

Analyse structurale de données à distribution dissymétrique : exemple du hareng écossais

Philippe GUIBLIN⁽¹⁾, Jacques RIVOIRARD⁽¹⁾ et E.J. SIMMONDS⁽²⁾

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

(2) SOAFD, Marine Laboratory, PO Box 101, Victoria Road, Aberdeen, Scotland, UK

Résumé. Différentes manières d'estimer la structure variographique sur les données à distribution dissymétrique sont étudiées ici, sur l'exemple de six campagnes acoustiques sur le hareng écossais. Indépendamment de toute lognormalité, le passage par le logarithme translaté (avec formule de retour) donne une estimation plus robuste que le classique variogramme expérimental, mais nécessite une hypothèse de stationnarité. Il en est de même du passage par la covariance non centrée. La structure est également améliorée, si l'on suppose stationnarité et isotropie, en calculant la structure le long de la route du navire.

Abstract. Several ways of estimating the variographic structure on skewly distributed data are studied, with the example of six acoustic surveys on Scottish herring. Independently of lognormality, the use of log-translated data (with a back transformation) gives a more robust estimation than the classical experimental variogram, but requires a stationarity hypothesis. The same applies when the non centered covariance is using. The structure is also improved if stationarity and isotropy are assumed when computing the structure along the track of the ship.

Table des Matières

1 — INTRODUCTION	138
2 — PRÉSENTATION DES DONNÉES	138
3 — MODES DE CALCUL DES VARIOGRAMMES	140
3.1. Calcul des variogrammes directionnels	140
3.2. Utilisation du logarithme	140
3.3. Calcul du variogramme le long de la route du navire	141
3.4. La covariance non centrée	141
4 — ANALYSE VARIOGRAPHIQUE	142
4.1. Données brutes	142
4.2. Données log-translatées	142
4.3. Variogrammes calculés le long de la route du navire	142
4.4. Estimation du variogramme des données brutes à partir de celui des données log-translatées	143
4.5. Estimation du variogramme à partir de la covariance non centrée	143
5 — CONCLUSION	144
ANNEXE	145

1 Introduction

Dans les cas stationnaires et intrinsèques, le variogramme est nécessaire pour les calculs de variances (variance d'estimation globale, par exemple) et pour le krigeage.

Cependant, lorsque la variable régionalisée a une distribution dissymétrique, comme très souvent en halieutique, certaines statistiques comme le variogramme sont peu robustes.

Cet article étudie d'autres moyens d'estimer la structure variographique : l'utilisation du logarithme translaté (avec formule de retour pour le cas stationnaire), et la covariance non centrée (dans le cas stationnaire également).

Ceci est illustré par des données de densité de harengs obtenues au cours de campagnes de prospection acoustique effectuées par bateau. La modélisation du variogramme expérimental ne sera pas envisagée ici.

2 Présentation des données

Six campagnes de mesures acoustiques ont été effectuées de 1989 à 1994 au Nord-Est de l'Écosse. Elles sont représentées sur la figure 1. En réalité, les campagnes s'étendent plus à l'Ouest, mais pour des raisons d'homogénéité structurale, nous ne considérerons ici que la zone située à l'Est des îles.

La grandeur mesurée est un indice d'abondance (surface réfléchissante, unité: m^2/mn^2). Les campagnes se présentent de la manière suivante:

La prospection acoustique se fait le long de transects parallèles orientés Est-Ouest. Les données sont régularisées tous les 2.5 milles nautiques. La distance inter-transect est d'environ 15 milles nautiques. Un schéma régulier est à peu près respecté, excepté par endroit où une prospection en zig-zag a été effectuée. La figure 1 donne une représentation proportionnelle des données pour chaque campagne ainsi que des statistiques globales. Pour chaque campagne le cercle de rayon maximal représente la valeur maximale, les autres cercles s'en déduisent proportionnellement. Les points représentent les valeurs nulles strictement. On retrouve une caractéristique fréquente des données halieutiques: quelques valeurs fortes isolées au milieu de valeurs très faibles ou nulles. Cette caractéristique générale rend l'analyse structurale plus délicate. L'importance numérique des valeurs nulles ou sub-nulles est résumée par le tableau 1.

On ne remarque sur les cartes proportionnelles aucune zone d'occupation préférentielle du poisson. Chaque année, il existe des paquets isolés de valeurs fortes mais non situés forcément aux mêmes endroits.

Les échantillons situés entre les transects, notamment en bordure Est du champ, ont été retirés des analyses, ceci pour ne pas biaiser la recherche d'une structure dans la direction Nord-Sud. L'analyse structurale porte donc sur les plans d'échantillonnage de la figure 2. Les statistiques élémentaires, tableaux 1 et 2, montrent des variations d'une année sur l'autre: la proportion des valeurs nulles est variable, la valeur des maxima est très variable. Les coefficients de variation sont supérieurs à 1 et caractérisent une distribution groupée (ou aggrégative) [4]. On remarque que les plus fortes moyennes (années 1989 et 1990) sont associées aux coefficients de variation les plus faibles: diminution du stock et surdispersion semblent aller de pair. On retiendra, pour la suite de l'étude, la variance particulièrement élevée de l'année 1991.

De telles distributions sont très dissymétriques. L'utilisation des données log-translatées, ici $\text{Log}(1 + \frac{Z}{12})$, Z étant la variable brute (voir plus loin), permet de manipuler une variable moins dispersée, les valeurs nulles restent inchangées et l'importance des valeurs fortes est relativisée (tableau 3). Il est à remarquer que cette transformation est utilisée indépendamment de tout hypothèse lognormale. En effet, vue la proportion de valeurs nulles, la variable brute ne saurait obéir à une distribution lognormale, même après translation.

Année	89	90	91	92	93	94
valeurs = 0	39%	32%	51%	53%	40%	68%
valeurs \leq 1	42%	34%	59%	55%	49%	70%

TAB. 1 - Proportion des valeurs nulles et sub-nulles

Année	89	90	91	92	93	94
moyenne	16.83	18.01	11.95	9.06	8.87	6.33
variance	944.31	1239.21	3221.10	689.19	1378.52	537.39
minimum	0.00	0.00	0.00	0.00	0.00	0.00
maximum	281.21	271.83	907.13	340.66	533.36	305.31
Nombre d'observations	497	501	462	465	500	490
Coefficient de variation	1.83	1.95	4.75	2.90	4.18	3.66

TAB. 2 - Statistiques élémentaires - Données brutes

Année	89	90	91	92	93	94
moyenne	0.59	0.59	0.32	0.33	0.29	0.22
variance	0.44	0.48	0.37	0.30	0.26	0.23
minimum	0.00	0.00	0.00	0.00	0.00	0.00
maximum	3.20	3.16	4.34	3.38	3.82	3.27
Nombre d'observations	497	501	462	465	500	490
Coefficient de variation	1.11	1.16	1.92	1.66	1.78	2.17

TAB. 3 - Statistiques élémentaires - Données log-translatées

3 Modes de calcul des variogrammes

3.1 Calcul des variogrammes directionnels

On rappelle qu'une fonction aléatoire $Z(x)$ est intrinsèque si ses accroissements $Z(x+h) - Z(x)$ sont d'espérance nulle et si leur variance ne dépend que du vecteur distance h , ce qui définit le variogramme [3]:

$$\gamma(h) = \frac{1}{2} \text{Var}(Z(x+h) - Z(x)) = \frac{1}{2} E[(Z(x+h) - Z(x))^2]$$

En pratique, on cherche à estimer le variogramme régional, 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 de la fonction aléatoire dans la mesure où l'on distingue une portée et un palier [1], [2].

Dans le cas d'un échantillonnage (à peu près) régulier, on calcule le variogramme expérimental pour les distances h multiples de la maille élémentaire: le variogramme expérimental apparaît alors exactement comme une estimation, par discrétisation de l'espace, du variogramme régional.

Remarquons que, dans le cas d'une distribution dissymétrique, les couples dans lesquels interviennent les valeurs les plus fortes influent énormément sur le variogramme. En particulier, le long d'une ligne, une valeur forte intervient dans deux couples pour toute distance h inférieure ou égale à la distance de la valeur forte à chacune des deux extrémités de la ligne, puis intervient dans un seul couple, et enfin n'intervient plus aux distances supérieures. D'où un comportement irrégulier du variogramme [5].

3.2 Utilisation du logarithme

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. Mais ceci présente l'inconvénient d'exagérer les écarts entre les valeurs faibles, sans compter le problème des valeurs nulles pour lesquelles le logarithme n'est pas défini. Un moyen de contourner cette difficulté consiste à prendre le logarithme de la variable translatée par une constante b , soit $\text{Log}(b + Z)$, ou ce qui est équivalent mais plus commode (les données nulles restant nulles): $L = \text{Log}(1 + \frac{Z}{b})$. Ainsi, si Z est petit devant b , L est équivalent à $\frac{Z}{b}$, tandis que si Z est grand devant b , L est équivalent à $\text{Log}(\frac{Z}{b})$. En pratique, la constante b , par rapport à laquelle Z est considérée comme faible ou forte, est choisie de l'ordre de grandeur de la moyenne de Z (la valeur exacte de b ne jouant alors pratiquement pas sur la forme du variogramme). On a pris, ici, $b = 12$ (voisine de la moyenne de Z pour les différentes années).

Naturellement, la structure du logarithme est, en général, différente de la structure brute. Pour passer du variogramme γ_L de L au variogramme γ_Z de Z , on fera usage de la formule (pour le cas stationnaire):

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

où m est la moyenne de Z et

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

Cette relation se démontre dans le cas où le logarithme (népérien) L est bigaussien, mais également dans celui où L est une somme de jetons aléatoires valués (voir annexe). Ici, on se bornera à essayer la formule, en tant que moyen de robustifier la structure brute.

3.3 Calcul du variogramme 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 extrémités de lignes se trouvent réduits. Le procédé est intéressant si l'on suppose, a priori, stationnarité et isotropie. Mais il faut également supposer que les structures sont suffisamment courtes pour que les liaisons entre transects successifs ne génèrent pas de biais importants. Dans cette approche, des distances courtes impliquent des intervalles de temps courts; aussi peut-on s'attendre à ce que les éventuelles variations temporelles soient réduites.

3.4 La covariance non centrée

Dans le cas stationnaire d'ordre 2, la fonction aléatoire $Z(x)$ possède une espérance m indépendante de x et une covariance ne dépendant que de h :

$$\text{Cov}(Z(x), Z(x+h)) = E[(Z(x) - m)(Z(x+h) - m)] = E[Z(x)Z(x+h)] - m^2 = C(h)$$

Une telle fonction aléatoire est évidemment intrinsèque, et on a :

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

L'inverse n'est pas forcément vrai, une fonction aléatoire intrinsèque à variogramme non borné (linéaire, par exemple) n'étant pas stationnaire. C'est pourquoi, le variogramme est un outil théoriquement plus général que la covariance stationnaire. De plus, la covariance dépend du paramètre m , 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 évidente:

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

Une méthode intéressante pour estimer, sans biais, le variogramme consiste donc, une fois admise l'hypothèse de stationnarité, à estimer 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 :

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

A la différence du variogramme expérimental, les couples faisant intervenir une valeur forte ne donnent pas tous une contribution forte: ainsi, le produit par une valeur nulle donnera zéro et la covariance ne sera pas sensible aux distances entre valeurs fortes et extrémités de lignes.

4 Analyse variographique

4.1 Données brutes

Les variogrammes expérimentaux sont présentés figure 3, ils ont été calculés dans la direction des transects et dans la direction perpendiculaire aux transects à partir des données présentées figure 2. Un pas de 2.5 milles nautiques est utilisé dans la direction Est-Ouest et de 15 milles nautiques dans la direction Nord-Sud. Les variogrammes sont présentés simultanément pour chaque année avec des figurés différents.

Vue la représentation géographique des valeurs de densités, ce sont en grande partie des structures courtes que l'on cherche à mettre en évidence. Mais les fluctuations sont telles qu'aucune structure n'est clairement visible, ni dans l'une, ni dans l'autre direction. On observe surtout des effets de pépite, seules peut-être les années 89 et 90 présentent dans la direction Est-Ouest une structure. Les variogrammes calculés dans la direction Nord-Sud sont plutôt pépitiques. L'hypothèse de stationnarité est plausible dans la mesure où l'on devine un palier, sauf pour l'année 91 où le variogramme Est-Ouest prend une allure parabolique à partir de 20 milles nautiques, due à la valeur maximale située en bordure du champ. Il est classique de considérer que le variogramme n'est valable que jusqu'à une distance de l'ordre du tiers ou de la moitié du champ (autour de 60-80 mn). Il est risqué d'avancer à ce stade de l'analyse variographique qu'un palier clairement égal à la variance est identifiable.

Si une structure stable existe et se répète d'année en année, il est intéressant de regarder le variogramme moyen sur les 6 campagnes. Cela n'est pas le cas, néanmoins on a choisi de le représenter à titre d'exemple. Dans les deux directions Est-Ouest et Nord-Sud, on calcule pour chaque pas la moyenne, pondérée par le nombre n_h de couples, des 6 valeurs de $\gamma^*(h)$. Ceux-ci sont présentés sur la figure 11-a. On remarque surtout l'influence de l'année 91, la plus variante. Le variogramme moyen n'a un intérêt que si aucune année n'impose sa structure, ce qui n'est pas le cas ici. La structure du variogramme moyen dans la direction Nord-Sud reste pépitique.

4.2 Données log-translatées

Comme précédemment, un variogramme dans deux directions perpendiculaires est calculé pour chaque campagne (figure 4). Dans l'ensemble les variogrammes sont stabilisés. Les structures sont bien visibles. Dans la direction Est-Ouest, on distingue des portées de 5 à 20 mn et un palier de l'ordre de la variance. La direction Nord-Sud conduit à des effets de pépite. Vraisemblablement, s'il existe une structure spatiale, elle n'est pas visible sur le variogramme dans la direction Nord-Sud.

Le variogramme moyen sur les 6 campagnes est présenté figure 11-b. Dans la direction Est-Ouest, on observe une composante pépitique, puis une structure à 5 mn, enfin une croissance due à l'année 91. On observe un effet de pépite dans la direction Nord-Sud.

4.3 Variogrammes calculés le long de la route du navire

La variable pouvant être considérée comme stationnaire, n'importe quelle valeur peut se trouver n'importe où dans le champ. De plus, cette variable peut être considérée comme isotrope. En effet, d'après ce qui précède, la structure Nord-Sud n'est pas visible et la portée de la structure Est-Ouest est courte et inférieure à la distance inter-transect. Une méthode pour stabiliser le variogramme consiste à le calculer le long de la route du navire. Ceci ne peut être justifié que dans la mesure où la portée n'excède pas la distance inter-transects.

Ces variogrammes sont calculés comme des variogrammes à une dimension, le pas

d'échantillonnage étant de 2.5 milles nautiques. Ils sont calculés à partir des plans d'échantillonnages comprenant les données relevées entre les transects présentés figure 1.

On observe (figures 5 et 6) des structures beaucoup plus visibles qui restent courtes. Les pépites sont assez élevées (75% de la variance). Les paliers sont bien stabilisés sur la variance. Les variogrammes calculés sur les données log-translatées donnent des structures beaucoup plus nettes avec des effets de pépité relativement plus faibles.

En moyenne sur 6 années la stabilisation est aussi très nette (figure 12-a et b). L'allure générale est particulièrement satisfaisante sur la figure 12-b.

4.4 Estimation du variogramme des données brutes à partir de celui des données log-translatées

Ces dernières observations nous incitent à utiliser la formule de retour (formule 2). Connaissant le variogramme estimé de la variable log-translatée on revient à celui de la variable brute. Cette relation étant utilisée indépendamment d'un modèle, on la justifiera de manière empirique à partir des variogrammes calculés le long de la route du navire.

Sur les figures 7 et 8, on a représenté, superposés, les variogrammes expérimentaux calculés à partir des données brutes (traits pointillés) et ceux déduits des variogrammes expérimentaux calculés sur les données log-translatées (traits pleins). Sur la figure 7, on observe une bonne superposition des deux variogrammes. D'où une justification empirique de la formule de retour. Celle-ci est alors utilisée pour l'estimation des variogrammes dans la direction Est-Ouest (figure 8): on constate que le passage par le logarithme stabilise le variogramme et donne des structures mieux visibles.

4.5 Estimation du variogramme à partir de la covariance non centrée

Dans ce paragraphe, on considère un autre outil, la covariance non centrée, et on examine ce qu'il peut apporter en pratique.

On calcule dans chaque direction de l'échantillonnage un variogramme expérimental estimé à partir de la covariance non centrée expérimentale (formule 3 puis 4).

Sur la figure 9 sont présentés les variogrammes estimés par ce moyen, ils sont à comparer avec ceux de la figure 3. Les pics descendants observés correspondent à des distances entre valeurs fortes. On observe en général, une bonne stabilisation, avec des structures à petites distances en Est-Ouest et des variogrammes pépitiques en nord-sud. Passer par le calcul des covariances non centrées pour l'estimation n'a un intérêt que si l'on travaille sur les données brutes. Travailler sur les données log-translatées augmente le nombre de paquets de valeurs fortes et la robustification est nettement moindre. C'est pourquoi on a choisi de ne pas les représenter.

Le variogramme moyen sur les 6 années (figure 14) est toujours très influencé par l'année 91.

Les variogrammes calculés le long de la route du navire par l'intermédiaire de la covariance non centrée sont présentés sur la figure 10. Ils sont peu différents des variogrammes précédents (figure 9, sauf pour 90). Mais ils sont également extrêmement proches des variogrammes calculés directement le long de la route du navire (figure 5).

5 Conclusion

Nous avons expérimenté différentes manières d'estimer la structure variographique dans un cas où la distribution des valeurs est très dissymétrique. Le variogramme expérimental classique présente des fluctuations qui rendent difficile la mise en évidence de structures. Le passage par le logarithme translaté réduit ces fluctuations. Sous une hypothèse de stationnarité, justifiée ici, la formule de retour (2) permet une estimation plus robuste de la structure. Passer par la covariance non centrée permet de renforcer également la robustesse, mais soulignons que ceci nécessite au préalable la stationnarité.

Dans notre exemple, 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: 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.

Les structures apparaissent assez différentes d'une année à l'autre. Bien sûr, 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. Cependant, une année particulièrement variée peut imposer sa structure à ces variogrammes.

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".

Références

- [1] Chauvet P. - Aide mémoire de géostatistique linéaire. Cours de DEA de Géostatistique, Ecole de Mines de Paris, 1993.
- [2] Journel A.G., Huijbregts Ch.J. - Mining géostatistics. Academic Press, London, 1978.
- [3] Matheron G. - La théorie des variables régionalisées et ses applications. Les cahiers du centre de Morphologie mathématique, fasc. 5, Ecole de Mines de Paris, 1970.
- [4] Ramade F. - Éléments d'écologie, écologie fondamentale, Mc Graw-Hill, Paris, 1991.
- [5] Rivoirard J. - Computing variograms on uranium data, in Geostatistical case studies, G.Matheron and M.Armstrong ed., Reidel Pub., Dordrecht, pp 1-22, 1984.

ANNEXE

Une formule de retour pour le variogramme du log-translaté

Le cas lognormal

Sous une hypothèse de bilognormalité, on sait repasser de la covariance stationnaire des logarithmes à celle de la variable brute. Rappelons-en les éléments de démonstration.

Pour cela, partons de la fonction caractéristique d'un couple bigaussien (L_1, L_2) sous la forme :

$$E(e^{\alpha L_1 + \beta L_2}) = e^{\alpha\mu_1 + \beta\mu_2 + \frac{1}{2}\alpha^2\sigma_1^2 + \frac{1}{2}\beta^2\sigma_2^2 + \alpha\beta\sigma_{12}}$$

où $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12}$ sont respectivement les moyennes, variances et covariance de L_1 et L_2 , et où α et β sont des constantes (réelles ou complexes).

Appliquant cette formule au logarithme népérien $L(x)$ d'une Fonction Aléatoire Stationnaire $Z(x) = e^{L(x)}$, on obtient les moments :

$$E(Z(x)^\alpha Z(x+h)^\beta) = E(e^{\alpha L(x) + \beta L(x+h)})$$

En particulier, faisant $\beta = 0$, on trouve le moment d'ordre α de $Z(x)$:

$$m_\alpha = E(Z(x)^\alpha) = e^{\alpha\mu + \frac{1}{2}\alpha^2\sigma^2}$$

d'où en particulier la moyenne :

$$m = E(Z(x)) = e^{\mu + \frac{\sigma^2}{2}}$$

et le moment d'ordre 2 :

$$m_2 = E(Z(x)^2) = e^{2\mu + 2\sigma^2} = m^2 e^{\sigma^2}$$

et donc la variance :

$$\text{var} Z(x) = m^2(e^{\sigma^2} - 1)$$

Dans ces expressions μ est la moyenne de $L(x)$, et σ^2 sa variance, autrement dit la variance logarithmique de $Z(x)$. Désignant par $\rho(h)$ le corrélogramme de $L(x)$, sa covariance s'écrit $\sigma^2\rho(h)$. On obtient alors, en faisant $\alpha = \beta = 1$:

$$E(Z(x)Z(x+h)) = m^2 e^{\sigma^2\rho(h)}$$

et donc, pour la covariance stationnaire :

$$C_Z(h) = m^2(e^{\sigma^2\rho(h)} - 1)$$

Les jetons multiplicatifs

En réalité, la formule précédente, établie dans le cas bilognormal, est également valable dans au moins un autre modèle, celui des jetons multiplicatifs.

Partons, pour cela, de points Poissonniens x_i de densité θ (le nombre de points tombant à l'intérieur d'un domaine D est une variable de poisson de paramètre $\theta|D|$). Implantons en chacun de ces points un jeton (éventuellement aléatoire) de covariogramme $K(h)$, de valuation X_i indépendante et de même loi qu'une variable aléatoire $X > 0$.

Posons enfin:

$$Z(x) = \prod_{i=1}^{N(x)} X_i$$

produit des valuations X_i pour les $N(x)$ jetons recouvrant le point x (si $N(x) = 0$, on pose $Z(x) = 1$).

Ce nombre $N(x)$ est une variable aléatoire de Poisson de paramètre $\theta K(0)$. On en déduit le moment d'ordre α de $Z(x)$. En effet, à $N(x)$ fixé, on a: $E(Z(x)^\alpha) = [E(X^\alpha)]^{N(x)}$ et donc, en moyennant pour toutes les valeurs de $N(x)$:

$$m_\alpha = E(Z(x)^\alpha) = E[E(X^\alpha)^{N(x)}] = e^{\theta K(0)(EX^\alpha - 1)}$$

En particulier:

$$m = E(Z(x)) = e^{\theta K(0)(EX - 1)}$$

et

$$m_2 = E(Z(x)^2) = e^{\theta K(0)(EX^2 - 1)}$$

Considérons maintenant 2 points x et $y = x + h$. Fixons le nombre de jetons N_c recouvrant à la fois x et y , $N_p(x)$ celui des jetons recouvrant x mais non y et $N_p(y)$ celui des jetons recouvrant y mais non x , et donc $N(x) = N_c + N_p(x)$ et $N(y) = N_c + N_p(y)$. On a:

$$\begin{aligned} E[Z(x)^\alpha Z(y)^\beta] &= [E(X^\alpha)]^{N(x)} [E(X^\beta)]^{N(y)} \\ &= [E(X^{\alpha+\beta})]^{N_c} [E(X^\alpha)]^{N_p(x)} [E(X^\beta)]^{N_p(y)} \end{aligned}$$

N_c , $N_p(x)$ et $N_p(y)$ sont indépendants et Poissonniens, de paramètre $\theta K(h)$ pour le premier, et $\theta[K(0) - K(h)]$ pour les deux autres.

En moyenne pour tous les nombres possibles de jetons, on obtient:

$$E[Z(x)^\alpha Z(x+h)^\beta] = e^{\theta K(h)(EX^{\alpha+\beta} - 1) + \theta[K(0) - K(h)](EX^\alpha + EX^\beta - 2)}$$

En particulier pour $\alpha = \beta = 1$:

$$E[Z(x)Z(x+h)] = m^2 e^{\theta K(h)(EX^2 - 2EX + 1)}$$

qu'on peut encore écrire, avec $\rho(h) = \frac{K(h)}{K(0)}$ et σ^2 une constante positive:

$$m^2 e^{\sigma^2 \rho(h)}$$

d'où finalement la covariance stationnaire de $Z(x)$:

$$C_Z(h) = m^2 [e^{\sigma^2 \rho(h)} - 1],$$

relation formellement identique au cas lognormal, et dans laquelle $\rho(h)$ désigne toujours le corrélogramme du logarithme.

En effet celui-ci:

$$\text{Log} Z(x) = \sum_i^{N(x)} \text{Log} X_i,$$

est une somme de jetons aléatoires, méthode classique de simulation d'une Fonction Aléatoire de covariance proportionnelle à $K(h)$. Il obéit donc bien au corrélogramme $\rho(h) = \frac{K(h)}{K(0)}$.

Le log-translaté

Considérons maintenant le cas où ce n'est plus $Z(x)$, mais $1 + \frac{Z(x)}{b}$ qui est:

- soit bilognormal ($Z(x)$ pouvant alors prendre des valeurs négatives $> -b$);

– soit produit de jetons.

La covariance de $Z(x)$ s'écrit alors:

$$C_Z(h) = (m + b)^2(e^{\sigma^2 \rho(h)} - 1)$$

avec $\rho(h)$ corrélogramme de $\text{Log}(1 + \frac{Z(x)}{b})$, m moyenne de $Z(x)$, et σ^2 vérifiant:

$$\text{var}Z = (m + b)^2(e^{\sigma^2} - 1)$$

soit

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

Passant en variogramme, avec $\gamma_L(h)$ variogramme de $\text{Log}(1 + \frac{Z(x)}{b})$, mais toujours dans le cas stationnaire, on obtient finalement la formule de retour:

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

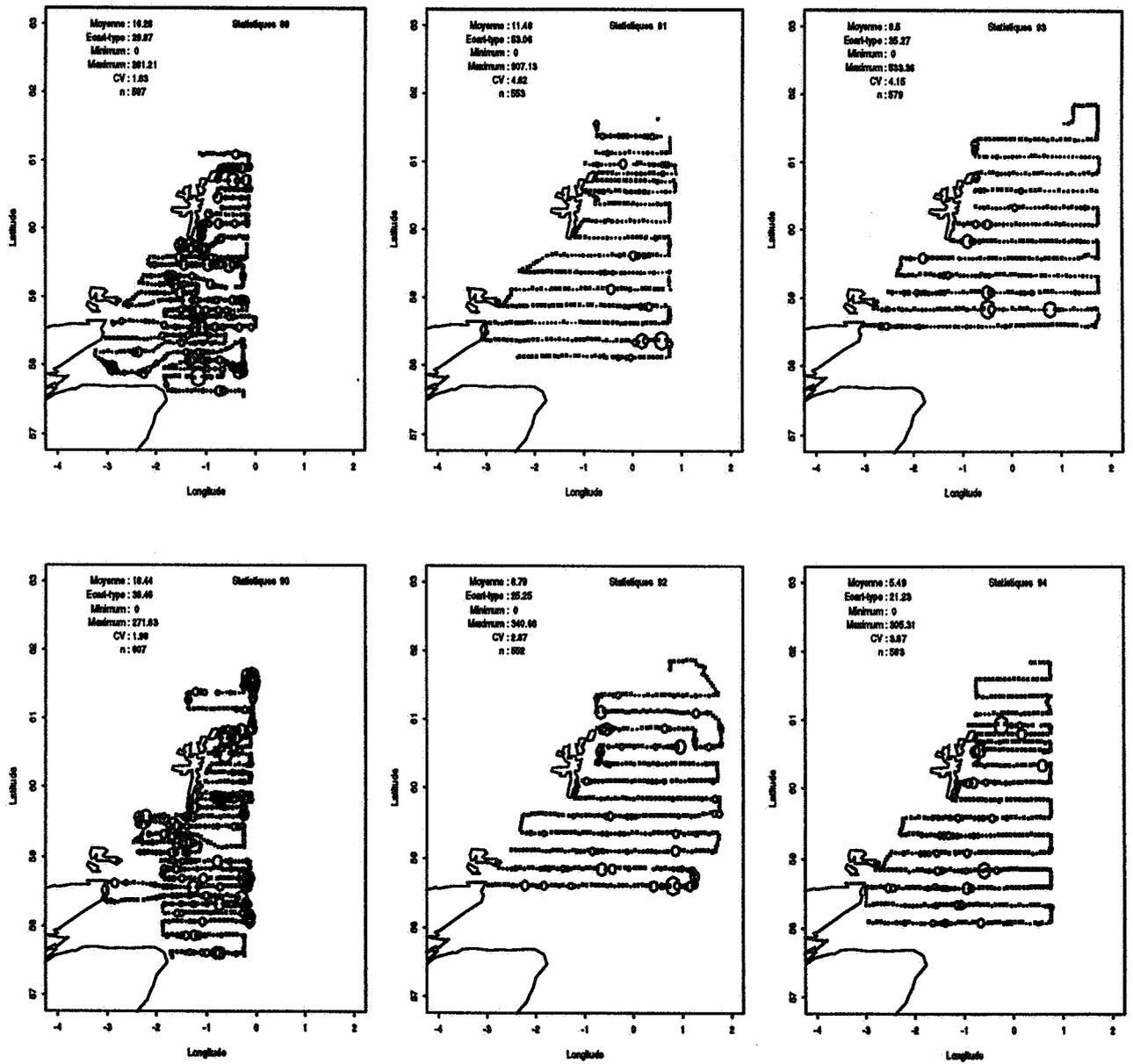


FIG. 1 - Représentations proportionnelles des données brutes - statistiques élémentaires

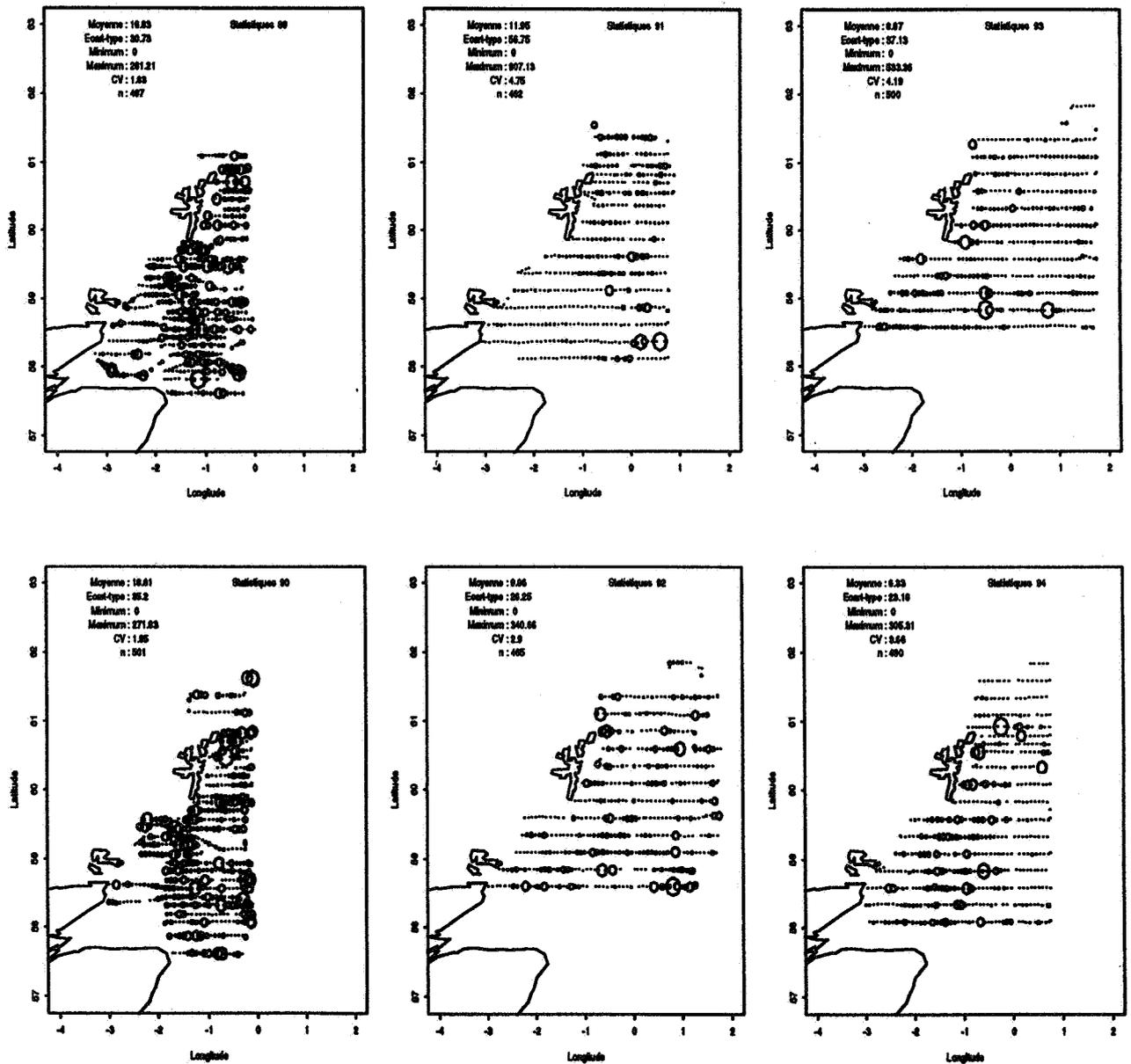


FIG. 2 - Représentations proportionnelles des données brutes - statistiques élémentaires (sans les données relevées entre les transects)

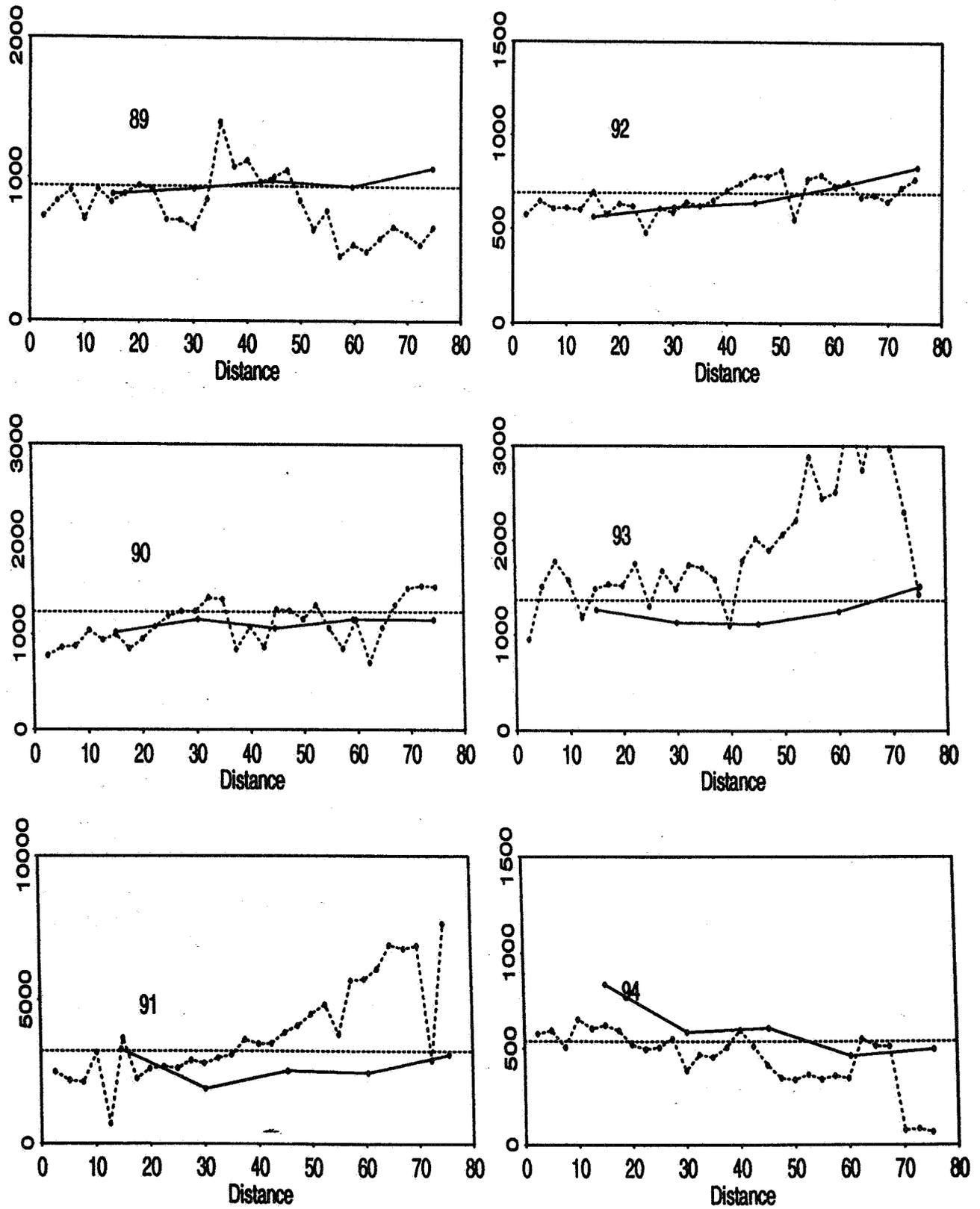


FIG. 3 - Variogrammes directionnels - données brutes
 —direction Nord-Sud -....direction Est-Ouest

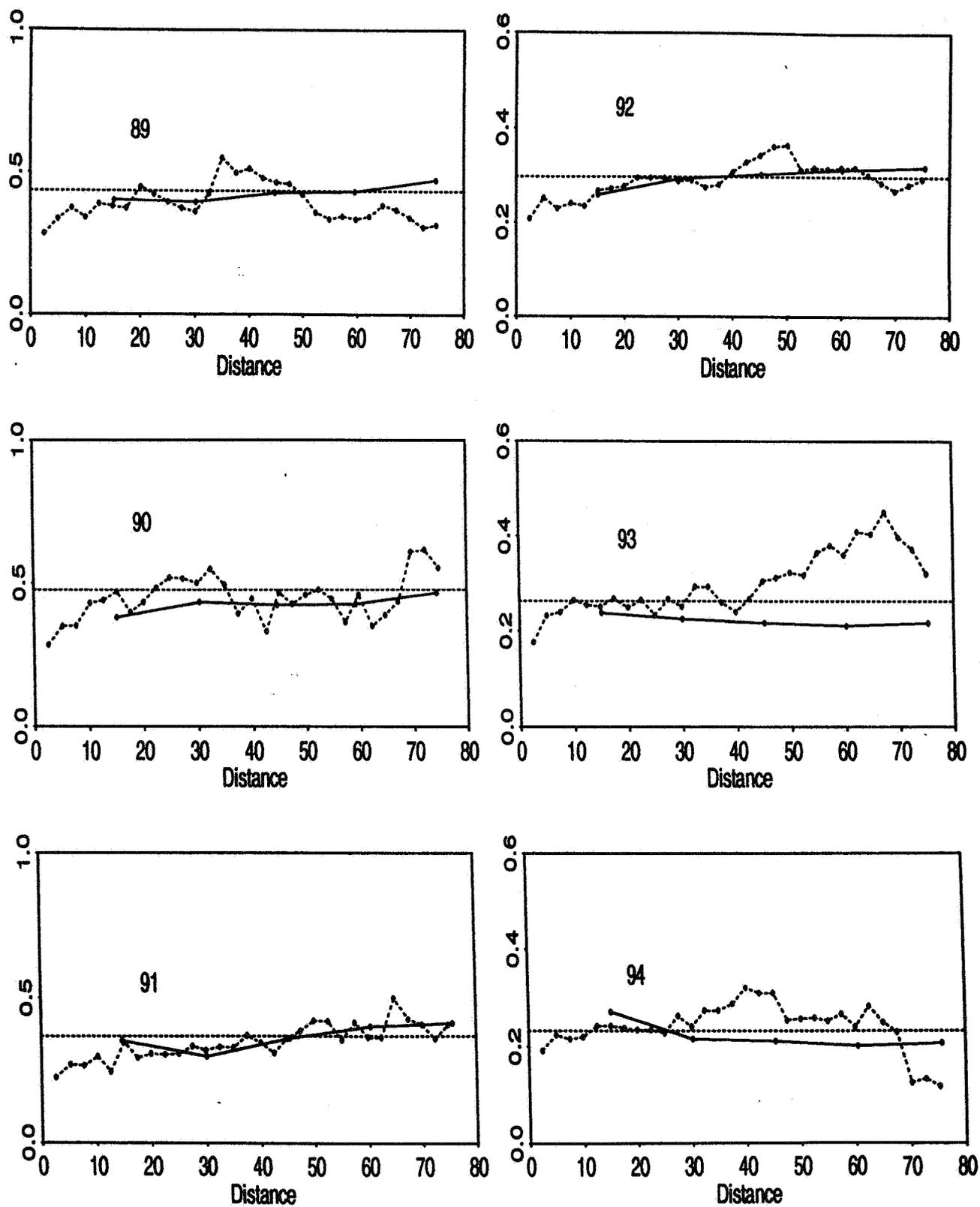


FIG. 4 - Variogrammes directionnels - données log-translatées
 —direction Nord-Sud -....direction Est-Ouest

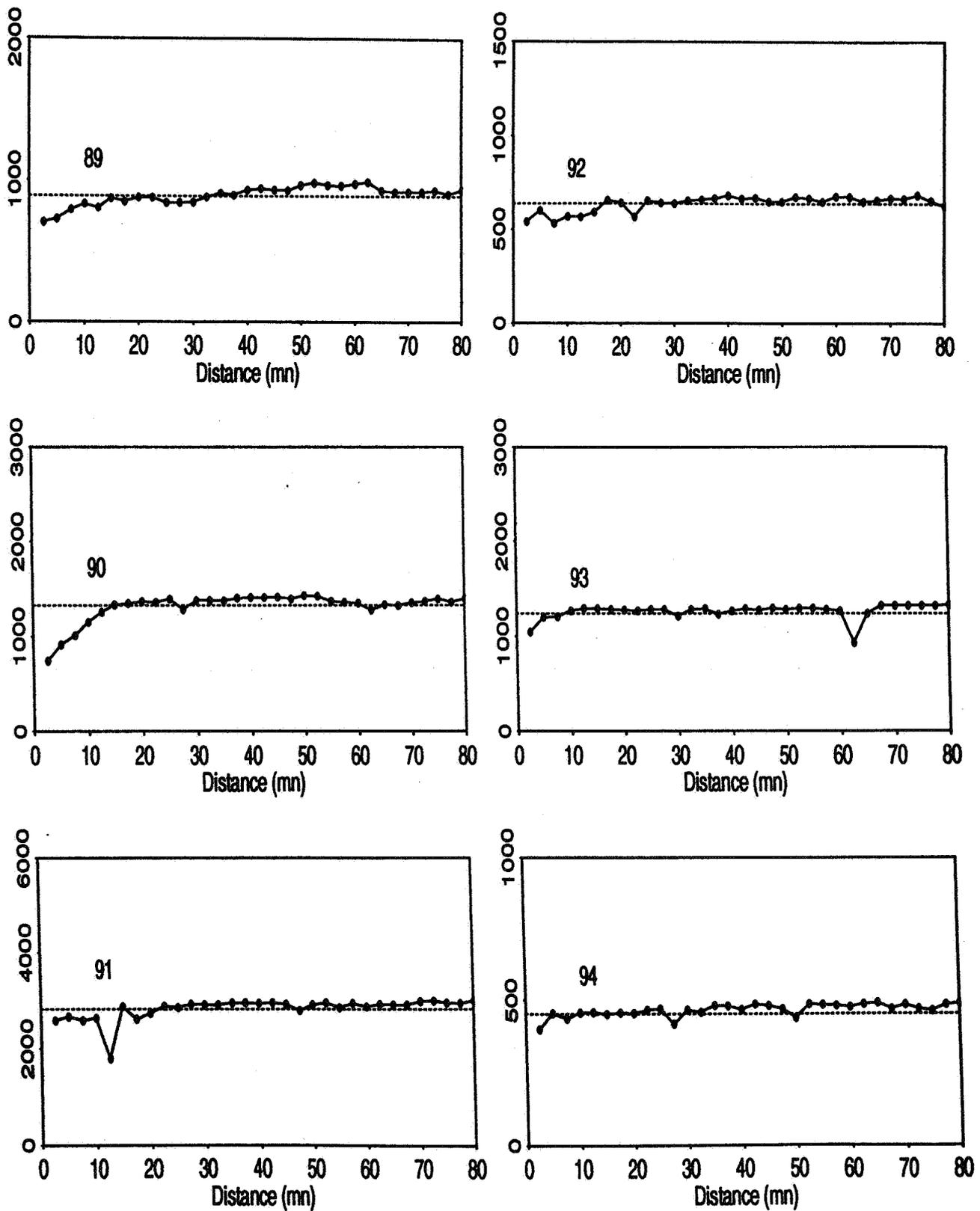


FIG. 5 - Variogrammes calculés le long de la route du navire - données brutes

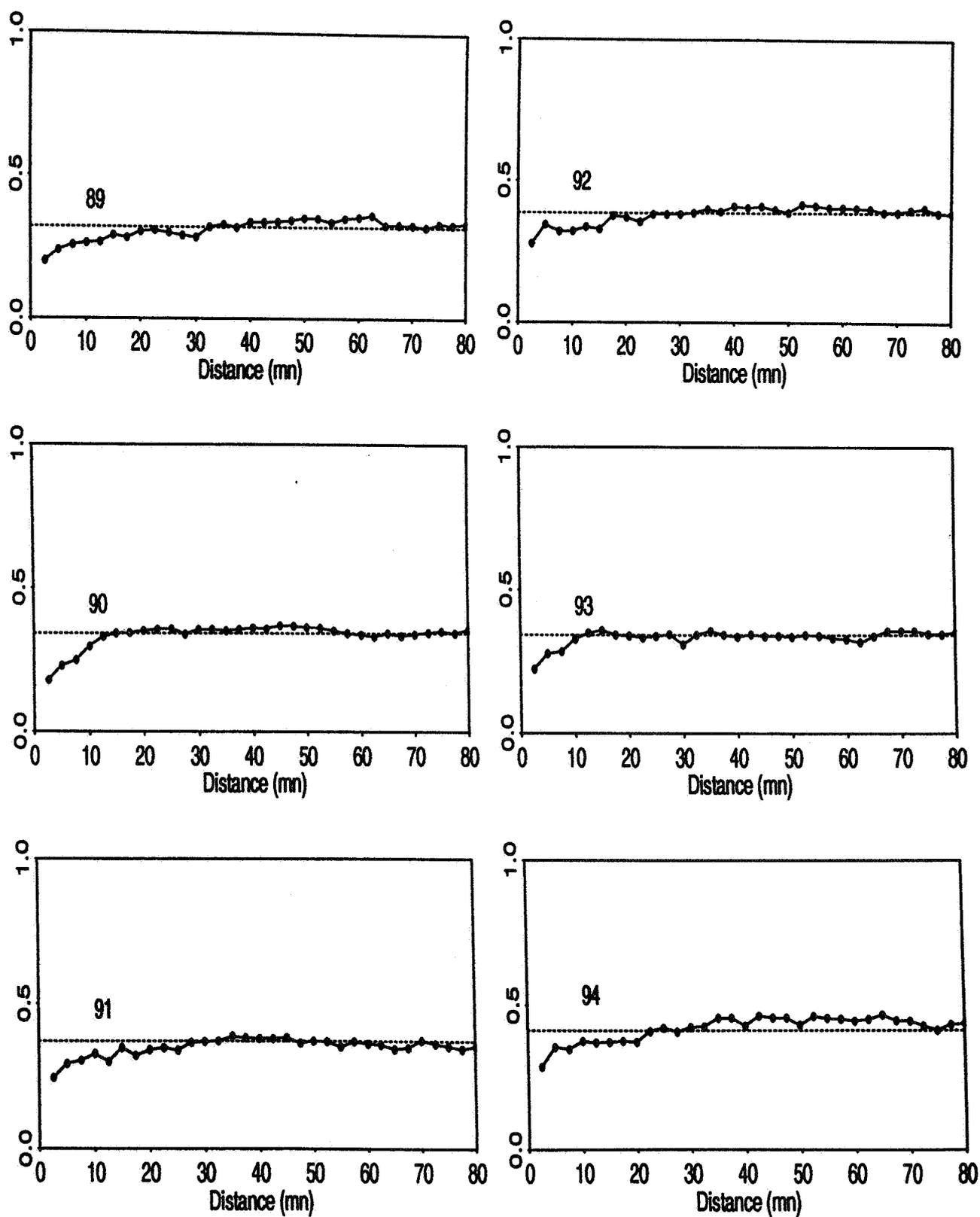


FIG. 6 - Variogrammes calculés le long de la route du navire - données log-translatées

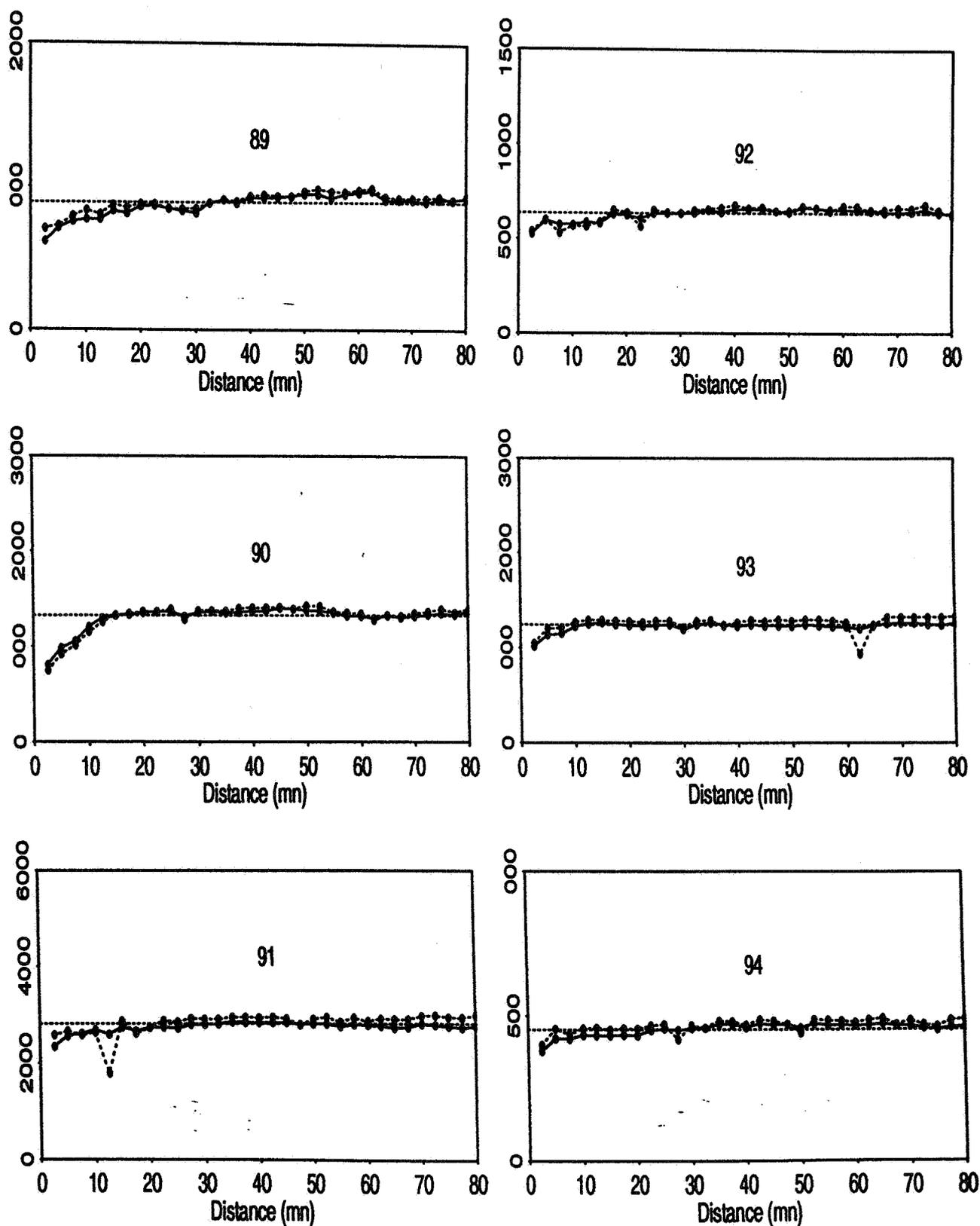


FIG. 7 - Estimation du variogramme des données brutes à l'aide de celui des données log-translatées - le long de la route du navire
 —variogramme retour -...variogramme direct

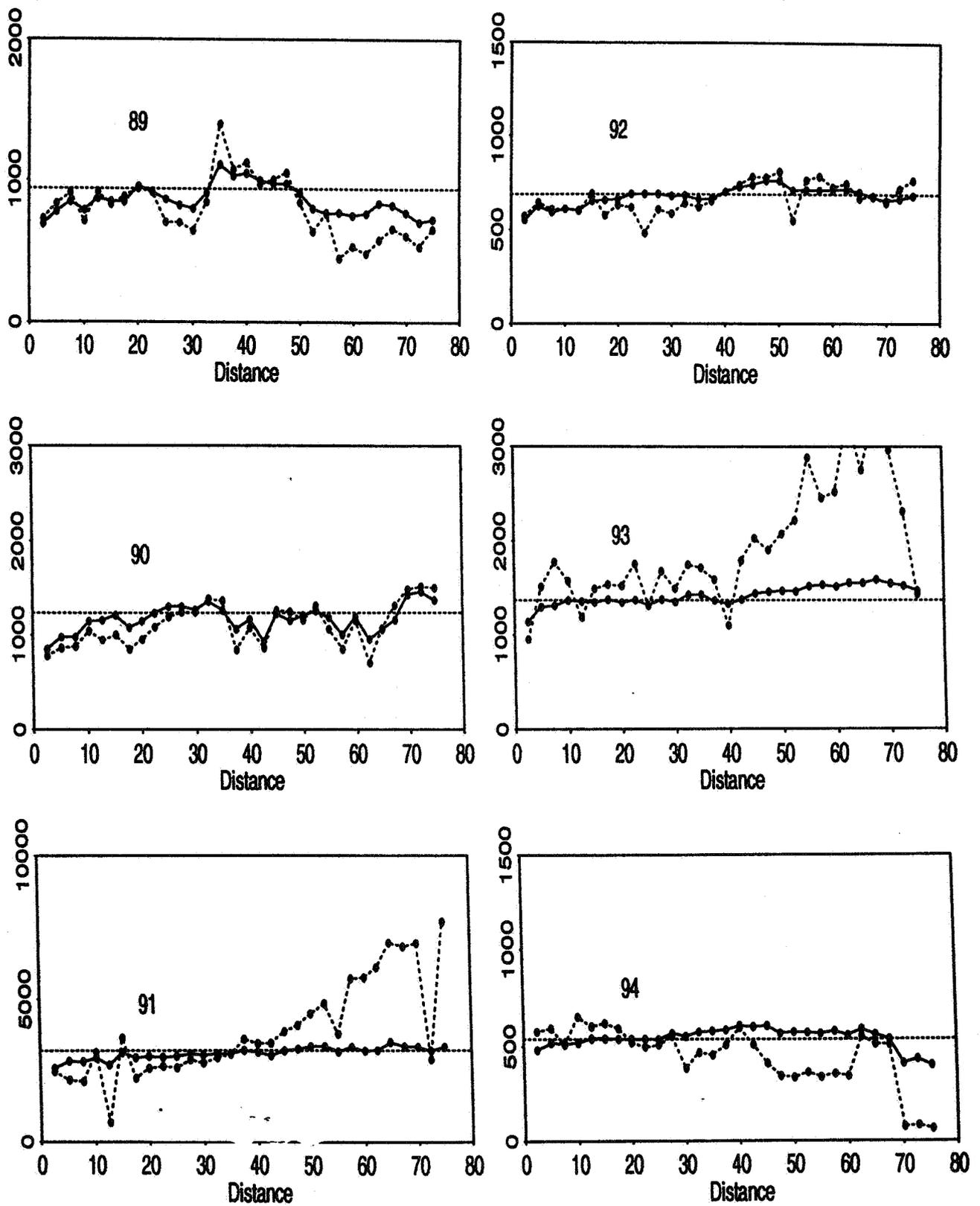


FIG. 8 - Estimation du variogramme des données brutes à l'aide de celui des données log-translatées - Direction est-ouest
 —variogramme retour -...-variogramme direct

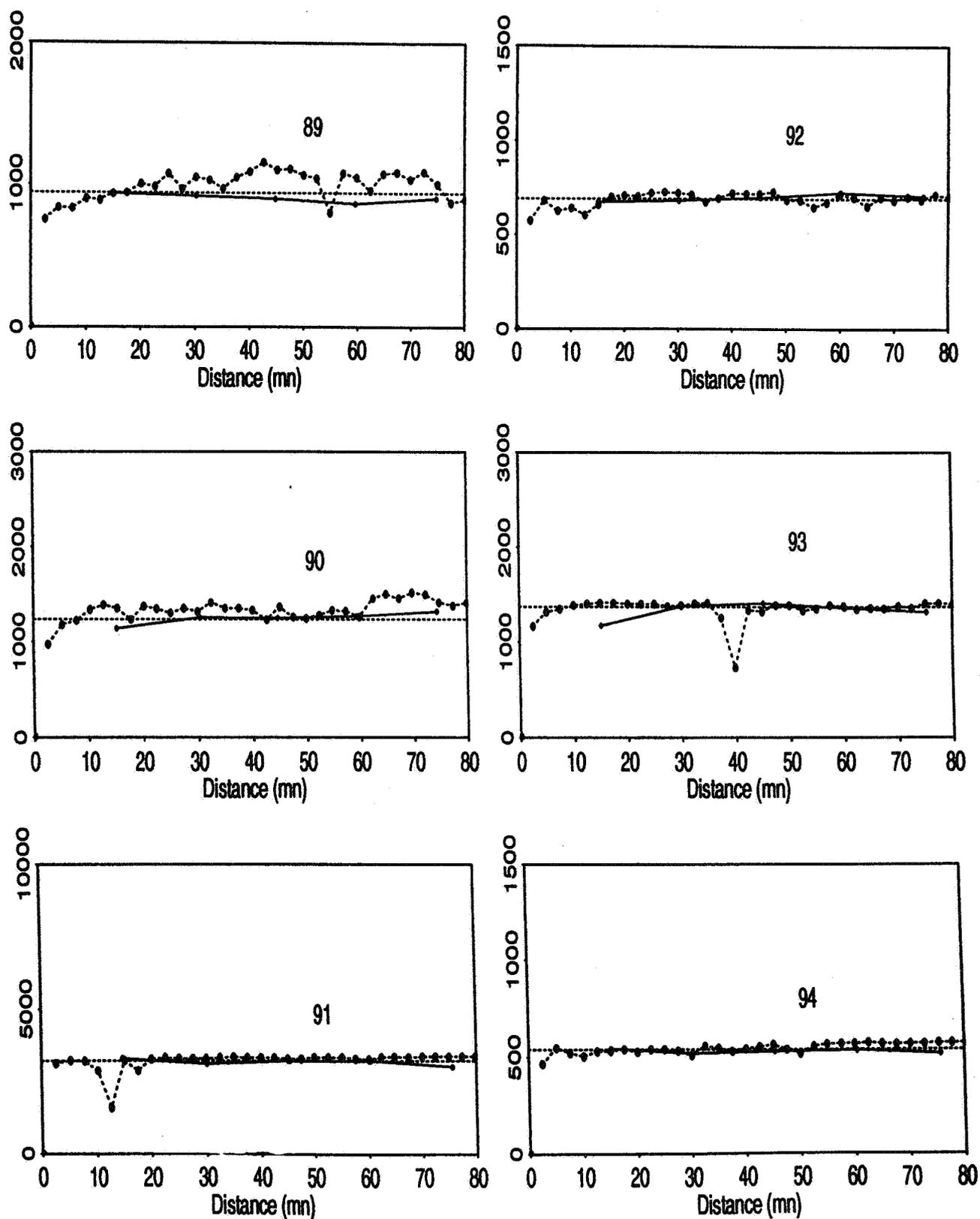


FIG. 9 - Variogrammes directionnels calculés à partir des covariances non centrées - Données brutes
 —direction Nord-Sud -....direction Est-Ouest

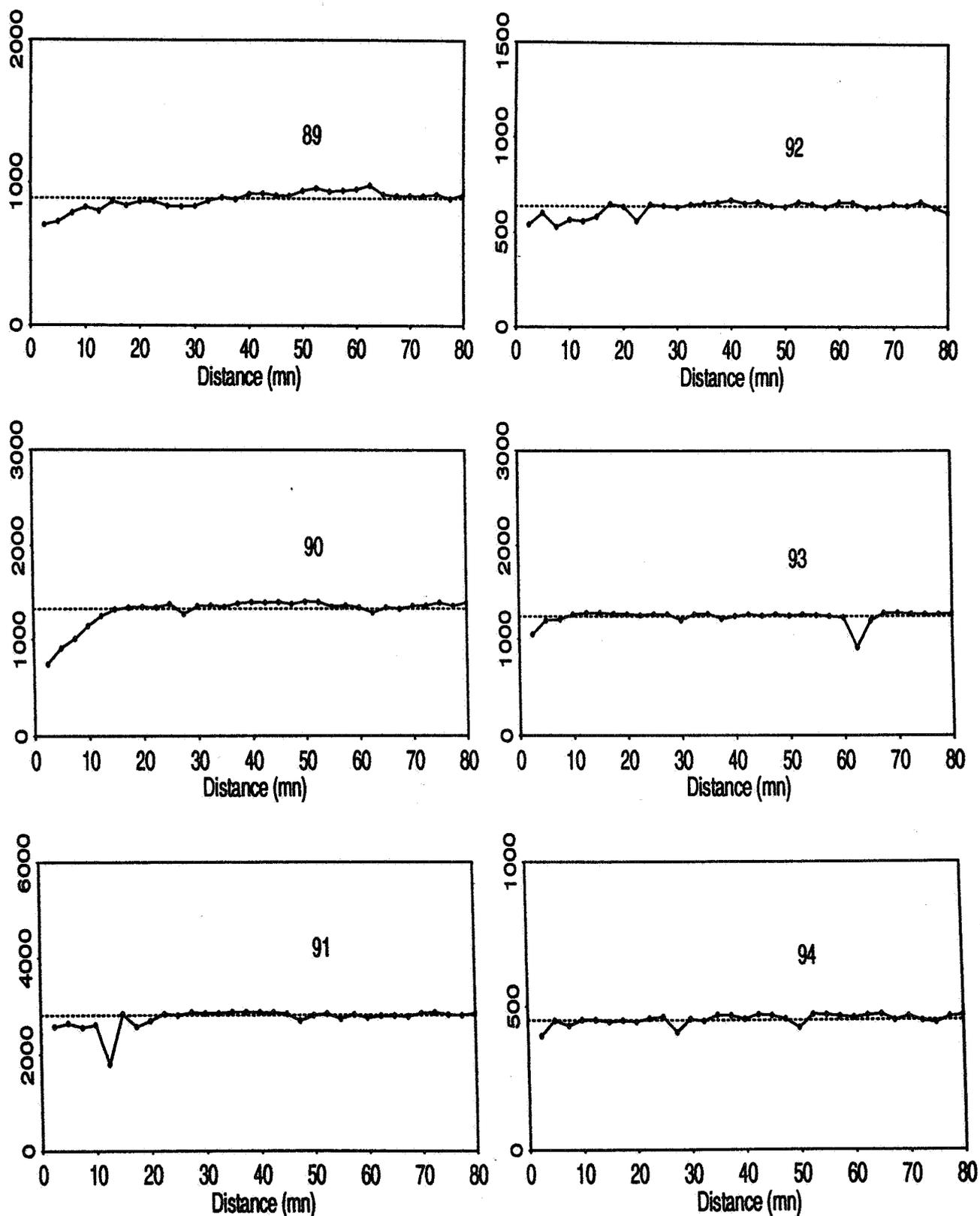


FIG. 10 - Estimation du variogramme des données brutes à l'aide des covariances non centrées - le long de la route du navire

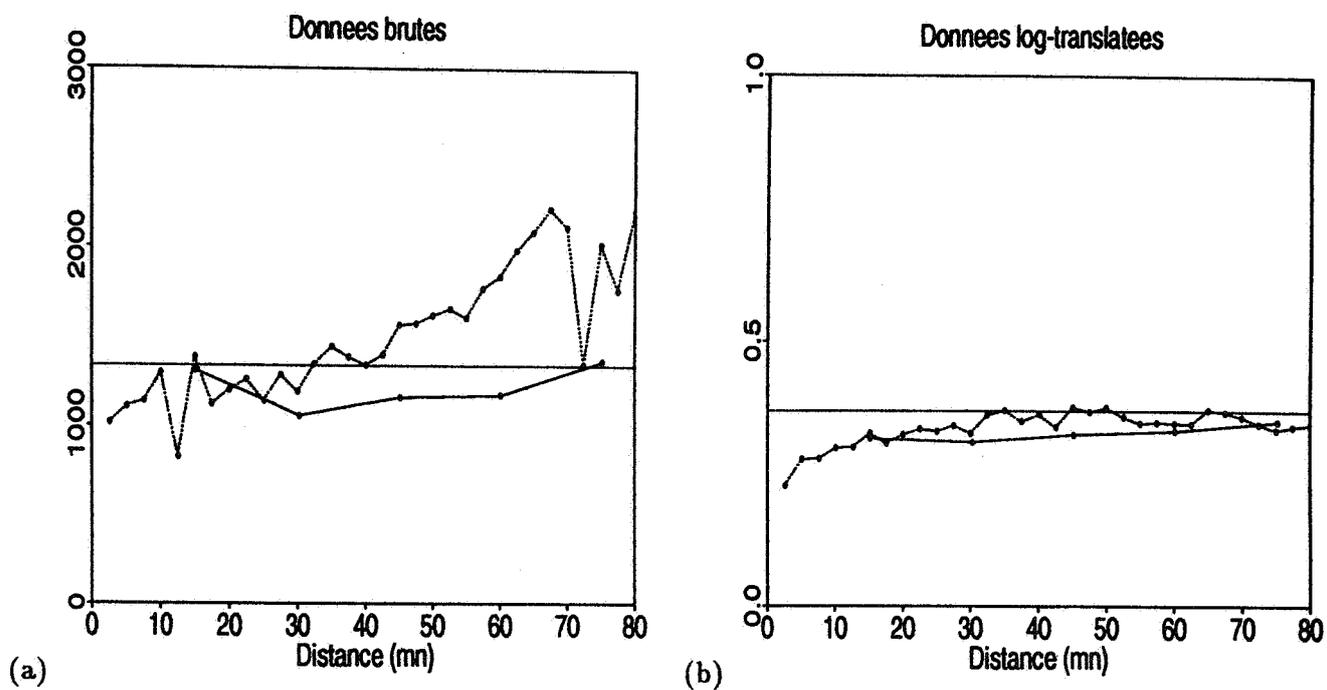


FIG. 11 - Variogrammes moyens sur 6 années
—direction Nord-Sud -....direction Est-Ouest

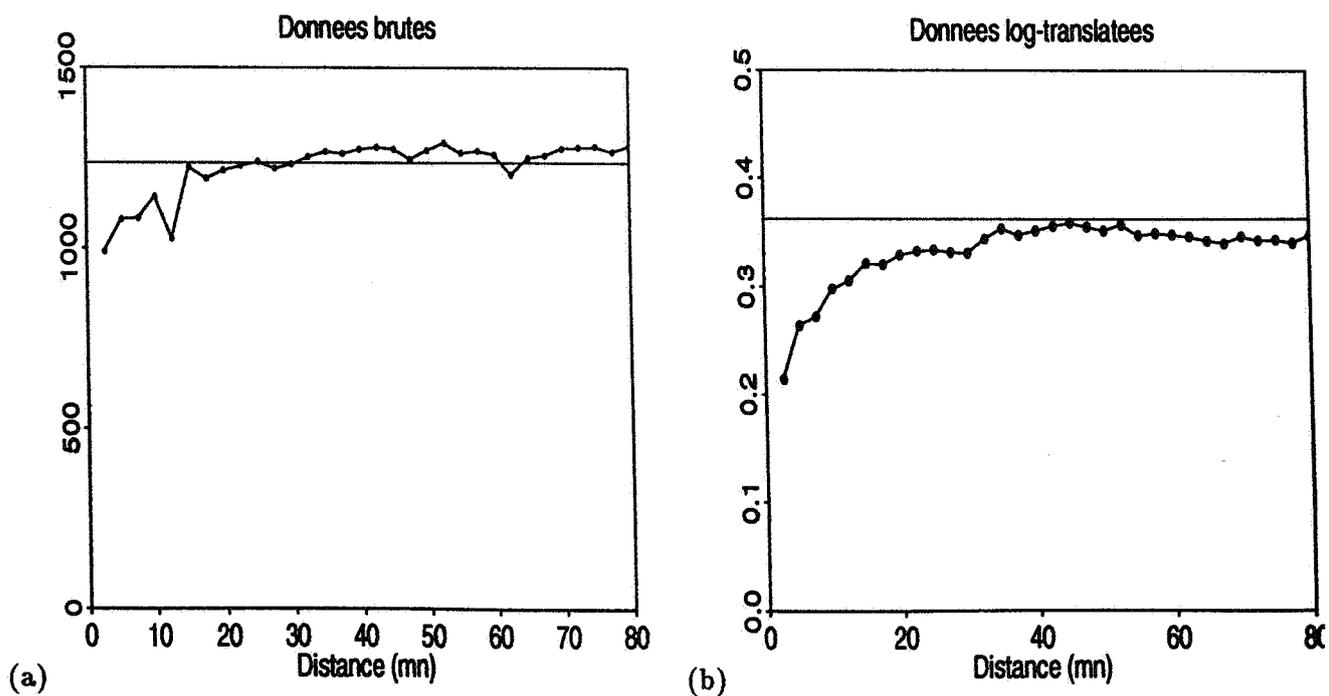


FIG. 12 - Variogrammes moyens sur 6 années - le long de la route du navire

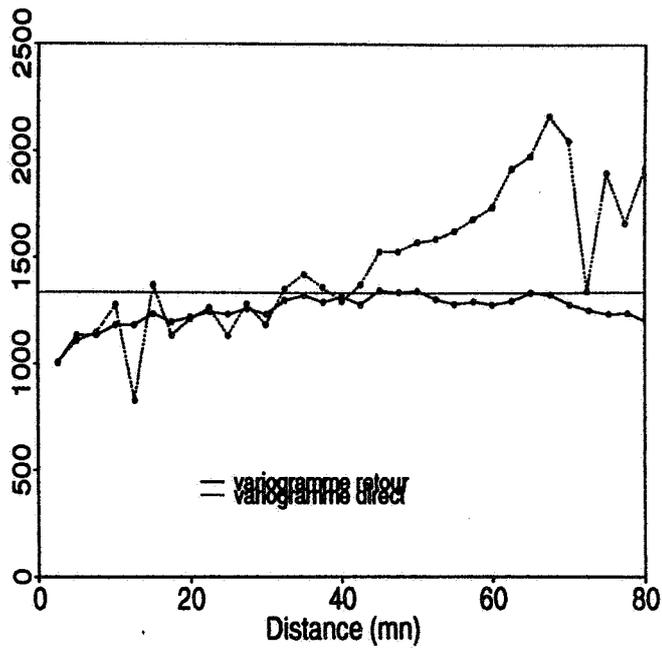


FIG. 13 - Estimation du variogramme des données brutes à l'aide de celui des données log-translatées - Variogrammes moyen sur 6 années - direction est-ouest

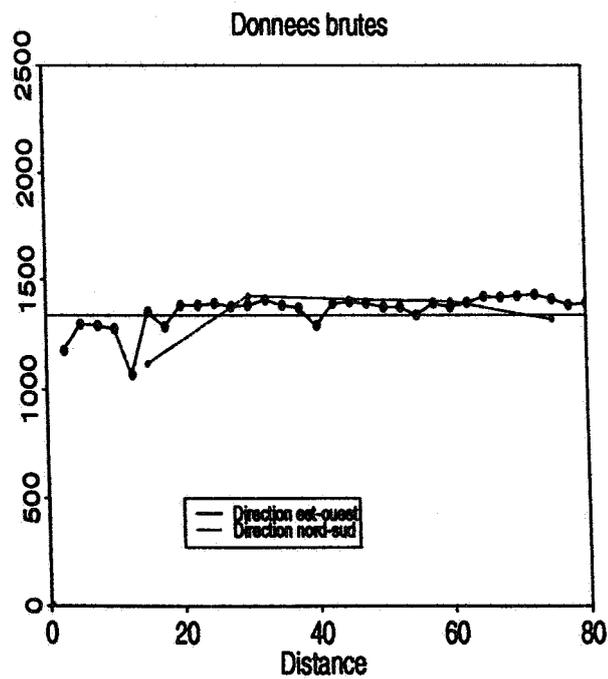


FIG. 14 - Estimation du variogramme des données brutes à l'aide de celui des covariances non centrées - Variogrammes moyen sur 6 années - données brutes