

Méthodes et exemples de construction de plans factoriels fractionnaires

A. Kobilinsky¹, Fliss M.², Carpentier B.³

Avril 1994

(1) Labo de Biométrie. INRA, Versailles.

(2) Malteries Soufflet, Quai Sarrail, Nogent-sur-Seine

(3) CNEVA, LERPAC, PARIS.

Journée d'étude du Groupe Français de Génie des Procédés

Stratégie expérimentale pour la recherche en
Procédés agro-alimentaires et biotechnologiques

Résumé

Parmi les méthodes de construction de plans fractionnaires, avec éventuellement un ou plusieurs systèmes de blocs, la méthode algébrique dite de la matrice clé est l'une des plus mal connues en France. Nous illustrons son intérêt au travers de deux exemples. Le premier est un plan en blocs relatif à l'optimisation de la qualité du malt, dont l'analyse fait intervenir la notion de décomposition en strate inter et intra bloc. Le second est une fraction de résolution 4 pour l'étude de la nettoyabilité des surfaces utilisées dans l'industrie agro-alimentaire.

I INTRODUCTION

Pour celui qui désire sélectionner une fraction ad hoc d'un plan factoriel complet et désire éventuellement répartir en blocs homogènes les traitements constituant cette fraction, il existe plusieurs méthodes.

— Consultation de catalogues, tels que ceux proposés dans (1), (2), (3). Il existe aujourd'hui de nombreux programmes facilitant l'accès à de tels catalogues et incluant des auxiliaires de calcul et randomisation permettant d'obtenir des plans "prêts à l'emploi". On pourra se référer à (4) pour avoir d'avantage d'information sur ces programmes.

— Utilisation d'une procédure automatique de recherche d'un plan optimum. Le critère utilisé pour définir l'optimum est par exemple le déterminant de la matrice $X'X$ des équations normales, critère D , qui caractérise la précision des estimations des paramètres du modèle. Un algorithme d'échange (5) ou une méthode de type gradient (6) peuvent être utilisés pour optimiser. Les plans obtenus sont cependant peu robustes. Si le modèle retenu pour optimiser est incomplet, ils peuvent conduire à des biais importants dans la réponse prédite. D'autre part ces procédures sont à l'heure actuelle incapable de produire des plans comme le plan de résolution 4 décrit plus loin, plan qui permet d'ajuster correctement un modèle polynomial de degré 1 même en présence d'une surface de réponse possédant une certaine courbure donc de type quadratique. Un article récent (7) laisse espérer qu'on pourra prochainement trouver des algorithmes modifiés corrigeant au moins en partie ces défauts.

— Utilisation d'une procédure algébrique de construction conduisant à un plan dit *régulier*. Ces procédures peuvent redonner une grande partie des plans figurant dans les catalogues. Mais elles permettent aussi de faire du sur mesure, pour peu que l'on connaisse les règles élémentaires gouvernant leur emploi. Il existe aussi des programmes permettant un choix automatique de la bonne procédure s'il en existe, à partir de la donnée d'un modèle et de termes à estimer dans ce modèle (8), (9), (10). C'est ce type de procédure algébrique, encore peu utilisée en France, que nous allons illustrer au travers de deux exemples: un plan en bloc utilisé au cours d'une étude sur l'optimisation de la qualité du malt, un plan fractionnaire de résolution 4 dont l'objet était d'améliorer l'évaluation de l'encrassement résiduel après nettoyage et désinfection de différents types de surface utilisées dans les industries agroalimentaires. En dehors de leur utilité propre, les plans construits par ces procédures peuvent être assemblés (11) pour apporter des solutions à une multitude de problèmes pratiques d'expérimentation. La pratique de ces assemblages ne sera pas abordée ici.

II PLAN EN BLOCS POUR OPTIMISER LA QUALITE DU MALT

II.1 Facteurs définissant les traitements et blocs

L'étude en question est décrite de façon précise dans (12). Nous nous intéressons ici à l'un des plans mis en place à la micromalterie. C'est un plan factoriel complet pour 6 facteurs A, B, C, D, E, F à 2 niveaux. Les trois premiers A, B, C concernent l'ajout d'une substance de croissance, la gibbérelline, après la trempe ou en début de germination. Le facteur D est le diagramme de trempe qui conduit à une plus ou moins grande humidité (40%, 45%) du grain. Les facteurs E et F sont la durée (5j, 6j) et la température (12°, 20°) de germination. Les $64 = 2^6$ traitements correspondants ne peuvent être expérimentés qu'à raison de 4 tous les deux jours. Les 4 lots de grain recevant ces traitements sont manipulés simultanément dans des conditions voisines: ils forment un bloc. La variabilité dite *intra-bloc* entre ces 4 lots est a priori plus faible que la variabilité *inter-bloc* entre lots expérimentés à des dates différentes. Aussi prend-on en compte ces blocs aussi bien lors de la conception du plan que lors de l'analyse des résultats. Les comparaisons entre traitements importantes, effets principaux et interactions entre deux facteurs, doivent si possible être effectuées *intra-bloc* c'est à dire par comparaison entre les unités d'un même bloc. Elles ont ainsi une précision accrue.

II.2 Répartition en blocs

Les quatre lots d'un bloc sont contenus dans une même enceinte et à la trempe dans une même cuve. Par suite les niveaux des facteurs D, E, F doivent être constants pour ces 4 lots. Chaque triplet de niveaux de ces facteurs D, E, F définit alors un ensemble de 8 traitements que l'on répartit en deux blocs en utilisant la valeur du produit ABC des trois autres facteurs, codés -1 ou 1 . Le choix du produit ABC est justifié plus loin.

Les blocs sont donc définis par les 4 *pseudofacteurs*

$$P_1 = D, \quad P_2 = E, \quad P_3 = F, \quad P_4 = ABC \quad (1)$$

Pratiquement on calcule pour chaque traitement les niveaux de ces pseudofacteurs comme cela est illustré dans le tableau 1. Dans ce tableau, le bloc est repéré par le quadruplet de niveaux de P_1, P_2, P_3, P_4 auquel on a associé un numéro de bloc compris entre 1 et 16.

A	B	C	D	E	F	$P_1 = D$	$P_2 = E$	$P_3 = F$	$P_4 = ABC$	n° bloc
1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	-1	1	1	-1	1	3
						...				
1	-1	-1	-1	-1	-1	-1	-1	-1	1	15
-1	1	1	1	1	1	1	1	1	-1	2
						...				
-1	-1	-1	-1	-1	1	-1	-1	1	-1	14
-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	16

TAB. 1 – Calcul des pseudofacteurs blocs donnant la répartition en blocs

II.3 Randomisation

Randomisation des blocs. L'ordre de réalisation des blocs doit si possible être *randomisé*, c'est à dire tiré au hasard. Ce tirage au hasard permet d'éviter qu'un dérèglement intempestif dans l'installation ou dans les procédés de mesure ne vienne superposer son

effet à celui d'un facteur tel que D , qui après tri sur les blocs et en l'absence de randomisation serait égal à 1 pour les 32 premières expériences, -1 pour les 32 dernières.

Cependant la micromalterie ne possède pour la germination qu'une seule enceinte fermée. Comme la trempe dure 2 jours tandis que la germination en dure 6, cette enceinte a été prévue pour recevoir 3 blocs simultanément. Ceci permet d'expérimenter un bloc tous les deux jours à condition de ne pas modifier la température de germination entre des blocs expérimentés à la suite. Pour profiter de cette possibilité, la randomisation a été restreinte de façon à ce que la température de germination F ne soit modifiée qu'une fois après les 8 premiers blocs. Pour simplifier l'organisation des essais, on a également restreint la randomisation pour la durée de germination E qui n'a été modifiée que tous les 4 blocs.

Ces restrictions dans la randomisation rendent non valides les tests des effets principaux température et durée de germination, ainsi que le test de l'interaction entre ces deux facteurs. On a ignoré dans la suite cette restriction dans la randomisation. Mais il faut être conscient que les tests de E , F et EF n'ont qu'une valeur indicative. Ils ne peuvent être interprétés qu'en faisant l'hypothèse, très plausible il est vrai, que la variabilité entre les groupes de 4 ou 8 blocs consécutifs est du même ordre que la variabilité entre blocs consécutifs.

Randomisation des unités dans les blocs. Les traitements doivent être placés au hasard dans chaque bloc, toujours pour éviter qu'un effet position, provenant par exemple d'une certaine hétérogénéité dans les températures ou l'aération du milieu, ne se superpose à l'effet d'un facteur.

Conception de la micromalterie. Cette micromalterie déjà ancienne n'a pas été conçue pour ce type d'étude et n'y est pas très bien adaptée. Ainsi si les 4 lots d'un bloc étaient immergeables dans des cuves séparées, il serait possible de randomiser pleinement le diagramme de trempe, gagnant ainsi une précision intra-bloc pour le facteur D et ses interactions avec E et F . Une telle organisation permettrait en outre de comparer des diagrammes de trempe différents sur une seule période de deux jours seulement. Un ensemble de trois enceintes séparées pour la germination aurait de même permis une randomisation pleinement valide des blocs et un gain de temps dans l'étude de facteurs tels que F .

Cet exemple illustre bien l'intérêt qu'il peut y avoir à considérer sérieusement les problèmes de planification expérimentale lors de la conception initiale des installations. Quelques dépenses supplémentaires peuvent parfois augmenter de façon considérable l'efficacité d'une installation.

II.4 La méthode de la matrice clé

La construction utilisée ici est un cas particulier d'une méthode beaucoup plus générale, dite de la matrice clé, qui trouve ses racines dans (13), (14), (15), (16). Programmée dès 1966 (17), cette méthode a été utilisée intensivement depuis cette date dans les expériences d'agronomie menées aux Royaumes Unis. Elle utilise des *facteurs de base* dont les combinaisons de niveaux définissent les unités expérimentales et forme à partir d'eux de nouveaux facteurs appelés *facteurs ajoutés*.

Dans l'exemple du malt, on crée à partir des facteurs de base A , B , C , D , E , F les facteurs ajoutés P_1 , P_2 , P_3 , P_4 définissant les blocs. Dans la présentation précédente, les

niveaux sont codés $-1, 1$ et P_4 est défini par le produit ABC . On peut aussi coder les niveaux 1 et 0 au lieu de -1 et 1 . Dans ce cas le produit $P_4 = ABC$ est remplacé par la somme modulo 2: $P_4 = A + B + C \pmod{2}$. La présentation utilisant le produit est dite multiplicative, celle qui utilise la somme est dite additive. La correspondance entre les codages additif et multiplicatif est définie par la fonction $x \mapsto (-1)^x$. En notation additive, les facteurs ajoutés peuvent aussi être définis à partir des facteurs de base en multipliant par une matrice appelée la *matrice clé*:

$$\begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} A \\ B \\ C \\ D \\ E \\ F \end{bmatrix} \pmod{2}$$

La présentation additive se généralise immédiatement à des facteurs à 3 niveaux et de façon plus générale à des mélanges de facteurs ayant des nombres de niveaux quelconques (18), (19), (11). Dans ce dernier cas il convient d'ajouter certaines restrictions au cadre général initialement proposé par Patterson (17) si l'on veut obtenir des propriétés simples.

II.5 Propriétés du plan et analyse des résultats

Avec les restrictions mentionnées ci-dessus, les plans construits par la méthode de la matrice clé ont une structure extrêmement simple et des propriétés qui se déduisent immédiatement des règles de construction (19), (20), (21).

Dans l'exemple on associe un effet factoriel à chacun des produits entre les facteurs traitements. Les produits réduits à une lettre sont les effets principaux, les produits de deux lettres les interactions de deux facteurs, et ainsi de suite. Une notation $e(\cdot)$ peut être utilisée pour distinguer un tel produit du pseudofacteur produit correspondant. Ainsi A ou $e(A)$ est l'effet principal du facteur A , AB ou $e(AB)$ l'interaction entre les deux facteurs A et B , etc... Des effets blocs sont de même associés aux 15 produits non vides entre les 4 pseudofacteurs blocs: $P_1, P_2, P_1P_2, \dots, P_1P_2P_3P_4$. Il est également utile d'introduire la moyenne générale notée $\mathbf{1}$ ou $e(\mathbf{1})$.

Les égalités (1) définissant le plan conduisent immédiatement aux 15 égalités suivantes:

$$\begin{aligned} P_1 &= D, \quad P_2 = E, \quad P_3 = F, \quad P_4 = ABC, \quad P_1P_2 = DE, \quad P_1P_3 = DF, \quad P_1P_4 = ABCD, \\ P_2P_3 &= EF, \quad P_2P_4 = ABCE, \quad P_3P_4 = ABCF, \quad P_1P_2P_3 = DEF, \quad P_1P_2P_4 = ABCDE, \\ P_1P_3P_4 &= ABCDF, \quad P_2P_3P_4 = ABCEF, \quad P_1P_2P_3P_4 = ABCDEF \end{aligned}$$

Règle: Les effets associés à des pseudofacteurs égaux sont *confondus*. C'est à dire qu'on ne peut pas les estimer séparément, mais qu'on ne peut estimer que leurs sommes $e(P_1) + e(D)$, $e(P_2) + e(E)$, \dots , $e(P_1P_2P_3P_4) + e(ABCDEF)$.

Parmi les effets factoriels des traitements, les 15 obtenus en effectuant les produits entre les facteurs D, E, F, ABC servant à définir les blocs sont donc confondus avec des effets factoriels des blocs. Les autres $e(A), e(B), e(AB), \dots, e(BCDEF)$ sont non confondus donc estimables séparément. Cette règle justifie le choix du produit ABC pour définir le quatrième pseudofacteur bloc P_4 . Si on avait pris un produit tel que AB comportant

seulement deux lettres, on aurait confondu avec les blocs les interactions AB , ABD , etc ... au lieu de ABC , $ABCD$, etc ... , donc on aurait perdu de l'information sur des interactions a priori plus importantes puisqu'elle implique chaque fois un facteur en moins.

Si l'on reparamétrise en fonction des sommes d'effets confondus et des effets non confondus, le modèle linéaire statistique décrivant le vecteur y des réponses est alors *orthogonal*. Plus précisément il s'écrit:

$$y = X\beta + \varepsilon = \mathbf{1}e(\mathbf{1}) + Ae(A) + Be(B) + \dots + P_1[e(P_1) + e(D)] + \dots + \varepsilon \quad \text{var}(\varepsilon) = \sigma^2\mathbf{I}$$

La matrice X , dont les colonnes sont les vecteurs $\mathbf{1}$, A , B , ... de dimension 64×1 , vérifie $X'X = 64\mathbf{I}$. Le vecteur des paramètres estimé est $\hat{\beta} = X'y/64$ et a pour coordonnées

$$\hat{e}(A) = A'y/64, \quad \hat{e}(AB) = (AB)'y/64, \quad \dots \quad \widehat{e(P_1) + e(D)} = D'y/64, \quad \text{etc} \dots$$

Pour estimer la variance d'erreur σ^2 , on fait l'hypothèse que les interactions de 4 facteurs et plus sont nulles. Il y a 15 interactions de ce type non confondues avec les blocs, donc 15 degrés de liberté pour estimer cette variance.

Pour tester les effets de D , E , F et leurs interactions, on fait l'hypothèse que les effets des blocs sont eux-même aléatoires d'espérance nulle et de variance σ_B^2 . Dans ces conditions, les 15 effets factoriels $e(P_1), \dots, e(P_1P_2P_3P_4)$ sont eux-même aléatoires d'espérance nulle et on peut estimer sans biais les effets confondus $e(D)$, $e(DE)$, etc ... par les combinaisons linéaires correspondantes $D'y/64$, $(DE)'y/64$. Ces dernières ont une variance $(4\sigma_B^2 + \sigma^2)/64$ que l'on peut estimer en faisant l'hypothèse que les 7 interactions de quatre facteurs et plus confondues – $ABCD$, $ABCE$, $ABCF$, $ABCDE$, $ABCDF$, $ABCEF$, $ABCDEF$ – sont nulles. Les résultats tels que ceux du tableau 2 montrent que cette variance est pour la plupart des variables observées nettement supérieure à la variance $\sigma^2/64$ des effets non confondus. Le fait d'ignorer les blocs conduirait donc ici à une variance estimée intermédiaire, trop forte pour tester les effets non confondus, trop faible pour tester les effets confondus et susceptible d'amener des erreurs importantes d'interprétation.

En pratique des programmes tels que ANOVA de GENSTAT (22), AOV de S Plus (23) permettent de traiter avec facilité ce type de données. On spécifie dans ce cas quel est le facteur bloc et ils trouvent automatiquement les effets confondus et la bonne variance d'erreur pour les tester (tableau 2). Ils permettent aussi d'estimer quelques données manquantes. Si le nombre de données manquantes devient trop important, il est préférable d'utiliser la méthode du maximum de vraisemblance restreint (Restricted Maximum Likelihood) qui estime d'abord les composantes de la variance σ_B^2 et σ^2 et utilise ce premier résultat pour estimer les effets factoriels des traitements. Cette méthode est programmée dans plusieurs logiciels très répandus (22), (24), (23). Les programmes en question sont cependant relativement récents, les sorties abondantes et pas toujours très claires et il est recommandé de s'assurer au départ le concours d'un statisticien pour vérifier que l'on comprend parfaitement les sorties utilisées.

III PLAN POUR L'ETUDE DE LA NETTOYABILITE DES SURFACES

Le comptage des bactéries restant adhérentes à une certaine surface après nettoyage nécessite une méthode efficace de détachement pour permettre les dénombrements ultérieurs. Le plan décrit ci-après est un des plans mis en place pour optimiser ce détachement. Il comprend 16 unités expérimentales et 8 facteurs à 2 niveaux (tableau 4) et (25)). C'est

Source de variation	d.f.	m.s.	v.r.	F pr.	Source de variation	d.f.	m.s.	v.r.	F pr.
strate bloc					strate bloc.*Unités*				
<i>D</i>	1	4.3733	8.05	0.025	<i>A</i>	1	2.7019	19.63	< .001
<i>E</i>	1	0.0490	0.09	0.773	<i>B</i>	1	8.3160	60.42	< .001
<i>F</i>	1	22.9082	42.16	< .001	<i>C</i>	1	8.5337	62.00	< .001
<i>D.E</i>	1	0.1903	0.35	0.573	<i>A.B</i>	1	1.0686	7.76	0.014
<i>D.F</i>	1	0.0631	0.12	0.743	<i>A.C</i>	1	0.1287	0.94	0.349
<i>E.F</i>	1	0.1453	0.27	0.621	<i>B.C</i>	1	0.0276	0.20	0.660
<i>A.B.C</i>	1	0.7988	1.47	0.265	<i>B.E</i>	1	0.9096	6.61	0.021
<i>D.E.F</i>	1	0.2364	0.44	0.531	<i>A.B.F</i>	1	1.5845	11.51	0.004
<i>Residuelle</i>	7	0.5433	3.95		<i>C.D.F</i>	1	1.7989	13.07	0.003
variance inter: $4\hat{\sigma}_B^2 + \hat{\sigma}^2 = 0.5433$					<i>Residuelle</i> 15 0.1376				
variance intra: $\hat{\sigma}^2 = 0.1376$									

TAB. 2 – Anova sur l'extrait (de protéines et sucres)

d.f. : degrés de liberté (degrees of freedom)
m.s. : carrés moyens (mean squares)
v.r. : rapport des variances (variance ratio)
F pr : probabilité qu'un *F* de Fisher Snedecor dépasse v.r.

donc une fraction 1/16 des $256 = 2^8$ traitements du plan factoriel complet. Il peut être défini en notation multiplicative par les égalités

$$E = BCD, \quad F = -ACD, \quad G = ABC, \quad H = ABD,$$

précisant les facteurs ajoutés *E, F, G, H* en fonction des facteurs de base *A, B, C, D*. Comme $A^2 = B^2 = \dots = H^2 = 1$, ces égalités conduisent à celles figurant à gauche du tableau 3 qui donnent les effets confondus avec la moyenne générale. La multiplication

1 = BCDE = -ACDF = -ABEF	A = ABCDE = -CDF = -BEF
ABCG = ADEG = -BDFG = -CEFG	BCG = DEG = -ABDFG = -ACEFG
ABDH = ACEH = -BCFH = -DEFH	BDH = CEH = -ABCFH = -ADEFH
CDGH = BEGH = -AFGH = -BCDEFGH	ACDGH = ABEGH = -FGH = -BCDEFGH

TAB. 3 – Effets confondus avec la moyenne générale et avec *A*

par chacun des 15 produits entre les facteurs de base *A, \dots, ABCD* donne ensuite les différents ensembles d'effets confondus. Par exemple, la multiplication par *A* donne la partie droite du tableau 3 où apparaissent les effets confondus avec *A*. Si les interactions de 3 facteurs et plus sont supposées nulles, le seul effet non nul parmi ces effets confondus est *A*. On peut donc estimer $e(A)$ par $\hat{e}(A) = A'y/16$. Le fait que *A* ne soit confondu qu'avec des interactions d'au moins trois facteurs vient du fait que les 15 effets confondus avec la moyenne générale à gauche du tableau 3 ont tous au moins 4 lettres. Une fraction vérifiant cette propriété est dite de *résolution* 4. Elle a la propriété fondamentale de permettre l'estimation des effets principaux dans un modèle incluant les interactions de deux facteurs. Par contre les interactions de deux facteurs n'y sont pas obligatoirement estimables. Ainsi la multiplication par *AB* conduit, si l'on ignore les produits de trois lettres et plus associés à des interactions négligeables, aux égalités

$$AB = -EF = CG = DH. \quad (2)$$

Seule la combinaison linéaire $e(AB) - e(EF) + e(CG) + e(DH)$ est estimable et on ne peut pas estimer séparément ces interactions confondues. Le tableau 4 donne les effets principaux et combinaisons linéaires d'interactions de deux facteurs ainsi trouvées estimables et leurs estimations, classées dans l'ordre des valeurs absolues décroissantes. Il n'y a pas ici comme dans l'exemple du malt d'interactions que l'on puisse négliger pour dégager des degrés de liberté pour l'erreur. On fait donc l'hypothèse que certains de ces effets ou combinaisons d'effets sont négligeables, savoir ceux dont les estimations sont les plus faibles en valeur absolue. Le graphique des quantiles –Half Normal Probability plot– (fig 1), (26) suggère de retenir les 9 derniers qui permettent d'obtenir une estimation de variance résiduelle et de tester les autres (tableau 4). Cette procédure conduit en règle générale à une sous estimation de la variance d'erreur et donc à des rapports de variance trop grands. Il n'y a cependant pas de doute ici que les effets principaux de A (puissance des ultra-sons) et B (nature du support, inox ou résine) sont très nettement significatifs, ce qui conduit à interpréter la combinaison linéaire significative $e(AB) - e(EF) + e(CG) + e(DH)$ comme étant l'interaction AB seule. Les effets apparaissant ensuite sur les lignes m, j du tableau 4, bien qu'ils donnent des résultats que l'on jugerait assez nettement significatifs dans une analyse de variance ordinaire, indiquent ici tout au plus des tendances. Ils seraient considérés comme non significatifs par une procédure plus rigoureuse comme celle de Zahn (27) version moderne de l'approche initiale de Daniel (26), qui fournit des limites (guardrails) que l'on peut facilement faire figurer sur le graphique des quantiles.

On notera que d'autres approches intéressantes pour analyser ce type de plan où on ne dispose pas a priori de degrés de liberté pour estimer la variance d'erreur ont été proposées (28), (29).

La température D trouvée antérieurement très influente entre 20° et 40° n'a pas ici d'effet décelable compte tenu de sa variation entre 40° et 45° . En fin de compte, cette expérience apporte avec seulement 16 essais des informations fiables sur les 8 facteurs étudiés. Il n'a pas été jugé utile compte tenu des résultats obtenus de chercher un complément d'information sur ces facteurs. Si l'interprétation avait été douteuse, on aurait pu définir des essais complémentaires pour *désaliaser* les effets confondus, c'est à dire en obtenir des estimations séparées.

	effet	est.	v.r.	F pr.
	1	69.80		
<i>a</i>	A : ultrason	17.76	72.7	< .001
<i>b</i>	B : support	17.52	70.7	< .001
<i>i</i>	$\underline{AB} - \underline{EF} + \underline{DH} + \underline{CG}$	-12.33	35.0	< .001
<i>m</i>	$\underline{CD} - \underline{AF} + \underline{GH} + \underline{BE}$	6.32	9.2	0.014
<i>j</i>	$\underline{AC} + \underline{EH} - \underline{DF} + \underline{BG}$	5.39	6.7	0.029
<i>l</i>	$\underline{BF} - \underline{AE} - \underline{CH} - \underline{DG}$	3.57	3.0	0.120
<i>e</i>	E : récipient	3.43		
<i>h</i>	H : EDTA	3.15		
<i>c</i>	C : souche	-2.62		
<i>d</i>	D : température	-1.76	$\hat{\sigma}^2 = 69.477$	
<i>k</i>	$\underline{AD} - \underline{CF} + \underline{EG} + \underline{BH}$	1.63	ddl=9	
<i>g</i>	G : détergent	1.49		
<i>o</i>	$\underline{BD} + \underline{CE} + \underline{AH} - \underline{FG}$	-1.35		
<i>n</i>	$\underline{BC} - \underline{FH} + \underline{DE} + \underline{AG}$	0.73		
<i>f</i>	F : $Na_4P_2O_4$ + PCSA	0.13		

TAB. 4 – Estimation et test des combinaisons linéaires d'effets estimables

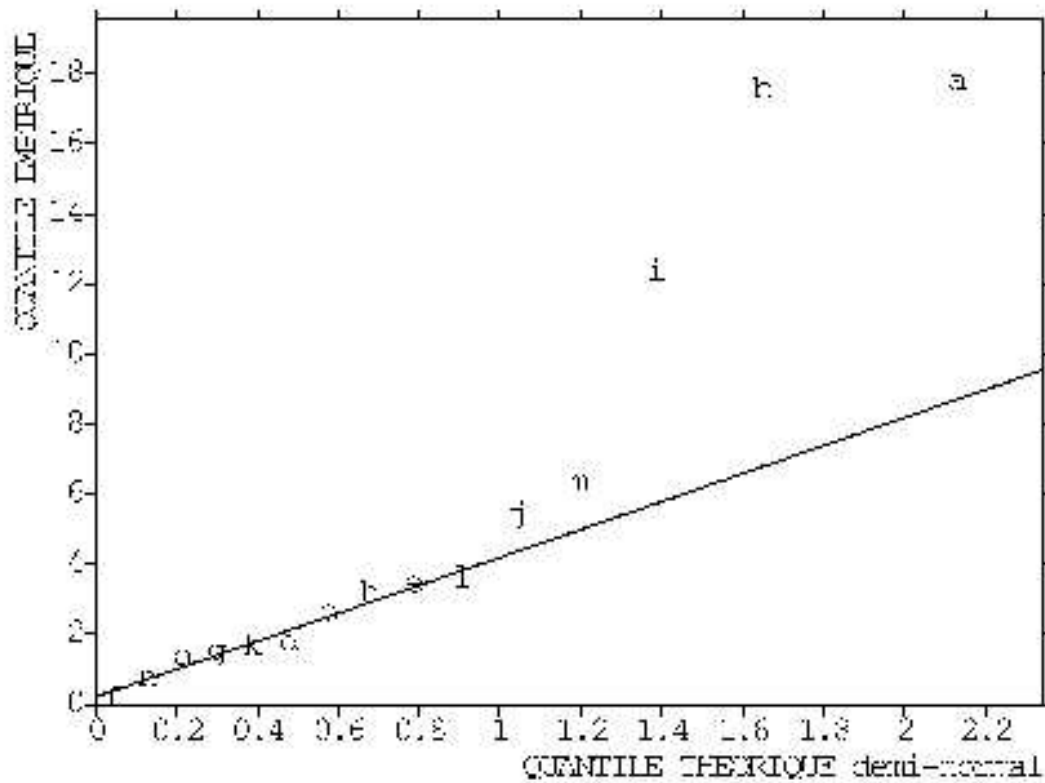


FIG. 1 – *Graphique des quantiles*

Dans cette étude, les 16 expériences correspondent au travail réalisable en deux semaines et les 8 facteurs épuisent en gros la liste de ceux que l'on pense susceptibles d'agir dans cette phase de la recherche. Le plan de résolution 4 est donc tout indiqué. Il est cependant intéressant d'examiner a posteriori ce qu'aurait donné un plan de résolution 3 avec 8 essais seulement et 7 de ces mêmes facteurs, par exemple A, B, C, D, E, F, G . Ce plan aurait confondu les effets principaux avec des interactions de deux facteurs. Une simulation montre qu'il aurait vraisemblablement détecté des effets significatifs de A, B et du facteur confondu avec l'interaction AB . Il aurait été alors nécessaire de poursuivre par 8 autres expériences pour désaliaser. Un tel plan aurait donc été désavantageux puisqu'il aurait conduit in fine à étudier avec un même nombre d'observations un facteur de moins. En fait, le plan de résolution 3 pour 8 unités et 7 facteurs ne semble vraiment intéressant que si le nombre de facteurs influents n'excède pas 1 ou 2 et qu'il n'y a pas d'interaction, cas qui se présente rarement en pratique.

REFERENCES

- (1) McLean R.A., Anderson V.L., Applied Factorial and Fractional Designs. Dekker, 373p, (1984)
- (2) Diamond W.J., Practical Experiment Design for Engineers and Scientists. LLP, 348p, (1981)
- (3) Taguchi G., System of Experimental Design. Unipub Kraus, 2 vol, (1987)
- (4) Rash D., Darius P., Computer Aided Design of Experiments. Technical Note 93-01, Labo. Stat. and Exp. Design. Faculty of Agricultural Sciences, K.U. Leuven (1993).
- (5) Mathieu D. Contribution de la méthodologie de la recherche expérimentale à l'étude des relations structure réactivité. Thèse de Docteur-ès-Sciences de l'université d'Aix-Marseille, (1981)
- (6) Hardin R.H., Sloane N.J.A., A new approach to the construction of optimal designs. *J. Statist. Plann. Infer.* **37**, 339-369, (1993).
- (7) DuMouchel W., Jones B., A simple bayesian modification of D -optimal designs to reduce dependence on an assumed model. *Technometrics*, **36**, 37-47 (1994)
- (8) Kobilinsky A., PLANOR: programme de génération automatique de plans d'expériences réguliers. Document interne, Labo. de Biométrie, INRA, Versailles (1994).
- (9) SAS/QC Software. The FACTEX procedure. SAS Institute Inc. (1992)
- (10) Zemroch, P.J., KEYFINDER - A complete toolkit for generating fractional-replicate and blocked factorial designs. In Dodge Y. and J. Whittaker (eds). *Computational Statistics Vol 2*, 263-268. Physica-Verlag (1992).
- (11) Kobilinsky A., Monod H., Juxtaposition of regular factorial designs and the complex linear model. A paraître dans *Scand. J. Statist.* (1994)
- (12) Fliss M., Delattre D., Brault R., Baret J.L., Kobilinsky A., Boudrant J., Les plans d'expériences en malterie. A paraître dans *Récents progrès en génie des procédés*, Vol 8, Ed. Groupe Français de Génie des Procédés., Nancy. (1994)
- (13) Yates, F., The design and analysis of factorial experiments. *Tech. Commonw. Bur. Soil Sci. No. 35*, Harpenden, (1937)
- (14) Fisher R.A., The theory of confounding in factorial experiments in relation to the theory of groups. *Ann. Eugen. Lond.*, **11**, 341-353, (1942)
- (15) Bose R.C., Mathematical Theory of the symmetrical factorial design, *Sankhya* **8**, 107-166 (1947).
- (16) Finney D.J., The fractional replication of factorial arrangements, *Ann. Eugen.*, **12**, 291-301, (1945)
- (17) Patterson H.D., Generation of factorial designs. *J. Roy. Statist. Soc. Ser. B* **38**, 175-179, (1976)
- (18) Bailey R.A., Dual abelian groups in the design of experiments, in *Algebraic Structures and Applications* (P.Schultz et al., Eds.), *Lecture Notes in Pure and Applied Mathematics*, **74**, Dekker, New York, 45-54 (1981)

- (19) Kobilinsky A. and Monod H., Experimental design generated by group morphisms: an introduction. *Scand. J. Statist.*, **18**, 119-134, (1991)
- (20) Cliquet S., Durier C., Kobilinsky A., Principle of a fractional design for qualitative and quantitative factors; application to the production of *Bradyrhizobium japonicum* in culture and inoculation media. A paraître dans *Agronomie*, (1994). La traduction en français de cet article est disponible au laboratoire de Biométrie INRA Versailles.
- (21) Goupy J., La méthode des plans d'expériences. *Dunod*, 303 p, 1988.
- (22) GENSTAT 5, Procédures ANOVA et REML. Reference Manual, *Oxford Science Publications, Clarendon Press*, (1987)
- (23) S-PLUS. Manuel de référence. SIGMA-PLUS, Toulouse, (1991)
- (24) SAS. Technical Report P-229, The MIXED Procedure. SAS Institute Inc. (1992)
- (25) Carpentier B., Kobilinsky A., Biofilms: outils pour l'étude de la nettoyabilité des surfaces utilisées dans l'industrie agro-alimentaire. *Actes de Contaminexpert 93*, (1993)
- (26) Daniel C., Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments, *Technometrics*, **1**, 311-341, (1959).
- (27) Zahn D.A., Modifications of and revised critical values for the half-normal plot. *Technometrics*, **17**, 189-200, (1975).
- (28) Box G.E.P, Meyer R.D. Finding the active factors in fractionated screening experiments. *J. of Quality Technology*, **25**, 94-105 (1993).
- (29) Lenth R.V. Quick and easy analysis of unreplicated factorials. *Technometrics*, **31**, 469-473 (1989).