

Les plans d'expériences sont très utilisés dans l'ordonnancement et l'analyse des essais. Ils permettent de déterminer les liens existant entre les paramètres d'entrée et la sortie d'un système. Plusieurs travaux dans la littérature ont utilisé les plans d'expériences pour étudier l'influence des paramètres procédé [55-60].

- Verran et al. [55] ont analysé la coulée sous pression d'une pièce automobile (bloc moteur à deux cylindres) en AlSi12Cu1,3 cherchant « la meilleure valeur de densité possible » en utilisant les plans d'expériences et la simulation numérique du procédé HPDC. Dans cette étude la vitesse 1^{er} phase (V1), la vitesse 2^{ème} phase (V2) et la pression 3^{ème} phase (P3) ont été variées dans l'objectif d'optimiser la teneur en porosités dans l'alliage. Les résultats des dix-huit essais réalisés avec différentes combinaisons des trois paramètres d'injection (V1 : 0,14 m/s ; 0,27 m/s ; 0,9 m/s, V2 : 1,3m/s ; 2,6 m/s et P3 : 15 MPa ; 22 MPa ; 30 MPa) ont permis d'évaluer l'influence de ces paramètres sur l'apparition des porosités de fonderie. Compte tenu des résultats expérimentaux obtenus ; des simulations numériques en fonderie sous pression ont été réalisées en considérant les paramètres d'injection qui ont donné la plus grande teneur en porosités (densité : 2,639 g/cm³) pour V1=0,27m/s, V2= 2,6 m/s, P3= 15 MPa et la plus petite teneur en porosités (densité : 2,7 g/cm³pour V1=0,14m/s, V2= 1,3 m/s, P3= 30 MPa). La comparaison des résultats expérimentaux et numériques a montré que la prédiction de la porosité par la simulation corrèle bien avec les résultats expérimentaux.
- Kim et al. [56] ont utilisé la méthode des plans d'expériences dans le procédé de fabrication des écrans en verre. Des modèles de régression ont été établis à partir d'un plan factoriel complet à quatre facteurs pour l'identification des effets des paramètres procédé et l'estimation des paramètres de forme de ces écrans ; les interactions entre les paramètres procédé ont été aussi étudiés et la technique de l'analyse de la variance a été utilisée pour la vérification de la robustesse des modèles de régression.

Cette partie a pour finalité d'introduire la méthode de plan d'expériences pour bien ordonnancer les essais qui seront mis en œuvre durant la thèse pour pouvoir les analyser par la suite.

La méthode des plans d'expériences

De manière générale, la méthode de plan d'expériences cherchera à établir les liens existant entre deux types de variables (figure 24) :



Figure 24 : Plan d'expériences

- Les facteurs (variables d'entrée) : grandeurs physiques modifiables ou non par l'expérimentateur, susceptibles d'influencer sur les variations de la réponse. On distingue :
 - Les facteurs contrôlables, qui dépendent directement du choix de l'expérimentateur (pression, température, matériau, etc.)
 - Les facteurs non contrôlables (conditions climatiques, environnement d'utilisation...)

Les facteurs d'entrée, dont on cherche à analyser l'influence font partie des facteurs contrôlables et non contrôlables.

- La réponse (variable de sortie) : grandeur physique étudiée, c'est une sortie du système. Elle est mesurée à chaque essai. Le plan vise à déterminer quels facteurs l'influencent ou quelle est son évolution en fonction de ceux-ci. Cette grandeur est qualitative ou quantitative.

Le plan doit permettre de bâtir un modèle, exprimant la réponse en fonction des facteurs d'entrée, à partir des résultats de séries d'expériences. L'utilisation du plan d'expériences peut se faire en deux étapes :

1)- La technique du screening

Cette technique permet de déterminer parmi les facteurs recensés ceux qui ont une influence statistiquement non négligeable sur les variations de la réponse ; le but est de chercher pourquoi la réponse varie (en fonction de quels facteurs).

2)- La méthodologie des surfaces de réponse

Les variations de la réponse sont calculées en fonction des facteurs précédemment jugés influents (par la technique de screening) ; le but est de déterminer comment la réponse varie.

Pour résumer, la Méthode de Plan d'Expériences (MPE) est donc un ensemble de techniques complémentaires aidant à choisir les expériences à réaliser ainsi qu'à comprendre, analyser et exploiter les résultats obtenus. Ces outils s'appuient essentiellement sur des bases statistiques et algébriques. Cette particularité induit la possibilité de connaître les erreurs concédées sur les données expérimentales et sur celles qui en sont déduites.

Une analyse par plan d'expériences est une démarche qui se fait en trois étapes :

- La formalisation du problème;
- L'étude qualitative du système (screening) ;
- L'étude quantitative du système (méthode de surfaces de réponse) ;

1. Formalisation du problème

L'étape de formalisation du problème commence par un brainstorming qui consiste à regrouper l'ensemble des acteurs. C'est une technique de créativité de groupes permettant de produire le plus d'idées possibles, dans un minimum de temps, sur un sujet donné. Cette étude est utilisée dans la plupart des étapes de résolution de problèmes, notamment pour recenser les questions suscitées lors du lancement d'un nouveau projet ou pour trouver les causes possibles d'une anomalie du fonctionnement d'un système. Il peut être utilisé aussi pour la formalisation et l'organisation d'un plan d'expériences.

La formalisation du problème permet de rassembler l'expérience accumulée, afin de la compléter et de la préciser grâce à l'expérimentation. En s'appuyant sur cette méthode, nous allons :

▪ **DEFINIR LE PROBLEME**

Pour définir le problème, il faut le poser clairement en précisant de quoi il s'agit, et en quoi le problème proposé est un problème. Il est recommandé de faire appel à la méthode du QQCOQP qui consiste à répondre aux questions :

Q : en Quoi consiste le problème ?

Q : Qui est gêné par le problème ? Qui est le demandeur de l'étude ?

C : Combien de cas ? Combien ça coûte ?

O : Où cela se passe-t-il ?

Q : Quand cela arrive-t-il ?

P : Pourquoi est-ce un problème ?

▪ **DETERMINER LES OBJECTIFS**

De manière générale, les deux objectifs principaux d'un plan d'expériences sont :

- Trier parmi un ensemble de variables d'entrée susceptibles d'influencer les variables de sortie et de les hiérarchiser.
- Estimer avec suffisamment de précision l'effet de l'ensemble des variables d'entrée influentes, de façon à pouvoir prévoir de manière assez fiable, la valeur de la variable de sortie que l'on obtiendrait pour l'ensemble des combinaisons des variables d'entrée.

▪ **DEFINIR LES REPONSES (SORTIES)**

La réponse correspond à un paramètre de sortie du système étudié. Une réponse doit être représentative, quantifiable et la moins dispersée possible pour des variables d'entrées maîtrisées et constantes.

Pour appliquer la méthodologie des plans d'expériences, il est préférable d'avoir une réponse sous forme quantitative. En effet, les méthodes d'analyses des résultats d'essais telles que l'analyse de variance ou l'analyse de régression au sens des moindres carrées, s'appuient sur des données quantitatives.

▪ **DEFINIR LES FACTEURS (ENTREES)**

Les paramètres d'entrées du système sont appelés facteur X. Un facteur est une cause possible de variation de la réponse Y ; le choix des facteurs se fait souvent en plusieurs étapes :

- Phase de recensement : cette phase consiste à recenser l'ensemble des entrées du système ;
- Phase de classement : consiste à classer les facteurs recensés par ordre d'influence sur les sorties ;
- Phase de sélection et de définition des modalités pour chaque facteur.

▪ **DEFINIR LE DOMAINE EXPERIMENTAL**

Le domaine expérimental peut être continu ou discret :

- Domaine expérimental continu : Un domaine expérimental continu est délimité par des facteurs quantitatifs. La figure 25 illustre le domaine d'étude de deux facteurs quantitatifs X1 et X2. Le facteur X1 varie dans $[X1_{\text{binf}}, X1_{\text{bsup}}]$ et le facteur X2 varie dans $[X2_{\text{binf}}, X2_{\text{bsup}}]$; le nombre total de combinaisons que l'on peut réaliser est infini dans ce domaine.

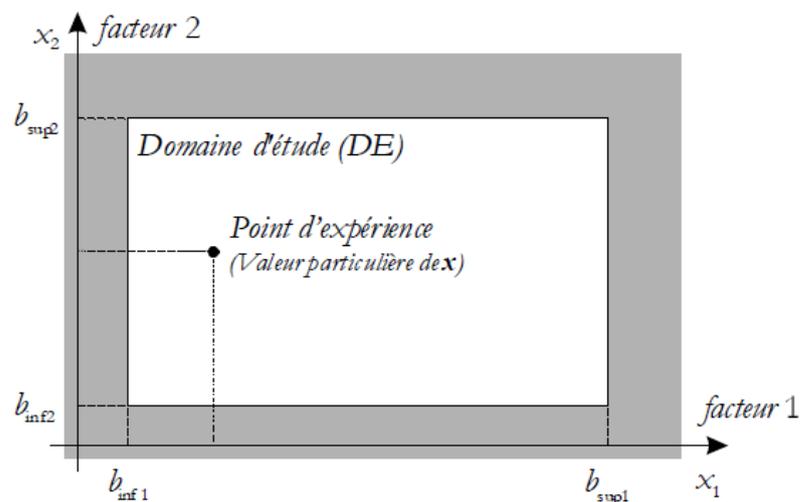


Figure 25: Représentation d'un domaine d'étude continu pour le cas de deux facteurs à deux modalités [61]

- Domaine expérimental discret : Ce type de domaine est délimité par des facteurs discrets ou qualitatifs. La figure 26 illustre le domaine expérimental de deux facteurs qualitatifs A et B ; les facteurs A et B prennent chacun, trois modalités A1, A2, A3 pour A et B1, B2, B3 pour B.

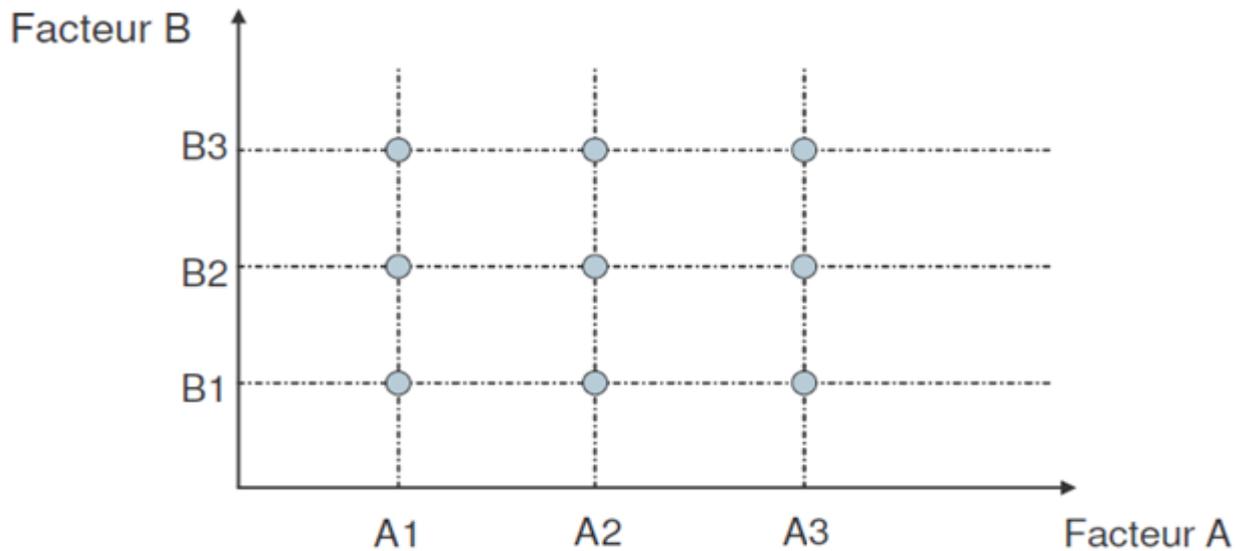


Figure 26: Représentation d'un domaine d'étude discret pour le cas de deux facteurs à trois modalités (domaine discret) [62]

▪ **DEFINIR LES CONTRAINTES**

Cette étape consiste à définir les contraintes en termes de temps, moyen ou d'infaisabilité de certaines configurations d'essais.

▪ **CHOISIR LA STRATEGIE EXPERIMENTALE**

La méthodologie des plans d'expériences consiste à construire un modèle qui apportera des éléments d'informations à l'expérimentateur encore appelé « modèle empirique ». Pour cela, le choix du plan d'expériences à mettre en œuvre dépend du problème posé. Dans cette thèse, les deux types de plans seront abordés, les plans de criblage et les plans de surface de réponse

- Les plans de criblage (ou de screening) : permettent de déterminer le poids des facteurs sur les réponses d'un système à partir d'un modèle.
- Les plans de surface de réponses : permettent à partir d'un modèle mathématique polynomial de déterminer les valeurs des facteurs influents correspondants à une réponse particulière du système.

▪ **CHOISIR LE MODELE**

La valeur de la réponse est recherchée pour l'ensemble des points constituant le domaine expérimental. Mais, pour des contraintes de coût et de temps, le nombre d'expériences nécessaires pour tout point du domaine expérimental n'est pas réalisable. C'est pourquoi nous allons utiliser un modèle empirique, qui nous permettra d'avoir le maximum d'informations en réalisant un minimum d'expériences. Ce modèle est de la forme générale : $f(\text{facteurs}) = \text{réponse}$.

Le tableau 4 suivant présente les modèles empiriques pour chacune des deux techniques du MPE, le modèle additif pour la technique de screening et le modèle polynomial pour la technique des surfaces de réponses :

Forme générale : $f(X)=Y$		
calculé	$Y = \sum_{i=1}^k W_i + \text{cte}$ $W_i : \text{poids du facteur } i$	$Y = \beta_0 + \sum_i \beta_i u_i + \sum_i \beta_{ii} u_i^2 + \sum_{ij} \beta_{ij} u_i u_j$ $\beta_0 : \text{moyenne arithmétique des réponses}$ $\beta_0, \beta_{ii}, \beta_{ij} : \text{coefficients du modèle}$
estimé	Etude des X : plan de criblage $Y = b_0 + \sum_i b_i x_i$ $b_0, b_i : \text{estimateurs des } \beta$	Etude des Y : plan à surface de réponses $Y = b_0 + \sum_i b_i x_i + \sum_i b_{ii} x_i^2 + \sum_{ij} b_{ij} x_i x_j$ $b_0, b_i, b_{ii}, b_{ij} : \text{estimateurs des } \beta$
	Modèle additif	Modèle polynomial

Tableau 4 : Modèles empiriques pour les deux techniques du MPE

▪ CHOISIR LE TYPE DE PLAN

- Plan factoriel complet

Le plan factoriel complet combine de manière exhaustive toutes les combinaisons possibles des variables avec un nombre de facteurs finis. Si on considère un nombre discret de variations par facteur, il permet d'envisager toutes les combinaisons possibles des modalités des facteurs. Le nombre d'essais N se calcule d'après la formule suivante $N=2^k$, avec k le nombre de facteurs. Les plans factoriels complets sont dits des plans sans risque car ils permettent de déterminer tous les effets et toutes les interactions. Le nombre d'essais nécessaire est au moins égal au nombre total de coefficients à déterminer. Les essais sont réalisés de telle sorte que les coefficients sont estimés avec une variance minimale. Leur simplicité d'exploitation assure un bon rendement par rapport aux résultats obtenus. Néanmoins, ils présentent une limite essentielle : le nombre d'essais augmente très rapidement avec le nombre de facteurs ce qui devient très vite difficile à réaliser dans la pratique.

- Plan factoriel fractionnaire

Le plan fractionnaire a été conçu pour remédier à l'inflation rapide du nombre d'essais dans les plans complets. L'objectif des plans fractionnaires consiste alors à réduire le nombre d'expériences à réaliser par rapport au nombre maximum donné par le plan complet. Les plans fractionnaires utilisent les matrices des effets des plans complets.

On parlera de plan 2^{k-p} (p un entier à choisir) pour indiquer un plan fractionnaire issu du plan complet 2^k avec k facteurs à 2 niveaux. Néanmoins les plans fractionnaires nécessitent une phase de conception plus longue car l'interprétation des résultats dépend essentiellement du choix de p. Plus le

nombre p augmente, plus la charge expérimentale va diminuer mais au détriment de la qualité des informations tirées du plan.

- Autres plans

Plan de Taguchi : Le plan de Taguchi a été développé dans une optique d'utilisation industrielle ; c'est un plan factoriel fractionnaire, prenant en compte certaines interactions jugées importantes et en abandonnant d'autres.

Plan de Plackett-Burman: Les matrices de Plackett-Burman ne sont autres que des matrices des plans factoriels fractionnaires de résolution III [62]. Leur construction a été simplifiée et systématisé par Plackett et Burman. Mais aucune interaction entre facteurs principaux n'est considérée pour ce type de plan.

Plan de Rechtschaffner : Les matrices proposées par Rechtschaffner (1967) [63] sont des matrices de résolution III qui permettent la quantification des effets principaux des facteurs b_i et des effets d'interactions entre les facteurs pris deux à deux b_{ij} , avec un faible nombre d'expériences (matrice saturée). Bien que ce nombre d'expériences soit minimal pour une matrice de résolution III, une augmentation du nombre de facteurs entraîne une forte augmentation du nombre d'expériences. Ainsi, pour étudier 5 facteurs, il faut réaliser 16 expériences, pour 9 facteurs, il faut 46 essais, etc.

2. Analyse de la variance

Les plans présentés plus haut donnent la définition des expériences dans le DE. Les valeurs des réponses en ces points doivent être analysées afin de mesurer l'influence des facteurs et des interactions sur les variations constatées de la réponse.

La principale méthode répondant à cet objectif est l'analyse de la variance. Cette méthode est appelée « Analysis of Variance » dans la littérature anglo-saxonne ; son appellation est couramment abrégée en ANOVA.

« 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 une composante résiduelle, la première devant être testée par rapport à la deuxième.» [62].

Les composantes factorielle et résiduelle seront mathématiquement représentées par des carrés moyens, c'est-à-dire des variances.

En définitive, l'intérêt de l'analyse de variance est de pouvoir tester de manière absolue l'influence des facteurs sur les variations d'une réponse donnée.

L'analyse de la variance est une généralisation de la comparaison de K populations (avec $K \geq 1$), cette dernière consiste à comparer les moyennes et les variances de K sous populations. Les hypothèses nulles et alternatives sont:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_K$$

H_1 : Au moins deux moyennes sont différentes

L'application de L'ANOVA suppose la vérification de deux critères :

- 1- Les sous populations sont normalement distribuées
- 2- Les variances dans les sous-groupes sont identiques

2.1 Technique de screening

La technique de screening est un procédé de sélection ou de criblage. Elle permet de déterminer, parmi un ensemble de facteurs, les éléments influents. Généralement, toute variation de facteur pris en compte induit une variation au niveau de la réponse mais à des échelles différentes.

On procède alors à la réalisation d'un test statistique permettant d'accepter ou non l'hypothèse selon laquelle un facteur induit des variations de la réponse significativement plus importantes que celles engendrées par des facteurs considérés comme constants. Un facteur est alors jugé influent si son action sur la réponse étudiée est statistiquement supérieure à un certain niveau.

Cette analyse permet donc de classer les facteurs entre eux, relativement à leur influence sur la réponse.

2.1.1 Carrés moyens des facteurs et des interactions

La variance des facteurs s'obtient en calculant la somme des carrés des écarts (SCE) que l'on divise par le nombre de degrés de liberté (ddl) associé au facteur f considéré. Le nombre de degrés de liberté ddl_f associé à un facteur f est le nombre de niveaux (de valeurs distinctes) qu'il prend lors de la réalisation du plan, minoré de 1. Dans tous les cas :

$$ddl_i = N_{n_i} - 1 \quad (1.2)$$

N_{n_i} : Nombre de niveaux du facteur i.

La somme des carrés des écarts associée au facteur f vaut :

$$SCE_f = y_f \cdot \sum_{i=1}^N (\bar{y}_i - \bar{y})^2 \quad (1.3)$$

Avec :

- N : le nombre total d'expériences ;
- $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$ la moyenne des réponses ;

- $y_f = \frac{N}{N_{n_i}}$ le nombre d'expériences pour lesquelles le facteur f prend un de ses N_{n_i} niveaux ;
- \bar{y}_i la moyenne des réponses observées pour les expériences où le facteur f prend son $i^{\text{ème}}$ niveau.

Le calcul des carrés des moyens pour les interactions se fait de la même manière.

On déduit alors la valeur des carrés moyens, associés au facteur ou à l'interaction considéré(e) x, comme étant :

$$CM_x = \frac{SCE_x}{ddl_x} \quad (1.4)$$

2.1.2 Variance résiduelle

Dans le cas des expériences réelles, la variance résiduelle est prise comme étant un estimateur de la variance expérimentale, qui traduit la variabilité inhérente des résultats sur plusieurs réalisations identiques. Cependant, l'utilisation d'expériences virtuelles exclut cette possibilité.

Dans le même cas, l'équation de variance doit toujours être vérifiée : la variance résiduelle (SCE_r) est dans tous les cas une composante de la somme des carrés des écarts totale (SCE_t).

La variance résiduelle est la somme des carrés des résidus, elle est calculée comme étant la somme des carrés des écarts entre réponses mesurées (y) et réponses calculées (y_{mod}) correspondantes [64,65] ; il s'agit donc de :

$$SCE_r = \sum_{i=1}^N (y(x^i) - y_{mod}(x^i))^2 \quad (1.5)$$

x^i est le vecteur des coordonnées du $i^{\text{ème}}$ point d'expériences du plan.

On calcule donc un carré moyen résiduel tel que :

$$CM_r = \frac{SCE_r}{ddl_r} = \frac{1}{N-p} \sum_{i=1}^N (y(x^i) - y_{mod}(x^i))^2 \quad (1.6)$$

Calculer SCE_r de cette manière permet de tester le caractère significatif des facteurs et des interactions et dans le même temps d'évaluer la qualité du modèle utilisé (y_{mod}).

Cette solution n'est pas applicable lors de l'utilisation de plans saturés (plans factoriels par exemple). Dans ces cas précis, certains auteurs Schimmerling et al. [66], Goupy [67] et Saporta [68] proposent la construction de la variance résiduelle à partir des interactions dont les variances (carrés moyens) sont les plus faibles. Leurs valeurs doivent être du même ordre de grandeur.

Le calcul de la variance résiduelle (ou carrés moyens résiduels) peut alors s'écrire comme :

$$CM_r = \frac{SCE_r}{ddl_r} = \frac{\sum_i SCE}{\sum_i ddl} \quad (1.7)$$

Les sommes des carrés des écarts (SCE) et les nombres de degrés de liberté (ddl) se rapportant aux interactions choisies.

Cette solution permet de retrouver la première écriture de l'équation de variance vue précédemment :

$$\begin{aligned} SCE_t &= \sum SCE_x \\ &= \sum SCE_{x^i} \end{aligned} \quad (1.8)$$

De manière générale, si une telle variance résiduelle ne peut être construite, il est inutile de calculer l'ANOVA : ses résultats seraient inexploitable.

2.1.3 Test de Fisher-Snedecor

Le test de Fisher-Snedecor permet d'utiliser la loi statistique dite de Fisher (ou loi F) pour comparer deux variances. Cette loi travaille sur un quotient de variances et prend en compte le nombre de degrés de liberté de chacune d'elles. Les variances concernées doivent être celles de variables aléatoires à distribution normale et à variances constantes.

On calcule alors pour chaque facteur (ou interaction) considéré (e), le ratio F_{obs} , tel que :

$$F_{obs} = \frac{CM_x}{CM_r} \quad (1.9)$$

La variance associée au facteur ou à l'interaction étudié(e) (CM_x) peut être considérée comme égale à la variance résiduelle (CM_r) si le rapport F_{obs} est faible, i.e. inférieur à une valeur seuil statistique. On définit ainsi l'hypothèse statistique H_0 , selon laquelle l'affirmation précédente est vraie. Si c'est le cas, F_{obs} est alors une valeur observée d'une variable F de Fisher-Snedecor, à ddl_f et ddl_r degrés de liberté.

L'hypothèse H_0 doit être rejetée au niveau α , lorsque :

$$P(F \geq F_{obs}) \leq \alpha \quad (1.10)$$

Ou, de manière équivalente, quand :

$$F_{obs} \geq F_{1-\alpha} \quad (1.11)$$

Avec $F_{1-\alpha}$ la fonction de répartition, pour une probabilité cumulée égale à $1-\alpha$. Ainsi, par exemple :

Si l'on obtient $F_{obs} = 62$ pour $ddl_f = 4$ et $ddl_r = 5$, on trouve une probabilité inférieure à 0.0002 pour que la variance du facteur soit du même ordre que la variance résiduelle et le facteur est significatif au niveau de 99.98%.

Il est courant d'utiliser un tableau réunissant les résultats des calculs précédents. Il peut prendre la forme suivante (tableau 5):

Sources de variation	ddl	Somme des carrés des écarts	Carrés moyens	F_{obs}	Probabilité P	Source influente ?
Facteur 1	ddl_1	SCE_1	$CM_1 = SCE_1 / ddl_1$	CM_1 / CM_r	$P(F \geq F_{obs})$	$P < \alpha ?$
...		
Facteur f	ddl_f	SCE_f	$CM_f = SCE_f / ddl_f$	CM_f / CM_r		
...		
Facteur k	ddl_k	SCE_k	$CM_k = SCE_k / ddl_k$	CM_k / CM_r		
Interaction fg	ddl_{fg}	SCE_{fg}	$CM_{fg} = SCE_{fg} / ddl_{fg}$	CM_{fg} / CM_r		

...		
Variation résiduelle	ddl _r	SCE _r	CM _r =SCE _r /ddl _r			
Totaux	ddl _t	SCE _t				

Tableau 5:Tableau d'analyse de variance de screening [63]

2.1.4 Rejet de facteurs

L'analyse de la variance évalue la probabilité que les variances des termes à évaluer soient significativement différentes de la variance résiduelle.

Ces résultats permettent donc de déterminer les facteurs et les interactions dont les probabilités sont inférieures à un niveau fixé arbitrairement. Selon ce niveau de signification, ces termes peuvent être rejetés, c'est-à-dire supprimés de l'étude et ne pas être considérés comme facteurs influents.

Cette opération est importante, car en diminuant le nombre de dimensions du problème, elle autorise et favorise l'utilisation de démarches coûteuses et généralement dépendantes du nombre de facteurs : il s'agit principalement de l'analyse de RSM et des optimisations par plans d'expériences.

Le test de Fisher opéré dans l'analyse de la variance (ANOVA) étant réalisé en prenant la variance résiduelle comme référence, ses résultats en dépendent. Ainsi, le nombre de facteurs rejetés peut être différent suivant le modèle calculé, et par voie de conséquence suivant le plan de screening réalisé.

2.2 Méthode de surface de réponse

La méthode de surface de réponse (MSR) est le second volet de la méthode des plans d'expériences. Elle vient ainsi s'appliquer à la suite de la technique de screening. Elle n'utilise que les facteurs précédemment jugés influents. Elle est utilisée aussi sous le sigle d'origine anglo-saxon RSM (Response Surface Methodology).

La technique MSR vise à déterminer d'une façon quantitative les variations de la fonction réponse vis-à-vis des facteurs d'influence significative [69,70].

2.2.1. Modèle à régression polynomiale

La modélisation par surface de réponses peut se faire en utilisant les modèles à régression polynomiale. Ces derniers sont destinés à analyser les résultats des plans d'expériences physiques, à créer des modèles empiriques à partir des valeurs observées et à interpréter statistiquement les résultats.

Dans le cadre de la régression polynomiale, on cherche à approximer la fonction Y (réponses du plan d'expériences) par une fonction polynomiale \hat{Y} tel que :

$$\hat{Y} = \beta_0 + \sum_i \beta_i x_i + \sum_i \beta_{ii} x_i^2 + \sum_{ij} \beta_{ij} x_i x_j \quad (1.12)$$

La détermination des coefficients nécessite au moins autant de couples distincts. Dans la pratique, on dispose généralement d'un nombre de mesures supérieur au nombre de coefficients. La détermination des coefficients se fait alors par la méthode des moindres carrés.

Estimation des coefficients des modèles polynomiaux

Considérons un modèle polynomial de la réponse, exprimée en fonction de k facteurs. Ce modèle fournit la valeur de la fonction réponse $y_{mod}(x^i)$ en tout point de coordonnées x^i , pour lequel on a :

$$y_{mod}(x^i) = f_x(x^i) \cdot \beta \quad (1.13)$$

Dans le cas de la réalisation d'une série de N expériences, la variable x prend des valeurs différentes. Notons les x^1, x^2, \dots, x^N .

On peut donc écrire :

$$y_{mod}(x^i) = X \cdot \beta \quad (1.14)$$

Avec :

$Y_{mod} = [y_{mod}(x^1) \ y_{mod}(x^2) \ \dots \ y_{mod}(x^N)]'$ le vecteur des N valeurs de réponses modélisée ;

$X = [f_x(x^1) \ f_x(x^2) \ \dots \ f_x(x^N)]'$ la matrice d'expériences bâtie à partir des N points x^1, x^2, \dots, x^N , et permettant le calcul d'un modèle polynomial à p coefficients. Le nombre de facteurs est k . Une telle matrice d'expériences de RSM n'est en général pas carrée.

Dans la pratique, le problème est différent, dans la mesure où :

- Le vecteur des réponses y est connu puisqu'il contient les réponses expérimentales ;
- Le vecteur des coefficients β est indéterminé et doit être estimé ;
- Il peut exister une erreur de mesure (e), pour chaque valeur de y .

Dans un premier temps, on doit alors considérer la relation :

$$y = X \cdot \beta + e \quad (1.15)$$

Etant données les caractéristiques propres aux expériences virtuelles, il résulte qu'il ne peut exister d'erreur de mesure ; le vecteur e est donc nul. En effet, lancer 2 fois la même simulation débouchera sur l'obtention de 2 valeurs de résultats identiques.

Par conséquent, on arrive à la relation $y = X \cdot \beta$, et donc finalement à $y = y_{mod}$. Cette dernière égalité signifie que le modèle passe exactement par les N points d'expérience. Mis à part tout cas fortuit, cela implique que les coefficients du modèle sont déterminés de façon univoque, et donc que ces p éléments constituent la solution unique des N relations linéaires données par $y = X \cdot \beta$. On en déduit que

$N = p$, la matrice X est donc ici saturée (carrée).

On remarque accessoirement que l'on doit toujours avoir $N \geq p$. Pour $N < p$, le système ci-dessus serait sous-déterminé.

Cependant, l'égalité $N = p$ n'est qu'un cas particulier. Ainsi, pour les situations plus courantes où

$N > p$, 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. Celui des moindres carrés est bien sûr le plus connu et le plus utilisé.

Pour chaque expérience existe alors un écart dit de modélisation, entre les valeurs mesurées et calculées de la réponse. Le vecteur ε représentera ces grandeurs.

On aboutit alors à la relation qui suit :

$$y = X \hat{\beta} + \varepsilon \quad (1.16)$$

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

$$\varepsilon = y - y_{mod} \quad (1.17)$$

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

Il est clair que l'estimation $\hat{\beta}$ doit être réalisée telle que l'erreur de modélisation soit minimale.

Le critère des moindres carrés traduit cette exigence par un objectif équivalent :

$$\|\varepsilon\|^2 = \|y - X \cdot \hat{\beta}\|^2 \text{ minimal}$$

Pour cela on cherche à trouver ε tel que : $\frac{\partial \varepsilon' \cdot \varepsilon}{\partial \hat{\beta}} = 0$

D'où on déduit l'expression connue de $\hat{\beta}$ déduite par la méthode des moindres carrés (pseudo-inverse de Moore-Penrose) :

$$\hat{\beta} = (X' \cdot X)^{-1} \cdot X' \cdot y \quad (1.18)$$

Avec X' : la transposée de X

On peut montrer les résultats suivants :

$\hat{\beta}$ constitue bien un minimum pour $\varepsilon' \cdot \varepsilon$;

$\hat{\beta}$ est un estimateur sans biais de β ;

$\hat{\beta}$ est, de tous les estimateurs sans biais de β de la forme $B \cdot y$ ($B = (X' \cdot X)^{-1} \cdot X'$), celui de la variance minimale.

2.2.2. Analyse de la variance sur modèle

L'analyse de la variance vue dans le cadre du screening peut également servir à tester la validité d'un modèle entier.

La valeur de $\hat{\beta}$ peut être obtenue par un raisonnement différent. On peut ainsi concevoir la grandeur $y_{mod} = \hat{\beta} \cdot X$ comme le résultat d'une projection D-orthogonale de y sur le sous espace W de dimension v , engendré par les variables correspondant à chacune des v colonnes de la matrice d'expériences X .

La définition mathématique d'un projecteur D-orthogonal étant $X \cdot (X' \cdot D \cdot X)^{-1} \cdot X' \cdot D$, on a donc :

$$y_{mod} = X \cdot (X' \cdot D \cdot X)^{-1} \cdot X' \cdot D \cdot y \quad (1.19)$$

Pour le projecteur particulier $D = \frac{1}{N} I_N$, on retrouve bien l'expression de $\hat{\beta}$ déjà trouvée.

Le fait que y_{mod} soit une projection orthogonale de y provoque l'apparition d'un triangle rectangle de côtés $y - 0 = y$, $y_{mod} - 0 = y_{mod}$ et $\varepsilon = y - y_{mod}$. 0 est le point d'origine appartenant au sous-espace W .

Le théorème de Pythagore peut donc s'appliquer :

$$y' \cdot y = y'_{mod} \cdot y_{mod} + \varepsilon' \cdot \varepsilon \quad (1.20)$$

Dans l'expression de la réponse mesurée y , le terme $X \cdot \beta$ appartient lui aussi à W mais est distinct de y_{mod} . On peut donc faire apparaître un second triangle rectangle de côtés $\varepsilon = y - y_{mod}$, $e = y - X \cdot \beta$ et

$X \cdot \hat{\beta} - X \cdot \beta$; on peut donc écrire de la même façon :

$$\varepsilon' \cdot \varepsilon = (y - X \cdot \beta)' \cdot (y - X \cdot \beta) - (X \cdot \hat{\beta} - X \cdot \beta)' \cdot (X \cdot \hat{\beta} - X \cdot \beta) \quad (1.21)$$

Le terme $\varepsilon' \cdot \varepsilon$ se révélera être important dans les investigations qui vont suivre. Il représente la somme des carrés des résidus, i.e. des écarts entre les réponses mesurées et les réponses calculées.

Suivant le contexte d'étude, on pourra ainsi être amené à noter cette quantité par l'écriture équivalente SCE.

L'analyse de la variance sur modèles complets repose entièrement sur la relation vue plus haut :

$$y' \cdot y = y'_{mod} \cdot y_{mod} + \varepsilon' \cdot \varepsilon \quad (1.22)$$

Le membre de gauche est appelé somme des carrés totaux ; il est aussi noté SCT.

Le terme $y'_{mod} \cdot y_{mod}$ correspond à la somme des carrés dus à régression (c'est-à-dire au modèle) ; il est noté également SCR.

On a donc l'égalité $SCT = SCR + SCE$, qui n'est autre qu'une écriture particulière de l'équation d'analyse de variance.

De ces sommes de carrés, on déduit les carrés moyens en divisant par les degrés de liberté correspondants.

Ainsi, les carrés moyens CMR dus à la régression sont :

$$CMR = \frac{SCR}{p} \quad (1.23)$$

Les carrés moyens CME associés aux écarts valent :

$$CME = \frac{SCE}{N-p} \quad (1.24)$$

CME est une estimation non biaisée de la variance expérimentale σ^2 .

Le nombre de degrés de libertés total est $(N - p) + p = N$, qui correspond bien au nombre de simulations du plan d'expériences.

On effectue alors le test de Fisher-Snedecor. Pour le modèle considéré, on calcule le ratio :

$$F_{obs} = \frac{CMR}{CME} \quad (1.25)$$

La variance due à la régression (CMR) peut être considérée du même ordre que la variance résiduelle (CME) si le rapport est F_{obs} inférieur à une valeur seuil statistique. L'hypothèse nulle H_0 sous-jacente revient à dire que $\beta = 0$. Sous cette hypothèse, F_{obs} est alors une valeur observée d'une variable F de

Fisher-Snedecor, à p et (N-p) degrés de liberté. L'hypothèse H0 doit être rejetée au niveau α quand :

$$P(F \geq F_{obs}) \leq \alpha$$

On utilise couramment le tableau 6 d'analyse de variance pour réunir ces informations.

Source de variation	ddl	Somme des carrés	Carrés moyens	Fobs	Probabilité P	Source influente ?
Régression (modèle)	p	SCR	CMR	CMR/CME	$P(F \geq F_{obs})$	$P < \alpha$
Résidus	N-p	SCE	CME			
Total	N	SCT				

Tableau 6: Tableau d'analyse de la variance sur modèle - Cas général

La qualité du modèle est mesurée par le calcul du coefficient de corrélation R^2 :

$$R^2 = \text{SCR} / \text{SCT} \quad (1.26)$$

2.2.3. Calcul des contributions de facteurs et interactions

Pour quantifier l'influence d'un(e) facteur (interaction) sur la réponse, le pourcentage de contribution d'un (e) facteur (interaction) x est calculé avec les coefficients de Fisher en utilisant la formule suivante :

$$C_x(\alpha) = \frac{F_x(\alpha)}{\sum F_x(\alpha)} \times R^2 \quad (1.27)$$

Avec :

α : Intervalle de confiance

F_x : Coefficient de Fisher du facteur ou de l'interaction x

R^2 : Coefficient de corrélation

2.3 Test de Kruskal-Wallis

Le test de Kruskal-Wallis est un test non paramétrique utilisé afin de déterminer si k échantillons indépendants proviennent d'une même population ou si au moins un échantillon provient d'une population différente des autres. Il est utilisé dans le cas où les critères de l'ANOVA ne sont pas vérifiés. Ce test consiste à :

- Poser l'hypothèse nulle H0 (les échantillons proviennent de la même population) et l'alternative H1 (les échantillons proviennent de populations différentes). Si on désigne par M_i la médiane du paramètre i, les hypothèses nulles H0 et alternatives H1 sont les suivantes :
 - H0 : $M_1 = M_2 = \dots = M_k$
 - H1 : il existe au moins un couple (i,j) tel que $M_i \neq M_j$
- Calculer la statistique de Kruskal-Wallis K et p-value ;
- Si p-value est supérieure à α (niveau de signification), on accepte H0. Sinon, on rejette H0.

Le calcul de la statistique K fait intervenir le rang des observations, une fois les k échantillons mélangés, K est défini par :

$$K = \frac{12}{N(N+1)} \times \sum_{i=1 \dots k} (R_i^2 - 3(N+1)) \quad (1.28)$$

Avec :

- $N = \sum n_i$; n_i est la taille de l'échantillon i
- R_i : la somme des rangs pour l'échantillon i
- p-value : est la probabilité que les k échantillons considérés n'appartiennent pas à la même population

V. Conclusion

Ce chapitre présente une revue bibliographique sur les paramètres matériau/procédé et leurs influences sur les propriétés des produits de fonderie sous pression, ainsi que les méthodes statistiques de plans d'expériences et les outils d'analyse de la variabilité. Nous constatons d'après cette synthèse que la plupart des travaux disponibles dans la littérature examinent l'influence des éléments d'alliage ou des paramètres procédé séparément. Aussi, ces travaux n'étudient pas la variabilité Matériau/Procédé en considérant un champ de variation bien établi. Ce qui explique notre choix de traiter la problématique productique en fonderie sous pression de la transformation des alliages d'aluminium liée à la variabilité matériau et la robustesse procédé en fabrication universelle. Les résultats attendus de cette étude permettront donc de prédire la convergence du couple matériau/procédé en relation avec le cahier des charges du produit, et par la suite d'améliorer les propriétés mécaniques des pièces coulées dès la phase de la conception.

Nous abordons maintenant les essais mécaniques et les procédures expérimentales réalisés tout au long de la thèse, en commençant par la présentation des plans d'expériences utilisés, puis la description des moyens et protocoles expérimentaux pour la caractérisation de la variabilité des caractéristiques mécaniques.