

Détermination de la taille des unités d'échantillonnage par analyse variographique de données satellitaires. Exemples d'inventaires d'occupation du sol en France et en Pologne.

GODARD V.¹

Mots clés : Analyse spatiale ; Échantillons ; France ; Inventaires ; Occupation du sol ; Pologne ; Sondages spatiaux ; Télédétection ; Variogrammes

Depuis le milieu des années 80, un grand nombre d'inventaires régionaux de statistiques agricoles utilisent les sondages spatiaux associés aux données satellitaires. Celles-ci sont utilisées pour améliorer les résultats d'estimation de superficie et / ou pour les cartographier. Dans le cadre des sondages spatiaux, l'unité d'échantillonnage est appelée segment. Elle a une taille non nulle, qui doit être définie en fonction de l'hétérogénéité du paysage.

En milieu agricole, l'expérimentation (FOURNIER 1988) a montré qu'un segment calé sur la taille moyenne des exploitations assurait un compromis satisfaisant entre : représentativité des différentes cultures, temps d'enquête et précision des estimations. En revanche, en milieux forestiers ou en milieux faiblement anthropisés, il n'existe pas d'information analogue permettant de calibrer le segment sur une quelconque organisation des paysages. Toutefois, certaines caractéristiques des secteurs à inventorier peuvent être perçues grâce à une analyse variographique des données télédéteectées.

Les résultats portent sur deux secteurs agro-forestiers, l'un dans le sud de l'Île-de-France, l'autre dans l'est de la Pologne. Ils permettent de déterminer la taille des unités d'échantillonnage, soit en ciblant l'analyse sur la composante agricole ou forestière du paysage, soit en prenant en compte la globalité de la zone d'étude.

Le variogramme et la dépendance spatiale

Présente depuis les années 70 dans les Sciences de la Terre, l'analyse variographique est un des outils de la géostatistique utilisé pour analyser la dépendance spatiale des échantillons entre eux. Son utilisation sur des données satellitaires est plus récente, elle date des années 80. Elle a

¹Université de Paris 8 & Centre de Biogéographie-Écologie UMR 180 CNRS - Le Parc, 92211 Saint Cloud Cedex, France

essentiellement servi à calibrer l'espacement des unités d'échantillonnage (CURRAN 1988) ou à restaurer les radiométries des images dégradées (N'KANZA et NAIZOT 1993).

Rappelons que les données satellitaires sont spatialement corrélées. Cela signifie que les voisins immédiats d'un pixel tiré au hasard lui sont en général plus ressemblant que des voisins plus lointains. Cette dépendance spatiale reflète l'extension des zones homogènes qui marquent les paysages à inventorier, son étude doit donc permettre de calibrer les segments d'enquête.

Pour connaître les distances caractérisant les changements de paysage, on tire un échantillon de points et l'on compare chaque point à son voisinage immédiat puis sur des distances de plus en plus grandes et cela dans un certain nombre de directions. On en tire une variance par distance et par direction que l'on représente en général à l'aide d'une courbe, le variogramme. La formule utilisée pour calculer cette variance est la suivante :

$$\gamma (D,d) = \frac{1}{n} \sum_{i=1}^n \frac{(y_{i+d} - y_i)^2}{2}$$

où y_i est la valeur radiométrique du premier point tiré au hasard, y_{i+d} la valeur radiométrique du deuxième point du couple pris à une distance "d" de y_i dans la direction "D" (DUPLAT 1981). La recherche des points d'inflexion sur la courbe informe des ruptures dans la dépendance spatiale, et donc dans la continuité des paysages. Des segments dont la taille est calée sur ces distances doivent permettre une prise en compte satisfaisante des paysages de la zone d'étude.

Les zones d'étude

Il nous a paru intéressant d'adjoindre aux résultats portants sur un secteur francilien, présentés lors du colloque Théo Quant de 1995 à Besançon (FOURNIER *et alii* 1995), les variogrammes réalisés lors d'un inventaire d'occupation du sol en Pologne, au nord de la forêt de Bialowieza. Les deux approches diffèrent. Les premières analyses variographiques (France) sont réalisées "post-inventaire" alors que les secondes (Pologne) sont un préalable à celui-ci.

La première étude sur laquelle ont été réalisés les variogrammes se situe au sud de l'Ile-de-France sur quatre sites : deux à dominante agricole et deux à dominante forestière. A cheval sur l'Essonne, la Seine-et-Marne et l'extrême nord du Loiret, la zone d'étude fait une trentaine de kilomètres de côté et s'étend du nord au sud entre Brétigny et Pithiviers et de l'ouest à l'est entre Étampes et Fontainebleau. Pour chaque site de cinq kilomètres de côté, nous avons étudié les données numériques des trois canaux bruts du capteur HRV2 de SPOT1 (KJ 40-252, 1er mai 1986) ainsi que deux néo-canaux, l'indice de végétation et l'indice de brillance.

Nous avons retenu cette image d'archives pour comparer la taille des segments estimée par analyse variographique, avec celle utilisée par le Service Central des Enquêtes et Études Statistiques avec ces mêmes données (FOURNIER 1988).

La seconde zone d'étude concerne un paysage agro-forestier de l'est de la Pologne dans la voïvodie de Bialystok. Le secteur d'étude est compris entre la retenue de Siemianowka, récemment

mise en eau (la construction du barrage est de 1986) et le nord de la forêt de Bialowieza. Ce lac artificiel semble contemporain de la construction de résidences secondaires sur des terres précédemment affectées à l'agriculture. Il s'agit d'étudier ici le changement d'affectation des terres pour un inventaire d'occupation des sols. L'analyse variographique a servi à calibrer la taille des segments pour l'inventaire. Pour ce secteur d'environ 20 km de côté, nous avons étudié les trois canaux bruts du capteur HRV1 de SPOT2 (KJ 81-242, 24 septembre 1994) ainsi que l'Indice de Végétation, l'Indice de Brillance les deux premières composantes principales.

Pour calculer les variogrammes, la dépendance spatiale est analysée sur quatre directions (N-S, E-W, NW-SE et NE-SW) et sur des distances qui n'excèdent pas la moitié des sites d'étude. De toute façon la partie la plus féconde en enseignement est la partie du variogramme la plus proche de l'origine (DALY 1991).

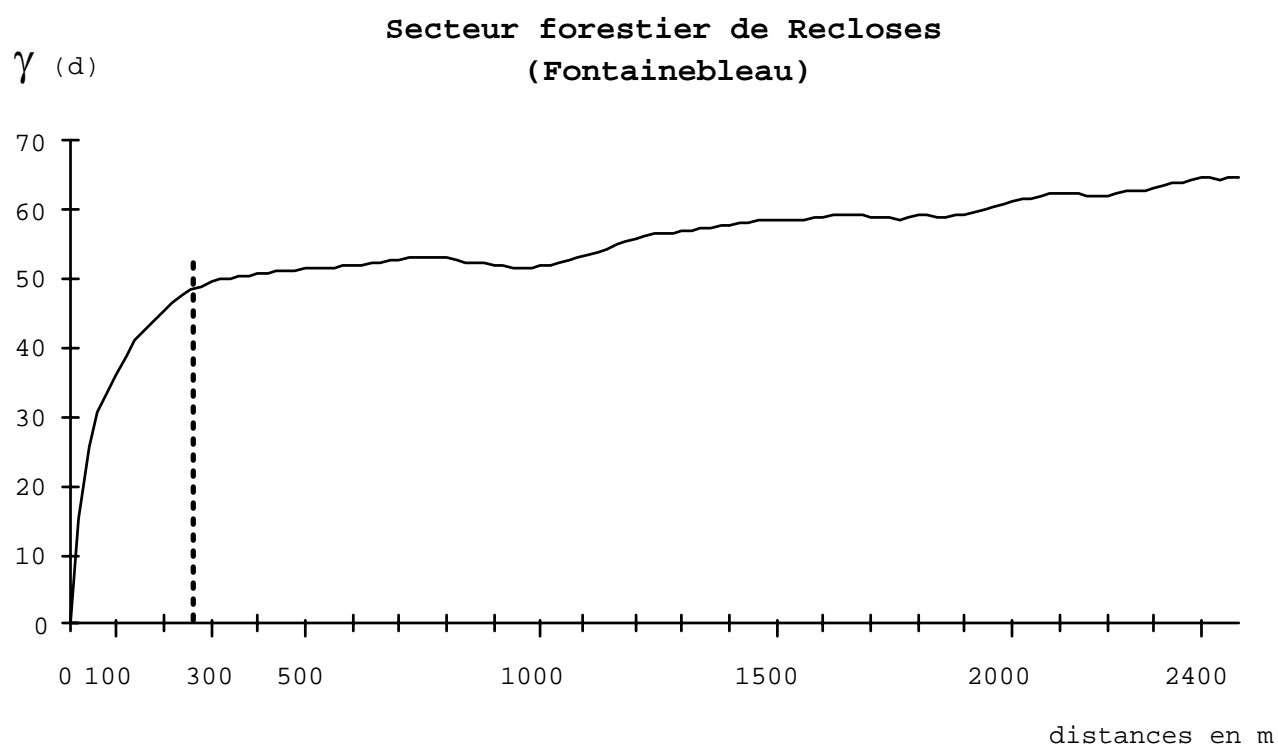
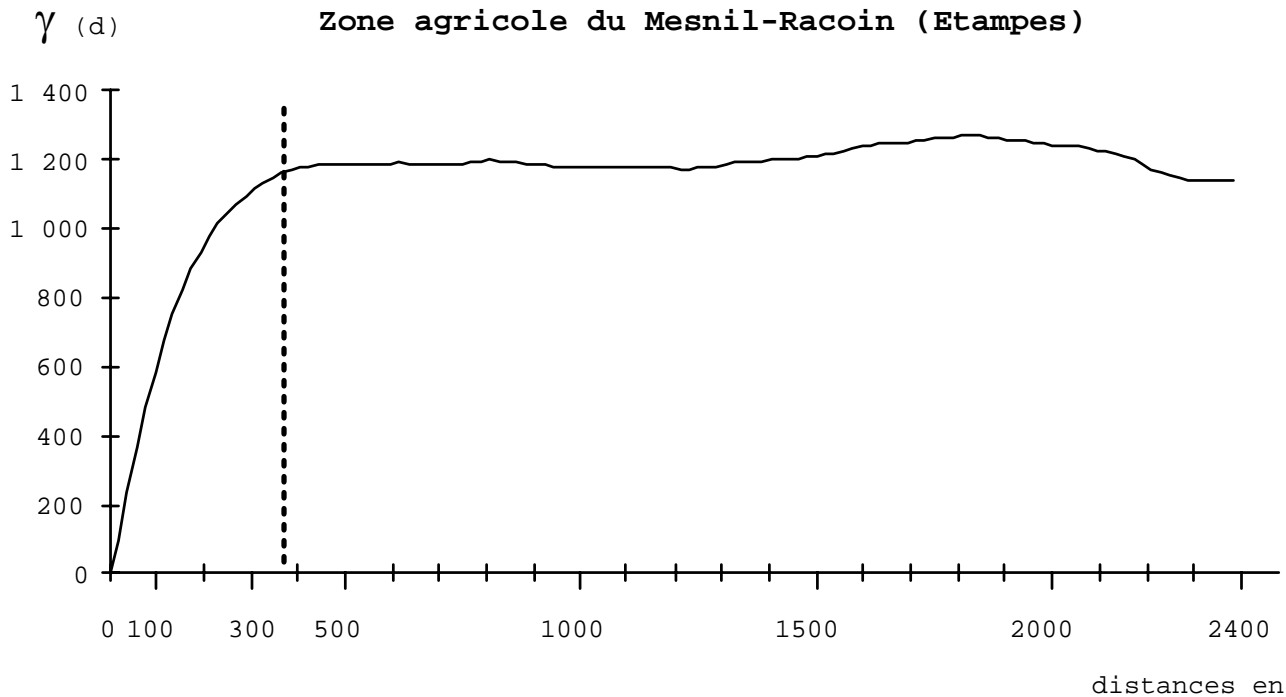
Résultats

Dans un premier temps, pour chaque site et par canal nous avons réalisé le variogramme moyen des quatre directions. Celui-ci permet d'éliminer rapidement les canaux inutilisables car spatiaux ou au contraire dotés d'une variance continue et indéfiniment croissante (CURRAN 1988) à notre échelle. Une analyse plus fine par direction peut être entreprise si la dépendance spatiale se révèle nettement anisotropique.

Pour la zone d'étude située au sud de l'Ile-de-France, il ressort de l'analyse du variogramme moyen de chaque canal, et pour cette date d'enregistrement, que pour les quatre sites, l'indice de végétation est le canal le plus systématiquement exploitable, car après une croissance rapide à l'origine, la courbe s'infléchit et décrit un plateau (cf. fig. 1). Cela signifie que la diversité radiométrique et donc thématique croît rapidement sur les premières dizaines voire centaines de mètres et qu'ensuite cette diversité s'estompe. Continuer de s'éloigner du point origine n'apporte pas un gain d'informations supplémentaires significatif. La rupture de pente sur le variogramme marque ainsi la dimension moyenne du parcellaire agricole ou des peuplements homogènes forestiers. C'est donc une taille pertinente pour caler le segment. La détermination de cette rupture de pente (portée du phénomène) est effectuée ici de manière visuelle (cf. fig. 1).

L'algorithme du variogramme "local", développé par S. N'KANZA, prend en compte l'ensemble de la population des pixels de la fenêtre d'étude. Il n'a été utilisé que sur les petites zones d'étude n'excédant pas 5 km (250 pixels SPOT) de côté. L'aspect des courbes est suffisamment "lisse" pour que la lecture ne requiert pas d'ajustement. Il en va tout autrement des variogrammes "globaux" dont l'algorithme a été développé par N. COQUILLE (Centre de coopération International pour la Recherche Agronomique et le Développement, CIRAD) puis H. GEROYANNIS.

figure n°1 - Variogrammes moyens d'une zone agricole et d'un secteur forestier



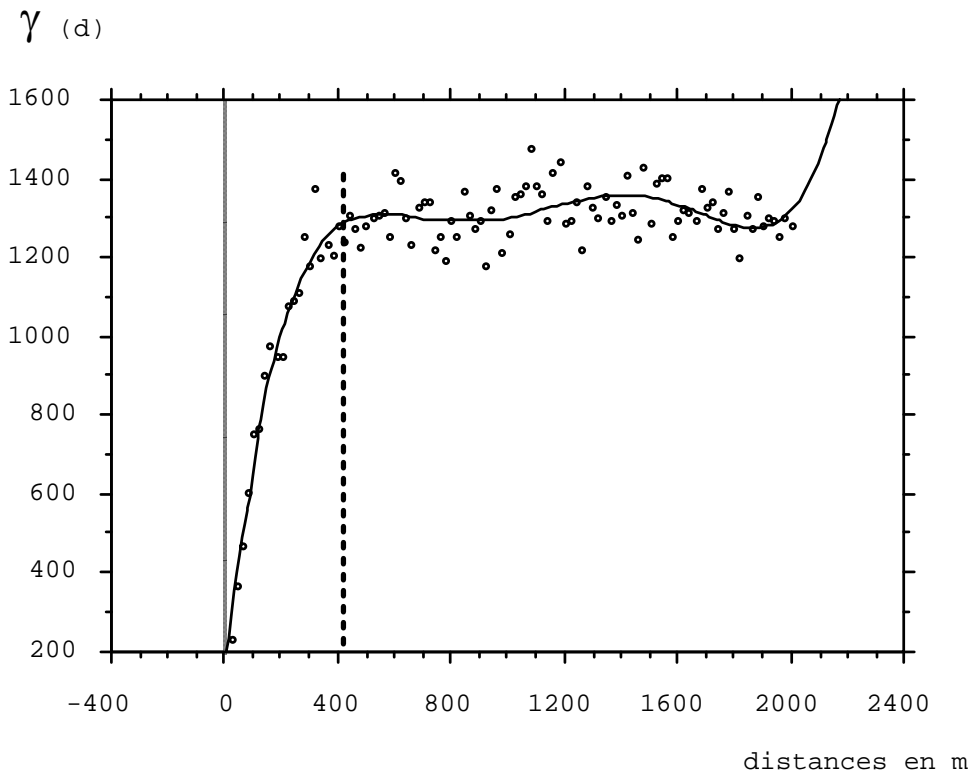
Moyenne des 4 directions calculée sur l'Indice de Végétation Normalisé.

Ceux-ci, calculés sur l'ensemble des secteurs d'étude (20 à 30 km de côté), imposent de recourir à un échantillon de pixels (200 par direction et par distance) pour limiter les temps de

calcul. L'aspect pointilliste qui en découle (cf. fig. 2 et 3) nécessite un lissage pour permettre la recherche de la portée. Plusieurs types d'ajustement sont possibles [sphérique, exponentiel, etc. (cf. Mc BRATNEY *et alii* 1986, CURRAN 1988, N'KANZA et NAIZOT 1993, ...)].

Nous avons retenu un ajustement par régression polynomiale directement accessible dans la plupart des logiciels. La recherche de la portée se fait alors sur cette courbe d'ajustement.

figure n°2 - Variogramme moyen global du secteur francilien



*Moyenne des 4 directions calculée sur l'Indice de Végétation Normal:
Ajustement par régression polynomiale d'ordre 5, avec :*

$$y = 155,42 + 132,412 * x - 5,66 * x^2 + 0,112 * x^3 - 0,001 * x^4 + 3,535 E-6 * x^5$$

$$R^2 = 0,926$$

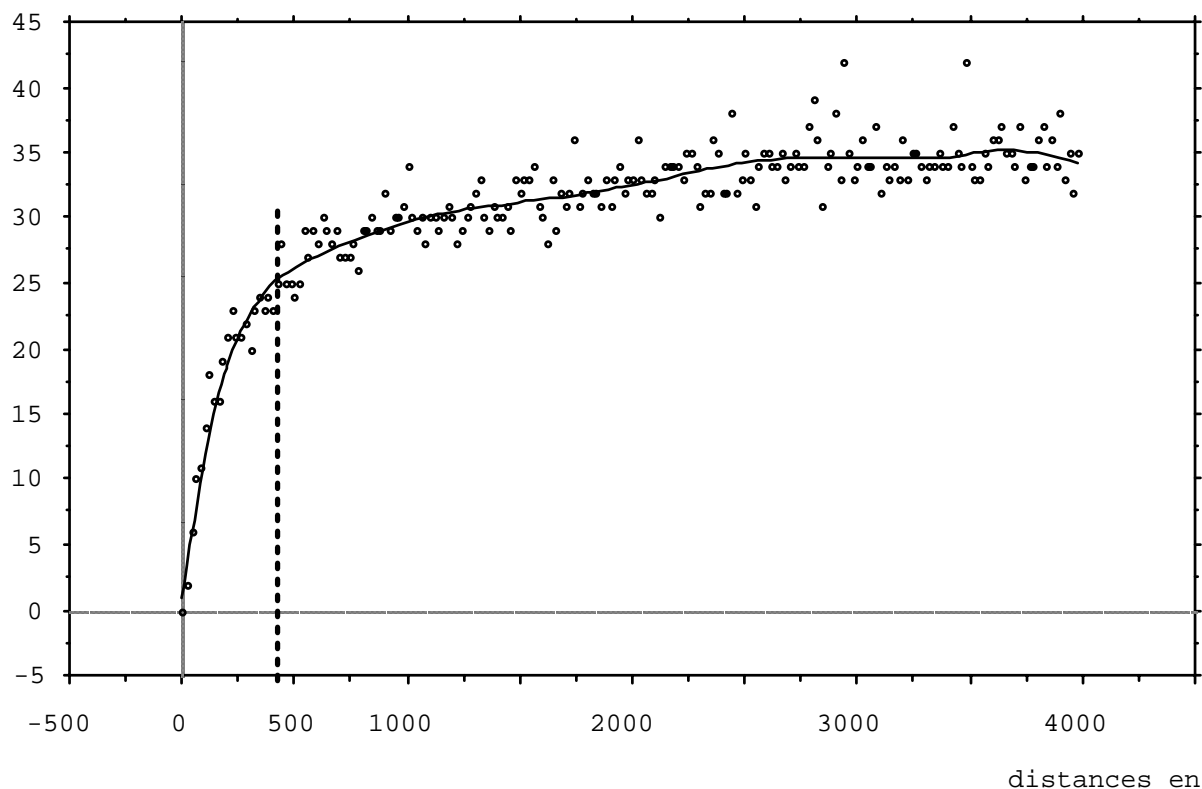
Pour le secteur francilien, comme on pouvait s'y attendre, les points d'inflexion caractérisant les milieux forestiers et les milieux agricoles sont légèrement décalés (cf. fig. 1). La distance significative est d'environ 350 m sur les secteurs agricoles, et 250 m sur les secteurs forestiers. Si l'on se réfère à un variogramme moyen global, établi sur l'ensemble de la zone d'étude (30 km de côté), il convient alors de caler le segment sur une distance d'environ 400 m (cf. fig. 2). Cet accroissement de la portée est dû à la prédominance d'un parcellaire agricole de grande taille au niveau de la zone d'enquête.

En analysant ce dernier résultat, on peut penser que l'inventaire agricole réalisé par le SCEES en 1986 s'est appuyé sur une taille d'unités d'échantillonnage (720 m de côté) nettement

surdimensionné (de + de 300m) par rapport à la diversité des paysages. Un "surdimensionnement" des unités d'échantillonnage a également été constaté sur un inventaire d'occupation du sol en milieu naturel tropical (GODARD 1994).

figure n°3 - Variogramme moyen global du secteur de Siemianowka (Pologne)

γ (d)



Moyenne des 4 directions calculée sur la deuxième composante principale
Ajustement par régression polynomiale d'ordre 9, avec :

$$y = 0,984 + 3,062 * x - 0,164 * x^2 + 0,005 * x^3 - 9,358E-5 * x^4 + 1,071E-6 * x^5 - 7,59E-9 * x^6 + 3,243E-11 * x^7 - 7,622E-14 * x^8 + 7,563E-17 * x^9$$

$$R^2 = 0,92$$

Pour le site polonais de Siemianowka, le variogramme global (cf. fig. 3) indique également une distance moyenne d'environ 400 m. Dans la mesure où aucune autre information de terrain n'est disponible, c'est cette distance de 400 m qui sera retenue pour l'inventaire qui doit avoir lieu au printemps 1996. En l'absence de ce résultat, il aurait été nécessaire de déterminer la taille de l'unité d'échantillonnage de façon empirique en se référant à des inventaires effectués sur d'autres milieux et jugés comparables.

Discussion

Les analyses variographiques de secteurs agricoles et forestiers ont montré qu'il était possible de caler la taille des unités d'échantillonnage d'un inventaire sur la complexité des paysages étudiés à l'aide des données satellitaires. Cependant, sur deux études et par rapport à d'autres méthodes de détermination de la taille des segments, le variogramme a semblé minorer la taille des unités d'échantillonnage.

Actuellement, il ne nous est pas possible de dire si un inventaire dont les segments ont été déterminés par analyse variographique est plus efficace qu'un inventaire dont les segments ont été déterminés par une autre méthode. Les outils d'évaluation, encore en développement, ne sont pas satisfaisants. De plus, il conviendrait de voir si les résultats obtenus sont stables en fonction de différentes dates de prises de vues et de différentes résolutions spatiales.

Remerciements

Nous remercions la société SPOT-Image et le Centre National d'Etude Spatial pour nous avoir fourni les données SPOT sur la Pologne dans le cadre du programme de soutien ISIS ainsi que H. GEROYANNIS et B. JULIEN-LAFERRIERE du Laboratoire PASTEL pour leur aide et leurs conseils.

Bibliographie

- CURRAN (P.J.) - 1988 - The Semivariogram in Remote Sensing: An Introduction. *Remote Sensing of Environment*. 24 : 493-507.
- DALY (C.) - 1991 - *Application de la géostatistique à quelques problèmes de filtrage*. Thèse de doctorat de géostatistique, ENSMP, Fontainebleau, 235 p.
- DUPLAT (P.), PERROTTE (G.) - 1981 - *Inventaire et estimation de l'accroissement des peuplements forestiers*. Paris, ONF, 432 p.
- FOURNIER (Ph.), GEROYANNIS (H.), GILG (J.P.) - 1988 - Évaluation des surfaces des cultures annuelles et inventaire du territoire dans le Bassin Parisien par combinaison des enquêtes statistiques classiques et des données SPOT (PEPS 127). *SPOT-1. Utilisation des images, Bilan, Résultats*. Paris 23-27 novembre 1987. Toulouse, Cépadués éd. : 251-259.
- FOURNIER (Ph.), GEROYANNIS (H.), GODARD (V.) - (à paraître) - Analyse variographique de données satellitaires pour déterminer la taille des unités d'échantillonnage. Exemple d'un inventaire d'occupation du sol en milieu agricole et forestier. *Deuxièmes Rencontres de Théo Quant*. Besançon, 4 - 5 octobre 1995, 5 p.
- GODARD (V.) - 1994 - Apport de l'analyse variographique pour déterminer la taille et l'espacement des unités d'échantillonnage lors d'un inventaire d'occupation du sol en milieu naturel. *Bulletin de la SFPT*. 94-4 (136) : 33-44.

GODARD (V.) - 1996 - Détermination de la taille des unités d'échantillonnage par analyse variographique de données satellitaires. Exemple d'inventaires d'occupation du sol en France et en Pologne. in : *Géopoint 96 - Espace et Nature dans la Géographie d'aujourd'hui*. Avignon 23 et 24 mai 1996 : 102-106.

McBRATNEY (A.B.), WEBSTER (R.) - 1986 - Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. *Journal of Soil Science*. 37 : 617-639.

N'KANZA (S.), NAIZOT (T.) - 1993 - Méthode d'analyse structurelle d'une image dégradée. *Photo-interprétation*. 93-1 (1), juillet : 5-8