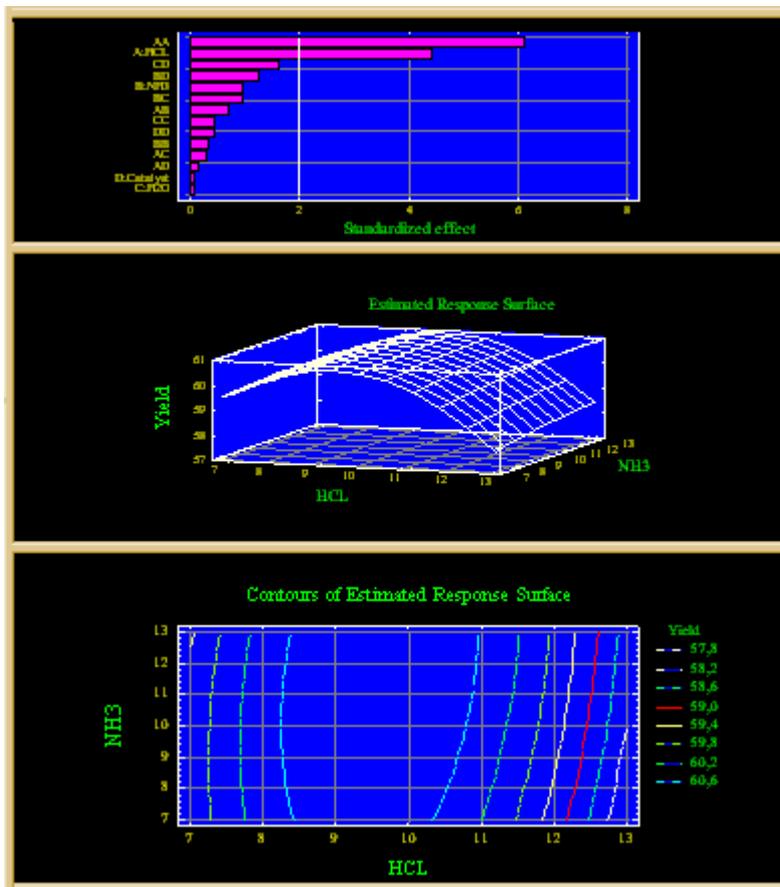




Initiation aux Plans d'Expériences



Par l'utilisation de StatGraphics

Christophe CORDELLA

(AFSSA Sophia Antipolis/ Laboratoire Arômes-Synthèses-Interactions)

Table des Matières

1) RAPPELS THEORIQUES	3
1.1- QUELLE EST LA MEILLEURE DEMARCHE POUR ETUDIER UN PHENOMENE ?	3
1.2- LES PLANS D'EXPERIENCES	4
1.2- APPLICATIONS DES PLANS D'EXPERIENCES	5
1.3- LES DEUX PRINCIPAUX TYPES DE PLANS D'EXPERIENCES	5
1.3.1- LES PLANS DE CRIBLAGES (« SCREENING DESIGN »)	5
1.3.2- LES PLANS DE SURFACE DE REPONSE (« <i>RESPONSE SURFACE DESIGN</i> »)	6
a) Plan central composite (« Central Composite Design »)	6
b) Plan Box-Behnken	7
c) Plan de Doehlert	7
d) Plan de mélange	8
1.4- METHODOLOGIE DES PLANS D'EXPERIENCES	9
2) UTILISATION DE STATGRAPHICS POUR METTRE EN PLACE UN P.E.	10
3) INTERET ET LECTURE DES PRINCIPAUX GRAPHES ET TABLEAUX DE DONNEES OBTENUS DANS LA PLUPART DES LOGICIELS	13
1)RESUME DU PLAN D'EXPERIENCES	13
2)LA MATRICE DE CORRELATION DU PLAN	13
3)ANALYSE DU PLAN	13
A. ESTIMATION DES EFFETS	13
B. ANALYSE DE LA VARIANCE (ANOVA)	14
C. COEFFICIENTS DE REGRESSION	14
D. TABLE DES POINTS INFLUENTS	14
4)LES GRAPHES ET LEUR INTERPRETATION	14
A. LE DIAGRAMME DE PARETO (<i>PARETO CHART</i>)	14
B. LE DIGRAMME DES EFFETS PRINCIPAUX (<i>MAIN EFFECTS PLOT</i>)	14
C. LE DIAGRAMME DES INTERACTIONS (<i>INTERACTIONS PLOT</i>)	14
D. LE DIAGRAMME DE LA LOI NORMALE (<i>NORMAL PROBABILITY PLOT OF EFFECTS</i>)	14
E. SURFACE DE REPONSES	15
F. LE DIAGRAMME DE DIAGNOSTIC (<i>DIAGNOSTIC PLOT</i>)	15
4) LES PRINCIPAUX LOGICIELS DE PLAN D'EXPERIENCES	15

1) Rappels théoriques

Les **plans d'expériences** (PE) font partie des outils statistiques appartenant à ce qui est courant d'appeler *l'Expérimentique*, qui n'est autre que l'ensemble des travaux regroupés sous la forme d'une méthodologie permettant de répondre à la question fondamentale de l'expérimentateur : « Est-ce que j'utilise la meilleure stratégie pour conduire mes recherches ? » (I).

1.1- Quelle est la meilleure démarche pour étudier un phénomène ?

(Positionnement du vocabulaire relatif aux PE)

La méthode classique pour étudier un phénomène physique ou chimique consiste à fixer certains paramètres pendant que l'on observe (enregistre) les variations d'un autre. On peut illustrer cela de la façon suivante : un agronome s'intéresse à une grandeur, par exemple le rendement en blé d'une parcelle de terre. Cette grandeur, le rendement, sera fonction de la nature du terrain, de la quantité d'engrais incorporé, de l'exposition au soleil, du climat local, de la variété de blé ensemencé, etc...

Sous forme mathématique, la grandeur qui intéresse notre agronome se note y et s'appelle également **grandeur réponse** ou **réponse**. Y est fonction de plusieurs variables x_i que l'on appelle **facteurs**.

$$y = f(x_1, x_2, x_3, \dots, x_k)$$

Pour étudier la fonction y entièrement et donc obtenir une bonne modélisation de la grandeur « rendement en blé de la parcelle de terre », l'agronome devra fixer le **niveau** de tous les facteurs choisis dans son expérimentation, sauf un et mesurer la réponse y en fonction de plusieurs valeurs du facteur non fixé x_1 .

Lorsque l'expérimentation sur x_1 est terminée, l'agronome pourra alors tracer la courbe représentative de $y = f(x_1)$, et devra recommencer les expériences sur les autres facteurs x_2, x_3, \dots

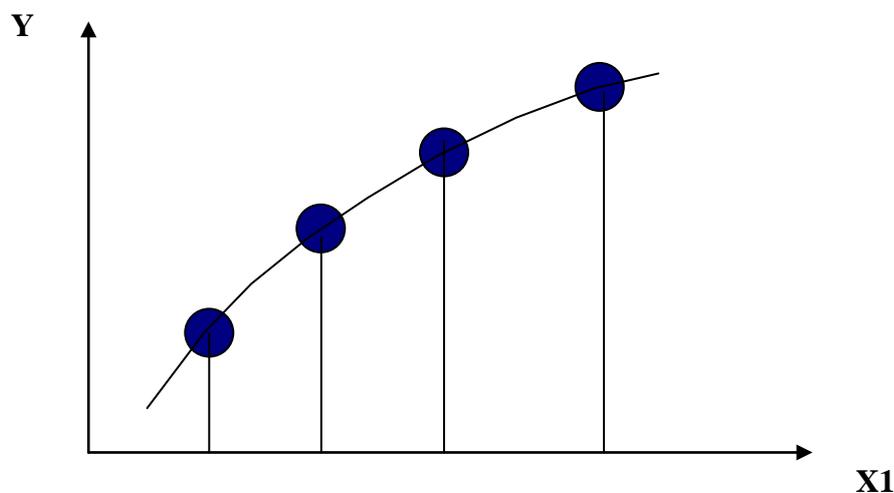


Figure 1 : Seule la variable x_1 prend des niveaux différents, les autres sont fixées.

(1) Introduction aux plans d'expériences. Jacques Goupy. 2^{ème} édition chez DUNOD, Série Industrie-Techniques. Paris, 2001. ISBN 2 10 005606 9

Un rapide calcul montre que si notre agronome souhaite faire une étude complète de tous les facteurs (7 par exemple) en prenant soin de réaliser 5 **points expérimentaux** par facteurs, il lui faut réaliser :

78 125 expériences (5^7). Ce qui n'est absolument pas réalisable. Pour réduire ce nombre, il n'y a que 2 possibilités :

Réduire le nombre de points expérimentaux

Au lieu de 5 points expérimentaux, on peut n'en prendre que 3 mais cela nous laisse tout de même $3^7 = 2187$ expériences à réaliser. **Encore beaucoup trop.**

Au lieu de 3 points expérimentaux, plus que 2, cela donne $2^7 = 128$ expériences. **Faisable mais quel travail !** Sans compter les incidents possibles qui peuvent nécessiter de refaire une expérience et le temps que cela représente.

Etant donné qu'il n'est pas possible de faire moins de 2 expériences par facteurs, notre agronome doit inévitablement réduire le nombre de facteurs étudiés (ce qui n'est pas sa volonté initiale et l'éloigne donc de ses objectifs).

Réduire le nombre de facteurs étudiés

Au lieu de 7 facteurs, l'agronome décide de n'en garder que 4 et de mesurer 3 niveaux pour ces facteurs, il devra réaliser $3^4 = 81$ essais.

Par cette approche le nombre d'expériences devient réaliste mais l'abandon de certains facteurs laisse toujours planer un doute sur les résultats de part le manque d'information que cela implique.

Adopter ou non cette approche dépendra des enjeux et des risques liés à la sécurité ou au montant financier propre à l'expérimentation en cours. Dans tous les cas, elle reste incomplète par définition et souvent on ne peut pas s'en satisfaire totalement. Il conviendra alors de se tourner vers les plans d'expériences.

1.2- Les plans d'expériences

Les plans d'expériences permettent de faire varier les niveaux de tous les facteurs à chaque expérience, mais de manière programmée et raisonnée.

Les avantages majeurs sont les suivants :

- diminution du nombre d'essais (plan factoriels complets ou fractionnaires),
- nombre de facteurs étudiés important,
- détection des interactions entre facteurs (plan de criblage),
- détection des optimaux (plan de surface de réponse),
- meilleure précision sur les résultats (points centraux),
- optimisation des résultats,
- modélisation des résultats (accès aux modèles mathématiques).

1.2- Applications des plans d'expériences

Il peut y avoir nécessité d'utiliser un plan d'expériences principalement pour 3 raisons :

- 1) Etudier l'influence de différents facteurs sur un signal analytique afin de tester la robustesse d'une méthode,
- 2) Evaluer parmi un ensemble de facteurs en interaction entre eux, ceux qui sont les plus influents,
- 3) Optimiser une grandeur expérimentale ou plusieurs en même temps. Exemple : un rendement chimique, une propriété physique et l'aspect d'un produit simultanément, etc...

1.3- Les deux principaux types de plans d'expériences

1.3.1- Les plans de criblages (« Screening Design »)

Ce sont des plans d'expériences à **2 niveaux** notés : 2^k avec k : le nombre de facteurs
 2 : le nombre de niveaux

Ce sont des plans factoriels complets, car leur forme mathématique prend en compte tous les facteurs étudiés.

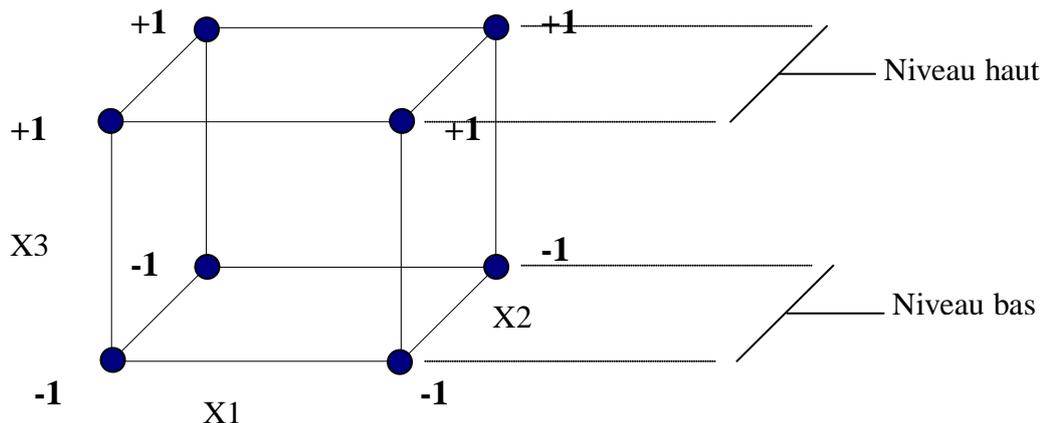


Figure 2 : Une façon de représenter les niveaux d'un plan factoriels complet à 2 niveaux.

Variante : les plans factoriels fractionnaires. Exemple : 2^{3-1} : 1 facteur de moins.

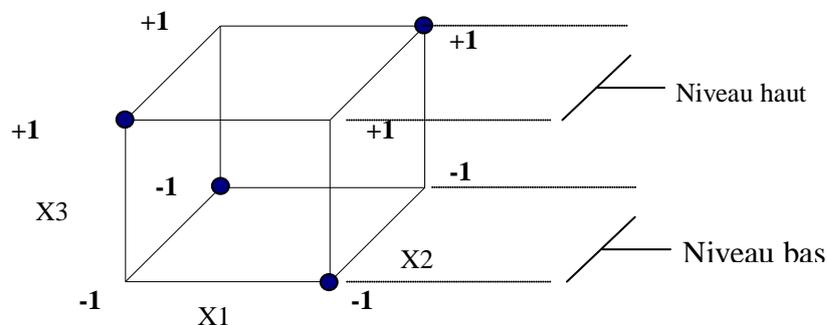


Figure 3 : Une façon de représenter les niveaux d'un plan factoriels à 2 niveaux.

Les plans de criblage sont principalement utilisés pour évaluer les effets des facteurs sur un process analytique.

1.3.2- Les plans de surface de réponse (« Response Surface Design »)

Les plans de surface de réponse sont des plans à 3 niveaux notés : 3^k

Cependant le nombre d'expériences pouvant devenir relativement important, des plans particuliers ont été développés :

a) Plan central composite (« Central Composite Design »)

Ce plan est une combinaison d'un plan factoriel complet ou fractionnaire et d'un plan factoriel supplémentaire, en général un plan en étoile.

Si le centre des 2 plans coïncident, on l'appelle **plan central composite**.

Ce plan est noté : $r = 2^{k-p} + 2k + n_0$. Avec r = nombre d'expériences
.k = nombre de facteurs .p = nombre entier pour la réduction du plan n_0 = nombre d'expériences réalisées au centre du plan.

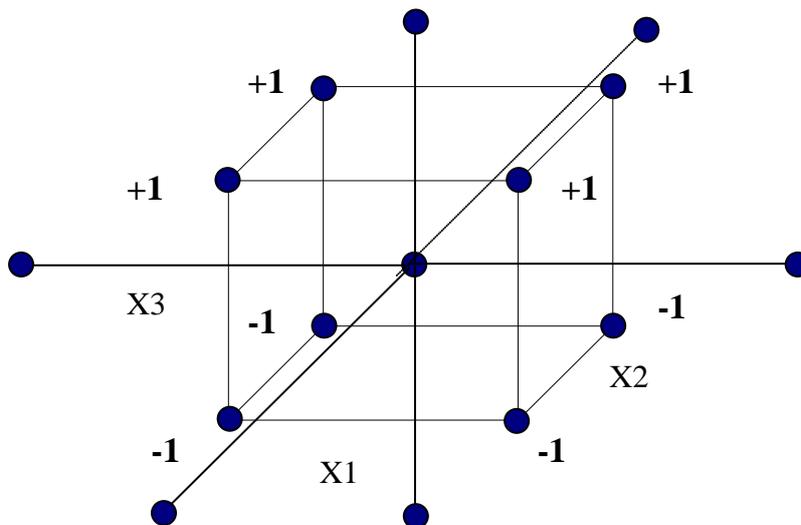


Figure 4 : Représentation d'un plan central composite en étoile.

b) Plan Box-Behnken

Dans ce plan, les points sont disposés sur une hypersphère donc à distance égale par rapport au centre.

C'est une autre façon de modéliser le domaine d'expérience.

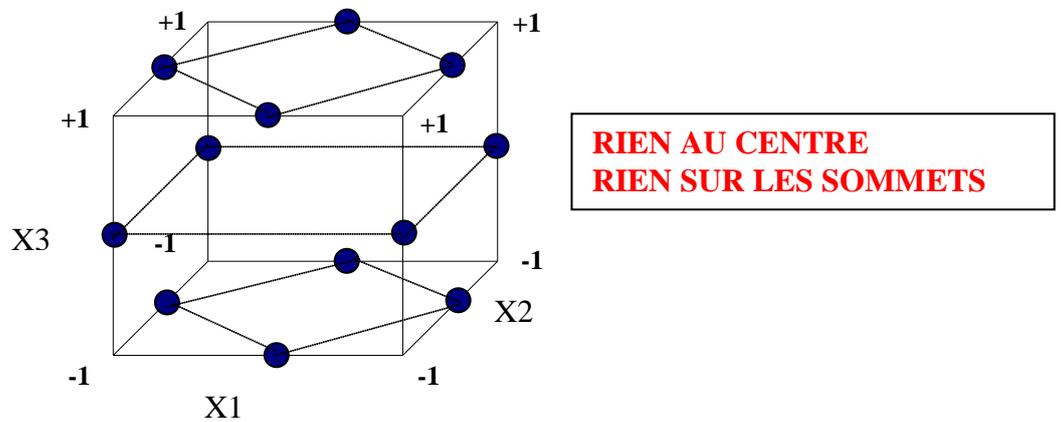
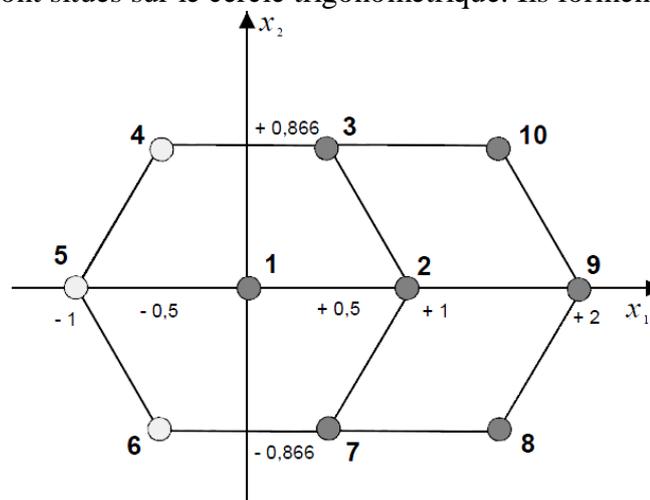


Figure 5 : Représentation d'un plan factoriel Box-Behnken.

c) Plan de Doehlert

La caractéristique principale des plans de Doehlert ⁽²⁾ est d'avoir une répartition uniforme des points expérimentaux dans l'espace expérimental. La Figure ci-dessous donne la disposition de ces points pour un plan à deux facteurs (essais 1 à 7). Tous les points sont à la même distance du centre du domaine d'étude et sont situés sur le cercle trigonométrique. Ils forment un hexagone régulier.



Les points 1 à 7 illustrent un premier plan de Doehlert. Les trois points 8, 9 et 10 illustrent les expériences supplémentaires. Les points 2, 7, 8, 9, 10, 3 et 1 illustrent un deuxième plan de Doehlert.

Si l'expérimentateur désire explorer le domaine expérimental, il peut facilement ajouter des points d'expériences supplémentaires et retrouver une disposition identique à celle de départ.

² DOEHLERT David H. «Uniform Shell Design» *Appl. Stat.* n°19, p 231. (1970).

La Figure montre qu'avec trois points d'expériences supplémentaires (essais 8, 9 et 10), on peut obtenir un nouveau plan de Doehlert (essais 2, 7, 8, 9, 10, 3 et 1).

Ce type de plans existe pour un nombre quelconque de facteurs.

d) Plan de mélange

Ces plans sont utilisés pour mettre au point des formulations et sont également des plans factoriels. Voir l'ouvrage suivant pour plus d'information : R. Kellner, J.-M. Mermet. *Analytical Chemistry*, Wiley – VCH. p 766.

1.4- Méthodologie des plans d'expériences

Voici la démarche générale lors de la mise en place et le développement d'un plan d'expériences. Cette démarche est généralement reprise dans la plupart des logiciels de création et d'analyse de plans d'expériences comme par exemple Statgraphics (<http://www.statgraphics.com/>) ou Modde (http://www.umetrics.com/default.asp/pagename/software_modde/c/2).

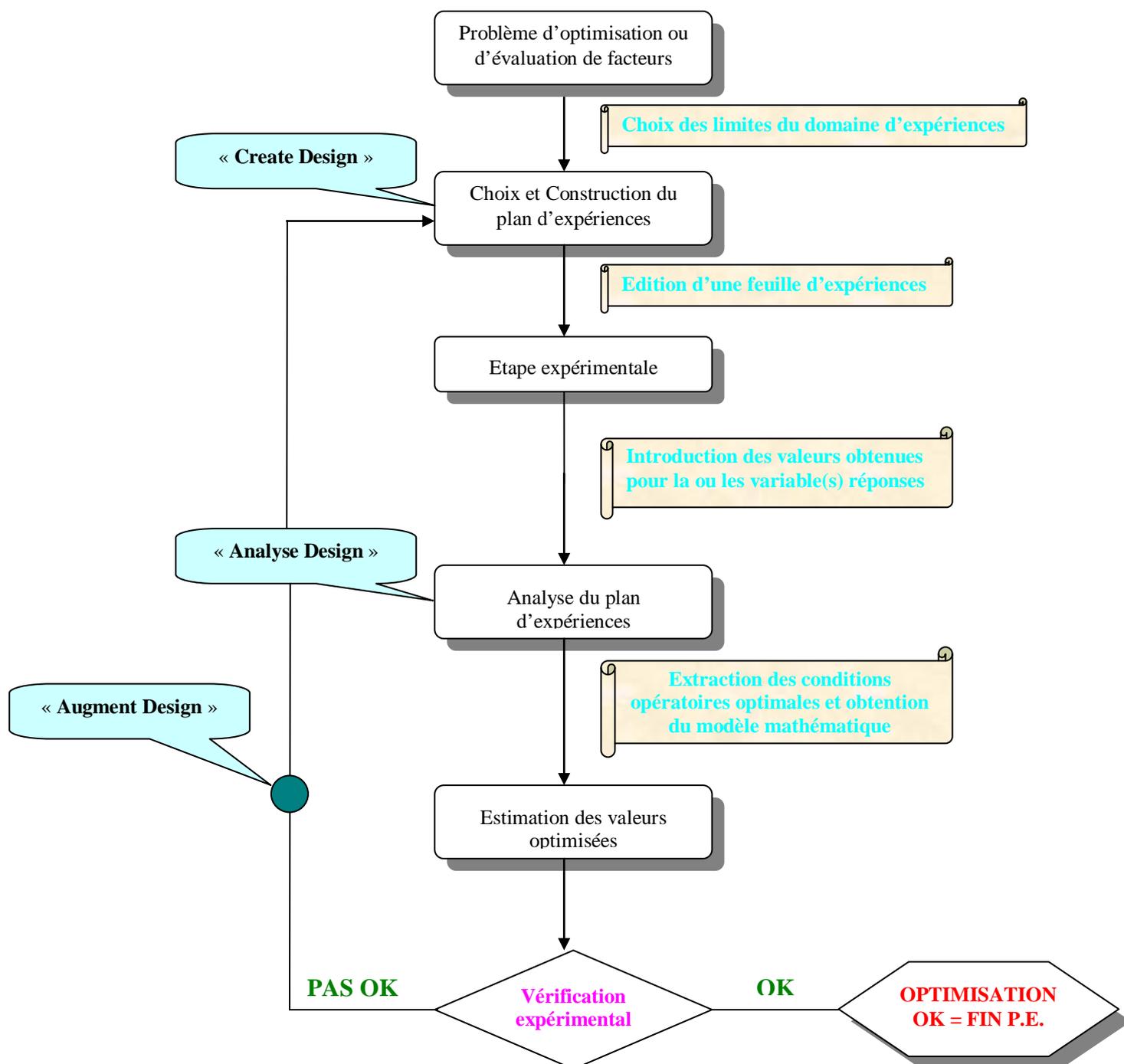
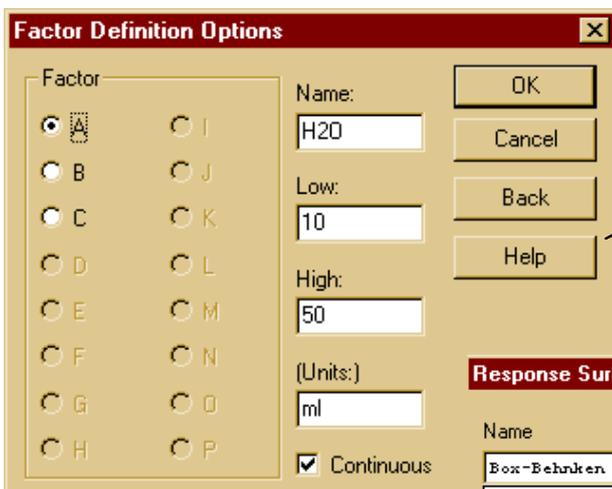
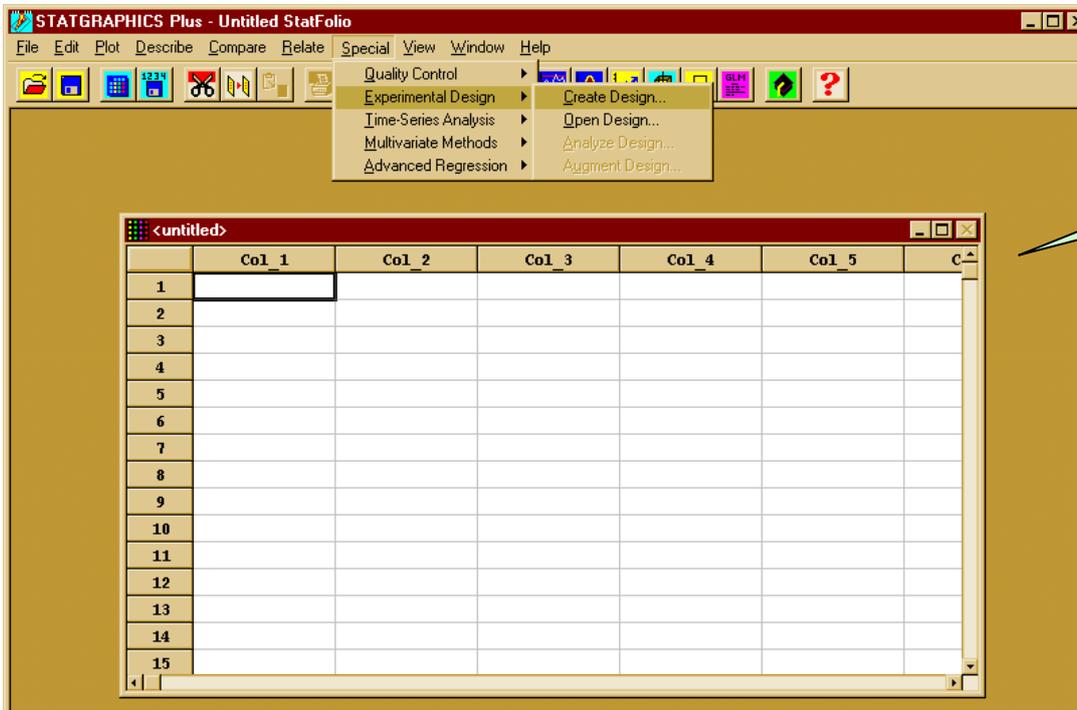


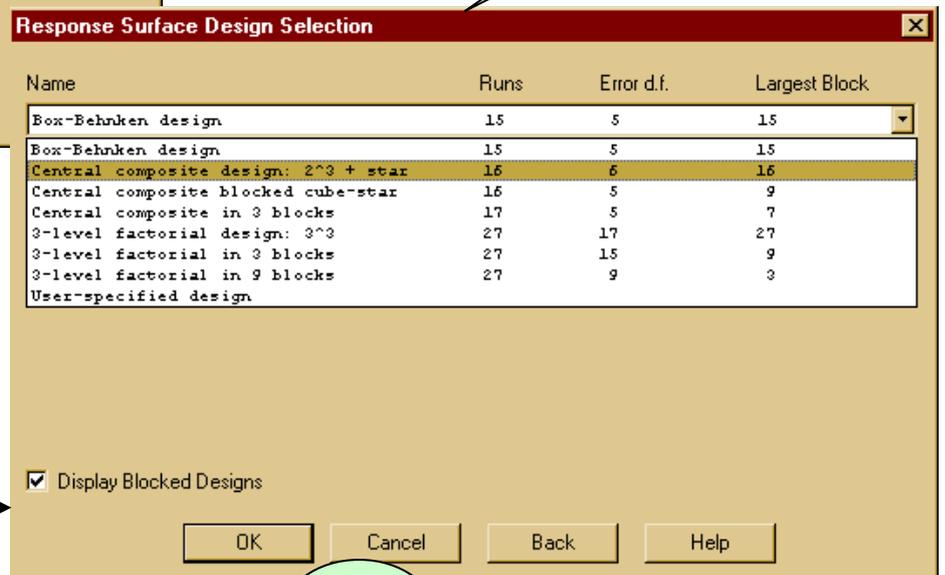
Figure 6 : Démarche générale pour la réalisation d'un plan d'expériences.

2) Utilisation de Statgraphics pour mettre en place un P.E.

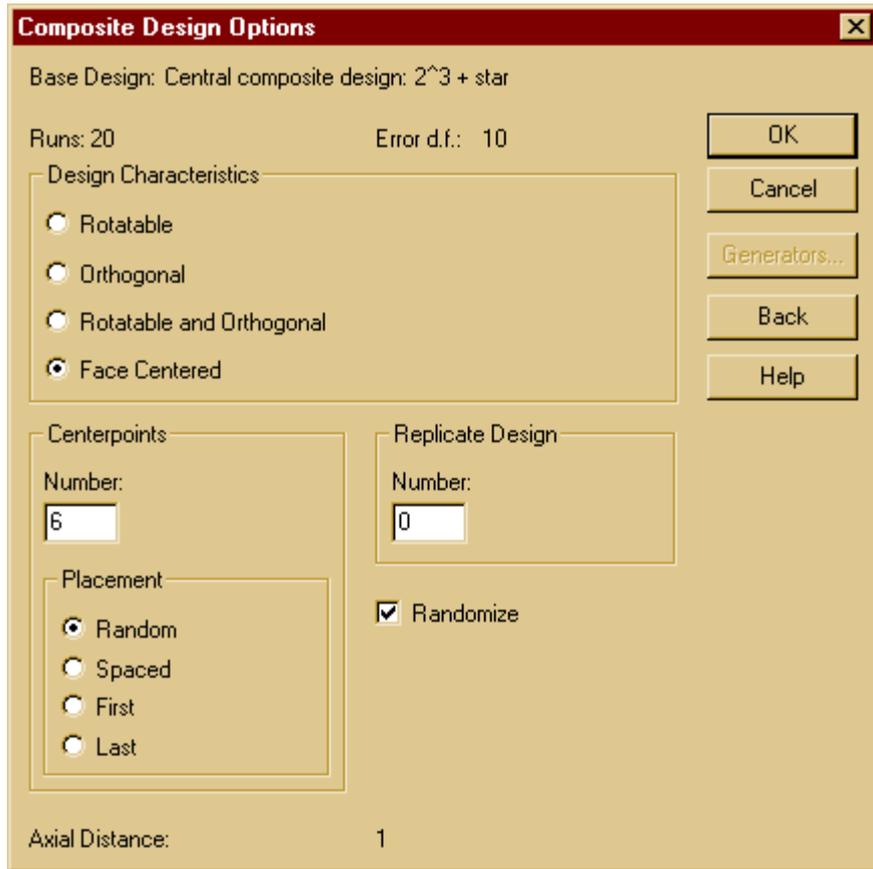
Create Design : création du plan



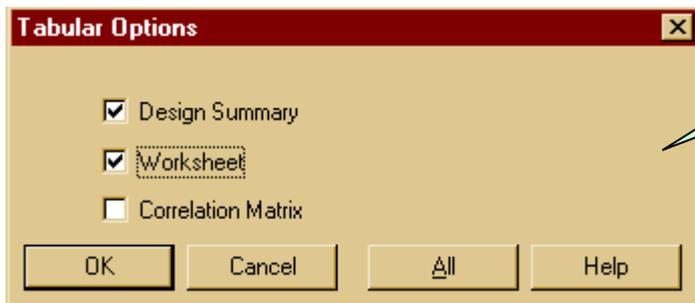
Définition des facteurs



Choix du modèle



Paramétrage du plan



Imprimer la feuille d'analyse

Analyse Design : Utilisation des données expérimentales pour l'obtention du modèle

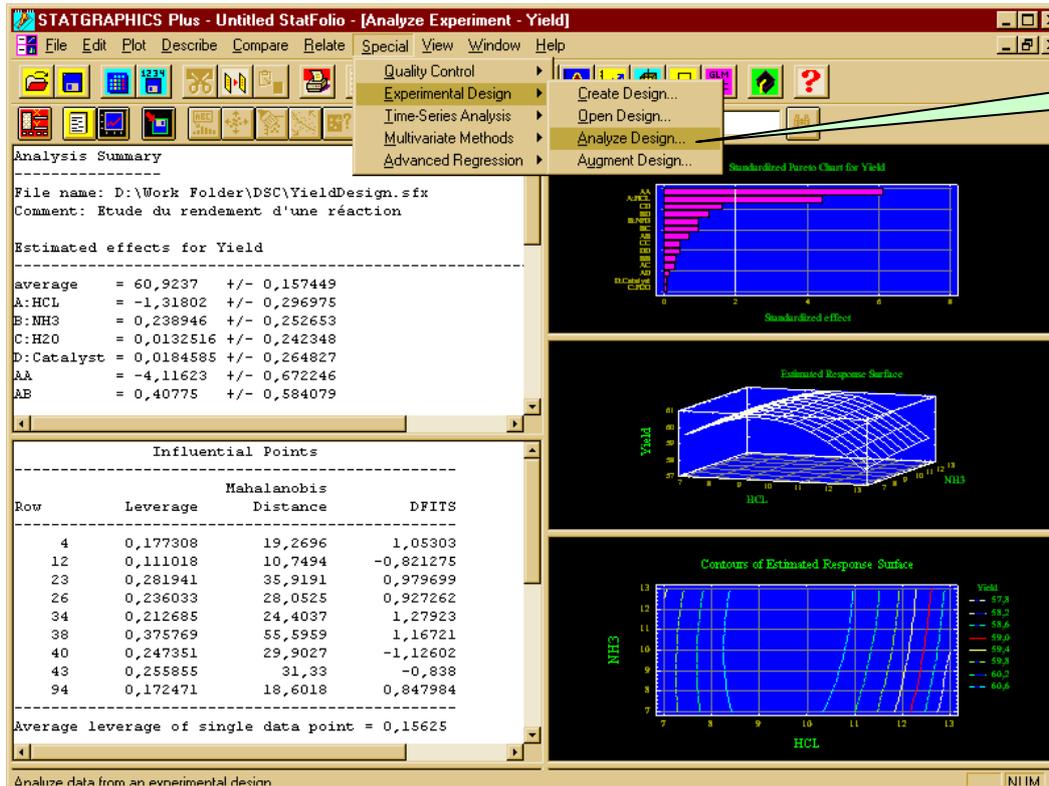


Figure 7: Pour obtenir la modélisation du domaine d'expériences, on procède à l'analyse du plan: option "Analyse Design" du menu "Special/Experimental Design".

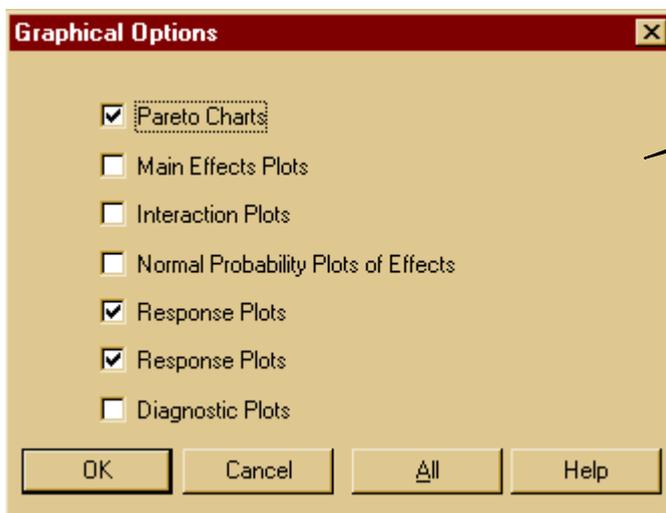


Figure 8: Affichage des principaux graphes intéressants issus de l'analyse d'un plan d'expériences.

3) Intérêt et lecture des principaux graphes et tableaux de données obtenus dans la plupart des logiciels

1) Résumé du plan d'expériences

Tous les logiciels fournissent un récapitulatif des paramètres étudiés. Voici ce récapitulatif sous Statgraphics Plus.

Design class: Response Surface

Design name: User-specified

File name: J:\Fourre tout temporaire\Plan experiences\Plan Exp 01.sfx

Comment: Plan d'experiences – optimisation d'un rendement

Base Design

Number of experimental factors: 3

Number of responses: 1

<i>Factors</i>	<i>Low</i>	<i>High</i>	<i>Units</i>	<i>Continuous</i>
<i>Toluene</i>	<i>5.0</i>	<i>20.0</i>		<i>Yes</i>
<i>Acylium</i>	<i>5.0</i>	<i>20.0</i>		<i>Yes</i>
<i>AlCl3</i>	<i>0.05</i>	<i>0.20</i>		<i>Yes</i>

Responses *Units*

Yield *%*

2) La matrice de corrélation du plan

Obtenu à partir des données expérimentales entrées par l'utilisateur lors de la définition du plan, la matrice de corrélation montre les corrélations existantes entre les colonnes de la matrice du plan d'expériences, c'est-à-dire entre les paramètres étudiés et leurs interactions d'ordre 1 et/ou 2. Un terme anglais revient souvent lorsqu'on étudie la matrice de corrélation : **confounding** ou **alias**. C'est une mesure de la redondance entre paramètres étudiés ou entre un paramètre et ses interactions d'ordre supérieur. Une forte corrélation entre 2 colonnes indique que l'une peut être entièrement expliquée par l'autre et réciproquement, aussi conserver les 2 devient inutile du point de vue information statistique.

Un plan orthogonal montrerait une matrice de corrélation diagonale (des 1 sur la diagonale de la matrice et des 0 partout ailleurs).

Une valeur non nulle ailleurs que sur la diagonale, indique que l'estimation des effets correspondant à cette ligne et cette colonne sont corrélés. Une valeur supérieure à 0.5 implique souvent des difficultés pour séparer les effets les uns des autres lors de l'analyse du plan. Lorsque cela est possible il peut alors être intéressant d'ajouter des expériences au plan afin de tenter de réduire ces corrélations.

3) Analyse du plan

a. Estimation des effets

La table *Estimation of effects* montre tous les effets calculés et leurs interactions ainsi que les écarts types associés (écart-type = mesure de l'erreur d'échantillonnage – *sampling error*).

L'importance de ces effets peut-être visualisée sous forme de graphe dans l'ordre inverse d'importance, c'est le but du diagramme de Pareto (Cf. diagramme de Pareto).

b. Analyse de la variance (ANOVA)

L'analyse de variance est résumée sous forme d'une table présentant de manière différenciée des informations sur la variabilité de la « variable réponse » étudiée et cela en fonction de chacun des effets.

Il est alors possible d'évaluer la significativité de chaque effet en comparant le carré de la moyenne à une estimation de l'erreur expérimentale. Il s'agit ici d'un test de Fisher dont la valeur est calculée par le logiciel et donnée pour chaque effet. ($F\text{-Ratio} = V_{intra} / V_{inter}$).

L'analyse de la variance permet aussi d'obtenir un paramètre important dans l'évaluation de la pertinence du modèle obtenu pour représenter les données expérimentales. Il s'agit du coefficient de détermination R^2 . Plus R^2 est proche de 1 et plus le modèle est adapté à la modélisation du domaine d'expérience choisi.

c. Coefficients de régression

Les coefficients de régression calculés par le logiciel sont ceux de l'équation du modèle. Selon le type de plan d'expériences choisi, on obtiendra plusieurs coefficients de régression (notés B) associés aux différents paramètres de l'étude ainsi qu'à leurs interactions d'ordre un (si le plan repose sur un modèle d'ordre 1 = polynôme ne faisant intervenir que la somme des paramètres de l'étude associés à leur coefficient de régression respectif ou la somme des produits de ces mêmes paramètres associés à d'autres coefficients spécifiques à chaque produit) ou d'ordre 2 si le plan repose sur un modèle faisant intervenir les paramètres et leurs interactions d'ordre 1 et 2.

d. Table des points influents

La table des points influents liste toutes les observations qui présentent un poids statistique ou « valeur levier » (*leverage value*) supérieur à 3 fois le poids statistique moyen calculé par observation. La valeur levier d'une observation mesure l'importance d'une observation dans la détermination des coefficients de régression du modèle choisi.

4) Les graphes et leur interprétation

a. Le diagramme de Pareto (*Pareto chart*)

Histogramme dont les barres sont proportionnelles à la valeur absolue de leurs effets estimés associés (ce sont les coefficients B) ou bien à leurs effets estimés standardisés. Il permet d'identifier les effets les plus influents sur la réponse étudiée.

b. Le diagramme des effets principaux (*main effects plot*)

Ce graphe affiche les facteurs primaires étudiés dans le plan. Il permet de voir comment évolue la réponse et les réponses étudiées dans le domaine d'expérience de chaque facteur, donc entre la valeur basse et la valeur haute du facteur considéré.

On peut ainsi mieux percevoir les effets antagonistes.

c. Le diagramme des interactions (*Interactions plot*)

Ce graphe montre l'influence de la combinaison de deux facteurs sur la réponse étudiée.

d. Le diagramme de la loi normale (*Normal probability plot of effects*)

Ce diagramme montre la distribution cumulative des effets. Le graphe affiche les effets standardisés. L'axe des X pour chaque point, correspond à l'effet estimé du facteur. Sur ce graphe, l'allure de la distribution normale prend la forme d'une droite.

Ce graphe permet de distinguer les effets du bruit spécialement lorsque le plan d'expériences n'est pas répliqué. Les points sur le graphe se répartissent de telle sorte que tous les effets

correspondant à du bruit se répartissent de manière aléatoire autour de la droite, c'est-à-dire approximativement le long de la droite ou sur la droite.
Les effets correspondant réellement à un signal distinct du bruit se trouvent relativement éloignés de la droite.

e. Surface de réponses

i. Surface 3D

Représentation en 3 dimensions (c'est-à-dire selon 3 facteurs) du modèle obtenu.

ii. Contour 2D

Représentation sous forme de graphe iso-réponse (même principe que les cartes géographiques qui sont des iso-courbes d'altitude)

Ces 2 types de graphes permettent de visualiser l'optimum ou les optima d'un plan d'expérience selon les facteurs sélectionnés.

f. Le diagramme de diagnostic (Diagnostic plot)

Ce diagramme permet de déterminer si le modèle calculé représente bien les données.

4) Les principaux logiciels de plan d'expériences

JMP	http://www.jmpdiscovery.com
Minitab	http://www.minitab.fr
Statistica	http://www.intesoft.com/produits/tech/statistica
Statgraphics	http://www.sigmaplus.fr
Unscrambler	http://www.camo.no
Pirouette	http://www.infometrix.com
Modde	http://www.umetrics.com