

Approche transitive et densités de poissons

Nicolas BEZ⁽¹⁾, Jacques RIVOIRARD⁽¹⁾ et Jean–Charles POULARD⁽²⁾

(1) Ecole des Mines de Paris, Centre de Géostatistique, 35 rue Saint–Honoré, 77305 Fontainebleau

(2) IFREMER, BP 1049, 44037 Nantes

Résumé. L'approche transitive peut être utilisée pour décrire la distribution spatiale d'une concentration, lorsque l'abondance provient de valeurs fortes et rares, et qu'il est difficile de délimiter un champ à l'intérieur duquel les variations sont homogènes. Divers développements permettent d'estimer et de raffiner la structure spatiale représentée par le covariogramme. On examine en particulier le cas d'un échantillonnage à maille irrégulière. Des données sur le merlu servent à illustrer l'approche.

Abstract. The transitive approach can be used to describe the spatial distribution of a concentration when the main contribution to the abundance comes from a few large values, and when delimitating a domain with homogeneous variations is difficult. Several developments are proposed to estimate and refine the spatial structure given by the covariogram. The case of an irregular sampling pattern is examined in particular. Data on hake are used as an illustration.

Table des Matières

1 — INTRODUCTION	162
2 — COVARIOGRAMME, DÉVELOPPEMENTS, UTILISATIONS	162
3 — ESTIMATION DU COVARIOGRAMME	166
4 — APPLICATION À DES DONNÉES DE MERLU	169
5 — CONCLUSION	174

1 Introduction

En halieutique, ou encore en environnement et en pollution, les concentrations présentent souvent une très forte sélectivité (dans notre étude sur le stock de merlu du Golfe de Gascogne, 10% des densités de poissons suffisent à expliquer 50% de l'abondance totale). Par ailleurs, dans certains cas, les valeurs les plus fortes semblent avoir une localisation préférentielle dans le champ. En biologie des pêches, ceci peut par exemple correspondre à la notion de nourricerie. Une nourricerie est une zone géographique où les juvéniles d'une espèce migratoire comme la sole, se concentrent pour une phase de croissance rapide après leur développement larvaire. La distribution des juvéniles de telles espèces est donc fortement influencée par la forme de la nourricerie (Petitgas, 1991; Koutsikopoulos, 1991). On retrouve également cette tendance dans la distribution spatiale de petits poissons pélagiques comme le maquereau au moment de la ponte. Celle-ci est en effet fortement reliée aux frontières du plateau continental européen (Walsh et Martin, 1986).

Quand l'implantation des densités est préférentielle, peut-on parler de structure interne dans le champ? Dans ces cas-là en particulier, il est difficile de supposer que les patches de densités auraient pu être observés n'importe où dans le champ. De plus, les frontières du champ sont parfois délicates à définir parce que la différence entre les densités nulles ou très faibles n'est pas très nette, ou encore parce que les densités nulles qui se trouvent à l'intérieur de la régionalisation peuvent être interprétées comme externes ou internes à la distribution.

Un contexte méthodologique adapté aux caractéristiques de telles variables est la démarche transitive (Matheron, 1970), où le covariogramme sert d'outil structural. Cependant, le formalisme transitif répond en général aux questions relatives à la variance d'estimation globale dans le cas d'un échantillonnage régulier. Notre propos vise à étendre la démarche transitive à un échantillonnage irrégulier et à étudier le profit qui peut en être tiré pour l'analyse structurale d'une densité de poissons. Ces développements sont illustrés à partir des captures de merlu (*Merluccius merluccius*) obtenues lors de la campagne annuelle organisée en 1990 par l'IFREMER dans le Golfe de Gascogne.

2 Covariogramme, développements, utilisations.

2.1 Définition du covariogramme.

Un phénomène qui se régionalise dans un espace à 1, 2 ou 3 dimensions peut être représenté par une fonction habituellement notée $z(x)$ où x est un point quelconque de cet espace (ayant 1, 2 ou 3 coordonnées respectivement). Cette fonction est encore appelée variable régionalisée (Matheron, 1970). Ce sera ici une concentration (densité de poissons par exemple), prenant donc des valeurs positives ou nulles. Son champ S , que nous supposons fini, correspond à l'ensemble géométrique où cette variable est différente de 0. On notera

$$Q = \int z(x)dx$$

l'abondance totale. Le covariogramme (dit transitif) est défini par (Matheron, 1970):

$$g(h) = \int z(x)z(x+h)dx$$

où h désigne une distance vectorielle. C'est une fonction symétrique en h , telle que : $g(h) \leq g(0)$. Il s'annule dans chaque direction à partir d'une certaine distance (portée). Le comportement plus ou moins régulier de $g(h)$ au voisinage de l'origine (continuité, dérivabilité) traduit la plus ou moins grande régularité spatiale de la variable régionalisée $z(x)$.

Appliqué à l'indicatrice du phénomène, $1(x) = 1$ si $z(x) > 0$ et $1(x) = 0$ sinon, le covariogramme, alors appelé covariogramme géométrique, permet de décrire la géométrie du champ :

$$K(h) = \int 1(x)1(x+h)dx = |S \cap S_{-h}|$$

Il correspond à la surface de l'intersection du champ S avec son translaté du vecteur $-h$, noté aussi S_{-h} (en 0, $K(0) = |S|$). Ainsi, pour un champ sphérique à trois dimensions, le covariogramme

géométrique est identique dans toutes les directions (isotrope), et son expression analytique, connue sous le terme de covariogramme sphérique, ne dépend que du module du vecteur distance :

$$K(h) = \begin{cases} C(1 - \frac{3|h|}{2a} + \frac{|h|^3}{3a^3}) & \text{si } |h| \leq a \\ 0 & \text{si } |h| \geq a \end{cases}$$

Les deux paramètres C et a donnent respectivement la hauteur du covariogramme et sa portée. Le covariogramme peut également servir à la description d'objets anisotropes. Dans le cas d'un champ ellipsoïdique par exemple, on obtiendrait un covariogramme de forme sphérique mais plus ou moins allongée selon la direction (le paramètre a dépendant de la direction).

Le covariogramme (transitif) de la variable régionalisée mélange les informations géométriques du champ et la répartition des densités à l'intérieur. Cet outil structural est donc approprié pour décrire les régionalisations qui dépendent de la forme du champ. Certains stocks halieutiques sont à cet égard de bons exemples dans la mesure où la forme d'un stock est probablement un facteur clef de la distribution spatiale des individus.

2.2 Covariogramme et covariance.

On peut écrire:

$$g(h) = K(h) \frac{\int z(x)z(x+h)dx}{K(h)}$$

expression dans laquelle apparaît, à côté du covariogramme géométrique $K(h)$ du champ, la moyenne de $z(x)z(x+h)$ quand x et $x+h$ appartiennent tous deux au champ, c'est-à-dire la covariance non centrée moyenne de z dans le champ.

A noter qu'en écrivant:

$$g(h) = S \frac{\int_S z(x)z(x+h)dx}{S}$$

le covariogramme lui-même apparaît, au facteur S près, comme une covariance non centrée moyenne à partir du champ, c'est-à-dire lorsque x parcourt le champ S , $x+h$ pouvant ou non appartenir à ce champ.

2.3 Covariogramme et régressions.

Partant de :

$$z(x) = \int_0^{z(x)} dz = \int_0^\infty 1_{z \leq z(x)} dz$$

on peut écrire

$$\begin{aligned} g(h) &= \int z(x)z(x+h)dx \\ &= \int_0^\infty dz \int 1_{z \leq z(x)} z(x+h)dx \\ &= \int_0^\infty dz \int_{S(z)} z(x+h)dx \end{aligned}$$

où $S(z)$ est l'ensemble des points x tels que $z(x) \geq z$. En intégrant par parties, si l'on désigne par $dS(z)$ l'ensemble des points x tels que $z(x) = z$, on obtient :

$$g(h) = \int_0^\infty z dS(z) m[z(x+h)|z(x)=z]$$

où $m[z(x+h)|z(x)=z]$ est la moyenne des valeurs de $z(x+h)$ quand $z(x)=z$. Ainsi, au lieu de sommer directement en x , on commence par sommer en z pour tous les x tels que $z(x)=z$.

La forme du covariogramme, donnée par le covariogramme normé :

$$\frac{g(h)}{g(0)} = \frac{\int z(x)z(x+h)dx}{\int z^2(x)dx}$$

s'écrit encore :

$$\frac{g(h)}{g(0)} = \frac{\int m \left[\frac{z(x+h)}{z(x)} | z(x) = z \right] z^2 dS(z)}{\int z^2 dS(z)} \quad \text{avec } z \neq 0$$

et apparaît comme une moyenne pondérée des régressions de $z(x+h)/z(x)$ sachant $z(x)$ (lesquelles ne sont pas nécessairement symétriques en h). Sa forme traduit donc, en moyenne, la chute relative des densités lorsqu'on s'éloigne de la distance h .

En pratique on aura plutôt recours aux régressions moyennes par classes. Les valeurs de z sont alors rangées dans des classes C_i . La régression moyenne dans la classe C_i s'écrit :

$$m \left[\frac{z(x+h)}{z(x)} | z(x) \in C_i \right] = \frac{\int_{C_i} m \left[\frac{z(x+h)}{z(x)} | z(x) = z \right] z^2 dS(z)}{w_i}$$

avec $w_i = \int z^2 dS(z)$, et le covariogramme normé est la moyenne, pondérée par les w_i , de ces régressions moyennes par classes :

$$\frac{g(h)}{g(0)} = \frac{\sum w_i m \left[\frac{z(x+h)}{z(x)} | z(x) \in C_i \right]}{\sum w_i}$$

2.4 Seuils, ensembles et fréquences conditionnelles.

Pour fixer les idées on supposera dans cette partie la variable $z(x)$ définie à 2D. On s'intéresse aux surfaces $K_{zz'}(h)$ des ensembles géométriques définis par les points x tels que les densités aux points x et $x+h$ dépassent respectivement les seuils (coupures) z et z' :

$$K_{zz'}(h) = \int 1_{z(x) > z} 1_{z(x+h) > z'} dx$$

Ces fonctions correspondent aux covariogrammes géométriques croisés des sous-ensembles du champ de la variable régionalisée définis par les deux coupures z et z' (ils ne sont pas symétriques en h). La somme (double) de ces surfaces pour toutes les valeurs possibles de densités donne :

$$\begin{aligned} \int \int K_{zz'}(h) dz dz' &= \int dx \underbrace{\int 1_{z(x) > z} dz}_{=z(x)} \underbrace{\int 1_{z(x+h) > z'} dz'}_{=z(x+h)} \\ &= \int z(x)z(x+h)dx \\ &= g(h) \end{aligned}$$

Ainsi le covariogramme transitif se décompose en toutes ces fonctions. Il est intéressant d'en regarder l'interprétation fréquentielle. Notons $K_z(h) = K_{zz}(h)$ les covariogrammes simples. Alors $K_z(0) = \int 1_{z(x) > z} dx$ quantifie le nombre de densités supérieures à la coupure z . Par ailleurs $K_{zz'}(h)$ correspond au nombre de densités $z(x)$ telles que à la fois $z(x)$ et $z(x+h)$ dépassent respectivement z et z' . Le rapport de ces deux quantités s'interprète donc comme la fréquence avec laquelle $z(x+h)$ excède le seuil z' sachant que la densité en x est supérieure à z , ce que nous noterons, dans le cadre transitif où nous nous sommes placés, fréquence conditionnelle (plutôt que probabilité conditionnelle qui fait référence à des variables aléatoires) :

$$\frac{K_{zz'}(h)}{K_z(0)} = f(z(x+h) > z' | z(x) > z)$$

Ces fréquences conditionnelles permettent de se placer du point de vue des fortes concentrations (sans se préoccuper des valeurs faibles ou nulles) et en particulier de décrire le caractère attractif ou répulsif des patches de densités.

Remarque : la somme (simple) des fonctions $K_z(h)$ donnerait :

$$\begin{aligned} \int K_z(h) dz &= \int dz \int 1_{z(x) > z} 1_{z(x+h) > z} dx \\ &= \int dx \int 1_{z(x) > z} 1_{z(x+h) > z} dz \\ &= \int dx \int_0^{\min(z(x), z(x+h))} dz \\ &= \int \min(z(x), z(x+h)) dx \\ &= K_{3D}(h) \end{aligned}$$

mesure de l'intersection, entre lui-même et son translaté de $-h$, de l'objet à trois dimensions délimité par :

- le champ de la variable régionalisée pour les deux premières dimensions,
- la variable régionalisée elle-même pour la troisième dimension.

Il s'agit donc de son covariogramme géométrique 3D pour la distance h à 2D.

2.5 Distance entre deux individus.

Supposons que la variable régionalisée $z(x)$ représente la densité d'une population d'individus (poissons par exemple). Dans ce cas $z(x)/Q$, de somme 1, représente la densité de probabilité de la position x d'un individu tiré au sort.

On peut alors montrer que la densité de probabilité du vecteur distance h entre deux individus tirés au sort indépendamment est précisément égale au covariogramme (normé) de $z(x)$, soit $g(h)/Q^2$ (Matheron 1970, exercice 16 p. 47). D'où une interprétation très biologique du covariogramme.

2.6 Estimation globale d'abondance à maille régulière

Dans la pratique, la variable régionalisée n'est pas connue exhaustivement, mais seulement en un certain nombre de points (ou éléments de surface ou de volume) de l'espace. Se pose alors le problème de l'estimation de la quantité globale Q présente dans le champ. Dans le cas d'une grille régulière à 2 dimensions, de maille (a_1, a_2) , où les échantillons z_i sont connus aux nœuds de la grille, l'estimateur :

$$Q^* = a_1 a_2 \sum_i z_i \quad \text{de la quantité totale} \quad Q = \int z(x) dx$$

est sans biais, à condition que l'emplacement de l'origine de la maille soit aléatoire et uniforme au sein de la maille élémentaire de la grille (Matheron, 1970), et que l'ensemble du phénomène soit échantillonné (c'est-à-dire que la variable régionalisée soit, ou puisse être supposée, égale à 0 partout ailleurs sur la grille). Si tel est le cas, Q^* peut être effectivement interprété comme une variable aléatoire, et la variance d'estimation est fournie par la formule suivante (Matheron, 1970) :

$$Var(Q - Q^*) = a_1 a_2 \sum_{k_1} \sum_{k_2} g(k_1 a_1, k_2 a_2) - \int g(h) dh$$

Dans cette formule, k_1 et k_2 décrivent l'ensemble des entiers positifs, négatifs et nul, et les $(k_1 a_1, k_2 a_2)$ représentent les distances vectorielles construites à partir de la maille (a_1, a_2) . La variance d'estimation apparaît comme la différence entre la valeur approchée et la valeur exacte de l'intégrale du covariogramme. Elle est en général d'autant plus petite que la maille est petite et que la fonction $g(h)$, et donc la variable régionalisée elle-même, est régulière. La forme du covariogramme autour de l'origine revêt donc une grande importance.

2.7 Krigeage en représentation externe.

Bien qu'il ne décrive pas le comportement local de la variable régionalisée, le covariogramme peut être utilisé pour une estimation locale (cartographie), précisément dans les cas où les frontières du champ sont inconnues et où seules les valeurs échantillonnées nulles témoignent que l'on n'est plus à l'intérieur du champ.

En chaque point x de l'espace à cartographier, on se propose d'estimer $z(x)$ à l'aide d'une moyenne pondérée des valeurs échantillonnées (zéros compris) à l'intérieur d'un voisinage donné de x , soit en des points $x + h_i$ (Matheron, 1968 et 1978) :

$$z^*(x) = \sum_i \lambda_i z(x + h_i)$$

Les poids λ_i sont choisis de sorte que, s'il était possible de translater partout dans l'espace la configuration formée par les points x et $x + h_i$, la somme des erreurs quadratiques entre les valeurs vraies et estimées $\int (z(x) - z^*(x))^2 dx$ serait minimisée (optimalité). Or :

$$\begin{aligned} \int (z(x) - z^*(x))^2 dx &= \int \left(z(x) - \sum_i \lambda_i z(x + h_i) \right)^2 dx \\ &= \int \left(z(x)^2 - 2 \sum_i \lambda_i z(x) z(x + h_i) + \sum_i \sum_j \lambda_i \lambda_j z(x + h_i) z(x + h_j) \right) dx \\ &= g(0) - 2 \sum_i \lambda_i g(h_i) + \sum_i \sum_j \lambda_i \lambda_j g(h_i - h_j) \end{aligned}$$

On montre que les λ_i qui minimisent cette expression sont solutions du système linéaire suivant :

$$\sum_j \lambda_j g(h_i - h_j) = g(h_i) \quad \forall i$$

Le krigeage en représentation externe peut également être conduit sous contrainte. On peut en effet choisir les poids λ_i sous la contrainte que l'abondance totale Q soit respectée :

$$\int z^*(x) dx = \int z(x) dx \quad \text{ce qui donne} \quad \sum_j \lambda_j = 1$$

Cette condition correspond à la recherche de non biais introduite en krigeage ordinaire. Ainsi, en introduisant le paramètre de Lagrange μ nécessaire à la minimisation sous contrainte, on aboutit au système suivant :

$$\begin{cases} \sum_j \lambda_j g(h_i - h_j) = g(h_i) & \forall i \\ \sum_j \lambda_j = 1 \end{cases}$$

Noter que, avec ou sans contraintes, les poids ne changent pas si l'on multiplie ou si l'on divise le covariogramme par une constante (en considérant par exemple le covariogramme normé).

3 Estimation du covariogramme.

En pratique, le covariogramme $g(h)$ est inconnu. Il faut donc en faire une estimation à partir des valeurs connues z_i . Le covariogramme expérimental ainsi obtenu doit être ensuite ajusté par une courbe, en adoptant le comportement à l'origine le plus approprié (mathématiquement, la fonction ajustée doit être choisie parmi les fonctions de type positif de façon à assurer une variance non négative). C'est ce modèle, au fond une estimation du vrai covariogramme inconnu, qui est utilisé pour un calcul de variance globale ou pour un krigeage.

Dans ce chapitre, on commence par rappeler le calcul du covariogramme expérimental à maille régulière pour ensuite proposer 3 modes d'estimation à maille non régulière.

3.1 Estimation du covariogramme à maille régulière

Matheron (1970) propose l'estimation suivante, écrite ici pour une maille régulière à 2 dimensions (a_1, a_2) d'origine (x_0, y_0) :

$$g^*(k_1 a_1, k_2 a_2) = a_1 a_2 \sum_i \sum_j z(x_0 + i a_1, y_0 + j a_2) z(x_0 + i a_1 + k_1 a_1, y_0 + j a_2 + k_2 a_2)$$

Dans cette estimation tout se passe comme si la valeur de chaque échantillon était étendue à sa surface d'influence, et le cas d'une maille régulière constitue un cas favorable pour l'estimation. Cependant, le formalisme transitif, et donc le recours au covariogramme, est motivé par les caractéristiques de la variable régionalisée (hétérogénéité, influence du champ sur la régionalisation, localisation préférentielle de certaines classes de densités,...), sans restriction a priori sur celles du plan d'échantillonnage. Il convient donc de proposer une estimation du covariogramme pour un échantillonnage non régulier.

3.2 Estimation du covariogramme par migration sur une grille.

Pour retrouver les conditions favorables de l'estimation à maille régulière, une première solution consiste à générer une grille régulière fine et à informer chacun de ses nœuds par l'échantillon le plus proche.

Ce procédé est simple mais la finesse du résultat ne doit pas faire illusion. En particulier cette migration vers une maille fine a tendance à linéariser le covariogramme à l'origine. Pour le voir prenons un exemple simple à une dimension. A partir d'une grille connue de maille a , informons une grille de maille $a/2$. Alors le covariogramme expérimental obtenu grâce à la grille la plus fine à la distance $a/2$ vérifie bien:

$$g\left(\frac{a}{2}\right) = \frac{g(0) + g(a)}{2}$$

Au facteur de discrétisation près (valeur de la maille), chaque échantillon se voit, dans cette méthode, étendu à sa surface d'influence. Lorsque la maille devient très petite, le covariogramme expérimental devient:

$$g^*(h) = \sum_i \sum_j z_i z_j |S_i \cap S_{j-h}|$$

où S_i est la surface d'influence de l'échantillon z_i et S_{j-h} son translaté du vecteur $-h$. Un produit de densités observé *in situ* pour une certaine distance, est donc affecté, dans ce calcul, à toute une gamme de distances, l'éventail des distances étant lié aux formes et positions relatives des surfaces d'influence. Tout se passe comme si la variable régionalisée s'écrivait sous la forme $z(x) = \sum_i z_i 1_{x \in S_i}$. Si maintenant on considère toutes les classes de distance en faisant la somme du covariogramme, on remarque que la contribution globale d'une paire de densités (z_i, z_j) pour l'ensemble des directions et distances possibles est égale au produit des surfaces d'influence $S_i S_j$:

$$\begin{aligned} \int g^*(h) dh &= \int \sum_i \sum_j z_i z_j 1_{x \in S_i} 1_{x+h \in S_j} dh \\ &= \sum_i \sum_j z_i z_j \int \int 1_{x \in S_i} 1_{y \in S_j} dx dy \\ &= \sum_i \sum_j z_i z_j S_i S_j \end{aligned}$$

3.3 Estimation du covariogramme par la covariance.

Cette estimation se fait en deux temps:

1) Estimation du covariogramme géométrique du champ (faite par le procédé précédent en informant l'indicatrice du champ sur une grille).

2) Estimation de la covariance à l'intérieur du champ.

La covariance étant une moyenne et non une somme, on peut pour celle-ci, et c'est l'intérêt de cette méthode, travailler par classes de distances.

Chaque paire (z_i, z_j) intérieure au champ sera alors rattachée à la classe correspondant à la distance entre ces échantillons. De plus, pour tenir compte de l'irrégularité de l'échantillonnage, on donnera à cette paire, dans sa classe, un poids proportionnel à $S_i S_j$. D'où finalement l'estimation:

$$g^*(h) = K^*(h) \frac{\sum \sum_{j \sim i+h} z_i z_j S_i S_j}{\sum \sum_{j \sim i+h} S_i S_j} \quad (j \sim i+h \text{ donne les paires à peu près distantes de } h)$$

Cependant, les classes de distance ne doivent pas être utilisées pour l'estimation à distance strictement nulle, donnée directement par :

$$g^*(h) = K^*(0) \frac{\sum_i z_i^2 S_i}{\sum_i S_i} = \sum_i z_i^2 S_i$$

Il faut s'assurer, dans cette approche (et dans la suivante d'ailleurs aussi), que l'échantillonnage n'est pas préférentiel. Si l'échantillonnage était plus dense à l'endroit des fortes concentrations, la covariance aux distances faibles pourrait n'être calculée qu'avec ces valeurs fortes et être biaisée.

3.4 Estimation du covariogramme par les régressions.

Ecrire le covariogramme:

$$g(h) = \int z(x) dx z(x+h)$$

suggère de l'estimer en sommant les $z_i S_i$ multipliés par la moyenne de $z(x+h)$ pour $z(x) = z_i$. Cependant, à maille irrégulière et si l'on travaille par classes de distances, on peut, à distance h d'un tel point, trouver plusieurs valeurs z_j (positives ou nulles, peu importe). A ces valeurs z_j on affecte des extensions S_j plus ou moins fortes, dont il est naturel de tenir compte. Ceci donnerait comme estimation du covariogramme à distance h :

$$\sum_i z_i S_i \frac{\sum_{j \sim i+h} z_j S_j}{\sum_{j \sim i+h} S_j}$$

Les classes de distances doivent être ainsi choisies que, pour une classe h , à partir de chaque $z(x) = z_i$, il y ait au moins une valeur z_j échantillonnée, à moins que l'on ne tombe hors du champ (ce qui revient à prendre $z_j = 0$). Cependant, le résultat de cette expression n'est pas symétrique en h . Pour estimer le covariogramme qui, lui est symétrique en h , il est préférable de faire la moyenne des résultats obtenus en h et $-h$:

$$g^*(h) = \frac{1}{2} \left(\sum_i z_i S_i \frac{\sum_{j \sim i+h} z_j S_j}{\sum_{j \sim i+h} S_j} + \sum_i z_i S_i \frac{\sum_{j \sim i-h} z_j S_j}{\sum_{j \sim i-h} S_j} \right)$$

Au facteur $g^*(0) = \sum_i z_i^2 S_i$ près, ceci revient à estimer le covariogramme par une moyenne pondérée de $z(x+h)/z(x)$.

Une formule analogue peut être utilisée pour estimer la régression de $z(x+h)/z(x)$ par classe de $z(x)$ (laquelle n'est pas nécessairement symétrique en h):

$$m^* \left[\frac{z(x+h)}{z(x)} \mid z(x) \in C_i \right] = \frac{\sum_{z_i \in C} z_i S_i \frac{\sum_{j \sim i+h} z_j S_j}{\sum_{j \sim i+h} S_j}}{\sum_{z_i \in C} z_i^2 S_i}$$

dans laquelle on ne retient alors que les z_i appartenant à la classe C .

4 Application à des données de merlu.

4.1 Présentation des données.

L'analyse porte sur des données de merlu (*Merluccius merluccius*) issues de la campagne de chalutage organisée, en 1990, par l'IFREMER dans le Golfe de Gascogne. L'échantillonnage est stratifié aléatoire et comprend 137 échantillons (Fig.1). Les densités de merlu d'âge 0 sont exprimées en nombre d'individus par heure de chalutage (ind/h). Elles sont réparties dans la partie centrale du plateau continental; l'ensemble du stock de la région semble avoir été observé (présence de 0 en périphérie). L'irrégularité spatiale apparaît importante puisque de fortes densités sont entourées de densités très faibles voire nulles. 30% des densités sont nulles, et moins de 10% des densités classées par ordre décroissant engendrent 50% du total (forte sélectivité), de sorte que leur distribution est très dissymétrique. La moyenne des 137 densités vaut 130 ind/h, leur médiane 10 ind/h et leur écart-type 284 ind/h (ce qui donne un coefficient de variation de 2.2).

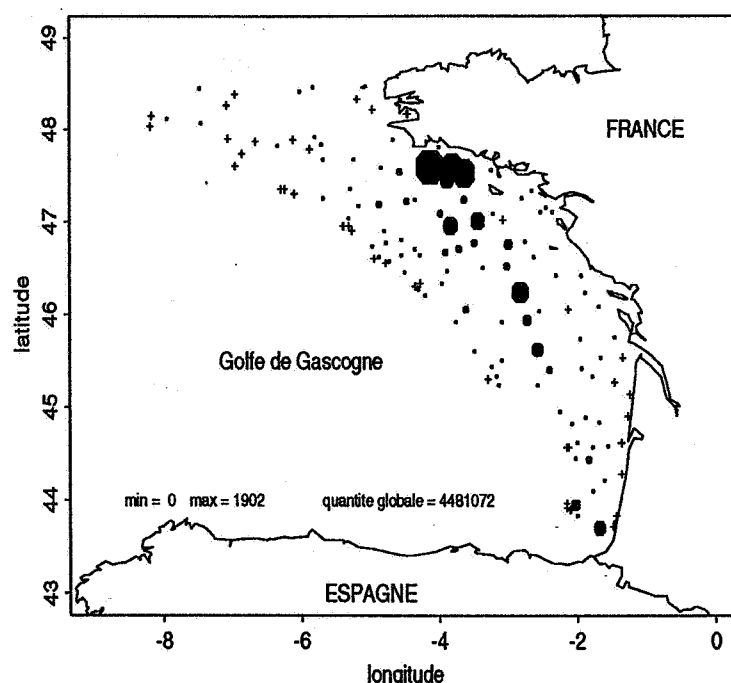


FIG. 1 - Représentation proportionnelle des densités de merlu d'âge 0 - Densités nulles représentées par une croix - Golfe de Gascogne 1990 - Minimum (min), maximum (max) exprimés en nombre d'individus par heure de chalutage - Quantité globale conventionnellement égale à $\sum_i z_i S_i$.

A chaque échantillon, on associe une surface d'influence déterminée par les médiatrices entre points de données et exprimée en milles nautiques au carré (m.n.²). Pour les échantillons de bordure, un cercle de rayon 24 m.n. est intersecté avec le polygone d'influence afin de rendre cohérent l'ensemble des surfaces d'influence. De plus, pour les échantillons les plus à l'Est, les polygones d'influence sont limités aux frontières continentales. L'examen du nuage de corrélation entre densité et surface d'influence (Fig. 2), indique que les concentrations supérieures à 500 ind/h correspondent à des surfaces de tailles variées (de 100 à 600 m.n.²). Il n'y a donc pas de lien important entre densités et surfaces d'influence.

Les résultats structuraux sont présentés pour les quatre directions suivantes, exprimées en degrés trigonométriques à partir de la direction Est-Ouest : 27°, 72°, 117° et 162°.

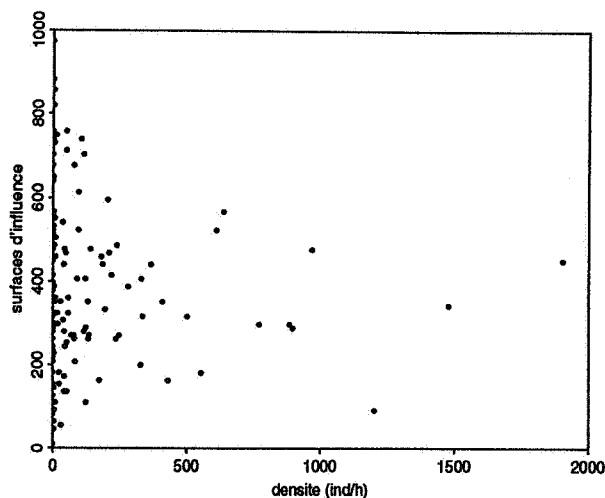


FIG. 2 - Nuage de corrélation entre les surfaces d'influence (m.n.^2) et les densités de merlu (ind/h).

4.2 Estimation du covariogramme.

Estimation du covariogramme par migration sur une grille.

Les données ont été migrées sur une grille de 1 m.n. x 1 m.n. dont les axes sont orientés à 27° et 117° . La valeur en chacun des nœuds de la grille correspond à celle de l'échantillon le plus proche s'il se trouve distant de moins de 24 m.n., ou bien à 0 pour les nœuds chevauchant le continent ou distants de plus de 24 m.n. de l'échantillon le plus proche. Le covariogramme transitif est obtenu en migrant la densité (Fig. 3).

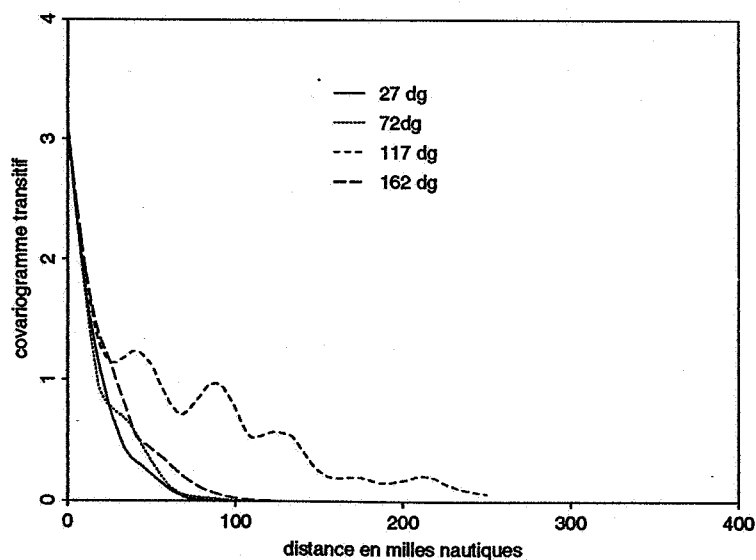


FIG. 3 - Covariogramme transitif expérimental (exprimé en 10^9) - Migration des densités vers une grille régulière 1 m.n. x 1 m.n.

Estimation du covariogramme géométrique par migration sur une grille.

De la même façon, on obtient le covariogramme géométrique (Fig. 4) en migrant l'indicatrice de la densité.

La surface du champ (ensemble des densités non nulles) est estimée à 26000 m.n.^2 (valeur du covariogramme géométrique en 0). Les portées s'échelonnent d'environ 100 m.n. à 350 m.n. selon la direction.

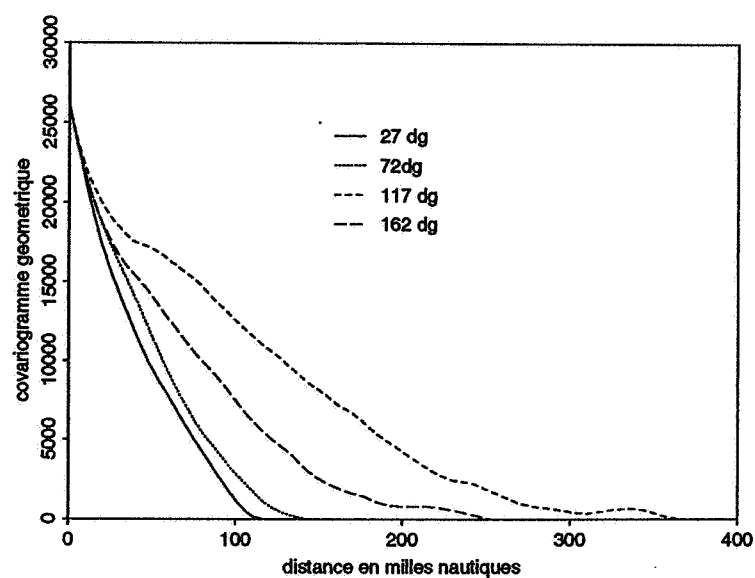


FIG. 4 - Covariogramme géométrique expérimental (exprimé en $m.n.^2$) - Migration des indicatrices des densités vers une grille régulière $1 m.n. \times 1 m.n.$

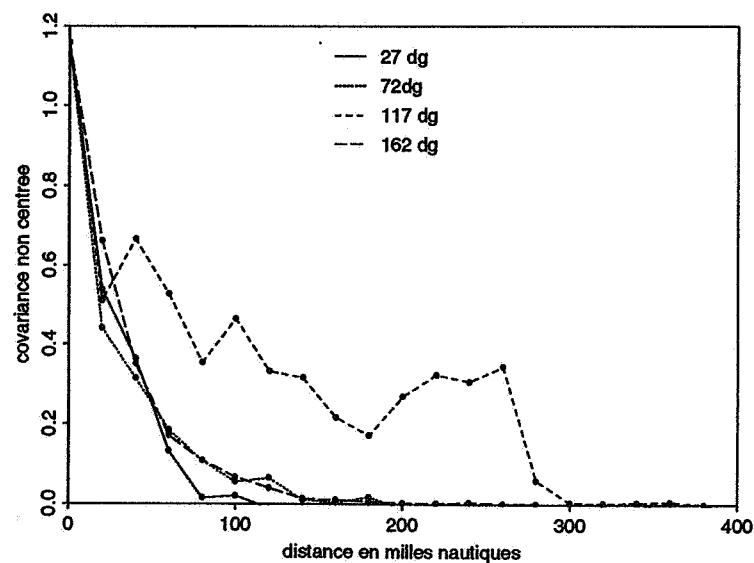


FIG. 5 - Covariance non centrée expérimentale (exprimée en 10^5).

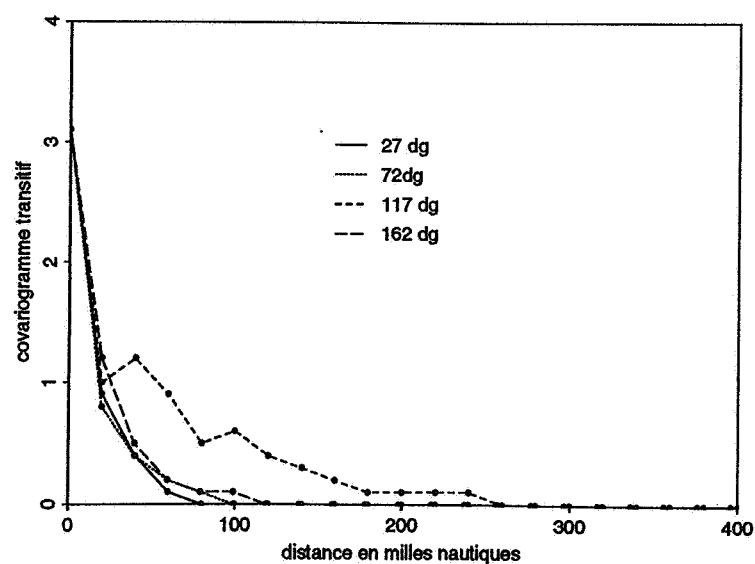


FIG. 6 - Covariogramme transitatif expérimental estimé par la covariance (exprimé en 10^9).

Estimation du covariogramme par la covariance.

La covariance non centrée a été calculée par classe de distance de 20 m.n. avec une tolérance angulaire de 22.5 degrés. La chute au premier pas de calcul est importante. Elle est pratiquement la même dans les quatre directions (Fig. 5). Au-delà, la covariance dans la direction 117° se distingue des trois autres directions par une décroissance lente jusqu'à 300 m.n.

Le produit de cette structure par le covariogramme géométrique fournit une estimation du covariogramme transitif par la covariance (Fig 6).

Estimation du covariogramme comme régression.

Pour ce mode d'estimation les classes de distances ont été choisies égales à 30 m.n. avec une tolérance angulaire de 22.5 degrés. Les résultats sont présentés figure 7.

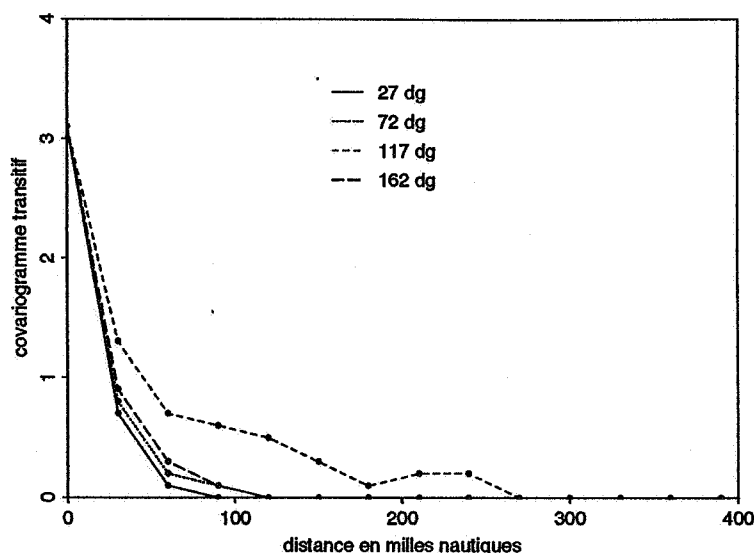


FIG. 7 - Covariogramme transitif expérimental comme régression (exprimé en 10^9).

Comparaison des résultats.

Les différentes estimations du covariogramme transitif montrent globalement les mêmes tendances : décroissance rapide et isotrope autour de l'origine et schéma anisotrope à grande échelle ensuite. La comparaison, pour chaque direction, des covariogrammes obtenus par les trois méthodes d'estimation, ne montre pas de différence importante.

Le covariogramme obtenu par migration, calculé avec un pas très fin, présente des ondulations, atténuées par les autres modes d'estimations, qui peuvent être interprétées comme des effets de trou.

Si la migration vers une grille régulière génère un comportement linéaire à l'origine dont l'interprétation en termes de régularité spatiale doit être faite avec précaution, on retiendra que l'estimation du covariogramme en tant que régression exige des classes de distance d'autant plus larges que l'échantillonnage est irrégulier. Par ailleurs, lors d'une estimation par la covariance, il convient de s'assurer que l'échantillonnage n'est pas trop préférentiel, sans quoi le comportement du covariogramme à l'origine, dont on a vu l'importance sur la variance d'estimation globale par exemple, serait fortement biaisé.

4.3 Estimation des régressions par classe.

Les régressions ont été calculées avec un pas de 30 m.n. et pour trois classes de densités $[0,50[$, $[50,200[$ et $[200,2000[$ d'effectifs respectifs 85, 28 et 24 échantillons (Fig. 8).

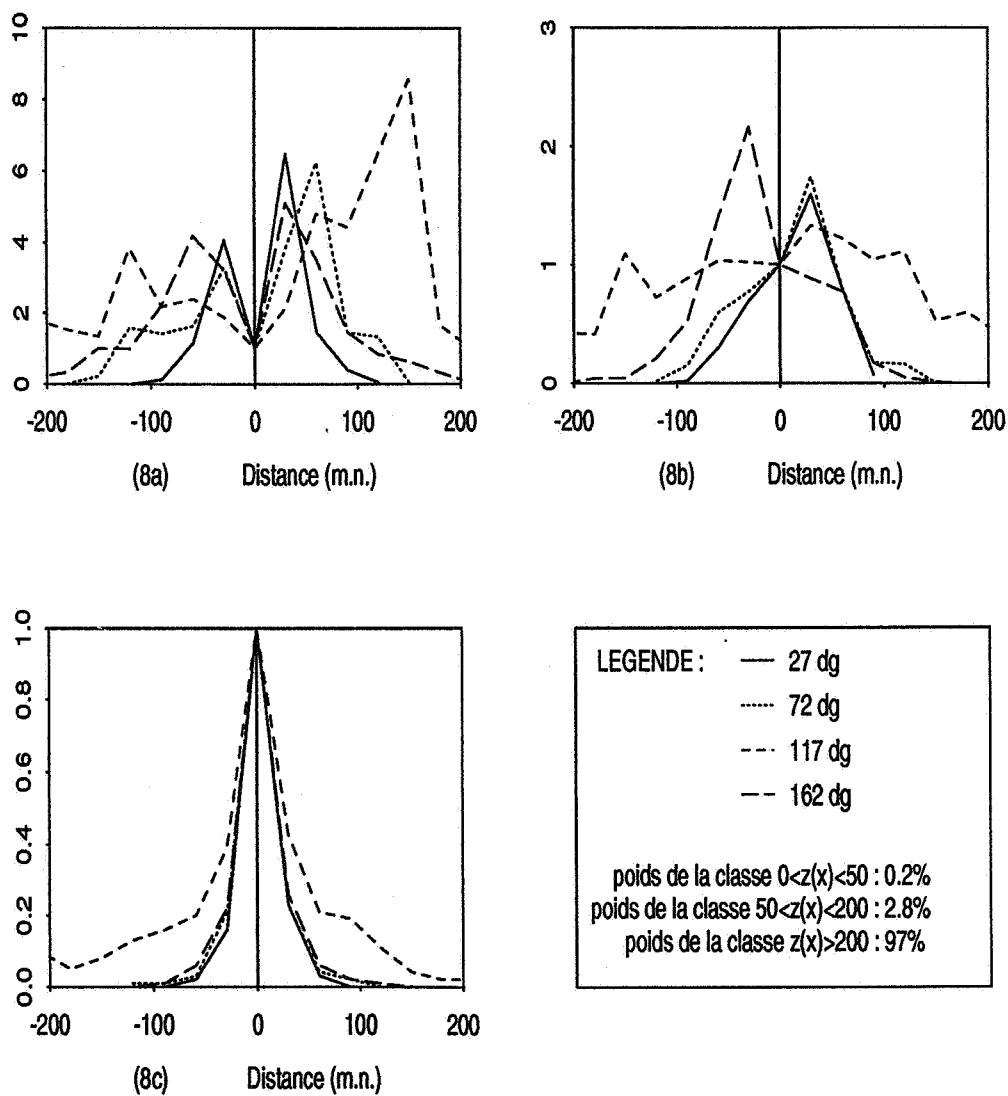


FIG. 8 - Forme de la régression $m \left[\frac{z(x+h)}{z(x)} \mid z(x) = z \right]$ pour trois classes de densités.

- (8a) : $z \in [0, 50[$
 (8b) : $z \in [50, 200[$
 (8c) : $z \in [200, 2000[$

Le poids de la classe des densités fortes dans le covariogramme normé est très élevé (97%). De fait, la forme du covariogramme est très influencée par la régression aux fortes concentrations, c'est-à-dire par la chute relative des densités au voisinage des valeurs élevées (Fig. 8c).

Les régressions concernant les deux premières classes de densités interviennent avec un poids relatif de 3% dans la définition de la forme du covariogramme transitif. Ce sont ces régressions qui présentent le plus de dissymétrie (Fig. 8a et 8b). On remarque en particulier que partant d'une concentration moyenne, la densité augmente pour des petites distances orientées vers l'Est-Nord-Est, et diminue en allant vers l'Ouest-Sud-Ouest (Fig. 8b).

4.4 Estimation des fréquences conditionnelles.

On constate (Fig. 9) une diminution rapide des fréquences relatives aux fortes coupures (>200 ind/h). C'est-à-dire que, sachant qu'on se situe en un endroit où la densité est au moins égale à 200 ind/h, la probabilité que la concentration soit également supérieure à 200 ind/h à 50 m.n. chute considérablement.

4.5 Cartographie par krigeage.

Un modèle a été ajusté sur le covariogramme expérimental obtenu par migration (Fig. 10). La structure choisie est composée d'un schéma sphérique isotrope de portée 20 m.n. et de palier 1.5×10^9 , et d'une structure anisotrope composée d'un schéma sphérique de portée 270 m.n. et de palier 0.55×10^9 et d'un schéma exponentiel de paramètre 100 m.n. (portée pratique 300 m.n.) et de palier 1.06×10^9 (le coefficient d'anisotropie étant de 4 dans la direction 117°). Disposant d'un modèle structural il est alors possible de réaliser une cartographie par krigeage. Le résultat est présenté sous la forme de lignes isodensités (Fig. 11).

5 Conclusion

L'approche transitive est bien adaptée à une variable régionalisée de type concentration, présentant quelques pics de valeurs fortes, beaucoup de valeurs faibles, et pour laquelle il serait difficile de définir un champ à variations homogènes. Cette situation se rencontre en particulier assez souvent en halieutique pour des densités de poissons.

L'utilisation du covariogramme ne se limite pas au calcul de la variance d'estimation à maille régulière. Il peut servir également à réaliser une cartographie par krigeage. De plus, il permet une description structurale de la variable, qui peut être raffinée par les décompositions proposées. En revanche, l'inférence de tous ces outils structuraux à partir d'une maille irrégulière est assez délicate.

Remerciements: Les développements présentés ici ont été motivés par le programme européen AIR 93 1105 SEFOS (Shelf Edge Fisheries and Oceanography Study), dont l'objectif est de relier la variabilité spatio-temporelle de certaines espèces de poissons aux paramètres hydrographiques. L'illustration simple qui en est faite ici porte sur les captures de merlu (*Merluccius merluccius*) obtenues lors de la campagne annuelle organisée en 1990 par l'IFREMER dans le Golfe de Gascogne.

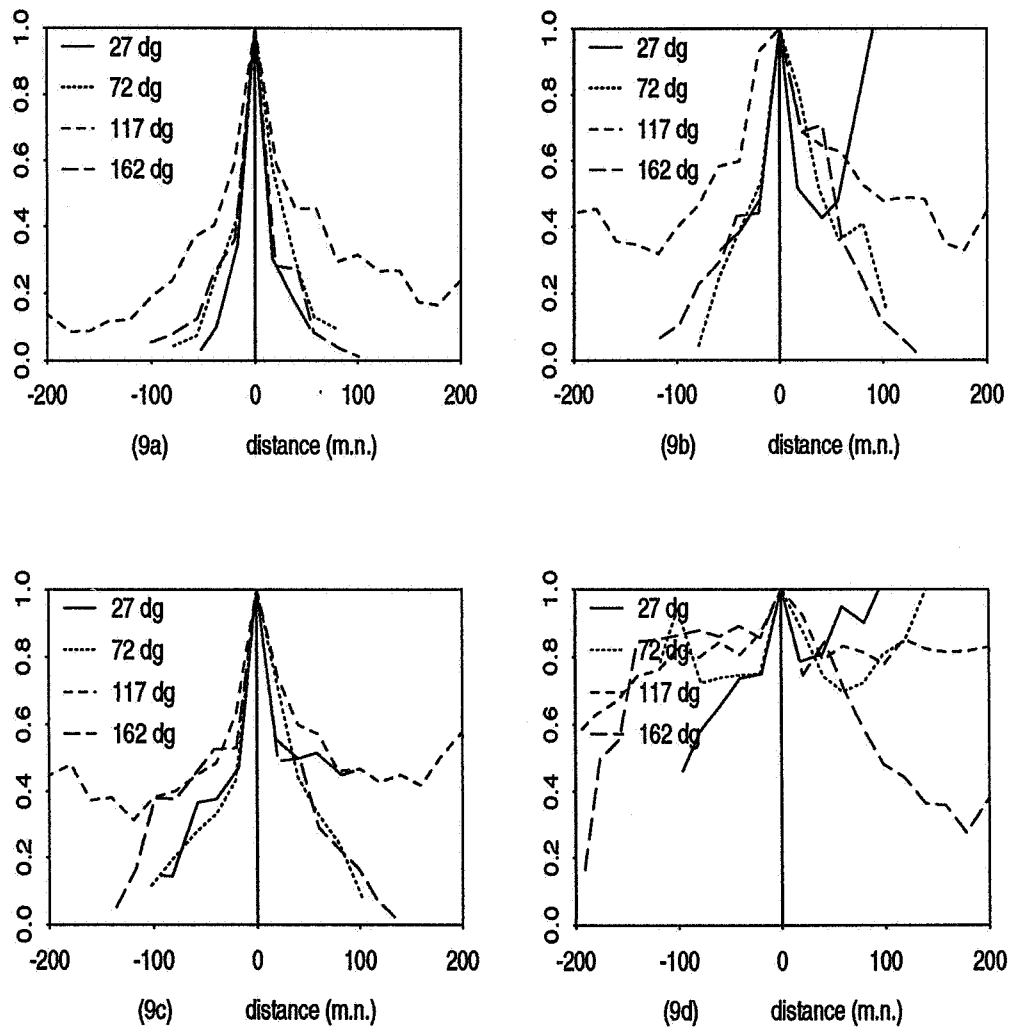


FIG. 9 - Fréquences conditionnelles pour différents couples de coupures.

$$(9a) : f(z(x+h) > 200 | z(x) > 200)$$

$$(9b) : f(z(x+h) > 50 | z(x) > 200)$$

$$(9c) : f(z(x+h) > 50 | z(x) > 50)$$

$$(9d) : f(z(x+h) > 0 | z(x) > 50)$$

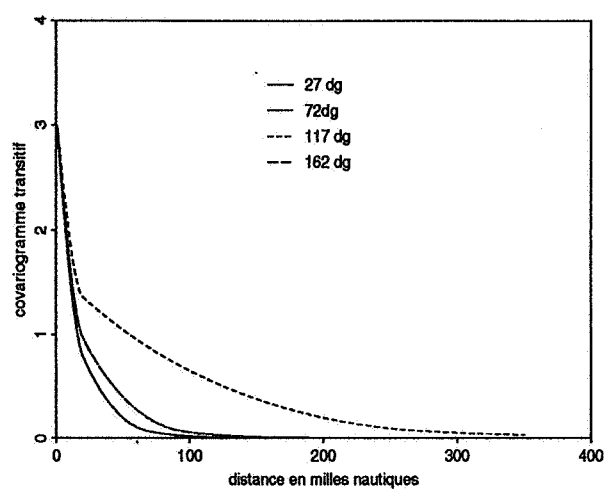


FIG. 10 - Covariogramme transitif (exprimé en 10^9) - Modèle dans 4 directions.

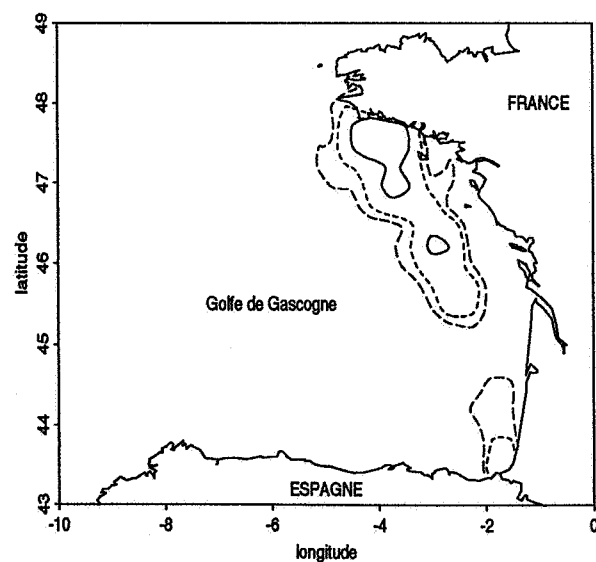


FIG. 11 - Krigeage en représentation externe sans contrainte - Courbes iso-densités de merlu d'age 0 (niveaux 100, 200, 500 ind/h).

6 Références

- Koutsikopoulos C., 1991. Recrutement de la sole (*Solea solea*, L.) du Golfe de Gascogne : influence de l'hydrologie et de l'hydrodynamisme. Thèse de Doctorat d'Océanographie Biologique, UBO.
- Matheron G., 1968. Le krigeage en représentations transitives. Note géostatistique N° 71, N-67, E.N.S.M.P., Fontainebleau, France, 14p.
- Matheron G., 1970. La théorie des variables régionalisées et ses applications. Les Cahiers du Centre de Morphologie Mathématique, Fascicule 5, E.N.S.M.P., Fontainebleau, France, 212p.
- Matheron G., 1978. Estimer et choisir. Les Cahiers de Morphologie Mathématique, Fascicule 7, E.N.S.M.P., Fontainebleau, France, 175 p.
- Petitgas P., 1991. Contributions géostatistiques à la biologie des pêches maritimes. Thèse Dr. en Géostatistique, E.N.S.M.P., Fontainebleau, France. 211 p.
- Walsh M., J.H.A. Martin, 1986. Recent changes in the distribution and migrations of the western mackerel stock in relation to hydrographic changes. ICES CM 1986/H:17.