

ÉCOLE NATIONALE SUPÉRIEURE DES MINES DE PARIS

**ANALYSE GÉOSTATISTIQUE DE CAMPAGNES
(ACOUSTIQUE ET CHALUTAGE)
SUR LE HARENG ÉCOSSAIS**

TOME I - TEXTE

THÈSE

présentée devant
l'École Nationale Supérieure des Mines de Paris
par

Philippe GUIBLIN

pour obtenir le titre de Docteur en Géostatistique

Soutenue le 19 Février 1997
devant le jury composé de :

M.	Schmitt	Michel	Président
Mme.	Duby	Camille	Rapporteur externe
M.	Petitgas	Pierre	Rapporteur externe
M.	Rivoirard	Jacques	Examineur

Remerciements

Je souhaite remercier, tout d'abord, le Professeur Georges Matheron pour m'avoir accueilli au Centre de Géostatistique en Février 1993 ainsi que Chantal de Fouquet, grâce à qui je suis venu au C.G. et avec qui j'ai pu travailler et apprendre les techniques géostatistiques.

Je remercie très sincèrement Jacques Rivoirard pour ce qu'il m'a apporté tout au long de ce travail de thèse qu'il a dirigé. J'ai beaucoup apprécié sa disponibilité et sa pédagogie.

Je remercie l'Union Européenne pour le support apporté dans le cadre du projet AIR2-94-1007 "Geostatistics for fish stock assessment".

Je remercie John Simmonds du Marine Laboratory of Aberdeen (SOAEFD) pour m'avoir permis de travailler sur ses données. Je le remercie aussi pour ses conseils ainsi que pour m'avoir invité à séjourner dans son laboratoire en Août 1993 et permis de faire connaissance avec son équipe. Je me réjouis d'avoir pu travailler avec Paul Fernandes. Je remercie de même Ken Foote et Marek Ostrowski de l'Institute of Marine Research de Bergen.

Nos réunions dans le cadre du projet "Geostatistics for fish stock assessment" et dans le cadre de séminaires scientifiques m'ont été très profitables.

Je remercie Dominique Pelletier du Centre Ifremer de Nantes et Pierre Petitgas du Centre Orstom de Montpellier pour s'être intéressés à ce travail.

Je remercie Camille Duby et Pierre Petitgas d'avoir accepté d'être rapporteurs et d'avoir participé au jury ainsi que Michel Scmitt pour avoir accepté d'en être le président.

Christian Lajaunie et Didier Renard m'ont ponctuellement aidé à résoudre des problèmes informatiques. J'envie leur aisance dans ce domaine et les remercie.

Ma reconnaissance va également à l'ensemble du personnel du Centre de Géostatistique ainsi qu'à toutes les personnes avec qui, ces quatre années passées à Fontainebleau, mes rapports furent aussi divers et enrichissants.

Résumé

Ce travail propose une méthodologie de calcul d'estimations globales et de variances dans le cadre de campagnes d'échantillonnage acoustique. L'approche géostatistique est utilisée. Deux problèmes majeurs sont abordés à l'aide de deux cas d'étude différents: l'estimation robuste d'un variogramme et l'estimation dans un cadre non stationnaire.

Différentes manières d'estimer la structure variographique sur des données à distribution dissymétrique sont étudiées ici, sur l'exemple de six campagnes acoustiques. 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.

Des données collectées par chalutages scientifiques fournissent des informations quant à la distribution des longueurs et des classes d'âge pour chaque station d'échantillonnage. En général, plus les poissons sont longs et âgés, plus ils s'éloignent des côtes. La distribution spatiale de la longueur moyenne n'est pas stationnaire et dépend de la bathymétrie. Elle est de plus stable d'années en années. Par ailleurs, les proportions d'âge sont beaucoup plus variables. Des liens non linéaires sont mis en évidence entre longueurs moyennes et classes d'âge. Des cartographies des longueurs et des proportions d'âge sont réalisées.

Ces analyses sont nécessaires à l'obtention de l'abondance (nombre de poissons) globale et par classes d'âge. Un calcul de variance globale supposant la longueur déterministe est alors proposé.

Summary

This work presents a methodology to obtain global estimations and global variances within the framework of fish acoustic surveys, using the geostatistical approach. Two major problems are considered: a robust estimation of the variogram and estimation in a nonstationary case.

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

The trawl data from acoustic surveys consist of a distribution of lengths and ages of the fish for each trawling station. As a rule, the larger and the older the fish, the more offshore they are. The spatial distribution of the mean length is not stationary but changes according to seabed depth. It is nevertheless stable through the years. Age proportions, on the other hand, vary from year to year. Nonlinear links between mean lengths and age classes are identified. Mapping of lengths and of age proportions are performed.

The trawl data are used to obtain the abundance in terms of number of fish - both globally and for each age class. The global variance is then computed, assuming the length to be a deterministic variable.

Table des matières

Introduction	1
I Analyse géostatistique de mesures acoustiques	3
1 Concepts et outils géostatistiques nécessaires au calcul d'une variance d'estimation globale: Rappels	5
1.1 Fonction aléatoire et variogramme	5
1.1.1 Fonction aléatoire	5
1.1.2 Stationnarité d'ordre 2	6
1.1.3 L'hypothèse intrinsèque et le variogramme	6
1.2 Variance d'estimation d'une moyenne globale	6
1.2.1 Définition générale	6
1.2.2 Expression dans le cas d'un échantillonnage à maille régulière (dans le plan)	7
1.3 Cas particulier: méthode de composition des variances (dans le plan)	8
2 Analyse variographique	11
2.1 Estimation classique du variogramme	11
2.2 Robustesse	12
2.2.1 Estimateur robuste de la moyenne	12
2.2.2 Estimation d'un variogramme robuste	12
2.2.3 Variogramme de données log-translatées: une formule de "robustification" (cas stationnaire)	13
2.2.4 Démonstration de la formule de "robustification"	14
2.2.5 Variogramme déduit de la covariance non centrée	17
2.2.6 Calcul du variogramme le long de l'échantillonnage	18
3 Application à un stock de hareng écossais	19
3.1 Quelques généralités sur les campagnes acoustiques	20
3.2 Présentation du jeu de données	21
3.2.1 Distribution spatiale des données	21
3.2.2 Problème des zéros	21
3.2.3 Distribution de la variable log-translatée	22
3.3 Analyse variographique	23
3.3.1 Données brutes	23
3.3.2 Données log-translatées	23
3.3.3 Variogrammes le long de l'échantillonnage	24

3.3.4	Estimation du variogramme des données brutes à partir de celui des données log-translatées	24
3.3.5	Estimation du variogramme à partir de la covariance non centrée	25
3.3.6	Conclusion de l'analyse variographique	25
3.4	Calcul de variance d'une moyenne globale: campagne 1993	26
3.4.1	Choix d'un modèle de variogramme	26
3.4.2	Estimation d'un indice acoustique global	27
3.4.3	Variance et CV d'un indice acoustique global	28
3.4.4	Conclusion	30
 II Analyse géostatistique de données mesurées par chautage		31
1	Objectifs généraux	33
1.1	Analyse classique des échantillons pêchés	34
1.1.1	Stratification	34
1.1.2	Construction des strates	34
1.1.3	Moyennes des échantillons	35
1.1.4	Calcul de la surface équivalente	36
1.2	Analyses géostatistiques	36
2	Méthodes géostatistiques de cartographie	37
2.1	Estimation locale: technique du krigeage	37
2.2	L'estimateur de krigeage et la prise en compte de dérives	38
2.2.1	La dérive est une fonction polynôme des coordonnées	39
2.2.2	L'estimateur de krigeage avec dérive externe	40
2.3	Approches multivariables: krigeage et simulation	41
2.3.1	Le système de cokrigeage	41
2.3.2	Principe des simulations conditionnelles géostatistiques	43
3	Application: cartographie de la longueur et des classes d'âge	45
3.1	Présentation du cas d'étude	45
3.2	Distribution spatiale de la longueur moyenne	46
3.2.1	Analyse de cartes et variographie	46
3.2.2	Cartographie: krigeage avec dérive externe et effet de pépite différentiel	47
3.2.3	Conclusion: stabilité inter-annuelle de la longueur	48
3.3	Distribution spatiale des fréquences d'âge	49
3.3.1	Analyse de cartes et variographie	49
3.3.2	Conclusion: variabilité inter-annuelle des fréquences d'âge	50
3.4	Lien âge/longueur et cartographie des proportions d'âge	50
3.4.1	Observation des nuages de corrélation	50
3.4.2	Approche déterministe	51
3.4.3	Approches probabilistes: calculs de proportions d'âges et variances globales	52

III	Combinaison de mesures acoustiques et des mesures obtenues par chalutage : calculs d'abondance	59
1	Estimation d'une abondance globale	61
1.1	Calculs classiques	61
1.1.1	Calculs d'abondance	61
1.1.2	Calculs d'abondances par groupes d'âge	62
1.1.3	Calcul des biomasses	63
1.1.4	Calculs des variances d'abondance globales	63
1.2	Estimation globale et prise en compte de la corrélation spatiale . .	64
1.2.1	Estimation globale d'un quotient	64
1.2.2	Couplage de données acoustiques et de mesures chalutées .	65
2	Application au hareng écossais: campagne 1993	67
2.1	Méthodologie	67
2.2	Applications numériques	68
2.2.1	Abondance globale	68
2.2.2	Variance d'estimation globale	69
2.2.3	Calculs de biomasses	70
2.2.4	Abondances par classes d'âge	70
2.2.5	Discussion	70
3	Problèmes ouverts	73
3.1	Variabilité due à la longueur	73
3.2	Variabilité dans le champ des paramètres biologiques	74
3.3	Variabilité des abondances	75
	Conclusion	77

Introduction

• Cadre général

La géostatistique est un ensemble de méthodes dérivées de la théorie des probabilités particulièrement adaptées à l'analyse des données spatialement structurées. Historiquement, ces méthodes ont été utilisées pour résoudre des problèmes d'estimation de réserves dans les domaines miniers et pétroliers. Des méthodes spécifiques ont été développées pour les calculs d'estimation et d'évaluation de la qualité de l'estimation (par la variance). Un intérêt de la géostatistique par rapport aux techniques statistiques plus classiques est la prise en compte des corrélations spatiales dans le calcul de variances d'estimations globales quel que soit le type d'échantillonnage, aléatoire ou non. En particulier, la géostatistique fournit des méthodes pour le calcul de la variance d'estimation dans le cadre de campagnes d'échantillonnage systématique.

Ce travail se propose de s'intéresser dans un cadre très précis (celui de l'estimation d'un stock de hareng dans la zone des îles Shetland) à l'application des méthodes géostatistiques à un type de données particulier. L'utilisation de prospections acoustiques dans le suivi routinier des stocks halieutiques a incité les biologistes des pêches à s'intéresser à ces méthodes pour le calcul d'estimations globales de stocks halieutiques. La confrontation des méthodes géostatistiques à des données biologiques dont la caractéristique est la très grande variabilité spatiale nous a amené à revoir ces méthodes. Ce travail est l'illustration de deux problèmes que l'on rencontre lors de la confrontation des méthodes géostatistiques aux données halieutiques: le manque de robustesse du variogramme et la limite des méthodes stationnaires. Le variogramme, outil de base de la géostatistique, peut ne pas être adapté à l'étude de données à distribution spatiale très erratiques. Son manque de robustesse dans certains cas extrêmes comme celui de l'étude de variables à distribution très dissymétrique est un premier obstacle pour le praticien en géostatistique. L'utilisation de la géostatistique stationnaire sur des zones larges est limitée par la présence de dérives, cause de non stationnarité: influence de facteurs du milieu physique (par exemple la profondeur). Comment en tenir compte pour l'évaluation des stocks et l'estimation de variances d'estimations globales?

• Description du mémoire de thèse

Les biologistes des pêches cherchent à obtenir des estimations globales de stocks par classes d'âges et des mesures de l'erreur d'estimation. Le travail de recherche appliquée, présenté ci-dessous, se situe dans cette problématique. Le travail effectué pendant cette thèse s'est particulièrement porté sur les problèmes soulevés

par l'analyse d'un jeu de données composé de deux sources d'informations: des mesures acoustiques et des mesures obtenues par chalutages.

Les données acoustiques se composent de résultats de mesures de six campagnes annuelles (de 1989 à 1994) concernant le stock de hareng écossais de la Mer du Nord. Chaque donnée fournit un indice d'abondance moyenné sur une distance de 2.5 milles nautiques. Les données obtenues par chalutages donnent des informations biologiques quant à la distribution des tailles et la proportion des âges de chaque trait de chalut. Les traits de chaluts sont effectués de manière opportuniste en fonction des résultats acoustiques.

La première partie de ce travail est consacrée à l'analyse des mesures acoustiques et aux problèmes méthodologiques posés lors de l'analyse variographique. Il existe des difficultés considérables dans l'application correcte des techniques d'analyse géostatistique aux variables rencontrées dans l'étude des stocks halieutiques: les distributions rencontrées sont très dissymétriques et posent des difficultés lors de l'analyse variographique. Le problème majeur rencontré dans ce travail a été de trouver une méthode d'obtention d'un variogramme robuste. Plusieurs techniques sont proposées qui s'affranchissent d'hypothèses sur la distribution des données.

La deuxième partie s'intéresse à la cartographie de variables biologiques collectées par chalutages comme la longueur et l'âge du poisson. La densité de poisson en un point donné s'exprime comme le quotient de la valeur de l'indice acoustique et d'un facteur fonction de la longueur du poisson. La densité de poisson d'un âge donné est le produit de cette densité par la proportion de l'âge considéré. Une cartographie des longueurs et des proportions d'âges est donc nécessaire comme préliminaire aux calculs d'estimations. Dans la zone étudiée et pour cette espèce de hareng, la distribution spatiale des longueurs semble plus stable, sur plusieurs campagnes, que celle des classes d'âge. La mise en évidence et l'exploitation de ce phénomène biologique constituent l'originalité de ce travail. Les poissons de petites tailles se situent près des côtes où la bathymétrie est faible, ceux de grandes tailles sont plus au large où la bathymétrie est plus importante. De ce fait, la distribution spatiale des longueurs est fortement non stationnaire. Cette observation nous a amené à utiliser une technique particulière d'estimation locale (le krigeage avec dérive externe) où la profondeur est prise en compte comme variable auxiliaire. D'autre part, la quantification de la variabilité inter-annuelle et la relative stabilité de la distribution des longueurs dans le temps nous ont permis de construire des cartes annuelles de krigeage des longueurs à l'aide des données de toutes les campagnes. La cartographie directe des proportions d'âges étant rendue trop difficile du fait de leur non stationnarité, il a été décidé d'exploiter le lien fonctionnel étroit entre l'âge et la longueur mis en évidence sur les valeurs mesurées. Une attention particulière a été prêtée à l'étude de la relation longueur/âge par des approches probabiliste et déterministe. Ces techniques ont été appliquées à une campagne donnée.

La troisième partie tente de mettre en relation les résultats des études séparées des deux types de données pour obtenir des abondances et variances globales et des abondances par classes d'âge. Une méthodologie de calcul d'estimations globales et de variances dans le cadre de campagnes de mesures utilisant les méthodes d'échantillonnage acoustique est ainsi proposée.

Première partie

**Analyse géostatistique de
mesures acoustiques**

Chapitre 1

Concepts et outils géostatistiques nécessaires au calcul d'une variance d'estimation globale: Rappels

Calculer une variance d'estimation globale dans un cadre géostatistique nécessite l'utilisation de quelques formules reposant sur des principes issus de la théorie des probabilités formulés par G. Matheron [22]. On se propose, dans ce chapitre, de présenter l'essentiel du formalisme et des notions utiles pour le calcul d'une variance. Un cas particulier est présenté et sera utilisé dans une application à des données réelles au chapitre 3.

1.1 Fonction aléatoire et variogramme

1.1.1 Fonction aléatoire

Une Fonction Aléatoire (F.A.) peut être définie comme une variable vectorielle infinie $Z(x) = (Z(x_1), Z(x_2), \dots, Z(x_n), \dots)$ où les composantes, $Z(x_i)_{i=1,2,\dots,\infty}$ sont des variables aléatoires non indépendantes associées à tout point x_i de l'espace [22].

On appelle réalisation de la fonction aléatoire $Z(x)$, une fonction particulière $z(x) = (z(x_1), z(x_2), \dots, z(x_n))$, où les valeurs $z(x_i)_{i=1,2,\dots,n}$ sont des valeurs numériques particulières, résultat d'un tirage au sort effectué selon la loi de probabilité de Z .

On définit la fonction aléatoire $Z(x)$ par sa loi spatiale, à savoir la donnée de toutes les lois de distribution:

$$P(Z(x_1) \leq z_1, Z(x_2) \leq z_2, \dots, Z(x_n) \leq z_n)$$

pour tous les entiers n et pour tous les points x_1, x_2, \dots, x_n . L'inférence statistique est en général impossible à partir d'une seule réalisation de la F.A. $Z(x)$ sans quelques hypothèses supplémentaires.

1.1.2 Stationnarité d'ordre 2

$Z(x)$ est stationnaire si sa loi est invariante par translation, $Z(x)$ a même loi spatiale que $Z(x+h)$ où h est un vecteur quelconque. $Z(x)$ est stationnaire d'ordre 2 si ses deux premiers moments sont stationnaires [8]. Son espérance existe et est égale à une constante

$$m = E(Z(x)) = E(Z(x+h))$$

et sa covariance centrée ne dépend que de h :

$$C(h) = E((Z(x) - m)(Z(x+h) - m)) = E(Z(x)Z(x+h)) - m^2$$

Cette hypothèse suppose l'existence d'une variance finie $C(0)$ [22].

1.1.3 L'hypothèse intrinsèque et le variogramme

$Z(x)$ vérifie l'hypothèse intrinsèque si l'accroissement $Z(x) - Z(x+h)$ est stationnaire d'ordre 2, de moyenne nulle:

$$E(Z(x) - Z(x+h)) = 0,$$

et sa variance:

$$Var(Z(x) - Z(x+h)) = 2\gamma(h),$$

où $\gamma(h)$ est le (semi-)variogramme. Sous cette hypothèse, $Z(x)$ peut ne pas avoir une variance finie.

Ainsi, une F.A. stationnaire d'ordre 2 vérifie l'hypothèse intrinsèque. Le variogramme d'une F.A. stationnaire d'ordre 2 est borné [22] et vérifie alors:

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

1.2 Variance d'estimation d'une moyenne globale

1.2.1 Définition générale

Le domaine V sur lequel est effectuée l'estimation étant connu, la moyenne globale s'exprime comme l'intégrale:

$$Z_V = \frac{1}{V} \int_V z(x) dx.$$

L'estimateur de cette quantité s'exprime comme une combinaison linéaire de n données prélevées sur le domaine V :

$$Z_V^* = \sum_{i=1}^n \lambda_i z(x_i).$$

Par exemple, la moyenne arithmétique des données $\frac{1}{n} \sum_{i=1}^n z(x_i)$.

La condition de non biais, $E(Z_V - Z_V^*)$ impose une condition sur les coefficients λ_i :

$$\sum_{i=1}^n \lambda_i = 1.$$

La variance d'estimation de la moyenne globale est:

$$\text{Var}(Z_V - Z_V^*) = \text{Var}\left(\frac{1}{V} \int_V z(x) dx - \sum_{i=1}^n \lambda_i z(x_i)\right)$$

ce qui conduit à:

$$\text{Var}(Z_V - Z_V^*) = \frac{1}{V^2} \int_V \int_V C(x, y) dx dy - 2 \sum_{i=1}^n \lambda_i \frac{1}{V} \int_V C(x, x_i) dx + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(x_i, x_j)$$

Dans le cas où seule l'hypothèse intrinsèque est acceptable, la variance d'estimation peut être calculée par l'expression:

$$\text{Var}(Z_V - Z_V^*) = -\frac{1}{V^2} \int_V \int_V \gamma(x-y) dx dy + 2 \sum_{i=1}^n \lambda_i \frac{1}{V} \int_V \gamma(x-x_i) dx - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(x_i - x_j)$$

1.2.2 Expression dans le cas d'un échantillonnage à maille régulière (dans le plan)

Dans le plan, les échantillons sont prélevés en série selon une grille régulière. On considère alors l'estimateur de la quantité globale :

$$Q^* = Z_V^* \cdot V$$

où Z_V^* est l'estimateur de la moyenne globale de la variable (densité de poissons, par exemple) Z_V sur tout le champ, calculée comme la moyenne arithmétique des données collectées ($\frac{1}{n} \sum_i z(x_i)$) et V , la surface du champ que l'on suppose fixé.

En interprétant, ici, $Z(x)$ comme la réalisation au point x d'une fonction aléatoire intrinsèque, la variance d'estimation de Z_V a pour expression

$$-\frac{1}{V^2} \int_V \int_V \gamma(x-y) dx dy + \frac{2}{n} \sum_{i=1}^n \frac{1}{V} \int_V \gamma(x-x_i) dx - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma(x_i - x_j)$$

que l'on symbolise par l'expression:

$$\sigma_E^2 = 2\bar{\gamma}(V, V') - \bar{\gamma}(V, V) - \bar{\gamma}(V', V') \quad (1.1)$$

où V' représente l'ensemble des échantillons.

On définit ainsi la notion de variance d'extension d'un domaine V à un domaine V' comme étant la variance de l'erreur que l'on commet en attribuant à V la moyenne calculée sur V' [22]. Ici, V est le champ total échantillonné et V' , l'ensemble des échantillons.

Dans le cas stationnaire d'ordre 2, la variance d'estimation peut s'exprimer à l'aide de la covariance $C(h)$ et a pour expression:

$$\text{Var}(Z_V - Z_V^*) = \frac{1}{V^2} \int_V \int_V C(x, y) dx dy - 2 \sum_{i=1}^n \frac{1}{V} \int_V C(x, x_i) dx + \sum_{i=1}^n \sum_{j=1}^n C(x_i, x_j)$$

que l'on symbolise par l'expression:

$$\sigma_E^2 = \overline{C}(V, V) + \overline{C}(V', V') - 2 \cdot \overline{C}(V, V')$$

En pratique, appliquer directement ces formules suppose la connaissance d'un modèle de variogramme valable sur toute la zone étudiée. Pratiquement, on ne peut pas toujours supposer la variable stationnaire sur toute la zone. De plus le variogramme n'est pas connu pour des distances de l'ordre des dimensions du domaine. Il est alors nécessaire de recourir à des approximations: C'est la méthode de **composition des variances**. Cette méthode ne suppose l'éventuelle stationnarité que sur des voisinages restreints [7].

Dans la théorie intrinsèque, on suppose l'indépendance entre le champ et la variable. Dans ce travail, le champ sera toujours considéré comme fixé. Il n'y a donc pas d'erreur due à l'erreur d'estimation de la surface du champ.

1.3 Cas particulier: méthode de composition des variances (dans le plan)

On suppose le plan d'échantillonnage composé d'une grille régulière de points, les échantillons sont répartis régulièrement le long de segments (lignes, transects) parallèles et équidistants. On note V le champ et V' , l'ensemble des échantillons.

La variance d'estimation est donnée, dans le cas stationnaire, par la formule:

$$\sigma_E^2 = \overline{C}(V, V) + \overline{C}(V', V') - 2 \cdot \overline{C}(V, V')$$

Le paragraphe suivant donne une méthode d'approximation à deux dimensions dont l'avantage est de n'utiliser que le comportement aux petites distances du variogramme (ou de la covariance) [29].

Le domaine d'étude, V , est découpé en n sous unités de surface v . Un échantillon est prélevé au centre de chaque surface v_i . La variance de l'estimation globale est décomposée en erreurs élémentaires que l'on suppose non corrélées [22], [18]:

$$\sigma_E^2 = T_l + T_s$$

où

- T_l est le **terme de ligne**: C'est la variance de l'erreur commise en estimant la moyenne des lignes par la moyenne des échantillons qu'elles contiennent. Elle est liée à l'erreur que l'on commet en estimant un segment élémentaire par son échantillon central:

$$T_l = \frac{1}{n} \sigma_{El}^2$$

où n est le nombre total d'échantillons et σ_{El}^2 est la variance d'extension d'un point dans un segment de longueur l (variance de l'erreur que l'on commet en estimant une moyenne sur le segment l par son échantillon central).

- T_s est le **terme de tranche** (ou de section): C'est la variance de l'erreur commise en estimant un plan par la moyenne pondérée des lignes qu'il contient. La surface totale peut être assimilée à la réunion de n_s rectangles $s_i = l_i \times h_i$. Soit V la surface totale des sections $V = \sum_{i=1}^{n_s} s_i$, le terme de section (ou de tranche) s'écrit:

$$T_s = \frac{1}{V^2} \sum_{i=1}^{n_s} s_i^2 \sigma_{Es_i}^2$$

où $\sigma_{Es_i}^2$ est la variance d'extension de la médiane l_i à son rectangle d'influence s_i (variance de l'erreur que l'on commet en estimant une moyenne sur un rectangle s_i par la moyenne estimée sur sa médiane l_i).

$$\sigma_{Es_i}^2 = 2\bar{\gamma}(s_i, l_i) - \bar{\gamma}(s_i, s_i) - \bar{\gamma}(l_i, l_i) \quad (1.2)$$

Finalement, dans un problème à deux dimensions la variance de l'estimation globale est donc [22, 29]:

$$\sigma_E^2 = \underbrace{\frac{1}{n} \sigma_{El}^2}_{\text{terme de ligne}} + \underbrace{\frac{1}{(\sum_{i=1}^{n_s} l_i)^2} \sum_{i=1}^{n_s} l_i^2 \sigma_{Es_i}^2}_{\text{terme de tranche}} \quad (1.3)$$

Le calcul du terme de ligne suppose que les erreurs que l'on fait en estimant la moyenne sur chaque segment par la valeur de l'échantillon central sont indépendantes. De même, le calcul du terme de tranche suppose indépendantes les erreurs commises lors de l'estimation des tranches de surface s_i par le cumul des lignes. Ceci n'est valable que si l'on considère que le réseau d'échantillonnage comporte une direction dont la densité des prélèvements est plus importante [22, 29]. Une application est proposée dans le chapitre 3 dans le cas d'une prospection acoustique effectuée le long de transects parallèles et équidistants.

Le calcul théorique d'une variance d'estimation globale fait intervenir des calculs intermédiaires utilisant la connaissance d'un outil de base: le variogramme. La connaissance du variogramme (calcul expérimental et modélisation) est l'objet de l'analyse variographique (ou analyse structurale). Le chapitre 2 revient sur l'estimation du variogramme au cours de l'analyse variographique.

Chapitre 2

Analyse variographique

L'efficacité d'une étude géostatistique, dépend, dans une large mesure, de la qualité de l'estimation du variogramme [4]. Le problème de l'estimation du variogramme a, par le passé, été abordé de nombreuses fois [4, 10, 11, 26]. Il ne s'agit pas, ici, de faire une revue des différentes méthodes possibles, ce travail ayant déjà été fait [26]. On se contente, dans ce chapitre, de rappeler le problème auquel se confronte le géostatisticien devant un cas pratique d'analyse variographique et de proposer un estimateur robuste que nous avons utilisé sur des données réelles.

2.1 Estimation classique du variogramme

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 [22]:

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

En pratique, on cherche à estimer le variogramme régional, demi-moyenne de $[z(x+h) - z(x)]^2$ sur le champ [31]:

$$\gamma(h) = \frac{1}{2|V \cap V_{-h}|} \int_{V \cap V_{-h}} [z(x+h) - z(x)]^2 dx$$

où V est le champ et V_{-h} est le champ translaté d'une distance $-h$. L'estimateur classique et sans biais [22, 10] du variogramme expérimental est calculé par l'expression:

$$\gamma^*(h) = \frac{1}{2n_h} \sum_{i=1}^{n_h} [(z(x_i+h) - z(x_i))^2] \quad (2.1)$$

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 [7], [18]. 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 [31].

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 [31].

2.2 Robustesse

2.2.1 Estimateur robuste de la moyenne

L'estimateur de la moyenne globale sur un champ calculé comme moyenne pondérée ou moyenne arithmétique des données mesurées risque de manquer de robustesse. C'est notamment les cas où coexistent dans l'espace de fortes valeurs en faible nombre entourées de faibles valeurs. Certaines statistiques sont plus robustes, médiane ou autres quantiles [16], statistiques sur les logarithmes dans le cas lognormal. L'abondance globale est cependant uniquement fournie par la moyenne. Travailler sur une statistique robuste, puis repasser à la moyenne nécessite une hypothèse forte sur la distribution. Ces hypothèses sont, en général, incontrôlables.

L'estimateur moyenne étant peu robuste, celui de la variance ou du variogramme comportant des moyennes de carrés l'est encore plus.

2.2.2 Estimation d'un variogramme robuste

Le manque de robustesse est responsable de grandes variations observées sur le variogramme expérimental ce qui peut rendre les structures difficiles à identifier (portées, anisotropies...).

Bien souvent, la présence de valeurs extrêmes dans l'histogramme ("outliers") rend très difficile l'analyse variographique. Supprimer la (ou les) valeur(s) forte(s) lors du calcul du variogramme peut permettre d'observer une structure mais alors un nouveau problème se présente: quel sens accorder à ce variogramme? Éliminer des valeurs peut être aussi sans intérêt si la structure réelle de la variable provient justement des valeurs éliminées [25]. Le problème est encore plus difficile dans le cas d'une distribution très dissymétrique, il n'y a pas d'extrema particulier à enlever. L'écrêtage (les valeurs au dessus d'un seuil sont rabattues à ce seuil) des données est parfois envisagé. Il apparaît nécessaire de chercher une statistique moins instable pour traduire la structure spatiale du phénomène étudié. Plusieurs tentatives ont été envisagées par le passé pour rendre interprétable la structure du variogramme expérimental [31], [4] et [10]: les problèmes distributionnels rencontrés ont incité les praticiens à rechercher un estimateur robuste du variogramme. Des formules de "robustification" du variogramme ont été proposées qui supposent connues la distribution des données. L'établissement de ces

estimateurs doit répondre à trois points:

- Trouver une transformation de la variable, ou de l'estimateur classique du variogramme (utilisation de fonctions puissances, de la médiane).
- Introduire un facteur correctif de normalisation pour annuler le biais.
- Admettre une loi de distribution de la variable régionalisée qui autorise ces corrections (loi multigaussienne, généralement).

Plusieurs transformations possibles ont ainsi été établies qui donnent satisfaction suivant les cas d'études proposés [11, 26]. Cependant ces méthodes ne peuvent être appliquées lorsque la distribution expérimentale des données est très dyssymétrique comme c'est le cas pour les distributions de densités de poissons que nous avons rencontrées. Aucune de ces approches n'est apte à fournir la structure brute sans hypothèse supplémentaire. Ainsi, le variogramme robuste de Cressie [10] suppose que les lois des couples $(Z(x), Z(x+h))$ sont bi-gaussiennes.¹

Une solution envisagée et mise en pratique dans ce travail est de travailler sur une transformation logarithmique de la variable qui s'inspire directement des résultats d'une étude réalisée dans le domaine minier [31].

2.2.3 Variogramme de données log-translatées: une formule de "robustification" (cas stationnaire)

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 $\text{Log}(b + Z(x))$, ou ce qui est équivalent mais plus commode (les données nulles restant nulles): $L(x) = \text{Log}(1 + \frac{Z(x)}{b})$. Ainsi, si $Z(x)$ est petit devant b , $L(x)$ est équivalent à $\frac{Z(x)}{b}$, tandis que si $Z(x)$ est grand devant b , $L(x)$

1. Dans le but de diminuer l'influence de valeurs extrêmes ("outliers") Cressie propose comme estimateurs robustes de variogrammes [10]:

$$- 2\gamma^*(h) = \left(\frac{1}{n_h} \sum_{i=1}^{n_h} |z(x_i + h) - z(x_i)|^{1/2}\right)^4 / (0.457 + 0.494/n_h)$$

et

$$- 2\gamma^*(h) = (\text{med}|z(x_i + h) - z(x_i)|^{1/2})^4 / 0.457 \text{ où la fonction med() désigne la médiane.}$$

L'idée est que pour des distributions gaussiennes $Z(x_i)$, la fonction $(Z(x_i + h) - Z(x_i))^2$ suit une distribution du χ^2 à 1 degré de liberté. La transformation puissance utilisée qui permet de se rapprocher le mieux d'une fonction gaussienne est la puissance quatrième de la racine carrée de la valeur absolue des différences (Cressie et Hawkins (1980 in [10]). La transformation est en partie justifiée par le fait que les termes $(|z(x_i + h) - z(x_i)|^{1/2})$ sont moins corrélés que les termes $(z(x_i + h) - z(x_i))^2$. Travailler avec une distribution du χ^2 permet aux auteurs de déterminer le facteur correctif de normalisation qui conduit à des estimateurs robustes non biaisés. Le contexte général de ce travail est celui d'une "distribution gaussienne contaminée". Le cas pratique auquel nous nous sommes intéressés s'en éloigne beaucoup et c'est pourquoi nous nous sommes orientés sur d'autres voies.

est équivalent à $\text{Log}(\frac{Z(x)}{b})$. En pratique, la constante b , par rapport à laquelle $Z(x)$ est considérée comme faible ou forte, est choisie de l'ordre de grandeur de la moyenne de $Z(x)$ (la valeur exacte de b ne jouant alors pratiquement pas sur la forme du variogramme).

Par le passé, il a été montré sur des données expérimentales que la structure des données log-translatées est bien plus stable. Aucune formule permettant de passer du variogramme de la variable transformée à celui de la variable d'origine n'avait été envisagé. L'intérêt de l'observation des variogrammes expérimentaux des données transformés résidait, alors, dans la comparaison de structures et dans l'étude des anisotropies [31].

Naturellement, la structure du logarithme est, en général, différente de la structure brute. Pour passer du variogramme γ_L de $L(x)$ au variogramme γ_Z de $Z(x)$, nous avons établi la formule (pour le cas stationnaire):

$$\gamma_Z(h) = ((b + m)^2 + \text{var}(Z(x)))(1 - e^{-\sigma^2 \frac{\gamma_L(h)}{\text{var}(L(x))}) \quad (2.2)$$

où m est la moyenne de $Z(x)$ et

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

Cette relation se démontre dans le cas où le logarithme (népérien) $L(x)$ est bigaussien [24], mais également dans celui où $L(x)$ est une somme de jetons aléatoires valués (jetons multiplicatifs).

L'utilisation et l'intérêt de cette formule de "retour" est présentée dans l'application du chapitre 3. Cette formule sera utilisée de façon heuristique indépendamment de tests de modèle, en tant que moyen de "robustifier" la structure brute. Si l'estimation classique du variogramme de la fonction $L(x)$ est sans biais celle du variogramme de la variable $Z(x)$ par cette formule peut poser un problème de biais.

2.2.4 Démonstration de la formule de "robustification"

Cette formule donne le moyen de passer du variogramme de la variable transformée à celui de la variable et est proposée dans deux cas: le cas lognormal déjà connu [24] et celui plus original des jetons multiplicatifs [13].

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, pour $\beta = 0$, 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:

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

ce qui donne en notant $\gamma_L(h)$ le variogramme de la fonction $L(x)$ et $\gamma_Z(h)$, celui de $Z(x)$:

$$\gamma_Z(h) = m^2 e^{\sigma^2}(1 - e^{-\gamma_L(h)})$$

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 $\theta|D|$, et les nombres de points tombant dans deux domaines disjoints sont indépendants). 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 $\rho(h) = \frac{K(h)}{K(0)}$.

Remarque:

Contrairement au cas lognormal, σ^2 n'est pas la variance logarithmique de $Z(x)$.

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:

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

soit

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

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:

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

2.2.5 Variogramme déduit de 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 :

$$K^*(h) = \frac{1}{n_h} \sum_{i=1}^{n_h} z(x_i)z(x_i + h) \quad (2.3)$$

puis le variogramme :

$$\gamma^*(h) = K^*(0) - K^*(h) \quad (2.4)$$

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.

2.2.6 Calcul du variogramme le long de l'échantillonnage

Les plans d'échantillonnage sur lesquels nous travaillons sont ceux fournis par le résultat de campagnes acoustiques. Un navire collecte un message acoustique tout au long d'une route prédéfinie (voir chapitre 3). Un moyen de calcul original du variogramme consiste à le calculer indépendamment des directions de l'espace, en continu le long de l'échantillonnage. 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.

Chapitre 3

Application à un stock de hareng écossais

Ce chapitre propose de présenter une application à un jeu de données réelles des méthodes géostatistiques exposées dans les deux chapitres précédents. L'application des méthodes géostatistiques à une campagne acoustique n'est pas nouvelle. Pierre Petitgas y a consacré une partie de sa thèse [29]. Comme lui, nous nous intéresserons à l'estimation d'un stock et à l'estimation de la variance d'estimation. La méthode de l'écho-intégration permettant l'obtention des mesures à bord des navires scientifiques a été largement décrite dans son travail. P. Petitgas propose des calculs prenant en compte, ou non, l'effet géométrique dû au champ. Le plan d'échantillonnage étant une maille régulière, le calcul transitif [29, 22] permet un calcul de variance d'estimation global s'affranchissant de l'effet géométrique. Le calcul par l'approche intrinsèque suppose la géométrie du champ fixée. L'erreur sur l'estimation globale due à l'estimation de la surface du champ peut être rajoutée. Dans ce travail, nous nous sommes surtout intéressés à améliorer la connaissance du variogramme. Son estimation par les différentes méthodes exposées dans les deux chapitres précédents a focalisé notre intérêt. Nous avons tenté, ensuite, de tirer quelques conclusions générales à l'aide de l'étude de six campagnes de prospection. Enfin, un calcul de variance d'estimation globale est présenté sur une campagne donnée.

La grandeur physique mesurée est un indice d'abondance (surface réfléchissante, unité: m^2/mn^2). L'indice acoustique d'abondance S mesuré au point de l'espace à deux dimensions x est relié à la densité ρ (nombre de poissons par unité de surface) par la relation:

$$\rho(x) = \frac{S(x)}{\sigma(x)}$$

où σ est la surface équivalente moyenne des poissons échantillonnés (section réfléchissante). La surface équivalente σ dépend de caractéristiques anatomiques du poisson, de sa longueur totale et des dimensions de sa vessie natatoire par rapport à la longueur d'onde émise [3]. Elle est étroitement reliée à la taille du poisson. Cette propriété sera exploitée dans les parties 2 et 3 de ce mémoire: la longueur du poisson est obtenue expérimentalement à l'aide de prélèvements par chalutage. Sa valeur en chaque point x n'est pas connue aux points de l'échantillonnage acoustique et sera l'objet d'une estimation.

3.1 Quelques généralités sur les campagnes acoustiques

L'intérêt des campagnes acoustiques est de couvrir de larges zones en peu de temps, c'est un outil d'observation idéal pour les ressources pélagiques. Les objectifs d'une campagne acoustique consistent dans l'observation des distributions de poissons et dans l'estimation des stocks. Les scientifiques des pêches (modélisateurs) ont besoin d'estimateurs d'abondance pour leurs modèles de dynamique de population (analyse des cohortes...) [1]. Les gestionnaires des pêches ont besoin d'estimateurs globaux pour définir des quotas de pêche d'espèces commerciales. Une bonne campagne doit incorporer toutes les connaissances disponibles sur le stock en question [1, 2, 3]. En général, le poisson a tendance à s'agréger formant ainsi des distributions groupées. Le degré d'agrégation varie suivant, entre autres, l'espèce et le stock, la biomasse totale, la saison, l'heure du jour (jour/nuit), la distribution de la nourriture, les conditions environnementales. Toutes ces informations doivent être considérées lors de la construction d'un plan de campagne d'échantillonnage. Le choix d'une zone à échantillonner repose sur l'un ou plusieurs des éléments suivants:

- les frontières internationales (zones ICES/CIEM),
- les frontières physiques (côtes),
- l'étendue des espèces cibles (plateau continental),
- les conditions océanographiques.

Parfois, les navires de recherche acoustique sont limités par la profondeur du fond marin. Certaines espèces se répartissent en eaux peu profondes dangereuses pour la navigation.

L'objectif est d'estimer un stock sur un champ donné. La méthodologie d'échantillonnage utilisée pour l'estimation du stock de hareng de la Mer du Nord (*Clupea harengus*), pendant les mois d'été, est celle décrite par Kirkegaard *et al.* [19]. La zone d'étude est divisée en "rectangles statistiques" (15 minutes de latitude par 30 minutes de longitude). Ceci est illustré sur la figure 1.1. La distribution du poisson à l'intérieur de ces rectangles est supposée homogène. La collecte des données se fait le long de segments parallèles orientés Est-Ouest appelés transects. Tous les carrés statistiques traversés par les transects sont échantillonnés sur une même distance. Les résultats des campagnes des années précédentes permettent de définir 2 niveaux d'échantillonnage. Pour les zones où la densité de poisson est forte, deux transects traversent chaque rectangle, pour les zones de faibles densité, un simple transect est alloué par rectangle statistique (au sud de $57^{\circ}45'N$, (fig.1.1)). Les auteurs [1] proposent un estimateur d'abondance et une variance pour chacun de ces rectangles statistiques dont la méthodologie sera décrite dans la partie 3. L'intérêt de ces campagnes est d'utiliser les résultats des estimations des campagnes antérieures.

Le paragraphe suivant présente les six campagnes de mesures acoustiques qui nous ont été fournies et que nous avons analysées.

3.2 Présentation du jeu de données

3.2.1 Distribution spatiale 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 1.2 à l'aide de figurés proportionnels. Les campagnes s'étendent de part et d'autre des îles Orkney et Shetland, mais pour des raisons d'homogénéité structurale, nous considérerons séparément les parties Est et Ouest. En effet, rien ne prouve que la structure des données est la même de part et d'autre des îles.

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 (chaque mesure correspond à la moyenne sur 2.5 milles nautiques du signal récupéré par les appareils de mesure, cette distance est l'Unité Élémentaire de Distance d'Échantillonnage). La distance inter-transect est d'environ 15 milles nautiques dans les zones pauvres et d'environ 7.5 milles nautiques dans les zones supposées plus riches. Chaque "rectangle statistique" est ainsi traversé deux fois dans les zones supposées plus riches. A l'intérieur des zones riches et pauvres, un schéma régulier est à peu près respecté, excepté par endroit où une prospection en zig-zag a été effectuée. La mesure acoustique fournie par le SOAEFD d'Aberdeen est l'indice acoustique divisé par un facteur 4π . C'est sur cette variable que s'effectueront les calculs dans cette première partie. La figure 1.2 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 strictement nulles. 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. 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.

L'analyse structurale porte sur les plans d'échantillonnage des figures 1.7 et 1.8. Les statistiques élémentaires (moyennes, variances et cv) montrent des variations d'une année sur l'autre: la proportion des valeurs nulles est variable et la valeur des maxima est très variable. Les coefficients de variation sont supérieurs à 1 et caractérisent une distribution groupée (ou agrégative) [32]. 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.

3.2.2 Problème des zéros

Les histogrammes de la figure 1.3 montrent très clairement la forte dissymétrie des distributions rencontrées et la forte proportion des valeurs nulles ou proches de zéro. L'importance numérique des valeurs nulles est résumée par les histogrammes présentés sur les figures 1.3 et 1.4. La proportion des valeurs nulles est variable

d'une année sur l'autre. Même sans les zéros (fig. 1.4), les distributions restent très dissymétriques. La présence de nombreux zéros pose deux problèmes:

- quel sort réserver aux valeurs nulles dans nos analyses?
- quelle loi de distribution inférer à partir des données?

Les figures 1.5 et 1.6 semblent montrer que les distributions des données privées des valeurs nulles peuvent être grossièrement assimilées à des distributions log-normales (test de la droite de Henry). Pour les mêmes quantiles, les valeurs de la variable transformée et de la gaussienne centrée réduite sont représentées: les points dessinent une droite si la variable est log-normale, ce qui semble être le cas ici. Seulement, les zéros font partie des données. On pourrait envisager de retirer des traitements les zéros en bordure du champ [29] correspondant à la limite d'occupation de l'espace par le poisson. Mais que faire des zéros intérieurs? Il n'existe pas de zone constamment inoccupées par le poisson. Ce qui gêne l'analyse variographique n'est pas la forte proportion de zéros mais la présence de forts écarts entre les données sur de courtes distances. Supprimer les zéros (dans notre cas) ne change pas le problème. Pourquoi, alors, ne pas retirer aussi les faibles valeurs, les 1 ou toutes les valeurs inférieures à 10, 20 etc...?

Une autre possibilité est de réserver un traitement particulier aux données nulles. Mac Lennan et Mac Kenzie (1988) [20], travaillant sur la même espèce de hareng dans la même zone géographique proposent de traiter les échantillons 0 de façon distincte ce qui permet d'éliminer une source de non stationnarité. Les données non nulles sont normalisées à l'aide d'une transformation non linéaire. Des estimateurs globaux et des variances sont calculés sous l'hypothèse d'absence de structure spatiale. Il semble alors préférable de supposer que les distributions envisagées soient log-normales. La même idée est reprise par Petitgas [29] qui utilise le modèle de Δ -distribution sur une distribution ne comportant que 8% de valeurs nulles.

Dans notre cas, on pourrait éventuellement supposer les distributions des données non nulles comme log-normales, seulement nous pensons pouvoir malgré tout extraire une structure de nos données et nous pensons que c'est la méthode de calcul des variogrammes qui est à améliorer. On se heurte, ainsi, à des problèmes distributionnels importants. Le parti pris, dans ce travail, a été de garder toutes les données et de tâcher d'en extraire une structure spatiale, au besoin, au moyen d'une transformation non linéaire des données conservant les valeurs nulles.

3.2.3 Distribution de la variable log-translatée

De telles distributions sont très dissymétriques. L'utilisation des données log-translatées, ici $\text{Log}(1 + \frac{Z}{12})$, 12 étant la moyenne expérimentale de toutes les données toutes campagnes confondues, permet de manipuler une variable moins dispersée, les valeurs nulles restent inchangées et l'importance des valeurs fortes est relativisée (fig. 1.9 et 1.10). Il est à remarquer que cette transformation est utilisée de façon heuristique, indépendamment de toute hypothèse lognormale. En effet, vue la proportion de valeurs nulles, la variable brute ne saurait obéir à une distribution lognormale, même après translation. Des paquets de données

moins hétérogènes sont ainsi mis en évidence sur les cartes proportionnelles et laissent imaginer une structure spatiale plus accessible. Le problème du passage de la structure spatiale des données transformées à celle des données brutes est présenté de manière théorique au chapitre précédent et sera illustré au paragraphe suivant.

3.3 Analyse variographique

3.3.1 Données brutes

Les variogrammes expérimentaux sont présentés figures 1.11 et 1.12, ils ont été calculés dans la direction des transects et dans la direction perpendiculaire aux transects à partir des données présentées figures 1.7 et 1.8. 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. Pour les variogrammes expérimentaux la formule 2.1 a été utilisée.

Vu 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. 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 (fig. 1.15). 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 1.15-(a) et (b). On remarque surtout l'influence de l'année la plus variante (89 pour la partie Ouest et 91 pour la partie Est). 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.

3.3.2 Données log-translatées

Comme précédemment, les variogrammes dans deux directions perpendiculaires sont calculés pour chaque campagne et de part et d'autre des îles (figures 1.13 et 1.14). 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. Dans la direction Nord-Sud l'espace et le temps sont mélangés, excepté sur les segments entre les transects, une variabilité temporelle vient donc se surajouter à la variabilité spatiale. Certaines campagnes montrent clairement la différence entre les parties Est et Ouest en ce qui concerne les structures spatiales (91, 92).

Les variogrammes moyens sur les 6 campagnes sont présentés figure 1.15-(c) et (d). Ils ne présentent aucune structure vraiment visible (effets de pépite).

3.3.3 Variogrammes le long de l'échantillonnage

La variable pouvant être considérée comme stationnaire, n'importe quelle valeur peut se trouver n'importe où dans le champ. De plus, cette variable sera 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 figures 1.18 et 1.19.

On observe des structures beaucoup plus visibles qui restent courtes. Les pépites sont assez élevées. Les paliers sont bien stabilisés sur la variance.

3.3.4 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 2.2). Connaissant le variogramme estimé de la variable log-translatée, on revient à celui de la variable brute à l'aide de la formule 2.2. 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 1.16, 1.17, 1.20 et 1.21 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). Ne sont considérés, ici, que les variogrammes expérimentaux calculés dans la direction Est-Ouest.

Sur les figures 1.20 et 1.21, 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 (figures 1.16 et 1.17): on constate que le passage par le logarithme stabilise le variogramme et donne des structures mieux visibles.

3.3.5 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 2.3 puis 2.4). Sur les figures 1.22 et 1.23 sont présentés les variogrammes estimés par ce moyen, ils sont à comparer avec ceux des figures 1.11 et 1.12. 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 pépitiqes 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 augmenterait le nombre de paquets de valeurs fortes et la robustification serait nettement moindre. C'est pourquoi on a choisi de ne pas les représenter.

3.3.6 Conclusion de l'analyse variographique

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 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 variante peut imposer sa structure à ces variogrammes.

Quel modèle privilégier? En fait, aucune méthode ne s'impose vraiment sur une autre, les résultats sont fluctuants d'une année sur l'autre et de part et d'autre des îles. Le choix du modèle dépendra de l'année prise en compte et de la méthode de calcul de variance globale choisie.

3.4 Calcul de variance d'une moyenne globale : campagne 1993

Ce paragraphe est consacré au calcul d'une variance d'estimation globale. Il est l'application directe du chapitre 1 §3.

La variance est estimée en utilisant le variogramme. Dans le cas d'un échantillonnage régulier comportant des lignes (transects) de plus fortes densités d'échantillonnage, la variance peut être calculée en utilisant la méthode de composition de lignes et de tranche. Cette méthode a l'avantage de n'utiliser que le comportement à petite distance du variogramme [22]. Ils suppose aussi la non corrélation des erreurs d'estimation entre les lignes. Il est donc nécessaire d'être dans de bonnes conditions d'application de la méthode. La campagne de l'année 1993 nous semble a priori remplir cette condition. L'échantillonnage de la campagne 1989, plus "accidenté" (nombreux zig-zag, échantillonnage moins régulier) rend ce calcul pratique plus délicat.

Nous considérerons la campagne de l'année 1993. La partie Est de la campagne est échantillonnée régulièrement à partir d'une maille élémentaire 2.5 mn x 15 mn. A l'Ouest, dans une zone plus riche, l'échantillonnage est plus resserré et les transects sont doublés (conformément à la méthode d'échantillonnage présentée au chapitre 2). On considérera 4 transects espacés de 7.5 mn dans une sous-zone "Centre". La partie Ouest sera donc estimée en considérant trois sous-zones (Sud, Centre et Nord) dont les surfaces respectives serviront de pondérateurs lors des calculs globaux. Un modèle de variogramme sera ajusté pour la zone Est et un autre pour la zone Ouest, considéré comme valable pour les trois sous-zones.

3.4.1 Choix d'un modèle de variogramme

Au regard des six campagnes, une hypothèse globale d'isotropie est adoptée ici. Les structures sont courtes et l'échantillonnage Nord-Sud ne permet pas d'inférer une structure dans cette direction.

Grâce à l'utilisation de méthodes d'estimation de variogrammes robustes, une hypothèse globale de stationnarité est supposée. Pour l'année 1993, le variogramme calculé sur les données brutes, dans la partie Est, semble présenter une anisotropie zonale (le palier dans la direction Nord-Sud est plus élevé que dans la direction Est-Ouest). Celle-ci disparaît si l'on observe le variogramme déduit de la covariance non centrée. Cette apparente anisotropie semble donc être l'effet de la présence de valeurs fortes dont l'importance s'estompe dans un calcul de variogramme plus robuste.

Compte tenu du fait que les structures sont courtes et que la dispersion des valeurs est grande, on peut considérer que les hypothèses de stationnarité et d'isotropie sont raisonnables, les variogrammes calculés le long du tracé du navire ont été choisis pour déterminer les structures à petites distances. Les modèles ajustés sur les variogrammes expérimentaux sont présentés sur la figure 1.24. Deux modèles sphériques semblent convenir conformément au comportement linéaire près de l'origine. Le variogramme de la partie Est a un palier légèrement supérieur à la variance, la structure est courte (10 mn) et l'effet de pépité correspond à 70% de la variance. Le variogramme de la partie Ouest est plus continu avec une

portée de 11 mn et un effet de pépite correspondant à 20% de la variance (on ne retient comme significatif dans ce variogramme expérimental que la partie à petites distances, le comportement linéaire après 25 mn n'est pas pris en compte dans le modèle). La méthode d'estimation de la variance utilisée par la suite ne considère que le comportement à petites distances.

Les modèles sont :

- Partie Ouest :

$$\gamma(h) = \begin{cases} 0 & \text{pour } h = 0 \\ 200 + 800 * (1.5 * \frac{h}{11} - 0.5 * (\frac{h}{11})^3) & \text{pour } 0 \leq h \leq 11 \\ 1000 & \text{pour } h > 11 \end{cases}$$

- Partie Est :

$$\gamma(h) = \begin{cases} 0 & \text{pour } h = 0 \\ 800 + 330 * (1.5 * \frac{h}{10} - 0.5 * (\frac{h}{10})^3) & \text{pour } 0 \leq h \leq 10 \\ 1130 & \text{pour } h > 10 \end{cases}$$

3.4.2 Estimation d'un indice acoustique global

L'indice acoustique moyen sur le champ ne donne pas l'abondance. Il n'est représentatif de l'abondance qu'à un facteur près dépendant de la distribution des longueurs. Cependant, le calcul de sa variance est un bon indicateur de la couverture spatiale d'une campagne acoustique.

À l'intérieur de chaque zone (ou sous-zone envisagée), où l'échantillonnage est régulier, l'indice acoustique moyen est estimé par la moyenne arithmétique des données :

$$Z_V^* = \frac{1}{n} \sum_{i=1}^n z(x_i)$$

L'estimateur à l'Est des îles est la moyenne arithmétique des données, celui de la partie Ouest est la moyenne pondérée par les trois surfaces de trois moyennes arithmétiques correspondant aux trois sous-zones. L'estimateur de la moyenne pour tout le domaine est lui aussi une moyenne pondérée des deux estimateurs Est et Ouest.

D'après les figures 1.7 et 1.8 nous obtenons :

- Partie Est : Surface = $2.3 \times 10^4 \text{ mn}^2$, $Z_V^* = 7.99 \text{ m}^2/\text{mn}^2$
- Partie Ouest : $Z_V^* = 9.3 \text{ m}^2/\text{mn}^2$

Sous zone	Moyenne m^2/mn^2	Surface mn^2	Effectif
Nord (espacement de 15 mn)	4.81	1457	73
Centre (espacement de 7.5 mn)	9.30	2192	74
Sud (espacement de 15 mn)	10.60	4448	142

Total Partie Ouest :

Moyenne = $9.30 \text{ m}^2/\text{mn}^2$

Surface = 8000 mn^2

- Total : $Z_V^* = 8.63 \text{ m}^2/\text{nm}^2$ (après pondération)

3.4.3 Variance et CV d'un indice acoustique global

Par définition, la variance d'estimation d'une valeur moyenne sur un domaine est la variance de la différence entre la vraie valeur inconnue Z_V et son estimateur Z_V^* (ici une moyenne pondérée des données). Un coefficient de variation peut être calculé (estimé):

$$CV = \frac{\sqrt{\text{Var}(Z_V^* - Z_V)}}{Z_V^*}$$

On le comparera à celui calculé sous l'hypothèse d'absence de corrélation spatiale.

En **statistique classique**, les mesures sont supposées non corrélées entre elles. Par conséquent, cette variance d'estimation se calcule directement à partir de la variance de dispersion des valeurs mesurées suivant un calcul exposé par Cochran (in Grosbras, [16]).

$$\sigma^2 = \frac{s^2}{n}$$

où n est le nombre d'échantillons et

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (z(x_i) - m^*)^2$$

est la variance expérimentale des mesures supposées non corrélées et où m^* est la moyenne arithmétique des données sur le champ.

En biologie des pêches, on apprécie usuellement la qualité d'une estimation à l'aide du coefficient de variation (estimé):

$$CV = \frac{\sqrt{s^2/n}}{m^*}$$

Le modèle de variogramme étant choisi, comportant un effet de pépite C_0 , et une composante sphérique $C_1(h)$, nous pouvons calculer la variance d'estimation à l'aide de trois termes (formule 1.3):

$$\sigma_E^2 = \frac{C_0}{n} + \frac{1}{n} \sigma^2(a) + \frac{\sum_i l_i^2 \sigma_{Es_i}^2}{(\sum_i l_i)^2}$$

- $\frac{C_0}{n}$ est la part de la variance due à l'effet de pépite.
- $\frac{1}{n} \sigma^2(a)$ est la variance d'extension d'un point dans son segment d'influence correspondant à la composante sphérique, c'est **le terme de ligne**: la variance de l'erreur faite en estimant les lignes à l'aide des échantillons ($a = 2.5$ nm). Dans cette étude, le terme de ligne est égal à 0, les transects sont échantillonnés en continu donc parfaitement connus.
- Le troisième terme est **le terme de tranche**: C'est la variance de l'erreur faite en étendant la valeur estimée à partir des lignes à leur rectangle d'influence et correspondant à la composante sphérique. Les l_i sont les longueurs des transects parallèles i .

Chaque tranche qui contribue proportionnellement à sa longueur à l'estimation de la moyenne, contribue proportionnellement au carré de sa longueur à la variance. De manière similaire, des variances pondérées, par le carré des surfaces correspondant à des espacements de transects différents, sont à prendre en compte (Partie Ouest, combinaison de la partie Ouest et Est).

Les variances d'extension mises en jeu dans le calcul du terme de tranche nécessitent l'utilisation de la formule 1.1 adaptée ici au calcul des variances d'extension d'un segment à son rectangle d'influence (formule 1.2). Ces variances ont été calculées à l'aide des abaques donnés par Matheron dans le cas du schéma sphérique [22]. Ces abaques permettent d'obtenir la variance d'extension d'un segment médian (l_i) dans son rectangle (s_i) connaissant la portée du modèle de variogramme, la longueur et la largeur du rectangle considéré.

Résultats numériques pour la partie Est :

Variance due à l'effet de pépite: $\frac{800}{654} = 1.22$

Terme de tranche: 0.380

$Var(Z_V - Z_V^*) = 1.60$

$$CV = \frac{\sqrt{Var(Z_V - Z_V^*)}}{Z_V^*} = 15.8\%$$

(Sans prendre la structure spatiale en compte, on a: $CV = 16.7\%$)

Résultats numériques pour la partie Ouest :

Sous-zone	Variance pépitique	Terme de tranche	Total
Nord	2.74	5.39	8.13
Centre	2.70	1.51	4.21
Sud	1.40	1.83	3.23

Total Partie Ouest

$$Var(Z_V - Z_V^*) = 8.13 \left(\frac{1457}{8.10^3}\right)^2 + 4.21 \left(\frac{2192}{8.10^3}\right)^2 + 3.23 \left(\frac{4448}{8.10^3}\right)^2 = 1.58$$

$$CV = \frac{\sqrt{Var(Z_V - Z_V^*)}}{Z_V^*} = 13.5\%$$

(Sans prendre la structure spatiale en compte, on a: $CV = 20.9\%$)

Pour tout le domaine:

$$Var(Z_V - Z_V^*) = \left(\frac{2.3 \times 10^4}{3.1 \times 10^4}\right)^2 \times 1.60 + \left(\frac{8 \times 10^3}{3.1 \times 10^4}\right)^2 \times 1.58 = 1.05$$

$$CV = \frac{\sqrt{Var(Z_V - Z_V^*)}}{Z_V^*} = 11.0\%$$

(Sans prendre la structure spatiale en compte, on a: $CV = 13.0\%$)

3.4.4 Conclusion

La prise en compte de la structure spatiale contribue à faire diminuer l'erreur sur l'estimation de la quantité globale. Les résultats utilisant un calcul intrinsèque donnent un coefficient de variation meilleur dans la mesure où le modèle traduit la structure à petite distance de la variable régionalisée. Lors des calculs en théorie intrinsèque, on prend en compte une hypothèse probabiliste sur la variable étudiée ainsi que la géométrie du champ que l'on suppose fixée.

Deuxième partie

Analyse géostatistique de données mesurées par chalutage

Chapitre 1

Objectifs généraux

L'indice d'abondance mesuré par les campagnes acoustiques, permet un calcul d'abondance à condition de tenir compte de la distribution des longueurs. La densité réelle de poisson en un point x de l'espace s'obtient comme l'indice d'abondance pondéré par un facteur dépendant de la longueur. La première partie de ce travail s'est intéressée à l'analyse géostatistique des mesures acoustiques et a proposé un calcul de variance d'estimation globale comme si l'indice acoustique pouvait être considéré comme une densité. L'intérêt de cette première partie réside surtout dans le développement de méthodes pouvant améliorer la connaissance du variogramme. Le calcul de variance proposé a surtout un intérêt pédagogique. Nous nous en servons à nouveau, néanmoins, dans la troisième partie lors de la présentation des calculs d'abondance finaux.

Il est nécessaire de compléter les campagnes acoustiques par des campagnes de mesures portant sur la distribution des longueurs du poisson dans le champ ainsi que sur la distribution des âges si l'on veut estimer de plus des abondances par classes d'âge. Une campagne visant l'estimation de la biomasse de poisson sur un domaine donné se compose donc en plus des mesures acoustiques de mesures biologiques effectuées directement sur des poissons pêchés. Chaque échantillon pêché se compose de poissons de tailles et d'âges différents. La distribution des longueurs et des âges est donc connue aux points de l'échantillonnage. À chaque échantillon correspondent donc une valeur moyenne des longueurs (longueur moyenne des poissons de l'échantillon) et une valeur moyenne de chaque classe d'âge considérée fournie sous forme d'une proportion d'âge. Ce sont ces deux types de données que l'on se propose d'analyser dans cette partie.

L'analyse faite par les praticiens [1, 35, 36, 37] sur la distribution des longueurs et des âges est assez rudimentaire: elle consiste à délimiter des zones dans le champ à l'intérieur desquelles on peut considérer la longueur comme constante, une valeur fixe correspondant à chaque proportion d'âge est affectée ainsi à chacune de ces strates. A partir de là, un calcul d'abondance globale sur tout le stock puis globale par classe d'âge peut être fait (voir la partie 3). La stratification du domaine intervient après l'analyse des échantillons pêchés. La (post-)stratification est donc différente suivant les campagnes. On présente, dans ce chapitre, les éléments nécessaires à l'établissement de ces strates. L'usage chez les praticiens est de regrouper entre eux des échantillons dont on peut considérer qu'ils correspondent à des distributions de longueurs homogènes. Le test statistique utilisé est le

test non paramétrique de Kolmogorov-Smirnov pratique lorsqu'on manipule de petits échantillons [33].

L'analyse que nous proposons sur ces données pêchées est différente. Elle utilise les techniques géostatistiques d'estimations locales pour effectuer des cartographies des longueurs et des classes d'âges. Le chapitre 2 présentera des techniques usuelles en géostatistique qui seront appliquées dans le chapitre 3. Dans l'approche géostatistique, la longueur moyenne sera considérée comme une fonction aléatoire non stationnaire. Une méthode particulière de krigeage prenant en compte une variable auxiliaire (ici la profondeur) permettra d'obtenir une cartographie plus précise que celle obtenue par un krigeage ordinaire.

La cartographie des âges sera déduite de celle des longueurs après une étude de la relation entre l'âge et la longueur.

1.1 Analyse classique des échantillons pêchés

La connaissance de la distribution de longueurs est nécessaire pour estimer la surface réfléchissante, facteur pondérateur de l'indice acoustique, et pour estimer des abondances par groupe d'âges, au besoin en utilisant une clé taille/âge. Les paragraphes suivant présentent succinctement la méthode utilisée par les praticiens pour estimer la longueur par strate du domaine échantillonné.

1.1.1 Stratification

Une bonne raison pour stratifier le domaine échantillonné est d'obtenir des variances plus faible. Les variances intra-strates sont plus faibles que les variances inter-strates. Ceci peut constituer un test pour mesurer la qualité de la stratification [1]. Les strates peuvent être un ensemble de rectangles statistiques, des zones définies à l'aide de connaissances a priori sur la distribution des groupes d'âges, des strates liées à la profondeur ou bien une combinaison de tous ces éléments. Ceci signifie que ce qui guide la définition des strates est le souhait d'obtenir des strates homogènes quand à la composition des longueurs. La définition des strates peut être basée sur des expériences de campagnes précédentes sur la même zone (pré-stratification) ou à partir de l'analyse de la campagne en cours (post-stratification). Il semble qu'il soit fréquent de définir les strates au regard des distributions des longueurs. Des tests basés sur les comparaisons (et similarités) des longueurs sont utilisés. Celui de Kolmogorov-Smirnov est présenté dans le paragraphe suivant et est utilisé par Kirkegaard *et al.* [19] et MacLennan et Simmonds [21].

1.1.2 Construction des strates

La construction des strates repose sur l'idée de regrouper dans de mêmes zones les échantillons contenant les distributions des longueurs les plus proches. Un test classique de comparaison des distributions statistiques est proposé par le test de Kolmogorov-Smirnov.

Le principe du test est le suivant [5, 32]:

Le test de Kolmogorov-Smirnov consiste à comparer les fonctions de répartition empiriques (F_1 et F_2) des longueurs de deux coups de chaluts d'effectifs n_1 et n_2 . Le test de l'hypothèse $H_0: "F_1 = F_2"$ contre l'hypothèse $H_1: "F_1 \neq F_2"$ est construit à partir de la statistique:

$$D_{n_1 n_2} = \max_i |F_1(x_i) - F_2(x_i)|$$

La quantité $n_1 n_2 D_{n_1 n_2}$ suit une loi de Kolmogorov. Une règle empirique donne :

- si $n_1 = n_2 \leq 25$, H_0 est rejetée pour $D_{n_1 n_2} \geq D_\alpha$ où D_α est déterminée à l'aide d'une table et α le seuil du test.
- si n_1 ou n_2 est ≥ 25 on peut utiliser une approximation utilisant $K_\alpha = \sqrt{\frac{-1}{2} \text{Ln}(\frac{\alpha}{2})}$ et $D_\alpha = K_\alpha \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$, l'hypothèse H_0 est rejetée si $D_{n_1 n_2} \geq D_\alpha$.

L'intérêt de ce test est qu'il n'impose aucune loi de distribution sur les échantillons et permet donc de travailler sur de petits échantillons.

Une autre règle empirique peut aussi être appliquée (Simmonds, 1993 [35])

- si $D_{n_1 n_2} \simeq 0$, les distributions peuvent être considérées comme similaires.
- si $D_{n_1 n_2} \gg 0$ les distributions peuvent être considérées comme très différentes.

L'exemple de la stratification de la campagne 1993 du FRV *Scotia* est présentée sur la figure 2.1 où trois strates (I, II et III) regroupant les stations de distribution des longueurs le plus homogènes possible ont été définies [33].

1.1.3 Moyennes des échantillons

Usuellement la variable utilisée est une longueur moyenne à l'intérieur d'une strate. En supposant qu'à l'intérieur d'une strate, il y a n échantillons la fréquence moyenne (f_i) du groupe de longueur i est calculé par l'expression [1]:

$$f_i = \frac{1}{\sum_j^n a_j} \sum_j^n a_j f_{ij}$$

où f_{ij} est la fréquence du groupe de longueur i de l'échantillon j et a_j est le pondérateur de l'échantillon j . Le pondérateur de l'échantillon peut, par exemple, être proportionnel à l'abondance mesurée à partir du coup de chalut. La valeur de la classe d'âge en est déduite à l'aide d'une clé taille/âge.

Un autre moyen est d'affecter par strate une valeur longueur moyenne constante, directement à partir des longueurs des poissons pêchés.

Remarque:

- Dans les calculs d'estimation globale par acoustique, la longueur moyenne utilisée par les praticiens est en fait la racine carrée de la moyenne des carrés des longueurs, $\sqrt{\overline{l(x)^2}}$, ("root mean square") qui au carré devient $\overline{l(x)^2}$. (On note, $\overline{l(x)^2}$ la moyenne des longueurs au carré mesurées au point x).

- Il s'avère qu'au niveau de chaque chalut, l'écart-type des longueurs mesurées est petit devant $\overline{l(x)}$ (de 1 à 2 cm), par conséquent : $\sqrt{\overline{l(x)^2}} \simeq \overline{l(x)}$.
- on pourra donc:
 - 1. résumer la distribution des longueurs par la longueur moyenne
 - 2. prendre $L_m^2 (= \overline{l(x)^2})$, au lieu de $\overline{l(x)^2}$ pour les calculs combinant l'acoustique avec les mesures pêchées.

1.1.4 Calcul de la surface équivalente

La surface équivalente σ peut être déterminée directement à partir des mesures *in situ*. Une longueur moyenne (L_m) ayant été calculée, une valeur moyenne σ_m est calculée par la formule [3]:

$$\sigma_m = 4\pi 10^{(a+bLn(L_m)/10)} \text{ (cm}^2\text{)}$$

où a et b dépendent de l'espèce du poisson et où L_m est mesurée en cm.

Dans le cas du hareng, on a plus simplement [21]:

$$\sigma_m = 6.3 \times 10^{-71.2} L_m^2$$

Il ressort de ceci que la connaissance de la distribution des longueurs est capitale pour mener à bout les calculs d'estimation de biomasse et a fortiori de variances globales de ces biomasses. On se propose, compte tenu des objectifs proposés, de fournir des améliorations à l'aide des méthodes géostatistiques.

1.2 Analyses géostatistiques

Dans cette partie, on se propose d'appliquer les méthodes d'analyse classiques de géostatistique à l'établissement de cartographies. Les calculs géostatistiques proposent des méthodes d'estimations locales tenant compte de la corrélation spatiale entre les mesures, c'est le **krigeage**. Des cartographies précises peuvent être envisagées. L'application proposée au chapitre 3 présentera des cartographies de la longueur déduite de chacune des campagnes de mesures. La longueur moyenne de chaque échantillon synthétisera l'information contenue dans chaque échantillon.

Le problème de la non stationnarité que l'on attribuera à l'influence de la profondeur sera abordé. En effet, la longueur moyenne du poisson s'accroît à mesure qu'il s'éloigne des côtes, en même temps que la profondeur augmente. Une technique de krigeage particulière le **krigeage avec dérive externe** permet l'utilisation d'une variable auxiliaire (la dérive externe) pour améliorer l'estimation. Une fois ces cartographies effectuées on en tirera une conclusion quant à la suite des calculs à adopter en vue d'une estimation globale.

De la même manière une cartographie des proportions d'âges sera proposée. On montrera qu'il existe un lien entre la longueur moyenne et les âges, la distribution des longueurs est liée à celle des âges. La structure spatiale des âges varie d'année en année alors que celle des longueurs moyennes est stable. Il apparaît ainsi nécessaire de faire une cartographie de la longueur puis d'en déduire celle des âges.

Chapitre 2

Méthodes géostatistiques de cartographie

Ce paragraphe présente succinctement les méthodes géostatistiques nécessaires aux cartographies qui nous intéresseront au chapitre 3 lors des cartographies de la longueur moyenne.

2.1 Estimation locale : technique du krigeage

Le krigeage est une technique d'estimation locale qui fournit le meilleur estimateur linéaire de la grandeur que l'on étudie [18]. On présente le krigeage ponctuel dans le cadre d'un modèle stationnaire d'ordre 2 où la moyenne est inconnue.

Soit $Z(x)$ une fonction aléatoire, de support ponctuel x de l'espace à deux dimensions, stationnaire d'ordre 2. Elle a pour espérance m (constante inconnue), pour covariance $C(h)$ et pour variogramme $\gamma(h)$.

On cherche à estimer la valeur moyenne $Z(x_0)$ implantée au point x_0 . Les valeurs connues mesurées par l'échantillonnage, aux points x_α , sont notées $z(x_\alpha)$. L'estimateur s'écrit comme une combinaison linéaire de n données d'un voisinage donné:

$$Z^K(x_0) = \sum_{\alpha=1}^n \lambda_\alpha z(x_\alpha)$$

Les pondérateurs λ_α dépendent de la corrélation spatiale (structure spatiale) entre les valeurs échantillonnées ainsi que de leur position dans l'espace. L'estimateur de krigeage est le meilleur estimateur linéaire sans biais.

– La condition de non biais s'écrit $E[Z(x_0) - Z^K(x_0)] = 0$ soit:

$$\sum_{\alpha=1}^n \lambda_\alpha = 1$$

La variance d'estimation (cas stationnaire) s'écrit:

$$\sigma_K^2 = E[Z(x_0) - Z^K(x_0)]^2 = E[Z(x_0)]^2 - 2E[Z(x_0)Z^K(x_0)] + E[Z^K(x_0)]^2$$

soit :

$$\sigma_K^2 = \int_V \int_V C(x-y) dx dy - 2 \sum_{\alpha=1}^n \lambda_\alpha C(x_\alpha - x_0) + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_\alpha \lambda_\beta C(x_\alpha - x_\beta)$$

où V est un voisinage autour du point x_0 .

- **La variance d'estimation doit être minimale** ce qui conduit à minimiser la fonction objectif:

$$\phi = \int_V \int_V C(x-y) dx dy - 2 \sum_{\alpha=1}^n \lambda_\alpha C(x_\alpha - x_0) + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_\alpha \lambda_\beta C(x_\alpha - x_\beta) - 2\mu \left(\sum_{\alpha=1}^n \lambda_\alpha - 1 \right)$$

Cette dernière équation est valable même si la covariance n'existe pas. S'il existe un variogramme mais pas de covariance une même équation peut être obtenue en changeant C en $-\gamma$.

On aboutit ainsi à la résolution d'un système où les inconnues sont les λ [7]:

$$\begin{cases} \sum_{\beta=1}^n \lambda_\beta C(x_\alpha - x_\beta) + \mu = C(x_\alpha - x_0), \forall \alpha \\ \sum_{\alpha=1}^n \lambda_\alpha = 1 \end{cases}$$

La variance d'estimation s'écrit donc :

$$\sigma_K^2 = \int_V \int_V C(x-y) dx dy - \sum_{\alpha=1}^n \lambda_\alpha C(x_\alpha - x_0) + \mu$$

Dans le but de résoudre ces équations, il est nécessaire de connaître un modèle de covariance. Si le variogramme est borné, variance et covariance sont reliées par l'expression $\gamma(h) = C(0) - C(h)$. Pour exprimer le système de krigeage en terme de variogramme il suffit de remplacer la covariance C par l'opposé du variogramme $-\gamma$ [22]. Si le variogramme n'est pas borné, le système n'a de sens qu'avec le variogramme. Les équations présentées ci-dessus avec la covariance sont valables dans le cas où la stationnarité peut n'être supposée que sur un voisinage V du point à estimer. On parle de quasi-stationnarité. Le krigeage effectué sur tout le champ échantillonné est dit alors à voisinage glissant. Chaque point du domaine à estimer est krigé avec un nombre réduit de points d'un voisinage où la fonction aléatoire peut être considérée comme "localement" stationnaire. En pratique, un variogramme inféré à petites distances est suffisant. On peut, par conséquent, très bien envisager une cartographie n'utilisant qu'un modèle de variogramme "local".

Les deux paragraphes suivant présentent des systèmes de krigeage où la quasi-stationnarité n'est plus envisageable. Il faut alors prendre en compte l'influence de dérivées.

2.2 L'estimateur de krigeage et la prise en compte de dérivées

L'inférence de la covariance, même sur un voisinage réduit, peut parfois être rendue impossible par la présence d'une dérivée. On peut admettre que cette dérivée

est au moins localement de forme polynômiale. On se limite à ne travailler qu'avec des combinaisons linéaires qui filtrent les polynômes de degré inférieur ou égal à k donné et on n'impose plus la condition de stationnarité qu'à ces combinaisons linéaires [8]. L'inférence statistique ne porte alors plus sur toute la covariance mais sur une partie de celle-ci [8].

Dans d'autres cas, on impute à une autre variable la responsabilité du comportement stationnaire de la fonction aléatoire qui nous intéresse. Cette variable auxiliaire, connue à peu près partout dans le domaine spatial, donne l'allure générale de la morphologie du phénomène régionalisé [39].

2.2.1 La dérive est une fonction polynôme des coordonnées

On considère ici que la fonction aléatoire s'écrit [7, 8]:

$$Z(x) = m(x) + \epsilon(x)$$

où $m(x)$ est la partie déterministe inconnue et $\epsilon(x)$ la partie probabiliste stationnaire d'ordre 2, d'espérance nulle et de covariance $K(h)$.

La dérive $m(x)$ est un polynôme de degré k de fonctions des coordonnées $f_l(x)$.

La dérive s'écrit:

$$m(x) = \sum_{l=0}^L a_l f_l(x)$$

où les f_l sont un jeu de fonctions de base monomiales $(1, x, y, x^2, \dots)$ et où les a_l sont des coefficients inconnus.

L'estimateur de $Z(x_0)$ en x_0 s'écrit:

$$Z^K(x_0) = \sum_{\alpha} \lambda_{\alpha} Z(x_{\alpha})$$

La condition de non biais conduit à:

$$\sum_{\alpha} \lambda_{\alpha} f_l(x_{\alpha}) = f_l(x_0), \forall l = 0, \dots, L-1$$

et la minimisation de la variance de l'erreur d'estimation $Z(x_0) - Z^K(x_0)$ conduit au système de krigeage suivant, où interviennent $L+1$ paramètres de Lagrange μ_l .

$$\begin{pmatrix} K(x_{\alpha} - x_{\beta}) & f_l(x_{\alpha}) \\ f_l(x_{\beta}) & 0 \end{pmatrix} \begin{pmatrix} \lambda_{\beta} \\ \mu_l \end{pmatrix} = \begin{pmatrix} K(x_{\alpha} - x_0) \\ f_l(x_0) \end{pmatrix}$$

Il n'est pas nécessaire de spécifier les coefficients a_l pour estimer la variable $m(x)$. Seul est requis le degré $L-1$ d'un polynôme s'ajustant localement, en moyenne sur le champ [7].

L'inférence statistique de la covariance est assez délicate. Le comportement "naïf" consisterait à estimer une dérive par une méthode de régression puis d'en déduire un résidu et sa structure. Malheureusement ceci conduit à un variogramme des résidus sous-estimé [22]. Pour estimer ce variogramme sans passer

par l'estimation de la dérive, il existe des méthodes d'inférence statistique assez sophistiquées [8, 17]. Les équations écrites plus haut restent valables à condition de remplacer la covariance stationnaire par une covariance dite "généralisée" correspondant à des combinaisons linéaires filtrant la dérive (théorie des fonctions intrinsèque d'ordre k [8]).

2.2.2 L'estimateur de krigeage avec dérive externe

Si la variable d'étude est connue pour suivre dans l'espace la forme de la distribution spatiale d'une variable explicative, la méthode de la dérive externe permet de réaliser une cartographie conforme au lien entre les deux variables [12].

Dans ce modèle, la fonction aléatoire $Z(x)$ est en moyenne proportionnelle à une variable régionalisée $s(x)$ à une constante près [39]:

$$Z(x) = a + bs(x) + \epsilon(x)$$

La fonction $s(x)$ décrit l'allure générale du phénomène observé et ϵ est une fonction aléatoire stationnaire d'ordre 2 de variance $K(h)$. Au point x_0 , l'estimateur $Z^K(x_0)$ est tel que:

$$\begin{aligned} E(Z^K(x_0)) &= \sum_{\alpha=1}^n \lambda_{\alpha} E(z(x_{\alpha})) = a + b \sum_{\alpha=1}^n \lambda_{\alpha} s(x_{\alpha}) \\ &= a + bs(x_0) \end{aligned}$$

Il y a donc, par rapport au système de krigeage ordinaire, une condition supplémentaire sur les pondérateurs, si l'on veut que la condition de non biais soit respectée:

$$s(x_0) = \sum_{\alpha=1}^n \lambda_{\alpha} s(x_{\alpha})$$

La fonction à minimiser devient:

$$\phi = \sigma_K^2 + \mu_1 \left(\sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right) + \mu_2 \left(\sum_{\alpha=1}^n \lambda_{\alpha} s(x_{\alpha}) - s(x_0) \right)$$

Ce qui conduit finalement au système de krigeage [39]:

$$\begin{pmatrix} K(x_{\alpha} - x_{\beta}) & 1 & s(x_{\alpha}) \\ 1 & 0 & 0 \\ s(x_{\beta}) & 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda_{\beta} \\ \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} K(x_{\alpha} - x_0) \\ 1 \\ s(x_0) \end{pmatrix}$$

La variance d'estimation au point x_0 s'écrit donc:

$$\sigma_K^2 = K(0) - \sum_{\alpha=1}^n \lambda_{\alpha} K(x_{\alpha} - x_{\beta}) - \mu_1 - \mu_2 s(x_0)$$

Un modèle considérant que la source de la non stationnarité de la fonction $Z(x)$ est imputable à deux phénomènes, une dérive "interne" et une dérive "externe" s'exprime ainsi:

$$\begin{pmatrix} K(x_\alpha - x_\beta) & f_i(x_\alpha) & s(x_\alpha) \\ f_i(x_\beta) & 0 & 0 \\ s(x_\beta) & 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda_\beta \\ \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} K(x_\alpha - x_0) \\ f_i(x_0) \\ s(x_0) \end{pmatrix}$$

La méthode de la dérive externe consiste ainsi à ajouter des contraintes à l'estimateur présenté au paragraphe précédent afin que celui-ci respecte en plus une forme spécifiée par une fonction $s(x)$ de la variable explicative. On doit connaître les valeurs de $s(x_\alpha)$ en tous les points de données x_α des échantillons de $Z(x)$, ainsi qu'aux noeuds de la grille d'estimation.

Le chapitre 3 se servira de tels estimateurs lors de l'estimation de la longueur dans laquelle la non stationnarité sera attribuée à la profondeur qui jouera le rôle de la dérive externe.

2.3 Approches multivariées: krigeage et simulation

2.3.1 Le système de cokrigeage

Il arrive que l'on soit amené à estimer une variable à l'aide d'autres variables régionalisées. D'un point de vue théorique, le cokrigeage ne diffère pas du krigeage.

Soit la corégionalisation caractérisée par l'ensemble des s fonctions aléatoires intercorrélées spatialement, $Z_s(x)$, $s = 1, \dots, S$. Sous l'hypothèse stationnaire: $E(Z_s(x)) = m_s = \text{constante}$. On s'intéresse à l'estimation au point x_0 de la fonction $Z_{s_0}(x)$, prise parmi les $Z_s(x)$. Cet estimateur est noté $Z_{s_0}^K(x_0)$. L'estimateur de cokrigeage est une combinaison linéaire de pondérateurs λ_α^s avec des variables situées en des points d'un voisinage du domaine [39]:

$$Z_{s_0}^K(x_0) = \sum_{s=1}^S \sum_{\alpha=1}^n \lambda_\alpha^s Z_s(x_\alpha)$$

La condition de non biais est remplie en choisissant des pondérateurs de somme 1 pour la variable d'intérêt, et de somme nulle pour les variables auxiliaires:

$$\sum_{\alpha=1}^n \lambda_\alpha^s = \delta_{ss_0} = \begin{cases} 1 & \text{si } s = s_0, \\ 0 & \text{sinon} \end{cases}$$

Le biais s'écrit:

$$\begin{aligned} E(Z_{s_0}(x_0) - Z_{s_0}^K(x_0)) &= E(Z_{s_0}(x_0) - \sum_{s=1}^S \sum_{\alpha=1}^n \lambda_\alpha^s Z_s(x_\alpha)) \\ &= E\left(\sum_{\alpha=1}^n \lambda_\alpha^{s_0} Z_{s_0}(x_0) + \sum_{s=1, s \neq s_0}^S \sum_{\alpha=1}^n \lambda_\alpha^s Z_s(x_0) - \sum_{s=1}^S \sum_{\alpha=1}^n \lambda_\alpha^s Z_s(x_\alpha)\right) \end{aligned}$$

$$= \sum_{s=1}^S \sum_{\alpha=1}^n \lambda_{\alpha}^s E(Z_s(x_0) - Z_s(x_{\alpha})) = 0$$

La variance d'estimation s'écrit :

$$\sigma_K^2 = E\left[\sum_{s=1}^S \sum_{\alpha=1}^n \lambda_{\alpha}^s Z_s(x_{\alpha}) - Z_{s_0}(x_0)\right]^2$$

Considérant l'existence d'une covariance croisée définie entre deux fonctions aléatoires $Z_{s_1}(x)$ et $Z_{s_2}(x)$ stationnaires [39], on a :

$$C_{s_1 s_2}(h) = E[(Z_{s_1}(x) - m_{s_1})(Z_{s_2}(x+h) - m_{s_2})]$$

la variance d'estimation s'écrit :

$$\sigma_K^2 = C_{s_0 s_0}(x_0 - x_0) - 2 \sum_{s=1}^S \sum_{\alpha=1}^n \lambda_{\alpha}^s C_{s s_0}(x_{\alpha} - x_0) + \sum_{s=1}^S \sum_{t=1}^S \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha}^s \lambda_{\beta}^t C_{st}(x_{\alpha} - x_{\beta})$$

Après minimisation, où les contraintes sur les pondérateurs nécessitent S paramètres de Lagrange on obtient le système de cokrigage suivant :

$$\begin{cases} \sum_{j=1}^S \sum_{\beta=1}^n \lambda_{\beta}^j C_{sj}(x_{\alpha} - x_{\beta}) + \mu_s = C_{s s_0}(x_{\alpha} - x_0), \forall s = 1, \dots, S; \alpha = 1, \dots, n \\ \sum_{\alpha=1}^n \lambda_{\alpha}^s = \delta_{s s_0}, \forall s = 1, \dots, S \end{cases}$$

Ce qui conduit à la variance de cokrigage :

$$\sigma_K^2 = C_{s_0 s_0}(x_0 - x_0) + \mu_{s_0} - \sum_{s=1}^S \sum_{\alpha=1}^n \lambda_{\alpha}^s C_{s s_0}(x_{\alpha} - x_0)$$

Une étude géostatistique multivariable utilisant le cokrigage nécessite donc le calcul et la modélisation de toutes les covariances (ou variogrammes).

Le cas qui nous servira dans le chapitre 3 est celui où on considèrera que les fonctions aléatoires sont en "corrélation intrinsèque". Les F.A. sont en corrélation intrinsèque si leurs covariances ou variogrammes simples et croisés sont tous proportionnels [18, 39] :

$$C_{s_1 s_2}(h) = b_{s_1 s_2} C_0(h), \forall s_1, s_2$$

Si les configurations d'information de chaque variable sont identiques, on montre que le cokrigage d'une variable particulière est identique au krigeage de cette variable.

L'ajustement des covariances (ou variogrammes) a été formalisé dans le cadre d'un modèle dit "modèle linéaire de corégionalisation". Les paliers des variogrammes croisés suivent la contrainte suivante garantissant un modèle autorisé [39] :

$$|b_{s_1 s_2}| \leq \sqrt{b_{s_1 s_1} b_{s_2 s_2}}$$

2.3.2 Principe des simulations conditionnelles géostatistiques

Chaque mesure $z(x)$ sur le champ V est interprétée comme une réalisation d'une certaine fonction aléatoire $Z(x)$ caractérisée par une loi de distribution et une covariance (ou un variogramme). L'idée des simulations consiste à tirer d'autres réalisations z_s de cette même classe de F.A. On parle de simulations conditionnelles quand on ne retient que les réalisations $z_{sc}(x)$ qui retrouvent les données expérimentales aux points d'échantillonnage x_α [18].

Le principe

La F.A. $Z(x)$ à simuler est considérée stationnaire multigaussienne d'espérance m et de covariance $C(h)$ (ou de variogramme $\gamma(h)$). Une simulation géostatistique cherche à construire une réalisation de $Z(x)$ présentant les mêmes espérance et covariance. Étant conditionnelle, elle doit retrouver les données expérimentales. Considérons au point x , la vraie valeur $Z(x)$ et son estimateur de krigeage $Z^K(x)$. On peut écrire:

$$Z(x) = Z^K(x) + [Z(x) - Z^K(x)]$$

Sous les hypothèses gaussiennes, le processus constitué par l'erreur de krigeage $[Z(x) - Z^K(x)]$ est indépendant du processus $Z^K(x)$. Le principe de la simulation conditionnelle consiste donc à simuler une erreur de krigeage $[Z_s(x) - Z_s^K(x)]$ et à lui ajouter la valeur krigée. Dans la pratique, on effectue une simulation z_s de la F.A.. La procédure de krigeage appliquée aux données $z_s(\alpha)$ conduit à une erreur de krigeage $[z_s(x) - z_s^K(x)]$ indépendante de $Z^K(x)$. La simulation conditionnelle s'écrit donc:

$$z_{sc}(x) = z^K(x) + [z_s(x) - z_s^K(x)].$$

La mise en oeuvre

Pour obtenir la simulation conditionnelle $z_{sc}(x)$ il suffit de simuler non conditionnellement une réalisation $z_s(x)$ de la F.A. $Z(x)$ de mêmes espérance et covariance. Simuler à deux ou à trois dimensions une F.A. à covariance imposée est une opération devenue usuelle en géostatistique. La méthode qui sera utilisée ici est celle dite "des tranches tournantes" proposée par G. Matheron [23] (in Journal, 1977 [18]) qui consiste à réduire toute simulation à trois (ou deux) dimensions à plusieurs simulations indépendantes à une dimension sur des droites que l'on fait tourner dans l'espace. Les détails de la méthode sont présentés par Journal (1977). Cette méthode est maintenant accessible usuellement dans le logiciel géostatistique ISATIS [17] qui fournit des simulations multidimensionnelles en des temps très faibles. Les réalisations issues de ces simulations ont des lois spatiales gaussiennes. Si la distribution expérimentale des données n'est pas gaussienne, il est alors nécessaire de travailler sur une variable gaussienne transformée de la variable brute.

En terme de variance d'estimation, le krigeage est deux fois meilleur que la simulation conditionnelle: l'estimation n'est pas l'objet d'une simulation [18]. L'intérêt d'une simulation est de reproduire un histogramme et une covariance (variogramme). Nous nous en servons dans le chapitre 3, où nous chercherons

à obtenir des cartes simulées de distributions de classes d'âge de poisson. Plusieurs cartes simulées permettent d'estimer une moyenne globale (moyenne des moyennes globales sur chacune des simulations) et une variance d'estimation de cette moyenne comme variance des moyennes des résultats de simulations. De plus, considérant que les régionalisations des différentes classes d'âges sont corrélées, nous procéderons à des simulations conditionnelles multivariées. Dans le cadre du "modèle linéaire de corégionalisation" [39] les variables à simuler sont décomposées linéairement sur une base de facteurs indépendants. Une simulation non conditionnelle multivariée des variables est d'abord obtenue par combinaison de simulations non conditionnelles indépendantes des facteurs. La simulation conditionnelle multivariée s'obtient alors par un cokrigage conditionnant (qui se réduit à un krigeage conditionnant de chaque variable si celles-ci sont en corrélation intrinsèque).

Chapitre 3

Application : cartographie de la longueur et des classes d'âge

3.1 Présentation du cas d'étude

L'objectif des campagnes d'échantillonnage est de fournir des informations permettant de calculer des estimations globales par classes d'âge. Ainsi, la structure en âge et en taille de chaque "coup de chalut" est mesurée. Ce chapitre propose comme préalable à l'estimation des quantités globales par classes d'âge de présenter l'analyse des différentes campagnes de mesures chalutées concernant la distribution des longueurs et la distribution des classes d'âge.

Les données utilisées sont celles fournies par le laboratoire de Biologie Marine d'Aberdeen concernant le stock de hareng de la Mer du Nord (*Clupea harengus*) autour des îles Orkney et Shetland (division CIEM IVa). A chaque station d'échantillonnage sont mesurées pour chaque coup de chalut les fréquences de taille et d'âge des poissons pêchés. Elles ont été mesurées lors de la collecte de mesures acoustiques au cours des campagnes de 1989 à 1994. Un chalutage est effectué quand la mesure acoustique indique une forte densité de poisson et est utilisé pour déterminer la distribution des longueurs et la proportion de chaque classe d'âge à chaque station. La classe d'âge 0 sera notée A0, celle d'âge 1, A1, 2 immature noté I2, 2 mature noté M2, 3 immature noté I3, 3 mature noté M3, 4 noté A4, 5 noté A5, ..., 8 noté A8 et 9 ans et plus noté A9.plus. Une variable longueur moyenne (L_m) est calculée qui synthétise l'éventail des longueurs mesurées.

Deux phénomènes biologiques ont été observés et ont guidé notre travail:

- en général, plus le poisson est grand et agé, plus sa position s'éloigne des côtes. La distribution spatiale de la longueur moyenne n'est pas stationnaire et dépend de la profondeur du fond. Cependant, elle était relativement stable d'années en années, entre 1989 et 1994.
- les distributions spatiales des proportions d'âge, elles, varient beaucoup d'une année sur l'autre. Cartographier directement les proportions d'âge est une tâche trop difficile pour être tentée directement. On a donc cherché à relier proportion d'âge et longueur moyenne à l'aide de relations non linéaires.

L'exploitation de ces phénomènes a constitué la suite de notre travail.

3.2 Distribution spatiale de la longueur moyenne

3.2.1 Analyse de cartes et variographie

Les cartes de la figure 2.2 donnent une représentation proportionnelle des longueurs moyennes pour les campagnes 1989 à 1994. Au total sur les six campagnes, 135 mesures ont été effectuées. Chaque valeur moyenne est représentée par un cercle de rayon proportionnel à sa valeur. Le cercle de rayon maximal correspond à la valeur maximale des longueurs moyennes pour les six campagnes. Les autres cercles s'en déduisent proportionnellement. Ainsi les six distributions spatiales peuvent être comparées. On est frappé par la similarité des distributions présentées. La figure 2.14 présente une représentation proportionnelle où les six campagnes sont superposées. La variabilité spatiale de la longueur est en moyenne stable d'année en année. Sur ces figures, on a représenté, de plus, la bathymétrie. Les isolignes de profondeur 100, 200, 300 et 400 mètres de profondeur sont rajoutées en figurés pointillés. La zone échantillonnée correspond donc au plateau de profondeur inférieur à 300 mètres. L'échantillonnage est même majoritairement effectué à proximité de l'isoligne 100 mètres. Seule est observable une dérive spatiale vers les zones de plus grande profondeur. En général, plus le poisson est grand plus sa position s'éloigne des côtes.

Un autre moyen de représenter ce phénomène et de traduire ces deux observations est de calculer les variogrammes de la longueur moyenne mesurant la variabilité moyenne entre 2 points comme fonction de la distance entre ces points. Deux types de variogrammes peuvent ainsi être calculés:

- Le variogramme calculé avec toutes les mesures sans distinction de l'année de la campagne (les couples de points échantillonnés la même année ne sont pas retenus).
- Le variogramme calculé comme moyenne pondérée des six variogrammes annuels. À chaque distance donnée, le variogramme est la moyenne pondérée par les nombres de couples des six variogrammes annuels.

Pratiquement, le pas choisi pour le calcul du variogramme est de 10 milles nautiques et la tolérance angulaire de 90 degrés. Les variogrammes ne sont donc pas directionnels, l'échantillonnage n'étant pas spatialement directionnel.

Ces deux variogrammes sont représentés superposés sur la figure 2.16. Celui correspondant à toutes les années mélangées est figuré en trait plein, celui correspondant à la moyenne de six variogrammes annuels est en trait pointillé. Le décalage visible à petite distance (environ 2) entre les deux variogrammes fournit une mesure de la variabilité inter-annuelle en chaque point de l'espace. L'effet de pépité correspondant au premier variogramme est assez faible, il est quasiment nul en ce qui concerne le deuxième variogramme. On peut admettre ainsi la grande stabilité spatiale inter-annuelle de la longueur moyenne. D'allures linéaires, les deux variogrammes traduisent de manières pratiquement similaires la

non stationnarité de la longueur. On peut donc considérer que la variabilité de la longueur pourrait être modélisée par la somme de deux composantes:

- un effet de pépite traduisant totalement la variabilité inter-annuelle.
- une composante linéaire traduisant la variabilité spatiale moyenne intra-annuelle et constante d'année en année.

Un tel modèle (variogramme linéaire + effet de pépite différentiel) pourrait être utilisé afin de réaliser une cartographie de la longueur moyenne. Ce n'est pas ce modèle qui a été choisi, nous avons voulu en effet tenir compte de l'influence de la profondeur intervenant comme dérive externe et responsable de la non stationnarité de la longueur moyenne.

3.2.2 Cartographie: krigeage avec dérive externe et effet de pépite différentiel

Comme la taille du poisson augmente avec son éloignement des côtes, il n'est pas surprenant d'observer une relation croissante, pratiquement linéaire, entre la longueur et la profondeur (fig. 2.18). On admet à partir de là que autour de tout point x , la longueur moyenne (L_m) a une forme locale moyenne s'exprimant comme une fonction linéaire de la profondeur.

$$E(L_m(x)) = a + b.s(x)$$

où $s(x)$ est la profondeur au point x . Les données de profondeur peuvent servir de données auxiliaires aux données de longueur donnant ainsi la forme générale de la dérive dirigeant la longueur qui peut alors être utilisée pour réaliser un krigeage de la longueur moyenne [12]: c'est la technique du **krigeage avec dérive externe** présentée au chapitre 2 §2.2. La covariance $K(h)$ mise en jeu modélise la structure spatiale du résidu. Le modèle choisi est celui que nous estimons compatible avec le variogramme des résidus de la régression Longueur/profondeur (fig.2.19).

Le calcul du variogramme des résidus de la régression Longueur/profondeur a été calculé de manière à isoler deux termes traduisant d'une part une variabilité spatiale stationnaire indépendant de l'année et d'autre part, une variabilité inter-annuelle. Il semble que les variogrammes expérimentaux calculés par ce moyen se prêtent à une modélisation de ce type:

- une composante pépitique différentiel de valeur 1.8 (variabilité inter-annuelle)
- une composante sphérique de portée 62 milles nautiques et de palier 7.5 (variabilité intra-annuelle)

Pratiquement, lors de la cartographie d'une année donnée, le voisinage étant donné les pondérateurs du système de krigeage sont fonctions de la valeur de la covariance du résidu. Celle-ci vaut donc la somme des deux composantes présentée plus haut sauf lorsqu'elle fait intervenir des couples de points appartenant à la même année où l'effet différentiel est alors nul. L'intérêt d'une telle modélisation est de pouvoir réaliser des cartographies annuelles de la longueur moyenne à

l'aide des données collectées toutes les années. La variabilité inter-annuelle donne une flexibilité qui fait que les cartes annuelles ne respectent pas exactement les données des autres années mais respectent uniquement celles de l'année.

Le système de krigeage est celui du krigeage avec dérive externe que l'on peut adapter comme suit où:

- $z(x_\alpha)$ représente l'année correspondant à la mesure x_α .
- $K(x_\alpha - x_\beta, z(x_\alpha), z(x_\beta))$ est la fonction de covariance à deux structures:

$$K(x_\alpha - x_\beta, z(x_\alpha), z(x_\beta)) = K_0 1_{[z(x_\alpha)=z(x_\beta)]} + K_1(x_\alpha - x_\beta)$$

- où K_0 est une constante (ici égale à 1.8) et $K_1(h)$ une fonction de covariance ne dépendant que de la distance entre deux points de l'espace à deux dimensions (ici un modèle sphérique de palier 7.5 et de portée 62 milles nautiques).
- $1_{[z(x_\alpha)=z(x_\beta)]}$ étant une fonction valant 1 si lorsqu'un couple de point appartient à la même année et 0 sinon.

$$\begin{pmatrix} K(x_\alpha - x_\beta, z(x_\alpha), z(x_\beta)) & 1 & s(x_\alpha) \\ 1 & 0 & 0 \\ s(x_\beta) & 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda_\beta \\ \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} K(x_\alpha - x_0, z(x_\alpha), z(x_0)) \\ 1 \\ s(x_0) \end{pmatrix}$$

3.2.3 Conclusion: stabilité inter-annuelle de la longueur

Avec cette technique, on obtient des cartes de krigeage de la longueur moyenne pour chaque année. Elles sont présentées à l'aide d'isoligne sur la figure 2.20 où les valeurs mesurées lors de l'échantillonnage sont superposées. Le krigeage est effectué à voisinage unique et a été réalisé à l'aide de routines informatiques conçues par C. Lajaunie et J. Rivoirard sur le logiciel Splus. La maille utilisée est une maille carrée de 5 mn sur 5 mn. Les isolignes en trait plein correspondent aux valeurs 25, 30 et 35 cm, ceux en trait pointillé correspondent aux isolignes 22, 27 et 32 cm.

On remarque que l'isoligne 27 cm correspond, en général, à la profondeur 100m (sauf pour l'année 1989 qui semble donner des résultats un peu différents). Les isolignes 30 et 32 cm correspondent à la chute du plateau continental dans la région (300, 400 m). On distingue très clairement une zone occupée chaque année par les poissons inférieurs à 27 cm. Il existerait ainsi une "boîte écologique" sur le plateau continental, entre les îles et le continent, de poissons de tailles inférieures à 27 cm.

Au nord de cette "boîte écologique", à partir de l'isoligne 30 cm, la distribution spatiale des longueurs semble plus variable selon les années. L'échantillonnage y est moins intense et en ce qui concerne le krigeage, la zone au delà de la latitude 61° est majoritairement une zone d'extrapolation.

On obtient, ainsi, six cartes de krigeage de la longueur obtenues à l'aide des données expérimentales. Une dernière carte peut être obtenue pour une année supplémentaire ne comportant pas de données expérimentales (1995, par exemple). Elle peut s'interpréter comme une cartographie moyenne de la longueur sur le champ. Elle est présentée sur la figure 2.22.

La figure 2.21 donne une cartographie des écarts-types d'estimation de la longueur krigée. Ceux-ci varient peu d'une année sur l'autre. Les isolignes en pointillés correspondent à la valeur 2 et les isolignes pleines correspondent à 3 et 5 (limite du champ échantillonné. L'essentiel des estimations conduit à des écarts-types d'estimation compris entre 2 et 3 (par exemple pour la campagne 1993, 70% des valeurs appartiennent à l'intervalle [2,3] et la moyenne des écarts-types vaut 2.5). Cette erreur d'estimation n'est pas négligeable, on s'interrogera plus loin (partie III) sur la manière dont on pourrait prendre en compte la variabilité des longueurs sur un calcul de variance globale. Néanmoins dans la suite de cette partie, nous conférerons un statut déterministe à la distribution des longueurs.

3.3 Distribution spatiale des fréquences d'âge

3.3.1 Analyse de cartes et variographie

À chaque station d'échantillonnage, la proportion des classes d'âge est mesurée. Au total on considère 11 classes d'âge de la classe 0 à la classe 9 ans et plus: 0 noté A0, 1 noté A1, 2 immature noté I2, 2 mature noté M2, 3 immature noté I3, 3 mature noté M3, 4 noté A4, 5 noté A5, ..., 8 noté A8 et 9 et plus noté A9.plus).

La représentation proportionnelle des distributions des proportions d'âge en comparaison avec celles des longueurs montre bien leur variabilité inter-annuelle figures 2.3 à 2.13. Ici, encore un figuré proportionnel a été choisi. Le figuré correspondant au maximum est calé sur le maximum de toutes les campagnes pour chaque classe d'âge. Ceci permet une visualisation très nette de la variabilité inter-annuelle de la distribution des âges. Globalement, les âges élevés se distribuent préférentiellement loin des côtes et les âges jeunes près des côtes. Seulement d'années en années on peut trouver de fortes variations en un même point de l'espace. Les cartes confondant toutes les classes d'âge sont très claires, sur des distances proches on note la présence de fortes valeurs (grands cercles) et de plus faibles valeurs (de petits voire de très petits cercles). (fig.2.14 et 2.15). Il semble d'autre part exister une zone préférentiellement occupée par les âges jeunes.

La variabilité inter-annuelle des âges comme celle des longueurs peut être mise en évidence à l'aide des variogrammes moyens calculés comme précédemment. Ces variogrammes expérimentaux sont présentés sur les figures 2.16 et 2.17.

La différence (décalage) entre les deux premiers points des deux variogrammes que l'on peut imputer à la variabilité inter-annuelle en chaque point du domaine est, en général, très élevée.

On se rend compte également de la difficulté de modéliser de tels variogrammes. Ils sont soit fortement non stationnaires soit totalement destructurés. Une cartographie directe des âges à partir des variogrammes des proportions d'âge est impossible. Il faut envisager une autre méthode.

3.3.2 Conclusion: variabilité inter-annuelle des fréquences d'âge

Écologiquement c'est la longueur qui est informante. Nos résultats confirment quelques faits biologiques bien connus des biologistes:

- le métabolisme du poisson dépend de sa taille ou mieux, de son poids (qui intègre trois dimensions). C'est la taille qui influence la reproduction beaucoup plus que l'âge.
- À un âge donné, il y a de plus des différences de croissance d'année en année [9].
- La formation des bancs chez les clupeidés (famille à laquelle appartient le hareng) n'est possible que pour des groupes de poissons dont les tailles sont comprises entre des limites relativement proches. Si dans une zone déterminée se trouvent plusieurs classes de tailles, elles formeront des bancs séparés [9].

Ces faits biologiques nous incitent donc à considérer la longueur moyenne comme paramètre conditionnant la distribution des âges.

3.4 Lien âge/longueur et cartographie des proportions d'âge

La distribution des fréquences d'âge nous apparaît, ainsi, fortement non stationnaire. Cette non stationnarité ne s'écrit pas sous la forme d'un polynôme des coordonnées; On cherche, ici, à exploiter la relation entre l'âge et la longueur pour aboutir à des cartographies des proportions des classes d'âges. La non stationnarité de la distribution des fréquences d'âge est due à l'existence d'un lien non linéaire avec la longueur moyenne.

Le contexte global de l'étude est d'obtenir des estimations globales d'abondance par classes d'âge. Il est par conséquent nécessaire de pouvoir obtenir dans un premier temps des cartographies des âges. La variabilité des âges étant beaucoup plus forte que celle des longueurs, nous avons fait le choix de considérer la longueur comme paramètre conditionnant la répartition des âges. Une année étant fixée, nous regardons la proportion des âges comme une fonction des longueurs moyennes. La suite de ce chapitre est donc consacrée à la modélisation de la relation âge/longueur. La longueur est considérée comme étant déterministe, on donnera à la variable "classe d'âge" le statut de variable déterministe où probabiliste à travers deux approches. L'intérêt d'une approche probabiliste est de pouvoir calculer des variances.

3.4.1 Observation des nuages de corrélation

Le lien âge/longueur est représenté sur des diagrammes de corrélations présentés sur les figures 2.25 et 2.24 où on a représenté pour toutes les années confondues,

les classes d'âge en fonction de la longueur moyenne. On remarque une forte dispersion autour de la moyenne. Pour une longueur donnée la proportion à un âge donné peut varier du simple au quadruple. Cette dispersion est considérablement réduite si l'on considère les nuages par années. Ces graphes donnent ainsi une autre caractérisation de la variabilité inter-annuelle des âges.

Si au lieu des proportions d'âge on considère les proportions à partir d'un âge donné, par exemple $T_4 = A_4 + A_5 + \dots + A_9$, plus, donnant la proportion d'âge supérieure ou égale à 4 ans. La relation avec la longueur devient encore plus étroite pour toutes les années mélangées (fig. 2.25 et 2.24). Ces mêmes diagrammes considérés pour chaque année séparée montrent des liens quasi déterministes (fig. 2.27 et 2.28). Ces liens peuvent être modélisés par une fonction non linéaire de type logistique et ajustés à l'aide de la méthode des moindres carrés non linéaire [6, 38]. On a cherché à ajuster des régressions du type :

$$T_i = \frac{e^{a_i L_m + b_i}}{1 + e^{a_i L_m + b_i}} + \epsilon_i$$

où a_i et b_i sont caractéristiques de la coupure (âge cumulé) T_i donnée. L'objectif pratique est d'utiliser ces liens mis en évidence pour obtenir des cartes de proportions d'âge en utilisant :

- la régression comme transformation $T = f(L_m)$
- puis par soustraction des seuils estimés, par exemple : $A_4 = T_4 - T_5$.

Certaines relations ne sont pas tout à fait déterministes (ex: T3.1, proportion supérieure à 3 ans) par conséquent la transformation $T_i = f(L_m)$ n'est pas satisfaisante et il nous semble intéressant de considérer une approche où l'on tient compte des résidus. On propose alors une approche probabiliste où la structure des résidus sera utilisée pour obtenir des simulations. Les résidus de toutes ces régressions sont a priori corrélés, l'approche doit être considérée sous un angle multivariable.

3.4.2 Approche déterministe

On considère pour l'année 1993 comportant 32 stations de mesures, les régressions :

$$T_i(x) = \frac{e^{a_i L_m(x) + b_i}}{1 + e^{a_i L_m(x) + b_i}}$$

Ces régressions représentent l'évolution de proportions cumulées qui, décollant de zéro, augmentent pour se stabiliser à 1 (sauf pour T_0 et T_1 qui ne sont pas représentées, quasiment égales à 1). Ces régressions étant considérées comme déterministes, on décide de négliger l'influence du résidu. Des cartes d'estimation des âges cumulés T_i peuvent alors être déduites de la carte de la longueur moyenne krigée de l'année 1993 ainsi que, par soustraction, des cartes des proportions d'âge. On a donc un moyen simple de résoudre le problème de la non stationnarité des classes d'âge. La non stationnarité des classes d'âge provient d'un lien non linéaire avec une variable auxiliaire, ici la longueur moyenne. Le résultat des ajustements

des régressions (à partir de l'âge 2) par la méthode des moindres carrés non linéaires est explicité tableau 3.1.

TAB. 3.1 - Régressions âges cumulés/longueur moyenne

$T_{2,1}(x) = \exp(1.22L_m(x) - 27.17)/(1 + \exp(1.22L_m(x) - 27.17))$
$T_{2,2}(x) = \exp(0.88L_m(x) - 20.86)/(1 + \exp(0.88L_m(x) - 20.86))$
$T_{3,1}(x) = \exp(0.56L_m(x) - 14.79)/(1 + \exp(0.56L_m(x) - 14.79))$
$T_{3,2}(x) = \exp(0.68L_m(x) - 18.40)/(1 + \exp(0.68L_m(x) - 18.40))$
$T_4(x) = \exp(0.77L_m(x) - 22.20)/(1 + \exp(0.77L_m(x) - 22.20))$
$T_5(x) = \exp(0.71L_m(x) - 20.85)/(1 + \exp(0.71L_m(x) - 20.85))$
$T_6(x) = \exp(0.61L_m(x) - 18.49)/(1 + \exp(0.61L_m(x) - 18.49))$
$T_7(x) = \exp(0.52L_m(x) - 17.20)/(1 + \exp(0.52L_m(x) - 17.20))$
$T_8(x) = \exp(0.57L_m(x) - 19.84)/(1 + \exp(0.57L_m(x) - 19.84))$
$T_9(x) = \exp(0.62L_m(x) - 22.30)/(1 + \exp(0.62L_m(x) - 22.30))$

Les résultats des cartographies pour l'année 1993 issues de cette méthode sont présentés sur les figures 2.29 et 2.30 pour les âges cumulés et 2.31 et 2.32 en ce qui concerne les proportions d'âge. Ces cartes sont présentées sous forme d'isoligne et sur un domaine restreint à une zone correspondant à la zone d'échantillonnage acoustique. La représentation des âges cumulés est faite à l'aide des isolignes 0.1, 0.5, 0.9 (en pointillés) et 0.3 et 0.7 (en traits pleins). Les valeurs superposées correspondent aux valeurs d'origine déduites des mesures expérimentales, elles permettent de noter que le conditionnement par l'intermédiaire de la régression par la longueur se passe bien. On note un fort lien avec la profondeur, c'est l'influence de la profondeur via le processus d'estimation utilisant une dérive externe. Ceci se remarque notamment dans les zones d'extrapolation (au Nord-Ouest de la carte) où la chute du plateau continental correspond aux zones de présence d'âges élevés. En ce qui concerne les cartes des âges déduites de celles des âges cumulés par soustraction ($A_i = T_i - T_{i+1}$) l'isoligne 0.1 est en trait plein et les isolignes 0.03, 0.05, 0.95, 0.5 sont en traits pointillés. De toutes ces cartes, il ressort le déplacement vers le nord (éloignement des côtes) des individus âgés. On retrouve de plus l'existence d'une zone en doigt de gant correspondant à la localisation d'âges jeunes.

3.4.3 Approches probabilistes: calculs de proportions d'âges et variances globales

On propose ici deux modèles liant la longueur moyenne aux âges cumulés qui exploitent différemment le nuage de corrélation entre la variable T_i et la longueur moyenne L_m . Le premier modèle propose de lire le nuage donnant l'âge cumulé en fonction de la longueur, considérée comme paramètre. Le deuxième propose l'approche inverse.

Approche directe

La régression est de la forme

$$T_i(x) = f_i(L_m(x)) + \epsilon_i(x)$$

où f est une fonction non linéaire de type logistique. La longueur moyenne L_m est considérée comme paramètre. Le nuage donnant $T_i(x)$ en fonction de L_m est lu verticalement. La régression n'explique pas toute la variabilité spatiale de $T_i(x)$. En effet, les résidus de la régression $T_i(x) - f_i(L_m(x))$ ont une distribution qui dépend de L_m : la variance des écarts $s_i^2(L_m(x))$ dépend de la valeur du paramètre. La variance des résidus est faible lorsque $f_i(L_m(x))$ est proche des valeurs 0 et 1. $T_i(x)$ est une proportion, donc comprise entre 0 et 1, on peut s'attendre à ce que la variance soit proportionnelle à $f_i(L_m(x))(1 - f_i(L_m(x)))$ (comme pour une indicatrice ou une variable binomiale). On peut alors supposer que la variable réduite soit stationnaire d'ordre 2 et chercher à déterminer sa structure spatiale par sa covariance stationnaire (ou son variogramme). Le modèle de régression s'écrit donc:

$$T_i(x) = f_i(L_m(x)) + s_i(L_m(x))R_i(x)$$

où $s_i(L_m(x))$ est l'écart-type des résidus et $R_i(x)$ le résidu réduit. La moyenne et la variance du résidu réduit ne dépendent plus alors de $L_m(x)$.

Dans notre étude, plusieurs variables interviennent. On considère, pour simplifier l'étude, trois coupures d'âge $T_{3.1}$, T_5 et T_7 somme des classes d'âge supérieures ou égales à trois ans, cinq ans et sept ans respectivement. Les régressions sont présentées sur les figures 2.33, 2.34 et 2.35. Toutes les années sont présentées. Les régressions sont quasiment déterministes, seule l'année 1993 présente quelques écarts et semble montrer plus de variabilité dans la répartition du stock. La régression correspondant à la campagne 1993 sera choisie pour les calculs d'estimation par la suite. Les variations de proportions d'âges sont reliées linéairement, les unes augmentant au profit des autres. Les coupures T_i sont ainsi reliées elles aussi linéairement $T_i = T_{i+1} + A_i$. Le problème est donc à considérer sous un angle multivariable. Un modèle multivariable ajusté sur les covariances ou variogrammes simples et croisés des différentes variables normées apparaît plus admissible.

La méthode suppose la régression parfaitement connue. Elle suppose aussi la régression permettant le calcul de la variance (puis l'écart-type) parfaitement connue. Cette variance a été estimée à l'aide d'un modèle du type

$$s_i^2(L_m(x)) = k_i f_i(L_m(x))(1 - f_i(L_m(x))).$$

Dans la pratique ces deux régressions sont estimées à partir des mesures expérimentales. L'estimation de la dérive ($f(L_m(x))$) par la méthode des moindres carrés conduit à des résidus dont la variabilité spatiale est sous-estimée [22]. Cet inconvénient est d'autant plus négligeable que la forme choisie pour la régression est simple et que la dispersion des valeurs autour de la régression est faible. Cet effet sera négligé dans la suite de notre application. Une autre source de biais provient de l'estimation de l'écart-type (conditionnel) issu de la racine carrée de l'estimation de la variance. Ici encore la simplicité du modèle utilisé nous incite à négliger ce problème.

Les résultats des régressions pour l'année 1993 sont exposés dans le tableau 3.2.

TAB. 3.2 - Régressions âge/longueur et variances conditionnelles

regression $f(L_m(x))$	variance
$T_3(x) = \frac{e^{0.56L_m(x)-14.79}}{1+e^{0.56L_m(x)-14.79}}$	$s_3^2(L_m(x)) = 0.05f(L_m(x))(1 - f(L_m(x)))$
$T_5(x) = \frac{e^{0.71L_m(x)-20.85}}{1+e^{0.71L_m(x)-20.85}}$	$s_5^2(L_m(x)) = 0.06f(L_m(x))(1 - f(L_m(x)))$
$T_7(x) = \frac{e^{0.52L_m(x)-17.20}}{1+e^{0.52L_m(x)-17.20}}$	$s_7^2(L_m(x)) = 0.02f(L_m(x))(1 - f(L_m(x)))$

Lorsque T_i se rapproche de 0 ou de 1, la variance conditionnelle peut être théoriquement très faible et contribuer à fournir des résidus normés très grands en valeur absolue. Cependant lors de notre application, aucune valeur de résidu normé ne nous est apparue comme nuisible à l'analyse géostatistique pouvant entraîner un manque de robustesse.

L'analyse variographique multivariable des trois résidus réduits $R_{3,1}$, R_5 et R_7 est présentée sur la figure 2.37. Pour augmenter la qualité des variogrammes expérimentaux, ce sont les variogrammes moyens calculés sur 6 années qui ont été calculés. Les structures ne sont pas évidentes. L'ajustement semble très difficile à faire, quelle importance, notamment, donner au premier point? L'ajustement proposé (et réalisé au mieux) tient compte des contraintes du modèle linéaire de corégionalisation. De plus, un modèle de variogramme à une structure de portée 35.6 milles nautiques semble convenir. Les variogrammes simples ont leur palier calé sur 1 et les variogrammes croisés ont leur palier déduit des équations du modèle linéaire de corégionalisation [39]. Aucun effet de pépite n'a été rajouté au modèle et nous avons considéré que l'on pouvait se placer dans un cas où les variables sont en corrélation intrinsèque. Tous les variogrammes sont proportionnels à un modèle sphérique de portée 36 mn et de palier 1 (tableau 3.3).

TAB. 3.3 - Analyse de corégionalisation - Approche directe

Variogrammes	Variogrammes croisés
$\gamma_{R_3-R_3}(h) =$ sphérique(36)	$\gamma_{R_3-R_5}(h) =$ -0.35 sphérique(36)
$\gamma_{R_5-R_5}(h) =$ sphérique(36)	$\gamma_{R_3-R_7}(h) =$ -0.40 sphérique(36)
$\gamma_{R_7-R_7}(h) =$ sphérique(36)	$\gamma_{R_5-R_7}(h) =$ 0.50 sphérique(36)

Les trois variables réduites peuvent être considérées comme (multi-)gaussiennes ce qui permet d'envisager des simulations de ces variables, on cherche alors à reproduire leurs distributions. Les tests graphiques (droite de Henry) visant à vérifier la distribution gaussienne des résidus sont présentés sur la figure 2.36. On a considéré sur ces figures tout l'ensemble des résidus calculés toutes mesures confondues, notre année 1993 constituant un sous échantillon multi-gaussien de l'ensemble des mesures. Par la suite les trois variables réduites sont simulées à l'aide de simulations conditionnelles.

On a, ainsi, réalisé dix simulations conditionnelles de résidus simulés, qui permettent d'obtenir pour chaque coupure, dix cartographies de coupures. Il n'a pas

été nécessaire de retirer de résidu "trop élevé" associé à une variance conditionnelle trop faible. Si cela avait été le cas, ne pas tenir compte de ces résidus lors du krigeage conditionnant n'est pas très grave du fait que la variance de $T_i(x)$ à ces endroits est très faible, $T_i(x)$ y est donc bien connue.

Pour chaque carte, une moyenne globale peut être calculée, la variance des dix moyennes obtenues peut être considérée comme la variance d'estimation globale des moyennes des coupures. Les âges cumulés $T_i^*(x)$ sont estimées en chaque point x par la formule:

$$T_i^*(x) = f_i(L_m(x)) + s_i(L_m(x))(R_{i, \text{simule}}(x))$$

Les résultats de neuf cartographies simulées sont présentées pour chacune des trois coupures sur les figures 2.38, 2.39 et 2.40. Les différences entre simulations d'une même variable sont très faibles. Les simulations conditionnelles ont été réalisées conformément à la méthode dite des "tranches tournantes" à l'aide du logiciel ISATIS conçu au Centre de Géostatistique [17]. On arrive ainsi à estimer quatre groupes de classes d'âge 1+2, 3+4, 5+6 et 7 et plus. Par soustraction, neuf simulations par groupe d'âge sont présentés sur les figures 2.41, 2.42, 2.43 et 2.44.

Les résultats des variances sont présentés dans le tableau 3.4.

TAB. 3.4 - *Estimations globales et variances (approche directe)*

Seuil	moyenne globale	variance globale	écart-type
T_3	0.761	10^{-4}	0.01
T_5	0.484	3×10^{-4}	0.017
T_7	0.204	10^{-6}	0.001
Classes d'âge	moyenne globale	variance globale	écart-type
1 + 2	0.239	10^{-4}	0.01
3 + 4	0.277	2×10^{-7}	0.0004
5 + 6	0.280	9×10^{-6}	0.003
7+	0.204	10^{-6}	0.001

Les résultats des cartes simulées des âges cumulés et de proportions d'âge sont quasiment identiques à ceux des cartes obtenues par l'approche déterministe. On n'a représenté des simulations que sur certaines classes d'âge. Ceci est dû au fait que l'analyse variographique multivariable de n variables conduit à l'estimation (calcul + ajustement) de 2^n variogrammes (simples et croisés). Ce travail est long et difficile du fait des contraintes qu'imposent le modèle de corégionalisation. L'intérêt des simulations est l'obtention de variances sur les estimations globales, celles-ci sont très faibles du fait du lien déterministe très fort avec la longueur.

Approche inverse

Une autre piste est proposée dans ce paragraphe. Cette approche, dite inverse, a l'avantage de ne considérer qu'une estimation, celle de la régression. Elle consiste à considérer le nuage de corrélation $(T_i(x), L_m(x))$ horizontalement. $T_i(x)$ est

considérée comme la transformée, par une fonction g , du paramètre $L_m(x)$ décalé d'une fonction aléatoire $R_i(x)$. Ceci revient à déterminer la régression monotone $f(T_i(x)) (= g^{-1}(T_i(x)))$ de $L_m(x)$ connaissant $T_i(x)$. Le décalage est le résidu de la régression:

$$R_i(x) = L_m(x) - f(T_i(x))$$

On suppose alors $R_i(x)$ stationnaire (ou intrinsèque) et gaussien.

Le modèle de régression estimé est de type logistique:

$$L_m(x) = a_i \text{Ln}\left(\frac{T_i(x)}{1 - T_i(x)}\right) + b_i + R_i(x)$$

Les coefficients a_i et b_i sont estimés à partir des méthodes de régression non linéaires [6]. Les régressions considérées sont présentées sur les graphes des figures 2.45 et 2.46. La variabilité autour de la moyenne semble plus importante dans ce modèle. La fonction g est une fonction strictement monotone tendant asymptotiquement vers 0 et 1. Ces deux valeurs jamais atteintes par le modèle sont en pratique très gênantes (plus gênantes que dans l'approche précédente). Dans notre cas d'étude, les valeurs inférieures à 0.04 et supérieures à 0.95 ont été exclues de l'analyse.

Les tests sur la distribution des résidus calculés pour toutes les campagnes sont présentés sur la figure 2.47. Ils montrent que les résidus des régressions peuvent difficilement être considérés comme multi-gaussiens hors de l'intervalle $[-1,1]$. C'est pourquoi il semblerait préférable de travailler sur des variables transformées gaussiennes de ces résidus.

Malheureusement, notre cas d'étude nous conduit à des résidus manifestement non gaussiens, c'est pourquoi il nous a semblé inutile de continuer dans cette voie. Cette approche semblait pratique par le fait qu'elle ne nécessitait que l'estimation de la régression $f(T_i(x))$. Elle se complique ici par des manipulations statistiques lourdes.

Mener à bout cette approche aurait consisté donc à:

- transformer les résidus $R_i(x)$ en variables gaussiennes $Y_i(x)$,
- effectuer une analyse de corégionalisation des variables $Y_i(x)$,
- simuler des variables $Y_{i_{simule}}(x)$,
- transformer à nouveau les variables gaussiennes en variables "d'origine" $R_{i_{simule}}(x)$,
- estimer les coupures $T_i^*(x)$ en chaque point x par la formule:

$$T_i^*(x) = g_i^{-1}(L_m(x) - R_{i_{simule}}(x)).$$

- estimer les âges $A_i^*(x)$ par soustraction des $T_i^*(x)$,
- estimer des variances globales (moyenner sur les simulations).

Sans être forcément une mauvaise piste, l'approche inverse ne peut pas ici nous donner satisfaction dans la mesure où elle ne représente pas une alternative efficace à l'approche directe. Nous nous contentons donc de la présenter sans la mettre en pratique jusqu'au bout.

Conclusion

Vu l'influence faible que semble représenter le résidu, les estimations contribuant à donner des cartographies proches de celles issues de l'approche déterministe semblent nous inciter à considérer les proportions d'âge comme quasiment déterministes. L'approche probabiliste directe bien que nécessitant l'estimation de deux régressions semble fournir rapidement de bons résultats. C'est une méthode qui, dans cet exemple, n'est pas indispensable au regard de la qualité excellente de la régression longueur/coupure d'âge (du strict point de vue de l'estimation d'une variance globale). Néanmoins, son intérêt est réel et donne une solution au problème du calcul de variance globale de variables non stationnaires liées par des relations non linéaires. Un autre intérêt de cette méthode probabiliste réside dans le fait qu'elle conduit grâce à l'utilisation des simulations conditionnelles à des cartographies des âges cumulés et des âges respectant les mesures ce que ne reproduit pas l'approche probabiliste. Les cartes obtenues par la méthodes probabilistes sont donc plus précise du point de vue du conditionnement.

Les problèmes de l'approche inverse sont liés à l'estimation de la régression qui doit négliger l'influence de quelques points incompatibles avec la transformation logarithmique ainsi qu'à la distribution des résidus qui, ici, est incompatible avec une hypothèse gaussienne.

Troisième partie

**Combinaison de mesures
acoustiques et des mesures
obtenues par chalutage : calculs
d'abondance**

Chapitre 1

Estimation d'une abondance globale

La dernière partie de ce travail a pour objectif de combiner les deux types de données présentées et analysées dans les deux parties précédentes afin de réaliser des calculs d'abondance globale et des variances. Le problème, on le rappelle, est de réaliser des calculs d'abondance à partir de deux types de données:

- un indice d'abondance obtenu par mesures acoustiques en continu le long de trajets parallèles et fourni sur une grille quasi-régulière $S(x)$,
- la surface réfléchissante $\sigma(x)$ elle-même calculée comme une fonction carrée de la longueur: $\sigma(x) = k.l(x)^2$.

On peut résumer le problème ainsi: comment réaliser des calculs d'abondance globale connaissant la donnée des deux grandeurs intervenant dans le calcul d'une densité de poissons à partir de mesures acoustiques. La densité réelle de poissons (individus au mille nautique carré) peut être symbolisée par l'expression:

$$\rho(x) = \frac{S(x)}{\sigma(x)} = \frac{S(x)}{k.l^2(x)}$$

où S est mesurée en mn^2 et l en cm . On abordera aussi le calcul des biomasses.

Une méthode originale nous servira pour mener à bout les calculs d'abondance. Comme dans les deux parties précédentes une campagne d'échantillonnage réelle servira d'illustration à ces calculs.

1.1 Calculs classiques

1.1.1 Calculs d'abondance

L'échantillonnage de la campagne 1993 est pris en exemple sur la figure 3.1, l'échantillonnage acoustique est représenté par des points, celui effectué par mesures pêchées est représenté par des triangles. Les calculs réalisés par les biologistes des pêches [33] s'opèrent conformément au découpage de la zone échantillonnée en rectangles statistiques de 15 minutes de latitude par 30 minutes de longitude: l'échantillonnage acoustique est effectué le long de transects parallèles

orientés Est-Ouest espacés d'environ 15 milles nautiques dans les zones "pauvres" et 30 milles nautiques dans les zones riches. Chaque rectangle statistique échantillonné est, ainsi, traversé une fois dans les zones pauvres (par sa médiane) et deux fois dans les zones riches. Chaque rectangle statistique appartient à une strate k correspondant à une zone d'homogénéité spatiale pour la longueur du poisson, définie après l'analyse des échantillons pêchés.

Du point de vue de l'estimation, chaque rectangle statistique j correspond donc à:

- Une valeur de l'indice acoustique moyenné sur la médiane (ou sur les deux segments le traversant) de ce carré S_j
- Une valeur moyenne de la variable σ , σ_j , correspondant à la strate k à laquelle appartient le rectangle statistique. Au sein d'une même strate tous les σ_j sont égaux..

La densité ρ_j correspondant à chaque rectangle statistique est donnée par l'équation:

$$\rho_j = \frac{S_j}{\sigma_j}$$

L'abondance sur ce rectangle de surface V_j vaut donc

$$Q_j = \frac{S_j}{\sigma_j} V_j$$

On en déduit l'abondance totale:

$$Q = \sum_j Q_j = \sum_j \frac{S_j}{\sigma_j} V_j$$

1.1.2 Calculs d'abondances par groupes d'âge

La proportion de chaque groupe d'âge est accessible directement à partir des résultats de chaque coup de chalut. Au point x de l'espace à deux dimensions, la densité du groupe d'âge i peut être définie par l'expression:

$$\rho_i(x) = \frac{S(x)}{\sigma(x)} p_i(x)$$

où $p_i(x)$ est la proportion de l'âge i au point x . Comme pour les longueurs, une valeur p_i moyenne est calculée pour chacune des strates k et une valeur p_i^j est affectée à chaque strate puis à chaque rectangle statistique j . L'abondance globale du groupe d'âge i est donc:

$$Q_i = \sum_j Q_i^j = \sum_j \frac{S_j}{\sigma_j} p_i^j V_j$$

1.1.3 Calcul des biomasses

Une autre clé longueur/poids:

$$W = cL^\alpha \text{ (g)}$$

donne par strate un poids moyen calculé à partir d'une longueur moyenne de poisson.

Par strate un poids moyen est calculé à partir de la valeur moyenne des longueurs au moyen d'une relation taille/poids. Chaque rectangle est affecté d'une valeur moyenne W_m^j (poids moyen) puis d'une biomasse W_j , et un poids total sur tout le domaine est calculé comme:

$$W = \sum_j W_j = \sum_j \frac{S_j}{\sigma_j} W_m^j$$

1.1.4 Calculs des variances d'abondance globales

Lorsque le domaine échantillonné est analysé par strates, les calculs exposés plus haut sont applicables au sein de chaque strate. Les calculs présentés ici ont été présentés par Cochran, 1977 (in Anon. 1992 [1]). On présente les résultats des calculs de variances globales dans le cas d'un échantillonnage aléatoire stratifié: on note k l'indice des strates. Les corrélations spatiales ne sont, ici, pas prises en compte.

La densité moyenne de la strate k est estimée par sa moyenne arithmétique:

$$\bar{\rho}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \rho(x_i)$$

où $\rho(x_i)$ est la densité mesurée par échantillonnage acoustique et n_k est le nombre d'échantillons dans la strate k .

L'abondance dans la strate k vaut

$$Q_k = \bar{\rho}_k V_k$$

Au sein de chaque strate k , ($k=1,2,\dots,K$), de surface V_k , on peut définir une variance d'estimation globale de la moyenne estimée sur la strate: dans chaque strate, la variance d'estimation σ_k^2 de la moyenne est estimée à l'aide de la variance de dispersion de la densité:

$$s_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (\rho(x_i) - \bar{\rho}_k)^2$$

ainsi:

$$\sigma_k^2 = \frac{s_k^2}{n_k}$$

La surface V_k étant connue la variance d'estimation de l'abondance totale sur la strate vaut:

$$var(Q_k) = V_k^2 \cdot s_k^2$$

La variance de la densité moyenne vaut:

$$\text{var}(\bar{\rho}_V) = \sum_k \left(\frac{V_k}{V}\right)^2 \frac{s_k^2}{n_k}$$

L'abondance totale vaut:

$$Q_V = \sum_k Q_k$$

et la variance d'estimation de l'abondance:

$$\text{var}(Q_V) = V^2 \text{var}(\bar{\rho}_V) = \sum_k V_k^2 \frac{s_k^2}{n_k}$$

1.2 Estimation globale et prise en compte de la corrélation spatiale

1.2.1 Estimation globale d'un quotient

Notre problème dans ce travail consiste à calculer une variance d'estimation globale d'une moyenne que l'on peut exprimer comme le quotient de deux grandeurs physiques mesurées séparément. Nous cherchons à coupler nos deux échantillonnages. Les équations ci-dessous sont issues de l'estimation globale d'un quotient présentées par Journal (1977) [18]. Nous cherchons à estimer:

$$\rho_V = \frac{S_V}{\sigma_V}$$

par

$$\rho_V^* = \frac{S_V^*}{\sigma_V^*}$$

En interprétant les estimateurs S_V^* et σ_V^* comme des variables aléatoires: $S_V = S_V^* + \epsilon_{S_V}$ et $\sigma_V = \sigma_V^* + \epsilon_{\sigma_V}$ avec $E(\epsilon_{S_V}) = E(\epsilon_{\sigma_V}) = 0$, l'estimateur de la moyenne globale peut s'écrire:

$$\rho_V^* = \frac{S_V^*}{\sigma_V^*} = \frac{S_V + \epsilon_{S_V}}{\sigma_V + \epsilon_{\sigma_V}} = \frac{S_V + \epsilon_{S_V}}{\sigma_V} \cdot \frac{1}{1 + \epsilon_{\sigma_V}/\sigma_V}$$

Si l'erreur relative $\frac{\epsilon_{\sigma_V}}{\sigma_V}$ est petite vis à vis de 1, on a le développement limité:

$$\rho_V^* = \left(\rho_V + \frac{\epsilon_{S_V}}{\sigma_V}\right) \cdot \left(1 - \frac{\epsilon_{\sigma_V}}{\sigma_V} + \frac{\epsilon_{\sigma_V}^2}{\sigma_V^2} - \frac{\epsilon_{\sigma_V}^3}{\sigma_V^3}\right)$$

d'où:

$$E(\rho_V^*) = \rho_V \left(1 + \frac{\sigma_{\sigma_V}^2}{\sigma_V^2} - \frac{E(\epsilon_{\sigma_V} \epsilon_{S_V})}{S_V \sigma_V} + \dots\right)$$

Il y a donc un biais relatif même si les 2 estimations sont indépendantes ($E(\epsilon_{\sigma_V} \epsilon_{S_V}) = E(\epsilon_{\sigma_V})E(\epsilon_{S_V}) = 0$) qu'à priori on négligera.

La variance relative d'estimation peut s'écrire:

$$\frac{\sigma_{\rho_V}^2}{\rho_V^2} = \frac{\sigma_{S_V}^2}{S_V^2} + \frac{\sigma_{\sigma_V}^2}{\sigma_V^2} - 2 \frac{E(S_V - S_V^*)(\sigma_V - \sigma_V^*)}{S_V \sigma_V} + \dots$$

Dans le cas où les 2 estimations peuvent se faire indépendamment l'une de l'autre, on peut utiliser la formule :

$$\frac{\sigma_{\rho_V}^2}{\rho_V^2} = \frac{\sigma_{S_V}^2}{S_V^2} + \frac{\sigma_{\sigma_V}^2}{\sigma_V^2}$$

Cependant ce calcul pose un problème. À supposer que les deux fonctions aléatoires $S(x)$ et $\sigma(x) = k.l^2(x)$ soient non corrélées, rien n'indique que S_V et σ_V le soient également. En effet, σ_V n'est pas une régularisée de $\sigma(x)$, mais dépend de $S(x)$:

$$\rho_V = \frac{1}{V} \int_V \rho(x) dx = \frac{1}{V} \int_V \frac{S(x)}{k.l^2(x)} dx$$

et

$$S_V = \frac{1}{V} \int_V S(x) dx$$

par ailleurs:

$$\rho_V = \frac{S_V}{\sigma_V} = \frac{S_V}{k.l_V^2}$$

d'où:

$$\frac{1}{\sigma_V} = \frac{1}{k.l_V^2} = \frac{\rho_V}{S_V} = \frac{1}{V} \int_V \frac{S(x)}{S_V} \frac{1}{k.l^2(x)} dx = \frac{1}{V} \int_V \frac{S(x)}{S_V} \frac{1}{\sigma(x)} dx$$

Le calcul de la variance ne serait praticable, dans cette approche qu'en supposant $\sigma_V = k.l_V^2$ indépendant de $S(x)$ (par exemple $l_V^2 \simeq \frac{1}{V} \int_V l^2(x) dx$).

Par conséquent l'estimation de ρ_V pose un problème, il nous faut envisager une autre méthode.

1.2.2 Couplage de données acoustiques et de mesures calculées

Estimation d'une densité et d'une moyenne globale

Considérant que la longueur est un paramètre très stable dans le temps, on la considèrera comme une variable déterministe. Par conséquent, nous négligerons le biais induit lors de l'estimation du facteur $\sigma(x) = k.l^2(x)$. La carte des longueurs moyennes peut ainsi être utilisée pour transformer chaque mesure acoustique $s(x)$ en une densité de poisson $\rho(x)$. Une valeur de la longueur puis de $\sigma(x)$ peut alors être estimée en chaque point de l'échantillonnage acoustique. Ainsi, la densité moyenne sur le domaine V peut être estimée de la même manière que l'indice acoustique moyen (voir partie I). Comme $s(x)$ est bien plus fluctuant que la longueur, la transformation ne change pas beaucoup les allures des variogrammes et un calcul de variance globale peut être effectué de manière similaire à celui de la partie I.

L'abondance, en nombre de poissons, est estimée en multipliant la densité moyenne estimée par la surface du champ. Le calcul de la variance globale doit

tenir compte éventuellement d'une pondération par le carré des surfaces (stratification).

Estimation de l'abondance par âge

Comme il a été présenté dans la partie II, des relations déterministes très fortes peuvent être utilisées pour convertir la longueur moyenne en proportions d'âge $p_i(x)$. Ainsi, les densités de poissons aux âges i en chaque point de l'échantillonnage acoustique peuvent être estimés par:

$$\rho_i(x) = \rho(x) \times p_i(x)$$

Des densités moyennes par âge sur tout le domaine peuvent alors être calculées comme précédemment. En les multipliant par la surface on obtient des abondances en nombre de poissons.

Chapitre 2

Application au hareng écossais: campagne 1993

2.1 Méthodologie

La méthodologie du calcul d'une abondance globale et de sa variance d'estimation a été exposée au paragraphe précédent. On rappelle néanmoins les grandes étapes de ce calcul.

La longueur étant considérée comme un paramètre stable, elle a été cartographiée à l'aide de la technique du krigeage. La bathymétrie étant considérée comme connue partout sur le champ et conditionnant la répartition des longueur, des contraintes sont ajoutées au système de krigeage de la longueur. Cette technique dite du krigeage avec dérive externe permet d'obtenir une carte de la distribution des longueurs. Un modèle de covariance particulier a permis d'utiliser toutes les données collectées sur six campagnes annuelles et de réaliser six cartes ne respectant que les données de l'année. Ceci a été effectué dans la partie II. La même méthode de krigeage peut être utilisée afin d'effectuer en chaque point de l'échantillonnage acoustique une estimation de la longueur moyenne ($L_m^*(x)$). Une densité de poissons en terme d'individus par mille nautiques carré peut alors être estimée en utilisant la transformation suivante:

$$\rho^*(x) = \frac{s(x)}{9.5 * 10^{-7} (L_m^*(x)^2)}$$

Le facteur $9.5 * 10^{-7}$ étant caractéristique de l'espèce de hareng considérée ici (*Clupea harengus*).

On rappelle (voir partie II, §1.1.3) que dans les calculs combinant les mesures acoustiques et les mesures chalutées, la longueur intervient sous la forme d'une moyenne des carrés des longueurs mesurées. La faible dispersion des longueurs mesurées au niveau de chaque chalutage nous permet d'une part, de résumer la distribution des longueurs par la moyenne des longueurs et d'autre part, de considérer le carré de la moyenne comme une bonne approximation de la moyenne des carrés.

Le lien étroit entre la longueur moyenne et les coupures d'âges $T_i(x)$ (proportion d'âge supérieur à l'âge i) caractérisé par des fonctions logistiques (voir partie II) permet d'obtenir pour la campagne considérée des estimations de ces coupures

aux points de l'échantillonnage acoustique:

$$T_i^*(x) = f_i(L_m^*(x))$$

On obtient donc des densités estimées correspondant aux coupures d'âge:

$$\rho_{age \geq i}^*(x) = \frac{s(x)}{9.5 * 10^{-7} (L_m^*(x))^2} \cdot T_i^*(x)$$

Des proportions d'âge $p_i^*(x)$ sont calculées à partir des coupures d'âge par soustraction:

$$p_i^*(x) = T_i^*(x) - T_{i+1}^*(x)$$

Des densités estimées par classes d'âge sont alors calculées:

$$\rho_i^*(x) = \frac{s(x)}{9.5 * 10^{-7} (L_m^*(x))^2} \cdot p_i^*(x)$$

2.2 Applications numériques

Le domaine échantillonné a, comme dans la partie I, été stratifié en quatre zones (fig. 3.1):

- une partie Est correspondant à un échantillonnage régulier comportant 13 transects espacés de 15 milles nautiques.
- une partie Ouest que l'on a subdivisée en 3 sous-zones:
 - la sous-zone Sud (7 transects espacés de 15 mn)
 - la sous-zone Centre (4 transects espacés de 7.5 mn)
 - la sous-zone Nord (4 transects espacés de 15 mn)

Les calculs d'abondances globales sont calculés séparément par strates puis additionnés. Les calculs de variance doivent tenir compte d'une pondération par les surfaces. La mesure acoustique fournie étant l'indice acoustique divisé par un facteur 4π , la densité estimée a donc été multipliée par ce facteur pour obtenir la vraie densité estimée en chaque point de l'échantillonnage acoustique.

2.2.1 Abondance globale

Pour chaque strate k l'abondance est donnée par: $Q_k^* = \bar{\rho}_k \times V_k$ où $\bar{\rho}_k$ est la moyenne arithmétique de la densité de poisson sur la surface V_k de la strate. Des calculs statistiques élémentaires sont présentés sur les figures 3.2 et 3.3. On obtient ainsi les résultats:

- Partie Est: $133\ 115 \times (2.3 \times 10^4) = 3.1 \times 10^9$ individus
- Partie Ouest:

Sous-zone	moyenne ind./mn ⁻²	Surface mn ²	nombre d'individus
Nord	64239	1457	9.4x10 ⁷
Centre	130979	2192	2.9x10 ⁸
Sud	154579	4448	6.9x10 ⁸

Total Partie Ouest :

Densité moyenne = 133581 ind./nm²

Surface = 8000 mn²

Abondance = 1.1x10⁹ individus

- Total: 4.2x10⁹ individus

2.2.2 Variance d'estimation globale

Les figures 3.2 et 3.3 présentent la répartition de la densité de poisson. Ces cartes ressemblent beaucoup à celles de l'indice acoustique présentées dans la partie I. Les mêmes problèmes distributionnels se posent à nouveau pour la densité et poseront aussi les mêmes difficultés lors de l'analyse structurale. Du fait des variations faibles de la longueur moyenne comparées à celles de l'indice acoustique, les variogrammes expérimentaux sont similaires à ceux de l'indice acoustique à un facteur près. Comme précédemment, nous avons gardé les variogrammes calculés le long de la route du navire. Le calcul de la variance utilise la même méthode d'approximation que dans la partie I (composition des variance). Ce calcul se fait en deux étapes:

- un calcul de variance d'estimation de la moyenne globale par strate (σ_k^2)
- un calcul de variance d'estimation globale de l'abondance tenant compte de pondération par les surfaces:

$$\sigma_{E,Q}^2 = V^2 \sum_k \left(\frac{V_k}{V}\right)^2 \sigma_k^2$$

Les résultats du calcul des variances pour la partie Ouest ont été regroupés et présentés dans le tableau ci-dessous:

	Partie Ouest	Partie Est
pépite	1.9x10 ⁸	3.1x10 ⁸
terme de tranche	1.6x10 ⁸	9.9x10 ⁷
σ_E^2	2.5x10 ¹⁶	2.2x10 ¹⁷
CV	14.4%	15.1%
CV classique	20.1%	15.2%

Total :

$$\sigma_{E,Q}^2 = 2.5 \times 10^{17}$$

$$CV = 12\%$$

Remarque: À titre de comparaison le coefficient de variation calculé sans tenir compte des corrélations spatiales est calculé (CV classique). Ne pas tenir compte des corrélations spatiales incite, ici, à surestimer la variance.

2.2.3 Calculs de biomasses

À partir de la carte des longueurs que l'on considère comme déterministe, un poids moyen peut être estimé à l'aide de la relation longueur/poids fournie par les biologistes [33]:

$$W_m(x) = 1.08 \times 10^{-3} \times L_m(x)^{3.62}$$

où W_m est le poids en grammes par individu. Un poids moyen peut alors être calculé par strate ainsi que des biomasses globales comme produit de l'abondance par le poids moyen:

On obtient ainsi les résultats:

- Partie Est: Poids moyen = 223.90 g/ind. soit une biomasse de:
 $3.1 \times 10^9 \times 223.90 = 6.94 \times 10^{11} \text{ g} = 694\,090 \text{ tonnes}$.
- Partie Ouest:

Sous-zone	poids moyen g/ind.	nombre d'individus	biomasse (tonnes)
Nord	269.05	9.4×10^7	2.5×10^4
Centre	471.42	2.9×10^8	1.4×10^5
Sud	229.59	6.9×10^8	1.6×10^5

Total Partie Ouest:

Biomasse = 325 000 tonnes

- Total: 1 019 090 tonnes (1.02×10^6 tonnes)

2.2.4 Abondances par classes d'âge

Les proportions des classes d'âge p_i^* ayant été estimées à partir de l'estimation des coupures d'âge, des densités moyennes par strates peuvent être calculées. Des abondances par strate puis pour l'ensemble des strates ont ainsi été calculées. Les abondances par classes d'âge ont ainsi été calculées pour la campagne 1993:

Age	Age 1	Age 2 immatures	Age 2 matures	Age 3 immatures	Age 3 matures
individus	4.0×10^8	1.9×10^8	6.9×10^8	1.8×10^8	5.4×10^8

Age	Age 4	Age 5	Age 6	Age 7	Age 8	Age 9 plus plus	total
individus	4.0×10^8	4.5×10^8	7.3×10^8	4.4×10^8	1.4×10^8	1.1×10^8	4.2×10^9

2.2.5 Discussion

L'estimation de l'abondance, présentée ici, pour l'année 1993 a l'avantage de prendre en compte la structure spatiale des données. Il est important de noter que cette estimation résulte d'un processus complexe qui tient compte des résultats de campagne réalisées d'autres années: (i) les hypothèses d'isotropie et de stationnarité de l'indice acoustique découlent de l'analyse de toutes les campagnes;

(ii) la carte des longueurs a été réalisée en utilisant la donnée de la longueur moyenne sur les six campagnes permettant d'admettre la stabilité temporelle de la longueur et de négliger sa contribution à la variance. La carte de la distribution des longueurs aurait été beaucoup moins précise si réalisée avec les seuls 32 échantillons de la campagne 1993.

Aucune variabilité provenant de la distribution des longueurs n'a été considérée ici. Seule celle issue de l'indice acoustique a été prise en compte dans le calcul de variances d'abondances globales. On a estimé, ici, que la contribution due à la longueur dans le calcul des variances était très faible et négligeable. Il est bien évident qu'une solution générale à notre problème d'estimation de variance doit tenir compte de la variabilité de la longueur.

En ce qui concerne les abondances par classes d'âge, le problème est encore plus délicat. Les distributions des longueurs et des proportions d'âge sont supposées bien connues. La transformation utilisée pour l'estimation de densités par âge conduit à des variables non stationnaires difficiles à aborder dans le cadre de calculs de variances d'estimation. Ce travail constituerait à lui seul un thème de recherche conséquent que nous n'avons pu traiter pendant la durée de ce travail. Les éventuelles déviations dues à l'estimation du lien longueur/âge peuvent aussi constituer une autre source d'incertitude.

Chapitre 3

Problèmes ouverts

3.1 Variabilité due à la longueur

Nous n'avons pu résoudre notre problème d'estimation globale qu'à la condition de supposer la longueur moyenne du poisson comme étant une variable déterministe. Cette hypothèse découle de l'observation des mesures expérimentales, des variogrammes et des cartes de krigeage:

- les cartes des mesures expérimentales nous indiquent la faible variation de la distribution spatiale de la longueur moyenne par rapport à celles des distributions des classes d'âge.
- les variogrammes de la longueur nous permettent de distinguer une stabilité dans la distribution spatiale répétée d'année en année ainsi qu'une variabilité inter-annuelle qui nous permet de réaliser des cartes annuelles différentes.
- le krigeage de la longueur moyenne (ici avec dérive externe) nous permet de cartographier la longueur pour chacune des campagnes et d'accepter l'hypothèse de la stabilité des longueurs.

Il est néanmoins manifeste que la variabilité de la longueur existe et doit être prise en compte dans les calculs de variances globales. La densité est une fonction de l'inverse de la longueur moyenne au carré. Il faut donc envisager le calcul d'une variance globale d'une variable en $\frac{1}{L_m^2(x)}$.

Nous nous sommes donc orientés en ce qui concerne, le calcul d'abondance globale, vers l'estimation globale de la variable $\rho(x) = \frac{s(x)}{\sigma(x)}$ sur un champ V à savoir l'estimation de $\frac{1}{V} \int_V \frac{s(x)}{\sigma(x)} dx$ ce qui est, en général, différent de l'estimation de $\frac{\int_V s(x) dx}{\int_V \sigma(x) dx}$. Si l'incertitude sur l'estimation des longueurs est forte, elle peut conduire à des biais importants sur l'estimation de $\rho(x)$. En effet, on a à estimer la densité $\rho^*(x)$ par:

$$\rho^*(x) = \frac{s(x)}{\sigma(x)^K}$$

où $\sigma(x)^K$ est l'estimateur de $\sigma(x)$ par krigeage. On a donc, en notant $\epsilon_K(x)$ l'erreur de krigeage sur $\sigma(x)$:

$$\rho^*(x) = \frac{s(x)}{\sigma^K(x)} \cdot \frac{1}{1 + \frac{\epsilon_K(x)}{\sigma(x)^K}} \simeq \rho^*(x) \left(1 - \frac{\epsilon_K(x)}{\sigma(x)^K} + \frac{\epsilon_K^2(x)}{(\sigma(x)^K)^2}\right)$$

d'où en espérance:

$$E(\rho^*(x)) = \rho^*(x) \left(1 + \frac{\text{var}^K(\sigma(x))}{(\sigma(x)^K)^2}\right)$$

où le terme $\frac{\text{var}^K(\sigma(x))}{(\sigma(x)^K)^2}$ mettant en jeu la variance relative d'estimation de $\sigma(x)$ est responsable du biais.

L'estimateur

$$\widehat{\rho(x)} = \rho^*(x) \left(1 - \frac{\text{var}^K(\sigma(x))}{(\sigma(x)^K)^2}\right)$$

pourrait tenir lieu éventuellement d'estimateur sans biais de $\rho(x)$.

La carte des écarts-types d'estimation nous indique des écarts-types compris majoritairement entre 2 et 3. Globalement sur l'ensemble du champ l'écart-type relatif est de l'ordre de 10% ce qui conduit, en ce qui concerne la longueur à une variance relative de l'ordre de 1 pour cent. L'impact de la longueur sur l'estimation globale semble donc assez faible. À l'opposé, la variabilité de la longueur peut avoir un impact important sur les estimations des âges. En effet, un décalage horizontal des courbes longueur/âges cumulés change la relation entre l'âge et la longueur. Ce décalage s'équilibre-t-il en moyenne sur le champ des mesures ou non? Rien ne permet de l'affirmer.

Quel est alors la conséquence de la variabilité de la longueur sur l'estimation des âges? La réponse est loin d'être immédiate et nécessite probablement un travail long et très approfondi.

3.2 Variabilité dans le champ des paramètres biologiques

Cette variabilité peut être examinée grâce à des simulations conditionnelles. Des simulations conditionnelles de la longueur donnent, en effet, la variabilité dans le champ de la longueur ou de toute fonction déterministe de la longueur, par exemple le poids moyen:

$$W_m(x) = 1.08 \times 10^{-3} L_m^{3.62}(x).$$

En enchaînant les simulations conditionnelles, on peut utiliser une simulation conditionnelle de la longueur comme point de départ pour une simulation d'une fonction non déterministe de cette longueur, comme les proportions d'âges cumulées. Par de tels procédés, on obtient la variabilité de la longueur, du poids moyen, des classes d'âges cumulées et simples à travers le champ.

Mais les variances que l'on peut obtenir par la répétition de simulations ne tiennent pas compte des variations de l'indice acoustique, donc ne correspondent pas à la population.

3.3 Variabilité des abondances

Les abondances (en nombre total, en nombre par classe d'âge ou en biomasse) sont obtenues en pondérant l'indice acoustique $S(x)$ par les paramètres biologiques précédents.

Ainsi la répétition de simulations conditionnelles de ces paramètres aux points échantillonnés par acoustique permet d'étudier l'influence de la variabilité de ces paramètres sur les calculs d'abondances. On peut donc obtenir la variance des abondances provenant de la variabilité des paramètres biologiques (qui ne prend pas en compte la variabilité de l'acoustique). C'est une variance d'une autre nature que la variance des abondances provenant de la variabilité de l'acoustique (ou de $\rho(x)$ et calculée à l'aide du variogramme) laquelle ne tient pas compte de la variabilité des paramètres biologiques.

La combinaison des deux variances semble actuellement hors de portée, à moins de savoir simuler conditionnellement également la variable acoustique.

Conclusion

L'objectif de ce travail était de proposer une méthodologie permettant de réaliser des estimations globales de stocks halieutiques et des variances dans le cadre de campagnes acoustiques. L'approche est celle des méthodes géostatistiques. Comme le souligne P. Petitgas [29], le "maître mot est l'analyse structurale". De part la nature des données, nous n'avons pas échappé aux difficultés qu'elle pose. Notre apport a donc consisté à proposer des moyens de rendre l'analyse structurale possible et d'avancer dans la connaissance du variogramme et donc d'une variance d'estimation. Ce qui nous a guidé, dans la première partie de ce travail est la recherche de la robustesse du variogramme. Les techniques proposées permettent l'inférence d'un variogramme nécessaire au calcul d'une variance globale. Il est à noter qu'aucune de ces méthodes ne s'impose de manière universelle. Chaque jeu de données présente des particularités auxquelles le praticien doit s'adapter. En définitive, le choix d'une technique d'estimation et d'un modèle repose sur des hypothèses propres à la distribution expérimentale des mesures pour une campagne donnée en un lieu géographique donné. Nous insistons sur le caractère local des résultats obtenus. Par contre, l'éventail des techniques proposés fait partie de la richesse que constitue l'apport de la géostatistique à la biologie des pêches.

La deuxième partie de ce travail nous a amené à réfléchir sur la manière de rendre compte de la variabilité de certains phénomènes biologiques: la distribution des longueurs et des classes d'âge au sein d'un stock halieutique. Des techniques simples d'investigations ont été utilisées dans le but d'identifier ce qui parmi nos variables peut être considéré comme stable ou très fluctuant dans l'espace et aussi dans le temps. On a ainsi pu établir une hiérarchie dans la dépendance de nos variables explicatives: la bathymétrie "dirige" la distribution des longueurs (en moyenne) du poisson qui elle même "dirige" celle des classes d'âge. Ce phénomène est analysé et utilisé d'un point d'un vue géostatistique. Il est lui aussi lié au caractère local de l'étude qui a orienté nos choix méthodologiques. Il est assez rare en géostatistique de proposer des estimateurs dans un cadre multivariable à l'aide de relations non linéaires. Ici, la variabilité biologique a pour conséquence la non stationnarité des fonctions aléatoires. Celle-ci s'explique par la non linéarité des liens entre variables. La non-stationnarité est, ainsi, héritée de la non linéarité des liens entre variables. L'utilisation de liens fonctionnels non linéaires nous a permis de nous ramener à des cas simples et de résoudre des problèmes d'estimation et de cartographie.

La troisième partie conclut le problème de l'estimation globale à l'aide d'une approximation importante mais néanmoins réaliste: la distribution des longueurs est un phénomène stable qui permet de donner à la longueur le statut de paramètre déterministe. L'estimation des variances des proportions d'âge et des biomasses restent cependant des problèmes ouverts.

Bibliographie

- [1] Anon. - Acoustic survey design and analysis procedure: A comprehensive review of current practice. ICES cooperative research report, No.187, 1992.
- [2] Anon. - Report of the workshop on the applicability of spatial techniques to acoustic survey data. ICES cooperative research report, No.195, 1993.
- [3] Anon. - Introduction à l'utilisation des systèmes sonar dans l'estimation de la biomasse en poissons. FAO, Document technique sur les pêches, No.191.
- [4] Armstrong A. - Improving the estimation and modelling of the variogram. In Verly et al. (eds.) *Geostatistics for Natural Resources Characterization*, Part 1, 1-19. Reidel, Dordrecht, 1984.
- [5] Capéraà P., Van Cutsem B. - *Méthodes et modèles en statistique non paramétrique, exposé fondamental*. Dunod, 1988.
- [6] Chambers J.M. and Hastie T.J. - *Statistical models in S*. Chapman and Hall, London, 1990.
- [7] Chauvet P. - Aide mémoire de géostatistique linéaire. *Cahiers de géostatistique*, Fasc.2, École de Mines de Paris, 1994. 210 p.
- [8] Chilès J.P. - Géostatistique des phénomènes non stationnaires (dans le plan). Thèse de docteur-Ingénieur, Université de Nancy, 1977.
- [9] Collignon J. - *Ecologie et biologie marines - Introduction à l'halieutique*. Masson, Paris, 1991.
- [10] Cressie N. - *Statistics for spatial data*. Wiley, 1991.
- [11] Dowd P.A. - The variogram and kriging: robust and resistant estimators. In Verly et al. (eds.) *Geostatistics for Natural Resources Characterization*, Part 1, 91-106. Reidel, Dordrecht, 1984.
- [12] Galli A. and Meunier G. - Study of a gas reservoir using the external drift method. In G. Matheron and M. Armstrong (eds.) *Geostatistical Case Studies*, 105-119. Reidel, Dordrecht, 1987.
- [13] Guiblin P., Rivoirard J., Simmonds E.J. - Analyse structurale de données à distribution dissymétrique, exemple du hareng écossais. *Cahiers de Géostatistique*, Fasc.5, ENSMP, Paris, 1995., p.137-159.

- [14] Guiblin P., Rivoirard J., Simmonds E.J. - Abundance and variance of Orkney-Shetland herring using geostatistics. ICES/CIEM, CM 1996/D:15.
- [15] Guiblin P., Rivoirard J., Simmonds E.J. - Spatial distribution of length and age for Orkney-Shetland herring. ICES/CIEM, CM 1996/D:14.
- [16] Grosbras J.M. - *Méthodes statistiques des sondages*. Economica, Paris, 1987.
- [17] Groupe Informatique. ISATIS the geostatistical key: ISATIS tutorial. Centre de Géostatistique, École des Mines de Paris, 1994.
- [18] Journel A.G., Huijbregts Ch.J. - *Mining geostatistics*. Academic Press, London, 1978.
- [19] Kirkegaard K., Aglen A., Degnbol P., Hagström O., Palmén L. and Simmonds E.J. - Report on the herring acoustic surveys in the North Sea, Skagerrak and Kattegat during summer 1989. ICES/CIEM, CM 1990/H:24.
- [20] MacLennan D.N., and MacKenzie I.G. - Precision of acoustic Fish stock estimates. 1988. *Can. J. Fish. Aquat. Sci.* 45:605-616.
- [21] MacLennan D.N., and Simmonds E.J. - *Fisheries Acoustics*. Chapman Hall, London, 1991.
- [22] Matheron G. - La théorie des variables régionalisées et ses applications. *Les Cahiers du Centre de Morphologie Mathématique*, fasc. 5, École de Mines de Paris, 1970.
- [23] Matheron G. - The intrinsic random functions and their applications. *Adv. Appl. Prob.*, 5, pp439-468, 1973.
- [24] Matheron G. - Effet proportionnel et lognormalité, ou le retour du serpent de mer. Note N-374, Centre de Géostatistique, ENSMP, Fontainebleau, 1974.
- [25] Matheron G. - Estimer et Choisir - Essai sur la pratique des probabilités. *Les Cahiers du Centre de Morphologie Mathématique*, fasc. 7, École de Mines de Paris, 1978.
- [26] Omre H., - The variogram and its estimation. In Verly et al. (eds.) *Geostatistics for Natural Resources Characterization*, Part 1, 107-125. Reidel, Dordrecht, 1984.
- [27] Pelletier, D., Parma, A. M. and Sullivan, P. J., 1993.- Combining different sources of information in estimating abundance maps for exploited fish populations. In *Proceedings of the American Fisheries Society Symposium, Portland, Ore., August 1993*. American Fisheries Society, Washington, D.C.
- [28] Pelletier, D., and A.M. Parma. - Spatial distribution of Pacific halibut (*Hippoglossus stenolepis*): An application of geostatistics to longline survey data. 1994. *Can. J. Fish. Aquat. Sci.* 51:1506-1518.

- [29] Petitgas P. - Contributions géostatistiques à la biologie des pêches maritimes. Thèse de doctorat en géostatistique, École des Mines de Paris, 1991.
- [30] Rivoirard J. - *Introduction to disjunctive kriging and non-linear geostatistics*. Clarendon Press, Oxford, 1994.
- [31] Rivoirard J. - Computing variograms on uranium data. In G. Matheron and M. Armstrong (eds.) *Geostatistical Case Studies*, 1-22. Reidel, Dordrecht, 1987.
- [32] Sherrer B. - *Biostatistique*. Gaëtan Morin, 1984.
- [33] Simmonds E.J., Fryer R.J. - Survey strategies for structured populations PT II: Precision of variance estimators. ICES/CIEM CM 1993/D:31.
- [34] Simmonds E.J., Dommasnes A., Aglen A., Corten A., and Reid D. - 1992 ICES coordinated acoustic survey of ICES divisions IVa, IVb and VIa. ICES/CIEM, CM 1993/H:11.
- [35] Simmonds E.J., Dommasnes A., Corten A., Pedersen J., and Reid D. - 1993 ICES coordinated acoustic survey of ICES divisions IVa, IVb and VIa. ICES/CIEM, CM 1994/H:22.
- [36] Simmonds E.J., Toresen R., Corten A., Pedersen J., Reid D., Fernandes P.G. - 1994 ICES coordinated acoustic survey of ICES divisions IVa, IVb, VIa and VIIb. ICES/CIEM, CM 1995/H:15.
- [37] Simmonds E.J. - Survey design and effort allocation: A synthesis of choices and decisions for an acoustic survey. North sea herring is used as an example. ICES/CIEM, CM 1995/B:9
- [38] Tomassone R., Lesquoy E., Millier C. - *La regression, nouveaux regards sur une ancienne méthode statistique*. Masson, Paris, 1983.
- [39] Wackernagel H. - *Multivariate geostatistics: an introduction with applications*. Springer, Berlin, 1995.

ÉCOLE NATIONALE SUPÉRIEURE DES MINES DE PARIS

**ANALYSE GÉOSTATISTIQUE DE CAMPAGNES
(ACOUSTIQUE ET CHALUTAGE)
SUR LE HARENG ÉCOSSAIS**

TOME II - FIGURES

THÈSE

présentée devant
l'École Nationale Supérieure des Mines de Paris
par

Philippe GUIBLIN

pour obtenir le titre de Docteur en Géostatistique

Soutenue le 19 Février 1997
devant le jury composé de :

M.	Schmitt	Michel	Président
Mme.	Duby	Camille	Rapporteur externe
M.	Petitgas	Pierre	Rapporteur externe
M.	Rivoirard	Jacques	Examinateur

Table des matières

Figures	1
1 Figures illustrant la partie I	3
2 Figures illustrant la partie II	29
3 Figures illustrant la partie III	77

Table des figures

1.1	Tracé de la campagne acoustique 1989	4
1.2	Représentations proportionnelles des mesures acoustiques	5
1.3	Histogrammes des mesures acoustiques	6
1.4	Histogrammes des mesures acoustiques (valeurs non nulles)	7
1.5	Test de lognormalité des mesures acoustiques – Droite de Henry (valeurs non nulles)	8
1.6	Test de lognormalité des mesures acoustiques – Droite de Henry (données log–translatées)	9
1.7	Représentations proportionnelles des mesures acoustiques – Partie OUEST	10
1.8	Représentations proportionnelles des mesures acoustiques – Partie EST	11
1.9	Représentations proportionnelles des mesures acoustiques (données log–translatées) – Partie OUEST	12
1.10	Représentations proportionnelles des mesures acoustiques (données log–translatées) – Partie EST	13
1.11	Variogrammes directionnels (données brutes) – Partie OUEST	14
1.12	Variogrammes directionnels (données brutes) – Partie EST	15
1.13	Variogrammes directionnels (données log–translatées) – Partie OUEST	16
1.14	Variogrammes directionnels (données log–translatées) – Partie EST	17
1.15	Variogrammes directionnels moyens sur 6 années	18
1.16	Estimation du variogramme des données brutes à l'aide de celui des données log–translatées (direction EST–OUEST) – Partie OUEST	19
1.17	Estimation du variogramme des données brutes à l'aide de celui des données log–translatées (direction EST–OUEST) – Partie EST	20
1.18	Variogrammes calculés le long de la route du navire (données brutes) – Partie OUEST	21
1.19	Variogrammes calculés le long de la route du navire (données brutes) – Partie EST	22
1.20	Estimation du variogramme des données brutes à l'aide de celui des données log–translatées (le long de la route du navire) – Partie OUEST	23
1.21	Estimation du variogramme des données brutes à l'aide de celui des données log–translatées (le long de la route du navire) – Partie EST	24

1.22	Estimation du variogramme des données brutes à l'aide de la covariance non centrée (direction EST-OUEST) – Partie OUEST	25
1.23	Estimation du variogramme des données brutes à l'aide de la covariance non centrée (direction EST-OUEST) – Partie EST	26
1.24	Variogramme expérimental calculé le long de la route du navire et ajustement – Campagne 1993	27
2.1	Stations d'échantillonnage – Campagne 1993 et (post-)stratification	30
2.2	Représentation proportionnelle de la distribution des longueurs moyennes	31
2.3	Représentation proportionnelle de la distribution de la classe d'âge 1	32
2.4	Représentation proportionnelle de la distribution de la classe d'âge 2 (immatures)	33
2.5	Représentation proportionnelle de la distribution de la classe d'âge 2 (matures)	34
2.6	Représentation proportionnelle de la distribution de la classe d'âge 3 (immatures)	35
2.7	Représentation proportionnelle de la distribution de la classe d'âge 3	36
2.8	Représentation proportionnelle de la distribution de la classe d'âge 4	37
2.9	Représentation proportionnelle de la distribution de la classe d'âge 5	38
2.10	Représentation proportionnelle de la distribution de la classe d'âge 6	39
2.11	Représentation proportionnelle de la distribution de la classe d'âge 7	40
2.12	Représentation proportionnelle de la distribution de la classe d'âge 8	41
2.13	Représentation proportionnelle de la distribution de la classe d'âge 9	42
2.14	Représentation proportionnelle de la distribution des longueurs et des âges pour toutes les années	43
2.15	Représentation proportionnelle de la distribution des longueurs et des âges pour toutes les années (suite)	44
2.16	Variogrammes de la longueur moyenne et des proportions d'âge	45
2.17	Variogrammes de la longueur moyenne et des proportions d'âge (suite)	46
2.18	Relation longueur moyenne/profondeur (toutes les années)	47
2.19	Variogrammes des résidus de la régression longueur/profondeur	48
2.20	Cartographies de la longueur sur une grille 5 mn x 5 mn (krigeage avec dérive externe)	49
2.21	Krigeage de la longueur : cartographie des écarts – types d'estimation	50
2.22	Cartographie moyenne de la longueur sur une grille 5 mn x 5 mn (krigeage avec dérive externe)	51
2.23	Nuages de corrélation âge/longueur moyenne	52
2.24	Nuages de corrélation âge/longueur moyenne (suite)	53

2.25	Nuages de corrélation âges cumulés/longueur moyenne	54
2.26	Nuages de corrélation âges cumulés/longueur moyenne (suite)	55
2.27	Régressions âges cumulés/longueur moyenne – Année 1993	56
2.28	Régressions âges cumulés/longueur moyenne – Année 1993 (suite)	57
2.29	Cartographie des âges cumulés par l'approche déterministe	58
2.30	Cartographies des âges cumulés par l'approche déterministe (suite)	59
2.31	Cartographie des proportions d'âge par l'approche déterministe	60
2.32	Cartographie des proportions d'âge par l'approche déterministe (suite)	61
2.33	Régression T3.1/longueur moyenne sur 6 années	62
2.34	Régression T5/longueur moyenne sur 6 années	63
2.35	Régression T7/longueur moyenne sur 6 années	64
2.36	Tests sur la distribution des résidus (approche directe)	65
2.37	Etude variographique croisée des résidus (Régression $T = f(l) + \epsilon$)	66
2.38	Simulations de la classe d'âges cumulés T3.1	67
2.39	Simulations de la classe d'âges cumulés T5	68
2.40	Simulations de la classe d'âges cumulés T7	69
2.41	Simulations de la proportion A1 + A2	70
2.42	Simulations de la proportion A3 + A4	71
2.43	Simulations de la proportion A5 + A6	72
2.44	Simulations de la proportion Age ≥ 7	73
2.45	Régressions âges cumulés/longueur moyenne (régression inverse)	74
2.46	Régressions âges cumulés/longueur moyenne (régression inverse) (suite)	75
2.47	Tests sur la distribution des résidus (approche inverse)	76
3.1	Echantillonnage acoustique – Stations d'échantillonnage – Campagne 1993	78
3.2	Représentation proportionnelle de la densité estimée – 1993	79
3.3	Représentation proportionnelle de la densité estimée – 1993 – Parties OUEST et EST	80
3.4	Variogramme de la densité estimée expérimental calculé le long de la route du navire et ajustement – Campagne 1993	81

Chapitre 1

Figures illustrant la partie I

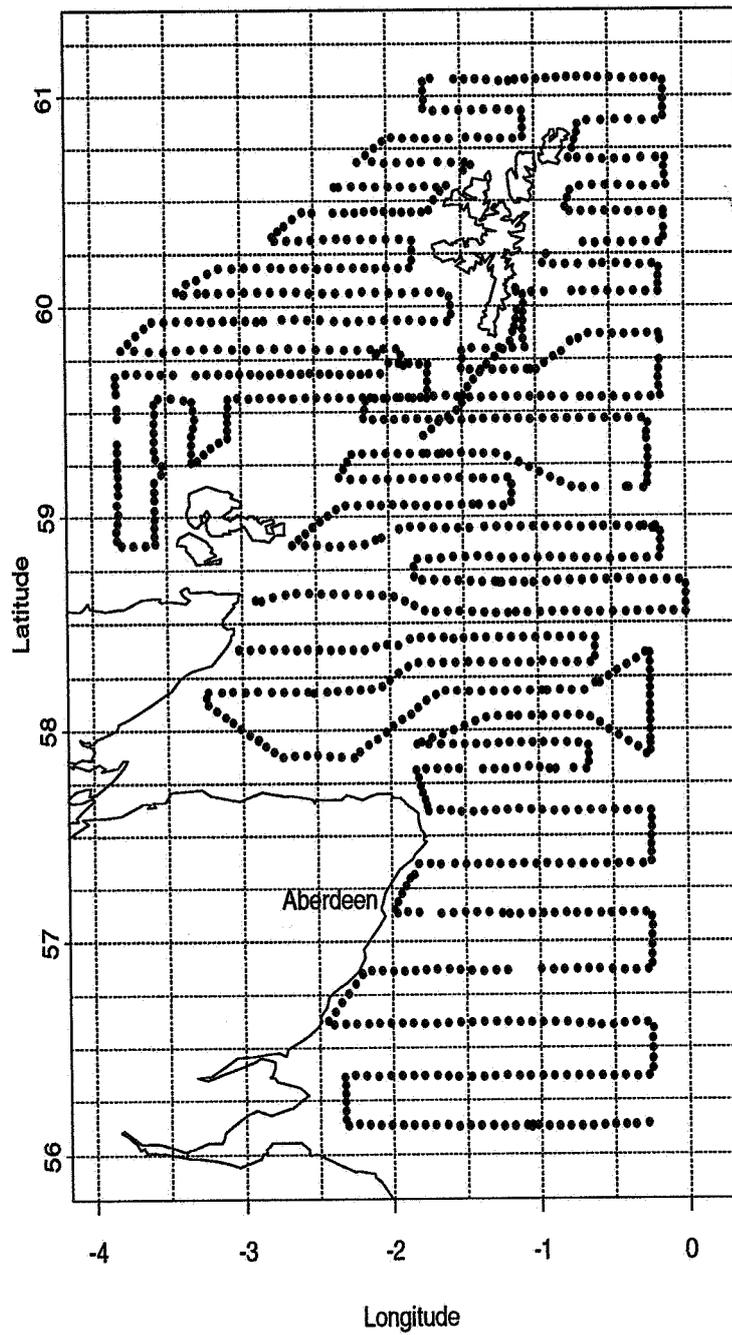


FIG. 1.1 - *Tracé de la campagne acoustique 1989*

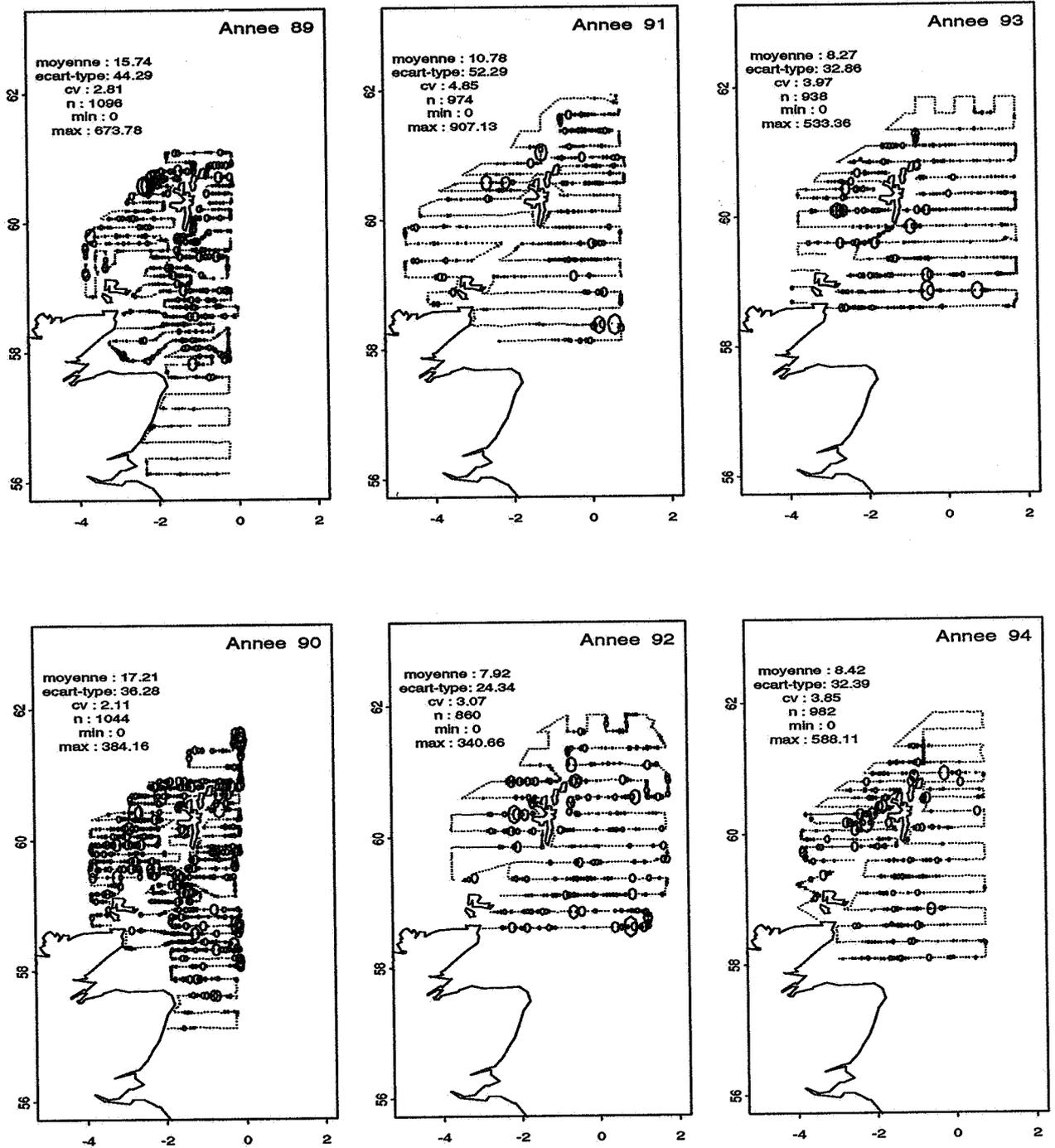


FIG. 1.2 - Représentations proportionnelles des mesures acoustiques

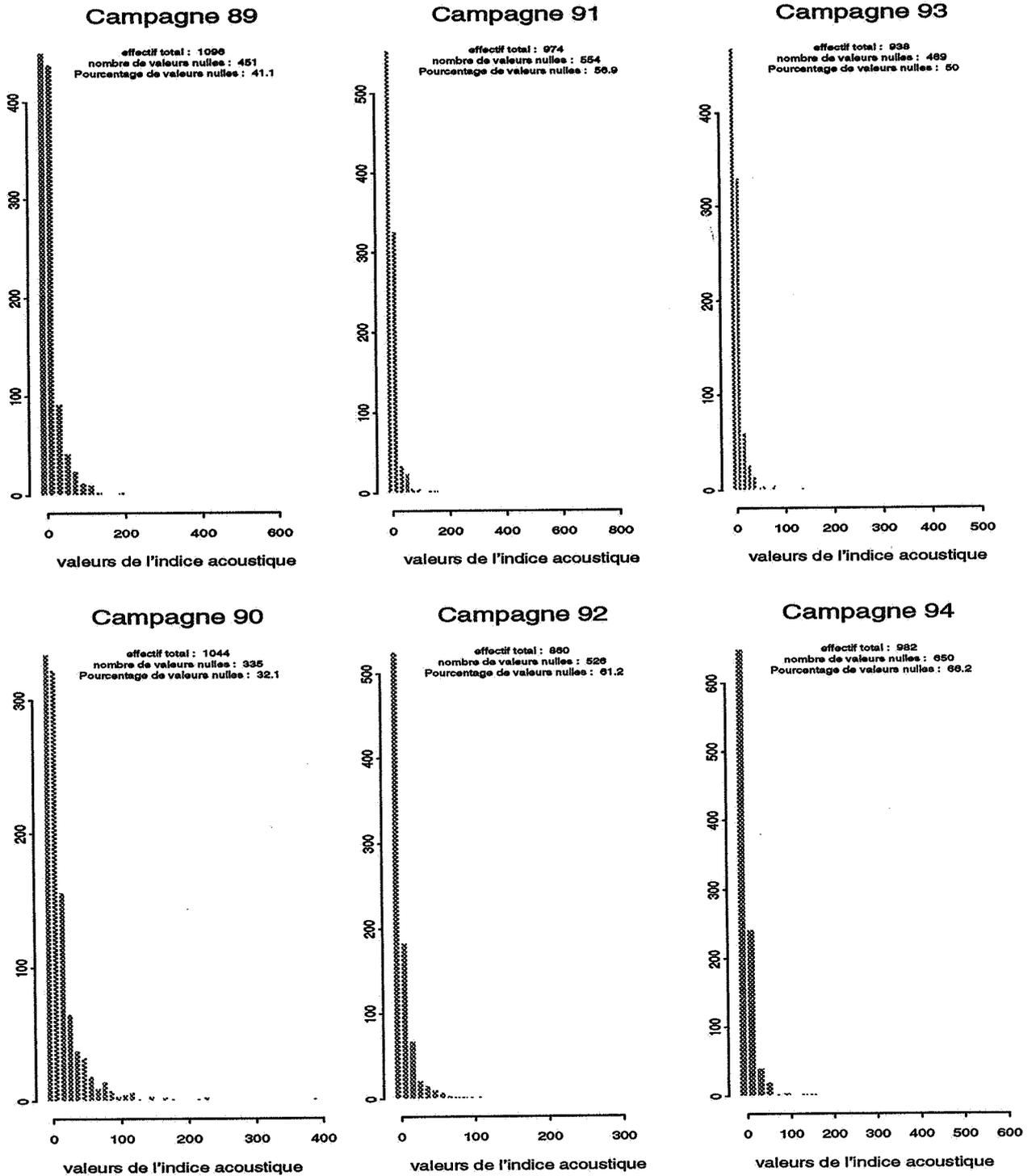


FIG. 1.3 - *Histogrammes des mesures acoustiques*

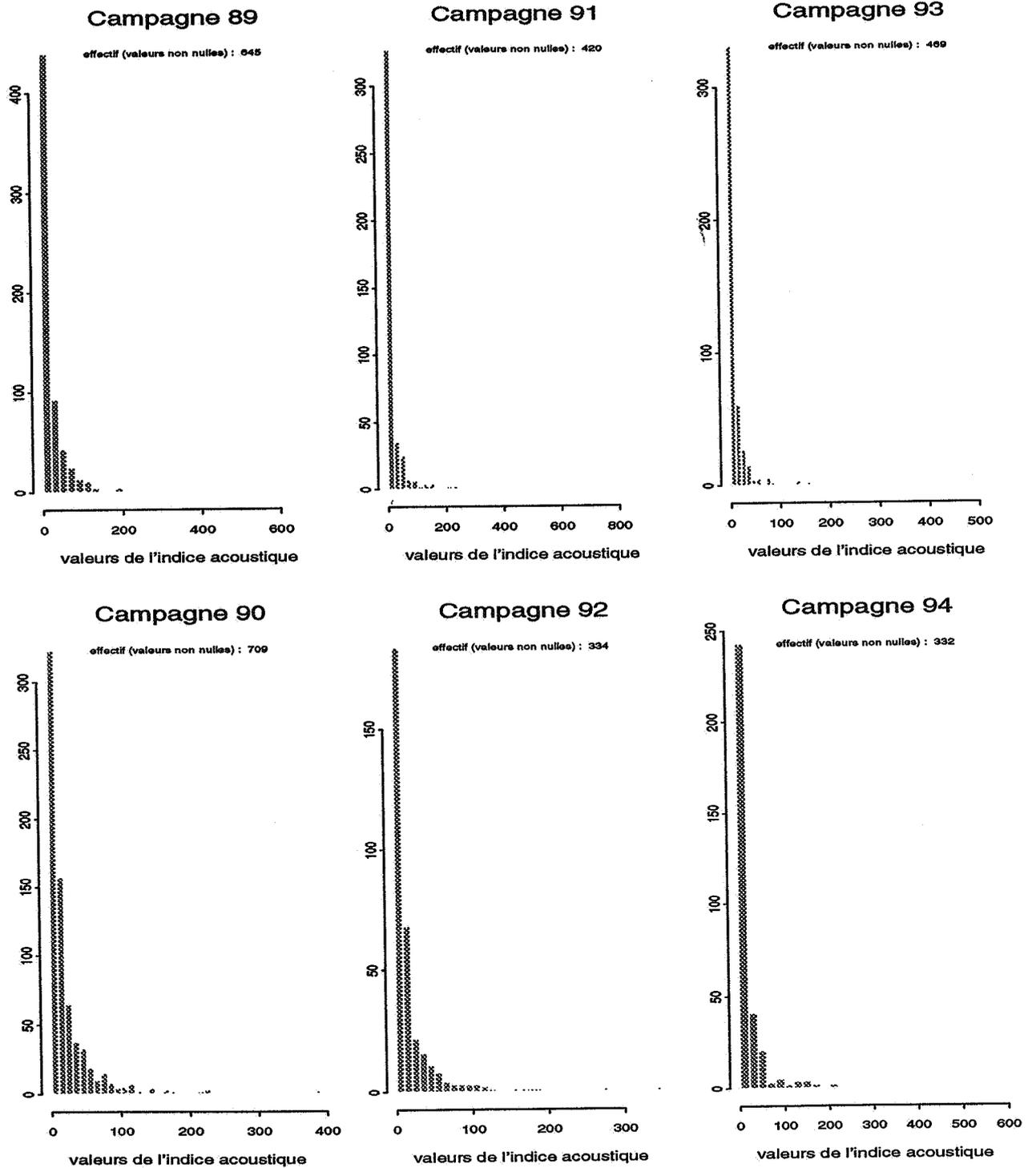


FIG. 1.4 - Histogrammes des mesures acoustiques (valeurs non nulles)

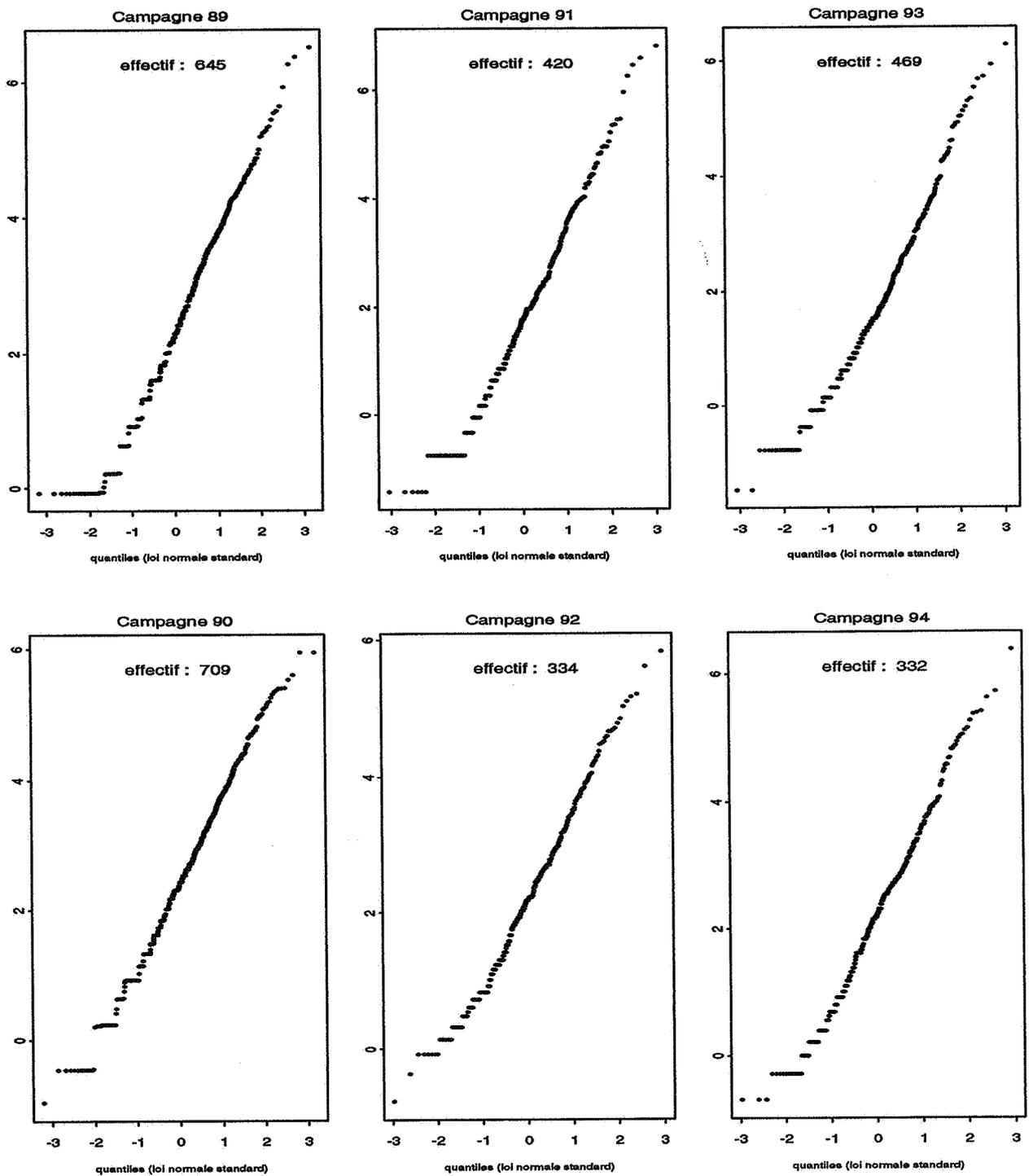


FIG. 1.5 - Test de lognormalité des mesures acoustiques - Droite de Henry (valeurs non nulles)

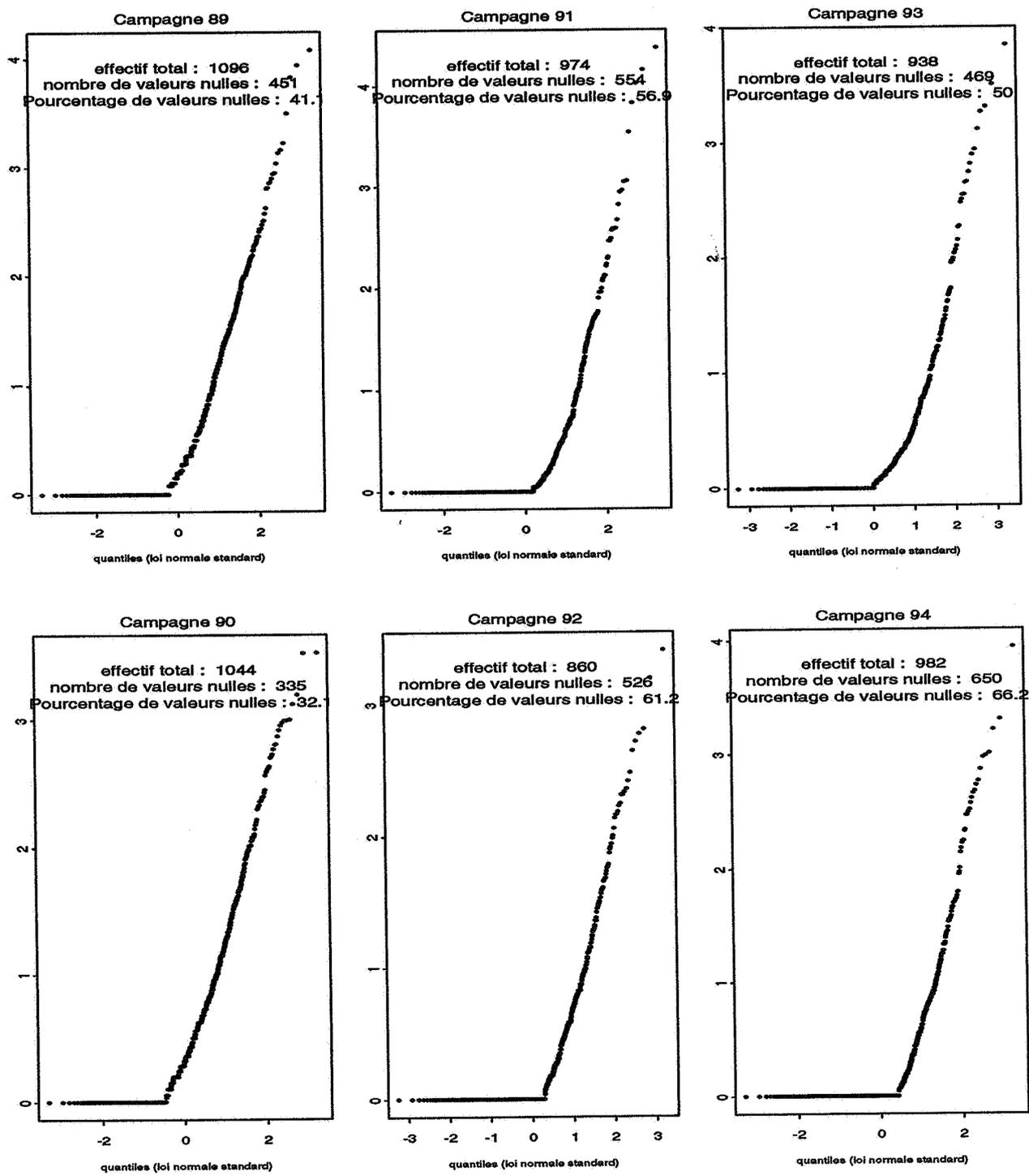


FIG. 1.6 - Test de normalité des mesures acoustiques - Droite de Henry (données log-translatées)

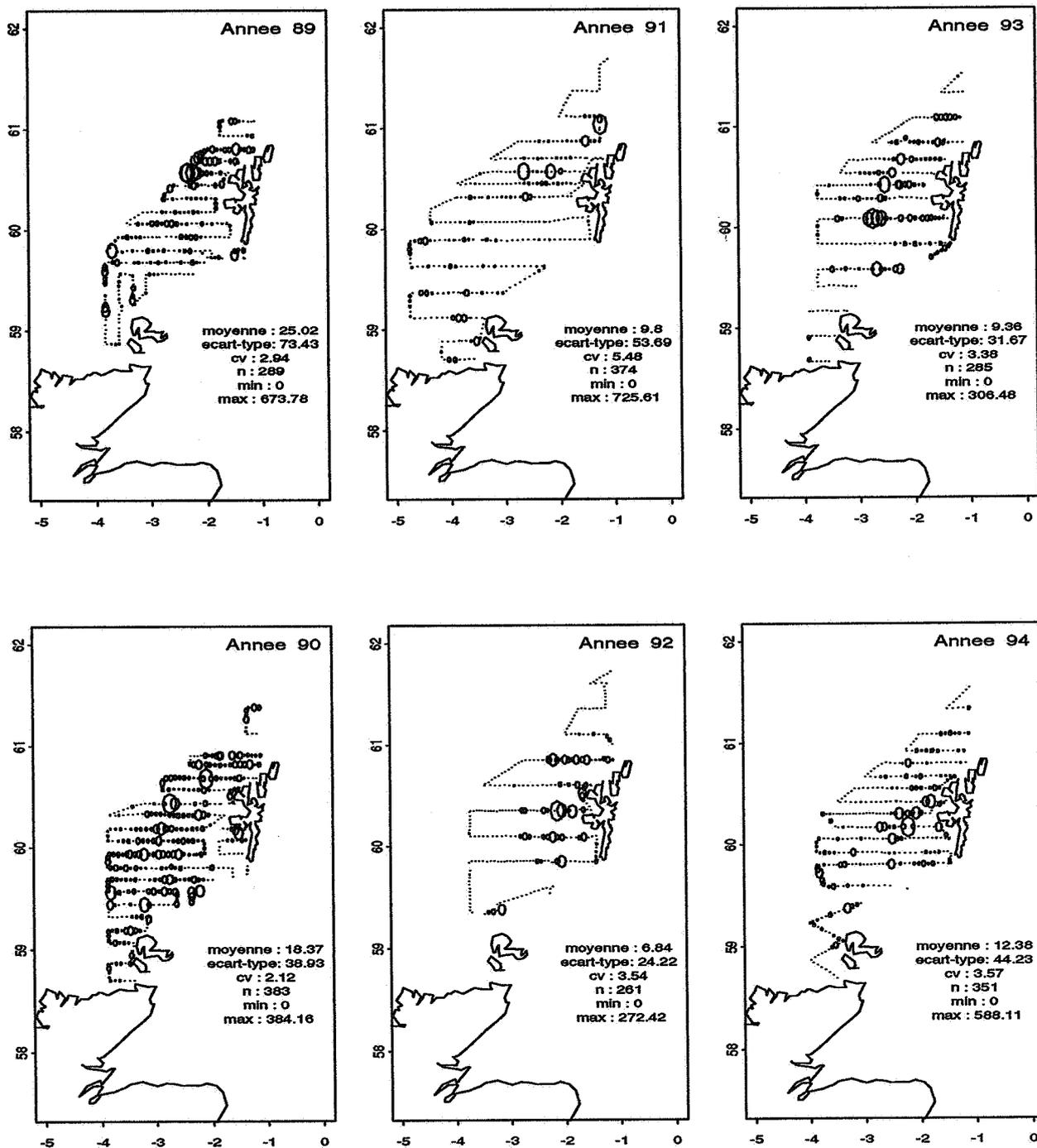


FIG. 1.7 - Représentations proportionnelles des mesures acoustiques - Partie OUEST

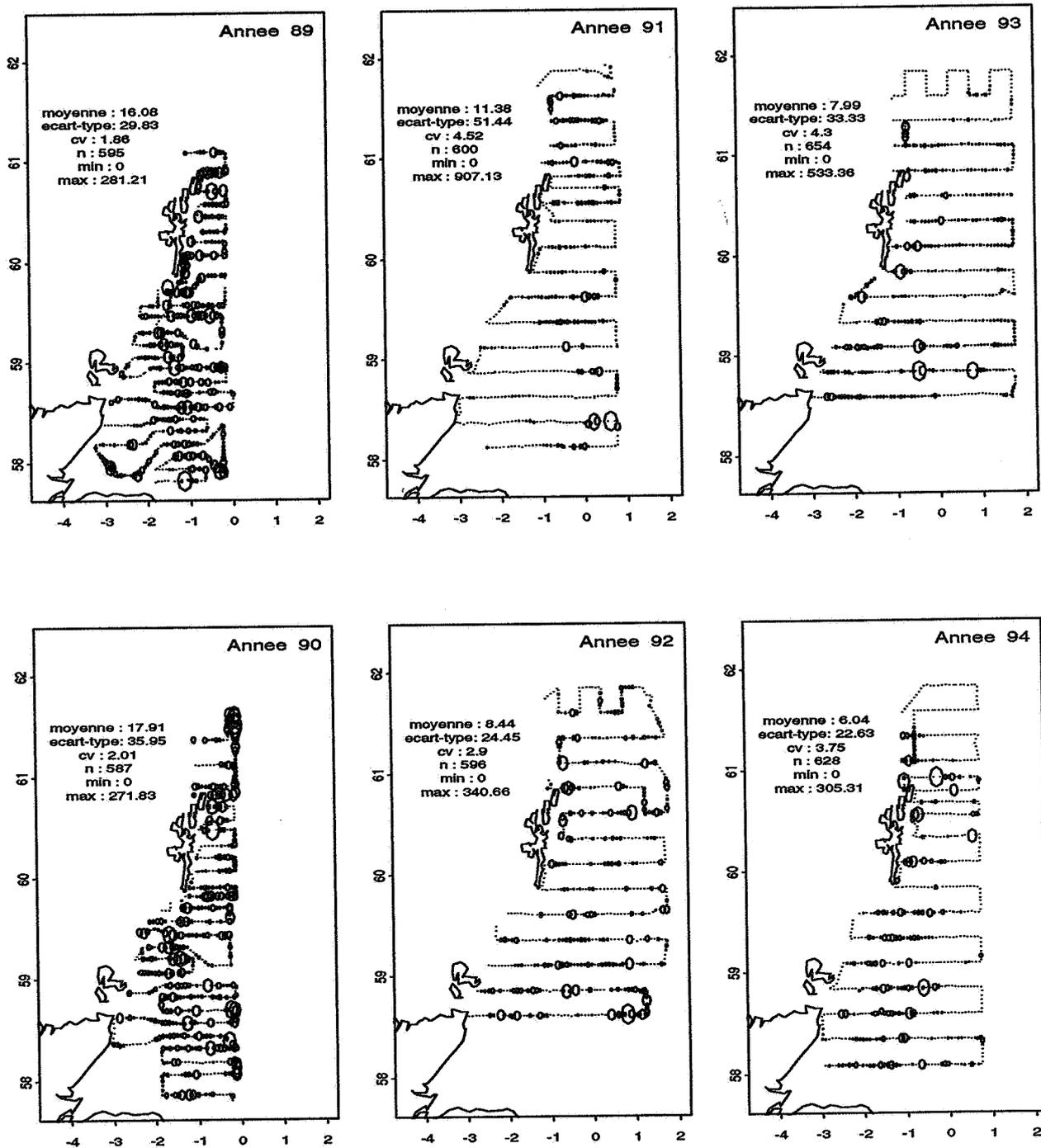


FIG. 1.8 - Représentations proportionnelles des mesures acoustiques - Partie EST

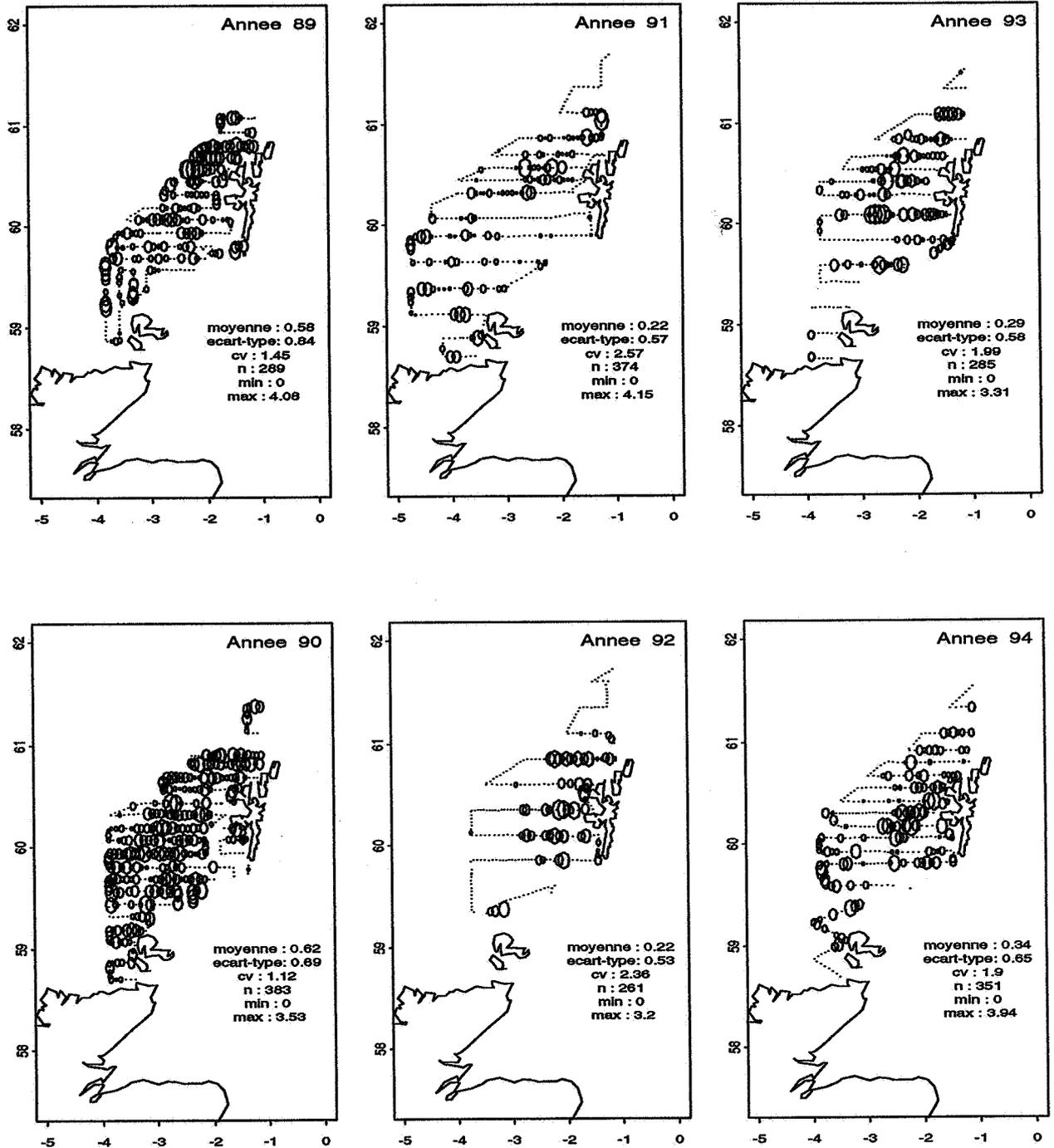


FIG. 1.9 - Représentations proportionnelles des mesures acoustiques (données log-translatées) - Partie OUEST

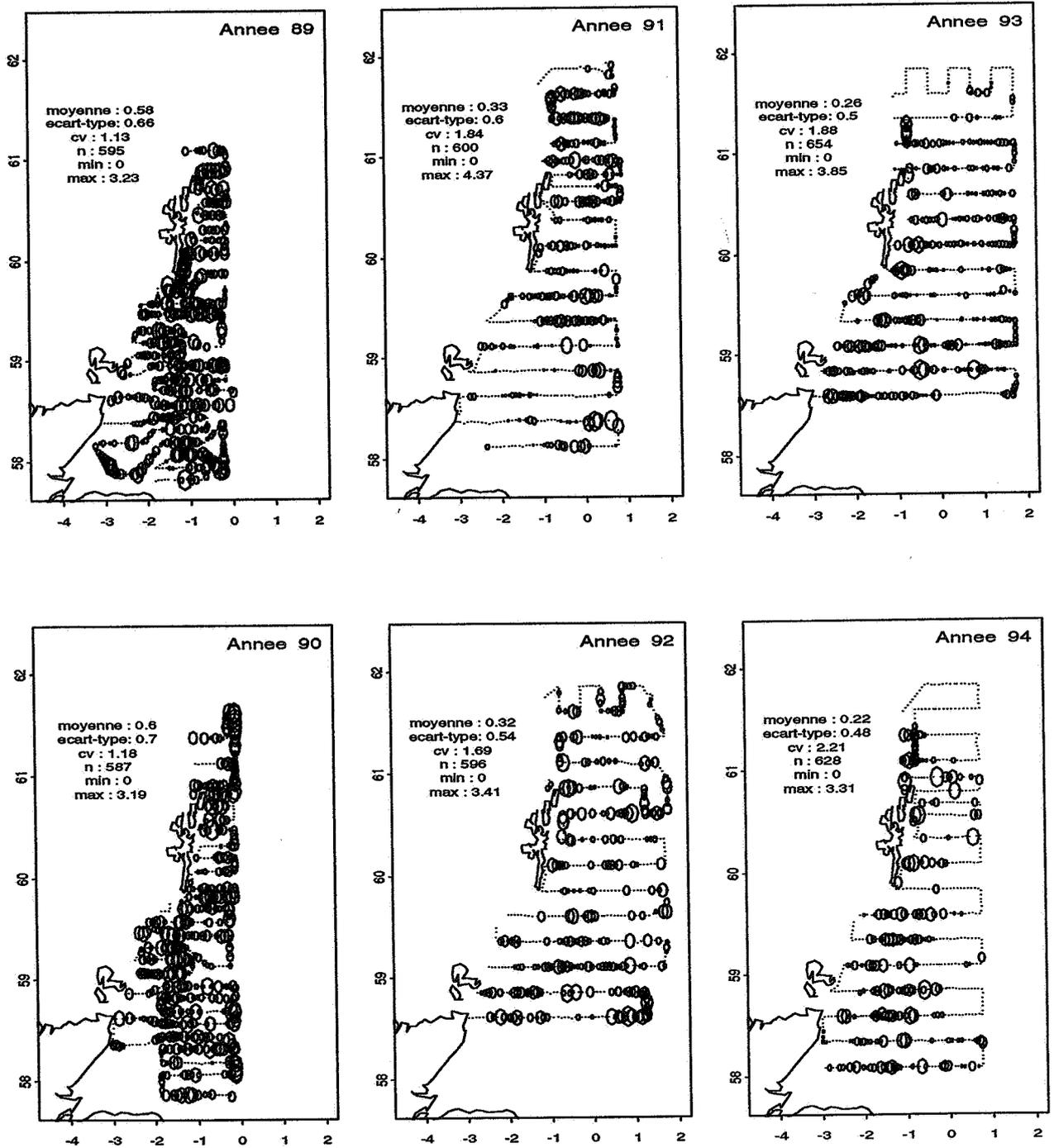


FIG. 1.10 - Représentations proportionnelles des mesures acoustiques (données log-translatées - Partie EST

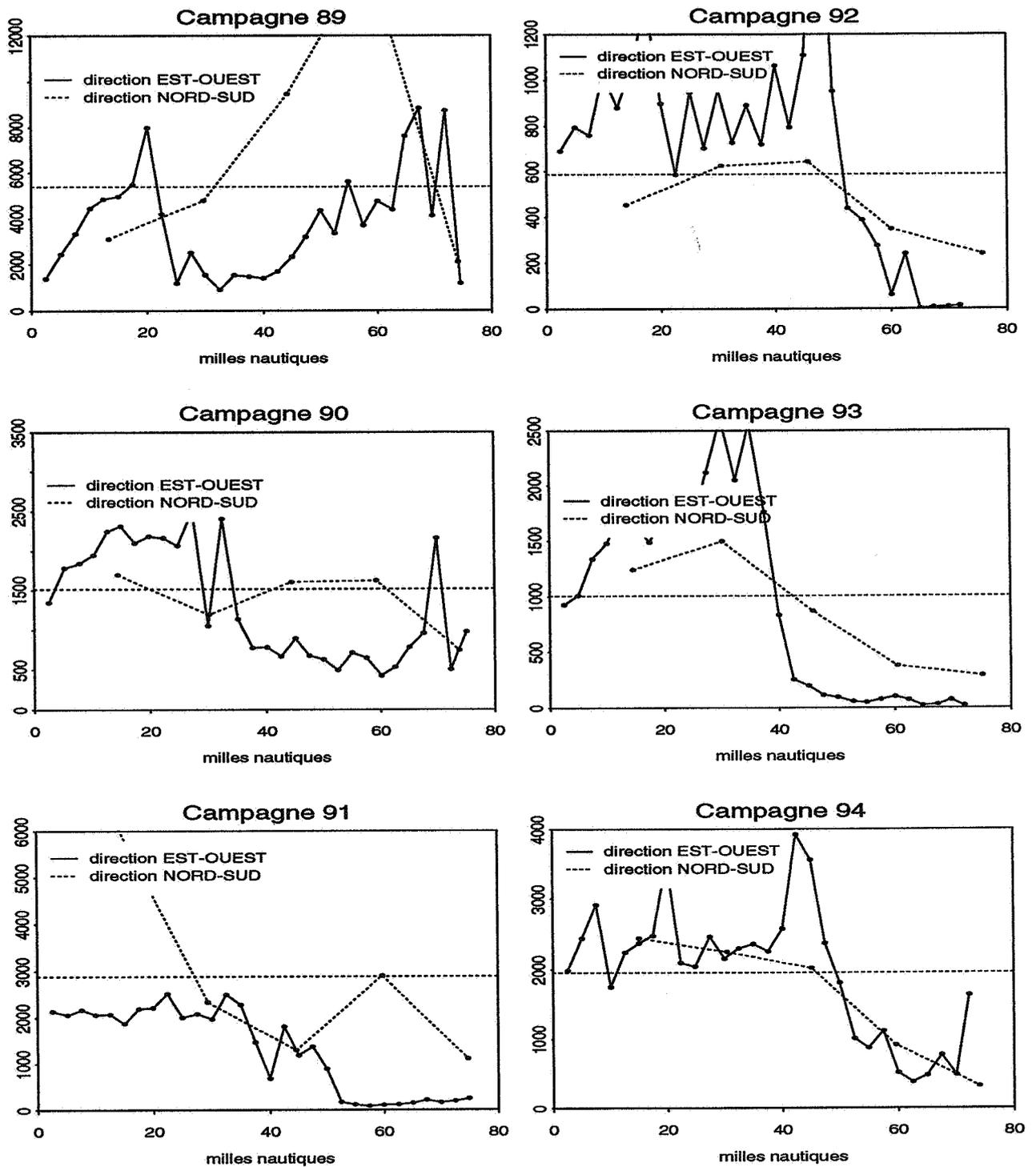


FIG. 1.11 - Variogrammes directionnels (données brutes) - Partie OUEST

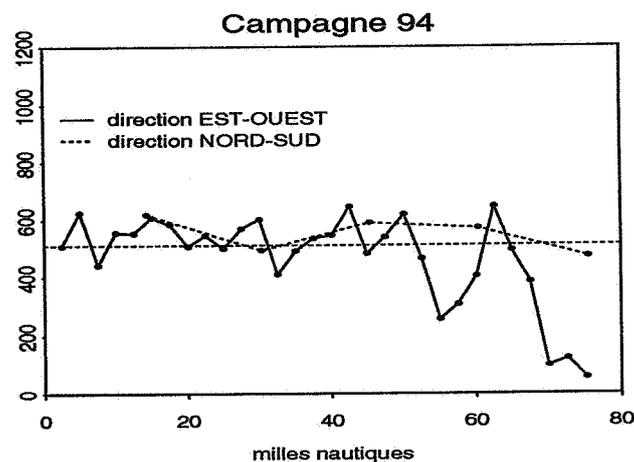
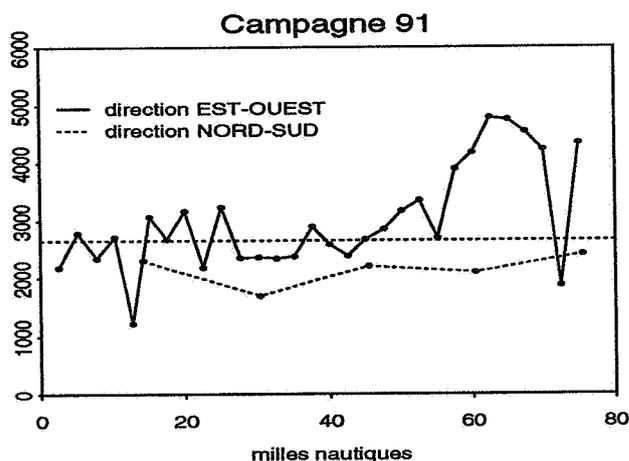
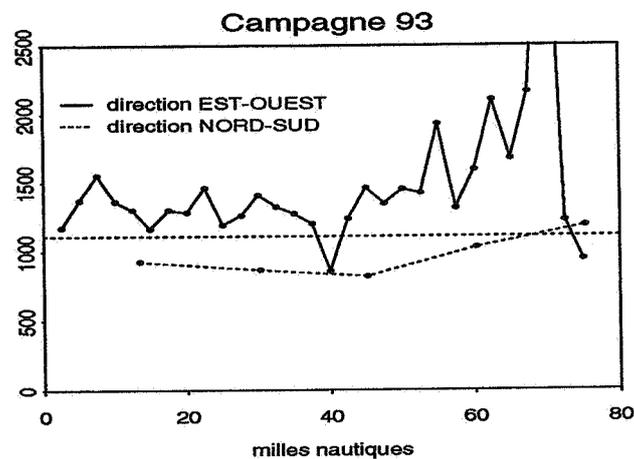
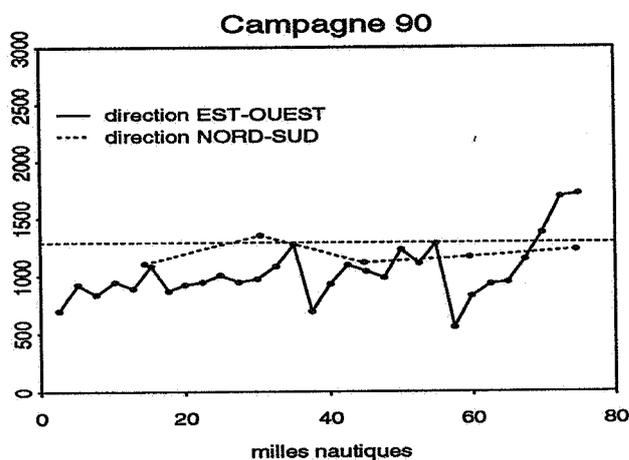
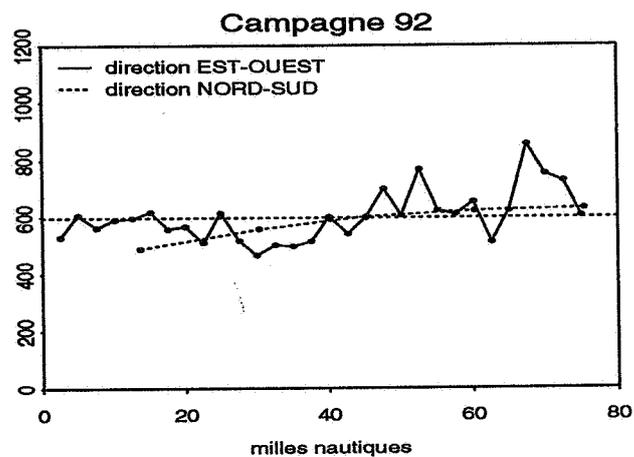
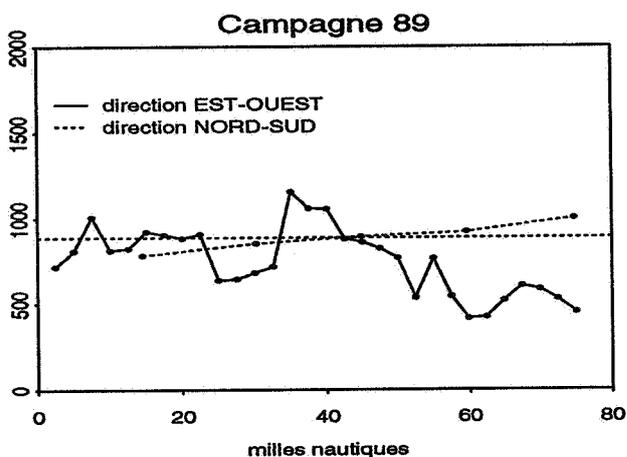


FIG. 1.12 - Variogrammes directionnels (données brutes) - Partie EST

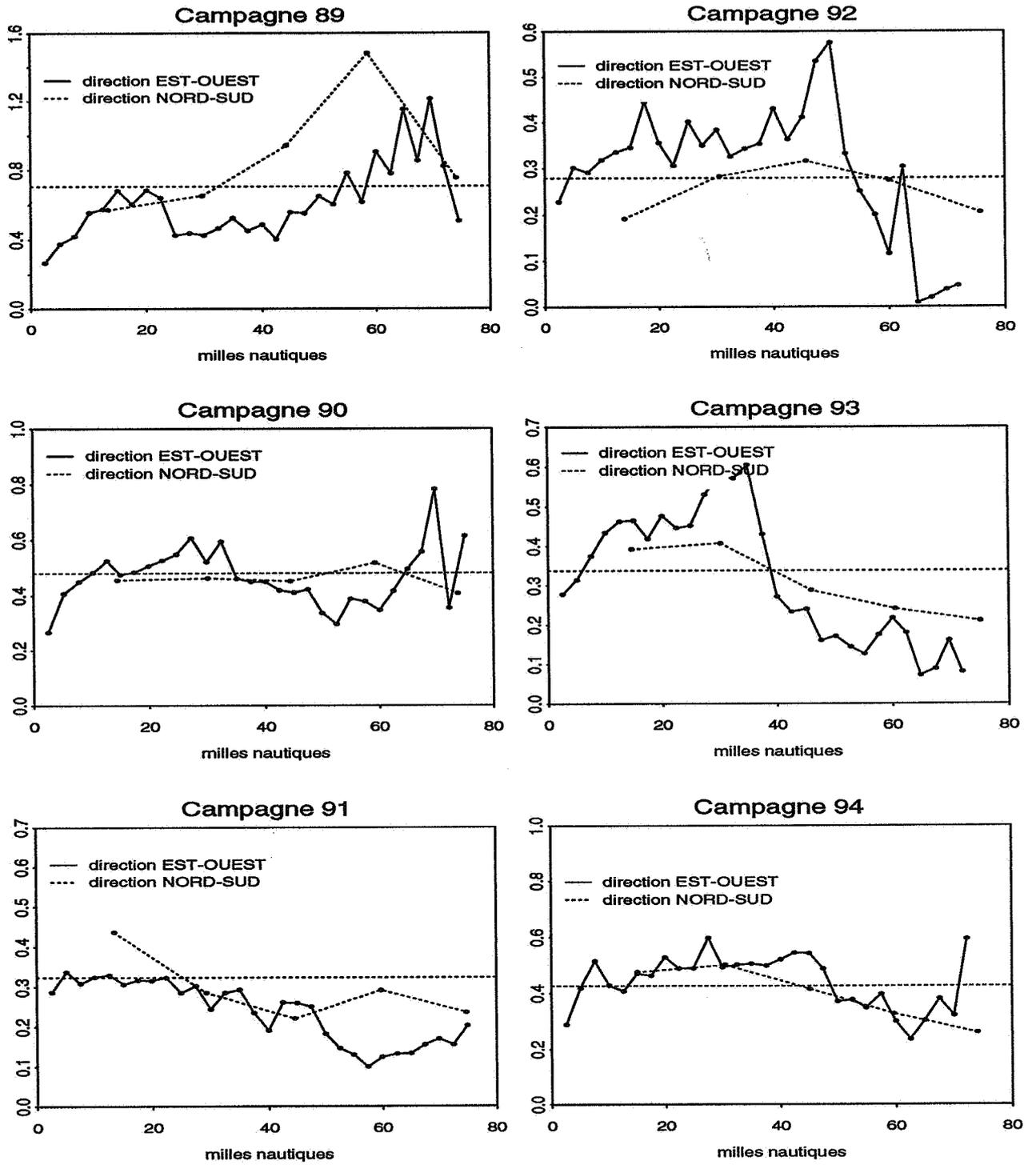


FIG. 1.13 - Variogrammes directionnels (données log-translatées) - Partie OUEST

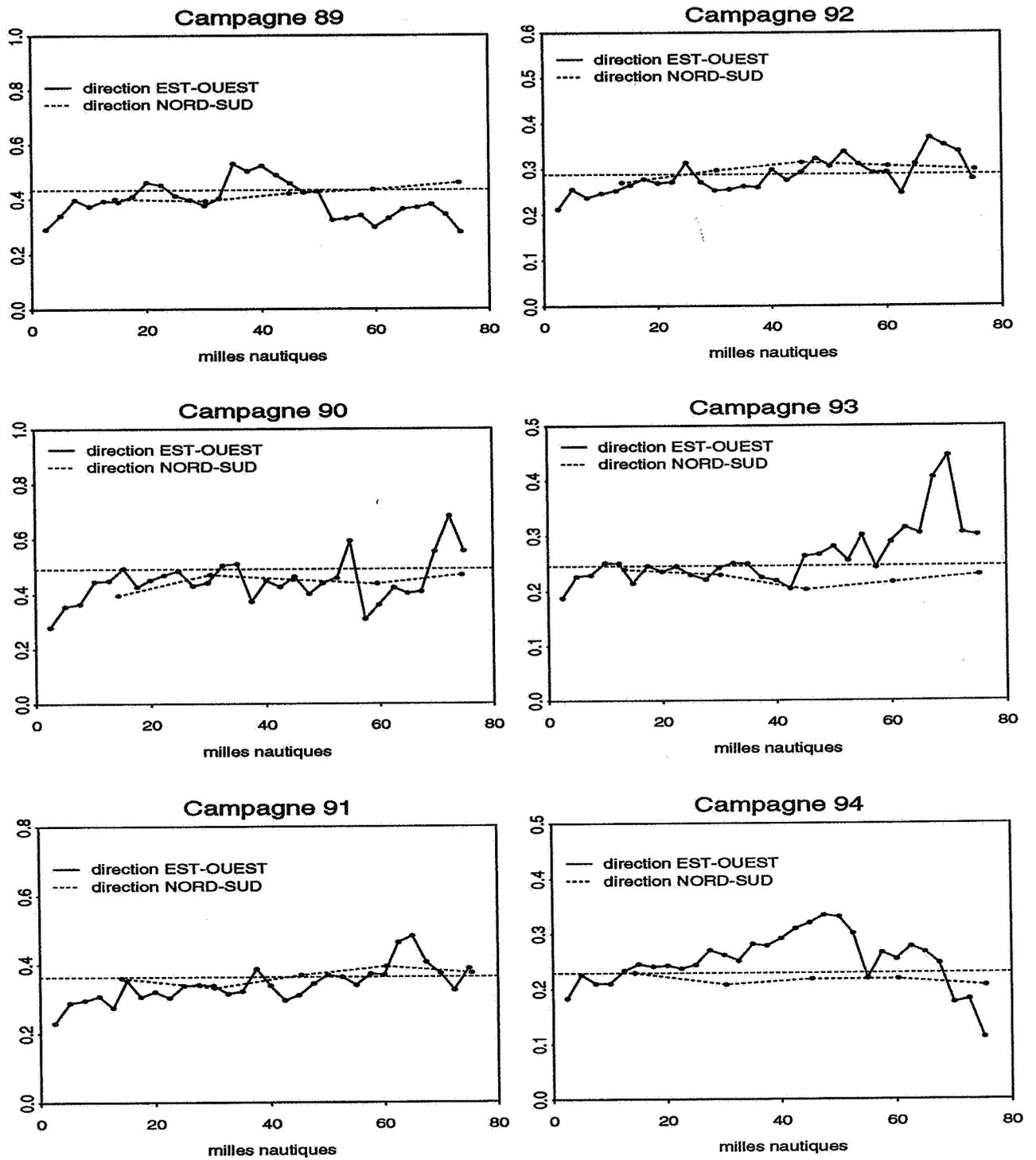


FIG. 1.14 - Variogrammes directionnels (données log-translatées) - Partie EST

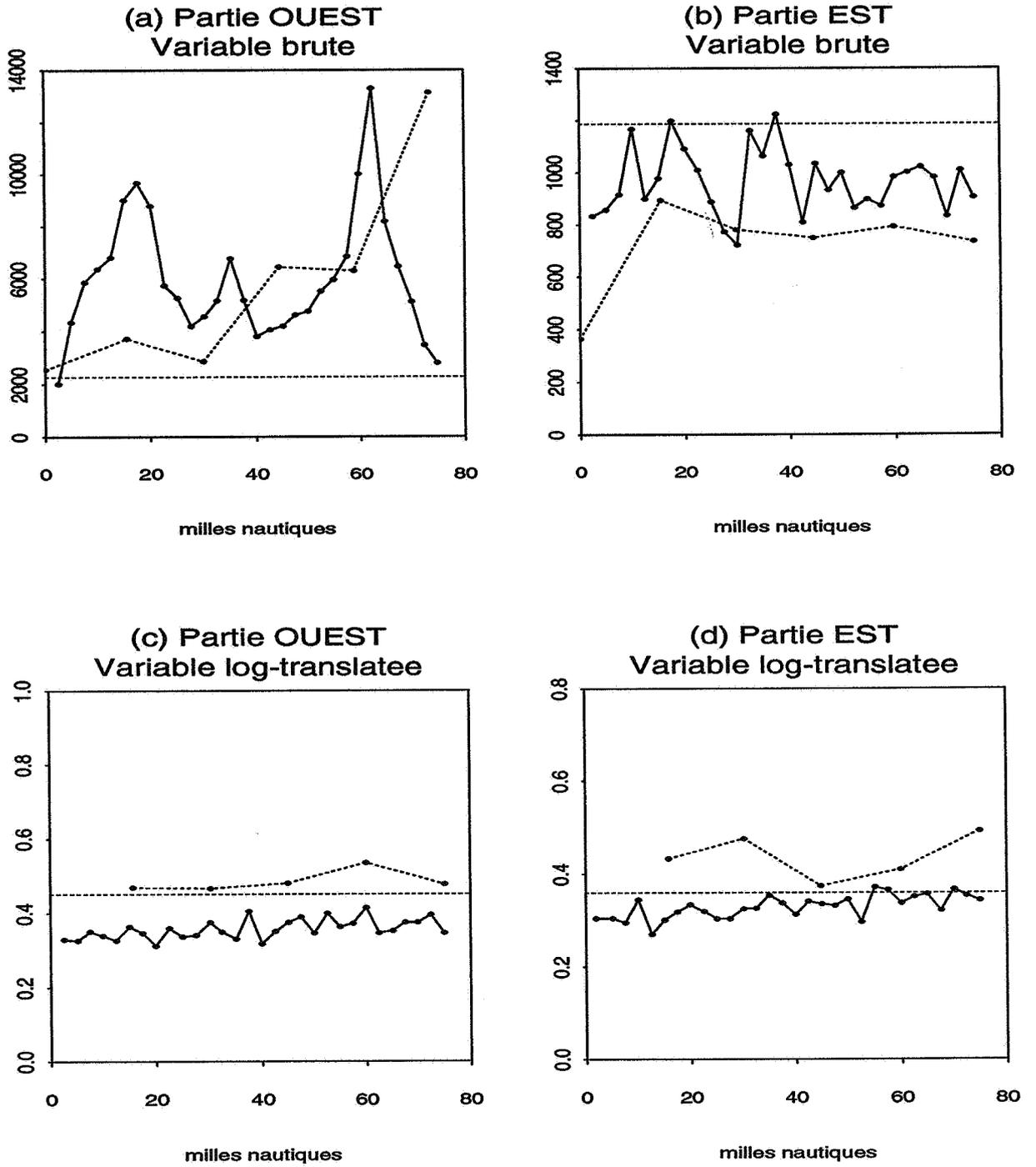


FIG. 1.15 - Variogrammes directionnels moyens sur 6 années

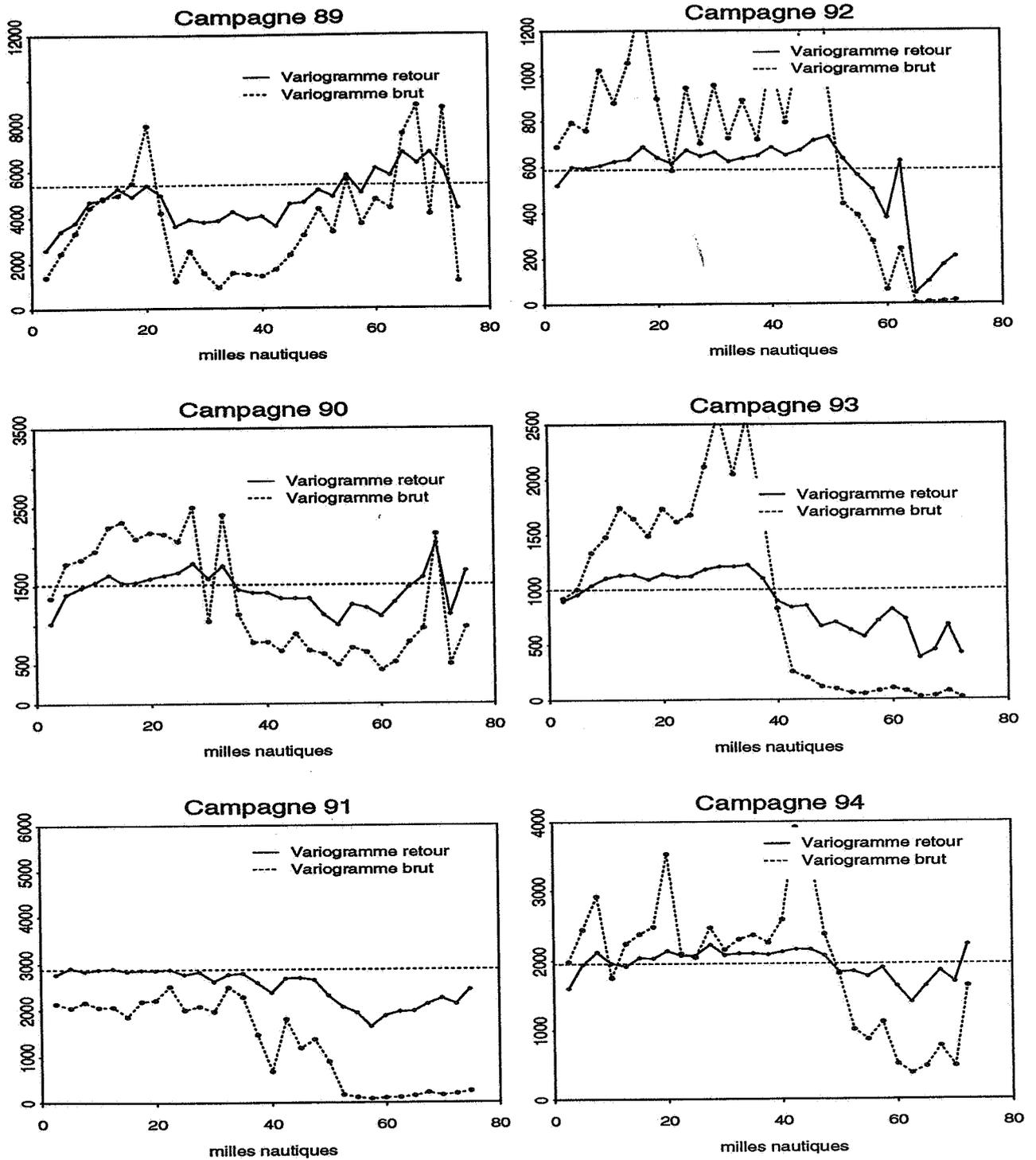


FIG. 1.16 - Estimation du variogramme des données brutes à l'aide de celui des données log-translatées (direction EST-OUEST) - Partie OUEST

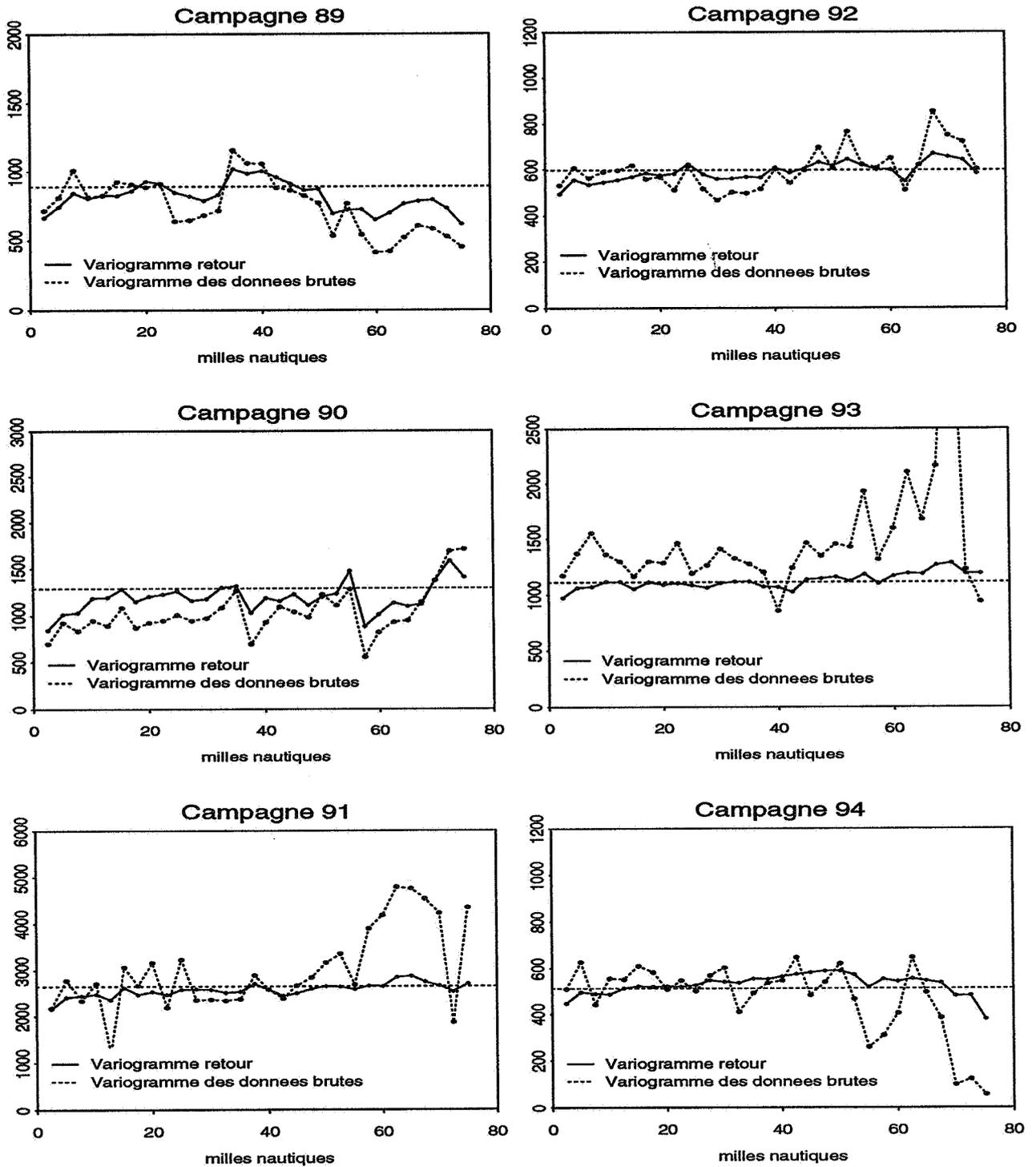


FIG. 1.17 - Estimation du variogramme des données brutes à l'aide de celui des données log-translatées (direction EST-OUEST) - Partie EST

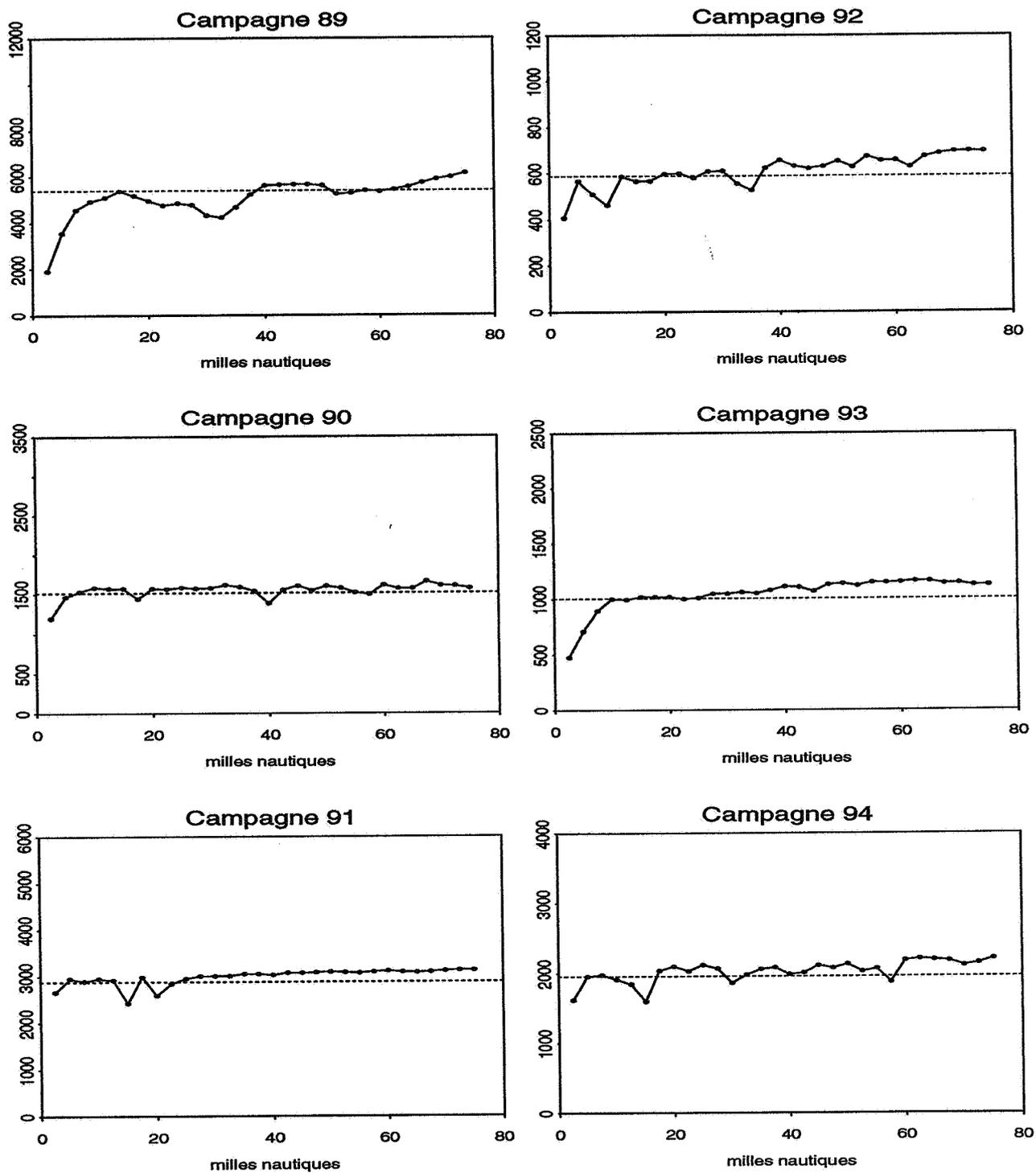


FIG. 1.18 - Variogrammes calculés le long de la route du navire (données brutes)
- Partie OUEST

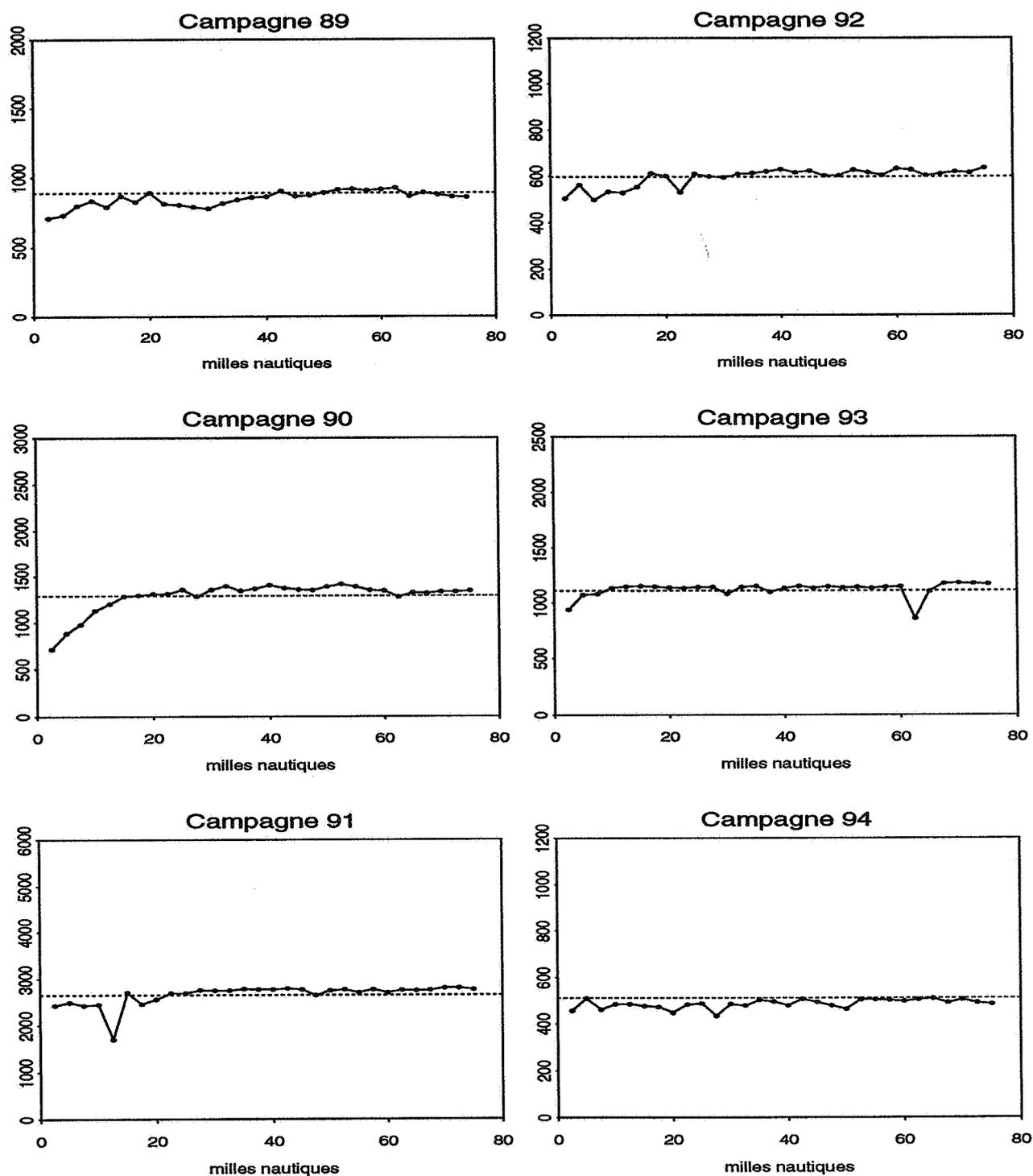


FIG. 1.19 - Variogrammes calculés le long de la route du navire (données brutes)
- Partie EST

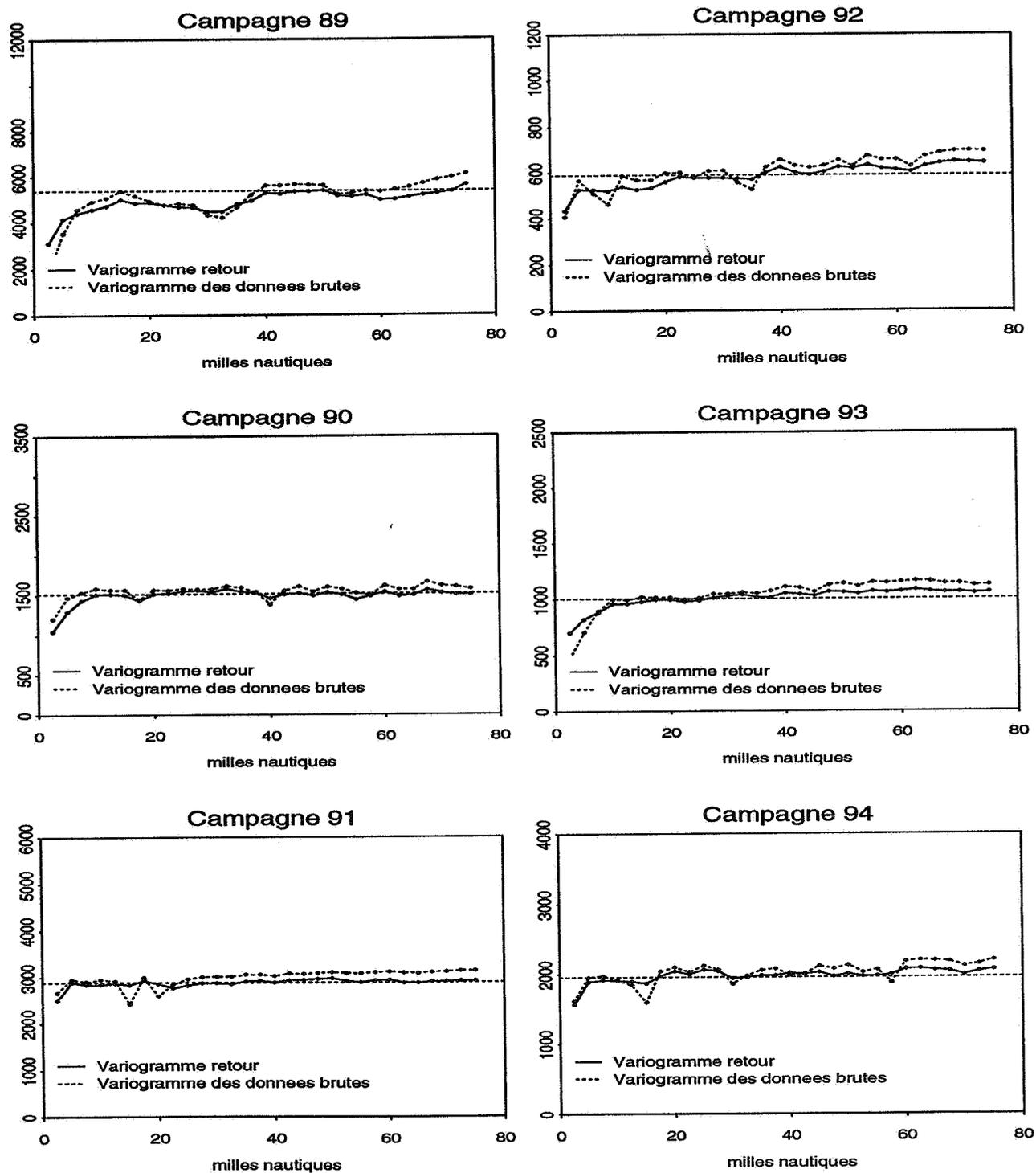


FIG. 1.20 - Estimation du variogramme des données brutes à l'aide de celui des données log-translatées (le long de la route du navire) - Partie OUEST

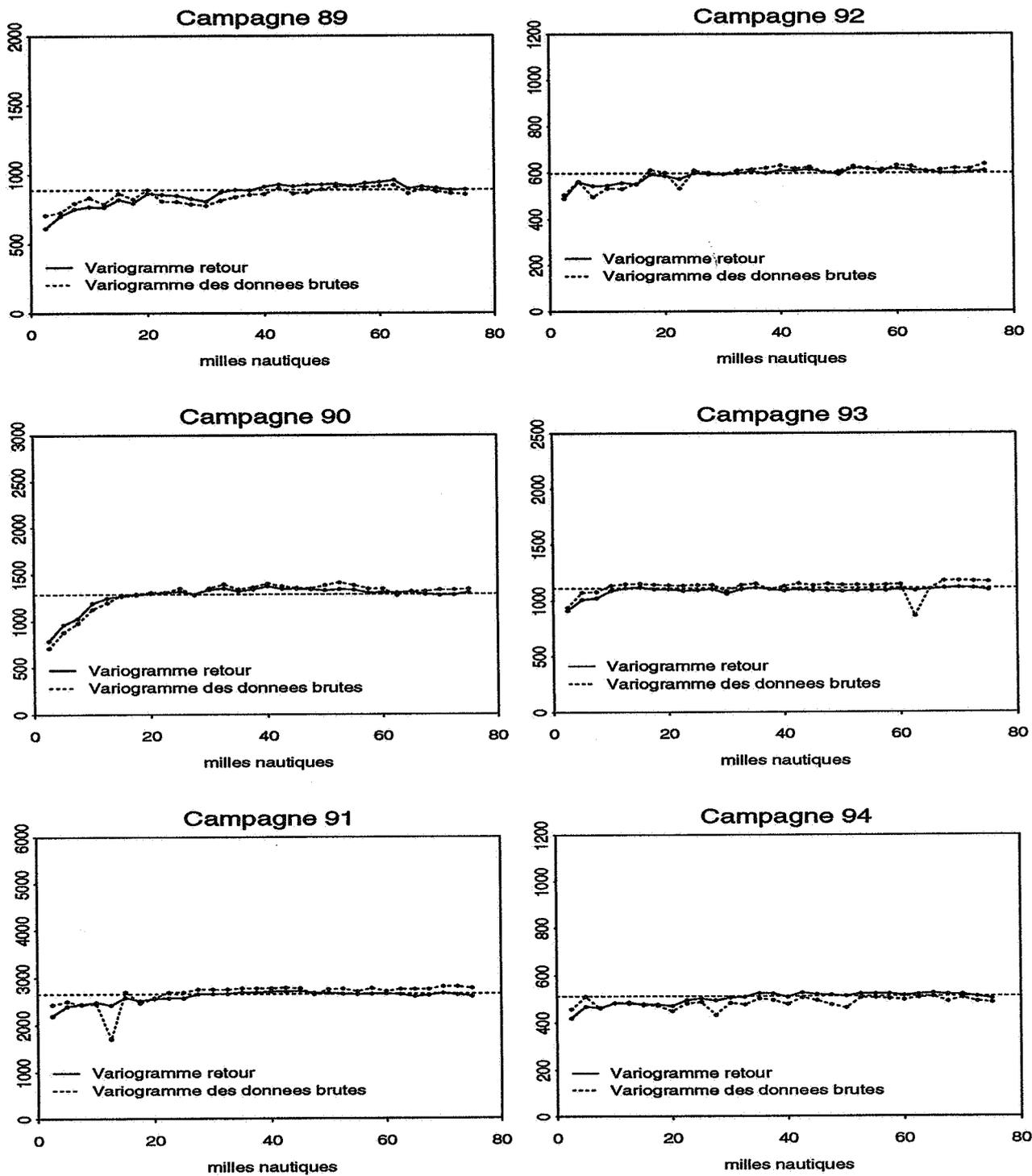


FIG. 1.21 - Estimation du variogramme des données brutes à l'aide de celui des données log-translatées (le long de la route du navire) - Partie EST

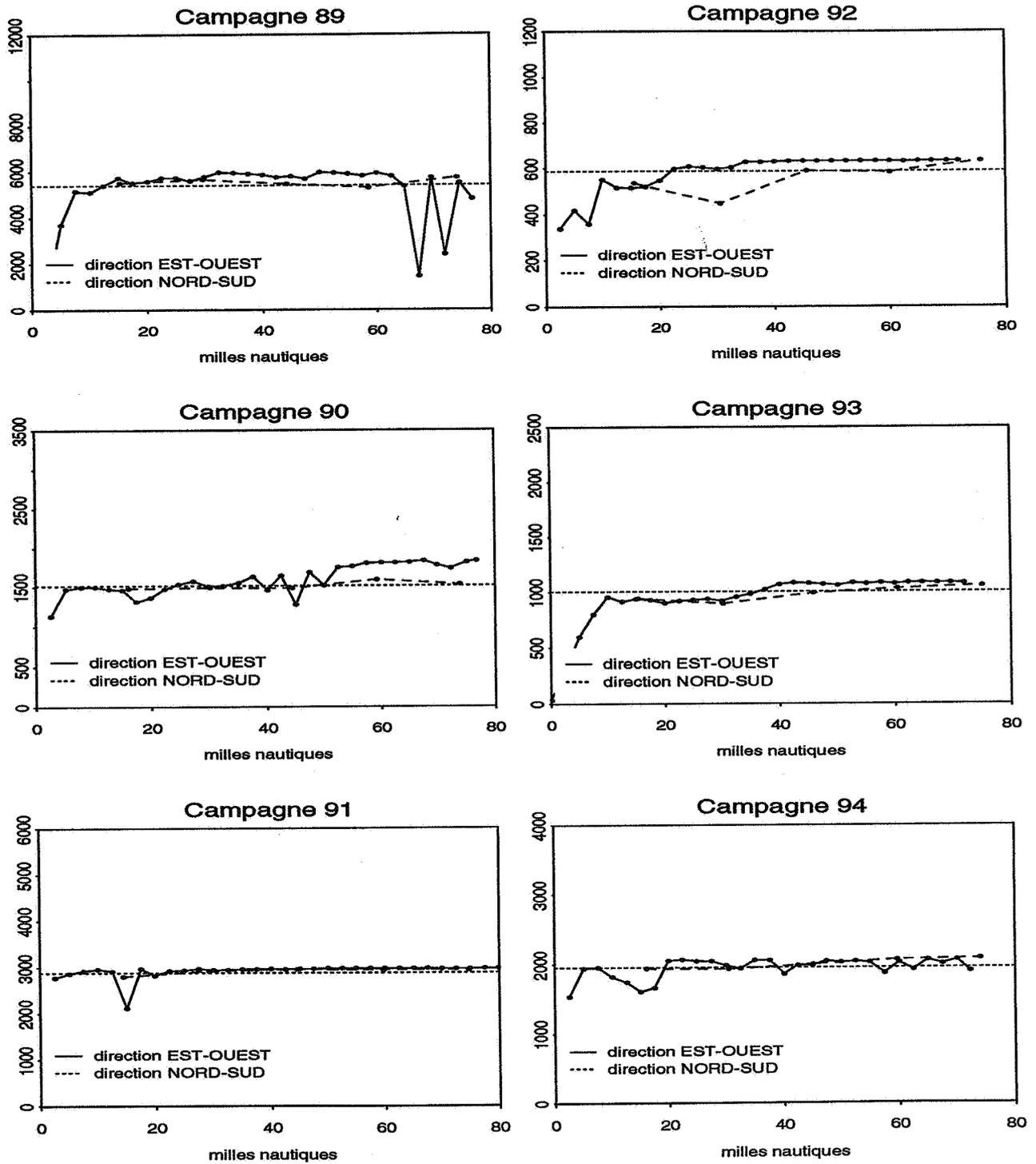


FIG. 1.22 - Estimation du variogramme des données brutes à l'aide de la covariance non centrée (direction EST-OUEST) - Partie OUEST

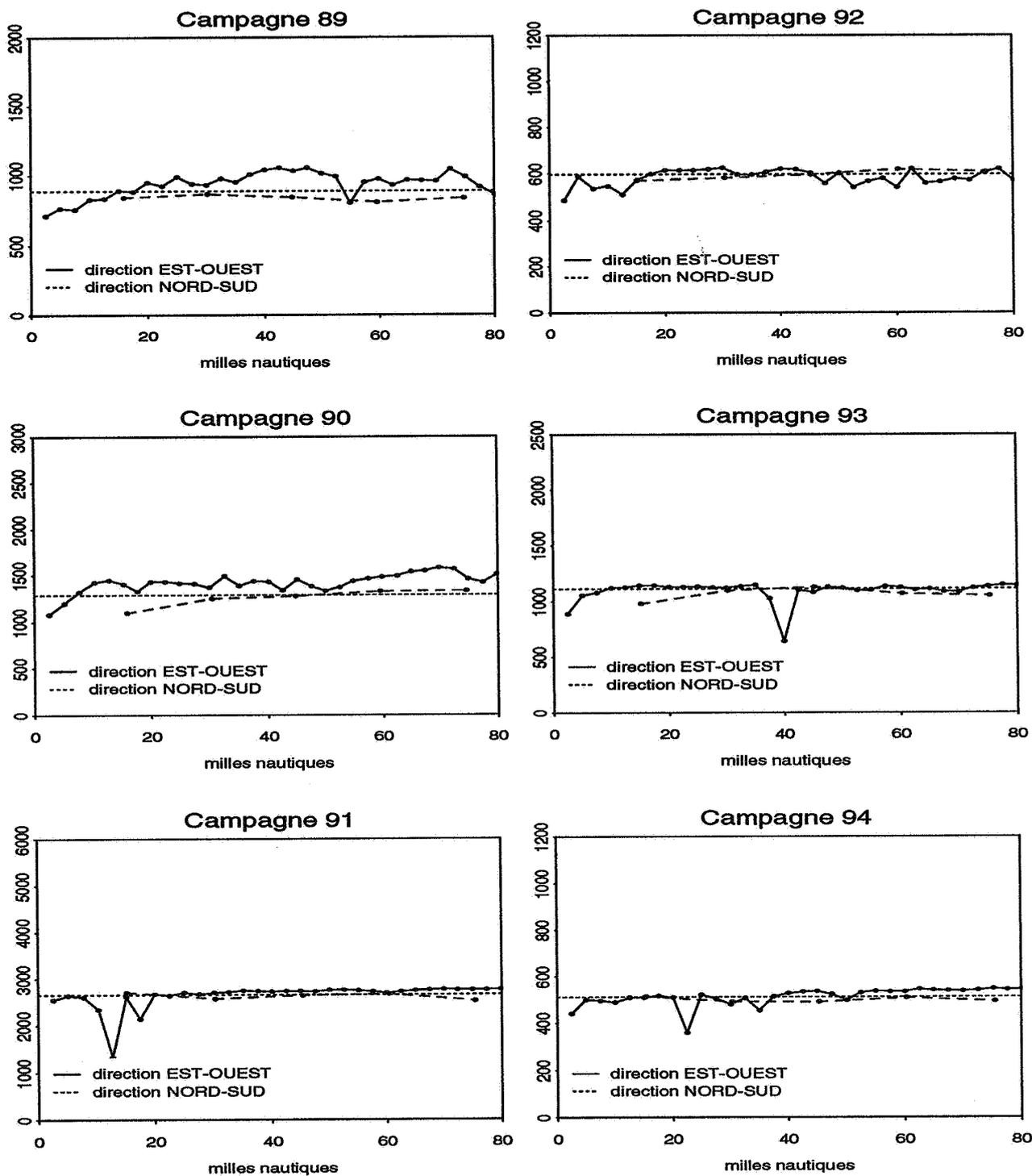


FIG. 1.23 - Estimation du variogramme des données brutes à l'aide de la covariance non centrée (direction EST-OUEST) - Partie EST

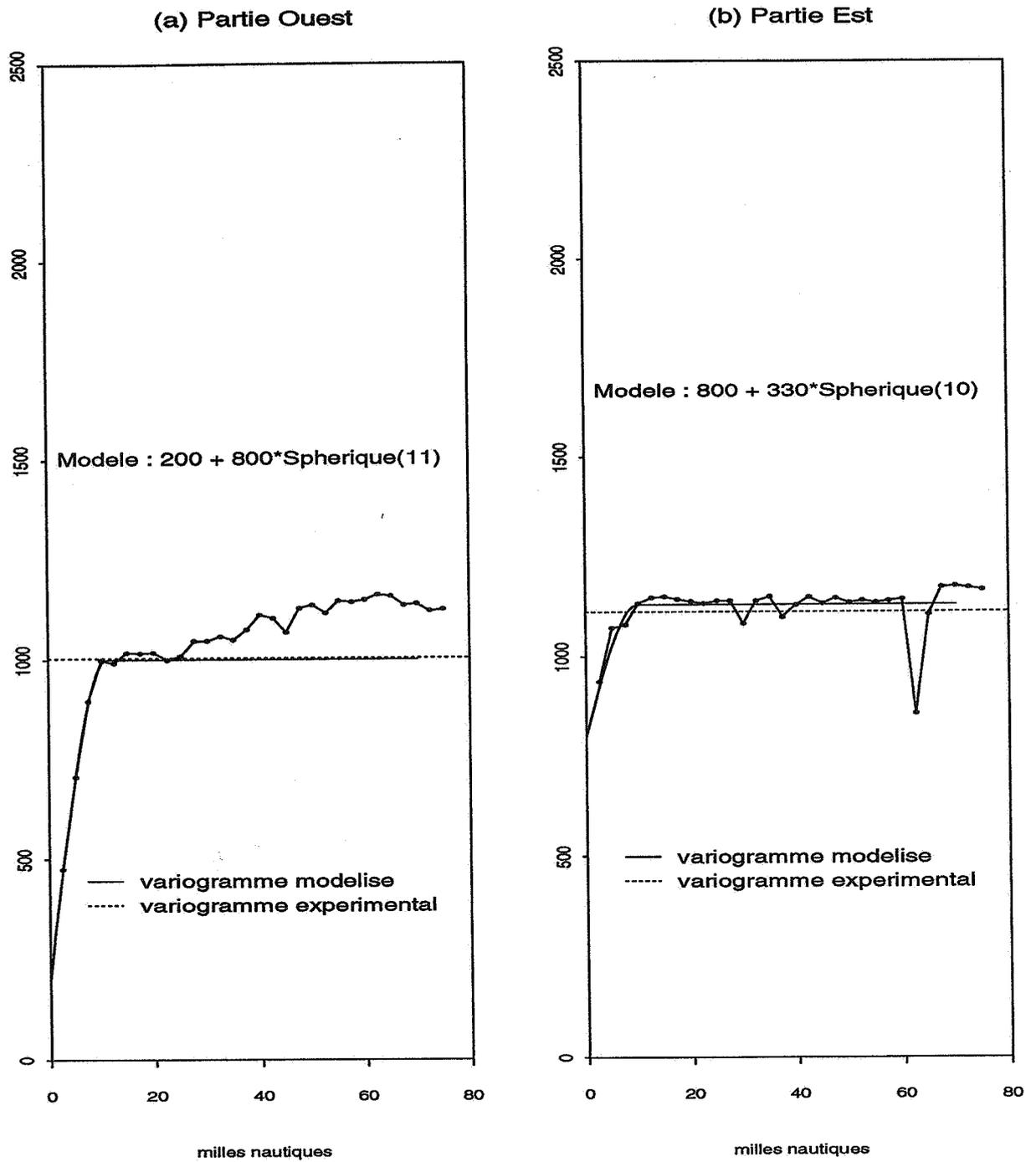


FIG. 1.24 - Variogramme expérimental calculé le long de la route du navire et ajustement - Campagne 1993

Chapitre 2

Figures illustrant la partie II

Campagne 1993

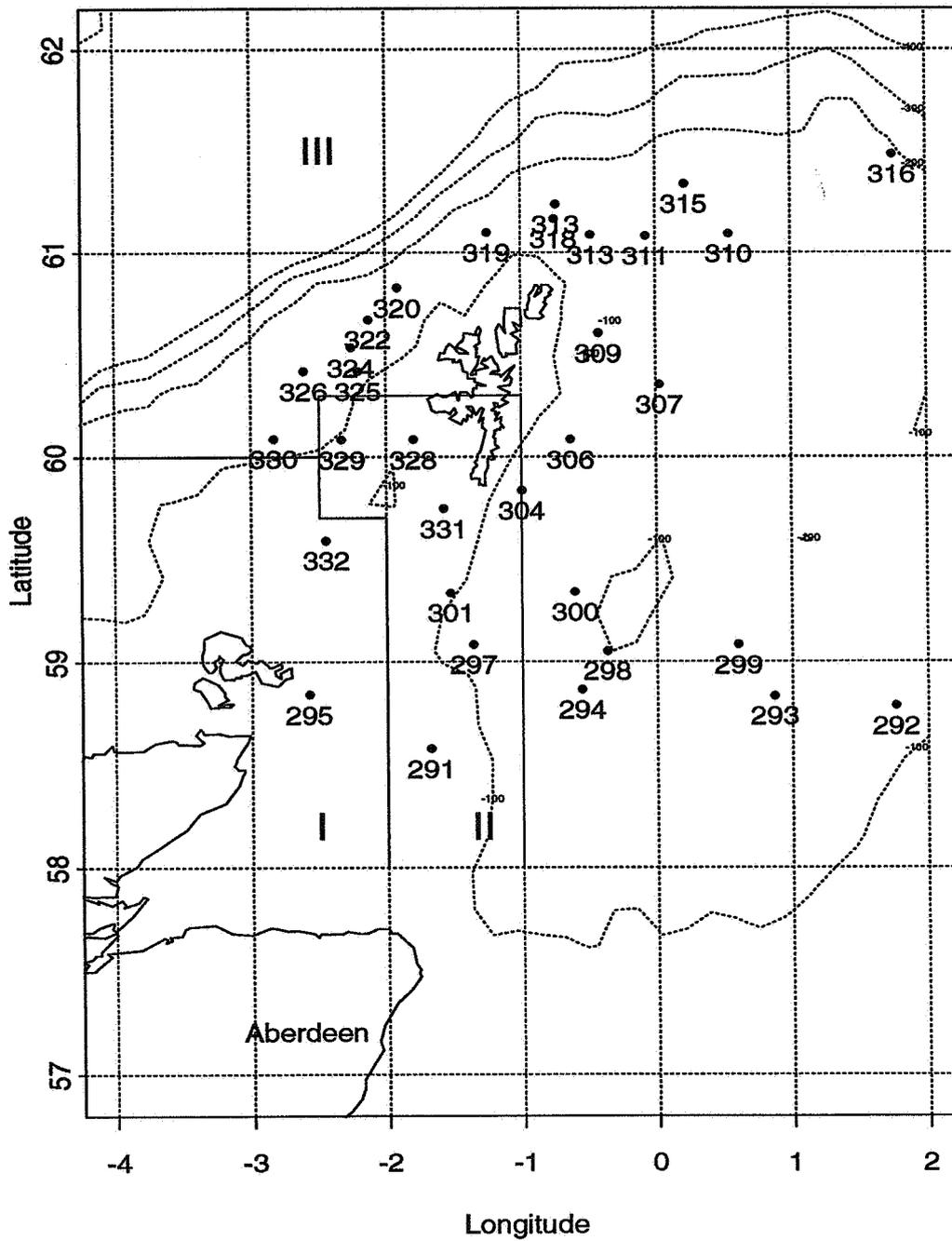


FIG. 2.1 - Stations d'échantillonnage - campagne 1993 et (post-)stratification

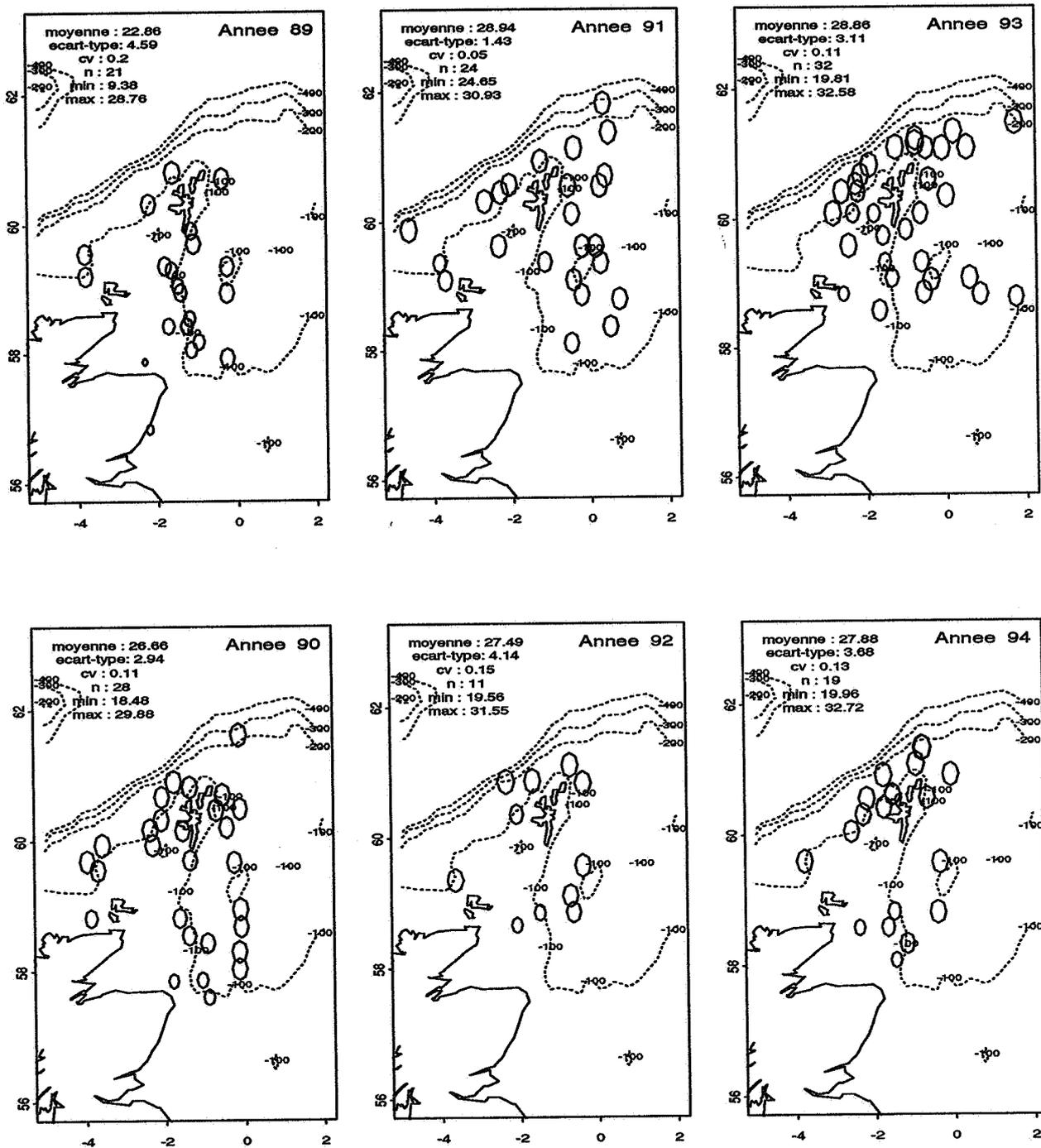


FIG. 2.2 - Représentation proportionnelle de la distribution des longueurs moyennes

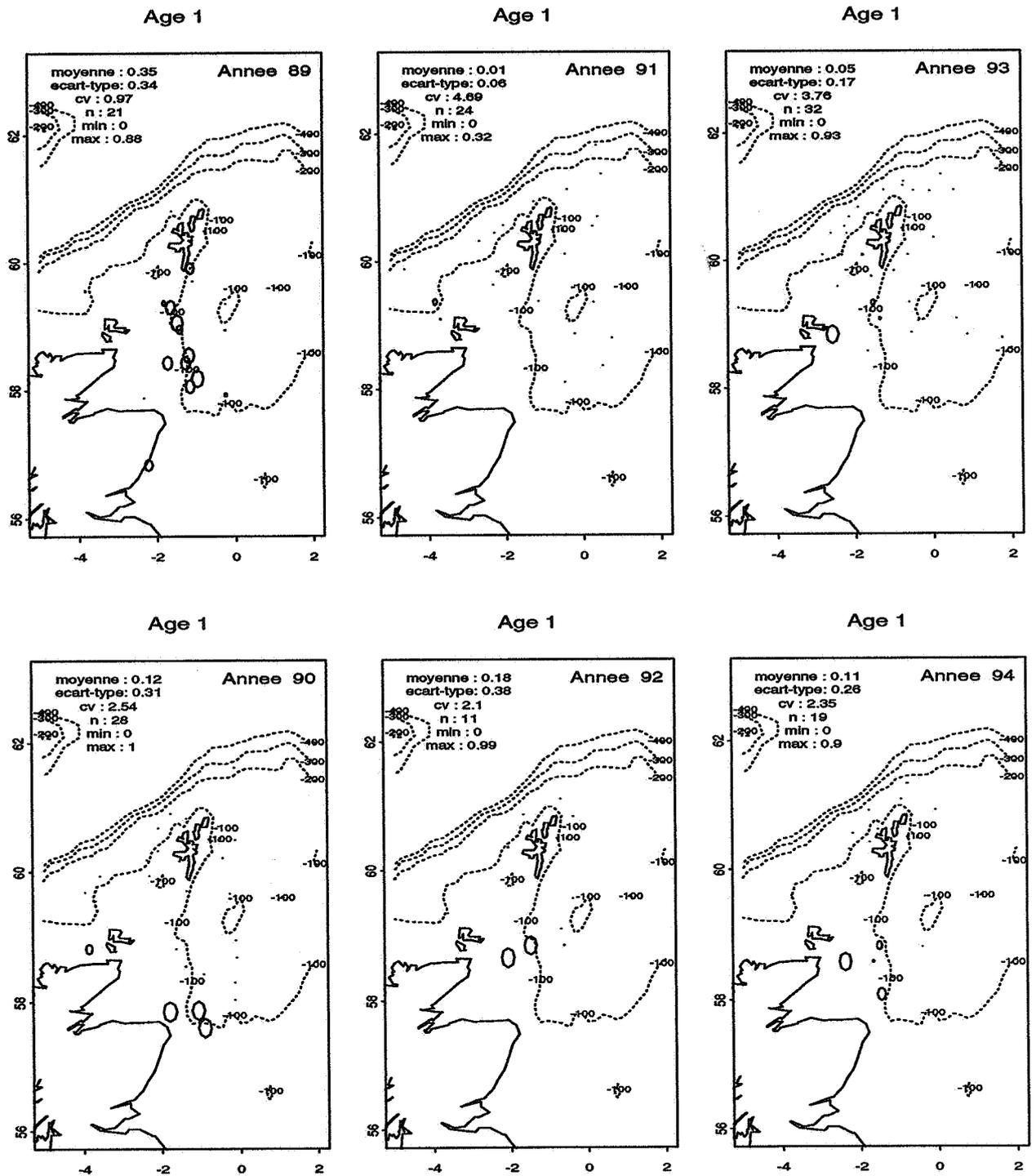


FIG. 2.3 - Représentation proportionnelle de la distribution de la classe d'âge 1

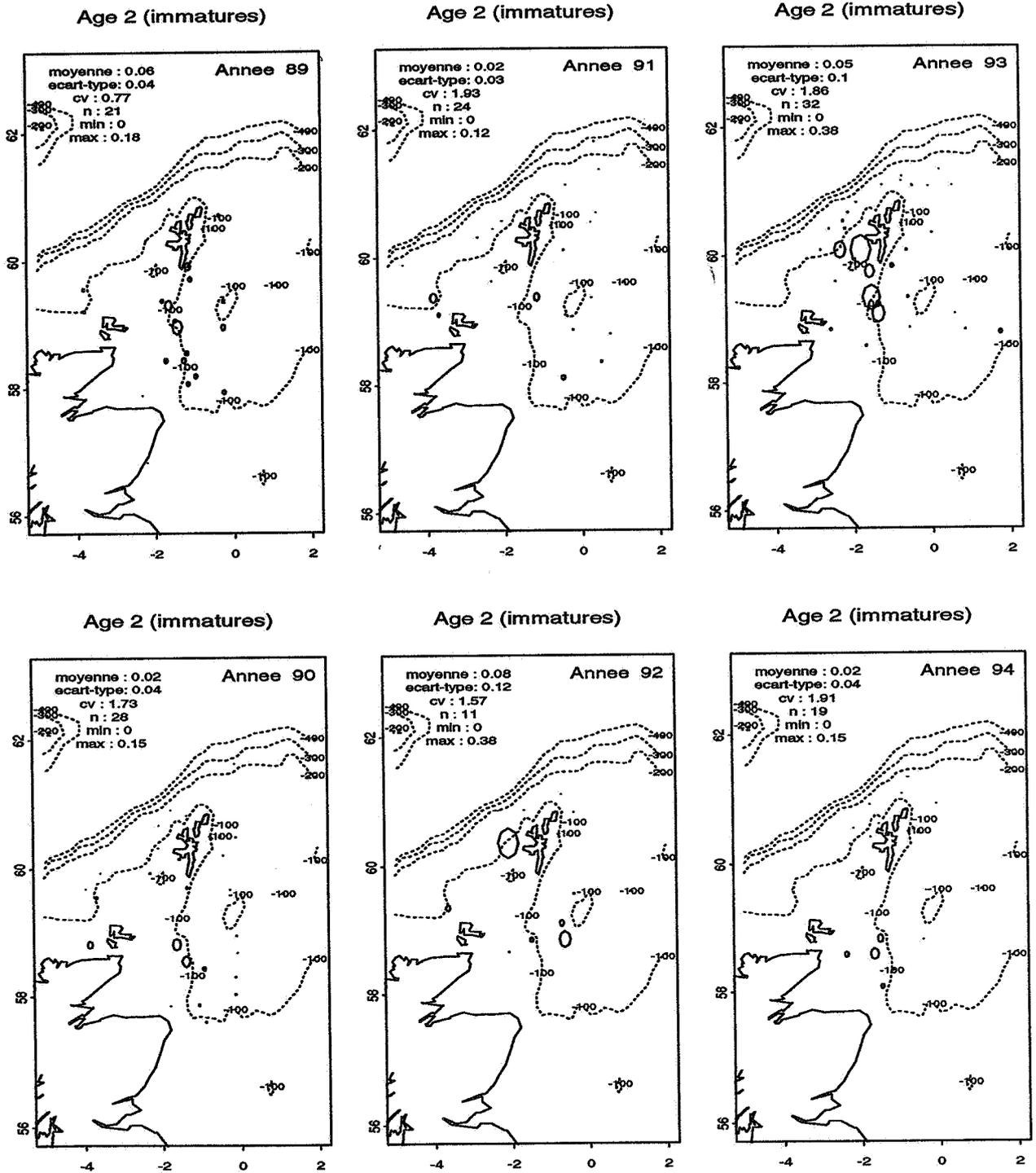


FIG. 2.4 - Représentation proportionnelle de la distribution de la classe d'âge 2 (immatures)

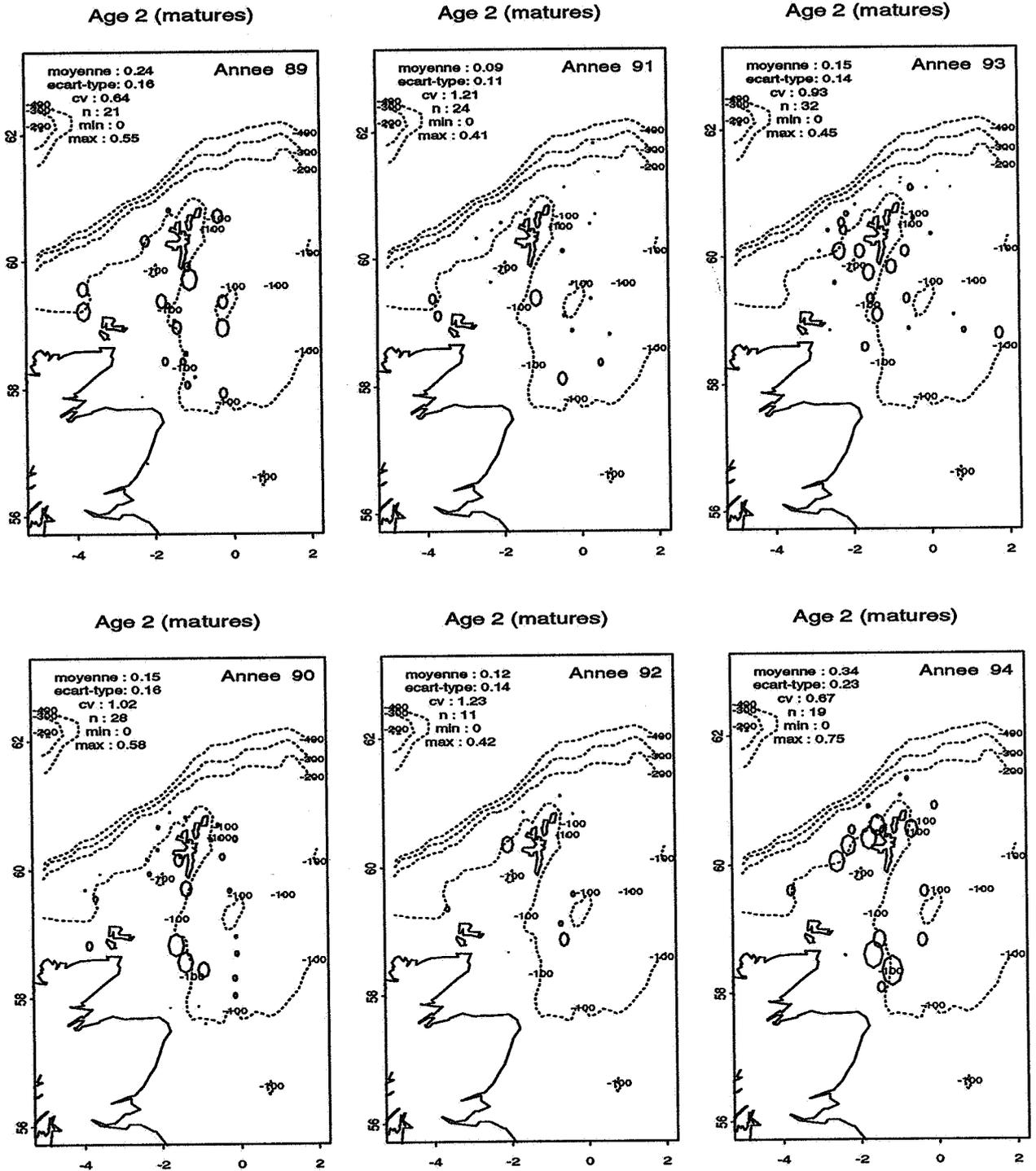


FIG. 2.5 - Représentation proportionnelle de la distribution de la classe d'âge 2 (matures)

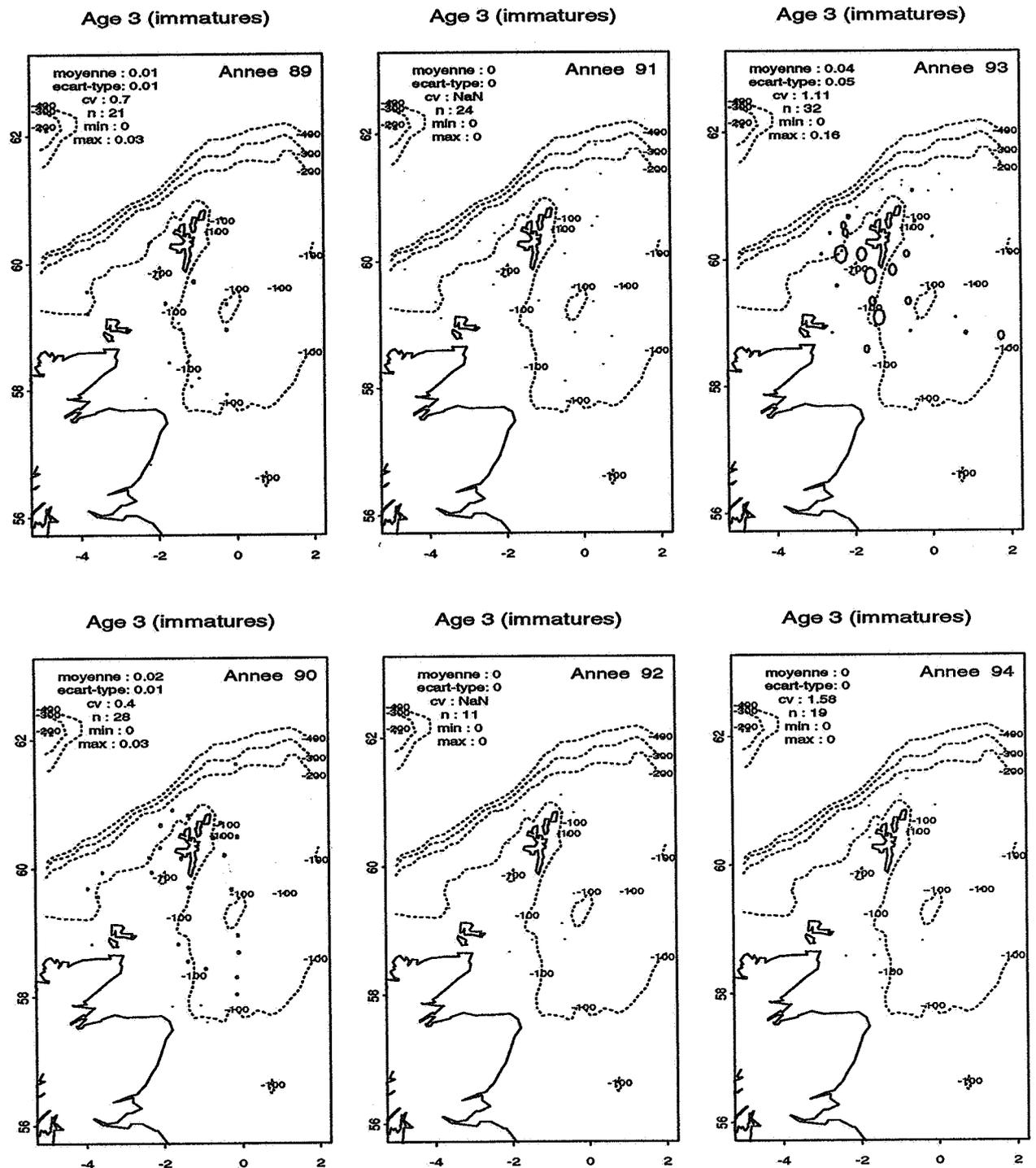


FIG. 2.6 - Représentation proportionnelle de la distribution de la classe d'âge 3 (immatures)

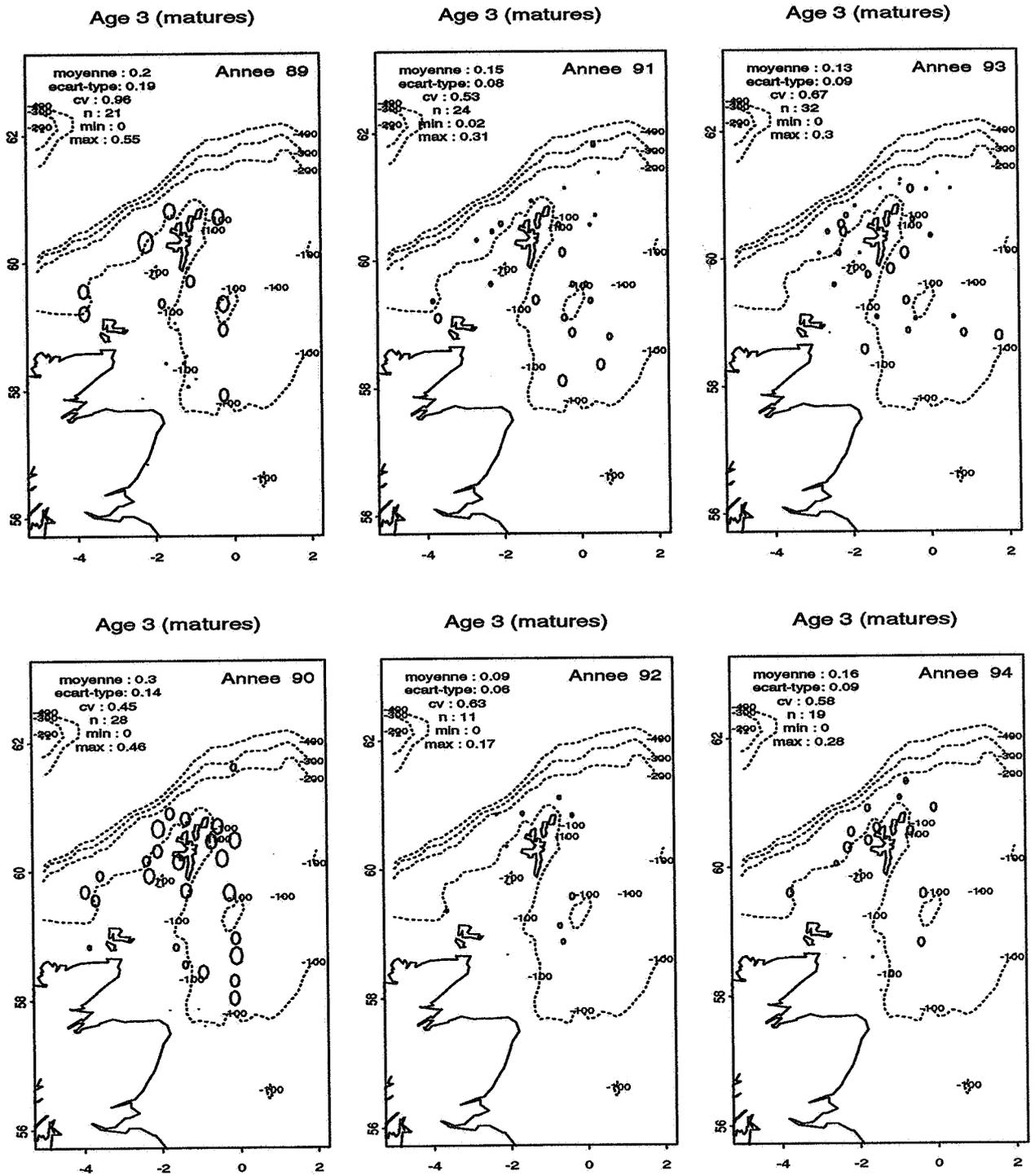


FIG. 2.7 - Représentation proportionnelle de la distribution de la classe d'âge 3

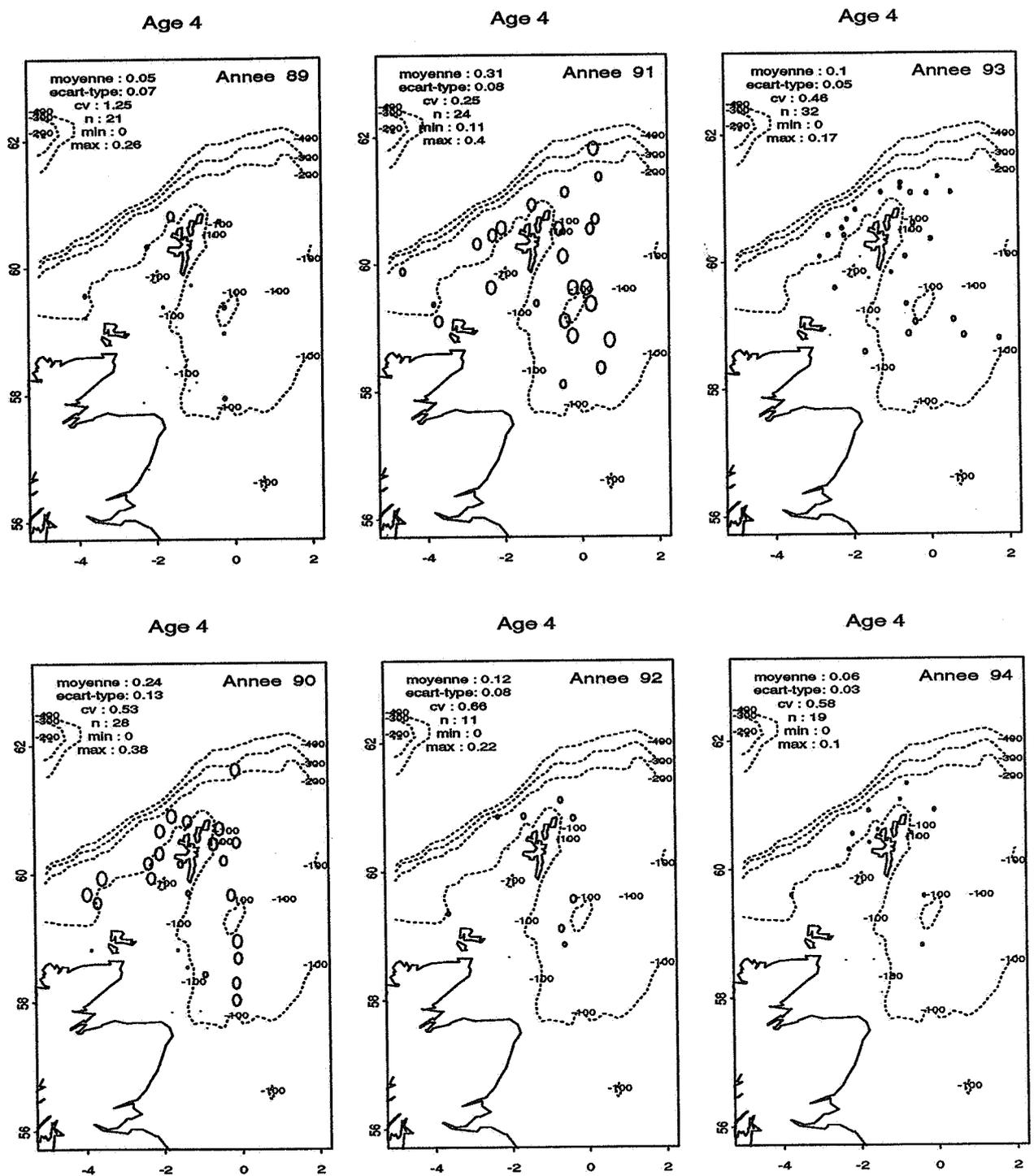


FIG. 2.8 - Représentation proportionnelle de la distribution de la classe d'âge 4

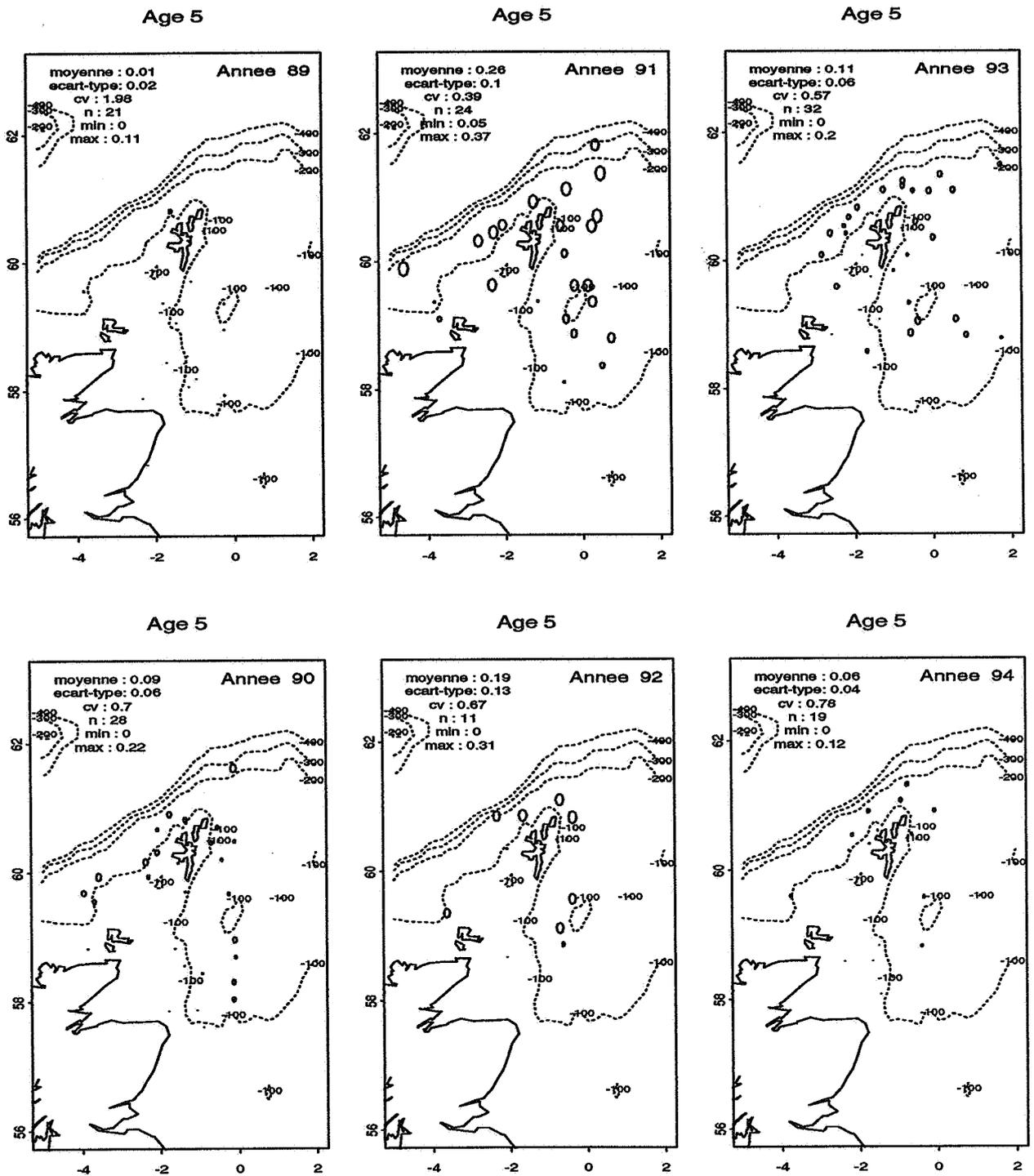


FIG. 2.9 - Représentation proportionnelle de la distribution de la classe d'âge 5

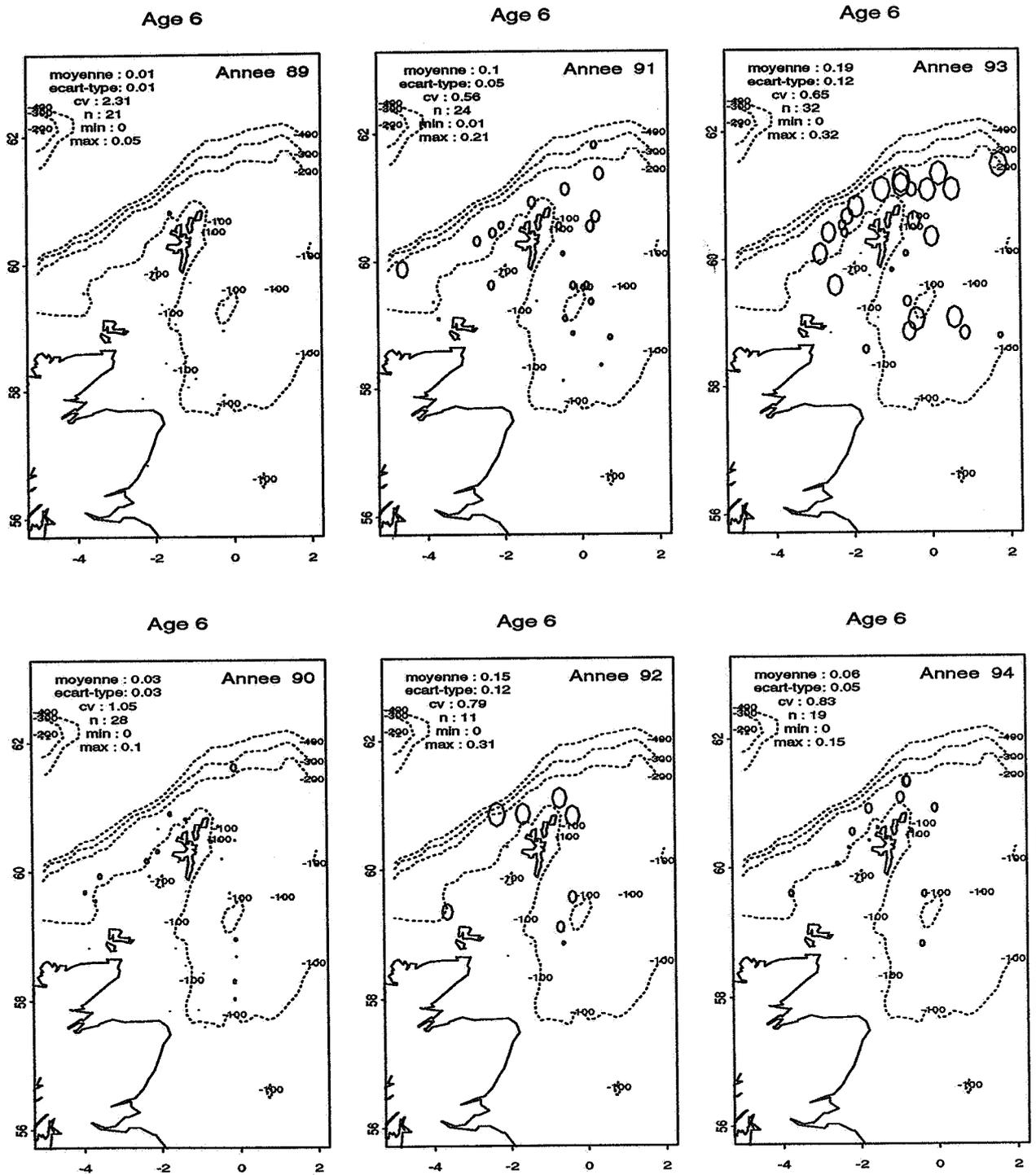


FIG. 2.10 - Représentation proportionnelle de la distribution de la classe d'âge 6

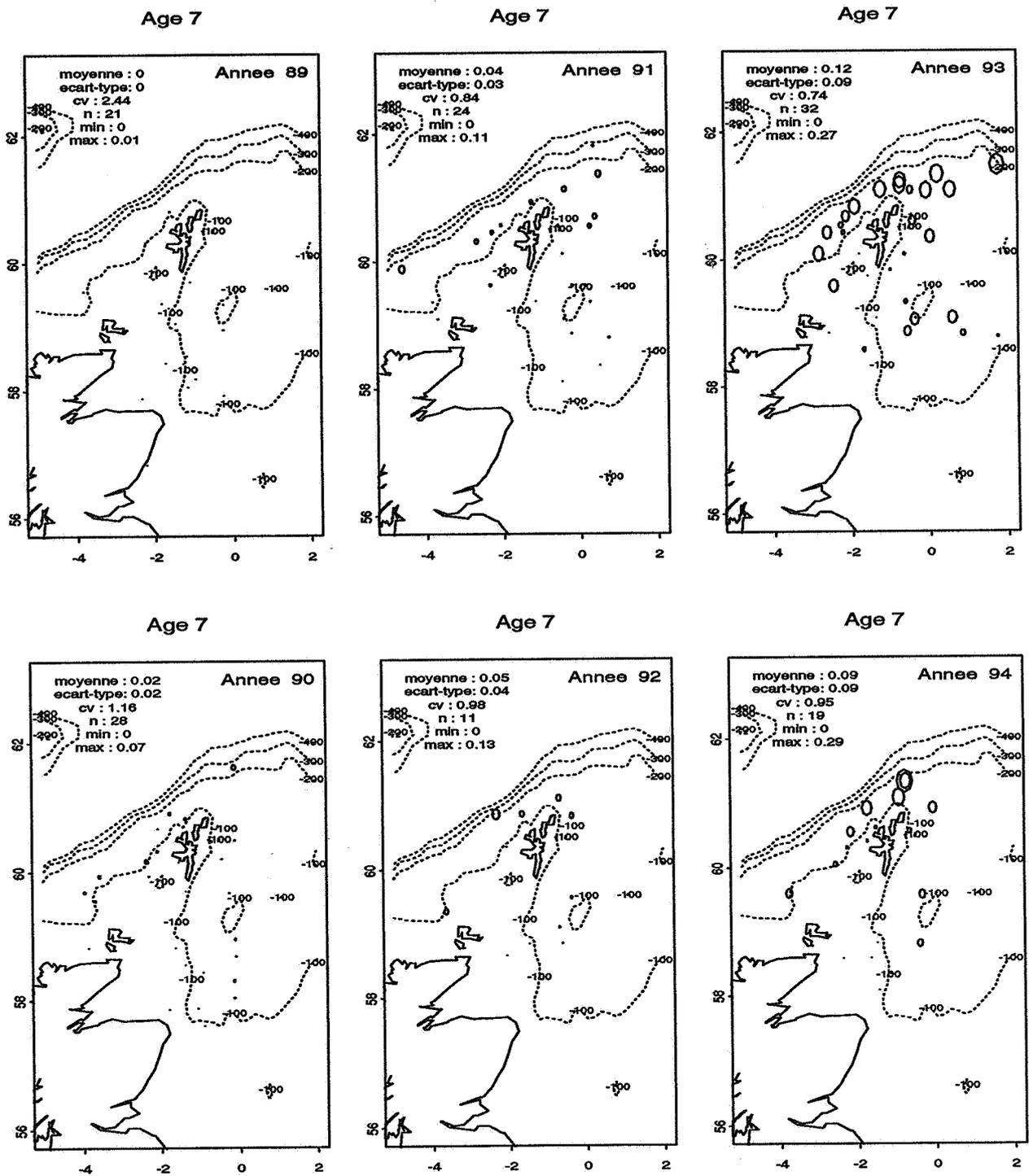


FIG. 2.11 - Représentation proportionnelle de la distribution de la classe d'âge 7

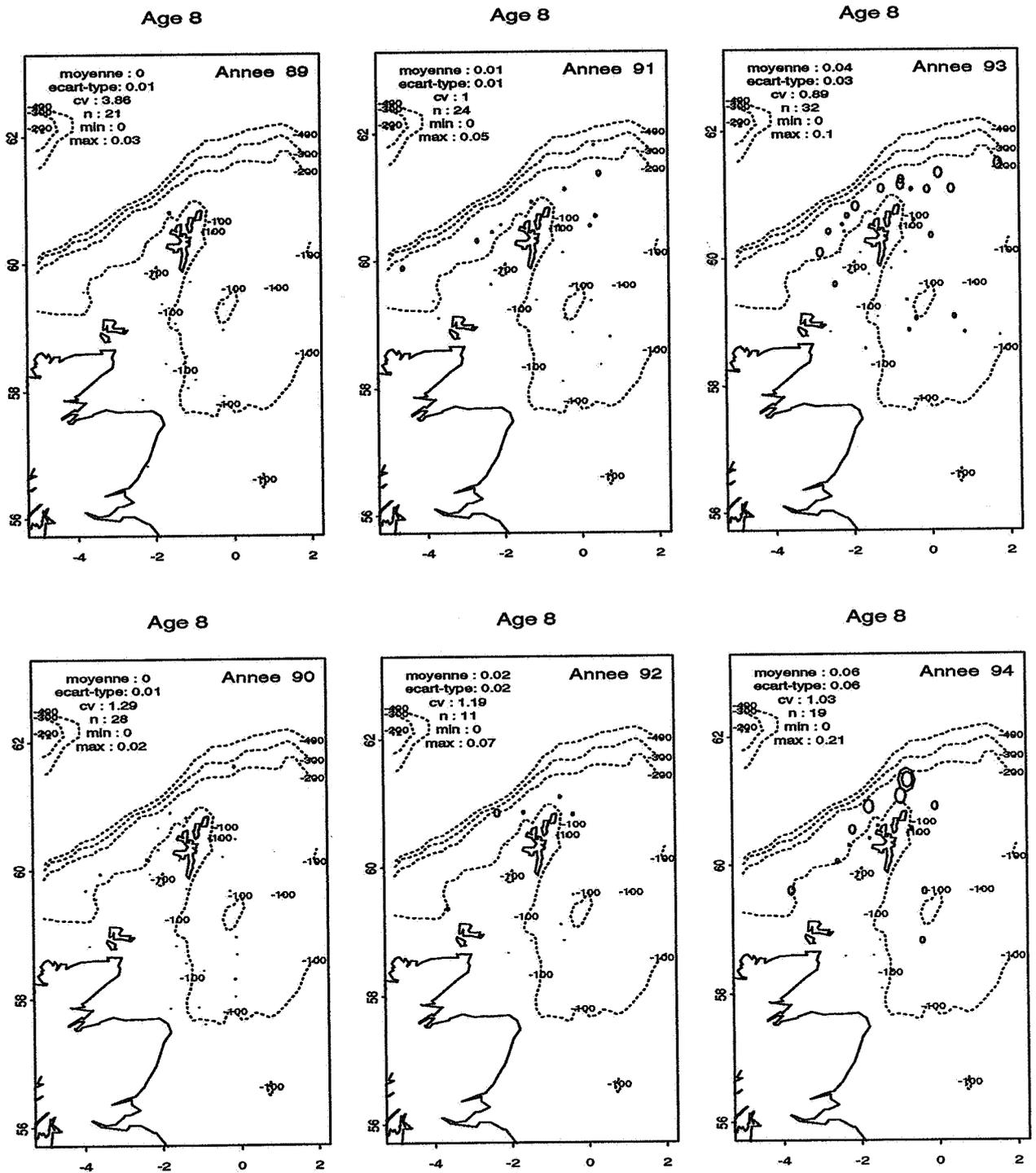


FIG. 2.12 - Représentation proportionnelle de la distribution de la classe d'âge 8

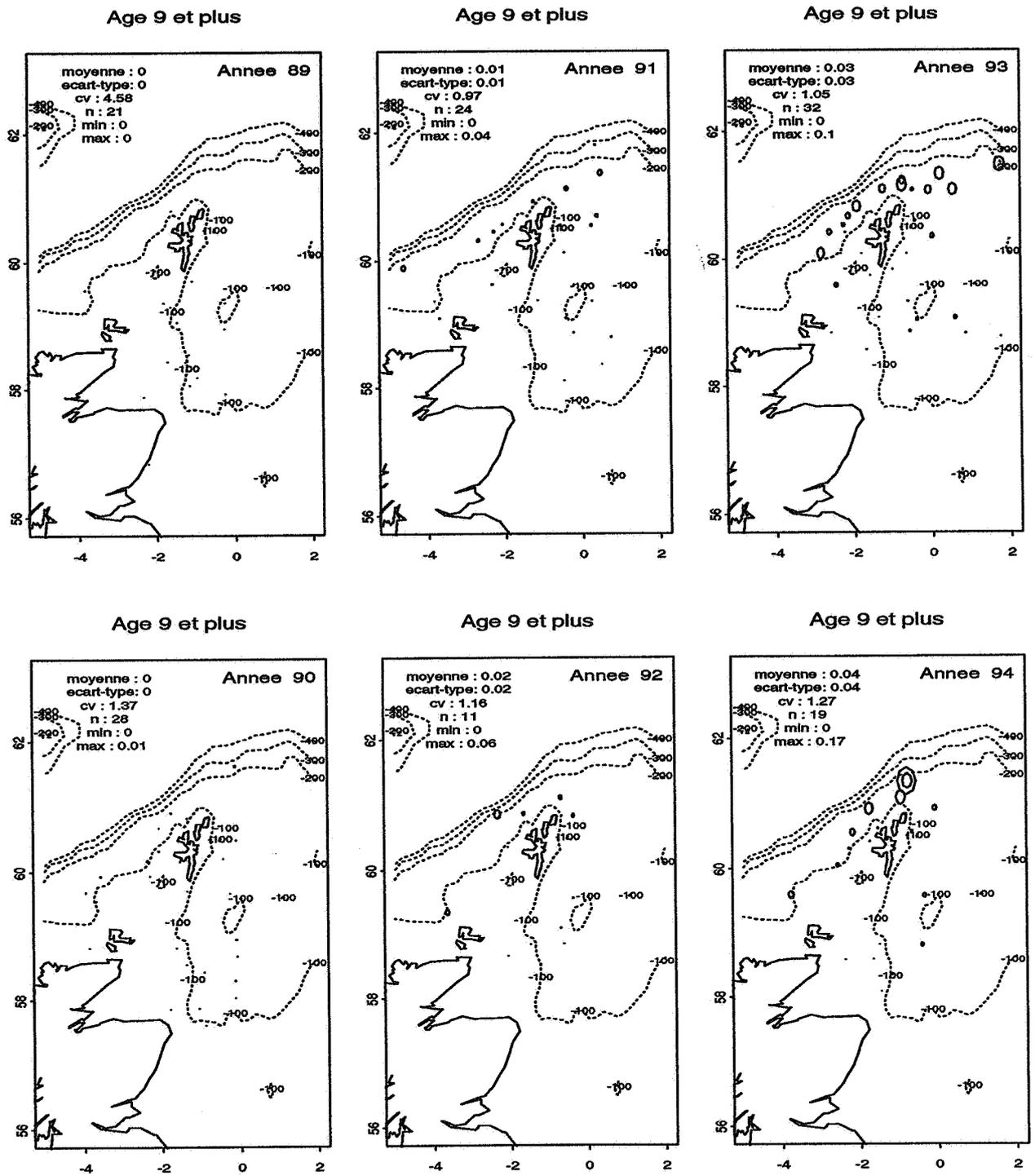


FIG. 2.13 - Représentation proportionnelle de la distribution de la classe d'âge 9

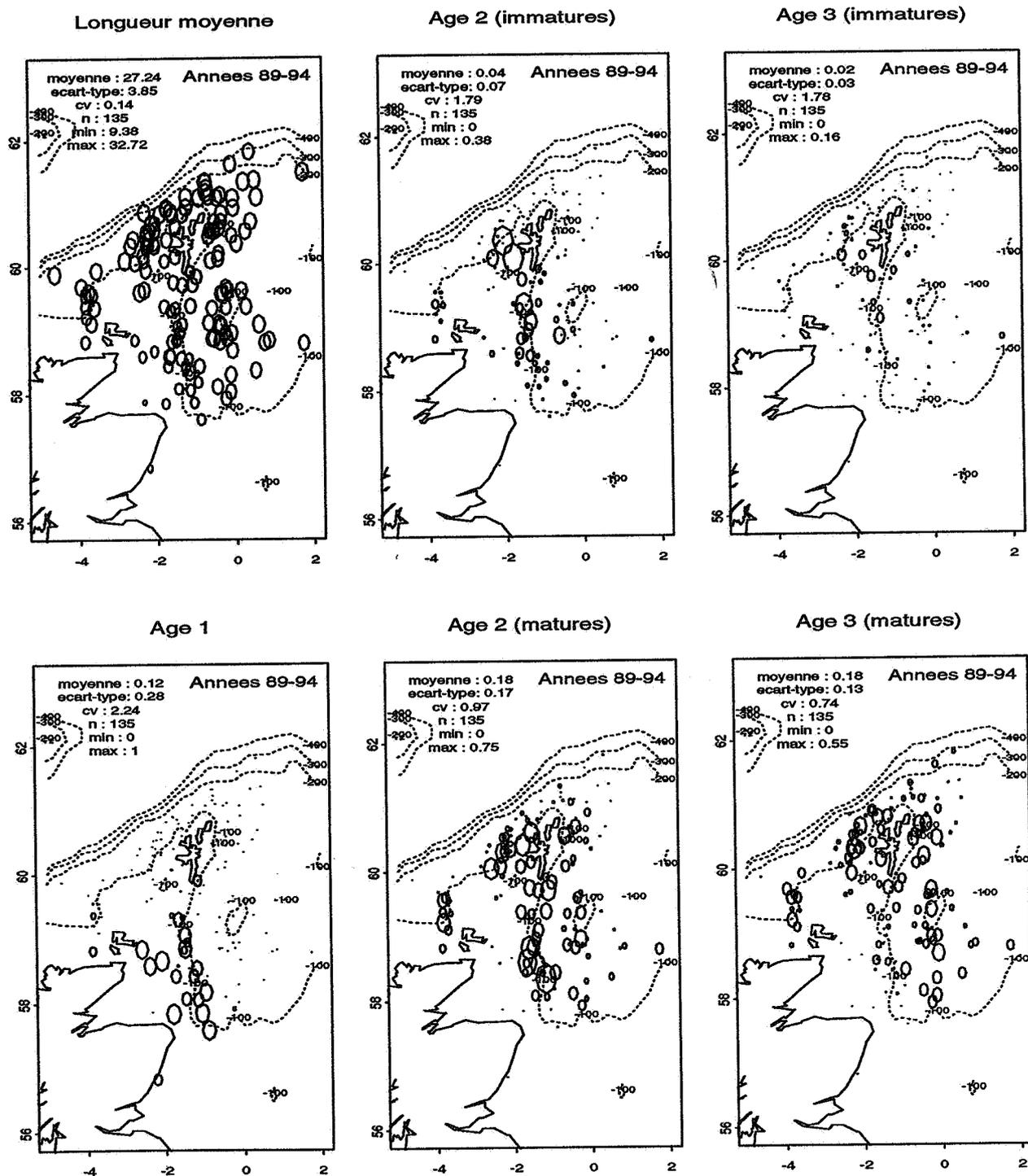


FIG. 2.14 - Représentation proportionnelle de la distribution des longueurs et des âges pour toutes les années

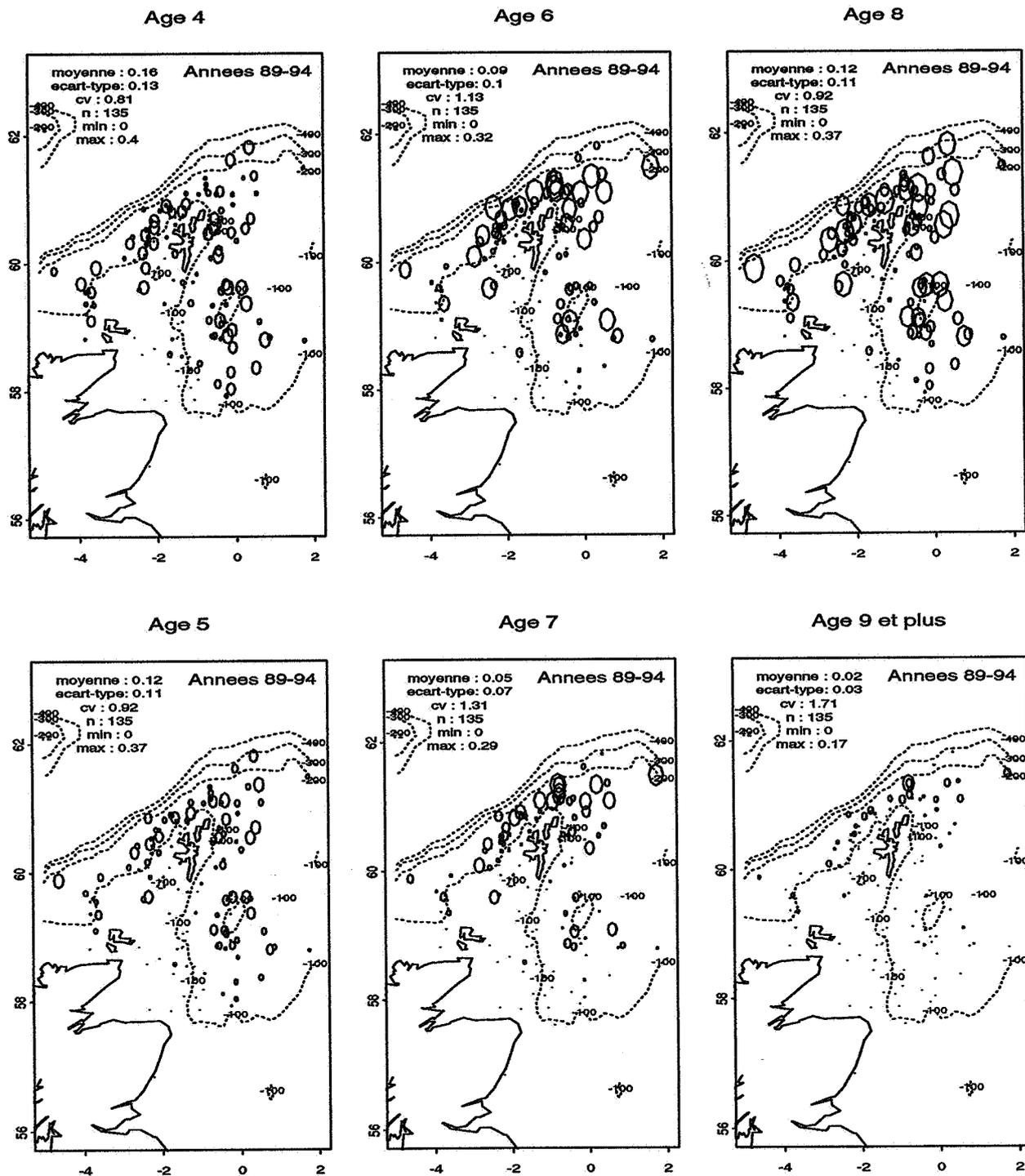


FIG. 2.15 - Représentation proportionnelle de la distribution des longueurs et des âges pour toutes les années (suite)

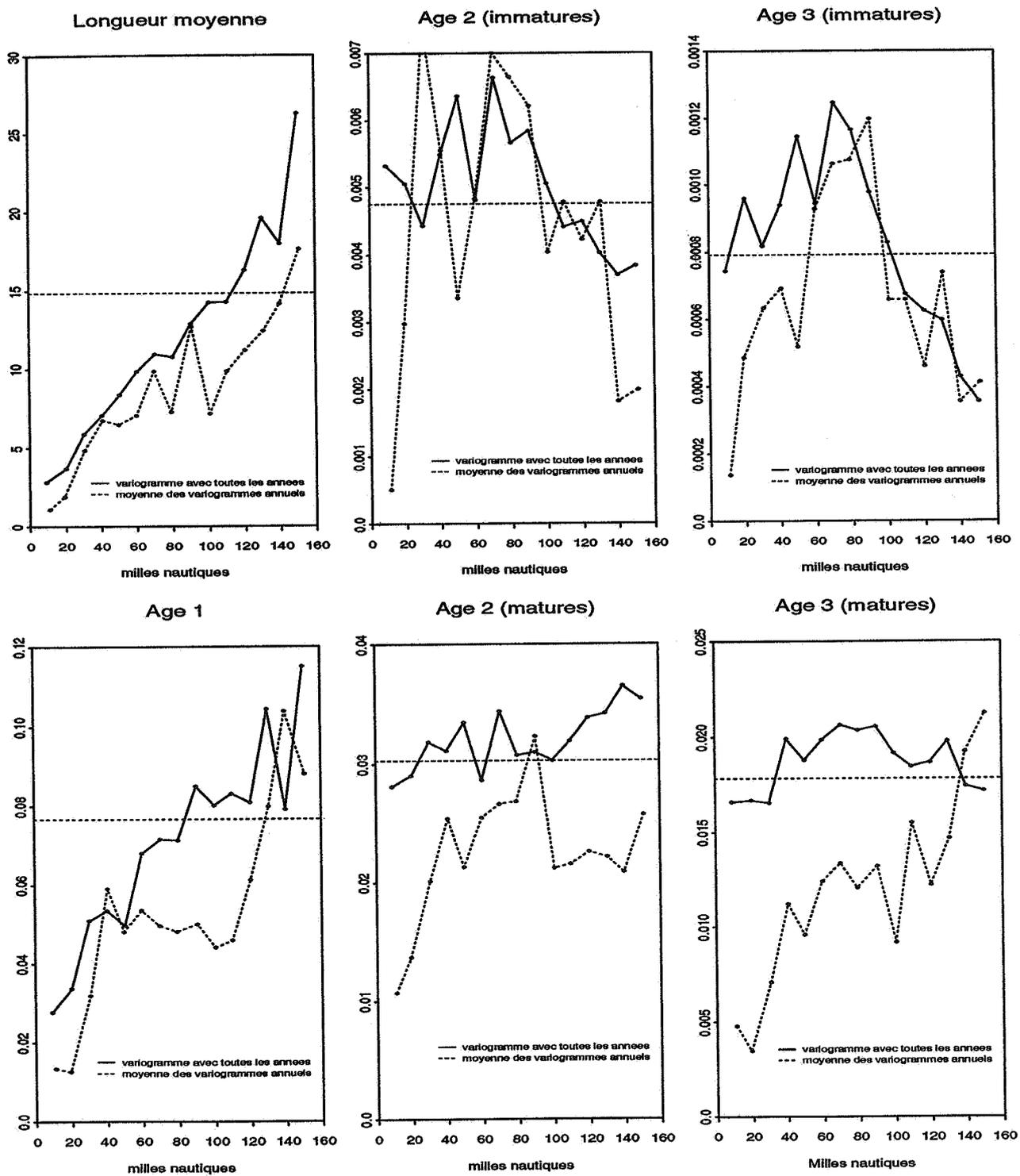


FIG. 2.16 - Variogrammes de la longueur moyenne et des proportions d'âge

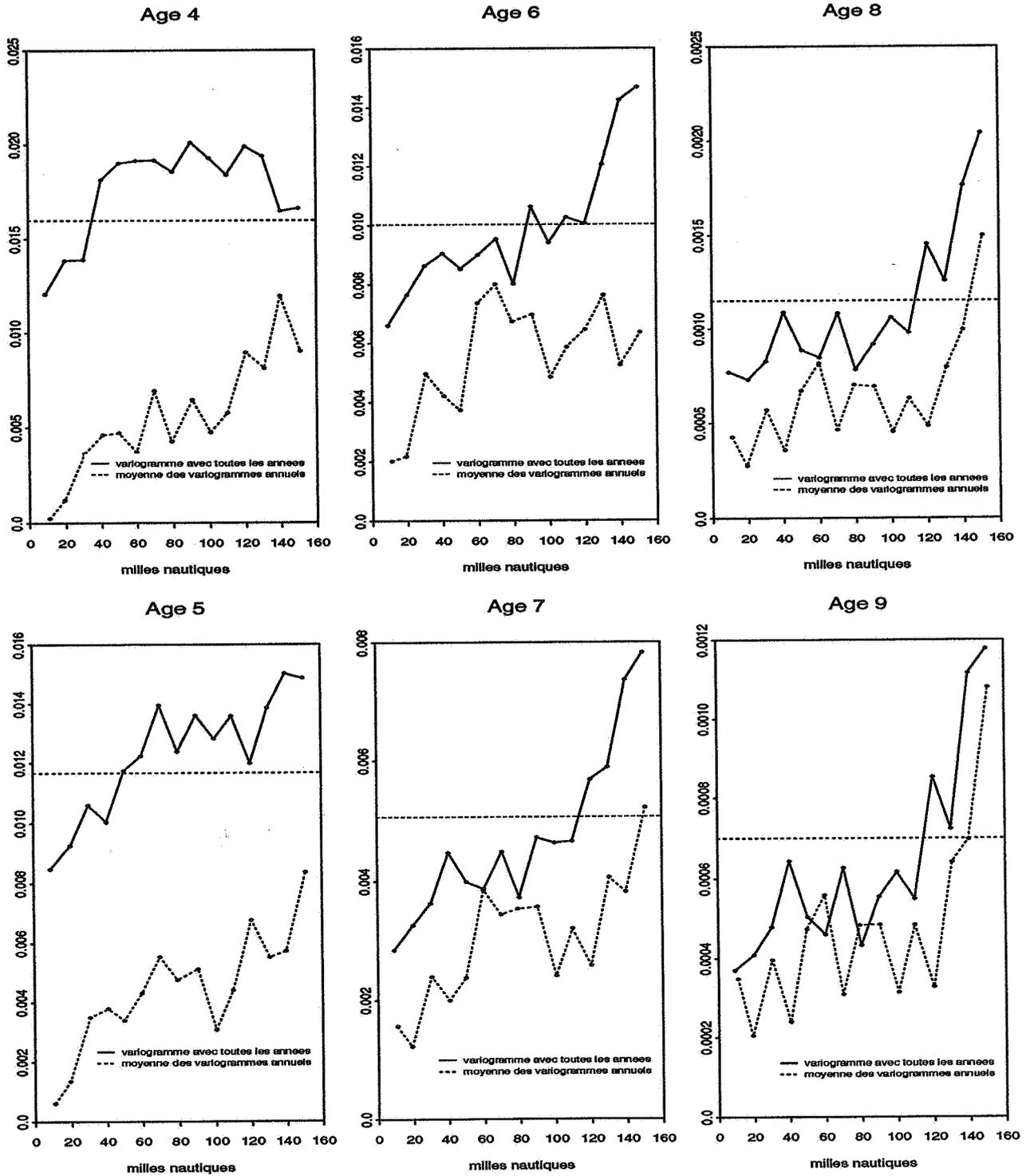


FIG. 2.17 - Variogrammes de la longueur moyenne et des proportions d'âge (suite)

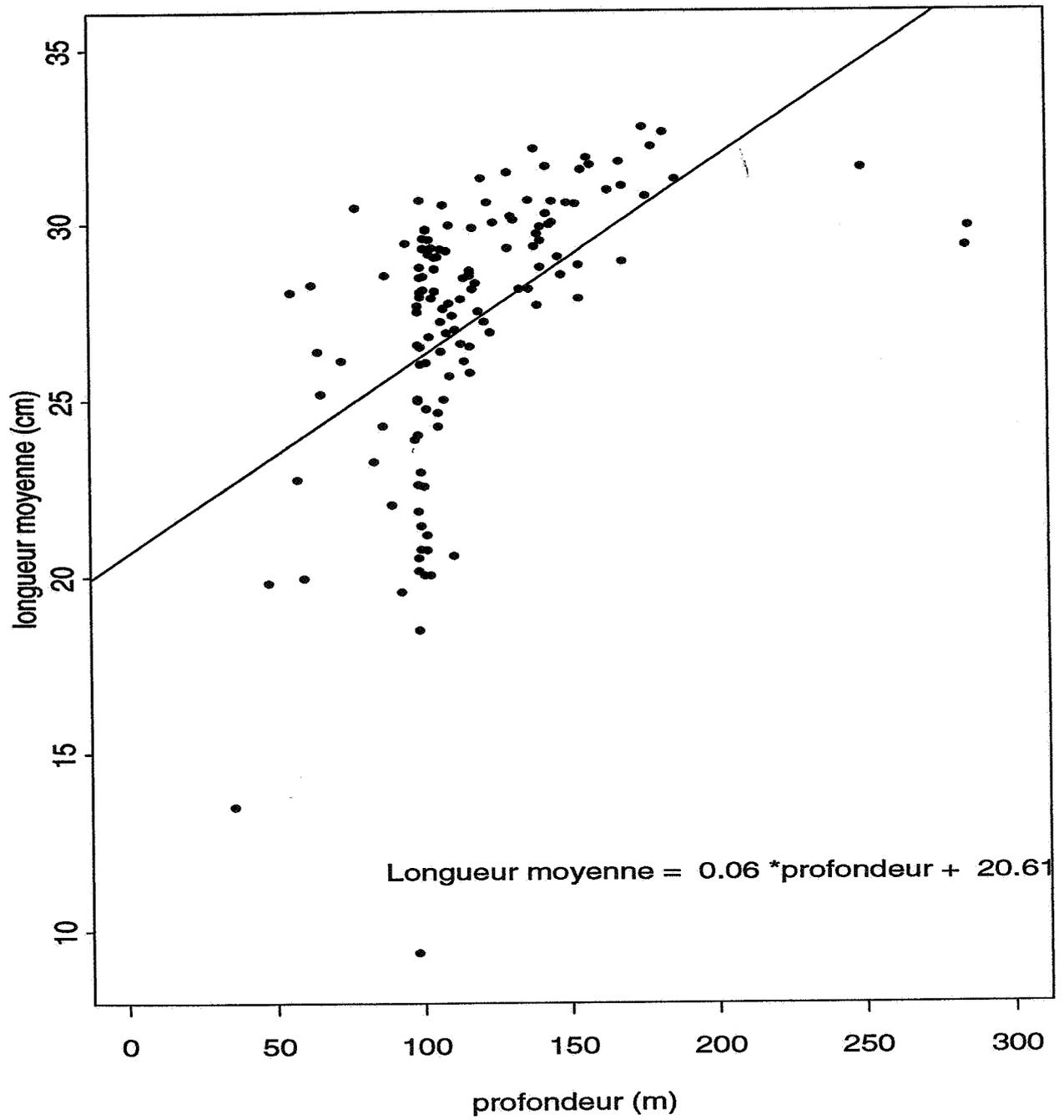


FIG. 2.18 - Relation longueur moyenne/profondeur (toutes les années)

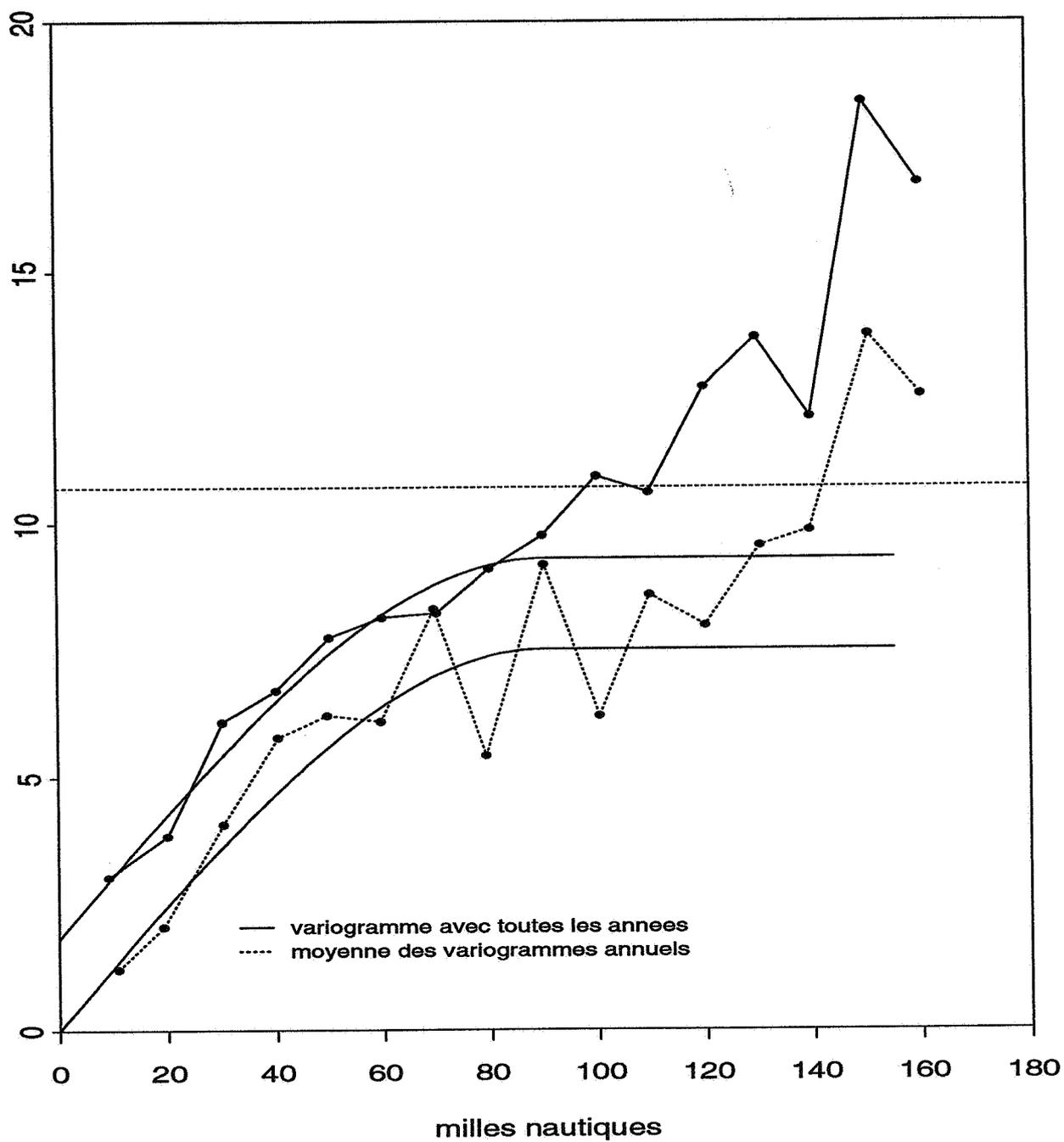


FIG. 2.19 - Variogrammes des résidus de la régression Longueur/profondeur

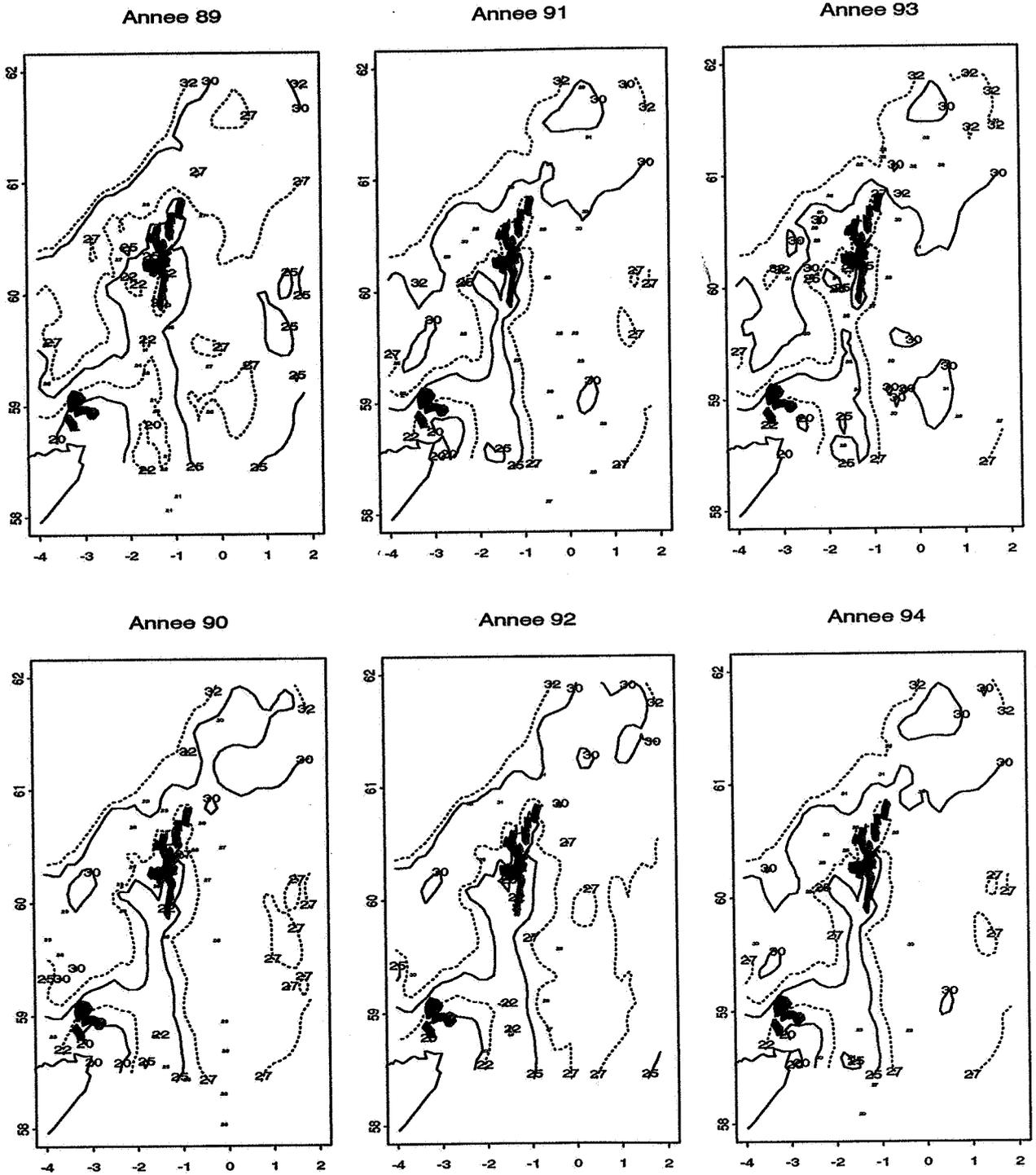


FIG. 2.20 - Cartographies de la longueur sur une grille 5 mn x 5 mn (krigeage avec dérive externe)

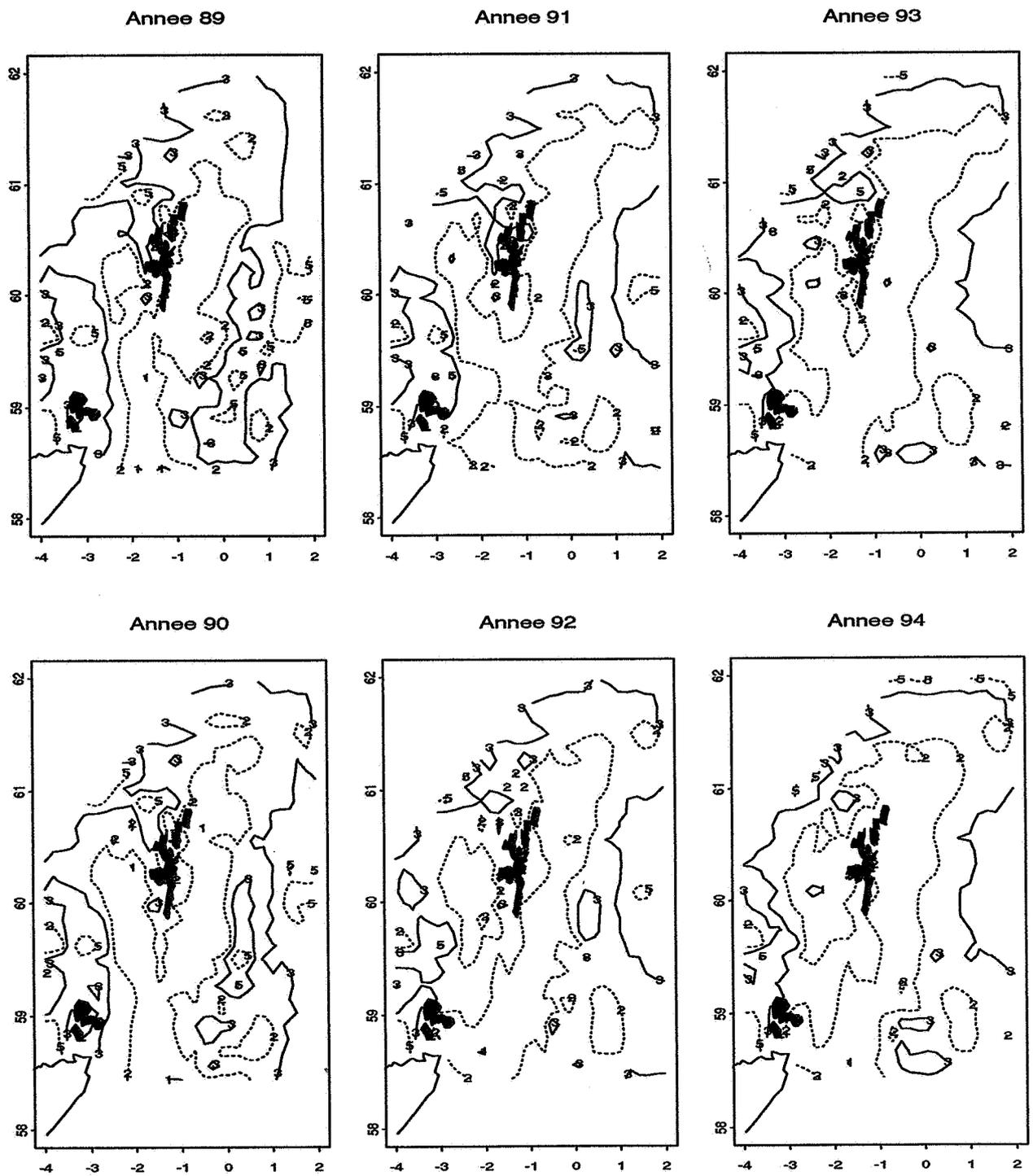


FIG. 2.21 - Krigage de la longueur: cartographies des écarts-types d'estimation

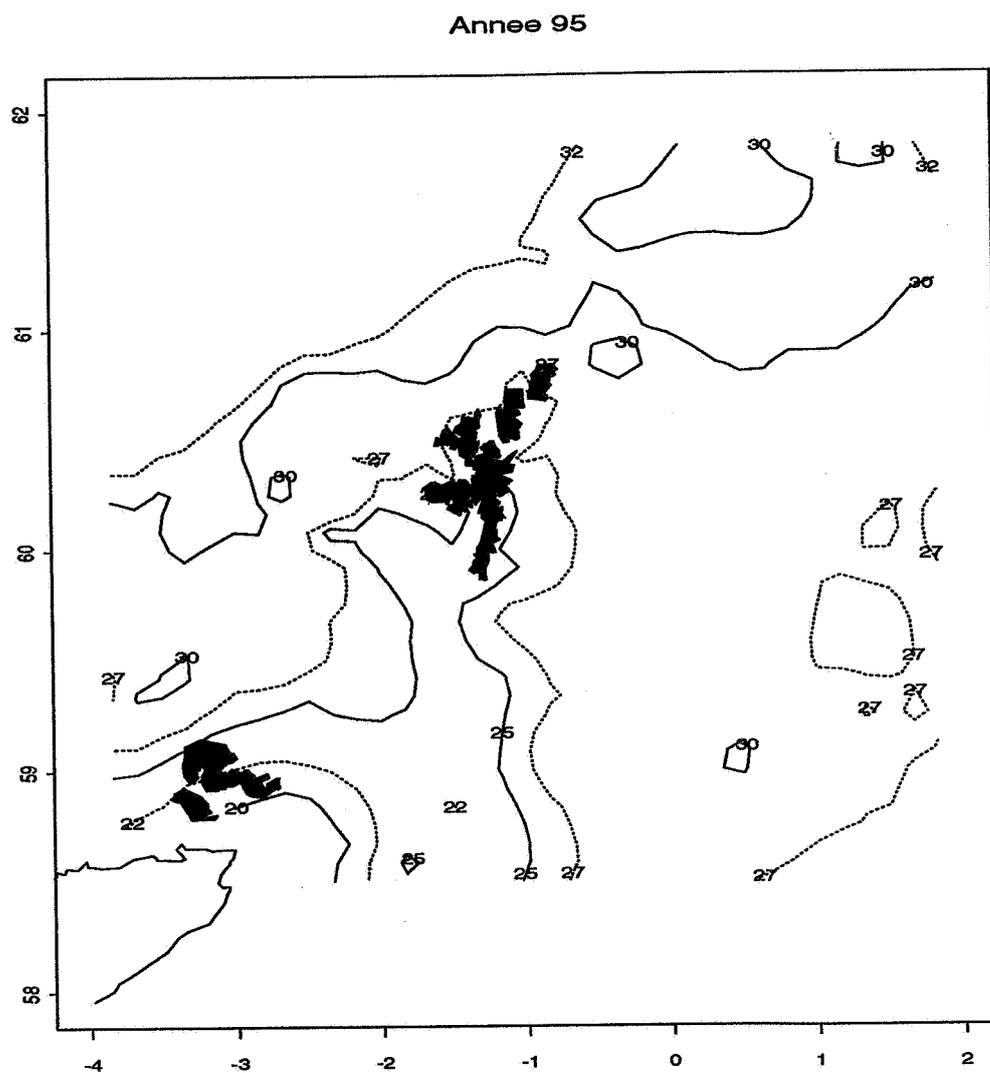


FIG. 2.22 - Cartographie moyenne de la longueur sur une grille 5 mn x 5 mn (krigeage avec dérive externe)

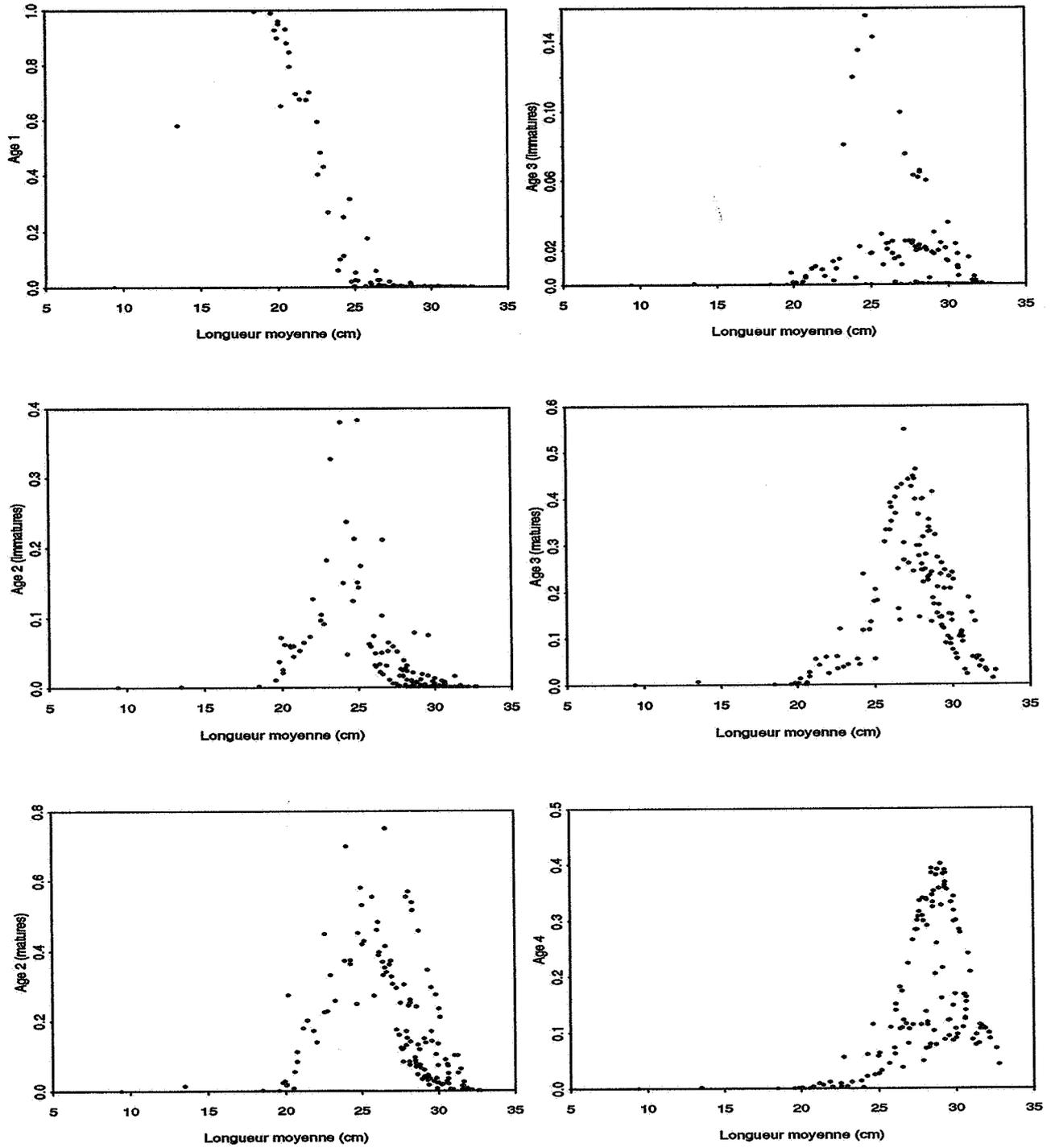


FIG. 2.23 - Nuages de corrélation $\hat{A}ge/Longueur\ moyenne$

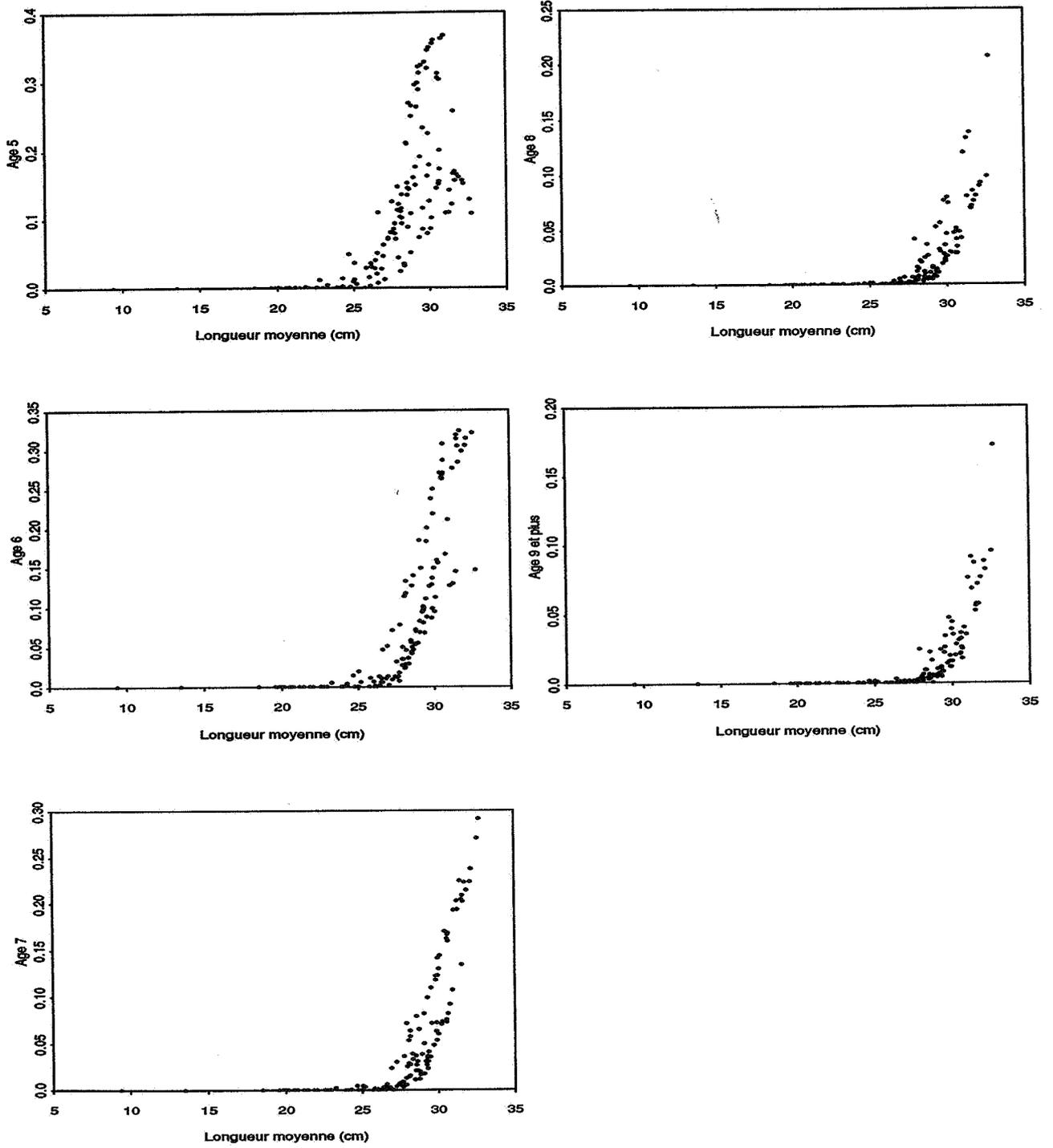


FIG. 2.24 - Nuages de corrélation Âge/Longueur moyenne (suite)

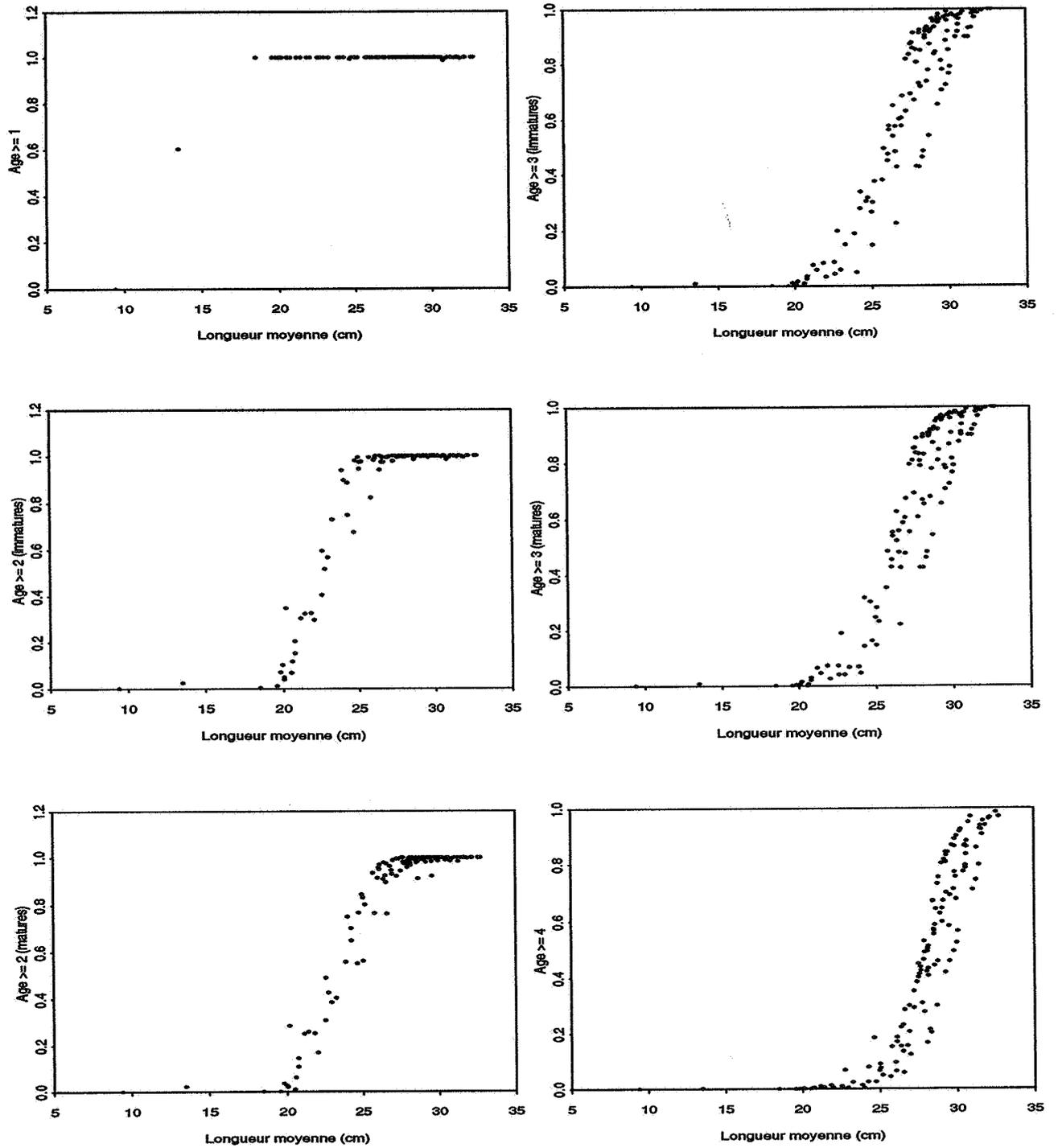


FIG. 2.25 - Nuages de corrélation âges cumulés/Longueur moyenne

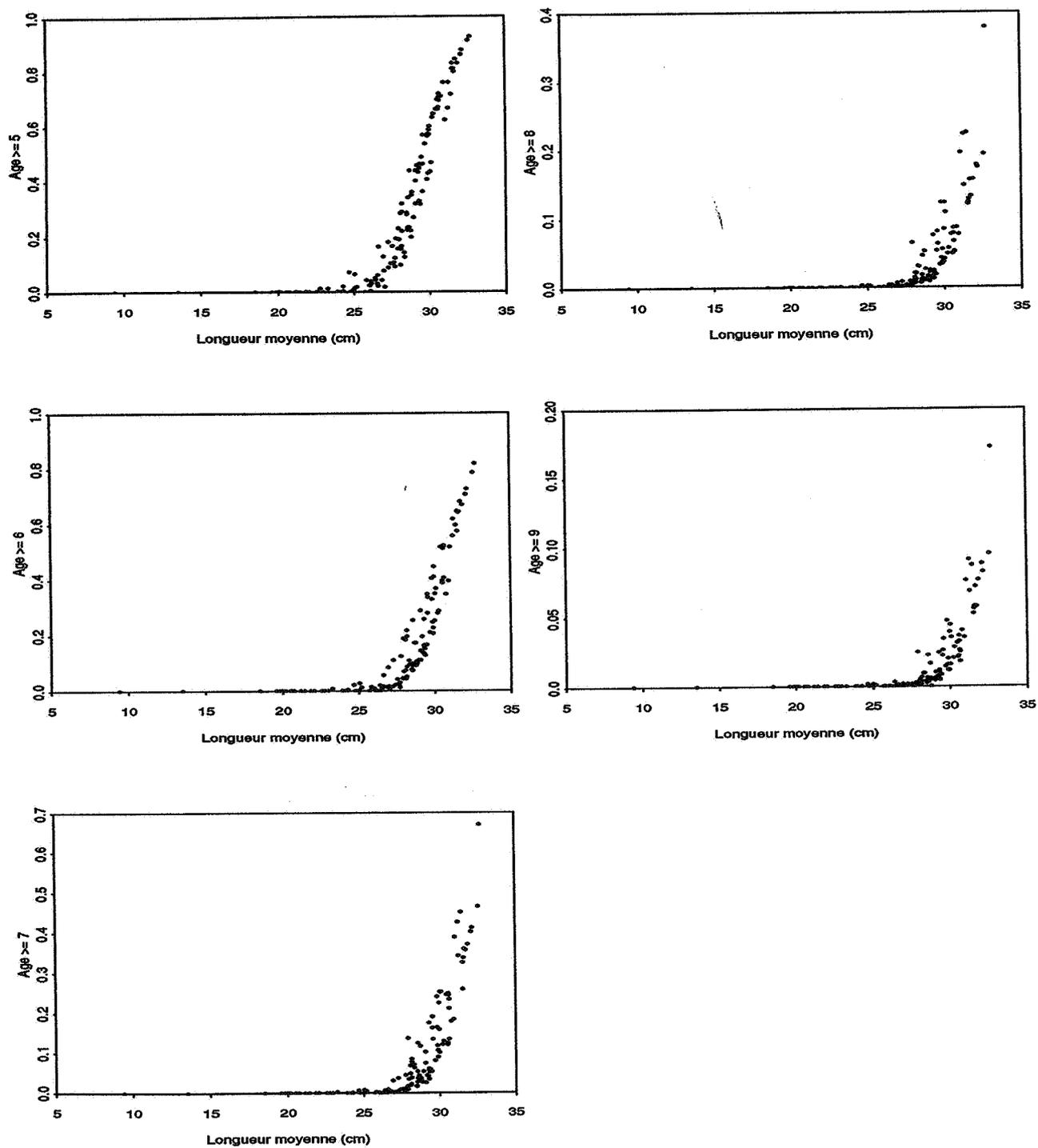


FIG. 2.26 - Nuages de corrélation âges cumulés/Longueur moyenne (suite)

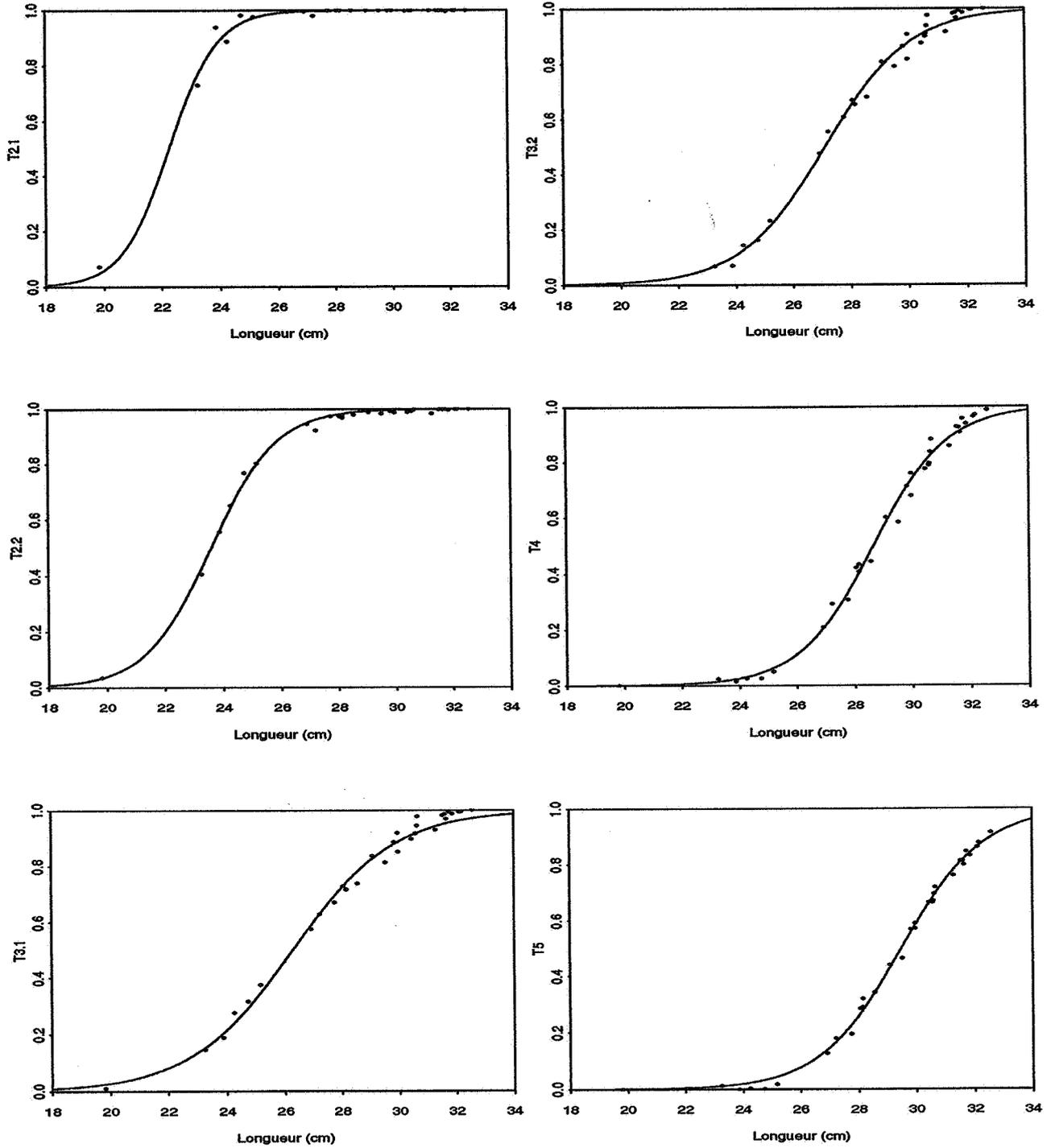


FIG. 2.27 - Régressions âges cumulés/longueur moyenne - Année 1993

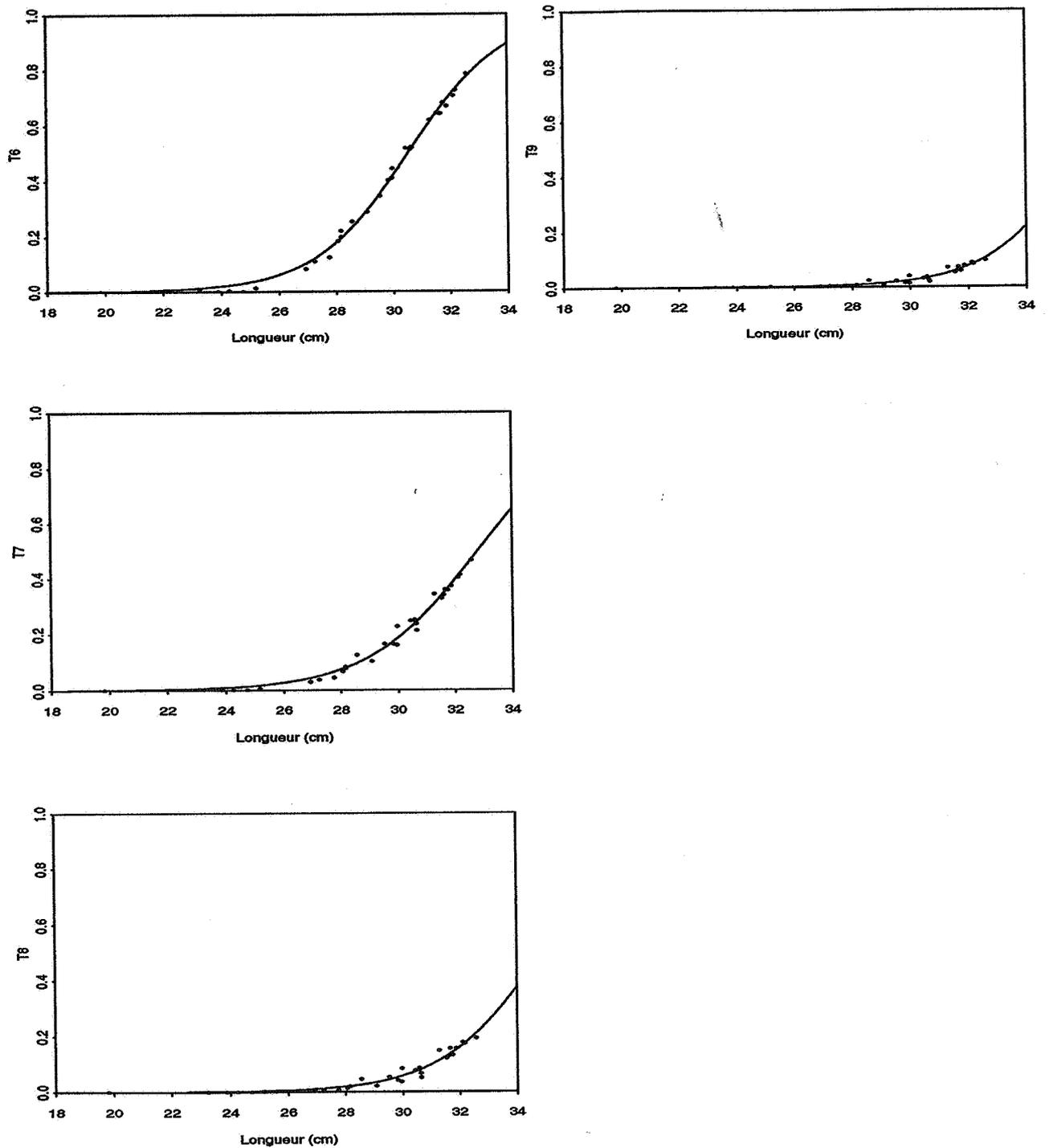


FIG. 2.28 - Régressions âges cumulés/longueur moyenne - Année 1993 (suite)

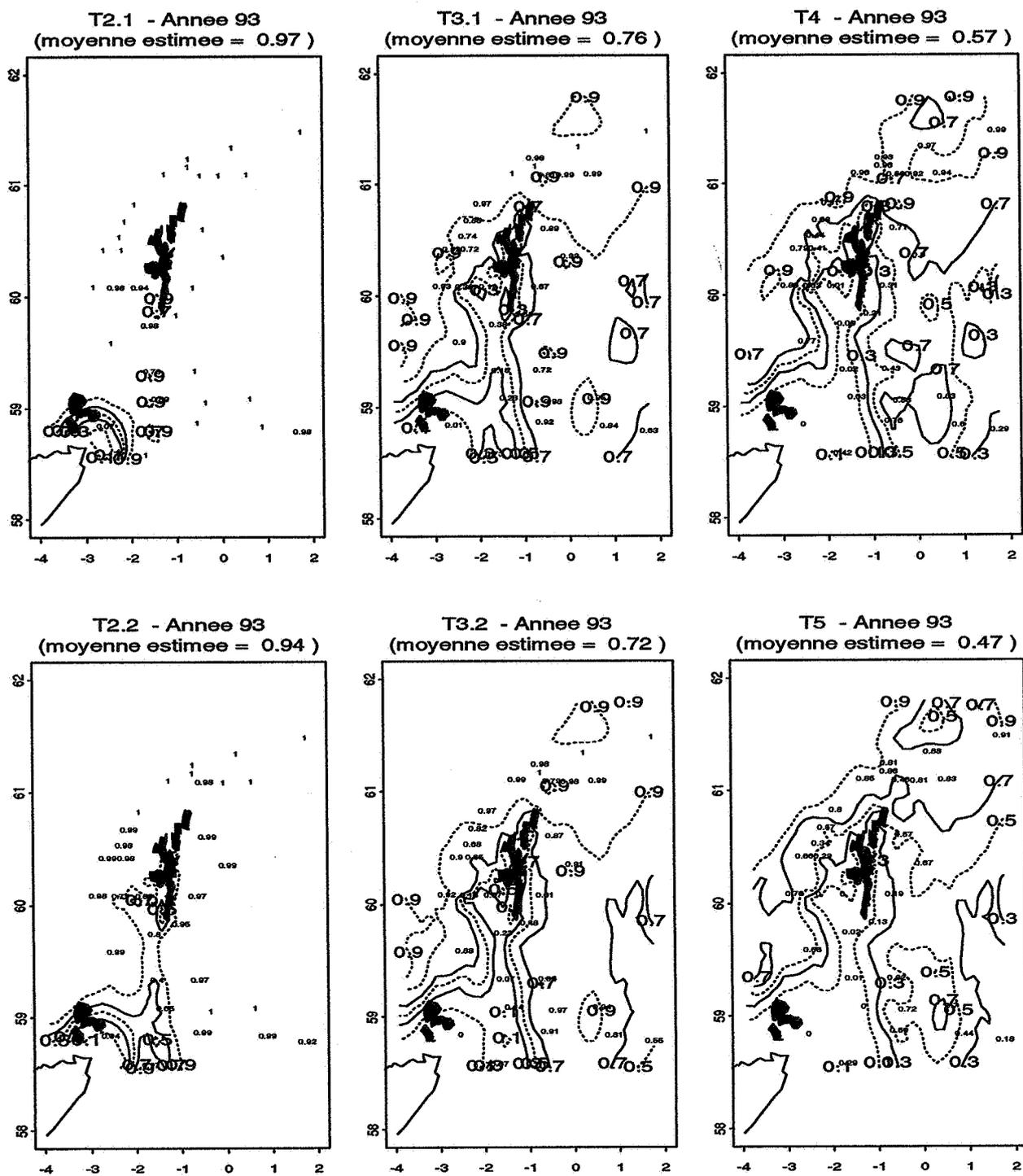


FIG. 2.29 - Cartographie des âges cumulés par l'approche déterministe

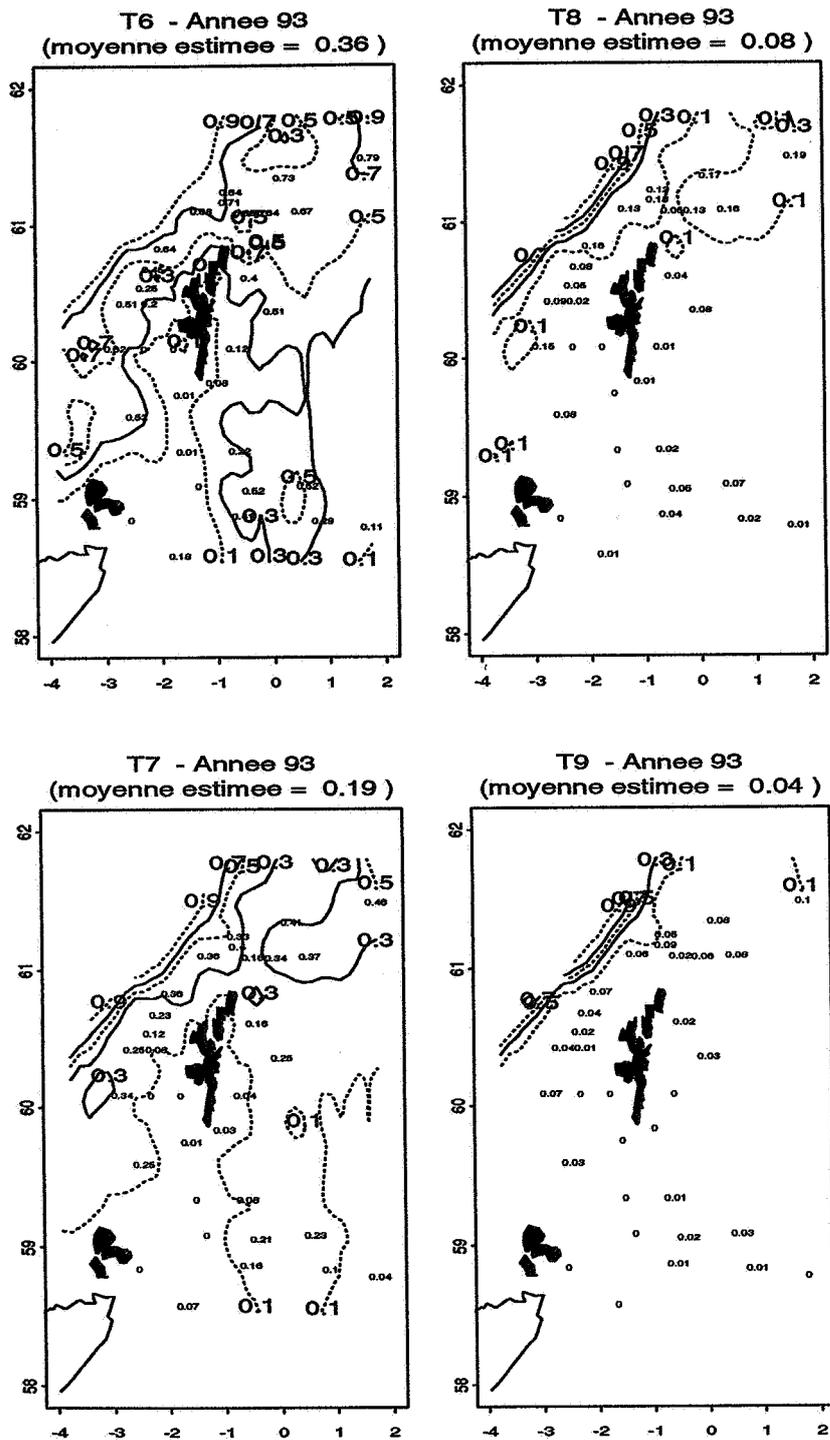


FIG. 2.30 - Cartographie des âges cumulés par l'approche déterministe (suite)

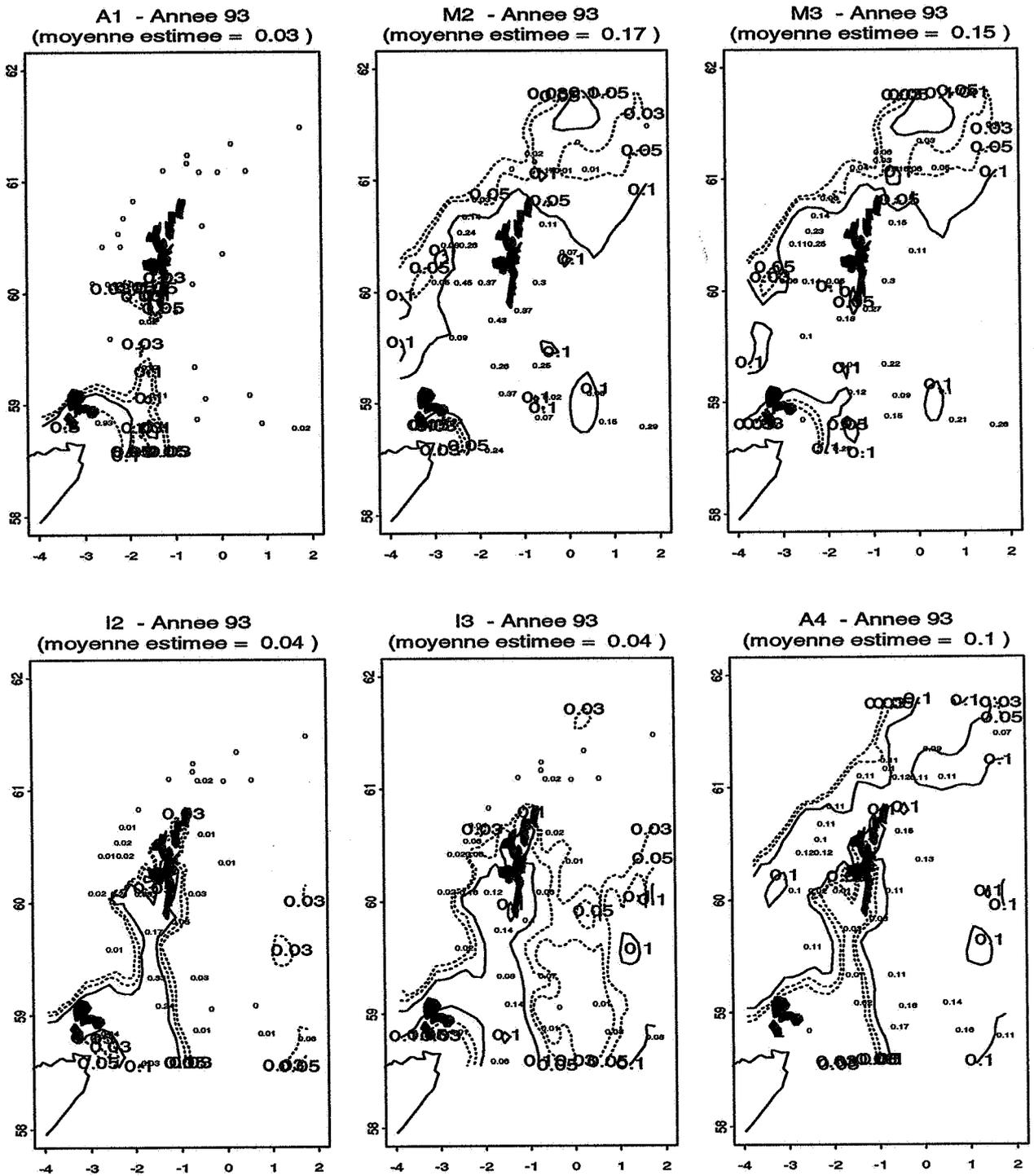


FIG. 2.31 - Cartographie des proportions d'âge par l'approche déterministe

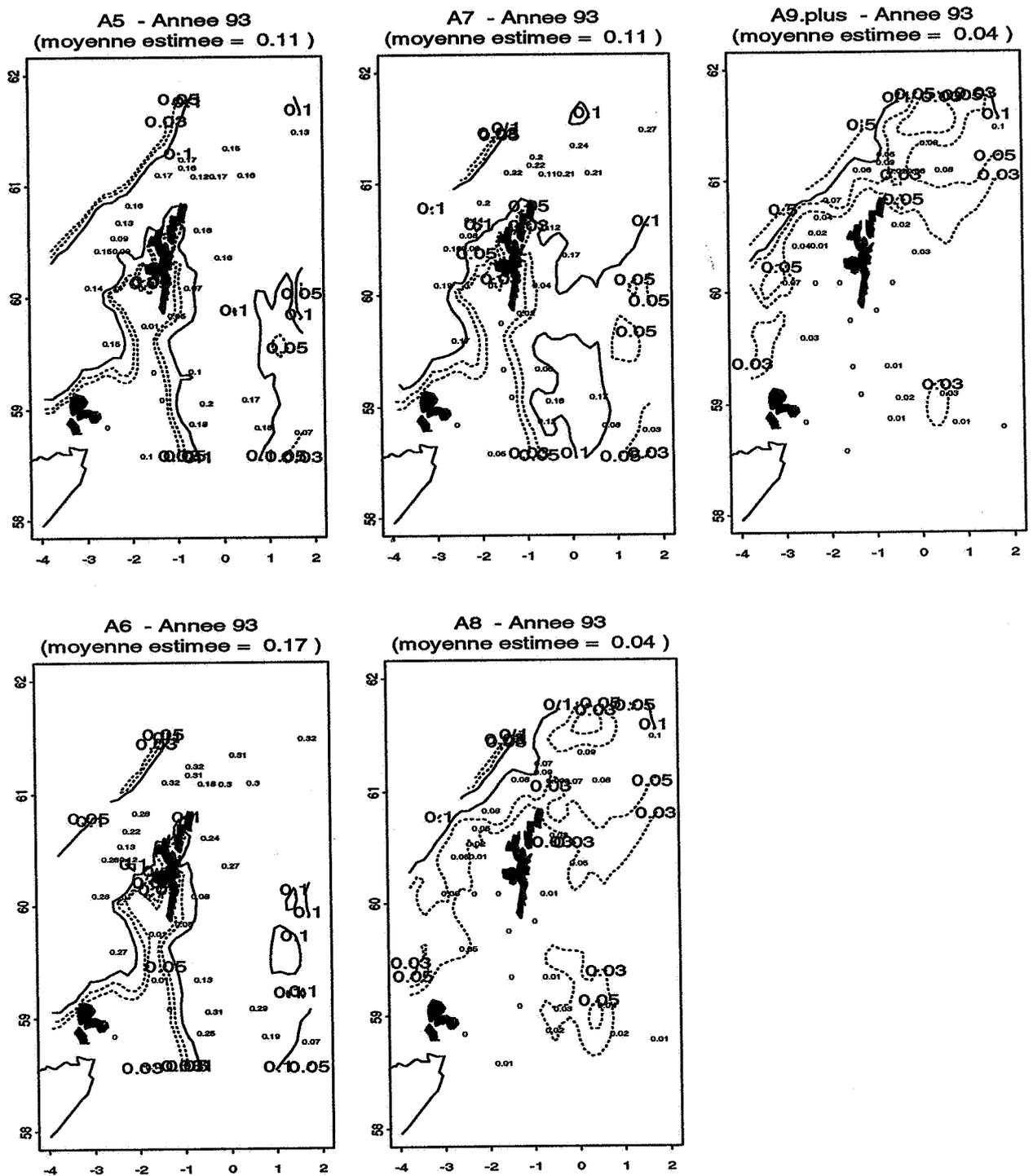


FIG. 2.32 - Cartographie des proportions d'âge par l'approche déterministe (suite)

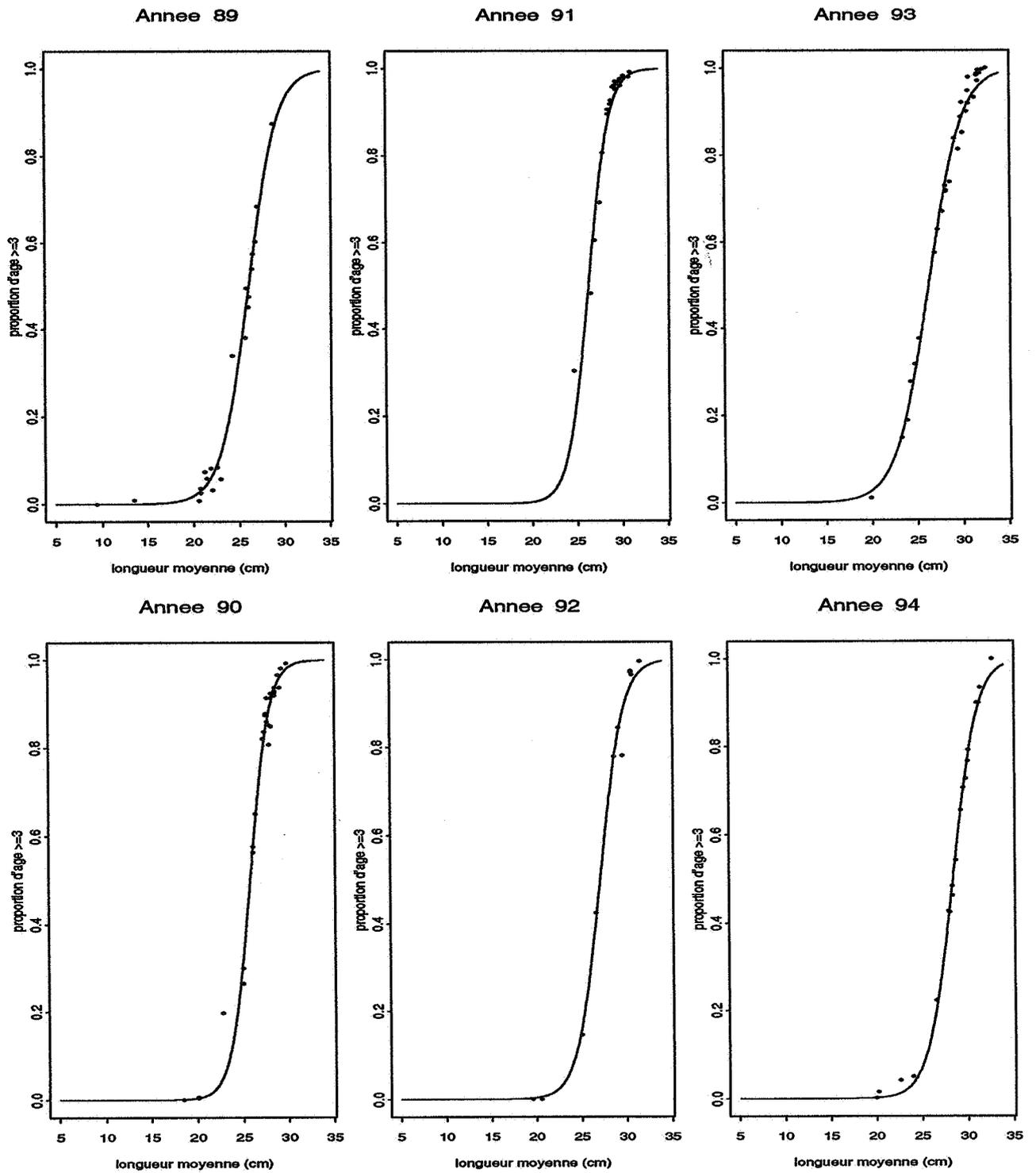


FIG. 2.33 - Régression T3.1/longueur moyenne sur 6 années

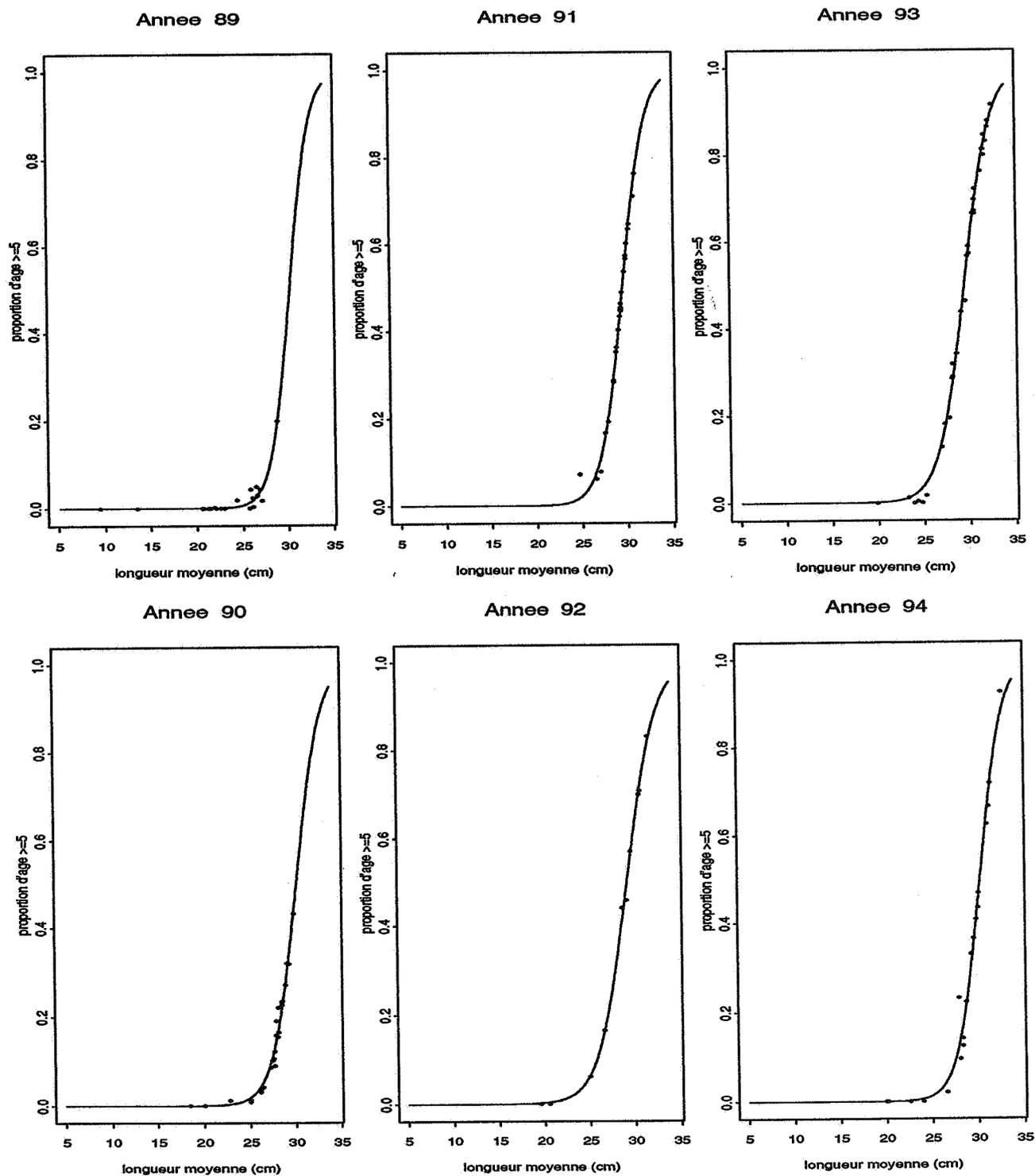
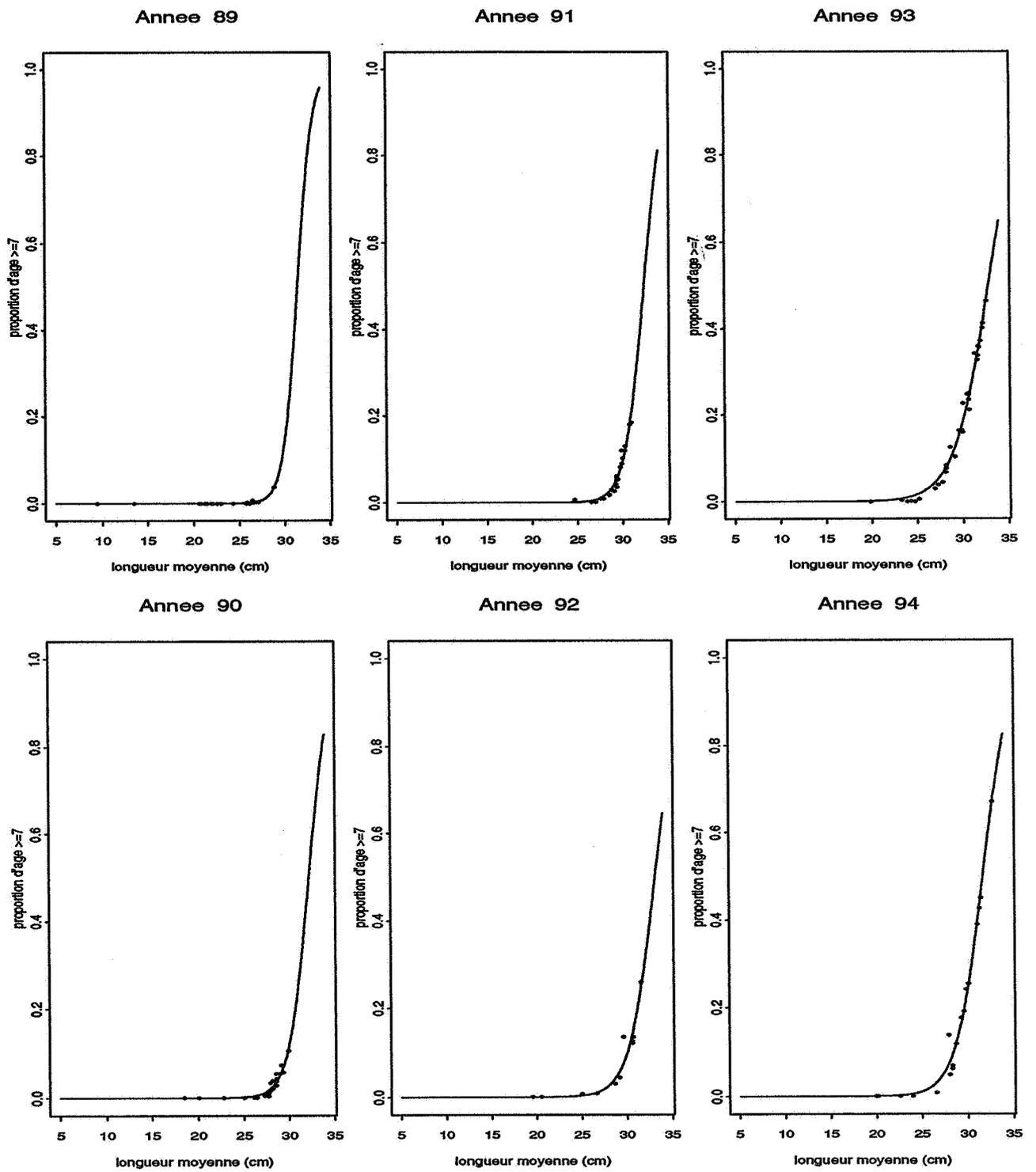


FIG. 2.34 - Régression T5/longueur moyenne sur 6 années

FIG. 2.35 - Régression $T7$ /longueur moyenne sur 6 années

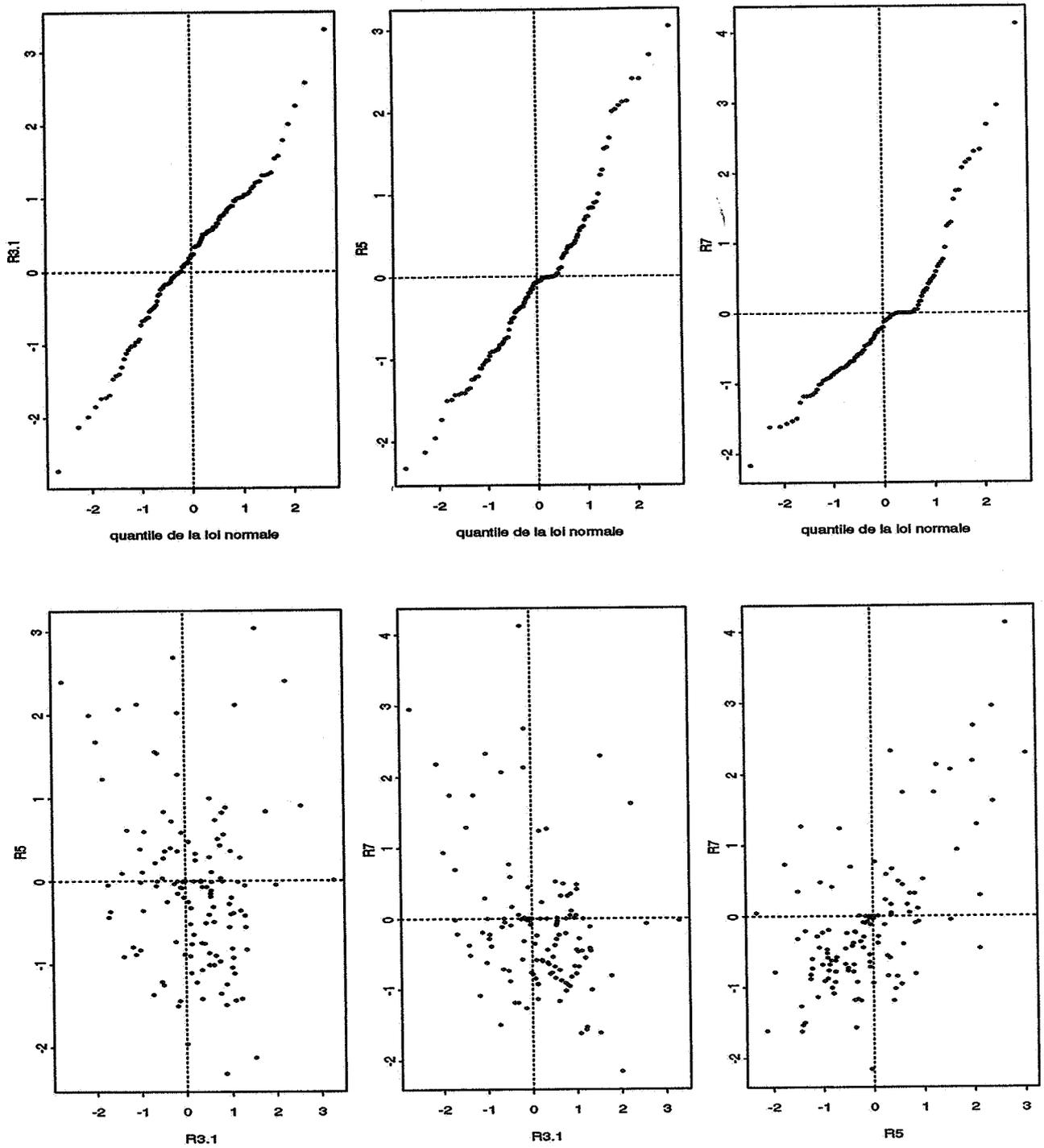


FIG. 2.36 - Tests sur la distribution des résidus (Approche directe)

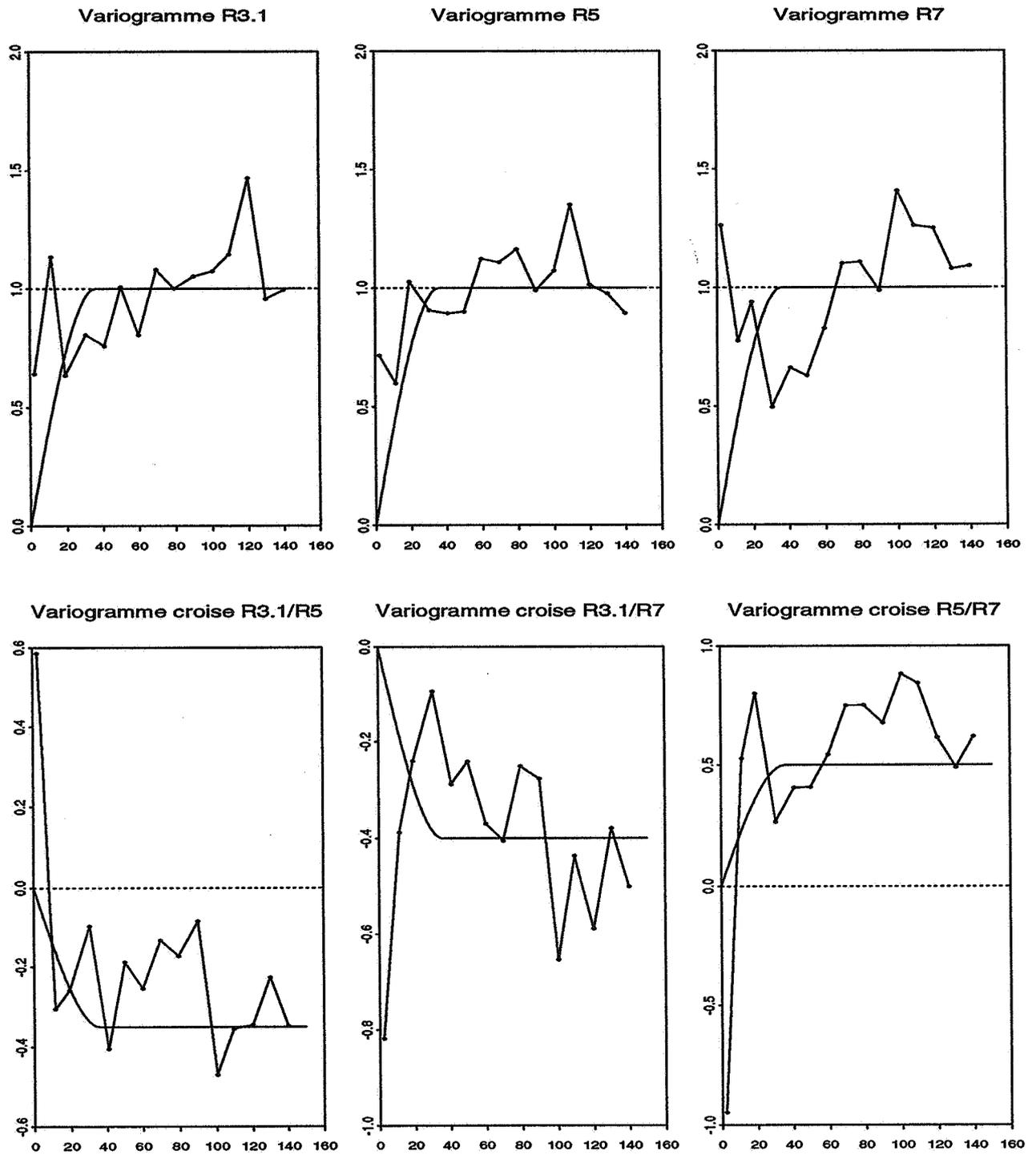


FIG. 2.37 - Étude variographique croisée des résidus (Régression $T=f(l)+\epsilon$)

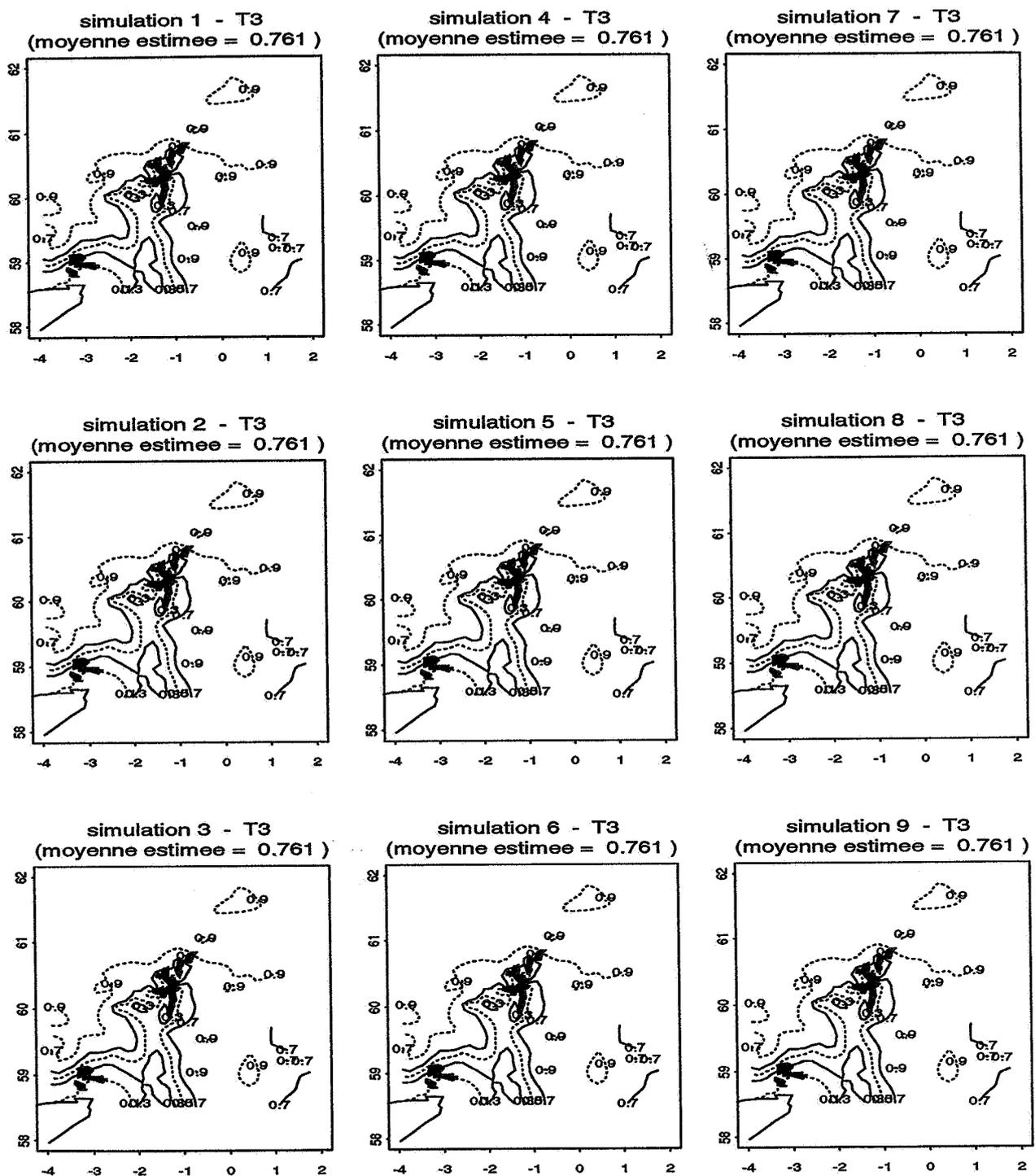


FIG. 2.38 - Simulations de la classe d'âges cumulés T3.1

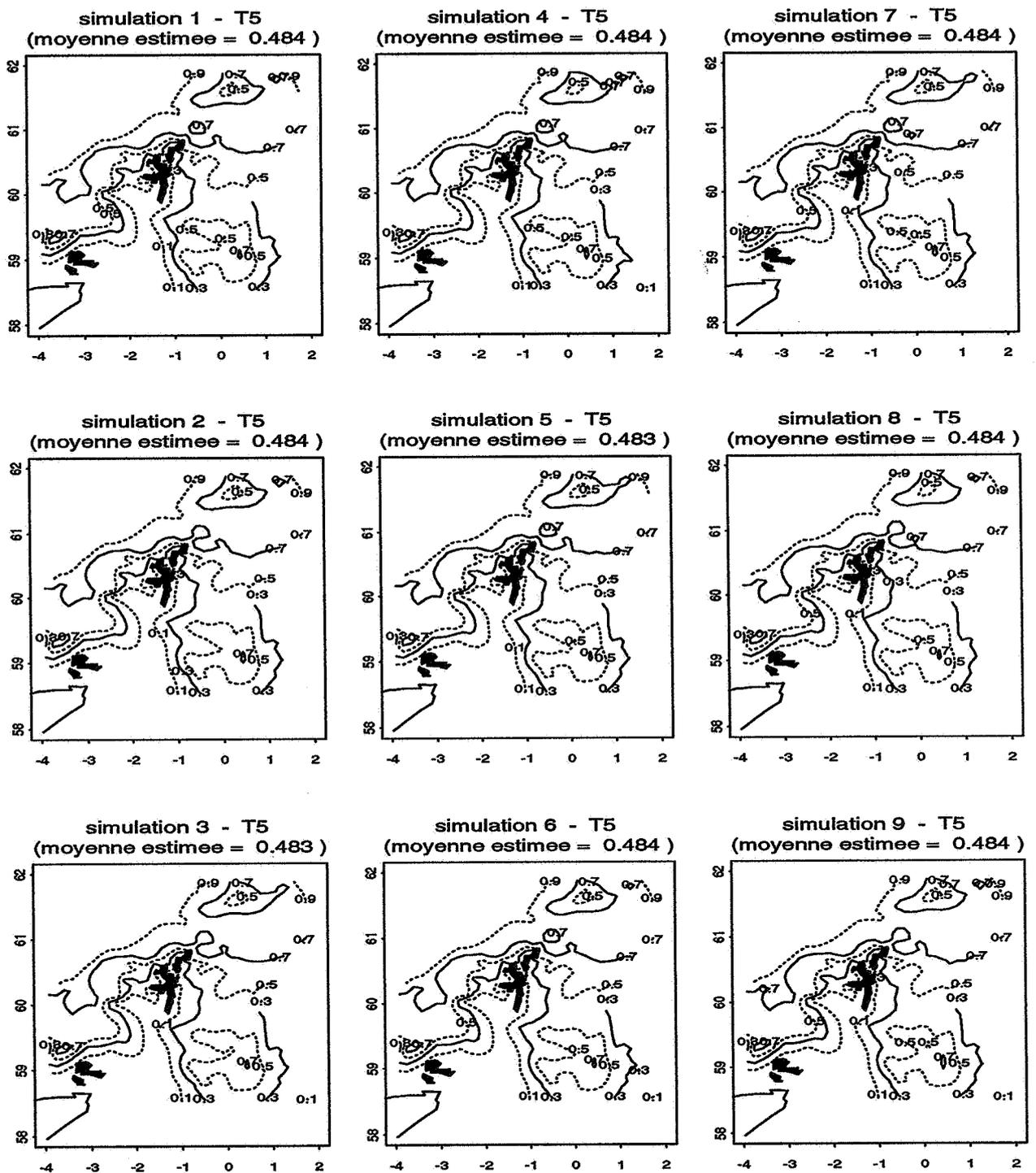


FIG. 2.39 - Simulations de la classe d'âges cumulés T5

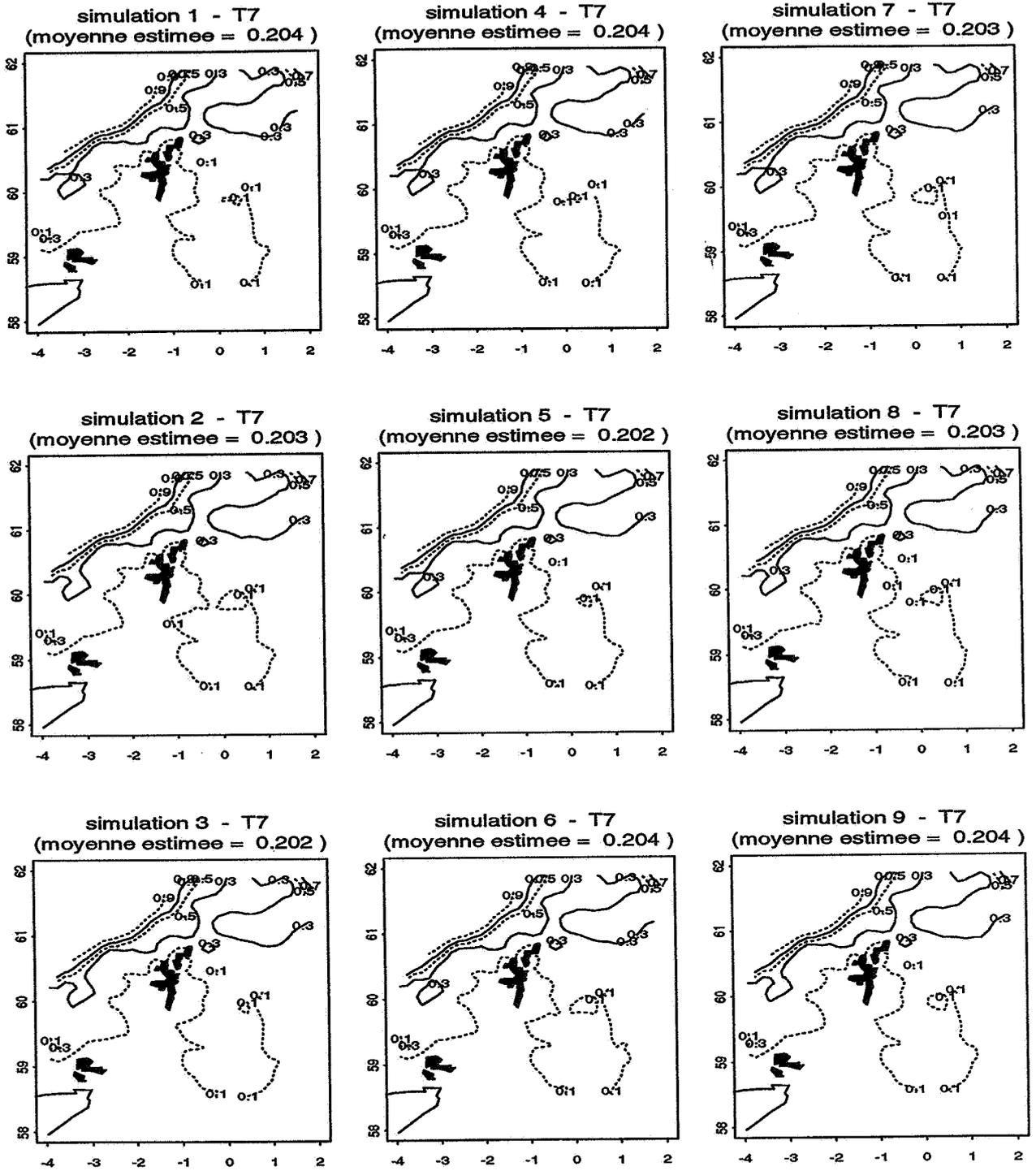


FIG. 2.40 - Simulations de la classe d'âges cumulés T7

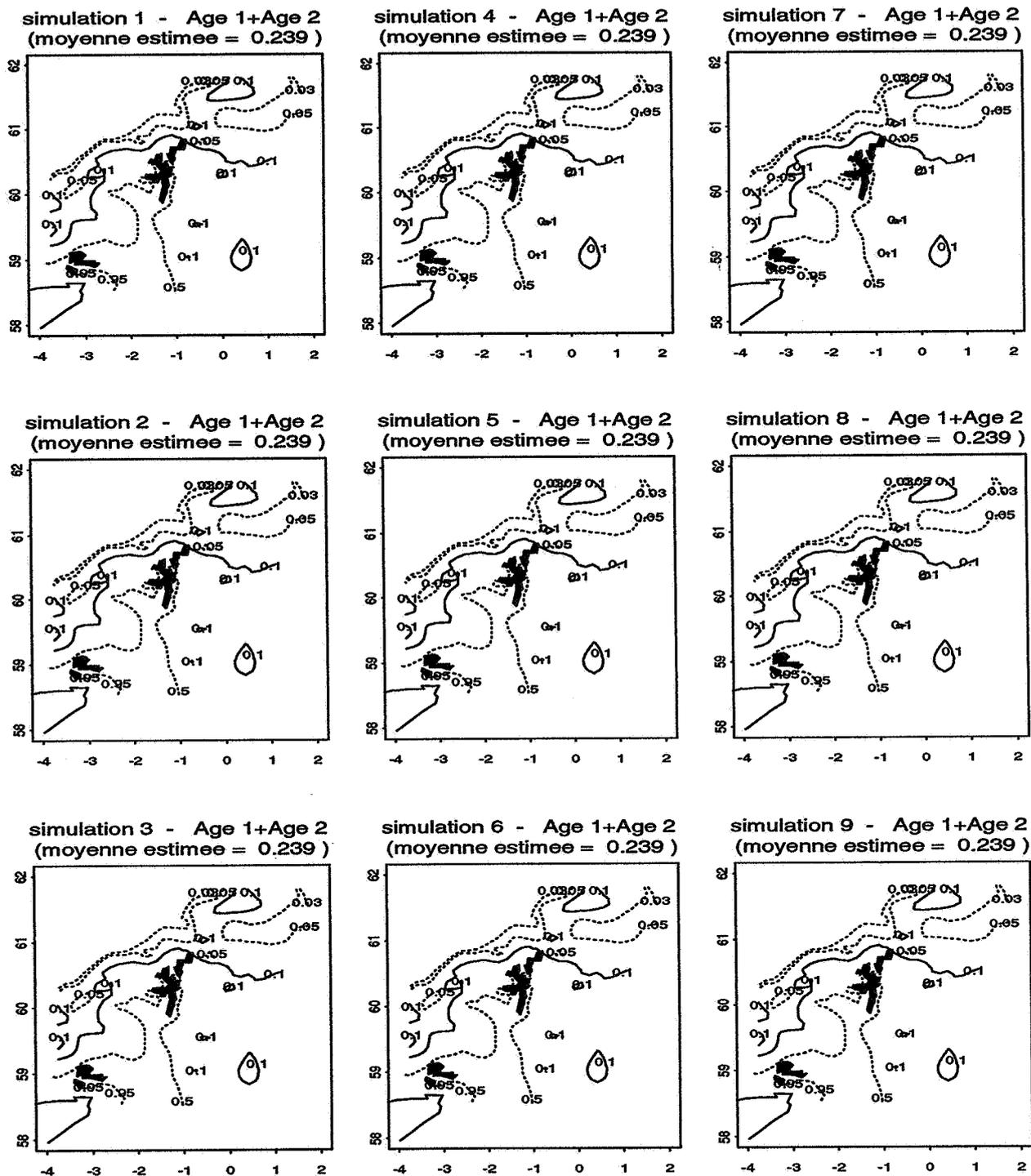
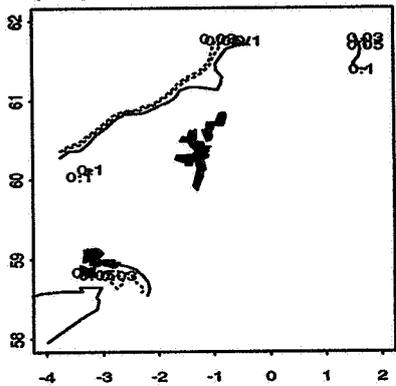
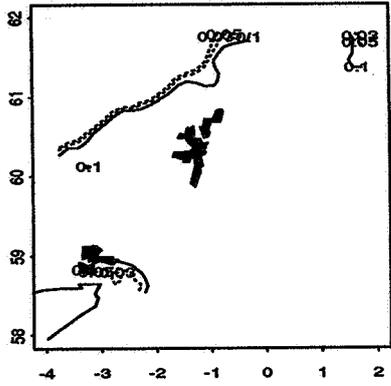


FIG. 2.41 - Simulations de la proportion A1+A2

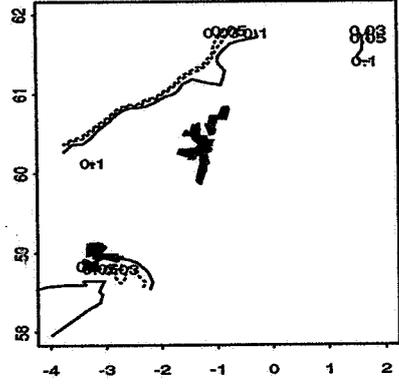
simulation 1 - Age 3 + Age 4
(moyenne estimee = 0.277)



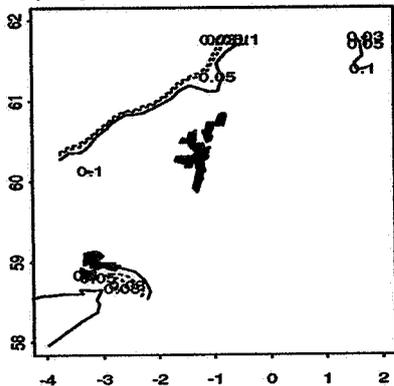
simulation 4 - Age 3 + Age 4
(moyenne estimee = 0.278)



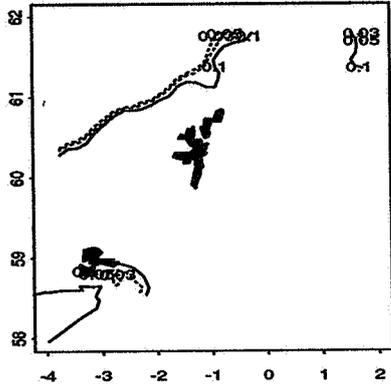
simulation 7 - Age 3 + Age 4
(moyenne estimee = 0.277)



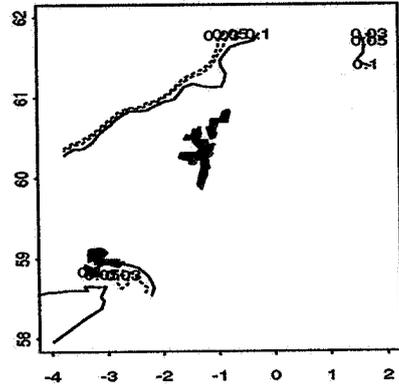
simulation 2 - Age 3 + Age 4
(moyenne estimee = 0.277)



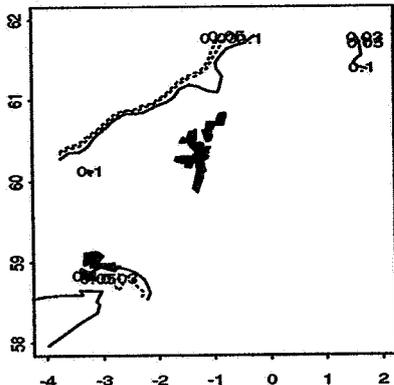
simulation 5 - Age 3 + Age 4
(moyenne estimee = 0.277)



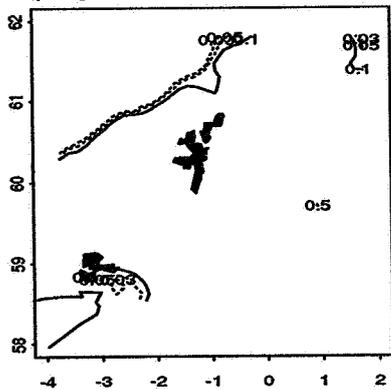
simulation 8 - Age 3 + Age 4
(moyenne estimee = 0.277)



simulation 3 - Age 3 + Age 4
(moyenne estimee = 0.277)



simulation 6 - Age 3 + Age 4
(moyenne estimee = 0.277)



simulation 9 - Age 3 + Age 4
(moyenne estimee = 0.277)

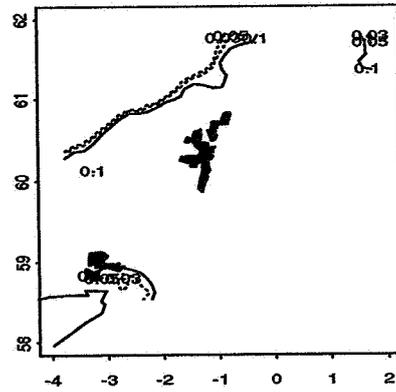


FIG. 2.42 - Simulations de la proportion A_3+A_4

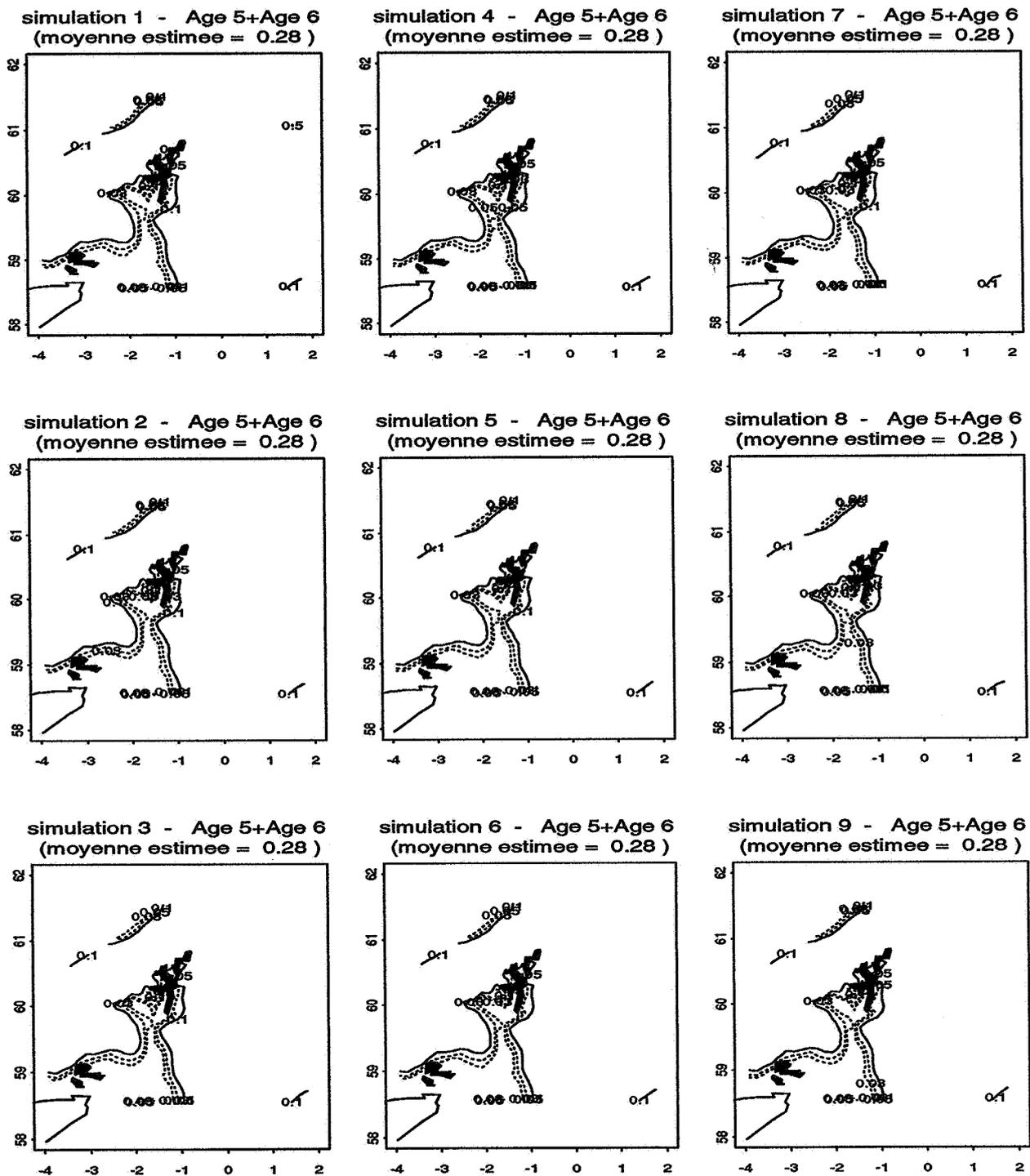


FIG. 2.43 - Simulations de la proportion A5+A6

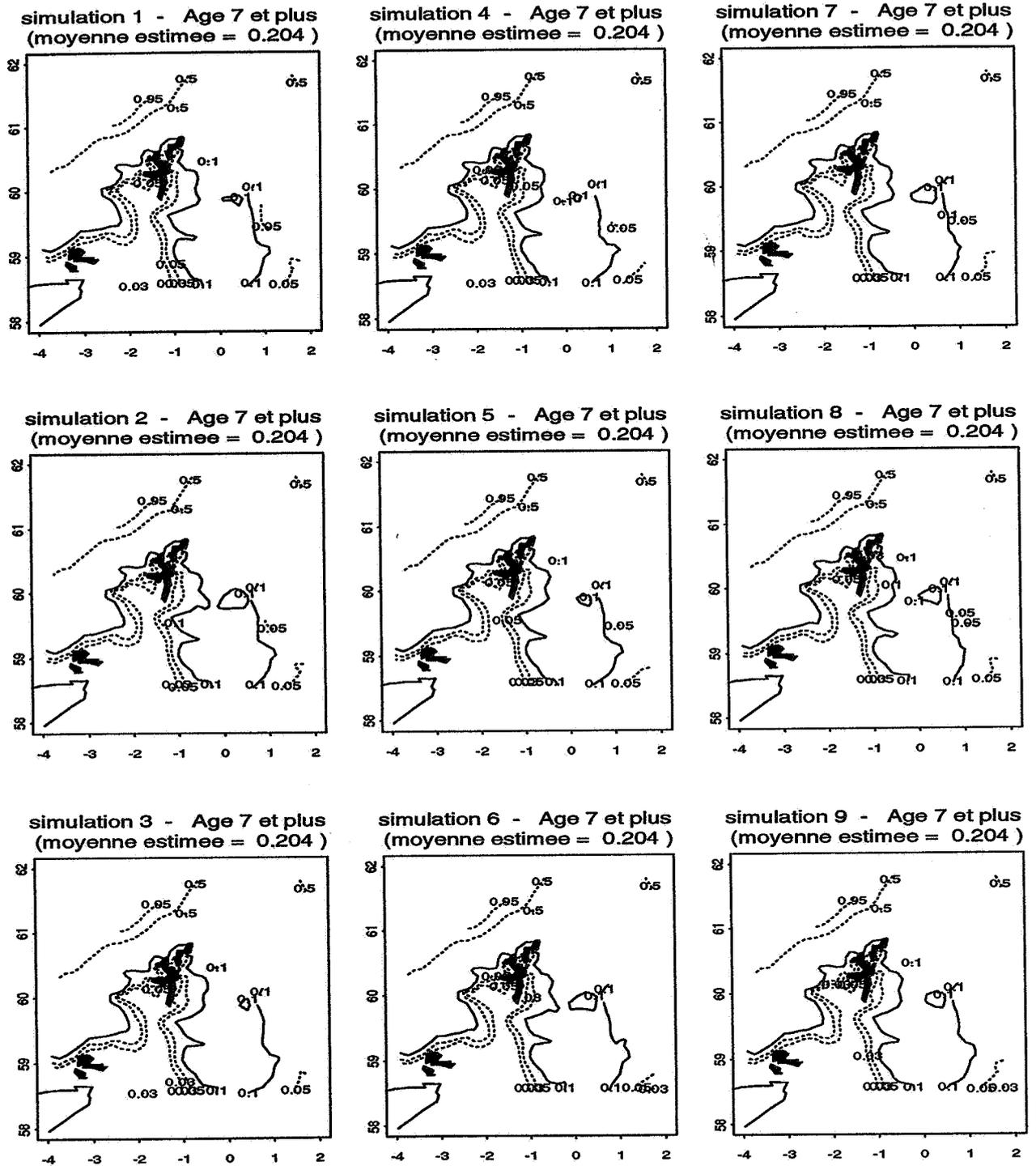


FIG. 2.44 - Simulations de la proportion Age ≥ 7

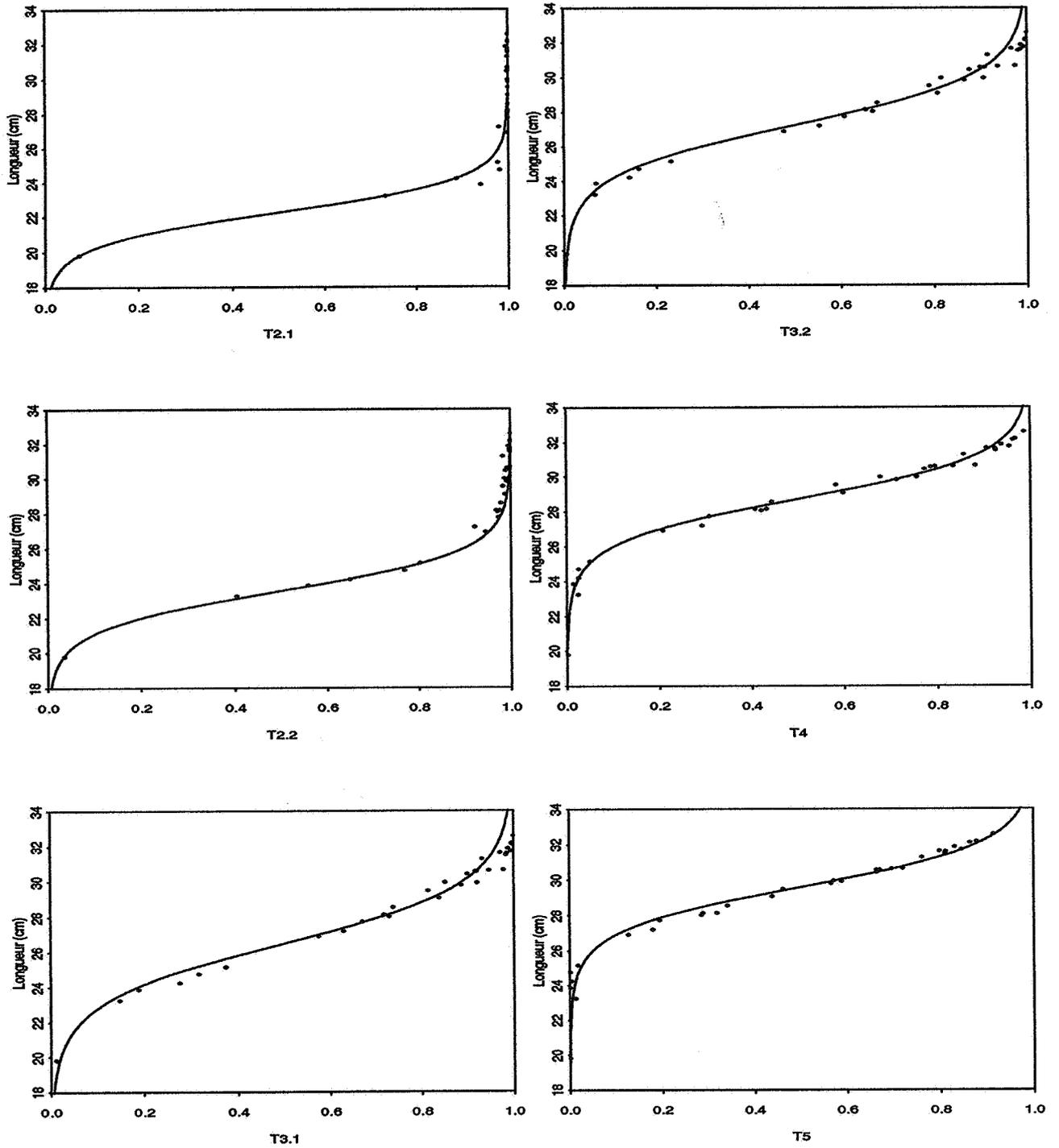


FIG. 2.45 - Régressions âges cumulés/longueur moyenne (Régression inverse)

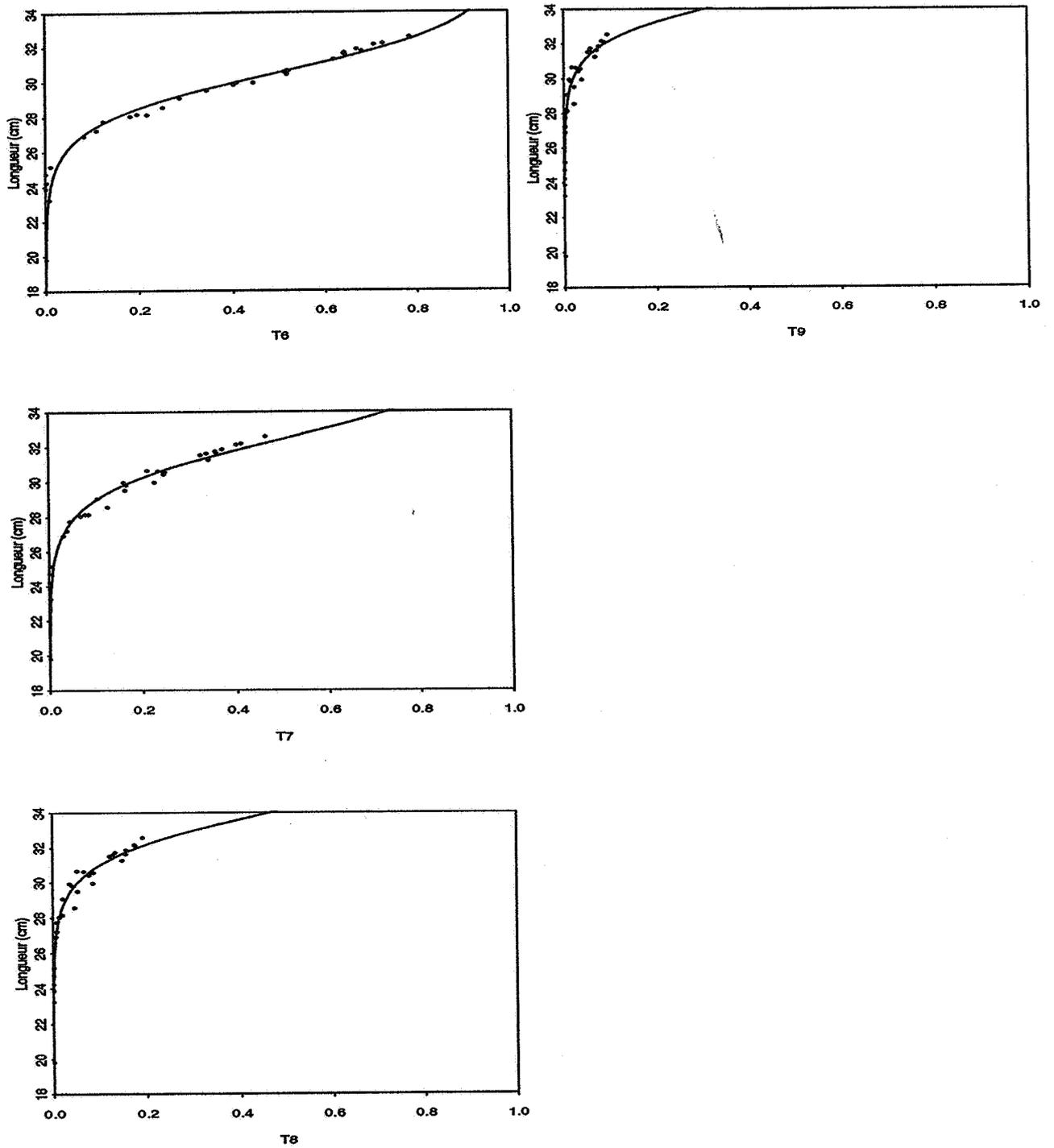


FIG. 2.46 - Régressions âges cumulés/longueur moyenne (Régression inverse - suite)

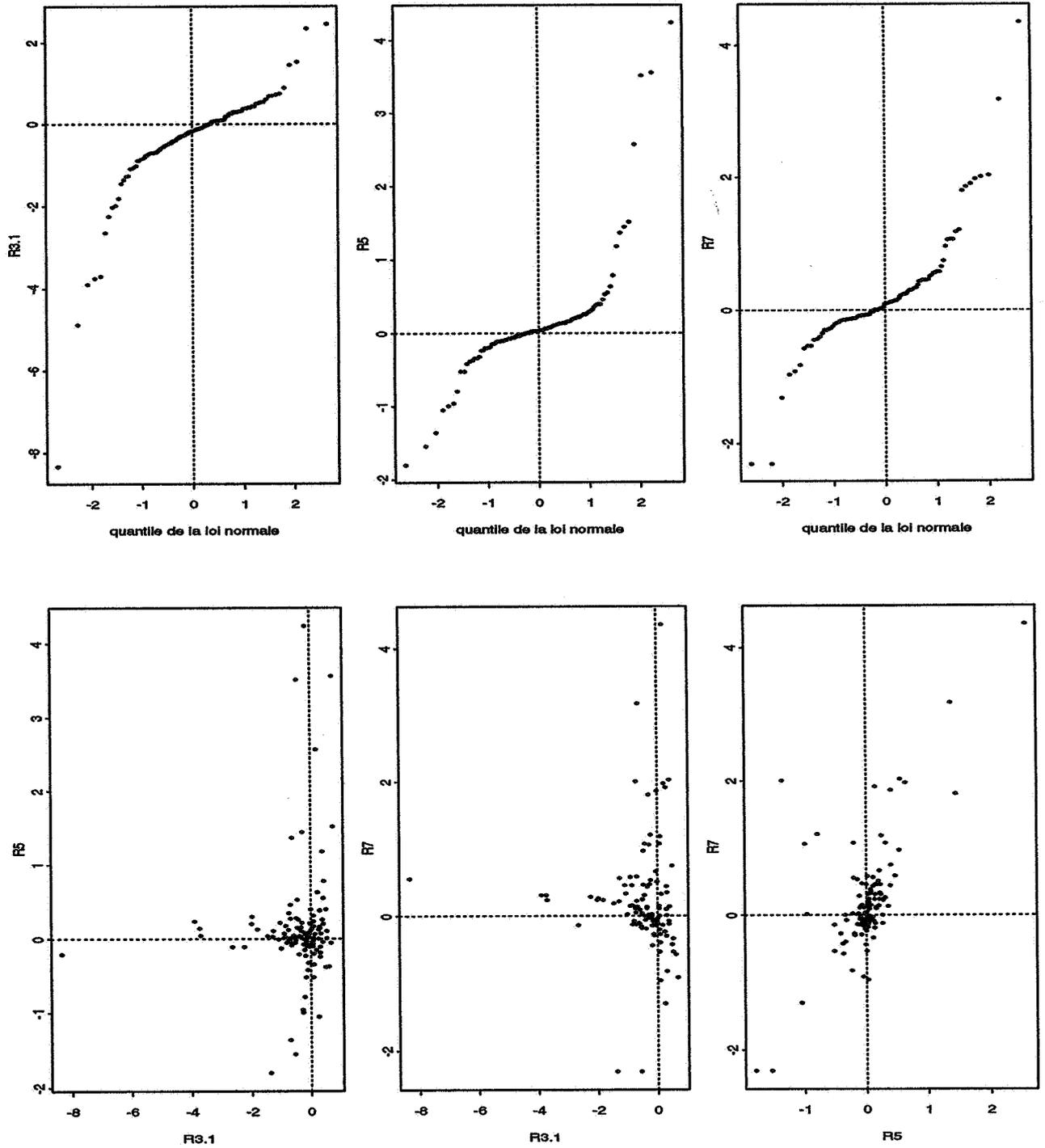


FIG. 2.47 - Tests sur la distribution des résidus (Approche inverse)

Chapitre 3

Figures illustrant la partie III

Campagne 1993

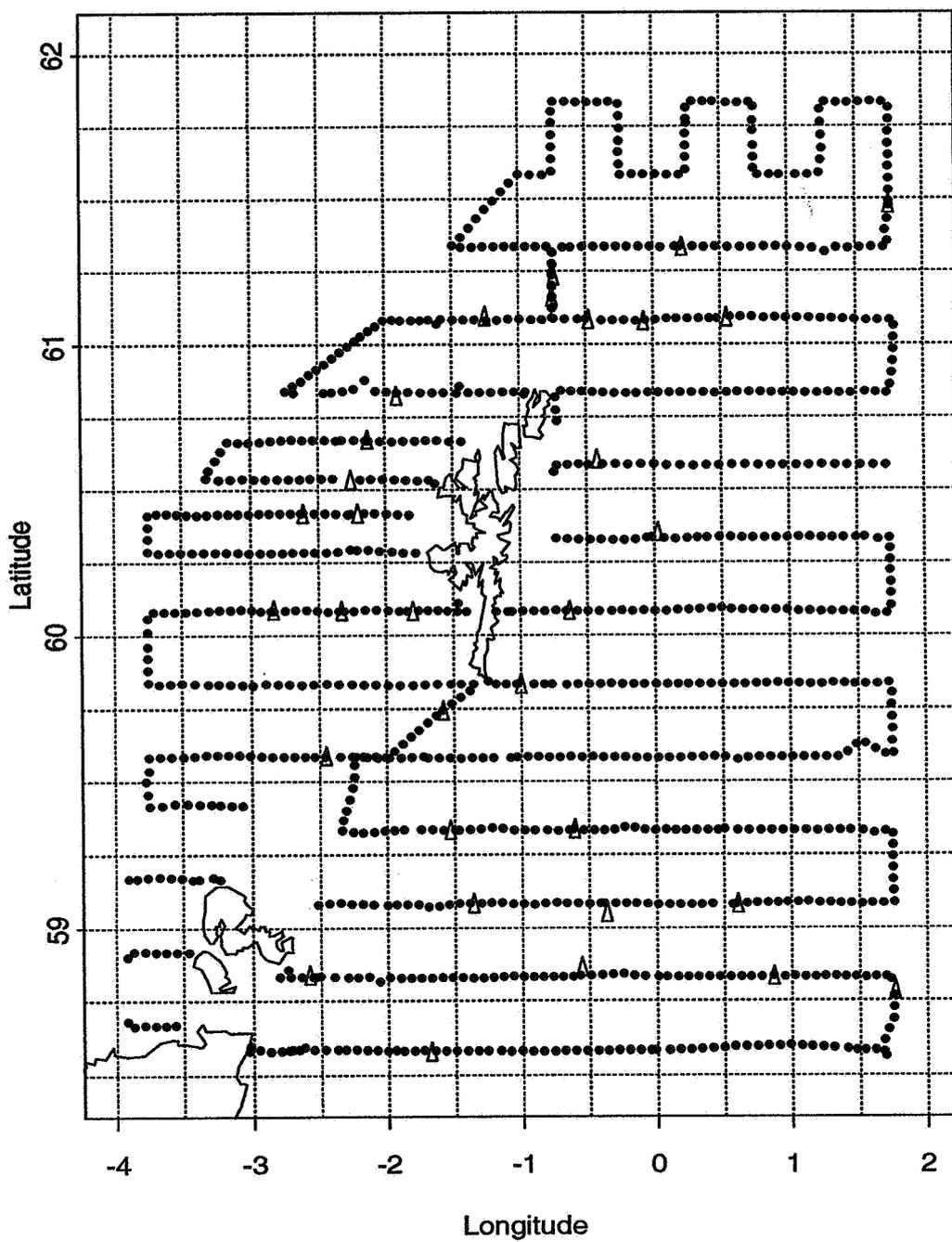


FIG. 3.1 - Échantillonnage acoustique - stations d'échantillonnage - campagne 1993

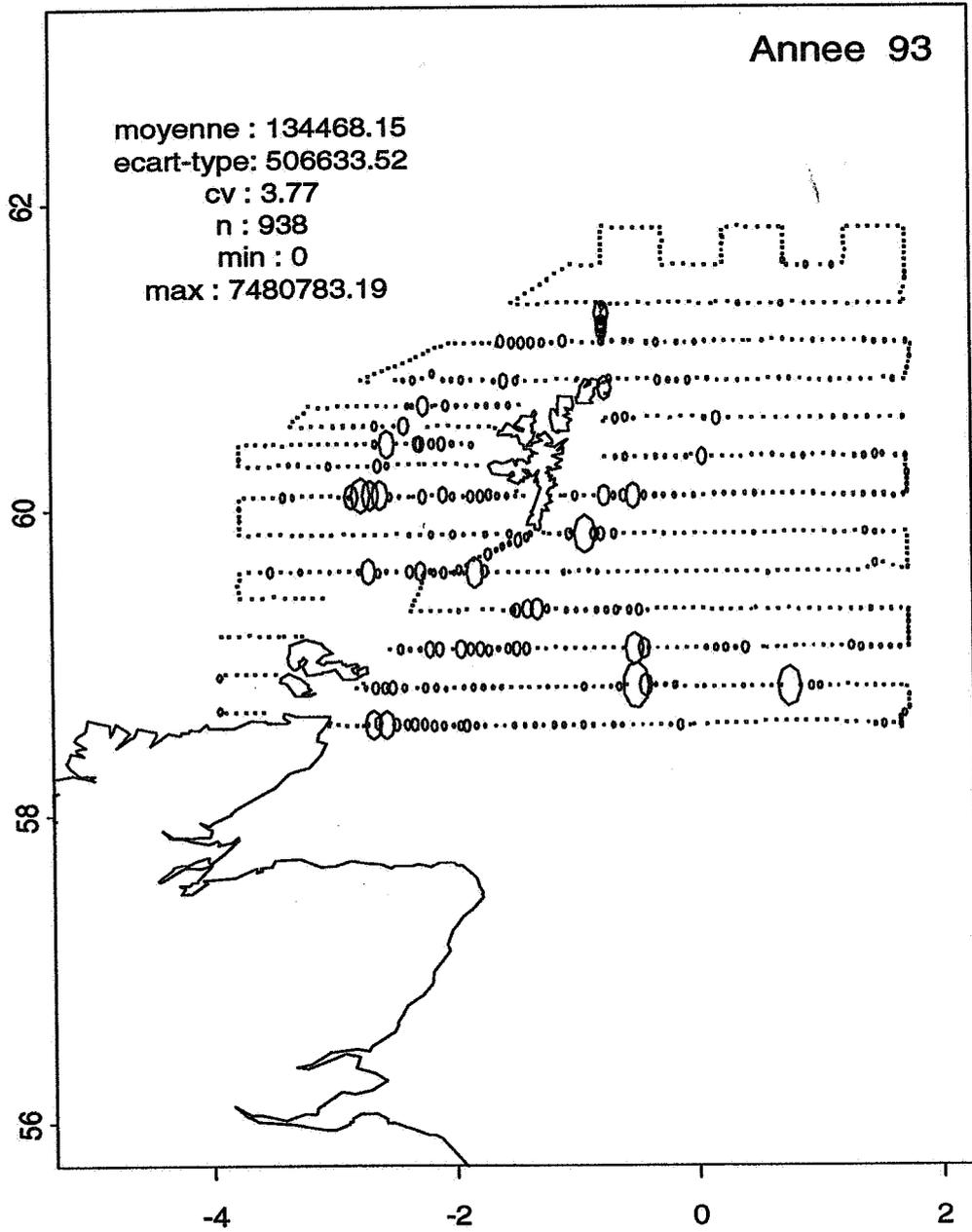
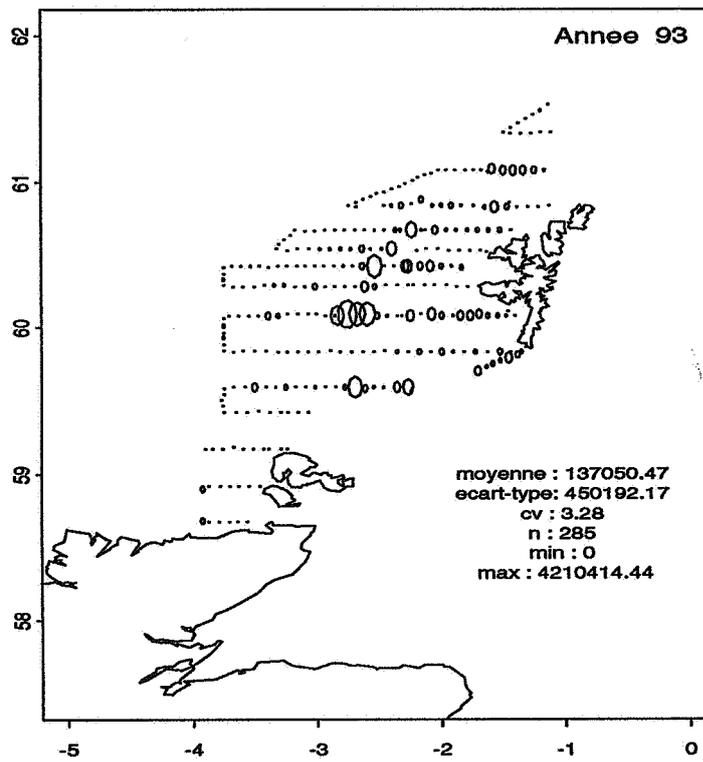


FIG. 3.2 - Représentation proportionnelle de la densité estimée - 1993

Partie OUEST



Partie EST

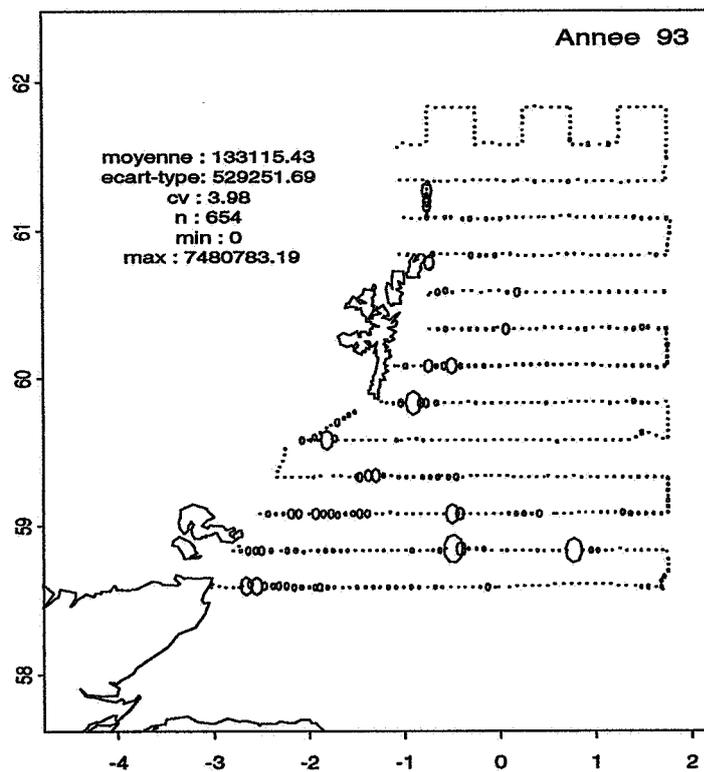


FIG. 3.3 - Représentation proportionnelle de la densité estimée - 1993- Parties OUEST et EST

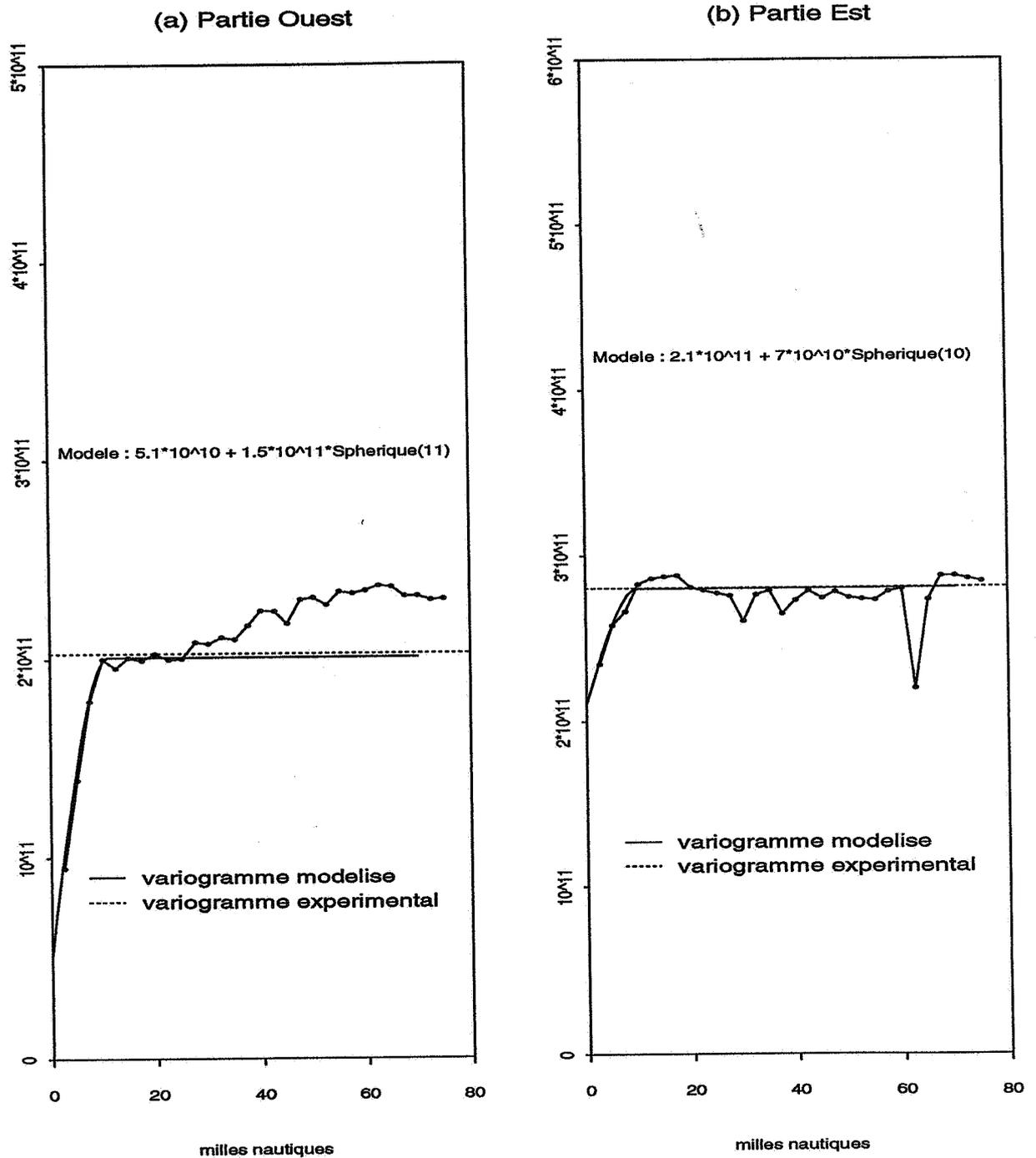


FIG. 3.4 - Variogramme de la densité estimée expérimental calculé le long de la route du navire et ajustement - Campagne 1993