Planification d'expériences en formulation

par **Didier MATHIEU**
Docteur ès sciences
Ingénieur de l'Institut national des sciences appliquées (INSA)
Ingénieur de l'Institut de pétroléochimie et de synthèse organique industrielle de Marseille (IPSOI)
Professeur à l'Université de la Méditerranée

et **Roger PHAN-TAN-LUU**
Docteur ès sciences
Ingénieur de l'École supérieure de chimie de Marseille (ESCM)
Professeur à l'université d'Aix-Marseille III

Bibliographie

Références

[1] LEWIS (G.A.), MATHIEU (D.) et PHAN-TAN-LUU (R.). – *Pharmaceutical experimental design, chapitre 1 : screening*. Marcel Dekker Inc., New York (1999).

[2] LEWIS (G.A.), MATHIEU (D.) et PHAN-TAN-LUU (R.). – *Pharmaceutical experimental design, chapitre 2 : factor influence study*. Marcel Dekker Inc., New York (1999).

[3] LEWIS (G.A.), MATHIEU (D.) et PHAN-TAN-LUU (R.). – *Pharmaceutical experimental design, chapitres 5 et 6 : Response Surface Methodology, optimization*. Marcel Dekker Inc., New York (1999).

[4] LEWIS (G.A.), MATHIEU (D.) et PHAN-TAN-LUU (R.). – *Pharmaceutical experimental design, chapitre 8 : exchange algorithms*. Marcel Dekker Inc., New York (1999).

[5] LEWIS (G.A.), MATHIEU (D.) et PHAN-TAN-LUU (R.). – *Pharmaceutical experimental design, chapitres 9 et 10 : mixtures*. Marcel Dekker Inc., New York (1999).

[6] CORNELL (J.A.). – *Experiments with mixtures. Designs, models and the analysis of mixture data* (2e éd.). John Wiley, New York (1990).

[7] BOX (G.E.P.) et DRAPER (N.R.). – *Empirical model-building and response surface*. John Wiley, New York (1987).

[8] MYERS (R.H.) et MONTGOMERY (D.C.). – *Response surface methodology*. John Wiley, New York (1995).

[9] SCHEFFE (H.). – *Experiments with mixtures*. J. Roy. Soc. Stat. Soc., Ser. B., *20*, p. 344-360 (1958).

[10] SNEE (R.D.) et MARQUARDT (D.W.). – *Screening concepts and designs for experiments with mixtures*. Technometrics, *18*, p. 19-20 (1976).

[11] PIEPEL (G.F.). – *Measuring component effect in constrained mixture experiments*. Technometrics, *24*, p. 29-39 (1982).

[12] COX (D.R.). – *A note on polynomial response functions for mixtures*. Biometrika, *58*, p. 155-159 (1971).

[13] LENTH (R.V.). – *Quick and easy analysis of unreplicated factorials*. Technometrics, *38* (4), p. 409-173 (1989).

[14] BOX (G.E.P.) et MEYER (R.D.). – *An analysis for unreplicated fractional factorials*. Technometrics, *28* (1), p. 11-18 (1986).

[15] BOX (G.E.P.) et HUNTER (J.S.). – *The $2^{k\text{-}p}$ fractional factorial designs*. Technometrics, *42* (1), p. 28-47 (2000).

[16] MATHIEU (D.), PHAN-TAN-LUU (R.), LANTERI (P.) et LONGERAY (R.). – *Chimie Industrielle, chapitre 5 : matrices de criblage*. Scharff (J.P.) et Perrin (R.), éd. Masson, Paris (1993).

[17] MATHIEU (D.), PHAN-TAN-LUU (R.), LANTERI (P.) et LONGERAY (R.). – *Chimie Industrielle, chapitre 6 : étude quantitative des facteurs*. Scharff (J.P.) et Perrin (R.), éd. Masson, Paris (1993).

[18] STARKS (T.A.). – *A note on small orthogonal main-effect plans for factorial experiments*. Technometrics, *6*, p. 220-222 (1964).

[19] GUPTA (V.K.), NIGAM (A.K.) et DEY (A.). – *Orthogonal main-effect plans for asymmetrical factorials*. Technometrics, *24* (2), p. 135-137 (1982).

[20] BROUDISCOU (A.). – *Génération de matrices d'expériences asymétriques à haut degré de saturation*. Thèse, université d'Aix-Marseille III (1994).

[21] SERGENT (M.), MATHIEU (D.) et PHAN-TAN-LUU (R.). – *Méthodologie de la recherche expérimentale appliquée aux mélanges de vins provenant de différents cépages*. Revue Française d'Œnologie, *2*, p. 36-43 (1985).

[22] MATHIEU (D.) et PHAN-TAN-LUU (R.). – *Plans d'expériences. Applications à l'entreprise, approche méthodologique des mélanges. Chapitre 4*. Éd. Droesbeke (J.J.), Fine (J.) et Saporta (G.), Éditions Technip, Paris (1997).

[23] MATHIEU (D.) et PHAN-TAN-LUU (R.). – *Plans d'expériences. Applications à l'entreprise, approche méthodologique des mélanges. Chapitre 5*. Éd. Droesbeke (J.J.), Fine (J.) et Saporta (G.), Éditions Technip, Paris (1997).

[24] BOUTEVIN (B.), RIGAL (G.), ROUSSEAU (A.), SCHAEFFNER (P.), MATHIEU (D.) et PHAN-TAN-LUU (R.). – *Optimisation des propriétés d'un revêtement à base d'un mélange d'émulsions aqueuses et de charges minérales*. Double liaison. Chimie des peintures, *306*, 59/17-66/24 (1981).

[25] MARCOS (M.), MELENDEZ (E.), SERRANO (J.L.), ELGUERO (J.), PHAN-TAN-LUU (R.) et MATHIEU (D.). – *Induced S_A-mesophases in binary, ternary, and quaternary mixtures of nematic' α-α'-dimethylbenzalazine derivatives*. Can. J. Chem., *62*, p. 2192-2201 (1984).

[26] McLEAN (R.A.) et ANDERSON (V.L.). – *Extreme vertices design of mixture experiments*. Technometrics, *8*, p. 447-454 (1966).

[27] CROSIER (R.B.). – *The geometry of* constrained *mixture experiments*. Technometrics, *26*, p. 209-216 (1986).

[28] DE AGUIAR (P.F.), BOURGUIGNON (B.), KHOST (M.S.), MASSART (D.L.) et PHAN-TAN-LUU (R.). – *D-Optimal designs*. Chem. Intell. Lab. Syst., *30*, p. 199-210 (1995).

[29] PERILLON (J.L.), LANTERI (P.), LONGERAY (R.), DREUX (J.), MATHIEU (D.), FENEUILLE (D.) et PHAN-TAN-LUU (R.). – *Optimisation des caractères rhéologiques de mélanges de copolymères d'acrylate et de méthacrylate d'alkyles et d'ester résinique*. Angew. Makromol. Chemie, *2170*, p. 99-118 (1985).

[30] BAUDUIN (G.), MOUANDA (J.), TAHA (M.), SERGENT (M.), MATHIEU (D.) et PHAN-TAN-LUU (R.). – *Mise au point d'un nouveau béton léger de polystyrène expansé. Études des constituants principaux*. European Pol. J., *23*, n° 6, p. 441-445 (1987).

[31] BAUDUIN (G.), MOUANDA (J.), TAHA (M.), SERGENT (M.), MATHIEU (D.) et PHAN-TAN-

LUU (R.). – *Mise au point d'un nouveau béton léger de polystyrène expansé. Études des additifs*. European Pol. J., *23*, n° 6, p. 447-453 (1987).

[32] CAMPISI (B.), CHICCO (D.), VOJNOVIC (D.) et PHAN-TAN-LUU (R.). – *Experimental design for a pharmaceutical formulation : optimisation and robustness*. J. Pharm. Biomed. Anal. *18*, p. 57-65 (1998).

[33] DERRINGER (G.) et SUICH (R.). – *Simultaneous optimization of several response variables*. J. Qual. Tech. *12* (4), p. 214-219 (1980).

[34] BOUTTEMY (S.), AUBRY (J.M.), SERGENT (M.) et PHAN-TAN-LUU (R.). – *Optimisation par la méthode du simplexe d'une microémulsion comme milieu réactionnel pour la peroxydation du citronellol par une source minérale d'oxygène singulet*. New J. Chem, *21*, p. 1073-1084 (1997).

[35] CHARDON (J.), NONY (J.), SERGENT (M.), MATHIEU (D.) et PHAN-TAN-LUU (R.). – *Experimental research methodology applied to the development of a formulation for use with textiles*. Chemometrics and Intelligent Laboratory Systems, *6*, 313-321 (1989).

[36] WALTERS (H.W.), PARKER (L.R.), MORGAN (S.L.) et DEMING (S.N.). – *Sequential simplex optimization*. CRC Press, Boca Raton (1991).

Ouvrages généraux

Ces ouvrages sont consacrés aux statistiques élémentaires : régression, etc.

[37] WONNACOTT (T.H.) et WONNACOTT (R.J.). – *Statistique* (4e éd.) Éd. Economica (1991).

[38] MONTGOMERY (D.C.) et PECK (E.A.). – *Introduction to linear regression analysis* (2e éd.). John Wiley & Sons (1992).

[39] DRAPER (N.) et SMITH (H.). – *Applied regression analysis* (2e éd.). Wiley-Interscience (1981).

Dans les Techniques de l'Ingénieur

[40] GOUPY (J.). – *Plans d'expériences*. P 230. Traité Analyse et Caractérisation, vol. P1 (1997).

[41] DELACROIX (A.) et PORTE (C.). – *Méthodes d'optimisation en chimie analytique*. P 225. Traité Analyse et Caractérisation, vol. P1 (1987).

Sites Internet

Pour tout contact avec les auteurs des articles [J 2 240] et [J 2 241] :

Didier MATHIEU : [mathieu@romarin.univ-aix.fr]

Roger PHAN-TAN-LUU : [roger.phan-tan-luu@LMRE.u-3mrs.fr]

On pourra probablement trouver à terme des informations intéressantes sur la planification d'expériences dans l'un des très nombreux sites de chimiométrie dans le monde. Les liens que proposent les trois suivants permettent d'accéder à un très grand nombre d'autres :

• Société Chimique de France : [*www.scifrance.org/*] et son groupe de Chimiométrie.

• Société de Chimiométrie belge : [*sch-www.uia.ac.be/chemomet/*].

• North American Chapter of the International Chemometrics Society (NAmICS) [*www.iac.tuwien.ac.at/NAmICS/WWW/welcome.html*].

Le site historique d'un des fondateurs de la planification d'expériences, G.E.P. Box, dirigé ensuite par le Pr William G. Hunter peut être consulté chez :

• Center for Quality and Productivity Improvement – University of Wisconsin-Madison : [*www.engr.wisc.edu/centers/cqpi/*].

Le site lié à un congrès international sur les mélanges (6-8 juin 2000) peut être consulté à : [*www.statease.com/clas_mix.html*].

Quelques-uns des sites suivants proposent des articles, des logiciels, des formations sur la planification d'expériences :

• C.F. Kavanaugh & Associates : [*www.kavanaugh.com/*].

• Macomb Intermediate School District : [*www.macomb.k12.mi.us/math/DOEmain*].

• Logiciel MIXSOFT : [*members.aol.com/mixsoft/overview.htm*].

• MultiSimplex : [*www.multisimplex.com/mixture.htm*].

• LPRAI : [*www.nemrodw.com*].

• Statistical Designs : [*www.statisticaldesigns.com/*].

• Walters, Parker, Morgan, Deming : [*www.multisimplex.com/simplexbook/index.htm*].

Logiciels

JUMP (SAS Institute)
NEMROD-W (LPRAI)
RS DISCOVER (Domain Solutions SA)
STATGRAPHICS + (Uniware)

Critères de qualité d'une matrice d'expériences

Il existe de nombreux critères, mathématiques ou qualitatifs, permettant de choisir une matrice d'expériences plutôt qu'une autre. Ils peuvent être classés en trois grandes catégories, selon qu'ils sont liés :

— au plan d'expérimentation : nombre d'expériences, coût de l'expérimentation, nombre et organisation des niveaux et des changements de niveaux, séquentialité, etc. ;
— à la qualité prévisionnelle du modèle mathématique postulé [7] isovariance par rotation, précision uniforme, variance de la prévision, etc. ;
— à la qualité des coefficients du modèle : précision, justesse (biais), indépendance, facteurs d'inflation, etc.

Rappelons que, si les résultats expérimentaux sont collectés au moyen de matrices d'expériences adéquates, les estimations des coefficients des modèles mathématiques polynomiaux, utilisés tout au long des articles [J 2 240] et [J 2 241], peuvent être calculés par la méthode des moindres carrés au moyen de l'équation matricielle :

$$\boldsymbol{B} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$$

avec
$\boldsymbol{Y}$ vecteur des valeurs mesurées de la réponse,
$\boldsymbol{B}$ vecteur des estimateurs b_i des effets β_i,
$\boldsymbol{X}$ matrice du modèle,
$\boldsymbol{X'}$ matrice transposée de $\boldsymbol{X}$,
$\boldsymbol{X'X}$ **matrice d'information**,
$(\boldsymbol{X'X})^{-1}$ inverse de $\boldsymbol{X'X}$, **matrice de dispersion** [37] [38] [39].

La plupart des critères de qualité des matrices d'expériences sont liés à la matrice $(\boldsymbol{X'X})^{-1}$ qui ne dépend que de la structure du modèle mathématique postulé et de la matrice d'expériences. Les valeurs mesurées y_i étant entachées d'une erreur expérimentale, les estimateurs b_i sont eux-mêmes entachés d'une incertitude que l'on caractérise soit par leurs variances σ_i^2, estimées par s_i^2 et proportionnelles aux termes diagonaux de $(\boldsymbol{X'X})^{-1}$, soit par un intervalle de confiance :

$$b_i - t_{v\,,\,\alpha/2} \times s_i \leqslant \beta_i \leqslant b_i + t_{v\,,\,\alpha/2} \times s_i$$

où $t_{v,\ \alpha/2}$ désigne le t de Student à v degrés de liberté pour un niveau de confiance de $1 - \alpha$.

La figure **A** montre les intervalles de confiance calculés pour deux coefficients selon cette formule.

En réalité, elle n'est valable pour le coefficient β_i que si tous les autres estimateurs b_i sont exactement égaux à leur valeur vraie β_j. Il faut donc utiliser la notion **d'intervalle de confiance joint**, qui, pour une valeur de α donnée, est un ellipsoïde de confiance.

Optimalité-D

Plus le volume de l'ellipsoïde est faible, plus les coefficients sont précis. Or le volume est proportionnel au déterminant de la matrice de dispersion, noté

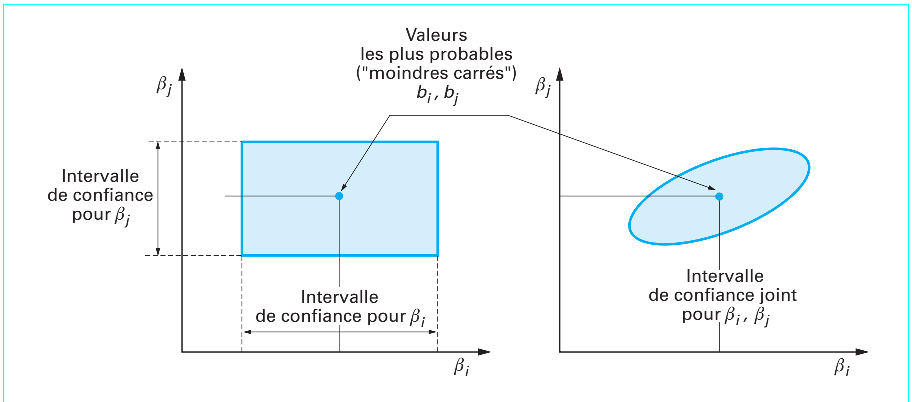

Figure A – Intervalles et ellipsoïde de confiance

$|(\boldsymbol{X'X})^{-1}|$, donc inversement proportionnel à $|(\boldsymbol{X'X})|$, ce qui conduit à la définition du critère de **l'optimalité-D** :

> de deux matrices d'expériences, la meilleure est celle qui conduit à estimer le plus précisément les coefficients du modèle postulé. La meilleure de toutes les matrices possibles est dite **optimale-D**. C'est elle qui conduit au déterminant de la matrice d'information maximal.

Efficacité-D – Rapport Qualité/Prix

Sachant qu'il suffit d'augmenter le nombre d'expériences pour augmenter le déterminant de la matrice d'information, donc la précision des coefficients, l'optimalité-D est un critère insuffisant. On lui préfère une sorte de rapport qualité/prix, défini par $|\boldsymbol{M}|$, où la matrice $\boldsymbol{M}$, appelée **matrice des moments**, est elle-même définie par $\boldsymbol{M} = \boldsymbol{X'X} / N$. C'est un critère qui tient compte du nombre d'expériences N de la matrice :

> une matrice d'expériences a la meilleure efficacité-D, donc le meilleur rapport qualité/prix si elle apporte une plus grande quantité d'information par expérience, donc si elle conduit à $|\boldsymbol{M}|$ maximal. Rappelons aussi que l'isovariance par rotation dépend aussi de la matrice des moments.

Optimalité-A

La trace de la matrice de dispersion, notée Trace $(\boldsymbol{X'X})^{-1}$, est la somme des termes diagonaux, donc des estimations des variances des coefficients. Minimiser Trace $(\boldsymbol{X'X})^{-1}$ revient à minimiser la variance moyenne des coefficients, donc la moyenne des longueurs des axes de l'ellipsoïde. Ce critère est appelé **l'optimalité-A** (de l'anglais *Average*).

Orthogonalité

Une matrice d'expériences est dite avoir la propriété d'**orthogonalité** lorsqu'elle permet d'obtenir des estimateurs des coefficients indépendants (non corrélés). Cela se caractérise par des axes de l'ellipsoïde parallèles aux axes des coefficients. Cette propriété est obtenue quand $(\boldsymbol{X'X})^{-1}$ (ou $\boldsymbol{X'X}$) est diagonale, donc quand les covariances des coefficients sont nulles.

Les critères -D, -A et l'orthogonalité sont différents, mais très liés : à volume égal, Trace $(\boldsymbol{X'X})^{-1}$ est minimisée lorsque toutes les variances sont égales, ce qui implique que l'ellipsoïde est une sphère, donc que la matrice est orthogonale. Il faut aussi noter que, lorsque $(\boldsymbol{X'X})^{-1}$ est diagonale, $|(\boldsymbol{X'X})^{-1}|$ est proportionnel au produit des variances des coefficients.

Optimalité-G

Les estimations des coefficients étant entachées d'une incertitude, il en est de même de toute prévision calculée grâce au modèle mathématique. L'incertitude de la prévision (réponse calculée par le modèle) est, elle aussi, caractérisée par la variance de la réponse prédite. On peut écrire :

$$\mathrm{Var}(\hat{y}_A) = d_A \sigma^2$$

avec $\hat{y}_A$ valeur de la réponse calculée en un point A quelconque du domaine expérimental,

σ^2 variance de la réponse expérimentale.

On peut montrer que d_A, appelée **fonction de variance** de la réponse calculée, dépend exclusivement des coordonnées du point A et de la matrice $(\boldsymbol{X'X})^{-1}$. On note d_{max} la valeur maximale de la fonction de variance dans le domaine expérimental, obtenue au(x) point(s) au(x)quel(s) la prévision est la moins précise. On dit d'une matrice d'expériences qu'elle est **optimale-G** si elle minimise la valeur maximale d_{max} de la fonction de variance :

> en d'autres termes, entre plusieurs matrices d'expériences différentes, la matrice **optimale-G** est celle pour laquelle d_{max} est minimale. Pour cette raison, ce critère est parfois appelé le critère ***minimax***.

C'est le critère-D qui est utilisé dans l'algorithme d'échanges présenté dans le paragraphe 1.3 de l'article [J 2 241]. Une des principales raisons est qu'il nécessite beaucoup moins de calculs que les autres. Il faut cependant remarquer que les critères $|(\boldsymbol{X'X})^{-1}|$, Trace $(\boldsymbol{X'X})^{-1}$ et d_{max} ont pour valeur limite théorique 0, et que, pour les deux premiers critères, il n'existe aucune valeur indicative au-dessous de laquelle la matrice d'expériences pourrait être considérée acceptable. Il s'ensuit que la meilleure matrice selon ces critères peut en réalité être la moins mauvaise. Il faut donc s'assurer que la matrice $(\boldsymbol{X'X})^{-1}$ est « suffisamment » diagonale afin de s'approcher de l'orthogonalité. De même, si la valeur optimale de $|\boldsymbol{M}|$ est 1, il ne s'agit que d'un rapport « qualité/prix » et non d'une qualité absolue.

Seule la fonction de variance d_{max} possède une valeur caractéristique absolue : la valeur 1 signifie que, au(x) point(s) du domaine expérimental pour le(s)quel(s) la prévision du modèle sera la moins précise, celle-ci sera cependant de même qualité que la mesure expérimentale $(\mathrm{Var}(\hat{y}_A) = d_A \sigma^2)$. En d'autres termes, l'intervalle de confiance de la valeur calculée par le modèle sera identique à l'intervalle de confiance de la valeur expérimentale, si celle-ci est effectuée.

Tous ces critères sont dits **critères a priori**, car ils peuvent être calculés avant que les expériences soient effectuées.