

**CONTRIBUTIONS GEOSTATISTIQUES
A LA BIOLOGIE DES PECHEES MARITIMES**

THÈSE
présentée à

l'Ecole Nationale Supérieure des Mines de Paris

par

Pierre PETITGAS

pour obtenir le titre de
Docteur en Géostatistique

Soutenue le 28 Octobre 1991

devant le jury composé de :

M. G. MATHERON	Président
M. Ph. GROS	Rapporteur externe
M. J. RIVOIRARD	Rapporteur interne
M. P. CHAUVET	Examineur
M. J.-Y. LE GALL	Examineur
M. E. MARCHAL	Examineur

Remerciements

Je voudrais remercier le Professeur G. Matheron, Directeur du Centre de Géostatistique de l'Ecole des Mines de Paris, Directeur de ma thèse et Président de mon jury, de m'avoir invité au Centre de Géostatistique et de m'avoir donné la possibilité d'y travailler dans de bonnes conditions. Ses travaux scientifiques ainsi que son intérêt pour les applications pratiques nouvelles m'ont toujours stimulé et enrichi.

Je souhaite exprimer ma reconnaissance à tous les chercheurs du Centre de Géostatistique pour l'amicale atmosphère de travail qu'ils y font régner, m'ayant laissé libre de mes intentions et m'ayant toujours, avec disponibilité et bonne humeur, aidé à les réaliser. Parmi mes collègues du Centre de Géostatistique, Jacques Rivoirard tient une place particulière car il s'est intéressé de très près à mon travail, l'a suivi et m'a fait bénéficier de ses réflexions comme de son expérience. Je voudrais rendre hommage à mes camarades de bureau, Xavier Freulon et Colin Daly, avec lesquels les discussions ont toujours été nombreuses, intéressantes mais surtout complices.

Une partie de mon travail était dans mon idée, extérieure au Centre de Géostatistique: c'était en mer pour des échantillonnages avec l'Ifremer et l'Orstom, c'était en réunions de travail au Conseil International pour l'Exploration de la Mer. A chaque fois, je demandais à participer. Je tiens à remercier les biologistes responsables des campagnes d'avoir bien voulu m'embarquer sur les navires océanographiques pour ces confrontations au réel. Je voudrais remercier encore une fois le Professeur G. Matheron, le Centre de Géostatistique et l'Ecole des Mines de Paris pour la liberté et la confiance qu'ils m'ont accordées afin que je puisse développer mon travail selon mes intentions.

Ces fugues naturalistes nécessaires n'auraient pas été possibles sans mes collaborateurs océanographes et biologistes. Parmi tous ceux que j'ai côtoyés à l'Ifremer–Nantes, –Brest, –Lorient, à l'Orstom en acoustique et qui m'ont apporté beaucoup, je voudrais remercier tout particulièrement Constantin Koutsikopoulos de l'Ifremer. Grâce à notre entente complice dans la vie et dans le travail, j'ai pu, entre Belle Ile et l'Ile d'Yeu, vivre toutes les étapes successives de la belle recherche océanographique, depuis la pêche en mer jusqu'au traitement des données. Je voudrais exprimer toute ma reconnaissance à Emile Marchal et François Gerlotto de l'Orstom pour leur ouverture, leur intérêt porté à mon travail alors qu'ils ne me connaissaient pas et pour la formation à la méthode acoustique qu'ils m'ont amicalement offerte.

Je dois beaucoup au Conseil International pour l'Exploration de la Mer (Ciem), cette Internationale de Biologistes des Pêches fondée par eux au début du siècle, qui est entre autres choses, aussi une tribune libre où j'ai pu faire valoir mon travail. J'ai plaisir à remercier tous ceux qui en Norvège, Islande, Ecosse, Canada, France et Caraïbes ont participé au lancement d'une dynamique géostatistique au sein de la communauté scientifique du Ciem.

Je remercie les membres du jury d'avoir accepté d'en faire partie et de m'avoir critiqué judicieusement. Je remercie particulièrement Philippe Gros de l'Ifremer et Jacques Rivoirard du Centre de Géostatistique pour leur lecture très approfondie du manuscrit.

Je remercie tous ceux qui de près ou de loin, du point de vue technique, logistique, intellectuel et affectif, m'ont permis de mener à bien le présent travail.

Résumé

SOME GEOSTATISTICAL CONTRIBUTIONS TO FISHERIES BIOLOGY

Geostatistics is an application of probability theory for the estimation of natural processes that develop in the geographical space. Some aspects of fisheries biology are typically spatial: the assessment of the population total biomass by sampling surveys; the manner in which the organisms disperse in space, year after year, as a response to changes in their total biomass. These two subjects are investigated using geostatistics.

In comparison to the commonly used classical statistics, geostatistics accounts for spatial correlation (structure). This has two consequences. Firstly, the method enables us to calculate the variance of estimation of the total biomass for any type of sampling whether it be random or not. Secondly, the structural analysis enables us to characterise the spatial distribution in geometrical terms which can then be used to appropriately plan future sampling surveys according to the spatial structure.

The temporal variability of a spatial distribution is a biological subject new to geostatistics. Interannual variations of seasonally stable spatial distributions are studied. A test is proposed for classifying three types of spatial strategies, i.e. local response of fish density to year after year variations of total biomass. Only one of these has been observed on data. A spatio-temporal model is proposed for this case.

CONTRIBUTIONS GEOSTATISTIQUES A LA BIOLOGIE DES PECHES MARITIMES

La Géostatistique est une application de la théorie des probabilités à l'estimation de grandeurs distribuées dans l'espace géographique. Certains sujets de la biologie halieutique se posent dans l'espace géographique: l'estimation des quantités totales par campagnes d'échantillonnage; la façon dont d'année en année, les organismes d'une population naturelle se répartissent dans l'espace, en réponse aux variations de leur quantité totale. On a donc utilisé la méthode pour analyser ces deux questions.

L'originalité de la Géostatistique par rapport à la théorie classique de l'échantillonnage aléatoire, est l'utilisation explicite des corrélations spatiales (structure). Ceci a deux conséquences. D'une part, la méthode permet de calculer la variance de l'estimation globale quel que soit le type d'échantillonnage, aléatoire ou non. D'autre part, l'analyse structurale dégage les critères géométriques de la distribution spatiale essentiels à la planification d'échantillonnages ultérieurs.

L'évolution temporelle de la distribution géographique est une nouveauté biologique pour la géostatistique. Les évolutions inter-annuelles de distributions stables saisonnièrement sont étudiées. Un test est établi pour distinguer trois stratégies d'occupation de l'espace quand la quantité totale varie d'année en année. Une de ces stratégies seulement a été observée. Un modèle spatio-temporel est proposé pour ce cas.

INTRODUCTION	1
---------------------------	----------

PARTIE I

ESTIMATION DIRECTE DES STOCKS	5
--	----------

A. GÉOSTATISTIQUE DE LA VARIANCE D'ESTIMATION	7
--	----------

A.1. Plan systématique	7
A.1.1. Théorie transitive	7
A.1.2. Théorie intrinsèque	9
A.1.3. Principes d'approximations	14
A.2. Plan aléatoire stratifié	19
A.3. Le variogramme près de l'origine	24
A.3.1. Quasi-stationnarité	24
A.3.2. Les variogrammes	25
A.3.3. Le choix du modèle	26

B. APPLICATION A UNE PROSPECTION ACOUSTIQUE A MAILLE	
---	--

RÉGULIÈRE	29
------------------------	-----------

B.1. L'échantillonnage	30
B.1.1. La méthode de l'écho-intégration et les données	30
B.1.2. Stratégie d'échantillonnage suivie	34
B.2. Description du jeu de données	35
B.2.1. Statistiques globales	35
B.2.2. Définition du champ	36
B.2.3. Statistiques dans le champ	37
B.2.4. L'estimation	37
B.3. Calcul de la variance d'estimation à 2D	38
B.3.1. Analyse variographique	38
B.3.2. Variance d'estimation à champ fixé	40
B.4. Calcul de la variance à 1D	42
B.4.1. Calcul en intrinsèque	43
B.4.2. Calcul en transitif	46
B.5. Variance d'estimation intrinsèque et limites du champ	47
B.5.1. Variance de l'erreur géométrique sur l'estimation à 2D	47
B.5.2. Variance de l'erreur géométrique sur l'estimation à 1D	50
B.5.3. Conclusion	51
B.6. Discussion : rapports entre les variances d'estimation de la Statistique Classique et de la Géostatistique	52

C.	EFFICACITÉ DE L'ÉCHANTILLONNAGE	55
C.1.	Choix de la distance inter-transects dans une reconnaissance aveugle	57
C.1.1.	Distance inter-transects et variance d'estimation	58
C.1.2.	Variogramme géométrique des très fortes valeurs	58
C.1.3.	Commentaires sur la distance inter-transects	59
C.1.4.	Granulométrie induite des très fortes valeurs sur les transects	60
C.2.	Application d'un modèle de krigeage disjonctif sans effet de bords	60
C.2.1.	Propriétés géométriques	61
C.2.2.	Le modèle et son ajustement : dissection des données	63
C.2.3.	Cartographie de l'aire du gisement	66
C.3.	Intervalle de confiance pour le nombre de grains à teneur très élevée	67
C.3.1.	Schéma de Poisson pour l'implantation des très fortes valeurs	68
C.3.2.	Résultat :	70
C.4.	Stratégie d'échantillonnage : discussion	71
D.	EFFICACITÉ DES ESTIMATEURS ET FOURCHETTES D'ERREUR	73
D.1.	Le formalisme lognormal	74
D.2.	Estimateur arithmétique ou lognormal ?	75
D.2.1.	Plan aléatoire pur	76
D.2.2.	Plan aléatoire stratifié	77
D.2.3.	Plan systématique	77
D.2.4.	Echantillonnage non orthodoxe	78
D.3.	Estimation lognormale du hareng norvégien	80
D.4.	Intervalle de confiance et fourchette d'erreur	81

PARTIE II

	ANALYSE DES CARTES	85
A.	ETUDE DE LA DISPERSION , RELATIONS TONNAGE-TENEUR .	89
A.1.	Graphes Q(T) et sélectivité	89
A.2.	La sélectivité pour valider le modèle lognormal du recrutement	90
A.3.	Les courbes Q(T) et P(T) pour caractériser trois stratégies d'occupation de l'espace	91
B.	APPLICATION AUX DONNÉES	95
B.1.	Le Golfe de Gascogne et ses caractéristiques	95
B.2.	Le merlu-0 de 1987 à 1990	96
B.2.1.	Echantillonnage	97
B.2.2.	Sélectivités	98
B.2.3.	La composante spatiale invariante dans le temps	101
B.3.	Les œufs de sole de 1986 à 1990	106
B.3.1.	Les données	107
B.3.1.2.	Variable étudiée	108
B.3.2.	Sélectivités	108

C. DIFFERENCES SIGNIFICATIVES ENTRE LES ESTIMATIONS

GLOBALES	111
C.1. Le merlu-0	111
C.2. Les œufs de sole	115
C.3. Discussion	116

D. UN MODÈLE MULTIPLICATIF POUR LA RÉGIONALISATION DES ŒUFS

DE SOLE	117
D.1. Le modèle	118
D.2. La dérive $m_w(x)$ et son déterminisme biologique	120
D.3. Les résidus et le modèle probabiliste spatio-temporel	123
D.3.1. Structure spatiale	123
D.3.2. Structure temporelle	124
D.3.3. Interactions espace-temps	124
D.4. Contrôle de l'inférence du modèle	125
D.4.1. Variance d'estimation de la dérive	126
D.4.2. Conséquences sur le variogramme des résidus estimés	127
D.4.3. Conséquences sur les variations spatiales de la dérive estimée	127
D.4.4. Partition de la variance spatiale des données	128
D.5. Krigeage de la dérive	129
D.5.1. Une présentation du krigeage en FAI-k	130
D.5.2. Cartographie de la dérive estimée	132
D.5.3. Cartes de la dérive estimée et de la dérive vraie	135
D.6. Krigeage des résidus	136
D.6.1. Cartographie des résidus stationnaires	136
D.6.2. Une interprétation de la ponte	137
D.7. Variance de l'estimation globale	139
D.7.1. Variance de l'estimation spatiale instantanée	139
D.7.2. Variance de l'estimation spatio-temporelle	142
D.8. Conclusion	145

CONCLUSION	147
-------------------------	------------

BIBLIOGRAPHIE	151
----------------------------	------------

ANNEXE : FIGURES	159
-------------------------------	------------

INTRODUCTION

La pêche peut être définie comme l'exploitation de ressources vivantes qui se renouvellent naturellement. Du point de vue biologique, l'exploitation s'effectue sous la contrainte de la croissance (gain de poids) et de la reproduction (renouvellement des générations). Ainsi, indépendamment de la culture et de l'histoire du monde de la pêche, on peut, en termes uniquement biologiques, formuler les conditions d'exploitation optimales de la ressource lorsque sa dynamique naturelle de production de biomasse est paramétrée. C'est le sujet des modèles de production de dynamique des populations exploitées (théorie des pêches). Nous rattachons notre travail à ce contexte.

Le modèle de production peut être global ou analytique. Dans l'approche globale, c'est la loi de croissance de la population en fonction de sa propre biomasse qui est modélisée par une courbe logistique de Verhulst-Pearl (SCHAEFFER, 1954). L'approche analytique se fonde sur la démographie : les populations de poissons sont structurées en classes d'âge car chacune est issue d'une reproduction qui est en général annuelle dans l'Atlantique Nord. Dans l'approche analytique, les différentes composantes de la production sont séparées : croissance individuelle, mortalité, loi de renouvellement. Quel que soit le taux de reproduction, en raisonnant par tête de poisson, BEVERTON et al (1957) déterminent les courbes de capture et les points d'optimalité d'exploitation, en fonction de la croissance pondérale et de la mortalité des classes d'âges. Couplant ce modèle avec un modèle de renouvellement, SHEPHERD (1982) établit un modèle de production.

Dans l'Atlantique Nord-Est, et tout particulièrement en Mer du Nord, les modèles analytiques sont très utilisés. Si la croissance peut être mesurée¹ et que la mortalité naturelle peut être estimée, le lien entre la population et son taux de renouvellement à chaque tour du cycle biologique reste encore aujourd'hui autant un sujet d'étude qu'un sujet d'étonnement. En effet, la capacité que montrent les populations de poissons de fond à supporter une mortalité par pêche atteignant jusqu'à 7 fois la mortalité naturelle (Mer du Nord, SHEPHERD et al, 1990), suggère des mécanismes compensatoires propres à la population et régulant son taux de reproduction. Or paradoxalement, ceux-ci ne sont pas visibles. Si l'on reporte, en ordonnées, la quantité totale estimée des progénitures issues de la reproduction de l'année n et, en abscisses, la quantité totale estimée des géniteurs de l'année n , on obtient un nuage de corrélation entre le niveau d'abondance de la population et son taux de renouvellement (recrutement): la relation entre la population et son recrutement est en général très bruitée et aucune relation ne se dégage clairement. De nombreux modèles ont été proposés, certains plus dangereux que d'autres en rendant impossible l'extinction de la population, mais en définitive, la façon dont les populations de poissons bouclent leurs

1. bien que les paramètres des courbes de croissance présentent des variations écologiques décennales et nécessitent régulièrement l'observation et le calibrage (VEEN, 1978).

cycles biologiques n'a pas été théorisée de manière satisfaisante. Si le renouvellement des générations dépend d'une loi compensatoire propre à la population, il est aussi déterminé par l'environnement.

HJORT (1926) a décrit la structuration en classes d'âge des populations de poissons et l'a reliée au problème du renouvellement. Chaque reproduction donne naissance à une classe d'âge. Les effectifs dans chaque classe peuvent être très différents, variant sur une échelle logarithmique, et ces différences se maintiennent pendant toute la durée de vie de la classe. La mortalité naturelle, disons après la première année, se stabilise et ne remet pas en cause les différences d'effectifs. Par contre, elle agit différenciellement d'année en année pendant les premières phases de la vie des poissons et détermine chaque année l'effectif des classes d'âges. En général, les œufs millimétriques contiennent des gouttelettes lipidiques et après la ponte, ils montent dans les couches de surface. A l'éclosion, ils donnent naissance à des larves planctoniques qui devront trouver leur nourriture dans le plancton. Après la métamorphose, les tout jeunes juvéniles constituant la classe d'âge-0 (ils sont nés de l'année) commencent à coloniser leur habitat qui leur est en général spécifique (zones de nourricerie). L'unité de mesure en nombre des quantités d'œufs pondus sur les frayères est le billion (10^{12}) ; celle des juvéniles est le million (10^6). Chacun des œufs a une probabilité très faible (environ 10^{-6}) de donner un jeune poisson. C'est au cours de la phase planctonique qu'a lieu la régulation des abondances. L'importance des conditions hydro-climatiques dans les couches de surface de l'océan est déterminante et affecte la démographie des populations marines par le truchement du développement planctonique de leurs œufs et larves. Pour boucler le cycle biologique des populations de poissons et construire le modèle complet de production, il manque un modèle de la survie planctonique. Dans ce contexte, il est important d'entreprendre régulièrement des campagnes d'échantillonnage afin d'estimer les quantités totales des œufs, des classes d'âge 0 ou 1 et des adultes sexuellement mûrs. Dans notre travail, on a fait l'estimation globale à chacune de ces 3 étapes, proposant un calcul géostatistique pour la variance d'estimation.

L'estimation de la quantité totale est au centre de la Science des Pêches. Non seulement elle est un paramètre économique, mais surtout elle est un paramètre biologique central de la dynamique de la population. L'essentiel de l'information quantitative pour l'estimation des quantités totales est fournie par les débarquements des pêches qui nourrissent les modèles de dynamique. Toutefois, l'estimation directe, c'est-à-dire par des campagnes d'échantillonnage en mer, est nécessaire non seulement dans le contexte du recrutement, mais aussi plus généralement pour fournir une information non biaisée et indépendante de la pêche. Le calcul de la variance de l'estimation globale pour les différents plans d'échantillonnage nous est apparu comme le premier grand apport de la Géostatistique à la Science des Pêches. Nous avons consacré la moitié du présent travail à ce sujet. Nous avons donc fait choix d'analyser par la Géostatistique les données issues des campagnes d'échantillonnage. Toutefois, l'application à d'autres types de données pourrait être fructueuse, comme par exemple aux

données des " log books " des patrons pêcheurs car elles sont implantées dans l'espace de façon très agrégée. Nous ne traiterons que des observations directes par campagnes d'échantillonnage.

Pourquoi faire appel à la Géostatistique ? Nous y voyons deux raisons principales. D'une part, les poissons présentent un certain degré de structuration spatiale et les valeurs des échantillons sont spatialement corrélées. D'autre part, les échantillons ne sont pas forcément prélevés indépendamment les uns des autres. La statistique classique suppose un modèle structural spatial très particulier et requiert l'indépendance des prélèvements. Nous utiliserons donc tout particulièrement la Géostatistique pour les échantillonnages à maille régulière. Pour l'estimation globale, c'est surtout le calcul de la variance d'estimation qui nous retiendra ; pour la cartographie, ce sera l'estimation. L'estimateur que propose le krigeage au point x_0 non échantillonné est une moyenne pondérée des échantillons pris aux points x_α : $Z(x_0) = \sum_\alpha \lambda_\alpha Z(x_\alpha)$. Les poids λ_α sont fonction des corrélations spatiales entre x_0 et les x_α d'une part et d'autre part, de la position relative de x_0 avec les x_α . Les poids λ_α sont choisis de telle sorte que l'estimateur de krigeage soit sans biais et de variance minimale. On appelle structure spatiale, les corrélations spatiales telles que décrites par une fonction de covariance ou un variogramme. Le variogramme mesure la variabilité moyenne entre 2 points x et $x+h$ en fonction de leur distance

vectorielle h : $\gamma^*(h) = \frac{1}{2n_h} \sum_i (Z(x_i) - Z(x_i+h))^2$. On infère un modèle de variogramme,

en moyenne sur toute la zone prospectée. Puis à chaque endroit de la zone, ce modèle moyen est appliqué. Une hypothèse d'homogénéité structurale est donc utilisée. Pourtant, on peut l'affaiblir au minimum en ne supposant l'homogénéité que pour les petites distances. On infère alors un modèle moyen valide que pour les petites distances, modèle dit local. Un modèle local suffit pour la cartographie comme pour le calcul de la variance de l'estimation globale.

Dans la première partie, on traite de l'estimation globale à partir de campagnes d'échantillonnage. On peut estimer sous les hypothèses convenables, la moyenne des échantillons par krigeage. C'est une moyenne pondérée sans biais et dont les poids minimisent la variance d'estimation. A chaque échantillon, on attribue un poids qui est fonction de la corrélation entre les points et de la position des échantillons dans le volume à estimer. Une telle pondération est intéressante quand l'implantation des échantillons est agrégée, sans être préférentielle. Or, pour les plans d'échantillonnage stratifiés ou à maille régulière, une telle situation n'a pas lieu. On utilisera donc la moyenne arithmétique des données comme estimateur de la moyenne sur le champ et on s'intéressera au calcul de la variance d'estimation. Il nécessite d'inférer un modèle structural d'où l'intérêt de la Géostatistique.

Dans un second temps, on aborde le problème des fluctuations d'échantillonnage. Bien qu'en Géostatistique linéaire aucune hypothèse ne soit faite sur l'histogramme, sa dissymétrie marquée rend les fluctuations d'échantillonnage

importantes tant sur l'estimateur que sur le modèle structural inféré. La variance d'estimation calculée dépend alors de l'échantillonnage spécifique effectué et dans une moindre mesure, de la régionalisation. On aborde ce problème de deux façons. Disposant du modèle structural vrai, on donne des critères pour caler un effort d'échantillonnage approprié. D'autre part, on discute de l'intérêt d'utiliser un estimateur non linéaire de la moyenne plutôt qu'arithmétique car ce dernier est moins robuste aux fluctuations d'échantillonnage de la queue de l'histogramme. Mais la quantité totale est une grandeur arithmétique. Les approches arithmétique et spatiale sont les plus satisfaisantes avec un effort d'échantillonnage approprié aux structures spatiales.

Dans la seconde partie, on étudie les cartes d'un point de vue biologique. On pose le problème de la relation dynamique entre la quantité totale et la régionalisation des densités (dispersion des valeurs et localisation des zones riches). On a choisi de situer cette étude au niveau des œufs planctoniques et de la classe d'âge-0 afin d'aborder le déterminisme du recrutement de façon spatiale. On a travaillé sur deux bases de données spatio-temporelles pour lesquelles un modèle de fonction aléatoire est inféré. Grâce aux courbes de sélectivité géostatistique, on propose un test simple afin de distinguer trois types de stratégie d'occupation de l'espace par les poissons, en fonction de la quantité totale.

En dernier lieu, on conclut sur les variations temporelles dans le contexte desquelles on situe notre travail.

PARTIE I

ESTIMATION DIRECTE DES STOCKS

On s'intéresse à l'estimation directe de l'abondance totale d'une population de poissons sur son aire de distribution, à l'aide de campagnes d'échantillonnages et indépendamment des débarquements de la pêche. Il s'agit d'estimer une intégrale

d'espace $Q = \iint f(x, y) dx dy$ où $f(x, y)$ est la densité de poissons au point (x, y) . En

dehors de son domaine de présence (habitat du stock), le poisson a une densité $f(x, y)$ nulle.

L'échantillonnage est effectué à un moment déterminé du cycle de vie, sur une aire maritime donnée, pendant une durée donnée. Cette fenêtre spatio-temporelle de l'observation aura été définie au préalable selon des critères biologiques. Cette connaissance qualitative biologique et préalable est d'une grande importance car elle détermine la validité et l'intérêt de nos calculs. Si elle n'est pas notre sujet, on cherchera à s'assurer de l'adéquation spatio-temporelle de l'échantillonnage au cycle de vie du poisson.

Un plan d'échantillonnage permet de partager l'espace en blocs, chacun étant prospecté par un (ou plusieurs) échantillon(s). La valeur de l'échantillon du bloc v_i , ou de la moyenne des échantillons du bloc, estime la densité moyenne z_{vi} dans le bloc. Q est estimée par une somme spatiale discrète : $Q^* = \sum_i v_i z_{vi}^*$.

De plus, le plan d'échantillonnage permet de spécifier comment sont construites les erreurs d'estimation dans l'espace. La qualité de l'estimation est mesurée par une variance $\sigma_{est}^2 = E(Q - Q^*)^2$. Pour calculer σ_{est}^2 on ne dispose que d'une valeur Q^* , celle que nous donne un échantillonnage. Dans la théorie de l'échantillonnage, comme exposée dans MATHERON (1965, 1970) ou COCHRAN (1977), l'implantation aléatoire des prélèvements joue un rôle important : elle permet de probabiliser l'erreur d'estimation $Q - Q^*$. La variance $E(Q - Q^*)^2$ représente la moyenne du carré des erreurs d'estimation de tous les échantillonnages possibles. Si la distribution spatiale du poisson est très erratique, les implantations différentes des échantillons lors de deux échantillonnages différents donneront des valeurs très différentes, et l'on s'attend à ce que les estimations données par chaque échantillonnage soient très différentes. La variance σ_{est}^2 est une mesure de cette différence. Elle est calculée à partir d'un échantillonnage unique. Si les échantillons sont prélevés indépendamment les uns des autres selon une loi uniforme dans l'espace, la théorie classique (COCHRAN 1977) permet de calculer une variance d'estimation sans qu'il soit nécessaire de prendre en compte les corrélations spatiales (i.e. structure spatiale). L'information structurale présente dans les données est ignorée. La Géostatistique propose d'en tenir compte. Si les échantillons ne sont pas implantés

indépendamment les uns des autres, le calcul de la variance d'estimation σ_{est}^2 requiert nécessairement un modèle de structure spatiale, d'où l'intérêt de la Géostatistique.

Techniquement, historiquement et statistiquement (calcul de la variance d'estimation), on peut séparer les méthodes d'échantillonnage en deux groupes. Techniquement d'abord, on sépare les méthodes d'échantillonnage où les échantillons sont prélevés séparément les uns des autres de celles où les échantillons ne le sont pas. Au premier groupe figurent le grand nombre des campagnes de chalutage de fond sur les plateaux continentaux : chaque coup de chalut est effectué indépendamment d'un autre. Figurent aussi dans ce groupe les campagnes de coups de filet à ichtyoplancton. Au deuxième groupe appartiennent les échantillonnages de densité de poissons par ondes acoustiques, les échantillonnages de plancton par pompage ou par filets de type LHPR (Longhurst Hardy Plankton Recorder). Ces échantillonnages se font en continu le long de la route du navire, selon des radiales ou transects, ou selon des lignes brisées.

Historiquement, les méthodes d'investigation de l'océan appartenant au second groupe ne sont pas issues directement des techniques ancestrales de la pêche, comme le sont les chaluts, mais viennent du monde scientifique. Ce sont, à ce titre, des méthodes "modernes" dont la validité a dû être établie. Elles permettent la prospection rapide de larges aires marines et sont, de ce fait, adaptées à fournir une image la plus synoptique des mondes planctoniques et pélagiques qui en général se distribuent en petits essais très denses dispersés sur de vastes zones.

Sur le plan statistique du calcul de la variance de l'estimation globale, la séparation précédente ne vaut que par le plan d'échantillonnage qu'impose la technique de prélèvement. Les méthodes du premier groupe sont compatibles avec les principaux plans d'échantillonnage : plan aléatoire simple (terminologie classique) ou plan aléatoire pur (terminologie géostatistique), plan aléatoire stratifié, plan systématique (terminologie classique) ou plan à maille régulière (terminologie géostatistique). En revanche, les méthodes du second groupe sont incompatibles avec les plans aléatoires simples et stratifiés. Leur plan d'échantillonnage orthodoxe est le plan systématique.

Lorsque les localisations des prélèvements sont corrélées par la prise de mesure en série, comme dans un plan systématique, le calcul de la variance σ_{est}^2 doit prendre en compte la structure spatiale. Ce calcul est resté un sujet de controverse dans la science des pêches. A notre connaissance, dans le domaine des estimations globales par méthode acoustique, deux auteurs (WILLIAMSON, 1982 et LALOE, 1985) ont proposé des solutions partielles qui prennent en compte les structures spatiales le long des radiales. Pourtant, ce calcul a été résolu par MATERN (1960) dans les sciences agronomiques, et d'une manière plus formelle et générale par MATHERON (1965) dans les Sciences de la Terre : la variance d'estimation s'exprime en fonction de la couverture spatiale réalisée par l'échantillonnage (localisation des échantillons), et en fonction de la régularité de la distribution spatiale du poisson (structure spatiale).

A. GÉOSTATISTIQUE DE LA VARIANCE D'ESTIMATION

La variance d'estimation calculée dans cette partie ne reflète que la variance de l'erreur due au fait que l'on ne prélève qu'une information fragmentaire de la réalité. Elle ne prend nullement en compte les biais d'échantillonnage inhérents aux techniques d'échantillonnage en mer qui peuvent être importants. Toutefois, la variabilité provenant d'erreurs non systématiques dans les mesures contribue à l'hétérogénéité des échantillons et se mélange à l'hétérogénéité spatiale stricte. Si des hypothèses sur l'erreur de mesure sont réalistes, ce qui peut être le cas pour certaines campagnes de chalutage de fond, un filtrage de cette erreur pourrait être envisagé. Nous n'avons pas abordé cette possibilité.

On se place ici dans le cas d'un échantillonnage effectué selon un plan bien défini qui discrétise l'espace. La quantité totale est estimée par sommation spatiale directement à partir des échantillons. On fournit les formules géostatistiques (MATHERON, 1965, 1970 et 1989) des variances d'estimation pour différents plans d'échantillonnage, en présentant les concepts mis en jeu. On applique ensuite la méthodologie à une prospection acoustique effectuée le long de transects parallèles et équidistants.

A.1. Plan systématique

Dans le plan systématique, les échantillons sont prélevés en série selon une grille régulière dont l'origine x_0 est implantée aléatoirement et uniformément dans la maille élémentaire. L'échantillonnage est sans biais car il est indépendant des valeurs prises par la variable échantillonnée. On présente deux classes de méthodes géostatistiques à cause des propriétés géométriques qu'elles mettent en jeu : la théorie transitive et la théorie intrinsèque.

A.1.1. Théorie transitive

La théorie transitive a été développée (MATHERON, 1965) pour l'estimation globale par échantillonnage à maille régulière. Elle fournit une expression exacte de la variance de cette estimation, ainsi que des principes d'approximation pour le calcul pratique.

Soit $f(x)$ la variable régionalisée de la densité de poissons au point x . $f(x)$ désigne une valeur déterministe sans interprétation probabiliste. On expose les calculs à 1 dimension par simplicité et parce que nous n'utiliserons que le calcul à 1D dans l'application qui suivra.

$$\text{La quantité totale est : } Q = \int_{-\infty}^{+\infty} f(x) dx$$

$$\text{L'estimateur de } Q \text{ est : } Q(x_0)^* = a \sum_{k=-\infty}^{+\infty} f(x_0 + k a)$$

où x_0 est l'origine de la grille d'échantillonnage, a la longueur de la maille. En dehors de sa zone de présence, le poisson a une densité $f(x)$ nulle. Les zéros extérieurs auront dû être échantillonnés.

Relativement à l'implantation aléatoire et uniforme de x_0 dans la maille élémentaire $[0, a]$, l'estimateur $Q^*(x_0)$ et sa variance $\sigma^2(a)$ sont des variables aléatoires.

$$Q^*(x_0) \text{ est sans biais : } E(Q^*) = \frac{1}{a} \int_0^a Q^*(x_0) dx_0 = Q$$

$\sigma^2(a)$, la variance d'estimation, s'écrit comme une moyenne de l'écart quadratique

$$(Q - Q^*(x_0))^2 : \sigma^2(a) = \frac{1}{a} \int_0^a (Q^*(x_0) - Q)^2 dx_0.$$

En exprimant $Q^*(x_0)$ et Q à l'aide de $f(x)$, le calcul (MATHERON, 1965 pp 72-73) fait apparaître une fonction structurale $g(h)$. La variance s'écrit finalement :

$$\sigma^2(a) = a \sum_{k=-\infty}^{+\infty} g(ka) - \int_{-\infty}^{+\infty} g(h)dh \quad (I.1)$$

où $g(h) = \int_{-\infty}^{+\infty} f(x) f(x+h) dx$ est le covariogramme transitif. $g(h)$ est une sorte de covariance non centrée : c'est l'outil structural.

Expérimentalement, le covariogramme transitif se calcule en translatant la grille d'échantillonnage sur elle-même d'un vecteur h et en sommant tous les produits non nuls $f(x) f(x+h)$. Pour calculer $\sigma^2(a)$, le recours à un modèle de covariogramme transitif est

nécessaire à cause du terme $\int_{-\infty}^{+\infty} g(h) dh$ qui sans modèle n'est pas calculable.

La variance de l'estimation totale est une différence entre une intégrale du covariogramme transitif et son approximation discrète sur la grille d'échantillonnage. La formule montre que la variance ne dépend que de la taille de la maille et de la régularité de la distribution spatiale paramétrée par $g(h)$.

Dans la théorie transitive, l'approche est la même que dans la théorie de l'échantillonnage. Aucune hypothèse probabiliste n'est faite sur la régionalisation de la variable étudiée. Ce sont, selon le type d'échantillonnage, les propriétés statistiques du réseau des prélèvements qui permettent de mener les calculs, quelle que soit la régionalisation de la variable étudiée. Deux propriétés des grilles sont utilisées : d'une part l'implantation aléatoire et uniforme de l'origine dont on a déjà vu l'importance méthodologique, et d'autre part la régularité de l'implantation des échantillons dans l'espace. Cette seconde caractéristique permet le calcul du covariogramme transitif. Il faut en effet que le réseau des points informants soit invariant par translation.

L'expression du covariogramme transitif est très proche de celle d'une covariance stationnaire, et l'inférence d'un modèle soulève des difficultés statistiques similaires. MATHERON (1965, 1989) explique les rapports entre les deux fonctions structurales.

Appelons champ la zone de présence du poisson lors de l'échantillonnage. Soit $k(x)$ l'indicatrice du champ : $k(x)$ vaut 1 à l'intérieur du champ et 0 à l'extérieur. Soit $K(h)$ le covariogramme transitif de $k(x)$: c'est le covariogramme géométrique de la variable $f(x)$. On nomme $S(h)$ l'intersection du champ avec sa translaturée de vecteur h , $K(h)$ mesure la

surface de $S(h)$. Dans $S(h)$, les densités de poissons $f(x)$ et $f(x+h)$ sont toutes deux non nulles. Appelons $C_R(h)$ la covariance non centrée expérimentale de $f(x)$ et $g(h)$ le covariogramme transitif de $f(x)$.

Expérimentalement, $C_R(h)$ s'exprime comme une moyenne :

$$C_R(h) = \frac{1}{K(h)} \int_{S(h)} f(x) f(x+h) dx = \frac{g(h)}{K(h)} \text{ et } g(h) = C_R(h).K(h)$$

Le covariogramme transitif est donc une covariance non centrée dans laquelle intervient la géométrie du champ. Le covariogramme transitif $g(h)$ reflète les propriétés géométriques du champ –par $K(h)$ et $S(h)$ – ainsi que les propriétés structurales de la régionalisation propre au poisson –par $C_R(h)$ – sans que les deux effets ne soient séparés.

Quand la géométrie du champ et la régionalisation propre au poisson sont intimement liées, la séparation entre des deux n'est pas possible. C'est le cas des régionalisations (d'un type statistique non stationnaire) présentant des zonéographies. Par exemple, la géométrie d'un sédiment favorable, ou celle particulière d'une isobathe ou d'un fjord, peut dicter au poisson une répartition spatiale zonée. La méthode transitive peut être un outil très utile et puissant pour le suivi routinier de la quantité totale d'un stock à partir de campagnes à la mer selon une même grille de surveillance facile à mettre en œuvre, sans que la description intime des régionalisations propres au poisson ne soit un sujet principal de préoccupation.

A.1.2. Théorie intrinsèque

A.1.2.1. Hypothèses

Dans la théorie intrinsèque, la séparation est effectuée entre la géométrie du champ et la régionalisation propre –ou "intrinsèque"– au poisson. Physiquement, les modèles intrinsèques rendent compte de régionalisations où les limites du champ semblent découper une distribution spatiale qui pourrait se poursuivre au-delà de cette frontière. Au voisinage des bordures du champ, la régionalisation qui peut être zonée ou non, n'est pas modifiée par un effet de bordure (Figure 1).

Par rapport à l'analyse en transitif, dans l'analyse probabiliste de la théorie intrinsèque, deux hypothèses concernant la répartition spatiale propre à la variable sont introduites. En premier lieu, il s'agit de l'indépendance interne entre la variable et son champ. Cette propriété physique, qui conditionne l'utilisation de la théorie intrinsèque, est difficile à vérifier rigoureusement dans la pratique. MATHERON (1965, chap. V) en donne une expression qui peut servir de test (p. 100) :

" $m = \frac{1}{2} [m(h) + m(-h)]$. Il y a indépendance interne entre une variable et son champ si

la moyenne des valeurs, prises par la variable dans les intersections du champ V avec ses translatées h et $-h$, reste constante et égale à la moyenne générale m ".

La séparation fonctionnelle entre la variable et son champ –certes fructueuse car elle permettra l'interpolation– pose ici, en ce qui concerne l'estimation globale, le problème

supplémentaire du calcul de l'erreur faite sur les limites du champ, en plus de celle faite sur la moyenne. On devra calculer un terme d'erreur géométrique qui représente l'effet, sur l'estimation de la moyenne, de l'erreur faite sur l'estimation des limites du champ. Le calcul est possible grâce à l'hypothèse d'indépendance interne.

La seconde hypothèse introduite est celle d'un certain degré de stationnarité de la régionalisation. Elle permettra, à champ fixé, de calculer la variance d'estimation de la moyenne. Cette hypothèse est liée à l'interprétation probabiliste de la variable régionalisée. La distribution spatiale est interprétée comme une réalisation d'une fonction aléatoire $Z(Z(x_1), \dots, Z(x_n), \dots)$, à l'intérieur d'une géométrie fixée². Comme un échantillonnage ne fournit qu'une réalisation de la fonction aléatoire, l'inférence des paramètres de Z n'est pas possible sans faire des hypothèses de stationnarité sur la loi spatiale. En effet, si l'on considère que la loi spatiale se répète égale à elle-même dans différents endroits de l'espace, alors les observations aux différents endroits sont utilisées comme autant de réalisations différentes du même phénomène. La stationnarité est une hypothèse méthodologique nécessaire, mais la stationnarité n'est pas indispensable. Il y a différents degrés de stationnarité: stationnarité de la variable elle-même (cas stationnaire), stationnarité des accroissements (cas intrinsèque, MATHERON 1965,70), stationnarité des accroissements d'ordre k (cas intrinsèque d'ordre k , MATHERON 1973). On verra au paragraphe A.3. comment l'hypothèse de stationnarité peut être réduite au strict nécessaire méthodologique (quasi-stationnarité).

A.1.2.2. Formule générale

Le modèle que nous utiliserons abondamment est le modèle intrinsèque d'ordre 0 : les incréments $Z(x) - Z(x+h)$ sont stationnaires, ont une moyenne nulle et possèdent une variance dépendant uniquement du vecteur h , non des points d'appui x et $x+h$:

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

$$E(Z(x) - Z(x+h))^2 = 2\gamma(h)$$

$\gamma(h)$ est le variogramme, l'outil structural. Le variogramme expérimental se calcule en prenant toutes les paires d'échantillons $(z(x), z(x+h))$ pour un vecteur h donné et en calculant la demi-moyenne des différences carrées $(z(x) - z(x+h))^2$.

Plutôt que des intégrales d'espace, on manipule, dans la théorie intrinsèque, des moyennes. La quantité à estimer est la moyenne de Z sur le champ V :

$$Z_V = \frac{1}{V} \int_V Z(x) \, dx . Z_V \text{ est estimée par une moyenne pondérée des échantillons. Dans}$$

le cas où l'échantillonnage est effectué selon une grille régulière, on donne à chaque échantillon une même zone d'influence : pour estimer Z_V on prend la moyenne

2. La notation $z(x)$ permet de différencier la variable régionalisée $f(x)$ de son interprétation en tant que réalisation d'une fonction aléatoire $Z(x)$. $Z(x)$ dénomme la fonction aléatoire, le modèle dont les $z(x)$ sont une réalisation.

arithmétique des données. On note Z_V^* cet estimateur afin de souligner qu'il s'agit de l'estimateur d'une moyenne d'espace. La variance de cette estimation est : $\sigma_E^2 = E(Z_V - Z_V^*)^2$.

Pour différentes réalisations de Z , le réseau des prélèvements restant le même, les estimateurs Z_V^* ont des valeurs différentes. σ_E^2 a la signification de la moyenne sur toutes les réalisations de la fonction aléatoire des écarts carrés $(Z_V - Z_V^*)^2$. En intrinsèque, on n'a pas à supposer que l'implantation des échantillons est "aléatoire" pour calculer la variance d'estimation. Mais l'implantation "aléatoire" confère à l'échantillonnage son caractère non préférentiel. Le calcul de la variance d'estimation n'est possible que grâce à l'hypothèse de stationnarité de $Z(x)$ ou des accroissements $(Z(x) - Z(x+h))$.

Dans le cas stationnaire, σ_E^2 se développe :

$$\sigma_E^2 = E(Z_V - Z_V^*)^2 = E(Z_V^2) - 2E(Z_V Z_V^*) + E(Z_V^*)^2$$

Chacun des trois termes s'exprime en fonction de la covariance.

$$\sigma_E^2 = \bar{C}(V, V) - 2\bar{C}(S, V) + \bar{C}(S, S)$$

(I.2)

où :

$\bar{C}(V, V)$ est la covariance moyenne entre 2 points décrivant V indépendamment l'un de l'autre,

$\bar{C}(S, V)$ est la covariance moyenne entre 2 points décrivant indépendamment l'un l'ensemble S des positions des échantillons et l'autre le volume V ,

$\bar{C}(S, S)$ est la covariance moyenne entre 2 points décrivant indépendamment l'ensemble S des positions des échantillons.

Dans le cas stationnaire, le variogramme présente un palier et il est lié à la covariance : $\gamma(h) = C(0) - C(h)$. En réécrivant la formule (I.2) avec le variogramme, la constante $C(0)$ s'élimine:

$$\sigma_E^2 = -\bar{\gamma}(V, V) + 2 \bar{\gamma}(S, V) - \bar{\gamma}(S, S) \quad (I.3)$$

Cette formule générale reste vraie dans le cas intrinsèque quand la variance $C(0)$ n'est plus définie dans le modèle. On a :

$$\bar{\gamma}(V, V) = \frac{1}{V^2} \int_V \int_V \gamma(x-y) dx dy$$

$$\bar{\gamma}(S, V) = \frac{1}{nV} \sum_a \int_V \gamma(x_a - x) dx$$

$$\bar{\gamma}(S, S) = \frac{1}{n^2} \sum_a \sum_\beta \gamma(x_a - x_\beta)$$

où V est l'aire du champ, (α, β) les indices des échantillons et n le nombre d'échantillons dans V .

La formule (I.3) un peu plus compliquée que celle de la théorie transitive, lui ressemble pourtant beaucoup: elle a la même signification. La formule (I.3) se différencie de la formule (I.1) en particulier par la présence du terme $\bar{\gamma}(S, V)$. Il représente l'effet sur l'estimation, de la disposition des échantillons par rapport à la géométrie du champ V car ici les implantations des échantillons sont considérées fixes. La variance σ_E^2 est une différence entre une intégrale double de la fonction de structure ($\bar{\gamma}(V, V)$) et ses deux approximations discrètes sur la grille d'échantillonnage ($\bar{\gamma}(S, V)$ et $\bar{\gamma}(S, S)$). $\bar{\gamma}(V, V)$ et $\bar{\gamma}(S, V)$ ne sont pas calculables sans recours à un modèle de variogramme. La variance σ_E^2 dépend de la structure spatiale au travers de trois géométries : de la géométrie du champ –par $\bar{\gamma}(V, V)$ –, de la disposition des échantillons les uns par rapport aux autres (maille du réseau) –par $\bar{\gamma}(S, S)$ –, et de la position du réseau dans le champ –par $\bar{\gamma}(S, V)$.

A.1.2.3. Champ inconnu

La formule (I.3) est générale et ne s'applique pas qu'au plan systématique : elle donne la variance de l'estimation de Z_V par la moyenne arithmétique des échantillons dont les implantations sont fixes dans une boîte V aux contours connus. Or lors de l'échantillonnage d'un stock de poisson, on se trouve dans une situation où les limites de l'aire de présence du poisson (limites du champ V) ne sont pas connues exactement avant l'échantillonnage. Elles sont, au contraire, estimées à partir des échantillons. Le réseau de prélèvement a une implantation "flottante" par rapport à la géométrie de V (plan systématique). Considérons que la régionalisation se laisse interpréter par un schéma intrinsèque à l'intérieur du champ V . En dehors, les valeurs sont nulles. Pour les grilles à implantation "flottante", en toute rigueur, la variance d'estimation intrinsèque de Z_V par Z_V^* se calcule grâce à l'approche transitive en utilisant la formule reliant le covariogramme transitif au variogramme (MATHERON, 1965). Sous l'hypothèse d'indépendance interne entre la variable et son champ, les estimations de Z_V et de V sont indépendantes l'une de l'autre. En intrinsèque, la variance d'estimation de la moyenne sur V s'écrit comme la somme de deux termes :

$$\sigma_m^2 = D_{(O/V)}^2 \frac{\sigma_V^2}{V^2} + \sigma^2 \quad (I.4)$$

$D_{(O/V)}^2$ est la variance de dispersion d'un point dans V . Dans le modèle, $D_{(O/V)}^2$ est égale à la valeur moyenne du variogramme pour toutes les distances dans V :

$$D_{(O/V)}^2 = \bar{\gamma}(V, V) \cdot (3)$$

σ_V^2 est la variance transitive d'estimation de la surface de V . Les zéros extérieurs auront dû être échantillonnés.

$D_{(O/V)}^2 \frac{\sigma_V^2}{V^2}$, variance relative d'estimation de V en transitif multiplié par la dispersion moyenne dans V des teneurs, traduit un effet de bordure : c'est le terme d'erreur géométrique. Il mesure la répercussion de l'estimation des limites de V , sur l'estimation de la teneur moyenne dans V . En général, le terme d'erreur géométrique ne sera pas négligeable.

3. Expérimentalement, on a :

$\bar{\gamma}(S, S) = \frac{1}{n^2} \sum_i \sum_j (z_i - z_j)^2 = \frac{1}{n} \sum_i (z_i - \frac{1}{n} \sum_j z_j)^2 = s^2$. La variance dans l'échantillon est égale à la moyenne du variogramme pour les distances entre les échantillons.

σ^2 est une variance d'estimation à champ connu. MATHERON (1965, chap XII) montre que σ^2 ne diffère pas de la variance σ_E^2 que l'on calculerait avec la formule (I.3) en considérant que la grille d'échantillonnage a une implantation fixe et préférentielle dans le champ (terminologie de MATHERON). Ayant échantillonné les zéros extérieurs, on peut définir les limites d'un polygone à partir des échantillons en ne retenant que les échantillons non nuls. La grille a alors une implantation "préférentielle" par rapport aux limites définies : par exemple, les échantillons sont au centre de leur zone d'influence (dispositif centré) et le champ défini est la réunion des mailles élémentaires dans lesquelles les échantillons sont positifs.

Dans la pratique, on peut donc procéder en trois étapes :

- définir un champ V en reconnaissant les zéros extérieurs,
- calculer la variance d'estimation avec la formule (I.3) comme si V était connu,
- ajouter enfin le terme d'erreur géométrique.

Ces calculs se font facilement grâce aux formules d'approximation développées par MATHERON (1965).

Pour une grille à implantation flottante, sous réserve que l'hypothèse d'indépendance interne soit physiquement acceptable, le calcul en intrinsèque selon la procédure précédente donnera un résultat similaire au calcul en transitif. C'est bien ce que nous constaterons sur nos données dans l'application numérique au chapitre I.B. suivant.

A.1.3. Principes d'approximations

Les formules des variances d'estimation (I.1) et (I.3) peuvent être calculées par intégration numérique. Toutefois, quand la maille est petite vis-à-vis du champ, les sommes discrètes présentent un grand nombre de termes. Les intégrales sont aussi calculées numériquement par discrétisation, si bien que la précision et la rapidité du calcul ne sont pas forcément assurées. Ce problème se rencontre pour les échantillonnages en continu le long de la route du navire, comme par exemple dans la méthode acoustique où le pas d'échantillonnage peut être très fin. On verra de plus au paragraphe suivant (A.3.) que dans le cas intrinsèque, on ne dispose pas, en général, d'un modèle de variogramme valide pour les grandes distances dans le champ. Ces raisons donnent aux approximations une grande utilité pratique. De plus, justifiés théoriquement, les principes d'approximation ont une grande importance méthodologique. Ils permettent d'affirmer que pour calculer la variance d'estimation, on n'a besoin que d'inférer le comportement de la fonction de structure ($g(h)$ ou $\gamma(h)$) pour les petites distances.

La taille de la maille de la grille, relativement au champ et à la portée des structures spatiales, définit les limites de validité des approximations.

A.1.3.1. Approximations dans \mathbb{R}^1

Les formules (I.1) et (I.2), en transitif et en intrinsèque, sont parentes et ce sont les mêmes principes qui fondent les règles d'approximation dans chaque cas. La variance est une différence entre valeurs exactes et approchées d'une intégrale.

MATHERON (1965) exprime la variance d'estimation comme le reste dans la formule d'Euler Mac Laurin. Pour ce faire, toutefois, les modèles de $g(h)$ et de $\gamma(h)$ ne sont pas assez réguliers partout : il concentrent les irrégularités analytiques au voisinage de l'origine ($h=0$) et au voisinage de la portée ($h=b$). La formule d'Euler Mac Laurin s'applique par intervalle et la variance peut s'écrire comme la somme de deux termes :

$$\sigma^2(a) = T_0(a) + T_Z(a)$$

où a est la maille du réseau régulier de l'échantillonnage.

$T_0(a)$ est le terme d'extension, il dépend du comportement de la fonction de structure au voisinage de l'origine ($h=0$) ; $T_Z(a)$ est le terme fluctuant, il dépend du comportement de la structure au voisinage de la portée ($h=b$).

En transitif (MATHERON 1965, p.79), le terme fluctuant est indéterminé. Mais comme il est nul en moyenne, il ne sera pas considéré. En intrinsèque (MATHERON 1965, p.201), le terme fluctuant est négligeable. Il reste donc à calculer le terme d'extension.

A.1.3.1.1. Principes de correspondance : maille très petite

En transitif comme en intrinsèque, on retrouve la même façon de procéder. Les modèles courants admettent un développement limité au voisinage de $h=0$. Quand la maille a est petite, ce développement est utilisable jusqu'en $h=a$. MATHERON (1965) montre que le terme T_0 admet alors un développement limité en $a=0$. Chacun des termes en $|h|^\lambda$ du développement limité de la fonction de structure apporte au développement limité de la variance T_0 un terme en a^k . Une règle de correspondance terme à terme permet de former en pratique le développement de T_0 . Précisons que la partie "régulière" du modèle de $\gamma(h)$ ou $g(h)$ –termes correspondant aux puissances paires h , dans le développement limité– ne contribuent en rien à la variance. Ce n'est que la partie "irrégulière" du développement limité de la fonction de structure –termes en h^{2k+1} et en $h^{2k} \log h$ – qui a une contribution à la variance d'extension $T_0(a)$.

Voilà donc où réside l'importance du comportement analytique du modèle choisi de variogramme ou de covariogramme transitif, au voisinage de l'origine $h=0$. Un modèle gaussien, parabolique à l'origine, et un modèle sphérique, linéaire à l'origine (moins "régulier"), donneront des variances d'estimation très différentes. Nous reviendrons sur ce point très concret (paragraphe A.3.3.).

A.1.3.1.2. Principe de composition des variances élémentaires : maille pas très petite

Quand la maille a n'est plus très petite, les développements limités précédents de $\gamma(h)$ et de $g(h)$ ne sont plus utilisables jusqu'en $h = a$. En transitif, le nombre de termes de la somme dans la formule exacte n'est plus si important et le calcul direct est raisonnable. En intrinsèque, MATHERON (1965, chap. XII) justifie une autre approximation valable pour les grilles à dispositif centré : la composition des variances élémentaires. La maille a reste inférieure à la portée.

Dans chaque zone d'influence v_i , l'échantillon x_i est au centre et tous les v_i sont identiques. Soit ε_i l'erreur commise dans v_i . L'erreur commise dans l'estimation de la

moyenne globale est : $\frac{1}{N} \sum_i \varepsilon_i$. D'où, en termes de variance :

$$\sigma_N^2 = \frac{1}{N} \sigma_E^2 + \frac{2}{N^2} \sum_{i < j} E(\varepsilon_i \varepsilon_j)$$

$\sigma_E^2 = E(\varepsilon_i^2)$ est la variance élémentaire. C'est la variance d'estimation de la moyenne de v_i par la valeur de l'échantillon x_i (extension de x_i à v_i). σ_E^2 se calcule en utilisant la formule (I.2) : $\sigma_E^2 = 2\bar{\gamma}(x_i, v_i) - \bar{\gamma}(v_i, v_i)$

Les $E(\varepsilon_i \varepsilon_j)$ sont des covariances d'erreur. L'approximation consiste à les négliger (MATHERON, 1965 p. 203). Quand la maille n'est pas très petite mais inférieure à la portée, les erreurs d'estimation élémentaires commises en étendant la teneur de chaque échantillon à sa zone d'influence, peuvent être considérées comme non corrélées.

Quand la maille est grande, supérieure à la portée, on devra utiliser d'autres formules d'approximation (MATHERON, 1970 p. 107) que l'on ne détaille pas ici et qui sont valables dans R^2 .

A.1.3.2. Approximations dans R^2

En transitif, MATHERON (1965, pp 84–87) établit le principe de composition des termes de ligne et de tranche. Considérons que le réseau d'échantillonnage présente, dans une direction, une densité plus grande de prélèvements. La variance d'estimation à 2D s'exprime comme la somme de deux variances à 1D :

$$\begin{array}{ccc}
 0 & 0 & 0 \\
 0 & 0 & 0 \\
 0 & 0 & 0 \\
 0 & 0 & 0 \\
 a_2 \left| \begin{array}{ccc} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right. & & \sigma^2(a_1, a_2) = \sigma_1^2(a_1) + a_1 \sigma_2^2(a_2) \\
 \xrightarrow{a_1} & &
 \end{array}$$

$\sigma_1^2(a_1)$ est le terme de tranche. C'est la variance d'estimation dans les rectangles $a_1 \times n_i a_2$, comme si leurs lignes médianes i étaient exhaustivement reconnues. $\sigma_1^2(a_1)$ se calcule à 1D avec la maille a_1 en utilisant le covariogramme des données cumulées le long des lignes (covariogramme monté).

$a_1 \sigma_2^2(a_2)$ est le terme de ligne. $\sigma_2^2(a_2)$ est la variance d'estimation le long des lignes à partir des prélèvements. $\sigma_2^2(a_2)$ se calcule à 1D avec la maille a_2 en utilisant le covariogramme de $f(x)$.

Dans les échantillonnages en continu le long de transects parallèles et équidistants, comme c'est le cas dans les prospections acoustiques de poissons, on peut considérer qu'il n'y a pas d'erreur le long des lignes, car leur échantillonnage a lieu en continu. Il reste donc uniquement le terme de tranche à calculer. Comme la distance inter-transects n'est généralement pas très petite, le calcul exact de la formule (I.1) est rapide car il n'y a pas trop de termes. La méthode transitive est ici bien adaptée et facile à mettre en œuvre. On l'utilisera dans l'application qui suit.

En intrinsèque, on utilise le principe précédent et celui de la composition des variances élémentaires. Les échantillonnages en continu le long de transects conduisent à estimer des rectangles par un échantillonnage continu de leur ligne centrale. La composition des variances élémentaires donne :

$$\sigma^2 = \frac{\sum_i b_i^2 \sigma_i^2}{(\sum_i b_i^2)^2}$$

où b_i est la longueur du transect i et σ_i^2 la variance élémentaire d'estimation du rectangle i .

Pour les échantillonnages à prélèvements séparés selon un réseau présentant dans une direction, une densité de prélèvements plus dense, on utilisera le principe de composition des termes de ligne et de tranche. Chacun des termes se calcule en composant des variances élémentaires :

$$\sigma^2 = \underbrace{\frac{1}{N} \sigma_{e_1}^2}_{\text{terme de ligne}} + \underbrace{\frac{\sum_i b_i^2 \sigma_i^2}{(\sum_i b_i)^2}}_{\text{terme de tranche}}$$

où N est le nombre total d'échantillons et $\sigma_{e_1}^2$ la variance élémentaire d'estimation d'un segment par son centre.

Les échantillonnages à prélèvements séparés selon un réseau à maille carrée, conduisent à estimer des carrés par leur point central. La composition des variances donne :

$$\sigma^2 = \frac{1}{N} \sigma_{e_2}^2$$

où N est le nombre de prélèvements et $\sigma_{e_2}^2$ la variance élémentaire d'estimation du carré.

Les variances élémentaires ($\sigma_{e_1}^2$, $\sigma_{e_2}^2$, σ_i^2) se calculent grâce à la formule générale (I.3) appliquée dans la maille élémentaire v: $\sigma^2 = 2\bar{\gamma}(s, v) - \bar{\gamma}(v, v) - \bar{\gamma}(s, s)$. Pour calculer la variance d'estimation, il suffit donc d'inférer un variogramme pour les distances inférieures à la taille de la maille d'échantillonnage. Le calcul requiert les variances $\bar{\gamma}$ dans la maille v et non dans tout le champ V.

Les géométries des mailles précédentes sont toujours les mêmes. Dans chaque cas, les intégrales des fonctions $\bar{\gamma}$ ont été calculées pour différents modèles de variogrammes, et pour différentes valeurs de maille et de portée, et les variances élémentaires précédentes sont données sous formes d'abaques (MATHERON, 1970 ; JOURNEL, 1978). En pratique, il suffit donc de lire les abaques pour calculer la variance d'estimation globale. On trouvera dans JOURNEL (1978, pp 133-169) les détails du calcul des fonctions $\bar{\gamma}$.

Les réseaux réguliers à maille carrée ou à maille plus dense dans la direction d'hétérogénéité maximale (généralement la direction côte large) sont bien adaptés à l'exploration d'une zone mal connue géographiquement, ou à la surveillance pluriannuelle d'une zone bien délimitée. Les formules données sont donc applicables à des campagnes de chalutage de poissons de fond comme à celle de filets à ichtyoplancton. Dans la partie II, nous traiterons le cas de la surveillance d'une frayère de sole à partir de l'échantillonnage à maille carrée des œufs planctoniques. Les échantillonnages en continu le long de transects parallèles et équidistants représentent un bon nombre de campagnes acoustiques. Nous appliquerons les méthodes intrinsèque et transitive à une

telle prospection, sur du hareng. D'autre part, ces formules permettent facilement, connaissant la fonction de structure, de prévoir l'évolution de la variance d'estimation en fonction du nombre des échantillons ainsi que de leur emplacement.

A.2. Plan aléatoire stratifié

Dans la terminologie géostatistique (MATHERON 1965, 70), l'échantillonnage aléatoire stratifié consiste à implanter un échantillon aléatoirement dans chaque maille d'une grille régulière. Cette définition est différente de celle de COCHRAN (1977) à laquelle il est principalement fait référence en biologie des pêches. Dans la terminologie de la statistique classique (COCHRAN 1977), la stratification découpe la population à échantillonner en unités homogènes. Les limites des strates correspondraient à des frontières ou cassures naturelles dans la régionalisation. Dans chaque strate, la structure spatiale serait pépétique. Le modèle spatial sous-jacent au plan stratifié de la terminologie classique est une mosaïque où seules les moyennes de strate donnent l'aspect général de la carte. L'intérêt du plan stratifié (classique ou géostatistique) réside dans sa précision: il permet de ne manipuler que des variances intra-strate et jamais de variance inter-strates, d'où une réduction de la variance d'estimation par rapport au plan aléatoire pur (COCHRAN 1977, MATHERON 1965, 70). L'identification de strates homogènes conduit à réduire la variabilité spatiale intra-strate par rapport à la variabilité inter-strates et à renforcer l'effet précédent. Les limites des strates sont choisies en fonction d'une variable auxiliaire, grâce à une connaissance préalable de la régionalisation.

Le plan aléatoire stratifié de la terminologie classique est très couramment utilisé pour l'estimation des poissons de fond par chalutages. A cela, nous voyons trois raisons :

- la relative stabilité inter-annuelle des distributions spatiales à cause de leur lien avec des paramètres physiques stables sur le fond, ce qui permet de connaître qualitativement des zones riches et pauvres ;
- la possibilité de réduire la variance d'estimation facilement en tenant compte d'un effet proportionnel, grâce à une orientation de l'effort d'échantillonnage en implantant plus de prélèvements dans les strates supposées riches par expérience ;
- la simplicité du calcul de la variance d'estimation à cause de l'indépendance d'une strate à l'autre, des erreurs d'estimation.

Le découpage de l'espace (stratification) est opéré généralement en fonction de variables de l'environnement physique dont la régionalisation ne varie pas dans le temps. La stratification est, en général, effectuée hiérarchiquement en fonction de deux variables, d'abord selon la profondeur ou les sédiments, puis selon la latitude. Ces variables sont connues pour influencer en général la régionalisation des biomasses, toutes espèces confondues, sur le plateau continental. On ne rencontre pas ici le problème de la définition des limites du champ : les limites sont les limites extérieures des strates. En général, les strates n'ont pas la même surface, ni la même géométrie. Dans chaque strate, les stations de chalutage sont implantées aléatoirement et uniformément ; leur nombre

dépend de la richesse présumée de la strate. On analyse par la Géostatistique le plan stratifié de la terminologie classique. La prise en compte des corrélations spatiales à l'intérieur des strates permet d'améliorer l'inférence d'un effet proportionnel entre la moyenne et la variance de strate.

L'estimateur de la moyenne est : $Z_V^* = \sum_i w_i Z_{v_i}^*$, où $Z_{v_i}^*$ désigne la moyenne arithmétique des n_i échantillons dans la strate i et $w_i = v_i/V$ le poids dans le champ V de la strate i de surface v_i . Par construction, l'erreur d'estimation dans une strate est indépendante de celle effectuée dans une autre, car les échantillons sont implantés indépendamment d'une strate à l'autre. La variance d'estimation globale s'obtient en composant les variances d'estimation de chaque strate :

$$\sigma^2 = \sum_i w_i^2 \sigma_i^2 \quad \text{avec} \quad w_i = \frac{v_i}{V}$$

où σ_i^2 est la variance d'estimation de la moyenne de la strate v_i par n_i échantillons implantés dans v_i indépendamment les uns des autres, aléatoirement selon une loi uniforme. La variance d'estimation s'écrit donc:

$$\sigma^2 = \sum_i w_i^2 \frac{D_i^2}{n_i}$$

où D_i^2 est la variance de dispersion d'un échantillon ponctuel dans la strate i . Comme l'implantation des n_i échantillons de la strate i est aléatoire dans v_i , d'après la théorie classique (COCHRAN 1977), D_i^2 peut être estimée par $S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (z_j - z_{v_i}^*)^2$. Du

point de vue géostatistique, D_i^2 est la moyenne du variogramme pour toutes les distances du volume v_i de la strate i . Selon la coïncidence ou non de la stratification avec la régionalisation, D_i^2 se calcule dans chaque cas avec un variogramme qui lui est spécifique.

On considère d'abord que la stratification correspond effectivement à la régionalisation. Dans un premier temps, on se place dans le cas où chaque strate présente un variogramme spécifique $\gamma_i(h)$. Par exemple, les strates sont grandes et peu nombreuses et correspondent à deux ou trois grandes zones dans le champ V de présence du poisson. La variance de la strate i peut être définie comme la variance d'estimation de v_i par un échantillon ponctuel d'implantation aléatoire x_k . Pour calculer D_i^2 , on suit le raisonnement de MATHERON (1970, p 70). A tirage fixé, conditionnellement à l'implantation fixe de x_k dans v_i , la variance d'estimation est donnée par la formule (I.3) qui s'écrit pour un échantillon x_k fixe et ponctuel : $\sigma_i^2(\text{ext.}) = 2\bar{\gamma}_i(x_k, v_i) - \bar{\gamma}_i(v_i, v_i)$. Mais x_k est aléatoire. Pour calculer D_i^2 , la variance de dispersion dans v_i , il faut déconditionner

$\sigma_i^2(\text{ext.})$ par rapport à l'indice k : on prend l'espérance de $\sigma_i^2(\text{ext.})$ sur toutes les implantations possibles de x_k dans v_i .

Le calcul donne :

$$D_i^2 = \bar{\gamma}_i(v_i, v_i) = \frac{1}{v_i^2} \int_{v_i} \int_{v_i} \gamma_i(x-y) dx dy \quad (4)$$

D_i^2 est la valeur moyenne du variogramme $\gamma_i(h)$ pour toutes les distances dans v_i . La localisation des échantillons n'intervient plus. La variance ne dépend que de la structure spatiale moyenne dans la géométrie de la strate. La variance d'estimation s'écrit donc :

$$\sigma^2 = \sum_i w_i^2 \frac{\bar{\gamma}_i(v_i, v_i)}{n_i} \quad (I.4a)$$

Quand la stratification découpe le champ V en un nombre important de petites strates reconnues chacune par un petit nombre d'échantillons, il ne sera en général pas réaliste de prétendre inférer un variogramme dans chaque strate. Considérant toujours que la stratification coïncide avec une composante structurale de la régionalisation, en l'absence de lien marqué entre moyenne et structure spatiale de strate, on peut se contenter d'inférer un variogramme moyen intra-strate, $\gamma_{in}(h)$. Chaque strate possède un même variogramme; il ne fait pas intervenir la variabilité spatiale inter-strates. La variance d'estimation s'écrit alors :

$$\sigma^2 = \sum_i w_i^2 \frac{\bar{\gamma}_{in}(v_i, v_i)}{n_i} \quad (I.4b)$$

Si un effet proportionnel entre la moyenne et la variance de strate doit être pris en compte, on peut essayer de regrouper des strates pour inférer dans chaque groupe, un variogramme moyen intra-strate.

Les 2 cas précédents sont assez particuliers. En général en effet, les limites de strates ne correspondront pas à des cassures naturelles dans la régionalisation. On considère maintenant que pour chaque distance h , le variogramme calculé sur les couples de points intérieurs aux strates est du même ordre de grandeur que le variogramme calculé sur les couples de points à cheval sur 2 strates différentes. La structure spatiale recoupe les limites de strates sans être influencée. La stratification est arbitraire par rapport à la régionalisation. On infère alors un variogramme moyen global sur tout le champ, $\gamma_m(h)$.

4. $D_i^2 = \sigma_i^2(\text{disp.})$ s'appelle variance de dispersion d'un point (aléatoire) dans v_i .
 $\sigma_i^2(\text{ext.})$ s'appelle variance d'extension d'un point fixe dans v_i .

En l'absence d'effet proportionnel marqué entre moyenne et variance de strate, la variance d'estimation s'écrit maintenant:

$$\sigma^2 = \sum_i w_i^2 \frac{\bar{\gamma}_m(v_i, v_i)}{n_i} \quad (I.4c)$$

Le variogramme moyen global $\gamma_m(h)$ fait intervenir les variabilités moyennes intra-strate comme inter-strates. Quand la stratification est arbitraire par rapport à la régionalisation et qu'il y a un effet proportionnel entre la moyenne et la variance de strate, comme c'est le cas en général, il est particulièrement intéressant de tenir compte de l'effet proportionnel car seule compte la variabilité intra-strate, dans le calcul de la variance d'estimation. Dans chaque strate on cherche à utiliser le variogramme moyen dont le palier serait calé à la variance de la strate. Sous une hypothèse plausible on se propose maintenant de tenir compte d'un effet proportionnel, dans le cas où la stratification recoupe la régionalisation un peu arbitrairement.

Le modèle structural que l'on cherche à utiliser est localement stationnaire (stationnaire pour les distances mises en jeu à l'intérieur des strates) et présente, de strate à strate, un effet proportionnel marqué (lien entre la moyenne par strate et la variance par strate). On fait donc l'hypothèse que le variogramme $\gamma_i(h)$ de la strate i est proportionnel au variogramme moyen global:

$$\gamma_i(h) = \frac{\bar{\gamma}_i(V, V)}{\bar{\gamma}_m(V, V)} \gamma_m(h)$$

où $\bar{\gamma}_i(V, V)$ et $\bar{\gamma}_m(V, V)$ sont les variances de dispersion sur tout le champ $V = \sum_i v_i$ calculées avec $\gamma_i(h)$ et $\gamma_m(h)$.

On en déduit alors le rapport de proportionnalité k_i entre variances de dispersion calculées avec γ_i et γ_m :

$$k_i = \frac{\bar{\gamma}_i(v_i, v_i)}{\bar{\gamma}_m(v_i, v_i)} = \frac{\bar{\gamma}_i(V, V)}{\bar{\gamma}_m(V, V)}$$

On cherche la variance de dispersion D_i^2 dont une estimation expérimentale est S_i^2 et dont la valeur dans le modèle géostatistique inféré est $\bar{\gamma}_i(v_i, v_i)$. Or à partir des données, on peut calculer $\bar{\gamma}_m(v_i, v_i)$ et S_i^2 mais non $\bar{\gamma}_i(v_i, v_i)$. En général S_i^2 sera une estimation imprécise de D_i^2 car on ne disposera que d'un petit nombre d'échantillons dans les strates. Comme la moyenne Z_{v_i} de strate est mieux estimée que la variance de strate D_i^2 à partir des données, il est intéressant de régresser S_i^2 sur $Z_{v_i}^*$ pour améliorer l'estimation de D_i^2 . Or, les strates sont de tailles et de géométries variables et la variance de dispersion est une grandeur qui dépend de la surface sur laquelle elle est calculée. Ainsi, deux strates de même moyenne ont des variances de dispersion différentes si elles sont de géométrie et de taille différentes. La géostatistique nous permet ici de tenir compte de l'effet géométrique

des strates sur la valeur de S_i^2 et ainsi d'inférer un effet proportionnel entre variance et moyenne par strate qui soit dégagé de l'effet géométrique des strates. On diminuera donc toujours la variance résiduelle de la régression en régressant, non pas S_i^2 sur Z_{vi}^* mais en régressant le rapport $k_i^* = \frac{S_i^2}{\bar{\gamma}_m(v_i, v_i)}$ sur Z_{vi}^* . En effet, deux strates de même moyenne mais de tailles très différentes auront des rapports k_i^* qui seront moins différents que ne le seront leurs variances S_i^2 . L'effet proportionnel entre la variance et la moyenne de strate, dégagé de l'effet géométrique des strates, est donc inféré avec une meilleure précision.

Finalemnt, on propose de calculer la variance d'estimation du plan stratifié

$\sigma^2 = \sum_i w_i^2 \frac{D_i^2}{n_i}$ par la formule suivante :

$$\sigma^2 = \sum_i w_i^2 k_i^* \frac{\bar{\gamma}_m(v_i, v_i)}{n_i} \quad (I.4d)$$

où k_i^* est estimé par sa régression sur les moyennes de strate et où le variogramme de chaque strate est supposé proportionnel au variogramme moyen global.

La formule de la variance d'estimation $\sigma^2 = \sum_i w_i^2 \frac{D_i^2}{n_i}$ est bien sûr inchangée par

la géostatistique. COCHRAN (1977) propose d'en faire l'estimation par $\sum_i w_i^2 \frac{S_i^2}{n_i}$. Sauf cas particulier de coïncidence entre stratification et régionalisation, nous proposons d'en faire l'estimation par la formule (I.4d) car la précision de l'estimation de la variance de strate peut être améliorée grâce à la Géostatistique lorsque le nombre d'échantillons par strate est faible et que les strates diffèrent en taille.

On calcule la variance d'estimation selon la formule (I.4d) dans le cas du merlu dans le Golfe de Gascogne, à la section II.C.1.

A.3. Le variogramme près de l'origine

Dans la théorie intrinsèque, la réalité observée par un échantillonnage est interprétée comme une réalisation d'un modèle, une fonction aléatoire Z . Il a été introduit une hypothèse de stationnarité sur la régionalisation, indispensable à l'inférence du modèle de variogramme à partir d'une réalisation unique de Z . La stationnarité a le statut d'une contrainte méthodologique en intrinsèque comme en transitif. En effet, la stationnarité s'introduit soit comme une caractéristique du réseau d'échantillonnage sans hypothèse sur la réalité biologique (théorie transitive), soit comme une caractéristique de la réalité biologique sans contrainte de stationnarité sur l'implantation des échantillons dans l'espace (théorie intrinsèque)⁵. La variance d'estimation ne s'exprime qu'en fonction d'une covariance moyenne ou du variogramme et dépend pour l'essentiel du comportement analytique de cette fonction pour les petites distances. Pour le calcul de la variance de l'estimation globale, l'hypothèse de stationnarité et l'inférence du modèle structural peuvent se limiter aux petites distances (quasi-stationnarité).

A.3.1. Quasi-stationnarité

MATHERON (1970, pp 97-100) montre que la quasi-stationnarité est l'hypothèse minimale requise pour le calcul de la variance avec les formules de la théorie intrinsèque (Géostatistique linéaire et stationnaire).

Soit $Z(x)$ une fonction aléatoire définie comme suit : $Z(x) = m(x) + \omega(x) Y(x)$. $m(x) = E[Z(x)]$ est une variable régulière lentement variable dans l'espace à l'échelle de l'échantillonnage : $m(x)$ peut être considérée constante dans la maille élémentaire du réseau régulier ou dans chacune des strates du plan stratifié. $Y(x)$ est une fonction aléatoire stationnaire sur tout le champ, de moyenne nulle ($E[Y(x)] = 0$). $\omega(x)$ est un effet proportionnel lentement variable dans l'espace, considéré comme constant dans des voisinages de dimensions égales à la maille. $Z(x)$ est quasi-stationnaire, ou localement stationnaire, c'est-à-dire de moyenne et de covariance stationnaires dans des voisinages aux dimensions de l'ordre de la maille. $Z(x)$ a pour covariance centrée : $C(x, y) = \omega(x)\omega(y)C_0(h)$ où $h = |x-y|$ et $C_0(h)$ est la covariance stationnaire de $Y(x)$. Pour les petites distances h inférieures à la maille, la covariance (pseudo-stationnaire)

moyenne de Z , est peu différente de : $\bar{C}(h) \approx C_0(h) \frac{1}{V} \int_V \omega^2(x) dx$.

5. Du point de vue pratique, pour la programmation d'une campagne à la mer comme pour le traitement ultérieur des données, suivre rigoureusement un plan d'échantillonnage n'est pas absolument nécessaire pourvu que l'échantillonnage ne soit pas biaisé. Les données issues de prospections non conformes à des plans d'échantillonnage orthodoxes pourront quand même être traitées par la théorie intrinsèque. On pense, en particulier, aux prospections acoustiques où bien souvent, pour des raisons de temps et de connaissance biologique, le bateau a fait route, couvrant bien la zone mais ne respectant pas de plan d'échantillonnage standard.

Considérons l'estimation de $Z_v = \frac{1}{V} \int_V Z(x) dx$ par la moyenne arithmétique

des données réparties sur une grille régulière. MATHERON (1970) montre, en utilisant les principes d'approximation, que la variance de cette estimation s'écrit :

$$\sigma_{\text{est}}^2 \approx \sigma_{\text{est}}^2 \frac{1}{V} \int_V \omega^2(x) dx \text{ où } \sigma_{\text{est}}^2 \text{ est la variance d'estimation de la moyenne de } Y(x).$$

La variance d'estimation σ_{est}^2 d'une fonction quasi-stationnaire, est la même que celle que l'on calculerait pour une fonction aléatoire stationnaire dont la covariance serait justement $\bar{C}(h)$.

En général, la stationnarité de la régionalisation sur tout le champ n'est pas physiquement acceptable. En revanche, la stationnarité locale est dans un grand nombre de cas "suffisamment faible pour être physiquement admissible" (MATHERON, 1970 p. 41)⁶. Physiquement, la stationnarité locale veut dire que sur de petits domaines spatiaux, n'importe où dans le champ, c'est bien le même processus de structuration du poisson qui est observé. Le variogramme aux petites distances en est une mesure moyenne. De par les principes d'approximation, la stationnarité locale suffit pour le calcul de la variance d'estimation de la moyenne spatiale globale.

A.3.2. Les variogrammes

On distingue trois variogrammes: régional, expérimental et théorique. Supposons que la régionalisation réelle soit connue en tous points. Le variogramme régional $\gamma_R(h)$ est une grandeur déterministe et s'écrit:

$$\gamma_R(h) = \frac{1}{2K(h)} \int_{V_{-h}} (f(x) - f(x+h))^2 dx \text{ où } V_{-h} \text{ est l'intersection du champ } V \text{ avec son}$$

translaté de vecteur $-h$ et où $K(h)$ est le covariogramme géométrique ($K(h)$ mesure la surface V_{-h}). Le variogramme expérimental, $\gamma_w(h)$, que l'on calcule sur les données relatives à l'échantillonnage w , est une estimation du variogramme régional $\gamma_R(h)$. D'autre part, en interprétant la régionalisation ($f(x_1), \dots, f(x_n), \dots$) comme une réalisation ($z(x_1), \dots, z(x_n), \dots$) d'une fonction aléatoire Z , le variogramme régional $\gamma_R(h)$ devient dans le modèle une variable aléatoire $\Gamma_R(h)$. Le variogramme $\gamma(h)$ de Z , i.e. le modèle que l'on prétend inférer, correspond dans le modèle de fonction aléatoire, à l'espérance de $\Gamma_R(h)$ sur toutes les réalisations de Z : $\gamma(h) = E[\Gamma_R(h)]$. Ainsi, l'interprétation probabiliste s'accompagne dans le modèle, de fluctuations du variogramme régional autour de sa version théorisée $\gamma(h)$. Dans la pratique γ_w estime γ_R puis on ajuste un modèle γ qui, dans le modèle probabiliste Z , a la signification de $E[\Gamma_R]$. Il y a deux variabilités de nature différente, d'une part les fluctuations d'échantillonnage du variogramme expérimental autour du régional et d'autre part, les fluctuations dans le modèle du variogramme régional autour de sa version probabilisée théorique. Dans le cas

6. Pour la cartographie, on utilise souvent un modèle localement stationnaire et l'interpolation en x_0 est effectuée avec les échantillons avoisinant x_0 (krigeage en voisinage glissant).

d'un effort d'échantillonnage faible sur une régionalisation très hétérogène, si la variance des fluctuations d'échantillonnage $E(\gamma_w - \gamma_R)^2$ est grande, on risque de modéliser un variogramme qui soit spécifique de l'échantillonnage w et qui ne corresponde pas au variogramme régional γ_R . Ce problème est celui du biologiste qui se doit de définir un effort d'échantillonnage suffisant. Ce problème fait l'objet de la partie C. On suppose maintenant que γ_w estime précisément γ_R . On s'intéresse aux fluctuations dans le modèle.

MATHERON (1965, 1970) développe dans le cadre gaussien, la variance $E(\Gamma_R - \gamma)^2$ des fluctuations dans le modèle de $\Gamma_R(h)$ autour de sa version théorisée $\gamma(h)$. Cette variance est une fonction croissante de la distance h . Quand la distance augmente, les fluctuations dans le modèle peuvent théoriquement être importantes. Le comportement du modèle de variogramme aux grandes distances ne sera donc pas biologiquement significatif, le modèle aux grandes distances n'aura qu'un rôle conventionnel. La variance relative $\frac{\text{Var}(\Gamma_R(h) - \gamma(h))}{\gamma(h)^2}$ tend vers zéro quand h tend vers

zéro. L'inférence d'un modèle est donc théoriquement possible pour les petites distances (micro-ergodicité). Dans la pratique, le modèle près de l'origine $h=0$ aura un sens biologique et la modélisation pourra être argumentée.

Dans la théorie transitive, le modèle de covariogramme paramètre le covariogramme régional réel. Or celui-ci ne sera jamais connu et est très compliqué. Le modèle qui lui est substitué, lui, est plus simple et selon MATHERON (1965, 70, 89) peut être interprété comme une espérance. L'aspect probabiliste du modèle transitif est implicite, caché et on ne travaille que sur des grandeurs régionales. En intrinsèque, le sens probabiliste du modèle est explicite. Toutefois, c'est toujours une grandeur régionale, une intégrale d'espace γ_R dont le modèle de variogramme rend compte. Ainsi les deux représentations différentes, transitive d'une part et fonction aléatoire intrinsèque d'autre part, conduisent à partir d'un même jeu de données, à des variances d'estimation globale similaires.

A.3.3. Le choix du modèle

Dans le calcul de la variance d'estimation, on doit calculer les intégrales du modèle de variogramme. Par exemple, dans l'intégrale

$$\bar{\gamma}(v, v) = \frac{1}{V^2} \int_V \int_V \gamma(x - y) \, dx \, dy, \text{ toutes les distances interviennent et en particulier}$$

les très petites distances. Or, expérimentalement, entre l'origine $h=0$ et le premier point expérimental du variogramme expérimental, aucune information n'existe dans les données pour nous guider dans la modélisation. Et la variance d'estimation dépend largement du comportement analytique du modèle pour ces distances. Il faut faire un choix: on injecte alors de l'information non contenue dans le jeu de données. Ce choix peut être argumenté biologiquement car il correspond à une interprétation de la

structuration du poisson. Le choix d'un modèle aux petites distances peut être guidé expérimentalement par un échantillonnage supplémentaire, à mailles très fines dans différentes directions, à différents endroits dans le champ. Dans les campagnes acoustiques, des croix d'échantillonnage sont facilement envisageables et ont déjà été recommandées (GOHIN, 1985). Dans les campagnes ichtyoplanctoniques à coups de filet à plancton, il est pratique courante de répéter la mesure sur un certain nombre de positions ; ces réplicats estiment la valeur du variogramme très près de l'origine. Bien souvent, un effet de pépité est une réalité expérimentale, non seulement à cause de micro-structures (structures spatiales à une échelle inférieure au pas d'échantillonnage), mais aussi à cause du dynamisme dans le temps de la structuration à très petite échelle.

B. APPLICATION A UNE PROSPECTION ACOUSTIQUE A MAILLE RÉGULIÈRE

Nous mettons en œuvre la méthodologie précédente sur une prospection acoustique de hareng, devant la côte norvégienne. A cela il y a des raisons statistiques et biologiques.

Les raisons biologiques concernent la fenêtre spatio-temporelle d'observation, qui ici est bien définie, ainsi que la nécessité du suivi de la quantité totale par l'échantillonnage. La population de hareng norvégien dont il est question, fait partie d'une entité plus vaste, le groupe "Atlanto-scandinave" que l'on trouve dans l'Atlantique nord depuis les Iles Faeroe jusqu'au Spitzberg. L'argumentation biologique est issue de DRAGESUND et al. (1980). Dans l'ensemble Atlanto-scandinave, des sous-groupes ont été identifiés : faroéen, islandais et norvégien. Le cycle migratoire dans l'espace de la vie du poisson au cours d'une année est schématisé sur la Figure 2. Pendant l'été, toutes les populations du groupe Atlanto-scandinave se mélangent et tous les poissons sont dispersés sur une aire d'alimentation dont l'extension au milieu de l'Atlantique nord est variable et peut être considérable. Après une première concentration sur des zones d'hivernage, les différents sous-groupes s'individualisent et opèrent, avant la fin de l'hiver, des migrations de ponte. Chacun migre vers ses propres lieux de ponte. Des marquages ont montré que les individus nés dans un groupe ne participent, pour la plupart, qu'à la reproduction de ce groupe. Chaque sous-groupe représente donc une unité de reproduction, donc de gestion, un stock. Devant la côte norvégienne, le hareng se concentre en quelques zones de frai connues, où il pond au début du printemps océanique en février-mars (Spring spawning herring). La fenêtre spatio-temporelle est donc définie : c'est au printemps, sur les zones de ponte, qu'il est possible de quantifier la biomasse féconde en place (standing stock) du hareng norvégien. A un autre moment de l'année, cette estimation est impossible.

Le suivi de la biomasse féconde par des campagnes à la mer est ici nécessaire. Le hareng est un poisson pélagique et grégaire. Les débarquements de la pêche ne reflètent pas forcément la quantité en place dans la mer car la pêche est fonction de la densité. Un stock pélagique, dont la biomasse totale diminue, peut être amené à se concentrer sans modifier beaucoup sa densité mais réduisant sa zone de présence. Ce fut effectivement le cas du hareng Atlanto-scandinave (DRAGESUND et al, 1980) lors de sa quasi-extinction au cours des années 1970. Les campagnes à la mer sont dans cette situation indispensables pour prévenir l'extinction. En 1977, le Conseil International pour l'Exploration de la Mer (C.I.E.M./I.C.E.S.) recommanda l'arrêt de la pêche à cause d'une surexploitation de tout le groupe Atlanto-scandinave à tous les stades de son cycle biologique. La pêche a été ré-ouverte en 1985, mais la ressource reste sous surveillance. Chaque année, dans chaque zone, un échantillonnage permet de mesurer l'état de santé des stocks (biomasse totale, fécondité, démographie, extension spatiale). Dans leur suivi routinier, les norvégiens utilisent la prospection acoustique selon un échantillonnage à maille régulière (transects parallèles et équidistants). L'échantillonnage par ondes acoustiques permet une

couverture très rapide par rapport à la durée du frai pendant laquelle le poisson reste sur zone.

Pourtant, en acoustique, l'utilisation de mailles régulières est très critiquée par les statisticiens classiques des pêches. JOLLY et HAMPTON (1987) proposent de "randomiser" la position des transects parallèles dans des strates définies au préalable. En effet, la méthodologie du calcul de la variance d'estimation pour les mailles régulières n'est pas bien établie dans la science des pêches⁷. Dans ce contexte, nous avons voulu d'abord montrer l'utilité de la Géostatistique telle qu'elle existe.

De plus, les données d'acoustique pélagique fournissent un cadre statistique extrême. On travaillera un peu en limite de validité des méthodes, ce qui constitue un test supplémentaire de leur validité opératoire. Le poisson grégaire se concentre dans l'espace en un petit nombre de paquets très denses. La dispersion des valeurs est très forte, l'histogramme très dissymétrique. L'estimation dépend de la rencontre ou non à l'échantillonnage d'un petit nombre d'agrégats très denses. On utilisera le modèle spatial géostatistique pour mieux définir l'effort d'échantillonnage à fournir afin d'assurer une couverture efficace de la zone. Enfin les données acoustiques sont intéressantes à cause du pas très fin des mesures de densité de poissons le long de la route du navire qui nous permet d'estimer le comportement du variogramme près de l'origine.

B.1. L'échantillonnage

L'échantillonnage concerne l'une des principales zones de frai du hareng, *Clupea harengus*, devant la côte norvégienne en février 1989 (Figure 3). Il s'agit d'une prospection acoustique d'écho-intégration le long de transects parallèles et équidistants.

B.1.1. La méthode de l'écho-intégration et les données

L'écho-intégration est une méthode acoustique utilisée depuis la fin des années 60 pour mesurer quantitativement la biomasse des poissons (ou des planctontes) situés dans la colonne d'eau sous le navire, à partir de la détection de ceux-ci à l'aide d'ondes sonores. La mesure quantitative a lieu en continu le long de la route du navire. Une description détaillée de la méthode peut être trouvée dans BURCZYNSKI (1982) et/ou JOHANNESSON et MITSON (1984). On expose brièvement les principes de la mesure et le prétraitement du signal dont nos données sont issues. Pour ce faire, on se réfère à BURCZYNSKI (1982). On replace ensuite la mesure physique dans son cadre biologique.

Un transducteur, placé sur la coque du navire sous la ligne de flottaison, transforme les impulsions électriques d'un émetteur en ondes mécaniques (ultrasons). Elles insonifient verticalement sous le navire et jusqu'au fond, un volume que l'on peut

7. Un atelier de travail du C.I.E.M., "Workshop on the Applicability of Spatial Statistical Techniques to Acoustic Survey Data", auquel nous participons travaille sur cette question (ANON, 1990, 91).

considérer comme conique. Après chaque émission sonore, le transducteur reste en attente (écoute) pour recevoir les échos réfléchis par les cibles se trouvant dans le volume insonifié par cette émission. Les échos reçus sont alors transformés par le transducteur en tension électrique. Celle-ci est mesurée à chaque instant par le sondeur et peut être visualisée (échogramme).

Pendant le temps d'écoute, on considère que les échos reçus sont uniquement ceux des cibles situées dans le cône insonifié par l'émission immédiatement précédente. La réponse acoustique est immédiate. En effet, on peut considérer que le bateau reste fixe pendant un cycle émission-réception du transducteur : faisant route en général à dix nœuds (18,52 km/h), sa vitesse est 300 fois inférieure à la vitesse moyenne très grande des ultrasons dans l'eau de mer ($c = 1500\text{m/s}$). Le taux de sondage (fréquence des émissions) est calibré en fonction de la profondeur d'eau sous le navire car le temps d'écoute doit permettre de recevoir les réponses de toutes les cibles dans le cône insonifié avant d'émettre une nouvelle émission. Les détails des lois physiques utilisées dans la mesure acoustique sont donnés sur les Figures 4 et 5.

Après chaque émission, les échos des cibles situées à différentes profondeurs arrivent au transducteur en écoute à différents instants, selon la distance que le son a parcouru. L'intensité de l'écho décroît avec la distance parcourue donc avec l'intervalle de temps qui sépare son arrivée au transducteur du départ de l'émission. Chaque écho reçu par le transducteur est multiplié par une fonction de gain variable dans le temps (GVT) qui compense les pertes d'intensité sur le parcours direct du son, depuis la descente de l'impulsion jusqu'à la remontée de l'écho directement réfléchi. Soit τ la durée de l'impulsion (durée pendant laquelle le transducteur vibre pour créer l'onde sonore). Le front d'onde qui descend a une épaisseur de $c\tau/2$: c'est l'épaisseur verticale élémentaire de l'échantillonnage. On montre que, à tout instant, la tension électrique mesurée après correction par une fonction GVT est proportionnelle à la racine carrée de la densité de poisson présent dans la couche d'épaisseur $c\tau/2$ située à la distance r du transducteur : $u^2(t) \sim \rho(r)\sigma(r)$ où $\rho(r)$ est la densité du poisson, en nombre, et $\sigma(r)$ est l'index de réflexion sonore des cibles ("target strength" ou TS). De façon à ne garder que les réponses directes dues au poisson, on ne retient brutalement que les tensions $u^2(t)$ supérieures à un seuil fixé par l'utilisateur. Les tensions en dessous du seuil sont considérées comme du bruit, dû au navire, aux vagues, aux cibles de réflectivité faible (plancton), aux échos plusieurs fois réfléchis... Pour l'estimation, les valeurs $u^2(t)$ supérieures au seuil sont intégrées deux fois. D'abord sur la verticale dans chaque cône pendant la durée ΔT d'écoute du transducteur (intégration par émission) : $u_{\Delta T}^2 = \int_0^T u^2(t)dt = \int_0^R \rho(r)\sigma(r)dr$, où R est la hauteur de la couche intégrée. Chaque émission insonifiant un cône, le volume échantillonné par l'onde d'impulsion croît comme le cube de la profondeur. La mesure de densité $\rho(r)$ est donc plus précise en profondeur qu'en surface, mais les différences de volumes échantillonnés ne biaisent pas la mesure de l'intégrale $u_{\Delta T}^2$. Les $u_{\Delta T}^2$ sont situées

dans l'espace aux sommets x_i de chaque cône d'émission. Les valeurs sont moyennées par segments L (Unité Élémentaire de Distance d'Echantillonnage) le long de la route du navire (intégration dite par mille) : $M_{uede} = \frac{1}{m} \sum_{i=1}^m u_{\Delta T}^2(x_i)$, où m est le nombre de cycles émission-réception le long de chaque segment L.

Le nombre m d'échantillons sur L est très grand. Au milieu du plateau continental sur des fonds de 100m, le son met 137ms pour faire l'aller retour depuis le bateau jusqu'au fond. En prenant 200ms pour la durée d'un cycle émission-réception (valeur classique), le bateau à la vitesse de 10 noeuds parcourt 1m dans ce laps de temps. Sur des segments de 1 mille nautique (1852m) M_{uede} est la moyenne de plus de 1850 données dont les supports d'échantillonnage coniques verticaux se recouvrent. On considère donc que les valeurs M_{uede} sont les teneurs moyennes réelles sur les segments u.e.d.e. le long de la route du navire. Comme ces segments sont contigus (ils se recoupent en fait par les cônes des extrémités), on considère que la prospection acoustique est exhaustive le long de la route du navire.

L'index de réflexion dépend en particulier de la taille du poisson. Une bibliographie peut être trouvée chez BURCZYNSKI (1982). Le poids des poissons est lié à leur taille par une fonction puissance, donc M_{uede} est proportionnel à une densité pondérale (biomasse) de poisson. On écrit : $M_{uede} = cd$, où d est la densité pondérale moyenne dans les couches insonifiées (tonnes/mille nautique) et où c est une constante d'écho-intégration. Celle-ci est liée au matériel utilisé qui devra être étalonné et à l'index de réflexion sonore σ (Target strength, TS). La constante d'intégration c peut être mesurée expérimentalement in situ en mettant en cage des quantités connues de poisson vivant et en sondant dessus.

Physiquement, l'index de réflexion sonore (TS) est fonction de la différence de densité entre les cibles et l'eau de mer. La TS des poissons varie donc en fonction de nombreux paramètres biologiques que l'on peut hiérarchiser : d'abord en fonction de l'espèce puis, pour chaque espèce en fonction de la taille, pour chaque taille en fonction de la physiologie dans le temps (rythmes circadiens, lunaires, saisonniers et sexuels) et/ou du comportement instantané (orientation du poisson dans l'eau de mer). L'intégration des tensions électriques sur la verticale puis sur l'horizontale correspond à une régularisation nécessaire : son effet est de ne garder qu'une mesure moyenne plus indépendante de toutes les variations biologiques instantanées du TS. La validité de la conversion des tensions électriques M_{uede} en biomasse absolue de poisson dépend de la possibilité de connaître c avec précision ou d'en connaître ses variations. Pour les principales espèces commerciales, le coefficient c est connu et sa relation avec la taille aussi. Toutefois appliquer un coefficient c peut être soumis à de nombreux biais. L'approche classique est de considérer que ces biais sont constants et que l'estimation fournit un index d'abondance dont les valeurs sont comparables d'année en année.

Bien souvent, en milieu tempéré et tropical, il n'est pas possible de définir une boîte spatio-temporelle dans laquelle un stock d'une espèce vienne s'isoler pendant une partie de son cycle annuel de vie. En général, il faut prospector les grandes étendues marines où le stock est dispersé, traquer quelques espèces d'intérêt parmi plusieurs, sur tout ou partie du plateau continental. On échantillonne alors acoustiquement différentes espèces en proportions variables dans l'espace. Comme on ne mesure que globalement M_{uede} , il se pose un problème difficile d'identification des échos des espèces dont on veut spécifiquement estimer la biomasse. Un certain nombre de pêches des cibles peut permettre de révéler les proportions entre espèces et les tailles de façon régionalisée ou moyenne sur toute la zone (MASSÉ, 1988). On applique ce coefficient aux réponses globales M_{uede} . Toutefois si les espèces d'intérêt présentent sur l'échogramme des signatures acoustiques reconnaissables avec un taux de réussite suffisant, la décomposition du signal électrique est possible et les intégrations de différents échos sont allouées à différentes espèces (DINER et al., 1989). A l'identification délicate par la pêche ou sur l'échogramme, s'ajoutent des possibilités de biais dont deux sources ont été établies : réaction de fuite du poisson à l'approche du navire modifiant la valeur de la TS comme le nombre des cibles insonifiées (OLSEN et al., 1983 ; GERLOTTO, 1989), dérive au cours de la journée de la valeur de la TS (MASSÉ, 1988).

Nous avons mentionné le cadre biologique difficile de l'utilisation de la méthode acoustique en climat tempéré et chaud, pour mieux mettre en valeur maintenant une situation biologique commune sous les hautes latitudes, très favorable à l'estimation quantitative de biomasse par écho-intégration dont le hareng norvégien nous fournit un exemple. Le hareng s'isole dans une petite boîte de temps et d'espace où il est possible d'échantillonner spécifiquement toute la population. L'échantillonnage acoustique est monospécifique. Les frayères ont des extensions spatiales relativement limitées et le hareng reste sur ses zones de frai pendant une durée de l'ordre de un à deux mois. Les prospections acoustiques rapides des frayères peuvent être considérées instantanées. La prospection dont nous nous occupons a duré 4 jours. A cette échelle de temps, on considère que la distribution spatiale générale sur la frayère ne varie pas au cours de l'échantillonnage. La frayère est située entre 62° et 64° de latitude nord. Début mars lors de l'échantillonnage, les caractéristiques physiques de l'habitat sont telles (eau froide de température constante, alternance jour-nuit peu marquée) que le poisson est peu vif sur zone et réagit peu aux stimuli provoqués par le navire (ONA et al, 1989). Deux difficultés majeures peuvent se rencontrer toutefois. Les échogrammes montrent le hareng formant une couche plus ou moins décollée du fond. Lorsqu'elle est très étalée sur le fond, les échos de ce poisson ne sont pas dissociables de ceux du fond et la biomasse est inaccessible à l'échantillonnage acoustique. Lorsque la couche est très dense, des ombrages acoustiques peuvent rendre invisible une partie de la biomasse, mais la baisse d'intensité des échos du fond permet de faire des corrections (TORESEN, 1990). Ces difficultés n'ont pas gêné l'échantillonnage dont nous traitons (I. Rottingen, comm.pers.). Les causes de biais sont donc très réduites ici. La pêche des cibles a révélé que le hareng était composé

en moyenne sur la zone de 91 % de la classe d'âge 6 ans et que l'histogramme total des tailles était resserré autour d'un seul mode de 32cm. Les caractéristiques acoustiques des cibles (une seule espèce, un seul mode de taille) sont très homogènes et la conversion des tensions électriques intégrées M_{uede} en biomasse de hareng se fait dans de bonnes conditions.

Nos données sont les valeurs des tensions électriques intégrées M_{uede} . La régularisation horizontale est effectuée sur des segments de un mille nautique (m.n). Les données sont exprimées en "backscattering units/nautical mile square", elles n'ont pas été converties en biomasse. Elles ont été fournies par Drs K. FOOTE et I. ROTTINGEN du Marine Research Laboratory de Bergen (Norvège) et correspondent à un des jeux tests proposés aux ateliers de travail du C.I.E.M. pour l'utilisation de la Géostatistique en écho-intégration (ANON, 1990a, 91a) auxquels nous avons participé.

B.1.2. Stratégie d'échantillonnage suivie

On donne sur la Figure 3 le plan d'échantillonnage avec une représentation proportionnelle des données. Le cercle de rayon maximal représente la donnée maximale, les autres s'en déduisent proportionnellement. Les cercles pleins et noirs marquent les zéros.

La prospection a lieu le long de transects orientés Nord-Sud, régulièrement espacés, de la côte vers le large. L'approche de l'échantillonneur est celle formulée dans la théorie transitive. Les limites exactes de présence du poisson ne sont pas connues avant l'échantillonnage, mais on sait que le stock est concentré sur la zone. On recherche donc les limites des zéros extérieurs. On les nomme extérieurs par opposition aux zéros dits intérieurs représentants (malgré les régularisations sur la verticale et sur les segments de 1 mille) des trous de densité de poisson à l'intérieur du champ de présence. Quand de nombreux zéros sont rencontrés le long d'un transect, s'aidant des cartes des années précédentes, l'échantillonneur comprend que les limites du champ sont atteintes : le transect est arrêté et un autre entamé. Le schéma régulier n'est pas respecté dans la partie Nord-Est. Là les limites du champ ne sont pas clairement marquées. Une prospection rapide en zig-zag a été effectuée afin de voir si une partie importante du stock ne serait présente en bordure Nord-Est. La réponse est négative car aucune valeur forte n'est rencontrée par l'échantillonnage. Cette région ne représente que 3 % de la moyenne globale calculée sans les zéros. On s'intéressera à la partie régulière de la prospection. On considère que cette grille est implantée indépendamment de la régionalisation du hareng.

B.2. Description du jeu de données

Les données sont des régularisées sur la verticale et sur des segments horizontaux de 1m.n. Elles sont implantées le long de la route du navire aux centres des segments et sont distantes entre elles de 1m.n. En revanche la distance inter-transect est de 10 minutes de longitude.

Afin de calculer des distances, il faut que les unités en latitudes et en longitudes représentent les mêmes distances réelles. Or un degré de longitude n'égal un degré de latitude qu'à l'équateur. Assimilant simplement la terre à une sphère, on a multiplié les longitudes par le cosinus de la latitude moyenne : $\cos(63^\circ)=0.454$. Cette correction simple nous suffit car la zone de prospection est petite sur le globe : environ 4000 km² sur 2° de latitude. Toutes les distances seront exprimées en milles nautiques (m.n). Rappelons qu'un mile nautique est la distance sur un méridien correspondant à une minute d'angle au centre de la terre et équivaut à 1852 m. La distance inter-transect dans la partie régulière de l'échantillonnage est de 10 minutes de longitude, soit 4.54 m.n

B.2.1. Statistiques globales

Le jeu total comprend 986 données, une large proportion de zéros et une faible proportion de fortes valeurs : moyenne arithmétique = 774., variance = $12.7 \cdot 10^6$. Le maximum et le minimum sont : 68 000 et 0.

Sur les Figures 6 et 7, on représente les contributions relatives des différents quantiles de l'histogramme à la moyenne et à la variance. Les valeurs f_i sont rangées par ordre croissant, moyenne et variance sont calculées sur les sous-ensembles $f(x) < f_i$. Leur pourcentage dans respectivement la moyenne et la variance globale, donnent leur contribution relative à celles-ci. La fréquence dans les données des 13 plus fortes valeurs ($f(x) > 10^4$) est environ de 1 %. Cette queue d'histogramme représente 45 % de la moyenne et 90 % de la variance. On peut déjà dire que l'estimation de la biomasse moyenne et la valeur de sa variance d'estimation dépendront énormément des quelques valeurs très fortes. On posera ce problème en terme d'efficacité de l'effort d'échantillonnage à estimer avec une précision suffisante la fréquence des valeurs fortes qui représentent une partie importante de la biomasse totale. On fera d'abord l'estimation sans se préoccuper de l'histogramme, comme c'est généralement le cas en géostatistique linéaire, puis on resituera l'estimation dans ce contexte statistique difficile.

B.2.2. Définition du champ

Pour étudier la régionalisation intrinsèque au poisson, il faut définir le champ V de sa présence (Figure 8). C'est une étape toujours sujette à discussion et très importante car le calcul de la variance d'estimation en intrinsèque se fera sur la surface de ce champ. En transitif, ce problème ne se pose pas. Les zéros représentent 52 % des données. La plus grande part sont des zéros extérieurs qu'il a fallu échantillonner pour définir les limites de V. Les limites du large et du Sud-Ouest sont marquées par des zéros qu'il est facile de reconnaître comme extérieurs. Tous les zéros aux extrémités des transects sont interprétés comme des zéros extérieurs à la régionalisation, i.e. comme des valeurs déterministes. Les transects Sud-Ouest numérotés 1 et 2 sont couverts de zéros que l'on interprète comme extérieurs. Le long des transects, on observe des successions de zéros entourés de valeurs non nulles. S'agit-il de zéros intérieurs à la régionalisation ou extérieurs ? Ces zéros doivent-ils être interprétés comme des réalisations de la fonction aléatoire, au même titre que les réalisations non nulles et marquant des trous aléatoires de densité à l'intérieur de la régionalisation, ou doivent-ils être interprétés comme des zéros déterministes marquant l'absence de poisson en dehors du champ sur lequel la fonction aléatoire prend ses valeurs ? La distinction n'est pas facile et est sujette à une erreur difficile à quantifier par une variance. On a procédé qualitativement. On s'est aidé de la carte de l'année précédente en faisant l'hypothèse biologique plausible d'une certaine reproductibilité des limites de la frayère entre les 2 années. On a ainsi considéré les zéros le long des transects qui sont encadrés par des valeurs non nulles comme des zéros intérieurs sauf sur les transects 5, 7, 8 et 9. Sur ces transects, on a reconnu des indentations de zéros extérieurs, d'une part parce que ces successions sont longues donc peu compatibles avec l'idée de trou de densité dans la régionalisation, d'autre part parce que la carte de l'année précédente présente aussi à ces endroits des indentations de zéros.

La limite vers la côte n'a pas pu être clairement échantillonnée, probablement pour des raisons de sécurité de navigation. Mais le poisson n'est pas présent jusqu'à la côte. Comme preuve, nous observons que les extrémités de transect les plus proches de la côte (sur les transects numérotés 5,7,8,13) montrent des répétitions successives de zéros. Au Nord-Est dans la partie irrégulière de l'échantillonnage, on se trouve en bordure du champ. Elle se présente de façon moins tranchée, plus diffuse. C'est la partie régulière de la prospection qui échantillonne la frayère de hareng. On définit le champ V de présence du poisson comme donné sur la Figure 8.

Les limites de V ont un statut mixte emprunté des méthodes transitive et intrinsèque. Vers la côte et dans la partie Nord-Est, on a coupé la régionalisation en arrêtant les limites de V au bout des transects alors qu'il existe un peu de poisson au delà (3% de la biomasse environ). Vers le large et sur la bordure Sud-Ouest, on fait passer les limites de V là où les zéros extérieurs ont été reconnus. On calcule en B5 le terme de variance de l'estimation de la moyenne du à l'erreur faite sur les limites de V.

B.2.3. Statistiques dans le champ

Dans le champ, on a les statistiques élémentaires suivantes qui sont propres à la régionalisation : nombre de données = 435 ; surface $V = 1975 \text{ m.n}^2$; moyenne arithmétique $Z_V^* = 1690$; variance $s^2 = 27.3 \cdot 10^6$; pourcentages de zéros intérieurs = 8 %. L'histogramme dans V est donné sur la Figure 9 : il est très dissymétrique, d'un type lognormal mais avec un petit atome de zéros à l'origine et une très longue queue. Le coefficient de variation est très élevé : $s/m = 3.1$. Les 13 plus fortes valeurs représentent ici 3% des données et 40% de la moyenne arithmétique.

L'hypothèse d'indépendance interne entre la variable et son champ nous apparaît qualitativement compatible avec l'observation des données (Figure 8). On n'observe pas de zonéographie prononcée. Les transects numérotés 2,4,8,12 montrent des paquets de valeurs très fortes. Elles se localisent aussi bien en bordure de champ (transects 3 et 4), que plus en milieu du champ (transects 8 et 12). L'hypothèse d'indépendance interne entre la variable et son champ a déjà été faite dans la science des pêches. Sur le hareng de la Mer du Nord autour de l'Ecosse et des îles Shetland, Mc LENNAN et al. (1988) séparent la variable de la géométrie de son champ. L'aire de dispersion de cette population de hareng est beaucoup plus vaste que la nôtre, le poisson y forme de grandes zones disjointes de présence. La distribution spatiale est interprétée par ces auteurs comme étant uniquement sous le contrôle de la présence/absence du poisson. Dans les zones de présence (habitats), dont la localisation géographique est constante dans le temps, la distribution spatiale du poisson est supposée stationnaire. PENNINGTON (1983) a une analyse similaire pour les oeufs de Maquereau sur le Banc George (USA) : il sépare les pêches nulles des pêches non nulles et fait l'hypothèse stationnaire et pépétique dans les zones des valeurs non nulles.

B.2.4. L'estimation

Comme le plan d'échantillonnage est systématique, on prend comme estimateur de la moyenne dans V , la moyenne arithmétique des données. L'estimateur de la quantité totale est : $Q^* = Z_V^* V = 3.3 \cdot 10^6$ "backscattering units".

Les campagnes d'écho-intégration effectuées le long de transects parallèles et équidistants donnent la possibilité de calculer la variance d'estimation selon deux procédures, à 1D et à 2D. Le calcul à 2D est mené dans le plan longitude-latitude à l'intérieur du champ V défini préalablement et avec le modèle de variogramme inféré dans V . L'idée du calcul à 1D vient du principe en transitif de composition des termes de ligne et de tranche. Comme les lignes sont reconnues exhaustivement, il ne reste que le terme de tranche. Or on a vu que son calcul se mène à 1D sur la droite perpendiculaire aux transects avec comme maille la distance inter-transects et comme covariogramme transitif celui des valeurs cumulées le long des lignes (covariogramme monté). Nous appliquons successivement les deux méthodes.

B.3. Calcul de la variance d'estimation à 2D

L'estimateur de $\frac{1}{V} \int_V Z(x) dx$ est la moyenne arithmétique des données dans

$V : Z_V^* = 1690$. La notation Z_V^* pour la moyenne arithmétique des données rappelle qu'il s'agit de l'estimateur d'une moyenne spatiale. On calcule la variance de cette estimation, $\text{Var}(Z_V - Z_V^*)$.

B.3.1. Analyse variographique

L'échantillonnage préférentiel dans la direction N-S des transects ne permet pas d'accéder expérimentalement au comportement du variogramme près de l'origine $h=0$ dans les autres directions. Le variogramme est calculé à l'intérieur du champ V précédemment défini, dans les quatre directions sur les données brutes (Figure 10). Le long des transects (N-S), il croît très rapidement puis chute et se stabilise sous la variance. Ce type de comportement a déjà été observé et analysé sur des données d'Uranium aux histogrammes très dissymétriques (RIVOIRARD, 1986) : l'effet de quelques valeurs très fortes se manifeste pour certaines distances en fonction de leur implantation dans le champ. D'où des sauts dans le variogramme expérimental dès que les valeurs fortes interviennent ou n'interviennent plus dans les couples $(z(x), z(x+h))$ dont on prend les différences carrées. Le variogramme, par son expression mathématique, est très sensible aux fortes valeurs. Dans chacune des autres directions, le variogramme est stabilisé autour d'un palier assez différent pour chaque direction, dès la première distance de 5 m. Le variogramme ne présente pas de portée plus longue dans les directions inter-transect que le long des transects. Une portée plus courte est possible mais aucun argument expérimental n'est disponible. En général, l'hétérogénéité est plus importante le long du gradient côte large que le long des isobathes. Le contraire (portée plus courte perpendiculairement aux transects) serait une situation très particulière. Une hypothèse d'isotropie est raisonnable.

Les fortes valeurs extrêmes ont un effet destructurant car elles contribuent très fortement à la valeur du variogramme quand elles sont prises dans les couples $(z(x), z(x+h))$, pour un vecteur h donné. Afin de stabiliser la variance des incréments $z(x) - z(x+h)$, on a transformé les données. La transformation utilisée est une normalisation par anamorphose gaussienne empirique (LANTUEJOL et RIVOIRARD, 1984). Cette transformation ϕ consiste à assigner à la valeur brute z_i sa transformée gaussienne y_i dont la fréquence cumulée sur la courbe gaussienne est égale à la fréquence cumulée expérimentale de z_i : $Z(x) = \phi(Y(x))$ où Y est la gaussienne centrée réduite. Le variogramme calculé dans les quatre directions sur les données transformées est donné sur la Figure 11. Le variogramme a une allure compatible avec un modèle sphérique. La stationnarité de la variable, même aux grandes distances, est confirmée. Dans toutes les directions, on observe une stabilisation du variogramme autour du même palier, à partir

d'une même distance entre 5 et 10 m.n L'isotropie reste une hypothèse raisonnable. Un modèle sphérique isotrope a été ajusté avec un effet de pépite. L'effet de pépite $C_0 \delta$ (avec δ la mesure de Dirac) est l'équivalent d'un bruit blanc de variance C_0 ; il modélise des transitions très rapides dans la régionalisation à très petite échelle spatiale et se marque par une discontinuité à l'origine sur le variogramme (MATHERON, 1970). Le modèle sphérique a des propriétés qui sont bien adaptées ici à rendre compte de la structure. Il croît avec la distance jusqu'à la portée r , après quoi il est constant et égal au palier C : $\gamma(h) = C(1, 5h/r - 0, 5h/r) \quad h \leq r$; $\gamma(h) = C \quad h \geq r$. Sa pente, à l'origine, est une droite ; elle modélise de ce fait une structuration pas trop régulière. La portée est la distance moyenne entre deux points $x, x+h$ après laquelle il n'y a plus de corrélations spatiales. Le palier mesure le degré maximal d'hétérogénéité entre points non corrélés. Les paramètres du modèle ajusté sont : $C_0 = 0,20$; $C = 0,78$; $r = 7$ m.n La portée est un peu plus grande que la distance inter-transect.

La modélisation du variogramme des données brutes n'est pas immédiate à cause de l'effet destructurant des fortes valeurs extrêmes. Dans cette tâche, le variogramme des données normalisées nous sert de guide qualitatif, car il nous assure des points suivants. Le variogramme (arithmétique) des données brutes est isotrope. Il présente une montée réelle qui n'est pas due au seul effet d'une valeur forte. Il atteint un palier unique et présente une portée entre 5 et 10 m.n qui n'est pas due à un simple effet de lognormalité (MATHERON, 1974). On a ajusté un modèle sphérique avec un effet de pépite : $C_0 = 0,15 \cdot 10^8$; $C = 0,12 \cdot 10^8$; $r = 7$ m.n

On argumente le choix de la pépite et de la valeur du palier. On interprète la pépite en termes de micro-structures et non en termes d'erreurs de mesure que nous pensons faibles. Sur le variogramme expérimental, le premier point est à 1 m.n Sa valeur est déjà très élevée et représente 59 % de la variance des données. 59 % de l'hétérogénéité totale est constituée par des structures spatiales aux transitions très rapides sur des distances inférieures à 1 m.n La mesure d'écho-intégration régularise les structures spatiales pour les distances inférieures à la distance u.e.d.e. (ici 1 m.n), et crée des discontinuités entre les données. On est typiquement dans le cas de la genèse d'un effet de pépite par la régularisation de structures aux portées plus courtes que le support des données. MATHERON (1970) montre que les composantes structurales de petites échelles se comportent alors comme une composante complètement aléatoire et discontinue. Précisons que le taux de recouvrement entre les supports des données $M_{u.e.d.e}$ est faible par rapport aux dimensions du support. L'effet régularisant à l'origine du variogramme dû à l'autoconvolution géométrique du support est donc négligé. On a calé la pépite très près du premier point expérimental : elle représente 56 % du palier ($\frac{C_0}{C + C_0} = 56\%$). Les transitions rapides aux petites distances s'observent généralement très bien sur les échogrammes. Malheureusement nous n'en disposons pas pour nos données.

Pour ce qui est du palier, sa valeur est mal connue à cause des différences de palier dans les différentes directions. Celles-ci sont interprétées comme un effet statistique causé par les valeurs fortes extrêmes et non comme une réalité biologique, où l'hétérogénéité serait fonction de la direction (anisotropie zonale). Le variogramme de la gaussienne ne présente qu'un palier, le variogramme des données brutes n'en a qu'un aussi. On a calé le palier, $C_0 + C$, sur la variance expérimentale, par mesure de sécurité. Ainsi, lors de la modélisation, on cherche à rendre compte d'une structure sous-jacente moyenne, plausible, en s'extrayant des effets statistiques dûs aux fortes valeurs. Le modèle est un résumé de la lecture des données. Il a un sens physique et biologique et est contrôlable par des échantillonnages (MATHERON, 1989).

B.3.2. Variance d'estimation à champ fixé

Le modèle choisi de variogramme est une somme de deux structures : un effet de pépite et un modèle sphérique. L'expression de la variance d'estimation est linéaire par rapport au variogramme. La contribution de chaque structure peut donc être individualisée et calculée pour elle-même.

Pour la composante pépitique, la variance d'estimation est donnée par la formule C_0/n , où C_0 est la constante de pépite et n le nombre d'échantillons. On est dans le cas de la statistique classique. On a donc : $\sigma_0^2 = 34483$.

La contribution de la structure sphérique à la variance d'estimation correspond au terme de tranche et est calculée en utilisant le principe d'approximation par composition des variances élémentaires. La variance élémentaire est la variance d'extension de la ligne centrale dans son rectangle d'influence. La largeur des transects est celle de la base du cône insonifié, à chaque profondeur. Son ordre de grandeur est au maximum quelques centaines de mètres. On néglige donc l'épaisseur des transects et on les considère comme des lignes. Le principe d'approximation est valable pour une variable ponctuelle et son variogramme. Considérant le support de régularisation très petit devant les surfaces mises en jeu, on néglige le fait que le variogramme est celui d'une variable régularisée sur des segments. D'ailleurs, on a déjà négligé l'effet de la régularisation dans le choix du modèle de variogramme à l'origine $h=0$. Cette façon de faire a déjà été utilisée par GOHIN (1985) pour un échantillonnage similaire. Le terme de tranche dû à la composante sphérique s'écrit donc :

$$\sigma_1^2 = \frac{\sum_i b_i^2 \sigma_i^2}{(\sum_i b_i)^2} \text{ où } b_i \text{ est la longueur (en m.n) du transect } i \text{ et } \sigma_i^2 \text{ la variance élémentaire } i.$$

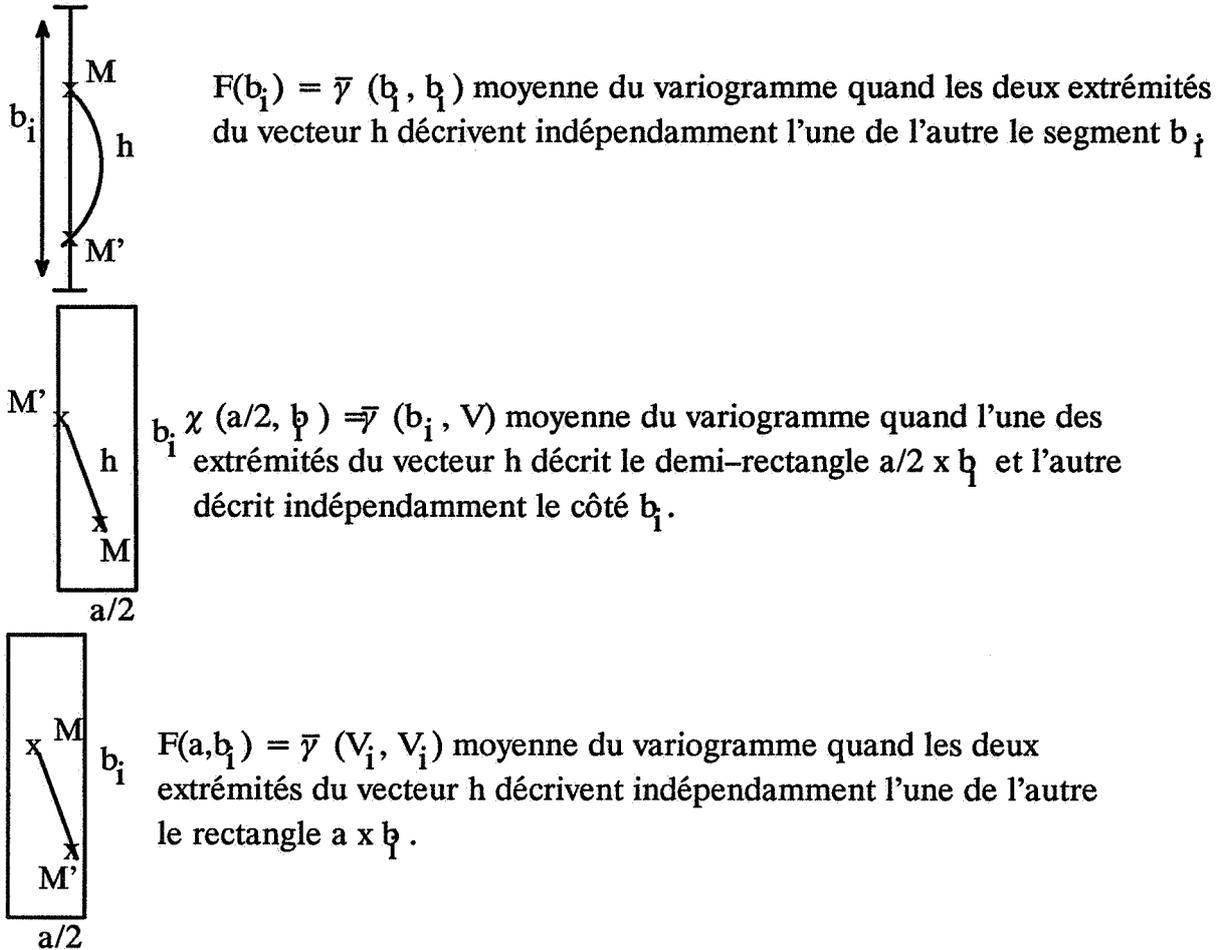
Les transects numérotés 7,8,9 sont coupés en deux morceaux par l'indentation des zéros extérieurs. L'erreur d'estimation élémentaire ne dépend pas dans le modèle de sa localisation géographique. On a considéré que, de chaque bord des zéros extérieurs, les erreurs d'estimations dans les deux rectangles sur chacun de ces transects étaient

I.B. Application à une prospection acoustique à maille régulière

indépendantes. Elles ont été composées comme si elles étaient situées sur des transects différents. On calcule donc 18 variances élémentaires σ_i^2 .

Le calcul fait intervenir les fonctions auxiliaires F à 1D et F et χ à 2D :

$$\sigma_i^2 = 2\chi(a/2, b_i) - F(a, b_i) - F(b_i).$$



Pour le modèle sphérique, ces intégrales du variogramme ont été résolues formellement. GUIBAL (1973) en donne les expressions. Elles sont assez lourdes et on ne les reproduit pas ici.

On a donc programmé ces expressions. Les entrées du programme sont le nombre de portions b_i de transects, les longueurs b_i , la distance inter-transect a , la portée et le palier du variogramme. Chacune des 18 variances σ_i^2 aurait pu être lue sur l'abaque établie à cette fin (MATHERON 1970, JOURNEL 1978). Toutefois, une campagne pouvant présenter un nombre élevé de transects, on a préféré programmer les formules données par GUIBAL (1973). On trouve : $\sigma_i^2 = 9200$.

La variance d'estimation est finalement :

$$\sigma_{2D}^2 = \frac{C_0}{n} + \frac{\sum_i b_i^2 \sigma_i^2}{(\sum b_i)^2} = 43683$$

L'écart-type relatif de l'estimation est : $\frac{\sigma_{2D}}{Z_V^*} = 12,4 \%$

L'effet de pépité contribue pour 79 % à la variance d'estimation. Il est une caractéristique importante de la régionalisation. La variance d'estimation obtenue sans tenir compte des structures spatiales, c'est-à-dire calculée comme si la structure spatiale

était une pépité pure, donne un coefficient de variation égal à : $\frac{S/\sqrt{n}}{Z_V^*} = 13,8 \%$. Bien

que légèrement moins bonne, cette précision est du même ordre de grandeur que celle obtenue en tenant compte des corrélations spatiales. Celles-ci sont assez faibles : pépité importante (56 % du palier), portée petite (7 m.n) vis à vis du champ (1975 m.n²). Leur impact sur la variance est en conséquence.

B.4. Calcul de la variance à 1D

Les données sont des régularisées sur des supports particuliers. Ils sont d'épaisseur négligeable dans les directions inter-transects, mais sont des segments de 1 m.n le long des transects. De plus, les supports des mesures sont contigus de telle sorte que l'échantillonnage des lignes est exhaustif. On ne fait pas d'erreur d'estimation le long des lignes. On met à profit cette circonstance particulière. Dans la méthode transitive, d'après le principe d'approximation à 2D, la variance d'estimation se met sous la forme d'un terme de ligne et d'un terme de tranche (§ A.1.3.2.). Ici le terme de ligne n'intervient pas, il ne reste que le terme de tranche. Or, on a vu que celui-ci se calcule avec le covariogramme de la variable cumulée (montée) le long des lignes. On se propose donc de calculer la variance de cette façon. L'intérêt de cette méthode de calcul est sa grande simplicité.

Les transects sont parallèles à l'axe des y. La variable ponctuelle cumulée sur les y

est : $q(x) = \int_{-\infty}^{+\infty} f(x, y) dy$. Comme les données ne sont pas ponctuelles mais des

moyennes exactes sur des segments contigus, on a : $q(x) = a_2 \sum_{k=-\infty}^{+\infty} M_{uede}(x, ka_2)$ avec a_2

la distance u.e.d.e. (1 m.n). En cumulant simplement les valeurs le long des transects, on obtient un jeu de données $q(x)$ monodimensionnel où chaque donnée est proportionnelle à la biomasse de poisson par ligne. On a 17 valeurs $q(x)$. Elles ont un support ponctuel sur la droite et y sont implantées régulièrement tous les 4,54 m.n Les deux accumulations les plus à l'Ouest sont nulles. On ne travaille que sur les 15 données non nulles. Leurs statistiques élémentaires sont : moyenne arithmétique : $q_L = 49006$; variance $S_q^2 = 3,11 \cdot 10^9$; coefficient de variation : $S_q/q_L = 1,14$. Les accumulations sont moins

hétérogènes que les données dans le plan. La Figure 12 donne une représentation des accumulations $q(x)$.

Un petit appauvrissement est sensible vers l'Est, la bordure Ouest est très franche. On n'observe pas de zonéographie marquée. On pourra faire, sur les $q(x)$, l'hypothèse intrinsèque.

L'estimateur de la quantité totale s'obtient en multipliant la moyenne arithmétique des 15 valeurs $q(x)$ par la longueur des 15 segments de 4,54 m.n : $Q^* = 3,3.10^6$. L'estimation n'est bien entendu pas changée. On calcule la variance d'estimation en transitif et en intrinsèque, où l'on étudie l'effet de la géométrie du champ sur le variogramme.

B.4.1. Calcul en intrinsèque

B.4.1.1. Variographie

Le variogramme des $q(x)$ est calculé à 1D (Figure 13). Sa modélisation n'est pas immédiate. Le variogramme de la variable monté sur l'un des axes est lié au variogramme à 2D. Dans la théorie intrinsèque, comme la régionalisation de la variable est indépendante de la géométrie du champ, une formule relie des deux variogrammes dans le cas où l'accumulation a lieu sur des hauteurs identiques le long de chacune des lignes (montée sous puissance constante). Dans la situation présente, les transects ont des longueurs différentes. La relation entre variogramme monté 1D et variogramme 2D n'est donc pas simple car elle fait intervenir un effet géométrie du champ. Toutefois, l'étude théorique de la montée (MATHERON, 1965, 1970) nous donne des arguments pour modéliser le variogramme des valeurs $q(x)$ près de l'origine et aux grandes distances.

Comme toujours, pour le variogramme expérimental, on ne dispose pas de point en $h=0$. Ici le premier point se trouve pour la distance inter-transect à 4,54 m.n A l'origine pourtant, le variogramme ne peut présenter un effet de pépité important. En effet, la pépité disparaît à la montée. Pour le voir, il suffit de remplacer dans la formule de montée, le variogramme 2D par un effet de pépité. Prenons par exemple la formule en transitif qui est plus simple qu'en intrinsèque :

$$g_1(h) = \int_{-\infty}^{+\infty} g_2(\sqrt{h^2 + x^2}) dx ; g_2(x) = C_0 \delta \text{ avec } \delta \text{ la mesure de Dirac } (\delta = 1 \text{ si } x=0$$

et $\delta = 0$ si $x > 0$). On a : $g_1(0) = [C_0 x]_0^0 = 0$ et $g(h) = 0$ pour $h > 0$.

Dans notre exemple, on n'a donc pas modélisé d'effet de pépité.

Pour un processus à 2D pépitique, l'accumulation sous puissance constante génère un processus 1D pépitique mais très peu variable. En effet, la moyenne le long des lignes est constante. Si l'accumulation a lieu sous puissance non constante, la structure spatiale 1D est alors engendrée par celle des puissances. Dans le cas du hareng, les accumulations $q(x)$ recèlent de l'information géométrique venue du champ. En effet, comme les longueurs totales de valeurs non nulles le long des transects (traversées de

présence de poissons) sont différentes de transect à transect, les accumulations gardent une réminiscence de l'épaisseur Nord-Sud de la frayère. Cet effet géométrique est fonction de la liaison qui existe biologiquement entre la densité en poisson et l'extension vers le large de la frayère (largeur N-S). La Figure 14 nous permet de visualiser cette relation : elle donne les accumulations $q(x)$ en fonction de la longueur totale des traversées de poisson le long des lignes. On observe deux extrêmes : une très forte accumulation de biomasse se développe sur une petite traversée (transect 3) et trois longs transects portent des accumulations faibles (transects 6,9,10). Entre ces deux extrêmes, on perçoit une certaine tendance à ce que les transects les plus long cumulent des biomasses croissantes. Ceci est en accord avec l'hypothèse stationnaire à 2D révélée par l'existence d'un palier sur le variogramme 2D : à densité moyenne constante, les lignes les plus longues ont tendance à supporter plus de biomasse. Ainsi, les valeurs moyennes de $q(x)$ montrent un certain lien à la géométrie N-S du champ. Il est sans doute masqué numériquement sur le variogramme expérimental à cause de l'existence sur le bord du champ de la très forte valeur de $q(x)$ (Figure 12) qui contribue à la croissance du variogramme aux grandes distances.

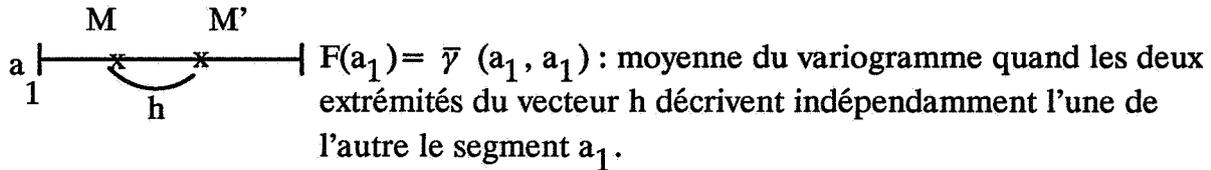
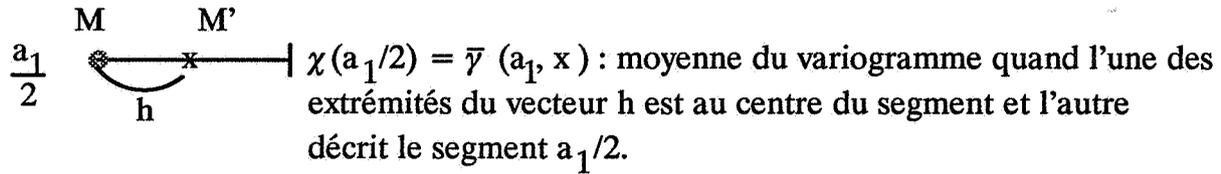
Le modèle choisi est une somme de deux structures : un modèle sphérique de courte portée et un modèle linéaire. On fait choix d'un comportement linéaire à l'origine du variogramme, alors que théoriquement MATHERON (1965, 1970) montre l'effet régularisant de la montée : un variogramme linéaire à l'origine à 2D donne par montée d'ordre 1 un comportement parabolique à 1D. L'absence de pépite nous suffit ici pour modéliser la régularité du variogramme à l'origine. Les paramètres de l'ajustement sont les suivants : le sphérique a un palier $C=0,21.10^{10}$ et une portée $r=5$ m.n ; le linéaire a une pente $A=0,3.10^8$.

B.4.1.2. Calcul de la variance d'estimation

On estime la moyenne d'espace $\frac{1}{L} \int_L q(x) dx$ ($L=15 \times 4,54$ m.n) par la moyenne arithmétique des 15 valeurs expérimentales $q(x)$. La variance d'estimation est calculée par composition de variances élémentaires bien que la maille $a_1 = 4,54$ m.n soit très proche de la portée du variogramme sphérique. Chaque variance élémentaire est la variance d'estimation de l'estimation de la moyenne du segment a_2 par la valeur $q(x)$ en son point central. Chacune des structures sphérique et linéaire apporte sa contribution à la variance d'estimation. On a :

$$\sigma_{cs}^2 = \frac{\sigma_{elem}^2}{15} \quad \text{avec} \quad \sigma_{elem}^2 = \sigma_{sph}^2 + \sigma_{lin}^2$$

Chacune des variances élémentaires se calcule grâce à la formule générale (I.3) : $\sigma^2 = 2 \bar{\gamma}(a_1, x) - \bar{\gamma}(a_1, a_1)$. Les intégrales $\bar{\gamma}$ du variogramme sont les fonctions auxiliaires F et χ à 1D :



La fonction χ est une intégrale simple ; la fonction F est une intégrale double. La procédure de Cauchy permet de la réduire à une intégrale simple. Les modèles sont des polynômes de h, l'intégration en est simple. Le détail du calcul peut être trouvé dans JOURNAL (1978, chap. II.E). Le calcul donne, avec les mêmes notations que

précédemment, : $\sigma_{\text{lin}}^2 = \frac{Aa_1}{6}$ et $\sigma_{\text{Sph}}^2 = C\left(\frac{1}{4} \frac{a_1}{r} + \frac{3}{160} \frac{a_1^3}{r^3}\right)$.

Finalement, on obtient : $\sigma_{\text{es}}/q_L^* = 12, 12\%$.

Ce résultat s'accorde avec la précision calculée avec le modèle structural à 2D. Ceci n'est pas étonnant car les deux modèles sont liés. Toutefois, l'écart-type relatif d'estimation est un peu plus faible qu'à 2D. On verra au paragraphe B.5 que la prise en compte du terme d'erreur géométrique rapproche les deux précisions intrinsèques.

Les accumulations $q(x)$ ne peuvent pas être pépitiqes à 1D, à cause du lien entre les structures 2D et 1D. La montée a un effet sur le palier du variogramme (variance) et sur le comportement à l'origine (structuration). Le palier est diminué, la pépite a tendance à disparaître. A l'origine, le variogramme est plus régulier. Le jeu de données obtenu par montée le long des transects est moins hétérogène et plus régulier. Cet effet tend à rendre plus facile le traitement des données, alors qu'à 2D les très fortes valeurs nous avaient obligé à transformer les données. Le variogramme est ici en-dessous de la variance pour les distances inférieures à 30 m.n Le coefficient de variation, calculé comme si les accumulations étaient pépitiqes, est $\frac{Sq/\sqrt{n_q}}{q_L^*} = 29, 4\%$ et est erroné.

B.4.2. Calcul en transitif

Le calcul est encore plus simple en transitif et aussi plus général car l'effet géométrique du champ peut entraîner, dans certains cas, une zonéographie sur les accumulations.

La quantité à estimer est l'intégrale : $Q = \int_{-\infty}^{+\infty} q(x) dx$. Son estimateur est ici la

somme discrète : $Q^* = a_1 \sum_{k=1}^{15} q(x_0 + ka_1)$. On a bien évidemment $Q^* = 3,3 \cdot 10^6$

"backscattering units".

B.4.2.1. Analyse structurale

Le covariogramme expérimental des accumulations $q(x)$ et sa modélisation sont présentés sur la Figure 15. Le covariogramme expérimental présente les mêmes caractéristiques que le variogramme : une décroissance très rapide des corrélations puis une décroissance moins rapide. Le modèle choisi est une somme de deux structures sphériques, l'une de portée $r_1 = 7$ m.n et de palier $C_1 = 0,21 \cdot 10^{12}$; l'autre de portée $r_2 = 70$ m.n et de palier $C_2 = 0,15 \cdot 10^{12}$. Sur le covariogramme expérimental, on dispose d'une

valeur à l'origine : $g(0) = \sum_{k=1}^{15} q^2(x_k)$. On a choisi ici aussi un comportement linéaire à

l'origine et pas de pépite. Le covariogramme expérimental présente deux bosses, l'une autour de 20 m.n, l'autre autour de 40 m.n Elles sont dues aux deux fortes valeurs sur les transects numérotés 3 et 12 (Figure 12). Le modèle choisi ne passe pas par ces fluctuations statistiques.

B.4.2.2. Calcul de la variance d'estimation

La variance d'estimation s'écrit :

$$\sigma_{cs}^2 = \text{Var}(Q - Q^*) = a_1 \sum_{-\infty}^{+\infty} g(ka_1) - \int_{-\infty}^{+\infty} g(h) dh$$

Le modèle sphérique de covariogramme est un modèle de covariance qui s'écrit :

$$g(h) = C(1 - 1,5h/r + 0,5h^3/r^3) \text{ si } h \leq r ; g(h) = 0 \text{ si } h \geq r$$

où C et r sont le palier et la portée. L'intégrale $\int_{-\infty}^{+\infty} g(h) dh$ d'un sphérique se calcule

facilement car la fonction $g(h)$ est nulle après r . L'intégrale se réduit à $\frac{3}{4} Cr$. Le nombre

de termes dans la somme discrète est fini. Il dépend de la portée et de la maille a_1 . Ici ce nombre est faible car la distance inter-transect $a_1 = 4,54$ m.n est importante : il y a trois termes pour le modèle de portée 7 m.n et 17 termes pour le modèle de portée 70 m.n La formule de la variance se calcule donc directement et facilement.

Le calcul donne : $\frac{\sigma_{cs}}{Q^*} = 12,5\%$.

Ce résultat s'accorde bien avec les deux précédents. La variance est facile à calculer et se calcule sans approximations.

La méthode est générale car elle peut permettre de traiter des données $q(x)$ zonées. En effet, dans un cas où l'épaisseur du champ le long des transects présente une zonéographie marquée, on peut s'attendre, si la régionalisation à 2D est stationnaire, à ce que les accumulations $q(x)$ présentent aussi cette zonéographie. Pour le calcul de la variance de l'estimation globale, la méthode transitive sur les accumulations des transects se pose comme la méthode privilégiée attachée aux campagnes d'écho-intégrations le long de transects parallèles et équidistants. Un premier calcul de variance peut être envisager très simplement sur le bateau avant que de débarquer de la campagne. La formule transitive va nous permettre de quantifier l'évolution de la variance d'estimation en fonction de la distance inter-transect (paragraphe C).

La valeur, en transitif, de l'écart-type relatif d'estimation est un peu supérieure à celles calculées en intrinsèque. En intrinsèque, la variable est étudiée indépendamment de son champ, et le calcul de variance a été effectué à champ fixé. On calcule maintenant le terme de variance géométrique représentant l'effet sur l'estimation de la moyenne, de l'incertitude sur les limites du champ.

B.5. Variance d'estimation intrinsèque et limites du champ

Dans l'analyse intrinsèque, on sépare l'étude de la variable de celle de son champ. L'étude de la variable est probabiliste (théorie intrinsèque), celle du champ est déterministe (théorie transitive). On calcule ici le terme géométrique sur l'estimation de la moyenne dans le modèle à 2D, puis dans le modèle monté à 1D.

B.5.1. Variance de l'erreur géométrique sur l'estimation à 2D

La définition d'un champ V soulève des problèmes d'interprétation des zéros en zéros extérieurs et intérieurs. Pourtant, la reconnaissance des zéros extérieurs directement sur les échantillons est importante. Tout d'abord parce que c'est la régionalisation intrinsèque au poisson qui est sujet d'étude. La variance des données calculée avec les zéros extérieurs n'est pas la dispersion expérimentale de la régionalisation du poisson, et le variogramme calculé avec les zéros extérieurs mélange la structure intrinsèque du poisson à la géométrie du champ de sa présence. En second lieu, et de façon tout aussi importante, la reconnaissance des zéros extérieurs définit la surface à l'intérieur de laquelle les erreurs d'estimation ont bien lieu. En dehors, il ne peut y avoir

d'erreur d'estimation, car il n'y a pas de poisson de façon certaine. De par son expression (I.3), la variance d'estimation est liée à la taille et à la géométrie du champ sur laquelle elle est calculée.

Les limites du champ jouent donc sur la variance d'estimation pour deux raisons. D'une part, la distinction entre les zéros extérieurs et intérieurs laissent différentes interprétations possibles donnant autant de champs différents. D'autre part, pour chaque choix de champ, l'erreur d'estimation de ce contour par la maille d'échantillonnage crée un effet de bordure se répercutant sur l'estimation de la moyenne (paragraphe A.1.2.3.). Dans notre cas, comme le pas des mesures est très petit le long des transects, c'est plutôt l'erreur commise dans la distinction entre zéros extérieurs et intérieurs qui est prépondérante le long des transects. En revanche, dans les directions inter-transects, c'est plutôt l'effet de bordure qui prévaut. Ainsi, l'erreur faite sur les limites du champ a un statut mixte. On se propose de mesurer l'influence de l'interprétation des zéros le long des transects, sur la variance d'estimation, par une analyse de sensibilité succincte pour différents choix de champs. Cette analyse est effectuée avec le variogramme inféré précédemment, c'est à dire à variogramme fixé. Ensuite, on calcule le terme d'effet de bordure qui s'ajoute à la variance d'estimation dans chaque choix de champ. On encadre le choix précédent de V (Figure 8) par deux interprétations extrêmes des zéros. Tout d'abord, on considère tous les zéros comme extérieurs. Les transects sont découpés en petits rectangles tous reconnus par leur segment central. Il y a 28 variances élémentaires σ_1^2 à composer. L'estimation donne : $n = 400$; $Z_V^* = 1838$; $\sigma_{2D}^2 = 47233$ et $\frac{\sigma_{2D}}{Z_V^*} = 11,8$

%. La moyenne est augmentée naturellement. La variance est aussi augmentée car la variance élémentaire croît quand la longueur de la ligne médiane diminue, à distance inter-transect et portée de variogramme fixées. Le coefficient de variation est plus faible que celui calculé précédemment : la surface de V est plus faible. En second lieu, on considère maintenant tous les zéros extérieurs encadrés par des valeurs non nulles comme des zéros intérieurs. Il y a 15 variances élémentaires à comparer. Les statistiques de l'estimation sont : $n = 480$; $Z_V^* = 1532$; $\sigma_{2D}^2 = 39704$ et $\frac{\sigma_{2D}}{Z_V^*} = 13,0$ %.

La moyenne est diminuée, naturellement. La variance aussi car les lignes centrales estimant les rectangles sont plus longues, donc les rectangles mieux estimés. Le coefficient de variation est plus élevé : la surface de V est plus importante.

La distinction entre zéros extérieurs et intérieurs le long des transects ne joue que sur 1 % de l'écart-type relatif de l'estimation de la moyenne. Celui-ci se situe entre 12 et 13 %. La sensibilité de la précision aux différentes interprétations de zéros est faible car les successions de zéros douteux (ceux qui sont encadrés par des valeurs non nulles) restent peu nombreuses, et car ces zéros ne représentent que 8 % de la surface du champ (Figure 8) sur lequel on a travaillé. On calcule maintenant le terme géométrique quantifiant l'effet de bordure sur l'estimation de la moyenne.

On considère que le réseau d'échantillonnage est flottant par rapport au champ V de présence du poisson. La variance d'estimation de V, σ_V^2 , se calcule en transitif en utilisant la variable indicatrice de V. La variance σ_V^2 se décompose en un terme de ligne et en un terme de tranche. Comme les supports u.e.d.e. sont contigus, on considère que les limites de V sont connues sans erreur le long des transects. Ainsi, on néglige le terme de ligne. La grille d'échantillonnage ne flotte par rapport au champ de présence du poisson que dans les directions inter-transects. Il n'y a à calculer que le terme de tranche. Pour ce faire, on doit modéliser le covariogramme monté de l'indicatrice du champ V de présence du poisson. La variable indicatrice vaut 1 à l'intérieur du champ V de la Figure 8 et 0 en dehors. On la cumule le long des transects. Son covariogramme géométrique est donné sur la Figure 16. Il présente à l'origine un comportement très régulier compatible avec un modèle au comportement parabolique. En effet, à 2D le covariogramme géométrique d'une indicatrice est nécessairement linéaire à l'origine (MATHERON, 1965, 1970). Par montée le long des transects, le covariogramme monté d'ordre 1 doit présenter un comportement parabolique (MATHERON, 1965, chap. II ; 1970). C'est bien ce que nous observons expérimentalement. Le modèle ajusté est une somme de deux modèles gaussiens. Le modèle de covariance gaussienne a un comportement parabolique à l'origine et présente une allure générale en S qui s'ajuste bien au covariogramme expérimental : $g(h) = Ce^{-h^2/r^2}$. Les paramètres C et r sont appelés respectivement palier et portée, bien que le modèle n'en possède pas de réels car le palier est asymptotique (portée pratique à $r\sqrt{3}$). Les paramètres de l'ajustement sont : $C_1 = 1318$ et $r_1 = 15$ m.n. pour la première structure, et $C_2 = 10500$ et $r_2 = 38$ m.n. pour la seconde. On calcule ici encore la variance d'estimation σ_V^2 , directement avec la formule générale transitive (I.1) :

$$\sigma^2 = a_1 \sum_{-\infty}^{+\infty} g(ka_1) - \int_{-\infty}^{+\infty} g(h) dh .$$

L'intégrale du modèle gaussien se calcule en utilisant

le fait que l'intégrale de la densité de Gauss est égale à 1 : $\int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}$. Le

changement de variable $\frac{h^2}{r^2} = \frac{x^2}{2}$ conduit à résoudre l'intégrale du modèle gaussien :

$$\int_{-\infty}^{+\infty} e^{-h^2/r^2} dh = r \sqrt{\pi} .$$

Ici comme le modèle n'a pas de portée, le nombre de termes

dans la somme discrète est infini, mais leur valeur tend vers zéro. On arrête la sommation dès que le terme à rajouter est inférieur strictement à 1 % de la somme déjà calculée. Pour le modèle de portée $r_1 = 15$ m.n., la somme comporte 8 termes. Pour le modèle de portée $r_2 = 38$ m.n., la somme en comporte 15. Le calcul donne finalement : $\sigma_V^2 = 43\,588$.

La variance de l'erreur géométrique sur l'estimation de la moyenne Z_V s'écrit :

$$\frac{\sigma_V^2}{V^2} D^2(o/V) .$$

On prend pour variance de dispersion $D^2(o/V)$, la variance S^2 des données

dans le champ de présence du poisson ($S^2 = 27,3 \cdot 10^6$). Le champ V , qui intervient sous le dénominateur, n'est pas le champ de présence du poisson, mais la surface reconnue par l'échantillonnage. C'est le rectangle dont la largeur est le nombre de transects multiplié par la distance inter-transect et dont la longueur est la longueur moyenne des transects : $V = n_{\text{tot}} \cdot a_1 \cdot a_2$ (m.n²) où n_{tot} est le nombre totale des données et (a_1, a_2) la maille. En considérant les 15 transects positifs, on obtient : $\frac{\sigma_V^2}{V^2} = 0,221 \cdot 10^{-4}$. Le terme d'erreur

géométrique est donc : $D^2(o/V) \frac{\sigma_V^2}{V^2} = 602$. Il s'ajoute à la variance déjà calculée mais ne représente que 1,4 % de la variance totale d'estimation. L'effet de bord sur l'estimation de la moyenne est dans notre cas négligeable à 2D. On a :

$$\sigma_{\text{tot}}^2 = \frac{C_o}{n} + \frac{\sum_i b_i^2 \sigma_i^2}{(\sum_i b_i)^2} + D^2(o/V) \frac{\sigma_V^2}{V^2} \text{ et } \frac{\sigma_{\text{tot}}}{Z_V^*} = 12,45\%$$

B.5.2. Variance de l'erreur géométrique sur l'estimation à 1D

En intrinsèque à 1D, on a séparé les accumulations de leur encaissant qui est un segment unique. Sa longueur est $L = n_t \cdot a_1$ où $a_1 = 4,54$ m.n est la distance inter-transect, et $n_t = 15$ est le nombre des accumulations non nulles. Ce segment mesure la largeur de la frayère sur les longitudes. Le réseau des points de maille a_1 est flottant par rapport aux extrémités du segment L . De même que précédemment, on calcule ici le terme $D^2(o/L) \frac{\sigma_L^2}{L^2}$. On prend comme valeur numérique de la dispersion $D^2(o/L)$ la variance S_q^2 des 15 accumulations positives ($S_q^2 = 3,11 \cdot 10^9$).

La variance σ_L^2 de l'estimation de la longueur du segment à partir du réseau de maille a_1 , peut se calculer directement. Les 15 accumulations positives sont bordées à l'Ouest par deux transects entièrement nuls, les transects 1 et 2. On considère qu'un transect 18 fictif, implanté à l'Est, aurait aussi donné une accumulation nulle. Pour calculer σ_L^2 , on suit le raisonnement indiqué par MATHERON (1970, p.46) qui traite exactement ce cas. Les extrémités du segment peuvent tomber entre les transects 2 et 3 ou 17 et 18, indépendamment l'une de l'autre avec une probabilité uniforme. La longueur L est une variable aléatoire. Son espérance est : $E(L) = n_t \cdot a_1$. On peut donc écrire L comme suit : $L = (n_t - 1) a_1 + \epsilon_1 + \epsilon_2$, où ϵ_1 et ϵ_2 sont deux variables uniformes et indépendantes sur $[0, a_1]$. La variance $D^2(L)$ est la moyenne, sur toutes les implantations possibles des deux extrémités, des écarts carrés de L à sa moyenne : $D^2(L) = \frac{1}{2a_1} \int_0^{a_1} \epsilon^2 d\epsilon = \frac{a_1^2}{6}$.

Ainsi le terme d'erreur géométrique prend la valeur : $S_q^2 \frac{D^2(L)}{L^2} = 2,3037 \cdot 10^6$. Ce terme se rajoute à la variance d'estimation de la moyenne des accumulations à segment

I.B. Application à une prospection acoustique à maille régulière

fixé : $\sigma_{\text{tot}}^2 = \frac{\sigma_{\text{elem}}^2}{15} + S_q^2 \frac{D^2(L)}{L^2}$. Il représente 6 % de la variance totale d'estimation et n'est donc pas négligeable. L'écart-type relatif de l'estimation en intrinsèque à 1D est finalement : $\frac{\sigma_{\text{tot}}}{q_L^*} = 12,5\%$.

B.5.3. Conclusion

Le tableau suivant résume les résultats des calculs effectués :

Ecart-type d'estimation

	Structure spatiale ignorée		Prise en compte de la structure spatiale		
	2D	1D	2D	1D	
			Intrinsèque	Intrinsèque	Transitif
	$\frac{S_{2D}/\sqrt{n_{2D}}}{Z_V^*}$	$\frac{S_{1D}/\sqrt{n_{1D}}}{q_L^*}$	$\frac{\sigma_{2D}}{Z_V^*}$	$\frac{\sigma_{1D}}{q_L^*}$	$\frac{\sigma_{1D}}{Q^*}$
Sans l'erreur géométrique	13.8%	29.4%	12.4% (80% dûs à l'effet de pépite)	12.1%	
avec l'erreur géométrique			12.45%	12.50%	12.50%

En tenant compte des effets de bords sur l'estimation de la moyenne en intrinsèque, les trois estimations ont des précisions identiques : 12,5 %. Cette accord dans les résultats en transitif et en intrinsèque conforte bien a posteriori l'hypothèse d'indépendance interne entre la variable et son champ.

Sur les données à 2D, à cause de la faible structuration (fort effet de pépite, portée faible relativement au champ), la variance d'estimation calculée sans tenir compte de la structure spatiale est proche de celle donnée par la géostatistique. Sur les accumulations à 1D au contraire, la prise en compte de la structure spatiale est cruciale. L'accumulation génère un jeu de données moins hétérogène et bien structuré (pas d'effet de pépite et longue portée relativement au champ), facile à traiter. La structure des accumulations à 1D est d'autant plus marquée que la structure 2D est faible et que la géométrie du champ est régulière.

Le calcul de la variance en transitif sur les accumulations le long des transects s'impose comme la méthode la plus simple et la plus générale car la séparation entre la variable et son champ ne sera pas forcément toujours possible. C'est la plus appropriée aux campagnes d'écho-intégration le long de transects parallèles et équidistants. L'accumulation le long des transects met à profit la particularité du support des données d'écho-intégration et elle permet de travailler sur un jeu de données beaucoup plus régulier.

En intrinsèque comme en transitif, on demande à l'échantillonnage de révéler les zéros extérieurs aux extrémités des transects, de même que la position des transects bordiers entièrement nuls.

B.6. Discussion : rapports entre les variances d'estimation de la Statistique Classique et de la Géostatistique

Soit σ^2 la variance de dispersion sur tout le champ de présence du poisson. C'est la moyenne du variogramme pour toutes les distances dans V. On a : $\sigma^2 = \overline{\gamma_{VV}} = C(0) - \overline{C_{VV}}$. (γ désigne un variogramme à palier et C la covariance). Physiquement, la variance de dispersion est fonction du champ tant que celui-ci est petit devant la portée des corrélations spatiales.

Ainsi, le paramètre σ^2/n donne la variance d'estimation de la moyenne de la loi des échantillons quand ceux-ci peuvent être considérés comme des observations indépendantes d'une même loi, sur un champ infini. Appelons σ_E^2 la variance d'estimation calculée par la Géostatistique. C'est la variance d'estimation d'une intégrale d'espace sur un champ fini. On se propose d'examiner les rapports entre σ^2/n et σ_E^2 , en particulier on se propose d'examiner les situations dans lesquelles on peut s'attendre à ce que l'une des variances soit supérieure à l'autre. La portée des corrélations spatiales et la taille du champ vis-à-vis de la portée jouent un rôle essentiel.

On a vu dans l'étude précédente que à 2D la pente du variogramme est importante, que la portée est petite vis-à-vis de la longueur des transects, mais pas vis-à-vis de la distance inter-transect. On a observé que σ_E^2 et σ^2/n étaient du même ordre de grandeur. En revanche à 1D le variogramme des accumulations est très structuré, les corrélations spatiales persistent pour des distances de l'ordre de la moitié du champ. σ^2/n à 1D est très élevée, bien supérieure à la variance d'estimation σ_E^2 . Pour mieux comprendre les rapports entre σ_E^2 et σ^2/n , on réécrit σ_E^2 en terme de covariance : $\sigma_E^2 = \overline{C_{\alpha\beta}} - 2\overline{C_{\alpha V}} + \overline{C_{VV}}$ où V désigne le champ sur lequel on estime l'intégrale d'espace et où α et β sont les indices des échantillons.

Si le champ V est très grand vis-à-vis de la portée de la covariance, les termes $\overline{C_{VV}}$ et $\overline{C_{aV}}$ sont très petits. On les néglige. On est alors typiquement dans le cas dont traite la Statistique Classique. Il vient alors :

$$\sigma^2 \approx C(0) \quad \text{et} \quad \sigma_E^2 = \overline{C_{a\beta}} = \frac{C(0)}{n} + \frac{n^2 - n}{n^2} \sum_{a \neq \beta} \overline{C_{a\beta}}$$

La double somme sera en général positive. Il faut donc s'attendre à ce que l'on ait :

$$\frac{\sigma^2}{n} < \sigma_E^2$$

Quand la portée des corrélations est faible vis-à-vis des dimensions du champ, alors l'estimation de l'intégrale d'espace revient à estimer la moyenne de la loi des échantillons prélevés dans un champ infini. Quand les échantillons sont corrélés, il y a redondance d'information et l'on dispose de moins d'information pour estimer la moyenne que si l'on disposait de n échantillons indépendants. La variance σ^2/n est trop optimiste et sous-estime la variance d'estimation réelle. Ceci ne vaut que pour un champ très grand vis-à-vis de la portée petite.

Quand la portée n'est pas très petite vis-à-vis du champ, les termes de la variance dans lesquels intervient le champ, $\overline{C_{aV}}$ et $\overline{C_{VV}}$, ne sont plus négligeables. On ne se trouve plus dans la situation dont traite la Statistique Classique. La variance de dispersion σ^2 est alors intimement liée au champ ($\sigma^2 = C(0) - \overline{C_{VV}}$) et elle est inférieure au palier $C(0)$ du variogramme. Considérons une maille d'échantillonnage régulière constitué de n échantillons telle que chaque échantillon, x, soit au milieu de son carré d'influence, v, et telle que la réunion des carrés constitue le champ. Comme la portée est grande dans le champ, on considère sans peine que les dimensions de la maille élémentaire, v, sont inférieures à la portée. D'après le principe d'approximation, la variance d'estimation s'écrit : $\boxed{x}_v \sigma_E^2 = \frac{1}{n} (C_0 - 2\overline{C_{xv}} + \overline{C_{vv}})$. En faisant apparaître dans cette formule la variance de dispersion σ^2 , il vient :

$$\sigma_E^2 = \frac{\sigma^2}{n} - \frac{1}{n} (2\overline{C_{xv}} - \overline{C_{vv}} - \overline{C_{VV}})$$

$\overline{C_{vv}}$ fait intervenir des distances qui sont en moyenne plus grandes que celles que $\overline{C_{xv}}$ fait intervenir car le point échantillon x est au centre de son carré d'influence. Et $\overline{C_{VV}}$, lui, fait intervenir des distances qui sont en moyenne plus grandes que celles que $\overline{C_{vv}}$ fait intervenir. On a donc la relation d'ordre : $\overline{C_{VV}} < \overline{C_{vv}} < \overline{C_{xv}}$ et le terme entre parenthèse est positif.

Il faut donc s'attendre à ce que l'on ait :

$$\sigma_E^2 < \frac{\sigma^2}{n}$$

La formule σ^2/n ne représente une variance d'estimation que dans le cas d'un champ infini lorsque les données sont indépendantes. Lorsque la régionalisation est régulière et très structurée dans un champ petit devant la portée des structures, la prise en compte des corrélations spatiales est très importante dans le calcul de la variance, car l'estimation de l'intégrale d'espace se fait dans de bien meilleures conditions que ne le prévoit le paramètre σ^2/n puisqu'il surestime la variance d'estimation réelle. C'est bien ce que l'on observe sur les données montées à 1D de notre étude précédente sur le hareng.

Entre les deux situations précédentes, on passe par une taille de champ vis-à-vis de la portée telle que les variances σ_E^2 et σ^2/n sont sensiblement égales. C'est ici que nous situons notre jeu de données sur le hareng à 2D.

C. EFFICACITÉ DE L'ÉCHANTILLONNAGE

La réalité de la régionalisation n'est connue qu'au travers d'un échantillonnage, c'est-à-dire d'une information fragmentaire. Le jeu des données est la seule réalité sur laquelle il nous est donné de travailler et de ce fait, le modèle structural inféré collera aux données. Comme on ne perçoit la réalité que par elles, on est incapable de savoir, à partir d'un seul échantillonnage, si le jeu de données reflète la réalité ou s'il s'en écarte beaucoup. Les calculs géostatistiques de variance ne font pas intervenir l'histogramme des données, mais uniquement la structure spatiale. C'est la dispersion des valeurs dans le champ qui nous préoccupe maintenant car, lorsqu'elle est importante, elle soulève un difficile problème d'échantillonnage. Si la régionalisation réelle est très hétérogène (forte dispersion des valeurs) et que l'effort d'échantillonnage est trop faible par la couverture spatiale trop lâche qu'il réalise, alors les jeux de données spécifiques qui pourraient chacun être prélevés par des échantillonnages différents seront très différents. Les fluctuations expérimentales seront très importantes. Dans cette partie, nous proposons de faire la synthèse des informations historiques disponibles afin d'avoir une idée moyenne de la régionalisation réelle. Puis on donne des critères afin de choisir un effort d'échantillonnage approprié à cette régionalisation réelle.

La queue de l'histogramme (i.e. la fréquence des très fortes valeurs, leur moyenne, la valeur maximale) est connue avec une précision d'échantillonnage qui dépend de l'interaction entre celui-ci et la régionalisation réelle. Les conséquences de cette situation se manifestent sur l'estimation de la moyenne et sur le modèle structural inféré. La moyenne arithmétique est l'estimateur spatial logique pour un plan systématique. Il est toutefois très sensible aux très fortes valeurs rencontrées. Rappelons qu'elles concentrent une large fraction de la biomasse totale. Dans notre jeu de données sur le hareng norvégien, les valeurs supérieures à 10^4 représentent 3 % des données positives et 40 % de la biomasse ; celles supérieures à $5 \cdot 10^3$ représentent 8 % des données positives et 80 % de la biomasse. PENNINGTON (1983) sur les œufs de maquereau du Banc Georges (USA), rapporte une dispersion similaire : 10 % des données contribuent pour 85 % à la quantité totale des œufs. On peut donc dire d'un échantillonnage qu'il est efficace s'il permet d'estimer, avec une précision suffisante, la queue de l'histogramme. Les valeurs fortes se développent sur de petites surfaces dans l'espace (faible fréquence d'apparition dans les données). L'effort d'échantillonnage peut être dit approprié ou efficace s'il permet, par la couverture spatiale réalisée, d'échantillonner suffisamment les agrégats à forte densité.

La variance d'estimation mesure l'écart quadratique moyen entre la valeur vraie et les estimations effectuées par des campagnes répétées quand le modèle structural inféré sur les données, colle à la réalité. Le modèle structural dépend aussi de la rencontre ou non de très fortes valeurs à l'échantillonnage. Les très fortes valeurs, en petit nombre, agissent sur la structure observée : leur présence impose leur structure à l'ensemble ; elles ont tendance à masquer numériquement la structure plus régulière des plus faibles

valeurs; elles destructurent le variogramme expérimental. Ainsi, si peu de valeurs fortes sont rencontrées, la distribution spatiale nous apparaîtra régulière ; au contraire, si des valeurs fortes sont rencontrées, le modèle spatial sera moins structuré. L'effort à fournir croît avec l'irrégularité de la régionalisation réelle. Toutefois, l'inefficacité de l'échantillonnage, qui se traduit par la possibilité plus grande que le modèle spatial inféré s'écarte de la réalité, est un problème pratique et non théorique. En effet, la fourchette d'erreur autour de la moyenne estimée dépend de la valeur de celle-ci : pour un même effort et pour différents échantillonnages d'une même régionalisation, les modèles structuraux inférés seront différents, les fourchettes seront plus larges pour les fortes moyennes estimées et deux moyennes devraient avoir des fourchettes d'erreur qui se recourent. En revanche, en pratique comme on prendra toujours un jeu de données spécifique pour la réalité, on ne veut pas que ni la moyenne estimée ni le modèle spatial inféré ne reflètent uniquement que l'échantillonnage spécifique qui a permis leur estimation. Il est donc important, en calant un effort approprié, de diminuer les fluctuations expérimentales des grandeurs estimées, donc de rendre l'estimation et le modèle spatial plus indépendants des conditions spécifiques d'un échantillonnage. Mieux vaut effectivement ne pas aller en mer pour rien lorsque le risque est grand de passer à coté du stock sans le voir. Aucun moyen par la suite n'existera à partir du jeu de données pour se rendre compte de la situation.

En premier lieu, pour casser ce cercle vicieux, il nous faut avoir une idée sur la régionalisation réelle. On considère que, dans la plupart des cas, celle-ci est connue qualitativement pour des raisons historiques, soit parce que d'autres stocks similaires de la même espèce sont connus, soit parce que le suivi sur des décennies de la pêche professionnelle a permis d'accumuler des informations spatiales, soit parce que des échantillonnages du stock ont eu lieu sur plusieurs années à différentes saisons. Comme on n'a pas disposé dans ce travail de telles bases de données, on gardera le même jeu de données que précédemment, et l'on fera comme si ses caractéristiques correspondaient à ce qui se dégage, en moyenne, des jeux de données issus de plusieurs années d'observation directe du poisson en mer. Ce comportement moyen serait pour nous la réalité. On ferait donc l'hypothèse d'une certaine permanence de la régionalisation du poisson correspondant à un habitat.

Dans la partie D, nous discuterons de l'emploi de l'estimateur lognormal faisant intervenir la distribution statistique des données. Dans la présente partie, on étudie l'agencement dans l'espace des très fortes valeurs sur le jeu de données du hareng norvégien, afin de dégager des critères pour choisir un effort d'échantillonnage efficace. Dans un premier temps, on considère le cas d'une reconnaissance aveugle par transects parallèles et équidistants, et l'on donne un critère pour choisir la distance inter-transects. Dans un second temps, l'ajustement d'un modèle de krigeage disjonctif nous permet d'interpréter l'agencement des fortes valeurs par rapport aux valeurs plus faibles par un modèle de processus sous-jacent. Un schéma de Poisson pour l'implantation des grains à très fortes densités nous permet de construire un intervalle de confiance de la biomasse

globale sur un base uniquement géométrique, l'estimation de la fréquence des grains à partir de leur rencontre par l'échantillonnage. Dans la reconnaissance aveugle quand la distance inter-transects est diminuée, on augmente l'effort dans les zones riches comme dans les zones pauvres. Le modèle de krigeage disjonctif donne des arguments pour détecter, lors de l'échantillonnage, l'entrée dans une zone hébergeant de très fortes valeurs, et alors pour orienter spécifiquement l'effort d'échantillonnage sur ces zones potentiellement riches. Une telle stratégie moins coûteuse est souvent suivie en mer par les biologistes acousticiens, malgré un risque de biais. Elle est bien sûr violemment critiquée par certains statisticiens des pêches (JOLLY et al, 1987). Le débat du choix de la stratégie d'échantillonnage en acoustique reste encore ouvert (SIMMONDS et al, 1991).

C.1. Choix de la distance inter-transects dans une reconnaissance aveugle

Pour un plan systématique, l'effort peut être mesuré par la dimension de la maille. En général dans les campagnes d'écho-intégration, les transects ne sont pas tous de même longueur. Les longueurs des transects ne sont pas exactement prévisibles avant l'échantillonnage car elles se déterminent au cours de ce dernier : le bateau fait route jusqu'à pénétrer dans les zéros extérieurs bordiers qui marquent les limites de la régionalisation du poisson. L'effort d'échantillonnage ne peut donc se quantifier précisément avant la campagne d'écho-intégration qu'à l'aide d'un seul paramètre : la distance inter-transects.

Le choix de la distance inter-transects se pose en termes géométriques. Si la distance inter-transects est égale au diamètre moyen des agrégats très denses, alors les transects recouperont tous les agrégats, des fois sur leur bord, des fois en leur milieu, mais tous seront visités et la queue de l'histogramme sera échantillonnée avec une précision suffisante. Mais la taille des agrégats très denses peut être très petite (quelques m.n). Convenons de les appeler alors des grains.

On choisira alors la distance inter-transects en fonction de l'agencement de ces grains dans l'espace selon le même raisonnement. On a donc besoin de deux paramètres. L'un est la taille moyenne à 2D des grains. Elle est inférée grâce au recoupement des grains à 2D par des lignes parallèles et équidistantes implantées indépendamment de la position des grains. On utilisera pour ce faire un formule de stéréologie. L'autre est la taille moyenne de la zone d'influence des corrélations entre grains. On inférera la portée du variogramme géométrique des très fortes valeurs. On travaille donc sur une variable géométrique, l'indicatrice des très fortes valeurs : $i_4(x) = 1$ si $f(x) \geq 10^4$; $i_4(x) = 0$ sinon.

C.1.1. Distance inter-transects et variance d'estimation

Disposant du modèle de covariogramme monté le long des lignes, la formule transitive (I.1) à 1D de la variance d'estimation convient bien pour étudier la croissance de la précision quand l'effort d'échantillonnage augmente, car la variance d'estimation n'est fonction que de la distance inter-transects. La distance inter-transects de notre échantillonnage $a = 4,54$ m.n est prise comme référence. On exprime l'effort d'échantillonnage par le rapport a/d où d est une autre distance inter-transects. C'est elle que l'on fait varier. On calcule la variance d'estimation σ_{cs}^2 en transitif avec le covariogramme inféré en B.5, pour différentes valeurs de d . L'évolution de la précision relative σ_{cs}/Q^* , en fonction de l'effort d'échantillonnage relatif a/d , est donnée sur la Figure 17.

La courbe ne présente pas d'optimum. La vitesse de décroissance de la précision relative σ_{cs}/Q^* diminue quand l'effort a/d augmente. La portion "en coude" de la courbe est la partie la plus intéressante : c'est là en effet que le rapport précision de l'estimation sur effort investi est le plus avantageux, car pour des efforts pas trop grands encore, la chute de précision est encore importante. L'effort d'échantillonnage ($a/a = 1$) de notre prospection est justement dans cette partie de la courbe.

On cherche maintenant à expliquer ce fait, en analysant la géométrie des zones à forte densité de poisson.

C.1.2. Variogramme géométrique des très fortes valeurs

Sur la Figure 18, on donne une représentation de l'indicatrice $i_4(x)$ des très fortes valeurs de la queue de l'histogramme. Les cercles matérialisent les valeurs $i_4(x) = 1$ et les points les valeurs $i_4(x) = 0$. Il n'y a que 13 valeurs supérieures à 10^4 . On considère que chacune constitue une petite agrégation très dense. On voit que ces très fortes valeurs sont des fois isolées (transects 3,4,12), se regroupent parfois en amas très denses (transect 3) ou moins denses (transect 12). Le variogramme de $i_4(x)$ nous renseigne sur la zone d'influence des corrélations entre grains.

Le variogramme de $i_4(x)$ est donné sur la Figure 19. On observe qu'un même palier est atteint dans toutes les directions à partir d'une portée commune entre 5 et 8 m.n. On fait l'hypothèse d'isotropie de la structure pour les distances inférieures. Le premier point expérimental du variogramme (pour $h = 1$ m.n) a une valeur très élevée. Or, le variogramme mesure un ensemble géométrique amputé de son intersection avec son translaté de vecteur h . Il ne peut donc pas présenter d'effet de pépite et ne peut avoir qu'une croissance linéaire à l'origine (MATHERON, 1970). La croissance du variogramme de $i_4(x)$ est donc ici très rapide pour les distances inférieures à 1 m.n. Comme on observe sur la Figure 18 des grains isolés et des amas de grains, on a choisi un modèle gigogne composé de deux structures sphériques, l'une de portée 1 m.n., l'autre de portée 6

m.n marquant l'extension spatiale maximale des corrélations entre grains. Les paramètres de l'ajustement sont : $C_1 = 0,012$, r_1 , 1 m.n et $C_2 = 0,014$ et $r_2 = 6$ m.n.

On ne manque pas de remarquer ici que la portée du variogramme de l'indicatrice $i_4(x)$ est très proche de celle du variogramme moyen de la variable. Dans la structure spatiale moyenne, on ne voit que la présence des amas de grains à très forte densité de poisson. GOHIN (1985) a proposé de caler la distance inter-transects sur la portée du variogramme moyen. Ceci semble donc tout à fait simple et raisonnable. La pépite du variogramme moyen est engendrée par l'apparition des grains $i_4(x) = 1$ isolés dans la masse des valeurs plus faibles.

C.1.3. Commentaires sur la distance inter-transects

On a étudié l'agencement spatial des très fortes valeurs ($f(x) \geq 10^4$) indépendamment des autres valeurs. De cette étude, on peut dire que la distance inter-transects a été bien calibrée car elle est du même ordre de grandeur que la portée sur les variogrammes. Ce résultat n'est pas le fait de la maille d'échantillonnage. On connaît, en effet, la structure à petite échelle le long des transects. On est obligé de faire une hypothèse d'isotropie pour toutes les distances inférieures à la distance inter-transects. Sous cette hypothèse, notre résultat tient. En effet, si la distance inter-transects était plus grande, on inférerait la même portée le long des transects et l'hypothèse d'isotropie (seule tenable) nous amènerait aux mêmes modèles spatiaux et à la même conclusion. On peut donc dire que puisque le taux de couverture réalisé par l'échantillonnage est approprié à la structure spatiale des très fortes valeurs, les modèles structuraux de variogramme et covariogramme inférés en B ne dépendent pas de l'échantillonnage. Le variogramme expérimental n'est pas trop éloigné du variogramme régional réel. Ainsi d'autres échantillonnages effectués avec la même distance inter-transects, variations temporelles mis à part, nous auraient permis d'inférer des modèles très similaires. On considère donc que les modèles structuraux inférés en B ont un sens objectif, qu'ils paramètrent la régionalisation réelle et pas uniquement les données spécifiques issus d'un échantillonnage.

On considère que lorsque la distance inter-transects est égale aux paramètres géométriques de la régionalisation à 2D des très fortes valeurs (taille des grains, portée des corrélations entre grains), l'effort d'échantillonnage est suffisant pour que le modèle spatial inféré ne dépende plus des échantillons spécifiques prélevés. On observe alors dans le cadre du modèle objectif, qu'un tel effort d'échantillonnage conduit effectivement à un rapport avantageux de la précision de l'estimation sur l'effort investi.

C.1.4. Granulométrie induite des très fortes valeurs sur les transects

On se propose ici de dégager un ordre de grandeur de la taille des grains $i_4(x) = 1$, à 2D. Contrairement aux sections précédentes, il est nécessaire pour effectuer ce calcul de faire quelques hypothèses peu contrôlables sur la géométrie des grains eux-même. En effet, on ne dispose que de 13 valeurs. D'une part, on considère que les transects recoupent les grains de façon non préférentielle. D'autre part, on considère qu'il n'y a qu'une population de grains, que ceux-ci sont séparés et de taille fixe. Sous ces hypothèses, on cherche à reconstituer la dimension moyenne à 2D des grains $i_4(x) = 1$ à partir de celle des grains induits à 1D sur les transects. Si l'on considère de plus que les grains originaux à 2D sont convexes, des petits disques par exemple, alors la surface, S , du grain à 2D est en relation simple avec l'espérance $E(l)$ de la granulométrie induite à 1D

(MATHERON, 1970, p.44) : $E(l) = \pi \frac{S}{2\bar{z}}$, où $2\bar{z}$ est le périmètre de S . Ainsi, en

prenant pour S un disque, son rayon r est : $r = \frac{2}{\pi} \bar{l}$, où \bar{l} est la longueur moyenne des traversées des grains $i_4(x) = 1$ mesurée le long des lignes. On a : $\bar{l} = 1,5$ m.n., donc $r \approx 1$ m.n. On peut donc dans un premier temps, interpréter la répartition des valeurs $f(x) \geq 10^4$ comme celle de disques séparés de taille fixe (rayon moyen 1m.n.), implantés selon un processus aléatoire où ils sont des fois solitaires, des fois agrégés.

C.2. Application d'un modèle de krigeage disjonctif sans effet de bords

Dans la partie B on a étudié la structure spatiale moyenne de toutes les valeurs. En C.1 on a étudié la structure spatiale des valeurs très fortes indépendamment de toutes les autres. Il convient maintenant d'analyser les rapports structuraux qui les lient à la toile de fond des valeurs plus faibles. En effet, dans cette partie, la variable est codée en indicatrices de chaque classe sur l'histogramme. Chaque classe est étudiée pour elle-même, puis le sont ses rapport deux à deux avec toutes les autres. C'est bien cette dissection structurale des données que modélise le krigeage disjonctif.

L'estimateur de krigeage disjonctif en x_0 inconnu est non linéaire. Dans le krigeage disjonctif ou K.D., plutôt que la variable, on estime une fonction de celle-ci par une somme de fonctions d'une seule variable : $f^*(z_0) = \sum_a f_a(z_a)$ où a est l'indice des échantillons. Pour mettre en œuvre le K.D., on a besoin des lois bivariées (z_0, z_a) et (z_a, z_β) du modèle (MATHERON, 1973, 1978). Plutôt qu'une covariance, l'outil structural est la loi bivariable $(Z(x), Z(x+h))$ pour différents vecteurs h . La régionalisation doit présenter une stationnarité même pour les grandes distances dans le champ. Si l'on peut écrire la fonction $f(z)$ à estimer, comme une somme finie ou convergente de "facteurs" $H_p(z)$ orthogonaux deux à deux, $f(z) = \sum_p H_p(z)$, alors MATHERON (1973, 1978) montre que l'estimation K.D. est possible et s'effectue par

krigeage de chacun des $H_p(z)$. Comme modèle isofactoriel, on utilisera ici le modèle à résidus orthogonaux d'indicatrices (RIVOIRARD, 1988, 1989).

En effet, ce modèle présente deux propriétés géométriques particulièrement adaptées aux données. Le modèle présente une destructuration des hautes teneurs. La portée des variogrammes des indicatrices ($1_{Z(x) \geq z_i}$) successives diminue avec l'ordre croissant des coupures z_i . Plus les teneurs sont fortes, plus les zones d'influence de leurs corrélations spatiales sont petites. La seconde propriété du modèle est son absence d'effet de bord. Ici, quand partant d'une zone pauvre, on chemine vers une zone riche, on ne traverse pas nécessairement de valeurs intermédiaires, contrairement à d'autres modèles comme par exemple des modèles de diffusion. Il y a indépendance entre les valeurs $Z(x) \geq z_i$ et le contour de la surface définie par les valeurs $Z(x) \geq z_i$. Cette seconde propriété du modèle semble particulièrement apte à rendre compte des transitions spatiales très rapides dans les ordres de grandeur des densités de hareng.

Dans un premier temps, on mesure l'adéquation aux données du modèle à résidus orthogonaux d'indicatrices, par le test des deux propriétés géométriques. Puis on utilise le modèle à deux fins. L'ajustement du modèle nous donne une interprétation des rapports structuraux entre les très fortes valeurs et les plus faibles. On cartographie alors la surface sur laquelle les très fortes valeurs ont une probabilité d'apparition élevée. Il s'agit bien d'une estimation non linéaire car on devra estimer une indicatrice. D'autre part, le modèle inféré est compatible avec une implantation des grains à très fortes densités sur cette surface, selon un schéma booléen. On utilise alors ce processus d'implantation pour établir un intervalle de confiance sur le nombre estimé des grains à très fortes densités, donc sur l'estimation globale. On ajuste le modèle le long des transects et l'on fera jouer l'hypothèse d'isotropie pour toutes les coupures.

C.2.1. Propriétés géométriques

Soit un couple de coupures (z, z') de la fonction aléatoire $Z(x)$ dont z' est la plus haute ($z \leq z'$). Le modèle à résidus orthogonaux d'indicatrices est construit sur les régressions linéaires de l'indicatrice haute sur l'indicatrice basse, pour tous les couples (z, z') de la fonction $Z(x)$. On reprend les notations de RIVOIRARD (1988).

La loi marginale s'écrit : $T_z = \text{Prob} (Z(x) \geq z) = E(1_{Z(x) \geq z})$.

La loi bivariable s'écrit : $T_{zz'}^h = \text{Prob} (Z(x) \geq z, Z(x+h) \geq z')$.

Le résidu de la régression linéaire s'écrit : $R(x) = 1_{Z(x) \geq z'} - \frac{T_{z'}}{T_z} 1_{Z(x) \geq z}$

Dans le modèle, $R(x)$ est sans corrélation avec $1_{Z(x) \geq z}$, d'où la relation

fondamentale : $T_{zz'}^h = \frac{T_{z'}}{T_z} T_{zz}^h$.

La loi bivariable est déterminée par la loi du minimum entre les deux variables et par la loi marginale.

Dans le modèle, la covariance non centrée, T_{zz}^h , se déstructure quand z augmente. Cette déstructuration s'observe expérimentalement sur les covariances ou de façon équivalente sur les variogrammes des indicatrices $1_{Z(x) \geq z}$ quand augmentent les coupures z . Les variogrammes de différentes indicatrices calculés le long des transects sont donnés sur la Figure 20. On observe une déstructuration importante, particulièrement entre les indicatrices $1_{Z(x) \geq 10^3}$ et $1_{Z(x) \geq 10^4}$ où la portée chute de 15 m.n à 5 m.n environ.

RIVOIRARD (1988) révèle la signification géométrique de la relation fondamentale (I.3) du modèle : "Conditionnellement à $Z(x) \geq z$, on a indépendance entre $Z(x) \geq z'$ et $Z(x+h) \geq z$, quels que soient $z'(\geq z)$ et h ". Il y a indépendance entre les valeurs supérieures à z et le contour de la surface A_z définie $Z(x) \geq z$ (pas d'effet de bord dans A_z). De la relation fondamentale, un test expérimental peut être tiré.

En effet, on écrit : $\sigma_{zz'}(h) = \frac{T_{z'}}{T_z} \sigma_z(h)$ où $\sigma_{zz'}(h)$ et $\sigma_z(h)$ sont respectivement les

covariances croisée et simple des indicatrices pour les coupures z' et z ($z \leq z'$). Le rapport du variogramme croisé des deux indicatrices sur le variogramme simple de l'indicatrice basse doit être constant. On a donc calculé un tel rapport de variogrammes pour chacun des six couples d'indicatrices que l'on peut former avec les coupures précédentes prises sur l'échelle \log_{10} (10, 10^2 , 10^3 , 10^4). Ils sont donnés sur la Figure 21. Un test statistique n'est pas envisageable ; on se contente de visualiser qualitativement le comportement des courbes. Toutes les courbes présentent un aspect similaire. Le rapport de variogramme croît très rapidement aux petites distances, puis manifeste une tendance à se stabiliser. Il existe donc un léger effet de bord sur les surfaces A_z ($z \geq 10$) dont l'extension spatiale est de l'ordre de 1 à 3 m.n. On considère donc que la taille moyenne des zones de transition d'une puissance de 10 à une autre est très petite, très abrupte. Le modèle à résidus d'indicatrices, dans lequel il n'existe pas de transition, est donc bien adapté aux données. On va désormais l'ajuster.

C.2.2. Le modèle et son ajustement : dissection des données

On discrétise la fonction $Z(x)$ en $n+1$ classes par n coupures z_1, \dots, z_n : $C_0 = [0, z_1[$, $C_1 = [z_1, z_2[$, ..., $C_n = [z_n, \infty [$. Soient f_0, \dots, f_n les valeurs prises par une fonction de $Z(x)$, $f(Z(x))$, sur les $n+1$ classes. On a : $f(Z(x)) = \sum_{j=0}^n f_j 1_{Z(x) \in C_j}$.

Les résidus des régressions linéaires entre indicatrices successives sont les facteurs du modèle. Les résidus $R_j(x)$ sont en fait définis ainsi : $R_j(x) = \frac{1_{Z(x) \geq z_j}}{T_j} - \frac{1_{Z(x) \geq z_{j-1}}}{T_{j-1}}$ où z_{j-1} et z_j sont deux coupures successives. Ils sont de moyenne nulle et on leur rajoute le résidu $R_0(x) = 1$. RIVOIRARD (1988) montre que les résidus (R_0, \dots, R_n) forment une famille de fonctions non corrélées deux à deux et qui engendrent les indicatrices. On a : $\text{cov}(R_i(x), R_j(x+h)) = 0$; $1_{Z(x) \geq z_j} = T_j \sum_{i=0}^j R_i(x)$.

Le cokrigeage d'un résidu égale donc son krigeage. Pour estimer chaque indicatrice par K.D., il suffit d'estimer par krigeage chacun des résidus $R_j(x)$ indépendamment des autres. Toute fonction $f(Z(x))$ admet donc un développement à partir des résidus : $f(Z(x)) = \sum_{j=0}^n f'_j R_j(x)$ et son estimation par K.D. s'en déduit. On spécifie maintenant le modèle spatial qui sert à effectuer le krigeage.

Pour effectuer le K.D., il suffit de modéliser le variogramme de chacun des résidus R_1, \dots, R_n . En effet, en vertu de la relation fondamentale du modèle, la loi bivariable est déterminée par la loi marginale et la covariance non centrée T_{jj}^h qui, elle, s'exprime grâce aux covariances des résidus selon la relation suivante (RIVOIRARD,

1988) : $T_{jj}^h = T_j^2 \sum_{i=0}^j \sigma_{R_i}(h)$. Pour assurer la cohérence du modèle, deux conditions pèsent sur les covariances $\sigma_{R_i}(h)$ des résidus. Premièrement, le modèle de covariance de résidu $\sigma_{R_i}(h)$ doit être autorisé. Le modèle sphérique convient. L'effet de pépité est permis. Si le résidu R_j est pépitique, il indique que, sachant que $Z(x_\alpha)$ dépasse ou non z_{j-1} aux points x_α , on n'est pas en mesure de prévoir si $Z(x_0)$ dépasse z_j en x_0 inconnu. Deuxièmement, les modèles $\sigma_{R_i}(h)$ doivent être cohérents entre eux afin de permettre une destructuration des T_{jj}^h quand la coupure j augmente. Cette seconde contrainte se traduit par l'inégalité

suivante : $\sigma_{R_j}(h) \leq \left(\frac{T_{j-1}^2}{T_j^2} - 1 \right) \left(1 + \sum_{i=1}^{j-1} \sigma_{R_i}(h) \right)$. On voit donc que la destructuration

des résidus élevés n'est pas nécessaire pour assurer la destructuration des teneurs élevées.

On a considéré tout le jeu des données. On a pris 10 coupures (0, 10, 50, 100, 200, 500, 1000, 2000, 5000, 10000). On considère donc une partition de $Z(x)$ en 11 classes. Il y a 10 résidus $R_0(x)$ à $R_9(x)$ qui sont explicités dans le tableau suivant.

R_j	R1	R2	R3	R4	R5	R6	R7	R8	R9
Z_j	10.	50.	100.	200.	500.	1000.	2000.	5000.	10^4
Z_{j-1}	0.	10.	50.	100.	200.	500.	1000.	2000.	5000.

Table des résidus utilisés. R_j est le résidu de la régression de l'indicatrice de $Z(x) \geq Z_j$ sur l'indicatrice de $Z(x) \geq Z_{j-1}$. (Z_{j-1} et Z_j sont 2 coupures successives)

Les variogrammes de $R_1(x)$ à $R_9(x)$ sont calculés le long des transects. Chaque modèle est un sphérique ajouté d'un effet de pépité (Figures 22 et 23). Les paramètres des 9 modèles sont consignés dans le tableau 2 suivant.

résidus : R_j	R1	R2	R3	R4	R5	R6	R7	R8	R9
pépité : C_0	0.20	0.50	0.28	0.60	1.50	2.50	4.00	20.00	70.00
palier : C	1.30	0.12	0.14	0.10	0.70	0.70	1.00	5.00	30.00
portée (m.n.)	35.	7.	8.	2.	5.	2.	2.	2.	5.
rapport : $C_0 / (C_0 + C)$	13.3%	81.3%	67.5%	85.7%	68.2%	78.1%	80.0%	80.0%	70.0%

Paramètres des variogrammes sphériques utilisés pour modéliser les variogrammes des résidus d'indicatrices.

De $R_1(x)$ à $R_5(x)$, les variogrammes présentent une destructuration progressive. Pour les résidus d'ordre supérieur, de $R_6(x)$ à $R_9(x)$, les variogramme ne présentent plus de destructuration mais une stabilisation de leurs paramètres : la portée oscille entre 2 et 5 m.n, la pépité se stabilise autour de 80 % de la variance. On n'a pas vérifié explicitement

l'inégalité $\sigma_{R_j}(h) \leq \left(\frac{T_{j-1}^2}{T_j^2} - 1 \right) \left(1 + \sum_{i=1}^{j-1} \sigma_{R_i}(h) \right)$, assurant la cohérence entre les

modèles. Toutefois, la chute de T_j quand j augmente, doit suffire à l'assurer même si les

modèles ne diffèrent plus trop en portée et pépite au-delà de $R_6(x)$. La loi marginale T_j est donnée par le tableau 3 suivant.

Z_j	10.	100.	500.	1000.	5000.	10^4
T_j	0.48	0.37	0.18	0.12	0.03	0.01

Table de la loi marginale T_j : évolution du pourcentage de valeurs supérieure à Z_j quand Z_j augmente. On a considéré tout le jeu de données.

Du point de vue de la description structurale quantitative des données, on peut considérer que les résidus R_6 à R_9 sont pépitiques. Ceci s'interprète physiquement. Pour ce faire, on revient à l'expression de l'estimateur K.D. d'une indicatrice dans le modèle :

$$1_{Z(x) \geq Z_j}^{K.D.} = T_j \sum_{i=0}^j R_i^K(x) \text{ où les exposants K.D. et K signifient estimé par krigeage}$$

disjonctif et par krigeage. Lorsque à partir d'une coupure z_p tous les résidus R_i sont pépitiques pour $p \leq i \leq j$, leur krigeage en x inconnu vaut leur moyenne qui est nulle puisque ce sont tous des résidus. Les résidus pépitiques n'apportent donc aucune contribution à la somme ; les indicatrices $1_{Z(x) \geq z_i}$ ($p \leq i \leq j$) n'apportent aucune information pour l'estimation de l'indicatrice $1_{Z(x) \geq z_j}$. Ainsi dans notre cas, sur l'aire A_{500} définie par $Z(x) \geq 500$, on ne sait pas prévoir quelle valeur supérieure à 500 $Z(x)$ peut dépasser en x inconnu. Ce peut être 600, 1000, 10000 ou plus. En revanche, l'aire A_{500} , c'est-à-dire l'indicatrice $1_{Z(x) \geq 500}$, peut être estimée en x sachant que aux points x_α échantillonnés les valeurs $z(x_\alpha)$ dépassent ou non les coupures inférieures à 500. En effet, les résidus R_1 à R_5 sont structurés.

Les valeurs supérieures à 500 représentent 90 % de la biomasse : l'aire A_{500} correspond au "gisement" du stock de hareng. L'inférence du modèle sur les données nous indique ce qu'il est envisageable d'estimer à partir des données. On peut estimer les contours du gisement car les résidus d'ordre inférieur sont structurés. Or on ne sait plus prévoir les teneurs à l'intérieur du gisement car les résidus d'ordre supérieur sont très déstructurés. A l'intérieur du gisement A_{500} , toute valeur supérieure à 500 peut apparaître n'importe où, en particulier de très fortes valeurs. Si maintenant, on considère le modèle tel qu'inféré comme reflétant non pas les données mais la régionalisation réelle, alors il permet de nous faire envisager l'échantillonnage. Tant que les résidus sont structurés, les données déjà acquises nous donnent des indices sur l'entrée dans une zone supportant potentiellement de très fortes valeurs. Mais une fois entrés dedans, les résidus d'ordre supérieur étant pépitiques, les valeurs rencontrées ne nous apportent aucun indice ni sur la valeur ni sur la localisation des très fortes valeurs ; on les rencontre en aveugle. Si dans le gisement la fréquence d'apparition des très fortes valeurs est faible, comme elles

concentrent une large part de la biomasse, on court le risque de prélever un jeu de données très écarté de la réalité. Le modèle donne un argument pour élaborer une stratégie d'échantillonnage adaptative : déceler les contours des zones riches comme A_{500} et les sur-échantillonner spécifiquement afin de mieux en estimer la moyenne à l'intérieur.

JAKOBSSON (1983) argumente pour une telle stratégie d'échantillonnage sur le stock de hareng islandais qui pond d'été ("summer spawning herring"). Ce stock fait partie de l'ensemble Atlanto-scandinave. Il s'isole de novembre à janvier sur une zone d'hivernage au Sud-Sud-Ouest de l'Islande, dont l'aire est environ 100 m.n^2 . Il est là dans une fenêtre spatio-temporelle favorable à son estimation totale par écho-intégration. Sur sa zone d'hivernage, le poisson se regroupe et forme deux à trois agrégations denses d'envergure moyenne $15 \text{ m.n} \times 5 \text{ m.n}$ ("herring schools"). Le gisement est donc fractionné en deux ou trois agrégations, petites vis-à-vis de l'aire à prospecter. Une reconnaissance aveugle par un plan systématique sur toute la zone d'hivernage est critiquée par JAKOBSSON (1983) : il faut connaître la biologie spatiale du poisson et trouver coûte que coûte ces taches de présence sur zone. Ce n'est que sur ces dernières que commence alors ce que l'on peut appeler l'échantillonnage du stock de hareng sur son gisement. Un plan systématique ne permet que d'estimer les contours des deux ou trois agrégations de présence dense, mais non d'estimer une moyenne globale suffisamment précise à moins de devoir développer un effort d'échantillonnage très important.

Grâce au modèle inféré, on va maintenant cartographier l'aire A_{500} du gisement puis donner un intervalle de confiance sur la fréquence des fortes valeurs rencontrées dedans par notre échantillonnage régulier.

C.2.3. Cartographie de l'aire du gisement

On veut estimer l'indicatrice $1_{Z(x) \geq 500}$ aux nœuds d'une grille à maille régulière. Son krigeage disjonctif s'effectue par krigeage successif de chacun des résidus $R_1(x)$ à

$R_5(x)$, selon la formule déjà donnée : $1_{Z(x) \geq 500}^{K.D} = T_{500} \left(1 + \sum_{i=1}^5 R_i^K(x) \right)$.

Les krigeages ont lieu en voisinage glissant, en utilisant 12 points de voisinage. L'implantation des échantillons le long des transects n'est pas propice à une bonne répartition des points informants dans les voisinages. C'est pourquoi, on a pris un petit nombre de données dans les voisinages pour effectuer le krigeage afin que des informations trop éloignées le long des transects ne participent pas à l'estimation. Les krigeages sont effectués en imposant sur les poids la condition d'universalité $\sum_a \lambda_a = 1$. RIVOIRARD (1988) montre que, dans ce modèle, la condition d'universalité permet de respecter la loi locale des biomasses.

Les valeurs krigées de l'indicatrice varient entre 0 (un peu moins) et 1 (un peu plus). Elles mesurent une probabilité locale de dépasser le seuil :

$1_{Z(x) \geq z_i}^{K.D.} = \text{Prob}(Z(x) \geq Z_i) = P_i(x)$. Le contour de l'aire A_i définie par $Z(x) \geq z_i$ dépend d'un seuil choisi sur $P_i(x)$. On a choisi la valeur P_i de façon à respecter globalement la fréquence expérimentale T_i (dans le jeu complet des données), qui est une mesure de la surface A_i . On a : $T_{500} = 0,18$ et $P_{500} = 0,30$.

L'aire A_{500} estimée par krigeage disjonctif est matérialisée par des hachures sur la Figure 24. Le gisement A_{500} occupe en deux ensembles une grande partie de la zone prospectée. Le plan d'échantillonnage l'aura donc bien investiguée. Notre cas n'est pas celui de JAKOBSSON (1983) où le gisement était fractionné en trois taches petites relativement au champ à prospector. Il apparaît donc que le choix d'une stratégie d'échantillonnage est fonction de la taille des zones supportant les paquets de valeurs très fortes, relativement au champ à prospector. On mesure maintenant les fluctuations d'échantillonnage de la fréquence des très fortes valeurs dans le cadre du modèle.

C.3. Intervalle de confiance pour le nombre de grains à teneur très élevée

On reprend l'étude de la queue de l'histogramme (valeurs supérieures à 10^4). Le rapport de ces très fortes valeurs aux autres valeurs a été spécifié en C.2.2. : les très fortes valeurs ne se développent que dans l'aire A_{500} et s'y développent dedans indépendamment du contour de A_{500} et sans corrélation avec les autres valeurs de A_{500} (car tous les résidus d'indicatrices supérieures à 500 sont pépitiqes). On se propose maintenant de mesurer les fluctuations expérimentales de la fréquence observée des très fortes valeurs, afin d'en mesurer les conséquences sur l'estimation de la biomasse moyenne globale.

L'estimateur moyenne arithmétique peut s'écrire sous une forme telle qu'on y fait apparaître la fréquence de chaque classe sur l'histogramme :

$$m^* = \frac{1}{n} \sum_i z(x_i) = \frac{1}{n} \sum_i n_i m_i = \sum_i f_i m_i$$
 où m_i est la moyenne des valeurs dans la classe i et f_i la fréquence de cette classe. Dans le plan systématique, le taux d'échantillonnage est le même dans les zones pauvres que dans les zones riches. Comme les zones riches sont plus hétérogènes, les fluctuations d'échantillonnage des f_i et des m_i de la queue de l'histogramme sont plus importantes que celles des autres classes. On ne

s'intéresse qu'à la dernière classe que l'on individualise : $m^* = \sum_{i=1}^{n-1} f_i m_i + f_n m_n$. Les

fluctuations expérimentales de m^* résident essentiellement dans la rencontre ou non de paquets de valeurs très fortes à l'échantillonnage. On se propose de mesurer l'effet sur m^* des fluctuations d'échantillonnage de f_n quand m_n est considérée connue bien que, en pratique, quand les fluctuations d'échantillonnage sur f_n sont importantes, celles sur m_n le sont aussi. On veut dégager un ordre de grandeur des fluctuations en construisant un intervalle de confiance. Pour ce faire, on considère la dernière classe de l'histogramme : $C_{11} = [10^4, \infty[$. Sa moyenne est $m_{11} = 26512$. La fréquence observée des valeurs supérieures à 10^4 dans le champ V de présence du hareng est : $f_{11} = 0,030$. Pour calculer un

intervalle de confiance, il faut disposer d'un modèle d'implantation des valeurs $Z(x) \geq 10^4$ sur l'aire A_{500} . Sur la Figure 25 on a représenté les valeurs $i_4(x) = 1$ dans l'aire A_{500} . Les limites sont définies sur la figure par coupure sur les échantillons.

C.3.1. Schéma de Poisson pour l'implantation des très fortes valeurs

On interprète l'implantation des valeurs $Z(x) \geq 10^4$ dans A_{500} par un schéma booléen car ce modèle est compatible avec l'absence d'effet de bord constaté dans A_{500} . Dans un schéma booléen (MATHERON, 1967), on implante sur une surface A des germes selon un schéma de Poisson et on fait se développer en chacun des germes un grain primaire aléatoire de covariogramme géométrique $K(h)$. Les grains du schéma booléen sont les réunions des grains primaires car certains peuvent se recouper. On considère ici un schéma simplifié : les grains primaires des valeurs $Z(x) \geq 10^4$ sont tous identiques, ce sont tous des disques de rayon 1 m.n. ; ils sont tous remplis d'une même densité constante de hareng $m_{11} = 26512$; ils ne se recoupent pas. Les germes des grains primaires ($Z(x) \geq 10^4$) sont implantés dans l'aire A_{500} selon un schéma de Poisson de densité θ .

L'interprétation des très fortes valeurs par un tel schéma n'est toutefois pas vraiment satisfaisante. En effet, l'implantation poissonnienne des germes n'est pas tout à fait compatible avec l'observation. L'implantation poissonnienne ne tient pas compte de la corrélation entre les valeurs $i_4(x) = 1$ de portée 6 m.n.. D'autre part, la taille fixe des grains du schéma booléen de même que leur non recouvrement sont des simplifications importantes. Leur validité pour représenter l'implantation des très fortes valeurs n'est pas vraiment contrôlable. C'est le petit nombre de grains qui nous y oblige. Le modèle n'est toutefois pas dénué de sens. Sa principale qualité est de nous permettre de dégager facilement un ordre de grandeur des fluctuations de l'erreur d'estimation du nombre de grains puis d'en mesurer les répercussions sur l'estimation de la densité moyenne du poisson.

Disposant donc d'un modèle pour l'implantation des grains $i_4(x) = 1$, on déduit la loi de l'erreur d'estimation ($N - N^*$) du nombre de ces grains. N est le nombre de grains et N^* son estimation. Comme les grains sont tous de même rayon 1 m.n., dans les bandes de largeur 1 m.n. centrées sur les transects, tous les grains auront été comptés sans omission. Les grains ne sont présents que dans A_{500} . On peut donc séparer A_{500} en deux compartiments. L'un est la surface effectivement explorée et connue dans laquelle tous les grains ont été comptés. On nomme cette surface, S_c , et N_c le nombre de grains qu'elle contient. L'autre compartiment est la surface non explorée et non connue, S_{nc} , sur laquelle on estime le nombre, N_{nc} , de grains. N_{nc} est estimé par N_{nc}^* . On a :

$$N = N_c + N_{nc}$$

Dans le modèle, les grains sont implantés dans A_{500} selon un schéma de Poisson de densité θ . L'estimateur N_{nc}^* suit donc une loi de Poisson de paramètre θS_{nc} . On a :

$$N^* = N_c + N_{nc}^* \quad \text{avec } N_{nc}^* : P(\theta S_{nc}) \quad \text{et } N_c = \theta S_c .$$

Tout naturellement, l'erreur d'estimation sur N est égale à celle faite sur N_{nc} et elle dépend de la surface S_{nc} , c'est à dire de la taille des grains et de la distance inter-transects. La variance d'estimation s'écrit : $\text{Var}(N - N^*) = \text{Var}(N_{nc} - N_{nc}^*) = \theta S_{nc}$. Plus les grains sont gros, plus le compartiment S_{nc} est petit, à distance inter-transects fixée. Si le rayon des grains circulaires est égale à la distance inter-transects, la variance d'estimation est logiquement nulle puisque tous les grains de A_{500} seront dénombrés à chaque échantillonnage.

Le modèle nous permet de mettre en relief la différence entre la variance de l'estimation du nombre de grains (variance de l'erreur) et la variance d'échantillonnage de l'estimateur du nombre de grains. En I.B.6, la taille du champ relativement à la portée était un paramètre crucial pour la variance d'estimation. Ici, c'est la taille du compartiment S_{nc} . La variance de l'estimateur représenterait ici la variance de l'estimation dans le cas où le rayon des grains serait si petit que S_c devenant alors négligeable, S_{nc} correspondrait à toute l'aire A_{500} .

Toutefois, variance d'échantillonnage de N_c et variance de l'erreur ($N_{nc} - N_{nc}^*$) sont intimement liées car la densité θ n'est pas connue. En effet, N_c est une variable aléatoire relativement aux différents échantillonnages: c'est une variable de poisson de paramètre θS_c . Chaque échantillonnage w dans un compartiment $S_c(w)$ fournit un estimateur θ_w^* de la densité θ . Ensuite, θ_w^* est utilisée pour construire la loi de N_{nc}^* . On va considérer que θ est estimée sans erreur à partir d'un échantillonnage w . On construit un intervalle de probabilité pour N_{nc}^* autour de N_{nc} , uniquement à partir de la loi de Poisson $P(\theta_w^* S_{nc})$. Cet intervalle sous-estimera les fluctuations de l'erreur d'estimation de N .

Soient D la distance inter-transects, r le rayon des grains et L la longueur totale des portions de transects intérieures à l'aire A_{500} .

Dans S_c , on a : $N_c(w) = 13$; $S_c = r L$; $\theta_w^* = \frac{N_c(w)}{r L}$

Dans S_{nc} , on a : $S_{nc} = (D - r) L$; $N_{nc}^*(w) : P(\theta_w^* S_{nc})$

Comme $r = 1$ m.n. , $L = 170$ m.n. , $D = 4.54$ m.n. ,

on trouve : $\theta_w^* S_{nc} = 46$ et $N^* = 59$.

Un intervalle de probabilité à 95 % pour les fluctuations de N_{nc}^* autour de N_{nc}

s'en déduit facilement. On a : $\text{Prob}(X \leq k) = e^{-46} \sum_{i=1}^k \frac{46^i}{i!}$. On calcule cette somme pour différentes valeurs de k et l'on garde les valeurs k_1 et k_2 qui respectivement donnent les valeurs $P(X \leq k_1)$ et $P(X \leq k_2)$ les plus proches de 0,025 et 0,975. On trouve : $P(X \leq 33) = 0,028$ et $P(X \leq 60) = 0,98$. Le nombre de grains dans S_{nc} est donc 46 compris entre 33 et 60. Les répercussions de cette fluctuation expérimentale sur la moyenne globale s'en déduisent.

C.3.2. Résultat :

On a :

$$f_{11}^* = \frac{N_c + N_{nc}^*}{V} = 0.030 \text{ comprise entre } 0.023 \text{ et } 0.037 \text{ (} V = n \text{ D est le champ prospecté) :}$$

$$m^* = \sum_{i=1}^{10} f_i m_i + f_{11} m_{11} = 895 + f_{11} \times 26512 .$$

On ne fait varier que f_{11} . L'intervalle de variation qui se déduit de l'intervalle de probabilité sur le nombre de grains est : $m^* = m = 1690$ compris entre $m_i^* = 1505$ et $m_s^* = 1876$. Les variations résultantes sur m^* sont donc de 11 % à gauche comme à droite. Elles sont de l'ordre de la moitié de celles données par la fourchette d'erreur gaussienne à 95% calculée à partir de la variance d'estimation de la partie B : $\pm 2 \frac{\sigma_{cs}}{Z_V^*} = 25\%$.

Notre procédé de calcul sous-estime la variance d'estimation du nombre de grains, $\text{Var}(N - N^*)$. Ainsi, la moitié ou plus des erreurs d'estimation sur la moyenne réside dans l'erreur d'estimation du nombre de grains à très forte densité. C'est bien la rencontre des paquets de valeurs extrêmes et l'erreur d'estimation dans leur dénombrement qui engendrent une grande partie de l'erreur d'estimation sur la biomasse globale moyenne.

Sur le variogramme moyen de la variable à 2D, la pépite est importante et elle contribue à 79 % de la variance d'estimation de la moyenne. Il était important d'en comprendre la genèse : les grains à très forte densité sont implantés dans A_{500} sans corrélation apparente avec les valeurs supérieures à 500. Cette implantation génère un effet de pépite car les transitions spatiales dans les ordres de grandeurs n'existent pas. La variance d'estimation dépend de l'adéquation dans A_{500} de la distance inter-transects à la structure spatiale des grains à très forte densité.

C.4. Stratégie d'échantillonnage : discussion

GERLOTTO et al (1983) utilisent un modèle très proche du modèle à résidus d'indicatrices orthogonaux pour effectuer des simulations. Il semble donc que la description des données pélagiques de hareng, telle qu'effectuée grâce au modèle précédent, contient un aspect plus général, et nous voulons l'utiliser pour classifier les types d'échantillonnages pélagiques qu'il est intéressant de mettre en œuvre. Les histogrammes des données pélagiques sont en général très dissymétriques avec une queue très longue. Les transitions spatiales dans les données (effets de bords) sont en général très petites. On considère que le modèle précédent recèle une valeur génétique assez générale et qu'il existera toujours une coupure z_p au-delà de laquelle les résidus d'indicatrice seront pépitiés. Dans les aires A_p définies par $Z(x) \geq z_p$, les très fortes valeurs, dont la rencontre détermine l'estimation de biomasse, nous apparaissent implantées au hasard peut-être de façon structurée mais sans relation avec les autres valeurs de A_p . La stratégie d'échantillonnage puis l'effort à fournir nous semblent logiquement liés à l'agencement géométrique des zones A_p dans le champ V (de présence du poisson) qu'il faut prospecter. On pense à classer en trois groupes les rapports possibles de A_p à V , selon une diminution progressive de la taille des zones A_p dans V . Le premier groupe correspond à l'exemple traité du hareng norvégien. L'aire totale A_p occupe une bonne partie de V ($\frac{A_{500}}{V} = 40\%$).

A_p est une zone unique ou est morcelée en un petit nombre de surfaces d'extension importante. Un échantillonnage à maille régulière est adapté. L'effort est calé sur les paramètres de la structure spatiale des grains très denses dans les zones A_p . Le second groupe correspond au cas du hareng islandais. La zone A_p dans V représente peut-être une surface aussi grande mais elle est morcelée en un petit nombre de surfaces plus petites relativement à V . Un réseau régulier adapté à la structure spatiale des fortes valeurs dans A_p peut conduire à développer un effort d'échantillonnage trop important car la surface à couvrir dans les zones pauvres est grande. Il peut être alors nécessaire de modifier la stratégie d'échantillonnage et orienter alors spécifiquement l'effort sur les zones A_p qui devront chacune être recherchées et prospectées. Le troisième cas de figure serait celui où l'aire A_p est très morcelée dans V en une multitude de petites aires. C'est plutôt le cas de figure qui se rencontre expérimentalement en milieu tempéré et tropical. Un réseau régulier redevient alors une stratégie adéquate : on n'investiguera pas toutes les petites zones A_p mais une proportion de celles-ci. L'effort d'échantillonnage sera calé sur la structure spatiale de l'indicatrice $1_{Z(x) \geq z_p}$ et non plus sur celle des très fortes valeurs. Les fluctuations expérimentales seront vraisemblablement importantes, et si le jeu de données ne s'écarte pas trop de la réalité, l'erreur d'estimation sera élevée. Les cas 1 et 2 apparaissent plus propices.

On a essayé de rechercher la prospection de l'espace qui permette d'estimer avec une bonne précision la queue de l'histogramme. Le tirage au sort de la position des

transects parallèles (plan aléatoire simple) sur la zone totale à prospector ou le tirage au sort (aléatoire stratifié) dans des strates d'espace comme proposés par JOLLY et al (1987), ne semblent pas permettre une prospection aussi efficace de l'espace que la maille régulière.

D. EFFICACITÉ DES ESTIMATEURS ET FOURCHETTES D'ERREUR

L'estimation effectuée précédemment est-elle optimale ? Jusqu'à présent, on a utilisé un estimateur uniquement spatial. Il était arithmétique. On attribuait aux échantillons une zone d'influence conformément au découpage de l'espace défini par le plan d'échantillonnage. Cette somme discrète estime l'intégrale d'espace $m = \frac{1}{V} \int_V Z(x) dx$. Un échantillonnage w conduit à une estimation a_w^* et à un modèle structural $\gamma_w(h)$. Appelons γ_R le variogramme régional de la variable régionalisée : c'est le modèle que l'on inférerait si tous les points étaient connus. On a vu que, comme la loi de distribution statistique de la variable régionalisée (fréquence d'occurrence des valeurs dans l'espace) est dissymétrique, l'estimation arithmétique peut être très imprécise ($E(a_w^* - m)^2$ élevée) sans que l'on puisse mesurer cette précision réelle car le modèle structural peut ne refléter que l'échantillonnage effectué ($E(\gamma_w - \gamma_R)^2$ élevée). Ce problème nous a amené à calibrer l'effort d'échantillonnage en fonction de ce qui est connu de la structure spatiale réelle γ_R afin de réduire les fluctuations d'échantillonnage $E(\gamma_w - \gamma_R)^2$ du modèle inféré.

Pour un effort d'échantillonnage modéré, doit-on préférer un estimateur spatial à un estimateur de moyenne de la loi ? L'estimateur spatial est arithmétique : on estime la moyenne sur un champ défini sans considérer la loi statistique de la variable régionalisée. L'estimateur de moyenne de loi estime la moyenne de la loi dissymétrique de la variable sans considérer le champ spatial. On l'appelle pour simplifier estimateur distributionnel. On considère un estimateur distributionnel non linéaire. Si tous les points de l'espace étaient connus, on verrait que les deux estimateurs ne coïncident pas. L'estimateur distributionnel non linéaire est donc biaisé. Toutefois, lorsque l'effort d'échantillonnage est modéré, il est intéressant dans la mesure où il est moins sensible que l'estimateur arithmétique aux fluctuations d'échantillonnage des fortes valeurs. On suppose que la loi réelle est du même type que celle des échantillons mais de paramètres différents.

La statistique classique formule comme suit le problème de l'estimation. Un échantillonnage spatial doit fournir n observations indépendantes de la même loi réelle. A partir des échantillons, on cherche à estimer les paramètres de la loi réelle dont sont issus les données. Dans ce problème, la méthode du maximum de vraisemblance (dans les limites de son application) permet de développer un estimateur optimal répondant à la question. Il faut faire choix d'une loi théorique de distribution pour la variable régionalisée. On traitera de la loi lognormale qui est un modèle heuristique.

Dans le problème du choix de l'estimateur, l'effort d'échantillonnage joue un rôle central. Chaque échantillonnage conduit à calculer une valeur de l'estimateur de la moyenne. La loi statistique de celui-ci, que l'on pourrait construire en répétant plusieurs échantillonnages, dépend entre autres de l'effort d'échantillonnage. Comme la loi de la variable régionalisée est dissymétrique et dispersée, celle de l'estimateur le sera aussi pour des efforts faibles. Les performances entre estimateurs différents peuvent être

grandes ici. Mais l'effort augmentant, la loi de tout estimateur tendra à se symétriser avec une dispersion qui tendra à se réduire. Ainsi, l'estimateur arithmétique redevient intéressant quand l'effort d'échantillonnage est bien calibré. Ce que propose la Géostatistique est un estimateur spatial arithmétique car la moyenne spatiale vraie est une grandeur arithmétique. La variance d'estimation se calcule grâce à un modèle structural car les valeurs des échantillons sont spatialement corrélées. Ce que propose la statistique classique est un estimateur non linéaire de la moyenne car la loi de la variable régionalisée est dissymétrique. Dans le cas d'un échantillonnage à maille régulière, sa variance d'estimation n'est pas toujours réaliste.

Dans cette partie, on rappelle les principaux résultats de la théorie lognormale de l'estimation quand les échantillons sont indépendants et quand ils ne le sont pas. On justifie alors l'emploi de la moyenne arithmétique sur les données d'acoustique, et on donne les fourchettes d'erreur conventionnelle que l'on peut construire autour.

D.1. Le formalisme lognormal

La loi des valeurs expérimentales $Z(x)$ est dissymétrique. On cherche un type de modèle théorique de loi qui soit compatible avec les données et qui reflète la loi réelle. C'est la moyenne de cette loi théorique que l'on cherche à estimer. Les estimations directes de la moyenne et de la variance par les estimateurs habituels sont très imprécises. En effet, elles sont sous l'influence de la queue de l'histogramme dont les fluctuations d'échantillonnage sont élevées. On cherche à réduire leur influence en changeant d'échelle de mesure pour les valeurs $Z(x)$ afin de se ramener par une transformation adéquate $\varphi(Z(x))$ à une loi gaussienne pour laquelle les estimations de la moyenne et de la variance sont efficaces, par les estimateurs habituels. Dans le formalisme lognormal, les passages des moments de $Z(x)$ à ceux de $\varphi(Z(x))$ sont spécifiés. De plus, le modèle lognormal a une justification physique. En effet, la référence à la loi lognormale est très fréquente en biologie comme dans d'autres sciences car elle est liée à l'effet proportionnel. C'est la variation relative $\frac{dZ}{Z}$ du phénomène Z étudié et non la variation élémentaire dZ , qui est indépendante de la valeur de Z . L'addition d'un grand nombre d'effets proportionnels à peu près indépendants en chaque point de mesure conduit à la variable $\int \frac{dZ}{Z} = \text{Log } Z$ qui se doit d'être approximativement gaussienne en vertu du théorème central limite. Donc Z a une distribution statistique d'un type lognormal.

Soit $Z(x)$ une loi lognormale : $\text{Log } Z(x)$ est gaussienne. Soient γ et σ^2 les moyennes et variances arithmétiques de $\text{Log } Z(x)$. γ et σ^2 sont respectivement la moyenne géométrique (ou médiane) et la variance logarithmique de $Z(x)$. Les moments de $Z(x)$ sont donnés par : $m_n = \gamma^n e^{\frac{n^2 \sigma^2}{2}}$.

L'espérance m de $Z(x)$ est : $m = \gamma e^{\frac{\sigma^2}{2}}$. Ainsi, deux gaussiennes ayant même moyenne $\text{Log } \gamma$ mais deux variances σ^2 différentes engendrent deux lois lognormales aux médianes identiques mais aux moyennes et dispersions très différentes. La variance arithmétique \sum^2 de $Z(x)$ est : $\sum^2 = m^2 (e^{\sigma^2} - 1)$. Quand σ^2 n'est pas trop grand, on peut ne retenir que les deux premiers termes du développement de e^{σ^2} . D'où on a : $\sigma^2 = \frac{\sum^2}{m^2}$. La variance logarithmique σ^2 a la signification d'un indice de dispersion relative de la variable $Z(x)$. La variance \sum^2 est liée au carré de la moyenne.

D.2. Estimateur arithmétique ou lognormal ?

La valeur à estimer est la moyenne de la loi lognormale de $Z(x)$: $m = \gamma e^{\frac{\sigma^2}{2}}$. On compare l'estimation arithmétique à l'estimation lognormale du point de vue du biais et de la variance d'estimation. On considère différentes situations correspondant à différents plans d'échantillonnage.

On note : $m^A = \frac{1}{n} \sum_i Z(x_i)$ l'estimateur arithmétique. Comme estimateur lognormal, on prend l'estimateur m^L obtenu en remplaçant dans la formule de m la médiane et la variance logarithmique par leurs estimations :

$$m^L = \gamma^* \frac{e^{\sigma^{2*}}}{2} \quad \text{Log } \gamma^* = \frac{1}{n} \sum_i \text{Log } Z(x_i)$$

$$\sigma^{2*} = \frac{1}{n-1} \sum_i (\text{Log } Z(x_i) - \text{Log } \gamma^*)^2$$

Outre sa simplicité, l'avantage de m^A est son non biais. En effet, si en tout point de l'espace la valeur de $Z(x)$ était connue, la moyenne m serait la moyenne arithmétique des valeurs. L'estimateur m^L est simple à former, mais il est biaisé pour deux raisons : d'une part, parce que la moyenne à estimer est la moyenne spatiale et non la moyenne de la loi ; d'autre part, m^L est un estimateur biaisé de la moyenne de loi quand le nombre n d'échantillons est faible.

D.2.1. Plan aléatoire pur

Les échantillons sont implantés au hasard dans le champ V et indépendamment les uns des autres de telle sorte que chaque partie de V ait la même probabilité d'être échantillonnée. On peut trouver dans MATHERON (1962, chap. XV) les résultats classiques de l'estimation lognormale quand les n échantillons sont indépendants. Ce sont ici des tirages indépendants d'une même loi lognormale.

La variance de l'estimateur m^L est toujours inférieure à celle de l'estimateur m^A . La précision de m^L est d'autant plus grande devant celle de m^A que la variance logarithmique σ^{2*} est grande. C'est là tout l'intérêt de l'estimation lognormale.

L'estimateur lognormal m^L est un estimateur biaisé de la moyenne de la loi quand le nombre d'échantillons est faible. Le terme de ce biais est connu et peut être corrigé (théorie de Sichel). Ce biais diminue avec le nombre d'échantillons et dans les campagnes de pêche, on le négligera.

L'estimateur m^L de variance plus faible que m^A n'est pas optimal. Le plan aléatoire simple correspond aux hypothèses de la statistique classique. FINNEY (1941) a développé par une méthode équivalente au maximum de vraisemblance, un estimateur optimal de $m = \gamma e^{\frac{\sigma^2}{2}}$, m^O , sans biais et de variance minimale :

$$m^O = \gamma * G_n \left(\frac{1}{2} \sigma^{2*} \right)$$

avec : $G_n(t) = 1 + \frac{n-1}{n} t + \sum_{j=2}^{+\infty} \frac{(n-1)^{2j-1}}{n^j(n+1)\dots(n-3+2j)} \frac{t^j}{j!}$; n est le nombre d'échantillons.

La série $G_n(t)$ converge ; AITCHINSON et BROWN (1957) l'ont tabulée. La variance de m^O , $D^2(m^O)$ s'exprime à l'aide du moment d'ordre 4. HOYLE (1968) par la même méthode fournit un estimateur de la variance de m^O :

$$S^2(m^O) = \gamma^{*2} \left[G_n^2 \left(\frac{1}{2} \sigma^{2*} \right) - G_n \left(\frac{n-2}{n-1} \sigma^{2*} \right) \right]$$

La variance d'estimation de m par m^O peut donc se calculer directement.

Toutefois en pratique, on est très rarement dans la situation d'indépendance entre les observations.

D.2.2. Plan aléatoire stratifié

Le champ V est découpé en strates. Dans chaque strate des échantillons sont implantés selon un plan aléatoire simple. Doit-on, vu les résultats précédents, utiliser un estimateur lognormal dans chaque strate ? La variance de dispersion dans chaque strate est inférieure à la dispersion totale. Ainsi la stratification conduit à réduire les variances logarithmiques par strates (strates homogènes) et à rendre l'estimateur arithmétique intéressant malgré le petit nombre de données. CHEVALIER (1985) compare sur un exemple les précisions de m^A et de l'estimateur m^O de Finney, par strate. Elles sont comparables. L'estimateur arithmétique m^A , beaucoup plus simple, est finalement plus intéressant. Ce d'autant plus que l'adéquation au modèle lognormal n'est jamais parfaite.

D.2.3. Plan systématique

Les échantillons sont implantés sur une grille à maille régulière dont l'origine est tiré selon une loi uniforme dans la maille élémentaire. MATHERON (1962, chap. XVIII) donne les caractéristiques de l'estimation lognormale quand les n échantillons sont liés par un système de $\frac{n(n-1)}{2}$ covariances. Les conditions statistiques sont ici moins favorables à l'estimation lognormale que lorsque les n échantillons sont indépendants. D'une part, l'estimateur m^L est un estimateur biaisé de m pour les faibles échantillons. Le terme de biais est connu et peut être corrigé. D'autre part, les estimations de γ et σ^2 ne sont plus indépendantes. L'effet de la non indépendance des échantillons sur l'estimation de la médiane est l'opposé de celui sur l'estimation de la variance de dispersion. MATHERON (1962, chap. IV) met en valeur ces deux effets sur un exemple très démonstratif. Or l'estimateur m^L mêle estimation de la médiane γ et de la variance de dispersion logarithmique σ^2 .

Alors que le plan systématique donne une estimation plus précise de la moyenne spatiale que ne le fait un plan aléatoire simple, ce dernier donne une estimation plus précise de la variance de dispersion (MATHERON, 1962 ; COCHRAN, 1977)⁸. On est dans l'impasse car les variations sur la valeur estimée de σ^2 "s'exponentient" lors de l'estimation de m et conduisent à rendre l'estimateur lognormal m^L très délicat d'utilisation pour les grilles régulières. C'est l'effort d'échantillonnage qui permet d'établir un compromis.

MATHERON (1962, chap. XVIII) montre que si l'on connaissait la variable régionalisée en tous points, la variance de l'estimateur logarithmique $D^2(m^L)$ serait

8. La variance de dispersion sur un volume V se calcule dès que l'on a un variogramme pour toutes les distances dans le champ : $\bar{\gamma}(V,V)$. Dans le cas des grilles régulières, pour calculer la variance d'estimation, on n'a pas besoin en Géostatistique d'estimer la variance de dispersion totale des valeurs. Seule suffit, suivant le principe d'approximation, la variance de dispersion dans la maille élémentaire : $\bar{\gamma}(v,v)$.

beaucoup trop forte. En effet, la moyenne m est alors la moyenne arithmétique des valeurs. Les deux estimations, de la médiane γ^* et de la variance logarithmique σ^2^* ne sont plus indépendantes mais liées par une corrélation négative à cause de la condition $m = \frac{1}{n} \sum_i Z(x_i)$. Or on ne sait pas tenir compte de cette anti-corrélation dans le calcul de D^2 (m^L). Quand le nombre d'échantillons augmente, les lois des estimateurs se symétrisent et leur dispersion diminue ; la moyenne arithmétique redevient intéressante et s'avère moins dangereuse à utiliser.

Ainsi, quand l'effort d'échantillonnage est modéré, l'estimateur m^L peut être intéressant. Toutefois, les précisions antagonistes sur l'estimation de γ et de σ^2 rendent l'utilisation de m^L délicate. Quand l'effort d'échantillonnage est élevé, l'estimateur arithmétique est justifié. En acoustique, les échantillons sont prélevés en continu et en grand nombre. Si l'effort est bien calibré relativement à la structure spatiale des très fortes valeurs, l'estimation arithmétique sera assez précise même si la loi statistique de la variable régionalisée est dissymétrique. C'est le parti que nous avons pris.

La pertinence de l'estimation lognormale repose aussi sur une bonne adéquation de la réalité au modèle lognormal. Les écarts à la lognormalité peuvent entraîner des biais importants. L'approche spatiale, avec un effort d'échantillonnage bien calibré, est finalement plus solide.

D.2.4. Echantillonnage non orthodoxe

En acoustique, sous la contrainte du temps, la prospection est souvent menée (comme dans la bordure N-E de la frayère de hareng) sans échantillonnage préférentiel mais sans non plus respecter strictement un plan d'échantillonnage théorique : le bateau fait route en zigzag couvrant bien la zone en faisant des aller et retours. Dans ce type d'échantillonnage, c'est la rapidité avec laquelle la méthode acoustique permet de prospecter les vastes zones marines qui est mise à profit. En effet, cette propriété est particulièrement importante pour l'échantillonnage de poissons pélagiques. A l'issue d'un tel échantillonnage, on peut construire soit un estimateur du type arithmétique et spatial, soit du type non linéaire et distributionnel.

Le C.I.E.M. coordonne régulièrement, entre différents pays, des campagnes d'estimation du hareng (*Clupea harengus*) sur toute la Mer du Nord jusqu'à la latitude 62° N. A l'issue des campagnes d'écho-intégration de 1989, KIRKEGAARD et al (1990) proposent un estimateur spatial. La zone est découpée en blocs égaux aux dimensions grandes de l'ordre de 30 minutes de longitude sur 30 minutes de latitude. Dans chaque bloc, v_i , la moyenne Z_{v_i} est estimée par la moyenne arithmétique $Z_{v_i}^*$ des données intérieures au bloc. Puis les moyennes de blocs sont combinées pour former l'estimateur spatial Z_V^* de la moyenne sur tout le champ V : $Z_V^* = \frac{1}{n} \sum_i Z_{v_i}^*$ avec $V = \sum_i v_i$. Les auteurs n'assignent pas de variance à leur estimation. Nous proposons ici un calcul. Du

point de vue géostatistique, la variance s'écrit :

$$\sigma^2 = E\left(\left(\frac{1}{N} \sum_i Z_{v_i}^* - Z_V\right)^2\right) = E\left(\frac{1}{n} \sum_i (Z_{v_i}^* - Z_{v_i})^2\right) \text{ où } Z_{v_i} \text{ est la moyenne vraie du}$$

bloc v_i . Comme les blocs sont grands, on suggère de considérer que la corrélation entre les échantillonnages de blocs est faible. En considérant donc que les erreurs d'estimation sont indépendantes entre blocs, on obtient : $\sigma^2 = \frac{1}{n} \sum_i E(Z_{v_i}^* - Z_{v_i})^2$. Ainsi, quand les blocs sont grands et que leurs échantillonnages sont faiblement corrélés, c'est la structure spatiale intra-blocs qui interviendra dans le calcul de la variance d'estimation et non la structure inter-blocs. Tout se passe comme si l'on était dans le cas du plan stratifié avec des strates toutes identiques et régulièrement espacées et dont la taille est grande vis à vis de la portée.

Par contre, Mc LENNAN et al (1988) utilisent un estimateur distributionnel pour l'estimation du hareng autour de l'Ecosse et des Îles Shetland, à l'issue de la campagne acoustique de 1987. Après calcul de la covariance spatiale sur leurs données, ils choisissent une distance u.e.d.e. (Unité Elementaire de Distance d'Echantillonnage, paragraphe B.1.1.) du même ordre de grandeur que la portée de la structure. La régularisation sur cette nouvelle distance u.e.d.e. est effectuée le long de la route du navire. Les segments u.e.d.e. sont exhaustivement reconnus et il n'y a pas d'erreur sur la moyenne de ces segments. Les données régularisées sont sans corrélations le long de la route du navire. Des corrélations spatiales à 2D peuvent subsister car le bateau est revenu à proximité d'un trajet précédent, mais les segments proches dans l'espace, appartenant à deux trajectoires différentes, sont peu représentés dans le jeu de données si bien que à 2D les données sont aussi non corrélées. Leur histogramme reste dissymétrique et un estimateur non linéaire de la moyenne de la loi est proposé. Ce cas de figure nous ramène à la discussion du paragraphe B.6. Cette façon de procéder est aveugle vis-à-vis du champ et ne s'applique que pour des couvertures spatiales assez lâches pour lesquelles les données u.e.d.e. régularisées peuvent être considérée comme des tirages indépendants sur un champ infini. Quand la couverture spatiale est moins lâche, des corrélations spatiales à 2D apparaissent entre segments appartenant à des trajectoires différentes, et l'on devra utiliser un calcul géostatistique sous peine de sous-estimer la variance d'estimation.

D.3. Estimation lognormale du hareng norvégien

On se propose maintenant d'effectuer l'estimation lognormale qui est classique en pêche, sur notre jeu de données norvégiennes, puis d'en discuter le résultat. Dans le champ V de présence du poisson, on a considéré 8 % de zéros intérieurs que l'on a interprétés comme des trous de densité sur le champ de présence. La loi de la variable régionalisée ne peut être modélisée par une loi lognormale. On utilise le modèle de Δ -distribution construit par AITCHINSON et BROWN (1957). La variable prend des valeurs nulles avec une probabilité $1-p$, et ses valeurs non nulles sont distribuées lognormalement. On soumet les valeurs non nulles au test graphique de lognormalité qui est celui de la droite de Henry.

Pour les mêmes quantiles q , soient ξ_q et ν_q les valeurs de la variable lognormale $Z(x)$ et de la gaussienne centrée réduite $y(x)$. On a la relation : $\xi_q = \gamma e^{\nu_q \sigma}$ avec : $q = \text{Prob}(Z(x) \leq \xi_q) = \text{Prob}(y(x) \leq \nu_q)$, soit : $q = F(\xi_q) = G(\nu_q)$.

Si l'on reporte sur un graphe en ordonnées les valeurs du logarithme de la variable $Z(x)$ ($\text{Ln } \xi_q = \text{Ln } F^{-1}(q)$) et en abscisses les valeurs de la gaussienne $N(0,1)$ pour les mêmes quantiles ($\nu_q = G^{-1}(q)$), on doit obtenir une droite si $Z(x)$ est lognormale. Le graphe est fourni sur la Figure 26. Les points s'alignent très bien sur une droite, sauf aux deux extrémités. Pour les faibles valeurs, $\text{Ln } (\xi_q)$ est inférieur à la droite : pour les faibles valeurs, l'histogramme expérimental de $Z(x)$ est en-deçà de la loi lognormale. Au contraire, aux grandes valeurs, les 13 plus fortes valeurs s'individualisent très nettement et s'écartent de la droite en se plaçant au-dessus d'elle : la queue de l'histogramme expérimental est plus importante que celle de la loi lognormale. Le modèle lognormal s'ajuste au grand nombre des valeurs moyennes, s'ajuste mal à l'origine et à l'infini. Le carré de la pente de la droite de Henry est égal à la variance logarithmique. Le calage de la droite a une grande importance si l'on veut en faire une estimation graphique.

On estime la médiane γ et la variance σ^2 des valeurs non nulles de $Z(x)$ graphiquement. On a calé la droite de Henry pour $\nu_q = -1$ et $\nu_q = 1$, car au-delà les fréquences observées sont soumises à des fluctuations expérimentales importantes. On a :

$$\text{Log } \nu^* = \text{Ln } \xi_{50\%} = 5,953$$

$$\text{Ln } \xi_{16\%} = \text{Ln } \xi_{50\%} - \sigma$$

$$\text{Ln } \xi_{84\%} = \text{Ln } \xi_{50\%} + \sigma$$

d'où :

$$\sigma^* = \frac{1}{2} \left(\text{Ln } \frac{\xi_{50\%}}{\xi_{16\%}} + \text{Ln } \frac{\xi_{84\%}}{\xi_{50\%}} \right) = 1,774$$

Les moments de la Δ -distribution qui prend ses valeurs nulles avec la probabilité $1-p$ sont : $m_n = p \gamma^n e^{\frac{n^2 \sigma^2}{2}}$.

L'estimateur de la moyenne m que l'on forme est : $m^\Delta = \frac{n_1}{n} \gamma^* e^{\frac{\sigma^{*2}}{2}}$, où n_1 est le nombre de valeurs non nulles. On trouve : $m^\Delta = 1703$. Ce résultat est très proche de l'estimateur arithmétique $m^A = 1690$.

On a estimé la variance en utilisant la formule de HOYLE (1968) que PENNINGTON (1983) a réécrite dans le cas d'une Δ -distribution. L'estimateur $S^2(m^\Delta)$ n'est pas optimal car les échantillons sur la grille régulière ne sont pas indépendants :

$$S^2(m^\Delta) = \frac{n_1}{n} \gamma^{*2} \left[G_{n_1}^2 \left(\frac{1}{2} \sigma^{*2} \right) - \frac{n_1 - 1}{n - 1} G_{n_1} \left(\frac{n_1 - 2}{n_1 - 1} \sigma^{*2} \right) \right]$$

On arrête la sommation de la fonction $G_n(t)$ quand le terme à rajouter est inférieur à 2 % de la somme déjà calculée. La somme comprend 10 termes. Le coefficient de variation de la Δ -estimation est égale à : $\frac{S(m^\Delta)}{m^\Delta} \simeq 13,9\%$. Dans notre cas, l'estimation lognormale n'apporte aucun gain de précision par rapport à l'estimation arithmétique dans le modèle structural pépitique. PENNINGTON (1983) au contraire améliore la précision de l'estimation des œufs de maquereau en utilisant la Δ -distribution. Son effort d'échantillonnage est très modéré : une centaine de prélèvements sur toute l'étendue du Banc Georges. La calibration de l'effort d'échantillonnage sur la structure spatiale des très fortes valeurs apparaît suffisant, au vu de notre exemple expérimental, à rendre la moyenne arithmétique précise.

D.4. Intervalle de confiance et fourchette d'erreur

On indice par w l'estimation de m que donne l'échantillonnage w . On note de façon générale m_w^* l'estimateur de m , qu'il soit m_w^A ou m_w^L ou m_w^O . m_w^* est une variable aléatoire. Les différentes valeurs m_w^* que l'on pourrait obtenir par des échantillonnages w indépendants sont indépendantes. On cherche à construire un intervalle autour de m_w^* dans lequel la moyenne m inconnue se trouve avec une probabilité donnée. En effet, la finalité à laquelle est destinée l'estimation (que ce soit pour comparer des années entre elles, pour valider des estimations faites par d'autres méthodes, ou pour caler un modèle de dynamique de population), nous amène à construire si ce n'est un intervalle de confiance au moins une fourchette d'erreur.

Pour ce faire, il faut connaître la loi de l'estimateur m_w^* et ses quantiles. Or on ne la connaît pas. On ne connaît que ses deux moments : sa moyenne est la valeur m ($E(m_w^*) = m$) si l'estimateur est non biaisé ; sa variance est la variance d'estimation. La

variance d'estimation est aussi relative à l'échantillonnage w car elle est calculée grâce au variogramme inféré, γ_w , sur les données de l'échantillonnage w . Le type de loi de m_w^* est lié à la loi de la variable régionalisée et à l'effort d'échantillonnage. On peut donc considérer que la loi de m_w^* est lognormale quand l'effort d'échantillonnage est modéré et qu'elle tend vers une gaussienne quand l'effort d'échantillonnage augmente. Connaissant donc les quantiles de la loi de m_w^* , dans l'hypothèse gaussienne ou lognormale, on pourrait construire un intervalle de probabilité autour de la valeur m dans lequel m_w^* se trouve. Or ce n'est pas ce que nous cherchons. Nous cherchons un intervalle de confiance contenant la valeur m inconnue, c'est un intervalle aléatoire construit autour de la variable aléatoire m_w^* .

Dans l'hypothèse gaussienne, la variable aléatoire m_w^* s'écrit: $m_w^* = \sigma_w Y + m$ où σ_w est la variance d'estimation et Y la gaussienne centrée réduite. L'intervalle de probabilité au risque 5% s'écrit alors : $\text{Prob}(m - 2\sigma_w \leq m_w^* \leq m + 2\sigma_w) = 0.95$. On en déduit l'intervalle de confiance gaussien : $\text{Prob}(m_w^* - 2\sigma_w \leq m \leq m_w^* + 2\sigma_w) = 0.95$.

Dans l'hypothèse lognormale, on note σ_w^2 non plus la variance de la loi de m_w^* (variance d'estimation) mais la variance logarithmique de m_w^* . La variance de m_w^* est notée $\sigma_{cs/w}^2$. La variable aléatoire m_w^* s'écrit : $m_w^* = m e^{\sigma_w Y - \sigma_w^2/2}$ où Y désigne toujours la gaussienne centrée réduite. L'intervalle de probabilité au risque 5% contenant m_w^* s'écrit: $\text{Prob}(m e^{-2\sigma_w - \sigma_w^2/2} \leq m_w^* \leq m e^{+2\sigma_w - \sigma_w^2/2}) = 0.95$. On en déduit l'intervalle de confiance lognormal : $\text{Prob}(m_w^* e^{-2\sigma_w + \sigma_w^2/2} \leq m \leq m_w^* e^{+2\sigma_w + \sigma_w^2/2}) = 0.95$.

Dans l'hypothèse lognormale, on a besoin de la variance logarithmique σ_w^2 de la loi de m_w^* . LAJAUNIE et al. (1987) proposent un modèle dans lequel σ_w^2 est calculable à partir du variogramme γ_w inféré sur les données. Ici, on procède directement à partir de l'expression mathématique de la variance logarithmique σ_w^2 : $\sigma_w^2 = \text{Log}(\frac{\sigma_{cs}^2}{m^2} + 1)$. Or, ni m ni σ_{cs}^2 ne sont connus. Par notre échantillonnage w , on ne peut former que m_w^* et $\sigma_{cs/w}^2$. De même que ces auteurs, on fait l'hypothèse plausible que les variances relatives vraies et d'échantillonnages sont égales : $\frac{\sigma_{cs/w}^2}{m_w^{*2}} = \frac{\sigma_{cs}^2}{m^2}$. On calcule donc la variance logarithmique σ_w^2 ainsi : $\sigma_w^2 = \text{Log}(\frac{\sigma_{cs/w}^2}{m_w^{*2}} + 1)$.

Sur nos données, on déduit l'intervalle de confiance lognormal conventionnel en prenant pour m_w^* la moyenne arithmétique m^A et pour $\sigma_{cs/w}^2$ la variance d'estimation

géostatistique calculée en B. La variance d'estimation relative est $\frac{\sigma_{cs/w}^2}{m^{A2}} = 0.0156$.

Comme elle est petite devant 1, on prend: $\sigma_w^{2*} \approx \frac{\sigma_{cs}^2}{m^{A2}}$. Le calcul donne : $m^A = 1690$

compris entre 1326 et 2187. La fourchette d'erreur lognormale est de 22% à gauche et de 29% à droite. Elle n'est pas trop dissymétrique et elle est proche de la fourchette gaussienne ($1690. \pm 25\%$), ce qui est compatible avec l'effort d'échantillonnage développé.

PARTIE II

ANALYSE DES CARTES

Logiquement dans une étude minière, après l'estimation globale vient l'estimation locale (la cartographie), car elle sera un outil pour l'exploitant. Mais pour le biologiste des pêches, la cartographie au moment précis de la campagne d'échantillonnage n'est pas vraiment intéressante. Il y a sur ce point comme une séparation des objectifs entre le scientifique et l'exploitant pêcheur. C'est le métier du marin-pêcheur que de savoir où se trouve le poisson à un moment donné pour le pêcher, et c'est celui du biologiste des pêches que de savoir quelle est la quantité totale en place et combien on peut en retirer de la mer sans mettre le stock en danger. En quelque sorte : la cartographie au pêcheur, l'estimation globale au scientifique. Du point de vue biologique, plutôt qu'une carte particulière, ce sont les lois de l'occupation de l'espace par le stock de poisson qui vont nous intéresser. Elles relèvent de l'auto-écologie du stock, c'est-à-dire d'une part de ses rapports avec les environnements physique et biologique définissant son habitat, et d'autre part de facteurs intrinsèques au stock qui constituent ce que l'on pourrait appeler sa stratégie d'occupation de l'espace. Certes, cette préoccupation n'est pas nouvelle. Il a déjà été observé, tout particulièrement chez les poissons pélagiques (clupéidés), que la diminution de l'aire d'extension du stock était concomitante d'une réduction de sa taille (abondance totale) alors que les densités restaient relativement constantes. La relation entre l'abondance du stock et sa dispersion spatiale peut agir dans les deux sens. DRAGESUND et al. (1980) attribuent la réduction de l'aire d'extension du hareng Atlanto-scandinave à la chute considérable des effectifs sous la pression de la pêche dans les années 70. Après l'arrêt total de la pêche en 1977, des campagnes d'échantillonnage ont montré que, le groupe reconstituant ses effectifs, des habitats délaissés étaient recolonisés. JUNQUERA (1986) décrit une contraction de l'habitat de l'Anchois du Golfe de Gascogne de 1960 à 1975 attribuable à des modifications écologiques qui auraient alors entraîné une chute de la taille du stock, voire une disparition du stock galicien. Selon l'auteur, l'effort de pêche actuel interdirait au poisson une recolonisation d'un habitat dont la taille a pu ré-augmenter. Il y a donc bien une relation entre démographie et régionalisation pour le stock tout entier.

On va, dans cette partie, chercher s'il existe une relation entre l'effectif total des tous jeunes individus issus de la reproduction chaque année et les caractéristiques de leur régionalisation. On sait que ces effectifs varient d'une année sur l'autre. Qu'en est-il des régionalisations associées ? On veut décrire ainsi, de façon spatiale, les énigmatiques mécanismes du recrutement. Notre idée part du fait que les histogrammes sont d'un type lognormal. On analyse donc la nature grâce aux propriétés du modèle lognormal : des petites variations inter-annuelles de la variance logarithmique suffisent à faire varier énormément la moyenne (effectif total) même si la médiane ne change pas. La variance logarithmique mesure la dispersion relative de la variable régionalisée. Ainsi, à partir

d'échantillonnages sur plusieurs années des jeunes issus de la reproduction de l'année, on devrait pouvoir tenter de relier la fonction d'hétérogénéité spatiale à la quantité totale, c'est-à-dire à la qualité du recrutement. Cette analyse que l'on va entreprendre sur les premiers stades de la vie des poissons s'accorde tout à fait avec l'analyse de la relation stock-recrutement effectuée par SHEPHERD et al (1990). Rappelons que la relation stock-recrutement est un nuage de corrélation entre, en ordonnée, la biomasse des enfants (recrues d'âge 0 : groupe 0) produits l'année n et, en abscisse, celle de leurs parents de l'année n . L'observation des relations stock-recrutement pour quelques stocks importants de Mer du Nord (Gadidés) sur trois décennies, montre que les lois des recrues-0, conditionnellement à leurs géniteurs, sont dissymétriques et très dispersées. Analysant ces lois par des modèles lognormaux, les auteurs proposent le mécanisme suivant de renouvellement : la variance logarithmique de la loi des recrues connaissant leurs géniteurs, augmente quand la biomasse des géniteurs diminue. Ils montrent alors que, sans faire intervenir des mécanismes compensatoires très stricts, les stocks de gadidés ne peuvent disparaître totalement puisque, même sous une pression de pêche importante, une année apportera un recrutement important reconstituant le stock ⁹. Cette année-là aura-t-elle été celle d'une hétérogénéité spatiale particulière ?

Si l'on peut identifier des variations dans les dispersions spatiales d'années en années et des variations dans la localisation des concentrations de recrues (ou d'œufs pondus), si l'on peut relier ces variations spatiales à celles des quantités totales associées, alors on est bien sur le chemin qui devrait permettre de cerner de façon plus précise les mécanismes plus biologiques et éco-physiologiques qui sont en jeu dans le déterminisme du taux de recrutement chaque année. Pour cette recherche, on a travaillé sur deux bases de données ; l'une se situant à la fin des processus de recrutement, l'autre au début. Il s'agit de l'échantillonnage pendant quatre années du groupe-0 du merlu dans le Golfe de Gascogne. Ces données nous ont été aimablement mises à disposition par Mr J.C. POULARD de l'IFREMER-Lorient. L'autre jeu de données consiste en la surveillance pendant quatre années de la frayère de sole (échantillonnage des œufs et larves) dans le Golfe de Gascogne. Le traitement de ces données a fait l'objet d'un contrat passé avec le Laboratoire Oeufs-Larves-Juveniles (Mr Y. DESAUNAY) de l'IFREMER-Nantes et s'est inséré dans le programme de recherche national sur le déterminisme du recrutement (PNDR).

Dans un premier temps, on propose un test grâce aux courbes de sélectivité minière, pour choisir entre trois types de stratégie d'occupation de l'espace. L'application à nos données montre que le merlu-0 comme les œufs de sole occupent chaque année tout leur habitat sur les mêmes lieux, et que la dispersion spatiale relative est toujours la même quelque soit la quantité totale. Ce résultat inattendu s'interprète logiquement dans la

9. La vie d'un gadidé est de l'ordre de la décennie. Ainsi, ces stocks sont composés de nombreuses classes d'âge et peuvent, dans l'analyse de Shepherd et al (1990), attendre l'année à fort recrutement. En revanche, l'Anchois (clupéidés) ne vit au maximum que quatre années. Effectivement, les stocks d'Anchois sont connus pour leurs très grandes variations d'abondances et leurs extinctions (JUNQUERA, 1986).

mesure où les quantités totales estimées varient sur une échelle arithmétique et ne sont pas toujours significativement différentes. Toutefois, ce résultat qui est une caractéristique physique des données, nous permet de construire un modèle spatio-temporel multiplicatif de la ponte de la Sole.

A. ETUDE DE LA DISPERSION PAR LES RELATIONS TONNAGE-TENEUR

On étudie d'année en année le lien qui existe entre la dispersion des valeurs échantillonnées et leur moyenne (quantité totale). Pour ce faire, on dispose des histogrammes de plusieurs campagnes d'échantillonnage effectuées sur plusieurs années. Il s'agit de comparer les dispersions sur les histogrammes. Or leurs moyennes et variances diffèrent. Une comparaison directe n'est donc pas immédiate. En Géostatistique minière et stationnaire, un certain nombre d'outils ont été développés afin de caractériser les lois de probabilité et leur dispersion, en particulier pour résoudre le problème de l'effet de support lors de la prévision des quantités récupérables (MATHERON, 1981 ; LANTUEJOUL, 1990). Les courbes dites Q(T) nous sont ici très utiles. En effet, elles mettent en relation la fréquence d'apparition d'une classe de valeurs (tonnage : T) à la quantité de poisson (ou de métal) que ces valeurs représentent (quantité : Q). On expose le principe de ce graphe et la notion de sélectivité, puis on développe un test simple grâce aux courbes Q(T) expérimentales afin de caractériser trois types de relations entre quantité totale et hétérogénéité spatiale.

A.1. Graphes Q(T) et sélectivité

On note Z la variable aléatoire représentant la densité de poisson en chaque point de l'espace, m sa moyenne, F sa fonction de répartition ($F(z) = \text{Prob}(Z < z)$).

La fonction tonnage est définie par : $T(z) : \int_z^{+\infty} dF(u) = 1 - F(z) = E(1_{Z \geq z})$. Le

tonnage T(z) est la fréquence cumulée des valeurs supérieures au seuil z. Le tonnage T(z) mesure la surface occupée par les valeurs supérieures à z, en pourcentage par rapport à l'aire totale prospectée. La fonction quantité de métal est définie par :

$Q(z) : \int_z^{+\infty} u dF(u) = E(Z 1_{Z \geq z})$. Q(z) est la quantité de métal ou de poisson (intégrale

d'espace) qui se trouve sur la surface T(z) définie par $Z \geq z$. Le graphe Q(T) paramètre la dispersion en fonction des surfaces couvertes, T, et non pas en fonction des seuils z. Q varie de 0 à m. T varie de 0 à 1. La concavité de la courbe nous renseigne sur la dispersion des valeurs.

L'aire qui sépare la courbe Q(T) de la droite mT s'appelle la sélectivité :

$S = 2 \int_0^1 (Q(T) - mT) dT = \int_0^{+\infty} F(u)(1 - F(u)) du$ (Figure 27). La sélectivité comme

la variance mesure la dispersion de la loi. Mais la variance ne se lit pas sur la courbe. On n'en a pas besoin et elle peut très bien, dans le modèle, ne pas exister. Ceci est un avantage pratique dans la mesure où, d'après sa définition, la sélectivité est moins sensible aux fortes valeurs que ne l'est la variance des données. On est maintenant en mesure de définir une relation d'ordre entre deux lois de probabilité (MATHERON, 1981). Soient deux lois F_1 et F_2 possédant la même moyenne m. F_2 est dite "plus sélective" que F_1 si la courbe $Q_2(T)$ est au-dessus de $Q_1(T)$ pour chacune des valeurs de T :

$Q_2(T) > Q_1(T)$, $\forall T \in]0, 1[$. Cela veut dire que sur chacune des surfaces $T(z)$, la quantité de poisson $Q(z)$ de l'année 2 est plus importante que celle de l'année 1 alors que les quantités totales des années 1 et 2 sont identiques.

Ainsi, en comparant les sélectivités ou tout simplement l'allure des graphes $Q(T)$ calculés directement sur les données, on peut facilement comparer les dispersions entre différentes années associées à des quantités totales différentes. Plutôt que $Q(z)$, on calculera alors $P(z) = Q(z)/m$, c'est-à-dire le pourcentage de biomasse totale que constituent les valeurs supérieures à z . On en déduit le graphe $P(T)$. Pour différentes années, P et T varient chacun entre 0 et 100 % et la sélectivité mesure la dispersion des valeurs relativement à la quantité totale de l'année. Grâce aux graphes $P(T)$, on peut donc lire simplement de visu si, relativement aux quantités totales, l'année 2 est plus sélective ou non que l'année 1. La sélectivité de la courbe $P(T)$ est donc un paramètre analogue à la variance logarithmique. Mais l'allure de la courbe recèle une information beaucoup plus riche que le simple paramètre de dispersion qu'est le coefficient de variation Σ/m .

A.2. La sélectivité pour valider le modèle lognormal du recrutement

La sélectivité des graphes $P(T)$ fournit un test graphique simple permettant de décider si les variations observées du recrutement peuvent effectivement être analysées grâce au modèle lognormal. Ce modèle nous dit que, une année donnée, les phénomènes locaux propices à la survie des jeunes poissons (œufs ou larves planctoniques ou juvéniles-0 colonisant leur habitat) engendrent une explosion des densités de façon multiplicative en des lieux privilégiés. Ceci se traduit alors sur l'histogramme de la variable densité par une queue de distribution particulièrement longue et par une grande dispersion. Par rapport aux autres années, il suffit dans le modèle que la variance

logarithmique $\left(- \frac{\sum^2}{m^2} \right)$ soit augmentée sans que la médiane ne le soit. L'hétérogénéité

spatiale plus ou moins grande de la fonction de survie suffit à générer dans le modèle lognormal des variations importantes du recrutement. Dans un tel modèle descriptif, que l'on peut qualifier d'analyse lognormale du recrutement, la courbe $P(T)$ de l'année à fort recrutement est alors nécessairement plus sélective que celles des années aux recrutements plus faibles. Le test graphique de sélectivité des courbes $P(T)$ permet de soumettre aux données le modèle lognormal du recrutement.

Si l'analyse lognormale du recrutement s'accorde avec la lecture des relations stock-recrutements effectuée par SHEPHERD et al (1990), elle est aussi compatible avec la classique notion de période critique (CUSHING, 1972 ; MAY 1974). La période critique est la période au cours de laquelle, pendant le développement de l'organisme, la mortalité est très élevée. Elle fonctionnerait comme un régulateur d'abondance. La période critique de MAY (1974) est sous la dépendance de facteurs intrinsèques à l'espèce : un stade de développement est par constitution plus fragile qu'un autre. La

période critique de CUSHING (1972) est éco-physiologique. Cet auteur définit la période critique comme le moment du bon couplage nécessaire entre le moment où les larves doivent commencer à se nourrir dans le plancton et le moment du démarrage de la production planctonique dont le développement printanier ("bloom") de phytoplancton est l'origine. S'il y a déphasage dans le temps et/ou l'espace entre la demande nutritive des larves et l'abondance des proies planctoniques ("mismatch"), des mortalités auront lieu. Au contraire, s'il y a concomitance ("match") la survie des larves sera possible en certains endroits propices.

Toutefois, la période critique et le modèle de "match/mismatch", que nous interprétons comme des formulations écologiques de l'analyse lognormale, restent heuristiques car ils n'ont jamais été observés en mer à notre connaissance. Le fait de ne pas avoir observé un tel mécanisme ne prouve pas qu'il n'agit pas. Il peut être simplement très difficile à observer. Par contre, un autre mécanisme aux échelles spatiales et temporelles moins fines a été mis en évidence. ILES et al (1982) décrivent une situation où les facteurs physiques hydrodynamiques sont prépondérants dans le déterminisme du recrutement. Les auteurs essaient de comprendre pourquoi dans l'Atlantique Nord il y a de grands stocks et de petits stocks de harengs. Ils montrent que la taille des stocks étudiés (leur biomasse) est proportionnelle à la zone moyenne sur laquelle les larves chaque année sont présentes. La taille de l'habitat des larves limite le recrutement. De cette comparaison entre stocks, les auteurs proposent le même mécanisme pour interpréter les variations du recrutement à l'intérieur d'un stock : les pulsations géométriques de l'habitat des larves, sous la dépendance de conditions physiques, sont liées aux variations du recrutement.

L'analyse lognormale du recrutement n'est donc pas la seule. On va s'attacher à caractériser quelques types de relations entre la dispersion et la quantité totale. Certaines de ces stratégies d'occupation de l'espace permettent peut-être aux quantités totales de plus amples fluctuations.

A.3. Les courbes $O(T)$ et $P(T)$ pour caractériser trois stratégies d'occupation de l'espace

On construit des outils d'investigation que l'on appliquera à nos données de recrutement dans le paragraphe B. On se propose de caractériser, grâce aux courbes $Q(T)$ et $P(T)$, les grands types de comportements spatiaux quand la quantité totale varie. On a répertorié trois types de réponses locales aux variations de la quantité globale.

Quand la quantité totale augmente, il y a augmentation importante des densités en quelques lieux spécifiques sans autre modification de la régionalisation des densités. Cette hypothèse (H_1) est tirée de l'analyse lognormale précédente (CUSHING, 1972). La seconde hypothèse s'énonce ainsi : quand la quantité totale augmente, il y a augmentation de l'aire de distribution avec une densité constante. Cette hypothèse (H_3) est tirée de l'observation de ILES et al (1982). Elle n'est pas spécifique du recrutement. Elle correspondrait aussi à une stratégie de type pélagique chez des poissons grégaires (DRAGESUND et al, 1980 ; JUNQUERA, 1986). Il convient d'ajouter une troisième

stratégie intermédiaire : quand la quantité totale augmente, il y a augmentation de la densité en chaque point, proportionnellement à la quantité totale sans que la régionalisation des pourcentages ne varie. C'est l'hypothèse notée (H₂). Dans les hypothèses (H₁) et (H₂), la géométrie de l'habitat est fixe et le poisson occupe tout son habitat ; c'est la densité qui varie quand la quantité totale varie. Dans l'hypothèse (H₃), la densité est fixe, c'est la géométrie de l'habitat qui varie ou c'est le poisson qui n'occupe qu'une partie de son habitat quand la quantité totale varie.

L'hypothèse (H₂) correspondrait plutôt à une stratégie d'occupation de l'espace de type benthique ou démersale chez des poissons de fond solitaires. De fait, MYERS et al (1989) l'ont déjà rencontrée pour certains gadidés en Mer du Nord.

Sur la Figure 28, on a représenté graphiquement chacune des hypothèses H₁, H₂, H₃ ainsi que les comportements spécifiques des courbes Q(T) et P(T) qui permettent de caractériser chacun des trois cas de figure. On a représenté deux années, l'une où la quantité totale de la classe d'âge est importante (m₂), l'autre où elle est faible (m₁) (m₂ > m₁). Dans l'hypothèse (H₃), les courbes Q(T) se superposent à l'origine et montrent une croissance linéaire. En effet, comme la densité est constante sur les zones habitées, Q est proportionnelle à T. Les courbes ne se séparent que pour la valeur de T correspondant à la surface habitée la plus petite. La courbe correspondante 1 s'arrondit pour atteindre la moyenne m₁. La courbe 2, elle, continue à croître linéairement jusqu'à ce que, à son tour, elle s'arrondisse pour atteindre la moyenne m₂. Dans l'hypothèse (H₂), les courbes Q(T) sont proportionnelles. Comme la régionalisation des pourcentages ne varie pas dans l'hypothèse (H₂), c'est donc pour cette hypothèse que les courbes P(T) sont entièrement superposées alors que les courbes P(T) dans l'hypothèse (H₃) ne sauraient l'être. Dans l'hypothèse (H₁), il n'y a jamais de superposition ni des courbes Q(T) ni des courbes P(T). Par contre, dans l'hypothèse (H₁), la courbe P₂(T) de l'année riche 2 est plus sélective que la courbe P₁(T) de l'année pauvre 1. Ce cas se différencie bien du cas (H₃) dans lequel la sélectivité des courbes P(T) est inversée. C'est pour l'année pauvre que la courbe P(T) est plus sélective. Ainsi, on propose de tracer les courbes Q(T) d'abord. Si l'on constate une superposition à l'origine, on est plutôt dans le cas (H₃). Sinon, on trace les courbes P(T). Si l'on constate une superposition des courbes, on est plutôt dans le cas (H₂). Sinon, il faut regarder le sens de la sélectivité des courbes P(T). On peut ainsi aisément détecter la logique spatiale d'un poisson.

Les hypothèses précédentes sont comme des cas de référence. Dans la réalité on pourra trouver des cas intermédiaires comme d'ailleurs des combinaisons des hypothèses (H₁), (H₂), (H₃). Les relations tonnage-teneur ne permettent nullement de supposer que les zones riches sont au même emplacement dans l'espace d'une année sur l'autre, comme le laisseraient croire les schémas représentatifs de la Figure 28. Il faudra s'aider d'une représentation des données faisant office de carte sommaire.

Les courbes Q(T) et P(T) se calculent facilement. Il suffit de ranger les données de chaque campagne par ordre décroissant et, partant de la plus forte valeur, de parcourir les données une à une, ou classe à classe. On effectue alors, d'une part, un cumul soit des

II.A. Etude de la dispersion par les relations tonnage-teneur

valeurs soit des pourcentages qui donne soit Q soit P, et d'autre part, on incrémente un compteur de fréquence qui donne la valeur correspondante de T. On a :

$$Q(z_p) = \sum_{i=p}^n \frac{n_i}{n} z_i ; \quad T(z_p) = \sum_{i=p}^n \frac{n_i}{n}, \text{ où } n_i \text{ est le nombre de données dans la classe } i, z_i \text{ la}$$

moyenne de la classe i et n le nombre total des données.

B. APPLICATION AUX DONNÉES

Dans cette partie, on restera dans le Golfe de Gascogne où, grâce à l'IFREMER, nous avons pu disposer de bases de données spatio-temporelles ainsi que de la problématique scientifique du programme national de recherche sur le recrutement. On plante le cadre biogéographique et sédimentaire du Golfe de Gascogne, où l'on situe l'habitat privilégié des espèces étudiées. On traitera du groupe d'âge 0 du merlu et des œufs planctoniques de la sole. On décrit les échantillonnages, on applique les tests précédemment construits afin de révéler la réponse spatiale locale aux variations des quantités totales. Pour interpréter biologiquement les résultats, on a besoin de calculer des fourchettes d'erreur autour des estimations globales afin de mettre en évidence des variations significatives entre les années.

B.1. Le Golfe de Gascogne et ses caractéristiques

Le Golfe de Gascogne, tel qu'on l'entend ici, correspond au plateau continental de la côte Atlantique française. Cette dernière a pour limites au nord, la latitude de l'île d'Ouessant par 48°30'N et au sud le Gouf de Cap Breton par 43°30'N. Le plateau présente deux zones morphologiquement différentes : l'une, septentrionale depuis la Bretagne jusqu'au sud de la Gironde par 45°N, est orientée NW-SE large de 60 à 100 miles nautiques ; l'autre, méridionale devant la côte landaise, est étroite d'une trentaine de miles, d'orientation N-S. La zone entre la Gironde et Arcachon forme une transition. Les isobathes sont parallèles à la côte.

Dans la partie septentrionale, l'orientation moyenne des isobathes est N225°. L'axe côte-large est orienté au N55°. Sur le plateau landais, les orientations sont cardinales.

GLÉMAREC (1977) fournit une carte synthétique biosédimentaire de l'ensemble du Golfe de Gascogne que nous reproduisons sur la Figure 29. A cette échelle, les sédiments montrent une gradation côte-large et s'agencent en trois bandes parallèles à la côte. Des fonds de 10 mètres jusqu'à la sonde des 80 mètres environ, des sables graveleux et propres dominant. Ils sont parsemés çà et là d'îlots de roches, de vases ou de sables dunaires. Plus au large, on rencontre à cheval sur la sonde des 100 mètres, une bande de sables fins envasés entre 80 et 140 mètres environ qui courent depuis la pointe de Penmarch jusqu'au Gouf. Cette bande de sables fins envasés présente dans la partie nord du Golfe un envasement plus important en trois endroits. Elle s'étend particulièrement en largeur sous la mâchoire sud de la Bretagne autour de l'Archipel des Glénans où son orientation est E-W. Dans le bloc septentrional du Golfe de Gascogne, cette bande de sables fins envasés occupant le milieu du plateau continental est connue sous le nom de "la Grande Vasière".

Occupant le milieu du plateau et constituée du dépôt de particules fines, la Grande Vasière est à relier à l'hydrodynamisme général du Golfe. VINCENT et al. (1969) décrivent l'individualisation d'un minimum thermique en été sur le fond, le bourrelet froid dont l'emplacement correspond à celui de la Grande Vasière. Il forme parfois une langue

en continuité avec l'eau de fond des plateaux celtique et irlandais. Il peut aussi parfois être isolé formant une poche nordique dans des eaux plus chaudes. PINGREE et al. (1982) mesurent des courants moyens advectifs (courants résiduels) très faibles sur le plateau dont l'orientation est parallèle aux isobathes, du SE vers le NW. Advection faible, sédiments fins, eau de fond peu variable à affinité plus nordique, voilà en quelques mots les caractéristiques physiques principales du milieu du plateau continental dans le nord du Golfe de Gascogne ¹⁰. Enfin, entre 140 mètres et l'accore, on trouve seulement dans la zone nord plus au large, une troisième bande différente, formée de sables moyens, dont la faune fouisseuse est spécifique. Ce sont les deux bandes plus côtières qui vont nous retenir.

La Grande Vasière est l'habitat principal du jeune merlu pendant ces deux premières années (DARDIGNAC, 1988). Les sables graveleux autour des îlots rocheux entre 50 et 80 mètres constituent l'habitat privilégié de la sole adulte l'hiver et en particulier celui de ses frayères (ARBAULT et al, 1969 ; DARDIGNAC, 1988 ; KOUTSIKOPOULOS, 1991).

B.2. Le merlu-0 de 1987 à 1990

Dans le Golfe de Gascogne, le merlu rencontré est l'espèce *Merluccius merluccius*, L. La biologie spatiale des adultes y est incomplètement connue. En revanche, celle des jeunes l'est beaucoup mieux (DARDIGNAC, 1988). La bande sablo-vaseuse du large entre 80 et 140 mètres héberge les petits merlus dans leurs deux premières années. Après le frai, qui aurait lieu à proximité des accores en pontes fractionnées à la fin de l'hiver, les jeunes nés au printemps (groupe d'âge 0) commencent à coloniser leur habitat spécifique dès juillet. On considère qu'à l'automne, en octobre, la classe d'âge est complètement recrutée sur son habitat. Les jeunes y restent jusqu'à leur troisième année où le groupe II durant son troisième été se disperse sur tout le plateau et commence à avoir un comportement saisonnier d'adulte. La Grande Vasière et la bande sablo-vaseuse du Sud Gascogne constitue les nourriceries du merlu. La Grande Vasière est l'une des plus importantes nourriceries de tout l'Ouest européen (DARDIGNAC, 1988). Des rapports entre les adultes du Golfe de Gascogne et ceux des plateaux celtique et irlandais étant soupçonnés, on ne peut considérer que le merlu du Golfe constitue un stock unitaire pour la reproduction (stock unité). Cela n'est pas grave pour l'étude de la dispersion des recrues sur leur nourricerie en fonction de leur quantité totale. Toutefois, les estimations globales ne peuvent servir d'indice de recrutement à la population dont l'extension spatiale est mal connue.

10. Le réchauffement estival important des eaux, principalement sur le plateau landais, atteint progressivement au cours de l'été les eaux du Nord du Golfe de Gascogne et conduit dans les eaux cotières à une stratification thermique marquée sur la verticale. C'est la seconde caractéristique physique importante du Golfe de Gascogne car d'une part, elle peut affecter la répartition spatiale de poissons pélagiques méridionaux comme la sardine (VINCENT et al, 1969). Elle peut, d'autre part affecter de façon différentielle dans l'espace la croissance estivale des jeunes poissons du group-0. C'est le cas du merlu (GUICHET, 1988).

B.2.1. Echantillonnage

Depuis 1987, l'IFREMER réalise saisonnièrement un programme d'échantillonnage de l'ensemble du Golfe de Gascogne destiné à augmenter les connaissances biologiques et spatiales des espèces commerciales de ce secteur comme décrit dans GUEGUEN (1988) et POULARD (1990). Nous nous sommes intéressés à la régionalisation de la classe d'âge-0 du merlu dont la régionalisation est bien délimitée et entièrement échantillonnée. Les campagnes qui nous intéressent sont celles d'automne. Elles sont conduites sur l'ensemble du Golfe depuis le Gouf de Cap Breton jusqu'à la latitude d'Ouessant, depuis les sondes de 15 mètres jusqu'à celles de 600 mètres.

Le plan d'échantillonnage est stratifié selon 8 strates de latitude que recouperont 7 strates de profondeur. Sur la Figure 30, on donne l'échantillonnage de la campagne d'automne 1988. Dans les strates, certaines stations sont fixes et visitées à chaque campagne du programme evhoe ; d'autres sont mobiles. On considérera que la disposition des stations d'échantillonnage est aléatoire dans les strates. 130 à 140 stations sont ainsi réparties sur l'ensemble du Golfe de Gascogne à chaque campagne. L'engin de pêche est un chalut de fond à grande ouverture verticale dont le type est le standard adopté par les pays membres du Conseil International pour l'Exploration de la Mer dans les échantillonnages de jeunes poissons de fond. On peut trouver dans GUEGUEN (1988) le plan du chalut et les détails de son gréement. Les pêches sont réalisées de jour et ont une durée standard de 30 minutes. Le support moyen des échantillons est évalué à 0,02 m.n² (POULARD, 1990). A bord, après chaque coup de chalut, le poisson est trié par espèce, mesuré, pesé, sexué. Les pièces calcifiées sont prélevées (otolithes) pour la détermination de l'âge. En effet la lecture de la succession des anneaux de croissance sur ces pièces peut permettre de déterminer l'âge exactement.

Or, la croissance du merlu se lit mal sur les otolithes (pièces calcifiées de l'oreille interne). La détermination de l'âge repose alors sur la séparation de l'histogramme polymodal des tailles, en populations statistiques unimodales. GUICHET (1988) a entrepris ce travail pour la constitution d'une clé taille-âge. Ce travail est délicat pour deux raisons. D'une part l'histogramme des tailles de la classe d'âge-0 peut présenter à l'automne plusieurs modes. En effet, la ponte serait fractionnée et l'arrivée des jeunes recrues sur les vasières du large se ferait par vagues successives de juillet à octobre. La croissance est hétérogène dans l'espace à cause des différences thermiques entre le nord et le sud, et elle est hétérogène dans le temps à cause du réchauffement estival. D'autre part, d'année en année, les modes varient. Les variations interannuelles peuvent atteindre plusieurs centimètres. A partir de 7 années d'observation, GUICHET (1988) donne un mode interannuel moyen à 13,8 cm pour la Grande Vasière et à 16,9 cm pour le sud du Golfe. La taille maximale définissant le groupe 0 pour les campagnes qui nous concernent a été fixée à 19 cm comme conseillé par Mr GUICHET (IFREMER-Lorient). Cette définition de l'âge est appliquée à l'ensemble du Golfe de Gascogne. Des poissons du groupe 1 sont comptabilisés dans nos données comme appartenant au groupe 0. Leur

proportion est supposée faible. La variable d'étude est une densité : c'est le nombre de merlu de longueur inférieure ou égale à 19 cm par coup de chalut standard.

B.2.2. Sélectivités

On donne d'abord sur la Figure 31 une représentation proportionnelle des données qui nous sert de carte descriptive de la régionalisation. Pour chaque représentation annuelle, le cercle de rayon maximal correspond à la valeur maximale de la campagne. Les représentations se comparent en densités relatives. Les zéros sont symbolisés par des étoiles. Les régionalisations des pourcentages se ressemblent énormément d'année en année, le poisson occupant toujours les mêmes lieux. Les zéros sont abondants devant la pointe occidentale de la Bretagne, dans le secteur A₁ et sur la bordure ouest du plateau. Dans le nord du Golfe de Gascogne, les 200 mètres sont les limites de la régionalisation du merlu-0, dans la partie sud du Golfe, le merlu-0 déborde jusqu'aux 400 mètres. Les densités relatives sont toujours plus importantes au milieu du plateau nord, sur la Grande Vasière. En 1988 et 1989, la distribution s'étale de part et d'autre de l'isobathe 100 mètres, alors qu'en 1987 et 1990 les densités relatives importantes se concentrent uniquement dans le nord de la Grande Vasière sous la mâchoire sud de la Bretagne. La partie sud du Golfe abrite aussi du merlu-0, apparemment sur toute la largeur du plateau landais. Les densités relatives sont plus faibles que dans le nord. En 1987 et 1990, le sud héberge une quantité relative très faible, alors qu'en 1988 et 1989, celle-ci est plus importante. Une zone de transition entre les concentrations du nord et du sud devant la Gironde n'abrite pendant ces quatre années jamais de merlu-0. Ainsi, on observe que la Grande Vasière du nord constitue une nourricerie principale, et celle du plateau landais une nourricerie secondaire dont l'importance relative varie. La biomasse peut être bien étalée sur tout son habitat (1988-1989) ou au contraire plus concentrée dans le nord de la Grande Vasière (1987-1990). Les zones occupées sont très semblables d'année en année. On étudie maintenant la stratégie d'occupation de l'espace en fonction de la quantité totale.

Les stations d'échantillonnage étant bien réparties sur l'ensemble du Golfe de Gascogne, on a jugé que la moyenne arithmétique des données était un indicateur raisonnable de la quantité totale.

Année	Nombre de chalutages	Moyenne arithmétique	Coefficient de variation : S/m
1987	131	44	2,41
1988	142	83	2,07
1989	134	57	1,97
1990	136	131	2,16

II.B. Application aux données

La gamme de variation des quantités totales est de 1 à 3 au cours de ces 4 années. Les dispersions relatives mesurées par le coefficient S/m sont similaires sauf pour 1987 où la dispersion relative est plus élevée, alors que cette année-là la classe d'âge est la moins importante.

Les courbes $Q(T)$ permettent de comparer plus précisément les dispersions. On calcule comme exposé en II.A les courbes $Q(T)$ et $P(T)$. Le calcul est effectué pour chaque campagne sur l'ensemble de son fichier. Les graphes sont donnés sur la Figure 32. On observe que les graphes $Q(T)$ ne se superposent pas à l'origine. Par contre en pourcentage, les graphes $P(T)$ montrent une superposition très nette. L'occupation du Golfe de Gascogne par le merlu-0 appartiendrait donc plutôt au schéma (H_2).

La géométrie de l'habitat est fixe, la densité du poisson en chaque point y est variable d'année en année mais varie proportionnellement à la quantité totale. Chaque année, sur l'ensemble du Golfe de Gascogne, le poisson occupe tout son habitat sur les mêmes lieux et s'y disperse toujours de la même façon en pourcentage de la quantité totale.

Pour une analyse plus fine, on sépare maintenant le Golfe de Gascogne en 3 zones. L'une d'elles est la zone périphérique au champ de présence du merlu-0. Elle est constituée par le secteur A_1 et la bordure ouest du plateau, soit la zone 200–800 m dans les secteurs A_2, B, C et 400–800 m dans le secteur D . Dans le champ de présence du merlu-0 (les secteurs A_2, B, C de 15–200 m et le secteur D de 15–400 m), on distingue deux zones, l'une dite nord constituée par les secteurs A_2, B, C de 15–200 m, l'autre dite sud constituée par le secteur D de 15–400 m. (Fig. 30).

Les statistiques élémentaires dans la zone nord sont les suivantes :

Année	Nombre de chalutages	Moyenne arithmétique	Coefficient de variation : S/m
1987	80	70	1,85
1988	81	109	1,79
1989	82	78	1,83
1990	82	188	1,82

On commente d'abord les moyennes puis les coefficients S/m . La gamme des variations de la quantité totale dans le nord du Golfe est de 1 à 2,7. L'année 1987 n'est pas très pauvre en merlu-0. En 1987 sur tout le Golfe de Gascogne, la quantité totale est faible car contrairement aux autres années, le sud héberge une quantité totale négligeable. En pondérant chaque zone par le nombre de ses échantillons, on peut estimer la contribution de la zone sud à la moyenne pondérée dans le champ de présence

du poisson: $m_p = \frac{n_n m_n + n_s m_s}{n_n + n_s}$, où n et m désignent le nombre d'échantillons et la moyenne arithmétique de la zone nord (indice n) et de la zone sud (indice s).

	1987	1988	1989	1990
m_p	54	101	73	159
m_s	6	79	59	79
$n_s m_s / (n_s + n_n) m_n$	1,5 %	20 %	21 %	13 %
n_s	27	28	29	30

En 1987, la nourricerie secondaire sud Gascogne est vide alors que pour les autres années, elle est remplie avec un niveau d'abondance peu variable. La partie sud héberge de 0 à 20 % de la quantité totale du merlu de l'ensemble du Golfe de Gascogne. Les rapports entre le remplissage de la zone sud et celui de la zone nord ne sont pas clairement déchiffrables au vu de ces seules données. On peut dire que les variations d'abondance sur la zone sud affectent aussi la quantité totale du Golfe de façon non négligeable.

Dans la partie nord du Golfe de Gascogne et selon le critère S/m, la dispersion relative est peu variable sur les 4 années d'observation. Pourtant, sur l'ensemble du Golfe, le coefficient S/m de 1987 est le plus élevé des 4 années. L'absence de merlu dans la partie sud génère un coefficient S/m global qui est élevé pour cette année là. Cependant, comme la zone sud a un poids moins élevé dans le jeu global, les courbes P(T) et Q(T) globales sont influencées par la dispersion dans le nord du Golfe. Les courbes P(T) calculées sur l'ensemble du golfe sans pondération des données se superposent donc.

La courbe P(T) est recalculée dans la zone nord uniquement restreinte aux strates de présence du merlu-0 (A₂, B, C entre 15 et 200 mètres). Le graphe P(T) est donné sur la Figure 33. On observe une très bonne superposition des courbes jusqu'au point (T=0,1, P=0,7). Chaque année, en pourcentage de la quantité totale, quelque soit celle-ci, les fortes densités se dispersent sur le plateau nord Gascogne toujours avec les mêmes gradients occupant 10 % de la surface de la zone A₂,B,C entre 15 et 200 mètres. Le lieu géographique de cette surface peut varier un peu.

En 1987 et 1990, il se situe dans le nord de la Grande Vasière autour de l'archipel de Glénan, alors qu'en 1988 et 1989, il se localise un peu plus au milieu de la Grande Vasière à la latitude de Noirmoutier. Pour les coupures plus faibles, c'est-à-dire pour les tonnages plus élevés (0,2 à 0,5), on observe une séparation des courbes P(T) où l'on note un regroupement des années 1988 et 1989 d'une part, 1987 et 1990 d'autre part. Les courbes P(T) en 1988 et 1989 sont plus sélectives que celles de 1987 et 1990 pour les plus grands tonnages (0,2 à 0,5). Comme la régionalisation des pourcentages est zonée (sa

forme grossière est assimilable à un dôme), les valeurs plus faibles se trouvent sur les bordures. La dispersion relative des valeurs au centre du dôme est la même d'année en année, mais les flancs du dôme ne le sont pas. Quand le centre du dôme est situé au milieu de la Grande Vasière vers la latitude de Noirmoutier, alors les flancs du dôme sont en pente douce. Quand le centre est situé sous la mâchoire sud de la Bretagne dans le nord de la Grande Vasière, le flanc du dôme est plus abrupt. Chaque situation a été observée deux fois. Le centre du dôme héberge 70 % de la quantité totale, les flancs en hébergent 30 %. La position du dôme, comme la pente des flancs, ne montrent pas de lien avec la quantité totale. Les années 1987 et 1990 présentent des régionalisations et des dispersions relatives très semblables alors que les quantités totales réalisent les deux extrêmes des 4 années d'observation.

Nous avançons l'hypothèse que la dispersion spatiale relative des valeurs les plus fortes au centre du dôme ($T=0,1$) est une caractéristique propre du poisson car elle est reproductible d'année en année. En revanche, la pente des flancs du dôme et la position géographique du centre seraient plutôt liées à un environnement variable ayant présenté deux états différents ; en particulier, on pense à la répartition spatiale du bourrelet froid.

En définitive, le merlu-0 répond à l'hypothèse (H_2) avec quelques fluctuations.

B.2.3. La composante spatiale invariante dans le temps

L'étude précédente a révélé que, d'une part, la distribution des valeurs est à peu près semblable d'année en année, au facteur multiplicatif près égal à la quantité totale, et que d'autre part, les lieux de présence du poisson ne varient pas beaucoup d'année en année. La régionalisation des pourcentages de la quantité totale recèle une composante invariante du temps. C'est cette composante que l'on se propose ici de décrire structurellement et de cartographier. On travaille sur la variable $p(x, t) = \frac{z(x, t)}{m(t)} 100$,

$z(x, t)$ est le nombre de merlu-0 par trait de chalut standard au point x l'année t et $m(t)$ est la moyenne des $z(x, t)$ pour l'année t . On étudie d'abord les variogrammes de $p(x, t)$ pour les 4 années. On observe qu'ils sont semblables quel que soit t . On propose alors une carte représentant l'espérance dans le temps des $p(x, t)$, $E_t[p(x, t)]$, qui correspond à la composante spatiale invariante du temps.

B.2.3.1. Analyse structurale

Pour calculer des distances, il nous faut d'abord transformer les coordonnées de latitude-longitude. Ici encore, 1° de longitude ne représente pas la même distance en fonction de la latitude. Le Golfe de Gascogne s'étend de $43^\circ 30' N$ à $48^\circ 30' N$, sur 5° . Comme la zone est ici assez importante, on a préféré transformer les latitudes par la fonction des latitudes croissantes de Mercator, ce qui nous permettra de cartographier le merlu-0 sur une carte marine. La projection de Mercator est telle que les loxodromies sur le globe sont rectifiées sur la carte : une courbe sur le globe coupant les méridiens à cap

constant est représentée sur la carte marine de Mercator par une droite, ce qui est particulièrement utile pour la navigation. Les longitudes sont exprimées en degrés et les latitudes sont transformées en coordonnées Mercator selon la fonction des latitudes croissantes suivante qui est celle du service de cartographie de l'IFREMER-Brest et qui nous a été aimablement mise à disposition par Dr C. KOUTSIKOPOULOS de

$$\text{l'IFREMER-Nantes : } y = \left[\text{Ln}\left[\text{tg}\left(\frac{\varphi}{2} + \frac{\pi}{4}\right)\right] + \frac{t}{2} \left[\text{Ln}\left(\frac{1 - t \sin \varphi}{1 + t \cos \varphi}\right)\right] \right] \frac{180}{\pi} \text{ où } \varphi \text{ est la}$$

latitude exprimée en radians et t est une constante ($t = \sqrt{6,72267 \cdot 10^{-3}}$). On appellera l'unité de distance, l'unité Mercator notée u.m. . 0,25 u.m correspondent en moyenne sur l'ensemble du Golfe de Gascogne à 10 m.n : 0,25 u.m \approx 10 m.n.

L'analyse structurale est menée sur la zone de présence du merlu-0, c'est-à-dire dans les secteurs A₂,B,C de 15 à 200 mètres et D de 15 à 400 mètres. Bien que la description de la régionalisation en B.2.2. révèle un aspect non-stationnaire en dôme de la régionalisation, on calcule des variogrammes afin d'inférer un modèle quasi-stationnaire suffisant pour la cartographie.

On a calculé les variogrammes des pourcentages $p(x,t)$ pour chacune des 4 années. Les deux directions principales d'anisotropie sont déterminées par la géométrie du plateau continental nord-Gascogne. Ainsi, on imposera dans la partie sud, moins représentée dans le jeu de données, une structure moyenne sur tout le Golfe et fortement influencée par les orientations du nord. Les deux directions principales sont la direction moyenne des isobathes à 125° comptée à partir de l'axe des x et la direction perpendiculaire côte-large à 35° comptée toujours à partir de l'axe des x. Sur la Figure 34, on a superposé les variogrammes des $p(x,t)$ pour les 4 années dans les deux directions d'anisotropie maximale. On observe une bonne superposition des variogrammes jusqu'à la distance 2 u.m dans la direction 35° et pour toutes distances dans la direction 125°. On considère donc que la structure spatiale des pourcentages $p(x,t)$ est invariante du temps. On a calculé de ce fait la moyenne des 4 variogrammes pour chaque distance dans chaque direction. Ce variogramme représente la structure moyenne interannuelle qui, dans notre analyse, est invariante dans le temps (Figure 35).

Le variogramme temporel moyen (Figure 35) grimpe plus vite et plus haut dans la direction côte-large (35°) que dans la direction des isobathes (125°). Dans la direction 125°, un palier est atteint autour de la variance entre 1 et 2 u.m. Un comportement stationnaire est observé dans la direction 125° alors que dans la direction 35°, le variogramme croît très rapidement, traverse la variance et chute ensuite. Ce comportement est interprété comme le fait du dôme de poisson, dans cette direction. Un modèle quasi-stationnaire est ajusté avec anisotropie géométrique et zonale. Le modèle est une somme de deux structures comme proposé par JOURNEL et al (1978) dans une telle situation. On a choisi deux schémas sphériques sans effet de pépite. Dans l'anisotropie géométrique, la portée est anisotrope, le palier est isotrope. Les portées dans les différentes directions se répartissent sur une ellipse. Son grand axe a ici pour

direction 125° , son petit axe 35° . Dans le modèle ajusté, le rapport des portées du grand axe sur le petit axe est de 1,3. Le palier est égal à $0,55 \cdot 10^5$. Dans l'anisotropie zonale, le palier est anisotrope. Ici, la variance dans la direction côte-large est plus importante que le long des isobathes. Les transitions côte-large sont non seulement plus rapides, elles sont aussi plus hétérogènes que dans la direction des isobathes. On rajoute donc au modèle précédent un schéma sphérique uniquement dans la portion de l'espace autour de la direction côte-large : $35^\circ \pm 22,5^\circ$. La portée de ce modèle est courte, 0,3 u.m (12 m.n), et renforce la croissance rapide du variogramme dans cette direction. Le palier est pris à $0,22 \cdot 10^5$. La variance dans la direction côte-large est 1,4 fois plus grande que dans la direction des isobathes. Le modèle n'est valide que pour des distances inférieures à 1,25 u.m (50 m.n).

B.2.3.2. Le modèle de fonction aléatoire

Les poids $p(x,t)$ montrent donc une même dispersion des valeurs et une même structure spatiale pour les 4 années d'observation, ainsi que des répartitions spatiales très similaires. Ces caractéristiques nous conduisent à interpréter la régionalisation du merlu-0 comme la réalisation chaque année d'une même fonction aléatoire quasi-stationnaire $P(x)$. L'effet année se marque uniquement par une proportionnalité en

chaque point x à la quantité totale de l'année : $\frac{Z(x, t)}{m(t)} = P(x)$. Les majuscules désignent

des fonctions aléatoires non une réalisation.

Suivant MATHERON (1970), on appelle dérive la moyenne des réalisations de $P(x)$ au point x : $pm(x) = E[P(x)]$. D'après les représentations proportionnelles (Figure 31), on n'a pas à regrouper différentes formes de dérive correspondant à différents niveaux d'abondance de la quantité totale, même si on l'a observé une fois que l'abondance globale est faible quand la zone sud est vide. Dans notre modèle, la dérive $pm(x)$ n'est pas fonction du temps. Elle représente la composante spatiale invariante du temps. Elle n'est pas nécessairement une composante spatiale régulière de grande échelle. Elle représente ce qui, d'année en année, se répète identiquement. Elle correspond biologiquement à l'équilibre adaptatif du poisson avec son environnement. Il apparaît donc que dans notre cas, sur un poisson démersal de stratégie (H_2), il est nécessaire de filtrer la composante spatiale invariante dans le temps (invariante de la quantité totale) avant que d'aborder la question du lien entre le comportement local et l'abondance globale. Ce seraient plutôt les écarts à la dérive, la variance de P en x , qui pourraient présenter un lien à la quantité totale.

B.2.3.3. *Cartographies*

On se propose de cartographier la composante invariante du temps. Chaque année, les échantillons ne sont pas prélevés aux mêmes stations. On reconstitue donc aux nœuds d'une grille régulière, la régionalisation par krigeage pour chacune des 4 années. On a inféré pour chaque année t le modèle spécifique de variogramme de la densité $Z(x,t)$ (nombre de poissons par coup de chalut). On a effectué alors un krigeage de bloc pour chaque année. Puis on a calculé les pourcentages de masses suivants :

$$P_V^K(x, t) = \frac{Z_V^K(x, t)}{\sum_i Z_V^K(x_i, t)} \times 100 \text{ où } Z_V^K(x, t) \text{ est l'estimateur par krigeage du bloc } V \text{ en } x.$$

Enfin, on a calculé en chaque x , la moyenne arithmétique des poids $P_V^K(x, t)$ sur les 4 années : $pm^*(x) = \frac{1}{n_t} \sum_t P_V^K(x, t)$. Cette dernière carte estime la composante spatiale invariante du temps. La prise de moyenne correspond à un filtrage. En effet, on peut écrire : $P_V^K(x, t) = pm(x) + R(x,t)$ où $pm(x)$ est la dérive que l'on cherche à estimer. La prise de moyenne filtre les résidus $R(x,t)$. La dérive $pm(x)$ est quasi-stationnaire.

Les variogrammes pour chaque année sont présentés sur la Figure 36. Ils sont calculés dans les deux directions principales d'anisotropies et présentent tous la même allure à l'exception de celui de 1987 qui est isotrope. En effet, cette année là, les fortes valeurs dans le nord de la Grande Vasière s'étendent largement dans la direction parallèle à la côte qui est E-W. Des schémas gigognes sphériques modélisent les variogrammes sur lesquels on a reconnu des anisotropies zonale et géométrique. Les paramètres des modèles sont consignés dans le Tableau 1 (p. 34 des Figures).

Les données ne sont pas implantées sur une grille à maille régulière si bien qu'il est impossible de respecter une configuration donnée de krigeage. Pour la cartographie, on a utilisé la procédure de recherche de voisinage implémentée dans le logiciel Bluepack (GALLI et al, 1991). Autour de chaque point à estimer, un cercle est tracé et l'estimation est une combinaison linéaire des échantillons situés à l'intérieur du cercle. Le nombre de points intérieurs et le diamètre du cercle sont des paramètres que l'on peut choisir et faire varier. Les modèles quasi-stationnaires inférés imposent de n'utiliser que les distances inférieures à 1 u.m (40 m.n). La zone krigée a pour limites à l'Est la côte, au Nord la latitude d'Ouessant, à l'Ouest l'isobathe 1000 mètre et au Sud le Gouf de Cap Breton. Ce polygone est discrétisé en blocs v de taille $0,5 \times 0,5$ u.m² (20×20 m.n²). On estime par krigeage ordinaire la moyenne Z_v dans chacun de ces blocs v . La carte aura donc un aspect lisse ce qui n'est pas un désavantage pour la représentation de l'aspect moyen. Le krigeage est effectué en voisinage glissant avec au minimum six échantillons dans le voisinage dont le diamètre est fixé à 1 u.m (40 m.n). Cette taille de voisinage est un peu grande pour la partie sud du Golfe de Gascogne.

L'estimateur de krigeage s'écrit dans chaque voisinage : $Z_V^K(x) = \sum_a \lambda_a Z(x_a)$.

Il est non biaisé si et seulement si: $E(Z_V - Z_V^K(x)) = 0$. Dans chaque voisinage la moyenne est inconnue. La condition de non biais conduit à imposer la contrainte suivante sur les poids λ_a : $\sum_a \lambda_a = 1$.

La variance de l'estimation s'écrit avec la covariance σ comme suit:

$$\text{var} = E((Z_V - Z_V^*)^2) = \sigma_{vv} - 2 \sum_a \lambda_a \sigma_{av} + \sum_a \sum_\beta \lambda_a \lambda_\beta \sigma_{a\beta}$$

avec $\sigma_{a\beta} = \sigma(|x_a - x_\beta|)$. C'est la formule numérotée (I.2) dans la partie I.A. où figurent maintenant les poids λ_a . Le minimum de var est obtenu en annulant ses dérivés partielles en λ_a .

La minimisation de var sous la contrainte de non biais $\sum_a \lambda_a = 1$ par le formalisme de Lagrange conduit au système de krigeage (ordinaire) dont les poids λ_a sont solution :

$$\begin{cases} \sum_\beta \lambda_\beta \sigma_{a\beta} = \sigma_{av} + \mu, \quad \forall a \quad \text{dans le voisinage} \\ \sum_a \lambda_a = 1 \end{cases}$$

(μ est le paramètre de Lagrange).

Les systèmes de krigeage ponctuel et de bloc se distinguent par leur membre de droite: le terme σ_{ax_0} du krigeage ponctuel, covariance entre l'échantillon en x_α et le point x_0 à estimer, est remplacé dans le krigeage de bloc par le terme σ_{av} qui est la covariance moyenne entre l'échantillon x_α et tous les points du blocs v à estimer. La condition de non biais $\sum_a \lambda_a = 1$ cale l'estimateur autour de la moyenne des échantillons dans le voisinage.

Les modèles de variogrammes sont gigognes : il y a deux structures. Pour 1987, on filtre l'effet de pépite lors de la cartographie en ne considérant pas cette composante pour le calcul de σ_{av} dans le membre de droite du système de krigeage. Pour les autres modèles, on considère les deux structures. A cause de l'anisotropie géométrique, le calcul de σ_{av} et $\sigma_{a\beta}$ requiert une transformation des coordonnées afin de se ramener à un espace isotrope. Une première rotation d'angle φ (ici $\varphi = +55^\circ$) est effectuée afin de faire coïncider le grand axe de l'ellipse des portées avec l'axe des x , puis les y sont alors multipliés par le rapport (> 1) des portées du grand axe sur le petit, de façon à transformer l'ellipse en cercle, enfin une rotation d'angle $-\varphi$ est appliquée pour revenir aux orientations initiales. Le variogramme calculé pour les distances ainsi transformées est

isotrope. Le variogramme modélisant l'anisotropie zonale n'est utilisé que dans la portion de l'espace $35^\circ \pm 22,5^\circ$.

Sur la Figure 37, on donne les cartes obtenues pour les 4 années, de la distribution de la densité du merlu-0. Sur la Figure 38, on a représenté la carte des pourcentages de biomasse $P_V^K(x)$ qui est l'estimation de la composante spatiale moyenne dans le temps. Sur cette carte moyenne, les blocs au sud de la Gironde ($y \leq 51$ u.m) représentent 13% de la quantité totale sur l'ensemble du Golfe de Gascogne. A partir de cette carte, en sélectionnant les blocs les plus riches, on peut représenter sur la Figure 39 l'aire sur laquelle se distribue en moyenne 80% de la classe d'âge. La coupure sur krigeage donne une estimation biaisée de la surface $A_p (Z(x) \geq z_p)$. Toutefois cette représentation nous suffit car on ne cherche qu'à visualiser l'emplacement moyen de la zone de densité maximale. Le contour de cette surface est superposé au fond de carte de la Grande Vasière (Figure 39). La nourricerie nord est bien centrée en moyenne sur la Grande Vasière de part et d'autre de l'isobathe 100 mètres. La nourricerie sud est en moyenne située entre Arcachon et le Gouf de Cap Breton de la côte jusqu'aux 200 mètres.

B.3. Les œufs de sole de 1986 à 1990

Dans le Golfe de Gascogne, la sole est l'espèce *Solea solea*, L. La biologie spatiale y est bien connue (DARDIGNAC, 1988 ; KOUTSIKOPOULOS, 1991). Elle est présente dans tout le Golfe mais plus particulièrement entre Quiberon et la Gironde. Les adultes (individus sexuellement mûrs) opèrent saisonnièrement une migration côte-large. En été, on les trouve sur les petits fonds côtiers inférieurs à 10 mètres ; en hiver, on les trouve sur la bande des sables graveleux propres entre 30 et 120 mètres. Contrairement au scénario de la Mer du Nord, dans le Golfe de Gascogne c'est au large qu'a lieu la ponte. Elle a lieu de janvier à mars sur les sables graveleux. Des marquages ont montré que les soles sur les zones de frai sont issues des différentes zones côtières : il y a donc mélange (KOUTSIKOPOULOS, 1991). La sole du Golfe de Gascogne est vue comme une population homogène qui est identifiée à une unité de reproduction : elle constitue un stock unitaire (ANON, 1991). Les principales régions de ponte se situent au large de l'estuaire de la Loire, entre Belle-Ile et l'Ile d'Yeu, et au large des Pertuis Charentais (ARBAULT et al 1984). Dès juillet, les eaux côtières sont colonisées par les larves âgées et les tout petits juvéniles (classe d'âge 0). Ces poissons resteront à la côte pendant leurs deux premières années, après quoi ils entreprendront les migrations saisonnières adultes. Les zones côtières sablo-vaseuses constituent les nourriceries de la sole.

Par l'analyse des débarquements de la pêche côtière et du large, la sole du Golfe de Gascogne peut être considérée comme un stock à recrutement stable marqué par quelques fortes classes d'âge (ANON, 1991b). Une telle dynamique est assez générale chez les poissons plats et a déjà été décrite en Mer du Nord (VEEN, 1978). Le stock de sole du Golfe de Gascogne a été retenu en France comme une des populations cibles dans les études sur le recrutement. Le programme Soldet (Déterminisme du Recrutement de la

Sole) de l'IFREMER–Nantes de 1985 à 1989 (Mr Y. DESAUNAY) avait pour but d'analyser les mécanismes intervenant dans le recrutement de la sole. Le programme devait permettre d'améliorer la connaissance de la frayère par le suivi de la distribution spatiale des œufs et larves planctoniques de la sole (description des régionalisations, relations avec l'environnement, variations intra et interannuelles), ainsi que l'étude de la colonisation des zones côtières par les larves âgées en métamorphose, lors du passage de la vie planctonique à la vie benthique. Nous nous sommes intéressé à la régionalisation dans l'espace et le temps des œufs dans leur premier stade de développement. A la température moyenne de l'eau de mer dans la zone d'étude (9°C), la durée du premier stade ovulaire est de quelques jours (FONDS, 1979), soit une durée d'un même ordre de grandeur que celle d'une campagne d'échantillonnage. Notre étude sera celle de la ponte instantanée.

B.3.1. Les données

B.3.1.1. Echantillonnage

Le secteur maritime compris entre Belle-Ile et l'Ile d'Yeu depuis les sondes de 30 mètres jusqu'à celles de 120 mètres a été reconnu comme un centre très actif de ponte (ARBAULT et al, 1984). Il a fait l'objet de 13 campagnes d'échantillonnage de 1986 à 1989. A chaque campagne, un même réseau de stations est visité (Figure 40). Le réseau est presque régulier et comprend 35 stations dont 11 sont répliquées. Sur ces stations, après un premier prélèvement, le bateau revient sur sa position et un second prélèvement est effectué. Ces 11 stations répliquées nous permettront d'évaluer un effet de pépité (variance intra-station).

Un prélèvement de plancton s'opère ainsi : le bateau faisant route à vitesse constante (2 nœuds), le filet à plancton est descendu depuis la surface jusqu'au fond puis est remonté jusqu'à la surface. La profondeur du filet dans l'eau est contrôlée par un système sonar. Les œufs de sole se dispersent dans toute la colonne d'eau de façon assez homogène et uniforme, comme l'ont révélé des échantillonnages fins sur la verticale (KOUTSIKOPOULOS, 1991). Un tel prélèvement en V sur la verticale assure un échantillonnage de toute la colonne d'eau et permet le calcul d'une densité moyenne (nb/10 m³). Sur les petits fonds (20–30 m), le prélèvement est effectué en W : après une première descente et remontée, un second V est effectué. En effet, cette procédure est imposée par la grande dilution des œufs et larves de sole dans le plancton qui requiert un échantillonnage de volumes importants.

L'engin de pêche utilisé a été un filet de type Bongo en 1986 et 1987 ; un autre engin à ouverture carrée, plus maniable dans le mauvais temps de février–mars, a été utilisé en 1988 et 1989. Les mailles du filet à plancton sont de 500 µm. Dans la gueule du filet, deux volucompteurs à hélice sont grées afin de mesurer le volume filtré lors de chaque pêche. Des détails plus précis peuvent être trouvés dans KOUTSIKOPOULOS (1991). A bord, après le trait de filet, la récolte de plancton est fixée. Ce n'est qu'au

laboratoire que le dépouillement a lieu. Chaque échantillon a été dépouillé dans sa totalité (pas de sous-échantillonnage). Les œufs et larves de sole sont triés et dénombrés par stade de développement.

B.3.1.2. Variable étudiée

Le stade 1 du développement ovulaire sur lequel on travaille correspond aux stades Ia et Ib de RILEY (1974). Dans nos données, une erreur de détermination du stade de développement est présente mais n'a pas été explicitée. Elle est supposée faible.

La gamme des profondeurs est petite et s'étale de 30 à 100 m. L'échantillonnage en W sur les petits fonds nous permet de considérer que le support des échantillons (volume d'eau de mer filtré par le filet) ne varie pas en fonction de la profondeur. Le volume moyen standard est de 544 m³ (Figure 41).

On a donc le choix de travailler sur l'une des deux variables : la densité (nb/10 m³) ou l'accumulation (densité x sonde : nb/10m² de surface). Ces deux variables sont toujours bien corrélées. Comme nous le verrons, les densités importantes se trouvent toujours au centre d'un dôme dont le grand axe suit la sonde 60 m. Les différences entre grandes et petites sondes sont faibles comparées aux différences entre faibles densités des bords et fortes densités centrales. Il en résulte que les différences de sonde ne "réussissent" pas à rendre les accumulations très différentes des densités.

Nous avons retenu pour l'étude la variable densité : nb. œufs de stade 1 /10 m³.

B.3.2. Sélectivités

Dans le temps, si l'on reporte en ordonnées les quantités totales pondues, la ponte se représente sur une courbe en cloche : début, pleine ponte, fin de ponte. Chaque échantillonnage du réseau se situe à un moment de cette courbe (Tableau 2 , p.39 des Figures). On sépare les moments de pleine ponte (moyenne arithmétique par campagne élevée ≥ 1) de ceux de début et fin de ponte (moyenne arithmétique faible $\leq 0,5$).

On donne d'abord sur les Figures 42 et 43 les représentations proportionnelles des données correspondant aux moments de pleine ponte. Elles font office de cartes qualitatives. Les représentations se comparent en densité relative car les valeurs de chaque campagne sont rapportées à la valeur maximale de la campagne. Ici encore, les régionalisations des pourcentages se ressemblent énormément de campagne à campagne, à l'intérieur d'une même année comme entre des années différentes. La distribution spatiale des œufs présente l'aspect d'un dôme elliptique de grand axe parallèle aux isobathes. Il est toujours centré à peu près au même endroit, au milieu de la zone et de part et d'autre de l'isobathe 60 mètres.

Le réseau de prélèvement étant presque régulier, on a pris la moyenne arithmétique comme estimateur de la quantité totale.

Sur la Figure 44, on présente les courbes Q(T) et P(T). Les courbes P(T) de chacune des 13 campagnes sont superposées sur la Figure 44a. Sur ces courbes, on observe

une individualisation des moments de début et fin de ponte par rapport à ceux de pleine ponte. Le début de la ponte n'a été échantillonné qu'une fois, par la campagne codée 90. Sa courbe P(T) est très sélective car toute la ponte est concentrée sur un petit nombre de stations. En fin de ponte, les dispersions sont soit plus concentrées ou soit du même ordre de concentration que pendant la pleine ponte. La grande différence en 1987 entre les campagnes 71 et 72 montre la grande variabilité des dispersions spatiales de fin de ponte. Soit la régionalisation des faibles densités est très dynamique spatialement, soit les très faibles densités sont échantillonnées avec de grandes fluctuations. Par contre, pendant la pleine ponte, on note une superposition des courbes bien marquée montrant une grande reproductibilité des dispersions spatiales. Ainsi, en début et fin de ponte, celle-ci peut être concentrée sur une petite partie de la frayère alors que pendant la pleine ponte, on observe que tout l'habitat est toujours occupé. On négligera les début et fin de ponte. D'abord car les quantités pondues sont très faibles ; ensuite pour des raisons relatives à la physiologie de la ponte (LE BEC, 1985). En effet, il nous semble raisonnable de considérer qu'une population de soles adultes constituée de plusieurs classes d'âge pond, dans son ensemble, régulièrement au cours d'une pleine saison de ponte plutôt que sporadiquement. On se représente donc la ponte comme un phénomène en tout ou rien, pas de ponte ou ponte régulière. Dans ce schéma, un modèle spatial de pleine ponte est le seul dont nous ayons besoin. Dans ce schéma, un processus aléatoire stationnaire paraît raisonnable pour représenter l'évolution temporelle de la quantité totale pondue à chaque instant sur la frayère, pendant la période de pleine ponte.

Les courbes des Figures 44 b et c sont celles des moments de pleine ponte. Elles sont compatibles avec l'hypothèse (H₂). Pendant la pleine ponte, les œufs de stade 1 occupent à chaque instant chaque année tout leur habitat avec les mêmes gradients relativement à la quantité totale, sur les mêmes lieux. La situation présente est analogue à celle du merlu-0. Devant la répétition dans le temps d'un même dôme de ponte, notre idée est de construire un modèle où la répartition spatiale des masses est invariante dans le temps. Soient $z(x,t)$ la densité en x à l'instant t et $m_q(t)$ la moyenne spatiale à l'instant t .

$\frac{z(x, t_1)}{m_q(t_1)}$ et $\frac{z(x, t_2)}{m_q(t_2)}$ représentent les pourcentages de biomasses en (x, t_1) et (x, t_2) . On

considère qu'ils ont même histogramme et même structure spatiale, indépendamment du

temps. $\frac{z(x, t_1)}{m_q(t_1)}$ et $\frac{z(x, t_2)}{m_q(t_2)}$ sont interprétés comme deux réalisations d'un même

phénomène, la ponte. Ainsi la distribution spatiale des biomasses est invariante dans le temps, deux instants différents par la valeur de leur quantité totale dont le modèle ne sait pas expliquer les différences. On développe ce modèle dans la partie D.

On a observé que la ponte de la sole comme le merlu-0 répond de la même façon dans l'espace aux fluctuations de la quantité totale : la régionalisation des pourcentages ne varie pas. Les fluctuations de 1 à 3 dans les quantités totales ne sont peut être pas suffisantes pour modifier qualitativement le comportement spatial local. Les variations

dans les quantités totales observées sont-elles significatives ? On aborde maintenant cette question.

C. DIFFERENCES SIGNIFICATIVES ENTRE LES ESTIMATIONS GLOBALES

On se propose ici de comparer les fluctuations inter-annuelles (i.e. inter-campagnes) de la moyenne aux erreurs d'estimation de celle-ci, pour chaque campagne. Si les erreurs d'estimation sont plus importantes que les fluctuations inter-campagnes, alors on ne pourra pas dire que les différences observées dans les abondances globales sont significatives. On calcule la variance d'estimation de la moyenne à chaque campagne pour le merlu-0 puis pour les œufs de sole. On construit aussi les fourchettes d'erreur autour des estimations globales de chaque campagne et on observe si ces fourchettes se recoupent. On considère la fourchette conventionnelle gaussienne à 95 %, i.e. $\pm 2 \sigma_{\text{estim}}$.

C.1. Le merlu-0

L'estimation est effectuée sur le polygone V de présence du poisson. Ce polygone est constitué de la réunion des 37 strates suivantes. Dans les secteurs latitudinaux A₂, B₁, B₂, B₃ et C, les strates de profondeur sont définies par les isobathes suivantes : 15-30 mètres, 30-80 mètres, 80-120 mètres, 120-160 mètres, 160-200 mètres. Dans les secteurs D₁ et D₂, aux 5 strates de profondeurs précédentes, on a rajouté la strate 200-400 mètres. Les surfaces v_i de chaque strate sont calculées par discrétisation par la maille 0,1 x 0,1 u.m² (4 x 4 m.n²) : c'est le nombre de carrés dont le centre est intérieur au polygone. Le poids w_i de la strate i est : $w_i = \frac{v_i}{\sum_i v_i}$. La moyenne sur V et la variance d'estimation sont calculées comme si le plan d'échantillonnage était aléatoire stratifié.

L'estimateur de la moyenne Z_V sur le polygone $V = \sum_i v_i$ est : $Z_V^* = \sum_i w_i Z_{v_i}^*$

où $Z_{v_i}^*$ est la moyenne des n_i échantillons dans la strate i. $Z_{v_i}^* = \frac{1}{n_i} \sum_j z_{ij}$.

Pour calculer la variance d'estimation $\text{var}(Z_V - Z_V^*)$ on cherche à se ramener à l'une des situations exposées dans la partie I.A.2., afin de définir quel variogramme utiliser dans les strates. On spécifie donc d'abord le rapport qui existe entre stratification et régionalisation. Ici, la stratification ne correspond pas à la régionalisation. On a calculé pour l'année 1988 deux types de variogrammes: l'un dit intra-strate est calculé avec les paires de points appartenant à une même strate, l'autre dit inter-strates est calculé avec les paires de points à cheval sur deux strates différentes. Si régionalisation et stratification coïncidaient, le variogramme inter-strates devrait systématiquement être au dessus du variogramme intra-strate, aux petites distances.

Les valeurs des variogrammes inter-strates et intra-strate sont consignées dans le tableau suivant:

h	direction 35		direction 125	
	$\gamma(h)$	Nombre of couples	$\gamma(h)$	Nombre of couples
0.25	$\gamma_{kk} = 16283.$	28	$\gamma_{kk} = 3972.$	36
	$\gamma_{kq} = 20846.$	59	$\gamma_{kk} = 4196.$	47
0.50	$\gamma_{kk} = 51593.$	18	$\gamma_{kk} = 20112.$	34
	$\gamma_{kq} = 29305.$	102	$\gamma_{kk} = 13288.$	124
0.75	$\gamma_{kk} = 81743.$	22	$\gamma_{kk} = 28872.$	25
	$\gamma_{kq} = 40933.$	134	$\gamma_{kk} = 15609.$	204
1.00	$\gamma_{kk} = 82066.$	5	$\gamma_{kk} = 17824.$	18
	$\gamma_{kq} = 51177.$	206	$\gamma_{kk} = 22303.$	239

Variogrammes intra-strate (γ_{kk}) et inter-strates (γ_{kq}) pour l'échantillonnage stratifié d'automne 1988. Les distances h sont exprimées en unités mercator (0.25 u.m. \approx 10m.n.)

Un test de significativité n'est pas envisagé car on ne connaît pas la distribution statistique des différences carrées $(Z(x_\alpha) - Z(x_\beta))^2$. Mais on lit que les valeurs et les fluctuations des deux variogrammes sont du même ordre de grandeur. Ainsi, deux points séparés d'une même distance h sont autant corrélés quand ils sont dans la même strate que lorsqu'ils sont chacun dans une strate différente. Les corrélations intra-strates sont identiques aux corrélations inter-strates pour les distances inférieure à 1 u.m., i.e. inférieures à la portée. On dispose d'un petit nombre de données par strate (8 au maximum). On utilisera donc dans chaque strate le variogramme moyen global. Les dimensions des strates font intervenir des distances intra-strate inférieures à la limite de validité des modèles de variogrammes quasi-stationnaires inférés pour chaque année.

De plus, il existe un effet proportionnel entre la variance par strate et la moyenne par strate dont il est important de tenir compte car les strates pauvres même si de géométrie différente des strates riches, étant moins variables que celles-ci, contribuent moins à la variance d'estimation. On se trouve donc dans le cas de la formule (I.4.d) proposée dans la partie I.A.2. On fait l'hypothèse que dans chaque strate la structure spatiale se laisse décrire par un variogramme proportionnel au variogramme moyen global.

II.C. Différences significatives entre les estimations globales

On calcule donc la variance d'estimation $\text{var}(Z_V - Z_V^*)$ selon la formule géostatistique proposée:

$$\sigma_g^2 = \sum_i w_i^2 \frac{k_i}{n_i} \bar{\gamma}_m(v_i, v_i)$$

$$\text{avec } k_i^* = \frac{S_i^2}{\bar{\gamma}_m(v_i, v_i)} \quad \text{et} \quad S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Z_{ij} - Z_{vi}^*)^2$$

où S_i^2 est calculée expérimentalement et où le rapport k_i est donné par la régression de k_i^* sur Z_{vi}^* (espérance conditionnelle). On a vu que cette manière de procéder conduisait à modéliser un effet proportionnel qui soit dégagé de l'influence géométrique des strates. La variance résiduelle de la régression s'en trouve réduite.

Sur la Figure 45, on donne les droites de régression du logarithme de k_i^* sur le logarithme des moyennes des strates Z_{vi}^* . La droite de régression conduit à modéliser l'effet proportionnel par une loi puissance de la forme: $k_i = b Z_{vi}^{*a}$. L'exposant a est la pente de la droite de régression en log. Afin d'assurer le non-biais de l'estimateur $k_i = b Z_{vi}^{*a}$, on a pris b égale au rapport des moyennes $\bar{k}_i^* / \overline{Z_{vi}^{*a}}$.

année coefficients	1987	1988	1989	1990
a	1.769	1.757	1.565	1.691
b	5.72 E-4	1.08 E-4	4.17 E-4	7.51 E-5

Les variogrammes, γ_m , utilisés sont ceux déjà calculés et modélisés pour la cartographie (section B.2.3.3.). Chacun est spécifique de son année. Les modèles sont donnés sur la Figure 36 et sur le Tableau 1 (p.34 des Figures). Pour chaque année, le variogramme moyen dans la strate i , $\bar{\gamma}_m(v_i, v_i) = \frac{1}{v_i^2} \int_{v_i} \int_{v_i} \gamma_m(|x-y|) dx dy$, est calculé par intégration numérique sur la maille $0,1 \times 0,1 \text{ u.m}^2$ ($4 \times 4 \text{ m.n}^2$) selon l'algorithme proposé par JOURNEL et al (1978) : l'une des extrémité du vecteur $|x-y|$ décrit le réseau des nœuds de la grille de discrétisation, pendant que l'autre décrit indépendamment l'ensemble des points dont chacun est tiré aléatoirement dans chaque maille de discrétisation. En évitant que x et y ne soient au même endroit, ce procédé aléatoire stratifié évite de moyenner un nombre trop important de zéros, ce qui réduirait la valeur du $\bar{\gamma}_m(v_i, v_i)$. Ayant digitalisé le contour de chaque strate, on a utilisé l'algorithme donné par ANDERSON (1976) pour sélectionner d'une part les échantillons intérieurs et d'autre part, les centres de maille de discrétisation intérieurs aux limites des strates. En effet, ces polygones ne sont pas forcément convexes. Lors du calcul, les anisotropies sont traitées comme déjà expliqué dans la section B.2.3.3 .

On trouve :

	Nb de strates où $n_i < 2$	Z_V^*	σ_g/Z_V^*
1987	2	46	12.5 %
1988	2	68	13.3 %
1989	2	90	8.9 %
1990	4	125	10.6 %

Dans le cas présent, la régression du logarithme de k_i^* sur le logarithme de Z_{vi}^* n'a pas une variance résiduelle plus faible que celle du logarithme de S_i sur le logarithme de Z_{vi}^* . La raison en est que les strates riches sont au centre du plateau et toutes de grande taille tandis que les strates pauvres sont à la côte ou au large et toutes de petite taille (excepté dans le secteur A1). Dans chaque classe de moyenne, la variabilité de la taille des strates est faible. L'intérêt de la formule (I.4.d) ne peut donc pas s'exprimer pleinement dans le cas présent.

Le coefficient de variation ($\frac{S}{\sqrt{n}}/\overline{Z_V^*}$) des fluctuations inter-annuelles de la moyenne est de l'ordre de 21%. Il est donc bien supérieur à l'écart-type d'estimation. Les observations sont donc suffisamment précises pour que l'on puisse distinguer effectivement une différence numérique, d'année en année, entre les quantités totales des classes d'âge-0.

Les fourchettes d'erreurs gaussiennes conventionnelles calculées avec σ_g^2 sont :

	$Z_V^* - 2\sigma_g$	Z_V^*	$Z_V^* + 2\sigma_g$
1987	35	46	58
1988	50	68	86
1989	74	90	106
1990	99	125	152

D'année en année, les fourchettes ne se recoupent que du bout de la pointe.

C.2. Les œufs de sole

La moyenne spatiale à chaque campagne est estimée par la moyenne arithmétique des données, Z_V^* . La variance de cette estimation à chaque campagne est calculée dans le cadre du modèle structural de pleine ponte, dans la partie II.D.7 suivante.

Comme on le verra, on trouve une précision relative égale à : $\frac{\sigma_g}{Z_V^*} = 16,6\%$. Les

fourchettes d'erreurs gaussiennes conventionnelles sont :

	$Z_V^* - 2\sigma_g$	Z_V^*	$Z_V^* + 2\sigma_g$
1986, 1	1,5	2,2	2,9
1986, 2	1,9	2,8	3,7
1986, 3	1,9	2,8	3,7
1986, 4	0,9	1,4	1,9
1987	2,1	3,1	4,1
1988	1,7	2,6	3,5
1989	1,7	2,5	3,3

Les fourchettes d'erreur sont très grandes et se recoupent largement. Le coefficient de variation ($\frac{S}{\sqrt{n}} / \overline{Z_V^*}$) des fluctuations inter-annuelles de la moyenne est de l'ordre de 7%. Il est donc largement inférieur à l'écart-type relatif de l'estimation de la moyenne à chaque campagne ($\frac{\sigma_g}{Z_V^*} = 16,6\%$). Les observations ne sont pas suffisamment précises pour que l'on puisse distinguer effectivement une différence numérique entre les quantités totales.

C.3. Discussion

Les variations des quantités totales observées ont lieu sur une échelle arithmétique. Ces différences sont-elles réellement significatives pour la population?

Pour les oeufs de sole, ces variations ne peuvent pas être distinguées des erreurs d'estimation. L'hypothèse (H2) pour l'occupation de l'espace est donc logique. Pour le merlu-0, les variations inter-annuelles de la force du recrutement correspondent à des variations naturelles réelles. L'hypothèse (H2) pour l'occupation de l'espace a donc une signification biologique.

Dans le cas du merlu du Golfe de Gascogne, on a rencontré l'hypothèse (H2) pour des variations de 1 à 3 dans les effectifs de la classe d'âge 0. Celles-ci se présentent sur une échelle arithmétique. Bien que non observées, si des variations des effectifs sur une échelle logarithmique sont possibles (l'analyse des débarquements des pêches pourrait donner des indications), observerait-on toujours l'hypothèse (H2) d'occupation de l'espace? Comme les très fortes classes d'âge sont assez rares, une telle observation ne peut se faire que sur des bases de données longues de l'ordre de plusieurs décennies, comme par exemple celle constituée par les campagnes annuelles IYFS (International Young Fish Survey) en Mer du Nord, coordonnées par le C.I.E.M. Deux situations sont envisageables. Tout d'abord, on peut envisager la situation suivante: ce ne seraient uniquement que des variations d'abondance sur une échelle logarithmique qui se manifesteraient par des variations dans la stratégie d'occupation de l'espace et aussi par des modifications de la dérive. La dérive présente alors deux ou plusieurs états d'équilibre. Cette situation n'a pas été observée. La seconde situation est celle où il n'y aurait que des variations arithmétiques d'année en année. L'occupation de l'espace serait compatible avec l'hypothèse (H2), il n'y aurait qu'un seul état pour la dérive. Les variations des quantités totales sont alors à mettre en relation avec les résidus, non avec la dérive.

Sur les œufs de sole, nous allons nous contenter de modéliser les variations résiduelles des régionalisations autour d'un état d'équilibre.

D. UN MODÈLE MULTIPLICATIF POUR LA RÉGIONALISATION DES ŒUFS DE SOLE

Le type d'occupation de l'espace qui a été observé en fonction de la quantité totale correspond à l'hypothèse (H₂). Les caractéristiques structurales des données nous ont conduit à interpréter la régionalisation dans l'espace et le temps grâce au modèle suivant de fonction aléatoire : $\frac{Z(x, t)}{m_q(t)} = m_w(x) + R(x, t)$ où $m_q(t)$ est la moyenne d'espace

au temps t et $m_w(x) = E \left[\frac{Z(x, t)}{m_q(t)} \right]$ correspond à la dérive, que l'on définit comme la

composante spatiale moyenne des pourcentages, invariante dans le temps. Dans le cas du merlu-0, un modèle quasi-stationnaire était suffisant pour rendre compte de $m_w(x)$. Ici, on travaille à une autre échelle spatiale, le dôme de ponte occupe tout le champ prospecté. On a une analyse non stationnaire de $m_w(x)$.

On se propose d'explicitier sur les œufs de sole le modèle précédent. Il est dit multiplicatif car les effets temps et espace se multiplient. On utilise la convention habituelle qui est de représenter par une majuscule la fonction aléatoire et par des minuscules les valeurs des réalisations.

Comme le même réseau d'échantillonnage est visité aux mêmes stations à chaque campagne on peut, en utilisant les répétitions temporelles, estimer la dérive $m_w(x)$ en chaque station d'échantillonnage, directement à partir des différentes réalisations temporelles de la fonction aléatoire $\frac{Z(x, t)}{m_q(t)}$. On sépare alors l'étude de la dérive de celle

des résidus. Pour chaque composante, on entreprend une analyse structural qui nous conduit à construire d'une part, un modèle spatial et invariant dans le temps pour la dérive et d'autre part, un modèle spatio-temporel pour les résidus. On montre que la quantité totale pondue est plus sous la dépendance de la dérive spatiale que des résidus spatio-temporels.

Disposant d'un modèle structural, on calcule la variance de l'estimation spatiale de la quantité totale pondue sur toute la frayère, à chaque instant. Le calcul géostatistique est inévitable car l'échantillonnage est effectué sur un réseau régulier de stations fixes. Un calcul est proposé pour estimer la variance de l'estimation spatio-temporelle, de la quantité totale des œufs émis sur toute la frayère, pendant toute la saison de ponte.

D.1. Le modèle

Le modèle est construit grâce aux propriétés de la ponte relative instantanée.

Cette variable se laisse séparer en une dérive et des résidus : $\frac{Z(x, t)}{m_q(t)} = m_w(x) + R(x, t)$

où $R(x, t)$ sont les résidus en (x, t) et où $m_w(x)$ représente le pourcentage moyen de biomasse en x .

La définition de la dérive, $m_w(x) = E \left[\frac{Z(x, t)}{m_q(t)} \right]$, correspond à celle de la dérive

en Géostatistique non-stationnaire (Matheron, 1970, 1980). La dérive traduit une permanence temporelle. Elle marque la répétition dans le temps d'une même régionalisation sous-jacente. Elle ne décrit pas nécessairement, comme dans le Krigeage Universel, une composante spatiale régulière et de grande échelle. En effet, ce que nous appelons dérive peut, comme en théorie des FAI-k, recouvrir une réalité plus complexe de petite échelle, si cette complexité se reproduit identiquement au cours du temps.

On estime $m_w(x)$ par la moyenne arithmétique des réalisations $\frac{z(x, t)}{m_q(t)}$ sur toutes

les répétitions temporelles (7 campagnes), en chaque point x .

La moyenne spatiale $m_q(t)$ est estimée par la moyenne arithmétique des $n_x = 35$ stations d'échantillonnage:

$$m_q^*(t) = \frac{1}{n_x} \sum_{i=1}^{n_x} z(x_i, t)$$

La dérive est estimée aux 35 stations:

$$m_w^*(x_i) = \frac{1}{n_t} \sum_{t=1}^{n_t} z_p^*(x_i, t) \quad \text{avec} \quad z_p^*(x_i, t) = \frac{z(x_i, t)}{m_q^*(t)} \quad (11)$$

Dans ces deux calculs, on a utilisé la moyenne des répliquats aux stations répliquées (Ce sont les stations r \bar{e} -échantillonnées. Les détails de l'échantillonnage ont été exposés au paragraphe II.B.3.2.) . Les résidus se déduisent facilement. Mais ici, aux stations répliquées, on a mis dans les résidus les différences entre répliquats :

$$m_w^*(x) - \frac{z(x_R, t)}{m_q^*(t)} = r^*(x_R, t) \quad \text{et} \quad m_w^*(x) - \frac{z(x, t)}{m_q^*(t)} = r^*(x, t)$$

11. Dans les calculs numériques, on a pris non pas la moyenne $m_q^*(t)$ mais la somme des données. Sur les Figures, les valeurs indiquées pour $m_w(x)$ sont égales à celles du texte multipliées par 35/100.

II.D. Un modèle multiplicatif pour les oeufs de sole

Les résidus sont par construction de moyenne nulle dans le temps: $\sum_i r^*(x, t_i) = 0 \quad \forall x$. Ils sont aussi de moyenne nulle dans l'espace :

$$\sum_i r^*(x_i, t) = 0, \quad \forall t, \text{ car la somme des masses est égale à l'unité : } \sum_i \frac{z(x_i, t)}{m_q^*(t)} = 1, \quad \forall t.$$

A ce stade, on met en évidence un effet proportionnel entre la variance temporelle des résidus $E_t[r^2(x, t)]$ et valeur de la dérive $m_w(x)$. La relation entre l'écart-type et la dérive (Figure 46) se laisse modéliser par une droite (d'où un effet proportionnel en $m_w^2(x)$). Ce lien moyenne-variance rend compte du fait qu'au centre du dôme les variations autour du dôme moyen sont en valeur absolue plus fortes que sur les bords : les densités fortes sont les plus variables. Quelle droite choisir pour rendre compte de ce lien ? La droite de régression est à rejeter car elle ne passe pas par l'origine (abscisse à l'origine 0,12). En effet, on ne peut pas avoir en moyenne une valeur $m_w(x)$ nulle et des valeurs $\frac{z(x, t)}{m_q(t)}$ non toutes nulles et toutes positives. La droite des moindres rectangles

représente l'axe principal du nuage de points. Elle a une pente $a = 0,58$ et son abscisse à l'origine est $b = 0,02$. La droite passant par l'origine et le point moyen a une pente ne variant de celle des moindres rectangles qu'à partir de la 3^{ème} décimale. Cette droite nous semble très cohérente avec les données.

On utilise donc la relation :

$$r(x, t) = a \cdot m_w(x) \cdot u(x, t) \quad a = 0,58$$

On vérifie maintenant, sur la Figure 47, la non-corrélation de la variance temporelle de $U(x, t)$, $E_t[U^2(x, t)]$, avec les valeurs de la dérive $m_w(x)$. Les variances temporelles en chaque x des résidus $U(x, t)$ ne sont pas liées à la dérive $m_w(x)$. Elles fluctuent autour de 1, ce qui est cohérent avec la relation moyenne-variance modélisée. Ainsi, la droite extrait bien l'effet proportionnel.

Sur la Figure 48, on regarde le lien entre la variance spatiale des $u(x, t)$, $E_x[U^2(x, t)]$ et les valeurs des moyennes spatiales par campagnes, $m_q(t)$. Les campagnes à forte moyenne spatiale présentent-elles des fluctuations résiduelles plus importantes ? Une corrélation positive n'est pas claire. Elle ne reposerait que sur une campagne, la campagne 50.

On considère que les résidus $u(x, t)$ ne varient ni en fonction de $m_w(x)$, ni en fonction de $m_q(t)$. Le modèle se factorise entièrement en composantes non corrélées :

$$\frac{Z(x, t)}{m_q(t)} = m_w(x) + a \cdot m_w(x) \cdot U(x, t)$$

où $U(x, t)$ est un résidu stationnaire de moyennes spatiale et temporelle nulles et de variances spatiale et temporelle proches de 1. On étudie plus loin leurs structures spatiale et temporelle.

On écrit :

$$Z(x, t) = m_q(t) * m_w(x) * Y(x, t)$$

avec $Y(x, t) = 1 + a U(x, t)$: résidus stationnaires

Ce modèle construit à partir de l'observation des données est en accord avec l'expérience des biologistes sur les poissons de fond. Celle-ci les conduit à utiliser très fréquemment les modèles G.L.M (Generalized Linear Models) afin d'extraire des données de chalutages commerciaux différentes composantes, et les cartographier (ROBSON 1966, GAVARIS 1980, ANON 1990). En effet, en log. notre modèle est linéaire. Dans l'utilisation des G.L.M, trois effets linéarisés par les log. sont généralement considérés: bateau, année, strate spatiale, auxquels s'ajoutent des interactions et des résidus gaussiens et mutuellement indépendants. L'ajustement se fait par la méthode des moindres carrés. Sont choisis a priori, le modèle multiplicatif, les facteurs, l'absence de structure des résidus et leur loi.

Notre modèle exprime aussi que deux effets, espace et temps agissent en se multipliant. Mais ici les deux effets sont bien spécifiés physiquement. L'effet temps marque une proportionnalité des densités ponctuelles à la quantité totale de l'instant t, l'effet espace marque sur les pourcentages de biomasse, une répartition spatiale déterminée invariante dans le temps.

Dans l'analyse classique, le résidu correspond à un bruit blanc, une erreur : c'est la partie du phénomène qui ne peut être expliquée. Dans l'analyse géostatistique, puisque le krigeage est un interpolateur exact, toutes les irrégularités spatiales peuvent être reproduites. Les résidus correspondent alors à la composante structurale aléatoire et leur structure spatio-temporelle est essentielle.

D.2. La dérive $m_w(x)$ et son déterminisme biologique

L'analyse de la distribution spatiale des pourcentages de masse $\frac{z(x, t)}{m_q(t)}$ nous a fait opter pour une analyse de la régionalisation des œufs où l'on distingue une dérive non-stationnaire, de résidus aléatoires spatio-temporels. Les $m_w(x)$ représentent un signal invariant dans le temps. En effet, nous avons des arguments pour interpréter les $m_w(x)$ de cette façon. Une représentation proportionnelle des $m_w(x)$ est fournie sur la Figure 49.

Une régression multiple des $m_w(x)$ par la méthode des moindres carrés sur des variables dites explicatives permet d'interpréter biologiquement la dérive $m_w(x)$:

$$m_w(x) = \sum_{\ell} a_{\ell} d^{\ell}(x) + \varepsilon(x)$$

où les $d^{\ell}(x)$ sont des variables explicatives (fonctions de dérive) et où les $\varepsilon(x)$ sont des résidus au sens classique (de moyenne nulle, sans structure spatiale, non corrélés à $m_w(x)$).

$$R^2 = \frac{\text{variance due à la régression}}{\text{variance totale}} = \frac{\text{Var} [\sum_{\ell} a_{\ell} d^{\ell}]}{\text{Var} [m_w]} = \frac{\sum_{\ell} a_{\ell} \text{Cov} [d^{\ell}, m_w]}{\text{Var} [m_w]}$$

Le coefficient R^2 indique la proportion des différences spatiales entre les $m_w(x)$ que le modèle régressif explique.

On a considéré trois régressions. Dans un premier temps, comme la variable $m_w(x)$ présente une régionalisation en forme de dôme, on choisit les variables d^{ℓ} correspondant à une dérive quadratique en X et Y, soit : $X^2, X, Y^2, Y, XY, 1$. Le R^2 de cette régression est important : 63 % .

Dans un second temps, on introduit une autre variable explicative : la qualité du fond notée qf. En effet, les sables sur le fond constituent l'habitat privilégié de la sole. Les sédiments sont codés sur une échelle arithmétique de 0 à 9 depuis les rochers (0) en passant par les sables jusqu'aux vases (9). Sur la Figure 50, on donne l'emplacement des sables (qf de 2 à 8) susceptibles d'être des lieux favorables à la ponte de la sole. Ils ont effectivement le même emplacement que le dôme de ponte des $m_w(x)$. Sur la Figure 51, on donne le nuage de corrélation entre le code des sédiments et $m_w(x)$. On voit que les codes centraux au milieu de la gamme des valeurs de qf codent les sédiments les plus favorables à la ponte. On rend compte tout d'abord de cette liaison par un modèle à deux boîtes, c'est à dire par une régression de $m_w(x)$ sur l'indicatrice des sables favorables ($1_s(x) = 1$ si qf est compris entre 2 et 8 , $1_s(x) = 0$ sinon). On a :

$$m_w(x) = m^+ 1_s(x) + m^- (1 - 1_s(x)) + \text{res}(x)$$

et
$$E_x[m_w(x)] = pm^+ + (1 - p)m^-$$

où p est la proportion des sables favorables sur zone ($E_x[1_s(x) = 1] = p = 0.54$) et où m^+ et m^- sont les moyennes de $m_w(x)$ sachant respectivement que $1_s(x) = 1$ et que $1_s(x) = 0$: $E_x[m_w(x)/1_s(x) = 1] = m^+ = 4.6$; $E_x[m_w(x)/1_s(x) = 0] = m^- = 0.7$.

Le R^2 de cette régression est proche de 45% et atteste clairement que la répartition des sables favorables détermine deux niveaux d'abondance de la ponte.

En troisième lieu, on rend compte de la liaison entre $m_w(x)$ et qf par une régression quadratique et on ajoute aux six variables d^{ℓ} du dôme de ponte en X et Y, quatre nouvelles variables explicatives relatives à un dôme de ponte en qf: $X^2, X, Y^2, Y, XY, qf^2, qf, qfX, qfY, 1$. Le R^2 de cette troisième régression est 77 % (Figure 52). Il est très élevé et atteste que la régionalisation des oeufs est liée à la sonde (ou longitude), à la latitude et aux sédiments.

Dans ce genre d'analyse, plus on considère de variables d^{ℓ} , même si elles sont corrélées, plus on augmente le R^2 . De plus, l'interprétation des coefficients a_{ℓ} peut être douteuse (TOMASSONE, 1983) du fait de colinéarités entre variables d^{ℓ} . Nous nous contentons de montrer que 77 % des variations spatiales de $m_w(x)$ peuvent être expliqués par trois variables très simples : latitude, longitude, sédiment. La régionalisation de ces

variables n'est pas variable dans le temps. La ponte dépendant de celles-ci, il est cohérent d'observer une grande reproductibilité de la régionalisation des œufs dans le temps. La régression est aussi influencée par l'échelle arithmétique choisie pour coder qf. Mais on ne dispose pas d'arguments physiques susceptibles de nous faire opter pour une autre échelle afin de coder le passage progressif des pierres aux sables puis aux vases. Mais à supposer que l'on change l'échelle de codage, alors on changerait aussi les fonctions de dérive relatives à qf de façon à reproduire au mieux $m_w(x)$.

Qu'en est-il des 23 % des variations spatiales des $m_w(x)$ non expliquées ? Dans notre modèle, toutes les variations spatiales des $m_w(x)$ et en particulier aussi ces 23 % sont déterministes. Sur la Figure 52, on donne le nuage de corrélation entre les valeurs des $m_w(x)$, prédites par la régression multiple et les valeurs expérimentales $m_w(x)$. On observe une très nette sous-estimation des valeurs fortes. Les 23 % inexpliquées concernent donc des détails au centre du dôme de ponte, par exemples des pics locaux de $m_w(x)$. Pour les interpréter, on peut penser à la répartition des géniteurs sur la frayère ou à la circulation des eaux regroupant les œufs.

On remarque que l'on retrouve sur le nuage de corrélation de la figure 52, les deux niveaux d'abondance m^+ et m^- . La fonction de prédiction $\sum_{\ell} a_{\ell} d^{\ell}(x)$ semble expliquer plutôt une indicatrice $1_{m_w(x) \geq z}$ que $m_w(x)$ elle-même, avec z proche de 2.

Même en ajoutant d'autres variables explicatives d^{ℓ} , la fonction de prédiction $\sum_{\ell} a_{\ell} d^{\ell}(x)$ ne pourra jamais reproduire à 100 % les variations spatiales des $m_w(x)$. L'estimateur de krigeage lui peut le faire : disposant d'un modèle spatial pour les résidus $\varepsilon(x)$, il est un interpolateur exact (MATHERON, 1970).

Mais, les $m_w(x)$ ne sont que des estimations des moyennes temporelles vraies. Les erreurs d'estimation, plus grandes en valeur absolue au centre du dôme de ponte à cause de l'effet proportionnel, engendrent une variabilité spatiale supplémentaire. Toutefois, les 23 % inexpliqués ne correspondent pas entièrement à du bruit. On calcule la part de bruit dans la dérive estimée au paragraphe D.4.

D.3. Les résidus et le modèle probabiliste spatio-temporel

Les résidus sont définis ainsi :

$$u(x, t) = \frac{m_w(x) - z_p(x, t)}{a m_w(x)} \quad \text{avec :} \quad z_p(x, t) = \frac{z(x, t)}{m_q(t)}$$

On s'intéresse à leurs structures spatiale et temporelle. Afin de calculer des distances dans le plan longitude (x) – latitude (y), il est nécessaire de transformer les coordonnées (x,y) exprimées en degrés et minutes d'angle. Une minute d'angle au centre de la terre correspond sur tout méridien et sur l'équateur à une distance de 1 mile nautique (1 mi.n = 1 852 m). Sur un parallèle de latitude φ , 1 minute d'angle de longitude correspond en distance à $\cos \varphi$ mi.n. On a multiplié les longitudes par le cosinus de la latitude moyenne de la zone afin de définir une même échelle de distance sur chacun des axes. Les x et y sont exprimés en minutes, donc les distances sont exprimées en mi.n.

D.3.1. Structure spatiale

Pour calculer un variogramme, on dispose à chaque campagne de 46 stations (35 + 11 répliquats), ce qui est peu. On se contentera d'inférer une structure isotrope. On a choisi pour le calcul, un pas de 5 mi.n qui assure un nombre de couples suffisant pour calculer le variogramme.

On calcule le variogramme expérimental des valeurs $u(x,t)$, pour chaque campagne. On a 7 variogrammes. On en prend la moyenne pour chaque distance. On infère un modèle où la structure spatiale de $U(x,t)$ est invariante dans le temps.

Le variogramme moyen des résidus $U(x,t)$ est donné à la Figure 53. Il est bien structuré et se laisse modéliser très bien par un modèle sphérique (portée = 11 mi.n ; palier = 1) ajouté d'un effet de pépite de 0,2 δ :

$$\gamma_u(h) = 0,2\delta + \text{sph}(11)$$

Les répliquats nous permettent d'estimer la pépite. Elle est faible. Ceci indique que la méthode de prélèvement est adaptée à l'échantillonnage des oeufs et/ou que la variance intra-station est petite devant la variance inter-stations.

La structuration du variogramme des résidus s'interprète ainsi : autour du dôme moyen des $m_w(x)$, on observe à chaque campagne des fluctuations, bosses et creux dont la dimension isotrope moyenne est de 11 mi.n., et dont la localisation géographique n'est pas définie. La distance moyenne entre deux stations d'échantillonnages est de 8 mi.n. Ces fluctuations résiduelles sont donc bien locales.

D.3.2. Structure temporelle

Le variogramme temporel des $U(x,t)$ est proportionnel à celui des $Z_p(x,t)$. En chaque x , on écrit :

$$\gamma_{z_p}(d) = E[Z_p(x, t) - Z_p(x, t + d)]^2$$

$$\gamma_{z_p}(d) = a^2 m_w^2(x) \gamma_u(d)$$

On a calculé le $\gamma_{z_p}(d)$ en chaque x et on en a pris la moyenne sur les 35 stations. Le pas choisi est l'année. Le variogramme temporel moyen des pourcentages de masse $z_p(x,t)$ est le suivant:

h (années)	Variogramme
0	0,9
1	1,2
2	1,0
3	1,1

La structure temporelle, en moyenne dans l'espace, est pépitiq. Ainsi, en chaque point x de la zone, la variation temporelle intra-annuelle est du même ordre de grandeur que la variation inter-annuelles.

D.3.3. Interactions espace-temps

Les résidus sont structurés dans l'espace. On analyse la relation dans le temps entre les variations spatiales des résidus u , prises aux mêmes endroits $x, x+h$ mais à des instants différents $t, t+d$. En moyenne dans l'espace, le mont présent dans une zone donnée à l'instant t reste-t-il un mont à l'instant $t+d$, est-il remplacé par un creux ou y a-t-il recoupement ? On calcule des variogrammes spatiaux croisés entre les résidus d'une campagne i et ceux d'une campagne j , sans considérer de translation spatiale :

$$\gamma_{ij}(h) = E_x[(U(x, t_i) - U(x + h, t_i)) (U(x, t_j) - U(x + h, t_j))]$$

Le calcul des variogrammes est isotrope, par pas de 5 mi.n. Comme il y a 7 campagnes, on dispose de 21 variogrammes croisés. On en a pris la moyenne, par pas de temps de 1 an (Figure 54). On observe globalement que dès 5 mi.n (1^{ère} distance inter-stations), les variogrammes croisés atteignent en moyenne un palier identique. Les interactions espace-temps intra-annuelle et inter-annuelles sont du même ordre de grandeur. Pour les petites distances, la valeur des variogrammes est proche de 0 : absence de relation dans le temps entre les variations spatiales courtes, prises aux mêmes endroits mais à deux instants différents. Le palier est négatif (-0,1). Cette anti-corrélation est

faible. On l'interprète comme due à l'inférence du modèle. En effet, la dérive $m_w(x)$ est estimée par une moyenne temporelle. Donc des résidus positifs au temps t en x , puisqu'ils sont de moyenne nulle, sont forcément contrebalancés par des résidus négatifs en x à un autre instant. On considère que les variogrammes croisés calculés sont pépitiqes.

On n'a pas calculé de variogrammes croisés entre campagnes et décalés dans l'espace, c'est-à-dire :

$$\gamma_{ij}(h) = E_x[(U(x, t_i) - U(x + h, t_i)) (U(x + \tau, t_j) - U(x + \tau + h, t_j))]$$

En conclusion, les résidus ne sont structurés que dans l'espace. Il n'y a pas de structure dans le temps et il n'y a pas d'interaction entre l'espace et le temps. Les monts et les creux résiduels autour du dôme moyen de ponte sont de rayon moyen isotrope 11 mi.n, et se déplacent dans l'espace au cours du temps, sans mémoire. Le modèle est simple. Dans la classification des modèles G.L.M que propose le Conseil International pour l'Exploration de la Mer (ANON, 1990b), il correspond au cas de "structure spatiale simple et persistante".

Ce modèle structural, outre son intérêt descriptif, nous permettra de calculer des variances d'estimation globale et de tracer des cartes.

D.4. Contrôle de l'inférence du modèle

Dans la situation classique en Krigeage Universel, on ne dispose que d'une réalisation de la fonction aléatoire $Z(x) = m(x) + R(x)$. Alors, estimer une dérive fonctionnelle $m^*(x) = \sum_{\ell} a_{\ell} d^{\ell}(x)$ par la méthode des moindres carrés et à la fois les résidus par rapport à celle-ci, $R(x) = Z(x) - \sum_{\ell} a_{\ell} d^{\ell}(x)$, soulève un problème inextricable de biais (MATHERON, 1970). En effet, les résidus seront sous-estimés et leur covariance sera biaisée.

Notre situation est différente. Chaque année la ponte a lieu et un même dôme de densité d'oeufs est observé, dont les caractéristiques spatiales peuvent être considérées invariantes dans le temps. On interprète ces répétitions temporelles comme autant de réalisations d'une même fonction aléatoire. On estime en chaque station d'échantillonnage la valeur de la dérive par une moyenne arithmétique prise sur les répétitions temporelles :

$$E_t[Z_p(x, t)] = m_w(x) \quad \text{avec} \quad Z_p(x, t) = \frac{Z(x, t)}{m_q(t)}$$

$$m_w^*(x_i) = \frac{1}{n_t} \sum_t z_p^*(x_i, t) \quad \text{avec} \quad z_p^*(x_i, t) = \frac{z(x_i, t)}{m_q^*(t)} = \frac{z(x_i, t)}{\frac{1}{n_x} \sum_j z(x_j, t)}$$

L'estimation de $m_w(x)$ par $m_w^*(x)$ est non biaisée car il n'y a pas de dérive dans le temps. Les résidus estimés $r^*(x, t) = m_w^*(x) - z(x, t)$ ne le sont pas non plus. Toutefois, la

variance d'estimation de $m_w(x)$ se répercute sur celle des résidus estimés et sur leurs variogrammes.

Notre approche sépare directement sur les données une dérive, de résidus. Leurs cartographies peuvent s'effectuer indépendamment l'une de l'autre. Considérer que cette séparation est possible est un peu une profession de foi car le contrôle de sa qualité est difficile. Il faut disposer de suffisamment d'échantillons dans l'espace pour estimer $m_q(t)$ donc $z_p(x, t)$ avec précision. Il faut ensuite suffisamment de répétitions dans le temps pour estimer $m_w(x)$. Notre approche simple ne paraît pas tenable quand les estimations de $m_q(t)$ et $m_w(x)$ sont trop imprécises.

D.4.1. Variance d'estimation de la dérive

La variance de l'estimation de $m_w(x)$ par $m_w^*(x)$ se calcule comme suit :

$$\sigma^2(x) = E_t[(m_w(x) - \frac{1}{n_t} \sum_{t=1}^{n_t} z_p(x, t))^2] = \frac{1}{n_t^2} E_t[(\sum_{t=1}^{n_t} R_{xt})^2]$$

Comme les résidus ne sont pas structurés dans le temps, on a :

$$E_t[(\sum_{t=1}^{n_t} R_{xt})^2] = \sum_{t=1}^{n_t} E_t[R_{xt}^2]$$

Et à cause de l'effet proportionnel, il vient : $E_t[R_{xt}^2] = a^2 m_w^2(x)$. D'où finalement :

$$\sigma^2(x) = \frac{a^2 m_w^2(x)}{n_t}$$

L'écart type relatif d'estimation dépend de l'effet proportionnel et du nombre de réalisations : $\frac{a}{\sqrt{n_t}} = 0,22$

L'estimation de $m_w(x)$ est peu précise. Il y a une certaine composante aléatoire dans les variations de $m_w^*(x)$. L'erreur d'estimation de la dérive est multiplicative puisqu'il y a un effet proportionnel. On écrit :

$$m_w^*(x) = m_w(x) + b m_w(x) \varepsilon(x)$$

où : $b = \frac{a}{\sqrt{n_t}}$ et $\varepsilon(x)$ est un bruit de moyenne nulle et variance unité, sans corrélation avec les $m_w(x)$.

D.4.2. Conséquences sur le variogramme des résidus estimés

L'imprécision sur l'estimation de $m_w(x)$ se marque sur le variogramme des résidus estimés. On cherche le variogramme des résidus U vrais.

Partant des résidus R^* estimés, on écrit, en omettant les indices t :

$$\begin{aligned} R^*(x) - R^*(x + h) &= \left(\frac{Z(x)}{m_q} - m^*(x) \right) - \left(\frac{Z(x + h)}{m_q} - m^*(x + h) \right) \\ &= R(x) - R(x + h) + \frac{a}{\sqrt{n_t}} \left(m_w(x) \epsilon(x) - m_w(x + h) \epsilon(x + h) \right) \end{aligned}$$

Les résidus $u^*(x, t) = \frac{r^*(x, t)}{a \cdot m_w^*(x)}$ sont dégagés de l'effet proportionnel. L'erreur sur U est donc additive :

$$U^*(x) - U^*(x + h) = U(x) - U(x + h) + \frac{1}{\sqrt{n_t}} \left(\epsilon(x) - \epsilon(x + h) \right)$$

Dans le cas d'une erreur $\epsilon(x)$ additive, pépitique et sans corrélation avec les valeurs de U, le variogramme des U(x,t) vrais qui nous intéresse, doit donc être égale à celui des résidus U^* estimés moins un effet de pépité égale à $\frac{1}{n_t} = \frac{1}{7} = 0.14$. Sur le variogramme des résidus U^* (Figure 53), on lit que les répliquats donnent une valeur expérimentale du variogramme à l'origine de 0,25. On a choisi finalement d'ajuster une pépité égale à 0.20 .

D.4.3. Conséquences sur les variations spatiales de la dérive estimée

La décomposition précédente du signal estimé en la somme du signal vrai et d'un bruit multiplicatif, $m_w^*(x) = m_w(x) + b m_w(x) \epsilon(x)$, conduit à calculer la part de bruit qui existe dans le signal spatiale que représentent les valeurs $m_w^*(x)$ dans l'espace. On a :

$$\text{Var}_x[m_w^*(x)] = \text{Var}_x[m_w(x)] + b^2 E_x[m_w^2(x)]$$

L'indice x signifie que variance et espérance sont calculées dans l'espace. On veut calculer $\text{Var}_x[m_w(x)]$ et le rapport $\text{Var}_x[m_w(x)] / \text{Var}_x[m_w^*(x)]$.

De la relation $m_w^*(x) = (1 + b\epsilon(x)) m_w(x)$, en prenant des moyennes spatiales, on tire :

$$E_x[m_w^{*2}(x)] = (1 + b^2) E_x[m_w^2(x)]$$

Expérimentalement, on a des estimations des termes suivants : $\text{Var}_x[m_w^*(x)] = 8.6$;

$$b = 0.22 ; E_x[m_w^{*2}(x)] = 16.$$

Il vient :

$$\frac{\text{Var}_x[m_w(x)]}{\text{Var}_x[m_w^*(x)]} = 0.91$$

Il y a donc dans le signal spatial des $m_w^*(x)$ environ 9% de bruit (multiplicatif) imputable à l'estimation.

D.4.4. Partition de la variance spatiale des données

Pour développer la variance dans l'espace de $z_p^*(x, t) = m_w^*(x) (1 + a u^*(x, t))$, on utilise la non-corrélation entre $u^*(x, t)$ et $m_w^*(x)$.

$$\text{Var}_x[z_p] = \text{Var}_x[m_w^*] + a^2 E_x[m_w^{*2}] \text{Var}_x[u^*]$$

$\text{Var}_x[m_w^*]$ représente la part de variance due à la dérive estimée et $a^2 E_x[m_w^{*2}] \text{Var}_x[u^*]$ celle due aux résidus estimés.

Tous ces termes sont accessibles expérimentalement.

On a :

$$\text{Var}_x[u^*] = 1,2$$

$$a^2 E_x[m_w^{*2}] = \frac{1}{35} \sum_i a^2 m_w^*(i) = 5,4$$

$$\text{Var}_x[m_w^*] \approx 9$$

$$\text{Var}_x[z_p] \approx 15$$

La dérive estimée $m_w^*(x)$ prend donc en charge 60 % de variations spatiales des $z_p(x,t)$, quelque soit t . Or il y a 9 % de bruit dans le signal spatial $m_w^*(x)$. Donc 54 % du signal spatial de $z_p(x,t)$ est vraiment déterministe (i.e. reproductible dans le temps), 6 % correspond à un bruit d'estimation autour de ce signal et 40-43 % aux fluctuations résiduelles estimées. Notre modèle est assez déterministe car on met 60 % des variations spatiales des z_p dans la dérive.

Dans notre cas, on considère que la carte de la dérive est accessible directement à partir des valeurs estimées $m_w^*(x)$. Pour une régionalisation similaire (modèle multiplicatif, dérive invariante dans le temps), mais dans un cas où un trop petit nombre de réalisations temporelles sont disponibles, la dérive pourrait être considérée comme l'une des composantes spatiales et son estimation pourrait relever d'un krigeage filtrant les autres composantes.

D.5. Krigeage de la dérive

$$z_p^*(x, t) = m_w^*(x) + a m_w^*(x) u^*(x, t)$$

Les cartes des valeurs $z_p(x, t)$ ne sont pas intéressantes en soi pour l'interprétation de la ponte. On veut produire la carte de la dérive $m_w^*(x)$ et les cartes des résidus stationnaires $u^*(x, t)$.

Les plus petites distances inter-stations sont de l'ordre de 5 mi.n. L'interpolation est effectuée au noeuds d'une grille dont la maille est de 2×2 mi.n², dans un polygone centré sur les donnés. A partir de la position des données périphériques, on place les limites à une distance correspondant en moyenne à une demi-maille de la grille d'échantillonnage. Le polygone d'estimation s'étend entre les isobathes 10m et 110m, depuis le nord de Belle Ile jusqu'au sud de l'Ile d'Yeu.

On a vu que la qualité du sédiment détermine en partie la régionalisation. C'est pourquoi on cherche à ce que la carte reproduise des caractéristiques importantes de la régionalisation des valeurs $m_w^*(x)$ dues à la répartition des sédiments favorables : la faible valeur de $m_w^*(x)$ autour de la station n° 12 sur les vases du large, les faibles valeurs sur les roches côtières, l'avancée vers la côte de valeurs moyennes sur les sables au Sud-Est de Belle-Île (stations n° 25, 26), enfin les fortes valeurs centrales (en particulier sur la station n° 11). (Figures 49, 50, 40).

La dérive $m_w^*(x)$ représente un signal invariant dans le temps. Dans l'espace, elle se laisse interpréter comme une réalisation unique d'une fonction aléatoire. Comme la distribution géographique des $m_w^*(x)$ est non-stationnaire, l'estimation ponctuelle aux noeuds de la grille d'interpolation est effectuée en utilisant la méthode du krigeage des Fonctions Aléatoires Intrinsèques d'ordre k (FAI-k, MATHERON, 1973, 1980), ainsi qu'une version multivariable appelée "dérive externe" (GALLI, 1987). En effet, 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 aux corrélations entre les deux variables.

Le dérive est donc ici une fonction aléatoire. En toute rigueur, on la noterait par une majuscule : $\frac{Z(x, t)}{m_q(t)} = M_w(x) (1 + a U(x, t))$.

D.5.1. Une présentation du krigeage en FAI-k

Techniquement, la méthode procède ainsi. Dans un voisinage autour de tout point x , on sait que la forme de la distribution spatiale est d'un certain type. On appelle cette forme locale moyenne, $E_x[m_w^*(x)]$, dérivée de $m_w^*(x)$. C'est un plan (dérivée linéaire en X et Y) ; un dôme (dérivée quadratique en X et Y) ; une forme donnée par une variable explicative (dérivée externe). Considérons d'abord une situation sans dérivée externe. On écrit :

$$m(x) = \sum_{\ell=0}^{L-1} a_{\ell} f^{\ell}(x) + \varepsilon(x)$$

où les $f^{\ell}(x)$ sont des monômes de l'espace en X et Y , et où les $\varepsilon(x)$ sont des résidus de moyenne nulle. Pour estimer $m(x_0)$ en x_0 à partir des échantillons aux points x_{α} du voisinage, on forme une combinaison linéaire des échantillons :

$$m^K(x_0) = \sum_a \lambda_a m(x_a)$$

Cette combinaison linéaire est particulière (combinaison linéaire autorisée) car elle doit "filtrer" le polynôme d'espace $\sum_{\ell} a_{\ell} f^{\ell}(x_0)$. En effet, ceci assure deux choses : (1) que l'estimateur $m^K(x_0)$ se cale sur la valeur de $\sum_{\ell} a_{\ell} f^{\ell}(x_0)$ (non biais), (2) que la variance d'estimation est calculable et ne s'exprime qu'en fonction d'une fonction $K(h)$ appelée covariance généralisée, modélisant la structure spatiale des $\varepsilon(x)$.

La condition de non biais (1) et la variance d'estimation (2) s'écrivent :

$$(1) \quad \sum_a \lambda_a f^{\ell}(x_a) = f^{\ell}(x_0), \quad \ell = 0, \dots, L-1$$

$$\text{Var}[m(x_0) - m^K(x_0)] = \text{Var}[\varepsilon(x_0) - \sum_a \lambda_a \varepsilon(x_a)]$$

$$(2) \quad = K_{00} - 2 \sum_a \lambda_a K_{a0} + \sum_a \sum_{\beta} \lambda_a \lambda_{\beta} K_{a\beta}$$

$K_{a\beta}$ représente la covariance généralisée K pour la distance $|x_{\alpha} - x_{\beta}|$.

En minimisant la variance d'estimation (2) sous les L contraintes de non-biais (1), on obtient le système de krigeage des FAI-k dont les poids λ_{β} de la combinaison linéaire sont solution :

$$(A) \begin{cases} \sum_{\beta} \lambda_{\beta} K_{a\beta} - \sum_{\ell} \mu_{\ell} f^{\ell}(x_a) = K_{ax_0}, & \forall x_a \\ \sum_{\beta} \lambda_{\beta} f^{\ell}(x_{\beta}) = f^{\ell}(x_0), & \ell = 0, \dots, L-1 \end{cases}$$

où les μ_{ℓ} sont L paramètres de Lagrange.

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.

II.D. Un modèle multiplicatif pour les oeufs de sole

La méthode de la dérive externe consiste à ajouter des contraintes à l'estimateur FAI-k afin que celui-ci respecte en plus une forme spécifiée par L fonctions S^{ℓ} de la variable explicative.

$$(B) \begin{cases} \sum_{\beta} \lambda_{\beta} K_{\alpha\beta} - \sum_{\ell} \mu_{\ell} f^{\ell}(x_{\alpha}) - \sum_{\ell'} \eta^{\ell'} S^{\ell'}(x_{\alpha}) = K_{\alpha x_0} \quad , \quad \forall \alpha \\ \sum_{\beta} \lambda_{\beta} f^{\ell}(x_{\beta}) = f^{\ell}(x_0) \quad , \quad \forall \ell \\ \sum_{\beta} \lambda_{\beta} S^{\ell'}(x_{\beta}) = S^{\ell'}(x_0) \quad , \quad \forall \ell' \end{cases}$$

Quand la forme de la dérive est donnée uniquement par des fonctions $S^{\ell'}$ de la variable explicative, $f^0(x) = 1$ est la seule fonction f^{ℓ} qui figure alors dans (B). En posant $S^0(x) = 1$, le système (A) se ré-écrit uniquement avec des S^{ℓ} . Mais la fonction K est toujours choisie de façon à être compatible avec le polynôme d'espace $f^{\ell}(x)$ filtré. L'inférence du polynôme d'espace à filtrer (ordre k de la FAI-k), ainsi que l'inférence de la covariance généralisée $K(h)$ se font dans le logiciel BLUEPACK par une procédure automatique. Elle est décrite explicitement dans MATHERON et DELFINER (1980) comme dans la documentation du logiciel. On la résume rapidement.

L'inférence se fait à la suite de trois étapes.

a - Détermination de l'ordre k de la FAI-k

L'échelle d'étude est d'abord spécifiée par le choix de la taille d'un voisinage autour de chaque point. Ce voisinage est ensuite implanté autour des données. Dans chacun d'eux un polynôme d'espace de même degré k , $\sum_{\ell} a_{\ell} f^{\ell}(x_{\alpha})$, est ajusté par moindres-carrés. Trois polynômes sont essayés:

- pour $k=0$: 1 fonction f^{ℓ} , 1
- pour $k=1$: 3 fonctions f^{ℓ} , 1, x , y
- pour $k=2$: 6 fonctions f^{ℓ} , 1, x , y , x^2 , y^2 , xy

Est retenu le degré k dont les ajustements, sur l'ensemble des voisinages, sont les plus adaptés aux données. Les coefficients a_{ℓ} peuvent varier d'un voisinage à l'autre, mais pas k . L'inférence porte sur la forme moyenne d'une dérive locale.

b - Détermination de la covariance généralisée $K(h)$

Les modèles de covariance généralisée inférés sont des polynômes des distances h . Comme la variance d'une combinaison linéaire autorisée est une combinaison linéaire des $K(h)$, elle peut donc s'écrire comme une fonction des monômes en h . A partir de la construction de nombreuses combinaisons linéaires autorisées, les coefficients de $K(h)$ sont estimés par régression des variances des combinaisons sur les monômes en h . Lorsque l'ordre k est déterminé, seulement un petit nombre de $K(h)$ lui sont compatibles. Elles sont toutes inférées et ne sont retenues que les 3 qui redonnent au mieux, dans les voisinages, la variance observée des combinaisons linéaires autorisées.

c – Validation croisée des données

Chaque donnée est masquée et est ré-estimée avec les autres du voisinage. Cette ré-estimation se fait par krigeage en résolvant le système (A) avec les fonctions f^l sélectionnées en (a) (ou le système (B) si des s^l sont imposées) et les différentes covariances généralisées sélectionnées en (b). Est en définitive retenue la covariance généralisée qui permet la meilleure ré-estimation (erreur et variance d'estimation moyennes les plus faibles). Lors de la validation croisée, la ré-estimation doit tomber dans un intervalle de confiance autour de la valeur de l'échantillon. L'erreur de krigeage est considérée gaussienne et l'intervalle est : valeur observée $\pm 2\sigma_{\text{krigeage}}$.

D.5.2. Cartographie de la dérive estimée

On compare les cartes obtenues avec différentes contraintes de non biais. Pour simplifier, on ne fournit que les covariances généralisées retenues à l'issue des 3 étapes a, b, c.

Tout d'abord, l'ordre k de la FAI- k est lié à l'échelle d'étude. En effet, dans les petits voisinages (8 points, diamètre inférieur à 50 mi.n.) la dérive ajustée est linéaire ($k=1$). Dans des voisinages plus importants (16 à 28 points, diamètre supérieur à 50 mi.n.) la dérive retenue est quadratique ($k=2$). Pour l'inférence, on a retenu les conditions suivantes:

	k=1	k=2 et k=0
nb. de points dans le voisinage	12	28
diamètre moyen du voisinage (mi.n.)	55.	72.
nb. de voisinages implantés	840	391

II.D. Un modèle multiplicatif pour les oeufs de sole

Les statistiques d'inférence des FAI-k sont données dans les deux tableaux qui suivent :

	FAI-1	FAI-2	FAI-0 + dérive externe	FAI-2 + dérive externe
K(h)	0,418 h	0,539 h	0,317 h	0,460 h

Covariances généralisées

	FAI-1	FAI-2	FAI-0 + dérive externe	FAI-2 + dérive externe
em	- 0,04	- 0,10	- 0,16	0,30
qm	3,92	4,42	3,48	4,00
scal	1,02	1,04	1,45	0,97
station aberrante	12	12	24	-
$Z_{est} - Z_{obs}$ sur station aberrante	+ 2,77	+ 200	- 2,68	-

Validations croisées

em : $E(\text{val}_{est} - \text{Val}_{obs})$ = biais moyen : doit être proche de 0

qm : $E(\text{Val}_{est} - \text{Val}_{obs})^2$ = écart quadratique moyen : doit être le plus faible possible

scal = $qm/\sigma_{krigeage}^2$: doit être proche de 1

Premièrement, on infère un modèle de FAI-1. La covariance généralisée est linéaire. Lors de la validation croisée, une seule station est ré-estimée en dehors de l'intervalle admis : c'est la station 12 sur les vases du large, elle est fortement surestimée. Le krigeage est effectué en voisinage glissant en utilisant le même voisinage que pour l'inférence du modèle. La carte krigée donne un aspect de la frayère très compatible avec

les données à ceci près que l'indentation autour de la station 12 est très peu prononcée (Figure 55).

Le modèle de FAI-2 inféré donne des résultats très similaires. Il s'ajuste aux données avec des statistiques très semblables. La covariance généralisée est linéaire. Seule la station 12 est ré-estimée en dehors de son intervalle admissible, elle est fortement sur-estimée avec un ordre de grandeur très comparable à ce que donne le modèle précédent. Le krigeage est effectué en voisinage unique (les 35 stations sont utilisées pour l'estimation à chaque noeud de la grille). La carte krigée est dans l'ensemble fort semblable à la carte précédente mais en diffère sur 2 points. Au nord de Belle Ile, la FAI-2 prolonge le dôme central de ponton alors que la FAI-1 fournit une pente douce. Aussi, dans le quart N-W, la bordure de faibles valeurs est mieux rendue en FAI-2 qu'en FAI-1 (Figure 55).

Dans un second temps, on introduit 5 monômes s^l de la variable qf : qf , qf^2 , qfx , qfy , et $1 = f^0 = s^0$. Le polynôme $\sum_l a_l s^l(x)$ à filtrer représente un dôme calé sur les sables et des interactions avec les x et les y . Ici, l'ordre de la FAI- k est $k=0$ car il n'y a qu'un polynôme d'espace filtré, $f^0=1$. La covariance généralisée est linéaire. Les statistiques de validation croisée sont très semblables aux précédentes. Seule la station 24 (limite nord du centre de ponton) est estimée en dehors de son intervalle de confiance. Il y a sous-estimation. Ici, pour effectuer le krigeage il faut disposer des valeurs qf aux noeuds de la grille d'interpolation. Ceci pourrait être effectivement le cas car la carte des sédiments dans le secteur a été publiée (GLEMAREC, 1977). Malheureusement nous n'en avons pas disposé. On a donc estimé qf aux noeuds de la grille d'interpolation par krigeage ordinaire. Le variogramme de qf et la carte krigée sont fournis sur la Figure 56. Le krigeage est effectué en voisinage unique. La carte krigée des $m_w^*(x)$ en FAI-0 avec le dôme de qf comme dérive externe est satisfaisante sur la frayère elle-même, car les indentations et les valeurs centrales estimées reproduisent bien l'observation des données (Figure 55). Toutefois sur les bordures, en particulier à la côte et au nord de Belle Ile, la cartographie est irréaliste. Elle dépend là des valeurs des sédiments estimées par krigeage en bordure de champ.

Pour obliger l'estimateur d'une part à reproduire les indentations du dôme de ponton et d'autre part à décroître rapidement sur les bordures, on infère un modèle de FAI-2 avec le dôme de qf en dérive externe. La dérive externe nous permet de reproduire les indentations, la FAI-2 de caler l'estimateur à des valeurs faibles sur les bordures. Il y a 10 fonctions de dérive: les 6 f^l d'une FAI-2 et les 4 s^l précédentes. La covariance généralisée est ici encore linéaire. Ici encore, les statistiques de la validation croisée sont du même ordre que pour les modèles précédents. L'ajustement aux données est un peu meilleur car aucune station n'est ré-estimée en dehors de son intervalle de confiance. Le krigeage est effectué en voisinage unique. La carte krigée présente tous les objectifs que nous nous étions fixé de reproduire (Figure 55).

Sur chaque carte, les valeurs krigées sur les bordures du champ sont négatives. Cela tient au fait que la fonction de dérive $\sum_{\ell} a_{\ell} f^{\ell}(x)$ pour chacun des modèles n'est pas bien adaptée sur les bordures. Il y a un léger effet de bord.

Dans les différents modèles de FAI-k inférés, on note comment la variation spatiale de $m_w^*(x)$ se répartit entre la dérive $\sum_{\ell} a_{\ell} f^{\ell}(x)$ et la covariance $K(h)$. Dans le cas de la FAI-2, si on met dans la dérive un peu plus des variations spatiales de $m_w^*(x)$ en considérant une FAI-2 + dérive externe, on voit que la pente de la covariance généralisée diminue. Il y a un balancement entre la part déterministe et aléatoire de ces différents modèles.

Comparons les variances d'estimation des ces différents modèles. A covariance généralisée fixée, plus on ajoute de monômes f^1 et/ou s^1 à filtrer, plus la variance d'estimation augmente car on impose alors plus de contraintes à l'estimateur. Or dans les différents modèles, les covariances généralisées sont différentes. Les cartes de variance d'estimation sont très semblables. Le dernier modèle avec 10 fonctions de dérive fournit un estimateur aussi précis que les autres comme l'attestent les cartes des écarts-types de krigeage (Figure 57). Il reproduit mieux les détails de la régionalisation des $m_w^*(x)$.

D.5.3. Cartes de la dérive estimée et de la dérive vraie

La carte précédente reproduit tout des $m_w^*(x)$, donc pas exactement les $m_w(x)$ vrais. Tracer la carte des $m_w(x)$ vrais relève du filtrage de 9% de bruit sur les $m_w^*(x)$. On n'a pas effectué ce filtrage.

On considère que la valeur vraie $m_w(x)$ se situe dans un intervalle de probabilité "gaussien" autour de $m_w^*(x)$: $[m_w^*(x) - 2 \frac{a}{\sqrt{n_t}} m_w^*(x) ; m_w^*(x) + 2 \frac{a}{\sqrt{n_t}} m_w^*(x)]$.

La régression multiple établie au paragraphe I.D.2 donne une fonction de prédiction par les moindres carrés, $\sum_{\ell} a_{\ell} d^{\ell}(x)$, à 10 coefficients connus. La carte obtenue avec cette fonction de prédiction est donnée sur la Figure 58. Les indentations du centre de ponton sont reproduites de façon satisfaisante. Mais aucun détail au centre n'est visible. Les validations croisées des échantillons effectuées en voisinage unique, avec la fonction de prédiction des moindres carrés d'une part et le krigeage en FAI-2 avec le dôme qf en dérive externe d'autre part, fournissent des ré-estimations qui tombent, pour les 2 modèles, dans l'intervalle de confiance précédent. Les 2 cartes sont acceptables comme carte de la dérive vraie $m_w(x)$. Elles correspondent à 2 extrêmes, l'une sans détail l'autre avec beaucoup de détails. La réalité se trouve entre les deux. La carte obtenue par krigeage en FAI-2 en serait peut-être la représentation la plus proche. C'est elle que nous utiliserons dans la section D.7 concernant les calculs de variance d'estimation globale.

D.6. Krigeage des résidus

D.6.1. Cartographie des résidus stationnaires

Les résidus estimés $u^*(x,t)$ sont stationnaires dans l'espace et de moyenne nulle. Ils montrent un variogramme spatial sphérique, invariant dans le temps. Les cartes sont obtenues par krigeage sur la grille précédente, en voisinage unique. Bien que la moyenne soit connue, on résout quand même le système de krigeage ordinaire (krigeage à moyenne inconnue, FAI-0) de façon à caler vers 0 les estimations des bordures du champ, loin des données. Avec les mêmes notations que dans la partie précédente II.D.5, les équations s'écrivent avec le variogramme $\gamma(h)$ des résidus u :

$$u^K(x_0, t) = \sum_{a=1}^{35} \lambda_a u(x_a, t)$$

$$\text{condition de non-biais : } \sum_a \lambda_a = 1 \quad (1)$$

$$\text{Var. estimation : } \text{Var}[u_{0t} - u_{0t}^K] = 2 \sum_a \lambda_a \gamma_{a0} - \sum_a \sum_\beta \lambda_a \lambda_\beta \gamma_{a\beta} \quad (2)$$

La minimisation de (2) sous la contrainte (1) conduit au système de krigeage (ordinaire) ponctuel dont les poids λ_α sont la solution : (μ est le paramètre de Lagrange)

$$\begin{cases} \sum_\beta \lambda_\beta \gamma_{a\beta} + \mu = \gamma_{a0} & \forall a \\ \sum_\beta \lambda_\beta = 1 \end{cases}$$

La variance de krigeage est la variance minimale : $\sigma_K^2 = \mu + 2 \sum_a \lambda_a \gamma_{a0}$.

Lors du krigeage, l'effet de pépité est omis dans le membre de droite pour le calcul du γ_{a0} . On filtre ainsi cette composante structurale lors de la cartographie. On donne les 7 cartes de krigeage aux Figures 59 et 60. La carte des écarts-types de krigeage est donnée sur la Figure 61.

Les résidus $u(x,t)$ sont calculés ainsi :

$$u^*(x, t) = \frac{z_p(x) - m_w^*(x, t)}{a m_w^*(x)}$$

Donc des résidus négatifs indiquent des pics de ponte supérieurs à la moyenne $m_w^*(x)$. Sur les cartes, on voit que un à deux pics de ponte se développent, des fois près du centre de la zone, des fois près de Belle-Île et/ou de l'Île d'Yeu. On n'a pas réussi à décrire une dynamique dans le temps pour les cartes des résidus. On ne dispose que d'un très petit nombre de cartes pour chaque année. On a observé une grande variabilité temporelle et on a inféré en II.D.3 un modèle spatio-temporel moyen sans corrélation dans le temps. Les 7 cartes qui correspondent à 7 moments différents sont, en moyenne spatiale, non corrélées.

La dérive $m_w^*(x)$ prend en charge l'aspect répétitif de la régionalisation. Les résidus eux sont spécifiques de conditions particulières spatio-temporelles. Quantitativement, on a vu que 40% de la variance spatiale totale est due à la variance spatiale des résidus. C'est pourquoi, on recherche maintenant dans quelle mesure la quantité totale pondue à l'instant t dépend des conditions particulières spatio-temporelles.

D.6.2. Une interprétation de la ponte

La quantité totale pondue est-elle plus sous l'influence des processus imputables à la dérive ou sous ceux imputables aux résidus ? Pour répondre à cette question, on calcule le pourcentage de la quantité totale $m_q(t)$ que totalisent à l'instant t les stations où les résidus sont particulièrement importants. On définit qu'un résidu est important lorsque sa valeur est très différente de celles des autres résidus de la même campagne. On a donc besoin d'un seuil sur l'histogramme des résidus.

On calcule l'histogramme des résidus, toutes campagnes confondues (Figure 62). C'est l'histogramme, moyen dans le temps, pour les valeurs réparties dans l'espace. Il est dissymétrique avec une queue assez longue vers les valeurs négatives (pics de ponte). C'est donc que, sur les différentes réalisations, des pics résiduels de ponte d'une amplitude importante apparaissent aux moins une fois. Mais ils restent très localisés car leur fréquence d'apparition dans l'espace est faible. Leur fréquence d'apparition dans le temps n'est pas calculée ici. On considère comme résidus très importants ceux dont la valeur est inférieure au seuil u_s qui marque à gauche de l'histogramme 5 % des valeurs : $\text{Prob}(u(x, t) \leq u_s) = 0,05$. Le seuil u_s est calculé directement sur l'histogramme expérimental. On parcourt l'histogramme à partir de la gauche. Dès que la fréquence de 5 % est atteinte, on fixe le seuil : $u_s = -1,9$.

Les stations présentant ces pics de ponte sont localisées sur la Figure 63 toutes campagnes confondues. On observe qu'elles sont toujours situées en bordure du centre de ponte mais jamais sur la bordure Ouest du large. Cette observation est tout-à-fait cohérente avec le cycle biologique migratoire côte-large de la sole dans le secteur, décrit par KOUTSIKOPOULOS (1991).

On totalise le pourcentage de $m_q^*(t)$ qu'apportent les stations concernées, pour chaque campagne :

Année	1986				1987	1988	1989
Code campagne	40	41	42	50	70	80	91
N° station pic de ponte	2,3	1	19, 32	19, 27	7, 28, 29	8, 25	16, 25, 26, 30, 31
Contribution des pics (%)	14	0,3	1,4	1,1	8	0,2	16,5

Contribution des pics de ponte résiduels, à la quantité totale pondue.

$$\text{Contribution} = \frac{\sum_x z(x, t)}{m_q^*(t)} \times 100$$

On constate qu'à certaines campagnes, ces stations contribuent très peu à la quantité totale pondue $m_q^*(t)$. Elles sont alors situées près de la côte. Par contre, pour d'autres, la contribution de ces pics est faible, mais non négligeable (8 %, 14 %, 16 %). Ces pics sont alors situés près de l'Île d'Yeu (en 86) et près de Belle-Île (en 89).

La validité du calcul repose sur la qualité de l'estimation de la pente, a , de l'effet proportionnel (Figure 46). En effet, si la pente est mal adaptée aux fortes valeurs de la dérive et de l'écart-type $\sqrt{E_t(r_{xt}^2)}$, ou au contraire si elle est mal adaptée aux faibles valeurs, on risque de ne voir des pics résiduels de ponte qu'au centre de la frayère ou que sur les bords.

Sur la Figure 47, on note que les résidus $u^*(x, t)$ sont un peu plus variables pour les moyennes et faibles valeurs de la dérive $m_w^*(x)$. C'est principalement ce que notre calcul a chiffré. Mais les $u^*(x, t)$ ne dépendent pas des $m^*(x)$; le calcul n'est pas erroné.

Ainsi des concentrations locales de ponte, en particulier près des deux îles peuvent représenter environ entre 10 à 20 % de la ponte instantanée. Mais ces concentrations n'ont pas toujours lieu. Ainsi, la grande majorité (80 % et plus) de la quantité totale pondue à chaque instant est gouvernée par des phénomènes imputables à la dérive et elle ne dépend pas de conditions spatio-temporelles particulières.

Enfin, on soulève à nouveau la question de savoir quelle part revient aux géniteurs et quelle part à l'hydrologie, dans la régionalisation des œufs. L'habitat des géniteurs détermine la morphologie spatiale de la dérive. Les cartes des résidus se

laissent-elles classer en fonction des situations hydrologiques ou indiquent-elles des comportements particuliers des géniteurs ?

Les variations des quantités totales pondues ne sont pas liées aux pics de pontes résiduelles.

Le modèle spatial nous permet maintenant de calculer la variance de l'estimation de la quantité totale.

D.7. Variance de l'estimation globale

On doit faire deux estimations. L'une est une intégration spatiale à l'instant t . On estime la quantité totale présente dans l'eau à t : $m_q(t)$. L'autre est une intégration temporelle sous la courbe en cloche de la ponte. On estime ainsi la quantité totale pondue par le stock de soles pour une année. Comme la durée d'une campagne est de l'ordre de celle de la vie du stade I des oeufs, la moyenne des échantillons estime la ponte moyenne instantanée $m_q(t)$. D'autre part pour calculer la variance d'estimation spatio-temporelle, on aura besoin de la structure temporelle des moyennes spatiales instantanées $m_q(t)$.

D.7.1. Variance de l'estimation spatiale instantanée

La quantité à estimer est la quantité totale des oeufs pondus à chaque instant t :

$$Z_V(t) = m_q(t) = \frac{1}{V} \int_V Z(x, t) dx$$

où V est la surface de la frayère (champ).

Le réseau d'échantillonnage est presque régulier. On donne à chaque station le même poids et on estime l'intégrale d'espace $m_q(t)$ par la moyenne arithmétique des données :

$$Z_V^*(t) = m_q^*(t) = \frac{1}{n_x} \sum_{\alpha=1}^{n_x} Z(x_\alpha, t)$$

α désigne l'indice des stations et n_x le nombre de stations du réseau d'échantillonnage. L'estimateur $Z_V^*(t)$ est sans biais.

La variance d'estimation s'écrit :

$$\sigma^2(t) = E \left[\left(m_q(t) - \frac{1}{n_x} \sum_{\alpha=1}^{n_x} Z(x_\alpha, t) \right)^2 \right]$$

Or, comme dans le modèle on a : $\frac{Z(x, t)}{m_q(t)} = Z_p(x, t) = m_w(x) (1 + a U(x, t))$,

on est conduit à raisonner relativement à la quantité totale $m_q(t)$ qui est une constante à t fixé. Dans l'expression de la variance, $\sigma^2(t) = E[(m_q(t) - \frac{1}{n_x} \sum_{a=1}^{n_x} Z(x_a, t))^2]$, on met $m_q(t)$ en facteur. La variance relative d'estimation de $Z_V(t)$ est donc égale à la variance d'estimation du pourcentage moyen de biomasse $Z_{pV}(t)$:

$$\frac{\sigma^2(t)}{m_q^2(t)} = E[(\frac{1}{V} \int_V Z_p(x, t) dx - \frac{1}{n_x} \sum_{a=1}^{n_x} Z_p(x_a, t))^2]$$

A chaque t , c'est le même réseau d'échantillonnage qui est visité. D'autre part, le variogramme des résidus U et le dôme de ponte $m_w(x)$ sont invariants dans le temps. La variance d'estimation relative est donc invariante dans le temps. on la note :

$$\sigma_{spa}^2 = \frac{\sigma^2(t)}{m_q^2(t)}$$

Dans le calcul de σ_{spa}^2 on ne tiendra pas compte du fait qu'aux points échantillons x_a le pourcentage $Z_p(x_a, t)$ et la dérive $m_w(x_a)$ sont des valeurs estimées et non vraies. Notre calcul conduit à sous-estimer un peu la variance relative d'estimation.

Comme on a : $Z_p(x, t) = m_w(x) (1 + a U(x, t))$, l'erreur d'estimation de $Z_{pV}(t)$ est la somme de deux erreurs : $e_p = e_1 + e_2$ avec :

$$e_p = \frac{1}{V} \int_V Z_p(x) dx - \frac{1}{n_x} \sum_{a=1}^{n_x} Z_p(x_a)$$

$$e_1 = \frac{1}{V} \int_V m_w(x) dx - \frac{1}{n_x} \sum_{a=1}^{n_x} m_w(x_a)$$

$$e_2 = \frac{1}{V} \int_V a m_w(x) U(x) dx - \frac{1}{n_x} \sum_{a=1}^{n_x} a m_w(x_a) U(x_a)$$

Comme les résidus U sont d'espérance nulle, $E[e_1 e_2]$ égale zéro et la variance $\sigma_{spa}^2 = E[e_p^2]$ est la somme des variances de e_1 et e_2 . On en déduit facilement un ordre de grandeur pour σ_{spa}^2 .

La dérive $m_w(x)$ est considérée comme une fonction non aléatoire. On a :

$$\sigma_{spa}^2 = e_1^2 + E[e_2^2]$$

II.D. Un modèle multiplicatif pour les oeufs de sole

Calculons d'abord une approximation de e_1 : $e_1 \approx \frac{1}{n_k} \sum_{k=1}^{n_k} m_w^k(x) - \frac{1}{n_x} \sum_{a=1}^{n_x} m_w^*(x_a)$.

L'intégrale $\frac{1}{V} \int_V m_w(x) dx$ est estimée par la moyenne arithmétique des valeurs de la carte krigée en FAI-k. On a pris dans ce calcul le modèle de FAI-2 que l'on pense être le plus représentatif de la dérive vraie (section D.5.3). Le dôme de ponte est tronqué sur les bordures, les valeurs négatives krigées ne sont pas prises en compte dans le calcul de la moyenne. La moyenne $\frac{1}{n_x} \sum_{a=1}^{n_x} m_w(x_a)$ des valeurs vraies est estimée par celle calculée sur les valeurs expérimentales (celles-ci ne contiennent que 9% de bruit). On trouve :

$$e_1^2 \approx 0.0004$$

Calculons maintenant une approximation de $E[e_2^2]$. En développant le carré, on retrouve la formule classique de la variance d'estimation dans laquelle les valeurs de la covariance sont pondérées par l'effet proportionnel. Valeurs vraies et estimées de la dérive sont confondues. On a :

$$E[e_2^2] = \frac{a^2}{V^2} \int_V \int_V m_w(x) m_w(y) C_{u_{xy}} dx dy - 2 \frac{a^2}{n_x V} \int_V \sum_a m_x(x) m_w(x_a) C_{u_{xa}} dx + \frac{a^2}{n_x^2} \sum_a \sum_\beta m_w(x_a) m_w(x_\beta) C_{u_{a\beta}}$$

où n_x est le nombre d'échantillons (35) et $C_{u_{a\beta}}$ est la covariance des résidus U pour la distance $|x_\alpha - x_\beta|$.

Les trois termes sont facilement calculables, en approchant les intégrales par des sommes discrètes sur la grille d'interpolation du krigeage en FAI-k et en prenant pour valeurs vraies de la dérive les valeurs krigées. Toutefois, comme la portée de la covariance des résidus U est petite dans le champ V (11 m.n), les 2 termes où figurent les intégrales seront faibles devant la double somme discrète, puisqu'ils font intervenir des distances plus grandes et largement supérieures à la portée. On va donc les négliger. Il reste uniquement à calculer la double somme discrète : c'est la moyenne de la covariance des résidus U pour toutes les distances entre les échantillons, pondérée par les valeurs expérimentales de la dérive. On trouve :

$$E[e_2^2] \approx 0.0272$$

Finalement la variance relative de l'estimation instantanée est :

$$\frac{\sigma^2(t)}{m_q^2(t)} = \sigma_{spa}^2 \approx 0,0276$$

et l'écart-type relatif est : $\frac{\sigma(t)}{m_q(t)} \approx 16.6\%$

L'erreur carrée de l'estimation du dôme de pont ne contribue que pour 1.5% à la variance d'estimation totale. L'effort d'échantillonnage réalisé par le réseau régulier des 35 stations est donc tout à fait suffisant pour estimer la surface quadratique du dôme de pont. L'erreur relative de l'estimation de la surface quadratique par la maille d'échantillonnage n'est que de 2%. L'erreur d'estimation de $m_q(t)$ réside essentiellement dans l'erreur d'estimation de la moyenne des résidus aléatoires sur le champ V.

A titre de comparaison, prenons un autre modèle spatial a priori plus simple : les valeurs $z_p(x,t)$ sont considérées réparties sans structure dans l'espace. La variance d'estimation de la moyenne s'écrit alors : $\frac{\text{Var}_X[Z_p]}{n}$. Avec $n = 35$, on obtient un coefficient de variation de 22,9 %. Il est nettement plus élevé que celui donné par notre modèle. Dans le modèle spatial de la statistique classique, on n'a pas cherché à comprendre les variations spatiales, on les considère aléatoires sans structure. Dans notre modèle inféré, la variance d'estimation est moindre car on a mis en évidence non seulement 60 % de déterminisme dans les variations spatiales, mais aussi une structure spatiale sur les 40 % de variations aléatoires. Ce sont autant de variations spatiales expliquées qui n'entrent pas dans le calcul de la variance d'estimation.

D.7.2. Variance de l'estimation spatio-temporelle

La quantité à estimer est :

$$Z_{VT} = \frac{1}{VT} \int_T \int_V Z(x, t) dx dt = \frac{1}{T} \int_T m_q(t) dt$$

L'estimateur utilisé est la moyenne au cours de la saison de ponte, des moyennes spatiales instantanées. Il est sans biais.

$$Z_{VT}^* = \frac{1}{n_t} \sum_t \frac{1}{n_x} \sum_\alpha Z(x_\alpha, t) = \frac{1}{n_t} \sum_t m_q^*(t)$$

n_t : nombre de campagnes pendant la ponte

n_x : nombre de stations d'échantillonnages

α : indice des stations

t : indice des temps

Trois types d'erreurs sont effectués. En premier lieu, sur les valeurs vraies $m_q(t)$, on estime l'intégrale dans le temps par une somme discrète. En second lieu, on estime la somme discrète des quantités instantanées $m_q(t)$ par la somme discrète des quantités

II.D. Un modèle multiplicatif pour les oeufs de sole

estimées $m_q^*(t)$. En troisième lieu, la méconnaissance de la durée T de la pleine ponte se traduit par une erreur géométrique sur l'estimation de Z_{VT} . On note respectivement σ_1^2 , σ_2^2 , σ_3^2 , les variances de ces erreurs. On négligera le terme d'erreur géométrique dans l'estimation spatiale de la quantité instantanée $m_q(t)$. En effet, on considère que les limites spatiales de la ponte sont beaucoup mieux connues que les limites temporelles car dans le temps, on ne dispose pas de variable auxiliaire comme c'est le cas pour les sédiments dans l'espace, qui puisse aider à définir des limites.

On a besoin d'un modèle structural dans le temps pour la variable $m_q(t)$. On fait les trois hypothèses suivantes: chaque année, les quantités instantanées $m_q(t)$ oscillent autour d'une moyenne selon un processus stationnaire dans le temps. Celui-ci est sans structure (pépité) et sans effet de bord sur la durée T de la pleine ponte. D'après les résultats de LE BEC (1985), représenter la ponte de la sole par un processus temporel stationnaire pendant la durée de la pleine ponte n'est pas irréaliste. Toutefois les caractéristiques structurales de ce processus sont mal connues. On ne dispose pas d'un échantillonnage assez fin dans le temps pour déceler une structure et l'hypothèse d'absence d'effet de bord sur le segment temporel $[0, T]$ de la pleine ponte est assez forte. Ainsi, la courbe en cloche de la ponte est modélisée par une fonction en escalier à deux états, pleine ponte, début et fin de ponte.

On considère que les trois erreurs ne sont pas corrélées. On a donc :

$$\text{Var}[Z_{VT} - Z_{VT}^*] = \text{Var}\left[\frac{1}{T} \int_0^T m_q(t) dt - \frac{1}{n_t} \sum_t m_q(t)\right] + \text{Var}\left[\frac{1}{n_t} \sum_t m_q(t) - \frac{1}{n_t} \sum_t m_q^*(t)\right] + D^2 \frac{\sigma_T^2}{T^2}$$

où D^2 est la variance de dispersion sur T des valeurs vraies $m_q(t)$.

On calcule chacune des trois variances d'erreur :

premièrement, comme les valeurs $m_q(t)$ sont sans corrélation dans le temps sur $[0, T]$,

on a :

$$\sigma_1^2 = \text{Var}\left[\frac{1}{T} \int_0^T m_q(t) dt - \frac{1}{n_t} \sum_t m_q(t)\right] = \frac{D^2}{n_t} ;$$

deuxièmement, comme les erreurs des estimations spatiales ne sont pas corrélées dans le temps, on a :

$$\sigma_2^2 = \frac{1}{n_t^2} \text{Var}\left[\sum_t (m_q(t) - m_q^*(t))^2\right] = \frac{1}{n_t^2} \sum_t m_q^2(t) \sigma_{spa}^2 ;$$

troisièmement, dans le terme d'erreur géométrique intervient la variance relative de l'estimation de la longueur du segment temporel T qui correspond à la période de pleine ponte. Au paragraphe I.B.5.2, on a déjà calculé la variance relative de l'estimation d'un segment par une maille régulière a_1 , en supposant que chacune des deux limites du

segment flottent à chaque bout, indépendamment l'une de l'autre et uniformément dans la maille : $\frac{\sigma_T^2}{T^2} = \frac{a_1^2}{6T^2}$. Le terme d'erreur géométrique s'écrit donc :

$$\sigma_3^2 = D^2 \frac{\sigma_T^2}{T^2} = D^2 \frac{a_1^2}{6T^2}$$

L'année 1986 est la seule pour laquelle plusieurs campagnes d'échantillonnage tombent pendant la période de pleine ponte (Tableau 2, p.39 des Figures). On calcule la variance d'estimation de la quantité totale des oeufs pondus pendant la période de pleine ponte, pour cette année là. On estime les variances σ_1^2 et σ_2^2 facilement. En effet, on a :

$$n_t = 4, \sigma_{spa}^2 = 0.0276, \sum_t m_q^{*2}(t) = 22.793 \text{ et}$$

$$D^{*2} = \frac{1}{n_t - 1} \left(\sum_t m_q^{*2}(t) - \frac{(\sum_t m_q^*(t))^2}{n_t} \right) = 0.452,$$

d'où :

$$\sigma_1^2 = 0.113 \quad \text{et} \quad \sigma_2^2 = 0.039$$

Le calcul du terme d'erreur géométrique σ_3^2 pose un problème car les campagnes en 1986 ne se succèdent pas dans le temps selon une maille régulière. Le terme d'erreur géométrique est de ce fait sans doute élevé, mais on ne peut pas ici le calculer. On va considérer que l'échantillonnage dans le temps est régulier avec une maille de 15 jours car on estime la durée T de la pleine ponte à 2 mois (Tableau 2). On a donc : $\frac{a_1}{T} = \frac{1}{4}$. D'où :

$$\sigma_3^2 = 0.005$$

La variance d'estimation de la quantité totale des oeufs émis pendant la période de pleine ponte en 1986 est estimée par :

$$\sigma_{tot}^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 = 0.157$$

La variance σ_1^2 de l'erreur d'estimation de la moyenne spatio-temporelle Z_{VT} par la moyenne arithmétique des vraies valeurs $m_q(t)$, contribue pour 72% à la variance d'estimation totale. La variance σ_2^2 de l'erreur d'estimation de la moyenne arithmétique des vraies valeurs $m_q(t)$ par celle des valeurs estimées $m_q^*(t)$, contribue pour 25% à la variance d'estimation totale. La variance σ_3^2 de l'erreur faite sur Z_{VT} due à celle faite sur les limites de la durée de la ponte, contribue pour 3% à la variance d'estimation totale.

On remarque que les erreurs d'estimation des quantités totales instantanément pondues ($m_q(t)$) se répercutent de manière importante sur la variance de l'estimation

spatio-temporelle. L'effort d'échantillonnage dans l'espace a été privilégié dans les échantillonnages de 1986 à 1989. On mesure quantitativement que cette option était très raisonnable. 4 campagnes réparties régulièrement pendant la période de la pleine ponte (coefficient de couverture de 0.25), formées chacune de 35 stations dans l'espace sur une grille régulière, permettent d'estimer la moyenne spatio-temporelle avec une bonne précision.

Le écart-type relatif de l'estimation spatio-temporelle est en 1986 :

$$\frac{\sigma_{\text{tot}}}{Z_{\text{VT}}^*} = 17.1\%$$

D.8. Conclusion

Le modèle géostatistique est construit à partir de l'observation suivante: quelle que soit la quantité totale estimée à partir d'une campagne (ponte instantanée), les oeufs se répartissent sur la frayère toujours de la même façon, en pourcentage de la quantité totale. Le modèle est une somme de 2 composantes structurales: un dôme moyen des pourcentages de masse dont la forme et la localisation sont invariantes dans le temps et des variations résiduelles autour (déformations locales du dôme). Le dôme moyen extrait 60% de la variance totale. Le dôme moyen est lié aux sables sur le fond et peut être expliqué à 80% par ceux ci. L'estimateur de krigeage qui est un interpolateur exact peut reproduire les détails du dôme alors que la simple régression sur les sédiments ne le peut pas. L'analyse spatio-temporelle des résidus montre que les déformations résiduelles du dôme moyen apparaissent sans mémoire dans le temps. Dans l'espace, ces déformations correspondent à des pics et des creux de ponte de 11 milles nautiques d'envergure. Mais ces déformations locales ne représentent à chaque campagne jamais plus de 20% de la ponte instantanée. Ainsi, celle-ci est principalement sous la dépendance de facteurs généraux liés à la constance du dôme de ponte et non sous la dépendance de conditions particulières spatio-temporelles, intervenant à petite échelle de temps et d'espace.

Le modèle est multiplicatif. Il prend en compte deux effets, temps et espace qui se multiplient sans inter-action. L'analyse structurale a permis de les interpréter biologiquement: l'effet temps correspond à la proportionnalité à la quantité totale instantanée, l'effet espace au dôme de ponte lié aux sédiments.

Le modèle structural nous permet de calculer la variance de l'estimation de la quantité totale pondue à chaque instant sur toute la frayère. Pour calculer la variance de l'estimation spatio-temporelle, la structure temporelle de la quantité totale instantanée est nécessaire. Connaissant cette structure, l'expression de la variance de l'estimation spatio-temporelle peut permettre de répartir rationnellement l'effort d'échantillonnage dans l'espace et dans le temps.

CONCLUSION

Géostatistique et analyse structurale

Dans l'approche géostatistique, le maître mot est l'analyse structurale. Une fois inféré un modèle structural qui rend compte des données, estimateur, variance d'estimation, variance de dispersion sont calculés dans le modèle. Pour le calcul de la variance de l'estimation globale, l'inférence d'un modèle structural nous libère de contraintes trop lourdes pesant sur la stratégie d'échantillonnage, comme par exemple l'obligation de choisir des positions aléatoires pour les stations ou les transects. La force de cette démarche est aussi sa faiblesse. En effet, si le processus aléatoire inféré est une représentation très écartée de la réalité, alors les calculs menés ne la reflètent plus. Toutefois aucune description quantitative de la structure spatiale n'est possible sans ce risque. Ce risque renvoie à la planification de l'échantillonnage, au choix d'un effort d'échantillonnage suffisant de façon à diminuer les fluctuations expérimentales du modèle inféré.

Grâce aux outils de l'analyse structurale, on a pu décrire de façon naturaliste et quantitative la régionalisation des densités de poissons. Le modèle structural nous a permis d'entreprendre trois études. Il nous a permis de calculer la variance de l'estimation de la quantité totale, de façon réaliste, en fonction de l'information dont on dispose. Dans ce calcul, on a pu tenir compte du domaine spatial sur lequel est effectuée l'estimation. Le calcul de la statistique classique est aveugle sur ce dernier point. Le modèle structural inféré nous a aussi permis de réfléchir sur l'effort d'échantillonnage à fournir, en fonction des caractéristiques structurales des très fortes valeurs qui constituent l'essentiel de la biomasse totale. En troisième lieu, le modèle structural nous a permis d'effectuer la cartographie conformément à la structure spatiale des données.

Quand les campagnes d'échantillonnage sont réalisées régulièrement d'année en année, on dispose d'une base de données spatio-temporelle de la régionalisation du poisson. Aux outils classiques de l'analyse structurale, on propose d'ajouter celui de l'étude des courbes de sélectivité car elles permettent de comparer les dispersions de façon très simple, entre les différentes années. Ici encore, de la dissection naturaliste et quantitative des données grâce aux outils structuraux (dispersion et corrélation), un modèle quantitatif décrivant la régionalisation d'année en année a pu être établi.

Apports de la géostatistique à la biologie

Nous avons utilisé la géostatistique telle qu'elle existe actuellement, dans la situation biologique favorable où les déplacements du poisson sont faibles (celui-ci est donc considéré comme immobile) et où la structure spatiale peut être considérée comme stable saisonnièrement, ne variant donc pas pendant la durée de l'échantillonnage. Dans le temps, les échantillons d'une campagne sont traités comme des mesures synchroniques. La carte correspond alors à un instantané pendant la saison.

Dans cette situation, nous avons fait valoir la nécessité de l'inférence d'un modèle structural pour certains types d'échantillonnages. Dans une première partie sur l'estimation de la quantité totale, on a traité principalement du calcul de la précision de l'estimation pour les différents plans d'échantillonnage. Lorsque les échantillons ne sont pas prélevés indépendamment les uns des autres du fait de la technique des prélèvements ou du fait du plan à maille régulière, la précision ne peut se calculer qu'en tenant compte des corrélations spatiales. Le calcul géostatistique s'impose. La géostatistique nous apprend que deux paramètres structuraux déterminent la précision de l'estimation: la régularité de la distribution spatiale (comportement à l'origine du variogramme) d'une part et d'autre part la taille du domaine sur lequel s'effectue l'estimation relativement à la portée des corrélations spatiales. La statistique classique est aveugle sur ce dernier point et elle nécessite l'implantation au hasard des échantillons, ce qui n'est pas toujours possible ni forcément souhaitable. En particulier, la méthode d'échantillonnage acoustique par écho-intégration conduit à un échantillonnage en continu le long de la route du navire. Quand le plan d'échantillonnage est composé de transects parallèles, on propose un calcul très simple. Il suffit de cumuler les valeurs le long des transects et de travailler à 1D sur les cumuls. On montre que la structure spatiale à 1D des cumuls est très régulière: l'emploi de la géostatistique est alors particulièrement simple et pertinent. Une démonstration en a été faite sur le stock du hareng norvégien.

Dans une seconde partie, on a étudié l'évolution inter-annuelle de distributions spatiales stables saisonnièrement. Les courbes géostatistiques de sélectivité minière nous permettent de construire un outil d'investigation simple afin de différencier trois types d'occupation de l'espace quand la quantité totale varie. Dans le Golfe de Gascogne, sur le groupe-0 du merlu comme sur les oeufs planctoniques de la sole, on n'a observé que l'une de ces stratégies: proportionnellement à la quantité totale, la distribution spatiale des pourcentages est invariante dans le temps. Proportionnellement à la quantité totale, les organismes occupent chaque année tout leur habitat sur les mêmes lieux avec les mêmes gradients. Les distributions spatiales des pourcentages de biomasse sont alors interprétées comme autant de réalisations différentes d'une même fonction aléatoire. Un modèle multiplicatif est construit où interviennent 2 effets, temps et espace et des résidus. L'effet temps est une proportionnalité à la quantité totale de chaque instant, l'effet espace est une distribution spatiale non stationnaire reproductible d'année en année, liée à l'environnement et appelée dérive. Par rapport à l'analyse classique des modèles

log-linéaires, l'analyse structurale géostatistique permet non seulement d'interpréter biologiquement les effets temps et espace, mais encore de tenir compte de la structure spatio-temporelle des résidus.

D'autres aspects de la géostatistique peuvent être intéressants pour la biologie des pêches, que nous n'avons pas explorés et qui seraient susceptibles d'applications presque directes, c'est à dire dans les mêmes conditions relativement aux fluctuations temporelles des structures. Par exemple et sans vouloir être exhaustif, on mentionne quatre catégories de problèmes dont la formulation est assez typiquement géostatistique, qui intéressent la biologie halieutique :

- * l'estimation des réserves sur des zones inaccessibles à l'exploitation par la pêche, à partir de prélèvements ponctuels et d'une définition technique des zones non exploitables (ce peut être une donnée importante pour la dynamique de certaines populations de Bivalves exploitées à la drague);
- * l'estimation à partir d'un échantillonnage orienté par une information auxiliaire (c'est typiquement le cas lorsque les chalutages sont décidés au vue d'une prospection acoustique);
- * le filtrage d'erreurs de mesure non corrélées spatialement et dont l'amplitude est une fonction du nombre pêché (c'est la situation classique pour les chalutages de fond);
- * le choix du support des échantillons lorsque l'échantillonnage de la variable biologique s'opère à deux niveaux (il est fréquent de devoir estimer la moyenne spatiale d'un critère morphométrique ou histologique ou du régime alimentaire; il faut alors répartir l'effort entre le sous-échantillonnage dans le prélèvement, i.e. nombre de poissons mesurés, le nombre de prélèvements, i.e. nombre de coups de chaluts et le volume du prélèvement, i.e. le temps de traine).

Apports de la biologie à la géostatistique

Malgré la variété des types d'échantillonnage, des variables biologiques à mesurer en mer ou encore des problèmes d'exploitation dus à la particularité écologique régionale, nous pensons que ce sont les variations temporelles des régionalisations biologiques et des structures associées qui donnent à la géostatistique des axes de recherche fort nouveaux. Nous distinguons deux types de variations temporelles : les variations temporelles de la répartition des valeurs dans l'espace dues à la mobilité des poissons et les variations temporelles de leur structuration. En effet, la répartition des valeurs fortes et le comportement à l'origine du variogramme peuvent chacun varier sans l'autre. Dans le présent travail, on a choisi de se placer dans des situations biologiques où les déplacements sont de faible amplitude relativement à la zone de présence du poisson (champ). Les poissons restant en place, on a distingué deux types de variations temporelles de la structure, de natures différentes.

Nous avons cherché à relier les variations de la structure aux variations dans les quantités totales d'année en année : la structure ne varie dans le temps que dans la mesure où la quantité totale varie dans le temps. Les différences dans les structures d'une année

sur l'autre, sont vues comme des réponses aux variations des quantités totales (stratégies d'occupation de l'espace). On a situé cette étude spatio-temporelle dans le cas biologique suivant : les poissons restent en place, la structure spatiale est stable pendant une saison et ne varie pas pendant l'échantillonnage. Les variations temporelles sont les variations inter-campagnes et tous les échantillons d'une campagne sont des mesures synchroniques. Sur nos données, nous avons observé la permanence d'une année sur l'autre, d'une composante de la régionalisation qui a été nommée dérive. Cette permanence allait de pair avec des variations de la quantité totale sur une échelle arithmétique. Nous avons l'intuition que d'éventuelles modifications de la dérive d'une année à l'autre, seraient concomitantes de variations de la quantité totale sur une échelle logarithmique. Cette situation reste à observer et le modèle spatio-temporel adapté reste à construire.

Les poissons restant en place, leur structure spatiale peut varier au cours de l'échantillonnage si le variogramme est sous l'influence de déterminants écologiques dont le cycle de variation se déroule rapidement. C'est à notre avis la seconde cause de variations du variogramme dans le temps. En particulier, sous les latitudes tempérées et tropicales, l'alternance jour-nuit détermine des modifications de la structure spatiale des poissons pélagiques : ils forment des bancs le jour et se dispersent la nuit. L'échantillonnage en continu de jour et de nuit fournit sur des lieux différents, des informations aux caractéristiques structurales différentes. A la suite de discussions avec Xavier FREULON, nous avons proposé un modèle multivariable de co-krigeage, traitant le jour et la nuit comme deux variables distinctes et corrélées (PETITGAS, 1991). Mais comme on ne dispose pas à partir d'une campagne classique, des deux variables aux mêmes points, ce lien doit être défini théoriquement et doit être injecté dans les données. L'utilisation du modèle reste délicate car une connaissance expérimentale fait défaut. Des expérimentations spécifiques afin d'échantillonner les variations temporelles du variogramme, ont été entreprises (GERLOTTO et PETITGAS, 1991).

Plaçons nous maintenant dans un cas où l'amplitude des déplacements du poisson est grande par rapport au champ, sans que la structure ne varie pendant l'échantillonnage. C'est une situation fort plausible, par exemple pour des poissons pélagiques très actifs dans une petite zone marine comme un golfe ou un fjord. Sur une grille régulière, on échantillonne alors une population de poissons mobiles dont la structuration reste invariante dans le temps à petite échelle spatiale (comportement à l'origine du variogramme). La prise en compte de déplacements du poisson pendant la durée de l'échantillonnage est délicate et semble devoir passer par des simulations. Pour les poissons pélagiques, un schéma booléen où les grains de valeurs fortes seraient mobiles est envisageable et permettrait de réaliser facilement des simulations, mais la connaissance expérimentale sur les caractéristiques de la mobilité fait défaut.

BIBLIOGRAPHIE

AITCHINSON J., BROWN J.A.C., 1957

The lognormal distribution, Cambridge: Cambridge University Press

ANDERSON K.R., 1976

A simple algorithm for positioning a point close to a boundary,
Math. Geol., 8(1), pp 107–108

ANON, 1990a

Report of the Workshop on the Applicability of Spatial Statistical Techniques to Acoustic Survey Data, ICES/CIEM, CM 1990/D:34

ANON, 1990b

Report of the Working Group on Methods of Fish Stock Assessments, ICES/CIEM, CM 1990/Assess:15

ANON, 1991a

Report of the Workshop on the Applicability of Spatial Statistical Techniques to Acoustic Survey Data, ICES/CIEM, CM 1991/D:40

ANON, 1991b

Report on the North Sea Flat Fish Working Group, CIEM/ICES, CM 1990/Assess:5

ARBAULT S., CAMUS P., LE BEC C., 1986

Estimation du stock de soles (*Solea vulgaris*, Quensel 1806) dans le Golfe de Gascogne à partir de la production d'œufs, *J. Appl. Ichtyol.*, 4 , pp 145–156.

BEVERTON R.J.H., HOLT S.J., 1957

On the dynamics of exploited fish populations, Fish. Invest. Ser. II, 19

BURCZYNSKI J., 1982

Introduction à l'utilisation des systèmes sonar dans l'estimation de la biomasse en poissons, Version française préparée par E. Marchal, FAO Doc. Tech. Pêches, 191 , Rev. 1

CHEVALIER R. et al, 1985,

Réflexions sur la variabilité des résultats des campagnes œufs et larves, ICES/CIEM, CM 1985/L:20, Sess. X

COCHRAN W, 1977,

Sampling techniques, New York: John Wiley & Sons, Third Edition

CUSHING D., 1972

The production cycle and the number of marine fishes,
Symp. Zool. Soc. London, n°29, pp 213–232

DARDIGNAC J., 1988

Les pêcheries du Golfe de Gascogne – Bilan des connaissances, Rapports scientifiques et techniques de l'IFREMER, n°9

- DRAGESUND O., HAMRE J., ULLTANG Ø, 1980
 Biology and population dynamics of the Norwegian spring spawning herring,
Rapp. P.V. Réun. Cons. Int. Explor. Mer, 177, pp 43-71
- DINER N. et al., 1989
 A new acoustic data acquisition and processing system, ICES/CIEM, CM 1989/B:45
- FINNEY D.J., 1941
 On the distribution of a variate whose logarithm is normally distributed,
J.R. Statist. Soc. Suppl. 7, 151
- FONDS M., 1979
 Laboratory observations on the influence of temperature and salinity on
 developpment of eggs and growth of larvae of sole *Solea solea*,
Mar. Ecol. Prog., Ser. 1, pp 91-99
- GALLI A., MEUNIER G., 1987
 Study of a Gaz Reservoir Using the External Drift Method
 in *Geostatistical case-studies*, Matheron G. & Armstrong M. Eds., Holland: D. Reidel
 Publ. Co., pp 105-119
- GALLI A., RENARD D., 1991
 Notice d'utilisation du programme Bluepack 3D,
 Centre de Géostatistique, ENSMP, Fontainebleau
- GAVARIS S., 1980
 Use of multiplicative model to estimate catch rate from commercial data,
Can. J. Aquat. Sci., 37, pp 2272-2275.
- GERLOTTO F., PETIT D., FRÉON P., 1990
 Influence of the light of a vessel on TS distribution, ICES/CIEM,
 CM 1990/B:49 Sess. R.
- GERLOTTO F., PETITGAS P., 1991
 Some elements on time variability in acoustic surveys,
 ICES/CIEM, CM 1991/B:15 Sess.X
- GERLOTTO F., STEQUERT B., 1983
 Une méthode de simulation pour étudier la distribution des densités en poissons:
 application à deux cas réels, *In Proc. International Symposium on Fisheries Acoustics*,
 June 1982, Bergen (Norway), Nakken O. et Venema S.C. Eds, FAO Fish Rep. 300,
 pp 278-292
- GLEMAREC A. et al, 1977
 Carte sédimentaire du plateau continental du Golfe de Gascogne (1/500 000).
- GOHIN F., 1985
 Geostatistics applied to fish distributions as derived by acoustic surveys, ICES/CIEM,
 Report of the Acoustic Working Group Meeting, Tromsø (Norway), May 1985

GUEGEN J., 1987

Evaluation des ressources halieutiques de l'Ouest de l'Europe par campagnes de chalutages programmées, Première phase, Document IFREMER, n°2610 du 27/07/1987

GUIBAL D., 1973

Les fonctions auxiliaires à deux dimensions pour le schéma sphérique, Note N-347, Centre de Géostatistique, ENSMP, Fontainebleau

GUICHET R., 1988

Etude de la croissance du merlu européen (*Merluccius merluccius*, L.) au cours de ses premières années, ICES/CIEM, CM 1988/G:53

HJORT J., 1926

Fluctuations in the year classes of important food fishes, *J. Cons. Int. Explor. Mer*, 1, pp 5-38

HOYLE M.H., 1968

The estimation of variance after using a gaussianating transformation. *Annals of Mathematical Statistics*, 39, pp 1125-1143

ILES T.D., SINCLAIR M., 1982

Atlantic herring: stock discreteness and abundance, *Science*, 215, pp 627-633

JAKOBSSON J., 1983

Echo surveying of the icelandic summer spawning herring 1973-1982, *In Proc. International Symposium on Fisheries Acoustics*, June 1982, Bergen (Norway), Nakken O. et Venema S.C. Eds, FAO Fish Rep. 300, pp 240-249

JOHANNESSON K., MITSON R.B., 1984

Fisheries acoustics: a practical manual for aquatic biomass estimation. FAO Fish. Tech. Pap. n° 240

JOLLY G.M., HAMPTON I., 1990

Some common misconceptions in the statistical design and analysis of acoustic surveys to assess fish biomass, *In Proc. International Symposium on Fisheries Acoustics*, June 1987, Seattle (USA)

JOURNEL A., HUIJBREGTS Ch., 1978

Mining Geostatistics, London: Academic Press

JUNQUERA S., 1986

Pêche de l'anchois (*Engraulis encrasicolus*) dans le Golfe de Gascogne et sur le littoral atlantique de la Galice depuis 1920, *Rev. Trav. Inst. Pêch. Marit.*, 48(3 et 4), pp 133-142

KIRKEGAARD D., AGLÉN A., DEGNBOL P., HAGSTRÖM O., PALMEN L., SIMMONDS E., 1990

Report on the herring acoustic surveys in the North Sea, Skagerrak and Kattegat during summer 1989, ICES/CIEM, CM 1990/H:24

KOUTSIKOPOULOS C., 1991

Recrutement de la sole (Solea solea, L.) du Golfe de Gascogne : influence de l'hydrologie et de l'hydrodynamisme, Thèse de Doctorat d'Océanographie Biologique, Université de Bretagne Occidentale.

LAJAUNIE Ch., RIVOIRARD J., 1988

Une modèle d'intervalle de confiance pour la moyenne globale,
Note N-55/88/G, Centre de Géostatistique, ENSMP, Fontainebleau

LALOË F., 1985

Contribution à l'étude de la variance d'estimateurs de biomasse de poissons obtenus par écho-intégration, *Océanogr. Trop.*, 20(2), pp 161-169

LANTUÉJOUL Ch., RIVOIRARD J., 1984

Une méthode de détermination d'anamorphose,
Note N-916, Centre de Géostatistique, ENSMP, Fontainebleau

LANTUÉJOUL Ch., 1990

Cours de sélectivité, Document de cours C-140,
Centre de Géostatistique, ENSMP, Fontainebleau

LE BEC C., 1985

Cycle sexuel et fécondité de la sole (*Solea vulgaris*, Quensel 1806) du Golfe de Gascogne, *Rev. Trav. Inst. Pêches Marit.*, 47(3), pp 179-189.

MASSÉ J., 1988

Utilisation de l'écho-intégration en recherche halieutique. Document IFREMER, DRV 88030-RH/Nantes

MATERN B., 1960

Spatial Variation, Med. Fr. Statens, Skogsforskningsinstitut Band 49, n° 5, Stockholm

MATHERON G., 1962

Traité de Géostatistique appliquée, tome 1, Mémoires du BRGM, n° 14, Paris: Technip

MATHERON G., 1965

Les variables régionalisées et leur estimation, Paris: Masson

MATHERON G., 1967

Eléments pour une théorie des milieux poreux, Paris: Masson

MATHERON G., 1970

La théorie des variables régionalisées et ses applications,
Les Cahiers du Centre de Morphologie Mathématique, fasc. 5, ENSMP, Paris

MATHERON G., 1973

Le krigeage disjonctif, Note N-360, Centre de Géostatistique, ENSMP, Fontainebleau

MATHERON G., 1974

Effet proportionnel et lognormalité : le retour du serpent de mer,
Note N-374, Centre de Géostatistique, ENSMP, Fontainebleau

- MATHERON G., 1978a
Estimer et choisir – Essai sur la pratique des probabilités,
 Les Cahiers du Centre de Morphologie Mathématique, fasc. 7, ENSMP, Paris
- MATHERON G., 1978b
 Le krigeage disjonctif et le paramétrage local des réserves, Document de cours C-76,
 Bibliothèque du Centre de Géostatistique, ENSMP, Fontainebleau
- MATHERON G. et DELFINER P., 1980
 Les fonctions aléatoires intrinsèques d'ordre k , Document de cours C-80,
 Centre de Géostatistique, ENSMP, Fontainebleau.
- MATHERON G., 1981
 La sélectivité des distributions,
 Note N-686, Centre de Géostatistique, ENSMP, Fontainebleau.
- MATHERON G., 1989
Estimating and choosing – An essay on probability in practice, Berlin: Springer-Verlag
- Mc LENNAN D.N., Mc KENZIE I.G., 1988
 Precision of acoustic fish stock estimates, *Can. J. Fish. Aquat. Sci.*, 45, pp 605-616
- MAY R., 1974
 Larval mortality in marine fishes and the critical period concept, in *The Early Life History of Fish*, Blaxter J.H.S. Ed., Berlin: Springer-Verlag, pp 3-21
- MYERS R.A., STOKES K., 1989
 Density dependant habitat utilization of groundfish and the improvement of research surveys, ICES/CIEM, CM 1989/D:15, Sess. U
- OLSEN K., ANGELL J., LØVIK A., 1983
 Quantitative estimations of the influence of fish behaviour on acoustically determined fish abundance, in *Proc. International Symposium on Fisheries Acoustics*, June 1982, Bergen (Norway), Nakken O. and Venema S.C. Ed., FAO Fish Rep. 300, pp 139-149
- ONA E., 1990
 Physiological factors causing natural variations in target strength, in *Proc. International Symposium on Fisheries Acoustics*, June 1987, Seattle (USA)
- ONA E., TORESEN R., 1989
 Avoidance reaction of herring to a survey vessel studied by scanning sonar, ICES/CIEM, CM 1989/H:24
- PENNINGTON M., 1983
 Efficient estimators of abundance for fish an plankton surveys,
Biometrics, 39, pp 281-286
- PETTIGAS P., 1991
 Un modèle de co-régionalisation pour les poissons pélagiques formant des bancs le jour et se dispersant la nuit
 Note N-33/91/G, Centre de Géostatistique, ENSMP, Fontainebleau.

- PINGREE R.D., LE CANN B., 1989
Celtic and armorican slop and shelf residual currents, *Prog. Ocean.*, 23, pp 303–338
- POULARD J.C., 1990
Evaluation des ressources halieutiques de l'Ouest de l'Europe, Deuxième phase, Contrat IFREMER-CEE, n°88/1210834/BF
- RILEY J., 1974
The ditribution and mortality of sole eggs,
in *The Early Life History of Fish*, Blaxter J.H.S. Ed., Berlin: Springen-Verlag, pp 39–53
- RIVOIRARD J., 1986
Computing variograms on uranium data, in *Geostatistical case studies*, Matheron G. & Armstrong M. Eds., Hollande: D. Reidel Publ. Co., pp 1–22
- RIVOIRARD J., 1988
Modèles à résidus d'indicatrices auto-krigeables,
Sciences de la Terre, Série Informatique, n°28, pp 303–326
- RIVOIRARD J., 1989
Models with orthogonal indicator residuals, *In Proc. Third International Geostatistical Congress*, Sept. 1988, Avignon (France), *Geostatistics*, Matheron G. & Armstrong M. Eds, Vol. 1, Kluwer Academic Press, pp 91–108
- ROBSON D.S., 1966
Estimation of the relative fishing power of individual ships,
Res. Bull Int. Comm. N.W. Atlant. Fish, 3, pp 5–14.
- SCHAEFFER M.B., 1954
Some aspects of the dynamics of populations important to the managment of commercial marine fishes, *Bull. Inter-Amer, Trop. Tuna Comm.*, 1(2), pp 27–56
- SHEPHERD J.G., 1982
A versatile new stock-recruitment relationship of fisheries and construction of sustainable yield curves, *Cons. Perm. Int. Explor. Mer*, 40(1), pp 67–75
- SHEPHERD J.G., CUSHING D.H., 1990
Regulation in fish populations: myth or mirage?,
Phil. Trans. Roy. Soc. Lond., B330, pp 151–164
- SIMMONDS E.J., WILLIAMSON N.J., GERLOTTO F., AGLLEN A., 1991
Survey design and analysis procedures: a comprehensive review of good practice, ICES/CIEM, CM 1991/B:54
- TOMASSONE R., LESQUOY E., MILLIER C., 1983
La régression – Nouveaux regards sur une ancienne méthode statistique, INRA – Actualités Scientifiques et Agronomiques, 13, Paris: Masson
- TORSEN R., 1990
Absorption of acoustic energy in dense herring schools studied by the attenuation in bottom echo-signal, ICES/CIEM, CM 1990/B:2, Sess. R

VEEN J.F., 1978

Changes in the North Sea sole stocks (*Solea solea*, L.),
Rapp. P.V. Reun. Cons. Int. Explor. Mer, 172, pp 124–136

VINCENT A., KURC G., 1969

Les variations de la situation thermique dans le Golfe de Gascogne et leur incidence
sur l'écologie et la pêche de la sardine en 1968,
Rev. Trav. Inst. Pêches Marit., 33(2), pp 203–212

WILLIAMSON N.J., 1982

Cluster sampling estimation of the variance of abundance estimates derived from
quantitative echo-sounder surveys, *Can. J. Fish. Aquat. Sci.*, 39, pp 229–231

ANNEXE : FIGURES

CLE DES CORRESPONDANCES TEXTE-FIGURES

PARTIE I : ESTIMATION DIRECTE DES STOCKS

I.A. Géostatistique de la variance d'estimation :

Fig.1 **p.1**

I.B. Application à une prospection acoustique à maille régulière :

I.B.1 L'échantillonnage	Fig.2-5	p.2-6
I.B.2 Statistiques sur les données	Fig.6-9	p.7-9
I.B.3 Variance d'estimation à 2D	Fig.10-11	p.10
I.B.4 Variance d'estimation à 1D	Fig.12-15	p.11-13
I.B.5 Erreurs géométriques	Fig.8 et 16	p.8 et 14

I.C. Efficacité de l'échantillonnage :

I.C.1 Choix de la distance inter-transects	Fig.17-19	p.15-19
I.C.2 Krigeage Disjonctif du hareng	Fig.20-24	p.18-22
I.C.3 Nombre de grains à teneur très élevée	Fig.25	p.23

I.D. Efficacité des estimateurs :

Fig.26 **p.24**

PARTIE II : ANALYSE DES CARTES

II.A. Etude des dispersion par les relations tonnage-teneur :

Fig.27-28 p.25-26

II.B. Applications aux données :

II.B.1 Caractéristiques du Golfe de Gascogne

Fig.29 p.27

II.B.2 Le merlu-0

Fig.30-39 p.28-37

Tab.1 p.34

II.B.3 les oeufs de sole

Fig.40-44 p.38-42

Tab.2 p.39

II.C. Différences significatives entre les estimations globales :

Fig.45 p.43

II.D. Modèle multiplicatif pour les oeufs de sole :

II.D.1 Le modèle

Fig.46-48 p.44

II.D.2 Dérive: déterminisme biologique

Fig.49-52 p.45-46

II.D.3 Résidus: modèle probabiliste

Fig.53-54 p.47

II.D.5 Krigeage de la dérive

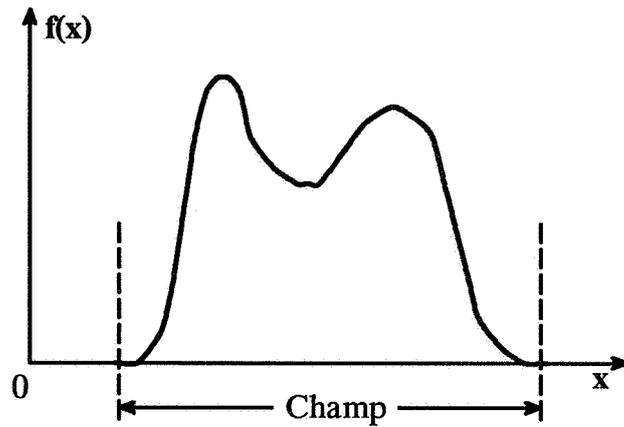
Fig.55-58 p.48-51

II.D.6 Krigeage des résidus

Fig.59-63 p.52-54

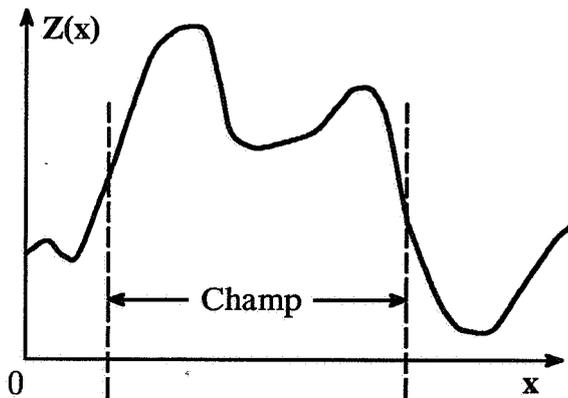
Variable régionalisée et limite du champ

Champ et variable intimement liés : Théorie transitive

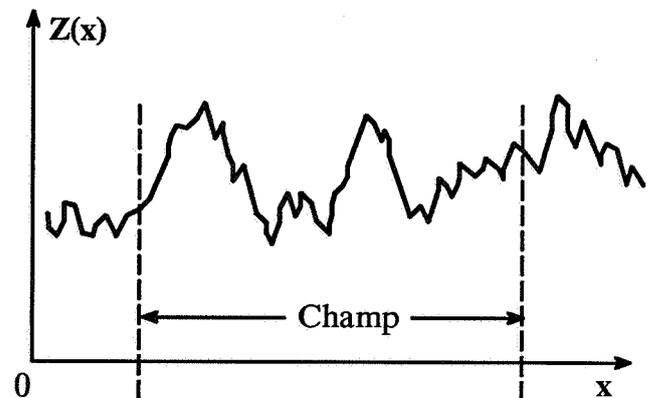


Indépendance interne entre champ et variable : Théorie intrinsèque

$f(x)$: Représentation partielle d'une fonction aléatoire $Z(x)$



FAI-k



FAI-0

Fig. 1

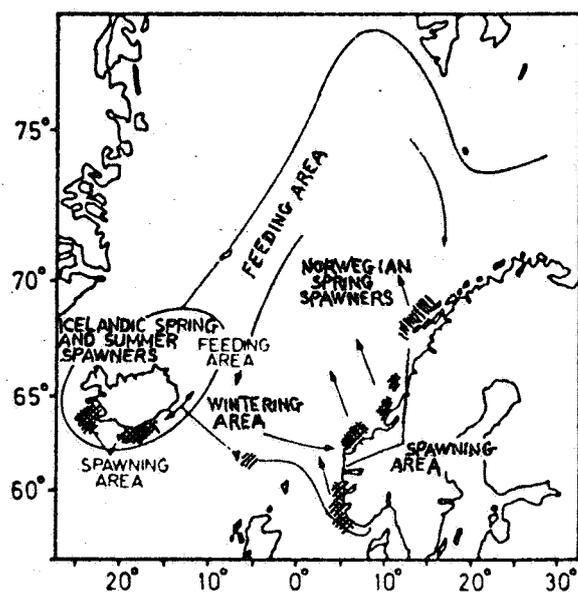


Fig. 2 : Localisation des stocks unitaires de harengs à l'intérieur de l'ensemble Atlanto-scandinave. (in Dragesund et al 1980)

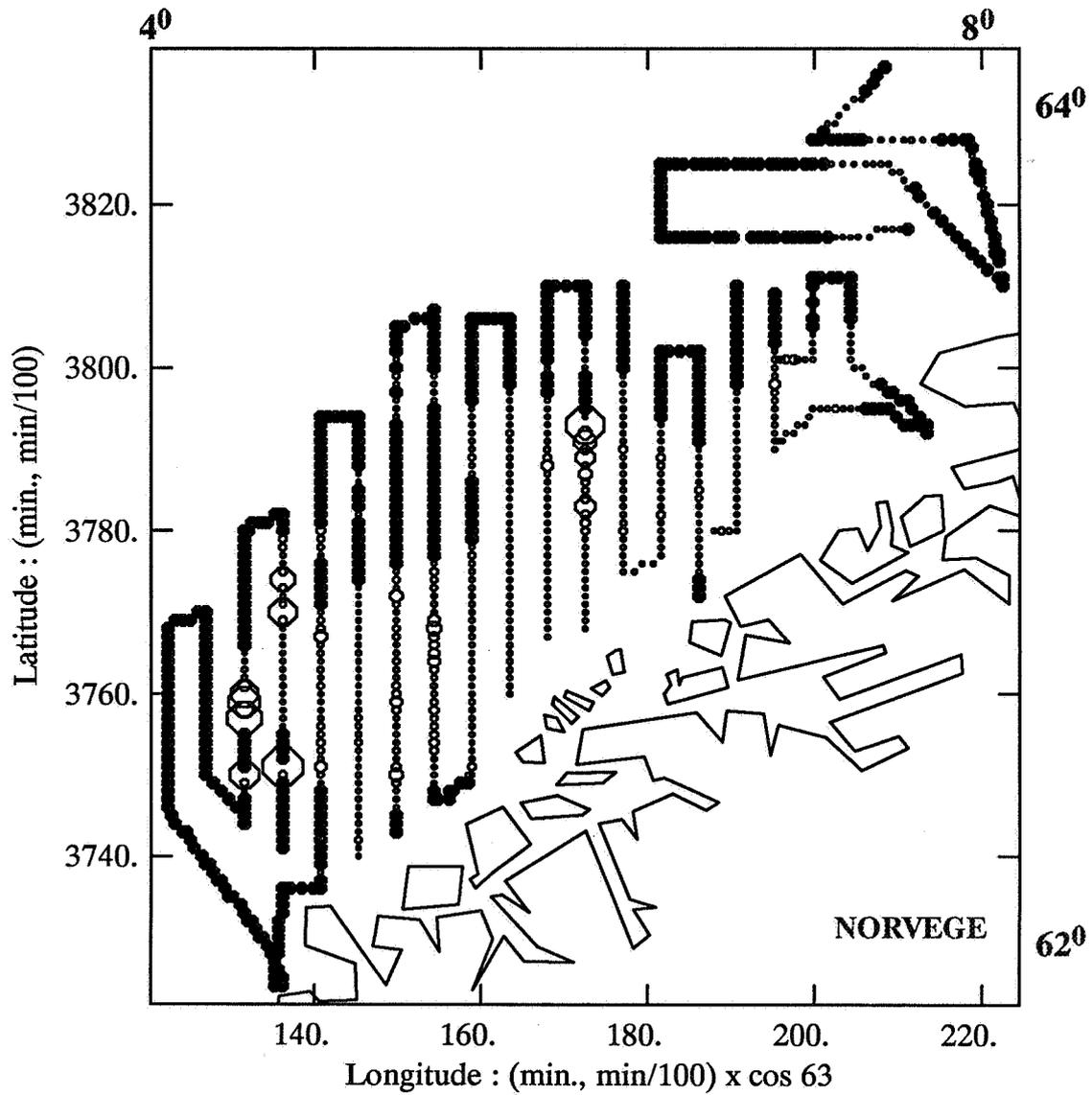


Fig. 3 : Prospection acoustique de février 1989 (N/O Eldjarn) sur une des frayères du hareng norvégien, avec la représentation proportionnelle des données. Les valeurs nulles sont représentées par des disques noirs. Le cercle de rayon maximal correspond à la valeur maximale. Les autres s'en déduisent proportionnellement.

EMISSION ↓		RECEPTION ↑	
		GVT	SIGNAL DE SORTIE $u \sim \sqrt{\rho \bar{\sigma} \tau}$
PUISSANCE EMISE $P = \text{const}$	TRANS- DUC- TEUR	SIGNAL ELECTR. $u_e \sim \sqrt{\frac{1}{R^2 \exp(2\beta R)}} \rho \bar{\sigma} \tau$	
SON EMIS I_0	ONDE SONORE	ECHO $I_e \sim n \bar{\sigma} \frac{I_0}{R^4 \exp(2\beta R)}$ $\sim \rho \bar{\sigma} \tau \frac{I_0}{R^2 \exp(2\beta R)}$	
SON INCIDENT $I_i \sim \frac{I_0}{R^2 \exp(\beta R)}$		SON REFLECHI $I_r \sim n \bar{\sigma} I_i \sim n \bar{\sigma} \frac{I_0}{R^2 \exp(\beta R)}$	

n poissons dans le volume $v = R^2 \psi \frac{c\tau}{2}$
 CIBLE: densité en poissons par unité de volume $\rho = \frac{n}{v} = \frac{n}{R^2 \tau \psi}$
 nombre de poissons dans le volume "v": $n = \rho v \sim \rho R^2 \tau \psi$

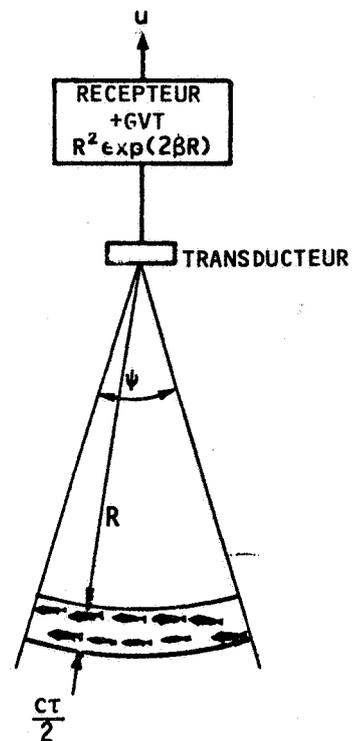


Fig. 4a : Réflexion par une cible multiple d'une impulsion émise par un sondeur (in BURCZYNSKI 1982)

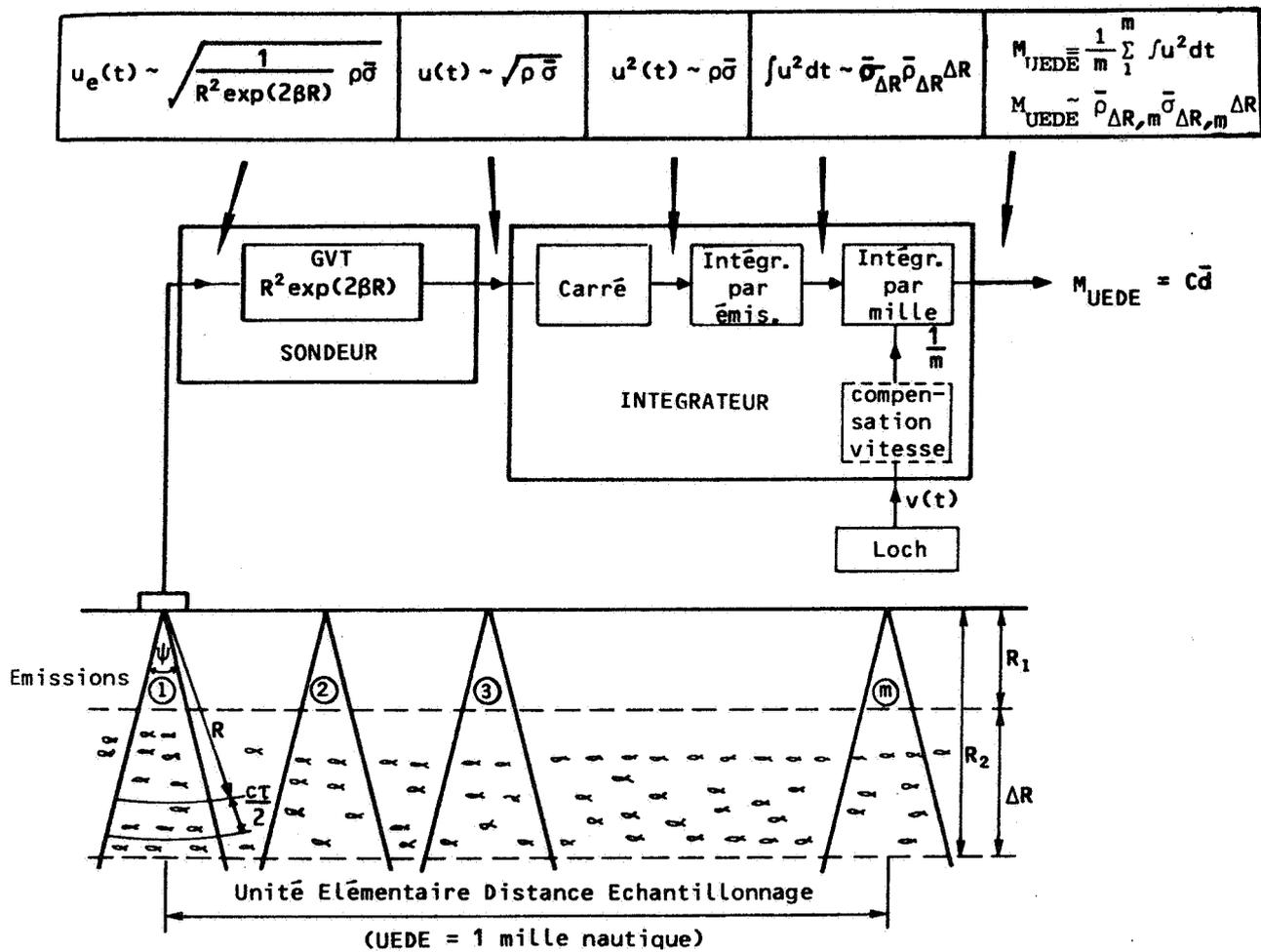


Fig. 4b : Fonctionnement d'un intégrateur d'échos
(in BURCZYNSKI 1982)

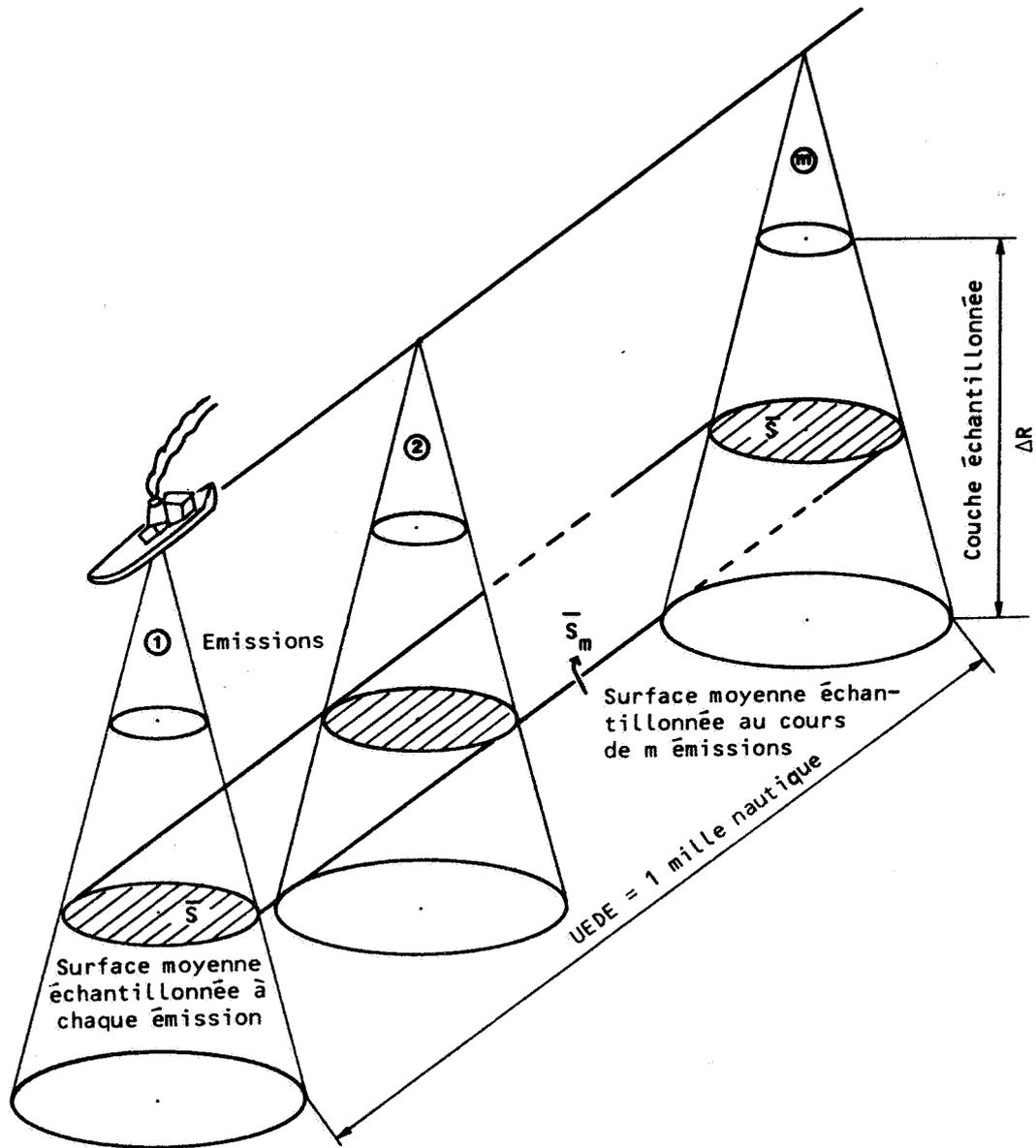


Fig. 5 : Surface moyenne de détection échantillonnée par un intégrateur d'échos
(in BURCZYNSKI 1982)

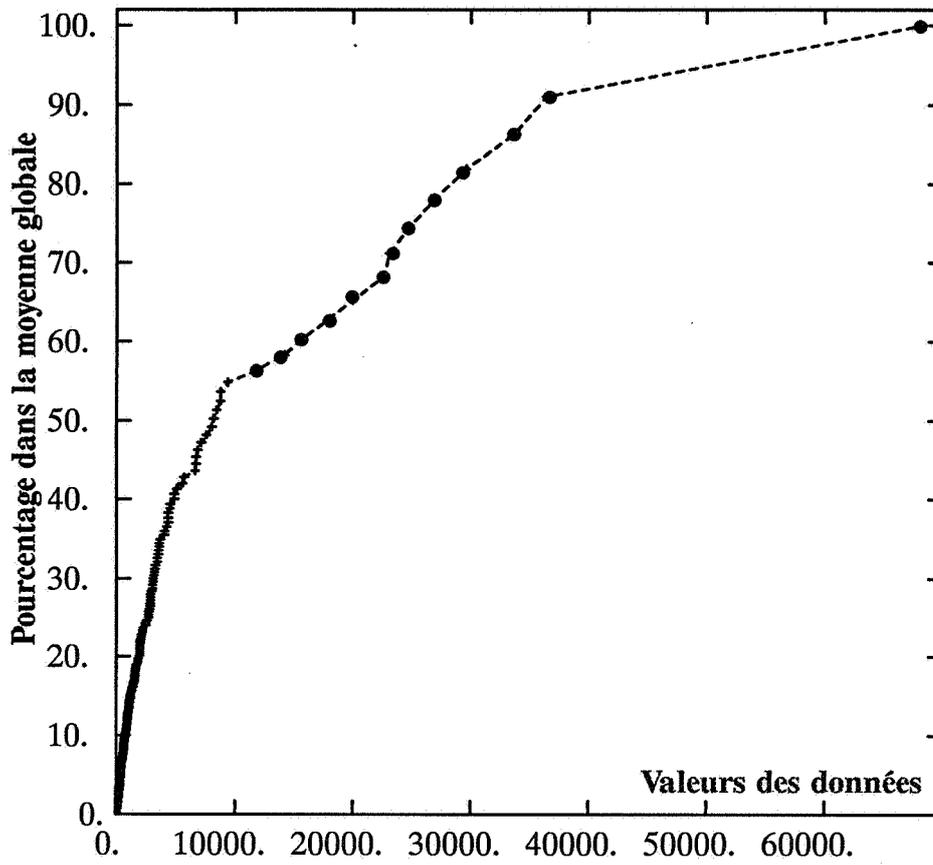
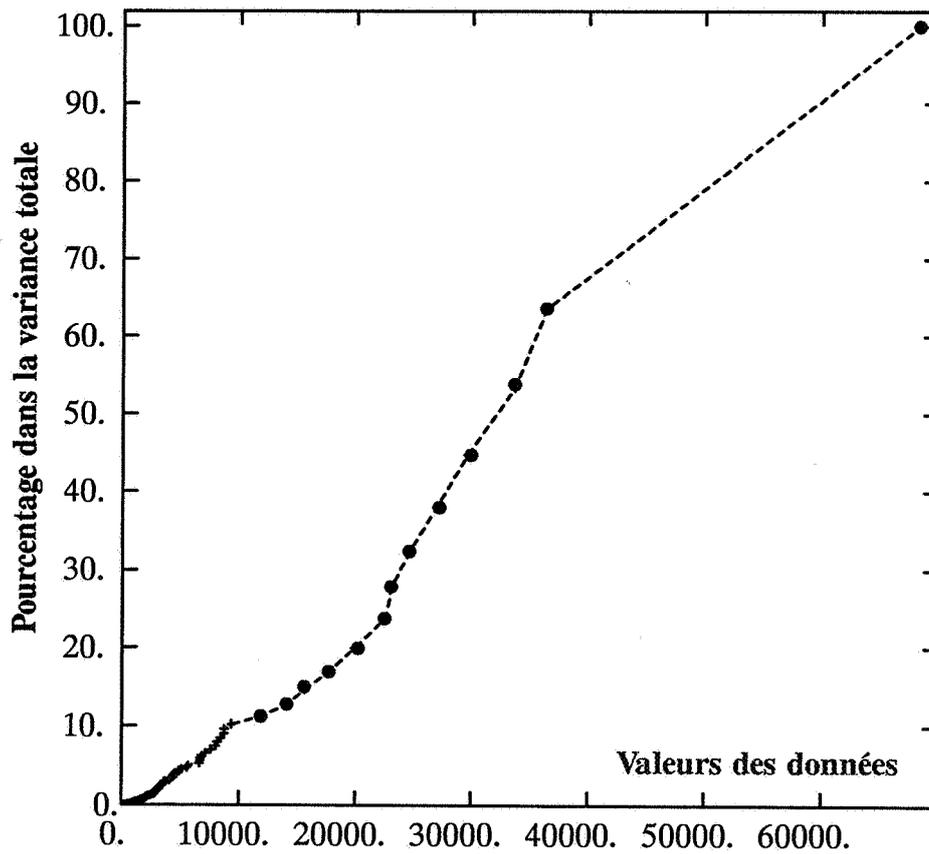


Fig. 6 et 7 : Contributions à la moyenne (en haut) et à la variance (en bas) des différentes portions de l'histogramme



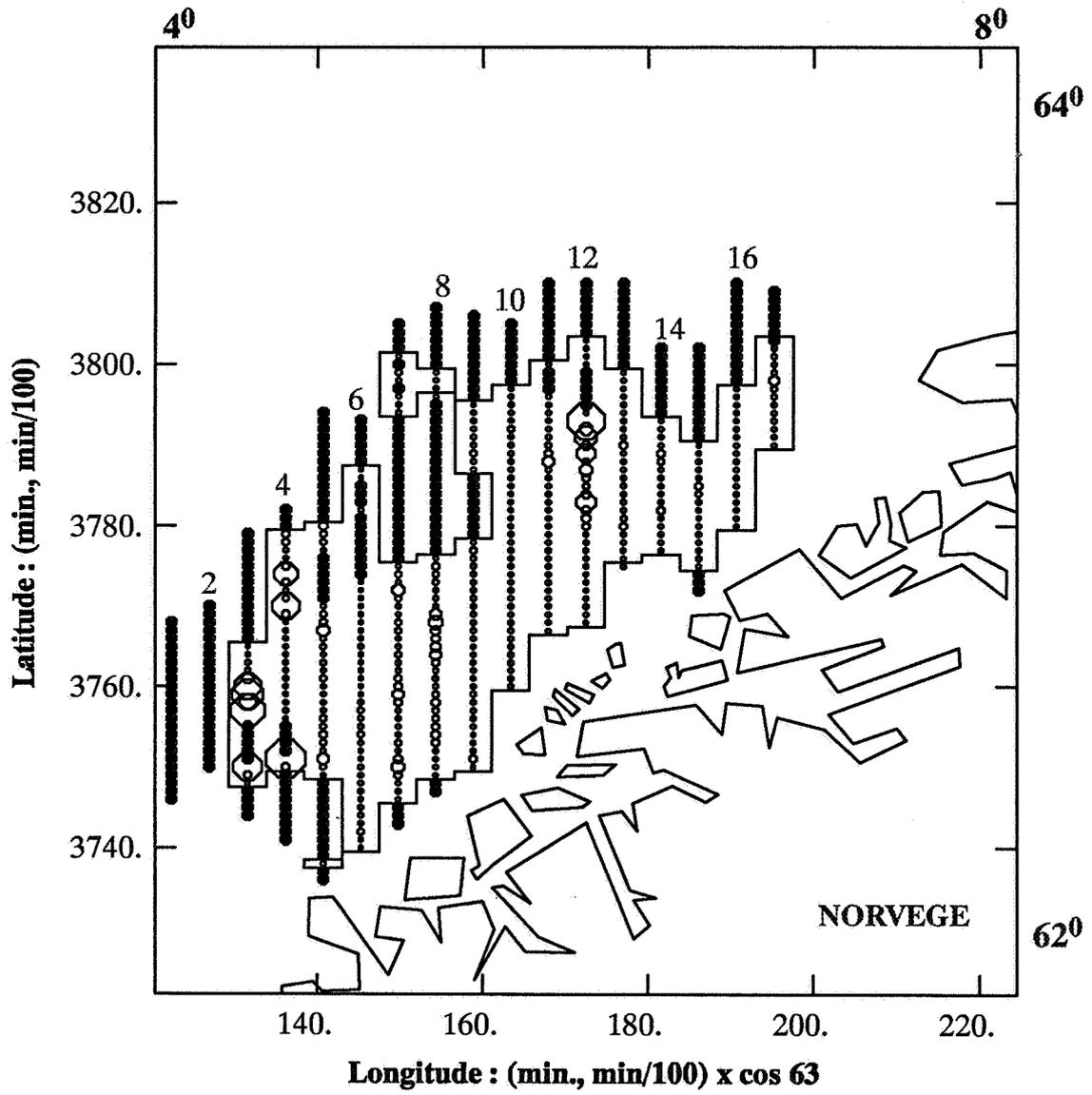


Fig. 8 : Le champ V de présence du hareng

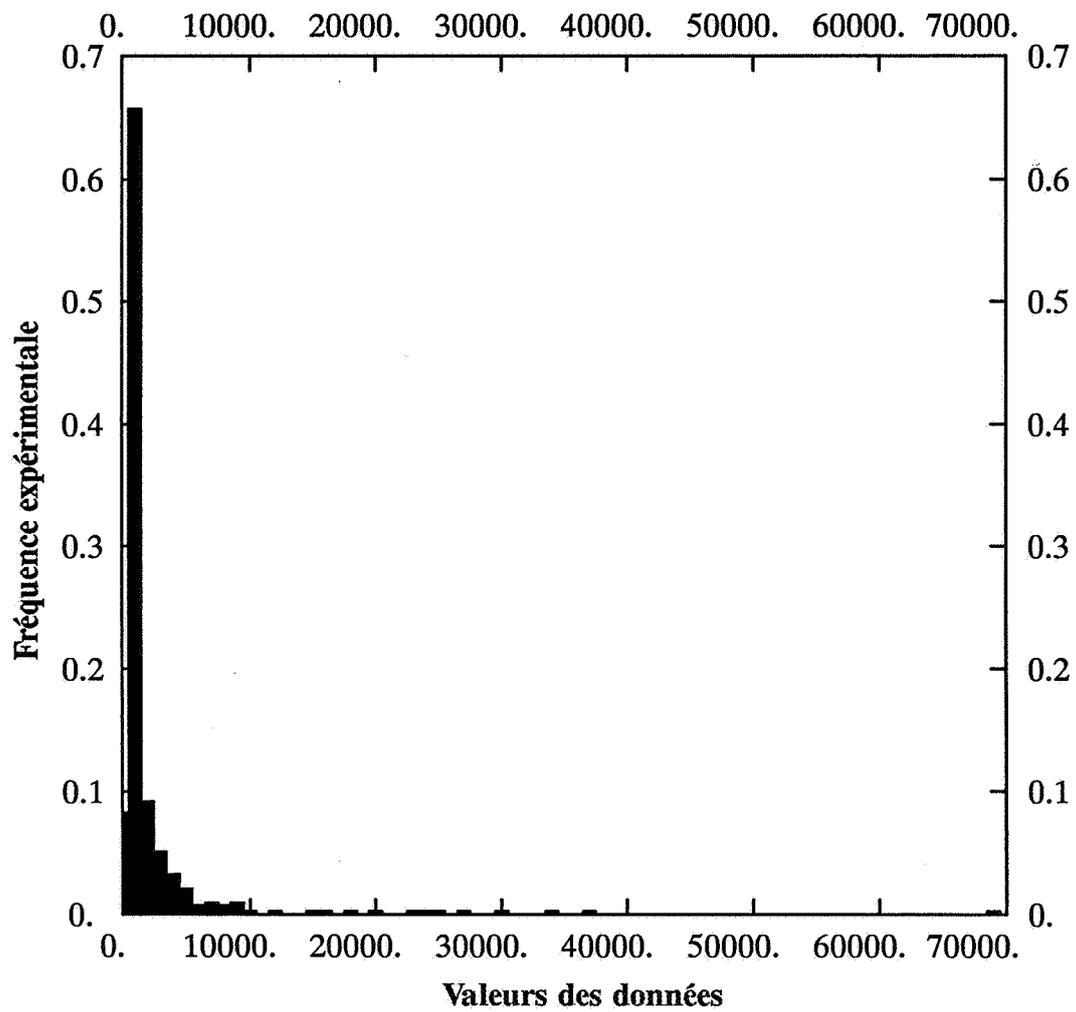


Fig. 9 : Histogramme des données dans V.

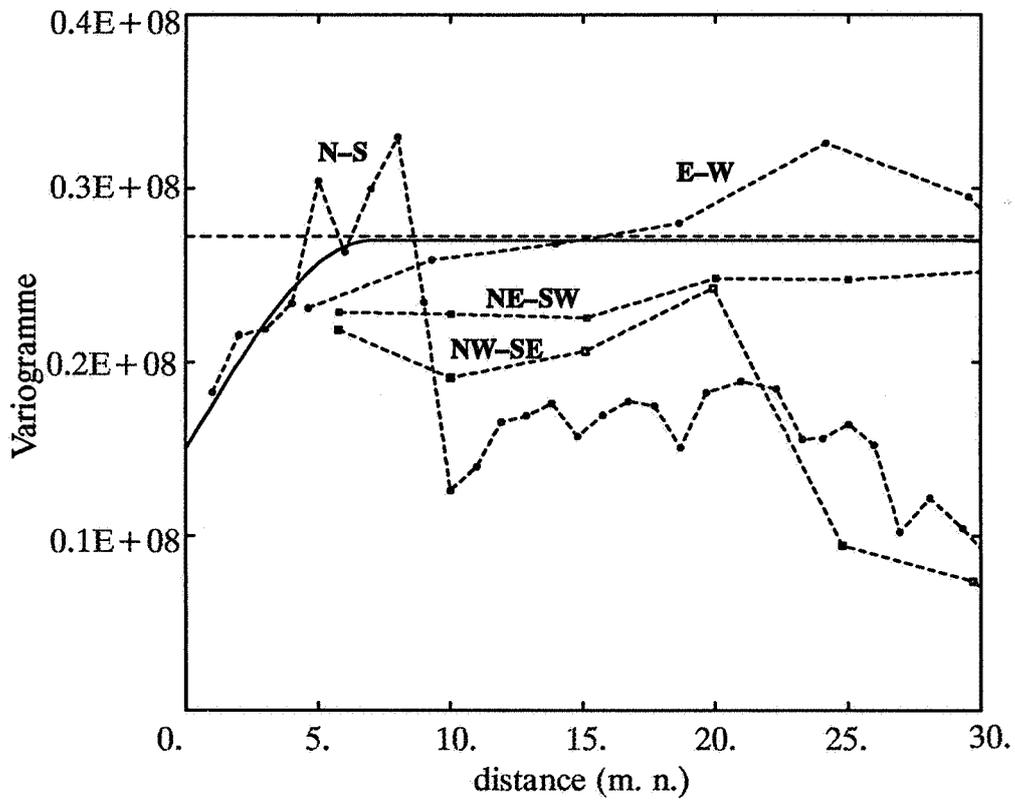


Fig. 10 : Variogramme des données brutes.
(le modèle est ajusté en s'aidant du résultat obtenu avec l'anamorphose gaussienne)

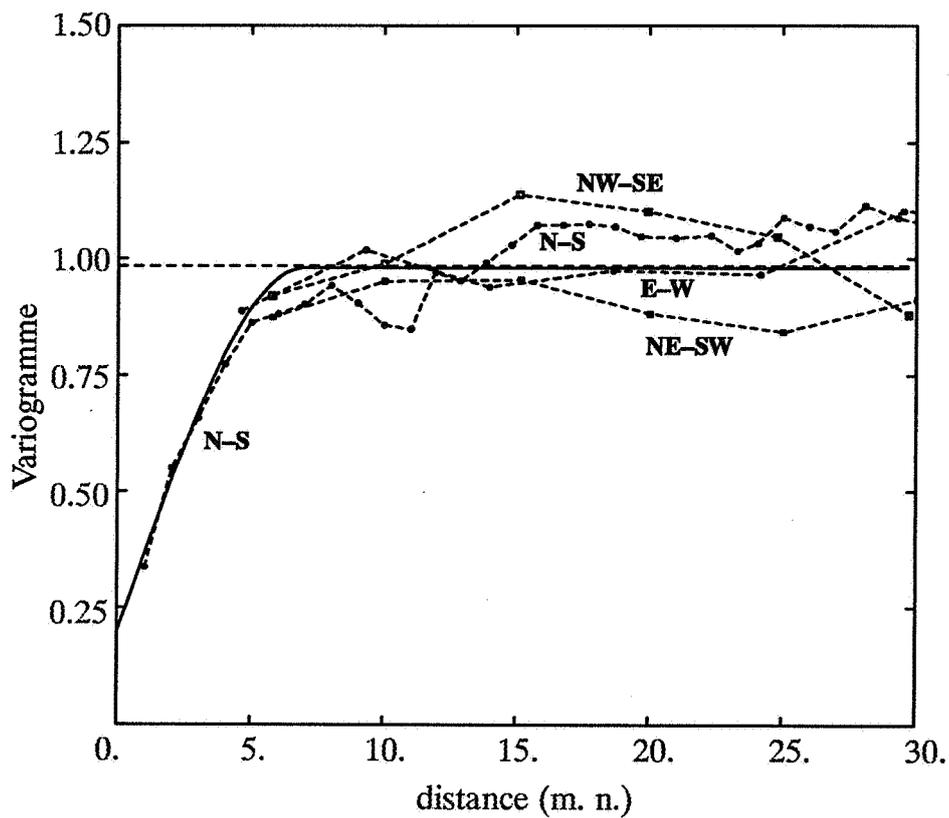


Fig. 11 : Variogramme des anamorphosées gaussiennes

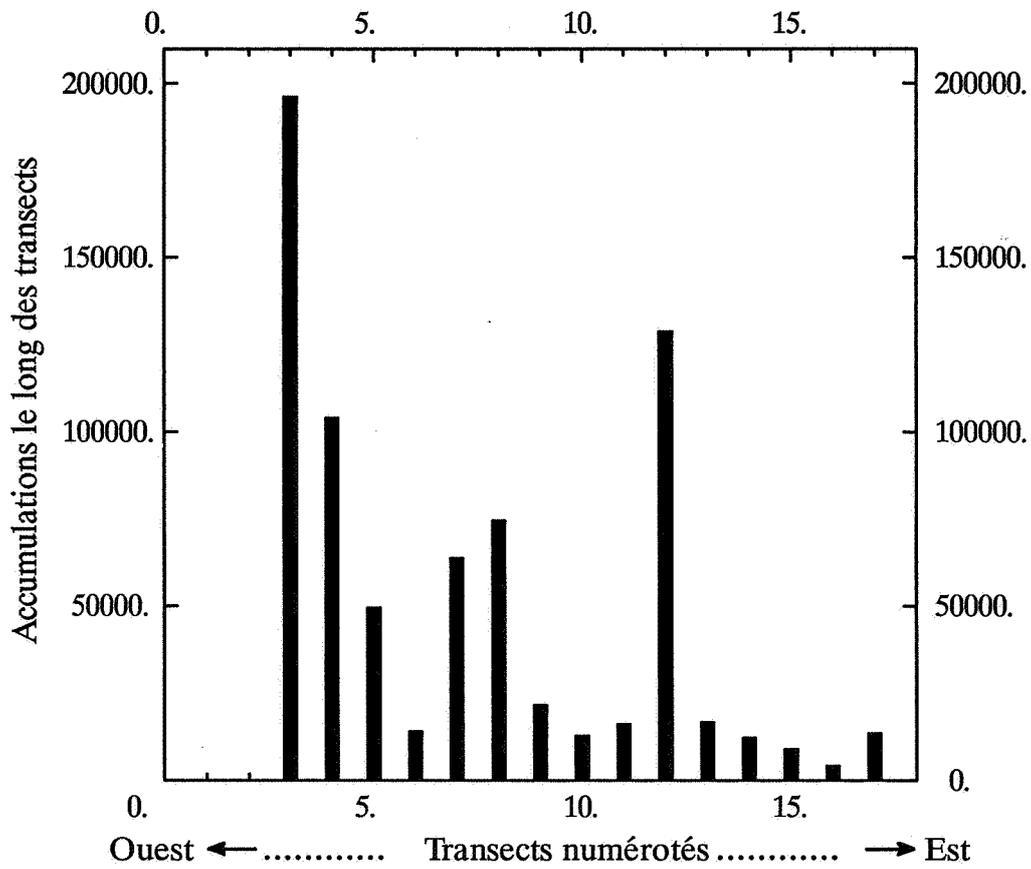


Fig. 12 : Représentation des cumuls $q(x)$ le long des transects

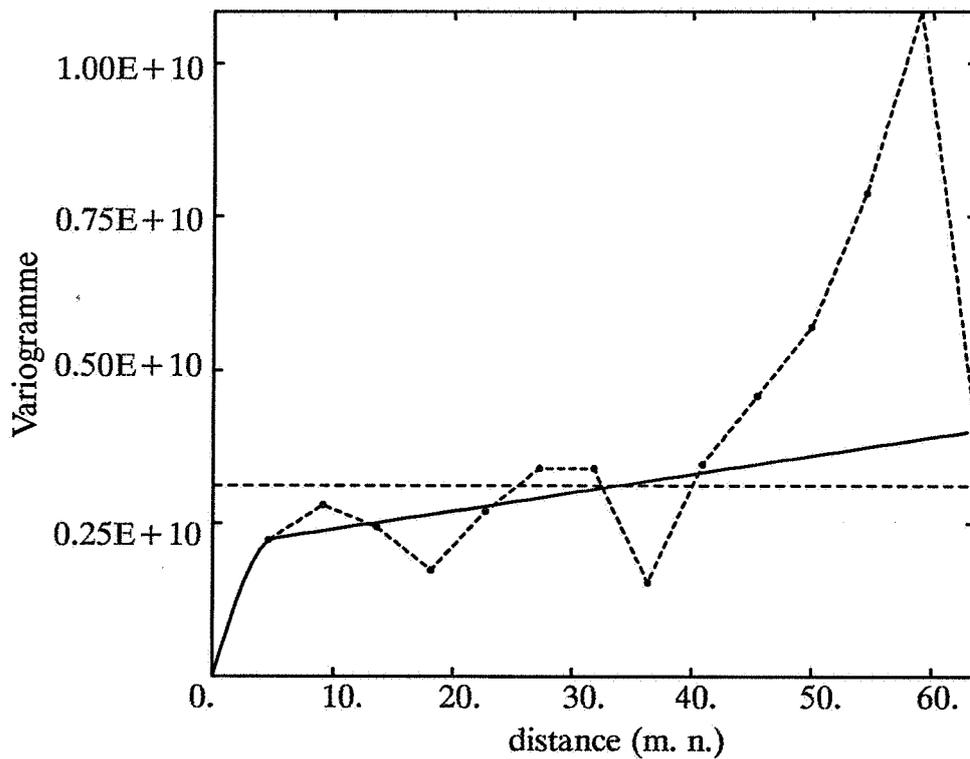


Fig. 13 : Variogramme 1D des $q(x)$

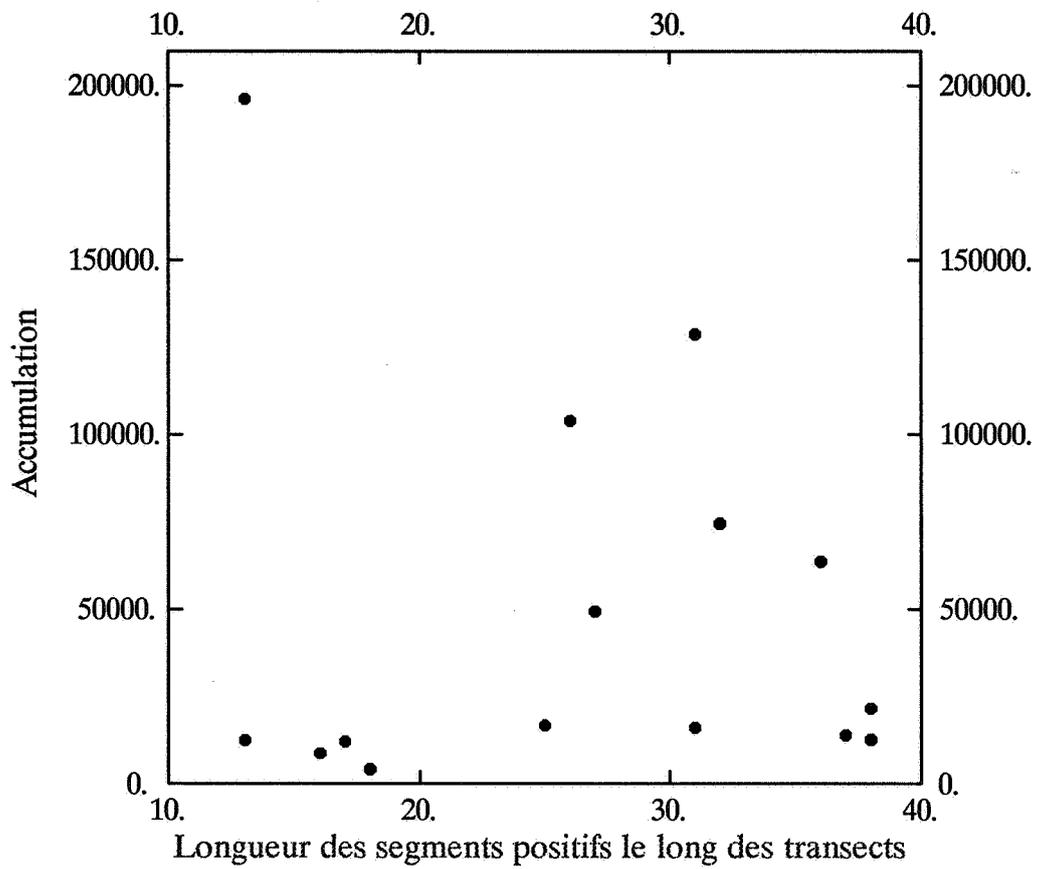


Fig. 14 : Nuage de corrélation entre les $q(x)$ et la longueur des traversées du champ V de présence

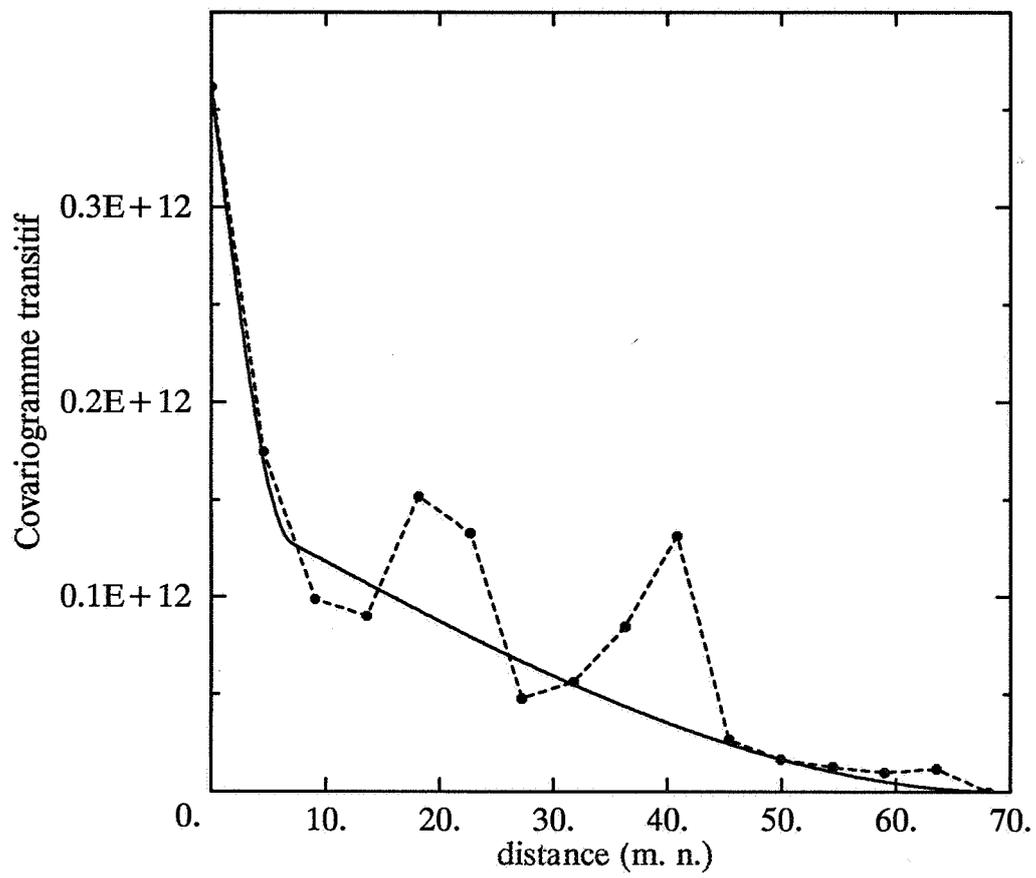


Fig. 15 : Covariogramme transitif des $q(x)$

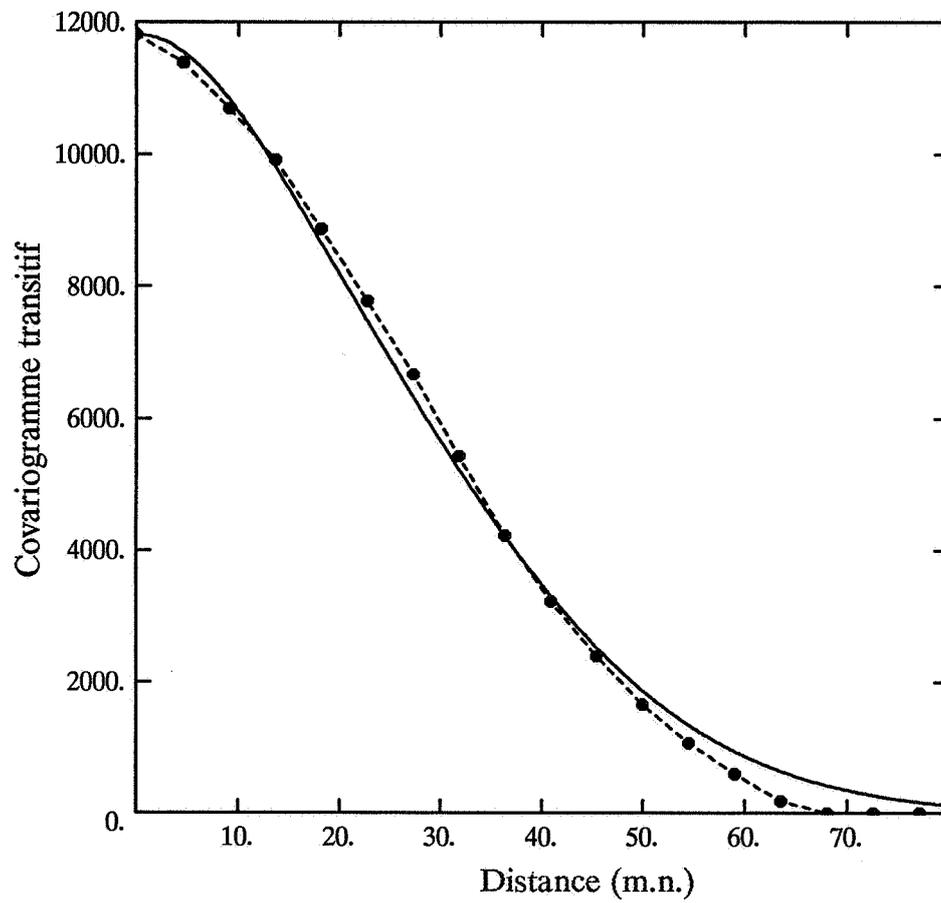


Fig. 16 : Covariogramme transitif de l'indicatrice du champ V montée le long des transects

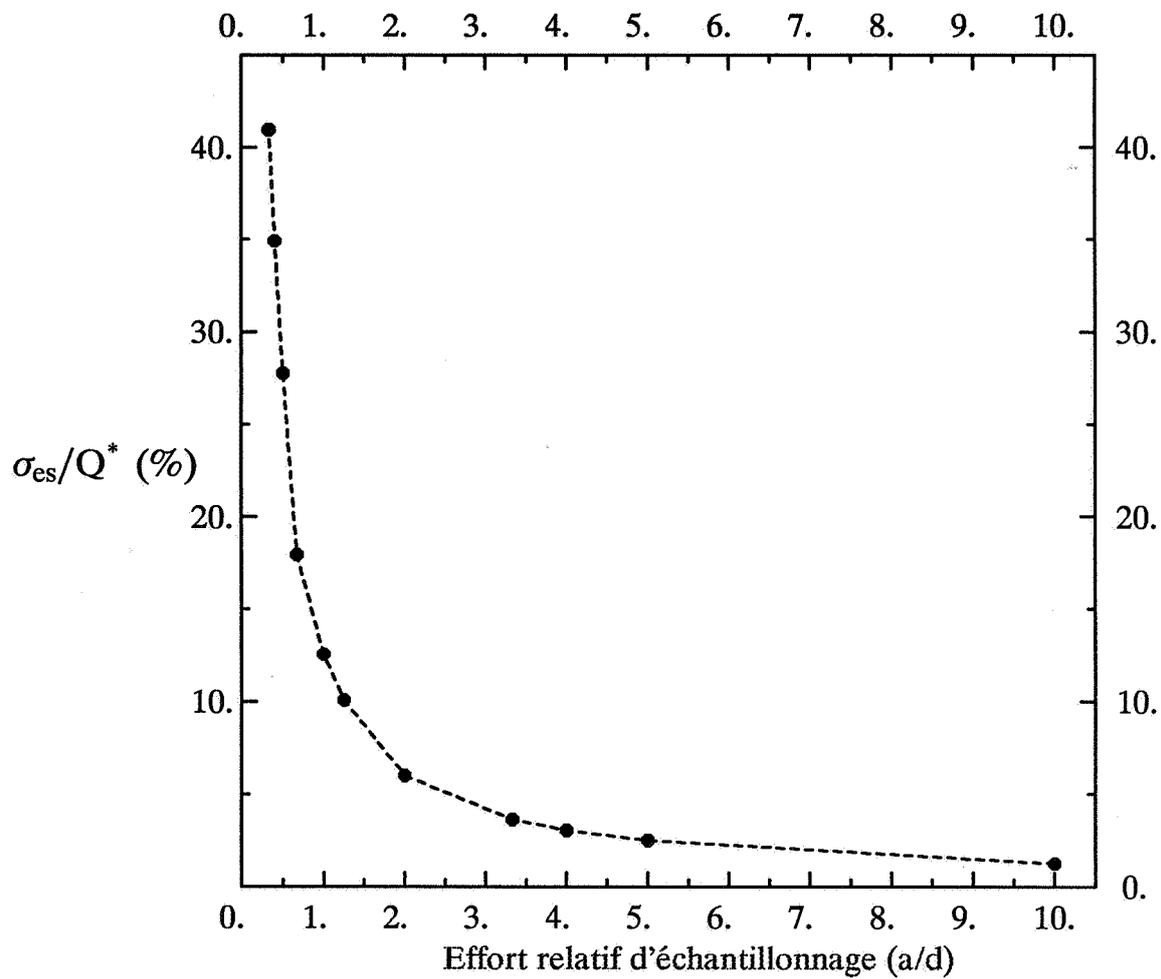


Fig. 17 : Evolution de la variance relative d'estimation en fonction de la distance inter-transect a. La distance d est celle de l'échantillonnage effectué.

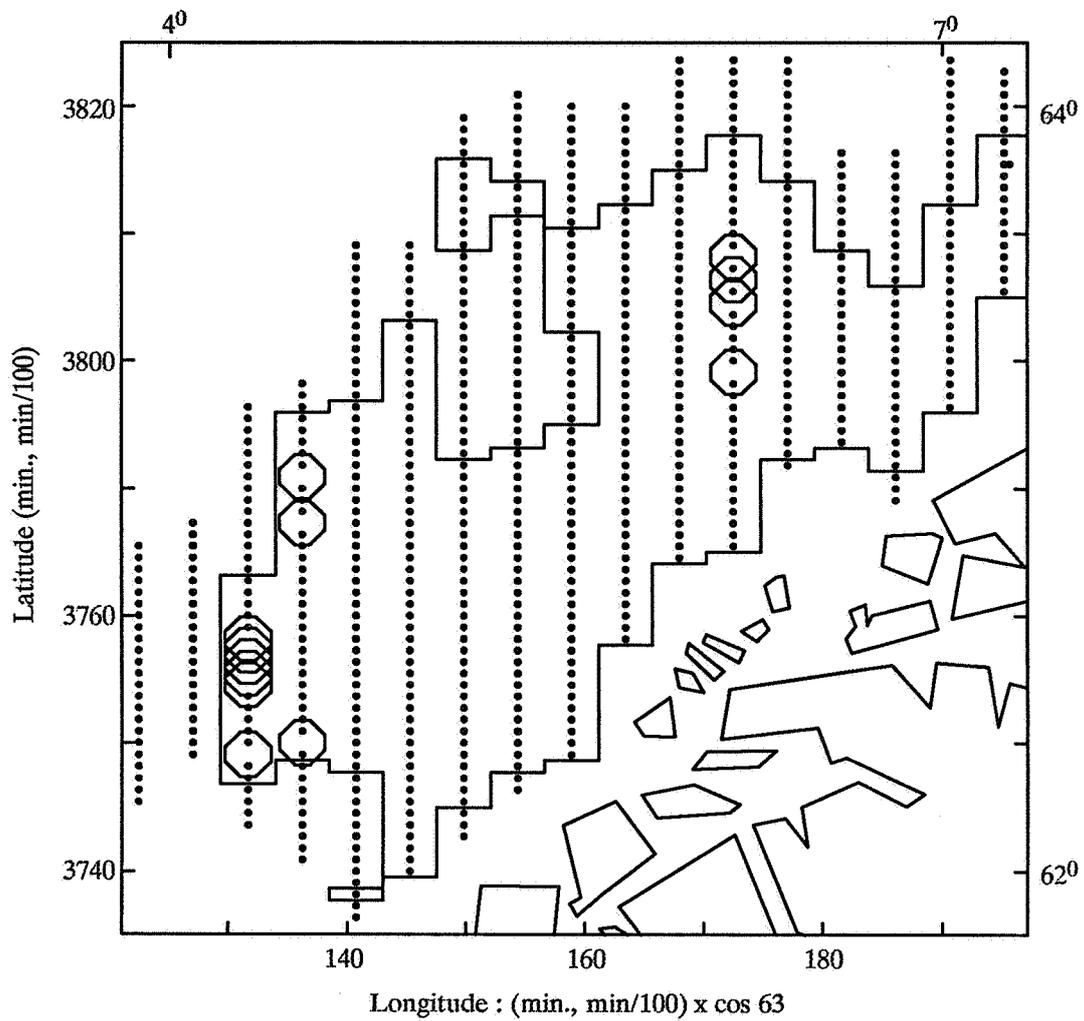


Fig. 18 : Représentation par des cercles des valeurs $z(x)$ supérieures à 10^4 dans le champ V

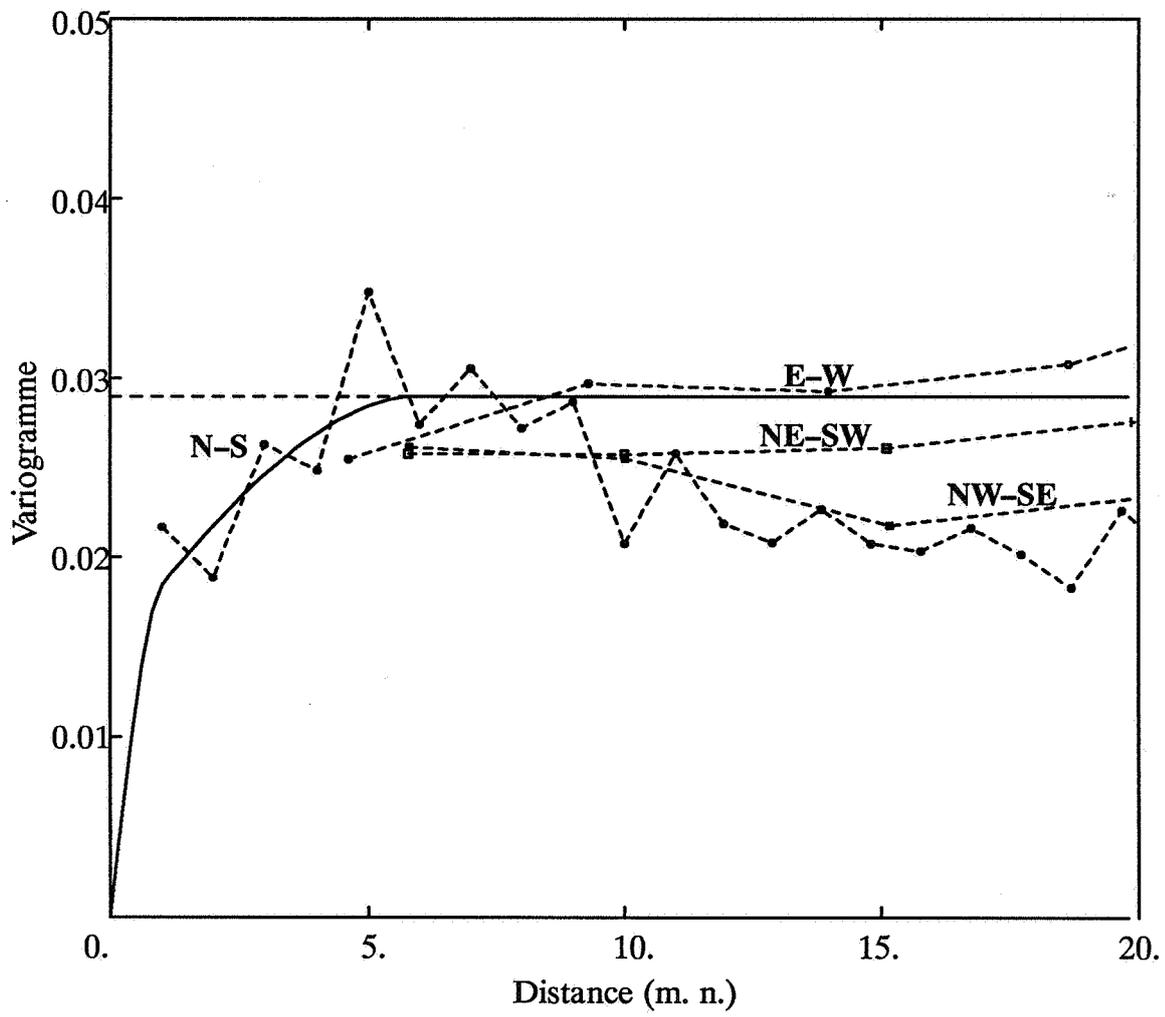


Fig.19 : Variogramme de $i_4(x)$

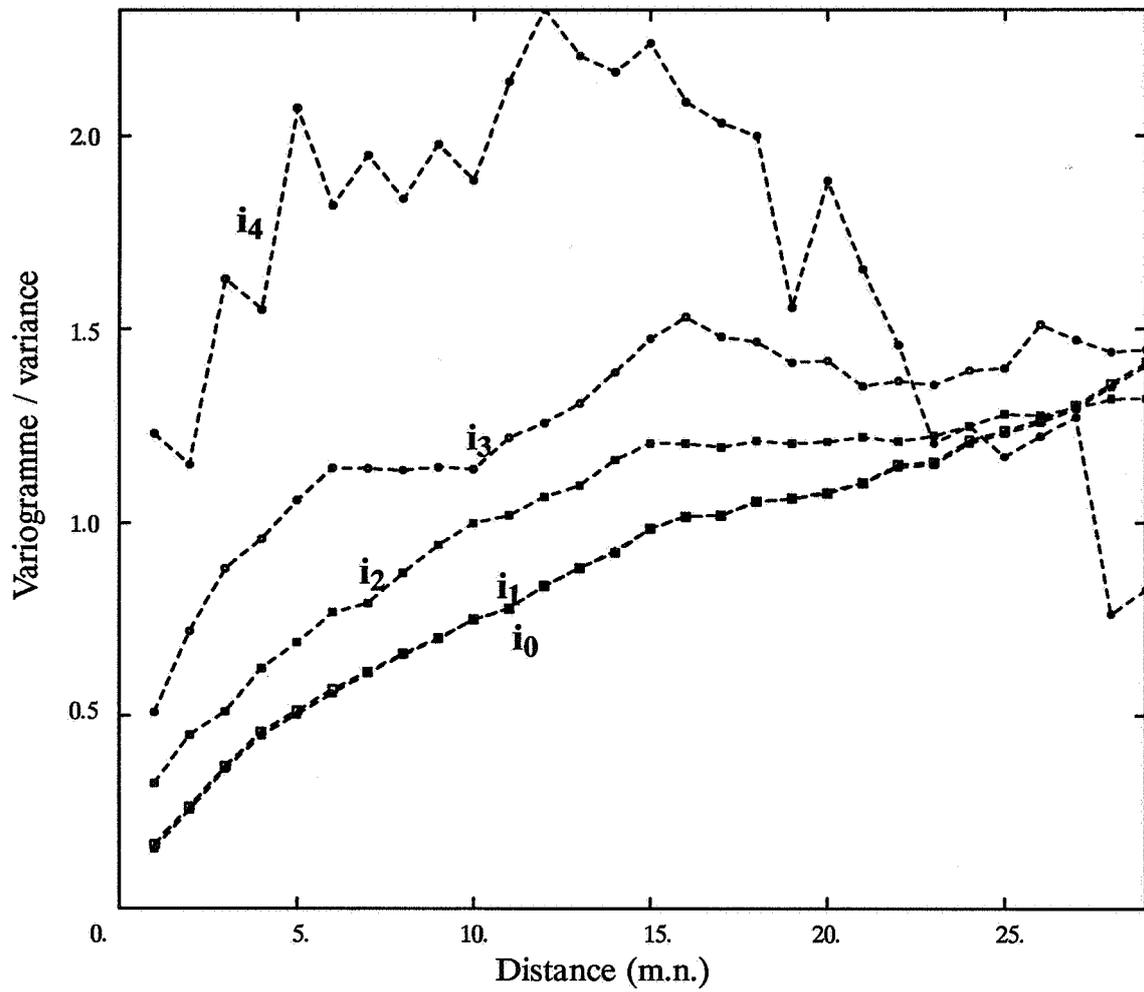


Fig. 20 : Variogrammes de 5 indicatrices. Ils sont calculés le long des transects (dir. N-S). Ils sont chacun normés par la variance de la variable. $i_k(x)$ est l'indicatrice pour la coupure 10^k : $i_k(x) = 1$ si $Z(x) \geq 10^k$; $i_k(x) = 0$ sinon .

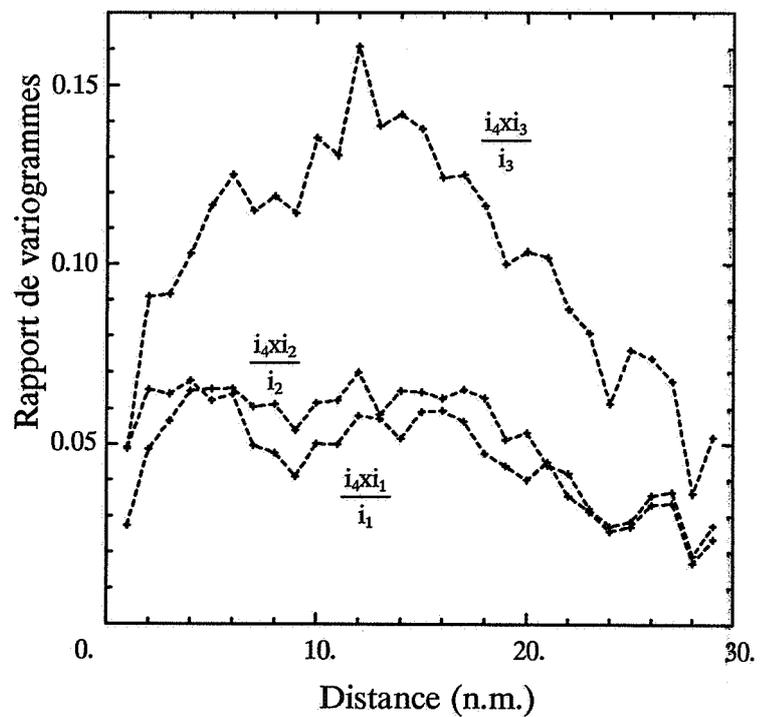
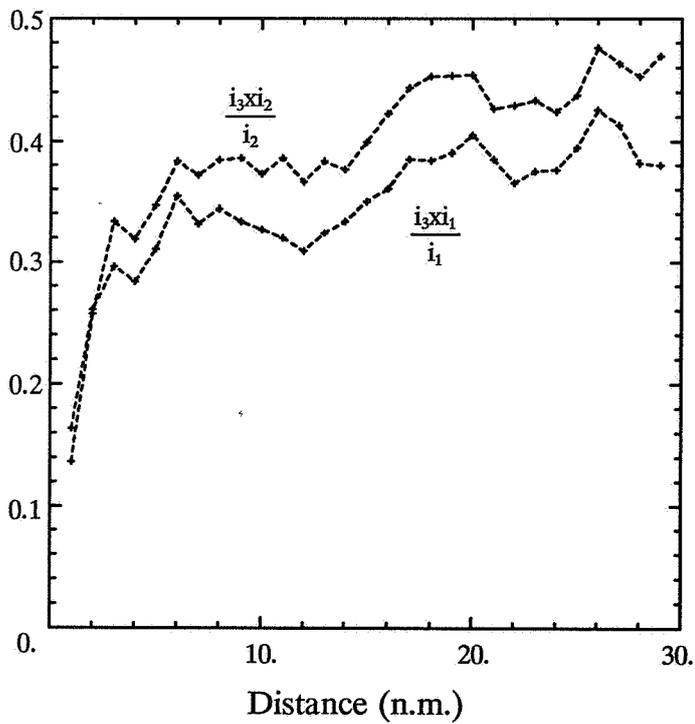
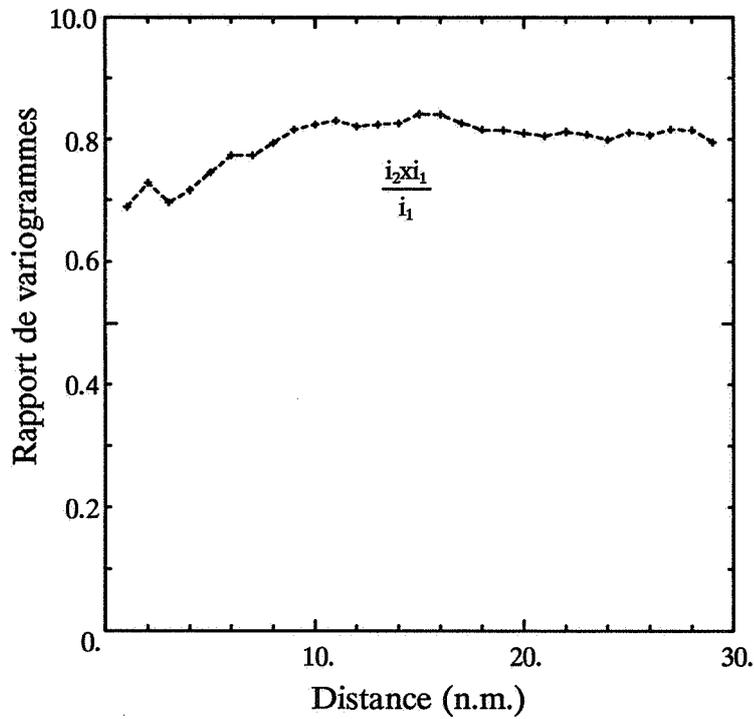


Fig. 21 : Test de l'existence d'effets de bord. Rapport du variogramme croisé entre 2 indicatrices sur le variogramme de l'indicatrice dont la coupure est la plus basse. (i_k est l'indicatrice de $Z(x) \geq 10^k$)

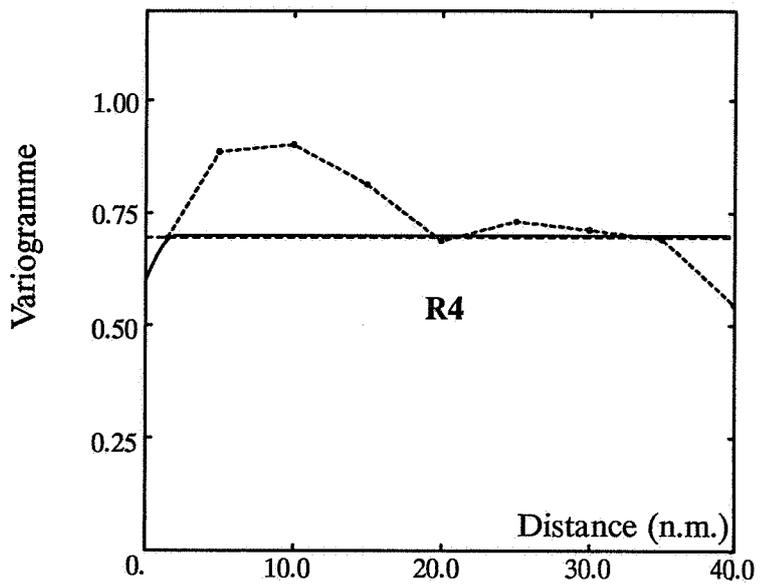
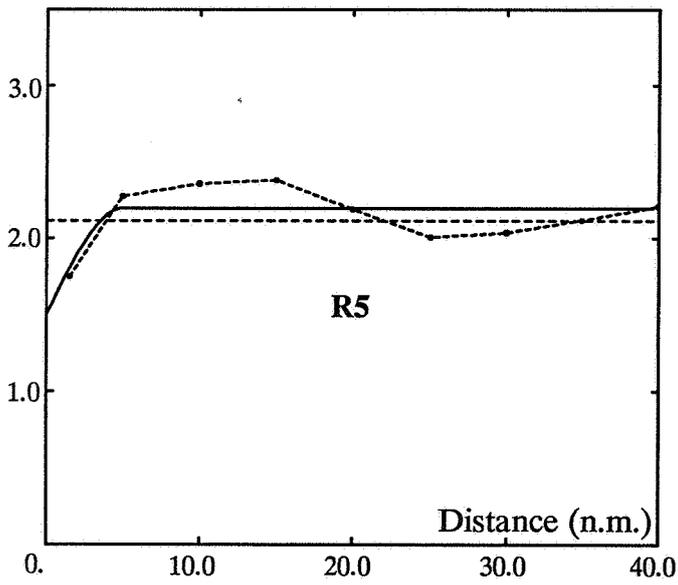
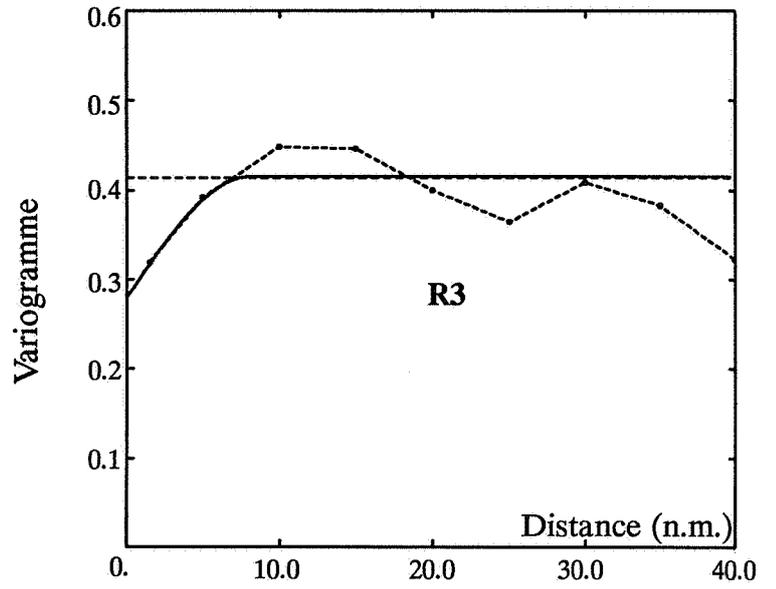
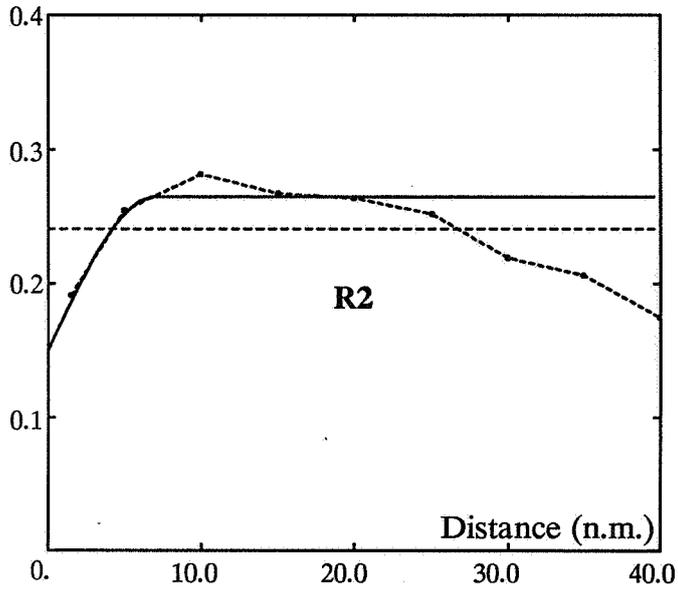
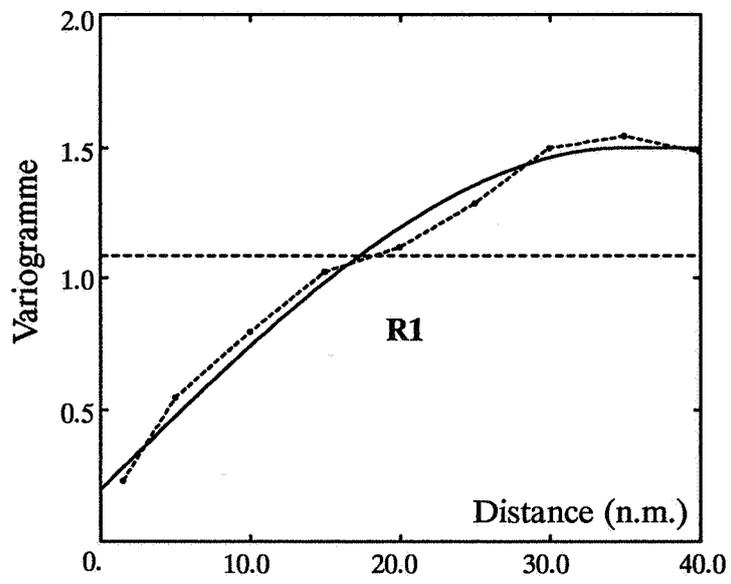


Fig. 22 : Variogrammes des résidus d'indicatrices (R1 à R5)

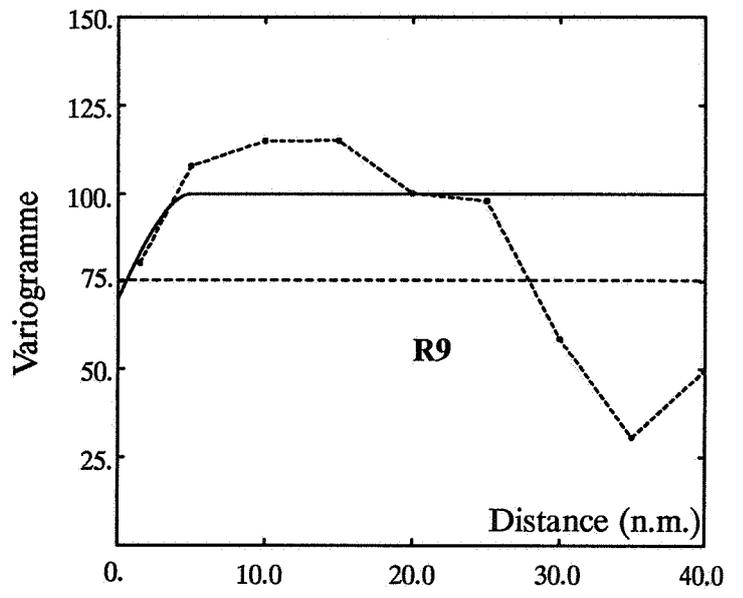
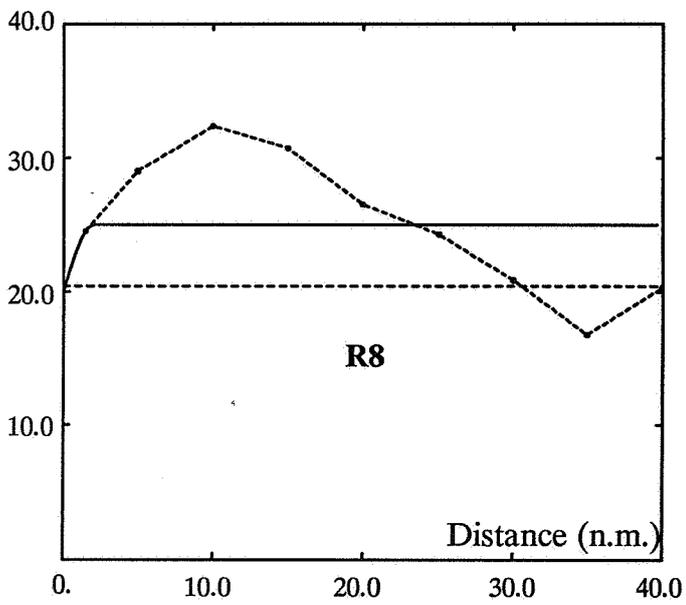
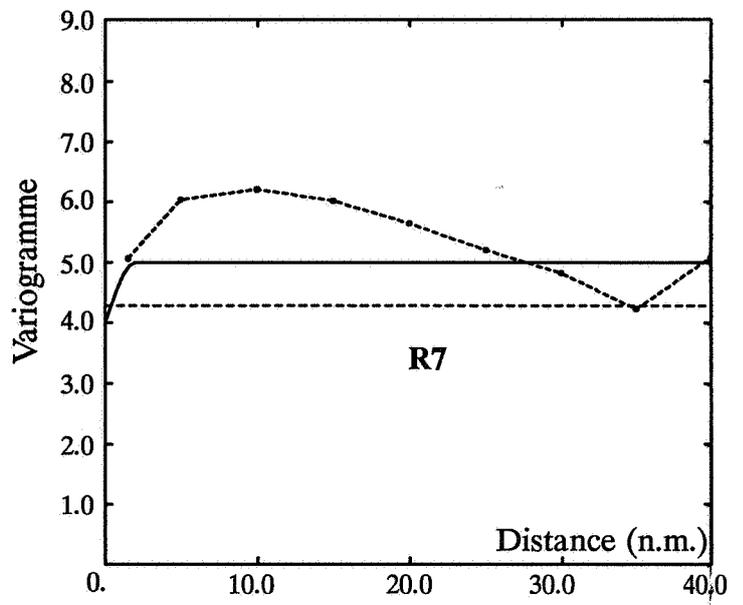
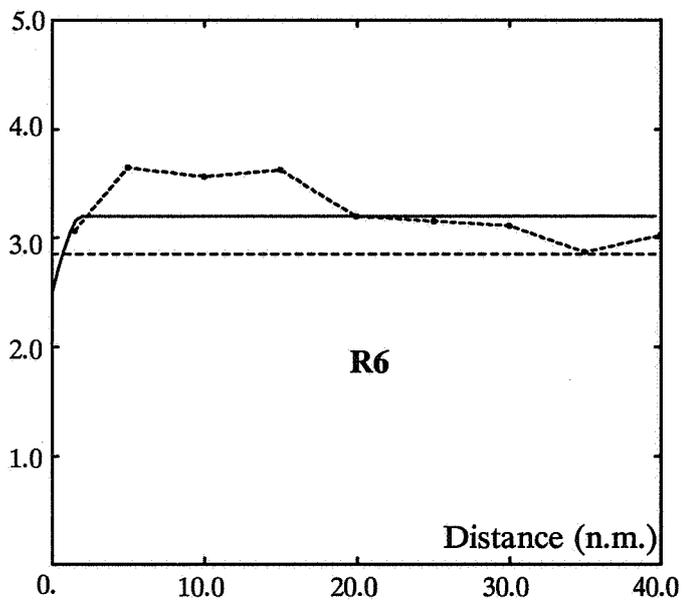


Fig. 23 : Variogrammes des résidus d'indicatrices (R6 à R9)

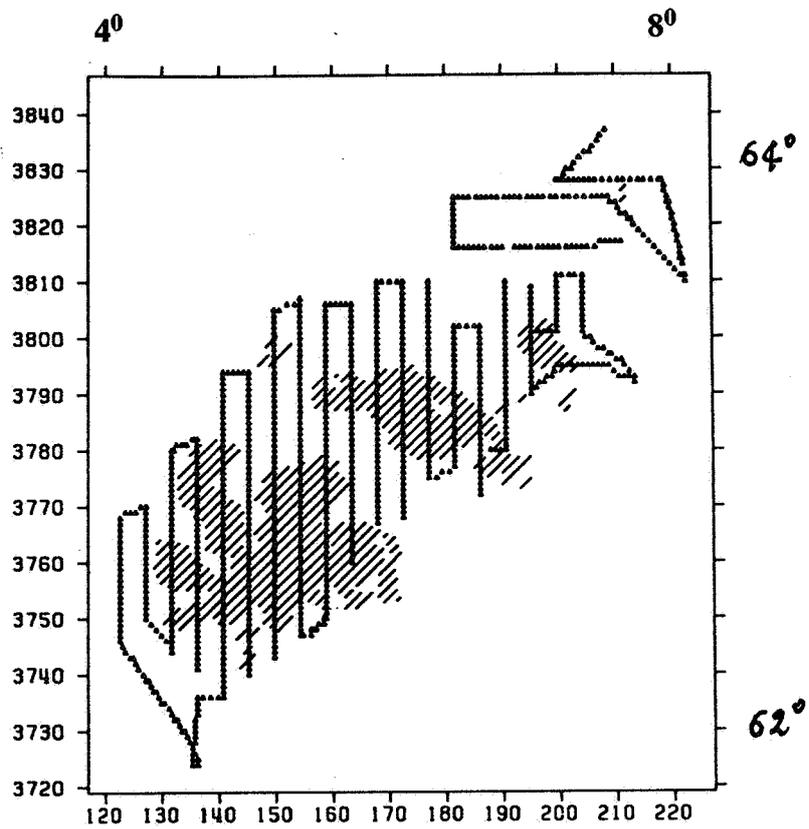


Fig. 24 : Estimation par krigeage disjonctif de l'aire A_{500} représentant le "gisement de hareng".

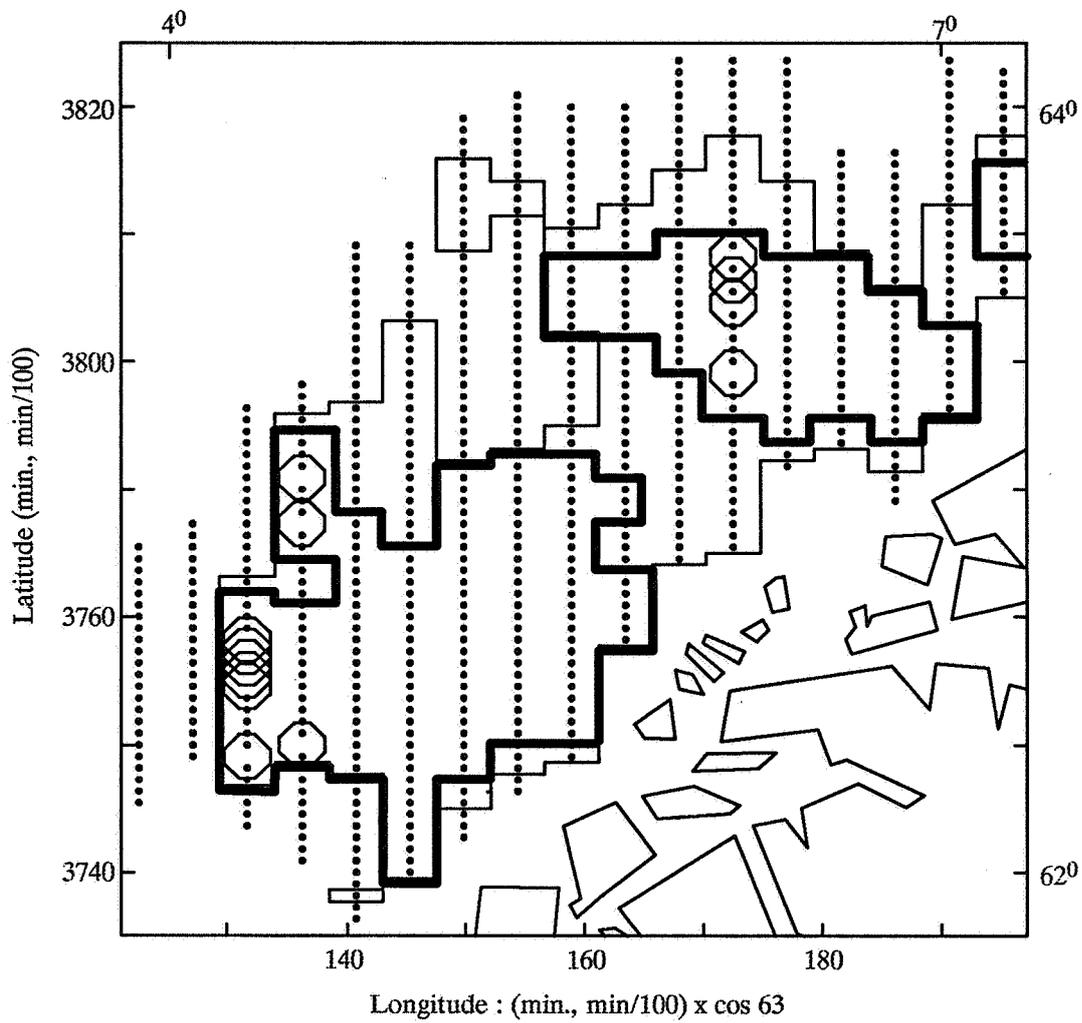


Fig. 25 : Représentation de l'indicatrice $i_4(x)$ dans le champ V et dans l'aire A_{500}

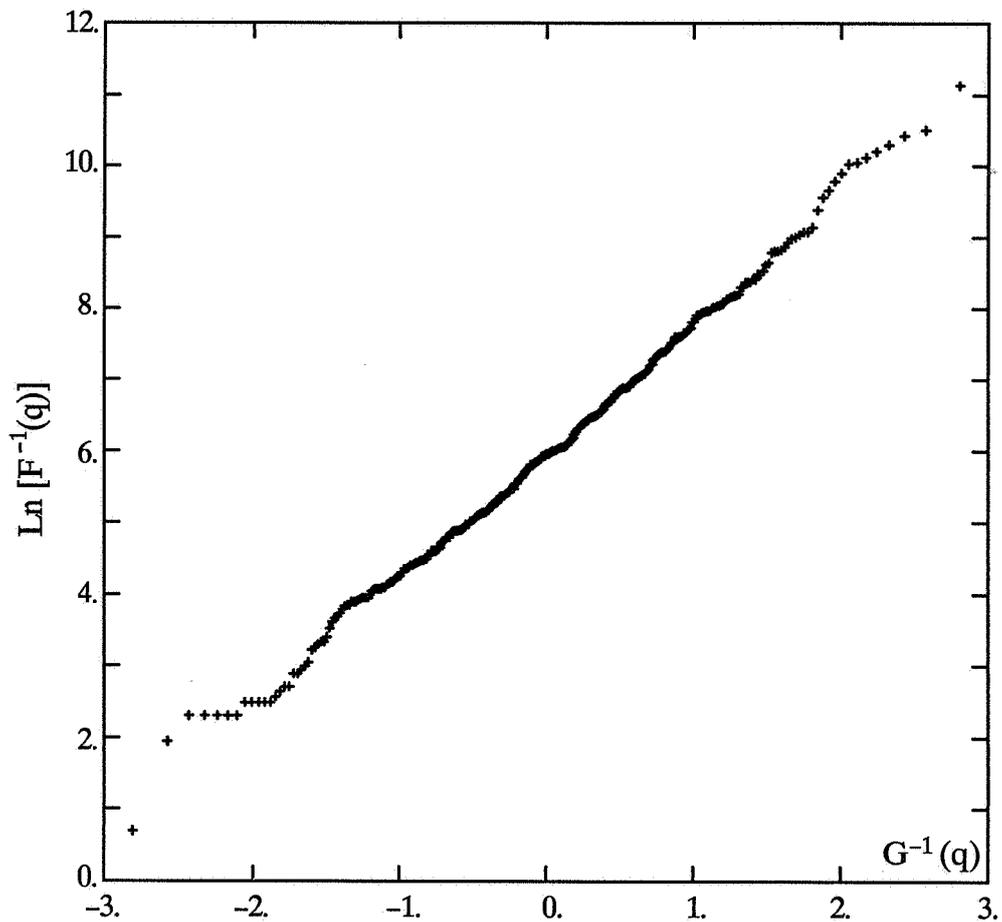


Fig. 26 : Droite de Henry sur les valeurs non nulles.
 G^{-1} est l'inverse de la fonction de répartition gaussienne (valeurs théoriques).
 F^{-1} est l'inverse de la fonction de répartition des données (valeurs expérimentales).

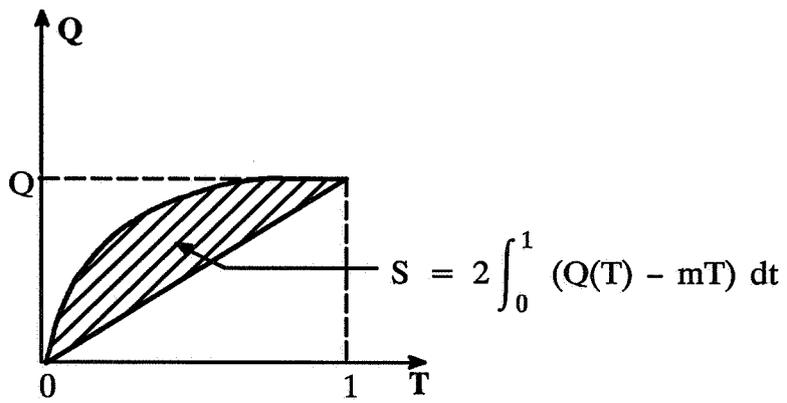


Fig. 27 : Courbe de sélectivité

Quantité totale et Occupation de l'espace

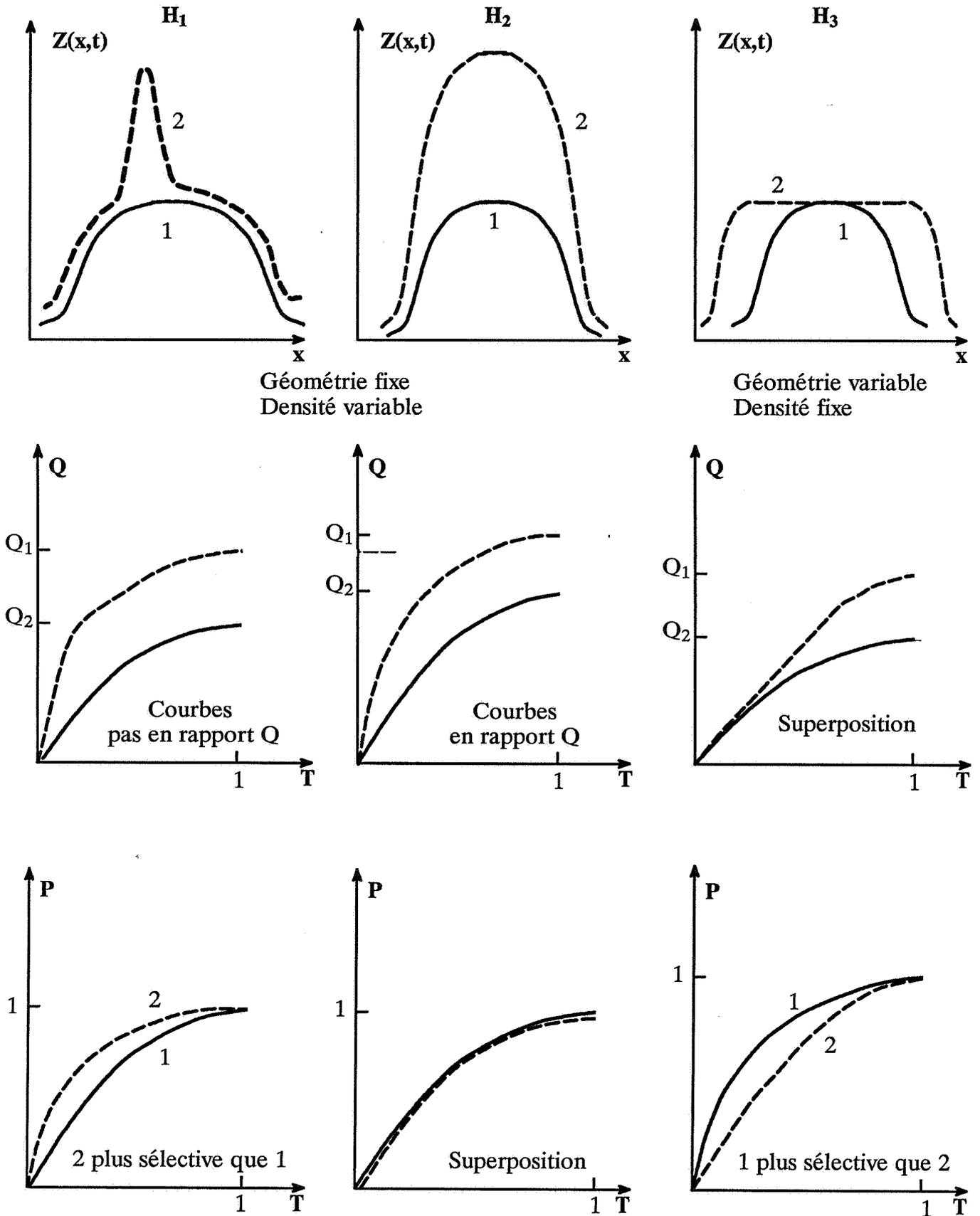


Fig. 28

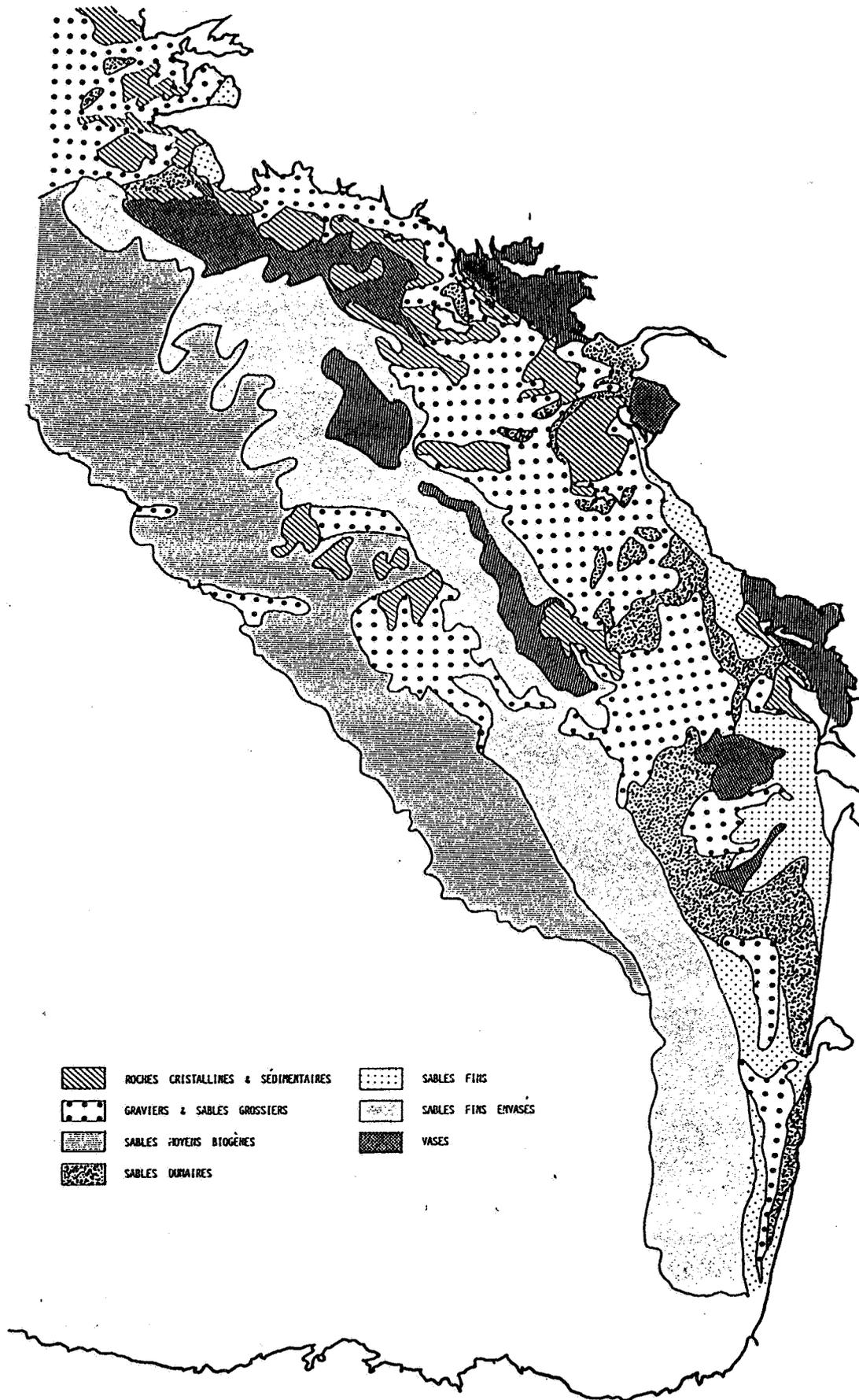


Fig. 29 : Carte biosédimentaire du plateau continental du Golfe de Gascogne.
 (d'après les données compilées par Glémarec en 1977) – in Poulard 1990

ANNEXE FIGURES

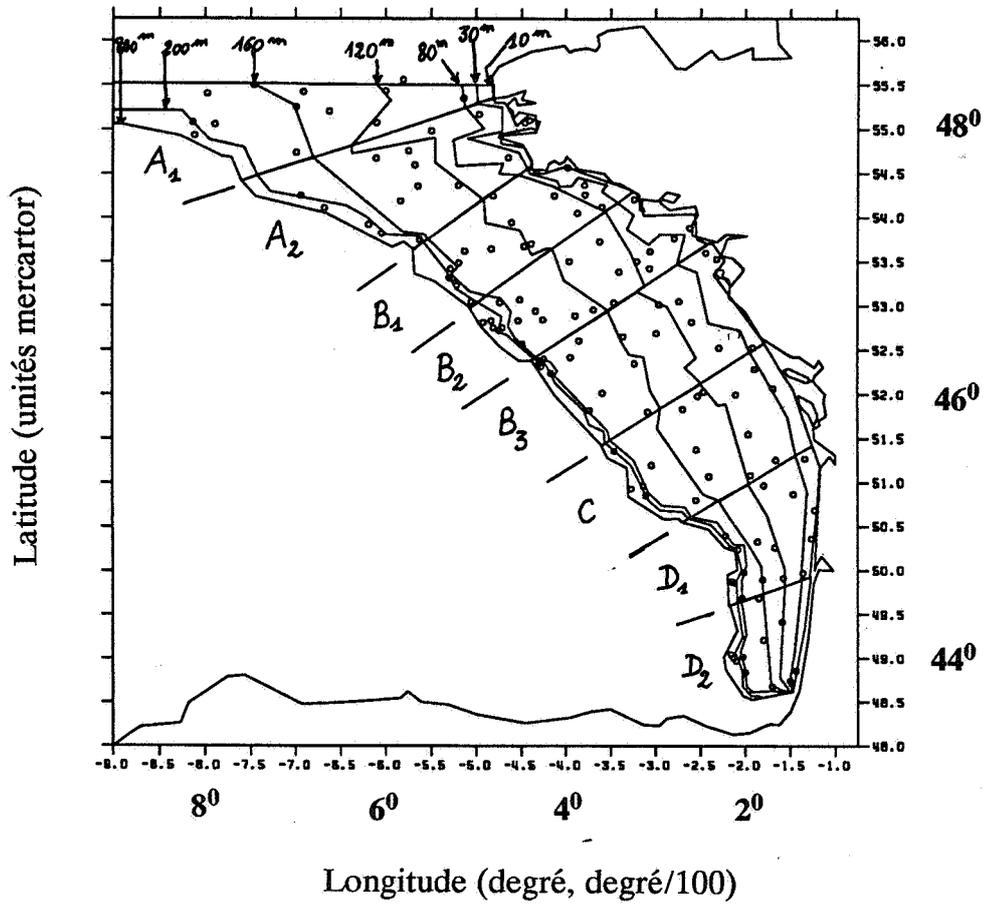


Fig. 30 : Le plan d'échantillonnage stratifié.
 (Les strates entre 200 et 800 m ont été regroupées pour la représentation)

Campagne de chalutages de fond du N/O Thalassa, automne 1988

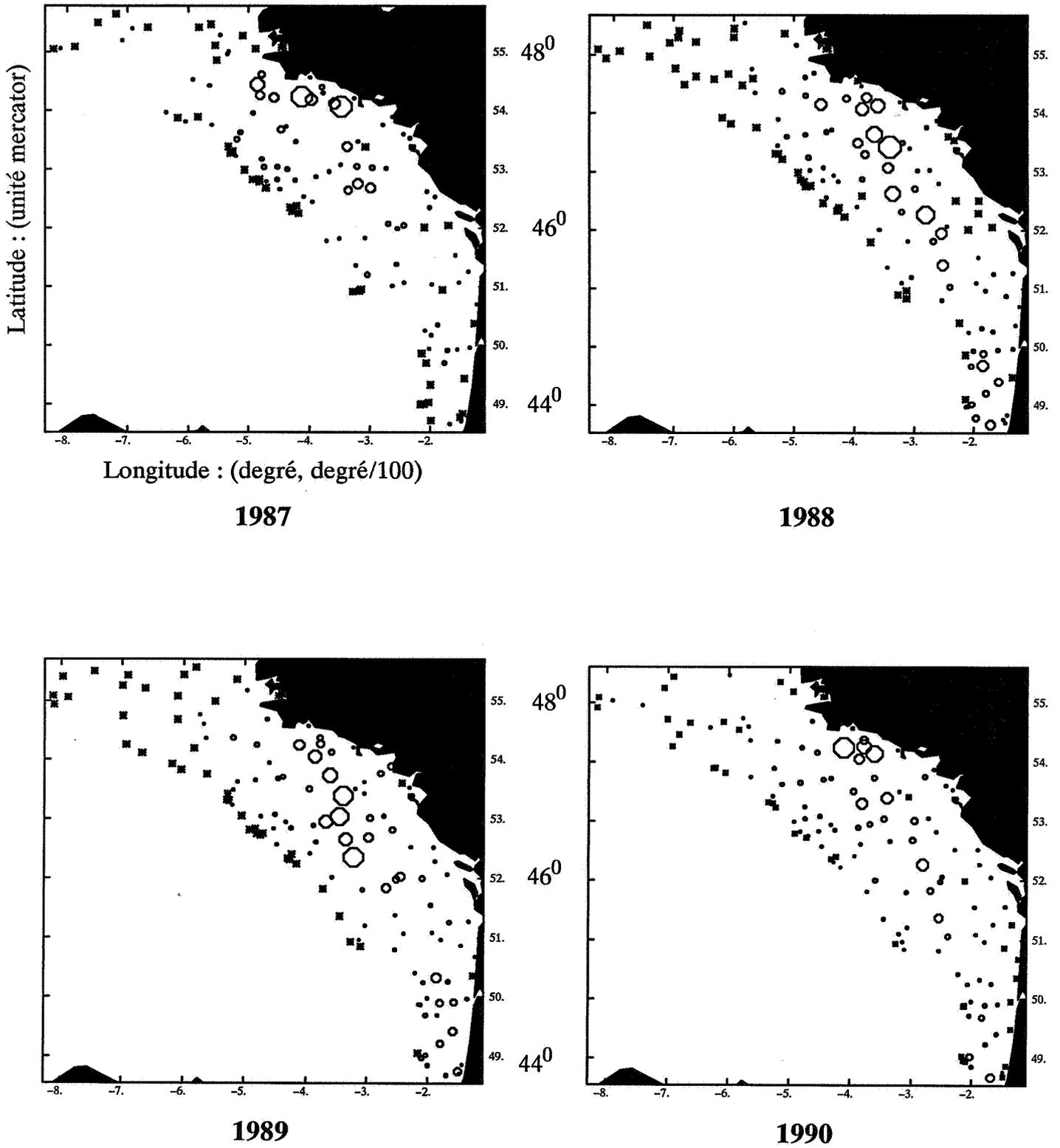


Fig. 31 : Représentation proportionnelle des pêches de merlu-0

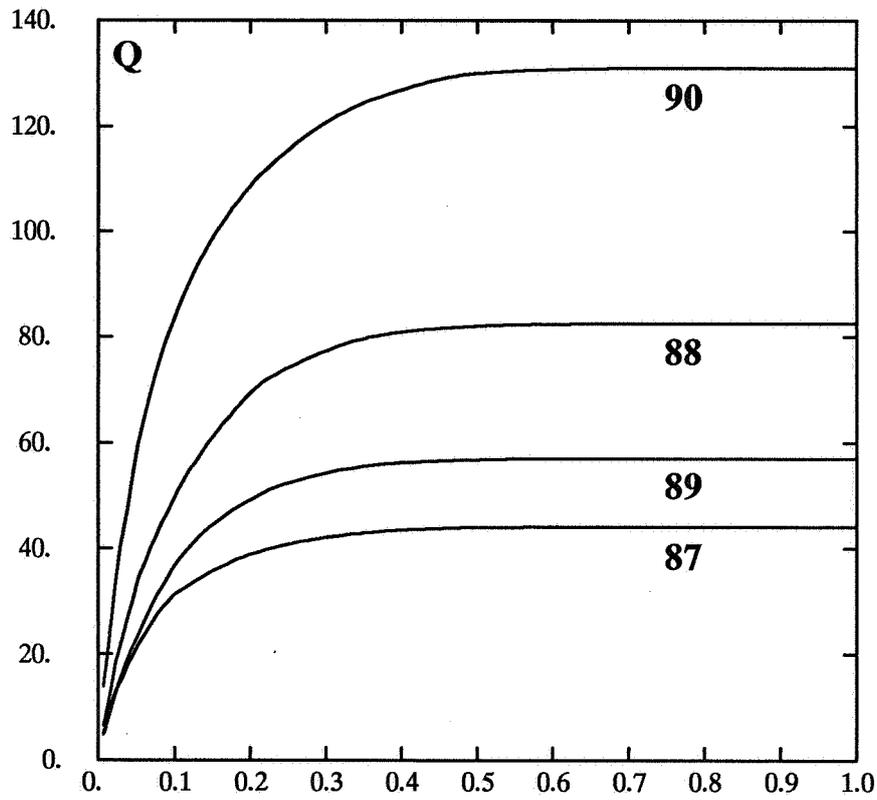
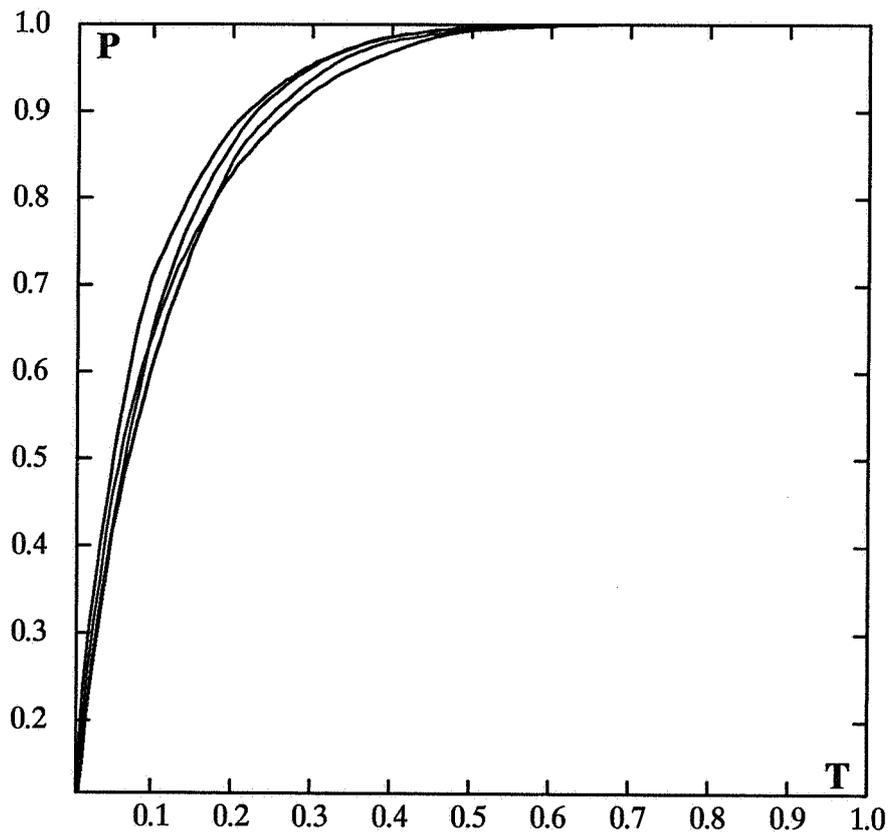


Fig. 32 : Courbes $Q(T)$ et $P(T)$ du merlu-0 sur l'ensemble du Golfe de Gascogne



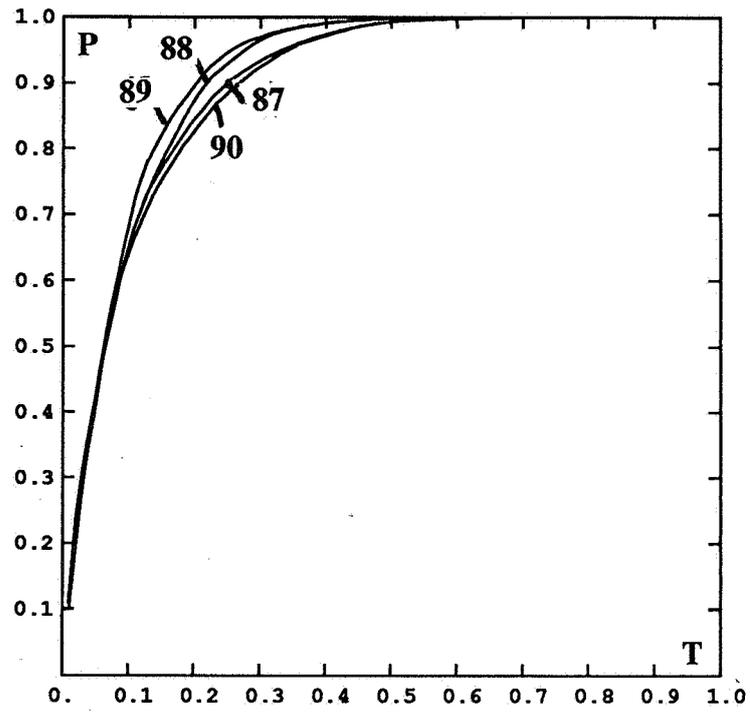


Fig. 33 : Courbes P(T) du merlu-0 dans le Nord du Golfe de Gascogne (secteurs A2,B,C, de 15 à 200m)

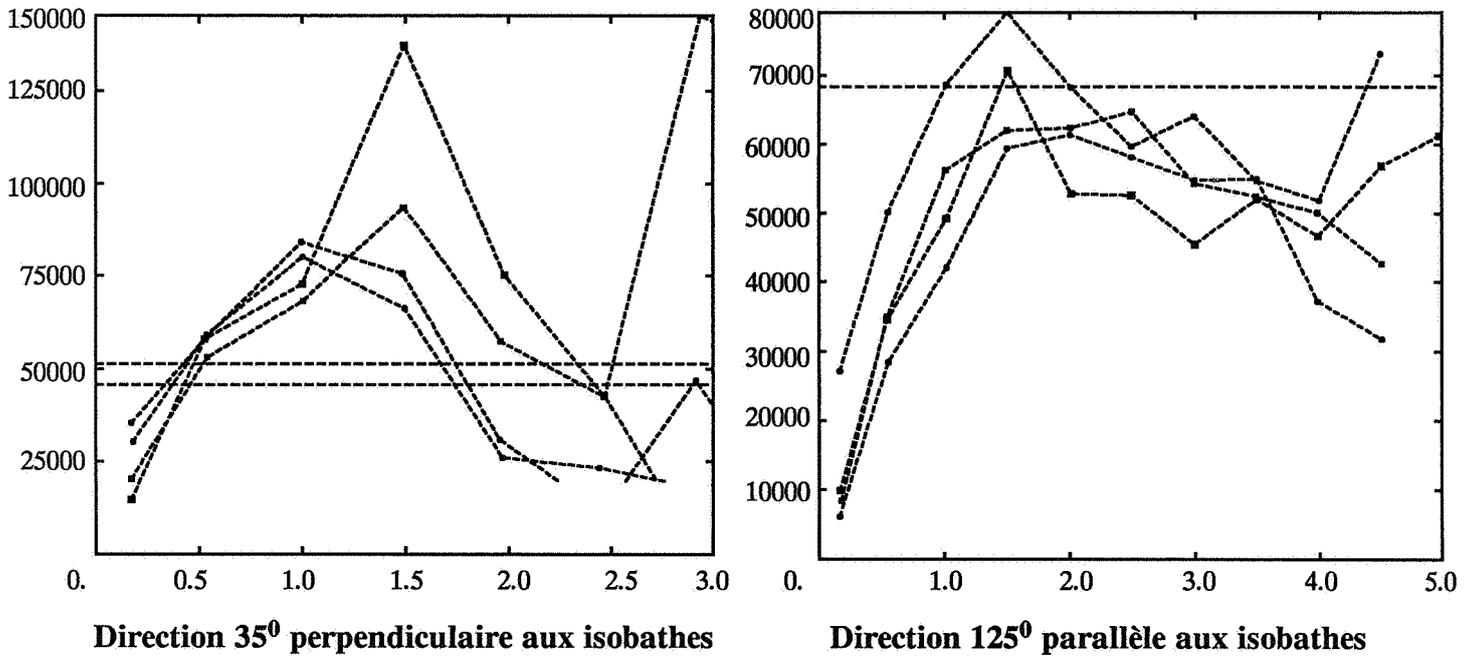


Fig. 34 : Merlu-0. Superposition des variogrammes annuels des pourcentages de biomasse $p(x,t)$

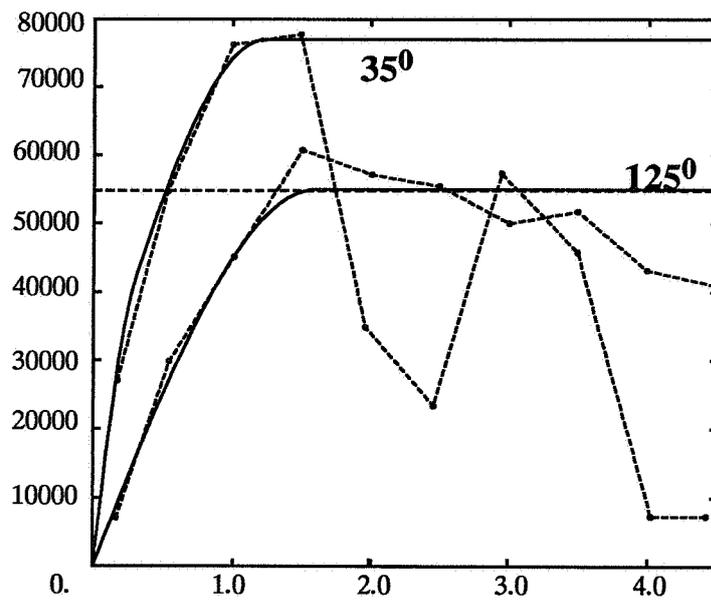


Fig. 35 : Merlu-0. Variogramme moyen de $P(x,t)$ sur les 4 années

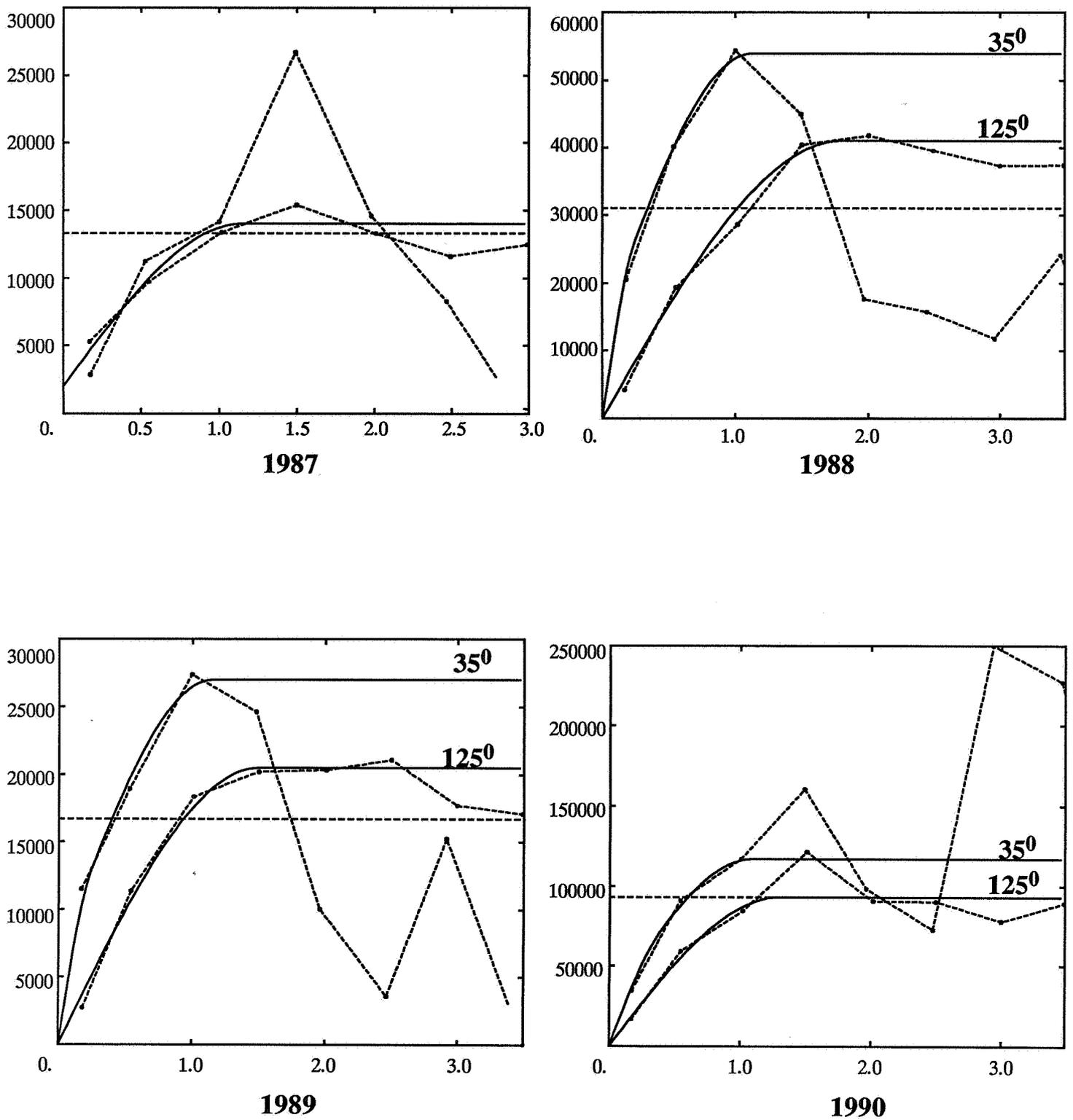


Fig. 36 : Merlu-0. Variogrammes expérimentaux et leurs modèles pour chaque année

Année	Pépite	Nb de modèles	Portée	Palier	Coefficient d'anisotropie		Rotation
					x	y	
87	0.2 E5	1	1.25	0.12 E+5	1.0	1.0	35°
88	-	2	1.8	0.41 E+5	1.6	1.0	35°
			0.25	0.13 E+5	1.0	0.0	
89	-	2	1.5	0.21 E+5	1.3	1.0	35°
			0.25	0.07 E+5	1.0	0.0	
90	-	2	1.3	0.93 E+5	1.2	1.0	35°
			0.4	0.24 E+5	1.0	0.0	
Variogramme moyen de $\frac{Z(x, t)}{m(t)}$	-	2	1.6	0.55 E+5	1.6	1.0	35°
			0.3	0.22 E+5	1.0	0.0	

(Portée en unités mercator : 0.25 = 10 mi.n.)

Tableau 1 : Paramètres des modèles sphériques des variogrammes du merlu-O, pour les différentes années

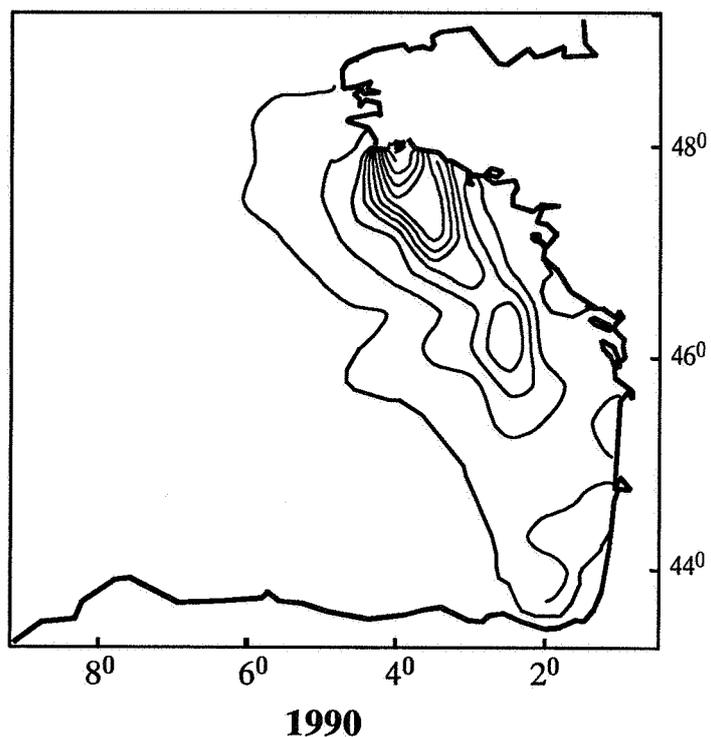
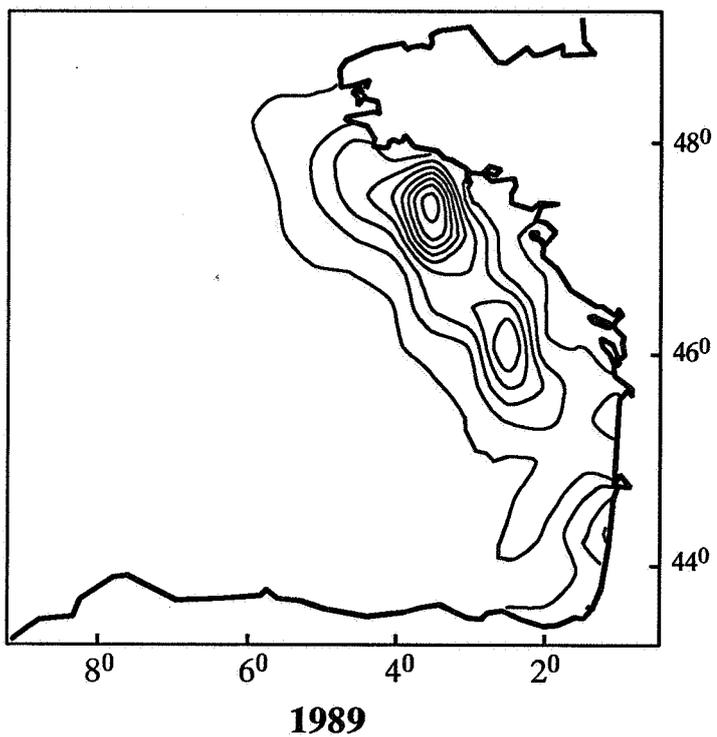
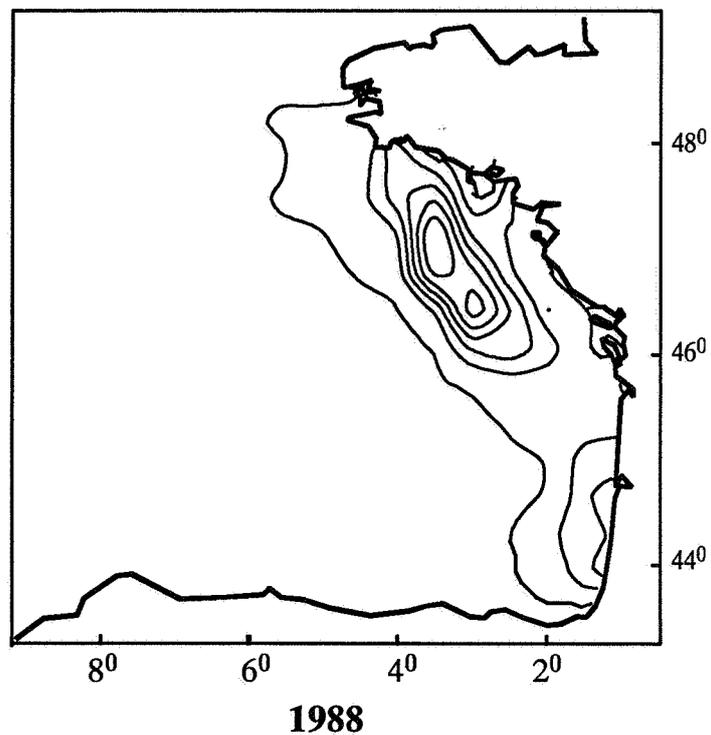
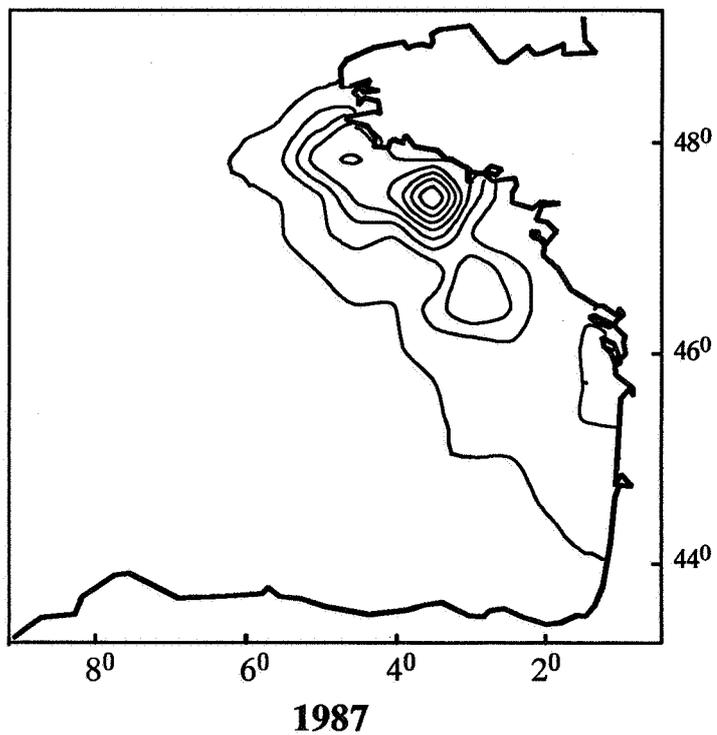


Fig. 37 : Merlu-0. Cartes des krigeages de la densité moyenne par bloc de 10 x 10 m. n.
 1987 et 1989 : isolines tous les 50
 1988 et 1990 : isoline tous les 100

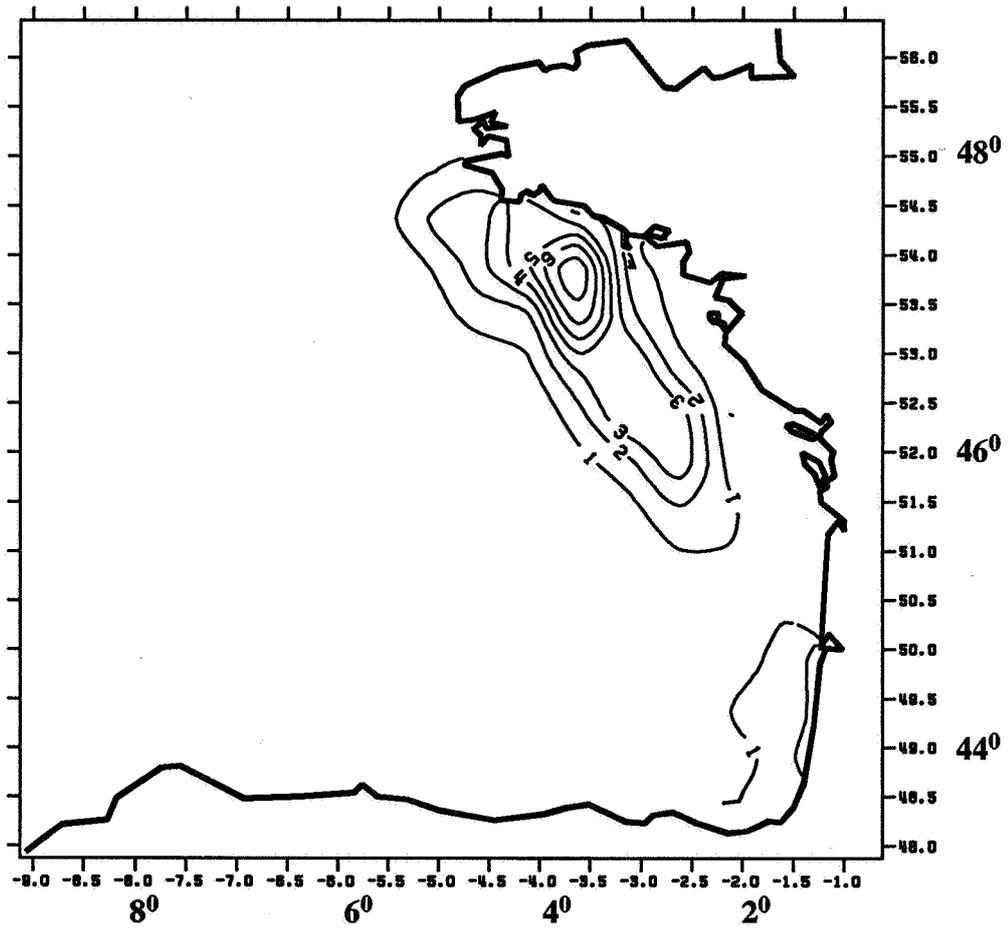


Fig. 38 : Merlu-0. Carte de la moyenne par bloc des pourcentages de biomasse

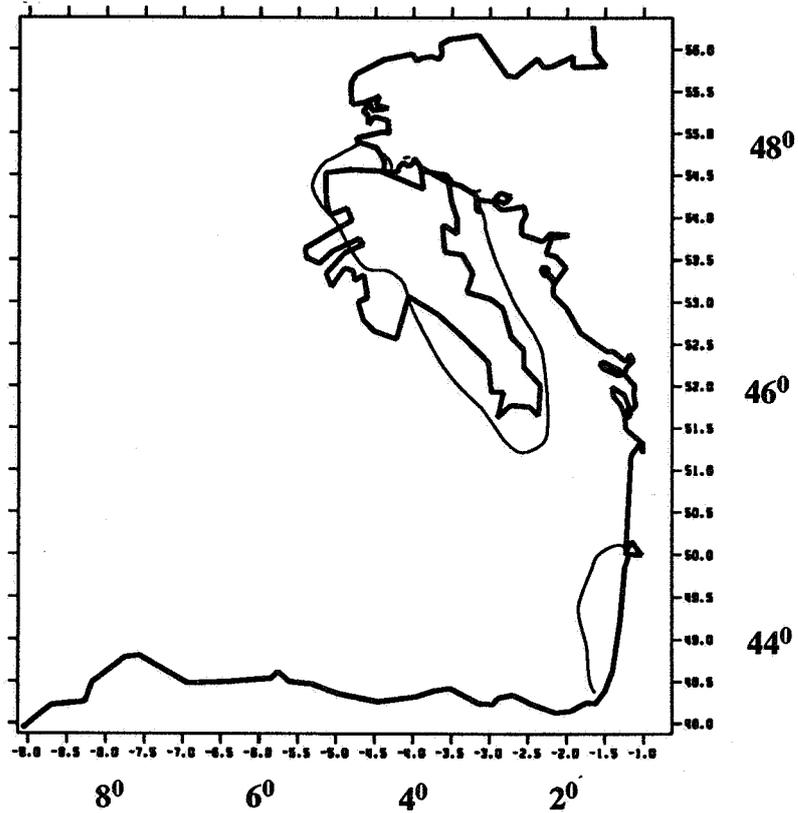


Fig. 39 : Aire occupée par 80% de la classe d'âge-0 du merlu, en moyenne chaque année.
Elle est superposée à l'emplacement de la Grande Vasière

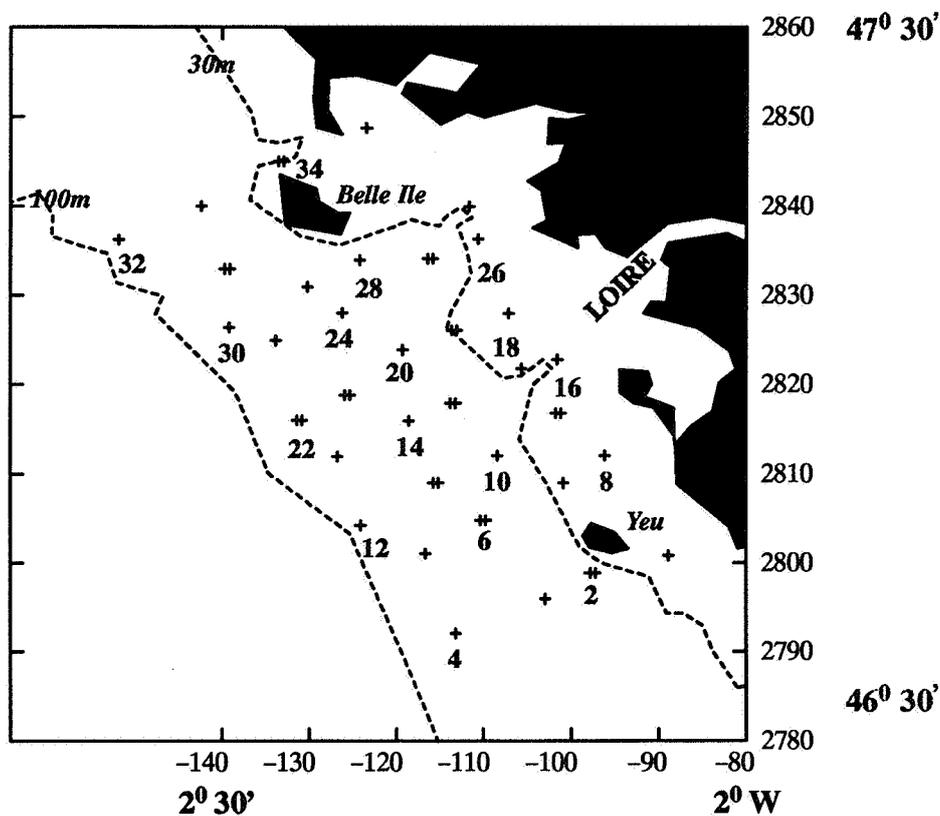


Fig. 40 : Le réseau des stations d'échantillonnage des oeufs de sole devant l'estuaire de la Loire (N/O Thalassa). Les numéros de code des stations sont spécifiés ; les latitudes Y sont données en minutes (i.e. milles nautiques) ; les longitudes X sont calculés ainsi : $X(mi.n) = X(minute) \times \cos(47^\circ)$

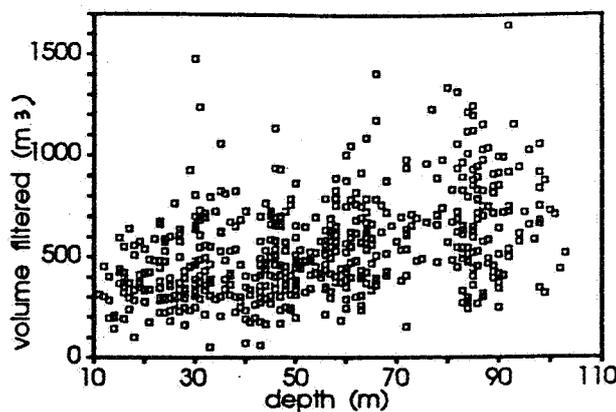
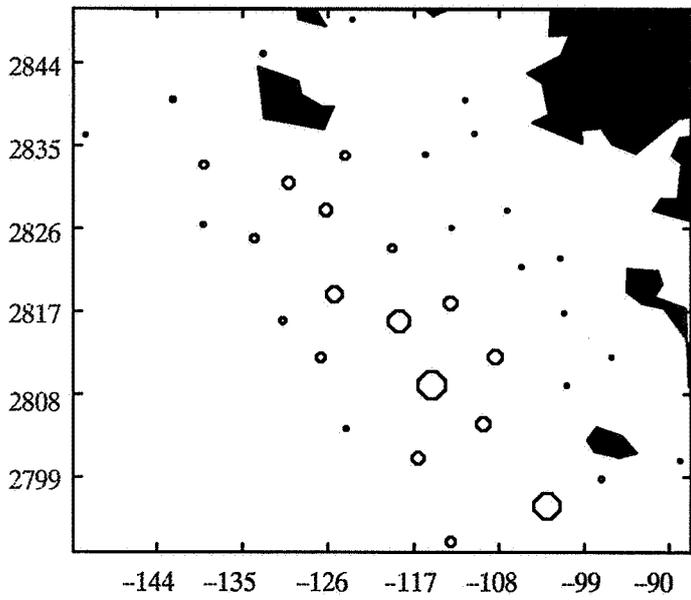


Fig. 41 : Relation entre la sonde et le volume filtré (support des échantillons), toutes campagnes confondues (d'après Koutsikopoulos, 1991)

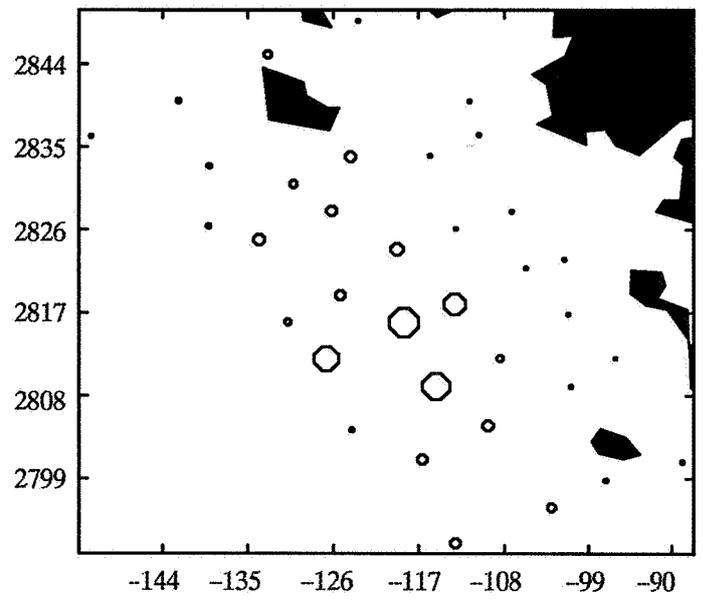
ANNEXE FIGURES

Code campagne	Année	Jour début campagne	Moyenne arithmétique des données
40 41 42 50 60	1986	02/03 06/03 10/03 18/03 07/04	2,22 2,81 2,83 1,40 0,41
71 72 73	1987	08/03 08/04 14/04	3,12 0,22 0,19
80 81	1988	04/03 30/03	2,64 0,12
90 91 92	1989	23/02 14/03 08/04	0,47 2,53 0,14

Tableau 2 : Caractéristiques des campagnes d'échantillonnage des oeufs de sole.

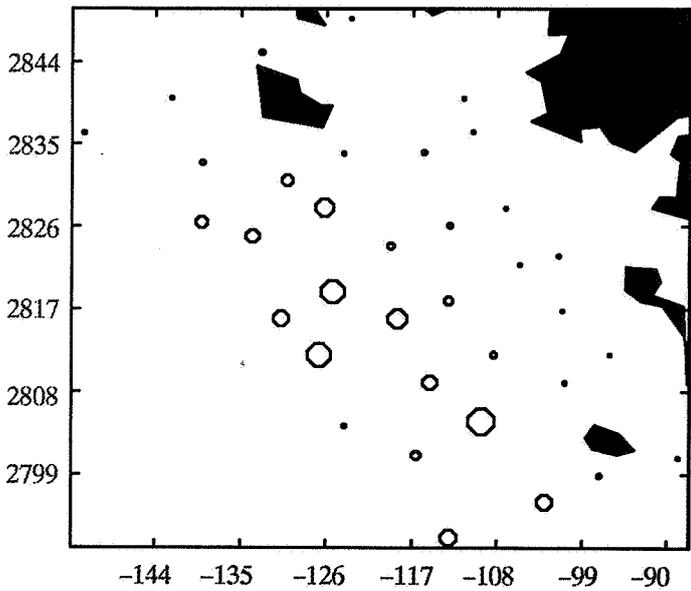


Soldet 40

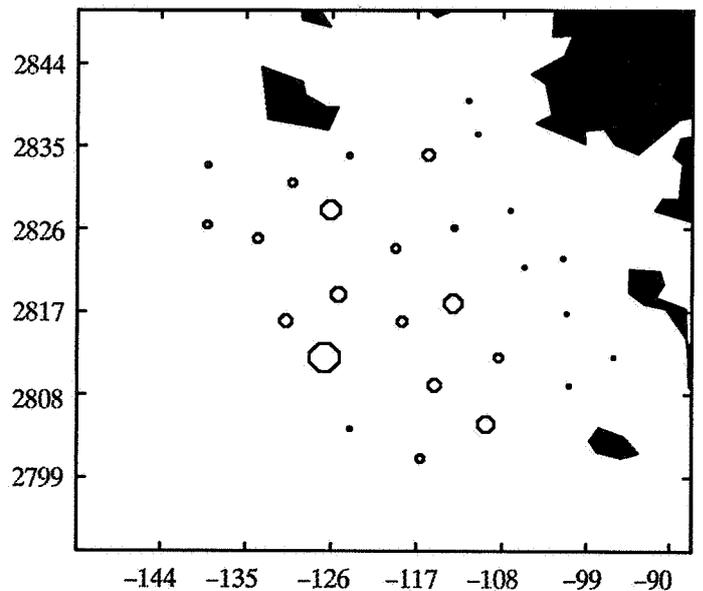


Soldet 41

1986

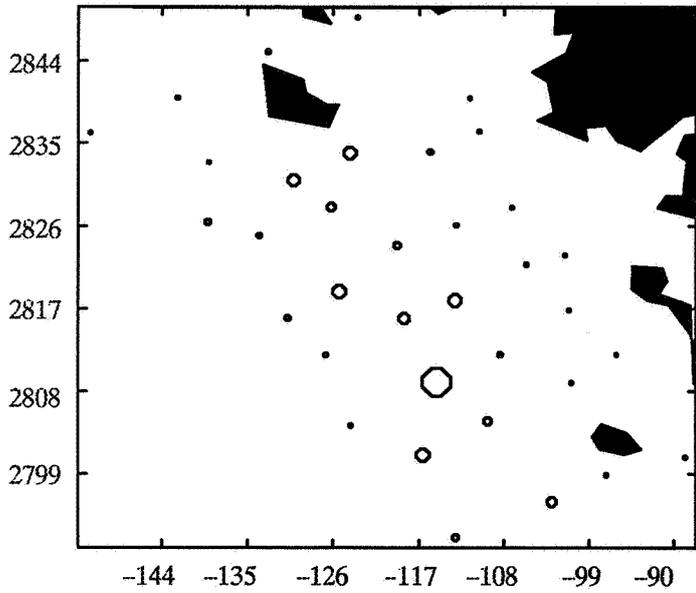


Soldet 42

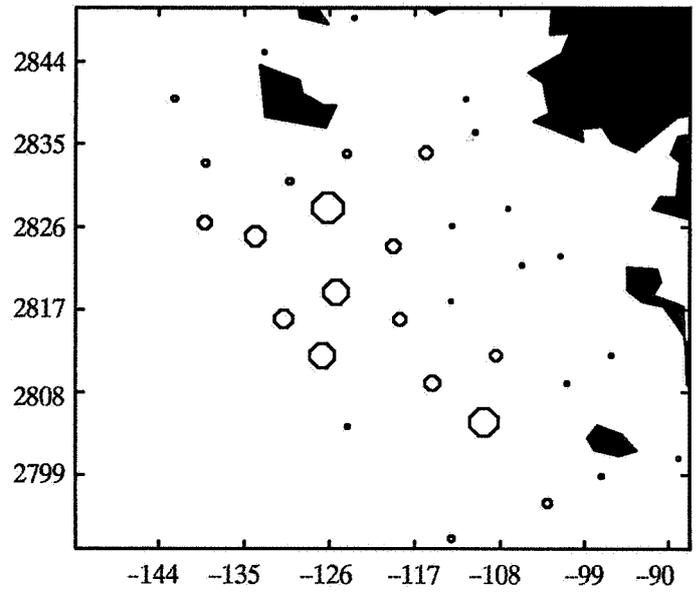


Soldet 50

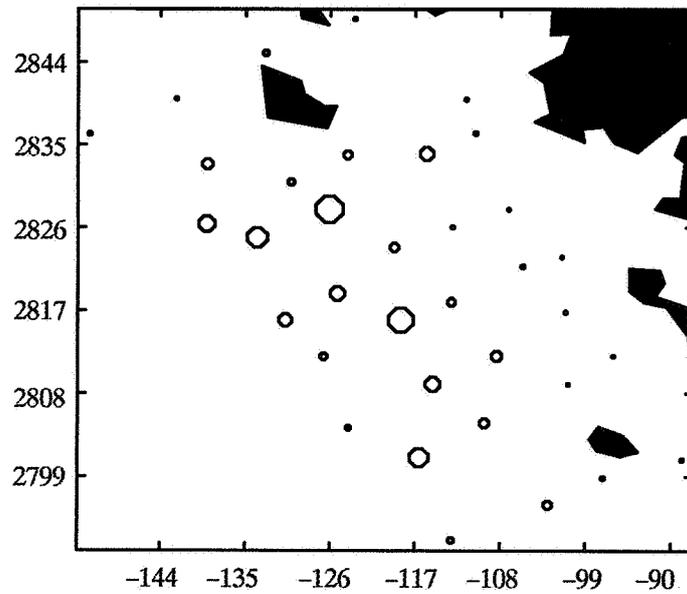
Fig. 42 : Les oeufs de sole. Représentation proportionnelle des pourcentages de biomasse, $z_p(x, t) = z(x, t)/m_q(t)$ en 1986. Le cercle de rayon maximal correspond à la valeur maximale sur chaque carte. On compare la répartition des pourcentages de biomasse et non celle des densités.



1987, Soldet 70



1988, Soldet 80



1989, Soldet 91

Fig. 43 : Les oeufs de sole. Représentation proportionnelle des pourcentages de biomasse en 1987, 88, 89.

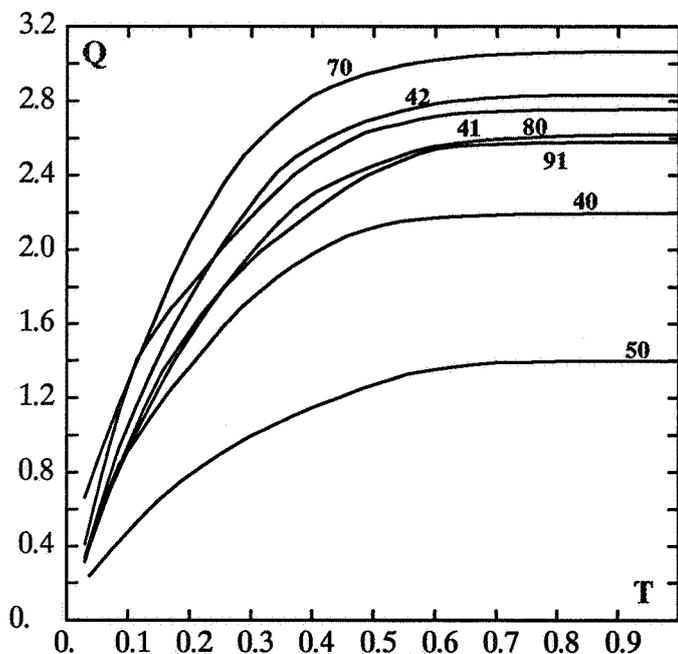


Fig. 44b

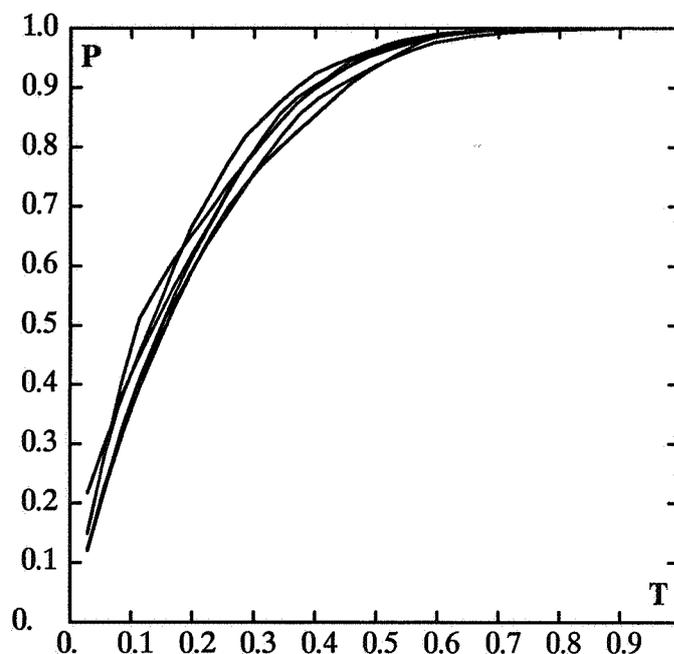


Fig. 44c

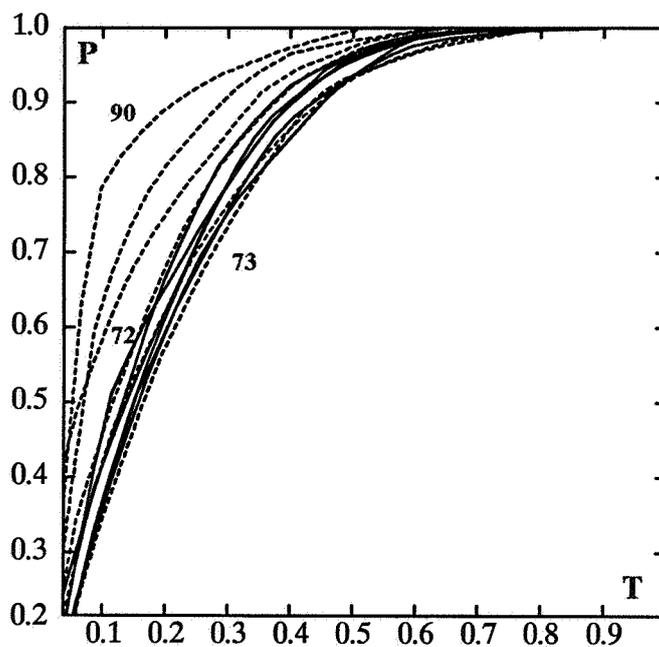
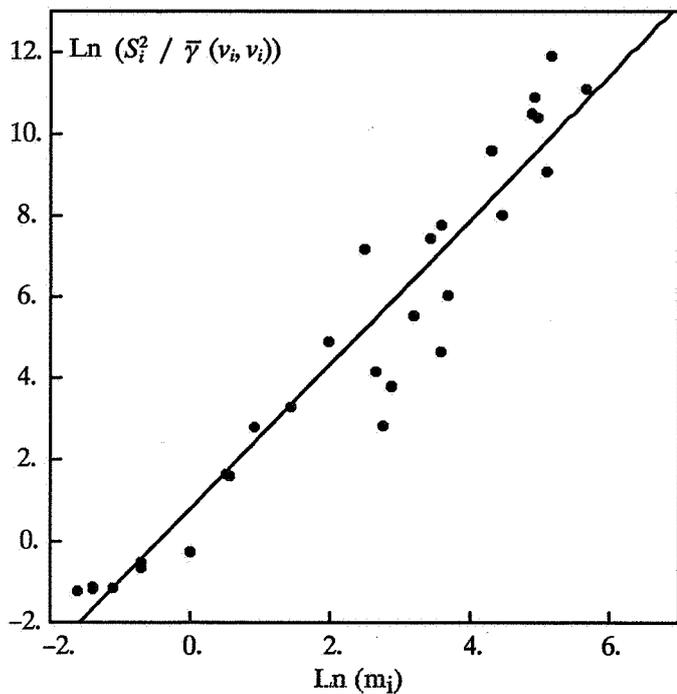
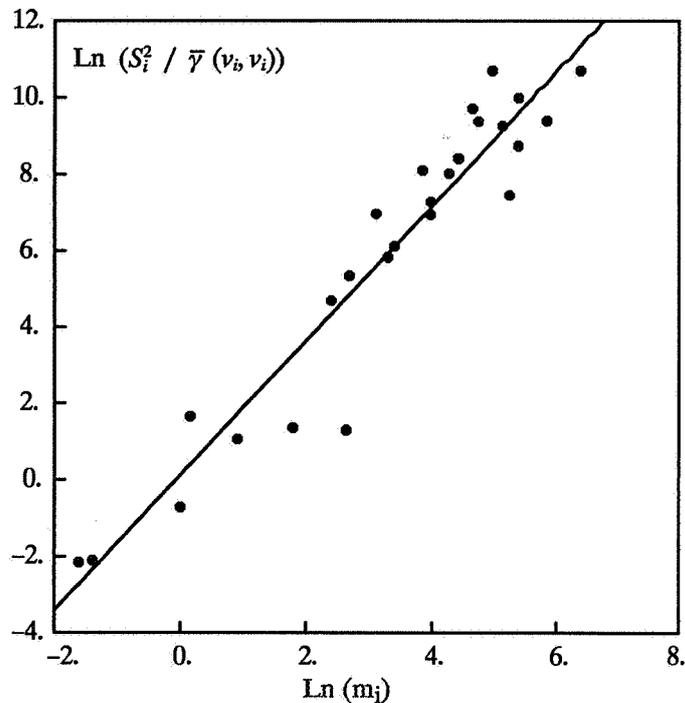


Fig. 44a

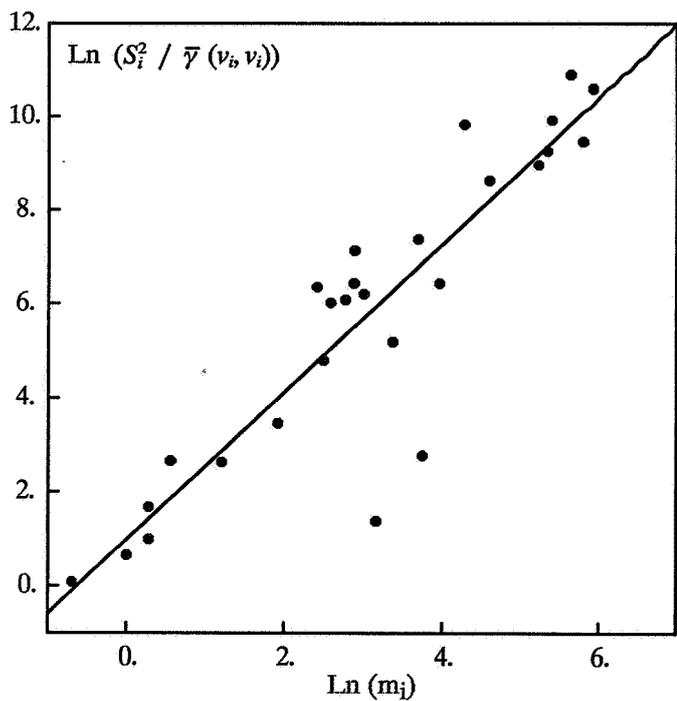
Fig. 44 : Les oeufs de sole. Courbes Q de T empiriques. Les traits pleins repèrent les campagnes de pleine ponte ; les traits pointillés celles de début et fin de ponte ; les numéros figurés désignent les codes des campagnes.



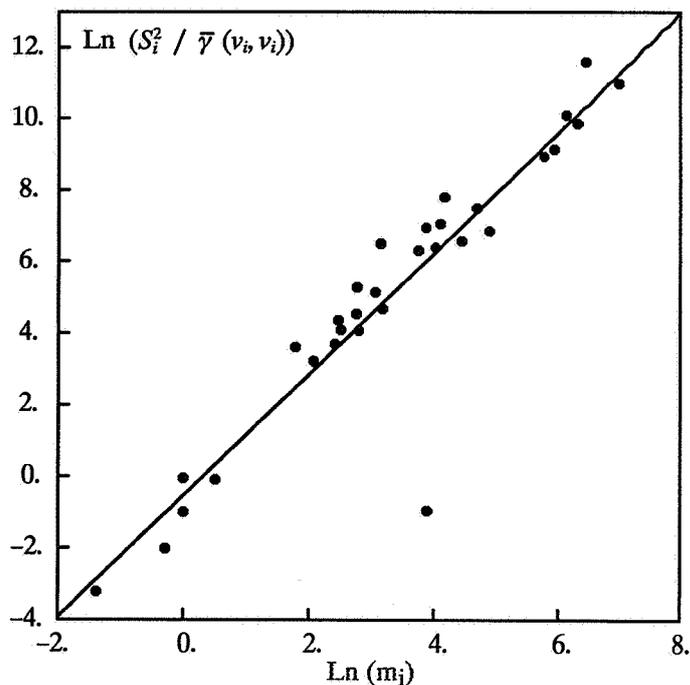
1987



1988



1989



1990

Fig. 45 : Merlu-0 . Relation entre la moyenne et un rapport de dispersion par strate, dégagée de l'effet géométrique des strates

$$m_i = \frac{1}{n_i} \sum_j z_{ij} ; S_i^2 = \frac{1}{n_i - 1} \sum_j (z_{ij} - m_i)^2 ; \gamma(v_b, v_i) = \frac{1}{v_i^2} \int_{v_i} \int_{v_i} \gamma(x-y) dx dy$$

ANNEXE FIGURES

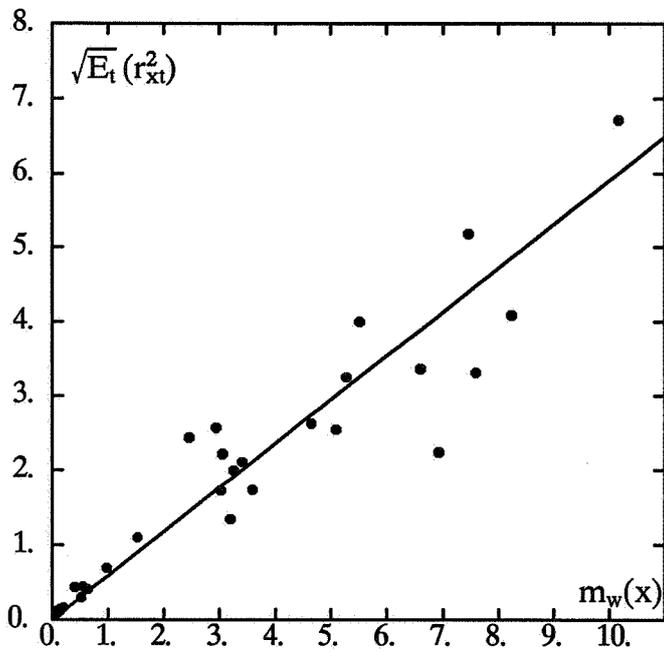


Fig. 46 : Relation entre la moyenne et l'écart-type temporel des résidus r en chaque station d'échantillonnage. Modélisation par une droite.

$$\sqrt{E_t(r_{xt}^2)} = a.m_x(x) ; a = 0.58$$

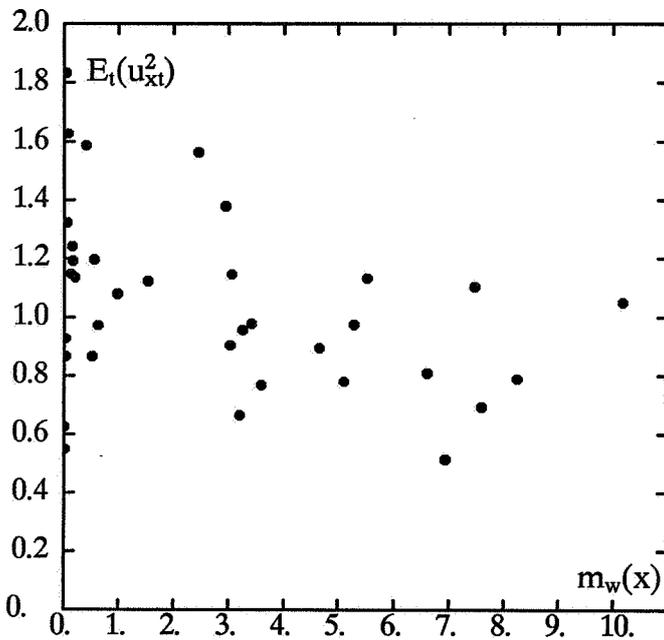


Fig. 47 : Relation entre la moyenne et la variance temporelle des résidus u, en chaque station d'échantillonnage : $u(x, t) = r(x, t)/a.m_w(x)$

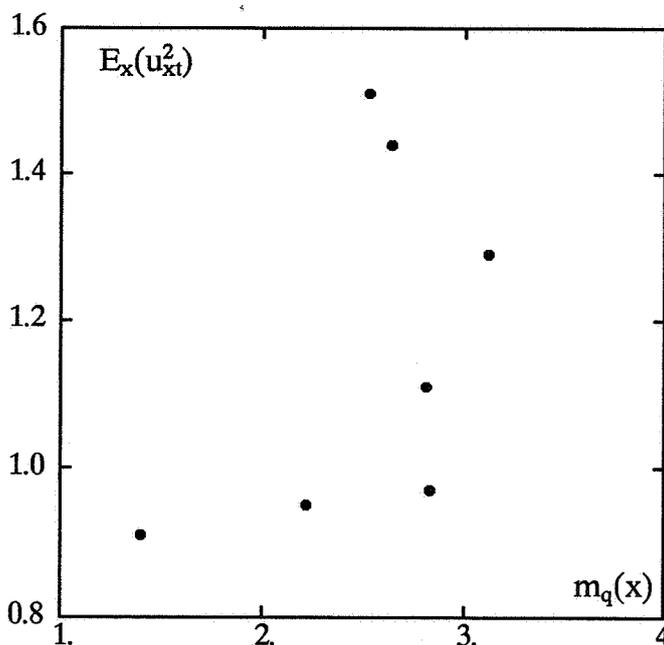


Fig. 48 : Relation entre la moyenne arithmétique des données à t ($m_q(t)$) et la variance spatiale des résidus u.

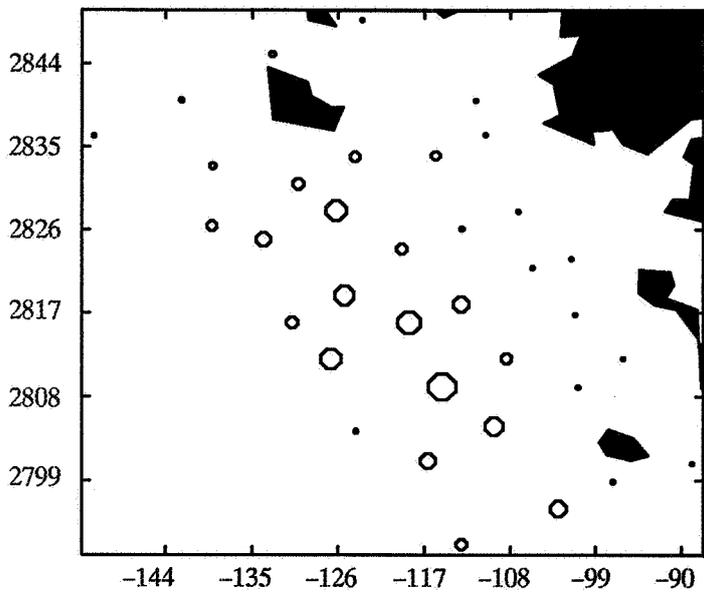


Fig. 49 : Représentation proportionnelle de la dérive estimée $m_w^*(x)$

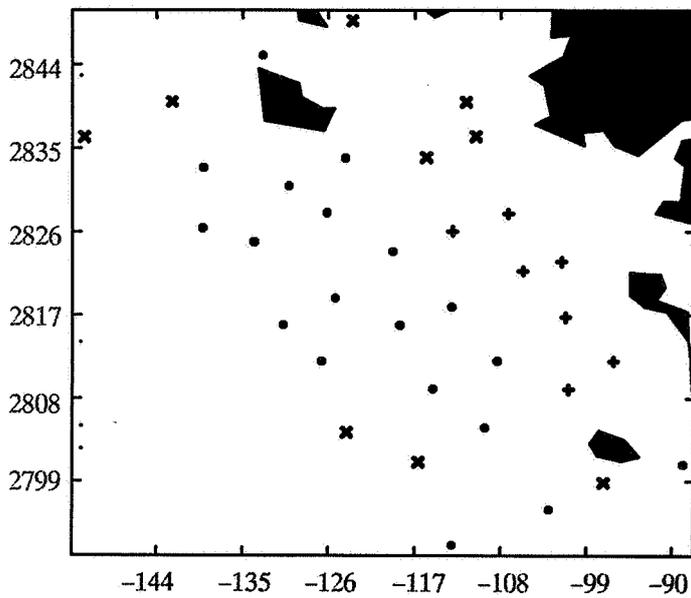


Fig. 50 : Valeurs des sédiments aux stations d'échantillonnage. Ils sont codés sur une échelle arithmétique de 0 à 9, selon la diminution progressive de la granulométrie

- + : rochers et graviers (0-1) ; vases (9) ;
- : sables grossiers à fins (2-8) représentant l'habitat favorable

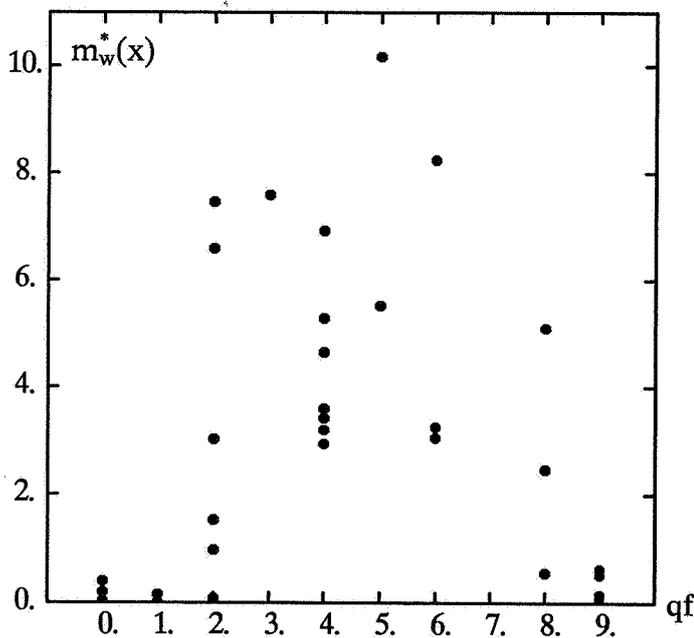
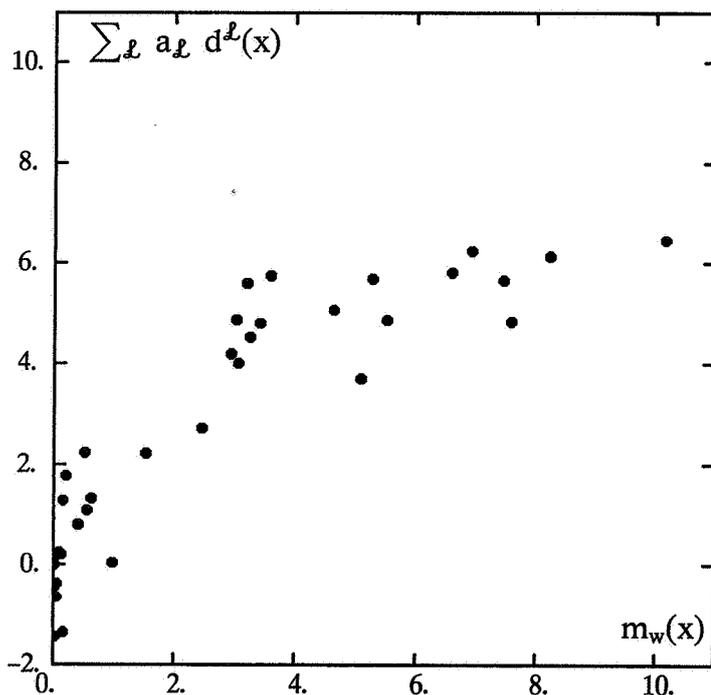


Fig. 51 : Relation entre la dérive $m_w^*(x)$ et la qualité du sédiment sur le fond, qf , codée de 0 à 9

Variabes d ^l	Coefficient a _l Régression quadratique en X et Y	Coefficient a _l Régression sur les 10 d ^l
1	- 4,56	- 3,08
X	- 36,95	- 30,78
Y	- 8,13	- 8,82
X ²	- 39,68	- 30,24
Y ²	- 21,03	- 17,94
XY	- 36,95	- 24,44
qf ²	-	- 13,39
qf	-	7,90
qfX	-	1,54
qfY	-	11,58
R ² de la régression	R ² = 0,63	R ² = 0,77

Régression multiple de la dérive sur les variables d^l par la méthode des moindres-carrés.
 Pour l'inversion matricielle, les variables explicatives sont standardisées :
 (Y-2790)/60 ; (X+81)/60 ; qf/10



Comparaison aux stations d'échantillonnage entre les valeurs observées de la dérive $m_w(x)$ et les valeurs prédites par la régression multiple à 10 variables explicatives.

Fig. 52 : Régression de la dérive $m_w(x)$ sur des variables explicatives

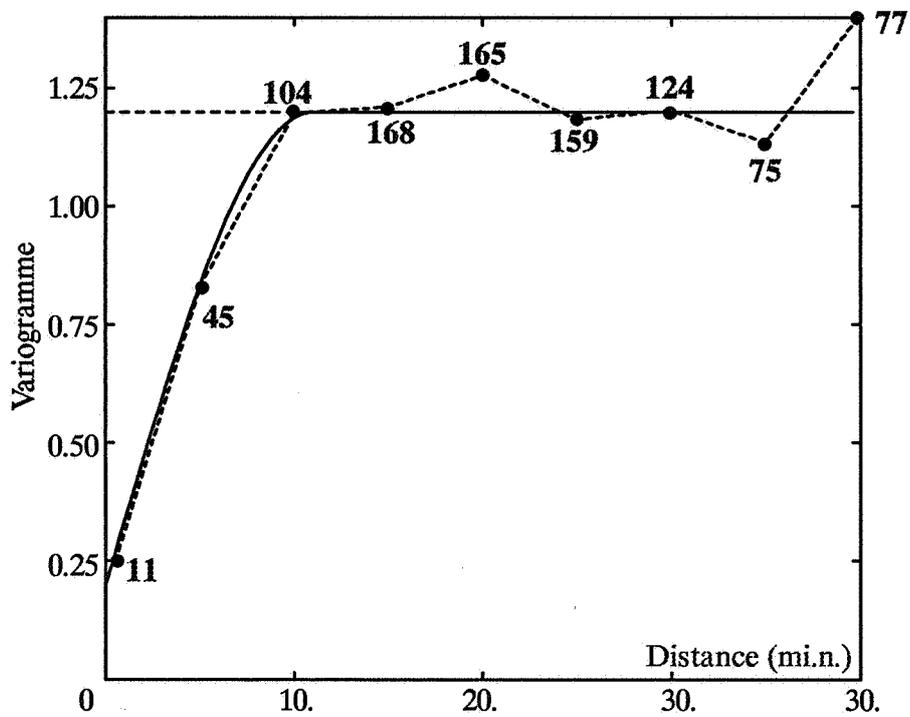


Fig. 53 : Variogramme spatial isotrope des résidus u ; c'est la moyenne temporelle des variogrammes spatiaux de chaque campagne. Le modèle est un sphérique (portée = 11 mi.n. ; palier = 1 ; effet de pépite = 0,2). Les nombres de couples sont figurés.

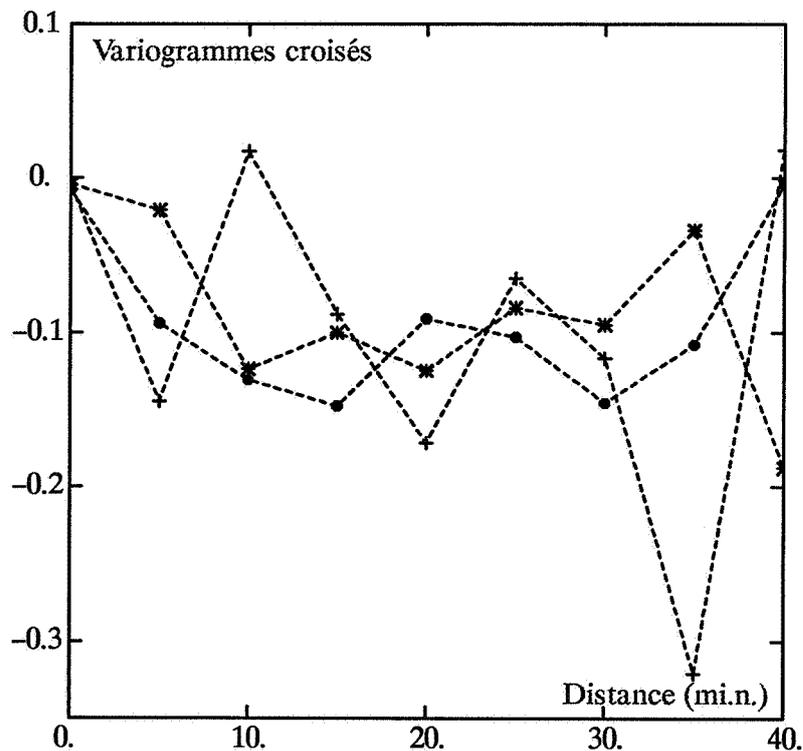
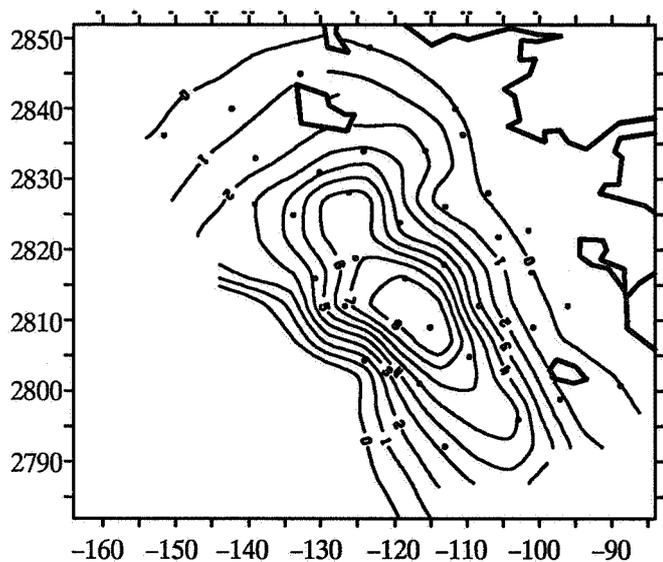
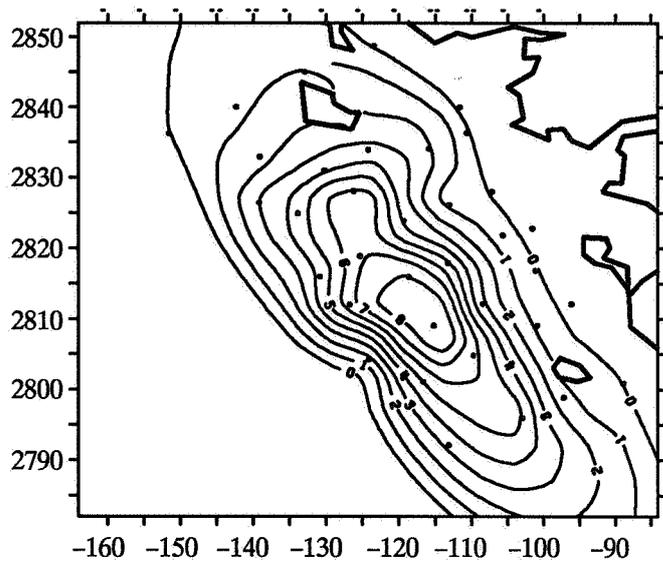


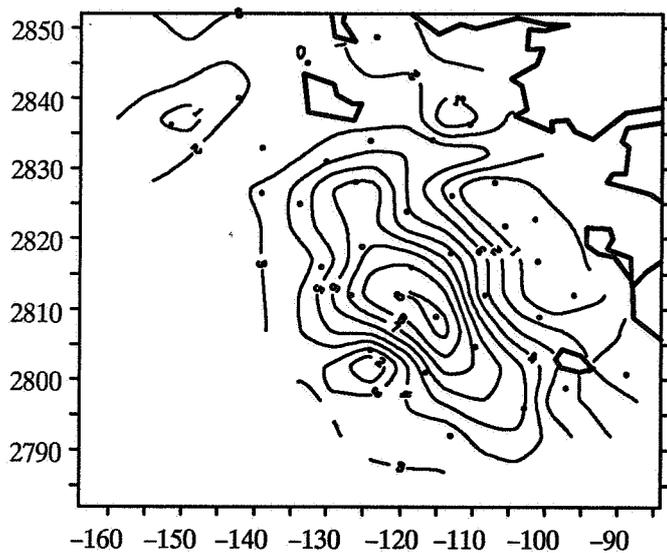
Fig. 54 : Variogrammes croisés entre les résidus u de 2 campagnes différentes ; ils sont calculés dans le temps par pas de 1 an • : 0 années ; * : 1 an ; + : 2 ans



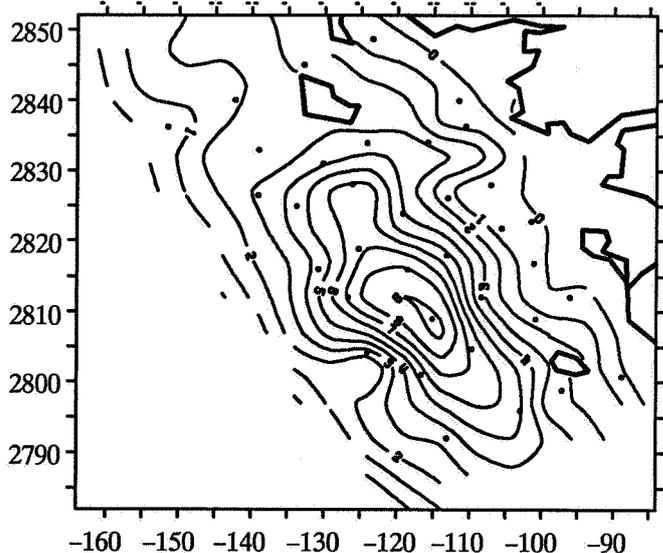
FAI-1



FAI-2

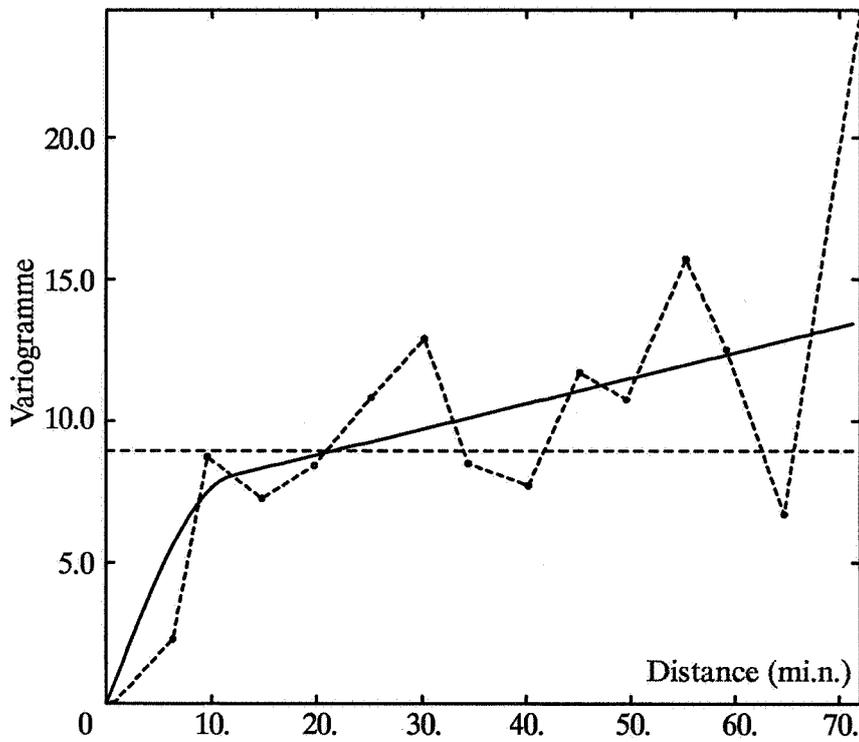


FAI-0 + dérive externe

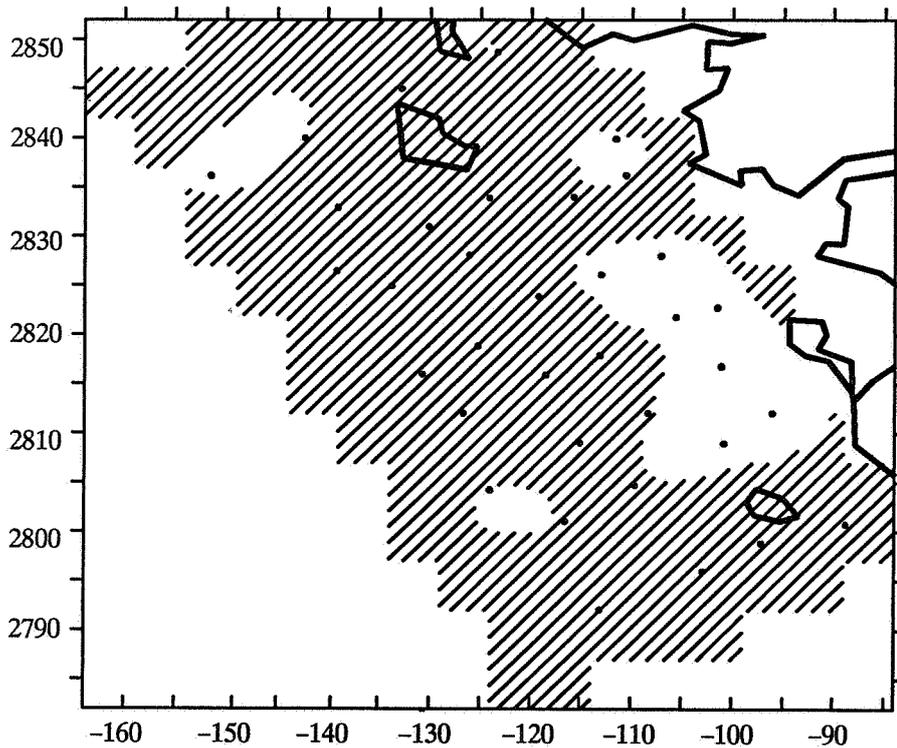


FAI-2 + dérive externe

Fig. 55 : Cartes du krigage en FAI-k de la dérive $m_w^*(x)$ pour chacun des modèles considérés



Variogramme isotrope
du code qf des sédiments



Localisation de l'habitat favorable ;
la zone hachurée représente
l'ensemble dans lequel les valeurs
krigés sont comprises entre 2 et 8

Fig. 56 : Estimation des codes du sédiment par krigeage sur la grille d'interpolation, afin de mettre en oeuvre la méthode de la dérive externe

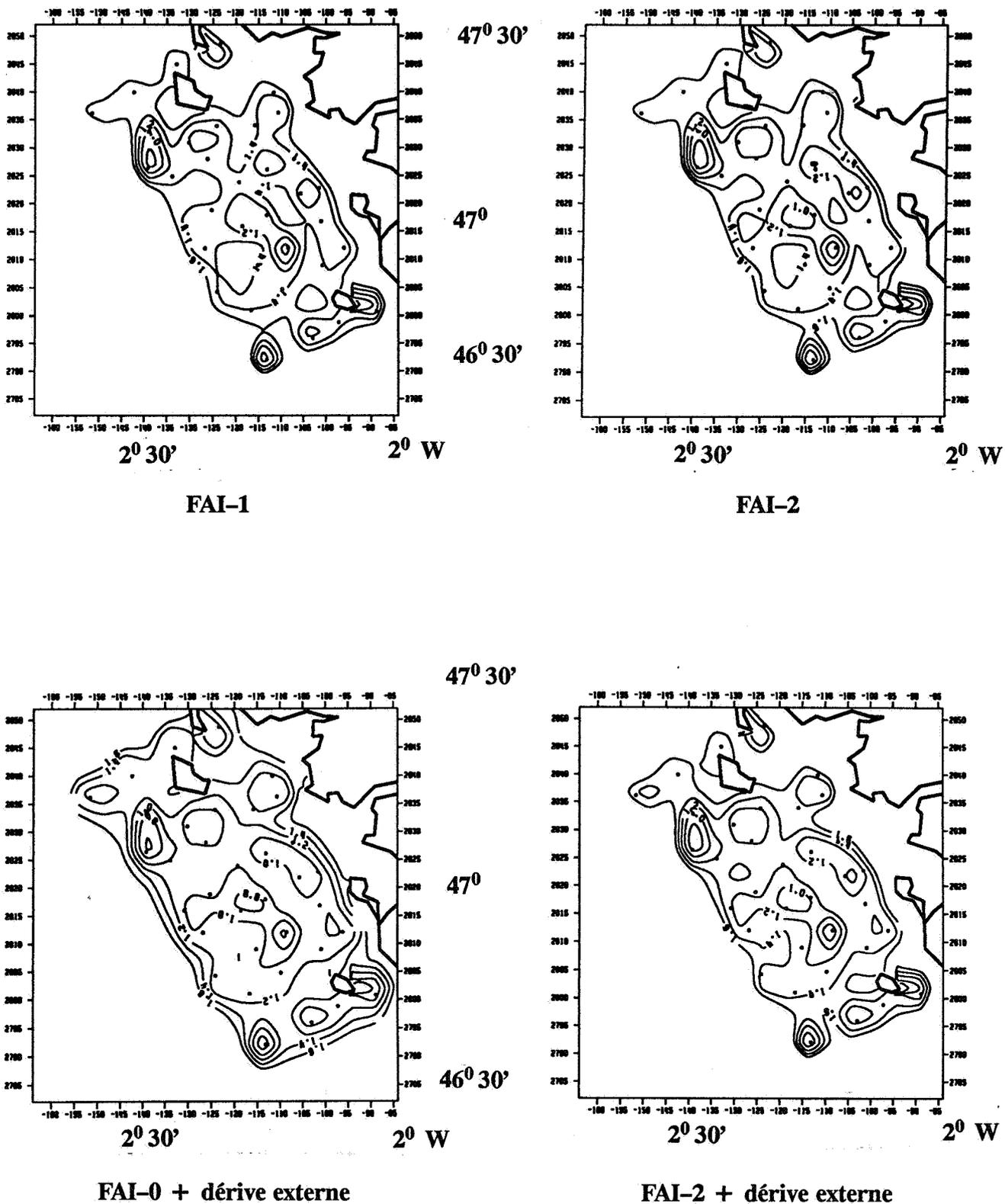
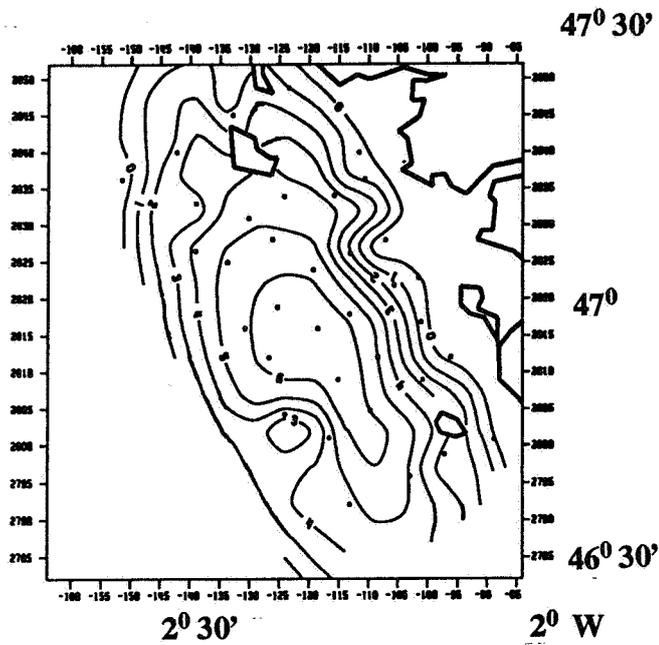
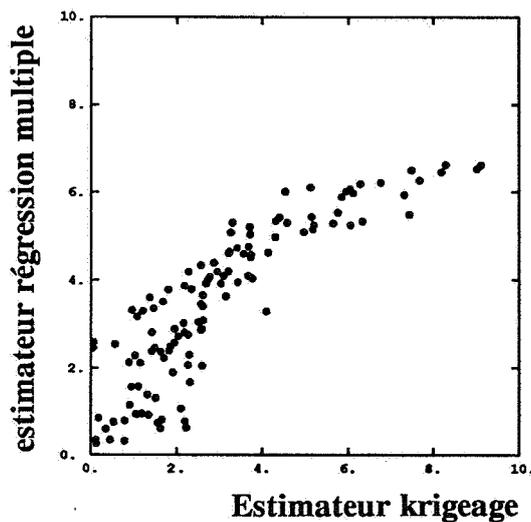


Fig. 57 : Cartes des écarts-type de krigeage en FAI-k de la dérive $m_w^*(x)$ pour chacun des modèles de FAI-k considérés



Carte de la dérive $m_w^*(x)$ obtenue par la fonction de prédiction de la régression multiple à 10 variables explicatives

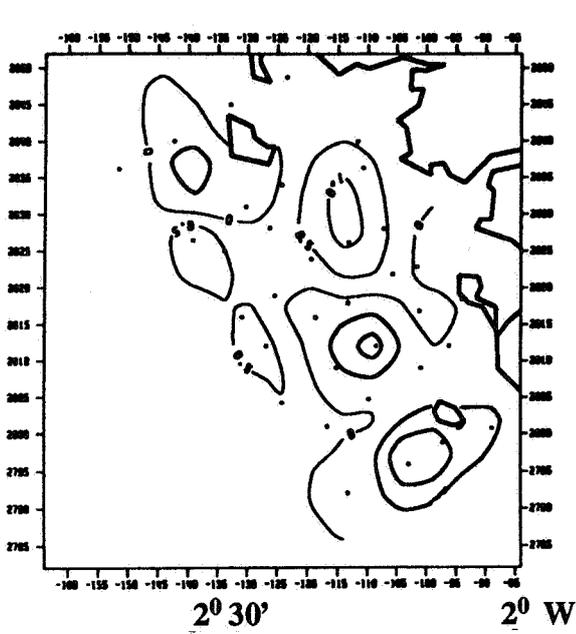
estimateur régression multiple



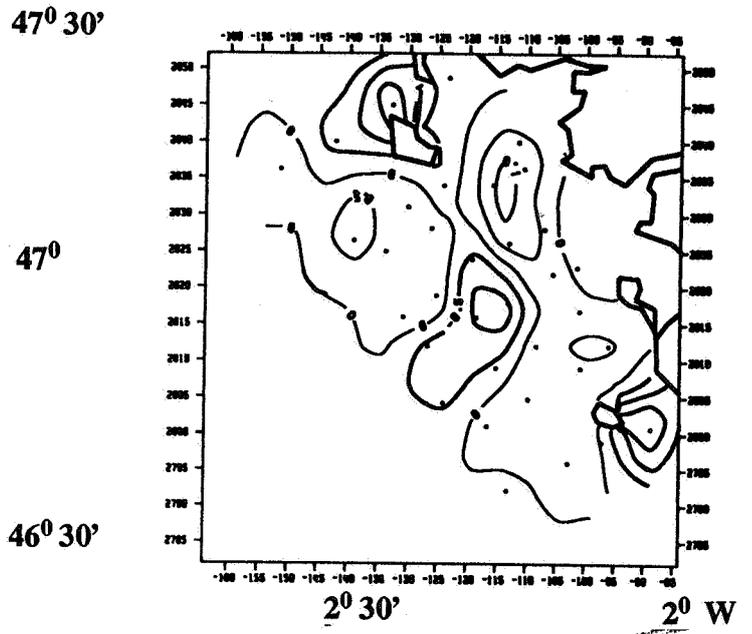
Estimateur krigeage

Comparaison aux noeuds de la grille d'interpolation, entre les valeurs estimées d'une part, par la fonction prédictrice de la régression multiple à 10 variables explicatives et d'autre part, par le krigeage en FAI-k avec les mêmes variables en conditions de non-biais (FAI-2 et dérive externe)

Fig. : 58

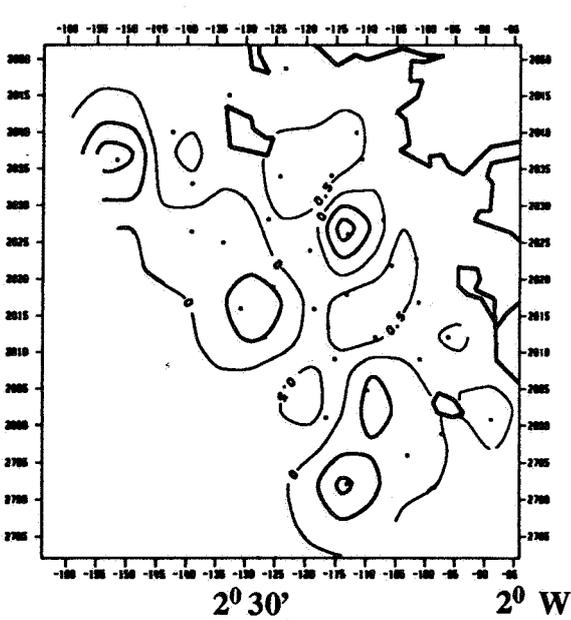


Soldet 40

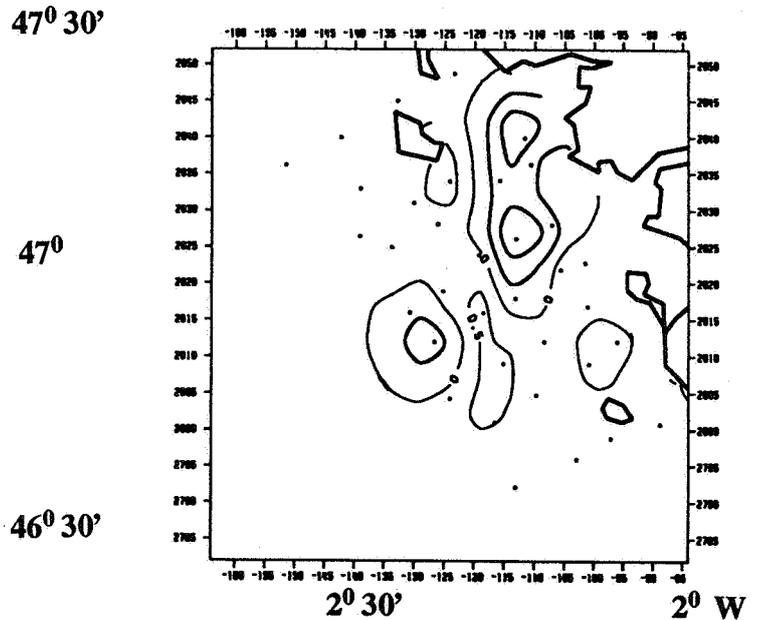


Soldet 41

1986

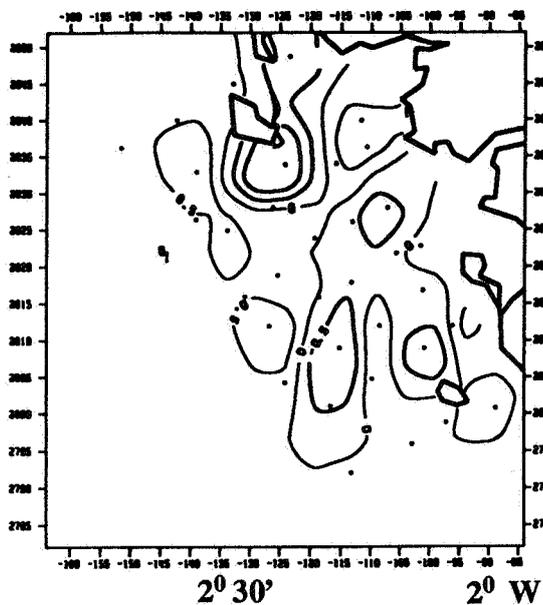


Soldet 42



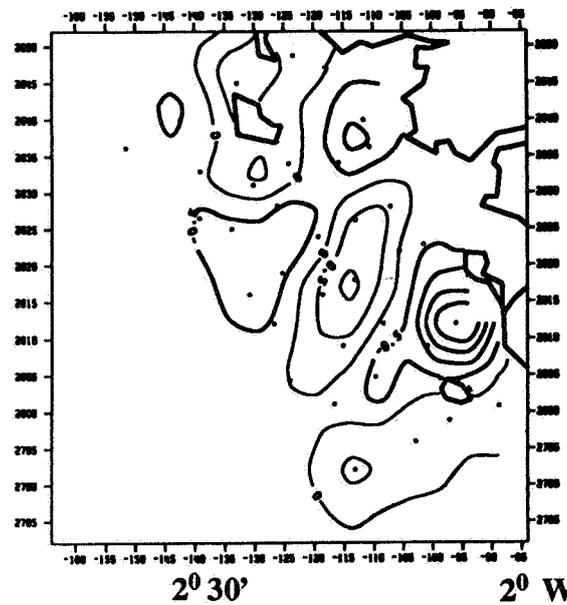
Soldet 50

Fig. 59 : Cartes du krigeage des résidus u, pour chaque campagne en 1986.
Les isolines grasses marquent les valeurs négatives

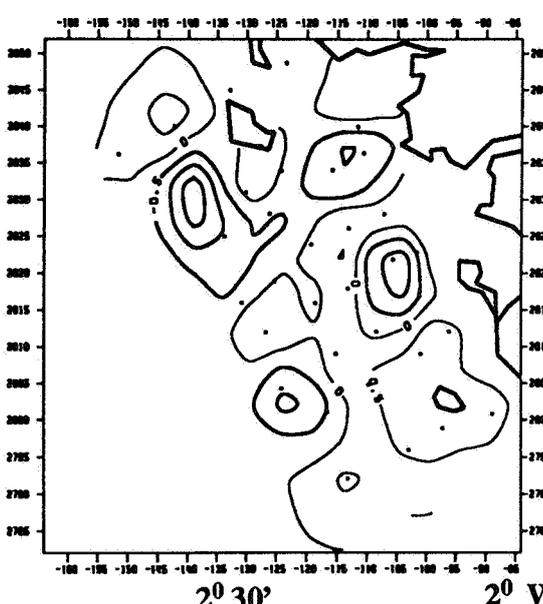


1987, Soldet 70

47° 30'
47°
46° 30'



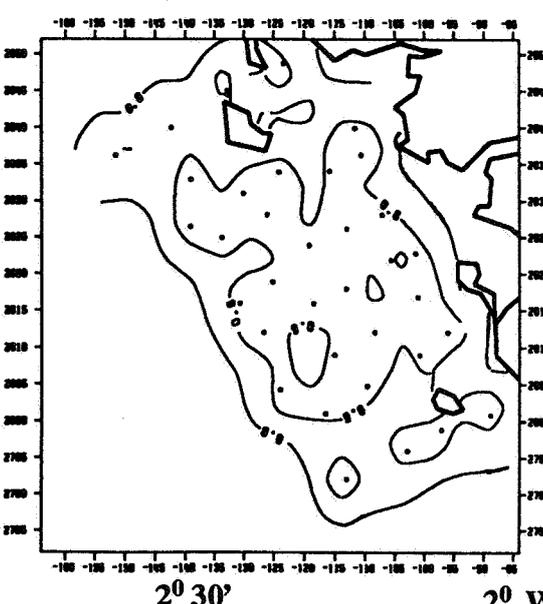
1988, Soldet 80



1989, Soldet 91

47° 30'
47°
46° 30'

Fig. 60 : Cartes du krigeage des résidus u, pour chaque campagne en 1987, 88 et 89. Les isolines grasses marquent les valeurs négatives.



2° 30' 2° W

47° 30'
47°
46° 30'

Fig. 61 : Carte de l'écart-type de krigeage des résidus u.

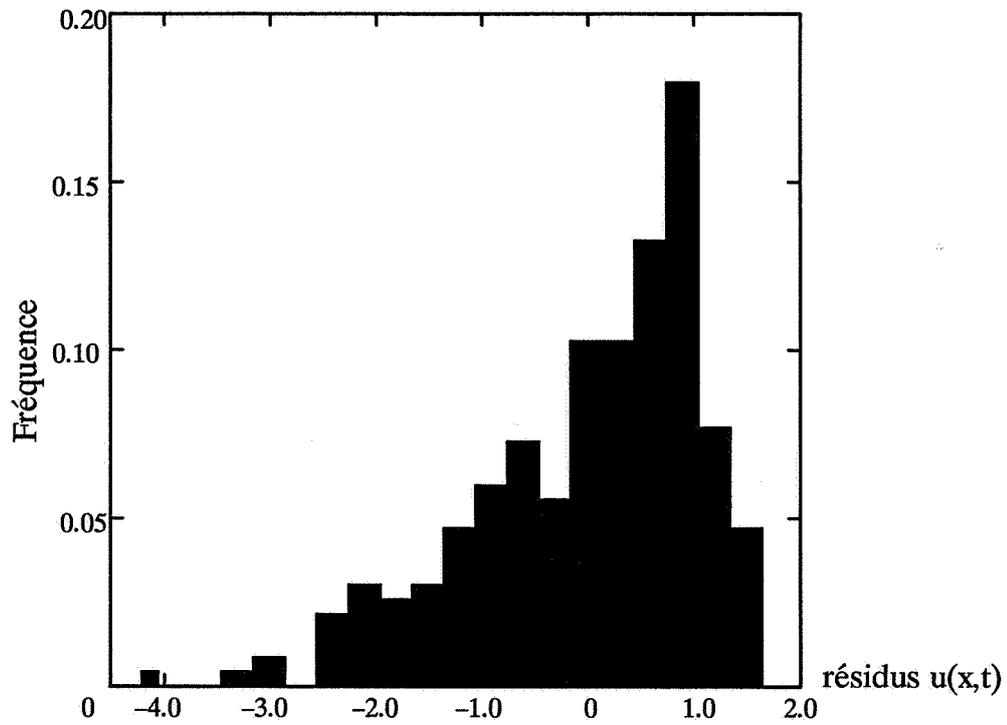


Fig. 62 : Histogramme des résidus u , toutes campagnes confondues. Les résidus négatifs marquent des pics de ponte par rapport à la valeur moyenne de la dérive $m_w^*(x)$

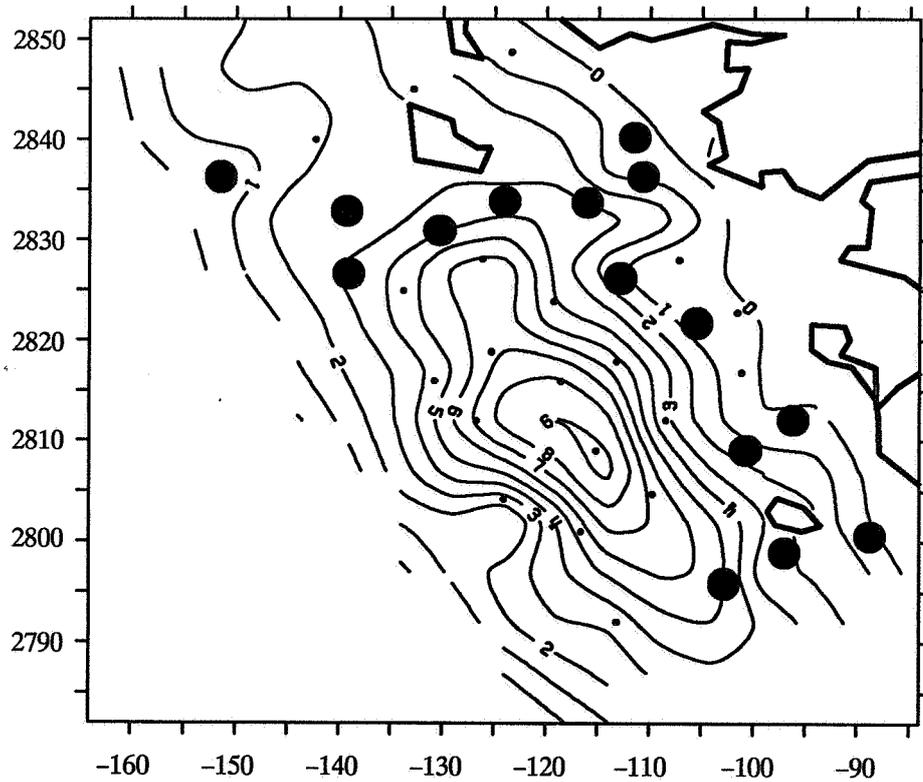


Fig. 63 : Localisation par des disques noirs sur la carte moyenne invariante dans le temps, des stations où les résidus u , toutes campagnes confondues, ont marqué au moins une fois un pic de ponte important $u \leq -1,9$