2,b=2, r=3\}$ . C'est-à-dire un facteur difficile à changer a deux niveau et un facteur facile à changer a deux niveau avec trois répétitions.

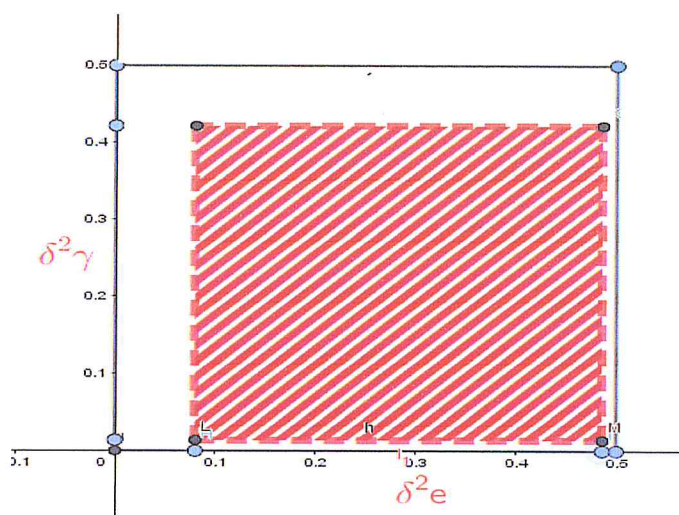


Figure 4.1a la région D-optimale de cas de 6 parcelles entières de taille 2

Tableau 4.3 : cas de 6 parcelles entières de taille 5

Nombre De la parcelle entière	Taille	Proportion initiale	Proportionnelle de composants de variance	D-Optimale conception	Région d'optimalité pour la conception D-optimale
6	5	MIVC=0.5 WSEVC=0.5	$\sigma_{\alpha}^2 = 0.05$ $\sigma_{\beta}^2 = 0.40$ $\sigma_{\alpha\beta}^2 = 0.05$ $\sigma_{\gamma}^2 = 0.30$ $\sigma_e^2 = 0.20$	{a=2,b=5, r=3}.	$0.02 \leq \sigma_e^2 \leq 0.49$ $0.01 \leq \sigma_{\gamma}^2 \leq 0.48$

Dans le cas de 6 parcelles entières de taille 5 le plan D-optimale est {a=2,b=5, r=3}. C'est-à-dire un facteur difficile à changer a deux niveau et un facteur facile à changer acinq niveau avec trois répétition

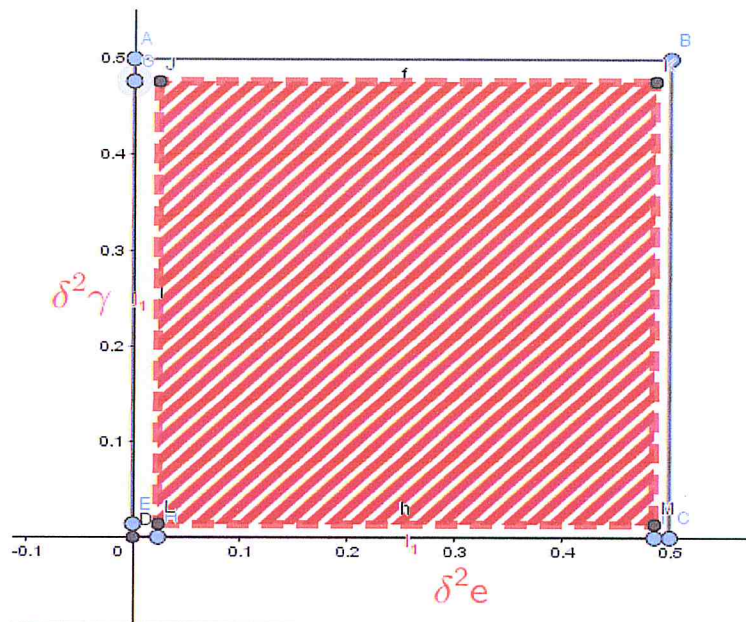


Figure 4.2 la région D-optimale de cas de 6 parcelles entières de taille 5

Tableau 4.4 : cas de 12 parcelles entières de taille 2

Nombre De la parcelle entière	Taille	Proportion initiale	Proportionnelle de composants de variance	D-Optimale conception	Région d'optimalité pour la conception D-optimale
12	2	MIVC=0.95 WSEVC=0.05	$\sigma_{\alpha}^2 = 0.05$ $\sigma_{\beta}^2 = 0.85$ $\sigma_{\alpha\beta}^2 = 0.05$ $\sigma_{\gamma}^2 = 0.01$ $\sigma_e^2 = 0.04$	{a=3,b=2, r=4}.	$0.03 \leq \sigma_e^2 \leq 0.04$ $0.01 \leq \sigma_{\gamma}^2 \leq 0.02$

Dans le cas de 12 parcelles entières de taille 5 le palm D-optimale est {a=3,b=2, r=4}. C'est-à dire un facteur difficile à changer a trois niveau et un facteur facile à changer a deux niveau avec quatre répétitions.

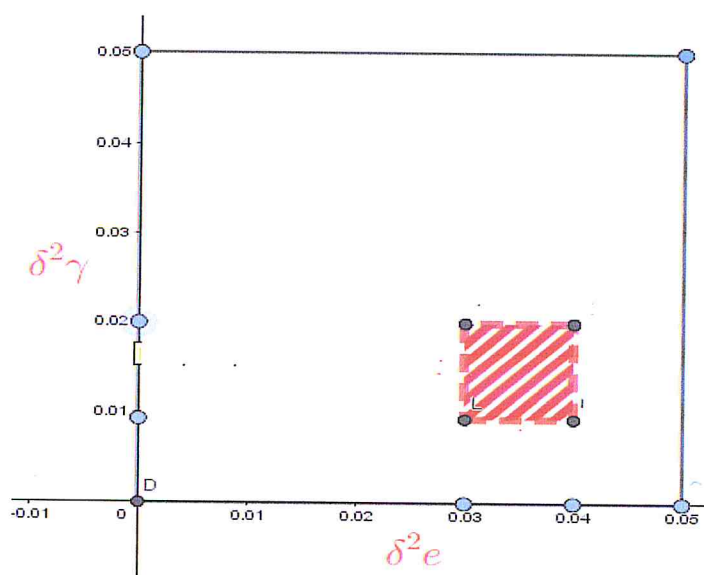


Figure 4.3 la région D-optimale de cas de 12 parcelles entières de taille 2

Après avoir comparé empiriquement des plans avec le même nombre et la même taille des toutes parcelles, ces deux déclarations de grande portée peuvent être faites

- (1) Pour des valeurs élevées de  $\sigma_{\alpha\beta}^2$  et  $\sigma_{\alpha}^2$ , les plans avec  $a \geq r$  sont optimales.
- (2) Le plan optimale est robuste lorsque on augmente la valeur de la taille de la parcelle. C.-à-d. si un plan est optimale pour une parcelle plus petite, le même plan est optimale pour des parcelles a des tailles plus élevée.

## CONCLUSION GÉNÉRALE

Les mathématiciens ont toujours été à l'avant garde de la recherche aussi bien fondamentale que pratique en développant plusieurs théories scientifiques. Ils les ont constamment améliorées et perfectionnées. Mais il s'est toujours écoulé un certain temps entre la découverte de ces théories et leur application pratique dans le monde industriel. La raison en est simple, car ces théories, lors de leur application pratique, nécessitent souvent un volume considérable de calculs. C'est évidemment l'outil informatique, venu à la rescousse qui a pu débloquent la situation en évitant au chercheur des calculs fastidieux et souvent impossible à effectuer. On peut citer par exemple l'Analyse des données dont les premiers développements théoriques furent établis pour la première fois en 1930 et leur vulgarisation et utilisation pratique dans le monde industriel ne fut effective que vers les années 1960. Il en est de même de l'application de la théorie des graphes et de beaucoup d'autres théories mathématiques.

La méthodologie de la recherche expérimentale, est une discipline mathématique faisant partie de la statistique inférentielle. L'une des principales difficultés d'utilisation des plans d'expériences était la réalisation des calculs. Mais l'arrivée de logiciel de plus en plus proches des besoins de l'utilisateur a permis d'effacer l'obstacle des calculs.

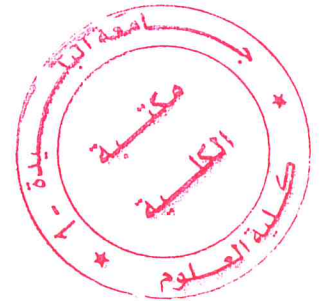
Le travail qui nous a été proposé avait deux objectifs :

- Nous nous sommes intéressés à l'étude des plans en parcelles divisées. Nous nous sommes intéressés plus particulièrement aux Plans à parcelles divisées pour un seul facteur facile à changer et un seul facteur difficile à changer.
- Nous nous sommes aussi intéressés à l'élaboration d'un code de programme écrit en Matlab pour déterminer la région D-optimale.

Comme perspective, il serait intéressant de développer l'étude de d'autres types de plans en parcelles divisées ou bien d'étudier le cas où le nombre de répétitions soit différents.



## Références



- [1] TRIBOULET. P. Notions de bases sur les plans d'expériences. Lycée Niepce – Chalon sur Saône, Paris, 2008.
- [2] J. GOUPY et L. CREIGHTON, "Introduction aux plans d'expériences", DUNOD, Paris, 2006.
- [3] L.S. Sergio Ivan, Optimal statistical design for variance components in multistage variability modele, These d'état, University of Manchester, 2014.
- [4] Site web : [www.statsoft.fr](http://www.statsoft.fr), Plans en parcelles divisées (split plot) dans un plan d'expériences .
- [5] O. NUGA, G.N. AMAHIA. et F. DEROGBA, Construction of Split plot Design for Estimating Variance Component, European Scientific Journal, vol. 10, No.18., 2014.
- [6] O.NUGA et G.N.AMAHIA et A.FAKOREDE A-Optimal Split Plot Design for Estimating Variance Components, American Journal of Applied Mathematics and Statistics, vol. 2, No. 4, 2014.
- [7] S. R. SEARLE, V. HAROLD ANDERSON, Dispersion matrices for Variance Components Models, Journal of American Statistical Association, vol. 74, No 366, 1979.