

Jean-Yves LAURENT  
ORSTOM, Fort-de-France

Jean-Pierre ROSSI  
ORSTOM, Bondy

## STRATÉGIES D'ÉCHANTILLONNAGE AU CHAMP. DÉTERMINATION DU NOMBRE ET DE LA DISPOSITION SPATIALE DES POINTS DE PRÉLÈVEMENTS

### 1. Introduction

Depuis 1992, les équipes des laboratoires « Matière organique des Sols tropicaux » et « Écologie des Sols tropicaux » développent des programmes d'étude du sol à des échelles de l'ordre de l'hectare. Ces travaux font appel à diverses méthodes de traitement de données parmi lesquelles l'outil géostatistique (ROSSI et LAVELLE, 1993) qui nécessite un important investissement (en termes de temps et de coût financier) lors de l'échantillonnage. Nous allons successivement envisager i) l'étude du nombre de mesures à effectuer en vue d'atteindre un niveau de précision donné et ii) la disposition des points de mesure à travers l'espace que l'on souhaite étudier.

C'est finalement par un compromis entre les objectifs souhaités et les différentes contraintes (budgétaires, humaines et temporelles) s'exerçant sur les équipes de recherche, que des stratégies d'échantillonnage pourront être mises au point.

### 2. Détermination du nombre de mesures nécessaires pour atteindre un degré de précision donné

La détermination du nombre de mesures à faire pour que la valeur moyenne de la série de données ait un intervalle de confiance donné se fait de façon simple pourvu que l'on dispose d'une estimation de la variance de la variable étudiée. En d'autres termes, il est nécessaire de réaliser un pré-échantillon à partir duquel on estimera la variance (voir par exemple la revue de CLINE, 1944). Soit un pré-échantillon de  $n$  mesures, et  $m$  la moyenne arithmétique de ces valeurs, on a :

$$m = \frac{1}{n} \sum_{1}^{n} x$$

La variance du pré-échantillon est définie par

$$\sigma^2 = \frac{1}{n} \sum (x - m)^2$$

et

$$\sigma^2 = \frac{1}{n-1} \sum (x - m)^2$$

si  $n < 100$ .

Le théorème central limite indique que la valeur moyenne ( $m$ ) de l'échantillon suit une loi Normale ayant pour moyenne l'espérance mathématique de  $x$

$$E[x] \text{ et pour variance la valeur } \frac{\sigma^2}{n} \text{ pourvu que } n > 30.$$

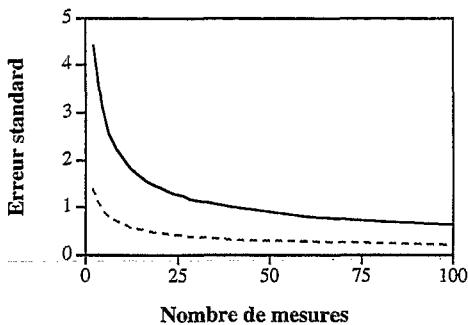
Ayant estimé  $\sigma^2$  on peut donc calculer la valeur de  $n$  correspondant à n'importe quelle valeur de l'erreur standard (c'est-à-dire l'écart type de  $m$ ).

Nous avons appliqué cette méthode à deux jeux de données formés de 60 dosages de la teneur en carbone d'un vertisol de la Martinique (ROSSI, 1992a). Les résultats présentés correspondent à deux profondeurs différentes, respectivement 0-10 et 30-40 cm.

Chaque échantillon permet l'estimation de la variance  $\sigma^2$  et l'application du théorème central limite permet le calcul de l'erreur standard pour chaque valeur de  $n$  comprise entre 1 et 100 (Figure 1).

Ces résultats montrent que i) l'erreur standard décroît avec les  $n$  croissant et ii) pour une même valeur de  $n$ , l'erreur standard est directement proportionnelle à la variance du pré-échantillon.

Figure 1 : Évolution de l'erreur standard de la teneur en carbone d'un vertisol en fonction du nombre de mesures effectuées. Les résultats sont présentés pour deux profondeurs : en trait plein pour la couche 0-10 cm et en pointillés pour la couche 30-40 cm (d'après ROSSI et LAVALLE, 1993).



Cet exemple indique que pour un échantillon de 60 mesures dans la couche 0-10 cm, l'erreur standard est de 0,81 alors que cette même précision est atteinte avec seulement 6 mesures pour la couche 30-40 cm (voir ROSSI et LAVELLE, 1993 pour une étude des 4 couches 0-10, 10-20, 20-30 et 30-40 cm). Ce résultat est expliqué par la différence des variances pour la couche 0-10 (39,03) et 30-40 (3,84).

Ce type d'approche permet de déterminer la taille † des échantillons, c'est-à-dire l'effort analytique requis pour atteindre un certain niveau de précision dans l'estimation de la moyenne (voir par exemple ONOFIOK, 1993). Cependant, les études portant sur de grandes surfaces entraînent un second problème qui est de déterminer la position des points de prélèvement au sein de la surface étudiée.

### 3. Disposition des stations d'échantillonnage dans l'espace étudié

Lorsque la surface étudiée est importante, on constate souvent la présence de structures spatiales des variables physico-chimiques par exemple des gradients liés à l'existence de pentes etc.

L'existence de ces structures entraîne l'apparition d'une hétérogénéité qui se traduit par une composante supplémentaire de la variance. De telles variables sont dites « variables régionalisées » (MATHERON, 1965) et peuvent être étudiées par des méthodes spéciales : les géostatistiques.

#### 3.1. Les géostatistiques

La géostatistique permet de rechercher et de décrire les structures spatiales des variables régionalisées. Ceci est réalisé par l'analyse du **semi-variogramme**. Une seconde phase permet ensuite de réaliser des estimations de la variable à des sites non échantillonnés par l'intermédiaire de diverses méthodes d'interpolation rassemblées sous le terme générique **krigeage**.

##### *Analyse de la structure spatiale : le semi-variogramme*

Soit  $n$  points de mesures, chaque point étant caractérisé par ses coordonnées dans l'espace à 2 dimensions ( $x$  et  $y$ ). Soit une valeur  $D$  de la distance, on dispose de  $n'$  couple de points séparés par une distance égale à  $D$ . La semi-variance permet de quantifier la relation spatiale existant entre les points séparés par une distance quelconque  $D$ . Pour une variable  $Z$  la semi-variance  $\gamma(h)$  est donnée par

$$2\gamma(h) = \text{VAR}\{z(x) - z(x+h)\}$$

avec  $h$  la distance, et  $Z(x)$  le valeur de  $Z$  au point  $x$  où  $\text{Var}$  signifie variance.

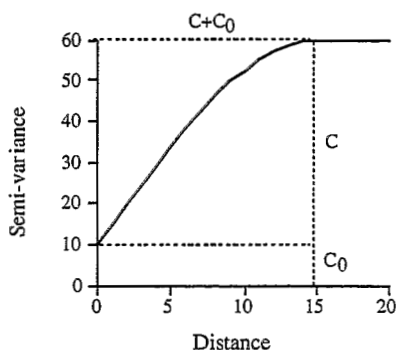
---

† Par taille d'un échantillon, on entend le nombre de mesures réalisées.

Le graphe obtenu en traçant  $\gamma(h)$  en fonction de  $h$  est appelé semi-variogramme ou variogramme. Il indique l'intensité de la structure spatiale et son évolution avec la distance  $h$ .

Au variogramme calculé à partir d'un échantillon est ensuite ajusté un modèle théorique qui sera utilisé lors de l'interpolation par krigeage. La figure 2 représente l'un des modèles les plus courants : le modèle sphérique (pour une revue détaillée voir McBRATNEY and WEBSTER, 1986).

Figure 2 : Semi-variogramme théorique de type « sphérique » illustrant la variance « pépîte » notée  $C_0$ , la variance spatiale notée  $C$ , le seuil ou sill-variance ( $C + C_0$ ) ainsi que la portée du variogramme correspondant à une valeur de 15. (d'après ROSSI et LAVELLE sous presse).



L'ordonnée à l'origine notée  $C_0$  est appelée *variance pépîte* et représente l'ensemble formé de la variabilité due aux erreurs de mesures, et les fluctuations aléatoires ainsi que d'éventuelles structures intervenant à des échelles inférieures au pas d'échantillonnage. La valeur  $C$  nommée variance spatiale rend compte de la proportion de la variance expliquée par la structure spatiale ; enfin, la somme  $C_0 + C$  correspond à la valeur maximale de la semi-variance que l'on nomme Sill-variance en anglais. Cette dernière valeur correspond théoriquement à la variance totale de la série.

Un dernier paramètre du modèle est particulièrement intéressant, il s'agit de la valeur ( $a$ ) de la distance à partir de laquelle la semi-variance atteint sa valeur seuil ( $C_0 + C$ ). Cette distance est la portée du variogramme. On considère que les points séparés par une distance supérieure ou égale à la *portée* du variogramme sont indépendants au sens statistique.

*Cartographie des variables régionalisées : le krigeage*

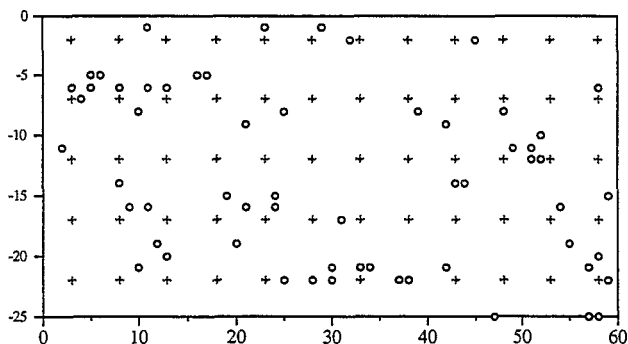
La cartographie des variables présentant une structure spatiale peut être réalisée par le krigeage. Le terme de krigeage désigne un ensemble de méthodes d'interpolation linéaire basées sur l'expression de la dépendance spatiale par le semi-variogramme. Les deux types de krigeage les plus communs sont le krigeage ponctuel (BURGESS and WEBSTER, 1980a) et le krigeage par blocs (BURGESS and WEBSTER, 1980b).

Le krigeage ponctuel correspond à l'estimation de la valeur d'une variable en un point non échantillonné, le krigeage par blocs estimant la valeur moyenne de cette variable sur une surface déterminée. Dans les deux cas, on procède à l'interpolation linéaire de la variable étudiée Z en un point où aucune mesure n'a été réalisée. Un certain nombre de points échantillonnés sont utilisés dans le calcul de la valeur estimée et le poids alloué à chaque point est une fonction de la distance le séparant du site où l'estimation est faite. Cette fonction est liée à la structure spatiale par l'intermédiaire du semi-variogramme (pour une revue détaillée le lecteur pourra consulter WEBSTER, 1985 ; WEBSTER and OLIVER, 1990).

**3.2. Effet de deux plans d'échantillonnage usuels sur l'analyse géostatistique**

Afin de cerner l'impact du mode de distribution des points de mesures sur les résultats de l'analyse géostatistique, nous allons examiner les résultats obtenus pour la variable « profondeur du sol » échantillonnée sur un même site mais selon deux plans d'échantillonnage différents. Le premier est un plan d'échantillonnage aléatoire simple et le second un plan d'échantillonnage régulier à maille carrée (Figure 3). Chaque plan d'échantillonnage présente le même nombre de points de mesures fixé à 60 en fonction des contraintes matérielles et temporelles et répartis sur une parcelle de 60 x 25 m.

**Figure 3 : Distribution spatiale de deux séries de mesures de la profondeur du sol. Les cercles figurent 60 stations distribuées aléatoirement dans le champ et les croix représentent 60 points distribués régulièrement selon une maille carrée.**



Les figures 4 et 5 illustrent les semi-variogrammes issus respectivement des plans aléatoire et régulier. La forme des courbes est différente et ceci se traduit par des ajustements à des modèles théoriques différents (sphérique pour le variogramme tiré de l'échantillonnage aléatoire et linéaire pour celui issu de l'échantillonnage régulier). Ces résultats indiquent donc clairement l'effet de la distribution des points de mesures sur notre perception de la structure spatiale de la profondeur du sol : les variogrammes sont influencés par le mode d'échantillonnage.

Figure 4 : Semi-variogramme de la variable « profondeur du sol » calculé à partir des mesures issues d'un échantillonnage aléatoire simple. En trait plein, le modèle théorique associé de type sphérique (d'après ROSSI et LAVELLE, 1993).

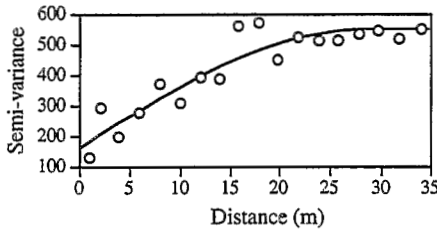
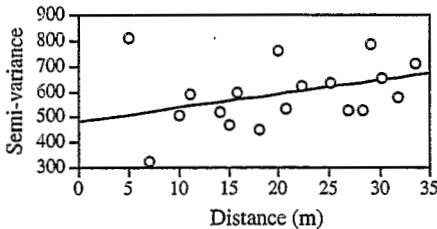


Figure 5 : Semi-variogramme de la variable « profondeur du sol » calculé à partir des mesures issues d'un échantillonnage régulier à mailles carrées. En trait plein, le modèle théorique associé de type linéaire (d'après ROSSI et LAVELLE, 1993).



Dans chacun des deux cas, la cartographie a été réalisée par krigeage par blocs. La taille des blocs est de 2 x 2 m. L'interpolation par krigeage fournit en un point, à la fois une estimation de la variable étudiée et un écart type d'estimation de cette valeur indiquant la fiabilité de l'estimation.

Les figures 6 et 7 présentent respectivement les cartes des estimations et des écarts types d'estimation pour les deux plans d'échantillonnage. On constate

tout d'abord (figure 6) que les estimations issues du krigeage varient fortement suivant que l'échantillonnage est de type aléatoire ou régulier. De même, les cartes de l'écart type d'estimation sont différentes, ce qui s'explique par ailleurs aisément si l'on considère qu'en un point quelconque cette valeur ne dépend pas des valeurs mesurées elles-mêmes mais uniquement de la variance « pépite » et de la distribution spatiale des points impliqués dans l'estimation. De fait, la figure 7 (en haut) montre clairement l'hétérogénéité spatiale de l'écart type d'estimation dans le cas d'un échantillonnage aléatoire, tandis que le plan régulier produit une carte homogène (figure 7 en bas). L'inégale distribution de l'incertitude se révèle gênante lorsque l'on cherche à comparer certaines zones d'une même carte. Elle révèle en outre l'existence de zone sous-échantillonnées dans le cas de l'échantillonnage aléatoire qui sont incompatibles avec une cartographie rigoureuse.

Figure 6 : Cartes de la variable « profondeur du sol » obtenues par krigeage par blocs de  $4m^2$ . En haut des estimations réalisées à partir d'un plan d'échantillonnage aléatoire simple et en bas, celles issues d'un plan régulier (d'après ROSSI et LAVELLE, 1993).

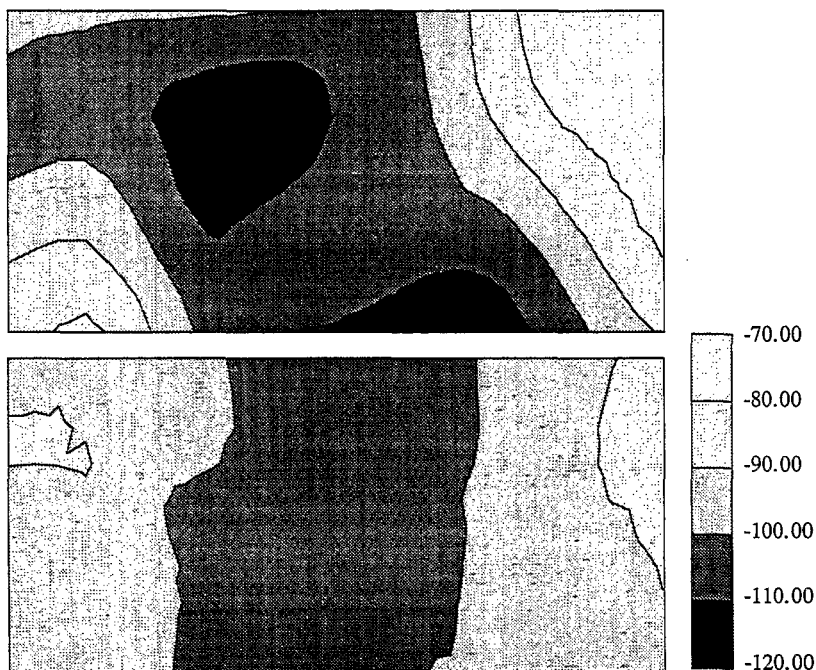
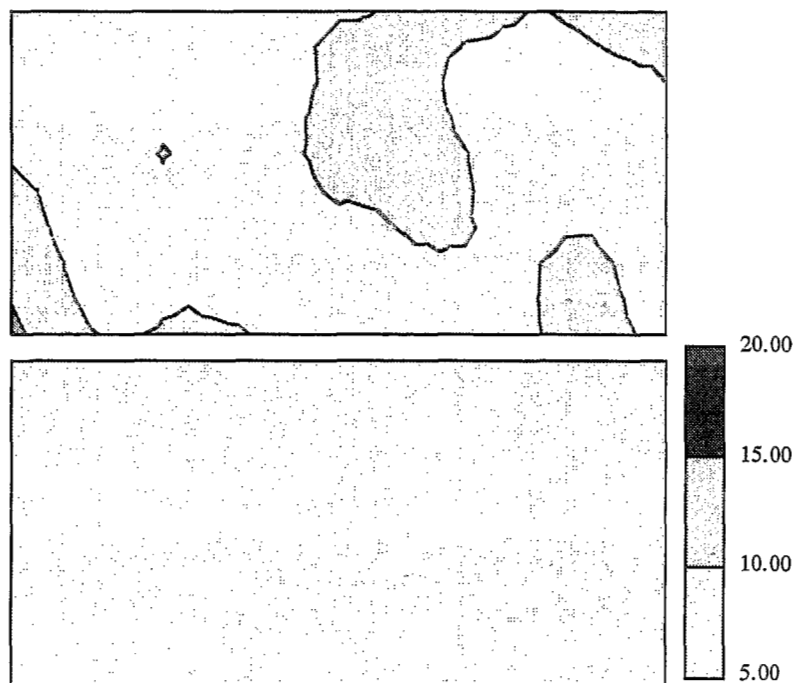


Figure 7 : Cartes des écarts types d'estimation de la variable « profondeur du sol » obtenues par krigeage par blocs de  $4\text{m}^2$ . En haut la carte réalisée à partir d'un plan d'échantillonnage aléatoire simple et en bas, celle issue d'un plan régulier (d'après ROSSI et LAVELLE, 1993).



#### 4. Discussion et conclusion

Il existe quantité de travaux traitant de l'échantillonnage des milieux naturels mais finalement relativement peu de références traitant explicitement des problèmes liés aux approches à l'échelle de la parcelle cultivée. Nous avons vu comment le théorème central limite pouvait apporter des informations sur le nombre de mesures nécessaires à l'estimation d'une valeur moyenne pour un seuil de fiabilité donné (CLINE, 1944 ; McBRATNEY and WEBSTER, 1983). Cependant, lorsque des études sont menées sur des surfaces importantes, il est très fréquent que des structures spatiales apparaissent et dans ce cas, il devient également indispensable de considérer avec la plus grande attention la disposition des points de mesure à travers l'espace étudié (BURROUGH, 1991).

L'utilisation d'un plan d'échantillonnage régulier ou aléatoire entraîne d'importantes variations dans notre perception de la variable étudiée, ce qui révèle l'importance du plan d'échantillonnage.



De nombreux auteurs soulignent que l'échantillonnage aléatoire simple est basé implicitement sur l'absence de structure spatiale, ce qui équivaut à considérer la variable comme homogène sur l'ensemble de la surface étudiée (DEBOUZIE and THIOULOUSE, 1986 ; FRONTIER, 1983).

Du point de vue du traitement géostatistique, l'échantillonnage aléatoire simple doit être évité car il constitue une importante source d'erreur. Intuitivement, il est clair que réaliser plusieurs mesures d'une même variable en des sites très proches consiste en fait à recueillir plusieurs fois des informations très proches (informations redondantes) tout en sous-échantillonnant d'autres surfaces de la zone étudiée. C'est ainsi que les pédologues utilisent généralement un plan d'échantillonnage régulier à maille carré (BURGESS *et al.*, 1981 ; McBRATNEY and WEBSTER, 1981 ; McBRATNEY *et al.*, 1981 ; WEBSTER and OLIVER, 1990).

## 5. Remerciements

Les auteurs remercient M<sup>lle</sup> F. CHARPENTIER et M. S. LOCATI pour leur lecture critique de la première version de ce texte ainsi que M. P. CLOCHARD pour son aide précieuse lors de la phase de collecte des données.

## 6. Références

- BURGESS T.M., and WEBSTER R., 1980a. Optimal interpolation and isarithmic mapping of soil properties. 1 The semi-variogram and punctual kriging. *J. Soil Sci.* 31 : 315-331.
- BURGESS T.M., and WEBSTER R., 1980b. Optimal interpolation and isarithmic mapping of soil properties. 2 Block kriging. *J. Soil Sci.* 31 : 333-341.
- BURGESS T.M., WEBSTER R. and McBRATNEY A.B., 1981. Optimal interpolation and isarithmic mapping of soil properties. 4 Sampling strategy. *J. Soil Sci.* 32 : 643-659.
- BURROUGH P.A., 1991. « Sampling designs for quantifying map unit composition ». In *Spatial variabilities of soils and landforms*. M. J. MAUSBACH and L. P. WILDING Eds. SSSA special publication n° 28 : 89-125.
- CLINE M.G., 1944. Principles of soil sampling. *Soil Sci.* 58 : 275-288.
- DEBOUZIE D., and THIOULOUSE J., 1986. « Statistics to find spatial and temporal structures in populations ». In *Pest control : Operations and systems analysis in fruit fly management*. Edited by M. MANGEL, J. M. CAREY and R. E. PLANT. pp. 263-282. Berlin, Springer Verlag.
- FRONTIER S., 1983. *Stratégies d'échantillonnage en écologie*. Paris, Masson.
- MATHERON G., 1965. *Les variables régionalisées et leur estimation*. Paris, Masson.

- McBRATNEY A.B. and WEBSTER R., 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables. 2. Program and examples. *Comp. Geosci.* 7 (4) : 335-365.
- McBRATNEY A.B. and WEBSTER R., 1983. How many observations are needed for regional estimation of soil properties ? *Soil Sci.* 135 : 177-183.
- McBRATNEY A.B. and WEBSTER R., 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. *J. Soil Sci.* 37 : 617-639.
- McBRATNEY A.B., WEBSTER R. and BURGESS T.M., 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables. 1. Theory and method. *Comp. Geosci.* 7 (4) : 331-334.
- ONOFIOK O.E., 1993. « Determining spatial and temporal variations of organic matter in a tropical soil using different sampling schemes ». In *Soil organic matter dynamics and sustainability of tropical agriculture* K. MULONGOY and MERCKX R (Eds.) pp. 65-74. Leuven, Belgium : John Wiley & Son.
- ROSSI J.P., 1992a. *Mise au point de méthodes d'échantillonnage de vers de terre en faible densité : optimisation par l'utilisation des géostatistiques*. Ministère de l'Environnement. NTIS, Rapport Final.
- ROSSI J.P. and LAVELLE P., 1993. « Statistical procedures for field surveys ». In *Conservation of soil fertility in low agricultural systems of the humid tropics by manipulating earthworm communities. (Macrofauna project 2)*. Edited by P. LAVELLE. pp. 63-86. Bondy, Orstom.
- WEBSTER R., 1985. Quantitative spatial analysis of soil in the field. *Adv. Soil Sci.* 3 : 2-70.
- WEBSTER R. and OLIVER M.A., 1990. *Statistical methods in soil and land resource survey*. Oxford University Press.