Planification d'expériences numériques

UP 4 - Exploitation mathématique de simulateurs numériques

Victor Picheny
INRA (Institut National de la Recherche Agronomique)
victor.picheny@toulouse.inra.fr

Déroulement

- Cours : 6h
  - 28/01 : plans géométriques
  - 29/01 : plans « orientés modèle »
- TP / Projet : 9h

- Support : présentations

- Évaluation : rapport de projet + présence & participation

Quelques repères historiques

- Fisher : Design of Experiments (1935)
- Kiefer : Optimum experimental designs (1959)
- Fedorov : Theory of optimal experiments (1972)
- McKay : Latin Hypercube Sampling (1979)
- Sacks : Design and analysis of computer experiments (1989)

- Traditionnellement : pharmaceutique, agronomie, procédés...

- Depuis 20+ ans : **expériences numériques**

Un exemple introductif... agricole

- Objet d'étude : plantation de **tournesol** sur une exploitation
- Actions possibles :
  - Placement : parcelle plus ou moins ensoleillée, qualité du sol
  - Arrosage, engrais, …
- Mesure : récolte (kg/an)
- Quels objectifs ?
  - Apprendre comment obtenir la meilleure récolte possible
  - Rechercher les variables influentes
  - Déterminer les relations entre variables et production pour choisir un compromis satisfaisant (prix, temps passé, …)
- Aléas :
  - Météo : ensoleillement, pluviométrie
  - Variabilité entre individus

Formalisation

Variables contrôlées

- Ensoleillement
  oui / non
- Engrais
  peu / beaucoup
  ***ou** kg*
- Sol
  type 1 / type 2
- Arrosage
  faible / moyen / fort
  ***ou** L / semaine*

Système
(plante)

Observation
Production (kg)

Variables non contrôlées

- Ensoleillement
- Pluviométrie
- Qualité de l'individu

Formalisation

Facteurs d'entrée
- Ensoleillement
- Sol
- Engrais
- Arrosage

Système
(plante)

Observation
Production (kg)

Facteurs non contrôlés
- Ensoleillement
- Pluviométrie
- Qualité de l'individu

Planification d'expériences :

- On dispose de *N* plans
- Quelles expériences choisit-on ?
- Comment exploiter les observations ?

Question subsidiaire : combien de plans pour chaque expérience ?

Embryons de solutions

- Quelques solutions intuitives
  - Un facteur à la fois
  - Toutes les combinaisons possibles
  - « Un peu de tout »

Plans géométriques

→ Plans remplissant l'espace

- Ajout d'information *a priori* :

  « toutes les variables contribuent linéairement à l'amélioration de la production »

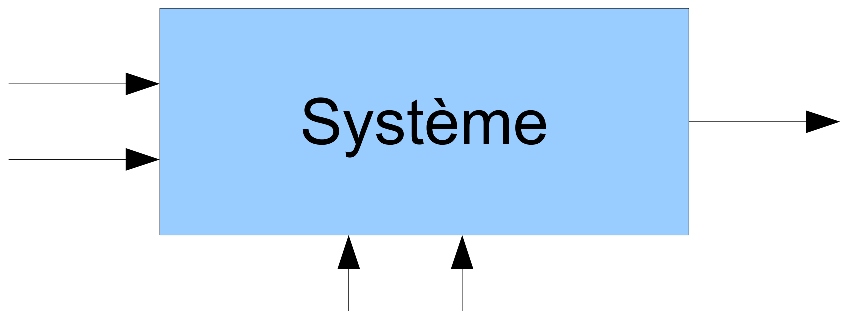
  - Modèle : $y=\beta_0+\beta_S\times S+\beta_P\times P+\beta_A\times A+\beta_E\times E+\epsilon$
  - Comment obtenir le modèle le plus précis possible ?

→ Plans optimaux

Vocabulaire et notation

Facteurs d'entrée

$\mathbf{x} = \{ x^1, ..., x^d \}$



Réponse(s)

$y(\mathbf{x}) \quad (+ \varepsilon\ ?)$

Facteurs non contrôlables

Plan d'expériences = un jeu de *N* facteurs d'entrée + les réponses (observations) correspondantes

$$\mathbf{X} = \{\mathbf{x}_1 , ..., \mathbf{x}_N\} \qquad \mathbf{Y} = \{y_1 , ..., y_N\}$$

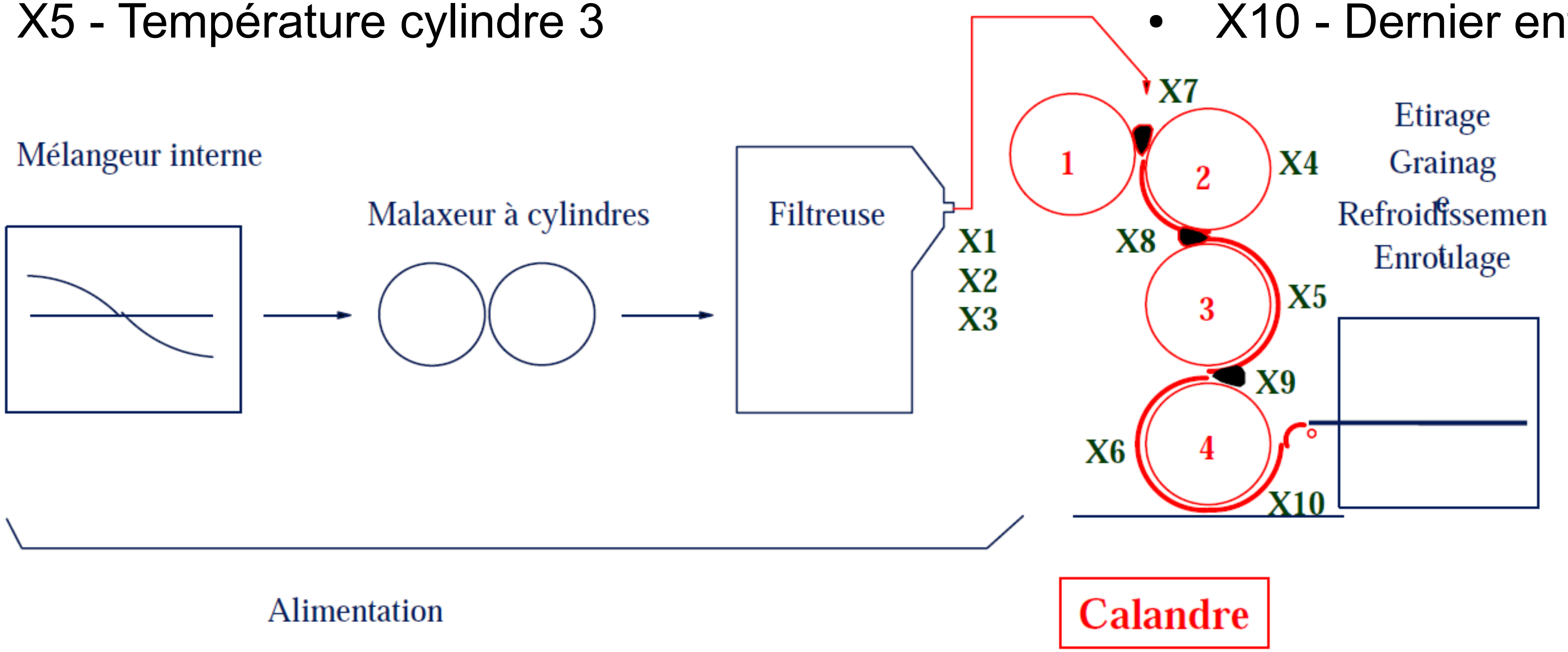
Quelques exemples où la planification d'expériences est nécessaire

Expériences « classiques »

- Matériaux : effet d'un traitement sur les propriétés physiques
  - Avec / sans traitement
  - Influence de certains paramètres (température, …)
- Pharmaceutique :
  - Effets de médicaments sur différentes populations
  - Effets combinés de traitements
- Production : amélioration de la qualité
  - Ex. : Etude d'un pistolet à peinture
  - Deux paramètres : ouverture et pression

Exemple : fabrication de feuilles de PVC

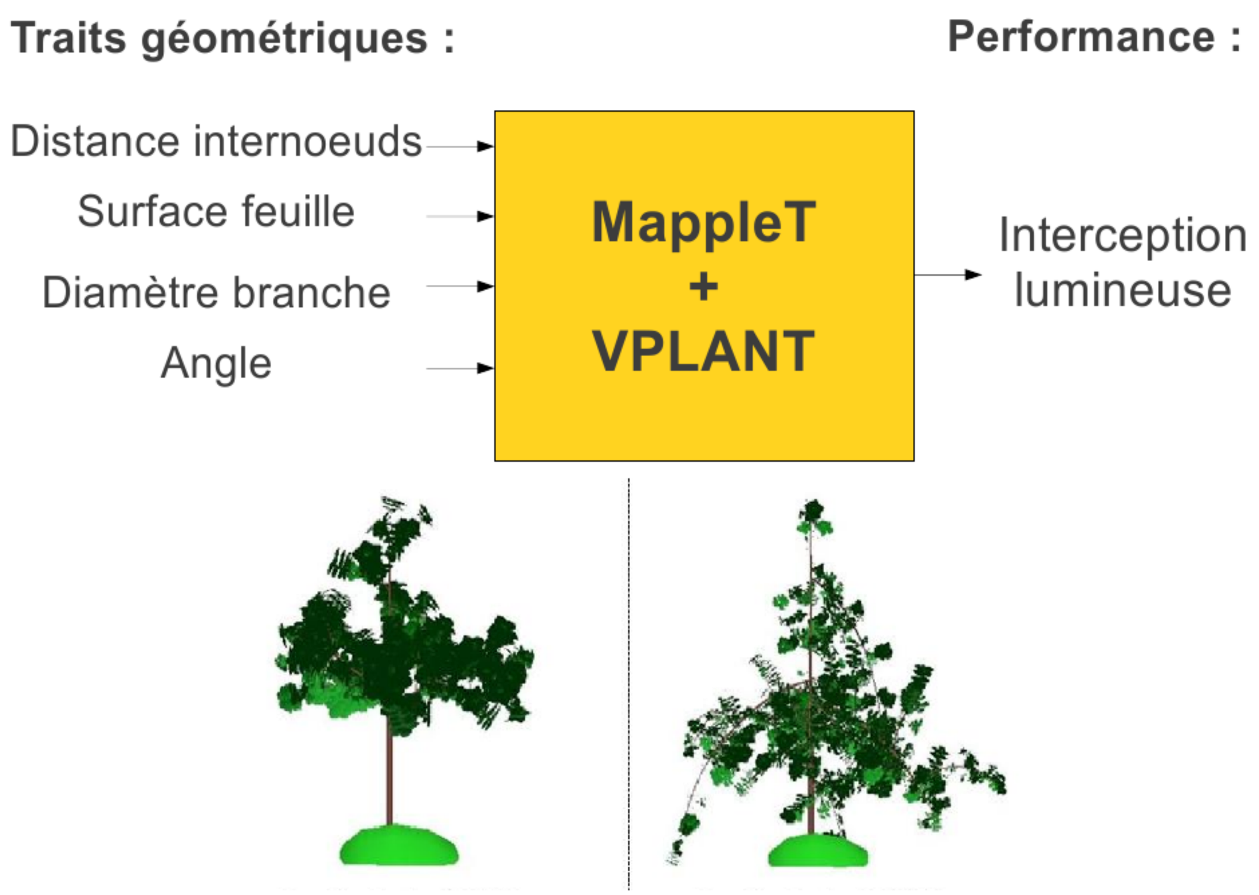
- X1 - Température matière entrée calandre
- X2 - Ouverture de la filière
- X3 - Débit matière à l'entrée
- X4 - Température cylindres 1 et 2
- X5 - Température cylindre 3
- X6 - Température cylindre 4
- X7 - Diamètre du bourrelet 1
- X8 - Diamètre du bourrelet 2
- X9 - Diamètre du bourrelet 3
- X10 - Dernier entrefer



- Objectif : améliorer la qualité (absence de bulles d'air, résistance)
- 1 essai = 1 jour, 2500 €
- Essayer toutes les combinaisons pour 2 valeurs de chaque Xi: 1024 essais = 3 ans

Source : B. Govaerts

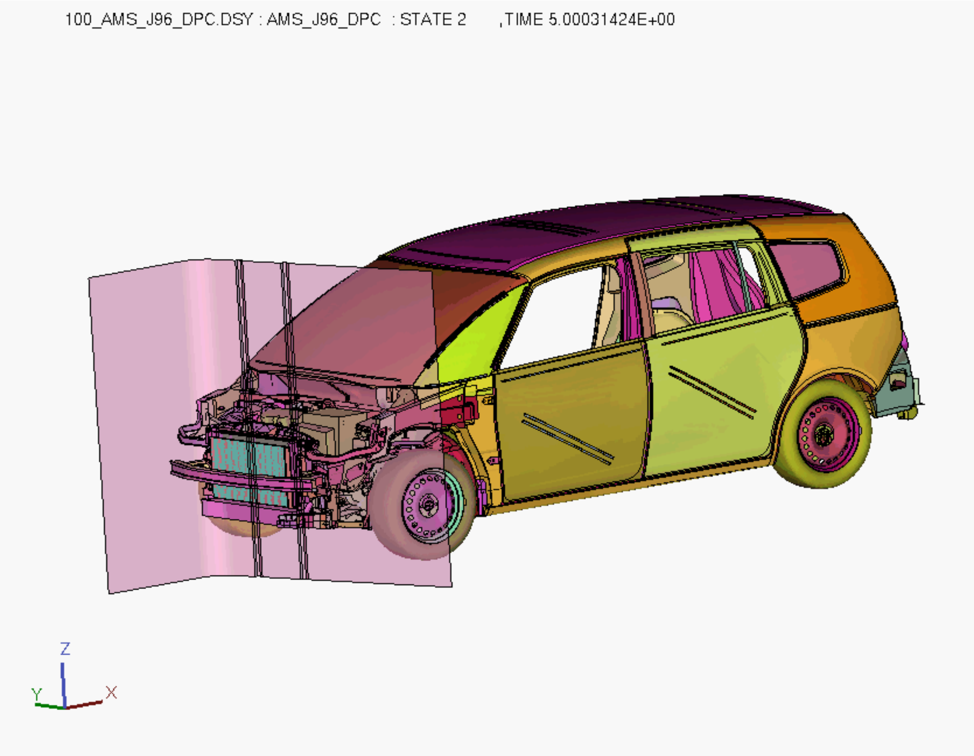
Exemple numérique : arboriculture

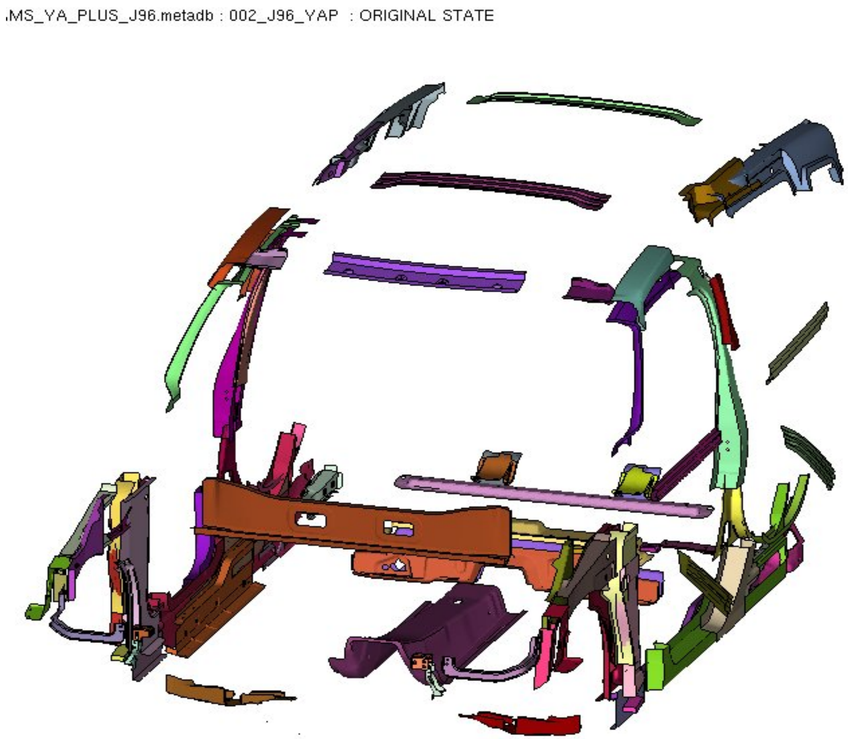


- Un calcul = 45 min
- Objectif : mesurer influence des traits sur l'interception

Exemple num. 3 : Crash test (60D)

- 53 épaisseurs de pièces de la caisse
- 7 facteurs présence / absence de pièces
- Objectif : meilleur compromis
  masse / prestations (optimisation)
- Temps de calcul : ~ 12 / 24h





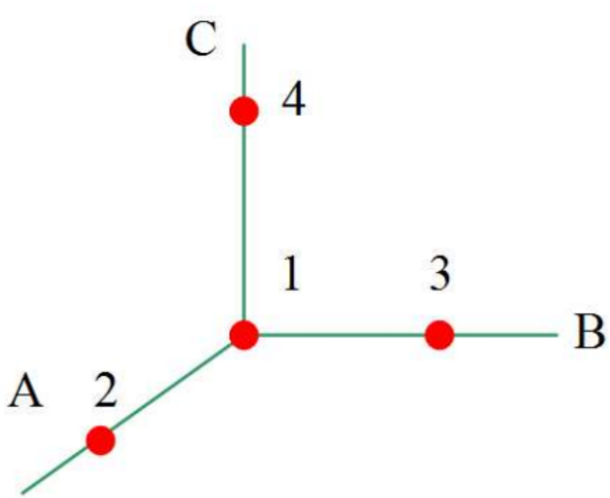
Essayer toutes les combinaisons pour 2 valeurs de chaque facteur :
2^60 essais >> extinction du Soleil

Planification : quelles objectifs ?

II- Plans pour facteurs discrets

Approche intuitive : « un à la fois »

- On cherche à savoir si un facteur a une influence sur la réponse
- On fixe tous les autres facteurs à leur valeur de référence
- On effectue une expérience de référence, et une pour la variation de chaque facteur étudié
- d + 1 expériences



exp	A	B	C
1	0	0	0
2	1	0	0
3	0	1	0
4	0	0	1

Exercice

- On considère deux facteurs A et B variant sur [0, 1]
- Déterminez la matrice d'expériences pour un plan en étoile
- Quel(s) modèle(s) de régression pouvons-nous utiliser ?
- Calculez les coefficients pour un modèle linéaire
- Essayez en ajoutant un terme d'interaction : que se passe-t-il ?

On rappelle que : $\beta = (F^T F)^{-1} F^T Y$

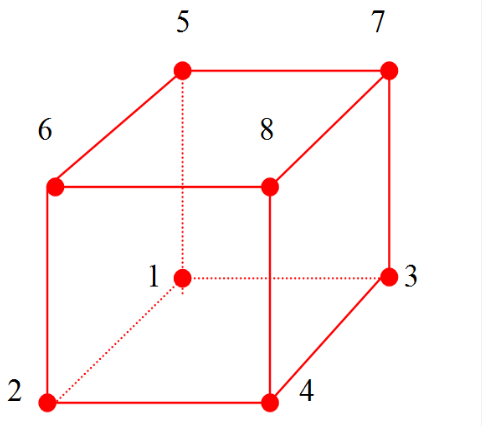
Plan « un à la fois »

- Lecture immédiate de l'effet d'un facteur :

$$y=\mu+\alpha_i+\beta_j+\gamma_k+\ldots+\epsilon$$

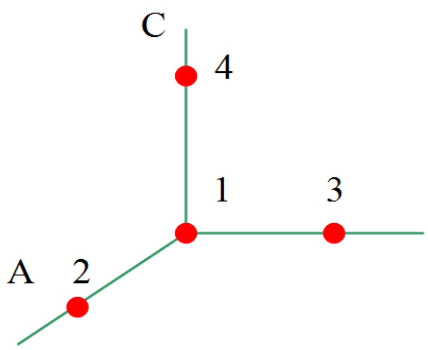
- Seules 2 expériences à la fois servent à mesurer un effet
- Pas de visibilité des interactions

Plan factoriel



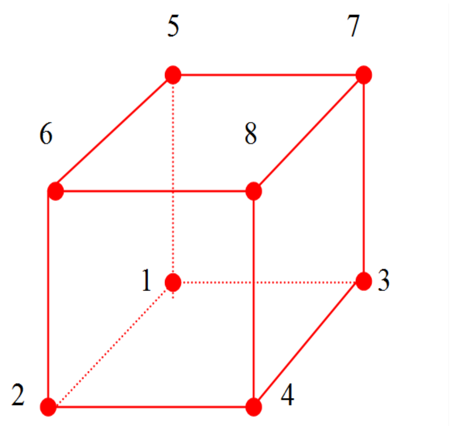
Exercice : mêmes questions que précédemment

Que peut-on voir ?



Effets linéaires d'un seul facteur

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + e$$



Effets principaux et interactions

$$y = a_0 + \Sigma\, a_ix_i + \Sigma_{i > j}\, b_{ij}\, x_ix_j + c\; x_1x_2\, x_3 + e$$

Termes « oubliés » : $\Sigma\, d_ix_i^2 + \ldots$

Permet une **analyse de sensibilité** (criblage) : détection des facteurs les plus importants

Plan factoriel

- Meilleur plan si effets linéaires avec interactions (cf. Partie III)
- 2^d expériences : rapidement impossible
- Peut-on « voir » des informations utiles avec moins d'expériences ?
- Quel prix à payer ?

→ Plans factoriels fractionnaires

Plans factoriels fractionnaires

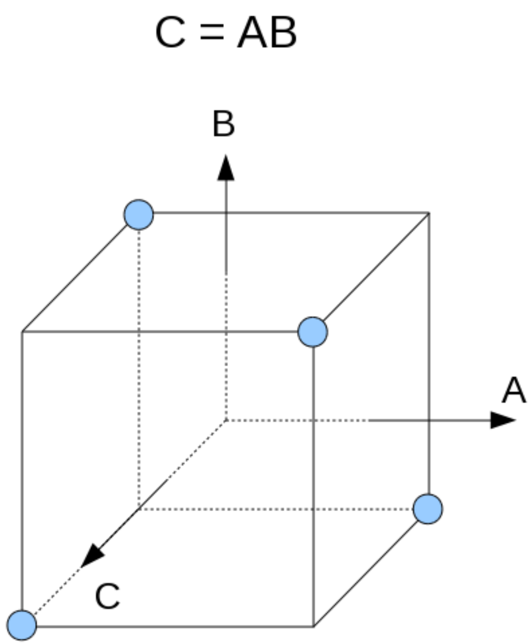
- Sous-ensemble d'un plan factoriel

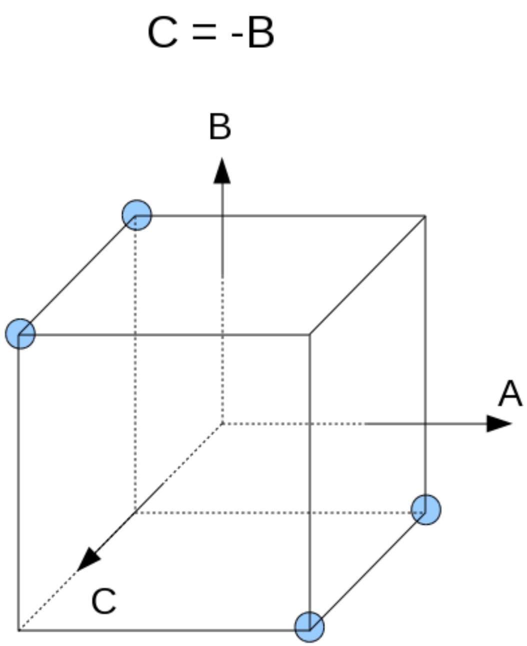
  Objectif : **réduire le nombre d'expériences**

- Fraction *régulière* : 2^{n-h} éléments
  - Fraction 1/2 : 2^{n-1}, fraction 1/4 : 2^{n-2}

- Construction : sélectionner 2^{n-h} éléments parmi 2^nn => *relations d'alias*

Fraction 1/2





exp.	A	B	C=AB
1	-1	-1	+1
2	-1	+1	-1
3	+1	-1	-1
4	+1	+1	+1

Confusion des effets

- En cas d'interactions :

$$y=\beta_0+\beta_A A+\beta_B B+\beta_C C+\beta_{AB} AB+\beta_{AC} AC+\beta_{BC} BC+\beta_{ABC} ABC+\varepsilon$$

- Tous les effets ne sont pas identifiables (4 obs., 8 coef.)
- On observe cependant :
  - A = BC
  - B = AC
  - C = AB
  - ABC = 1

exp.	A	B	C	AB	AC	BC	ABC
1	-1	-1	+1	+1	-1	-1	+1
2	-1	+1	-1	-1	+1	-1	+1
3	+1	-1	-1	-1	-1	+1	+1
4	+1	+1	+1	+1	+1	+1	+1

En fait : ABC = 1 implique les autres relations (fonction génératrice d'aliases)

- On obtient alors le modèle :

$$y=\beta_0(1+ABC)+\beta_1(A+BC)+\beta_2(B+AC)+\beta_3(C+AB)+\varepsilon$$

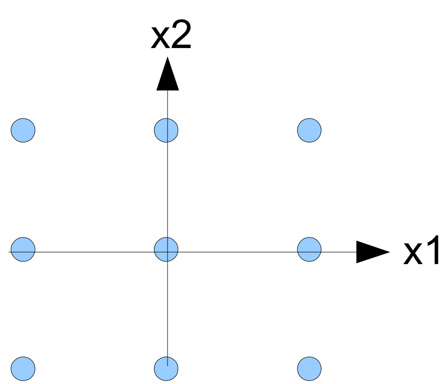
Utilité des plans fractionnaires

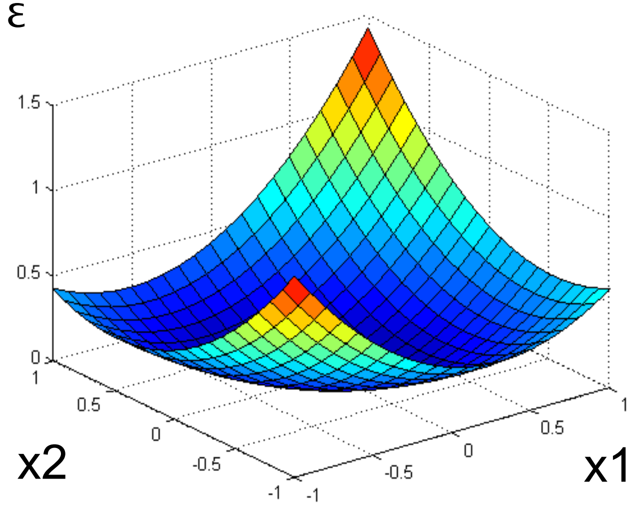
- Permet de gérer un grand nombre de facteurs
- Utile si peu de facteurs sont influents / beaucoup d'interactions peuvent être négligées

Facteurs à *k* niveaux

- Plans en étoile / factoriels / fractionnaires s'appliquent
- Plan factoriel à 3 niveaux : effets quadratiques

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_{11} x_1^2+\beta_{22} x_2^2+\beta_{12} x_1 x_2+\varepsilon$$





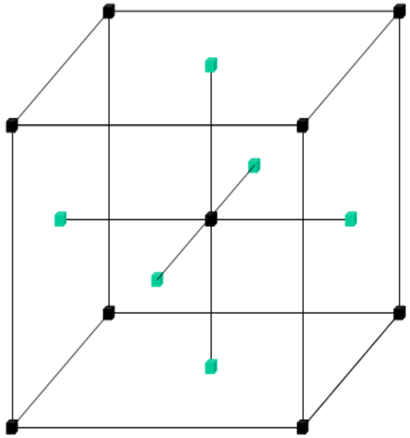
- Nombre d'expérience très élevé : $N = k^d$

Plans composites

- Moins grand qu'un plan factoriel à 3 niveaux
- Plan factoriel à 2 niveaux + 2*d point situés sur les axes à une distance α de l'origine + l'origine : $2^d + 2*d + 1$
- Modèle (1+2*d + d(d-1)/2 termes) :

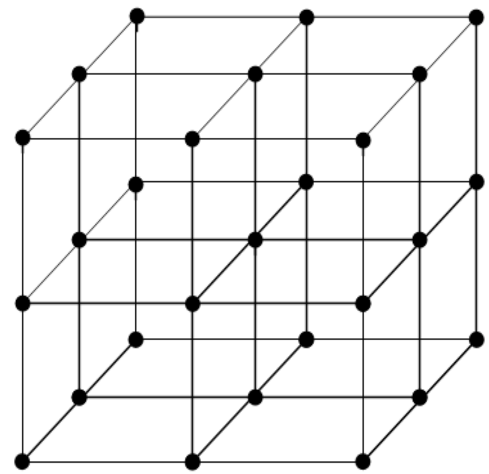
$$y = \beta_0 + \sum_i \beta_i x_i + \sum_i \beta_{ii} x_i^2 + \sum_{i>j} \beta_{ij} x_i x_j + \varepsilon$$

- Utilisable en petite dimension seulement !



Problèmes de projection

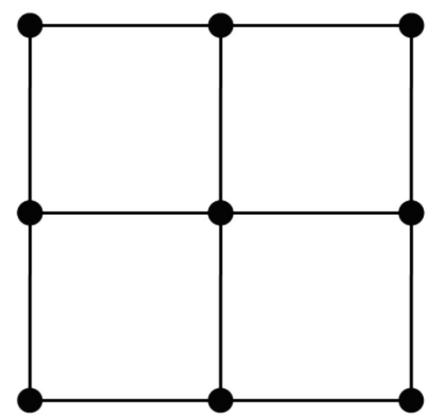
- Plan factoriel à 3 niveaux avec 3 facteurs :



- Après criblage : un des facteurs n'est pas influent

➡ On passe de 27 à 9 observations

➡ On ne pourra pas augmenter la complexité du modèle



Récapitulatif

- En Etoile : à éviter !
- Factoriels : coûteux
- Composite : coûteux
- Factoriels fractionnaires : OK en grande dimension mais confusion des effets

+	-
• Simplicité d'utilisation • Adaptés aux variables continues et discrètes • Adaptés aux modèles de régression simple et à l'ANOVA	• Très coûteux en grande dimension • Problèmes de projection • Pas de valeurs intermédiaires • Nb d'expériences rigide