

SONDAGES SPATIAUX EN MILIEU AGRICOLE PAR DES METHODES GEOSTATISTIQUES SUR LES IMAGES SATELLITAIRES

N'KANZA S.¹

RESUME

Nous définissons les variables régionalisées sur des données spatiales constituées par des images de télédétection classifiées selon une nomenclature "occupation du sol".

Nous étudions la structure de ces variables à travers le variogramme. La localisation et la configuration géométrique de l'échantillon nous permettent de proposer des estimateurs à pondération non uniforme: c'est le problème du krigeage. Enfin, par le calcul de la variance d'estimation, nous évaluons la précision du sondage spatial.

Ce projet de recherche reprend en partie les lignes présentées par Philippe BRION (Réf. 1 Biblio.).

Nous cherchons à modéliser la structure de l'espace à deux dimensions en tenant compte de la masse des données fournies par les images de télédétection. Des enquêtes par sondage effectuées en vue d'obtenir des statistiques sur les différentes catégories d'occupation du sol sont faites sous plusieurs formes. En particulier, elles sont menées à partir de grilles de points ou à partir de petites unités (segments).

Le nombre et la taille des unités à enquêter sont choisis en fonction de la capacité de l'enquêteur à parcourir une certaine superficie pendant un temps donné. La distance entre les points ou les segments est fixée uniformément selon les régions. Ceci ne permet pas de bien apprécier l'efficacité de tels sondages à cause du caractère systématique du tirage.

L'utilisation d'une image de télédétection offre un avantage certain. L'information fournie sur la structure du parcellaire reste riche en dépit des éventuelles confusions inhérentes à la classification des pixels. En comparaison, un fichier d'enquête réalisé à partir de points ou de segments ne donne que les valeurs au niveau de ces entités et ne renseigne pas sur ce qui existe entre elles.

¹L.A.T.E.S. - C.A.M.S. - E.H.E.S.S. 54, Bd Raspail 75006 PARIS

Nous proposons des estimateurs géostatistiques à pondérations non uniformes. La géostatistique est l'application de la théorie des variables régionalisées à l'estimation. Ces dernières présentent des valeurs réparties dans l'espace. Elles sont décrites par un modèle probabiliste $z(x)$ et définies comme des réalisations de fonctions aléatoires.

Cette théorie a été développée à l'Ecole des Mines par G. Matheron et son équipe. Elle est basée essentiellement sur deux hypothèses de travail.

I - HYPOTHESE DE STATIONNARITE D'ORDRE 2.

La variable régionalisée $z(x)$ est considérée comme une réalisation d'une fonction aléatoire stationnaire $Z(x)$.

$$E[Z(x)] = m,$$

$$\text{Cov}[Z(x), Z(x+h)] = c(h).$$

Ces deux quantités sont indépendantes de x du fait de la stationnarité.

En pratique, m est la moyenne globale et la covariance $c(h)$ est obtenue à partir du modèle du variogramme.

$$g(h) = \frac{1}{2} E [Z(x) - Z(x+h)]^2.$$

Cette grandeur ne dépend que de l'interdistance h (en module et en direction) et non pas du point x .

Par ailleurs, nous avons la relation liant le variogramme à la covariance:

$$c(h) = c(0) - g(h).$$

Le modèle de variogramme théorique lui-même est établi par un ajustement du variogramme expérimental (moyenne de $g(h) = (E [Z(x) - Z(x+h)]^2) / 2$). Il doit représenter la structure spatiale des données.

II - HYPOTHESE INTRINSEQUE.

Elle est beaucoup plus faible que l'hypothèse de stationnarité d'ordre 2.

La covariance $c(h)$ n'existe pas dans le sens évoqué plus haut. Mais, on suppose que les accroissements ont une variance finie $V [Z(x+h) - Z(x)]$.

Dans ce cas, on dit que la fonction aléatoire $Z(x)$ vérifie l'hypothèse intrinsèque, avec:

$$E [Z(x+h) - Z(x)] = m(h),$$

$$E [Z(x+h) - Z(x)]^2 = 2g(h),$$

où $m(h)$ est la dérive linéaire.

La connaissance du variogramme est la condition nécessaire pour estimer la précision d'un sondage spatial.

A partir de ces hypothèses, les estimations géostatistiques sont réalisées par krigeage.

Krigeage.

Le krigeage est un estimateur linéaire fournissant des poids optimaux, compte tenu, d'une part de la configuration géométrique des échantillons et, d'autre part de la structure du variogramme.

Une variable régionalisée $z(x)$ (V.R.) est connue en certains points: $(x_a)_{a=1, \dots, n}$:

$$z_a = z(x_a).$$

On estime alors sa valeur inconnue en un certain point x par une combinaison linéaire des z_a .

Pour cela, nous posons:

$$Z^*(x) = \sum_a |^a Z_a.$$

La valeur estimée est donc une moyenne pondérée des valeurs Z_a , la somme des poids $|^a$ valant 100%.

Le krigeage consiste donc à calculer les poids $|^a$ aboutissant à une bonne estimation.

Les échantillons les plus proches du point à estimer reçoivent en général un poids plus élevé mais dépendant du variogramme.

Les structures anisotropiques peuvent modifier ces constatations. Les anisotropies se manifestent par un comportement différent du variogramme dans les diverses directions de l'espace. Dans le cas idéal de l'isotropie, la variation du variogramme est indépendante de la direction.

Exemples de krigeage.

1°) Krigeage simple (K.S.) (ou à moyenne connue).

Hypothèse H_1 (stationnarité d'ordre 2):

$$E [Z(x)] = m,$$

$$E [(Z(x+h)-m) (Z(x)-m)] = c(h).$$

avec:

$$c(h) = c(0) - g(h).$$

On note:

$$Z^*(x) = \sum_a |^a Z_a. \quad \text{l'estimateur utilisé.}$$

Alors:

$$\text{Var}(Z^*(x)) = \text{Var} \left(\sum_a |^a Z_a. \right) = \sum_a \sum_b |^a |^b C_{ab}.,$$

avec $C_{ab} = \text{Cov} (Z_a - Z_b)$, par hypothèse de stationnarité.

Equations du krigeage.

La quantité $(Z(x) - m)$ étant d'espérance nulle, tout estimateur linéaire sans biais de $Z(x)$ s'écrit:

$$Z^*(x) - m = \sum_a |^a (Z_a - m).$$

L'erreur d'estimation :

$$Z(x) - Z^*(x) = (Z(x) - m) - \sum_a |^a (Z_a - m).$$

est bien d'espérance nulle, et a pour variance:

$$S_E^2 = \text{Var} (Z(x) - Z^*(x)) = \text{Var} \left(Z(x) - \sum_a |^a Z_a \right),$$

$$= \text{Var} (Z(x)) - 2 \sum_a |^a C_{ax} + \sum_a \sum_b |^a |^b C_{ab}.$$

$$= C(0) - 2 \sum_a |^a C_{abx} + \sum_a \sum_b |^a |^b C_{ab}.$$

Le krigeage consiste à trouver les λ^a minimisant cette variance d'estimation.

Nous avons:

$$ds_E^2 / d\lambda^a = 0 = \sum_a \lambda^a C_{ab} = C_{ax}, Y_a \quad (1) \text{ Equation du Krigeage}$$

et

$$S_K^2 = C(0) - \sum_a \lambda^a C_{ax} \quad (2) \text{ Variance du Krigeage}$$

Il est facile de voir par (1) que des échantillons non corrélés, ni avec les points à estimer, ni avec les autres échantillons, ont un poids nul et n'interviennent pas en krigeage simple (K.S.).

2°) Krigeage ordinaire (K.O.) (ou à moyenne inconnue).

Le paramètre m n'étant plus une moyenne connue représente une moyenne locale inconnue.

Ici, l'estimateur s'écrit:

$$Z^*(x) = m + \sum_a \lambda^a (Z_{a-m})$$

et $\sum_a \lambda^a = 1$ est la condition de non-biais.

Dans ce cas, les équations du Krigeage s'écrivent:

$$\sum_a \lambda^a C_{ab} = C_{ax} + m, Y_a$$

$$\sum_a \lambda^a = 1 \quad m \text{ est un multiplicateur de LAGRANGE.}$$

La variance du Krigeage est:

$$S_K^2 = C(0) - \sum_a \lambda^a C_{ax} + m.$$

Nous appliquerons cette théorie sur le département de l'Essonne. Les variables sont issues des fichiers de télédétection classifiés selon une nomenclature "d'occupation du sol". Nous en ferons des estimations locales et globales des différentes catégories de cultures (blé, orge, etc...) et nous en évaluerons la précision. Nous verrons par la suite dans quelle mesure une amélioration de l'utilisation conjointe des enquêtes terrain et de la télédétection en vue des statistiques agricoles s'avère possible.

CONCLUSION.

Notre choix du modèle probabiliste se justifie, non par un test statistique, mais par une confrontation a posteriori avec la réalité. Une estimation acceptable est donc confirmée par des résultats d'exploitation.

BIBLIOGRAPHIE

1. BRION, Ph. Projet de recherche: Sondages spatiaux. I.N.S.E.E.¹- S.C.E.E.S. ² Avril 1988.
2. BRION, Ph. et LABAT, Ph. Sondages: Application aux enquêtes agricoles. S.C.E.E.S. (Ministère de l'Agriculture), Novembre 1984.
3. FOURNIER, Ph. (*), GEROYANNIS H. (**) et GILG, J.P. (**). Evaluation des surfaces de cultures annuelles et inventaire dans le Bassin Parisien par combinaison des enquêtes statistiques classiques et des données S.P.O.T. (P.E.P.S. 127). (*) S.C.E.E.S. (Ministère de l'Agriculture). (**) C.A.M.S.³ - C.N.R.S. ⁴/E.H.E.S.S. ⁵
4. JOURNEL, A.G. Géostatistique Minière -Tomes 1 et 2. Centre de Géostatistique, E.N.S.M.P. ⁶, Fontainebleau 1977.
5. MATHERON, G. La Théorie des Variables Régionalisées et ses Applications. Fascicule n°5, Cahiers de Centre de Morphologie Mathématique. E.N.S.M.P., Fontainebleau 1970.
6. MARBEAU, J.P. Géostatistique Forestière. Thèse de docteur-ingénieur. E.N.S.M.P., Fontainebleau 1976.
7. RIVOIRARD, J. Le comportement des poids de Krigeage. Thèse de docteur-ingénieur. E.N.S.M.P., Fontainebleau 1984.

¹Institut National de Statistiques et des Etudes Economiques.

²Service Central des Etudes et Enquêtes Statistiques.

³Centre d'Analyse et de Mathématique et Sociales.

⁴Centre Nationale de Recherche Scientifique.

⁵Ecole des Hautes Etudes en Sciences Sociales.

⁶Ecole Nationale Supérieure des Mines de Paris.