

**ANALYSE VARIOGRAPHIQUE DE CAMPAGNES
ACOUSTIQUES SUR LE HARENG ÉCOSSAIS**
Philippe Guiblin^a, Jacques Rivoirard^a, E. J. Simmonds^b

I - INTRODUCTION

Le variogramme est, dans l'approche géostatistique, un outil nécessaire aux calculs de variance d'estimation globale. Son estimation préalable est particulièrement difficile lorsque la variable d'intérêt (ici, un indice d'abondance acoustique) a une distribution dissymétrique, comme c'est souvent le cas en halieutique. Différentes manières d'estimer la structure variographique sur des données à distribution dissymétrique sont étudiées ici, sur l'exemple de campagnes acoustiques sur le hareng écossais. Le passage par le logarithme translaté (avec formule de retour), ou par la covariance non centrée, donnent des estimations plus robustes que le classique variogramme expérimental, mais nécessite une hypothèse de stationnarité. La structure est également améliorée, si l'on suppose stationnarité et isotropie, en calculant la structure le long de la route du navire.

A - Matériel et méthodes : Modes de calculs des variogrammes

1) Variogramme des données brutes

En pratique, on cherche à estimer le variogramme de la variable $z(x)$, demi-moyenne de $[z(x+h) - z(x)]^2$ sur le champ à l'aide du variogramme expérimental:

$$\gamma^*(h) = \frac{1}{2n_h} \sum_{i=1}^{n_h} [(z(x_i+h) - z(x_i))^2] \quad (1)$$

où les x_i et $x_i + h$ sont les couples de points de données à distance h , en nombre n_h .

Classiquement, le variogramme est un bon indicateur de la stationnarité dans la mesure où l'on distingue une portée et un palier.

Le variogramme est sensible à la position d'une valeur forte par rapport aux extrémités des transects.

^aCentre de Géostatistique, ENSMP, 35, rue Saint Honoré 77305 Fontainebleau.

^bSOAFD, Marine Laboratory, PO Box 101, Victoria Road, Aberdeen, Scotland, UK.

B - Variogramme des données log-translatées

Les valeurs fortes d'une distribution dissymétrique ont tendance à rendre des statistiques comme le variogramme peu robustes. Un moyen de réduire l'influence des valeurs fortes est de passer au logarithme. Une façon de contourner la difficulté liée à la présence de nombreuses valeurs nulles consiste à prendre le logarithme de la variable translatée par une constante b , soit $Log(b + Z)$, ou ce qui est équivalent mais plus commode (les données nulles restant nulles) : $L = Log(1 + \frac{Z}{b})$.

Pour passer du variogramme γ_L de L au variogramme γ_Z de Z , on fera usage de la formule (pour le cas stationnaire), (Guiblin et al., 1995):

$$\gamma_Z(h) = ((b + m)^2 + var(Z))(1 - e^{-\sigma^2 \frac{\gamma_L(h)}{var(L)}}) \quad (2)$$

où m est la moyenne de Z et

$$\sigma^2 = Log(1 + \frac{var(Z)}{(b + m)^2})$$

C - Variogramme calculé le long de la route du navire

Ceci revient à calculer le variogramme en prenant comme distance entre deux points la distance parcourue par le navire pour aller du premier au second. Tout se passe alors comme si l'échantillonnage était réalisé le long d'une seule ligne: les problèmes générés par la distance des valeurs fortes aux extrêmités de lignes se trouvent réduits. Ce procédé approximatif suppose des portées courtes.

D - Variogramme déduit de la covariance non centrée

Dans le cas stationnaire, le variogramme γ et la covariance C sont liés par la relation :

$$\gamma(h) = C(0) - C(h)$$

Cependant, la covariance dépend de la moyenne de la variable sur tout le champ, et recourir en pratique à une estimation de ce paramètre engendre des biais dans l'estimation de la covariance.

Ce problème disparaît si l'on considère, toujours dans le cas stationnaire, la covariance non centrée:

$$E[Z(x)Z(x + h)] = K(h)$$

qui, elle aussi, ne dépend que de h , et est liée au variogramme par la relation:

$$K(0) - K(h) = \gamma(h)$$

On estime alors la covariance non centrée :

$$K^*(h) = \frac{1}{n_h} \sum_{i=1}^{n_h} z(x_i)z(x_i + h) \quad (3)$$

puis le variogramme :

$$\gamma^*(h) = K^*(0) - K^*(h) \quad (4)$$

II - RÉSULTATS ET DISCUSSION

Nous avons expérimenté différentes manières d'estimer la structure variographique dans un cas où la distribution des valeurs est très dissymétrique. Au vu de l'organisation de la campagne d'échantillonnage 1990 (figure 1a), on recherche des structures dans les directions est-ouest et nord-sud. Le variogramme expérimental classique présente des fluctuations qui rendent difficile la mise en évidence de structures (figure 1b). Le passage par le logarithme translaté réduit ces fluctuations (figure 2a). Sous une hypothèse de stationnarité, justifiée ici, la formule de retour (équation (2)) permet une estimation plus robuste de la structure (figure 2b). Passer par la covariance non centrée permet de renforcer également la robustesse (figure 3a), mais soulignons que ceci nécessite au préalable la stationnarité.

Dans notre exemple et au vu des résultats sur plusieurs années, une hypothèse d'isotropie peut-être raisonnablement admise. Dans ces conditions, et la structure étant a priori courte, l'estimation du variogramme est puissamment améliorée en cheminant le long de la route du navire (figure 3b): on observe des structures très stabilisées et très peu différentes selon le mode de calcul: variogramme classique, variogramme brut estimé à partir de celui du logarithme et estimation obtenue par la covariance non centrée. Une étude plus large sur plusieurs campagnes du même type montrent que les structures apparaissent assez différentes d'une année à l'autre. Ceci peut n'être que le reflet de l'échantillonnage. En admettant qu'en réalité les structures réelles doivent être identiques pour toutes les années, on peut, en principe, utiliser la structure provenant des variogrammes moyens.

Remerciements. Nous remercions l'Union Européenne pour le support apporté à cette étude, réalisée dans le cadre du projet AIR2-94-1007 "Geostatistics for fish stock assessment".

BIBLIOGRAPHIE

GUILBLIN (P.), RIVOIRARD (J.), & SIMMONDS (E.J.), 1995 - Analyse structurale de données à distribution dissymétrique, exemple du hareng écossais. *Cahiers de Géostatistique*, Compte-rendu des journées de Géostatistique 15-16 juin 1995, Paris, E.N.S.M.P. (A paraître)

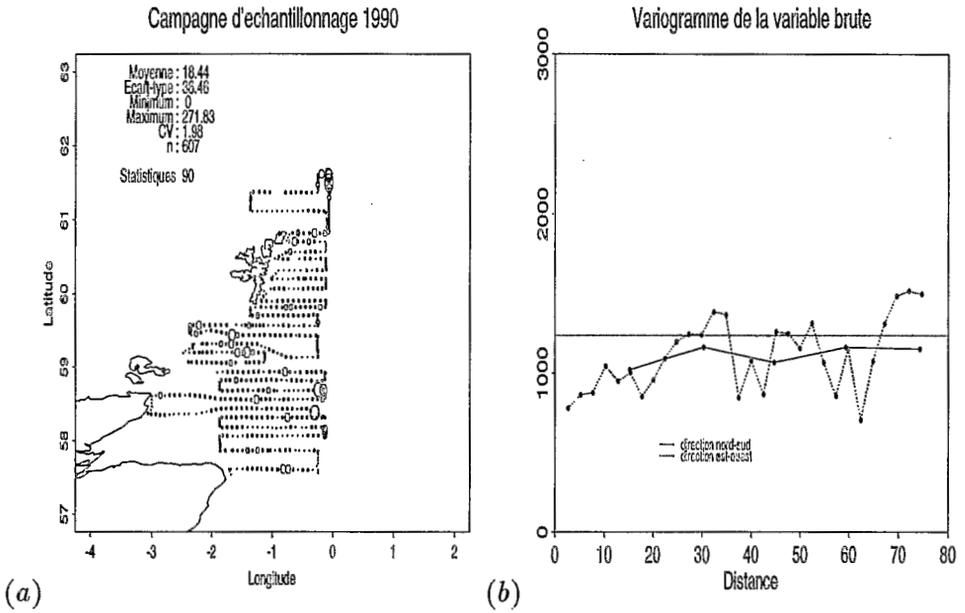


Figure 1. Plan d'échantillonnage (a) et variogramme expérimental calculé à partir des données brutes (b).

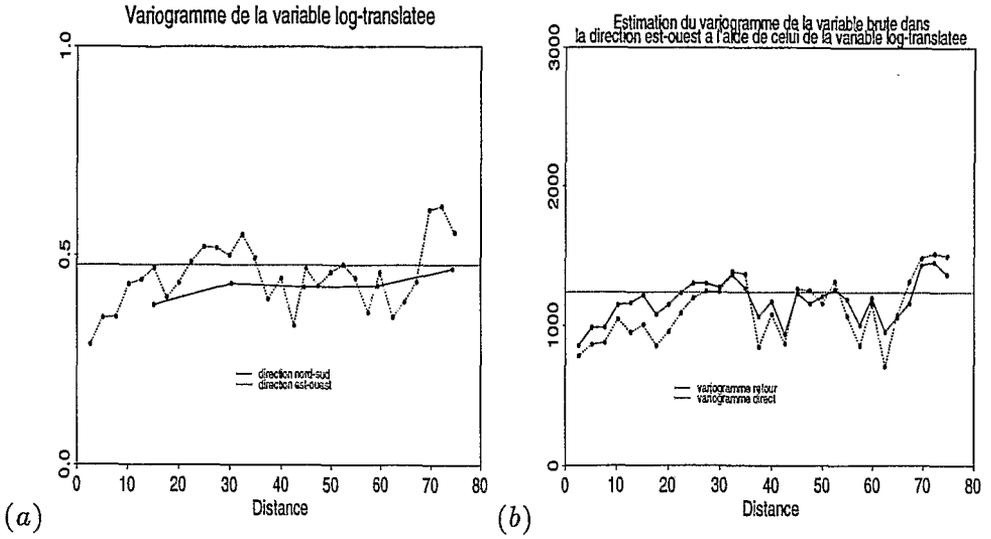


Figure 2. Variogrammes utilisant les données log-translatées.

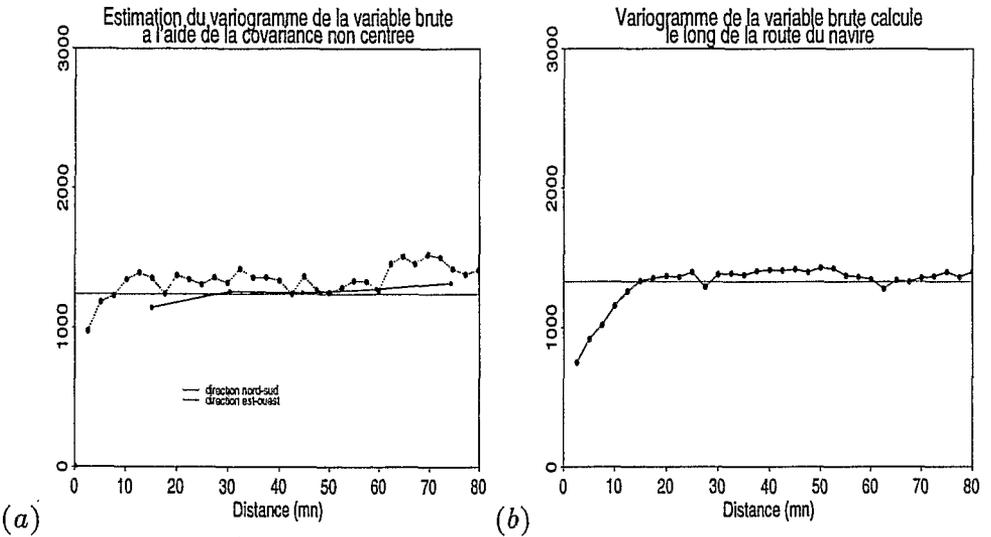


Figure 3. Variogrammes respectivement obtenu à partir de la covariance non centrée (a), et calculé le long de la route du navire (b).