

Méthode des plans d'expériences

Chapitre II

Méthode des plans d'expériences

II.1 INTRODUCTION

Dans le langage scientifique, expérience signifie, fait attendu ou provoqué de manière à vérifier une hypothèse, une loi, un modèle et parvenir ainsi à une connaissance théorique de la façon dont se déroulent des phénomènes. Lors d'études expérimentales multiparamétriques, le nombre de variables peut être élevé. Les stratégies couramment employées pour mener à bien ces expérimentations sont souvent informelles, parfois quelque peu inutiles, et elles peuvent conduire à un nombre de résultats difficile à exploiter. Pour optimiser l'organisation des expériences et exploiter efficacement les résultats obtenus, le scientifique peut avoir intérêt à recourir à des méthodes telles que les Plans d'Expériences (PE). Le principe général des PE consiste à n'étudier que certains points du domaine expérimental sous investigation, tout en réussissant pourtant à appréhender le phénomène physique étudié sur l'ensemble du domaine considéré. En s'inspirant de la norme ISO 3534-3 [Exp05], un PE peut être défini comme une organisation raisonnée d'essais. Une difficulté importante de la méthodologie réside alors dans la manière de choisir les points d'études de façon optimale. Dans l'industrie, la connaissance de la méthode des PE apparaît aujourd'hui comme un préalable d'une part à l'amélioration de la qualité des produits et des procédés, et d'autre part à la réduction des temps de développement. La méthode des PE permet en effet d'obtenir un maximum d'informations à un coût minimal. En résumé, les avantages les plus reconnus des PE sont :

- l'efficacité, car seules les expériences indispensables sont réalisées,
- l'exactitude : pour un effort expérimental donné, la plus grande exactitude possible sera atteinte,
- les interactions : les synergies existant entre les différents paramètres étudiés sont identifiées et mieux comprises.

D'une manière générale, un PE consiste à mettre en évidence et à quantifier l'influence existant entre deux types de variables :

- le **facteur** : une variable, ou un état, qui agit sur le système étudié,
- la **réponse** : la grandeur mesurée afin de connaître les effets des facteurs sur le système. Il convient bien sûr que la réponse soit représentative du phénomène observé.

Plus précisément, les PE visent à établir des relations liant la réponse tant avec les facteurs, qu'avec les facteurs entre eux (interactions). Cette technique des PE, employée dans des domaines aussi divers que l'agriculture, la chimie et la pétrochimie, la biologie, l'électronique, permet d'obtenir une modélisation expérimentale des phénomènes physiques et d'apprécier la confiance à accorder aux résultats. A ce stade, nous tenons à insister sur la différence entre la modélisation de l'expérimentation et la modélisation du phénomène physique. Modéliser l'expérimentation, c'est être capable de prévoir la réponse du phénomène physique uniquement en fonction des paramètres étudiés lors de l'expérimentation et variant dans le domaine d'étude choisi. Cela signifie qu'en général :

- la réponse du modèle de l'expérimentation ne pourra théoriquement pas être extrapolée à des valeurs de paramètres situées en dehors du domaine d'étude,
- les valeurs quantitatives des réponses seront liées à des configurations particulières de l'expérimentation, et ne seront pas non plus forcément extrapolables à d'autres conditions de fonctionnement (par exemple, cas d'un changement de machine).

Cette étape relative à la modélisation de l'expérimentation apparaît néanmoins très importante car elle permet tout de même d'appréhender localement le phénomène physique par la connaissance de la surface de réponse générée dans le domaine expérimental considéré.

Les contextes d'application de la méthode peuvent aller du traitement et de la résolution des problèmes qualité à l'optimisation d'un produit (sous l'angle de la conception) ou d'un processus destiné à fabriquer un produit. En outre, les PE peuvent être vus comme un ensemble de méthodes, d'outils qui peuvent être choisis en fonction des objectifs de l'étude à effectuer. Les principales méthodes des PE, liées à des utilisations et besoins différents sont :

➔ **la méthode comparative (en anglais : comparative design)**. Elle permet de trouver la relation cause - effet entre plusieurs facteurs de l'étude, et de déterminer le facteur important a priori. La question est ici de savoir si le facteur a un rôle "significatif" dans l'obtention de la réponse, s'il engendre ou pas un changement crucial de la réponse en fonction de ses différents niveaux.

➔ **la technique de criblage (screening design)**. Elle est mise à profit pour explorer un domaine expérimental inconnu. Les plans sont alors conçus de manière à isoler les facteurs influents et examiner les effets principaux des facteurs. La technique de criblage se nomme également technique principale d'effets.

➔ **la Méthode de la Surface de Réponse (MSR) (Response Surface Methodology, RSM)**. Elle est mise en œuvre pour élaborer des modèles descriptifs ou prévisionnels des phénomènes étudiés. Elle permet une étude qualitative des facteurs. Son utilisation est intéressante dès lors qu'il faut effectuer des optimisations, mettre au point des formulations permettant d'estimer les interactions entre facteurs ainsi que les effets quadratiques. La MSR permet, grâce à l'idée que l'on peut se faire de la forme de la surface de réponse, d'améliorer la qualité (la robustesse, la fiabilité) des produits.

➔ **la modélisation par régression (regression modeling)**. Elle est faite pour estimer les paramètres d'un modèle précis, en mesurant la dépendance de la ou des réponse(s) à l'égard des entrées du processus.

Pour résumer, les PE correspondent à une suite d'essais organisée à l'avance de manière à déterminer, en un minimum d'essais et avec un maximum de précision, l'influence de multiples paramètres sur une ou plusieurs réponses. Les PE permettent de répondre à deux familles de questions :

- Comment estimer et comparer les effets des paramètres de réglage d'un processus ?
- Comment affiner les paramètres de réglage d'un processus pour atteindre un optimum ?

Les travaux de cette thèse s'inscrivent pleinement dans ce double contexte, avec pour objet d'étude la pile à combustible (PàC). Pour illustrer ces différentes questions au travers du sujet qui nous intéresse, prenons quelques exemples.

- Comment estimer et comparer les effets des paramètres de réglage de la PàC tels que le niveau du courant imposé par la charge, la température du stack, les températures, pressions, débits, taux d'hydratation du combustible et du comburant... sur la tension de pile ?
- Comment régler les paramètres de fonctionnement de la PàC pour atteindre un optimum en tension, en puissance, en rendement ?

Les travaux présentés dans ce rapport ont été réalisés en respectant pour le mieux la méthodologie des PE, et en particulier ses différentes étapes et objectifs :

- Définition du problème,
- Application de l'analyse de criblage : étude qualitative et quantitative du système,
- Application de la méthodologie des surfaces de réponse,
- Optimisation.

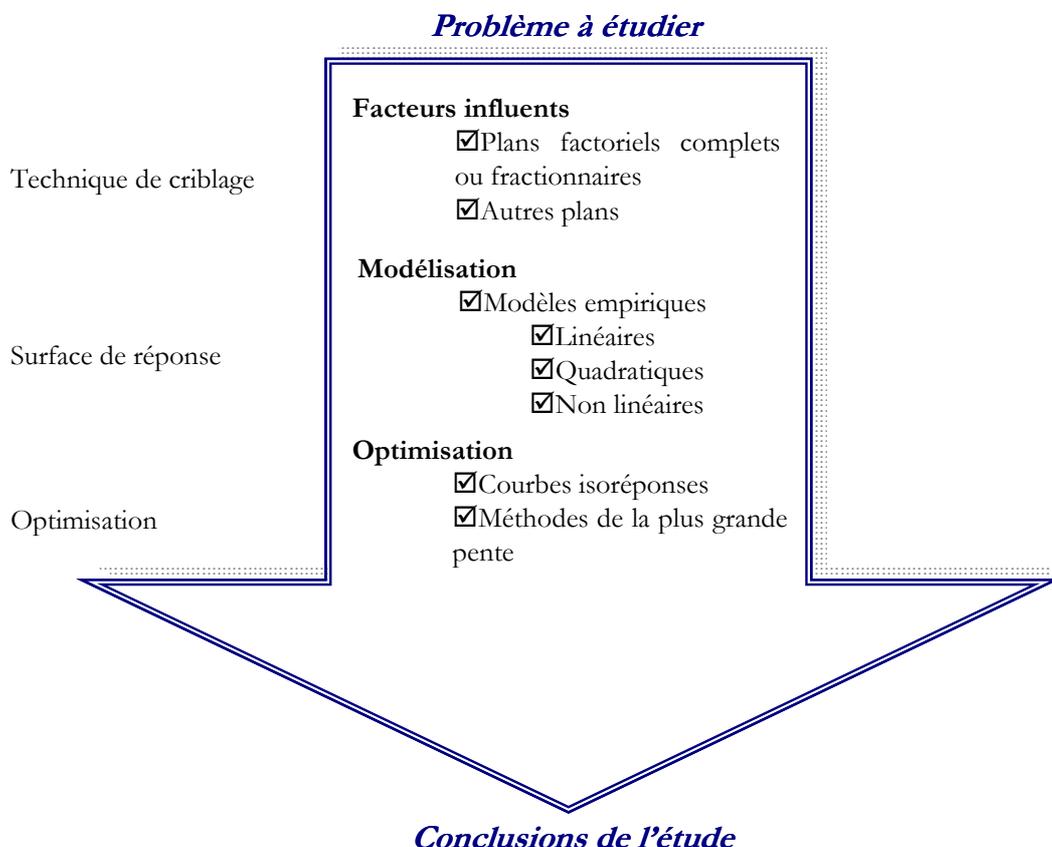


Figure II- 1 : Démarche méthodologique des PE

II.2 RAPPEL HISTORIQUE

Nous nous intéresserons à l'aspect historique des PE en ne considérant presque exclusivement que la période postérieure aux années 1920 [Dag00]. La notion d'expérimentation est bien sûr quant à elle antérieure à cette date [Dro97] [Lec60].

Le point de départ des principes modernes de l'expérimentation est souvent fixé à 1919, date de l'engagement du statisticien (également biologiste évolutionnaire et généticien) Ronald A. Fisher (1890 - 1962) par la Rothamsted Experimental Station (Centre de recherche agronomique) dans le cadre d'études agronomiques. Très rapidement, les travaux de Fisher font apparaître les notions de répétition, de répartition au hasard ou randomisation, de constitution de blocs, d'expériences factorielles, d'effets principaux et d'interactions, de confusions des effets [Fisher, 1925, 1926]. En 1931, Fisher est rejoint par Frank Yates (1902 – 1993). De leur collaboration résultent de nouveaux développements théoriques tels que les expériences factorielles fractionnaires, les notions d'expériences en blocs aléatoires incomplets [Yates, 1935, 1936, 1937].

Deux autres noms peuvent être cités : William G. Cochran (1909 – 1980) et David J. Finney (1917 -). On doit notamment à ces deux chercheurs des contributions importantes concernant les expériences croisées et portant aussi sur les expériences factorielles fractionnaires.

A partir des années 1935-1940, les notions d'expérimentation interviennent également dans le secteur industriel. Des concepts nouveaux apparaissent alors, tels que les plans de Plackett et Burman (1946), la notion de surface de réponse (1952), les plans optimaux (1959) et les plans Taguchi (1959, 1960, 1987) [Dag00].

Depuis, les PE ont pris un essor considérable avec le développement de l'informatique et la puissance de calcul qui l'accompagne.

II.3 ÉLÉMENT DE METHODOLOGIE ET DE TERMINOLOGIE

Avant d'aller plus loin, il est important de bien définir les principaux termes utilisés dans la méthodologie des PE.

II.3.1 VARIABLE

Variable : caractéristique susceptible de prendre plusieurs valeurs d'un ensemble auquel une mesure numérique peut être appliquée (par exemple : revenu, âge, poids). Dans la méthode des PE, des variables mathématiques sont mises en correspondance avec des grandeurs physiques (électriques, physiques, thermiques...) supposées intervenir dans le problème. Chaque variable peut être, selon le cas [VocabStat] [Dro97] :

- une variable quantitative Ses valeurs sont alors des nombres exprimant une quantité et sur lesquels les opérations arithmétiques (somme, soustraction, etc...) ont un sens. La variable peut alors être discrète ou continue selon la nature de l'ensemble des valeurs qu'elle est susceptible de prendre (valeurs isolées ou intervalle).
- une variable qualitative Ses valeurs sont dans ce cas des modalités, ou catégories, exprimées sous forme littérale ou au moyen d'un codage numérique et sur lesquelles des opérations arithmétiques n'ont aucun sens. On distingue des variables qualitatives ordinales et nominales, selon que les modalités peuvent être naturellement ordonnées ou pas.

- une variable continue Variable numérique qui peut prendre un nombre infini de valeurs réelles. L'âge, la distance et la température par exemple sont considérés comme des variables continues.
- une variable discrète Variable numérique qui prend uniquement un nombre limité de valeurs réelles.

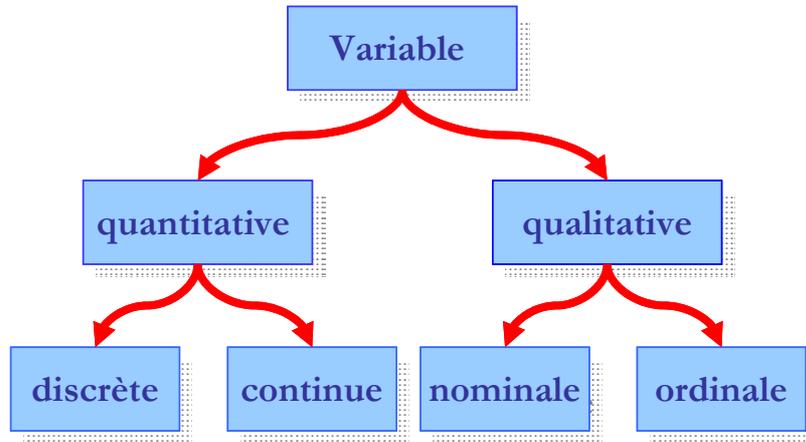


Figure II- 2 : Distinction fondamentale entre les différents types de variables

Remarque :

On note variables parasites ou variables parasites identifiables, les variables indépendantes non prises en considération dans l'expérience et qui sont celles que l'on peut totalement identifier [Pha93]. On cherche alors à les contrôler pour « effacer » leur effet. Les variables parasites non identifiables peuvent être contrôlées par randomisation (cf. II.3.6) [Dro97].

II.3.1.1 Facteur

Les facteurs, parfois appelés Variables Indépendantes (VI), sont les causes, supposées ou certaines, qui provoquent le phénomène. Tous les facteurs susceptibles d'avoir de l'influence sur le phénomène doivent être pris en compte, faute de quoi les résultats risquent d'être faussés et donc inutilisables. Un facteur peut être qualitatif ou quantitatif, continu ou discontinu, contrôlable ou non contrôlable.

- Les **facteurs contrôlés** sont ceux auxquels il est possible à l'avance d'imposer un état.
- Les **facteurs mesurables** sont ceux pour lesquels on ne peut pas imposer un état donné, mais ce sont aussi des facteurs dont on peut connaître précisément le niveau (température ambiante extérieure par exemple).
- Les **facteurs constants** sont maintenus dans un état fixe tout au long de la série d'expériences.
- Les **facteurs bruit** ont des effets supposés ou réels, et qui ne s'exercent que sur la dispersion ou la qualité des résultats (vieillessement du matériel, fatigue de l'opérateur...).

Un facteur varie entre deux bornes :

- la *borne inférieure* (niveau bas que l'on note le plus souvent -1),
- la *borne supérieure* (niveau haut que l'on représente généralement par +1).

L'ensemble de toutes les valeurs que peut prendre le facteur entre le niveau bas et le niveau haut s'appelle le domaine de variation du facteur ou plus simplement domaine du facteur. Un facteur peut prendre plusieurs niveaux à l'intérieur de son domaine de variation. Les différents éléments individuels qui constituent un même facteur sont appelés modalités, que ce soit pour les facteurs qualitatifs ou pour les facteurs quantitatifs.

Remarque :

- Pour qualifier un facteur, les termes de « facteur de variation » ou « facteur expérimental » sont aussi employés.
- Par usage, le terme de « variable indépendante » est utilisé dans un contexte méthodologique alors que le terme de facteur est utilisé dans un contexte statistique.

➔ **Précisions sur la notion de coordonnées centrées réduites**

Lorsque les valeurs -1 et +1 sont attribuées respectivement aux niveaux bas et haut d'un facteur, deux modifications importantes sont effectuées.

- L'unité de mesure est changée. Par exemple, si le niveau bas d'un facteur est fixé à 60°C et le niveau haut à 80°C alors 20°C séparent 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 10°C. Le nom de *pas* lui est donné.
- L'origine des mesures se trouve déplacée. Ainsi, dans l'exemple choisi, le centre de l'intervalle [-1, +1] correspond à la température de 70°C. La nouvelle origine, notée zéro, diffère donc de l'origine exprimée en unités courantes.

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. En anglais, on les trouve sous la dénomination de « coded values » (valeurs codées).

II.3.1.2 Réponse

Une réponse expérimentale, parfois appelée Variable Dépendante (VD), est une caractéristique mesurable d'un produit ou d'un processus, et dont la variation est analysée en fonction des variations des facteurs. La réponse doit être la plus représentative possible du phénomène observé. Le choix des réponses est un problème difficile qui ne relève pas directement de la théorie des PE. Ce n'est qu'après une analyse minutieuse des phénomènes, des enjeux, des objectifs et des contraintes que la ou les bonnes réponses peuvent être définies.

Remarque :

La réponse expérimentale peut être le résultat d'une expérience réelle mais aussi le résultat d'une simulation numérique [Viv02].

II.3.2 DOMAINE EXPERIMENTAL REEL OU DOMAINE D'ETUDE

Le domaine expérimental est l'espace défini par les variations des facteurs quantitatifs et / ou par les combinaisons des modalités des facteurs qualitatifs. Le domaine expérimental ou Domaine d'Etude (DE) peut être défini comme l'ensemble de tous les points de la surface délimitée par les niveaux bas et haut de chaque facteur, un espace k -dimensionnel, dans lequel chaque point représente une combinaison des valeurs possibles pour les k facteurs [Dro97].

Lorsque $k = 2$, un carré est obtenu pour le DE. Les points expérimentaux sont alors situés aux quatre sommets du carré (Figure II- 3).

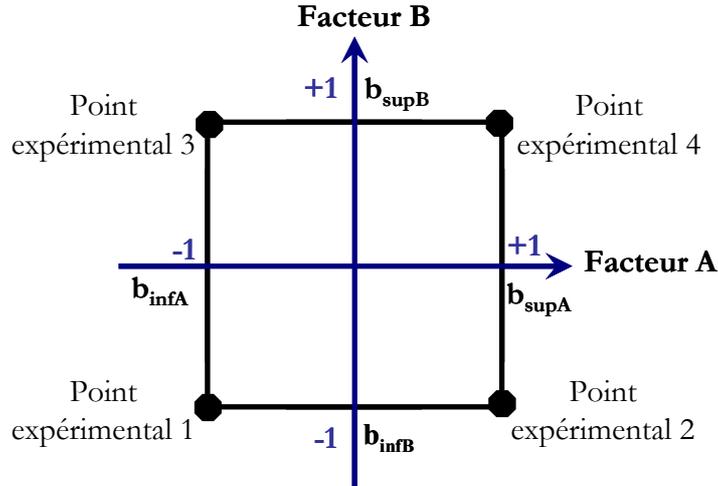


Figure II- 3 : Domaine d'étude d'un plan 2^2

Lorsque $k = 3$, nous obtenons un cube dont les huit sommets représentent les huit essais du plan (Figure II- 4).

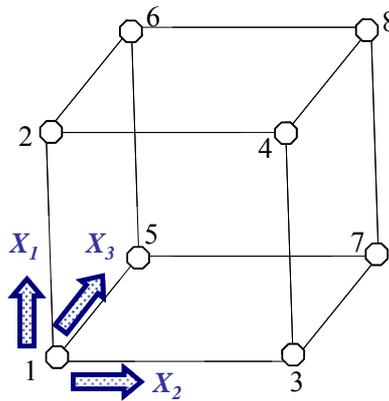


Figure II- 4 : Domaine d'étude d'un plan 2^3

On distingue :

- le *domaine continu* : les variables peuvent prendre n'importe quelle valeur dans le domaine d'étude,
- le *domaine discret* : les réponses sont étudiées relativement à des facteurs discrets ou qualitatifs. Cependant, cette définition peut être rendue plus générale dans la mesure où certains facteurs continus peuvent être discrétisés. Il est alors nécessaire d'introduire la précision (π_f) d'un facteur.

$$\pi_f = \frac{b_{\text{sup}} - b_{\text{inf}}}{N_{ni} - 1} \tag{II. 1}$$

avec :

N_{ni} : le nombre de niveaux du facteur considéré,
 b_{sup} et b_{inf} qui sont respectivement les bornes supérieures et inférieures du facteur.

Dans le cadre d'un problème discret, le domaine d'étude n'est pas un ensemble infini de valeurs mais une grille déterminée de points. (Figure II- 5).

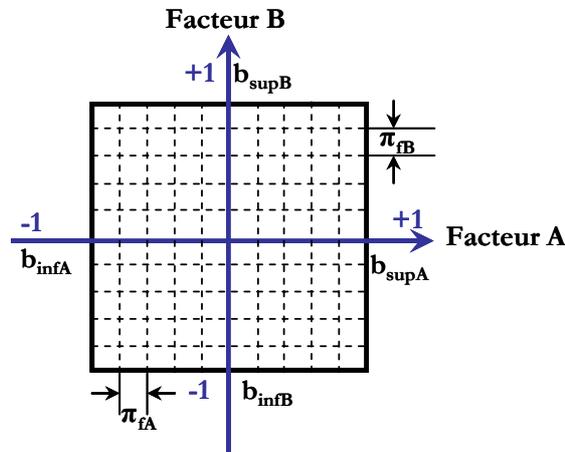


Figure II- 5 : Domaine d'étude discret d'un plan 2^2

- le *domaine mixte* comme étant relatif à une étude faisant intervenir des facteurs continus (non discrétisés) et des facteurs discrets et/ou qualitatifs.

II.3.3 CONTRAINTES

Ce sont les conditions pratiques particulières pour lesquelles il est exclu de réaliser des expériences. Il peut s'agir d'impossibilités manifestes ou plus simplement de limites que s'impose l'expérimentateur. Les contraintes sont de deux types :

- les contraintes en position. Ce sont celles qui ne concernent que les facteurs. Elles lient dans la majorité des cas les facteurs entre eux et ont l'avantage d'être connues avant la réalisation des expériences.
- les contraintes en valeurs atteintes. Elles font intervenir les réponses en plus des facteurs. [Viv02]

II.3.4 DOMAINE EXPERIMENTAL D'INTERET OU DOMAINE D'ETUDE POSSIBLE

Le Domaine d'Etude Possible (DEP) est le Domaine d'Etude (DE) soumis aux contraintes de positions. Il correspond donc à l'espace lié aux expériences réalisables, représenté par la zone grisée de la Figure II- 7.

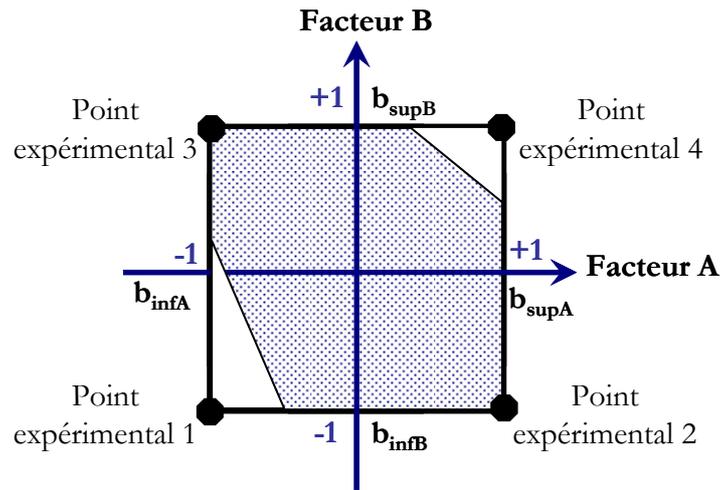


Figure II- 6 Domaine expérimental sous contraintes

Les expériences 1 et 4 ne sont plus réalisables ; la qualité du PE va maintenant s'en trouver dégradée. Le problème qui se pose à ce stade est donc de trouver, dans ce domaine sous contraintes, d'autres expériences qui pourront permettre le calcul des estimateurs des coefficients d'un modèle aboutissant à une précision jugée raisonnable [Dro97]. Une première solution pourrait consister à restreindre le domaine à étudier, de manière à retrouver un sous-domaine de même symétrie que le domaine original, et à y réaliser ensuite un plan classique possible (Figure II- 7 a). Une deuxième solution consiste à proposer un ensemble d'expériences candidates positionnées dans le domaine sous contraintes et à en extraire un sous-ensemble de taille modérée qui puisse permettre le calcul ultérieur du modèle dans les meilleures conditions possibles (Figure II- 7 b).

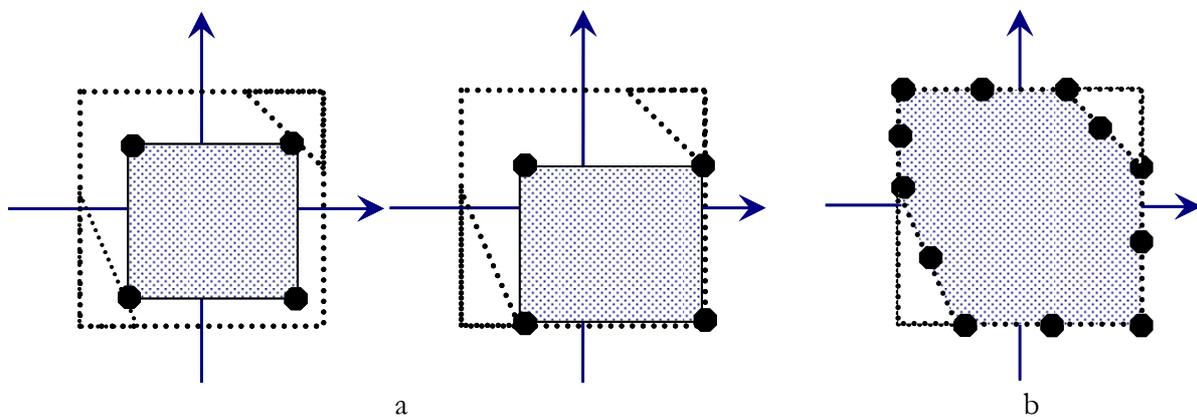


Figure II- 7 :

a- Quelques sous-domaines dans le domaine sous contraintes

b- Domaine expérimental sous contraintes avec des expériences candidates

On distingue :

- le *domaine asymétrique*, désignant un PE ou un domaine expérimental pour lequel les facteurs qualitatifs ne présentent pas le même nombre de modalités.
- le *domaine symétrique*, désignant un PE ou un domaine expérimental pour lequel tous les facteurs qualitatifs présentent le même nombre de modalités.

- le *domaine anisotrope*, qui correspond à un domaine expérimental pour lequel les niveaux des facteurs indépendants sont soumis à des contraintes relationnelles.
- le *domaine isotrope*, qui désigne un domaine expérimental pour lequel les niveaux des facteurs indépendants ne sont pas soumis à des contraintes relationnelles.

II.3.5 EXPERIENCES

Les deux définitions suivantes du mot expérience sont extraites, l'une d'un dictionnaire tout à fait classique, l'autre d'un dictionnaire de synonymes :

- Épreuve qui a pour objet d'étudier un phénomène.
- Fait provoqué ou attendu pour vérifier une hypothèse, une loi, et arriver ainsi à une connaissance théorique de la façon dont se passent les choses.

Comme cela a déjà été évoqué dans le paragraphe II.3.1.2, une distinction peut être faite entre l'expérience réelle et l'expérience virtuelle.

- L'**expérience réelle** est une expérience dans laquelle les phénomènes auxquels nous nous intéressons sont provoqués en maîtrisant certains facteurs, et en observant les conséquences qui en résultent. Les expériences réelles sont soumises aux erreurs expérimentales et aux erreurs de mesures. *L'erreur expérimentale* traduit l'erreur imputable à la conception et à la réalisation des expériences. Elle n'est liée qu'à l'expérimentation. *L'erreur de mesure* traduit la variabilité des réponses du fait de la prise en compte des résultats obtenus à l'issue du processus d'expérimentation.
- L'**expérience virtuelle**, ou issue d'une simulation, est une expérience effectuée non pas sur un système "réel" mais sur son modèle mathématique, traduit en un ou plusieurs programmes informatiques.

II.3.6 PRINCIPAUX DISPOSITIFS EXPERIMENTAUX

Le dispositif expérimental le plus simple est le dispositif complètement aléatoire (ou obtenu par randomisation). Il consiste à planifier et à réaliser les différentes expériences de manière tout à fait aléatoire. D'une manière générale, toute restriction à la répartition aléatoire doit être prise en compte lors de l'analyse des résultats, au risque sinon d'introduire des erreurs systématiques importantes, liées aux confusions qui peuvent exister entre les facteurs pris en considération dans l'expérience et d'autres facteurs, non maîtrisés.

De nombreux autres dispositifs peuvent être trouvés dans la littérature. Ils sont pour certains d'entre eux assez largement utilisés. Nous nous contenterons de citer les dispositifs en blocs aléatoires complets, les dispositifs en carré latin et dispositifs croisés (« cross-over »), les carrés gréco latin et d'autres généralisations des carrés latins, les réseaux ou « lattices », équilibrés et non équilibrés, les carrés latins incomplets ou carrés de Youden et les blocs incomplets non équilibrés ou partiellement équilibrés [Dro97] [Exp05].

II.4 DEMARCHE METHODOLOGIQUE D'UN PE

Toute expérience doit être l'objet d'une planification précise qui se concrétise sous la forme d'un plan d'expériences ou protocole expérimental. La démarche méthodologique d'un PE peut être décomposée en différentes étapes [Exp05] [Pil97].

II.4.1 ETAPE A : DEFINITION DES OBJECTIFS ET DES REPONSES

En tenant compte des objectifs à atteindre, il est d'abord nécessaire de faire la liste des réponses expérimentales qui peuvent être étudiées. Cette étape permet également de mettre en place les moyens et le budget nécessaires à l'étude.

II.4.2 ETAPE B : CHOIX DES FACTEURS ET DU DOMAINE EXPERIMENTAL

C'est de loin l'étape la plus importante dans la conduite d'un PE. Il faut :

- Sélectionner les paramètres, choisir les modalités et les interactions à étudier.
- Recenser les paramètres pouvant influencer la réponse.
- Identifier les interactions susceptibles d'être recherchées.
- Dissocier les facteurs principaux des facteurs bruits.
- Fixer le domaine d'étude pour chacun des facteurs.

II.4.3 ETAPE C : PROPOSITION D'UN MODELE

Le plan le plus adapté à la situation sera retenu. Le plan doit présenter les propriétés suivantes :

- Bien représenter la réponse expérimentale étudiée dans le domaine expérimental d'intérêt.
- Aboutir pour la valeur de la réponse étudiée à une estimation de qualité acceptable.

II.4.4 ETAPE D : ESTIMATION DES COEFFICIENTS DU MODELE

La valeur de la réponse expérimentale doit pouvoir être estimée avec une qualité acceptable en n'importe quel point du domaine expérimental d'intérêt.

II.4.5 ETAPE E : VALIDATION DU MODELE

Deux possibilités existent.

- Soit le modèle est validé, ce qui signifie qu'il représente suffisamment bien le phénomène étudié dans le domaine expérimental et dans ce cas, les objectifs sont atteints : nous pouvons utiliser ce modèle pour faire de la prévision en n'importe quel point du domaine expérimental.
- Soit le modèle n'est pas validé et alors son utilisation n'est pas possible. Il faut donc s'acheminer vers la proposition d'un modèle différent.

II.4.6 ETAPE F : MISE EN ŒUVRE ET SUIVI

- Le calcul de la réponse est possible en tout point du domaine expérimental.
- A partir du modèle calculé, nous pouvons prédire les résultats correspondant à la configuration optimale du produit ou du processus.

II.5 MODELISATION ET INTERPRETATION

II.5.1 OBJECTIF

En l'absence de modèle mathématique complet et éprouvé, la prédiction du comportement des systèmes complexes nécessite une étude multiparamétrique expérimentale. Les problèmes expérimentaux peuvent être synthétisés suivant le schéma de la Figure II- 8 où interviennent en entrée des facteurs susceptibles d'influer sur les réponses, jouant le rôle de causes potentielles, et en sortie un certain nombre de réponses assimilables à des conséquences [Pil97] [Sou94].

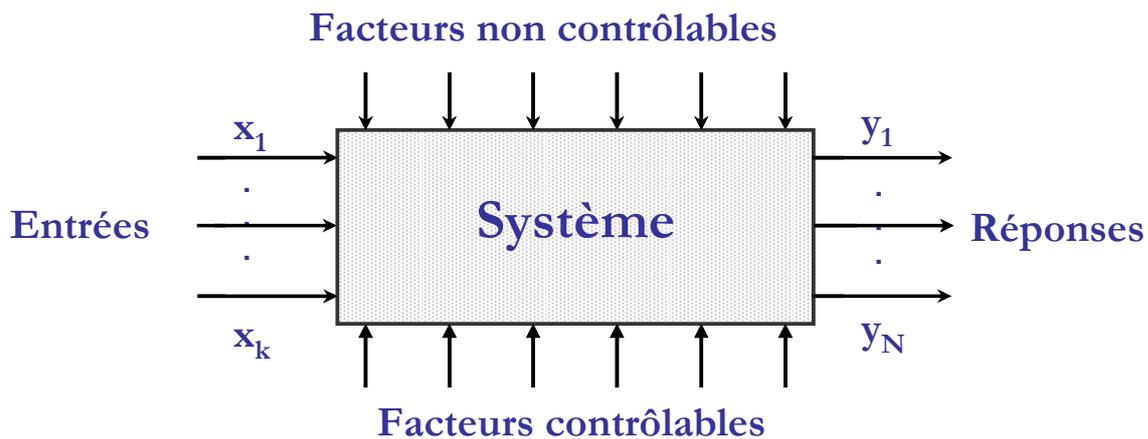


Figure II- 8 : Schéma simplifié de l'environnement d'un système donné

L'approche des PE présente souvent un intérêt majeur dans la mise au point d'une technologie complexe, telle que celle de la pile à combustible. Le but consiste à modéliser le comportement des procédés et/ou des produits afin de mieux prévoir et accroître leurs performances. La modélisation expérimentale doit permettre de définir les conditions d'utilisation optimale et de déterminer les facteurs à contrôler ou à piloter afin de maîtriser le procédé. A l'instar de la mise au point d'un procédé de fabrication, l'optimisation du fonctionnement d'un système complexe nécessite la construction d'un modèle expérimental.

II.5.2 MODELISATION

L'objectif est la mise en forme d'un modèle, le plus souvent polynomial, décrivant les variations de la fonction réponse y prenant les valeurs y_1, y_2, \dots, y_N relativement aux valeurs de k facteurs x_1, x_2, \dots, x_k .

$$y = f(x_1, x_2, \dots, x_k) \tag{II. 2}$$

avec :

N : le nombre d'expériences ;

p : le nombre de coefficients du modèle postulé ;

y : le vecteur colonne des réponses expérimentales y_i . $y^t = [y_1, y_2, \dots, y_N]$

La méthode de la régression multilinéaire est l'outil statistique le plus habituellement mis en œuvre pour l'étude de données multidimensionnelles. Une variable quantitative y dite à expliquer est mise en relation avec k variables quantitatives x_1, x_2, \dots, x_k dites explicatives.

Posons : ➤ X : la matrice $N \times p$, appelée matrice du modèle ou matrice des effets :

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1,p-1} \\ 1 & x_{21} & x_{22} & \dots & x_{2,p-1} \\ 1 & \vdots & . & \ddots & \vdots \\ 1 & x_{N1} & x_{N2} & \dots & x_{N,p-1} \end{bmatrix}$$

➤ β : le vecteur colonne des paramètres à estimer β_i :

$$\beta^t = [\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}]$$

➤ e : le vecteur colonne des erreurs expérimentales, aléatoires e_i :

$$e^t = [e_1, e_2, \dots, e_N]$$

Le modèle matriciel, s'écrit sous la forme classique suivante :

$$y = X \cdot \beta + e \tag{II. 3}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1,p-1} \\ 1 & x_{21} & x_{22} & \dots & x_{2,p-1} \\ 1 & \vdots & . & \ddots & \vdots \\ 1 & x_{N1} & x_{N2} & \dots & x_{N,p-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix} \tag{II. 4}$$

Les réponses calculées par les modèles utilisés ne sont généralement pas exactement égales aux réponses mesurées expérimentalement. Pour chaque expérience existe alors un écart dit de modélisation. Le vecteur ε représentera ces écarts.

$$y = X \cdot \hat{\beta} + \varepsilon \tag{II. 5}$$

Mathématiquement, le vecteur ϵ des résidus se définit comme suit :

$$\epsilon = y - \hat{y} \quad (\text{II. 6})$$

L'objectif est donc le calcul du vecteur $\hat{\beta}$, valeur estimée du vecteur β .

II.5.3 LES HYPOTHESES DE LA REGRESSION MULTILINEAIRE

Le calcul des coefficients des facteurs n'est possible qu'à la condition de faire les hypothèses suivantes :

- La réponse est la somme d'une quantité non aléatoire et d'une quantité aléatoire.
- Les paramètres inconnus $\beta_0, \dots, \beta_{p-1}$ sont supposés constants.
- Les écarts dits de modélisation (ϵ) sont purement aléatoires et ne contiennent pas d'erreurs systématiques.
- Les écarts ne sont pas corrélés entre eux.
- Les écarts sont normalement distribués.
- Les écarts (ou résidus) ont une moyenne nulle : $\text{Esp}(\epsilon_j) = 0$, où Esp représente l'espérance ou la fonction moyenne arithmétique.
- Les écarts sont issus d'une seule et même population et sont identiquement distribués. Ceci peut se traduire par l'équation : $\text{Var}(\epsilon) = \sigma^2 I$, où Var représente la fonction variance, σ l'écart type et I la matrice identité. La variance d'une variable est une mesure de la dispersion de ses valeurs autour de sa valeur moyenne.
- La distribution des écarts ne dépend pas des niveaux des facteurs.

II.5.4 TYPOLOGIE DES MODELES COURANTS

Il convient que la forme générale du modèle empirique soit adaptée aux objectifs de l'étude. Différents types de modèle sont envisageables [Exp05] :

➔ **Modèle additif sans couplage**

Lorsque l'objectif consiste à hiérarchiser les effets moyens des facteurs à partir d'une étude de criblage, la forme générale suivante, appelée modèle additif sans couplage, peut être adoptée.

$$y = \beta_0 + \sum_{i=1}^k \beta_i \cdot x_i + \epsilon \quad (\text{II. 7})$$

Le coefficient β_0 représente la moyenne arithmétique des réponses mesurées y à partir des N expériences du plan.

$$\beta_0 = \bar{y} = \frac{1}{N} \cdot \sum_{i=1}^N y \quad (\text{II. 8})$$

➔ **Modèle additif avec couplage**

Lorsque l'on veut préciser l'effet moyen des facteurs par des interactions d'ordre un (interaction entre deux facteurs), la forme générale suivante, appelée modèle additif avec couplage, est envisageable.

$$y = \beta_0 + \sum_{i=1}^k \beta_i \cdot x_i + \sum_{i=1}^k \sum_{j=i+1}^k \beta_{ij} \cdot x_i \cdot x_j + \varepsilon \quad (\text{II. 9})$$

Les termes β_i et β_{ij} symbolisent respectivement les effets moyens des facteurs x_i et les interactions d'ordre un entre les effets des facteurs x_i et x_j .

➔ **Modèle polynomial**

Lorsqu'un optimum doit être trouvé, la forme générale suivante, qui présente des termes du second degré, pourra être retenue.

$$y = \beta_0 + \sum_{i=1}^k \beta_i \cdot x_i + \sum_{i=1}^k \beta_{ii} \cdot x_i^2 + \sum_{i=1}^k \sum_{j=i+1}^k \beta_{ij} \cdot x_i \cdot x_j + \varepsilon \quad (\text{II. 10})$$

A chacune des formes générales précédentes est associée une équation permettant d'estimer le nombre d'inconnues, noté p .

- Pour un modèle additif sans couplage (II. 7), le nombre d'inconnues à estimer est défini à partir de la relation suivante :

$$p = 1 + \sum_{i=1}^k (m_i - 1)$$

avec m_i : nombre de modalités du facteur x_i .

- Pour un modèle additif avec couplage (II. 9), le nombre d'inconnues à estimer est défini à partir de la relation suivante :

$$p = 1 + \sum_{i=1}^k (m_i - 1) + \sum_{j \neq i} (m_i - 1)(m_j - 1)$$

- Pour un modèle polynomial du second degré (II. 10), le nombre d'inconnues est donné par :

$$p = \frac{(k+1)(k+2)}{2}$$

Remarque :

Pour prendre en compte dans un modèle les influences non linéaires de certains facteurs, il est possible :

- *de rajouter des monômes de degré plus élevé pour les variables explicatives ($\beta \cdot x^2, \beta \cdot x^3, \dots$),*
- *d'utiliser des changements de variables tels que $1/x, \log x$. Le modèle n'est alors plus linéaire par rapport aux coefficients mais le système d'équations demeure linéaire.*

➔ **Autres modèles**

L'utilisation de plans particuliers, autres que les plans factoriels complets implique dans la majorité des cas l'utilisation de modèles adaptés. Il faut en effet tenir compte des interactions prises en compte par de tels plans. Un cas également fréquent est celui où certains termes sont absents du développement limité habituel. La forme polynomiale (polynomiale par rapport aux coefficients β) incomplète suivante peut être donnée à titre d'illustration [Gau05]:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \beta_{11} \cdot x_1^2 + \beta_{23} \cdot x_2 \cdot x_3 + \beta_4 \cdot \text{Log}(x_3) + \varepsilon \quad (\text{II. 11})$$

On a $k = 3$ et $p = 7$, et les fonctions de régression sont :

$$f_0(x) = 1, f_1(x) = x_1, f_2(x) = x_2, f_3(x) = x_3, f_4(x) = x_1^2, f_5(x) = x_2 \cdot x_3, f_6(x) = \text{Log}(x_3).$$

La définition de ces modèles particuliers est assurée par le mode de construction des plans correspondants.

→ Modélisation non polynomiale

Il se peut que le modèle statistique soit trop simple eu égard à la complexité de la réponse étudiée et au niveau d'erreur souhaité. Il est alors nécessaire de faire appel à des approches statistiques plus sophistiquées telles que le krigeage, l'approche bayésienne, la tessellation irrégulière, les éléments diffus ... [Jou05] [Sou94].

La Figure II- 9 correspond à un schéma synoptique de l'étude des systèmes multidimensionnels complexes. Il fait apparaître de manière synthétique les différentes approches envisageables pour ce type d'étude, en même temps que les voies possibles pour représenter, modéliser les systèmes.

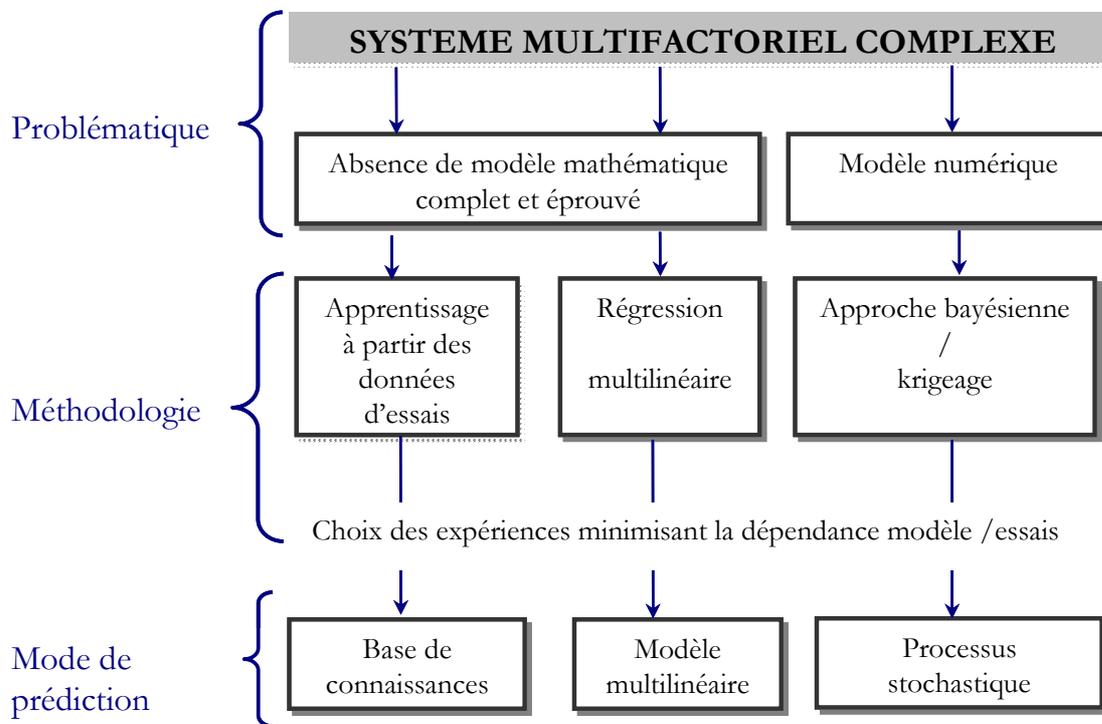


Figure II- 9 : Schéma synoptique de l'étude des systèmes multidimensionnels complexes [Sou94]

La résolution des systèmes multifactoriels complexes est un préalable à la prédiction et à la détermination des paramètres de mise en œuvre dans un processus d'optimisation. La modélisation expérimentale de la réponse d'un système peut s'insérer en amont d'une étude théorique dont le but est de comprendre et de formaliser les phénomènes physiques. Cette démarche constitue d'ailleurs pour des problèmes complexes la seule alternative réellement praticable. Dans ce cas, l'expérimentation doit permettre d'appréhender les différents phénomènes physiques intervenant, ainsi que de déterminer les influences des différents facteurs et les couplages entre ces facteurs.

Deux grandes familles d'objectifs permettent d'apporter des éléments de réponse. Il faut d'une part pouvoir comparer les effets des facteurs agissant sur une réponse. Nous pouvons dans ce cas recourir à des *plans de criblage* qui font appel à des modèles additifs avec ou sans interaction. L'analyse de variance est alors l'outil de dépouillement privilégié des résultats d'essais. D'autre part, il faut pouvoir modéliser la variation d'une réponse dans un domaine expérimental afin d'atteindre des conditions optimales, par *la méthode des surfaces de réponses* par exemple.

Il convient de bien distinguer ces deux familles d'objectifs, car de là découlent le type de modèle, la construction du plan ainsi que le choix d'une méthode d'analyse des résultats d'essais [Lou04] [Sou94].

II.6 TECHNIQUE DE CRIBLAGE

La technique de criblage (ou *screening design* en langue anglaise) permet d'étudier dans un domaine expérimental inconnu les facteurs affectant une variable de réponse, et d'isoler les plus influents. Il s'agit de la première étape de la démarche méthodologique qui consiste à bien décrire le problème et en particulier à bien préciser ce que l'on attend d'une campagne expérimentale.

Cette première technique et étape des PE permet de répondre à ces quelques questions :

- Quel est l'effet d'un changement de modalités d'un facteur sur la réponse observée ?
- Comment comparer le plus objectivement possible les effets de plusieurs facteurs ?
- Les effets d'un facteur sont-ils indépendants des modalités des autres facteurs ?

De façon plus générale, des tests statistiques qui permettent de classer les facteurs entre eux, relativement à leur influence propre, sont effectués. Cela permet ainsi de rejeter ou non l'hypothèse selon laquelle le facteur n'induit pas de variation de la réponse significativement plus importante que celles engendrées par le bruit.

II.6.1 PLAN FACTORIEL COMPLET

La réalisation d'un PE peut être très simple ou très compliquée suivant le nombre de facteurs étudiés et les hypothèses formulées. L'étude d'un plan complet consiste à étudier toutes les combinaisons possibles des facteurs pris en considération dans l'expérience. On note ce plan X^k , ce qui signifie que cette expérimentation concerne un système comportant k facteurs à X niveaux. Le principal inconvénient d'une telle méthode réside dans le nombre d'expériences nécessaires, vite dissuasif lorsque k devient important. Cependant, son grand avantage est qu'aucun facteur n'introduit de biais (ou erreur systématique selon la norme ISO/DIS 3534-2 [Sou94]) dans le calcul des effets des autres facteurs (effets indépendants).

Pour étudier une réponse y en fonction de k facteurs, les expériences correspondant à toutes les combinaisons possibles de facteurs, sont réalisées. Si chaque facteur A_i possède N_i niveaux, alors le nombre d'essais N à accomplir est donné par :

$$N = \prod_{i=1}^k N_i \tag{II. 12}$$

Par exemple, pour deux facteurs à deux niveaux, le nombre de combinaisons possibles entraîne $2^2 = 4$ configurations (Tableau II- 1). Dans le cas d'un plan comportant des facteurs ayant des niveaux différents, le calcul du nombre d'expériences du plan complet est effectué de manière similaire. Par exemple, pour un plan complet de 3 facteurs à 2 niveaux et 2 facteurs à 4 niveaux, $2^3 \times 4^2 = 128$ expériences sont nécessaires.

Tableau II- 1 : Plan factoriel complet 2^2 pour deux facteurs (A, B) à deux niveaux

N° essai	Facteurs		Réponse
	Facteur A	Facteur B	
1	-1 (niveau bas)	-1	y_1
2	+1 (niveau haut)	-1	y_2
3	-1	+1	y_3
4	+1	+1	y_4

II.6.1.1 Transformation des variables

Il est utile et souvent nécessaire de transformer les variables afin de comparer les influences des facteurs entre elles et de prendre en compte les facteurs qualitatifs.

1. les facteurs quantitatifs

Si l'on veut mesurer l'influence d'un facteur quantitatif A, en introduisant dans le modèle linéaire la variable x prenant la valeur du facteur, l'influence estimée correspondante dépend fortement des unités utilisées, et ne peut donc pas être comparée à l'influence d'un autre facteur. La solution est alors d'adimensionner les valeurs prises par les facteurs quantitatifs en utilisant les variables centrées réduites [Gou98]. Toutes les variables x varient alors dans un intervalle compris entre -1 et +1, comme cela a déjà été évoqué dans le paragraphe II 3.1.3.

2. les facteurs qualitatifs

Les facteurs qualitatifs ne prennent pas de valeurs continues et réelles. L'estimation des effets nécessite un codage des variables discrètes utilisées. Ainsi, la prise en compte d'un facteur qualitatif à 3 niveaux (ou modalités) peut-elle se faire à partir de 3 variables binaires V_i (i variant de 1 à 3). L'expression $a_1V_1+a_2V_2+a_3V_3$ est mise en œuvre dans le modèle, en introduisant une condition de centrage sur les coefficients telle que :

$$\sum_{i=1}^n a_i = 0$$

Le codage ne fait plus alors intervenir que deux variables $v_1 = (V_1-V_3)$ et $v_2 = (V_2-V_3)$, représentées dans le tableau ci-dessous :

Tableau II- 2 : Codage des facteurs qualitatifs

Niveau du facteur i	V_1	V_2	V_3	v_1	v_2
1	1	0	0	1	0
2	0	1	0	0	1
3	0	0	1	-1	-1

II.6.1.2 Matrice des expériences

L'introduction aux plans factoriels complets se fera ici par l'utilisation de l'approche matricielle.

La *matrice des expériences* (X_N) est une entité mathématique présentée sous forme de tableau comportant autant de colonnes que de facteurs (k), et autant de lignes que de combinaisons (N) de niveaux ou de modalités retenus dans le PE. La matrice s'exprime sous forme codée (au moyen des nombres -1 et +1). Ainsi, pour le plan factoriel complet 2^2 (Tableau II- 1), la matrice des expériences s'écrit :

$$X_N = \begin{bmatrix} -1 & -1 \\ +1 & -1 \\ -1 & +1 \\ +1 & +1 \end{bmatrix}$$

II.6.1.3 Construction de la matrice du modèle

La matrice X des effets, servant au calcul des coefficients du modèle, s'obtient en ajoutant à gauche de la matrice des expériences une colonne ne contenant que des 1.

Fisher et Yates ont montré qu'une matrice orthogonale, telle que la matrice X , conduit à l'indépendance des estimations des coefficients du modèle [Sou97].

Pour mémoire et selon la norme ISO 3534-3 [Exp05], un arrangement orthogonal est un ensemble de combinaisons de traitements tel que pour chaque paire de facteurs, chaque combinaison de traitements survient un même nombre de fois pour tous les niveaux possibles des facteurs. Un dispositif expérimental orthogonal est facile à construire, à partir de simples règles de permutation circulaire. Un dispositif expérimental orthogonal offre une incertitude minimale pour l'estimation des inconnues d'un problème, en particulier parce que les combinaisons retenues dans sa structure sont parfaitement équilibrées. Toutes les modalités apparaissent un même nombre de fois pour chacun des facteurs.

II.6.1.4 Calcul des effets factoriels et des interactions

Les facteurs qui pourraient avoir une influence sur la réponse choisie doivent être recherchés. La connaissance de leurs effets permet alors d'écrire la relation liant la réponse y aux valeurs des facteurs x [Pil97].

➔ Notion d'effet d'un facteur

L'effet principal du facteur A au niveau i , noté E_{Ai} , est calculé de la manière suivante :

$$E_{Ai} = \text{Moyenne des réponses lorsque } A \text{ est au niveau } i - \text{Moyenne générale}$$

L'effet moyen d'un facteur est défini comme étant la moitié de l'effet global.

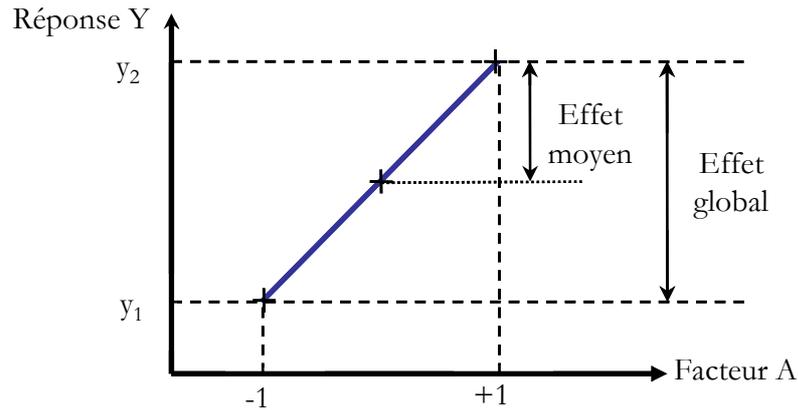


Figure II- 10 : Effet d'un facteur

On distingue :

- L'effet global : $y_2 - y_1$,
- L'effet moyen : $(y_2 - y_1) / 2$.

➔ **Notion d'interaction**

Lorsque l'effet d'un facteur dépend de la valeur prise par un autre facteur, cela se traduit par une *interaction* entre les deux facteurs. On appelle interaction d'ordre n une interaction entre n facteurs.

L'interaction $I_{A_i B_j}$ est recherchée à partir des résultats moyens des essais réalisés lorsque le facteur A se trouve au niveau i et le facteur B au niveau j. Dans tous ces essais, l'effet E_{A_i} joue sur la réponse, de même que l'effet E_{B_j} . Si M est la moyenne générale des essais, il est possible d'écrire :

$$I_{A_i B_j} = \text{Moyenne des réponses lorsque } (A=i, B=j) - M - E_{A_i} - E_{B_j}$$

Comme pour les effets principaux, nous pouvons représenter les interactions sur un graphe qui facilite l'interprétation des résultats (Figure II- 11). La présence d'une interaction apparaît sur ce graphe lorsque les deux droites ne sont pas parallèles.

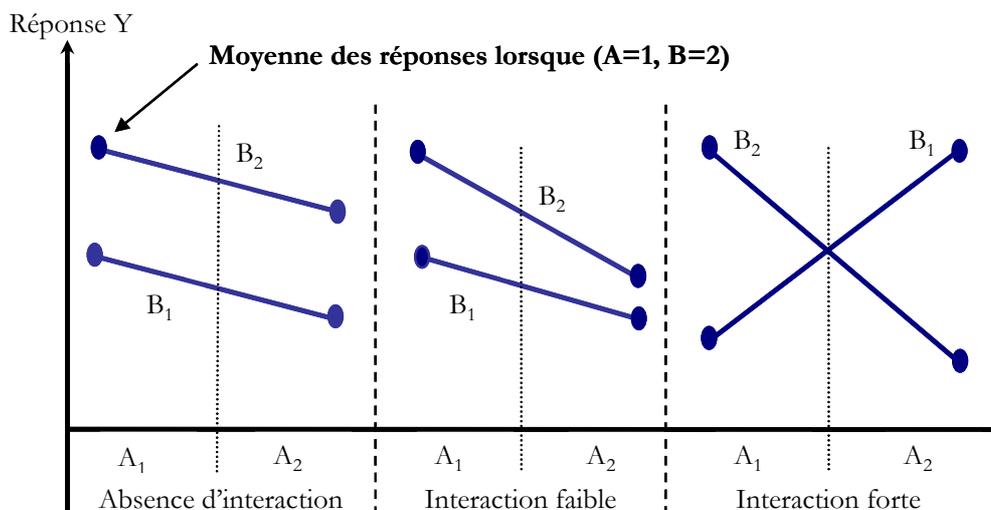


Figure II- 11 : Typologie des interactions

Dans la première configuration de la Figure II- 11, le remplacement de la modalité A_1 par la modalité A_2 occasionne une diminution de la réponse observée, ceci indépendamment de la modalité retenue pour le facteur B. D'un point de vue graphique, les effets réels sont matérialisés par des droites parallèles. L'effet moyen est égal aux effets réels : il y a absence de couplage ou d'interaction.

Dans la deuxième configuration de la Figure II- 11, le remplacement de la modalité A_1 par la modalité A_2 occasionne une diminution de la réponse observée, mais l'amplitude des effets réels dépend de la modalité retenue pour le facteur B. La présence d'une interaction faible se traduit graphiquement par des droites non parallèles. En pratique, la présence de couplages faibles perturbe peu l'additivité des effets moyens.

Dans la troisième configuration de la Figure II- 11, les droites se croisent et témoignent d'une interaction forte. Dans ce cas, les couplages forts perturbent de manière importante l'additivité des effets moyens.

➔ Notion de degrés de liberté d'un modèle

Le nombre de degrés de liberté (ddl) d'un modèle correspond au nombre de variables indépendantes qui le composent. Ainsi, le nombre de degré de liberté d'un modèle à k variables est de $k+1$ (k ddl associés aux variables plus un ddl associé à la constante β_0 du modèle qui est la moyenne arithmétique des réponses). Evidemment, la régression linéaire n'est possible que si le nombre de ddl du modèle est inférieur ou égal au nombre d'essais.

II.6.2 PLANS FACTORIELS FRACTIONNAIRES

Le principal inconvénient des plans factoriels complets est le nombre excessif d'essais à réaliser lorsque le nombre de facteurs devient important. La question est alors de savoir s'il est possible d'estimer les coefficients du modèle avec une précision acceptable sans pour autant réaliser tous les essais du plan factoriel.

Les travaux de Box et Hunter (1978) [Box78] d'une part et Taguchi (1987) [Pil97] d'autre part ont débouché sur les plans fractionnaires présentés sous forme de tables standards en fonction des éléments que l'on souhaite prendre en compte dans l'expérimentation et qui concernent les facteurs, les valeurs (ou niveaux) de ces facteurs, et les interactions entre facteurs [Gou96] [Dro97] [Box78].

L'hypothèse faite dans l'élaboration de plans fractionnaires est que certaines interactions sont insignifiantes et qu'elles peuvent donc être confondues avec des facteurs dont l'influence est significative. L'inconvénient majeur des plans fractionnaires est qu'ils occasionnent par conséquent des risques potentiels d'erreurs. Les avantages de ces plans résident dans le fait qu'ils sont beaucoup plus économiques, plus rapides et demandent moins d'essais.

II.6.2.1 Plans fractionnaires réalisables à partir d'un plan factoriel complet 2^p

Les plans fractionnaires utilisent seulement une fraction de toutes les combinaisons possibles contenues dans un plan factoriel complet. Pour illustrer ce propos, considérons l'exemple d'un PE de dimension 2^3 (Tableau II- 3) :

Tableau II- 3 : plan factoriel complet (3 facteurs à 2 niveaux)

	I	Facteurs			Interactions				Réponse	
		x_1	x_2	x_3	$x_1 \cdot x_2$	$x_1 \cdot x_3$	$x_2 \cdot x_3$	$x_1 \cdot x_2 \cdot x_3$	Y	
N° d'expérience	1	+1	-1	-1	-1	+1	+1	+1	-1	y_1
	2	+1	+1	-1	-1	-1	-1	+1	+1	y_2
	3	+1	-1	+1	-1	-1	+1	-1	+1	y_3
	4	+1	+1	+1	-1	+1	-1	-1	-1	y_4
	5	+1	-1	-1	+1	+1	-1	-1	+1	y_5
	6	+1	+1	-1	+1	-1	+1	-1	-1	y_6
	7	+1	-1	+1	+1	-1	-1	+1	-1	y_7
	8	+1	+1	+1	+1	+1	+1	+1	+1	y_8

Il existe $2^3-1=7$ manières possibles de couper le plan en deux parties (soit autant de fois que le nombre de colonnes associées aux facteurs et aux interactions). Dans notre cas, nous supposons que les interactions du troisième ordre sont négligeables, ce qui est une hypothèse souvent réaliste. Il suffit alors de conserver les essais pour lesquels l'interaction $x_1 \cdot x_2 \cdot x_3$ vaut par exemple +1 (ou alors -1). Le plan obtenu est noté 2^{3-1} et est représenté par le tableau suivant (Tableau II- 4) :

Tableau II- 4 : plan fractionnaire 2^{3-1}

	I	Facteurs			Interactions				Réponse
		x_1	x_2	x_3	$x_1 \cdot x_2$	$x_1 \cdot x_3$	$x_2 \cdot x_3$	$x_1 \cdot x_2 \cdot x_3$	Y
N° d'essai	2	+1	+1	-1	-1	-1	+1	+1	y_2
	3	+1	-1	+1	-1	+1	-1	+1	y_3
	5	+1	-1	-1	+1	+1	-1	+1	y_5
	8	+1	+1	+1	+1	+1	+1	+1	y_8

Il est possible de couper successivement un plan factoriel complet à partir de plusieurs colonnes mais nous allons voir que la diminution du nombre d'expériences entraîne des confusions dans la détermination des actions.

Dans le cas général, les plans fractionnaires sont notés 2^{p-q} (nombre correspondant au nombre d'essais), où p représente le nombre de facteurs étudiés et q le nombre de colonnes de coupure du plan.

II.6.2.2 Notions d'alias

En observant le demi plan obtenu précédemment en gardant les +1 de la colonne $x_1 \cdot x_2 \cdot x_3$, nous remarquons que les colonnes x_1 et $x_2 \cdot x_3$, les colonnes x_2 et $x_1 \cdot x_3$, les colonnes x_3 et $x_1 \cdot x_2$ sont identiques deux à deux. Le fait que les colonnes associées aux actions x_1 et $x_2 \cdot x_3$ sont identiques, signifie que l'effet du facteur x_1 sera confondu avec l'interaction $x_2 \cdot x_3$, et inversement. Nous appelons x_1 et $x_2 \cdot x_3$, x_2 et $x_1 \cdot x_3$, x_3 et $x_1 \cdot x_2$ des alias (des actions confondues). Ainsi, les effets estimés ne représentent pas réellement l'influence réelle du facteur et la présence de confusions induit un biais dans l'estimation de l'influence des facteurs. Par exemple, l'effet estimé E'_{x_1} du facteur x_1 correspond à l'estimation d'une combinaison linéaire de l'effet réel de E_{x_1} et de l'interaction $x_2 \cdot x_3$ notée $I_{x_2 \cdot x_3}$.

$$E'_{x_1} = E_{x_1} + I_{x_2 \cdot x_3}$$

De façon identique, nous avons pour les autres facteurs :

$$E'_{x_2} = E_{x_2} + I_{x_1 \cdot x_3}$$

$$E'_{x_3} = E_{x_3} + I_{x_1 \cdot x_2}$$

Nous allons montrer qu'il est possible d'utiliser les plans fractionnaires malgré la notion de confusion en attribuant aux actions du modèle des actions supposées non influentes (généralement des interactions d'ordre élevé). L'utilisation d'un plan fractionnaire nécessite de recenser les actions confondues sous peine d'attribuer des effets à des actions non influentes.

II.6.2.3 Générateurs d'alias

La construction d'un plan orthogonal fractionnaire à partir du plan factoriel complet n'est pas toujours aussi simple surtout lorsqu'il y a plusieurs colonnes de coupure [Sou94]. Il existe une règle de détermination des alias définie par Box et Hunter (1978) [Box78], dont le principe est de multiplier modulo 2 une action par la ou les colonnes de coupure pour obtenir ses alias. La colonne de coupure est égale à la colonne I correspondant au terme constant du modèle linéaire, définissant le générateur d'alias I. Si nous appliquons cette règle à l'exemple précédent avec la colonne de coupure notée $I=x_1 \cdot x_2 \cdot x_3$ ou GEN (comme générateur d'alias), l'alias de x_1 s'obtient en multipliant modulo 2 le facteur x_1 par le générateur d'alias. Soit :

$$x_1 \cdot I = x_1 \cdot x_1 \cdot x_2 \cdot x_3 = x_1^2 \cdot x_2 \cdot x_3 = x_2 \cdot x_3.$$

Ainsi l'alias de x_1 est l'interaction $x_2 \cdot x_3$. Le tableau des confusions est obtenu en appliquant cette règle aux autres actions :

Action	Alias
I	$x_1 x_2 x_3$
x_1	$x_2 x_3$
x_2	$x_1 x_3$
x_3	$x_1 x_2$
$x_1 x_2$	x_3
$x_1 x_3$	x_2
$x_2 x_3$	x_1

La stratégie permettant la construction du PE consiste à associer les facteurs aux colonnes du plan standard fourni par Box et Hunter [Box78] en fonction du tableau des alias et des interactions que l'on désire estimer ou que l'on suppose non influentes. Ce choix combinatoire conditionne la qualité du modèle obtenu en fonction de la validité des hypothèses formulées sur les actions prises en compte.

II.6.2.4 Notion de résolution

On appelle actions d'ordre I : les facteurs x_1, x_2, \dots , actions d'ordre II : les interactions de type $x_1 \cdot x_2, x_2 \cdot x_3 \dots$ et actions d'ordre III : les interactions de type $x_1 \cdot x_2 \cdot x_3 \dots$ [Pil99].

La notion de résolution définit pour les plans fractionnaires 2^{p-q} le degré des actions confondues et permet de choisir le PE permettant de déterminer les actions du modèle sans confusion. Une résolution de 3 signifie que les actions d'ordre 1 (facteurs) sont confondues avec les actions d'ordre 2 (interactions entre 2 facteurs).

Les plans de résolution IV sont plus intéressants. Les facteurs (d'ordre I) sont aliassés avec des actions d'ordre III ($I + III = IV$). Par contre, les actions d'ordre II seront aliassées avec d'autres actions d'ordre II ($II + II = IV$). Ces plans sont très intéressants car, comme les actions d'ordre III sont souvent insignifiantes, les facteurs principaux seront déterminés sans ambiguïté. Les plans de résolution V sont idéaux mais souvent peu fractionnaires et donc coûteux à réaliser.

En général, les plans fractionnaires de résolution r sont notés 2_r^{p-q} avec p le nombre de facteurs à deux niveaux et q le nombre de colonnes de coupure.

II.6.3 CONDITIONS DE CONSTRUCTION DES PLANS FRACTIONNAIRES

La construction d'un plan fractionnaire s'apparente à une optimisation du nombre d'essais N , fonction des facteurs et du nombre de niveaux par facteur que l'on souhaite étudier. Ce travail peut s'avérer difficile. Le choix du nombre d'essais doit respecter les conditions suivantes :

1. Condition d'orthogonalité du PE,
2. Condition sur le nombre de degrés de liberté.

II.6.3.1 Condition d'orthogonalité

Cette condition est indispensable pour pouvoir calculer les effets d'un facteur indépendamment des autres facteurs. Une méthode récente définie par M. Sisson et Vigier (1988) [Pil97] permet de déterminer le nombre minimal d'essais du plan orthogonal en fonction du modèle postulé par la règle du PPCM (plus petit commun multiple). Cette règle définit un minorant du nombre d'essais du plan orthogonal comme le PPCM de tous les multiples du plan calculés pour chaque combinaison de deux actions disjointes (action ne comportant pas de facteurs en commun).

Choisissons de nous intéresser à l'exemple suivant où l'objectif est de construire un plan orthogonal vis-à-vis d'un modèle en fonction de quatre facteurs et deux interactions.

- Les facteurs A, B, et D sont des facteurs à 3 niveaux.
- Le facteur C a 2 niveaux.
- Les interactions BC et CD comportent donc 6 niveaux.

Pour vérifier l'orthogonalité, on utilise le tableau à double entrée suivant :

Tableau II- 5 : Recherche de la condition sur le PPCM

Conditions sur les actions disjointes	A	3	*					
	B	3	3^2	*				
	C	2	2×3	2×3	*			
	D	3	3^2	3^2	2×3	*		
	BC	2×3	2×3^2	*	*	2×3^2	*	
	CD	2×3	2×3^2	2×3^2	*	*	*	*
			3	3	2	3	2×3	2×3
		A	B	C	D	BC	CD	
		Conséquences						

Pour construire ce tableau, on place chaque action intervenant dans le modèle en ligne et en colonne avec son nombre de niveaux décomposé en nombres premiers (l'interaction BC à 6 niveaux a été décomposée en 2×3). Ensuite, une étoile (*) est placée à chaque intersection

d'actions jointes et le produit des nombres de niveaux à chaque intersection entre deux actions est calculé ; le produit étant noté sous la forme de sa décomposition en nombres premiers.

Ici, le plus petit plan orthogonal que l'on puisse trouver est un plan comportant un nombre d'expériences égal au PPCM de 9, 6 et 18 ; soit un plan comportant 18 essais.

II.6.3.2 Condition sur le nombre de degrés de liberté

Pour identifier les deux coefficients a et b d'une droite d'équation $Y=ax+b$, il faut au moins deux essais (deux points). D'une façon générale, pour être capable de calculer p valeurs indépendantes, il faut introduire dans les calculs au moins p valeurs indépendantes. Ce qui se traduit par un modèle passant par tous les points d'essais mais qui ne prend pas en compte la variance naturelle du procédé. Pour prendre en compte cette variance, on réalise un nombre d'essais N supérieur au nombre p de coefficients à estimer, et on suppose que la variance résiduelle, à $N-p$ degrés de liberté, est un estimateur de la variance σ^2 (cf. II.6.5.B). Ainsi, le nombre de degrés de liberté de la variance résiduelle doit-il être au moins égal à 1.

Remarque :

Si le modèle complet doit être étudié, ce qui implique le calcul de toutes les interactions, il faut nécessairement effectuer le plan complet.

II.6.4 PLANS FRACTIONNAIRES PARTICULIERS

Il existe d'autres plans permettant de découvrir les facteurs les plus influents sur une réponse donnée. Nous proposons dans la suite du paragraphe de donner quelques informations et indications sur quelques uns de ces autres plans. Il n'est pas possible d'être exhaustif car le nombre de plans qui existent est considérable. Chaque année, de nombreuses recherches menées dans le monde viennent enrichir la liste déjà longue des plans ou des stratégies existants.

II.6.4.1 Plans de TAGUCHI

G. Taguchi [Tag87] a mis au point une méthode originale pour construire des plans fractionnaires. Dans son approche, un problème expérimental doit être synthétisé sous la forme de graphes dans lesquels :

- Les facteurs sont représentés par des points d'aspects différents qui témoignent de la facilité avec laquelle les niveaux du facteur considéré peuvent être changés,
- Les interactions entre facteurs sont représentées par des traits continus liant les points associés aux facteurs.

Une fois le graphe établi, la construction du PE est quasi immédiate à partir de tables orthogonales prédéfinies [Pil97].

II.6.4.2 Approche TAGUCHI et plans produits

Parmi l'ensemble des travaux de Taguchi qui sont aujourd'hui largement diffusés, on trouve les plans produits qui sont très originaux et intéressants pour l'optimisation des produits et des processus. L'idée originale, à la base des plans produits, est de décomposer la variabilité du procédé ou du produit étudié en :

- la variabilité intrinsèque du système étudié,
- la variabilité due à des facteurs non contrôlables, ou que l'on ne peut pas contrôler au cours des essais, appelés facteurs bruits.

Le but de la méthodologie des plans produits, est de trouver une configuration des facteurs contrôlables robuste vis-à-vis des facteurs bruits afin d'augmenter la qualité du produit ou du procédé.

Les recherches de Taguchi occupent une place très importante parmi les différents travaux et approches qui ont été proposés pour améliorer la qualité des produits au stade de la conception. L'essentiel des apports de Taguchi doit pouvoir se résumer ainsi :

- Introduction de la notion de robustesse des produits et introduction de la séparation des facteurs bruits et des facteurs principaux.
- Prise en compte de la variabilité produite par les facteurs difficilement contrôlables (facteurs de bruit) sur les caractéristiques de performance d'un produit ou d'un processus.
- Recherche des conditions opératoires que l'on doit donner aux facteurs contrôlés de façon à minimiser les effets des facteurs bruit.
- Taguchi suppose que la variance peut être une fonction des variables d'entrée, du temps au contraire de l'approche traditionnelle qui considère les dispersions comme indépendantes, avec une variance constante.
- Taguchi utilise principalement les PE orthogonaux simples le plus souvent à deux ou trois niveaux.
- Utilisation de méthodes statistiques dans le domaine de la production (travail novateur pour les années 1950 – 1960). Mise en place d'outils en visant une relative simplicité d'utilisation (Anova, représentations graphiques par exemple).
- Les facteurs sont souvent confondus avec des interactions. La configuration optimale est obtenue en fonction des effets marginaux de chaque facteur.

En règle générale, les scientifiques reconnaissent l'intérêt et la pertinence de la méthode proposée par Taguchi. Pourtant, cette démarche, qui a fait ses preuves dans le milieu industriel, est parfois critiquée par certains scientifiques tels que [Box88] [Nair92] [Lochner90]. En résumé, les raisons essentielles de ces critiques sont les suivantes :

- Faible résolution des PE proposés par les graphes linéaires de Taguchi.
- Faible intérêt porté pour les interactions de second ordre ou d'ordre supérieur dans la démarche proposée.
- L'avantage procuré par la simplicité voulue par Taguchi pour son approche (importance de la notion d'orthogonalité en particulier vis-à-vis des inversions de matrice) perd peut-être un peu de sa force avec le développement des moyens de calcul informatiques (calculs plus aisés maintenant).
- Eventuellement, nombre important d'essais générés par les plans produits.

Néanmoins, toutes ces critiques s'adressent sans doute finalement plus aux utilisations qui sont faites de cette approche, qu'à l'approche elle-même.

II.6.4.3 Plans d'expériences optimaux pour modèles linéaires

La mise en œuvre de PE orthogonaux est contraignante et leur utilisation est impossible dans les cas suivants :

- Il existe des combinaisons de facteurs interdites (contraintes).
- On veut prendre en compte des essais déjà effectués.

Une alternative aux contraintes engendrées par l'emploi des plans orthogonaux réside dans l'utilisation de plans optimaux [Fed72] [Sch98]. Un plan d'expérience est optimal (au sens d'un critère) s'il conduit à la meilleure précision sur l'estimation des coefficients du modèle.

L'optimalité du PE est déterminée en calculant la variance sur les coefficients β du modèle. Nous présentons ci-dessous la démarche permettant d'aboutir à l'expression de la matrice de variance – covariance à minimiser.

Nous reprenons tout d'abord la notation matricielle du modèle linéaire définie au paragraphe II-5 :

$$y = X \cdot \beta + e \quad (\text{II. 13})$$

avec y le vecteur des N réponses, X la matrice des N essais, β le vecteur à p coefficients et e le vecteur des erreurs.

L'ajustement des coefficients du modèle correspond à l'estimation des coefficients par la méthode de moindres carrés. Soit $\hat{\beta}$ l'estimateur des p coefficients, nous obtenons :

$$\hat{\beta} = (X^t \cdot X)^{-1} X^t y \quad (\text{II. 14})$$

L'estimation des coefficients du modèle dépend donc de la matrice des essais X et des valeurs des réponses.

En faisant les hypothèses de Gauss-Markov (hypothèses tout à fait réalistes : facteurs indépendants et variance de la réponse constante sur le domaine d'étude) [Dro97], on montre que le vecteur ε vérifie :

$$\text{Esp}(\varepsilon) = 0 \quad \text{et} \quad \text{Var}(\varepsilon) = \text{Esp}(\varepsilon\varepsilon^t) = \sigma^2 I_N \quad (\text{II. 15})$$

En utilisant les expressions (II. 14) et (II. 13) nous obtenons :

$$\hat{\beta} = \beta + (X^t \cdot X)^{-1} X^t \varepsilon \quad (\text{II. 16})$$

En calculant l'estimateur des coefficients, nous remarquerons que $\hat{\beta}$ est un estimateur sans biais de β , ce qui signifie que l'espérance de l'estimateur est égale à la variable :

$$\text{Esp}(\hat{\beta}) = \text{Esp}(\beta) + (X^t \cdot X)^{-1} X^t \text{Esp}(\varepsilon) = \beta \quad (\text{II. 17})$$

Le calcul de la variance sur l'estimation des coefficients se fait en calculant la variance de :

$$\hat{\beta} - \beta = (X^t \cdot X)^{-1} X^t \varepsilon \quad (\text{II. 18})$$

Comme son nom l'indique, la covariance est une mesure de la force du lien entre deux variables aléatoires (numériques). La matrice de variance-covariance des coefficients s'exprime donc par :

$$\text{Cov}(\hat{\beta}) = (X^t \cdot X)^{-1} \sigma^2 \quad (\text{II. 19})$$

où les termes diagonaux représentent les variances sur les coefficients ($\text{Cov}(\beta_i, \beta_i) = \text{Var}(\beta_i)$) et les termes non diagonaux les covariances entre les coefficients :

$$[\text{Cov}(\beta_i, \beta_j)] = \begin{bmatrix} \text{Var}(\beta_1) & \dots & \dots & \text{Cov}(\beta_1, \beta_{k+1}) \\ \cdot & \text{Var}(\beta_2) & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \text{Cov}(\beta_{k+1}, \beta_1) & \cdot & \dots & \text{Var}(\beta_{k+1}) \end{bmatrix} \quad (\text{II. 20})$$

La technique des PE consiste à combiner les niveaux des facteurs de façon à minimiser les termes de la matrice produit inverse. C'est ce principe qui permet d'explorer un maximum de facteurs, avec un maximum de précision et un minimum d'essais [Sou94].

Plusieurs critères permettent de minimiser les termes de cette matrice. Il n'est pas question de dresser une liste exhaustive ici mais uniquement de citer les plus fréquemment utilisés : les critères de D-optimalité, de G-optimalité, de A-optimalité... [Fed72].

II.6.4.4 Autres plans

Il existe d'autres PE pouvant assurer la fonction de criblage. Cependant, dans la majorité des cas, ces alternatives supplémentaires ne possèdent pas les avantages des plans vus précédemment [Kos33] [Kou04] [Exp05] [Pil97] [Gou99].

II.6.5 OUTILS D'ANALYSE DES RESULTATS

Rappelons l'objectif de la technique de criblage : les valeurs des réponses doivent être analysées afin de mesurer l'influence des facteurs et des interactions sur la variation constatée de la réponse. La principale méthode statistique répondant à cet objectif est l'analyse de la variance.

L'analyse de la variance (en anglais : ANalysis Of Variance, ANOVA) nous permet de déterminer à partir de quel seuil un effet peut être considéré comme significatif. Pour ce faire, il faut comparer la variation d'un facteur pris en compte dans le modèle avec la variance résiduelle. Lorsque le facteur n'est pas pris en compte dans le modèle, l'effet du facteur est alors contenu dans la variance résiduelle. La variation de la variance résiduelle permet de déterminer si globalement le facteur a une influence significative sur la réponse, c'est-à-dire si au moins une des variables associées à chacun des niveaux des facteurs a un effet non nul. La contribution de la variable x_i à la variation de y se détermine de différentes manières :

A. Calcul de la variance des facteurs et interactions

La variance est une mesure de dispersion qui est la somme des carrés des écarts (SCE) des observations par rapport à leur moyenne, divisée par un nombre égal au nombre d'observations moins un (nombre de degrés de liberté ddl). La variance de l'échantillon est un estimateur sans biais de la variance de la population (NF ISO 3534-1) [Exp05].

« D'une façon générale, en matière de régression, le principe de l'analyse de la variance est de subdiviser la variation totale en une composante factorielle relative à l'équation de régression

ou au modèle utilisé, et en une composante résiduelle, la première devant être testée par rapport à la deuxième » [Dro97].

A titre d'illustration, supposons que nous disposions de N observations désignées par y_{ijk} ($i=1, \dots, p$; $j=1, \dots, q$; $k=1, \dots, n$). Les deux premiers indices concernent les deux facteurs contrôlés (avec respectivement p et q modalités, et donc un nombre total d'objets égal à $p \times q$). Le troisième indice concerne les répétitions (n répétitions de chacun des $p \times q$ objets). Dans ces conditions, nous pouvons calculer $p \times q$ moyennes \bar{y}_{ij} relatives aux différents objets, p moyennes $\bar{y}_{i..}$ relatives aux différentes modalités du premier facteur, q moyennes $\bar{y}_{.j.}$ relatives aux différentes modalités du deuxième facteur, et une moyenne générale $\bar{y}_{...}$.

Les éléments successifs du raisonnement sont alors :

→ **Le modèle observé**

$$y_{ijk} - \bar{y}_{...} = (\bar{y}_{i..} - \bar{y}_{...}) + (\bar{y}_{.j.} - \bar{y}_{...}) + (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) + (\bar{y}_{ijk} - \bar{y}_{ij.}) \quad (\text{II. 21})$$

Cette relation est le modèle observé de l'analyse de la variance. Elle indique que les écarts entre les observations individuelles et la moyenne générale peuvent être considérés comme constitués d'une part d'une composante relative aux différences entre les moyennes observées pour les différentes modalités et la moyenne générale, et d'autre part d'une composante relative aux écarts entre les observations initiales et observées pour les différentes modalités.

→ **Les nombres de degrés de liberté**

Le nombre de degrés de liberté associé à un facteur x_i est le nombre de niveaux N_{n_i} qu'il prend lors de la réalisation du plan, minoré de 1. Les nombres de degrés de liberté sont additifs :

$$N_{n_i} - 1 = pqn - 1 = (p - 1) + (q - 1) + (p - 1)(q - 1) + pq(n - 1) \quad (\text{II. 22})$$

→ **L'équation de l'analyse de la variance**

En élevant au carré les deux membres de la relation (II. 21) et en sommant pour toutes les observations, on obtient l'équation de l'analyse de la variance :

$$\begin{aligned} \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (y_{ijk} - \bar{y}_{...})^2 &= qn \sum_{i=1}^p (\bar{y}_{i..} - \bar{y}_{...})^2 + pn \sum_{j=1}^q (\bar{y}_{.j.} - \bar{y}_{...})^2 \\ &+ n \sum_{i=1}^p \sum_{j=1}^q (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^n (\bar{y}_{ijk} - \bar{y}_{ij.})^2 \end{aligned} \quad (\text{II. 23})$$

En utilisant l'indice a pour le premier facteur et l'indice b pour le deuxième facteur, l'équation (II. 23) s'écrit :

$$SCE_t = SCE_a + SCE_b + SCE_{ab} + SCE_r \quad (\text{II. 24})$$

Les quantités SCE_t , SCE_a , SCE_b , SCE_{ab} et SCE_r sont appelées respectivement somme des carrés des écarts totale, somme des carrés des écarts factorielle a et b, somme des carrés des écarts de l'interaction ab et somme de carrés des écarts résiduelle.

Remarque :

Dans le cas des interactions d'ordre supérieur, le calcul se fait de façon analogue.

Enfin, en divisant les sommes des carrés des écarts par leurs nombres de degrés de liberté, on définit les carrés moyens, à savoir un carré moyen total, un carré moyen factoriel, un carré moyen relatif à l'interaction et un carré moyen résiduel.

$$CM_t = \frac{SCE_t}{(pqn - 1)}; \quad CM_a = \frac{SCE_a}{(p - 1)}; \quad CM_{ab} = \frac{SCE_{ab}}{(p - 1)(q - 1)}; \quad CM_r = \frac{SCE_r}{pq(n - 1)} \quad (\text{II. 25})$$

Ces quantités sont des mesures, respectivement, de la dispersion de l'ensemble des observations (variation totale), de l'importance des différences existant entre les moyennes relatives aux différentes modalités (variation factorielle), et de l'importance des variations apparaissant à l'intérieur des différentes modalités (variation résiduelle).

B. Calcul de la variance résiduelle

La variance résiduelle est directement liée aux écarts entre les réponses mesurées et les réponses estimées. Elle permet de juger de la qualité d'un modèle linéaire indépendamment du nombre d'essais réalisés. L'analyse de la variance permet de déterminer l'aptitude du modèle à décrire les variations de la réponse en comparant les variations expliquées par le modèle et les variations inexpliquées de la réponse :

- La variation expliquée de la réponse correspond à la variance induite par les facteurs et les interactions.
 - La variation inexpliquée de la réponse correspond à la variation résiduelle.
- La variance résiduelle σ^2 peut être décomposée en deux sources de variabilité :
- la variabilité intrinsèque du système étudié,
 - la variabilité due à des facteurs aléatoires non contrôlés au cours des essais.

La norme ISO 5725 [Sou94] définit deux estimateurs de la variance σ^2 :

- l'erreur de répétabilité ; c'est à dire la dispersion des résultats obtenus avec la même méthode de mesure, le même observateur, les mêmes instruments de mesure, le même lieu, les mêmes conditions d'utilisations, en effectuant les répétitions sur une courte période de temps,
- l'erreur de reproductibilité qui est la dispersion des résultats observée en faisant varier à chaque répétition les méthodes de mesure, les observateurs, les instruments de mesure, les conditions d'utilisation, dans différents lieux et en échelonnant les mesures dans le temps, les répétitions.

➔ L'hypothèse nulle

L'hypothèse H_0 , appelée hypothèse nulle, consiste à vérifier l'absence d'influence du facteur étudié.

➔ Test de Fisher-Snedecor

Le test Fisher-Snedecor est un test qui permet de comparer deux variances. Il est donc parfaitement adapté à notre problème puisque l'objectif est de comparer les deux variances expliquée et inexpliquée par le modèle. Cette loi a été tabulée par M. Snedecor qui lui a donné le nom de loi de Fisher-Snedecor en l'honneur du statisticien Fisher. Pour conclure que l'effet d'un facteur a, par exemple, est significatif, nous cherchons à montrer que CM_a est supérieure à CM_r .

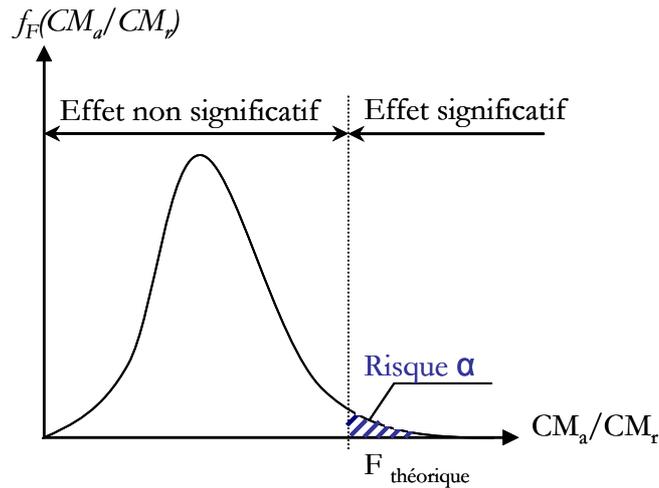


Figure II- 12 : Représentation de la loi de Fisher-Snedecor [Pil97]

La figure II- 12 nous montre que dans le cadre de l'hypothèse nulle (où le facteur n'est pas influent), le rapport peut être égal à l'infini. Cependant, la probabilité est très faible. Pour conclure sur un effet, il faut donc prendre le risque de conclure que cet effet est significatif alors que nous sommes dans le cadre de l'hypothèse nulle. Nous le noterons risque α . Il est généralement fixé à 5%. En fonction du risque choisi, la loi de Fisher-Snedecor déterminera un seuil de refus de l'hypothèse nulle que l'on appellera $F_{théorique}$.

Le test consiste donc à comparer le rapport entre CM_a et CM_r noté (F_{obs}) avec la valeur $F_{théorique}$ que nous trouvons dans le tableau lié à la loi de Fisher-Snedecor, en fonction :

- de ddl_a et ddl_r degrés de liberté du facteur étudié et de la résiduelle ;
- du risque que l'on choisit.

➔ Probabilité

La Probabilité ou « valeur p » est le fractile $F_{ddl_a,ddl_r,1-\alpha}$ de la loi de Fisher Snedecor centrée associée à une probabilité $1-\alpha$ de dépassement. Elle peut être utilisée comme repère de la confiance qu'on peut avoir dans un résultat particulier. Beaucoup de chercheurs utilisent une « valeur p » de moins de 0.05 comme limite de significativité statistique, ce qui revient à dire que le résultat observé dans une étude peut se produire par hasard moins d'une fois en vingt études différentes. La « valeur p » peut seulement prendre des valeurs comprises entre 0 et 1. Si elle est inférieure à 0.05, on conclut que l'effet est significatif et si elle est inférieure à 0.01, il est possible de conclure que le facteur est hautement significatif.

$$\text{prob}\{F_{obs} > f_{ddl_a,ddl_r;1-\alpha}\} = 1 - \alpha \quad (\text{II. 26})$$

C. Tableau d'analyse de la variance

Les résultats de l'analyse de la variance des PE sont généralement présentés sous la forme du tableau suivant :

Tableau II- 6 : Présentation générale d'un tableau d'analyse de la variance

Source de variation		ddl.	Sommes des carrés des écarts	Carrés Moyens	F_{obs}	$F_{théo}$	p-valeur
Facteurs	a	p-1	SCE_a	CM_a	CM_a/CM_r	$F_{(p-1; pq(n-1))}$	$F_{ddl_a, ddl_r, 1-\alpha}$

	b	q-1	SCE_b	CM_b	CM_b/CM_r	$F_{(q-1; pq(n-1))}$	$F_{ddlb, ddlr, 1-\alpha}$

Interactions
	ab	$(p-1) \times (q-1)$	SCE_{ab}	CM_{ab}	CM_{ab}/CM_r	$F_{((p-1)(q-1); pq(n-1))}$	$F_{ddlab, ddlr, 1-\alpha}$

Erreur		$pq(n-1)$	SCE_r	CM_r			
Total		$pqn-1$	SCE_t				

Remarque :

Cette méthode d'analyse est étroitement liée à la future qualité du modèle expérimental. En effet, si un facteur ou une interaction est oublié au cours de la formalisation du problème expérimental, son influence tend à faire augmenter la valeur de la variance résiduelle. Ainsi, des facteurs d'influence significative deviennent d'influence non significative.

II.7 METHODOLOGIE DES SURFACES DE REPONSE

La MSR constitue le second volet intéressant de la méthode des PE. Cette technique vise à déterminer d'une façon quantitative les variations de la fonction réponse vis-à-vis des facteurs d'influence significative [Gou99] [Box78] [Viv02].

Dans cette méthodologie, les modélisations de la fonction réponse peuvent également servir de base à la recherche de conditions optimales. On parle alors d'*optimisation indirecte*. L'optimisation d'une réponse ou la recherche d'un compromis entre plusieurs réponses consiste à définir, au sein du domaine expérimental, un réglage des facteurs permettant de satisfaire au mieux les exigences énoncées en termes de réponse.

La modélisation d'une surface de réponse s'appuie sur l'analyse de la variation des résultats d'essais obtenus suite à un PE. La surface de régression ainsi obtenue doit posséder une qualité descriptive et prédictive devant être la meilleure possible au sein du domaine expérimental.

Afin d'illustrer les propos théoriques figurant dans cette partie, une application de cette méthode à l'analyse d'un essai de vieillissement de PaC est faite au chapitre IV.

II.7.1 ESTIMATION DES COEFFICIENTS DES MODELES POLYNOMIAUX

Le but est la recherche de l'expression d'un modèle polynomial décrivant les variations de la fonction réponse relativement aux valeurs des k facteurs.

Le problème posé est différent s'il s'agit de mener des expériences réelles ou des expériences virtuelles [Viv02].

- Lors des expériences réelles, le vecteur y de réponse est connu puisqu'il contient les réponses expérimentales. Le vecteur des coefficients β est indéterminé et doit être estimé. Il peut exister des erreurs de mesure (cf II. 7. 2. 1) pour chaque valeur d'erreur.
- Lors des expériences virtuelles, il ne peut exister d'erreur de mesure. Le vecteur d'erreur est donc nul. En effet, procéder deux fois de suite à la même simulation débouchera sur l'obtention de deux valeurs de résultats identiques.

Cependant, avec la condition liée aux degrés de liberté $N \geq p$ (avec N le nombre d'expériences et p le nombre de coefficients du modèle), et pour les situations les plus courantes $N > p$ ($N = p$ n'est qu'un cas particulier, celui du plan saturé), les modèles utilisés ne passent plus exactement par les points d'expériences. Les coefficients de tels polynômes modélisateurs sont alors calculés (estimés) afin de minimiser un critère donné tel que celui des moindres carrés.

$$y = X \cdot \hat{\beta} + \varepsilon \quad (\text{II. 3})$$

L'estimation $\hat{\beta}$ du vecteur β doit être réalisée de telle façon que l'erreur de modélisation soit minimale. Le critère des moindres carrés traduit cette exigence par un objectif équivalent : par minimisation des carrés des écarts ou encore, en supposant la normalité de la valeur d'erreur ε ($N(0, \sigma^2)$).

$$\begin{aligned} \|\varepsilon\|^2 = \|y - X\hat{\beta}\|^2 \quad \text{minimal} \\ \text{tel que :} \quad \frac{\partial \varepsilon^t \varepsilon}{\partial \hat{\beta}} = 0 \end{aligned} \quad (\text{II. 27})$$

L'expression à minimiser sur β s'écrit :

$$\|y - X\hat{\beta}\|^2 = (y - X\hat{\beta})^t (y - X\hat{\beta}) = y^t y - 2\hat{\beta}^t X^t y + \hat{\beta}^t X^t X \hat{\beta}. \quad (\text{II. 28})$$

Par dérivation matricielle de la dernière équation, on obtient les "équations normales" dont la solution correspond bien à un minimum :

$$X^t y - X^t X \hat{\beta} = 0 \quad (\text{II. 29})$$

Nous faisons l'hypothèse supplémentaire que la matrice $X^t X$ est inversible. Alors, l'estimation des paramètres β est donnée par :

$$\hat{\beta} = (X^t X)^{-1} X^t y \quad (\text{II. 30})$$

On remarquera que $\hat{\beta}$ constitue une généralisation du concept d'effet vu lors de la présentation de la technique de criblage. Et les valeurs ajustées (ou estimées, prédites) de y ont pour expression :

$$\hat{y} = X\hat{\beta} = X(X^t X)^{-1} X^t y \quad (\text{II. 31})$$

On note ε le vecteur des résidus :

$$\varepsilon = y - \hat{y} = y - X\hat{\beta} = \left(I - X(X^t X)^{-1} X^t \right) y \quad (\text{II. 32})$$

Les estimateurs $\hat{\beta}$ sont des estimateurs sans biais : $\text{Esp}(\hat{\beta}) = \beta$, et ils sont de variance minimale (théorème de Gauss Markov).

II.7.2 QUALITE ET VALIDITE DU MODELE

Les valeurs expérimentales introduites dans le modèle sont entachées d'erreurs expérimentales (grandeurs aléatoires) qui se transmettent aux coefficients $\hat{\beta}$ du modèle ($\hat{\beta} = (X^t X)^{-1} X^t y$), puis aux valeurs calculées (\hat{y}) à l'aide de celui-ci ($\hat{y} = X\hat{\beta}$). Des tests statistiques permettent d'évaluer la qualité du modèle et la significativité des coefficients.

II.7.2.1 Les différents écarts

La figure II- 13 permet de distinguer trois types d'écarts différents.

- Le premier est l'écart entre une réponse mesurée et la moyenne des réponses mesurées au même point expérimental (avec $y_{x,1}$ et $y_{x,2}$ deux réponses mesurées dont la moyenne est \bar{y}_x). Il s'agit donc de l'erreur expérimentale, notée σ_y .

$$\sigma_{y,1} = y_{i,1} - \bar{y}_i$$

$$\sigma_{y,2} = y_{i,2} - \bar{y}_i$$

- Le second écart est celui que l'on constate entre la moyenne des réponses et la réponse prédite. Cette différence caractérise le manque d'ajustement. Elle est notée Δ_i :

$$\Delta_i = \bar{y}_i - \hat{y}_i$$

- Enfin, le troisième écart est celui qui existe entre chacune des réponses mesurées et prédites. Il s'agit donc de la somme des deux écarts précédents : liés au manque d'ajustement et à l'erreur expérimentale. Cet écart est le résidu et il a été noté ε_i :

$$\varepsilon_i = \sigma_{y_i} + \Delta_i$$

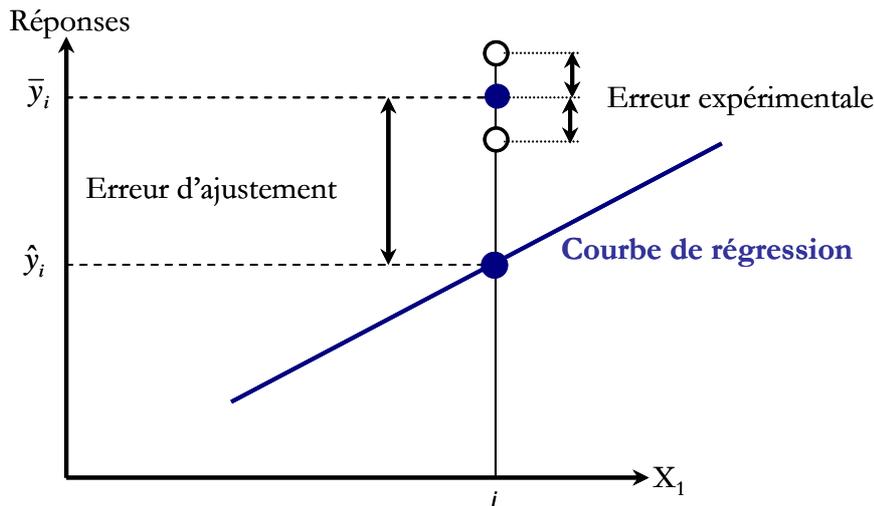


Figure II- 13 : Décomposition du résidu en deux écarts : écart expérimental et écart d'ajustement [Gou99]

II.7.2.2 Coefficients d'ajustement de modèles

Bien qu'à ce stade nous ayons des estimations des coefficients du modèle, nous ne pouvons pas les utiliser car nous ne savons pas si celui-ci représente la réponse expérimentale étudiée dans le domaine expérimental d'intérêt. Il est donc primordial de valider et de connaître la qualité de la modélisation obtenue. Nous pouvons envisager deux situations :

- Le nombre d'expériences distinctes est supérieur au nombre de coefficients du modèle ($N > p$). Nous pouvons utiliser un outil statistique comme l'analyse de variance, procurant des coefficients de détermination qui nous permettront de rejeter ou de ne pas rejeter le modèle en considérant un risque acceptable.
- Le nombre d'expériences distinctes est égal au nombre de coefficients du modèle. Dans ce cas, aucun outil statistique ne pourra nous aider. Il faudra accepter le modèle et le vérifier en faisant quelques expériences dans le domaine d'intérêt [Dro97].

➔ Analyse de la variance

La variance du modèle σ^2 n'est généralement pas connue mais nous pouvons en calculer une estimation. Posons :

- SCT : la Somme des Carrés Totaux ; $SCT = SCE + SCR$ et $SCR = y^t y$
- SCE : la Somme des Carrés des Ecart ; $SCE = (y - X\hat{\beta})^t (y - X\hat{\beta})$
- SCR : la Somme des Carrés due à la Régression ; $SCR = \hat{y}^t \hat{y}$

Nous déduisons les carrés moyens en divisant les sommes de carrés ainsi définies par les degrés de liberté correspondants.

- Ainsi, les carrés moyens CMR dus à la régression sont : $CMR = \frac{SCR}{p}$
- Et les carrés moyens CME associés aux écarts valent : $CME = \frac{SCE}{N - p}$
- CME est une estimation non biaisée de la variance expérimentale σ^2

Le nombre de degrés de liberté total est $(N-p) + p = N$, qui correspond au nombre d'expériences.

On effectue alors le test de Fisher-Snedecor. F_{obs} est une valeur observée d'une variable F de Fisher-Snedecor, à p et $(N-p)$ degrés de liberté. On calcule le ratio : $F_{obs} = \frac{CMR}{CME}$

On utilise ensuite le tableau d'analyse de la variance pour réunir ces informations.

Tableau II- 7 : Tableau d'analyse de la variance

Source de variation	d.d.l.	Sommes des carrés	Carrés Moyens	F_{obs}	$F_{théo}$
Régression (modèle)	p	SCR	CMR	CMR/CME	$F_{(p;N-p)}$
Résidus	$N-p$	SCE	CME		
Total	N	SCT			

En pratique, le modèle utilisé contient un terme constant β_0 , correspondant à la moyenne des réponses mesurées. Cette composante n'étant d'aucun intérêt dans l'analyse de la variance, elle est supprimée.

La somme des carrés due à la répression est décomposée en deux termes :

$$SCR = \hat{\beta}^t X^t y - N\bar{y}^2 + N\bar{y}^2 = SCR_0 + SCR_{(-0)}$$

où : $SCR_0 = N\bar{y}^2$ la somme des carrés due au terme constant β_0 ; ce terme prend 1ddl
 $SCR_{(-0)} = \hat{\beta}^t X^t y - N\bar{y}^2$ la somme des carrés due aux autres coefficients ; ce terme prend $p-1$ ddl

Le tableau d'analyse de la variance devient :

Tableau II- 8 : Tableau final d'analyse de la variance

Source de variation	d.d.l.	Sommes des carrés	Carrés Moyens	F_{obs}	$F_{théo}$
Régression (modèle)	$p-1$	$SCR_{(-0)}$	$CMR_{(-0)}$	$CMR_{(-0)}/CME$	$F_{(p;N-p)}$
Résidus	$N-p$	SCE	CME		
Total	$N-1$	SCT			

→ **Coefficients de détermination (R^2 , R^2 ajusté)**

Le coefficient de détermination R^2 est à la fois la fraction des variations de la réponse expliquée par le modèle seul et un indice de la qualité de la régression :

$$R^2 = \frac{\hat{y}^t \cdot \hat{y} - \bar{y}^t \cdot \bar{y}}{y^t \cdot y - \bar{y}^t \cdot \bar{y}} \quad (II. 33)$$

Le coefficient R^2 peut s'interpréter comme le quotient de la variance expliquée par la variance des réponses mesurées. D'après cette formule (II. 33), on voit que le rapport R^2 varie entre 0 et 1. Une valeur proche de 1 correspond à un modèle avec un très bon pouvoir prédictif.

Le coefficient de détermination ajusté R_a^2 est défini de façon analogue, comme étant la fraction des variations de la réponse expliquée par le modèle seul, relativement aux degrés de liberté correspondants :

$$R_a^2 = \frac{\frac{\varepsilon^t \cdot \varepsilon}{\text{ddl}_r}}{\frac{y^t \cdot y - \bar{y}^t \cdot \bar{y}}{\text{ddl}_m}} \quad (\text{II. 34})$$

avec ddl_r et ddl_m correspondant respectivement au degré de liberté des résidus et au degré de liberté de la somme de carrés des réponses mesurées corrigée de la moyenne ($p-1$).

Le coefficient R_a^2 est tout à fait similaire au R^2 . Ses valeurs s'interprètent de la même manière. Il peut cependant prendre des valeurs négatives si le R^2 est proche de 0.

Du fait de la prise en compte des degrés de liberté, on a toujours $R_a^2 \leq R^2$.

➔ Analyse des résidus

Il s'agit d'une représentation graphique des résidus. En fonction de l'allure des graphes obtenus, il est possible de se rendre compte visuellement s'il reste encore de l'information à extraire de l'ensemble des résidus. On regarde ainsi si les résidus semblent être distribués aléatoirement ou non.

Remarque :

De nombreux critères de choix de modèle sont présentés dans la littérature portant sur la régression linéaire multiple. Le choix du critère est déterminant lorsqu'il s'agit de comparer des modèles de niveaux différents.

II.7.3 DETERMINATION DES MEILLEURS SOUS MODELES

Un sous-modèle est déduit d'un modèle complet polynomial par suppression d'au moins un monôme, exception faite de la composante constante β_0 . Pour un modèle complet à p coefficients, on compte donc $(p-1)!$ sous-modèles.

Différentes stratégies sont donc possibles en fonction de l'objectif recherché et des moyens de calculs disponibles.

La méthode de simplification « pas à pas » en est une. A chaque pas, en partant du modèle complet, les coefficients d'ajustement sont calculés et classés dans un ordre décroissant de qualité.

Cela permet de déduire les facteurs principaux et les interactions correspondants, et d'éliminer ceux ayant le moins d'influence sur la perte de qualité du modèle initial. Ceux-ci peuvent être considérés comme peu influents vis-à-vis de la réponse puisque leur absence n'affecte que de manière limitée la qualité et l'aspect prédictif de la modélisation [Gou99].

Remarque :

De manière générale, il est souhaitable de ne pas réaliser d'analyse de MSR et/ou d'optimisation pour un nombre de facteurs supérieur à 5. L'étude préalable de criblage doit permettre de sélectionner et de rejeter les facteurs non influents.

II.8 CONCLUSION DU CHAPITRE II

Nous avons présenté dans ce chapitre des éléments bibliographiques essentiels relatifs à la Méthodologie des Plans d'Expériences (MPE). En tout premier lieu, les différents types de variables ont été recensés et présentés. Nous avons aussi donné la définition du domaine d'étude et du domaine d'étude possible soumis soit à des contraintes en position soit à des contraintes en valeurs atteintes. Le deuxième volet de ce chapitre a porté sur un aspect fondamental de la MPE : la modélisation, avec les principales techniques mathématiques utilisées pour exprimer les variations des réponses en fonction des valeurs des facteurs. Les deux principales composantes de la MPE, la technique de criblage et la méthodologie des surfaces de réponses, ont ensuite été décrites.

Les PE constituent avant tout un outil de gestion et d'organisation d'une campagne expérimentale dont le but consiste à apporter des éléments d'information, facilement interprétables. Pour ce faire, la définition des expériences à réaliser doit permettre de construire un modèle d'exploration du domaine expérimental, ce modèle étant l'élément clef de la stratégie. Il existe naturellement plusieurs familles de modèles, en étroite adéquation avec les objectifs des expérimentateurs.

La méthode des PE est constituée d'outils sûrs, pratiques permettant de conduire avec la meilleure efficacité possible une étude où interviennent de nombreux paramètres. Bien utilisée, elle doit conduire par exemple dans des phases d'élaboration de produits à des délais de conception et de production limités, à des coûts moindres, à une augmentation de la précision dans l'obtention des résultats, à une amélioration de la fiabilité. Ces avantages reposent sur le fait que la méthode s'appuie sur des fondements mathématiques rigoureux.

La démarche utilisée pour mener à bien et dans les meilleures conditions toute expérimentation peut être résumée dans un schéma de synthèse (Figure II- 14) qui reprend les trois grandes parties de l'expérimentation [Gou88] [Exp05] :

- le choix d'une méthode d'expérimentation,
- l'analyse des résultats,
- l'acquisition progressive des résultats.

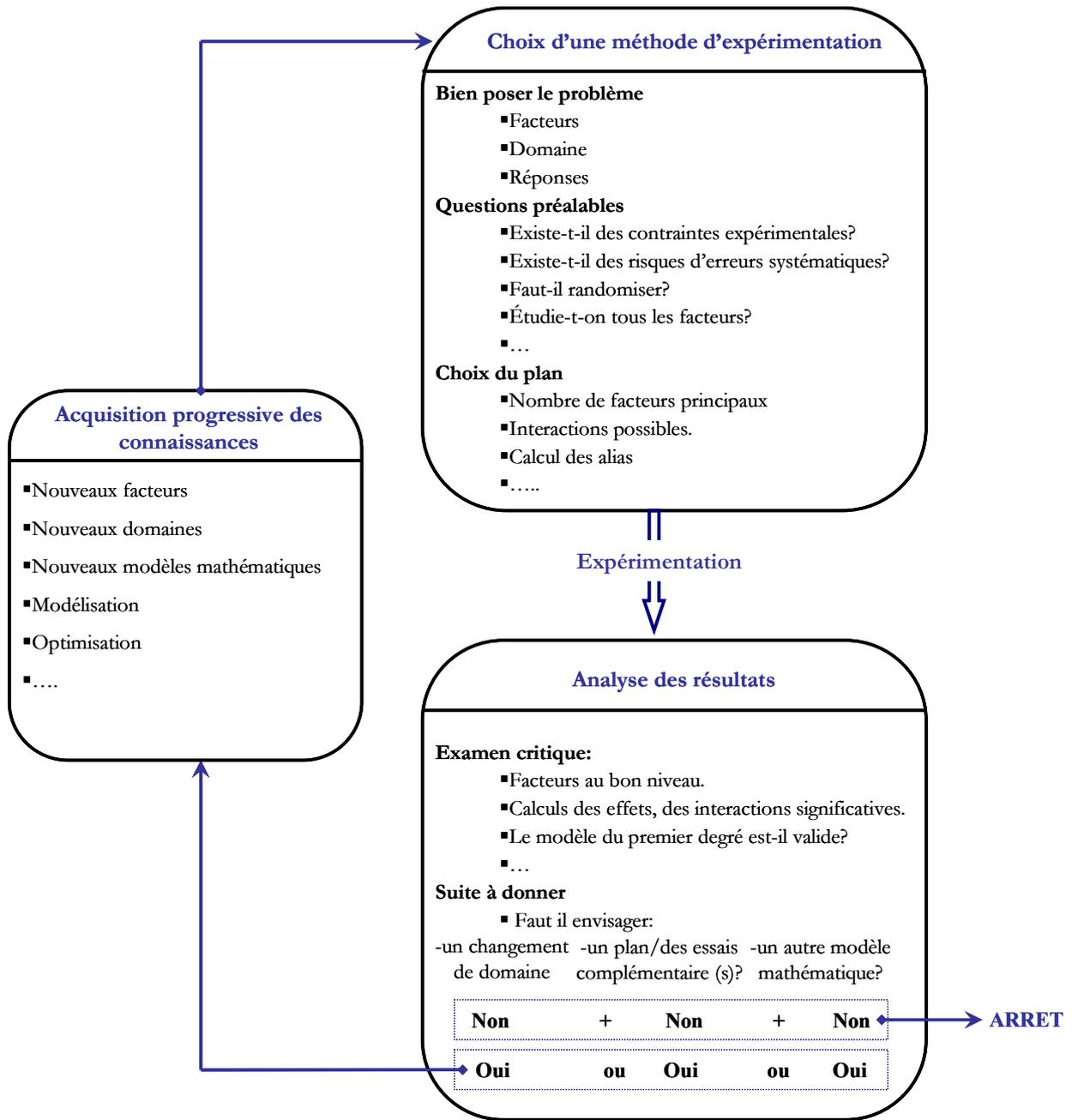


Figure II- 14 : Schéma général de conduite d'une étude selon la méthodologie de l'expérimentique [Gou88]