

# Étude de la structure d'un champ d'orge du service expérimental de la Minière

MONESTIEZ P. (1)

## INTRODUCTION

Le but de cette étude est de présenter une méthode d'analyse de structure basée sur la théorie des variables régionalisées.

Pour une première approche, nous avons pensé préférable de traiter un exemple; un exemple simple par les faibles moyens nécessaires à son traitement, mais ayant l'avantage de montrer les difficultés nées de la confrontation avec la réalité.

La méthode utilisée est connue, même si elle l'est surtout dans d'autres domaines et une présentation théorique n'aurait pas soulevé les problèmes propres aux plans d'expérience en agronomie.

Nous espérons que, par un exemple, les concepts de base seront mieux perçus, même si nous nous restreignons à une des multiples applications de la théorie.

## I – DÉFINITION DE L'ÉTUDE ET DE SON CONTEXTE

Préalablement, on cherche à estimer l'erreur apportée par l'hétérogénéité du terrain dans les plans d'expérience classiques en agronomie.

Cette étude passe obligatoirement par une approche spatiale des phénomènes car des variations non corrélées spatialement pourraient être traitées par les méthodes classiques.

L'utilisation des blocs dans les plans d'expérience prouve que les expérimentateurs ont conscience de ces problèmes d'hétérogénéité. Ceux-ci n'étant pas étudiés systématiquement, on a le plus souvent recours à des notions empiriques; cependant, ces plans d'expérience affinent déjà l'analyse par la prise en compte de l'hétérogénéité dans l'espace suggérée par des caractéristiques physiques du terrain.

Le problème que nous nous donnons à résoudre est de mettre au point des méthodes nous permettant d'appréhender suffisamment les corrélations spatiales afin, dans une première étape, de pouvoir déterminer de façon moins empirique, forme, taille, disposition et type du dispositif expérimental et, dans une seconde étape plus lointaine, de faire intervenir la connaissance du processus au niveau de l'analyse statistique elle-même.

Ce projet, déjà vaste, nécessite avant tout la connaissance des processus spatiaux que l'on rencontre en agronomie. Bien que l'on se restreigne à des modèles simples et à l'étude d'une seule fonction (le demi-variogramme), l'étude des processus spatiaux soulève de nombreuses difficultés par la faiblesse de l'information: la réalisation est unique, les structures sont complexes, le champ est petit et, surtout, le nombre de points où la variable est connue est très faible. C'est pour cela que l'on se rapprochera de la géostatistique qui, pour résoudre les problèmes d'estimation minière (problèmes d'étude de structure spatiale analogues aux nôtres), possède des outils mathématiques efficaces.

(1) C.N.R.F., Station de Biométrie, CHAMPENOUX 54370 EINVILLE.

## II – PRÉSENTATION DES BASES DE LA GÉOSTATISTIQUE

D'un point de vue théorique, on considère l'effet terrain de notre exemple comme une réalisation  $y(x)$  d'une fonction aléatoire  $Y(x)$  dans l'espace. Comme on ne dispose que d'une seule réalisation, l'inférence statistique n'est possible qu'avec une hypothèse supplémentaire.

L'hypothèse généralement adoptée est :

-  $Y(x)$  est une fonction aléatoire stationnaire d'ordre 2 définie par :

$Y(x_0)$  admet une espérance indépendante de  $x_0$  et  $\gamma h$

$C(h) = E[Y(x_0 + h) Y(x_0)] - m^2$  existe et ne dépend pas de  $x_0$ .

Cependant, il existe des cas réels où  $C(h)$  n'est pas définie alors que les accroissements  $Y(x_0 + h) - Y(x_0)$  ont une variance finie (mouvement Brownien). La géostatistique utilise donc une hypothèse plus faible appelée hypothèse intrinsèque

$$E[Y(x + h) - Y(x)] = m(h) \text{ indépendante de } x$$

$$E\{[Y(x + h) - Y(x)]^2\} = 2\gamma(h) \text{ existe et est indépendant de } x$$

La fonction associée à cette hypothèse est le demi-variogramme  $\gamma(h)$  que nous allons estimer à partir des données expérimentales.

$$\gamma^*(h) = \frac{1}{V} \int_V (Y(x + h) - Y(x))^2 dx$$

Lorsque nous disposons d'un maillage régulier de l'espace, l'intégrale précédente devient une somme discrète. Nous estimons le demi-variogramme pour des valeurs de  $h$  correspondant au pas du maillage et à ses multiples.

Nous utiliserons le demi-variogramme de préférence à la covariance car l'estimation de cette dernière peut apporter un biais très grave.

Remarque : par abus de vocabulaire, on dit le plus souvent «variogramme» à la place de «demi-variogramme». Dans la suite, nous parlerons toujours des demi-variogrammes en utilisant la notation  $\gamma(h)$  et le mot «variogramme».

## III – APPLICATION A L'EXEMPLE

### 1 – Présentation des données. (fig. 1)

L'expérience consiste en un «essai à blanc» d'orge betina réalisé en 1975 au Service d'Expérimentation de la Minière. Les prélèvements, récoltés sur une parcelle de 1 m sur 2,50 m, sont disposés selon une maille régulière de 7,50 m dans les deux directions. Le nombre de prélèvements est 153 (17 x 19) et la variable mesurée est le poids de grains en décagrammes pour 2,50 m<sup>2</sup>. Trois axes de parcelles mitoyennes (100 mesures dans la direction est-ouest, 2 x 27 mesures dans la direction nord-sud) nous permettent de travailler sur une échelle plus fine.

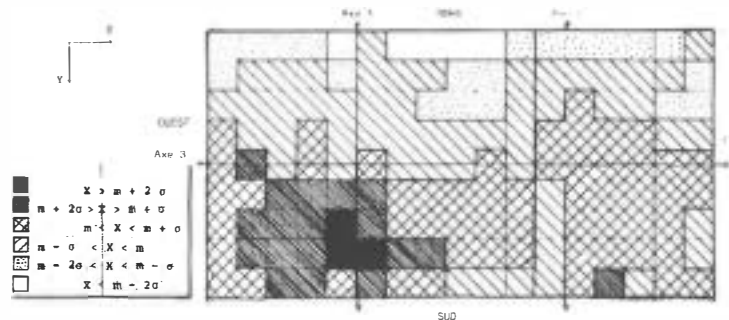


Figure 1 – CARTE DES DONNÉES. Chaque carré fait 7,5 m de côté, il est estimé par un prélèvement central.

## 2 — Analyse des résultats expérimentaux (fig. 2)

Nous nous bornerons ici à décrire quelques principes permettant l'analyse. La notion d'échelle est primordiale. La croissance très rapide au voisinage de l'origine ( $h$  petit) donne une idée des irrégularités de la fonction aléatoire. La croissance lente et régulière indique la liaison spatiale et la portée (distance à laquelle le palier est atteint) est la distance après laquelle il n'y a plus dépendance. La valeur du palier est un estimateur de la variance du phénomène. On rencontre aussi deux

types d'anisotropie : premièrement, la portée dépend de la direction, deuxièmement, des effets unidirectionnels se rajoutent à des phénomènes isotropes. Cela représente autant de caractéristiques qu'il faut déduire des courbes expérimentales. Cependant, on sera aidé par des études plus fines, sur des zones pour avoir une idée de la stabilité, sur les axes pour déterminer le début de la courbe, sur les diagonales pour vérifier si les hypothèses d'anisotropie ou d'isotropie sont vérifiées. C'est l'ensemble de cette analyse qui permettra la modélisation.

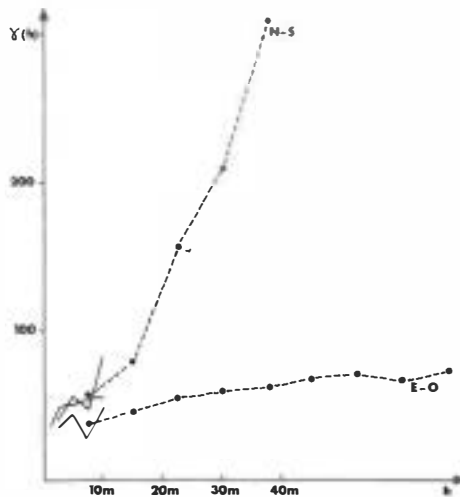


Figure 2 — VARIOGRAMMES EXPÉRIMENTAUX. NS variogramme expérimental NORD— SUD, EO variogramme expérimental EST—OUEST, traits pleins : début des variogrammes expérimentaux sur les axes.

## 3 — Modélisation de la structure

### a — Limites de cette modélisation

Il est évident que modéliser un processus à partir du variogramme seul, ferait appel à des hypothèses bien trop fortes pour pouvoir être vérifiées sur les quelques données que l'on a. De plus, deux processus différents peuvent très bien avoir le même variogramme.

Pour ne pas tomber dans ce piège, nous nous limiterons à la connaissance du variogramme.

Modéliser le variogramme reviendra à ajuster au variogramme expérimental une courbe théorique; ce sera celui d'un processus. (La seule condition est que la fonction soit de type positif conditionnel).

Apparemment, l'étude n'a pas fait un grand pas dans la connaissance du processus sous-jacent. Pourtant, l'ajustement d'une courbe continue sur une courbe estimée en certains points seulement permettra le calcul d'intégrales dont on tirera un grand nombre de résultats.

L'intérêt de cette méthode est de tirer le maximum des données sans faire d'hypothèses sur la définition du processus.

### b — Recherche d'un variogramme théorique

La méthode utilisée pour ajuster une courbe théorique au variogramme expérimental s'appuie sur le fait que l'on peut décomposer un modèle complexe en une somme de modèles simples seulement définis par quelques paramètres. Le variogramme final, ainsi que les variances calculées seront la somme d'expressions plus simples.

Nous ne rentrerons pas dans les problèmes pratiques d'ajustement, mais il est bon de noter qu'il existe une panoplie suffisamment abondante de modèles et que des ajustements plus grossiers, comme des fonctions linéaires, mènent à des résultats pratiques peu différents.

Dans l'exemple, nous avons choisi un modèle se décomposant en trois termes (fig. 3) :

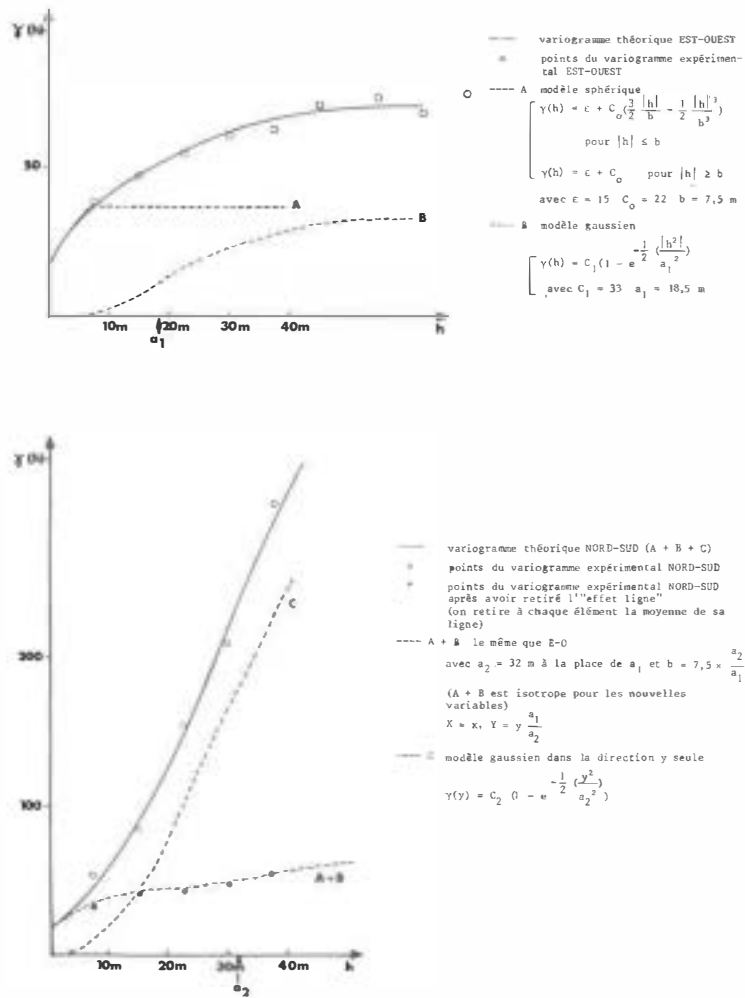


Figure 3 — MODELES CHOISIS COMPARES AUX VALEURS EXPERIMENTALES.

- un schéma sphérique (A) } isotropes pour  $X = x$
- un schéma gaussien }  $Y = \frac{a_1}{a_2} y$
- un terme indirectionnel ne dépendant que de y

$$\sigma^2(v|V) = \overline{\gamma_V(x, y)} - \overline{\gamma_V(x, y)}$$

$$\text{ou } \overline{\gamma_V(x, y)} = \frac{1}{V^2} \iint_V \gamma(x - y) dx dy$$

x et y deux points se déplaçant indépendamment.

#### 4 — Définition de la variance v dans V

Cette variance, notée  $\sigma^2(v|V)$  est la variance d'un échantillon v dans un domaine V. Elle n'est définie de façon rigoureuse que lorsque le domaine V apparaît comme la réunion de volumes  $v_i$  disjoints tous égaux à v et translétés les uns des autres.

Si on définit la variance  $s^2(v|V)$  comme la variance des valeurs des  $v_i$  à l'intérieur de V pour une réalisation fixée,  $\sigma^2(v|V)$  n'est autre que l'espérance de  $s^2(v|V)$  obtenue en déconditionnalisant par rapport à la réalisation.  $\sigma^2(v|V)$  s'écrit très simplement en fonction de  $\gamma$  lorsque les  $v_i$  et V ont une forme simple.

#### 5 — Application à des plans d'expérience

Si  $\hat{\alpha}_t$  est l'estimateur de l'effet traitement t, nous pouvons exprimer la variance due à l'hétérogénéité des parcelles de cet estimateur en fonction des  $\sigma^2(v|V)$  pour des plans d'expérience classiques.

Soit :

- v la parcelle unitaire du plan d'expérience,
- $V_B$  le bloc, V le champ,
- b le nombre de blocs,
- T le nombre de traitements.

Dispositif en blocs randomisés

$$\text{var} \hat{\alpha}_t = \frac{1}{b} \sigma(v|V_g)$$

Randomisation totale

$$\text{var} \hat{\alpha}_t = \sigma(v|V)$$

Carré latin

$$\text{var} \hat{\alpha}_t = \frac{1}{(T-1)} (\sigma(O|V_c) + \sigma(O|V_c) - \sigma(O|v) - \sigma(O|V))$$

$$\text{où } \sigma(O|V) = \gamma \sqrt{V(x, y)}$$

Supposons que nous ayons un dispositif expérimental comparant 10 traitements à implanter sur le champ précédemment étudié (on suppose que la structure est la même). Nous pouvons calculer les variances correspondant à différents plans ou types de découpage (fig. 4).

Nous avons laissé dans les expressions la décomposition de la variance en trois termes correspondant à ceux du variogramme. Nous pouvons ainsi voir l'influence;

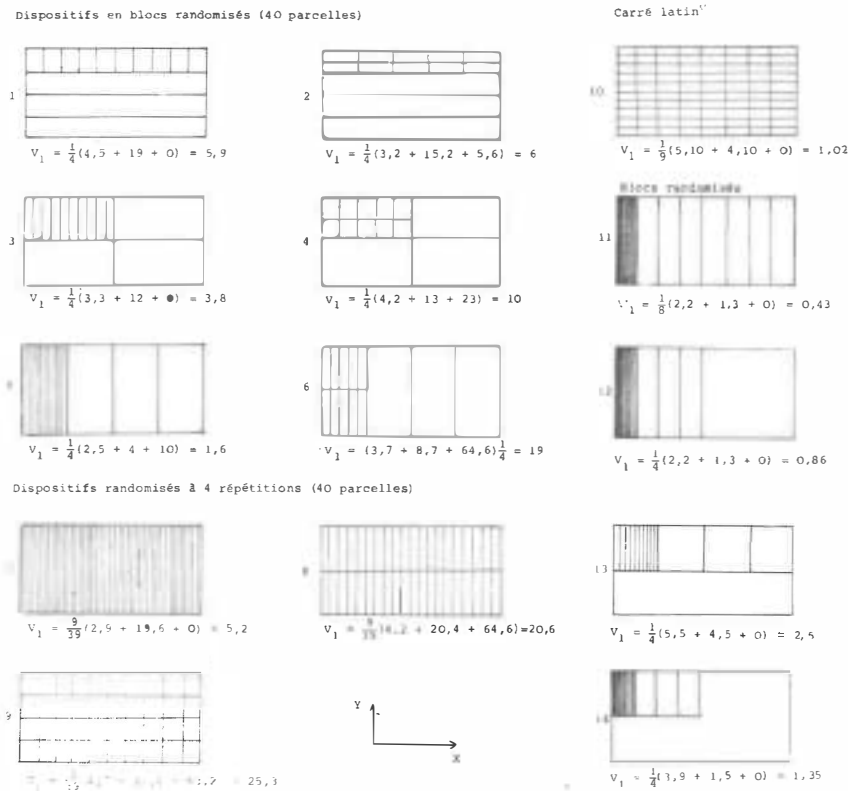


Figure 4 – VARIANCES CORRESPONDANT A DIFFÉRENTS PLANS D'EXPÉRIENCES OU TYPES DE DÉCOUPAGE.

— champ de (120 x 60) m ; 10 traitements.

1 à 6 : dispositifs en blocs randomisés (4 blocs)

7 à 9 : dispositifs randomisés à 4 répétitions

10 : carré latin

11 : blocs randomisés (8 blocs)

12 et 13 : blocs randomisés sur la moitié de la surface (4 blocs)

14 : blocs randomisés sur le quart de la surface (4 blocs)

— du terme A (variogramme à faible portée),

— du terme B (variogramme à longue portée),

— du terme C (effet unidirectionnel à fort coefficient).

Les résultats nous montrent immédiatement que :

— l'effet du 3ème terme doit être annulé du fait de son coefficient (prendre la même dimension pour v et  $V_g$  dans la direction y),

— on a avantage à choisir des parcelles de forme allongée,

— l'augmentation de la taille du champ n'a pas toujours un effet bénéfique.

Ces conclusions ne sont nullement générales; elles dépendent énormément de la structure et de sa portée. Il faut faire un compromis entre la variance, la taille du champ et les difficultés techniques. Un champ ayant une variance un peu supérieure au minimum réalisable est peut-être dix fois moins coûteux et grandement avantageux.

### CONCLUSION

Nous avons essayé de montrer comment la connaissance de la structure pouvait nous donner des

précisions importantes sur le choix du plan et surtout sur la forme et la taille des parcelles. Le problème qui reste est la stabilité de cette structure dans le temps (la réalisation pouvant être très différente) afin d'estimer *a priori* la variance de  $\hat{\alpha}_t$  du plan choisi.

Il est probable que les paramètres de la structure seront modifiés d'une année à l'autre, mais il suffirait que les grandeurs caractéristiques restent inchangées pour permettre un choix moins empirique et un gain d'information non négligeable.

