



HAL
open science

Statistiques individuelles et géostatistique transitive en écologie halieutique

Bez Nicolas

► **To cite this version:**

Bez Nicolas. Statistiques individuelles et géostatistique transitive en écologie halieutique. Statistiques [stat]. Ecole des Mines de Paris, 1997. Français. NNT : 1997ENMP0736 . tel-04649510

HAL Id: tel-04649510

<https://hal.science/tel-04649510v1>

Submitted on 16 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

ÉCOLE NATIONALE SUPÉRIEURE DES MINES DE PARIS

STATISTIQUES INDIVIDUELLES
ET
GÉOSTATISTIQUE TRANSITIVE
EN
ÉCOLOGIE HALIEUTIQUE

THÈSE

présentée par

Nicolas BEZ

en vue d'obtenir le titre de Docteur en Géostatistique

Soutenue le 10 Mars 1997,
devant le jury composé de :

MM.	Schmitt	Michel	Président
	Laloë	Francis	Rapporteur externe
	Gascuel	Didier	Rapporteur externe
	Rivoirard	Jacques	Examineur
	Reid	David	Examineur
	Chauvet	Pierre	Examineur

THE UNIVERSITY OF CHICAGO
PHYSICS DEPARTMENT
530 SOUTH EAST ASIAN AVENUE
CHICAGO, ILLINOIS 60607

1984

1985

1986

PHYSICS DEPARTMENT

PHYSICS DEPARTMENT

PHYSICS DEPARTMENT

Pour une thèse en partance,
flot de papiers;
minuscule.

POUR ANTONIN, FABIEN, ET
FLO DE MA VIE;
MAJUSCULE.

Remerciements

Pour classique et incontournable qu'il soit, l'exercice des remerciements est un moment qui m'est agréable. Je m'y attelle très volontier, ici, et dans la préface. Cet également un exercice important puisque je crois savoir que les pages de remerciements comptent beaucoup de lecteurs, étant parmi les premières à être lues, après le résumé.

Jacques RIVOIRARD est pour beaucoup dans l'aboutissement de mon travail. Il a su résister au flot de mes questions, rester disponible et orienter ma réflexion. Sa pédagogie et son expérience géostatistique m'ont été, pour le moins profitable, et me servent d'exemple. Sans ambages, mais avec beaucoup de sincérité: merci!

Je tiens à souligner que j'ai beaucoup apprécié les conditions humaines et matérielles de travail que le Centre de Géostatistique m'a offertes. Je fait en particulier référence aux conditions d'accueil (bureau, ordinateur), aux missions qu'il m'a été donné de faire, aux congrés que j'ai eu le loisir de suivre et aux publications que j'ai pu faire.

Je crois devoir remercier la providence d'avoir placé Christian LAJAUNIE et Jacques RIVOIRARD dans le même bureau. Par l'intermédiaire de ce dernier, j'ai pu profiter, une fois digérées, des idées pertinentes de Christian.

Je n'ai pas non plus laissé beaucoup le choix à Christian LANTUÉJOUL, Pierre CHAUVET, Chantal de FOUQUET, François GÉOFFROY, et Didier RENARD de répondre ou non à mes questions. Merci pour leur disponibilité.

Toutes mes excuses à Isabelle pour n'avoir pas su faire autre chose que des missions aux parcours biscornus. Mais, n'est-ce pas dans la difficulté qu'on reconnaît la vraie compétence?

Merci à Jacky d'avoir scanné l'ensemble des photos que je lui ai transmises. Qu'il se rassure, il y en a au moins une qui a été utilisée. Tout n'est pas perdu...

J'ai été par ailleurs très sensible au travail réalisé par les deux rapporteurs Didier GASCUEL et Francis LALOE. Je ne loue pas seulement la précision de leur lecture qui est d'ailleurs, je crois, assez largement reconnue, mais aussi l'ensemble des remarques constructives qu'ils ont formulées.

Merci à "Suzanne" d'avoir affrété des navettes suffisamment délabrées pour que le travail y soit impossible. Ce sas quotidien entre Fontainebleau et chez moi, m'aura permis de papoter et bavarder avec d'autres "navettomobilistes", mais surtout d'arriver chez moi l'esprit libre de maquereaux et de variogrammes.

Au commencement de mon travail, et peut-être aurait-il été bon de commencer par là, il y a un choix professionnel réalisé à deux. Merci à Florence d'avoir assumé ce choix avec moi.

Préface

Mon aventure halieutique a commencé en mer en tant qu'observateur pour le compte de l'IFREMER (Institut Français de Recherche pour l'Exploitation de la Mer) sur un chalutier de Douarnenez armé pour la pêche au germon (thon blanc). Bien que cette activité soit à bien des égards très intéressante, j'ai rapidement compris que je ne serai jamais marin pêcheur; trop dur! Je crois sincèrement que c'est un métier qu'on ne peut admirer qu'en ne le pratiquant pas.

Halieute, alors? Grâce à mes proches, et plus particulièrement à ma proche (bien lui en a pris), une fois la tempête passée, j'ai remis le cap, mais à terre cette fois, pour l'ENSAR (Ecole Nationale Supérieure d'Agronomie de Rennes) pour une année d'enseignement en halieutique. La difficulté de ce mois passé en mer m'est restée. Derrière les équations et les chiffres que je manie tous les jours, vous pouvez être sûrs que je n'oublie que rarement les poissons et jamais les quelques marins que j'ai rencontrés. La vie qu'ils mènent force mon admiration et me tient à terre dans mon statut plus confortable d'halieute.

Jean-Claude Le Guen, qui, en 1990, dirigeait le département halieutique de l'ENSAR, répétait à qui voulait bien l'entendre (mais peut-on ne pas entendre "Le Guen"?) qu'il y avait certainement quelque chose d'intéressant à faire en géostatistique. Intrigué mais ignorant encore des choses de la géostatistique, j'ai attendu de suivre une demie journée de cours délivrée par Francis Gohin pour découvrir ce qui se cachait derrière ce mot. Un peu moins ignorant mais toujours intrigué, j'ai fait alors route vers Fontainebleau. On s'éloigne de la mer me direz-vous! C'est dans l'espoir d'y revenir avec du matériel mieux affûté vous répondrais-je. Je tiens en tous les cas à remercier les halieutes de bon augure que j'ai rencontrés sur ma route et qui m'ont aidé à choisir ce cap: Jean-Claude Le Guen, Francis Gohin, Didier Gascuel et Pierre Petitgas.

Depuis quelques années, le port de Fontainebleau, c'est-à-dire les applications halieutiques de la géostatistique, ne cesse de se développer. Après les thèses de Francis Gohin (1987) et de Pierre Petitgas (1991), le Centre de Géostatistique de l'Ecole des Mines, participe aujourd'hui à deux contrats européens. L'un sur l'évaluation de stocks et l'autre sur l'étude conjointe de la variabilité de certaines espèces et de l'environnement marin (projet SEFOS). Il m'est d'ailleurs agréable de remercier chaleureusement le "patron de pêche", Margaret Armstrong, et son "second", Jacques Rivoirard, de m'avoir proposé de venir travailler avec eux à bord dans le cadre du projet SEFOS.

Si le terme de géostatistique est obscur, celui de SEFOS l'est encore plus. SEFOS sont les initiales de Shelf Edge Fisheries and Oceanographic Study. Ce qui, en "français" pourrait se tra-

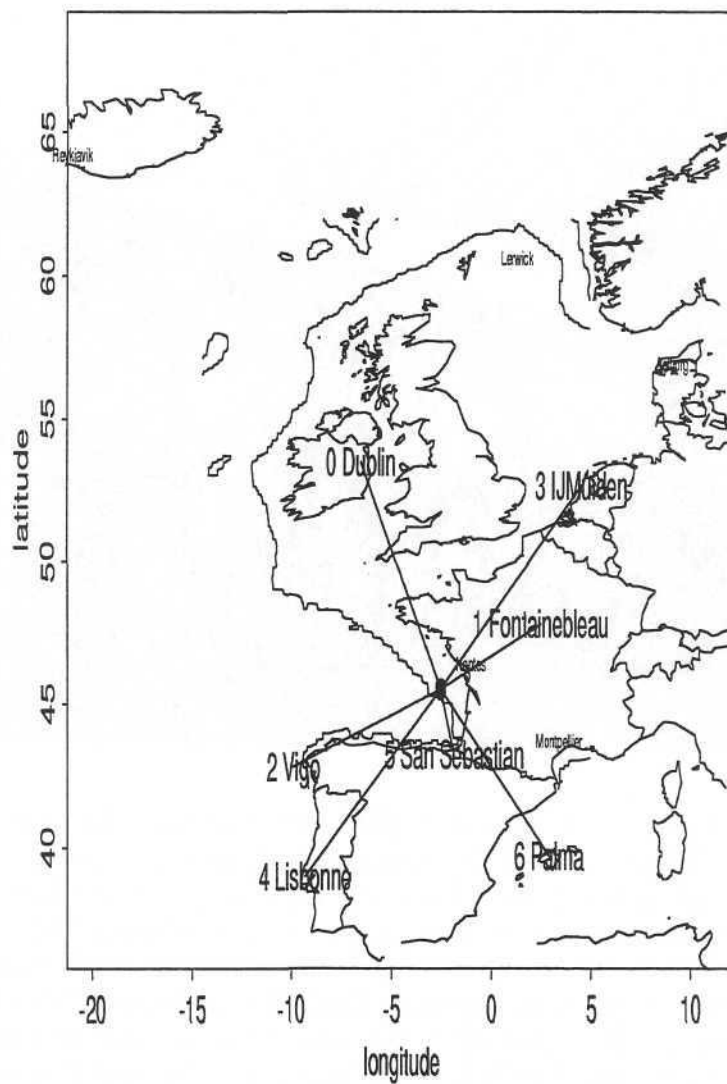
duire par : étude des pêcheries et de l'océanographie de la région des accores (bords du plateau continental). Ce projet européen, auquel participent huit pays, concerne 5 espèces commercialement importantes pour les pêcheries européennes : le maquereau, le merlu, le hareng, le merlan et le chinchard. Son objectif général est de contribuer à l'**analyse de l'influence de paramètres hydrographiques sur la distribution spatiale des espèces retenues**. L'objectif est de taille d'autant que l'ensemble des échelles possibles sont concernées dans ce projet; depuis le cycle biologique jusqu'à la distribution locale d'une phase particulière du cycle. On peut donc se demander ce qu'allait "faire le Centre de Géostatistique dans cette galère".

Mais, ne nous trompons pas. Il est des galères agréables. Celle de l'écologie halieutique, car c'est bien de cette galère dont il est question, l'est, sans contestation possible. Un des premiers objectifs du projet SEFOS était de réunir les données existantes pour décrire le plus complètement possible l'évolution des distributions de juvéniles et d'adultes de chaque "espèce SEFOS" à l'aide de cartes. Cet objectif aura été presque totalement atteint puisque l'ensemble des instituts de recherche des pays concernés ont mis en commun leurs données. Parallèlement à ce travail de compilation, des chercheurs allemands et norvégiens ont mis au point des modèles de courant pour l'ensemble de la "zone SEFOS". En marge de tout cela, la vocation du Centre de Géostatistique était de **développer des outils adaptés à la description conjointe de variables biologiques et hydrographiques**.

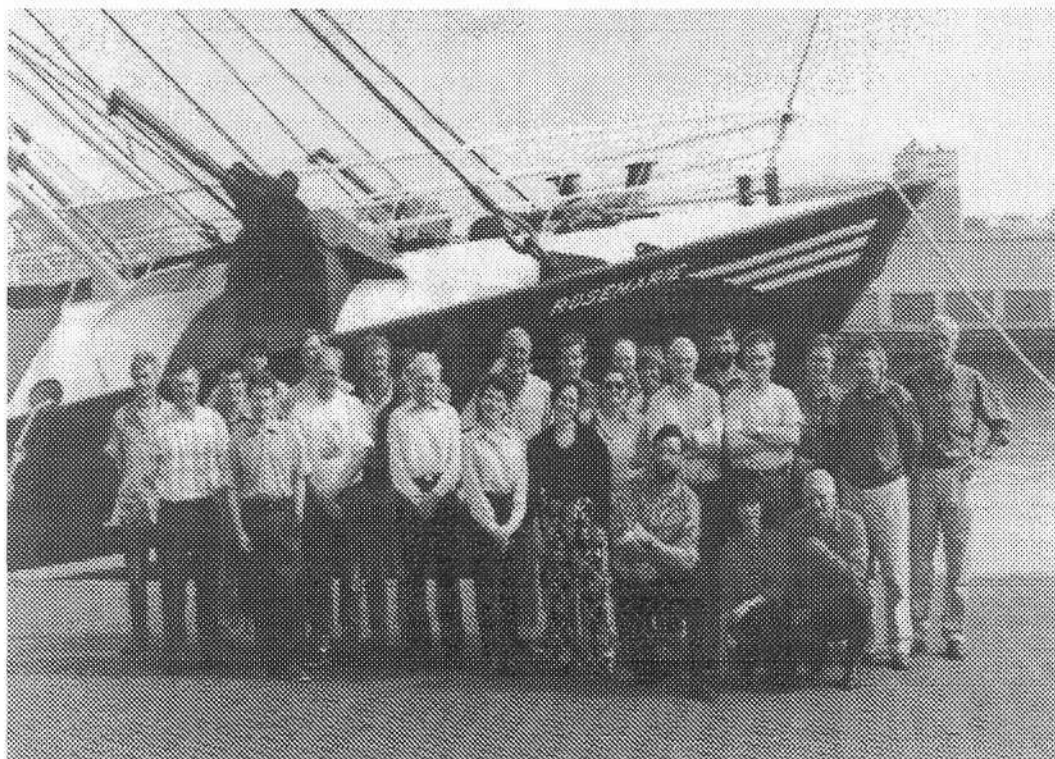
Les développements d'abord envisagés par le centre concernaient l'analyse de variables vectorielles comme les courants marins. Mais, rapidement, les questions formulées par les partenaires du projet, et par Martin Walsh du Marine Laboratory d'Aberdeen en particulier, nous ont orientés vers le traitement de densités de poissons. Rien de bien nouveau me direz-vous!

C'était sans compter que les outils classiques de la géostatistique comme le variogramme (maintenant bien connu dans le milieu halieutique) n'apportaient bien souvent que des résultats décevants voire inexploitable; je crois d'ailleurs que plusieurs halieutes ont déjà été confrontés à ce type de déception. Nos efforts ont donc porté sur le développement d'une méthodologie géostatistique adaptée à l'analyse de données halieutiques spatialisées. Ce que nous proposons dans ce travail consiste en fait à remettre au goût du jour un pan entier de la géostatistique, tombé aux oubliettes jusqu'à assez récemment puisque que Pierre Petitgas a commencé à en (re)parler dans sa thèse (1991). Utilisée jusqu'à présent dans le cadre très particulier du calcul de variance d'estimation pour un échantillonnage à maille régulière, la **géostatistique transitive** recèle en réalité les bases d'une approche méthodologique appropriée aux densités de poissons spatialisées. De plus, pour répondre aux attentes du projet SEFOS, les développements que nous avons menés concernent également la question, très importante en écologie, des **preferendums**. Existents-ils? et comment les détecter?

Il est un preferendum facile à deviner; c'est celui de Dave Reid (coordinateur principal du projet SEFOS) en matière de latitude. Pour assurer la collaboration entre les participants de ce gros projet européen, deux réunions ont eu lieu chaque année à des latitudes qui trahissent le preferendum en la matière du coordinateur principal du projet (voir figure ci-dessous)... Il est par ailleurs amusant de voir que le centre de gravité, dont on voit ici une première utilisation ludique (d'autres, un peu moins ludiques, encore que, suivront dans ce document), de ces réunions se trouve quasiment sur les accores (bord du plateau continental) au large de Bordeaux. C'est donc que Dave Reid suivait une logique secrète dans le choix des lieux de réunions (bravo!): celle d'avoir été en moyenne sur les accores au cours de ces 3 années de travail. Le choix de Bordeaux comme position moyenne des réunions n'est probablement pas non plus vide de signification ... à l'instar des bouteilles de cette région que beaucoup de contributeurs à ce projet ont eu du mal à boycotter.



Centre de gravité des réunions SEFOS... et lieux de rencontres induites par le projet (villes écrites en petit)



Partenaires présents à la réunion de IJMuiden (Hollande). De gauche à droite :

Jacques Rivoirard (France), Luis Valdes (Espagne), Nicolas Bez (France), Adolfo Uriarte (Espagne), Peter Miller (Angleterre), Steve Coombs (Angleterre), Morten Skogen (Norvège), Einar Svensen (Norvège), Alasdair MacIntyre (Ecosse), July Fives (Irlande), Gerrit van de Kamp (Hollande), Fatima Borges (Portugal), Anthony Walne (Angleterre), Carmela Porteiro (Espagne), Christian Mohn (Allemagne), Amor Sola (Espagne), David Reid (Ecosse), Terje Monstad (Norvège), Lorenzo Motos (Espagne), Alexandra Silva (Portugal), Joachim Bartsch (Allemagne), Hein von Westernhagen (Allemagne), Bill Turell (Ecosse), Jean-Charles Poulard (France), Ad Corten (Hollande).

Si la rédaction de ce document est le fruit d'un travail personnel, je souhaite insister sur le fait qu'il émane d'une collaboration étroite avec Jacques Rivoirard. La plupart des réflexions que vous trouverez dans ce document n'auraient pas vu le jour sans cette collaboration, que de nombreux thésards, je crois, souhaiteraient avoir avec leurs directeurs de thèse. C'est peu dire que Jacques a su répondre à chacune de mes sollicitations. Ce qu'il faut savoir, c'est qu'elles ont été très nombreuses ! très TRÈS nombreuses !!!

Pour faire un peu d'esprit dont chacun appréciera le niveau, je crois que cette thèse est, en fait, une histoire d'eau. De part son sujet bien sûr. Mais aussi, de part la nature même des acteurs. En photographie, l'eau n'est qu'un révélateur. Sans l'eau, la photo n'est rien. Elle ne devient jamais réalité. Mais, ses pigments, c'est-à-dire sa richesse, existent indépendamment de l'eau. Dans la collaboration que j'ai eu avec Jacques, je crois avoir tenu le rôle de l'eau, révélatrice d'une maturité géostatistique extérieure. Quand bien même Jacques voudrait-il s'en défendre, il ne le pourrait pas. C'est lui même qui m'a enseigné que j'étais, par mon signe astrologique j'entends, marqué par l'eau. Argument déterminant, non ? Mais, foin d'arguments. Ils me sont de toutes les façons inutiles pour adresser ma sincère et profonde reconnaissance à Jacques.

Enfin, cette préface ne serait pas complète si elle ne revenait pas aux origines. Et, en particulier à mon grand-père que je n'ai pas eu la joie de connaître et qui avait pourtant probablement inscrit ma route quelque part : petit-fils de Daniel Poisson, ingénieur des Mines, promotion 1925, que pouvais-je faire d'autre qu'une thèse sur les poissons à l'École des Mines...

Table des matières

Remerciements	i
Préface	iii
Introduction	1
I Considérations générales	3
1 Exemples introductifs	5
1.1 Estimation d'une densité moyenne	6
1.1.1 Délimitation du domaine. Formation d'un estimateur	7
1.1.2 Qualité de l'estimateur et besoin de modélisation	8
1.1.3 Remarques complémentaires	10
1.2 Analyse variographique des températures de surface	11
1.2.1 Nature des difficultés	11
1.2.2 Résultats	15
1.3 Analyse variographique de densités de merlu	16
1.3.1 Présentation des données	16
1.3.2 Des variogrammes expérimentaux "décevants"	17
1.3.3 Pourquoi de tels variogrammes?	19
1.4 Analyse structurale d'un courant marin	21
1.4.1 Présentation des données	21
1.4.2 Analyse structurale	25
1.5 Synthèse: nature des difficultés	26
2 Choix d'un concept opératoire et pertinent	29
2.1 De la moyenne à l'abondance pour une densité	30
2.1.1 Le problème du champ, des 0 et des ϵ	30
2.1.2 De la moyenne à l'abondance	31
2.1.3 Généralisation du principe: "ce qui était moyenne devient somme"	32
2.2 Régressions densité \times température	34
2.2.1 La régression dans un contexte spatial	34
2.2.2 Abondances de poissons par classe de température	37
2.3 Et si nous connaissions tout?	37
2.3.1 Pertinence d'un concept	38

2.3.2	Objectivité	38
3	Choix d'un modèle	41
3.1	Quel modèle et pour quoi faire?	42
3.1.1	Choix d'un modèle de loi de probabilité en statistique "classique"	42
3.1.2	Choix d'un modèle structural en géostatistique	43
3.1.3	Approches modèle-dépendantes?	44
3.2	Les 3 étapes du choix d'un modèle	44
3.2.1	Choix d'un modèle constitutif	44
3.2.2	Choix d'un modèle générique	46
3.2.3	Choix d'un modèle spécifique	46
3.3	Choix d'un modèle constitutif	47
3.3.1	Fonction aléatoire, stationnarité et champ	47
3.3.2	Causes possibles d'une non-stationnarité	48
3.3.3	Variable régionalisée, échelle et intégrale	50
II	Statistiques par individu et géostatistique transitive.	53
1	Loi d'un individu aléatoire	55
1.1	Définitions	56
1.1.1	Individu aléatoire	56
1.1.2	Loi de x_I	57
1.1.3	Centre de gravité et inertie	57
1.1.4	Commentaires	58
1.2	Estimation d'abondances totales	58
1.2.1	Echantillonnage à maille régulière ou systématique	58
1.2.2	Deux exemples d'échantillonnages par processus ponctuels	60
1.2.3	Echantillonnage quelconque	62
1.3	Application : estimation du centre de gravité et de l'inertie	63
1.4	Descriptions de séries historiques de cartes	64
1.4.1	Déplacement moyen d'une population	64
1.4.2	Axes principaux de la dispersion	64
1.4.3	Remarques	66
2	Caractéristiques d'un individu aléatoire.	69
2.1	Température de surface par individu	70
2.1.1	Formulation générale	70
2.1.2	Centre de gravité et inertie des températures par individu	71
2.1.3	Loi et histogramme expérimental	71
2.1.4	L'alternative entre régression et histogramme des abondances	72
2.2	Loi d'un couple de variables par individu	73
2.3	Preferendum dans un contexte spatial.	73
2.3.1	Maximum et preferendum	73
2.3.2	Critère monovariable pour diagnostiquer un preferendum	77
2.3.3	Difficultés pratiques et choix d'un paramètre conditionnant	78
2.3.4	Vers un modèle dynamique	81

3	Le covariogramme	83
3.1	Le covariogramme: définition et propriétés	85
3.1.1	Définition	85
3.1.2	Propriétés	86
3.1.3	Cas particulier : le covariogramme géométrique	89
3.2	Le covariogramme pour l'analyse structurale	91
3.2.1	Distance entre deux individus aléatoires	91
3.2.2	Covariogramme et covariance	93
3.2.3	Covariogramme et régressions	94
3.2.4	Seuils, ensembles et fréquences conditionnelles	96
3.3	Le covariogramme pour l'estimation globale et locale	97
3.3.1	Variance d'estimation globale à maille régulière	97
3.3.2	Krigeage transitif	100
3.4	Estimations et modélisation	103
3.4.1	A maille régulière	103
3.4.2	A maille irrégulière	104
3.4.3	Contraintes pour la modélisation	106
3.4.4	Covariogrammes: le vrai et le modélisé	108
3.4.5	Remarques sur les "simulations"	108
4	Densité par individu et effet de support	111
4.1	Effet de support: principe et définitions	112
4.1.1	Principe	112
4.1.2	Régularisation des variables	112
4.2	Densité par individu	114
4.2.1	Définitions	114
4.2.2	Densité par individu pour une régularisée	115
4.2.3	Les deux types de loi	115
4.3	Le bénéfice conventionnel	118
4.3.1	Définitions	118
4.3.2	Interprétation et propriétés	118
III	Analyse structurale et cartographie du merlu du Golfe de Gascogne	121
1	Présentation des données	123
1.1	Situation générale	124
1.2	Surfaces d'influences	126
1.3	Résultats préliminaires	127
1.3.1	Statistiques élémentaires	127
1.3.2	Distributions statistiques	127
1.3.3	Courbes de sélectivité	129
2	Analyse structurale	131
2.1	Estimations du covariogramme	132
2.1.1	Estimation du covariogramme transitif par migration vers une grille fine . . .	132
2.1.2	Estimation du covariogramme géométrique par migration vers une grille fine	132

2.1.3	Estimation du covariogramme par la covariance	135
2.1.4	Estimation du covariogramme par les régressions	135
2.1.5	Synthèse et interprétation structurale	135
2.2	Régressions bivariées par classe	138
2.3	Fréquences conditionnelles	141
3	Cartographie	145
3.1	Choix d'un modèle de covariogramme	146
3.2	Carte: méthode, résultat et discussion	146
IV	Etude de la ponte et des phases larvaires du maquereau	149
1	Biologie du maquereau	151
1.1	Importance commerciale du maquereau	151
1.2	Cycle biologique du maquereau	152
1.2.1	Description sommaire du cycle de vie	152
1.2.2	Remarques sur le projet SEFOS	154
2	Présentation des données et résultats préliminaires	155
2.1	Echantillonnage	157
2.1.1	Plan d'échantillonnage	157
2.1.2	Engins et modes de pêche	162
2.1.3	Variables mesurées	164
2.2	Traitements préliminaires	166
2.2.1	Echantillons aberrants	166
2.2.2	Passage à une densité par unité de surface	166
2.2.3	Taux journaliers de production	166
2.2.4	Homogénéité des mesures hydrographiques	168
2.2.5	Migration vers une grille régulière en degrés	173
2.3	Description d'une saison de ponte moyenne	174
2.3.1	Représentations proportionnelles	174
2.3.2	Covariogrammes moyens	177
2.4	Effet de support	177
2.4.1	Définition du support	177
2.4.2	Mesure de l'effet de support : campagne 1992 - période 2.	178
2.4.3	Support moyen par période	182
3	Analyse de la ponte de 1983 à 1995	185
3.1	Densité par individu et modèle possible pour la reproduction	186
3.1.1	Abondance, densité moyenne et surface équivalente	186
3.2	Localisation moyenne de la ponte	190
3.2.1	Persistance des déplacements et flux migratoires	190
3.2.2	Inertie	190
3.2.3	Déplacement vers l'Ouest	190
3.3	Structures spatiales	190
3.3.1	Niveau d'agrégation	190

3.3.2	Anisotropie	196
3.3.3	Stabilité du niveau du covariogramme en dehors de l'origine	196
3.4	Transformées de la profondeur	199
3.4.1	Log-translation	199
3.4.2	Laplacien	202
3.5	Température de surface	202
3.5.1	Température moyenne de la ponte	202
3.5.2	Y-a-t-il un preferendum; et de quelle nature?	207
3.6	Synthèse	215
4	Etude du développement embryonnaire et larvaire	217
4.1	Analyse du dernier stade embryonnaire (E5)	218
4.1.1	Abondance, densité moyenne et surface équivalente	218
4.1.2	Localisation et transport	221
4.1.3	Remarques sur la température de surface	225
4.1.4	Structures spatiales	228
4.1.5	Synthèse	228
4.2	Analyse du développement larvaire en 1989	229
4.2.1	Courbes de production et mortalité totale	229
4.2.2	Comparaisons de quelques statistiques individuelles pour la distinctions entre stades passifs et actifs	230
4.2.3	Transports à moyen terme (2-3 semaines)	234
4.2.4	Structures spatiales	236
4.2.5	Synthèse	238
	Conclusion	239
	Bibliographie	243
	Annexes	249
	Annexe 1: Analyse structurale d'un courant marin	249
	Annexe 2: Représentations proportionnelles	255

Introduction

Compte-tenu du savoir acquis sur plusieurs décennies en écologie halieutique et du développement des outils d'investigation (acoustique) et d'analyse (informatique), l'enjeu actuel des analyses statistiques et des essais de modélisation est souvent de s'approcher au plus près de la complexité supposée des systèmes halieutiques. Face à cette complexité, les écologistes, les biologistes ou encore les halieutes ont bien souvent une "vision" très large (pluri-disciplinaire) qui leur permet de dresser un schéma possible de fonctionnement des biotopes étudiés. Mais les moyens manquent parfois pour quantifier ne serait-ce que l'ordre de grandeur de certaines parties du schéma général. A l'extrême, en imaginant que nous connaissions tout sur tout (Cf. § I.2.3), il n'est pas sûr que nous sachions déterminer, s'il existe, le rôle conditionnant d'une variable de l'environnement sur la régionalisation d'une espèce.

D'un point de vue statistique, un écueil fréquent est celui de la distribution très dissymétrique des valeurs de densités, avec la présence de nombreuses valeurs nulles ou faibles dont le traitement pose problème. En particulier, la délimitation du domaine sur lequel doit être calculée une moyenne ou construit un nuage de corrélation devient une étape très délicate des analyses. On aura d'ailleurs noté, que cet outil de base des analyses multivariées qu'est le nuage de corrélation, n'a de sens qu'à l'intérieur d'un domaine précis de l'espace. N'est-ce pas là une limitation méthodologique, surtout quand la question posée cherche à replacer les individus dans leur contexte spatial? On a de plus de bonnes raisons de penser, comme cela sera exposé dans ce document (Cf. § II.2.3), que les courbes de régressions déduites de ces nuages de corrélations doivent être interprétées en fonction des structures spatiales de chacune des variables en jeu.

L'analyse de la complexité vers laquelle tendent la plupart des études d'écologie halieutique, va souvent de pair avec une modélisation elle-même complexe. Mais le prix de la complexité n'est-il pas parfois trop élevé? Un autre écueil fréquent est en effet celui de savoir si le modèle n'est pas trop "exigeant" par rapport à la réalité; soit parce qu'il repose sur un nombre trop important de paramètres, dont la spécification devient alors délicate voire impossible, soit parce que les hypothèses qui le fondent sont trop fortes par rapport à la réalité du système étudié. Le modèle peut alors progressivement perdre sa pertinence. Cette dernière objection a une incidence très précise en géostatistique. En particulier, la variabilité naturelle des systèmes halieutiques est-elle toujours compatible avec une analyse structurale basée sur le variogramme? Peut-elle expliquer que certains des résultats expérimentaux que nous obtenons soient difficilement interprétables?

Dans l'étude de la répartition spatiale des œufs de maquereau, ou encore pour l'analyse structurale de la régionalisation des merlus d'âge 0 aux alentours de la grande vasière du Golfe de Gascogne, j'ai été amené à revenir très en amont dans les choix de modélisation. Ce faisant, une "nouvelle"

méthodologie s'est construite, adaptée aux questions posées et me semble-t-il, à de nombreuses situations pratiques. C'est la raison pour laquelle, j'ai cherché à la présenter pour elle-même sous le nom de "statistique individuelle" avant d'en voir les applications pratiques (troisième et quatrième parties). Cette méthodologie est assez simple dans son principe et il m'a semblé aussi important de présenter les outils utilisés (deuxième partie) que d'expliquer pourquoi nous y avons eu recours (première partie); ce qui m'a conduit à suivre le plan suivant.

La **première partie** de cette thèse tente d'expliquer, en disséquant l'enchaînement des choix réalisés généralement en statistique, quelles sont les raisons qui ont poussé à ne pas utiliser des outils disponibles comme le variogramme, le variogramme croisé ou les régressions bivariées. C'est une partie assez peu mathématique qui traite essentiellement des concepts en jeu et qui constitue, en réalité, une (très) longue introduction au présent travail.

J'expose dans la **deuxième partie** les statistiques par individu. En dehors des statistiques simples que sont le centre de gravité et l'inertie, cela conduira d'une part vers le covariogramme dont nous verrons l'usage en estimation globale, en analyse structurale et en cartographie, et d'autre part vers la formulation d'un test pour juger de l'existence de préférendum dans un contexte spatial (inertiogramme).

La **troisième partie** permet d'illustrer le pouvoir descriptif du covariogramme en tant qu'outil structural pour aboutir à la cartographie par krigeage transitif des densités de merlus en 1990 dans le Golfe de Gascogne.

Enfin la **quatrième partie** concerne l'analyse de 5 années d'observations de la reproduction du maquereau. L'ensemble des statistiques individuelles sont alors sollicitées pour caractériser des éléments persistants ou à l'inverse, quantifier la variabilité de certaines caractéristiques du processus de reproduction.

Le lecteur trouvera un sommaire au début de chaque chapitre en français et en anglais qui devrait, je l'espère, aider à suivre le fil de l'exposé.

Par ailleurs, un certain nombre de pages sont laissées vierges à la fin du document pour d'éventuelles notes personnelles.

Première partie
Considérations générales

Chapitre 1

Exemples introductifs

Sommaire

Une série de 4 exemples (réels et inventés) permet d'illustrer les problèmes de choix de modèles et de concepts en statistique.

*Le **premier exemple**, emprunté à la statistique classique, est analysé sous l'angle du choix d'un modèle. On commence par illustrer la difficulté réelle de définir et de délimiter un champ. L'estimation de la densité moyenne sur le champ est ensuite l'occasion de préciser la différence sémantique entre un choix, une hypothèse et une estimation.*

*Le **deuxième exemple** concerne l'analyse structurale de la température de surface. Il montre les difficultés pratiques de calcul de variogramme dans le cas de variables présentant une dérive spatio-temporelle. C'est aussi et surtout l'occasion de présenter des variogrammes expérimentaux satisfaisants et exploitables.*

*Le **troisième exemple**, en revanche, concerne un cas où l'analyse variographique est décevante. Il s'agit de l'analyse structurale de la densité de merlu dans le Golfe de Gascogne. Les raisons de cet échec sont discutées.*

*Enfin, le **quatrième exemple**, est l'occasion de présenter quelques précautions nécessaires à l'analyse d'une variable vectorielle. L'illustration qui en est donnée, provient de l'analyse d'un courant marin sur la base de données acoustiques (ADCP).*

Exemples

Abstract

This chapter is made of 4 examples (real or made up) which help to illustrate problems like the choice of a model or a concept in statistics.

*The **first example**, coming from "classical" statistics, is looked at from the point of view of the choice of a model. We first underline and illustrate how difficult the delimitation of a field might be in fisheries. The estimation of the mean density over such a field is then the opportunity to clarify the difference we should make between a choice, an hypothesis and an estimation.*

The second example deals with the structural analysis of sea surface temperature. This example points out the kind of practical difficulties we encounter when the variable under study gets a strong spatio-temporal drift. But this example is also, and mainly, an occasion of showing satisfactory and usable experimental variograms.

On the other hand, the third example concerns, a case where the experimental variogram is not very good. This example is a matter of hake density in the Bay of Biscay. The reasons explaining the failure of the variographic analysis are analysed.

Then, the fourth example is just an illustration of some of the necessary precautions for the analysis of vectorial data. An illustration is given about acoustics measurements by ADCP of water motion.

Pour mieux comprendre la nature des difficultés rencontrées en statistique ou en géostatistique appliquée à l'écologie halieutique, je propose de présenter d'abord quelques exemples. Je m'attacherai dans cette première partie à distinguer, chaque fois que cela sera nécessaire, les statistiques dites classiques ou orthodoxes de la géostatistique; la statistique au sens large faisant référence à l'ensemble de ces deux familles de techniques.

Le premier exemple, complètement inventé est très simple. Il me permettra de préciser la nature des problèmes auxquels je me suis attaché pendant mon travail et de faire quelques remarques sur l'utilisation des statistiques "classiques".

Les trois exemples suivants ont pour but d'illustrer les difficultés que nous rencontrons assez souvent en géostatistique, tant du point de vue pratique que du point de vue théorique. Il s'agit, lors de l'analyse structurale, de la prise en compte : d'une dérive (analyse d'une température de surface), de la répartition dans le temps des échantillons ou encore du caractère vectoriel de la variable étudiée (courants marins). Mais surtout, la comparaison des résultats concernant la température de surface d'une part et le merlu d'autre part, sera le point de départ de notre réflexion sur le choix d'une méthodologie géostatistique adaptée aux données halieutiques. En somme, cette série d'exemples permettra de mieux comprendre les motivations qui ont conduit les développements méthodologiques présentés dans ce document.

1.1 Estimation d'une densité moyenne sur un domaine en statistique classique

La démarche poursuivie dans ce premier exemple emprunté à la statistique "classique", est inspirée de la deuxième partie de *Estimer et choisir* (Matheron, 1978). Sur la base d'un exemple très simple (et inventé d'ailleurs), nous allons porter notre attention sur la succession des choix que nous sommes amenés à faire pour procéder en routine à l'estimation d'une densité moyenne. Quelques uns de ces choix sont souvent faits de façon implicite et/ou automatique. Nous verrons qu'il convient dans certains cas, et en écologie halieutique en particulier, d'y apporter beaucoup de vigilance.

Supposons que nous étudions un phénomène à une seule dimension à l'aide d'un échantillonnage à maille régulière. Comme cela est souvent la règle, choisissons d'arrêter l'échantillonnage quand 3 densités nulles consécutives sont observées. Soit alors les données expérimentales suivantes où ϵ représente une valeur très faible :

0 - 0 - 0 - ϵ - ϵ - 0 - 0 - 1 - 1 - 10 - 1 - 1000 - 0 - 1 - 10 - 1 - ϵ - 0 - 0 - 0

Les valeurs numériques expriment par exemple des densités en nombre d'individus par mètre carré (ind/m^2). On retrouve quelques caractéristiques typiques d'une densité de poissons (ou encore d'un niveau de pollution). En particulier, on peut constater que sur les 20 échantillons générés, 9 ont une valeur nulle et un très grand nombre (17) une valeur très faible (≤ 1). On ne dispose par ailleurs que d'une seule (5%) donnée très forte, entourée de valeurs faibles illustrant par là une forte hétérogénéité spatiale.

1.1.1 Délimitation du domaine. Formation d'un estimateur

Pour décrire un ensemble de données comme celui là, il est commode et fréquent d'en construire l'histogramme expérimental. Ce dernier peut d'ailleurs être simplement réduit à la donnée d'une valeur moyenne et d'une variance. Or, chacun de ces concepts fait implicitement référence à un domaine. Ils ne peuvent être effectivement calculés que si on définit au préalable le domaine sur lequel ils doivent être évalués. Dans notre cas, plusieurs possibilités sont envisageables :

- Le domaine de calcul peut correspondre à l'ensemble des 20 échantillons disponibles. Dans ce cas, on concluerait à une densité moyenne de $51.25 \text{ ind}/\text{m}^2$. Cependant, il n'y a pas de raison particulière d'avoir décidé de stopper l'échantillonnage quand 3 zéros consécutifs étaient observés. Si on avait décidé d'arrêter l'échantillonnage après quatre densités nulles consécutives, nous aurions recueilli 22 échantillons, et par là une densité moyenne de $46.6 \text{ ind}/\text{m}^2$. A l'extrême, la population étant entourée de valeurs nulles, on pourrait étendre le domaine à l'infini et tendre alors vers une moyenne nulle.
- Cependant, en considérant que 3 densités consécutivement nulles signifie qu'on a dépassé la population, il est logique de limiter le domaine aux 14 échantillons "intérieurs".

0 - 0 - 0 - ϵ - ϵ - 0 - 0 - 1 - 1 - 10 - 1 - 1000 - 0 - 1 - 10 - 1 - ϵ - 0 - 0 - 0

Sur ce domaine la moyenne expérimentale est $73.2 \text{ ind}/\text{m}^2$.

- On peut également limiter le domaine aux seules valeurs strictement positives.

0 - 0 - 0 - ϵ - ϵ - 0 - 0 - 1 - 1 - 10 - 1 - 1000 - 0 - 1 - 10 - 1 - ϵ - 0 - 0 - 0

Ainsi, sur ce champ particulier, la moyenne des 11 données expérimentales non nulles est $93.2 \text{ ind}/\text{m}^2$.

Il est toutefois intéressant de noter qu'un autre statisticien pourrait décider d'inclure dans le champ l'une ou l'autre, ou encore l'ensemble des 3 valeurs nulles se trouvant à l'intérieur de la population. Soit qu'il décide de considérer qu'il existe naturellement dans la population des trous de densités. Soit qu'il juge que les moyens techniques d'échantillonnage ne permettent pas de dire avec certitude s'il n'y avait pas, en fait, une quantité minimale de poissons non détectée. Nous reviendrons plus en détails sur ce point dans le chapitre I.2.1.1.

Sur la base de cet exemple, certes caricatural, on retiendra que les concepts de moyenne, de variance, ou encore par exemple d'histogramme, sont définis relativement à un domaine dont la définition et la délimitation sont laissées à l'appréciation du praticien.

Ayant **choisi** le domaine, l'estimateur de la densité moyenne sur ce domaine est fourni simplement par la moyenne arithmétique des échantillons correspondants. Par le choix du domaine et du recours à la moyenne arithmétique, nous sommes donc parvenus à estimer la densité moyenne

sur le domaine. Jusque là, nous n'avons fait qu'utiliser le matériel expérimental disponible. Nous n'avons parlé ni de variables aléatoires ni de modèle d'aucune sorte.

Cependant, nous ne connaissons pas la qualité de notre estimation. Pour aller plus loin, et par exemple pour estimer le biais, la robustesse ou la variance de l'estimation, des choix de modélisation s'avèrent nécessaires.

1.1.2 Qualité de l'estimateur et besoin de modélisation

A ce stade, on choisit de considérer les données expérimentales comme des réalisations indépendantes d'une **variable aléatoire**. Une autre possibilité serait de considérer que les 20 échantillons constituent 20 réalisations de 20 variables aléatoires. Toutefois, dans l'un ou l'autre des cas, on choisit de recourir aux **méthodes probabilistes**. L'objet d'étude n'est plus la densité, mais un objet mathématique (une variable aléatoire) dont on espère que les propriétés vont nous permettre de répondre à la question posée.

Biais

Le recours aux techniques probabilistes est implicite dès lors qu'on parle du biais d'un estimateur. En effet, le biais est défini comme l'**espérance** de la différence entre l'estimateur et la valeur vraie inconnue. Or, parler d'espérance, c'est bien constituer l'estimateur en variable aléatoire, et donc se placer implicitement dans le contexte d'un modèle probabiliste.

Habituellement, dans l'estimation de la moyenne, le non biais est assuré par l'hypothèse que les 20 variables aléatoires sont indépendantes et de même loi, sans toutefois qu'il soit nécessaire de préciser le type de la loi de probabilité, ou même sa forme (symétrique, uni ou multimodale,...).

Ce premier exemple de critère de qualité d'un estimateur repose donc sur deux hypothèses relatives aux caractéristiques des variables aléatoires.

Robustesse

Par définition, *une estimation est robuste si son résultat est peu influencé par les erreurs que nous avons pu commettre soit dans le relevé des données expérimentales, soit dans le choix du modèle* (Matheron, 1978). On doit donc distinguer deux formes de robustesse. Une robustesse par rapport aux données mais également une robustesse par rapport au modèle.

- Robustesse par rapport aux données:

Dans l'exemple qui nous occupe, on cherche à estimer une densité moyenne avec des valeurs expérimentales présentant quelques valeurs très fortes et une large majorité de concentrations faibles. Dans ce contexte, la moyenne est réputée être un estimateur moins robuste aux valeurs extrêmes que la médiane. Et, de ce fait, on préfère parfois la médiane à la moyenne.

Une première objection me paraît utile ici. Dans l'hypothèse où l'effort d'échantillonnage pourrait augmenter à l'infini, c'est-à-dire dans l'hypothèse où nous disposerions d'une connaissance de plus en plus complète de la réalité, il est certain que la médiane s'écarterait de la moyenne et ne serait donc pas un bon estimateur. **Seule la moyenne arithmétique conviendrait**. Bien que basée sur une situation utopique, cette remarque constitue une objection certaine au recours à la médiane. Nous reviendrons plus loin sur l'intérêt et le besoin de supposer par moment que nous connaissons exhaustivement la réalité.

Par ailleurs, la robustesse de la médiane aux grandes valeurs de densités n'est démontrée que dans le cadre théorique d'une **loi de probabilité symétrique** (Saporta, 1990). Il faut donc, non

seulement spécifier la forme de la loi de probabilité, mais également s'assurer du bien fondé de cette nouvelle décision. On voit alors, que pour améliorer la robustesse par rapport aux données, on a dans le même temps détérioré la robustesse vis-à-vis du modèle.

Mais il y a une objection plus importante encore. **Est-il justifié de rechercher des estimateurs robustes quand les données semblent indiquer précisément une "réalité non robuste"?** Et ce particulièrement dans le domaine des pêches, où il est fréquent de rencontrer des régionalisations hautement variables. Dans ces conditions, la robustesse d'un estimateur tiendrait plus à quelques hypothèses concernant un modèle mathématique qu'au phénomène observé. Or, parmi les qualités attendues d'un modèle, on compte généralement le "respect des données"; c'est-à-dire le réalisme du modèle. Ainsi, en débordant parfois la réalité biologique, et en particulier en négligeant la variabilité intrinsèque des densités de poissons, la robustesse de la méthode est probablement au delà du seuil de réalisme. Je tiens à préciser que ce genre de remarques, malgré sa formulation, n'a pas vocation à être générale. Il est bien entendu que de nombreux phénomènes halieutiques présentent des niveaux de variabilité convenables, eu égard aux prérequisites des modèles probabilistes.

- Robustesse par rapport au modèle :

Je ne développerai pas ici ce type de robustesse qui est expliqué en détail dans la deuxième partie de *Estimer et Choisir* (Matheron, 1978). L'idée directrice est la suivante : la robustesse par rapport au modèle se détériore avec un niveau croissant de spécification. Cette idée peut également s'exprimer en termes de prise de risques. En effet, en progressant dans la modélisation, on prend le risque que les hypothèses qui permettent de spécifier le modèle relatives au type de loi ou à l'indépendance des variables aléatoires soient fausses. On conçoit alors assez bien que plus on prend des risques, plus on fragilise le résultat final de l'analyse.

A cet égard, les résultats obtenus par les techniques du maximum de vraisemblance, techniques qui permettent d'estimer les paramètres d'une loi de probabilité donnée (choisie par nous), n'ont qu'une robustesse relative vis-à-vis du modèle.

Variance

La variance de l'erreur d'estimation mesure la précision d'une estimation. Dans le cadre de notre modèle de variable aléatoire, connaître la variance d'estimation revient à spécifier le moment d'ordre deux de la loi de probabilité qui a été préalablement choisie. Ce moment d'ordre 2 est estimé puisque inconnu.

Qu'a-t-on choisi? Qu'a-t-on supposé? Qu'a-t-on estimé?

Au cours de cette estimation, il est donc apparu indispensable de spécifier de plus en plus le modèle de variable aléatoire pour obtenir de plus en plus d'informations sur la qualité de l'estimateur. Au delà de cette première remarque, il paraît intéressant, compte tenu de l'objectif poursuivi dans la présentation de cet exemple, de récapituler la succession des choix réalisés.

Malgré les apparences, **les deux premiers choix** concernent l'objectif et les moyens de l'étude. On a en effet décidé de s'intéresser à la **moyenne** dont l'utilisation conduit d'ailleurs à un problème de délimitation du domaine d'étude. Et il a également été décidé d'estimer cette moyenne en utilisant les outils de la **statistique classique**. Ces deux décisions n'ont pas été mises en exergue jusque là. Mais elles sont apparues intentionnellement dans le titre du paragraphe pour illustrer le

statut qu'on leur accorde en général. En effet, ces deux éléments sont souvent considérés comme des données du problème, quand, en réalité, ce sont deux choix très importants et très amont pris par le praticien. Ils conditionnent d'ailleurs largement la suite de l'étude.

Après avoir choisi le domaine d'étude, on a choisi de prendre comme estimateur la moyenne arithmétique. Ensuite, il a été décidé de considérer les valeurs expérimentales comme les réalisations d'une variable aléatoire, ce qui est un choix épistémologique. Puis, on a supposé que les variables aléatoires étaient indépendantes et de même loi. Enfin, on a choisi un type de loi de probabilité et procédé à l'estimation des quelques paramètres nécessaires à sa spécification.

S'il est nécessaire de bien savoir ce qui relève du choix du statisticien, il est aussi important de connaître la nature des choix réalisés. Certains choix sont de nature épistémologique. Ce qui signifie entre autre chose que ce ne sont pas des hypothèses contrôlables expérimentalement mais plus des décisions intuitives. On aurait pu par exemple choisir un autre concept que la moyenne, un autre champ que celui retenu et une autre méthodologie que la statistique classique sans qu'aucun argument objectif ne vienne trancher la question. C'est précisément sur ce type de choix que portera une large partie de mes réflexions.

D'autres choix correspondent par contre à des hypothèses (suppositions) sur la nature physique du phénomène, et sont donc à ce titre **susceptibles d'être réfutés ou non**.

1.1.3 Remarques complémentaires

J'ai pris soin dans l'exemple précédent de préciser que le besoin de modélisation provient du besoin de quantifier la qualité de l'estimation. C'est là un premier exemple de la règle suivante, démontrée par Matheron dans le cadre de la géostatistique transitive : **on ne peut pas tirer d'un unique matériel expérimental à la fois une estimation et la qualité de celle-ci**.

Par ailleurs, par le processus de modélisation, le statisticien apporte de l'information à l'ensemble constitué des valeurs expérimentales et du réseau d'échantillonnage. Cet apport d'information, qui permet justement de tirer des données plus qu'il n'y est réellement contenu, doit donc être réalisé avec le maximum de précautions, et en particulier avec objectivité.

Assez tôt dans la spécification du modèle, les données ont été considérées comme des réalisations indépendantes de variables aléatoires. Cela repose en général sur le choix d'une stratégie d'échantillonnage adaptée (points uniformes dans le domaine). Sinon, lorsque le processus d'échantillonnage est, soit inconnu soit non aléatoire, cela relève d'une hypothèse sur le phénomène étudié. Dans notre exemple, il était en fait difficile d'invoquer l'indépendance des réalisations puisqu'elles sont issues d'un échantillonnage régulier.

De plus, en choisissant la statistique classique, on a décidé de ne pas tenir compte de la localisation des points dans l'espace. C'est-à-dire que la situation observée :

$$0 - 0 - 0 - \epsilon - \epsilon - 0 - 0 - 1 - 1 - 10 - 1 - 1000 - 0 - 1 - 10 - 1 - \epsilon - 0 - 0 - 0$$

est considérée équivalente à une autre situation du type :

$$0 - \epsilon - 0 - 1000 - 10 - 1 - 0 - \epsilon - 1 - 0 - 1 - 0 - 1 - 10 - 0 - 1 - 0 - 0 - \epsilon - 0$$

Si effectivement la position des densités n'a, soit pas d'intérêt pour l'utilisateur, soit pas d'importance, alors l'histogramme est un bon outil pour décrire les données. Il est en effet identique dans l'une et l'autre de ces deux situations. Par contre, il n'est plus possible, sans hypothèse supplémentaire, d'interpréter cet histogramme dans un contexte spatial. L'histogramme modélisé ne représente pas, *a priori*, la probabilité d'obtenir une concentration donnée en un point non échantillonné de la population. Pour ce faire, il faudrait invoquer une hypothèse très forte de stationnarité, puisqu'il faudrait supposer que la loi de probabilité toute entière, et pas seulement ses deux premiers moments comme cela est pratiqué en géostatistique linéaire, est identique partout.

Si l'histogramme en tant qu'outil expérimental reste un moyen pratique pour décrire un ensemble de données **régionalisées**, aux inconvénients de délimitation de domaine près, sa modélisation et son utilisation en tant que loi de probabilité n'est possible que dans un cadre stationnaire.

Pour tenir compte de la localisation des points, une solution possible est de faire appel à la géostatistique. Les exemples qui suivent indiquent quelques applications possibles de la géostatistique en écologie halieutique.

1.2 Analyse variographique des températures de surface

Comme nous le verrons, la température de surface, comme d'autres paramètres physiques de l'environnement marin d'ailleurs, se prête relativement bien au traitement par la géostatistique. Dans cet exemple, dès le début, c'est-à-dire dans le titre, il est décidé de recourir à la géostatistique intrinsèque, en cela qu'on se propose d'utiliser le variogramme comme outil structural. Est-ce justifié? La variable le permet-elle? Ces deux questions seront en fait illustrées par le troisième exemple. Dans l'étude structurale de la température de surface de l'Atlantique Nord-Est, on s'attachera plus à analyser quelques difficultés techniques du calcul expérimental de variogrammes qu'à rechercher les causes d'une éventuelle inadaptation de la méthode. On se félicitera au contraire de la qualité des résultats.

1.2.1 Nature des difficultés

Les difficultés rencontrées dans cette analyse proviennent à la fois des caractéristiques intrinsèques de la température de surface et de la nature du plan d'échantillonnage.

Echantillonnage spatio-temporel

Les données proviennent de campagnes organisées par le Conseil International pour L'Exploration de la Mer (CIEM). On trouvera une présentation détaillée du mode d'échantillonnage plus loin dans le document (§IV-2). Organisé autour du bord du plateau continental entre le Sud du Golfe de Gascogne et le Nord de l'Ecosse, l'échantillonnage correspond à des transects parallèles parcourus en "S" (Fig. 1.1). Chaque année, pour couvrir la période de ponte du maquereau, plusieurs passages (périodes d'échantillonnage) sont réalisés. Selon la période, l'échantillonnage se fait du Sud vers le Nord, ou du Nord vers le Sud.

Compte-tenu du temps nécessaire au prélèvement de chaque échantillon biologique, il s'écoule un laps de temps d'environ trois semaines entre le premier et le dernier transect d'une même période d'échantillonnage. Toutefois, chaque transect est parcouru rapidement (quelques dizaines d'heures) de sorte qu'il est possible de considérer que les températures observées sur un même transect sont simultanées.

Dérive spatio-temporelle de la température de surface

Globalement, la température de l'eau, et surtout celle de la couche superficielle, décroît quand la latitude augmente. En termes géostatistiques, on parlera de dérive spatiale Nord-Sud. De plus, les campagnes scientifiques ont lieu au printemps (entre avril et juillet), au moment où le réchauffement de l'eau en surface est intense.

Pour illustrer les problèmes dont nous allons parler, cette combinaison de dérives spatiale et temporelle est traduit schématiquement (Fig. 1.2), par deux droites représentant l'évolution de la

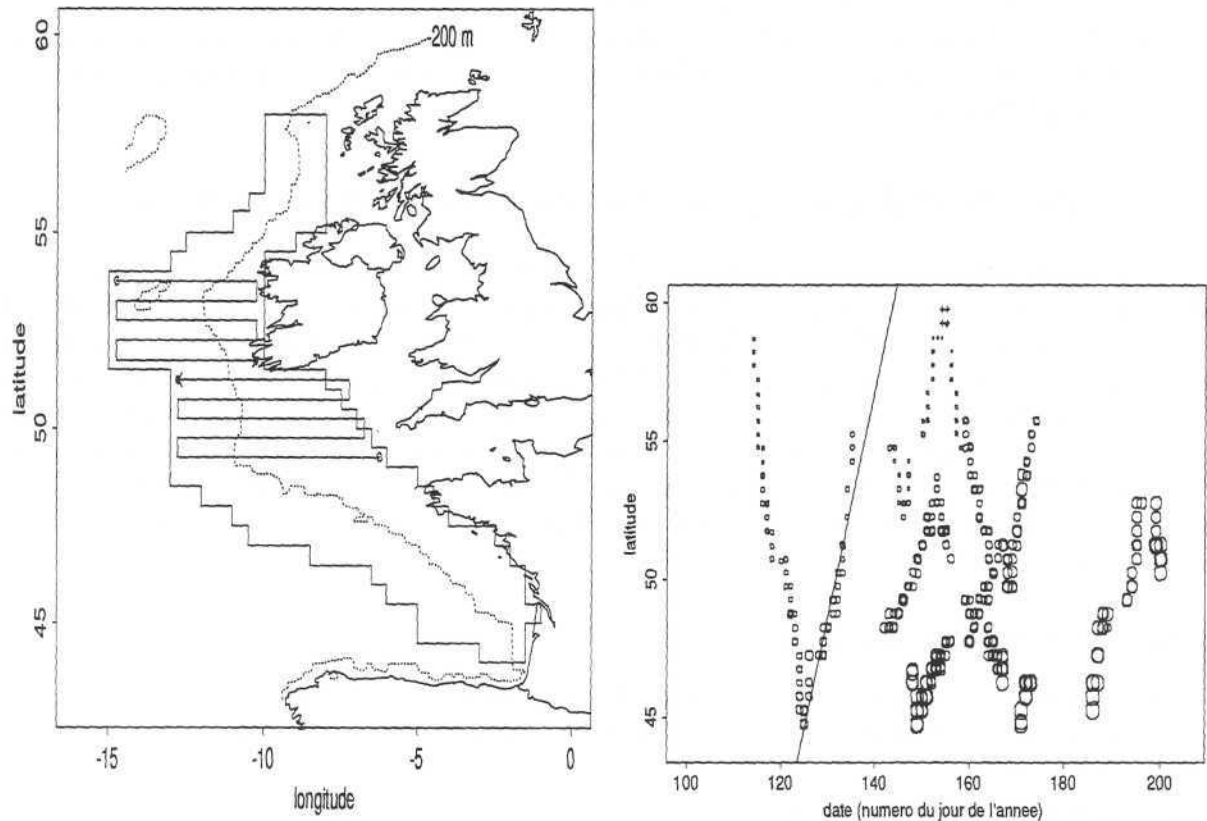


FIG. 1.1 - Plan d'échantillonnage. Représentation proportionnelle des températures de surface dans le plan latitude (mn) \times temps (numéro du jour de l'année). Direction moyenne de l'échantillonnage dans l'espace latitude \times temps.

température de surface en fonction de la latitude à deux moments différents t_1 et t_2 . La dérive en t_2 est identique en forme mais supérieure de quelques degrés à la dérive en t_1 .

Interférences entre l'échantillonnage et la dérive de la température de surface

Rappelons que le variogramme expérimental à la distance h est donné par la différence quadratique moyenne entre les températures distantes de h (à quelques tolérances près en distance et en direction). Là encore, le recours à une moyenne spatiale implique de délimiter, préalablement aux calculs expérimentaux, le domaine de calcul. Dans le cas de la température de surface, il n'y a pas trop de problèmes. Il n'y a pas de valeur nulle et l'extension du phénomène n'est limitée que par les continents. Il apparaît donc assez logique d'étendre la zone de calcul à l'ensemble de la zone échantillonnée.

Faire une analyse structurale à 2 dimensions reviendrait à négliger la dimension temporelle du phénomène physique lui-même mais également celle de l'échantillonnage. Il faut donc songer à inclure le temps comme troisième dimension. Mais, dans le cas d'une variable présentant une dérive spatio-temporelle nette comme la température de surface, et d'un échantillonnage réparti dans le temps, cela nécessite quelques précautions que la figure 1.2 va nous permettre de comprendre.

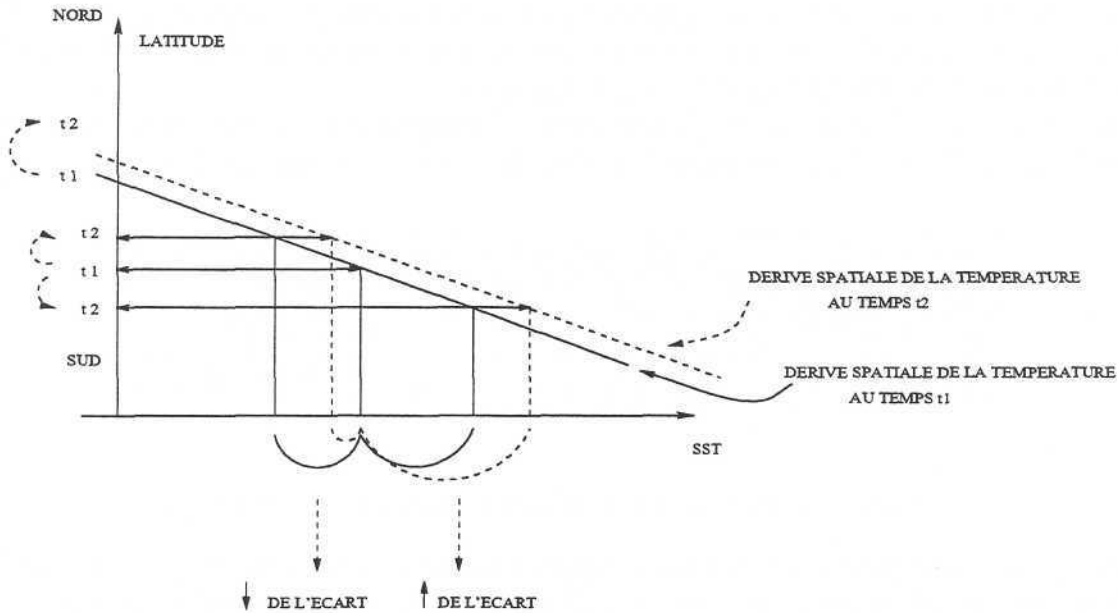


FIG. 1.2 - Schématisation des difficultés pratiques de l'analyse structurale de la température de surface. Accumulation ou compensation des effets dus à la dérive spatio-temporelle de la température de surface d'une part et au sens de progression de l'échantillonnage d'autre part.

Si l'échantillonnage se fait vers le Sud, alors l'augmentation de température à laquelle on devrait s'attendre se trouve amplifiée par le réchauffement qui s'est produit pendant le laps de temps nécessaire pour aller du point de départ au point d'arrivée (Fig. 1.2). Au contraire, si l'échantillonnage se fait vers le Nord, alors l'écart de température attendu sera réduit. En somme, le sens de parcours de l'échantillonnage (du Sud au Nord ou du Nord au Sud) compense ou augmente la dérive spatio-temporelle de la température de surface.

Il convient par conséquent de choisir avec précaution les directions de calcul pour l'analyse structurale.

Coordonnées et unités

Pour ces campagnes scientifiques, la grille d'échantillonnage est une grille régulière de 0.5 degré par 0.5 degré. La distance entre deux nœuds de même latitude dépend donc de la latitude. A 45°00N (Bordeaux), un degré de longitude correspond à 42 milles nautiques environ, alors qu'à 60°00N (Nord de l'Ecosse) il correspond à 30 milles nautiques seulement. Compte-tenu de l'extension en latitude de la zone d'étude, il est donc préférable, dès lors que les outils utilisés font intervenir la distance entre les valeurs, de transformer les coordonnées pour les exprimer en milles nautiques (mn). Une transformation incorrecte, mais simple, consiste à multiplier les longitudes par le cosinus des latitudes correspondantes et à multiplier le tout par 60 pour obtenir des milles nautiques.

Dans une analyse structurale à 3 dimensions, où le temps constituerait la troisième dimension, les distances se trouvent par voie de conséquence calculées à 3 dimensions. En toute rigueur, on travaille avec un vecteur distance. Ainsi, les directions de calcul pourraient être caractérisées par un écart de distance (à 2D) et un écart de temps (par exemple : 30 mn vers le Nord et 5 jours). Par

commodité pour la représentation graphique des résultats par exemple, on peut également chercher à exprimer l'ensemble des axes avec une unité commune de distance. Se pose alors le problème du choix d'une unité pour le troisième axe, l'axe temporel.

Pour orienter le choix des unités, j'ai comparé les variogrammes d'ordre 1 de la température de surface dans les 3 directions principales Nord-Sud, Est-Ouest et temporelle (Tab. 1.1).

direction	pas*	nombre de pas	tolérances sur les 2 autres axes
Nord-Sud	30 mn	15	25 mn et 3 j
Est-Ouest	21 mn	10	22 mn et 3 j
temps	15 j	8	24 mn et 31 mn

(*) mn : mille nautique j : jour

TAB. 1.1 - Paramètres de calcul des variogrammes d'ordre 1

Le variogramme d'ordre 1 correspond expérimentalement à la valeur absolue de la différence moyenne des températures en fonction de la distance; le variogramme classique étant de ce point de vue un variogramme d'ordre 2. Il permet de montrer que, en moyenne, la différence entre deux points distants de 100 mn est de 0.4°C dans la direction Nord-Sud comme dans la direction Est-Ouest (Fig. 1.3). On notera que le variogramme d'ordre 1 est identique dans ces 2 directions pour toutes les distances inférieures à 125 milles nautiques.

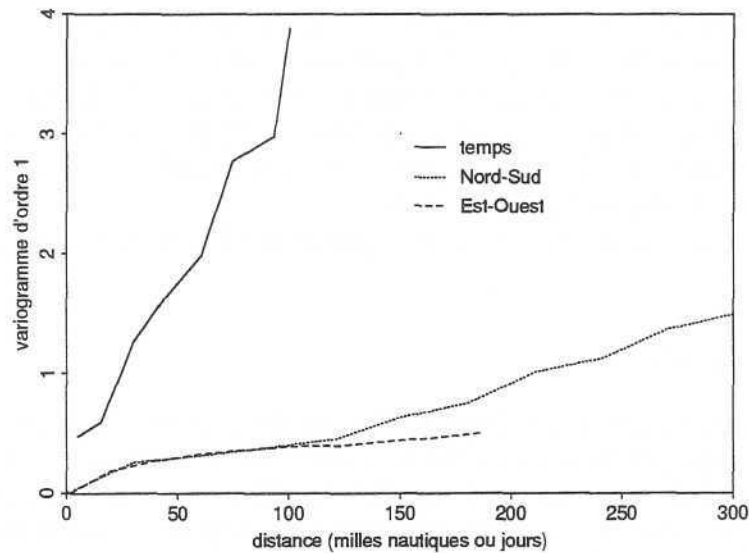


FIG. 1.3 - Variogramme expérimentaux d'ordre 1 pour la température de surface. Trois directions de calcul : Nord-Sud, Est-Ouest, temps.

Par contre, dans la "direction du temps", entre deux températures observées à peu près au même endroit, c'est-à-dire dans un carré de 24 sur 31 milles nautiques compte-tenu des tolérances de calcul, mais à des instants différents, l'écart moyen est sensiblement plus important. Il est en moyenne de 2°C pour un décalage de 2 mois, de 1.25°C pour un décalage de 1 mois et de 0.4°C

à l'origine, sensé représenter la différence moyenne qui existe entre des températures observées en un même point à moins d'un jour d'intervalle. Cet effet de pépité est probablement essentiellement dû aux nécessaires tolérances de calcul qu'on s'est imposées (entre 24 et 31 mn). Il peut donc être raisonnablement diminué de 0.25°C par exemple, qui est la valeur du variogramme d'ordre 1 pour une distance de 25 milles nautiques. Ainsi, on peut proposer l'équivalence suivante :

100 milles nautiques correspondent à 16 jours ce qui donne 6.25 mn/j

200 milles nautiques correspondent à 25 jours ce qui donne 8 mn/j

J'ai donc pris une conversion moyenne de 7 milles nautiques pour 1 jour afin de procéder à l'analyse structurale. On notera au passage que le variogramme d'ordre 1 nous a fourni les ordres de grandeurs des dérives dans les différentes directions spatiales et temporelle.

1.2.2 Résultats

On cherche ici à caractériser la structure spatiale de la température de surface entre avril et juin de 1989.

Directions de calcul

Comme il a déjà été précisé plus haut, j'ai considéré que les échantillons appartenant à un même transect étaient simultanés. Le variogramme dans la direction Est-Ouest fournit donc la structure purement spatiale de la température de surface en longitude.

Par contre dans la direction Nord-Sud, pour tenir compte de l'interférence entre l'échantillonnage et la dérive de la température de surface, j'ai distingué les deux cas suivants:

- une direction "Sud-Nord" où l'échantillonnage se fait en progressant vers le Nord. Les paires de points retenues pour le calcul sont telles que la température de surface la plus au Nord a été observée après l'échantillon le plus au Sud
- une direction "Nord-Sud" inverse de la première.

Dans le repère à deux dimensions latitude (mn) x temps converti en milles nautiques, ces deux directions correspondent aux pentes moyennes de l'échantillonnage, à savoir +/- 16 degrés, respectivement dans l'un et l'autre cas.

Interprétations des résultats

Le comportement du variogramme à l'origine (Fig. 1.4) est symptomatique d'une variable dite régulière, c'est-à-dire ayant des fluctuations lentes et homogènes dans l'espace. En effet, quelle que soit la direction, il n'y a pas de discontinuité à l'origine et l'augmentation du variogramme se fait très lentement jusqu'à des distances assez importantes (100 milles nautiques).

Dans la direction Est-Ouest, le variogramme se stabilise autour de 0.5 pour des distances comprises entre 100 et 200 milles nautiques. Ce serait le signe de la stationnarité de la fonction aléatoire dans cette direction si le variogramme n'augmentait pas au delà. Mais, l'interprétation de cette partie du variogramme est délicate. Elle concerne des distances qui sont de moins en moins bien représentées au sein de l'échantillonnage (demi-champ). Cependant, on observe une augmentation des températures en allant vers la côte le long des transects compatible avec une modélisation non stationnaire.

Le variogramme dans la direction "Sud-Nord" est superposé au précédent pour des distances inférieures à 200 milles nautiques. Il lui est inférieur au delà mais augmente régulièrement.

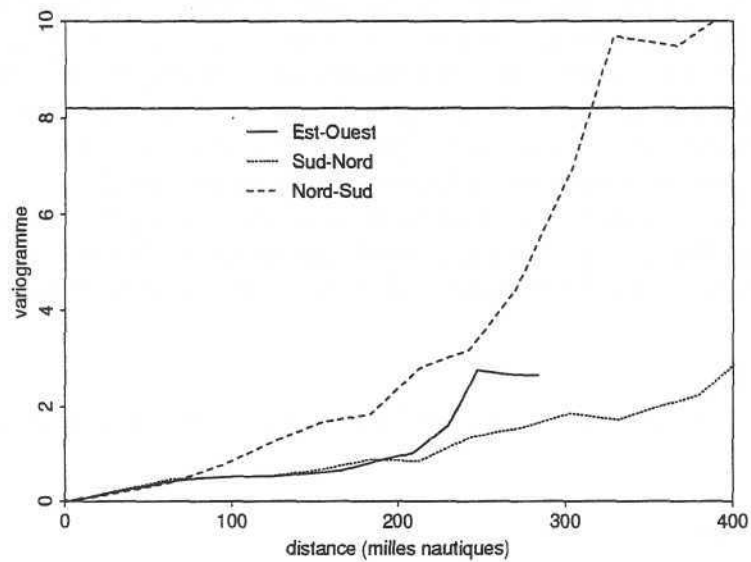


FIG. 1.4 - Variogramme expérimental pour la température de surface. Trois directions de calcul : Est-Ouest, Nord-Sud, Sud-Nord.

En revanche, dans la direction “Nord-Sud”, le variogramme expérimental est croissant. Le fait que le variogramme ne se stabilise pas est le signe de la non stationnarité à laquelle on s’attendait compte-tenu de la dérive connue de la température de surface dans cette direction. Par contre, on constate que les variogrammes Nord-Sud et Sud-Nord ne sont pas identiques. Conformément à nos attentes (Cf. §I.1.2.1), il semble qu’en allant du Sud vers le Nord, la dérive spatiale est quasiment contre-balançée par le décalage dans le temps entre les échantillons. Du Nord vers le Sud au contraire, on pourrait penser que la dérive, nettement visible, est accentuée.

L’aspect didactique de cet exemple concerne donc les deux points suivants :

- l’analyse structurale est confrontée à la difficulté de prendre en compte le temps dans la mesure où la variable présente une dérive spatio-temporelle susceptible d’interférer avec le sens de progression de l’échantillonnage.
- l’utilisation du variogramme conduit à des résultats satisfaisants (interprétables) dans l’analyse de la température de surface.

Dans l’exemple suivant, nous allons voir qu’il n’en est pas ainsi pour l’analyse variographique des densités de merlu.

1.3 Analyse variographique de densités de merlu

1.3.1 Présentation des données

L’analyse porte sur des concentrations de merlu (*Merluccius merluccius*) issues d’une campagne de chalutage de fond organisée, en 1990, par l’IFREMER dans le Golfe de Gascogne. L’échantillonnage est stratifié aléatoire et comprend 137 échantillons (Fig. 1.5). Les densités de merlu d’âge 0

sont exprimées en nombre d'individus par heure de chalutage (ind/h). Les grandes concentrations sont réparties dans la partie centrale du plateau continental. L'ensemble de la population de la région semble avoir été échantillonné puisqu'on a observé de nombreuses valeurs nulles en périphérie (les densités nulles sont représentées par une croix).

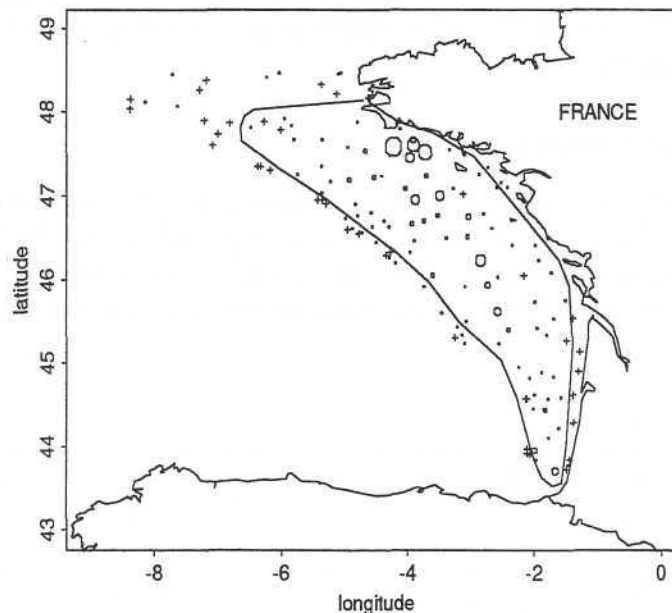


FIG. 1.5 - Représentation proportionnelle des densités de merlu d'âge 0. Golfe de Gascogne 1990. Densités nulles représentées par une croix. Le champ est délimité par le trait plein.

La moyenne des 137 échantillons vaut 130 ind/h, leur médiane 10 ind/h et leur écart-type 284 ind/h; ce qui donne un coefficient de variation élevé de 2.2. Un tiers des densités sont nulles, et moins de 10% des densités classées par ordre décroissant engendrent 50% du total (forte sélectivité). Ceci témoigne d'une distribution très dissymétrique. L'irrégularité spatiale apparaît importante puisque de fortes densités (maximum = 1902 ind/h) sont entourées de densités très faibles voire nulles. J'ai donc cherché à modéliser la variabilité spatiale de la densité de merlu par une analyse variographique.

1.3.2 Des variogrammes expérimentaux "décevants"

Choix des paramètres de calcul

Comme je l'ai déjà précisé au cours de l'exemple précédent, il faut, pour calculer des variogrammes expérimentaux, délimiter un domaine. La démarche est ici plus délicate que pour la température de surface. On voit en effet qu'en bordure Ouest de l'échantillonnage les densités sont toujours réduites mais constituées d'un mélange de densités nulles et de concentrations très faibles. De plus, on trouve quelques valeurs nulles à l'intérieur de la zone échantillonnée. Ainsi, se limiter au champ des valeurs strictement positives conduirait à un domaine dont la forme serait très découpée. Par ailleurs, se doter d'un seuil pour définir le domaine conduirait probablement à supprimer trop d'échantillons. J'ai donc opté pour une délimitation graphique par laquelle les échantillons

des bordures Ouest et Nord de la zone d'échantillonnage ont été retirés. Le champ résultant a grossièrement une forme d'ellipse et comprend 89 échantillons.

Compte-tenu des directions principales de la régionalisation qui ressortent de l'examen de la carte proportionnelle (Fig. 1.5), les résultats structuraux sont présentés pour les quatre directions suivantes exprimées en degrés trigonométriques à partir de la direction Est-Ouest : 27° , 72° , 117° et 162° . Les pas de calcul sont respectivement : 15, 20, 30 et 20 milles nautiques. La tolérance sur les distances est de 50% et de 22.5° pour la direction. Les coordonnées sont transformées en milles nautiques pour l'analyse structurale.

Variogrammes expérimentaux

Avant toute analyse plus fine des résultats, on voit clairement que les variogrammes expérimentaux sont très instables (Fig. 1.6). La différence avec la précédente analyse variographique est très nette. En particulier, quelle que soit la direction de calcul, la valeur au premier pas de calcul est au moins supérieure à la moitié de la variance expérimentale. La discontinuité à l'origine est donc très importante.

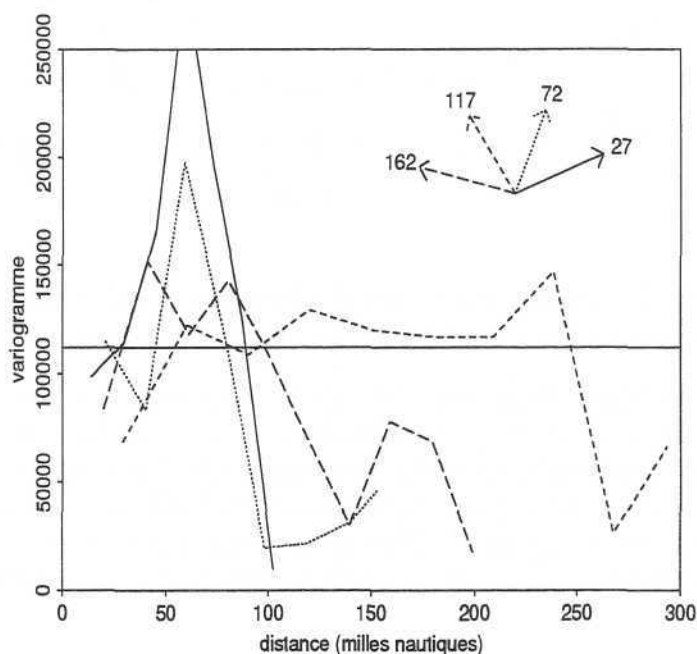


FIG. 1.6 - Merlu d'âge 0. Variogrammes expérimentaux dans quatre directions : 27° , 72° , 117° et 162° . La variance expérimentale des échantillons est matérialisée par le trait horizontal.

Toutefois, on distingue, comme le suggère la représentation proportionnelle des densités, une plus forte continuité dans la direction Nord-Nord-Ouest. Conformément aux dimensions du champ, les structures expérimentales sont disponibles pour des gammes de distances différentes selon la direction :

- dans les directions 27° et 72° , le variogramme commence au niveau de la variance, augmente très fortement jusqu'à 60 milles nautiques et décroît ensuite jusqu'à devenir nul.
- dans la direction 162° , le variogramme a la même forme en cloche que dans les deux

premières directions. Le maximum, légèrement au-dessus de la variance, correspond à une distance de 80 milles nautiques.

- dans la direction 117° le variogramme augmente de 30 à 60 milles nautiques et se stabilise ensuite autour de la variance. Il chute rapidement au-delà de 250 milles nautiques. Dans cette direction, la structure pourrait s'interpréter comme une discontinuité à l'origine de l'ordre de 2500 (le quart de la variance) combinée à une structure de portée 60 milles nautiques indiquant la portée des corrélations dans cette direction.

Les valeurs choisies pour les pas de calcul ne sont pas en cause dans la mauvaise qualité des résultats. En effet, plusieurs tentatives ont été réalisées avec des pas de valeurs différentes mais elles conduisent au même type de résultats.

Force est donc de constater que l'utilisation du variogramme expérimental serait ici très délicate voire impossible.

1.3.3 Pourquoi de tels variogrammes?

Instabilité numérique

Une première explication des fluctuations des résultats expérimentaux est de nature calculatoire. En effet, dans la pratique, on estime le variogramme par la moyenne **quadratique** de la différence des densités. Or, nous avons vu que les densités de merlu sont très contrastées à la fois en valeur mais également dans leur agencement spatial. Dès lors, les écarts de densités qui ont naturellement tendance à être importants se trouvent très fortement amplifiés par le passage au carré; et ce particulièrement pour les paires de points comprenant l'une des rares concentrations élevées.

Une façon de limiter cette instabilité consiste à transformer les données. Prendre simplement le logarithme de la densité pose un certain nombre de problèmes dont celui du traitement des valeurs nulles. Dans le cas d'une concentration, on lui préfère donc la "log-translation" qui, à z , fait correspondre la variable $\ln(1 + \frac{z}{m})$ où la constante m peut représenter la moyenne de z . Les avantages de cette transformation sont les suivants :

- les valeurs nulles restent inchangées
- l'écart entre les valeurs faibles est beaucoup moins amplifié que par une transformation logarithmique simple
- l'importance des valeurs très élevées relativement aux faibles est réduite.

Il faut tout de même garder à l'esprit que les résultats obtenus après transformation ne sont pas directement transposables à la variable brute. Cependant, on est fondé à juger l'anisotropie de la fonction aléatoire au vu des variogrammes de la variable transformée. Dans certaines conditions, il est toutefois possible, connaissant le variogramme de la variable log-translatée, d'estimer le variogramme de la variable de départ (Guiblin *et al.*, 1995).

Richesse du cœur et effet de bordure

La position des fortes densités dans le champ est également importante. Dans le cas du merlu du Golfe de Gascogne, les zones riches se trouvent globalement au milieu du champ, le long de son grand axe, avec des concentrations importantes en bordure Nord. C'est probablement cette localisation particulière de la partie dense de la population qui génère expérimentalement des variogrammes décroissants.

En effet, imaginons une distribution à 1D en forme de pyramide (Fig. 1.7a) qui schématiserait une section Nord-Nord-Est au travers de la régionalisation de merlu (Fig. 1.6). Le champ correspondant est alors limité aux frontières de la pyramide. Par conséquent, pour des distances supérieures au demi-champ, le cœur de la distribution ne participe plus aux calculs, induisant ainsi une diminution du variogramme (Fig. 1.7b).

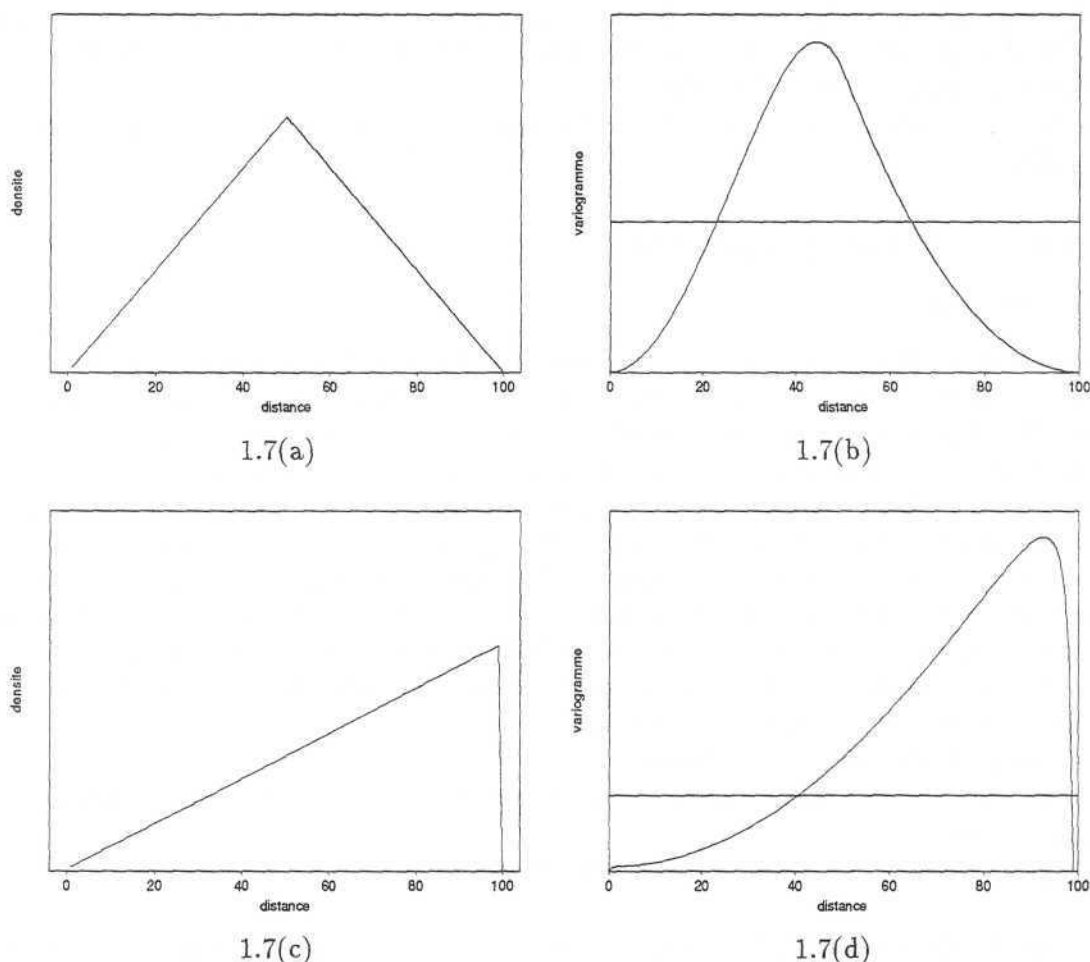


FIG. 1.7 - Exemples de distributions à une dimension et variogrammes correspondants. Le champ correspond à l'intervalle $[1,100]$.

Ce type d'explication est également valable si le cœur de la distribution se trouve en bordure du champ (Fig. 1.7c). Du fait de la variation brutale des densités en bordure Est du champ (entre les points d'abscisses 99 et 100), le comportement du variogramme à l'origine est linéaire (Fig. 1.7d).

En réalité, les problèmes engendrés par la position des valeurs fortes au sein du champ, sont d'une nature plus fondamentale que ne le laisse entendre les inconvénients numériques dont je viens de parler. C'est en fait la pertinence du recours aux fonctions aléatoires qui est probablement en cause (§I.3.3).

Pertinence du modèle

L'expérience acquise en géostatistique a conduit les praticiens à procéder d'emblée à une analyse variographique, sans précisément s'inquiéter des conditions requises par cette méthodologie. Dans cette démarche, c'est la cohérence des résultats qui permet de valider, a posteriori, le bien-fondé de l'emploi de la géostatistique intrinsèque. A contrario, la singularité des résultats structuraux invite à reconsidérer les choix de modélisation. Car, utiliser d'emblée le variogramme, sous-entend en fait qu'on a décidé de faire appel :

- aux méthodes probabilistes
- et à la théorie des fonctions aléatoires.

Etait-ce justifié? Le formalisme des fonctions aléatoires convient-il aux densités de merlu? Et en particulier, les hypothèses de stationnarité (de la variable ou de ses incréments) sous-jacentes à ces modèles.

Des différentes raisons qui ont été évoquées (instabilité numérique, position du cœur, modèle), c'est l'inadéquation du modèle aux caractéristiques du phénomène étudié qui est la plus profonde et qui a retenu mon attention. Nous verrons plus loin que la géostatistique intrinsèque nécessite des hypothèses qui peuvent se révéler trop lourdes par rapport à la réalité des données halieutiques alors que l'alternative transitive proposée en géostatistique permet de les éviter.

1.4 Analyse structurale d'un courant marin

Bien qu'il ne constitue pas précisément une aide à la réflexion sur le choix d'un modèle, l'exemple qui suit permet cependant d'aborder un autre type de problème rencontré en géostatistique. Il s'agit de savoir comment respecter, dans une analyse structurale, le caractère vectoriel du phénomène étudié. Qu'il s'agisse de l'étude d'un courant marin, ou encore du lien entre la distribution spatiale d'une espèce et d'un gradient (de température par exemple), il convient de prendre un certain nombre de précautions lors de l'analyse structurale, pour permettre, en particulier, l'interprétation physique des résultats.

On trouvera dans l'annexe 1 les éléments pratiques et théoriques de l'analyse structurale géostatistique d'un vecteur. Pour l'essentiel, il apparaît nécessaire de procéder pour chaque direction aux analyses des parties du vecteur qui sont colinéaires et orthogonales à cette direction.

L'exemple qui suit concerne des données qui m'ont été transmises par W.B. Turrell du Marine Laboratory d'Aberdeen en Ecosse, que je remercie chaleureusement.

1.4.1 Présentation des données

Rappels sur les ADCP

Un ADCP (Acoustic Doppler Current Profiler) est un outil de mesure acoustique. Son procédé de fonctionnement est donc identique à toute mesure acoustique. Un engin trainant émet des ondes acoustiques (4 trains d'ondes dans le cas des ADCP) qui sont réfléchies par l'ensemble des particules en suspension (plancton, sédiments, détritiques), recueillies par un récepteur et analysées ensuite par ordinateur. Cette analyse consiste en un certain nombre de filtres appropriés qui permettent de débarrasser le signal de l'ensemble des effets indésirés.

En supposant que les particules réfléchissantes vont à la même vitesse que la masse d'eau qui les contient, on dispose ainsi d'une mesure de courant par voix acoustique. Après traitement, les quatre

ondes incidentes fournissent la vitesse de l'eau par rapport au bateau. Dans les zones peu profondes (inférieures à 350 mètres), la réponse acoustique du fond peut être utilisée pour déterminer sans trop d'erreur la vitesse réelle du bateau. Connaissant par ailleurs le cap, l'ordinateur de bord transforme alors les 4 signaux acoustiques en une vitesse réelle le long du navire, perpendiculairement au navire et verticalement.

Les données

Les mesures d'ADCP sont réparties le long d'un transect de 21.7 km de long croisant le bord du plateau continental au Nord-Ouest des Shetland (Fig. 1.8). Les coordonnées des points de départ et d'arrivée sont : 61°00'N-1°30'W et 61°07'N-1°50'W. Au total, 18 passages (réplicats) ont été réalisés en 25 heures, du 25 septembre 1992 au lendemain, le 26 septembre 1992. On n'utilisera ici que le premier passage.

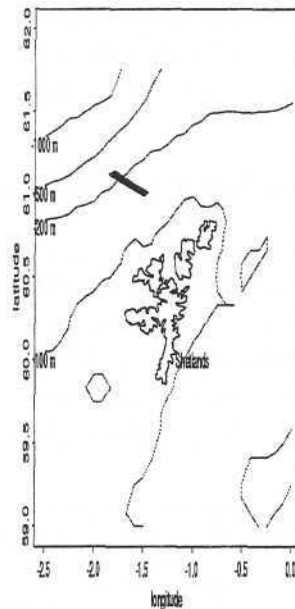


FIG. 1.8 - Localisation de la section échantillonnée par ADCP

Les réponses acoustiques sont moyennées selon une grille régulière. Le pas est de 10 minutes le long du transect de sorte qu'on dispose de 40 mesures de vitesse dans cette direction. Il est de 4 mètres en vertical. En chacun des points de cette section, le courant, exprimé en centimètre par seconde (cm/s) est décomposé selon une direction perpendiculaire au transect (V_{\perp}) et parallèle au transect ($V_{//}$). La composante verticale du courant n'est pas fournie. On dispose donc d'un courant horizontal à 2 dimensions connu sur une section à deux dimensions également, mais d'axes différents "transect x profondeur". Dans la représentation graphique que je donne des courants, je mélange au sein d'un même graphe ces deux repères. Graphiquement, la composante horizontale du courant ne correspond pas au mouvement selon la profondeur mais à la composante du courant perpendiculairement au transect, c'est-à-dire le long des isobathes. Compte-tenu de l'information disponible et des contraintes liées à l'analyse structurale, cette dernière ne pourra pas être complète.

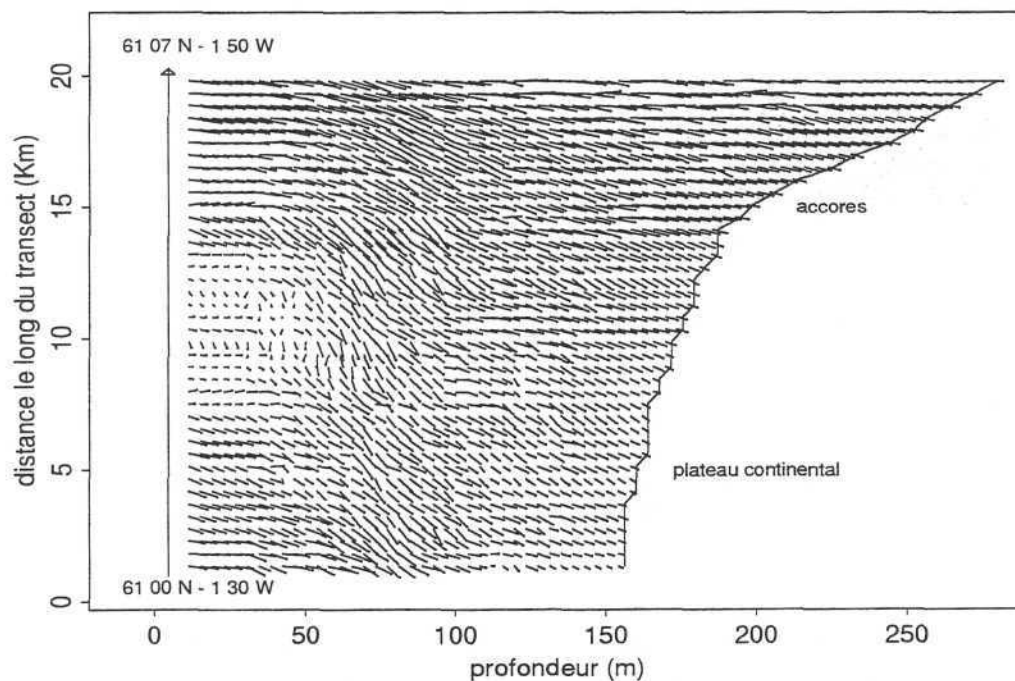


FIG. 1.9 - Présentation des mesures ADCP. Premier passage. En *abscisse*, la profondeur est donnée en mètre. En *ordonnée*, les distances sont en kilomètres le long du transect. En chaque point le *courant* horizontal est représenté dans un autre système d'axes avec la convention suivante: une flèche verticale donne le mouvement le long du transect (vers le haut = vers le Nord-Ouest); une flèche horizontale indique le mouvement le long des isobathes, c'est-à-dire perpendiculairement au transect (vers la droite = vers le Nord-Est).

La direction du transect est en effet la seule direction pour laquelle on dispose d'une partie colinéaire et d'une partie orthogonale.

Dans cette étude, on considère les mouvements de masses d'eau non corrigés par la marée. Des modèles mathématiques de marée permettent de connaître le coefficient de marée en tous lieux de cette zone et tous temps de l'année. Il serait ainsi possible de filtrer les composantes périodiques du mouvement total enregistré et de ne conserver que le courant résiduel seul responsable d'un déplacement effectif. Une autre solution plus simple consisterait à moyenner les mesures sur un pas de temps plus grand que la période de la marée. Cependant l'analyse qui suit n'a vocation qu'à **illustrer les concepts exposés en annexe**. En conséquence, aucun filtrage de la marée n'a été fait.

Résultats préliminaires

Il apparaît (Fig. 1.9) que le courant observé lors du premier passage est essentiellement orienté vers la droite sur le graphe, c'est-à-dire vers le Nord-Est, en surface et à proximité du fond, avec une intensité d'autant plus faible qu'on se situe au milieu du transect. Par contre, le courant mesuré entre 50 et 100 mètres de fond est plus fort et principalement orienté le long du transect vers le

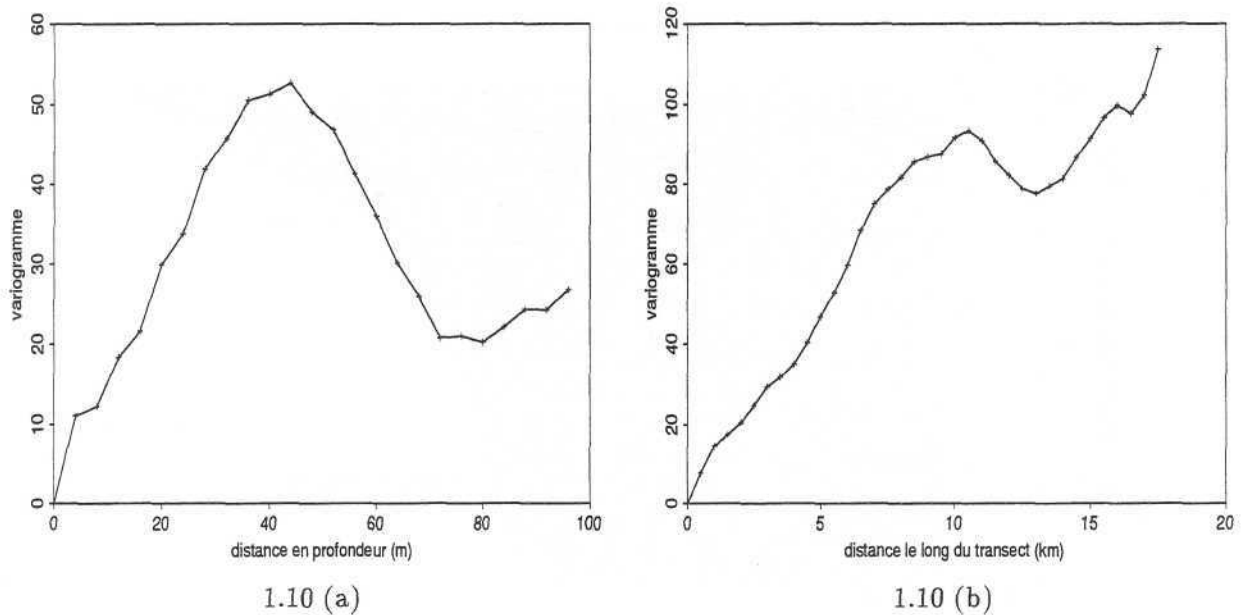


FIG. 1.10 - Description des caractéristiques spatiales des composantes par les variogrammes expérimentaux. (a) Composante $V_{//}$ le long de la verticale. (b) Composante V_{\perp} le long du transect.

Sud.

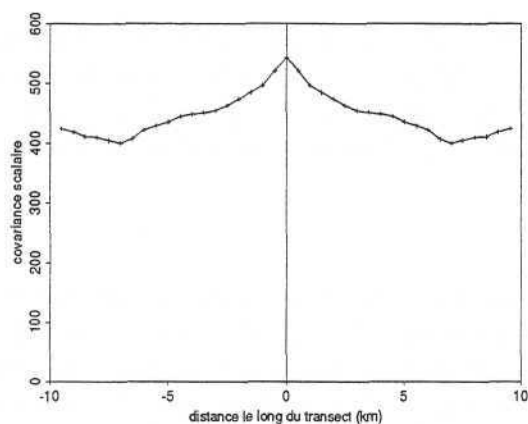
La composante $V_{//}$ augmente donc régulièrement avec la profondeur jusque vers les couches intermédiaires (50 - 100 mètres) et diminue ensuite. A cet égard, une approche non stationnaire serait probablement recommandée (fonction aléatoire intrinsèque d'ordre 2 au moins), mais l'utilisation du variogramme permet de synthétiser cette tendance, ainsi que la continuité spatiale de cette composante. Le variogramme expérimental de la composante $V_{//}$ en direction de la profondeur (Fig. 1.10a) présente un pic aux alentours de 40 mètres qui est relié, comme nous l'avons déjà expliqué (Cf. §I.1.3.3), à l'évolution en "cloche" de la variable dans cette direction. La position de ce pic indique que la tranche d'eau où la composante $V_{//}$ évolue en "cloche" est donc en moyenne épaisse de 80 mètres. L'impression visuelle de régularité spatiale qui se dégage de la carte (Fig 1.9) se traduit par une absence d'effet de pépite (Fig. 1.10). La pente assez importante du variogramme expérimental à l'origine indique cependant une continuité spatiale assez moyenne.

A la verticale du bord du plateau continental (accres), et dans une moindre mesure également au début du transect vers les faibles profondeurs, la composante orthogonale (V_{\perp}) est plus intense qu'au milieu du transect. Cette tendance à l'augmentation vers le large est traduite par une croissance régulière du variogramme expérimental de la composante V_{\perp} dans la direction du transect (Fig. 1.10b). La diminution du variogramme aux alentours de 13 km est à mettre en correspondance avec la diminution de V_{\perp} au centre du transect.

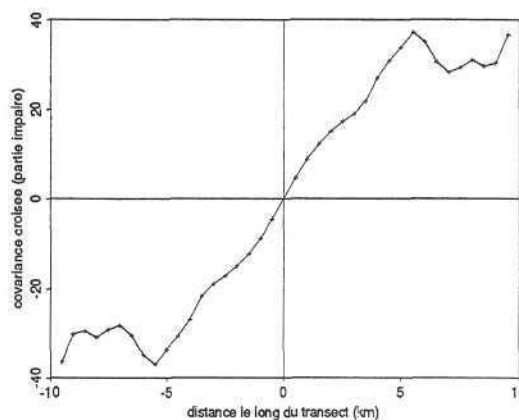
Jusque là, je me suis servi du variogramme pour décrire statistiquement les composantes du vecteur. Les résultats s'interprètent facilement et, à ce stade de l'analyse, rien ne semble contredire le choix de recourir aux techniques probabilistes de la géostatistique intrinsèque, même si les non stationnarités non pas été précisément prises en charge. La continuité spatiale du phénomène semble donc compatible avec notre modèle. L'analyse structurale du courant proprement dit n'en est pour autant pas effectuée.

1.4.2 Analyse structurale

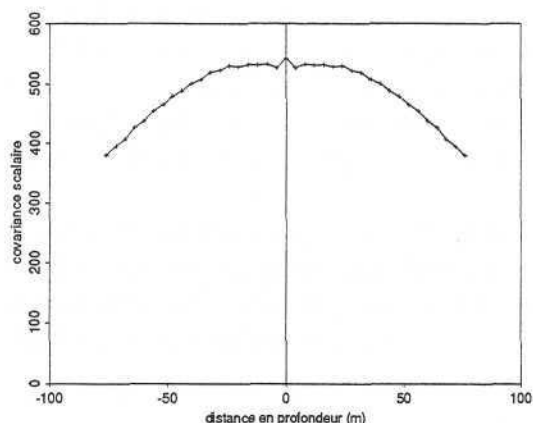
Compte-tenu des données disponibles et des outils pertinents pour l'analyse structurale d'une variable vectorielle (Cf. Annexe 1), on a été amené à s'intéresser à la covariance scalaire et au demi produit vectoriel dans la direction du transect et dans la direction verticale (profondeur). Les structures expérimentales ont été calculées jusqu'à des dimensions égales à la moitié du transect (10 km) et la moitié de la colonne d'eau (80 m).



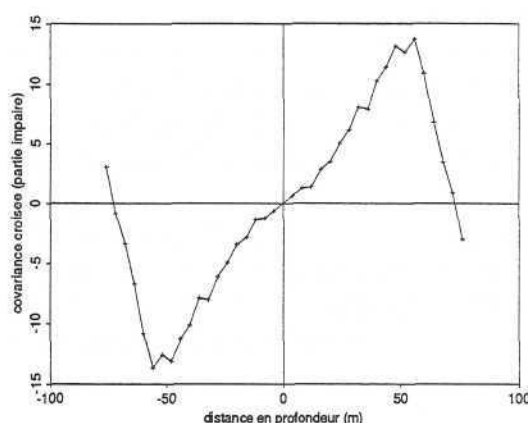
1.11 (a)



1.11 (b)



1.11 (c)



1.11 (d)

FIG. 1.11 - Analyse structurale du courant mesuré par ADCP. Dans la direction du transect: (a) covariance scalaire et (b) partie impaire de la covariance croisée. Dans le sens des profondeurs: (c) covariance scalaire et (d) partie impaire de la covariance croisée.

On remarquera tout d'abord (Fig. 1.11) que le produit vectoriel est, en moyenne, d'un ordre de grandeur inférieur, au produit scalaire ($10 \text{ cm}^2/\text{s}^2$ contre $500 \text{ cm}^2/\text{s}^2$).

A leurs échelles respectives, c'est-à-dire 10 mètres en profondeur et 1 kilomètre horizontalement, les covariances scalaires ne présentent pas le même niveau de continuité à l'origine (Fig. 1.11 a, c). La covariance scalaire est en effet quasiment constante pour des distances inférieures à 25 mètres vers la profondeur alors qu'elle décroît dans l'autre direction.

Pour qu'un produit scalaire soit constant il faut, soit que le courant ne change ni d'orientation ni de force, soit que des rotations compensent une augmentation de l'intensité du courant. Cette compensation, qui s'exprime analytiquement comme l'inverse du cosinus de la rotation, doit donc être d'autant plus marquée que la rotation est importante. Dans notre cas, la distance 25 mètres correspond plutôt aux dimensions des tranches d'eau de surface et de fond où le courant est à peu près constant le long de la colonne d'eau. Cette homogénéité en profondeur et en surface est d'origine technique puisque les extrémités de chaque colonne sont extrapolées à partir des dernières mesures correctes.

Par contre, en allant de la surface vers 60 - 80 mètres de fond, le courant a tendance à tourner et à s'intensifier. La chute de la covariance scalaire pour cette gamme de distances, provient du fait que l'intensification du courant n'est pas suffisante pour compenser l'importance de la rotation. On ne note pas de diminution équivalente pour des distances de l'ordre de la dizaine de kilomètres le long du transect.

La partie impaire des covariances croisées (produit vectoriel) traduit et mesure parfaitement les changements d'orientation du courant en fonction de la profondeur (Fig. 1.11d). En effet, on observe un maximum très net aux alentours de 50 mètres lié aux observations déjà faites sur le courant dans cette direction. Compte tenu des conventions habituelles, un produit vectoriel positif correspond à une rotation dans le sens trigonométrique (inverse des aiguilles d'une montre). Dans la mesure où les distances sont comptées positivement en allant du fond vers la surface, on doit interpréter les produits vectoriels moyens comme suit : en moyenne, le courant tourne progressivement dans le sens des aiguilles d'une montre lorsqu'on se déplace de la surface vers le fond (phénomène connu sous le nom de "courant de Ekman" - Crépon, 1993). En moyenne, le courant tourne de 90° en l'espace de 50 mètres (maximum du produit vectoriel). De façon identique, la partie impaire de la covariance croisée des composantes le long du transect (Fig. 1.11b) présente un maximum entre 5 et 6 km.

La présentation de cet exemple aura permis de montrer que la géostatistique, moyennant quelques précautions d'emploi, constitue un contexte méthodologique approprié au traitement d'une variable vectorielle. Les outils structuraux sont opératoires. Les difficultés concernent ici l'interprétation des résultats plus que la compatibilité du cadre méthodologique avec la réalité du phénomène.

1.5 Synthèse : nature des difficultés

Au cours de ce chapitre introductif construit autour de quatre exemples significatifs, j'ai cherché à définir la nature des difficultés habituellement rencontrées en statistique. Je ne reviendrai pas ici sur les difficultés de nature techniques et calculatoires. Il me semble plus important de réfléchir aux causes probables de certains mauvais résultats.

Le premier danger en statistique appliquée est probablement l'habitude. En effet, nous avons vu que la modélisation statistique repose sur une succession de choix. Nous avons vu également que l'habitude conduit à ignorer un certain nombre des étapes préliminaires du choix d'un modèle. C'est en tous les cas par habitude que j'ai abordé d'emblée, comme d'autres praticiens d'ailleurs, l'analyse des densités de maquereaux ou de merlus par le calcul des variogrammes. En d'autres termes, on choisit souvent un outil sans s'assurer que les conditions de son utilisation sont, au

moins, partiellement réunies.

Quand les résultats expérimentaux sont satisfaisants, comme par exemple pour la température de surface, l'analyse peut suivre son cours : on procède à la modélisation de la structure spatiale qui permet, en fonction des attendus de l'étude, de faire une cartographie par krigeage, une simulation ou encore une estimation de la variance d'estimation. La qualité des résultats est donc une invitation à poursuivre et, *a posteriori*, à ne pas se soucier de la compatibilité du phénomène étudié aux hypothèses implicites de la géostatistique intrinsèque. En somme, "il était possible" de brûler ou d'ignorer les premières étapes de la modélisation.

Mais il arrive souvent que les structures expérimentales soient décevantes. J'ai présenté, avec l'analyse structurale des densités de merlu, une telle situation. Cette étude de cas a d'ailleurs réellement valeur d'exemple puisque de nombreux travaux réalisés en halieutique aboutissent à des variogrammes dont la modélisation s'avère très délicate. Ayant décidé péremptoirement de calculer des variogrammes (ou de façon équivalente un histogramme), on cherche alors à améliorer les résultats par des transformations de données. J'ai cité la log-translation plus adaptée que le passage simple au logarithme, mais de nombreuses autres transformations sont utilisées. Poussée à l'extrême, cette recherche de robustification du variogramme expérimental peut ressembler à une sorte d'acharnement qui, à force de manipulations de données, perd tout lien physique avec la réalité, ou en tous les cas, trahit la nature réelle du phénomène. Il se peut "simplement" que les données ne soient pas compatibles avec un modèle de fonction aléatoire stationnaire. Cette fois-ci, contrairement au cas précédent, les résultats nous invitent à revenir en arrière et à reformuler quelques uns des choix fondamentaux de modélisation.

Nous avons commencé à voir à l'aide des quatre exemples précédents que ces choix fondamentaux renvoient en particulier à des questions de **délimitation de champ**, de **stationnarité**, ou encore de **statut des modèles**. Il me semble assez approprié de regrouper ces choix en deux catégories, et de les analyser séparément à l'occasion des deux prochains chapitres :

– **Choix d'un concept pertinent et opératoire.**

La moyenne spatiale, comme d'ailleurs l'espérance mathématique qui figure dans les modèles probabilistes dont la moyenne constitue la version expérimentale, est probablement un exemple de concept peu opératoire dans les cas où le champ est difficile à délimiter. Le chapitre suivant développera les avantages et inconvénients à **utiliser des sommes préférentiellement à des moyennes**. On proposera aussi quelques réflexions générales sur la difficultés de définir, en écologie, des concepts pertinents.

– **Choix d'un modèle.**

Dans le dernier chapitre de cette première partie, on sera amené, après avoir défini les trois étapes du choix d'un modèle, à s'intéresser plus particulièrement à **l'alternative entre la géostatistique intrinsèque et la géostatistique transitive** en référence aux caractéristiques générales des densités de poissons.

Chapitre 2

Choix d'un concept opératoire et pertinent

Sommaire

Dans ce chapitre, on cherche à définir des concepts pertinents et opératoires applicables à des variables régionalisées de type concentrations de poissons.

*Le **paragraphe 1** examine l'utilisation d'une somme à la place d'une moyenne. Pour l'analyse d'une densité, l'abondance, et plus généralement les intégrales d'espace, évitent les problèmes de délimitation de champ.*

*Le **paragraphe 2** permet d'étendre ce principe à l'analyse statistique du lien entre deux variables dont les champs ont des tailles très différentes. Cette approche est présentée comme une alternative aux régressions expérimentales dont quelques limitations d'utilisations dans un contexte spatial sont d'abord exposées.*

*Dans le **paragraphe 3**, on part d'une situation artificielle où tout est connu. On voit alors qu'en écologie halieutique la difficulté est autant de définir un concept pertinent que de disposer de beaucoup d'information. C'est aussi l'occasion de définir la notion d'objectivité utilisée en géostatistique.*

Choice of a relevant and operating concept

Abstract

The task of this chapter is to look for operating and relevant concepts for the analysis of regionalized variables like fish concentrations.

*The **paragraph 1** propose to substitute the usual statistics which are average kind of tools by tools based on sums. Dealing with fish densities, the abundance, and more generally space integrals, avoid the delimitation of a field.*

*The **paragraph 2** extends the principle of using sums rather than averages, for the analysis of the link between two variables with different size of fields. This approach is interpreted as an alternative to the usual experimental regressions. As a digression, I also discuss some of the limitation of the use of such regressions in a spatial context.*

In paragraph 3, we first of all imagine that we know everything everywhere. Even from this utopic situation, the choice of a relevant concept for answering ecological questions appear to be very difficult. Thus I underline that, most of the time, the limitation of our analysis is more due to the unrelevancy of the concept used than to the quantity of available informations. I also remind what objectivity means in geostatistics.

Dans le dictionnaire (Robert, 1993) on trouve la définition suivante : un **concept** est une “représentation abstraite d'un objet ou d'un ensemble d'objets ayant un caractère commun”. A cet égard, une moyenne, une variance ou encore un histogramme sont des concepts. Ils permettent de représenter un ensemble de valeurs de façon abstraite. Leur caractère abstrait ne fait aucun doute puisqu'il n'existe pas dans la nature de tels objets. Personne n'a jamais observé un histogramme.

Notre travail de recherche manipule donc couramment des concepts. La question que je souhaite évoquer ici concerne leur pertinence et leur caractère opératoire. Par **opératoire**, j'entends défini sans ambiguïté à partir d'une connaissance exhaustive de la réalité. Cette acception du mot opératoire est suggérée par Matheron qui, dans Estimer et Choisir (1978), conduit la reconstruction opératoire des concepts géostatistiques à l'aide de grandeurs régionales; c'est-à-dire à l'aide d'entités qui seraient disponibles si nous connaissions complètement le phénomène. Comme exemple de concept peu opératoire, nous avons vu au chapitre précédent la moyenne arithmétique dont l'utilisation pose quelques problèmes de délimitation de champ dans le cas de densités de poissons très hétérogènes. Dans le premier paragraphe de ce chapitre, j'expose l'intérêt pratique et théorique qu'il peut y avoir à remplacer les concepts faisant appel à des moyennes par des critères définis par des sommes. Le cas bivariable sera traité séparément dans le deuxième paragraphe.

Cette démarche nous conduira plus loin, à la géostatistique transitive et aux statistiques par individu (deuxième partie de la thèse).

Le troisième paragraphe de ce chapitre invite à réfléchir sur la **pertinence** des outils que nous utilisons dans la pratique courante des statistiques. En somme, la question que j'aborde est la suivante : les concepts employés sont-ils pertinents en regard des questions posées? L'exemple du lien entre deux variables (densité et température de surface par exemple) servira de base à notre réflexion et nous conduira à nous demander ce que nous ferions si nous connaissions exhaustivement la réalité. Cette question, sans fondement apparent puisque nous ne connaissons jamais exhaustivement la réalité, est en fait riche d'enseignements qu'il m'a semblé intéressant de préciser.

Les arguments qui sont utilisés dans ce chapitre sont d'inspiration halieutique. Il est toutefois important de noter que l'ensemble des remarques restent valables pour tout autre domaine d'application, comme la pollution par exemple, où la variable d'intérêt présente les mêmes caractéristiques qu'une densité de poissons.

2.1 De la moyenne à l'abondance pour une densité

2.1.1 Le problème du champ, des 0 et des ϵ

Que ce soit en statistique “classique” ou en géostatistique, on ne peut pas calculer une moyenne sans avoir au préalable distingué parmi les données disponibles, celles qui participent au calcul de celles qui sont ignorées. Il existe des cas où les limites géographiques de la population sont

clairement identifiées. La délimitation du domaine ne pose alors pas de réel problème. Il existe cependant des situations où la zone d'occupation d'une espèce n'est pas connue. C'est le cas par exemple de la plupart des espèces pélagiques où la seule certitude est que leur habitat est limité par les côtes continentales. Lorsqu'on veut connaître la densité moyenne d'une telle population, force est donc de s'en fixer les limites géographiques. Dans la pratique, ceci revient à se demander : quel échantillon est à l'intérieur de la population et quel échantillon est à l'extérieur ?

La principale difficulté provient alors du traitement des valeurs nulles ou très faibles (notées ϵ). Rappelons rapidement, au risque de pécher par excès de généralité, que les densités de poissons ont le plus souvent des distributions qui présentent de très longues queues et, surtout, une proportion de valeurs faibles ou nulles très importante.

- 0 intérieurs et extérieurs.

Lorsque l'échantillonnage est régulier (systématique) en transects parallèles, l'habitude est de considérer qu'on a dépassé la population quand plusieurs densités sont consécutivement nulles. Il est alors assez logique de considérer ces valeurs comme extérieures à la population étudiée. Il arrive cependant assez fréquemment de trouver des concentrations nulles au milieu de certains transects. Comment doit-on les classer ?

Si on estime, sur la base de connaissances biologiques antérieures, que l'espèce en question a naturellement tendance à se distribuer en générant des trous de densité, alors il convient de traiter ces valeurs nulles comme faisant partie de la distribution, c'est-à-dire intérieures au domaine. Au contraire, on peut décider qu'il n'est pas dans les caractéristiques intrinsèques de l'espèce que sa régionalisation présente des trous de densités. Dans ce cas, on choisira plutôt de considérer les concentrations nulles comme extérieures à la population en ne les conservant pas pour le calcul de la densité moyenne.

- 0 ou ϵ ?

Une autre difficulté, d'origine technique, concerne les **incertitudes** à propos, en particulier, des mesures nulles. On sait en effet, que les différents engins de captures ont des efficacités variables et pas toujours très bien connues. Le poisson peut par exemple échapper à la capture d'un chalut pour diverses raisons (rapidité du chalutier, taille de la maille, ...). Il convient alors de se demander si les 0 sont des 0 stricts, ou seulement l'indication que la densité était très faible. Dans ce dernier cas, la différence entre 0 et ϵ n'est pas significative et la délimitation du domaine s'en trouve encore un peu plus compliquée.

Ainsi, en fin de transect, la première de trois valeurs consécutivement nulles, peut très bien être jugée incertaine surtout si la densité précédente est très forte. Elle pourrait alors, bien que nulle, être incluse dans la population, ce qui ne serait pas sans effet sur la valeur de la densité moyenne.

Cette incertitude sur les mesures pourrait même raisonnablement justifier de ne sélectionner que les données supérieures à un certain seuil. En procédant de la sorte, on serait amené à supprimer une proportion importante de données et par là à modifier sensiblement la densité moyenne de la population.

2.1.2 De la moyenne à l'abondance

Nous avons vu la réelle difficulté qu'il y a à choisir un champ, et les conséquences que cela peut avoir sur le calcul d'une moyenne. Une solution consiste à ne plus s'intéresser à la densité moyenne

mais plutôt à l'abondance totale de la population.

Relativement au problème du traitement des valeurs faibles et nulles, l'abondance est un concept qui se révèle particulièrement adapté. La raison en est assez simple. Une **somme**, ou encore une abondance, contrairement à une moyenne, **est inchangée par l'addition d'un nombre quelconque de valeurs nulles**. Elle est également peu sensible à la prise en compte ou non des petites concentrations bien que nombreuses. On peut donc étendre le calcul à tout l'espace sans que cela change l'abondance. C'est ce qui fait de l'abondance un **concept indépendant du champ** (ce que n'est pas la moyenne) et par là, opératoire.

Passer de la moyenne à l'abondance, n'est pas qu'une question de traitement efficace des faibles densités. En s'intéressant aux sommes, on estime que l'abondance est un critère plus pertinent du point de vue halieutique que la densité moyenne, ce qui est par exemple le cas pour le suivi de l'état de santé des stocks exploités.

2.1.3 Généralisation du principe: "ce qui était moyenne devient somme"

Nous venons de remplacer la densité moyenne par l'abondance. Est-il intéressant de généraliser le principe selon lequel "ce qui était une moyenne devient une somme"? Peut-on construire une famille d'outils de type "somme" en remplacement des critères de type "moyenne"? Et, quel sens leur attribuer?

Covariance et covariogramme

Oublions un instant que la covariance (non centrée) ne concerne que des variables ou des fonctions aléatoires pour n'en conserver que sa version expérimentale, à savoir une moyenne de produits (en géostatistique, la covariance d'une densité pour une distance h est estimée par la valeur moyenne du produit des densités distantes géographiquement de h). En prenant, non plus la moyenne, mais l'intégrale de ces mêmes produits, on s'oriente vers le covariogramme (Fig. 2.1). On ne retrouve cependant pas complètement le covariogramme puisque, contrairement au covariogramme, la covariance n'est calculée qu'à partir des paires telles que à la fois x et $x+h$ sont inclus dans le champ. Nous verrons dans la deuxième partie le sens et l'intérêt du covariogramme (Cf. §II.3), mais d'ores et déjà on peut retenir que **le covariogramme est le pendant de la covariance quand on passe d'une moyenne à une abondance**.

Histogrammes

L'histogramme expérimental des fréquences de densités par classe de densités a aussi un équivalent en termes d'abondance. En effet, plutôt que de compter le nombre de données appartenant à une classe de densité qui, rapporté au nombre total de mesures, correspond à la proportion (fréquence) de la classe (densité de probabilité dans le modèle), on peut tout aussi bien calculer l'abondance associée aux densités d'une classe.

Compte tenu des remarques que nous avons déjà faites sur les caractéristiques générales d'une densité de poissons, il faut s'attendre à ce que ce "nouvel" histogramme ait une allure très différente de l'histogramme "classique". On peut voir en effet (Fig. 2.1) que dans le cas de la densité d'œufs de maquereau, les abondances respectives des différentes classes de densité sont beaucoup plus homogène que les fréquences par classe. Ces abondances sont exprimées en pourcentage de l'abondance totale de sorte qu'on peut interpréter facilement les deux histogrammes en même temps. Il apparaît alors que les concentrations les plus faibles représentent près de 80% des données (Fig. 2.1) mais seulement 20% de l'abondance totale. Ce type de résultats, où se croisent fréquences et

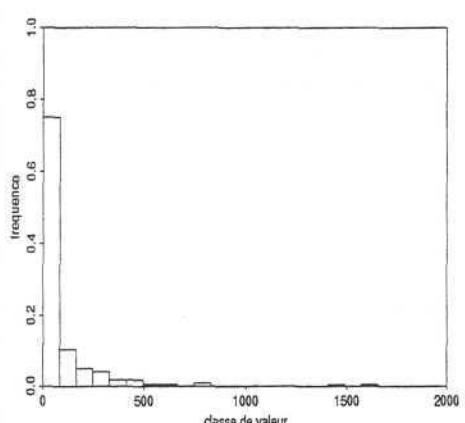
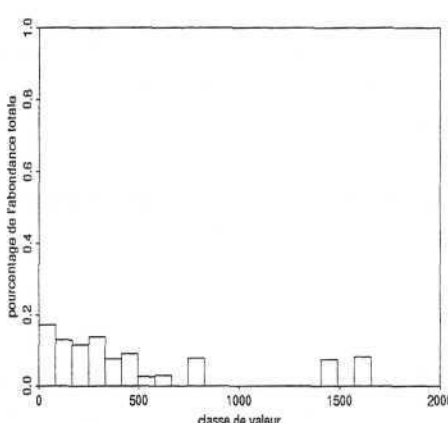
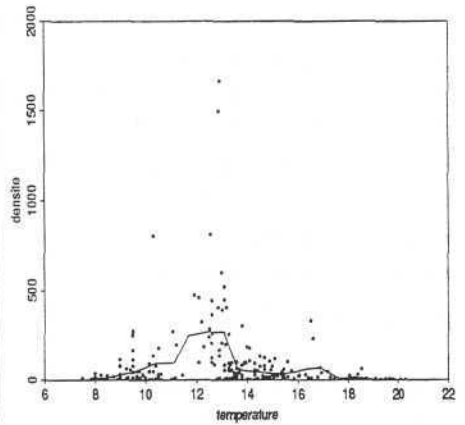
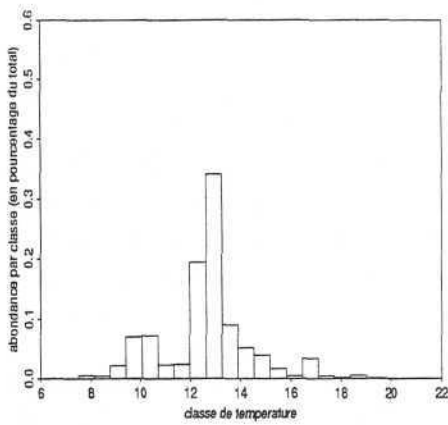
$m = \frac{1}{S} \int_S z(x) dx$ <p>moyenne</p>	<p>remplacé par</p>	$Q = \int z(x) dx$ <p>abondance</p>
$cov(h) = E[Z(x)Z(x+h)]$ <p>covariance non centrée</p>		$g(h) = \int z(x)z(x+h) dx$ <p>covariogramme</p>
 <p>fréquence par classe</p>		 <p>abondance relative par classe</p>
 <p>moyenne par classe</p>		 <p>abondance relative de poissons par classe de température</p>

FIG. 2.1 - Illustration du principe "ce qui était une moyenne (colonne de gauche) devient une somme (colonne de droite)". De haut en bas : densité moyenne / abondance - covariance non centrée / covariogramme - fréquence / abondance relative par classe - densité moyenne / abondance relative d'œufs par classe de température de surface (1989, deuxième période d'échantillonnage).

abondances, correspond aux courbes de sélectivité utilisées en géostatistique. Nous verrons dans la deuxième partie une autre interprétation possible des histogrammes d'abondances relatives par classe.

Enfin, on suggère également de décrire les liens bivariés à l'aide d'abondances relatives par classe. Mais, j'ai préféré réserver un paragraphe entier au cas bivarié afin de pouvoir revenir plus en détails sur les régressions expérimentales, et justifier ainsi l'usage des abondances par classe.

2.2 Régressions expérimentales entre une densité et une variable de l'environnement

Avant d'exposer le sens et l'intérêt d'un histogramme d'abondance par classe dans un cas bivarié, je propose d'abord de faire quelques remarques sur l'utilisation des régressions dans un contexte spatial.

2.2.1 La régression dans un contexte spatial

Domaine de définition

Pour juger l'effet d'une variable de l'environnement comme la température de surface sur une variable densité, on a souvent recours aux **régressions expérimentales**, c'est-à-dire aux **moyennes de densités par classe de température de surface**. Là encore, puisqu'il s'agit de calculer une moyenne, se pose le problème de la délimitation du domaine de calcul. Il s'exprime ici en des termes un peu différents puisque la température de surface est disponible sur toute la surface de l'océan. Son champ est à ce titre "infini"; tout du moins comparé à celui d'une population piscicole. Quel domaine faut-il donc prendre pour la régression? Celui de la température de surface? ou celui de la population?

Reprenons le premier exemple du chapitre précédent (échantillonnage régulier à 1D). Imaginons qu'en chacun des 20 échantillons on connaisse également la température de surface. Intéressons nous aux 6 échantillons (2x3) des extrémités droite et gauche du transect où la densité est nulle. Ayant montré que la température de surface est une variable souvent très régulière dans l'espace (deuxième exemple), on peut raisonnablement considérer que les 3 valeurs de température de chacun des 2 triplets sont identiques; ou, au moins, qu'elles appartiennent à une même classe. Inclure ces points dans le nuage de corrélation expérimental, revient donc à rajouter, à deux reprises, trois densités nulles dans une même classe de température. La régression expérimentale s'en trouverait alors modifiée.

Il est donc clair que **les régressions expérimentales sont définies relativement à un domaine de l'espace**. Elles n'ont donc pas de valeur intrinsèque par rapport à la question du lien entre deux variables de champs très différents. Il reste cependant qu'elles fournissent un moyen commode pour décrire, à l'intérieur d'un domaine fixé, l'évolution de la densité moyenne en fonction d'une variable "explicative".

Le terme "variable explicative" est d'ailleurs probablement usurpé. En effet, la courbe expérimentale des densités moyennes par classe de température ne permet pas d'expliquer pourquoi la population de poissons, prise dans son ensemble, s'est localisée là où on l'a observée. Si la courbe expérimentale montre un maximum autour d'une certaine valeur de température de surface (courbe en cloche), comme cela est souvent le cas en halieutique, on est impuissant à expliquer pourquoi il

n'y a pas de poisson dans une eau, de même température, mais quelques dizaines de milles nautiques plus loin. En somme, il y a loin entre le maximum d'une régression expérimentale, et la notion de *preferendum*. Cette distinction sera reprise plus loin dans le document (Cf. §II.2.3).

Une mauvaise bonne idée

Dans la mesure où chacune des deux variables serait compatible avec une approche par la géostatistique intrinsèque, on pourrait être tenté d'utiliser les outils structuraux multivariés comme le variogramme croisé. Le variogramme croisé est défini par l'espérance (la moyenne expérimentale) du produit des écarts de densité et des écarts de température pour une distance h donnée. Or, il arrive très fréquemment qu'en moyenne, une augmentation de température puisse être associée à une augmentation ou une diminution de densité selon le niveau de température considéré; c'est-à-dire selon qu'on se trouve à gauche ou à droite du maximum de la régression (Fig. 2.2). Les résultats structuraux sont donc inexploitablement et on doit retenir que **les variogrammes croisés, ou équivalents, sont applicables à des variables liées linéairement.**

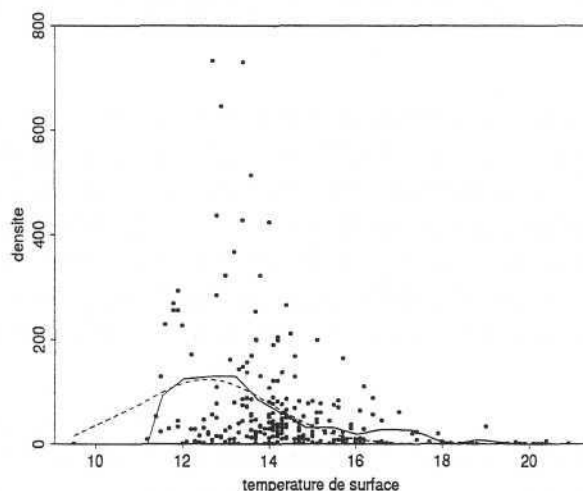


FIG. 2.2 - Nuage de corrélation entre la densité d'œufs de maquereau (ordonnée) et la température de surface (abscisse). Régression expérimentale (ligne continue) et ajustement non linéaire d'une exponentielle de Gauss (ligne pointillée). 1989. Troisième période d'échantillonnage.

Variabilité résiduelle. Modèles multiplicatifs

En plus d'être non linéaires, les régressions expérimentales ne rendent pas compte de la variabilité observée sur les nuages de corrélation. Quand on connaît l'importance, en terme d'abondance pour le stock, des quelques densités très élevées, on peut regretter que la régression échoue à les modéliser. Mais, les explosions de densités, dans notre exemple en tous les cas, apparaissent préférentiellement pour des températures voisines du maximum de la régression. Plus généralement, la variabilité autour de la régression semble d'autant plus importante que la régression est elle-même élevée. Expérimentalement, cela se traduit par une assez bonne **proportionalité entre la moyenne et l'écart-type des densités par classe de température de surface** (Fig. 2.3a).

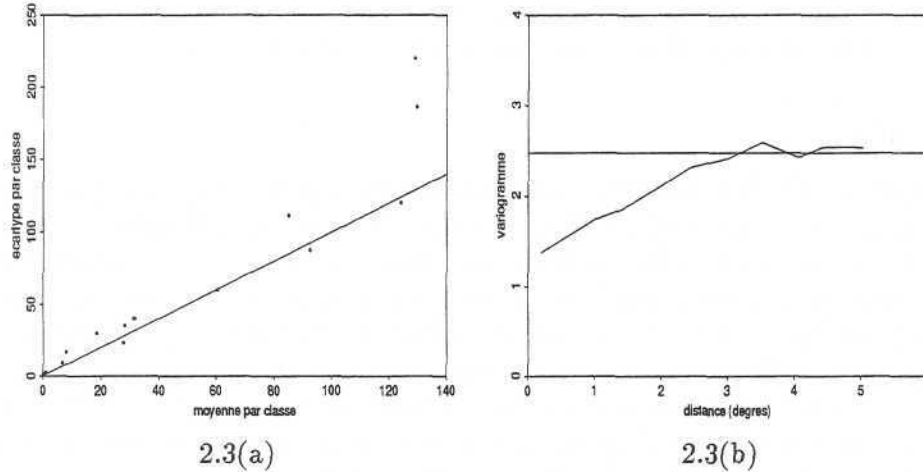


FIG. 2.3 - *Modèle multiplicatif. (a) Proportionnalité entre la densité moyenne et l'écart-type des densités par classe de température de surface. (b) Variogramme isotrope du résidu relatif*

Cette constatation suggère de considérer le résidu de la régression non pas de façon additive comme cela est généralement le cas ($Y = m(X) + R$), mais de façon **multiplicative** ($Y = m(X) \cdot R$). Le caractère multiplicatif du modèle autorise l'apparition de densités très élevées (explosions). On considère en géostatistique que le résidu (relatif), partie de la densité non expliquée par la température de surface, peut conserver une structure spatiale. Ce type de modèle constitue donc une autre approche géostatistique multivariable que celle déjà évoquée plus haut (structures croisées). Dans ce cas, on espère que le résidu relatif conduira à des résultats structuraux exploitables.

L'exemple que j'ai retenu porte sur le lien entre la densité d'œufs de maquereau et la température de surface. J'ai opté pour un ajustement non linéaire par moindres carrés d'une exponentielle de Gauss (Fig. 2.2). L'exponentielle, en plus de rendre compte de l'évolution générale de la corrélation (cloche), présente l'avantage de ne jamais s'annuler, ce qui facilite le calcul expérimental des résidus relatifs ($R = Y/Gauss(X)$).

Cependant des problèmes numériques se posent. Aux valeurs extrêmes de température la régression est très faible (asymptotiquement nulle). Le calcul expérimental du résidu relatif conduit donc à considérer le quotient de deux valeurs faibles dont le résultat est reconnu instable. J'ai par conséquent supprimé de l'analyse structurale tous les résidus relatifs supérieurs à 10. La densité maximum observée (≈ 800 œufs/m²) n'est guère que 4 fois supérieure à la valeur de la régression au même point (Fig. 2.2). Le seuil que j'ai fixé (résidu < 10) ne concerne donc que les marges du nuage de corrélation. D'autres difficultés existent, dont on pourra trouver un exposé détaillé dans Rivoirard et Guiblin (1996), mais je retiendrai que la mise en œuvre d'un tel modèle est, pour le moins, délicate.

Le variogramme expérimental du résidu relatif est calculé de façon isotrope avec des distances en degrés. L'effet de pépite, c'est-à-dire l'irrégularité locale, est important puisqu'il représente 50% de la variance (Fig. 2.3b). Le variogramme se stabilise autour de la variance expérimentale à partir de 3.5 degrés (soit environ 200 milles nautiques). Les résultats sont donc satisfaisants et semblent modélisables.

Echelle

Le haut niveau de variabilité d'une concentration de poissons génère donc des difficultés de modélisation bivariable. On ne doit d'ailleurs être que peu surpris dans la mesure où nous avons déjà évoqué les ennuis rencontrés lors de l'analyse monovariable d'une densité. Mais, cette forte variabilité n'est-elle pas aussi le signe qu'une approche trop fine, trop locale, est irréaliste? Si les systèmes biologiques naturels témoignent d'une très grande capacité à s'adapter aux changements environnementaux, c'est probablement grâce à une **variabilité intrinsèque** (intraspécifique) élevée de la sensibilité des poissons d'une espèce donnée à leur environnement. En conséquence, est-il sensé d'un point de vue biologique de construire un modèle qui prétende prédire le niveau de densité attendu pour une température donnée? Est-ce là une échelle pertinente pour l'analyse biologique? Ou bien, ne vaut-il pas mieux adopter un point de vue plus global?

Le paragraphe suivant propose une alternative pour décrire un lien bivariable qui, en recourant aux abondances, se place à une autre échelle que la régression et évite dans le même temps l'ensemble des problèmes liés à la délimitation du champ.

2.2.2 Abondances de poissons par classe de température

Comme pour l'analyse d'une densité, on suggère de remplacer la moyenne par l'abondance. Ceci conduit à décrire le lien entre deux variables en termes d'abondance de l'une par classe de l'autre (par exemple abondance de poissons par classe de température de surface). Relativement à l'abondance totale, cela amène à fournir le **pourcentage de poissons capturables par classe de température**. Cette notion présente l'avantage d'être indépendante de la définition d'un champ. En faisant la somme des densités sur l'espace des températures appartenant à une certaine classe, on récupère l'abondance recherchée quel que soit le nombre de densités nulles rencontrées.

Ainsi par exemple, on peut estimer que, pendant le mois de juin 1989, près de 40% des œufs de maquereau ont été pondus dans une eau entre 12.5°C et 13°C (Fig. 2.1). On remarquera également l'évolution symétrique des abondances par classe autour d'une valeur centrale de température de 13°C (histogramme en cloche).

Cependant, s'il permet de traiter efficacement le problème des zéros et de décrire un nuage de corrélation dont l'une des variables est une densité de poissons, le concept d'abondance par classe ne permet pas, pour des raisons voisines de celles déjà évoquées à propos de la régression, d'expliquer la nature du conditionnement d'une variable de l'environnement sur une densité de poissons. On s'est donc doté d'un concept opératoire, pertinent pour la description du lien entre une densité de poissons et une variable de l'environnement, mais impropre à expliquer le mécanisme du conditionnement. C'est la porte ouverte à une réflexion plus large que j'entend développer dans le paragraphe suivant sur la recherche de concepts pertinents en écologie halieutique.

2.3 Et si nous connaissions exhaustivement la réalité?

Que ferions nous si nous connaissions exhaustivement la réalité? Pourquoi est-il intéressant de se poser une telle question? On pourrait en effet douter de l'intérêt de se placer, par la pensée, dans une situation où tout serait connu. En définitive, cela permet d'envisager les deux points essentiels que sont l'objectivité des modèles statistiques et la pertinence des concepts utilisés en statistique.

2.3.1 Pertinence d'un concept

Il n'y a rien de tel pour juger de la pertinence (capacité) d'un concept à répondre à une question posée, que d'imaginer connaître exhaustivement la réalité. Imaginons par exemple que nous connaissions complètement la distribution spatiale d'une espèce de poissons ainsi que la température de surface. Les moyens d'investigations dont nous disposons actuellement (moyens acoustiques et satellitaires) et dont l'utilisation se développe de plus en plus, rendent une telle perspective "probable". C'est donc dans un "imaginaire proche" que je me place. Comment faire alors pour déterminer et quantifier l'effet de la température de surface sur la régionalisation des poissons? **Sait-on construire un concept pertinent pour répondre à cette question si importante en écologie halieutique?** Il nous faut bien malheureusement reconnaître aujourd'hui que nous sommes assez démunis face à ce genre d'interrogations.

C'est là un exemple important parmi d'autres. De nombreuses questions biologiques devraient d'abord être envisagées sous cet angle. On constaterait alors probablement que **la difficulté majeure n'est pas tant le manque d'information que le manque de concept pertinent**. J'insisterais volontiers sur ce point car il me semble qu'on est trop souvent enclin à penser que c'est en accumulant de plus en plus d'informations qu'on augmentera notre compréhension des systèmes marins. Cela est d'autant plus douteux que cet engrangement de données de toutes sortes n'est que rarement contrôlé. On ne sait en effet généralement pas précisément qu'elles sont les caractéristiques pertinentes à mesurer parmi la multitude des critères disponibles. Cette manière de faire s'apparente donc plus à une fuite en avant, qu'à une amélioration des connaissances.

Il reste toutefois indéniable qu'il est souhaitable, en théorie, d'augmenter autant que faire se peut l'effort d'échantillonnage. Mais, il me paraît tout autant fondamental, voire prioritaire puisque le temps est limité, d'approfondir la réflexion sur la conceptualisation des systèmes biologiques complexes. En outre, il me paraît inutile d'affiner à l'excès les méthodes d'ajustement d'une régression dont les résultats n'auront que peu de valeur en regard de la question posée.

On attend donc de la (géo)statistique :

- qu'elle définisse un concept clef par rapport au problème posé. C'est-à-dire qu'elle permette de "**faire parler les données**" emmagasinées.
- et qu'elle fournisse les moyens d'**estimer** les concepts retenus dans les meilleures conditions possibles, au titre desquelles on trouve l'objectivité.

2.3.2 Objectivité

En utilisant des concepts construits sur la base d'une connaissance exhaustive de la réalité, on assure également l'**objectivité** de leurs estimations statistiques. En effet, il ressort du travail de Matheron (1978) que l'objectivité peut être jugée, en statistique, par la **décidabilité**. Cette acception précise de l'objectivité conduit à ne considérer comme objectif que ce qui peut être décidé, c'est-à-dire infirmé ou confirmé, à partir d'une information exhaustive du phénomène étudié. A cet égard, la moyenne est un exemple de concept objectif car défini sans ambiguïté par la donnée d'un domaine et de la valeur de la variable en tous les points de l'espace. Par contre, l'espérance mathématique, c'est-à-dire le moment d'ordre 1 de la loi de probabilité associée au phénomène, n'a pas, de ce point de vue, de signification objective. On retiendra en particulier cette réflexion de Matheron : "**Il n'y a pas de probabilité en soi; il n'y a que des modèles probabilistes**". A

ce titre, l'énoncé "la variable régionalisée $z(x)$ est une réalisation d'une fonction aléatoire $Z(x)$ " est également dépourvu de signification objective. Rien ne permettra jamais de confirmer ou d'infirmier un tel propos.

Ainsi, tout concept, notion, modèle ou énoncé est objectif s'il peut s'exprimer en terme de **grandeurs régionales**, c'est-à-dire à l'aide d'entités qui seraient disponibles si nous connaissions exhaustivement la réalité.

Bien qu'objectif, un concept ainsi défini n'en est pas pour autant connu dans la pratique. Nous ne disposons en effet que d'une connaissance partielle, voire extrêmement limitée, de la réalité. Il faut donc se doter de moyens d'estimations adéquats. C'est la raison pour laquelle le chapitre suivant est consacré au choix d'un modèle.

Chapitre 3

Choix d'un modèle

Sommaire

Ce chapitre, largement inspiré du travail de Matheron (Estimer et choisir, 1978), analyse les étapes importantes du choix d'un modèle.

Le paragraphe 1 distingue le choix d'une loi de probabilité en statistique et le choix d'un modèle de fonction structurale en géostatistique. On montre que la modélisation affronte les mêmes écueils dans la pratique, qu'elle concerne un histogramme ou un variogramme. On discute alors le reproche fréquemment fait à la géostatistique d'être "modèle-dépendante".

Le paragraphe 2 décrit les trois étapes du choix d'un modèle et définit les critères de qualité d'un modèle. Il en ressort que le choix du cadre méthodologique qui correspond à la première étape de la modélisation, nécessite plus de précautions qu'on ne lui en accorde généralement. Le choix (systématique) de la géostatistique intrinsèque peut conduire dans certains cas, et en écologie halieutique en particulier, à des résultats peu satisfaisants.

Le paragraphe 3 examine le cas particulier du choix d'un modèle constitutif en géostatistique et développe les arguments du choix entre la géostatistique transitive et intrinsèque. Les définitions importantes sont rappelées et l'alternative entre les deux méthodes est décrite autour des deux éléments importants que sont le champ et la stationnarité. De là, il apparaît que la géostatistique transitive, bien que peu développée, est une méthodologie bien adaptée aux données halieutiques et aux problèmes posés.

Choice of a model

Abstract

This chapter, deeply inspired by one of the book written by Matheron (Estimating and choosing, 1989) analyses the important steps of the choice of a model.

Paragraph 1 makes the comparison between the choice of a probability law in statistics and the choice of a structural model in geostatistics. We indicate that in both case, the modelization faces the same kind of problems. In particular, the modelization of an histogram or a variogram get similarities. At the end, I recall and discuss the usual criticize made about geostatistics that is the model dependency.

The paragraph 2 describes the three steps of the choice of a model and defines also the criteria of the quality of a model. We then can see that the definition of the methodological context, which is in fact the first step of the modelization, is the most important one. That is why we must pay more attention to this step than we usually do. As an example, we quite systematically choose intrinsic geostatistics. But, in some cases, and in particular in fisheries, this can provide unsatisfactory results.

In the paragraph 3, I argue about the choice of a constitutif model in geostatistics, and in particular about the difference between intrinsic and transitive geostatistics. The major definitions are recalled and the difference between the two approaches is described in terms of field and stationnary considerations. Then, even if the transitive methodology is not widely developped very much, it will appear to be appropriate for the analysis of many fish.

Le choix d'un type de loi de probabilité, et ensuite sa spécification, permet en statistique "classique" de proposer une variance de l'estimateur de la densité moyenne sur un domaine fixé. Mais pour cela il faut au préalable avoir choisi de recourir aux techniques de la statistique. Pour sa part, l'analyse variographique se place dans le contexte des fonctions aléatoires, ce qui présuppose qu'on a choisi de recourir aux techniques de la géostatistique intrinsèque.

En somme, on perçoit bien que le choix d'un modèle comporte plusieurs étapes et en particulier, celle du choix d'une méthodologie. Dans ce chapitre, je reprend une partie des réflexions exposées par Matheron (1978) sur la différence de statut des 3 étapes du choix d'un modèle dans le but de souligner l'importance du choix d'un modèle constitutif. Dans ce cadre, j'insisterai sur la nécessité, en géostatistique appliquée à l'écologie halieutique, de se poser la question de savoir laquelle des deux approches intrinsèque et transitive est la plus appropriée.

3.1 Quel modèle et pour quoi faire?

On entend souvent dire que la géostatistique est "modèle-dépendante" en cela probablement que toute étude géostatistique nécessite la modélisation d'un variogramme (ou d'une structure équivalente). C'est l'occasion de rappeler ce qui est modélisé en géostatistique et dans quel but. Mais c'est aussi l'occasion de comparer les approches statistique et géostatistique du point de vue de leurs exigences respectives de modélisation.

3.1.1 Choix d'un modèle de loi de probabilité en statistique "classique"

Au travers du premier exemple introductif (§I.1.1), il a été mis en évidence qu'il faut, en statistique "classique", choisir un modèle de loi de probabilité pour estimer la variance d'une estimation et plus globalement sa qualité. Par contre, l'estimateur (de la moyenne) a été construit, lui, sans qu'il y ait besoin de se référer à un modèle probabiliste et *a fortiori* de choisir une loi de probabilité.

Mais ce n'est pas là une tendance générale. Il arrive en effet que le recours au modèle s'impose très tôt dans l'analyse, c'est-à-dire dès la phase d'estimation. C'est le cas par exemple de la technique du maximum de vraisemblance qui permet de construire un estimateur à loi de probabilité fixée.

Généralement, c'est le statisticien qui choisit le type de la loi de probabilité (Fig. 3.1). Ce qui signifie en particulier qu'il ne laisse pas le soin à une procédure automatique de le faire à sa place. C'est lui qui décide d'utiliser, par exemple, une loi normale, log-normale ou encore binomiale. Il est aidé dans son choix par l'histogramme expérimental bien sûr, mais également par d'éventuelles connaissances sur le phénomène; quitte d'ailleurs, au moment du choix, à négliger consciemment et délibérément une partie de l'histogramme qui pourrait ne pas être en accord avec le modèle retenu. C'est le cas en particulier des queues de distributions ou encore d'une classe de valeur mal représentée à cause du plan d'échantillonnage.

Qu'en est-il de la géostatistique? A quel moment intervient le choix d'un modèle? Et pour quoi faire?

3.1.2 Choix d'un modèle structural en géostatistique

Nous verrons dans la deuxième partie du document que l'estimation globale d'abondance en géostatistique transitive constitue une situation analogue à celle rencontrée en statistique pour l'estimation de la moyenne. C'est-à-dire que l'estimateur est d'abord construit en dehors de toute considération touchant au choix d'un modèle; la modélisation du covariogramme n'étant nécessaire que pour calculer la variance d'estimation. Nous verrons de plus que la probabilisation de l'information qui nous autorise à parler de biais ou de variance ne concerne que le réseau des points d'échantillonnage et non les valeurs de densités observées.

En cherchant à tenir compte de la structure spatiale du phénomène pour l'estimation locale (krigeage), on rend l'étape d'estimation elle-même tributaire du choix d'un modèle. La raison en est assez simple. Les distances qui séparent le point à estimer des échantillons qui sont dans son voisinage ne correspondent en général pas aux distances disponibles à l'échantillonnage. Il est donc nécessaire, sur la base de la structure spatiale expérimentale, de se doter d'une fonction structurale connue sur l'ensemble des distances intervenant dans le calcul. C'est-à-dire d'ajuster un modèle aux résultats expérimentaux. Mais on aura bien noté qu'il ne s'agit plus de choisir une loi de probabilité (on ne parle même pas d'histogramme), mais plutôt une fonction structurale, c'est-à-dire un variogramme, une covariance ou encore un covariogramme.

Pour ce faire, on préfère ne pas recourir à une procédure complètement automatique (on rejoint en cela la statistique "classique"). On veut pouvoir en particulier contrôler que le comportement de la fonction structurale à l'origine rende compte des résultats expérimentaux et des caractéristiques physiques de la variable régionalisée étudiée. Il est en effet démontré (Matheron, 1965) que le comportement analytique du modèle au voisinage de l'origine (faibles distances) est relié à la régularité spatiale du phénomène à cette même échelle. Il convient donc de garder à l'esprit ce principe au moment de la modélisation, ce que ne permet pas, toujours, de faire une procédure automatique.

Par ailleurs, les premiers points des variogrammes expérimentaux ont en général un poids statistique inférieur à celui des points suivants (à 2D : nombre inférieur de paires participant au calcul). Il est donc normal de ne leur accorder qu'une importance modérée. Cela n'est pas incompatible avec une procédure d'ajustement automatique puisqu'il suffit d'attribuer des poids différents aux points du variogramme avant l'ajustement. Toutefois, en procédant ainsi, on s'oriente vers des outils dont

on maîtrise de moins en moins l'interprétation physique.

Je retiendrais donc finalement que l'esprit de la démarche est le même qu'il s'agisse de choisir une loi de probabilité en statistique ou une fonction structurale en géostatistique.

3.1.3 Approches modèle-dépendantes?

La géostatistique est-elle donc si modèle-dépendante? En particulier, l'est-elle plus que la statistique "classique"?

Il me semble que le parallèle que j'ai tenté de faire entre la modélisation d'un histogramme d'une part, et la modélisation d'un variogramme d'autre part, est de nature à mettre les deux approches statistique et géostatistique au même niveau. La critique qui consiste à reprocher à l'estimation géostatistique de dépendre d'un modèle (de variogramme) est applicable également à toute estimation par maximum de vraisemblance qui est tout autant dépendante d'un modèle (de loi de probabilité).

De plus, dans les deux cas, le choix du modèle évite de faire appel à une procédure complètement automatique et aveugle.

Il me semble donc que ces **deux approches**, bien que cherchant à évaluer des concepts différents, **présentent le même niveau d'exigence de modélisation**.

3.2 Les 3 étapes du choix d'un modèle

Jusqu'à présent, on a parlé de modélisation uniquement pour le choix d'une fonction mathématique propre à représenter et combler "au mieux" des résultats expérimentaux et parcellaires. C'est oublier que la modélisation est en fait un processus beaucoup plus large que cette seule étape d'ajustement. En reprenant l'argumentaire développé par Matheron dans la deuxième partie de "Estimer et choisir" (1978), il apparaît qu'on peut distinguer trois étapes clefs dans le choix d'un modèle.

3.2.1 Choix d'un modèle constitutif

Au cours de chacun des quatre exemples introductifs (§I.1) nous avons successivement choisi de nous placer dans le cadre de la statistique classique, de la géostatistique intrinsèque, puis de la géostatistique transitive.

C'est le signe qu'il faut avant toute autre chose, décider de recourir ou non aux techniques probabilistes et/ou spatiales. **C'est une décision, et non une hypothèse**. C'est-à-dire qu'il n'existe pas d'expérience ou d'observation qui permettrait par exemple de réfuter une proposition du type "la distribution des merlus du Golfe de Gascogne est une réalisation d'un processus stochastique" de sorte que la sélection d'une méthodologie est une étape qui n'a pas de fondement objectif. Ce qui veut dire en particulier que le choix effectué n'est ni vrai ni faux. Il émane de ce que *nous avons de bonnes raisons de penser que c'est la meilleure décision possible, la seule peut-être qui fournisse un cadre conceptuel dans lequel nous puissions espérer formuler et résoudre au mieux le, ou les problèmes qui nous occupent* (Matheron, 1978).

Ce modèle est dit constitutif en tant qu'il fonde ou **constitue l'objet mathématique** (variable régionalisée, fonction aléatoire, ou bien encore variable aléatoire) que nous allons étudier par la suite. Il détermine aussi à l'avance les méthodes qui seront employées pour l'étude.

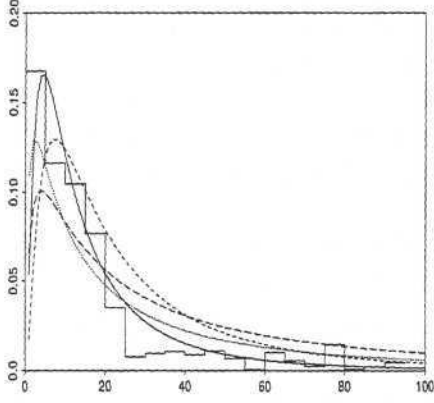
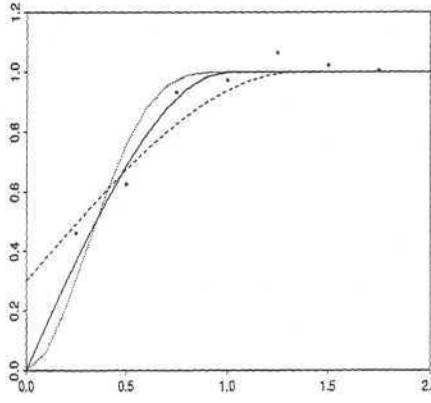
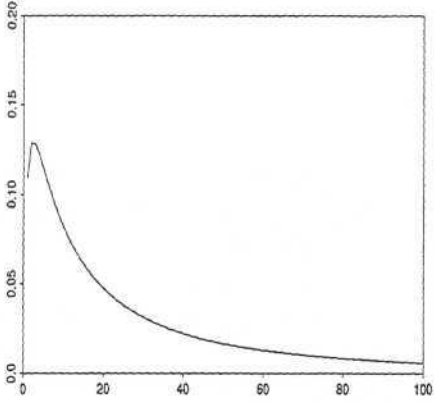
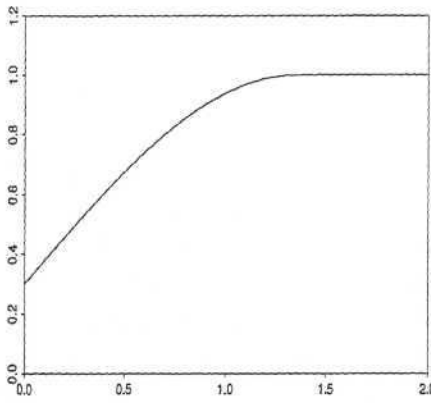
étape	modèle		
1	constitutif	variable aléatoire X	fonction aléatoire Z(x)
2	générique	 loi log-normale	 variogramme stationnaire
3	spécifique	 $m = 22.76; s = 66.3$	 0.3 pépite + 0.7 sphérique(h/1.3)

FIG. 3.1 - Les 3 étapes du choix d'un modèle. Deux choix constitutifs différents. Colonne de gauche : modèle de variable aléatoire en statistique. Colonne de droite : modèle de fonction aléatoire en géostatistique.

Cette étape n'est généralement **pas** considérée comme faisant partie de la modélisation. Or c'est bien ainsi qu'il faut l'interpréter: **le choix d'une méthodologie constitue un élément très amont mais fondamental d'un modèle statistique.**

3.2.2 Choix d'un modèle générique

Dans le cadre conceptuel défini par le modèle constitutif, on est ensuite amené à faire une **hypothèse de nature physique** sur le phénomène, c'est-à-dire contrôlable dans le cadre hypothétique d'une connaissance exhaustive du domaine réel. C'est donc une décision d'une toute autre nature que le choix d'un modèle constitutif. J'ai déjà mentionné un certain nombre d'exemples de ce type d'hypothèses dans le premier chapitre de cette partie. Ce peut être par exemple l'homogénéité dans l'espace qui conduit au modèle générique d'une fonction aléatoire stationnaire (Fig. 3.1). On notera qu'à cette étape, on a uniquement supposé la stationnarité (modèle générique) de la fonction aléatoire (modèle constitutif). Sur la base d'un autre modèle constitutif (variable aléatoire par exemple), on pourrait choisir comme modèle générique une loi de probabilité log-normale (Fig. 3.1), sans pour autant à ce stade en préciser les paramètres.

Cependant, nous ne sommes jamais capables que de contrôler l'adéquation du modèle aux données expérimentales - au moyen par exemple de tests statistiques qui, soit dit en passant, reposent eux-mêmes sur un certain nombre d'hypothèses. Mais, rien n'implique que le modèle soit compatible avec les données non disponibles (i.e. les parties inconnues ou inaccessibles du phénomène). En outre, *il ne s'agit pas simplement de "tester" une hypothèse à l'aide des données* (Matheron, 1978). Il faut décider, **faire l'hypothèse beaucoup plus forte, que le modèle générique est valable partout.** C'est d'ailleurs en cela que la modélisation, et donc plus particulièrement le choix d'un modèle générique, revient à **apporter de l'information** au matériel expérimental.

3.2.3 Choix d'un modèle spécifique

Une fois le modèle générique fixé, il ne reste plus qu'à le spécifier, c'est-à-dire à estimer les quelques paramètres qui manquent pour connaître parfaitement la fonction mathématique qui fera office *in fine* de "modèle". On aura pris soin de choisir un modèle générique qui ne dépende que d'un nombre minimum de paramètres. L'expérience a retenu des lois de probabilités ou des fonctions structurales qui ne dépendent que de 2 ou 3 paramètres. Une loi log-normale est par exemple connue sans ambiguïté par la connaissance du couple espérance-variance. Dans notre exemple (Fig. 3.1), on a spécifié le modèle générique "loi log-normale" par le choix d'une espérance ($m=22.76$) et d'un écart-type ($s=66.3$).

En géostatistique, l'habitude a "consacré" l'utilisation d'un petit nombre de schémas structuraux dont le sphérique (Fig. 3.1) défini par une portée et un palier auquel on peut ajouter par exemple un effet de pépite (discontinuité à l'origine).

En statistique "classique" cette étape de spécification d'un modèle générique est plus connue sous le nom d'inférence statistique. C'est d'ailleurs le plus souvent à cette seule troisième étape qu'il est fait référence lorsqu'on parle de modélisation. Cet abus de langage ne doit cependant pas cacher et, à force d'habitude amener à ignorer, que la modélisation commence par le choix d'un cadre méthodologique.

3.3 Choix d'un modèle constitutif: géostatistique transitive ou géostatistique intrinsèque?

Au cours des exemples introductifs, l'utilisation du variogramme, et donc de la géostatistique intrinsèque, n'est pas apparue très pertinente dans le cas des densités de merlu, alors que les résultats expérimentaux obtenus avec la température de surface étaient satisfaisants. Des réflexions précédentes sur les différentes étapes du choix d'un modèle, il ressort que l'utilisation du variogramme, et donc **la faiblesse des résultats, peut relèver du choix du modèle constitutif**. Plus généralement, on peut penser qu'une partie des difficultés rencontrées en géostatistique appliquée à l'halieutique provient du choix trop systématique d'un modèle de fonction aléatoire comme modèle constitutif quand la réalité physique ne s'y prête guère.

Plutôt que de remettre en cause la géostatistique dans son ensemble, une alternative consiste alors à envisager un modèle de variable régionalisée moins exigeant en hypothèses. Cette méthodologie connue sous le nom de géostatistique transitive, a été développée par Matheron (1965) dans le but de conduire des estimations globales d'abondances assorties d'une variance d'estimation, sur la base d'un échantillonnage régulier, en limitant au maximum le nombre des hypothèses requises. Bien que plus simple dans sa formulation et plus ancienne, cette méthode est moins utilisée et moins développée que la géostatistique intrinsèque.

Je me propose dans ce paragraphe de commencer par évoquer les exigences de l'approche intrinsèque en termes de stationnarité et de champ. Les raisons des non-stationnarités rencontrées en halieutique conduiront à la géostatistique transitive.

3.3.1 Fonction aléatoire, stationnarité et champ

Fonction aléatoire et réalisation

La **géostatistique intrinsèque**, c'est-à-dire la partie de la géostatistique la plus couramment utilisée, est développée dans le cadre conceptuel des **fonctions aléatoires**. On peut définir une fonction aléatoire comme vecteur aléatoire de dimension infinie (infinité de points d'appui), ou de façon équivalente, comme une fonction de l'espace dont la valeur en un point donné est une variable aléatoire.

Dans ce cadre, une variable régionalisée, comme par exemple une distribution de poissons, est interprétée comme étant une réalisation de la fonction aléatoire; plus précisément, comme l'une des réalisations possibles de la fonction aléatoire. Pour faire le parallèle avec la statistique classique, prenons l'exemple du jet répété d'un dé. La valeur 3 obtenue par exemple lors du premier jet est interprétée comme une réalisation de la variable aléatoire "valeur obtenue par le lancement d'un dé". Les valeurs obtenues lors des épreuves successives sont alors autant de réalisations de la même variable aléatoire. Dans le cadre de la géostatistique intrinsèque, la fonction qui représente la totalité d'une **régionalisation ne forme qu'une seule réalisation de la fonction aléatoire**. Les notations usuelles en géostatistique se fondent en partie sur le parallèle que je viens de faire:

- en statistique, on note x la valeur et X la variable aléatoire.
- en géostatistique, on note $z(x)$ la variable régionalisée et $Z(x)$ la fonction aléatoire (x étant un point de l'espace de coordonnées (x,y,z,\dots))

L'espérance d'une fonction aléatoire s'entend théoriquement comme la moyenne de la fonction en un point donné pour l'ensemble des réalisations possibles. Or, on ne dispose réellement que

d'une seule réalisation, ce qui, *a priori*, annule toute possibilité d'estimer l'espérance d'une fonction aléatoire.

Stationnarité

On dit qu'une fonction aléatoire est stationnaire si sa loi spatiale est invariante par translation. A cette définition "forte" correspond une version "faible" qui ne concerne que les deux premiers moments de la fonction aléatoire. Une fonction aléatoire est stationnaire (sous-entendu d'ordre 2) si son espérance et sa covariance sont stables par translation, c'est-à-dire identiques en tout point de l'espace. A noter que **la stationnarité est une propriété du modèle** de fonction aléatoire et non une caractéristique de la variable régionalisée.

En supposant que la fonction aléatoire est stationnaire (et ergodique), le problème du calcul de son espérance trouve une solution simple puisqu'on peut alors penser que les fluctuations spatiales de densités observées *in situ* sur une seule réalisation sont représentatives des fluctuations statistiques de la fonction aléatoire en un point donné. Ainsi, **sous couvert de stationnarité (et d'ergodicité), on remplace une moyenne ponctuelle sur l'ensemble des réalisations (espérance), théorique mais inaccessible, par une moyenne spatiale sur une seule réalisation, accessible mais reposant sur une hypothèse assez forte.**

Champ

L'approche intrinsèque conduit donc *in fine* à **choisir un champ** au sein duquel les fluctuations du phénomène sont jugées suffisamment homogènes pour justifier un modèle de fonction aléatoire stationnaire. En particulier, on cherchera à ce que les bordures du champ ne soient pas systématiquement plus pauvres que le cœur (absence d'influence du champ sur la régionalisation), ou encore à ce que les valeurs fortes ne soient pas localisées préférentiellement dans le champ. Ces deux points sont des traductions en termes physiques d'une non-stationnarité et ont une acuité particulière en biologie marine comme nous le verrons dans le paragraphe suivant.

Pour une fonction aléatoire stationnaire, le champ apparaît donc comme une fenêtre qu'on aurait découpée au sein d'une régionalisation plus vaste. Cela signifie en particulier que ce type d'approche a vocation à **analyser la structure interne, locale, ou encore intrinsèque des phénomènes régionalisés dont la distribution spatiale est indépendante de la géométrie du champ.** C'est pour cette raison qu'on parle de géostatistique intrinsèque par opposition à la géostatistique transitive que nous verrons plus loin. C'est également pour cela que l'approche intrinsèque convient bien à l'analyse de la température de surface et non à celle des densités de merlus.

3.3.2 Causes possibles d'une non-stationnarité

Régionalisations champ-dépendantes

Il n'est pas rare qu'à l'échelle d'un peuplement, d'une population ou d'une partie d'un stock, on observe que la densité d'individus soit importante au centre de la distribution et qu'elle s'étiolle sur les bords. Cette non-stationnarité globale peut toutefois contenir, à des échelles plus grandes, des situations compatibles avec un modèle stationnaire. Il est en effet possible que la distribution des densités soit à peu près homogène au sein des zones respectivement denses et dépeuplées, pour autant que ces zones soient géographiquement séparables. C'est en tous les cas l'illustration que la stationnarité (du modèle) dépend de l'échelle à laquelle se déroule l'étude.

Il est des cas cependant où les bordures ne semblent pas systématiquement plus faibles que le cœur. Au sein d'un banc de thons par exemple, une fois franchies les limites du banc, on ne remarque pas d'effet de bordure net. En somme, les zones denses d'un banc n'ont pas de localisation privilégiée et on peut alors supposer qu'elles sont stationnaires dans le banc.

Par ailleurs, dans la mesure où les individus d'une population inter-agissent entre eux, on peut penser que la géométrie du peuplement est un élément clef de la répartition des individus en son sein. Il est alors illusoire de vouloir séparer ce qui est imputable à la variable elle-même (partie intrinsèque) et à son champ (partie géométrique). Dans toutes ces situations où la régionalisation dépend du champ, la géostatistique intrinsèque, qui a vocation à caractériser la structure spatiale intrinsèque d'un phénomène, est inappropriée.

Localisation préférentielle

En utilisant un modèle probabiliste de fonction aléatoire on suppose que l'agencement des densités de poissons dans l'espace n'est pas fixe. En d'autres termes, on pense qu'il aurait été possible d'observer une autre régionalisation que celle effectivement rencontrée. En cela, nous entourons la situation réelle d'un halo d'états virtuels considérés comme possibles (probabilisation). Cette vision des choses me paraît tout à fait réaliste dans la mesure où aucune espèce marine ne se positionne dans l'espace de façon précise et pré-déterminée. Un poisson a été observé à tel endroit mais il aurait très bien pu se trouver un peu à côté.

Pour justifier la stationnarité, il faut, par contre, imaginer que les densités, et en particulier les valeurs fortes, auraient pu être observées n'importe où ailleurs dans le champ; ce qui n'est plus la même chose. Par exemple, dans le cas de la ponte du maquereau, de nombreuses références bibliographiques (Anon., 1977; Anon., 1995; Walsh et Martin, 1986; Bez *et al.*, 1995) indiquent que la distribution des œufs est systématiquement organisée d'année en année autour des accores. On ne peut alors pas supposer que les pics de ponte auraient pu être rencontrés sur le plateau continental ou bien encore, plus au large. Des programmes scientifiques comme celui dans lequel s'effectue mon travail (SEFOS) ont d'ailleurs pour but d'expliquer la localisation préférentielle de ce type de concentrations. Il est donc probable que dans bien des cas, les zones riches ont une localisation préférentielle incompatible avec l'hypothèse de stationnarité.

Retour au modèle constitutif

Qu'elle concerne la variable proprement dite ou bien ses accroissements, la stationnarité (modèle générique) est un passage obligé dans l'utilisation des fonctions aléatoires (modèle constitutif). Sans cette hypothèse, l'utilisation des fonctions aléatoires se révèle non-opératoire et donc inintéressante, même si, comme je l'ai indiqué, l'interprétation probabiliste sur laquelle elle est fondée est bien souvent raisonnable en halieutique. Lorsque le phénomène étudié ne se prête guère à une hypothèse stationnaire, il convient donc de reprendre le processus de modélisation dès le choix du modèle constitutif. On peut alors s'orienter vers la géostatistique transitive définie en terme de variables régionalisées et pour laquelle la probabilisation, qui s'exprime différemment, peut, dans certains cas, convenir aux données halieutiques.

3.3.3 Variable régionalisée, échelle et intégrale

Variable régionalisée et déterminisme

Une **variable régionalisée** est une fonction mathématique qui vaut en un point donné la valeur de la variable d'intérêt. C'est une fonction de nature déterministe qui constitue une version analytique de la réalité physique. On ne fait donc plus d'hypothèse à propos du phénomène et du modèle. On ne se pose plus la question de savoir si telle ou telle valeur aurait pu être observée ailleurs. On considère la réalité pour ce quelle est.

En plus de ne nécessiter à ce stade aucune hypothèse sur le phénomène, le caractère déterministe de ce modèle est totalement compatible avec d'éventuelles localisations préférentielles de densités.

Echelle

Contrairement à l'approche intrinsèque, on considère *en transitif*, qu'une variable $z(x)$ ne peut pas être considérée indépendamment de son champ (Matheron, 1965). Pour toutes les situations où la répartition spatiale dépend du champ, l'échelle pertinente de l'analyse est la totalité de la régionalisation, c'est-à-dire l'ensemble du champ de la variable régionalisée. On donne le nom de transitive à ce pan de la géostatistique par ce que précisément on souhaite tenir compte des zones de transition entre l'intérieur et l'extérieur de la régionalisation où apparaissent souvent les effets de bordure. On voit donc que les techniques transitives concernent une échelle différente et par là plus globale que celle des techniques intrinsèques.

Il s'agit par conséquent d'une méthodologie qui par son caractère global est encline à rendre compte de phénomènes à effets de bord (appauvrissement en bordure).

Intégrale

A l'extérieur de son champ, la variable régionalisée est supposée identiquement nulle. Pour cette raison, la géostatistique transitive utilise préférentiellement des intégrales d'espaces à des moyennes. On retrouve ici les mêmes arguments que ceux déjà évoqués à propos du traitement des valeurs nulles, c'est-à-dire qu'une somme est inchangée par une addition de valeurs nulles. C'est pour cela que les outils de base du transitif sont l'abondance et le covariogramme.

Cependant, l'estimation d'une somme pose de sérieuses difficultés et n'est réellement confortable que dans le cas d'un réseau d'échantillonnage régulier. C'est d'ailleurs dans ce contexte uniquement que la géostatistique transitive s'est développée. L'avantage qu'on en retire en retour est toutefois indéniable.

En effet, la régularité (stationnarité) du plan d'échantillonnage autorise une probabilisation élégante, efficace et pauvre en hypothèses. L'idée directrice est très simple. Elle consiste à imaginer que le réseau d'information se déplace dans l'espace. En particulier, *l'un, quelconque, des points expérimentaux choisi comme origine du réseau, est supposé se déplacer dans le rectangle défini par la maille, en entraînant avec lui l'ensemble du réseau. Ce déplacement définit l'ensemble des états virtuels nécessaires à la probabilisation en transitif.* Ainsi, en faisant porter la probabilisation sur la structure du réseau d'information plutôt que sur la variable régionalisée, on évite de "forcer" les données.

Conclusion

La réflexion sur les concepts d'une part et sur le choix d'un modèle d'autre part est donc convergente. En recourant aux techniques transitives on répond au souhait de travailler en somme plutôt qu'en moyenne.

Cependant, la géostatistique transitive élaborée initialement pour l'estimation globale à maille régulière, est peu utilisée dans la pratique et donc peu développée. Or ce cadre méthodologique convient bien aux traitements des données halieutiques.

Il apparaît donc souhaitable, dans le cadre de nos problématiques halieutiques, de commencer par "remettre au goût du jour" l'utilisation de la géostatistique transitive et de l'étoffer de quelques développements nouveaux comme celui des statistiques par individu. Cet ensemble est présenté dans la deuxième partie "Géostatistique transitive. Le point de vue de l'individu".

Deuxième partie

Statistiques par individu et géostatistique transitive.

Chapitre 1

Loi d'un individu aléatoire

Sommaire

*Dans le **paragraphe 1** on propose une interprétation probabiliste des variables régionalisées relativement à l'abondance totale. En considérant un individu aléatoire, c'est-à-dire un individu tiré au hasard dans une population, la variable régionalisée $z(x)/Q$ tient lieu de densité de probabilité pour la variable aléatoire "position d'un individu aléatoire". Les notions de moyennes et de variances, qui ne reposent plus sur la définition d'un champ, s'interprètent alors respectivement comme des centres de gravité et des inerties.*

*Le **paragraphe 2** traite de l'estimation d'abondances totales. Je rappelle tout d'abord l'estimation à maille régulière par la géostatistique transitive telle que développée par Matheron. On aborde ensuite l'estimation lorsque le plan d'échantillonnage est irrégulier (surfaces d'influence).*

*Le **paragraphe 3**, très court, concerne l'estimation du centre de gravité et de l'inertie. A maille irrégulière, on utilise, comme pour l'abondance, une pondération par les surfaces d'influence.*

*Dans le **paragraphe 4** je présente une façon très simple d'utiliser les centres de gravités et inerties pour décrire des séries historiques de cartes. Cette application intéressante du concept d'individu aléatoire sera illustrée dans la partie du document relative aux œufs de maquereau.*

Law of a random individual

Abstract

*In **paragraph 1** we give a probabilistic interpretation of a regionalized variable when it is divided by the total abundance of a population. In this respect we consider a random*

individual (noted I), that is a individual taken at random in the whole population. The location of such an individual (x_I) is then interpreted as a random vector whose probability density function is nothing but the relative regionalized variable $z(x)/Q$. In this approach, the mean and variance get the meaning of a center of gravity and an inertia.

*In the **paragraph 2** we deal with the estimation of global abundances. First of all, I will remind the methodology developed by Matheron (1965) for the global estimation with a regular sampling pattern (transitive geostatistics). Then we will see how to cope with irregular patterns (surfaces of influence).*

*The **paragraph 3** is very short as it is a simple application of the previous results to the estimations of centers of gravity and inertia in both case of regular and irregular sampling patterns.*

*In **paragraph 4** I indicate how the simple statistics per individual such as the center of gravity and the inertia give the opportunity to describe with few figures large series of surveys.*

La géostatistique transitive a été développée par Matheron en 1965 dans le but essentiel de proposer une variance d'estimation pour une estimation globale d'abondance à partir d'un réseau d'information à maille régulière. L'intérêt de cette méthode provient de ce qu'elle permet de tenir compte de la structure spatiale du phénomène régionalisé tout en faisant un nombre très limité d'hypothèses.

En s'inspirant de l'esprit de la démarche transitive, nous allons voir tout l'intérêt qu'il y a, en écologie et plus généralement pour toute variable de type "densité d'individu", à ne plus considérer la variable régionalisée $z(x)$ mais plutôt la variable régionalisée relative à l'abondance totale $z(x)/Q$. Un certain nombre de concepts trouvent alors une interprétation physique forte. De ce point de vue, **les développements qui suivent** ne présentent aucun sens pour des variables comme la température de surface ou la bathymétrie. Ils **ne sont réellement pertinents que pour des variables de type "densité" ou "concentration"**.

Dans cette partie, j'utiliserai abondamment le terme de population. J'entend lui donner le sens d'un groupe d'individus délimité dans l'espace, c'est-à-dire d'un ensemble statistique, d'un peuplement, d'un groupe, ou encore d'un stock. En aucun cas, je ne cherche à lui donner l'acception reconnue par les halieutes (Laurec et Le guen, 1981).

1.1 Définitions

1.1.1 Individu aléatoire

Imaginons que tous les individus d'une population soient numérotés de 1 à Q , où Q représente l'abondance totale. Le numéro I d'un **individu tiré au hasard** dans la population constitue alors une variable aléatoire uniforme dans l'intervalle $[1, Q]$:

$$I \stackrel{\mathcal{L}}{\equiv} \mathcal{U}[1; Q]$$

C'est donc un modèle constitutif dit d'individu aléatoire que nous proposons ici. Au modèle global de variable régionalisée est associée une variable aléatoire "centrée sur l'individu". C'est

précisément ce qui rend cette approche pertinente en écologie halieutique. Je prendrai l'habitude d'utiliser le terme "d'individu aléatoire" pour désigner un individu tiré au hasard dans une population.

1.1.2 Loi de x_I

Soit $z(x)$ la variable régionalisée associée à une population donnée. La position x_I d'un individu aléatoire doit donc être interprétée comme un vecteur aléatoire ayant autant de composantes que l'espace a de dimensions. Partant de la constatation triviale qu'un individu aléatoire a d'autant plus de chance de se trouver en un point donné que la densité en ce point est importante, la "probabilité" qu'un individu aléatoire se trouve en x est donnée par $z(x)/Q$. Plus exactement, la variable régionalisée $z(x)/Q$ doit être interprétée comme la densité de probabilité de x_I , que l'on notera $f_{x_I}(x)$:

$$f_{x_I}(x) = \frac{z(x)}{Q}$$

1.1.3 Centre de gravité et inertie

La "vraie" fonction $z(x)/Q$, c'est-à-dire celle définie en tout point de l'espace, peut être très riche et très compliquée. Au même titre qu'on résume une loi (histogramme) par la donnée d'une espérance (moyenne) et d'une variance, il est légitime de synthétiser la totalité de l'information contenue dans la densité de probabilité par un paramètre de centrage et un critère de dispersion.

Centre de gravité

L'espérance de x_I , qui correspond à la position moyenne d'un individu aléatoire, est par définition l'intégrale de l'ensemble des localisations possibles pondérées par leur (densité de) probabilité:

$$E(x_I) = \bar{x}_I = \int x \cdot f_{x_I}(x) dx = \int x \frac{z(x)}{Q} dx = \frac{\int x \cdot z(x) dx}{\int z(x) dx}$$

Ainsi, l'espérance a le sens du **centre de gravité** de la localisation des individus de la population (i.e. centre de gravité de la population) puisque chaque localisation a un poids proportionnel au nombre d'individus qui s'y trouvent. L'intégrale est étendue à tout l'espace sans référence à aucun champ. En particulier, la contribution d'une zone, ou d'un point, où la densité est nulle ne modifie pas le centre de gravité géographique de la population. Ainsi, **grâce à la pondération par $z(x)$, on dispose d'un paramètre de centrage, c'est-à-dire d'une moyenne, indépendamment de la définition du champ.**

Inertie

De façon équivalente, la variance de x_I qui s'écrit:

$$Var(x_I) = \sigma_{x_I}^2 = \int (x - \bar{x}_I)^2 \cdot f_{x_I}(x) dx = \frac{\int (x - \bar{x}_I)^2 \cdot z(x) dx}{\int z(x) dx}$$

a le sens d'une **inertie** et présente les mêmes avantages en terme de champ et de pondération que le centre de gravité.

Ce qui vient d'être fait avec la position des individus d'une population va progressivement devenir un principe général où le **couple moyenne / variance** d'une variable quelconque ayant un sens pour les individus, **sera remplacé par le couple formé du centre de gravité et de l'inertie** de cette variable.

1.1.4 Commentaires

Approche individu centrée

Y-a-t-il un rapport entre la loi d'un individu aléatoire, c'est-à-dire un modèle "centré sur l'individu" et les modèles dits "individu-centrés" d'après l'appellation anglaise Individual based models (IBM)? Les modèles individu-centrés visent à modéliser une population à partir des individus qui la composent (DeAngelis et Gross, 1992). Le plus souvent l'utilisation de ces modèles visent à partir des individus et à simuler les comportements qui émergent à l'échelle de la population. A l'inverse les statistiques par individu sont susceptibles de fournir des éléments statistiques concernant les individus à partir de l'état réel d'une population. En particulier, si cette approche permet de dire avec quelle probabilité un individu aléatoire se trouve à une certaine température, alors la règle correspondant au lien entre un individu et la température dans le modèle individu-centré peut être aisément probabilisée, avec, de surcroît, un contrôle expérimental du paramétrage de cette règle.

La capturabilité

La notion d'individu aléatoire trouve ici sa place dans un cadre méthodologique assez nouveau en halieutique. Il n'en reste pas moins vrai qu'elle est utilisée dans les concepts de base de la dynamique des populations. Ainsi, dans un article de 1993 présenté au premier forum de l'Association Française d'Halieumétrie, Gascuel définit "*la capturabilité d'un stock comme la probabilité d'être capturé pendant une unité de temps, par une unité d'effort de pêche nominal, pour un poisson pris au hasard dans le stock*" (Gascuel, 1993). Toutefois, la notion d'individu aléatoire est utilisée ici au sens de la statistique classique, c'est-à-dire sans prendre en compte la régionalisation des individus de la population. Mais rien n'empêche, sur la base de cette même définition, d'envisager une version géostatistique de la capturabilité, et par suite, de la dynamique des populations.

1.2 Estimation d'abondances totales

Les statistiques par individu nécessitent, par construction, la connaissance de l'abondance totale de la population. Je propose donc dans ce paragraphe de rappeler quelques éléments de l'estimation d'une telle quantité selon que le plan d'échantillonnage est régulier ou non. Ce sera l'occasion de donner la définition du covariogramme qui apparaît naturellement dans le calcul de la variance d'estimation à maille régulière.

1.2.1 Echantillonnage à maille régulière ou systématique

Ce paragraphe reprend brièvement les travaux de Matheron (1965). Pour faciliter les notations, je me place dans le cadre d'un échantillonnage à une seule dimension. La généralisation à des espaces de dimensions supérieures ne pose pas de problème.

L'estimateur

Le réseau d'information est une grille régulière de maille a d'origine x_0 . On suppose que la densité est nulle en dehors de la grille d'échantillonnage. Les N échantillons sont notés $z(x_0 + ka)$ avec $k \in [0, N-1]$ et l'estimateur de l'abondance totale Q est simplement la somme des échantillons multipliée par la maille :

$$Q^* = a \sum_{k=0}^{N-1} z(x_0 + ka)$$

Probabilisation

Pour probabiliser l'estimateur on imagine que l'origine de la grille se déplace dans le segment définissant la maille entraînant avec elle la totalité du réseau d'échantillonnage. Ce déplacement définit l'ensemble des états virtuels (i.e. possibles) et il suffit ensuite de considérer le point x_0 comme un point aléatoire uniformément distribué dans le segment de la maille pour achever la définition du modèle (Matheron, 1978). Cette hypothèse, qui ne concerne que l'implantation de l'origine de la grille d'observation, est tout à fait réaliste. En effet, si la forme et la taille de la maille sont en général définies de façon assez précise, la position du début des mesures est, elle, choisie "aléatoirement" en cela qu'il n'y a, dans la plupart des cas, aucune raison objective de "commencer" l'échantillonnage précisément là où il a commencé.

Ainsi, on a constitué l'estimateur en variable aléatoire. L'origine de la maille x_0 est considérée comme la réalisation d'une variable aléatoire X_0 de loi uniforme dans $[0, a]$. Et, l'estimateur devient lui aussi une variable aléatoire notée $Q^*(X_0)$:

$$Q^*(X_0) = a \sum_{k=0}^{N-1} z(X_0 + ka)$$

Biais

Dans ce modèle, on montre facilement que l'estimateur est **sans biais** (Matheron, 1965):

$$\begin{aligned} E(Q^*(X_0)) &= \int_0^a Q^*(x) \frac{dx}{a} = \int_0^a a \sum_{k=0}^{N-1} z(x + ka) \frac{dx}{a} \\ &= \sum_{k=0}^{N-1} \int_0^a z(x + ka) dx \\ &= \int_0^a z(x) dx \\ &= Q \end{aligned}$$



Variance d'estimation

Les échantillons étant répartis sur une grille régulière, il devient impossible de les supposer indépendants. Mais, sans plus d'hypothèse, le développement exact de la variance d'estimation est cependant possible. Les indications concernant le calcul de la variance d'estimation sont fournis

plus loin dans le document (Cf. § II.3.3). On retiendra seulement que le résultat final fait intervenir la fonction

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

baptisée **covariogramme** par Matheron (1965). La formule de la variance, et son interprétation, sont discutées plus loin (§ II.3.3.1)

La géostatistique transitive permet donc de tirer efficacement parti de la régularité du plan d'échantillonnage. En somme, la stationnarité, qui porte en géostatistique intrinsèque sur les caractéristiques physiques de la variable régionalisée, provient en géostatistique transitive des caractéristiques du réseau d'informations.

Il existe cependant beaucoup de cas où l'échantillonnage est irrégulier. Dans certains cas, l'échantillonnage, bien qu'irrégulier, n'est cependant pas incompatible avec une approche transitive de l'estimation globale pour autant qu'on connaisse le processus qui a généré les échantillons. Le cas d'un processus de Poisson à densité régionalisée est à cet égard très intéressant.

1.2.2 Deux exemples d'échantillonnages par processus ponctuels

Processus de Poisson à densité régionalisée

- Définitions.

Un ensemble de points peut être vu comme la réalisation d'un processus stochastique ponctuel. L'exemple le plus connu est le processus de Poisson (\mathcal{P}_θ). *Cette popularité n'est au demeurant guère imméritée car un processus de Poisson exprime exactement l'idée intuitive de points au hasard* (Lantuéjoul, 1993). Ce processus ne dépend que d'un seul paramètre θ constant dans l'espace qui fixe le nombre moyen de points par unité de "volume" (densité de points). Le paramètre θ est en quelque sorte une mesure de l'effort d'échantillonnage dans le cas aléatoire pur.

Cependant, la densité de points d'un processus ponctuel peut être régionalisée. Elle est notée dans ce cas $\theta(x)$ et on parle de processus ponctuel à densité régionalisée (Lantuéjoul, 1993). Ce genre de processus permet d'augmenter l'effort d'échantillonnage dans certaines sous-zones et correspond à la notion d'échantillonnage préférentiel. On notera bien que les points d'un tel processus conservent un caractère aléatoire.

L'intérêt en écologie halieutique est évident puisque cette approche permet de tenir compte de connaissances biologiques disponibles sur la population, mais également d'éventuelles variables conditionnantes souvent facilement accessibles. Ainsi, pour étudier une population dont la densité augmenterait avec la bathymétrie, il conviendrait d'accroître régulièrement la pression d'échantillonnage en allant vers les zones à forte profondeur, c'est-à-dire vers les zones riches. Ceci peut être fait en prenant comme densité régionalisée une fonction appropriée de la bathymétrie.

Les résultats qui suivent concernent le cas particulier de l'estimation d'abondance quand l'échantillonnage est issu d'un processus de Poisson à densité régionalisée. Ils proviennent d'un document qui m'a été fourni par Christian Lantuéjoul. Ils ne sont malheureusement pas publiés. Je me permets toutefois d'en rendre compte ici avec son accord.

- Estimateur.

Pour tenir compte de la régionalisation de la pression d'échantillonnage dans l'estimation de l'abondance, on suggère de pondérer la valeur observée en chaque point du processus $z(\alpha)$ par la densité du processus ponctuel $\theta(\alpha)$ en ces mêmes points :

$$Q^* = \sum_{\alpha \in \mathcal{P}_\theta} \frac{z(\alpha)}{\theta(\alpha)}$$

Le caractère aléatoire du processus ponctuel utilisé pour implanter l'échantillonnage confère à l'estimateur Q^* le statut de variable aléatoire. Son espérance nous fournit donc le biais.

- Biais et variance.

En utilisant la transformée de Fourier de la variable aléatoire Q^* et en la mettant sous une forme un peu particulière, on montre assez facilement que Q^* est un estimateur sans biais de Q :

$$E(Q^*) = \int z(x) dx$$

et, que la variance d'estimation $\sigma^2 = E(Q^{*2}) - Q^2$ vaut :

$$\sigma^2 = \int \frac{z(x)^2}{\theta(x)} dx$$

Par conséquent, la variance n'est pas accessible directement. Pour l'obtenir, il faut estimer la somme pondérée du carré de la variable régionalisée, qui ne se déduit malheureusement pas directement de Q^* . Mais on remarquera que seuls les produits $z(x)^2$ sont concernés. La variance est à ce titre de "type statistique" par opposition à une formulation de "type géostatistique" où les produits $z(x)z(x+h)$ remplaceraient $z(x)^2$. En utilisant un processus de Poisson on assure, par construction, l'indépendance mutuelles des échantillons, même si la densité du processus ponctuel est régionalisée.

- Commentaires.

J'aimerais attirer l'attention sur le fait que, bien que l'échantillonnage soit préférentiel, il a été possible de fournir une estimation non biaisée de la quantité totale. Ce résultat me paraît important et pourrait permettre d'améliorer l'efficacité de certaines campagnes scientifiques tout en maintenant le nombre total d'échantillons.

Cependant, on ne peut prétendre à cette amélioration que si on contrôle l'élaboration du plan d'échantillonnage. Il est en effet difficile d'inférer un modèle de processus ponctuel à partir de points déjà implantés dans un domaine. La démarche qui vient d'être présentée n'est donc pertinente que si l'implantation des échantillons est fidèle à la réalisation d'un processus ponctuel connu. C'est d'ailleurs parce que nous contrôlions la construction du plan d'échantillonnage que la probabilisation, qui porte là encore sur le réseau d'échantillonnage, a été possible.

Une précaution s'impose enfin. Les résultats (non biais et variance) sont connus **dans le modèle**. C'est-à-dire qu'ils s'entendent au sens d'une valeur moyenne qu'il serait possible de calculer après plusieurs échantillonnages réalisés sur la base de plusieurs simulations d'un même processus ponctuel d'échantillonnage. Or en pratique on ne dispose que d'une seule réalisation du processus et rien n'indique que les propriétés élaborées dans le modèle soient vérifiées pour cette réalisation particulière.

Une remarque en passant sur les échantillonnages aléatoires stratifiés

Un autre cas intéressant, souvent utilisé en géostatistique mais peu répandu en halieutique me semble-t-il (exemple possible : les campagnes International Young Fish Surveys-IYFS), concerne les mailles aléatoires stratifiées. Il s'agit de **tirer au hasard un échantillon dans chacune**

des cases d'une grille régulière. Cette acception est différente de celle envisagée en statistique "classique". Dans ce dernier cas, les strates correspondent à des zones supposées homogènes de la régionalisation et plusieurs échantillons y sont répartis de façon uniforme. Toujours est-il que dans la version géostatistique d'un échantillonnage aléatoire stratifié, on montre que la variance de Q^* est (Matheron, 1978, p 137) :

$$\text{Var}(Q^*) = a \left(g(0) - \overline{g(a)} \right)$$

où $\overline{g(a)}$ est la valeur moyenne du covariogramme lorsque les deux points x et $x+h$ décrivent le rectangle a définissant la maille. On notera au passage qu'on retrouve ici une expression géostatistique de la variance.

Conclusion

Au travers de ces différents cas d'estimation globale (maille régulière et processus ponctuel), il est apparu que *c'est la stationnarité du réseau qui nous a permis, en l'absence de toute hypothèse portant sur la variable régionalisée, de bénéficier des circonstances très avantageuses que l'on croit généralement liées à la stationnarité du phénomène lui-même, ou de la variable régionalisée qui le représente.* En somme, *la stationnarité peut souvent être introduite, non pas à titre d'hypothèse relative à la réalité physique, mais simplement comme une caractéristique de la méthode d'estimation choisie par nous* (Matheron, 1978, p 135).

Dans le cas d'un échantillonnage quelconque, il devient difficile de probabiliser l'estimateur.

1.2.3 Echantillonnage quelconque

Il est assez intuitif pour estimer une abondance à partir d'échantillons répartis de façon irrégulière dans l'espace, de chercher à pondérer chaque valeur par son importance relative par rapport au champ total. Dans le cas précédent, cette pondération provenait de la densité du processus ponctuel.

Je me place ici dans le cas différent où je ne connais pas le processus ponctuel. Ce genre de situations se rencontre assez souvent dans la pratique. Qu'il s'agisse du non respect du plan d'échantillonnage (tempête, avaries,...), qu'il s'agisse d'un sur-échantillonnage en cours de campagne orienté par les valeurs de densités importantes ou qu'il s'agisse tout simplement d'une ignorance du mode de construction de l'échantillonnage, il arrive que l'échantillonnage soit irrégulier sans plus de précision. Il convient alors de calculer la surface d'influence S_i de chaque échantillon z_i .

Surfaces d'influence

Je présente ici deux méthodes pour évaluer des surfaces d'influence. Une méthode rigoureuse dite des polygones d'influence, et une méthode approchée mais moins difficile à mettre en œuvre que la précédente.

- Une méthode rigoureuse : polygones d'influence.

Le pavage par des polygones d'influence est basé sur le fait que la médiatrice d'un segment est le lieu des points équidistants des extrémités de ce segment. Ainsi, en procédant de proche en proche, chaque point est comparé à tous les autres de sorte qu'on parvient à tracer autour de

chaque échantillon un polygone d'influence délimitant la zone de l'espace contenant l'ensemble des points qui lui sont le plus proches.

Outre les difficultés algorithmiques, cette méthode n'est réellement appropriée, d'un point de vue informatique, que pour réaliser le pavage d'un domaine fermé et convexe. Si tel n'est pas le cas, comme par exemple dans l'étude de la distribution des densités de merlu (§III), cela nous oblige, soit à définir le champ par plusieurs polygones convexes, soit à fixer un seuil de distance maximum entre un échantillon et les sommets de son polygone d'influence. Il faut par ailleurs s'assurer que les polygones d'influence ne chevauchent pas des zones stériles comme les continents. Pour cela, il est nécessaire de disposer d'un ou plusieurs polygones convexes qui longent grossièrement les côtes continentales.

- Une méthode approchée : migration vers une grille très fine.

Une solution plus commode d'un point de vue informatique (algorithmique) mais moins rigoureuse consiste à créer une grille régulière de maille très petite débordant le domaine échantillonné. Chaque nœud de cette grille est ensuite informé par l'échantillon le plus proche dans la mesure où la distance qui sépare l'échantillon du nœud n'excède pas un seuil fixé. Ainsi, en multipliant le nombre de nœuds occupés par chaque échantillon par la surface de la maille on obtient une valeur approchée des surfaces d'influence.

Estimateur

Connaissant les surfaces d'influence, la quantité totale est estimée par :

$$Q^* = \sum_{i=1}^N S_i \cdot z_i$$

qui relève de la même logique que l'estimateur précédent $Q^* = \sum \frac{z(x)}{\theta(x)}$. En effet, lorsque la densité d'un processus ponctuel augmente, la surface d'influence des échantillons correspondants diminue corrélativement. Ainsi, $1/\theta(x)$ et S_i évoluent dans le même sens.

Quelle probabilisation?

Pour aller plus loin et proposer par exemple le calcul du biais de l'estimateur, il faut tout d'abord définir la variable aléatoire dont on parle. Or, pour l'instant nous n'avons manipulé que des quantités déterministes. Peut-on, à l'instar des exemples précédents, "randomiser" le réseau d'échantillons?

Pour cela revenons sur le déplacement d'une grille régulière tel que réalisé précédemment pour l'estimation globale. Le non biais provient dans ce cas, de ce que le déplacement de l'ensemble de la grille permet de couvrir la totalité du champ. Rien de tel n'est envisageable à maille irrégulière.

Dès lors qu'on renonce à randomiser le réseau d'échantillonnage, seule une estimation globale est fournie, sans variance ni biais.

1.3 Application : estimation du centre de gravité et de l'inertie

Ce paragraphe est une application directe des notions exposées précédemment.

Dans le cas d'un échantillonnage régulier, on procède simplement par une moyenne des positions des échantillons pondérée par les densités observées.

Dans le cas d'un échantillonnage irrégulier, une pondération supplémentaire par les surfaces d'influence est réalisée.

x_I	Grille régulière	Grille irrégulière
\bar{x}_I^*	$\frac{\sum_i x_i z_i}{\sum_i z_i}$	$\frac{\sum_i x_i z_i S_i}{\sum_i z_i S_i}$
$\sigma_{x_I}^{2*}$	$\frac{\sum_i (x_i - \bar{x}_I^*)^2 z_i}{\sum_i z_i}$	$\frac{\sum_i (x_i - \bar{x}_I^*)^2 z_i S_i}{\sum_i z_i S_i}$

TAB. 1.1 - Estimations du centre de gravité et de l'inertie. Échantillonnage à maille régulière et à maille irrégulière. A chaque échantillon z_i est associée une surface d'influence S_i .

1.4 Descriptions de séries historiques de cartes

La plupart des stocks exploités sont l'objet d'un suivi annuel qui consiste en une ou plusieurs campagnes océanographiques. Ainsi, les différents instituts de recherche possèdent des séries importantes et très précieuses de cartes (séries historiques), les plus typiques étant des représentations proportionnelles des valeurs observées.

Cependant, l'abondance de ces cartes peut en rendre l'interprétation très difficile. Il convient alors de chercher le moyen de synthétiser l'essentiel de l'information qu'elles contiennent en un petit nombre de graphes. Le centre de gravité et l'inertie d'un individu aléatoire sont à cet égard de bons outils.

1.4.1 Déplacement moyen d'une population

La chose la plus simple est de suivre, de campagne en campagne, la position du centre de gravité de la population. Des déplacements significatifs de ce point moyen peuvent s'interpréter alors comme des déplacements ou des dérives de la population étudiée.

Cependant, ce n'est là qu'une description très sommaire, qu'on peut enrichir par la représentation graphique de la dispersion.

1.4.2 Axes principaux de la dispersion

Principe

La dispersion des individus d'une population autour de leur centre de gravité n'a aucune raison d'être identique dans toutes les directions de l'espace (isotrope). Il existe en général des directions privilégiées d'allongement ou au contraire de concentration (agrégation) conduisant la dispersion à être anisotrope.

Pour tenir compte d'éventuelles anisotropies on propose de décomposer l'inertie d'une population selon ses axes principaux. Le premier axe représente la direction d'inertie maximum, c'est-à-dire l'axe autour duquel le moment d'inertie est minimum. Rechercher les axes principaux revient donc à

faire une Analyse en Composante Principale (ACP) des coordonnées des individus de la population, c'est-à-dire une ACP des coordonnées des échantillons pondérées par les densités observées.

Les axes principaux correspondent par conséquent aux vecteurs propres de la matrice $[X]^t[Z][X]$ où $[X]$ est la matrice des coordonnées et $[Z]$ une matrice diagonale constituée des densités (Escoffier et Pagès, 1990). En prenant soin de centrer les coordonnées préalablement aux analyses, les valeurs propres indiquent l'importance relative de chaque axe par le pourcentage de l'inertie totale expliquée par cet axe.

Puisqu'on s'est fixé de représenter graphiquement l'inertie en même temps que le centre de gravité, il est souhaitable d'utiliser des **quantités homogènes à des distances**. C'est la raison pour laquelle on préférera se servir de la racine carrée de l'inertie plutôt que de l'inertie elle-même.

Système de représentation

En conséquence, on propose de représenter une série de cartes ou campagnes de la façon suivante. On supposera connus pour chacune d'elles, l'abondance totale, le centre de gravité, l'inertie et les axes principaux d'inertie.

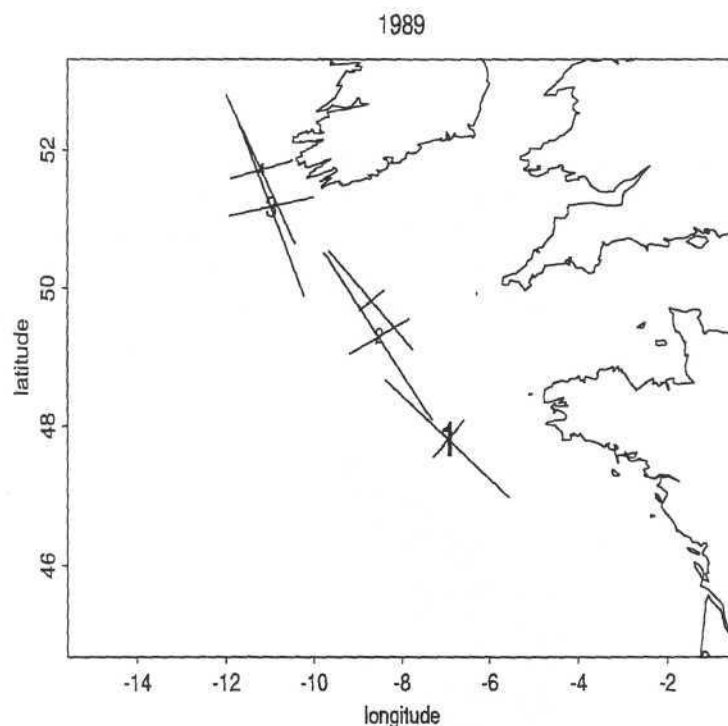


FIG. 1.1 - Exemple de représentation d'une série de 5 campagnes océanographiques. Cas de la distribution des œufs de maquereau en 1989 au cours de la saison de ponte.

Sur une carte géographique (Fig. 1.1), la position du centre de gravité de chaque campagne est indiquée par son numéro; la taille du numéro pouvant être d'autant plus importante que la quantité totale estimée pour cette campagne est importante. Centrés sur ce centre de gravité, des segments de droites représentent les axes principaux d'inertie. Leurs directions correspondent à celles des axes principaux et leurs tailles sont égales aux racines carrées des pourcentages d'inertie expliquée

par ces axes.

1.4.3 Remarques

Remarques bibliographiques

Des travaux récents, mais, à ma connaissance peu nombreux, utilisent ces notions. Dans une communication récente Ottersen *et al.* (1996) s'intéressent aux régions occupées sur plusieurs années par les centres de masse des différentes classes d'âge du stock de morue de l'océan Arctique. Ils n'utilisent par ailleurs pas de critère inertiel. Cette démarche les conduit également à considérer la "température ambiante" qui correspond à ce que je présente dans le prochain chapitre sous le terme de centre de gravité des température.

Perry et Smith (1994) ont adopté une démarche équivalente pour caractériser l'association entre une densité de poisson et une variable hydrographique à l'aide d'histogramme de températures pondérées par les captures ("catch-weighted cumulative distribution functions")

De leur côté, Watanabe *et al.* (1996) utilisent le même type de pondération par la densités pour décrire le déplacement vers le large des aires de ponte du pilchard du Japon (*Sardinops melanostictus*).

Interprétations

De même que la donnée d'une moyenne et d'une variance ne permet pas de rendre compte de la totalité d'un histogramme, l'utilisation du centre de gravité et de l'inertie d'un individu aléatoire ne décrit que sommairement une régionalisation. Le parallèle s'étend également au fait que la moyenne n'est un critère pertinent que dans le cas d'une distribution unimodale. L'inertie ne permet pas de savoir si la population était répartie en deux ou plusieurs sous-populations.

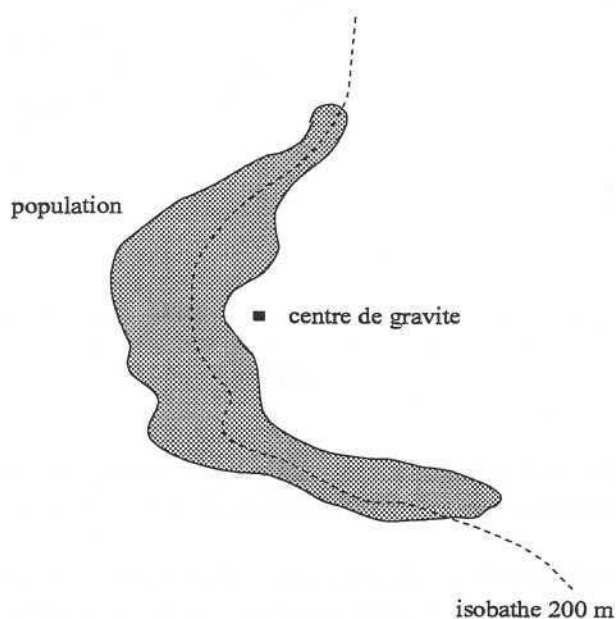


FIG. 1.2 - Exemple de piège d'interprétation. Différence entre la position géographique moyenne et la profondeur moyenne d'un individu aléatoire.

Sur l'exemple de la figure 1.1 j'ai intentionnellement omis d'indiquer la bathymétrie (par une courbe de niveau par exemple). La raison en est la suivante. Les centres de gravité représentent la position moyenne des œufs de maquereau dans le repère géographique longitude x latitude. Ils ne permettent cependant pas de savoir quelle est la bathymétrie moyenne à laquelle se trouve un œuf pris au hasard dans la population. On pourrait, en se basant uniquement sur la carte, penser que cette profondeur moyenne est inférieure à 200 m puisque les centres de gravité sont localisés sur le plateau continental. Mais ce n'est là que la conséquence des formes respectives de la distribution des œufs et des courbes isobathes (Fig. 1.2).

Chapitre 2

Loi des caractéristiques d'un individu aléatoire.

Sommaire

Ce chapitre est une généralisation du chapitre précédent. Il étend la notion de statistiques par individu à l'ensemble des caractéristiques qui peuvent être associées à un individu aléatoire comme par exemple la température de surface.

*Le **paragraphe 1** permet de définir le centre de gravité et l'inertie des températures par individu. Ces deux notions ne sont que les deux premiers moments de la loi des températures de surface par individu dont l'histogramme des abondances par classe est une estimation. Cette approche permet de décrire le lien entre densité et température du point de vue d'un individu. Elle est donc comparée aux régressions qui fournissent une description de ce lien à partir de la température.*

*Le **paragraphe 2**, assez court, s'intéresse à la prise en compte de plusieurs variables explicatives. Des raisons pratiques imposent de ne considérer que 2 ou 3 co-variables au maximum.*

*Le **paragraphe 3** est un paragraphe **important** de cette thèse. En se référant aux caractéristiques générales des structures spatiales des variables en jeu (densité de poissons et température de surface), on indique que les optima des modèles de régressions ou des densités de probabilité par individu peuvent être des artefacts, n'indiquant donc ni *preferendum* ni conditionnement. Sur la base de ces réflexions, un critère géostatistique, susceptible de diagnostiquer l'existence d'un conditionnement, est construit. Ce concept, par ailleurs très simple, pose quelques problèmes pratiques sérieux discutés par la suite.*

Characteristics of a random individual

Abstract

This chapter is nothing but the generalization of the previous one. We extend the principle of statistics per individual to any possible characteristics that can be associated to an individual of a population like for instance the sea surface temperature.

The paragraph 1 deals first with the center of gravity and the inertia of the temperature per individual. Then, the histogram of the abundances per class is interpreted as the estimation of the probability density function of the temperature per individual. Thus, the use of random individual and statistics per individual leads to a statistical description from the point of view of the individuals of a population. In this respect a traditional approach would give the temperature point of view.

The paragraph 2 is very short. We simply consider 2 or 3 variables rather than only one. Theoretically we could enlarge this approach to more than 3 co-variates but we might then face bad statistical conditions (not enough samples) for the estimations of abundances.

The paragraph 3 is an important section of this thesis. By referring to the spatial structures of the variables of interest (fish density and sea surface temperature) we point out that the optimum of the models of the regressions or the models of p.d.f. per individual might be an artefact. It is then no more possible to use such concepts for the analysis of the conditioning effect of the temperature on fish distribution, even if the p.d.f. is still an appropriate tool for the statistical description of correlations. Based on these reflexions, we propose a geostatistical criterion which enable to study the question of preferendum in a spatial context. This tool is very simple, but not so easy to use in practice. The third part of this paragraph is devoted to the analysis of these difficulties.

La mise en garde qui conclut le chapitre précédent m'amène à définir dans ce chapitre les statistiques concernant une ou plusieurs caractéristiques qui peuvent être associées à un individu pris au hasard dans une population. Il s'agit en fait de replacer des concepts déjà rencontrés antérieurement dans ce document, comme l'abondance de poissons par classe de température de surface, dans le cadre méthodologique plus vaste des modèles d'individu aléatoire. Pour rendre l'exposé plus parlant, on prendra dans le premier paragraphe l'exemple de l'analyse de la température de surface de l'eau dans laquelle se trouve un individu aléatoire. J'aborderai ensuite brièvement l'analyse d'un couple de variables par individu.

Mais, si l'approche par individu se révèle bien adaptée pour décrire statistiquement le lien entre une densité de poissons et une autre variable (ou encore un couple de variables), elle est impuissante à déterminer l'existence ou non d'un conditionnement. Après avoir analysée la distinction qui s'impose entre un optimum et un preferendum par des considérations touchant aux types de régionalisations des variables en jeu, je présenterai dans le dernier paragraphe un concept de géostatistique par individu qui devrait permettre une meilleure analyse des preferendum dans un contexte spatial.

2.1 Loi d'une variable par individu : exemple de la température de surface par individu.

2.1.1 Formulation générale

D'un point de vue mathématique, x_I (la position d'un individu aléatoire) est un vecteur aléatoire de densité de probabilité connue $f_{x_I}(x)$. L'espérance et la variance de toute fonction de ce vecteur

aléatoire $h(x_I)$, s'interprètent donc encore comme des centres de gravité et des inerties. Elles sont données par les expressions suivantes :

$$\begin{aligned} E(h(x_I)) = \bar{h}_I &= \int h(x) f_{x_I}(x) dx \\ &= \int h(x) \frac{z(x)}{Q} dx \\ &= \frac{\int h(x) z(x) dx}{\int z(x) dx} \end{aligned}$$

et

$$\begin{aligned} Var(h(x_I)) = \sigma_{h_I}^2 &= \int (h(x) - \bar{h}_I)^2 f_{x_I}(x) dx \\ &= \int (h(x) - \bar{h}_I)^2 \frac{z(x)}{Q} dx \\ &= \frac{\int (h(x) - \bar{h}_I)^2 z(x) dx}{\int z(x) dx} \end{aligned}$$

2.1.2 Centre de gravité et inertie des températures par individu

Prenons le cas particulier de la température de surface. Ainsi, pour chaque individu tiré au hasard dans la population, on ne regarde plus uniquement sa position x_I , mais la valeur de la température de surface en ce point. Cette fonction de x_I est une nouvelle variable aléatoire notée $t(x_I)$ ou encore plus directement t_I conformément aux notations habituelles où une minuscule indique ce qui est déterministe et une majuscule ce qui est probabiliste.

Dans ce cas, la moyenne \bar{t}_I et la dispersion $\sigma_{t_I}^2$ de la température d'un individu aléatoire traduisent le centre de gravité et l'inertie des températures de la population. En somme, le champ de température est "vu au travers" du champ de densité. L'avantage des statistiques par individu trouve ici une illustration simple puisque ces notions, basées sur une pondération par la variable régionalisée $z(x)$, sont "indépendantes" du champ. Leur définition fait en effet intervenir une intégration sur tout l'espace, sans limitation.

2.1.3 Loi et histogramme expérimental

L'expression analytique de la densité de probabilité de t_I passe par la fonction de répartition

$$\mathcal{P}(t_I < t) = \mathcal{P}(x_I \in \{x, t(x) < t\}) = \int \frac{z(x)}{Q} \mathbb{I}_{t(x) < t} dx$$

où $\mathbb{I}_{t(x) < t}$ représente l'indicatrice des températures inférieures au seuil t , soit :

$$\begin{aligned} \mathbb{I}_{t(x) < t} &= 1 \text{ si } t(x) < t \\ &= 0 \text{ si } t(x) \geq t \end{aligned}$$

Or, la définition analytique des ensembles géographiques de tous les niveaux de températures possibles ($\mathbb{I}_{t(x)<t}$) ne s'expriment pas de façon simple et commode et rend impossible le développement rigoureux de la loi de probabilité par cette voie. Par contre, une construction expérimentale de la loi de la température de surface d'un individu aléatoire est facilement réalisable.

Dans le deuxième chapitre de la première partie, j'ai suggéré d'examiner les relations statistiques entre deux variables par les abondances par classe plutôt que par les moyennes par classe afin d'éviter les problèmes de délimitation de champ.

En exprimant les résultats relativement à l'abondance totale, l'importance de chaque classe de température pour l'ensemble de la population était ainsi quantifiée. Mais l'abondance relative de poisson observée dans une classe de température donnée s'interprète aussi comme la "probabilité" que la température de surface associée à un individu tiré au hasard dans la population appartienne à la classe fixée. L'histogramme des abondances par classe constitue donc l'estimation de la densité de probabilité de la température de surface par individu.

L'ajustement d'une fonction adéquate à l'histogramme expérimental achève alors la spécification du modèle. Par ce biais là, on obtient une expression analytique de la loi de la température de surface par individu alors qu'une expression rigoureuse directe est impossible.

2.1.4 L'alternative entre régression et histogramme des abondances

Les deux points de vue

Dans l'alternative entre une régression expérimentale (moyenne par classe) et un histogramme des abondances relatives, l'enjeu ne s'exprime pas uniquement en terme de champ.

La modélisation d'une régression fournit une fonction mathématique qui indique, **dans un domaine donné**, la densité de poissons que l'on doit s'attendre à trouver en moyenne **connaissant la température de surface**. Une régression permet donc une description du lien bivariable du **point de vue de la température**.

A l'inverse, la loi de la température de surface par individu indique la probabilité de se trouver à une certaine température **connaissant un individu de la population**. Une telle loi fournit donc une description du lien bivariable du **point de vue d'un individu** d'une population.

Remarques sur le découpage en classes

Il existe aujourd'hui de nombreux moyens de lissages (voisinages glissants) qui permettent de modéliser la tendance qui se dégage d'un nuage de corrélation. Ces techniques de régression présentent l'avantage de ne pas nécessiter le découpage en classes de la variable explicative qui est une opération toujours un peu délicate.

Par contre, le découpage est nécessaire pour le calcul expérimental de la loi des températures par individu. Dans ce cas, il semble assez logique de choisir des classes régulières mais leur largeur a un effet évident sur la fonction de densité qui est inférée par la suite. Plus larges sont les classes et moins bonne est la modélisation. On cherchera donc à définir des classes les plus petites possibles tout en conservant des conditions statistiques d'estimation d'abondance convenables.

2.2 Loi d'un couple de variables par individu

La prise en compte de plusieurs co-variables ne pose pas de problème théorique. Cependant, les réserves évoquées plus haut concernant les effectifs par classe ont une acuité d'autant plus forte que le nombre de co-variables est important.

Soit par exemple le couple de variables hydrographiques comprenant la température de surface $t(x)$ et la salinité de surface $s(x)$. L'analyse statistique du lien entre une densité de poissons et ces deux variables conduit à considérer la loi du couple (t_I, s_I) par individu :

$$\begin{aligned} \mathcal{P}(t_I < t \text{ et } s_I < s) &= \mathcal{P}(x_I \in \{x, t(x) < t \text{ et } s(x) < s\}) \\ &= \int \mathbb{I}_{t(x) < t} \mathbb{I}_{s(x) < s} \frac{z(x)}{Q} dx \end{aligned}$$

Cette loi est estimée par l'abondance de poissons dans chacune des "cases" définies par le croisement d'une classe de température et d'une classe de salinité. On voit alors que le problème de l'effectif de chaque case se pose très rapidement et qu'il exclut pratiquement de considérer trois variables explicatives.

Sur la base des abondances par cases il est possible de représenter la loi bivariable par l'interpolation des lignes de même densité de probabilité dans le plan défini par les 2 co-variables. L'ajustement par une procédure automatique d'une fonction de type connu mais de paramètres inconnus permet, de son côté, de définir un modèle pour la loi bivariable par individu.

Je ne rentrerai pas dans le détail des différents ajustements possibles (en monovariable comme en multivariable). Il me paraît beaucoup plus important de préciser dans le paragraphe suivant les limites de l'interprétation des résultats fournis par ces deux outils que sont les régressions expérimentales et les abondances par classes. En particulier, peut-on, à partir de la loi des températures par individu, diagnostiquer le conditionnement de la température de surface sur l'espèce étudiée? En outre, l'optimum a-t-il valeur de *preferendum*?

2.3 *Preferendum* dans un contexte spatial.

Quel que soit le point de vue adopté, l'expérience montre que les régressions expérimentales et les histogrammes d'abondances relatives par classe présentent les mêmes caractéristiques générales, tout du moins lorsqu'il s'agit de mettre en relation une densité de poissons et une variable hydrographique. Les modèles auxquels on aboutit sont le plus souvent non linéaires et unimodaux (cloche). C'est à l'interprétation des maxima que je souhaite revenir ici, car il me semble qu'on les qualifie trop souvent, à tort, de *preferendum*.

2.3.1 Maximum et *preferendum*

Le maximum est une caractéristique du **modèle** de régression ou de densité de probabilité selon l'option choisie pour l'analyse des corrélations. On considèrera par la suite qu'il est unique. Le *preferendum*, quant à lui, quantifie un objectif supposé des poissons. Même s'il est justifié par la logique et les connaissances bio-physiologiques, rien ne prouve que le *preferendum* existe. Il s'agit donc d'un **concept biologique** par lequel on imagine que telle espèce (le maquereau par exemple) recherche une certaine (gamme de) température(s) à tel moment de son cycle biologique (la ponte

par exemple). En ce sens, on imagine qu'il existe une température optimale du point de vue du poisson; et que cette température optimale est indiquée par le maximum du modèle.

Reste alors à savoir dans quelle mesure l'optimum du modèle est un bon estimateur de l'optimum biologique, et en particulier, **dans quelle mesure l'existence d'un maximum indique ou non l'existence d'un preferendum.**

Position de la population dans le champ des températures

La régression et la loi des températures d'un individu aléatoire sont par construction des outils statistiques, ou encore ponctuels, en ce sens qu'ils ne portent que sur des paires de valeurs observées en un même point de l'espace. Cependant, les phénomènes auxquels on s'intéresse sont régionalisés et l'interprétation des résultats ne peut faire abstraction de cette dimension importante du problème. En particulier, ils ne permettent pas de savoir pourquoi, ailleurs, à quelques milles nautiques de la zone occupée par la population où la température de surface est pourtant optimale, il n'y a pas de poisson. Plus globalement, on ne sait pas non plus pourquoi la population dans son ensemble s'est localisée là où elle a été observée et non, à côté, où la température de surface était pourtant également favorable.

Ces considérations concernent essentiellement les tailles respectives des champs des deux variables : champ limité pour les poissons et infini pour la température de surface. Cependant, nous allons voir que les structures spatiales sont aussi la source de limitations importantes dans l'interprétation des résultats.

Structures spatiales

- Conséquences d'une différence de régularité spatiale.

Soit une population de petits pélagiques (maquereau par exemple). Bien que très irrégulière dans l'espace (contrastes importants en moyenne pour deux densités voisines) cette variable présente quand même une certaine structure spatiale. La tendance agrégative de l'espèce amène en effet les très fortes densités de poissons à se trouver grossièrement aux mêmes endroits, c'est-à-dire peu distantes en regard des dimensions de l'habitat de la population.

Soit, par ailleurs, le champ "infini" des températures de surface. Contrairement à une densité de poissons, cette variable régionalisée est très régulière dans l'espace. La forme du variogramme d'une telle variable (Cf. §I.2) témoigne du fait que la différence entre deux températures voisines est très faible même pour des distances importantes relativement au champ.

Ainsi, on doit s'attendre à ce que les densités élevées, assez regroupées dans l'espace géographique soient, par conséquent, assez regroupées également dans l'espace des températures. Quel que soit le mode de calcul (moyenne ou abondance par classe), la différence des structures spatiales des deux variables en jeux fait donc ressortir quelques classes privilégiées de températures même si il n'y a pas de lien entre la densité et la température. **La forme en cloche des résultats expérimentaux et par la suite des modèles, peut donc provenir de la différence des structures spatiales des variables régionalisées.**

- Simulations.

Pour s'en convaincre je propose de procéder à une simulation.

Les données de températures sont des observations réelles provenant d'une campagne scientifique.

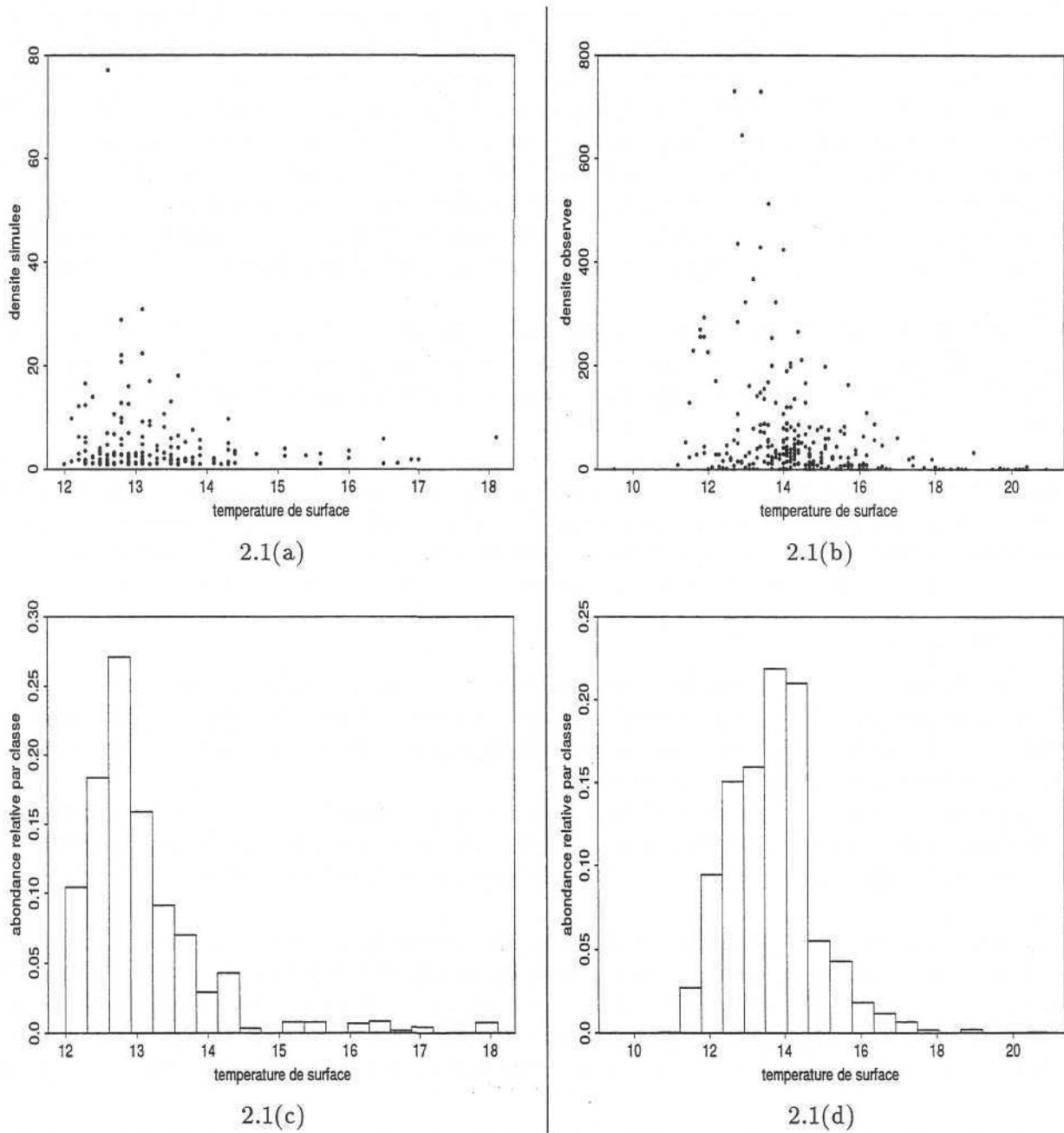


FIG. 2.1 - *Interprétation des optimums dans un contexte spatial. Comparaison entre les nuages de corrélations et les histogrammes d'abondances relatives par classe dans le cas d'une densité de poisson simulée (colonne de gauche) et observée (colonne de droite).*

Par ailleurs, pour obtenir des valeurs de densités aux mêmes points, j'ai réalisé une simulation d'une fonction aléatoire de structure spatiale connue dont je n'ai conservé que les points communs avec les températures. Pour espérer récupérer une simulation présentant les caractéristiques d'une densité de poissons, j'ai choisi une structure variographique composée d'un effet de pépite important et d'un schéma sphérique de portée assez courte relativement à la taille du champ à simuler. Dans cette simulation, je n'ai aucune prétention autre que celle d'obtenir *in fine* un ensemble de valeurs qui pourrait correspondre à des données halieutiques. La simulation a été faite par bandes tournantes (Matheron, 1973) de sorte que la distribution des valeurs simulées est à peu près gaussienne ($\mathcal{N}(0, 1)$). Pour retrouver une variable de type halieutique, c'est-à-dire positive et possédant beaucoup de valeurs nulles ou faibles et quelques valeurs fortes, j'ai pris l'exponentielle-translaté des résultats de la simulation ($z \rightarrow \exp(z + 0.5)$). Enfin, compte-tenu de notre objectif, il est bien entendu que cette simulation est par construction **indépendante** de la température de surface.

On dispose donc à ce stade de deux ensembles de valeurs correspondant à deux variables régionalisées "indépendantes" par construction, mais de structures spatiales bien différentes. Leur nuage de corrélation (Fig. 2.1a) présente une allure identique à celle des nuages de corrélation couramment obtenus à partir de données réelles (Fig. 2.1b).

Comme prévu, l'histogramme des abondances relatives par classe calculé à partir de la simulation présente un mode très marqué puisque 25% de l'abondance se trouve comprise entre 12.7°C et 13°C (Fig. 2.1c). L'histogramme correspondant à la situation réelle (Fig. 2.1d) possède à peu près les mêmes caractéristiques. En particulier, on remarquera que le pic de probabilité est du même ordre de grandeur que pour l'histogramme des valeurs simulées, à savoir environ 25%.

• Conclusion.

En utilisant, par simulation, deux variables indépendantes, on confirme donc que l'apparition d'un optimum tant sur les régressions expérimentales que sur les histogrammes expérimentaux peut être mis en relation avec la nature des structures spatiales des variables d'intérêts et non, nécessairement, avec leur éventuelle corrélation.

La présence d'un optimum dans les relations concernant ce type de paires de variables peut donc être interprétée comme un **artefact**. Cette situation est en quelque sorte analogue au paradoxe de Chayes exposé par Matheron (1963). Ce paradoxe s'exprime de la façon suivante.

Lorsqu'on effectue la composition minéralogique d'une roche en pourcentage des différents minéraux constitutifs, la somme des résultats reste égale à 100%. Cela entraîne nécessairement une corrélation négative entre les proportions : quand l'une augmente les autres diminuent. "*On ne peut donc conclure à une affinité ou à une répulsion entre les proportions que si la corrélation observée diffère significativement d'une certaine valeur considérée comme normale*" (Matheron, 1963, p 32).

La parallèle avec la corrélation bivariable qui nous intéresse s'exprime comme suit : on ne peut conclure à un lien entre une variable de type "densité de poissons" et une variable de type "température de surface" qu'à partir d'un "**certain niveau de cloche**".

Ce "niveau" à partir duquel il serait raisonnable de considérer qu'il y a un lien (ou un conditionnement) entre deux variables n'a cependant pas de définition rigoureuse. Pour détecter (diagnostiquer) la réalité du conditionnement de la température de surface sur une densité, il convient plutôt de se doter d'un autre outil statistique que la loi des températures par individu.

Rechercher un "certain niveau de cloche" nous oriente vers l'utilisation des inerties. Et, puisque notre réflexion a porté sur les structures spatiales, on suggère logiquement de s'orienter vers des outils géostatistiques, et plus précisément vers une inertie géostatistique.

2.3.2 Critère monovarié pour diagnostiquer un *preferendum*

Principe, définitions et notations

- Principe.

Le principe qui sous-tend notre démarche est le suivant : **une variable est dite conditionnante pour la répartition d'une espèce si l'ensemble des individus de cette espèce tend à minimiser naturellement son inertie par rapport à cette variable**; c'est-à-dire si la population se localise de façon à limiter les associations défavorables avec le paramètre en question. Par conséquent, pour juger du caractère conditionnant d'une variable, il faut chercher à savoir si son inertie est minimum, ou proche du minimum, pour la situation réelle, comparativement à toutes les autres implantations possibles de la population.

Enfin, le principe retenu nous amène à imaginer que la population est déplacée, d'un bloc, dans l'espace (un déplacement est défini par la combinaison d'une translation et d'une rotation autour du centre de gravité). Pour chaque "nouvelle" position de la population de poissons, on peut alors calculer l'inertie des températures et, auparavant, le centre de gravité.

- Centre de gravité.

Commençons par ne considérer que les translations. Seules les densités de poissons sont concernées par la translation; la variable explicative (température de surface) est inchangée. Une translation de vecteur h amène donc chaque densité $z(x)$ à la température $t(x+h)$. Le centre de gravité des températures de la population translatée de h est alors :

$$\bar{t}_I(h) = \frac{\int z(x)t(x+h)dx}{\int z(x)dx} = \frac{\int z(x)t(x+h)dx}{Q}$$

La prise en compte des rotations ne complique pas vraiment les notations. On suggère en effet, pour une rotation donnée d'angle α , de faire l'ensemble des translations possibles. Il s'agit donc de translations à rotation donnée qu'on notera $h|\alpha$; le centre de gravité s'écrivant alors :

$$\bar{t}_I(h|\alpha) = \frac{\int z_\alpha(x)t(x+h)dx}{\int z_\alpha(x)dx} = \frac{\int z_\alpha(x)t(x+h)dx}{Q}$$

où $z_\alpha(x)$ correspond à la variable régionalisée $z(x)$ après rotation autour du centre de gravité de la population.

- Inertie.

La calcul de l'inertie se déduit alors aisément de ce qui précède et s'exprime par :

$$\sigma_{t_I}^2(h|\alpha) = \frac{1}{Q} \int z_\alpha(x) [t(x+h) - \bar{t}_I(h|\alpha)]^2 dx$$

Partant d'un énoncé assez simple, on aboutit malheureusement à des notations légèrement complexes mais nécessaires pour conserver la totalité des idées sous-jacentes.

Par analogie avec la terminologie géostatistique où le variogramme désigne la fonction γ décrit l'évolution de la variance en fonction de la distance, l'inertie ainsi définie pourrait s'appeler **inertiogramme**.

- Dissymétrie

L'inertiogramme est une fonction dissymétrique. En particulier :

$$\sigma_{i_r}^2(h|\alpha) \neq \sigma_{i_r}^2(-h|\alpha)$$

2.3.3 Difficultés pratiques et choix d'un paramètre conditionnant

Si le concept est assez simple dans sa formulation théorique, c'est-à-dire lorsqu'il porte sur des variables régionalisées connues exhaustivement, un certain nombre de difficultés doivent être cependant contournées dans la pratique.

Nature des difficultés pour la densité et le paramètre et cas idéaux

La densité et le paramètre ne jouent pas un rôle symétrique dans l'inertiogramme. Les contraintes qui pèsent sur chacune de ces deux variables ne sont donc pas les mêmes dans la pratique.

- Situation idéale pour la densité.

Comme nous le verrons plus loin pour le covariogramme, l'estimation de l'inertiogramme n'est réellement commode que dans le cas où **la densité est connue sur une grille régulière**. C'est en effet la condition nécessaire pour estimer l'intégrale d'espace par une somme discrète.

L'utilisation de densités acoustiques, qui fournissent, aux erreurs de mesures près, une connaissance régulière et quasiment parfaite de la densité le long de transects, constitue bien sûr un cas favorable. Mais, ce qui compte pour la densité, reste bien la régularité du réseau d'informations.

- Situation idéale pour le paramètre

Dans une approche comme celle-ci où interviennent les valeurs du paramètre extérieures à la population, il faut bien évidemment disposer de mesures du paramètre au delà de la zone de présence du poisson. Le calcul d'un inertiogramme n'est par ailleurs possible que si on dispose, pour une translation (et une rotation) quelconque de l'ensemble des densités, de mesures correspondantes du paramètre aux nouveaux points de densité. La situation la plus confortable est donc celle d'un **paramètre connu (quasi)exhaustivement sur une zone débordant largement le champ de la population étudiée**.

C'est le cas par exemple de la température de surface lorsqu'elle est mesurée par satellite ou encore de la bathymétrie connue de manière quasi-exhaustive (toute les 5 minutes d'angle) dans la région qui nous intéresse (Nord Est Atlantique).

Les campagnes à la mer apparaissent alors malheureusement comme de mauvaises pourvoyeuses d'informations pour ce genre d'analyses. En effet, bien souvent dans les campagnes halieutiques, les mesures hydrographiques ne sont que des mesures d'accompagnement des relevés biologiques et ne sont donc connues que sur l'aire de répartition du poisson. Il devient alors très délicat de calculer l'inertiogramme sur la base de cette seule information.

Il se peut, de plus, que l'échantillonnage soit irrégulier. En l'absence d'informations complémentaires, le mieux est alors probablement de créer une grille régulière pour chacune des deux variables de façon *ad hoc*.

Méthodes appropriées pour créer des grilles régulières

Comment, dans l'optique du calcul d'un inertioGramme, générer une grille régulière à partir d'échantillons irrégulièrement répartis dans l'espace : par migration des échantillons les plus proches, par estimation, ou bien encore par simulation ?

- Remarques sur le krigeage

En retenant l'inertie comme critère, on a opté pour une approche basée sur un indice de dispersion. Or on sait qu'une procédure d'interpolation comme le krigeage conduit à des résultats lissés, ayant une variance plus faible que la variance initiale des données. Une étude inertielle réalisée sur des données interpolées par krigeage devrait donc être biaisée et il convient d'adopter d'autres solutions.

- Cas du paramètre

Dans le cas d'une variable de l'environnement qui, en général se prête assez bien à une modélisation par une fonction aléatoire (Cf. § I.1.2), il est envisageable, même si cela nécessite une étape préalable d'analyse structurale qui peut se révéler complexe, de simuler la variable. Les simulations présentent en effet l'avantage, c'est même un de leurs objectifs essentiels, de respecter la variabilité de la variable régionalisée.

- Cas de la densité.

Puisqu'on ne sait pas construire une variable régionalisée de covariogramme donné (Cf. § II.3.5.3), on suggère, pour ce qui concerne la densité, d'informer chaque nœud d'une grille régulière par l'échantillon le plus proche. L'analyse conserve alors sa pertinence s'il est légitime de randomiser, à grande échelle (i.e. petite distance), les densités de poissons, ou, ce qui revient au même, à supposer qu'une concentration donnée, forte ou petite, aurait pu être observée ailleurs mais dans un voisinage assez proche de là où elle a été mesurée. Cette conception, déjà exposée précédemment, me paraît être en accord avec la nature du phénomène. N'oublions pas en particulier que nous négligeons le plus souvent que les poissons se déplacent. Il n'est donc pas aberrant, de ce point de vue, de supposer qu'une valeur aurait pu être observée légèrement à coté de là où elle a été observée. Il aurait probablement suffi en effet que l'échantillonnage se déroule quelques heures après pour que les densités se soient légèrement déplacées. Enfin, l'incertitude sur les mesures de concentrations de poissons est beaucoup plus importante que celle sur la plupart des paramètres hydrographiques. Le traitement des densités de poissons peut donc se permettre d'être moins rigoureux.

Translations réalistes

Pour une population côtière, une rotation de 90 degrés ou une translation vers la côte va amener une partie importante de la population à se trouver à terre. Ainsi, en plus d'être par définition dissymétriques, il se peut que les parties "*h* positif" et "*h* négatif" de l'inertioGramme ne concernent pas les mêmes ensembles statistiques, entraînant des difficultés d'interprétations.

Il convient donc de ne pas trop systématiser les translations - rotations et de chercher à ne procéder qu'à des translations réalistes du point de vue du poisson.

La quatrième partie du document porte entre autre sur l'étude du preferendum des œufs de maquereaux en matière de température de surface. Les œufs ont la particularité de se répartir grossièrement le long du bord du plateau continental. Dans ce cas, pour obtenir des translations réalistes du point de vue de la population, nous nous sommes placés dans un repère centré sur le bord du plateau continental et procédé à des translations combinant des déplacements parallèles et perpendiculaires aux accores. Dans ce cas, on notera également qu'il devient assez peu intéressant de faire une quelconque rotation.

Instabilité numérique

L'inconvénient reconnu du passage au carré pour des densités de poissons est ici évité dès lors que les valeurs prises par la variable explicative ne sont pas trop contrastées. C'est le cas par exemple de la température de surface. Cependant, dans une région comme les accores du plateau continental européen, la bathymétrie évolue très rapidement passant de valeurs faibles (≈ 200 mètres) à des profondeurs très importantes (> 2000 mètres) en quelques dizaines de milles nautiques. On doit s'attendre à une certaine instabilité numérique lorsque de telles variables sont en jeu.

Choix d'un paramètre conditionnant

- Un bon paramètre

Compte-tenu de ce qui vient d'être exposé, une "bonne" variable conditionnante est une variable assez régulière (au sens de sa structure spatiale) et connue le plus finement possible sur un champ le plus grand possible. Les cartes de température de surface fournies par les satellites sont de ce point de vue de bonnes variables, aux incertitudes de mesure près.

Si les moyens techniques ne permettent pas de connaître aussi parfaitement un paramètre, il est alors souhaitable que la variable utilisée soit facilement simulable et, en particulier, que son analyse structurale ne pose pas trop de problèmes. Les dérivées spatio-temporelles dont j'ai parlé dans la première partie du document sont à cet égard des inconvénients notables.

- Le bon paramètre?

Il est possible que la bathymétrie à proprement parler, en plus d'être la source d'instabilités numériques, ne soit pas la variable appropriée pour expliquer, par exemple, le conditionnement de la régionalisation des œufs de maquereaux. On peut en effet envisager que ce soient les ruptures de pente du plateau continental qui constituent les éléments conditionnants ou encore les gradients de température de surface.

Le problème qui se pose ici est donc celui de rechercher, parmi la multitude des possibles, la bonne variable, i.e. la variable conditionnante. Avec une seule variable, notée $P(x)$ ("P" pour Paramètre) un grand nombre de critères sont envisageables :

- la variable elle-même : $P(x)$.
- le gradient : $\vec{grad}(P_x)$. Dans ce cas on suppose que ce sont les variations du paramètre qui sont importantes. Reste à savoir si le gradient intervient en tant que variable vectorielle, ou uniquement au travers de son module ou de son orientation.
- le laplacien : $\Delta = \text{div}(\vec{grad}(P(x))) = \frac{\partial^2 P}{\partial u^2}(x) + \frac{\partial^2 P}{\partial v^2}(x)$ où u et v sont les coordonnées du point courant x . Dans le cas de la bathymétrie, le recours à cette transformation se justifierait par le fait que ce n'est ni la profondeur ni la pente du plateau continental qui semble caractériser

l'aire de répartition du poisson, mais plutôt le changement de pente. C'est en tous les cas les changements de pentes qui génèrent des phénomènes hydrographiques turbulents et par là des conditions qui pourraient favoriser et conditionner la répartition de poissons à la verticale des accores.

Ce ne sont là que quelques exemples des transformations possibles à partir d'une seule variable. Ainsi, à la difficulté de choisir la bonne variable, s'ajoute la difficulté de choisir la bonne transformation.

2.3.4 Vers un modèle dynamique

Les raisons qui nous ont amenés vers la recherche d'un minimum d'inertie sont d'ordre statistique. Cependant, sous certaines hypothèses, on peut montrer que si la distribution spatiale des individus d'une population est telle que l'inertie d'un critère est minimum, alors cela est compatible avec un modèle dynamique où les modifications de distributions des individus sont liées au gradient du critère. En somme, les individus se distribuent dynamiquement en fonction du gradient de ce paramètre de sorte qu'à un instant donné, c'est l'inertie du paramètre qui est minimum.

Soit le cas de la température de surface. Si l'inertie de la température de surface est minimum, on peut alors supposer que les poissons ont besoin d'être dans un environnement à fort gradient de température pour pouvoir s'orienter vers les zones de températures favorables. C'est aussi dans des zones à forts gradients que la situation des poissons peut se détériorer rapidement s'ils se déplacent dans la mauvaise direction. Autrement dit ce n'est pas tant l'écart en température qui compte, mais comment il varie si "on" se déplace, soit le vecteur $(P(x) - P_0)\vec{grad}(P)$.

Chapitre 3

Le covariogramme

Sommaire

Ce chapitre, assez mathématique, est consacré au développement du covariogramme.

*Le **paragraphe 1** constitue le rappel de la définition du covariogramme et de ses principales propriétés : portées, anisotropies, régularité, positivité. Le cas particulier du covariogramme géométrique est abordé.*

*Le **paragraphe 2** s'intéresse au covariogramme en tant qu'outil statistique de description de structures spatiales. On déclinera alors le covariogramme de différentes façons. Tout d'abord comme la densité de probabilité que deux individus aléatoires soient distants d'une certaine distance h . Puis comme la combinaison des régressions donnant la valeur moyenne de $z(x + h)$ connaissant la densité au point x . Ou encore, comme la somme de fonctions qu'on interprétera comme des fréquences conditionnelles de dépassement de seuils.*

*Le **paragraphe 3** s'intéresse au covariogramme en tant qu'instrument d'estimation. On connaît en général, comme je le rappelle tout d'abord, son rôle dans le calcul de variances d'estimations globales. Mais, il peut également servir à des procédures de cartographie. On parlera alors de krigeage transitif.*

*Le **paragraphe 4** aborde le délicat problème de l'estimation du covariogramme. Après avoir rappelé le cas des échantillonnages réguliers, je propose quelques estimateurs dans le cas de plans d'échantillonnages irréguliers basés sur l'utilisation des surfaces d'influence des échantillons. Pour finir, quelques réflexions sur la construction d'une variable régionalisée de covariogramme connu sont proposées.*

Covariogram

Abstract

The focus of this chapter, a bit more mathematical than the previous ones, is the covariogram.

The paragraph 1 gives briefly the definition of the covariogram and remind some vocabulary and the basic properties: ranges, anisotropies, regularity, and positivity. The case of geometrical covariogram is also presented.

The paragraph 2 takes the covariogram as a tool for the description of spatial structures. As an extension of the statistics per individual, the covariogram is firstly presented as the probability density function of the distance between two independent random individuals. Besides this individual point of view, we move on to the decomposition of the covariogram into functions giving the average density of $z(x+h)$ knowing the density at point x . This leads, in fishery at least, to interpret the covariogram as the average decrease of density h apart large concentrations. Finally, we suggest another possible decomposition of the covariogram into functions based on density thresholds. By the end, these functions are seen as conditional frequencies that the density $z(x+h)$ exceeds a given threshold z' knowing that $z(x)$ exceeds another threshold z .

The paragraph 3 go back to the more traditional use of covariograms. That is a tool for estimation purposes. In the first part I remind the computation of global estimation variance as developped by Matheron (1965). Then I explain how the covariogram can be used for the cartography in a transitive context. This latest method is called by Matheron external or transitive kriging.

The paragraph 4 gives practical solutions for the estimation of covariograms. In particular when the sampling is not regular. This calls upon surfaces of influence like for the estimation of abundance. But, here the weighting is more complicated. By the end, some reflexions about the possible construction of a regionalised variable with a given covariogram are proposed.

Les deux chapitres précédents ont été consacrés à la définition d'outils statistiques adéquats pour l'analyse de données halieutiques. A l'occasion de la question fondamentale en écologie halieutique de la recherche et la quantification de preferendum, il est apparu nécessaire de construire des outils géostatistiques, c'est-à-dire tenant compte de la position des valeurs dans l'espace. En géostatistique transitive, l'outil de base des analyses spatiales est le covariogramme.

L'habitude a cependant consacré le variogramme comme outil structural mais nous avons vu que de nombreuses situations sont incompatibles avec ce type d'approche. L'alternative entre variogramme et covariogramme est aujourd'hui biaisée tant le contexte méthodologique du second est peu développé face à celui du premier. Ce chapitre, légèrement plus mathématique que les précédents, est par conséquent consacré au covariogramme et vise en outre à remettre son utilisation "au goût du jour". Il contribue au développement théorique mais surtout pratique du covariogramme et devrait favoriser l'utilisation de la géostatistique transitive dans les disciplines halieutiques.

Je m'intéresserai tout d'abord au covariogramme en tant qu'outil statistique de description. Je commencerai par rappeler sa définition qui trouve une place nouvelle dans le prolongement spatial des statistiques par individu. Nous verrons ensuite que le covariogramme, qui par définition est un outil global (intégrale d'espace), recèle des outils structuraux plus fins permettant d'affiner les analyses structurales descriptives.

Dans un deuxième temps, je rappellerai les travaux de Georges Matheron en matière de calcul de variances d'estimations globales à l'occasion desquels il a défini le covariogramme. J'exposerai

ensuite les principes d'une cartographie dans un contexte transitif, encore appelée krigeage transitif.

L'estimation du covariogramme et de ses dérivés (fonctions structurales plus fines), confortable à maille régulière, est assez délicate pour des échantillonnages irréguliers. Si l'on souhaite que le covariogramme soit un outil opératoire en halieutique, il est cependant nécessaire de proposer des solutions, même peu rigoureuses, au problème de son estimation à maille irrégulière. Parmi celles-ci, l'utilisation de surfaces d'influence se révèlera plus complexe que pour l'estimation d'abondance.

3.1 Le covariogramme : définition et propriétés

3.1.1 Définition

Le covariogramme, introduit par Matheron dès le début de la géostatistique en 1965, et parfois qualifié de **covariogramme transitif**, est une fonction de la distance géographique. Il est traditionnellement noté $g(h)$ où h est une distance vectorielle. Le covariogramme est défini par :

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

où $z(x)$ est une variable régionalisée. A 2 dimensions, chaque point de l'espace possède deux coordonnées (x, y) et le covariogramme devrait s'écrire comme l'intégrale double suivante :

$$g(\vec{h}) = \iint z(x, y)z(x + h_x, y + h_y)dx dy \quad \text{avec} \quad \vec{h} \begin{cases} h_x \\ h_y \end{cases}$$

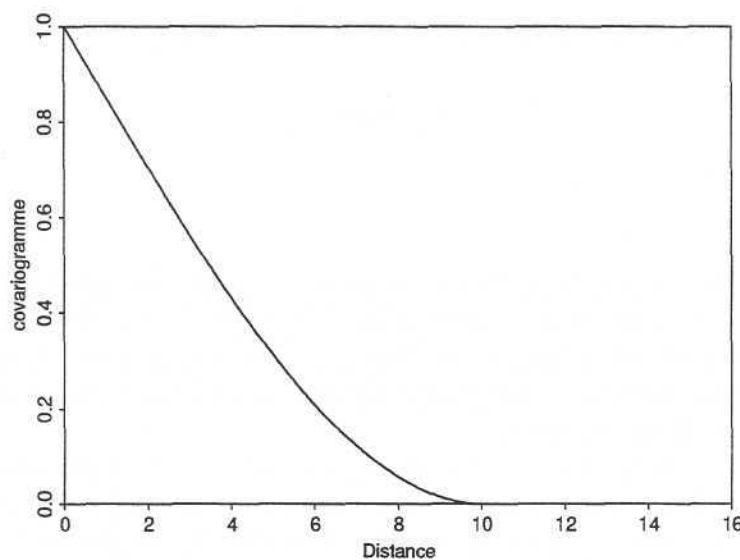


FIG. 3.1 - Exemple de covariogramme. Schéma sphérique de portée 10 et de palier 1 d'expression

$$g(h) = 1 - \frac{3h}{2 \cdot 10} + \frac{h^3}{2 \cdot 10^3} \quad \text{pour } 0 < h < 10$$

$$g(h) = 0 \quad \text{pour } h \geq 10$$

3.1.2 Propriétés

Signe, symétrie et inégalité.

De la définition, il découle que :

$$g(h) \geq 0 \quad \text{si} \quad z(x) \geq 0 \quad \forall x$$

Dans toute la suite, $z(x)$ représentant une **densité** de poissons, on considérera que $z(x)$ et $g(h)$ sont des fonctions **positives**.

Le covariogramme est une fonction **symétrique** en h :

$$g(h) = g(-h)$$

Par habitude, on ne représente, dans une direction donnée, que la partie "h positif" du covariogramme; la partie "h négatif" s'en déduisant directement par symétrie (Fig. 3.1) . Par analogie à l'inégalité de Schwarz :

$$g(h) \leq g(0)$$

où

$$g(0) = \int z(x)^2 dx \neq Q^2 = \left(\int z(x) dx \right)^2$$

On montre également (voir §II.3.2) que:

$$\int g(h) dh = Q^2$$

Portées.

Dans une direction donnée, la **portée** est la distance au-delà de laquelle le covariogramme est définitivement égal à 0. Dans notre exemple (Fig. 3.1), la portée du schéma sphérique est égale à 10. La portée correspond par définition au "diamètre" du champ, si le champ est convexe. Plus généralement, la portée mesure la plus grande dimension du champ dans chacune des directions. C'est donc une **quantité géométrique**.

Si la portée est une propriété du champ, il n'en reste pas moins qu'elle peut avoir une signification intrinsèque. En effet, la portée peut quantifier une limite biologique d'extension d'un ensemble d'individus. C'est le cas par exemple d'une espèce ne constituant jamais de banc d'une taille supérieure à une certaine limite; le banc se divisant en deux lorsqu'il devient trop grand. Dans ce cas, les portées sont une propriété de l'espèce.

Plus globalement, il faut bien voir qu'on a recours à la géostatistique transitive parce que les répartitions de densités de poissons sont en général influencées par la forme et la dimension du champ qu'elles occupent (champ-dépendance). Cela suppose en particulier qu'il n'est pas possible de distinguer dans la structure finale, et dans la valeur des portées en particulier, ce qui est dû à une contrainte géométrique extérieure et ce qui est intrinsèque à l'espèce étudiée.

Comportement à l'origine

“La régularité de $g(h)$ au voisinage de l'origine reflète les propriétés de continuité de la variable régionalisée dans sa répartition spatiale” (Matheron, 1970, p.11). Cela résulte de :

$$g(0) - g(h) = \frac{1}{2} \int [z(x) - z(x+h)]^2 dx$$

En effet, pour h petit, l'écart entre $g(0)$ et $g(h)$, et donc la pente du covariogramme à l'origine, est d'autant plus petit que $z(x)$ et $z(x+h)$ sont peu différents.

Parmi les situations caractéristiques, on retiendra que :

- si $z(x)$ est **dérivable**, $g(h)$ a un **comportement parabolique** au voisinage de l'origine (pente nulle).
- si $z(x)$ est **continue par morceau**, $g(h)$ a un **comportement linéaire** au voisinage de l'origine (pente plus ou moins forte). Le schéma sphérique en est un exemple (Fig. 3.1).
- si $z(x)$ présente des transitions très rapides, voire des discontinuités, $g(h)$ a un comportement non continu en $h = 0$ (pente infinie). On dit alors qu'il y a **effet de pépîte**. Plutôt qu'une véritable discontinuité, il s'agit le plus souvent d'une zone de transition très rapide qui, expérimentalement, se traduit par une discontinuité. Des erreurs de mesures peuvent engendrer des effets de pépîte.

Au vu des 3 cas évoqués plus haut, la pente du covariogramme à l'origine est d'autant plus fortement négative que la variable régionalisée est peu continue dans l'espace. Matheron donne une formulation plus précise de cette règle (1970, pp.12). Il envisage pour cela le développement limité du covariogramme en $h = 0$, c'est-à-dire l'équivalent polynomial du covariogramme pour des petites distances. Dans le cas isotrope (voir plus bas) le covariogramme s'exprime comme une fonction du module de la distance uniquement, donc comme une fonction d'une seule variable. Il ressort alors que le développement limité comprend une **partie régulière** ne comportant que les puissances paires et une **partie irrégulière** regroupant des termes en puissance impaires et non entières et des termes logarithmiques :

$$g(h) = g(0) + \underbrace{\sum_k a_k h^{2k}}_{\text{partie régulière}} + \underbrace{\sum_{k \neq \text{entier pair}} a_k h^k + \sum_k h^{2k} \ln(h)}_{\text{partie irrégulière}} \quad \text{pour } h \text{ petit}$$

De là, Matheron montre que le “le terme de plus bas degré de la partie irrégulière donne une mesure précise du degré d'irrégularité de la variable régionalisée dans sa variation spatiale” (1970).

Isotropie et anisotropie

Un covariogramme est **isotrope** s'il est identique dans toutes les directions de l'espace. La distance n'intervient alors que par son module. Cette situation est assez rare dans la nature, et on rencontre le plus souvent des structures **anisotropes**, c'est-à-dire différentes selon la direction (Fig. 3.2).

“On dit qu'il y a **anisotropie géométrique** lorsqu'une transformation linéaire des coordonnées suffit à rétablir l'isotropie” (Matheron, 1970). Cela signifie en particulier que les structures spatiales directionnelles diffèrent uniquement d'un facteur d'échelle géographique. Ainsi, en étirant (ou en

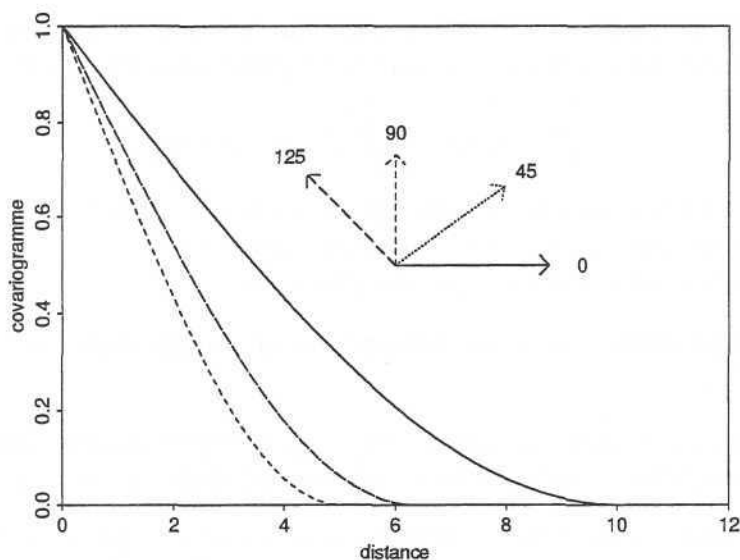


FIG. 3.2 - Exemple de covariogramme anisotrope. Schéma sphérique de palier 1. Anisotropie géométrique de coefficient 0.5 dans la direction 90° : portée 10 dans la direction 0° et portée 5 dans la direction 90° .

raccourcissant) les coordonnées dans la direction d'anisotropie, on doit retrouver des conditions isotropes.

Ce type d'anisotropie se manifeste par le fait que la rose des portées (i.e. la variation de la portée en fonction de la direction) s'inscrit dans une ellipse (Fig. 3.4). Les axes et l'excentricité de l'ellipse permettent de définir la transformation linéaire assurant une situation isotrope (Journel, 1977). En général, le grand axe de cette ellipse et le rapport du plus grand diamètre sur le plus petit fournissent respectivement la **direction** et le **coefficient d'anisotropie**.

Toutes les anisotropies ne peuvent cependant être traitées aussi facilement. Il existe des types plus complexes d'anisotropies.

Covariogrammes gigognes

Un covariogramme **gigogne** est composé de **plusieurs structures élémentaires** emboîtées (Fig. 3.3). Parmi les structures qui composent un covariogramme donné, certaines peuvent être isotropes et les autres anisotropes. En particulier, il est possible de rencontrer des répartitions spatiales telles que le covariogramme soit anisotrope jusqu'à une certaine distance et isotrope ensuite (le contraire est théoriquement impossible).

Dans une approche transitive, il faut prendre garde à l'interprétation qui est faite d'un tel modèle. L'expérience acquise dans l'utilisation des variogrammes est à ce titre "mauvaise conseillère". En outre, il n'est pas possible de construire un modèle dans lequel la variable régionalisée serait la somme de variables régionalisées dont les covariogrammes seraient les différentes structures reconnues dans le covariogramme.

Toutefois lorsqu'il y a plusieurs structures (modèles gigognes), l'anisotropie de la structure la plus longue (i.e. s'exprimant pour les distances les plus grandes) est nécessairement celle du champ

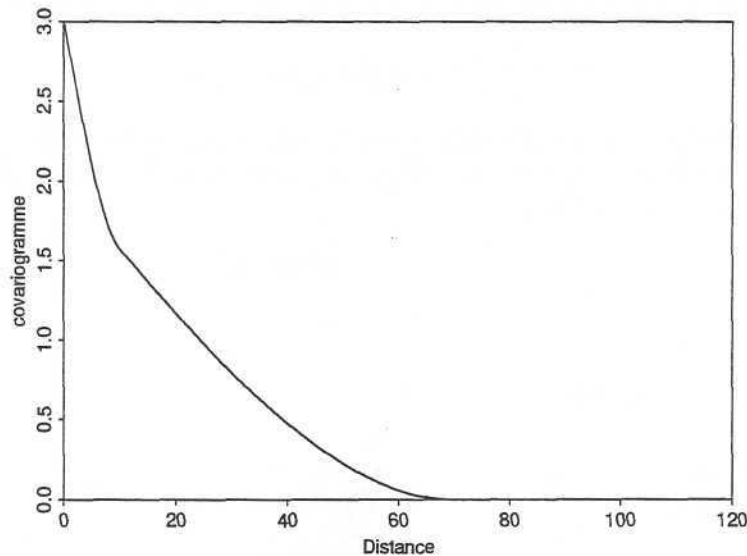


FIG. 3.3 - Exemple de covariogramme gigogne composé d'un schéma sphérique de portée 10 et de palier 1 et d'un schéma sphérique de portée 70 et de palier 2.

et donc nécessairement de nature géométrique.

Positivité.

On peut montrer (Matheron, 1970, p. 13) que $g(h)$ est une fonction de **type positif** (notion différente de celle évoquée plus haut concernant le signe positif de la fonction structurale). C'est-à-dire que pour tout entier k , tout ensemble x_1, x_2, \dots, x_k de points de l'espace et tout système $\lambda_1, \lambda_2, \dots, \lambda_k$ de nombres réels, on a :

$$\sum_{i=1}^k \sum_{j=1}^k \lambda_i \lambda_j g(x_i - x_j) \geq 0$$

On remarquera que les poids λ_i peuvent être négatifs. Cette propriété, très forte, puisqu'elle concerne une combinatoire extrêmement riche d'ensembles de points de l'espace et de pondérateurs, est difficile à contrôler dans la pratique. C'est la raison pour laquelle on choisira un modèle de covariogramme parmi les fonctions dont on sait qu'elles sont de type positif.

Nous verrons plus loin, que la positivité est la condition nécessaire et suffisante pour que les variances d'estimation (de combinaisons linéaires) soient positives.

3.1.3 Cas particulier : le covariogramme géométrique

Dans toute la suite de cet exposé **on notera indifféremment S une surface et la mesure de cette surface**. Il n'y a en général pas d'ambiguïté entre ces deux notions dans les formules.

Appliqué à l'indicatrice du phénomène, $\mathbb{I}(x) = 1$ si $z(x) > 0$ et $\mathbb{I}(x) = 0$ sinon, le covariogramme

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

est alors appelé **covariogramme géométrique**. Il correspond à la surface de l'intersection du champ S avec son translaté du vecteur $-h$ (Fig. 3.4), noté aussi S_{-h} (en 0, $K(0) = S$), et permet ainsi de décrire la géométrie du champ.

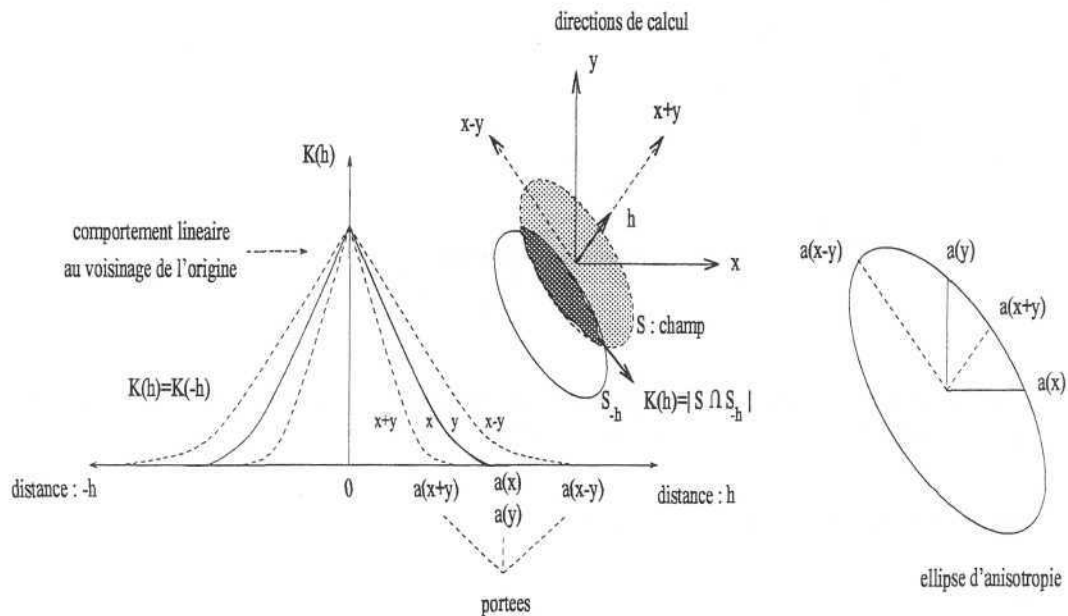


FIG. 3.4 - Covariogramme géométrique d'une ellipse (champ) dans quatre directions de l'espace : x , $x+y$, y et $x-y$. Propriétés du covariogramme : symétrie, portées, comportement linéaire à l'origine. Anisotropie géométrique et ellipse d'anisotropie (rose des portées).

Une situation remarquable est celle d'un champ en forme de sphère (3D). Dans ce cas, le covariogramme géométrique est identique dans toutes les directions (isotrope), et son expression analytique, connue sous le terme de covariogramme sphérique, ne dépend que du module du vecteur distance :

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

Les deux paramètres S et a donnent respectivement la hauteur du covariogramme (la surface du champ) et sa portée (ici le diamètre). Dans le cas d'un champ ellipsoïdique, on obtiendrait le même type de fonction, de forme sphérique, mais plus ou moins allongée selon la direction (Fig. 3.4), le paramètre a dépendant alors de la direction (structure anisotrope).

Parmi les propriétés d'un covariogramme géométrique, on retiendra que le comportement à l'origine est obligatoirement linéaire (Matheron, 1965, p. 105). C'est une propriété, propre au covariogramme géométrique, qu'il est bon d'avoir à l'esprit au moment de la modélisation afin d'éviter un manque de cohérence interne du modèle.

3.2 Le covariogramme pour l'analyse structurale

La plupart des éléments qui se trouvent dans ce chapitre ont déjà fait l'objets de publications : Bez N., J. Rivoirard et J-Ch. Poulard, 1995. Approche transitive et densités de poissons. *Cahiers de Géostatistique* 5, Ecoles des Mines de Paris, 161 - 177 pp.).

Bez N., J. Rivoirard, Ph. Guiblin and M. Walsh, 1996. Covariogram and related tools for structural analysis of fish survey data. Fifth International Geostatistics Congress, Wollongong, Australia. (*in press*).

3.2.1 Distance entre deux individus aléatoires

Algorithmes

Soient I et I' les numéros de deux individus tirés au hasard dans la population indépendamment l'un de l'autre. Ces deux variables sont des variables aléatoires indépendantes uniformes dans $[1, Q]$:

$$I \stackrel{\mathcal{L}}{\equiv} I' \stackrel{\mathcal{L}}{\equiv} \mathcal{U}[1 : Q]$$

Les positions respectives de ces deux individus aléatoires x_I et $x_{I'}$ sont des vecteurs aléatoires indépendants et de même loi :

$$x_I \stackrel{\mathcal{L}}{\equiv} x_{I'} \stackrel{\mathcal{L}}{\equiv} \frac{z(x)}{Q}$$

La distance qui les sépare $h_{I,I'} = x_I - x_{I'}$ est par conséquent un vecteur aléatoire. Dans notre modèle, cette distance est égale à la différence entre deux vecteurs aléatoires indépendants et identiquement distribués (i.i.d.). Sa densité de probabilité s'apparente donc à un produit de convolution.

Rappels: le produit de convolution permet de connaître la loi de la somme de deux variables aléatoires indépendantes (Saporta, 1980). Soient deux variables aléatoires X et Y de loi connues f et g . La densité de probabilité $h(z)$ de la somme $Z = X + Y$ est le produit de convolution des densités :

$$h(z) = \int_{\mathcal{D}_X} f(x)g(z-x)dx = \int_{\mathcal{D}_Y} g(y)f(z-y)dy$$

Dans notre cas, les lois de probabilité sont identiques. En développant le produit de convolution à une dimension et en modifiant la somme par une différence, on trouve que la densité de probabilité de la distance entre deux individus aléatoires est :

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

soit, le covariogramme relativement au carré de l'abondance totale. Ainsi, le covariogramme est une fonction de la distance qui donne, à un facteur Q^2 près, la densité de probabilité de la distance entre deux poissons tirés au hasard dans la population.

Interprétations structurales

Cette façon de présenter le covariogramme, proposée par Matheron en 1970 au travers d'un exercice (Matheron, 1970, exercice 16, p. 47), n'a eu qu'un essor limité. Elle est pourtant particulièrement pertinente en biologie marine.

Tout d'abord, elle confère une signification très physique au covariogramme basée sur la distance qui sépare les individus. La valeur à l'origine ($g(0)/Q^2$) correspond à la densité de probabilité d'une distance nulle entre deux poissons pris au hasard. Expérimentalement, on parlera plutôt de la probabilité pour que deux poissons pris au hasard, dans l'échantillonnage, se trouvent au même endroit, c'est-à-dire **agrégés**. Une augmentation de la valeur à l'origine, signifie donc que l'abondance totale est le résultat de poissons qui sont plus souvent les uns à côté des autres, ou encore agrégés. Ainsi, la valeur $g(0)/Q^2$ peut être interprétée comme une **mesure du niveau d'agrégation**.

Par construction, le covariogramme relativement à Q^2 est homogène à l'inverse d'une surface (ou d'un volume selon la dimension de l'espace de travail). La surface $Q^2/g(0)$ représentant l'inverse de la valeur de la loi de probabilité en $h = 0$ revêt un sens assez particulier. En effet :

$$\frac{Q^2}{g(0)} = \frac{\int g(h)dh}{g(0)}$$

est la **portée intégrale**. En géostatistique intrinsèque, cette grandeur représente la dimension du domaine à partir de laquelle il y a un sens de parler de moyenne (Matheron, 1980; Chauvet, 1994). Nous verrons dans le chapitre 4 le sens que l'on peut donner à la portée intégrale en géostatistique transitive.

Le comportement de $g(h)/Q^2$ à l'origine peut être rattaché à l'extension moyenne des agrégats : une forte pente est le signe que les zones à très fortes concentrations sont très peu étendues.

La moyenne du carré de la distance entre deux individus aléatoires d'une population est :

$$E(h_{I,I'}^2) = \int h^2 \frac{g(h)}{Q^2} dh$$

Or, puisque les positions aléatoires x_I et $x_{I'}$ sont indépendantes et de même loi, cette moyenne s'écrit encore :

$$\begin{aligned} E(h_{I,I'}^2) &= E([x_I - x_{I'}]^2) &= E(x_I^2 + x_{I'}^2 - 2x_I x_{I'}) \\ &= 2E(x_I^2) - 2E(x_I x_{I'}) \\ &= 2Var(x_I) \end{aligned}$$

On voit ainsi que le covariogramme est un outil de description de la dispersion d'une variable régionalisée beaucoup plus riche que l'inertie qui apparaît ici comme la demi moyenne de la distance quadratique entre individus.

La moyenne de la distance entre individus est plus délicate à développer à l'aide des variables x_I et $x_{I'}$ (problème de valeur absolue). On sait cependant, par analogie avec certaines des formules de sélectivité (Lantuéjoul, 1990), que cette moyenne est reliée à l'inertie par la relation suivante :

$$\frac{1}{2}E(\|\vec{h}_{I,I'}\|) \leq \sqrt{\frac{I}{3}}$$

Remarque complémentaire

Compte-tenu de sa définition (intégrale d'espace), l'utilisation du covariogramme n'est pertinente que si l'on fait la somme des produits de densités sur tout l'espace. Dans la pratique, il convient donc de s'assurer que l'échantillonnage a effectivement couvert l'ensemble de la régionalisation. Or il arrive qu'une partie de la population ne soit pas échantillonnée. Dans ces conditions, la quantité totale calculée sur la zone échantillonnée s'écarte de l'abondance totale vraie du stock.

Mais une partie du covariogramme reste peut-être exploitable, relativement à la quantité totale. Si on admet que le comportement de l'espèce étudiée, et le comportement agrégatif en particulier, n'a pas de raison de changer significativement dans la zone manquante, on peut étendre à l'ensemble de la population les résultats structuraux par individu obtenus à petite distance; c'est-à-dire le niveau de l'agrégation ($g(0)/Q^2$) et la "pente" à l'origine.

3.2.2 Covariogramme et covariance

Algorithmes

- Le covariogramme géométrique mesure la surface de l'intersection entre le champ et son translaté d'un vecteur h ou encore, le nombre le paire de densités où, à la fois x et $x+h$ appartiennent au champ de la variable régionalisée. Ainsi, de même que l'abondance divisée par la surface du champ ($S = K(0)$) donne la densité moyenne sur le champ, le covariogramme $g(h)$ divisé par $K(h)$ donne $c(h)$ la moyenne des produits de densités sur le champ :

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

- Par ailleurs, le covariogramme, en tant que somme de produits, est inchangé chaque fois que l'un au moins des éléments du produit est nul, soit, quand la densité $z(x)$ est prise en dehors du champ, quelque soit la valeur prise par $z(x+h)$ (positive ou nulle). Il suffit par conséquent d'assurer que x balaye le champ S pour obtenir la totalité du covariogramme qui peut alors s'écrire :

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

Sous cette forme, le covariogramme apparaît, au facteur S près, comme une covariance non centrée à **partir du champ**, et non plus dans le champ comme précédemment. Cette distinction justifie en partie le terme de géostatistique **transitive** en ce sens qu'on inclut dans l'analyse la transition entre l'intérieur et l'extérieur du champ.

Interprétation

Quel lien existe-t-il entre la moyenne de produits de densités $c(h)$ et la covariance non centrée d'une fonction aléatoire?

Imaginons que $z(x)$ soit une réalisation d'une **fonction aléatoire stationnaire** bornée à un certain champ S fixé *a priori*. On imagine en outre que le champ "a été découpé arbitrairement au sein d'une régionalisation homogène, qui aurait pu se poursuivre, sans changer de nature, bien au delà des frontières" du champ (Matheron, 1963, p 17). Il n'y a donc pas d'effet de bordure. La covariance non centrée serait évaluée, sur la base de cette réalisation, par la moyenne des produits de densités tels que les deux points d'appuis tombent dans le champ, et serait ainsi donnée par :

$$c(h) = \frac{\int_{S \cap S_{-h}} z(x)z(x+h)dx}{S \cap S_{-h}}$$

Si maintenant cette réalisation est prolongée en dehors du champ par des zéros, la covariance s'écrit :

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

Enfin cette covariance $c(h)$ calculée sur la base d'une réalisation unique, constitue une réalisation de la vraie fonction de covariance $C(h)$ de la fonction aléatoire $Z(x)$. Ainsi, sous couvert de stationnarité, il est de possible d'interpréter $c(h)$ comme la covariance non centrée moyenne de $Z(x)$:

$$E(C(h)) = c(h) = \frac{g(h)}{K(h)}$$

Indépendance entre une variable et son champ

Les développements précédents fournissent une première façon de tester expérimentalement l'indépendance entre une variable et son champ. Si la fonction $c(h) = g(h)/K(h)$ se stabilise autour d'un palier, cela signifie qu'il est possible de dégager un élément purement géométrique dans la forme du covariogramme et donc dans la régionalisation elle-même. Dans ce cas, il convient de se demander si un modèle de fonction aléatoire ne serait pas plus approprié qu'une approche transitive, ce qui sous-entendrait en outre l'indépendance entre la régionalisation et son champ.

3.2.3 Covariogramme et régressions

Densité moyenne à une certaine distance d'un individu aléatoire

Dans un modèle d'individu aléatoire, la fonction

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

s'interprète comme la moyenne de $z(x+h)$ pondérée par $z(x)$, et donc, comme la valeur moyenne de la densité à la distance h d'un individu tiré au hasard. Nous reviendrons plus loin sur cette fonction intéressante (Cf. § II.4.2).

Régressions bivariables

Dans la suite de ce paragraphe, on souhaite convertir l'intégrale d'espace en une intégrale sur les valeurs de la variable régionalisée. L'objet de cette transformation est de faire apparaître les régressions entre la densité en x et la densité en $x + h$.

Pour cela, l'idée de base est de proposer une autre écriture pour $z(x)$. Ainsi, en partant de :

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

on peut écrire

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

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

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

où

$$m(h,z) = m[z(x+h)|z(x)=z]$$

est la moyenne des valeurs de $z(x+h)$ quand $z(x) = z$ (puisque l'on manipule des variables régionalisées et non des variables aléatoires, la régression est notée avec la lettre "m" de moyenne et non avec le "E" d'espérance mathématique). Ainsi, mis sous cette forme, le covariogramme est tel que, au lieu de sommer directement en x , on commence par sommer en z pour tous les x tels que $z(x) = z$.

Cette expression rigoureuse est inutilisable dans la pratique et on lui préfère une version discrète basée sur N classes de valeurs de la variable régionalisée C_i avec $i \in [1; N]$:

$$g(h) = \sum_{i=1}^N q_i m_i(h)$$

avec

$$m_i(h) = m[z(x+h)|z(x) \in C_i] \quad \text{et} \quad q_i = \int_{C_i} z dS(z)$$

Chaque régression $m_i(h)$ intervient dans le covariogramme avec un poids égal à l'abondance de sa classe. La somme des poids correspond à l'abondance totale Q , de sorte que :

$$\frac{g(h)}{Q} = \frac{\sum_i q_i m_i(h)}{\sum_i q_i}$$

Commentaires

Les régressions ne sont pas nécessairement symétriques en h comme le covariogramme. Elles autorisent donc une approche légèrement plus détaillée de la structure spatiale que ne le permet l'analyse covariographique directe. Il est d'ailleurs remarquable que la combinaison de ces **régressions dissymétriques** aboutisse à une fonction symétrique.

Partant d'une valeur de densité élevée en x on doit s'attendre à ce que, en moyenne, la densité en $x+h$ diminue pour des petites valeurs de h . Inversement, la densité doit augmenter en moyenne dans le voisinage d'une concentration faible. Pour cette raison, les régressions des premières classes seront généralement croissantes autour de l'origine et celles relatives aux dernières classes décroissantes. Or par définition, un covariogramme est décroissant à l'origine ($g(h) < g(0)$), ce qui ne peut être vérifié que si la pondération favorise les classes de densités élevées.

C'est effectivement le cas puisque l'importance d'une classe est proportionnelle à l'abondance générée par les densités qu'elle contient. Le covariogramme est en outre d'autant plus marqué par la chute des densités au voisinage des fortes concentrations que ces dernières génèrent une part importante de l'abondance totale.

Enfin, la dissymétrie des régressions se manifestant essentiellement pour les classes de densités faibles ou moyennes, l'interprétation devra porter surtout sur les régressions des classes intermédiaires pour trouver un peu plus que ce qui se trouve déjà dans le covariogramme.

3.2.4 Seuils, ensembles et fréquences conditionnelles

Soient $K_{zz'}(h)$ les surfaces des ensembles géométriques définies par les points x tels que les densités aux points x et $x+h$ dépassent respectivement les seuils (coupures) z et z' :

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

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

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

Ainsi le covariogramme transitif se décompose en toutes ces fonctions $K_{zz'}(h)$ dont l'interprétation fréquentielle est intéressante.

Notons $K_z(h) = K_{zz}(h)$ les covariogrammes simples. Alors $K_z(0) = \int 1_{z(x) > z} dx$ quantifie le nombre de densités supérieures à la coupure z . Par ailleurs $K_{zz'}(h)$ correspond au nombre de densités $z(x)$ telles que à la fois $z(x)$ et $z(x+h)$ dépassent respectivement z et z' . Le rapport de ces deux quantités s'interprète donc comme la fréquence avec laquelle $z(x+h)$ excède le seuil z'

sachant que la densité en x est supérieure à z , ce que nous noterons, dans le cadre transitif où nous nous sommes placés, fréquence conditionnelle (plutôt que probabilité conditionnelle qui fait référence à des variables aléatoires):

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

Ces fréquences conditionnelles permettent de se placer du **point de vue des fortes concentrations** (sans se préoccuper des valeurs faibles ou nulles) et en particulier de décrire le caractère attractif ou répulsif des zones à fortes concentrations. Par rapport aux régressions présentées plus haut, elles présentent l'avantage de ne pas reposer sur un découpage en classes de valeurs de densités.

3.3 Le covariogramme pour l'estimation globale et locale

Ce qui fait l'intérêt du covariogramme par rapport à d'autres fonctions structurales, c'est qu'il permet de résoudre un certain nombre de problèmes d'estimation. C'est d'ailleurs pour calculer des variances d'estimations globales que Matheron a été amené à définir le covariogramme comme nous allons le voir dans la première partie de ce paragraphe.

Mais, l'utilisation du covariogramme est également possible en estimation locale, c'est-à-dire en cartographie. Cette partie de la géostatistique transitive est assez peu connue, bien que les éléments principaux figurent dans "Estimer et choisir" (Matheron, 1978). Je lui consacrerai la deuxième partie de ce paragraphe.

3.3.1 Variance d'estimation globale à maille régulière

Expression théorique

On se place ici dans le cas où les échantillons $z(x_0 + ka)$ sont répartis sur une grille régulière de maille a et d'origine x_0 . Ils ne sont donc pas indépendants. Compte tenu des caractéristiques de la variable étudiée, et cherchant à minimiser le nombre et l'importance des hypothèses, on choisit de considérer que l'origine de la grille a été implantée aléatoirement dans la maille élémentaire de la grille. Dans ce modèle, x_0 est une réalisation de la variable aléatoire X_0 . L'estimateur de la quantité totale $Q^* = a \sum_k z(x_0 + ka)$ est une variable aléatoire. C'est en fait une fonction de la variable aléatoire X_0 qu'on notera à cet effet $Q^*(X_0)$. On a déjà montré que cet estimateur était sans biais (Cf. § II.1.2.1).

Le biais étant nul et la quantité à estimer Q étant déterministe, la variance d'estimation $\sigma^2(a)$ s'écrit de la façon suivante :

$$\begin{aligned} \sigma^2(a) &= \text{Var}(Q^*(X_0) - Q) = \text{Var}(Q^*(X_0)) \\ &= E(Q^*(X_0)^2) - E^2(Q^*(X_0)) \\ &= E(Q^*(X_0)^2) - Q^2 \end{aligned}$$

Ce calcul fait apparaître d'une part, le carré d'une somme discrète qui se transforme en une somme de doubles produits

$$Q^{*2}(x_0) = \left(a \sum_k z(x_0 + ka) \right)^2$$

$$\begin{aligned}
&= a^2 \left(\sum_k z(x_0 + ka) \right) \cdot \left(\sum_l z(x_0 + la) \right) \\
&= a^2 \sum_k \sum_l z(x_0 + ka) z(x_0 + la) \\
&= a^2 \sum_k \sum_{k'} z(x_0 + ka) z(x_0 + ka + k'a)
\end{aligned}$$

et d'autre part, le carré d'une somme continue qui se transforme de la même manière en une somme de doubles produits :

$$\begin{aligned}
Q^2 &= \left(\int z(x) dx \right) \cdot \left(\int z(y) dy \right) \\
&= \int \int z(x) z(y) dx dy \\
&= \int \int z(x) z(x+h) dx dh \\
&= \int g(h) dh
\end{aligned}$$

On notera au passage que le carré de l'abondance est égal à l'intégrale du covariogramme. Pour finir les calculs, il reste à considérer l'espérance du carré de l'estimateur. Par définition, la variable aléatoire X_0 étant uniforme dans $[0, a]$:

$$\begin{aligned}
E(Q^{*2}(X_0)) &= \int_0^a Q^{*2}(x_0) \frac{dx_0}{a} \\
&= \int_0^a a^2 \sum_k \sum_{k'} z(x_0 + ka) z(x_0 + ka + k'a) \frac{dx_0}{a} \\
&= a \int \sum_{k'} z(x_0) z(x_0 + k'a) dx_0 \\
&= a \sum_k \int z(x_0) z(x_0 + ka) dx_0 \\
&= a \sum_k g(ka)
\end{aligned}$$

Pour cela, on a utilisé le fait que la combinaison de l'intégrale sur $[0, a]$ et de la sommation discrète sur les indices k de $z(x + ka)$ peut être remplacée par une intégration étendue à tout l'espace.

Sur la base de ces développements, la variance d'estimation s'écrit :

$$\sigma^2(a) = a \sum_{k=-\infty}^{k=+\infty} g(ka) - \int g(h) dh$$

Elle s'interprète comme la différence entre l'approximation discrète sur la grille d'échantillonnage de l'intégrale de la fonction $g(h)$ et sa valeur exacte.

Commentaires

- Nécessaire modélisation.

A ce stade, il "ne manque plus" que la donnée du covariogramme pour connaître la variance d'estimation. En effet, si on se contente des valeurs expérimentales pour calculer Q^2 et $g(ka)$, on obtient une variance identiquement nulle.

C'est la preuve qu'*"il est théoriquement impossible de déduire des mêmes données expérimentales à la fois une estimation et la précision de cette estimation"* (Matheron, 1965, p. 74).

- Estimation de la variance d'estimation

Sur la base du covariogramme expérimental calculé de façon *ad hoc* (§II.3.4) on choisit une fonction mathématique autorisée à 1, 2 ou 3 dimensions selon la dimensions de l'espace (i.e. de type positif et conduisant à des variances d'estimations positives) qui ajuste au mieux le covariogramme expérimental. Ce n'est donc pas le vrai covariogramme qu'on utilise mais sa modélisation, de sorte qu'on propose en fait une **estimation de la variance d'estimation**.

S'il est illusoire de vouloir calculer la variance de la variance d'estimation, on peut cependant proposer quelques réflexions sur la robustesse de cette variance.

- Robustesse.

Il est important de noter que l'élaboration de la formule de la variance d'estimation ne requiert aucune autre hypothèse que celle concernant le choix de l'origine de la grille d'échantillonnage. De ce point de vue, la variance en géostatistique transitive est très robuste. Il s'agit plus précisément d'une robustesse constitutive, c'est-à-dire relative au modèle constitutif (variable régionalisée et grille régulière à implantation aléatoire).

Par contre, puisque seul un modèle de covariogramme est accessible, les résultats ne valent que ce que vaut le modèle. La robustesse de type et la robustesse spécifique, c'est-à-dire la robustesse liée au choix du type de covariogramme et à sa spécification sont donc moins bonnes.

Néanmoins, on montre (Matheron, 1965) que la variance d'estimation est tributaire :

- du comportement du covariogramme au voisinage de l'origine et plus précisément dans la zone $h < a$, qui est justement la zone inaccessible expérimentalement (pas d'observation pour des distances inférieures à la maille)
- du raccord du covariogramme à la portée. On ne connaît en effet la portée qu'avec une précision égale à la taille de la maille.

Les points névralgiques de la modélisation sont donc bien circonscrits : ils concernent le **comportement du covariogramme à l'origine et le choix de la portée**.

N'oublions cependant pas que le covariogramme est aussi un outil structural. Cela signifie en particulier que sa forme et son expression analytique sont reliées à la régularité spatiale de la variable régionalisée étudiée. Cette propriété constitue une aide précieuse pour la modélisation. Le comportement du covariogramme à l'origine est en fait contraint par les connaissances directes disponibles sur le phénomène. C'est pour cette raison que les praticiens préfèrent choisir le comportement du covariogramme à l'origine plutôt que de le faire estimer par une procédure automatique nécessairement ignorante de la réalité physique du phénomène sous-jacent.

- Régularité.

La variance est égale à la différence entre la valeur exacte de l'intégrale de la fonction $g(h)$ et son approximation discrète sur la grille d'échantillonnage. Elle est par conséquent d'autant plus

faible que la maille est petite, ou encore, ce qui revient au même, que l'effort d'échantillonnage est important.

Mais elle est également d'autant plus faible que la fonction $g(h)$ est régulière. En effet, plus le covariogramme présente de fluctuation et plus le calcul approché de son intégrale est différente de la valeur vraie (Fig. 3.5).

Par conséquent, la **régularité du covariogramme conditionne, à effort d'échantillonnage fixé, le niveau de la variance d'estimation.**

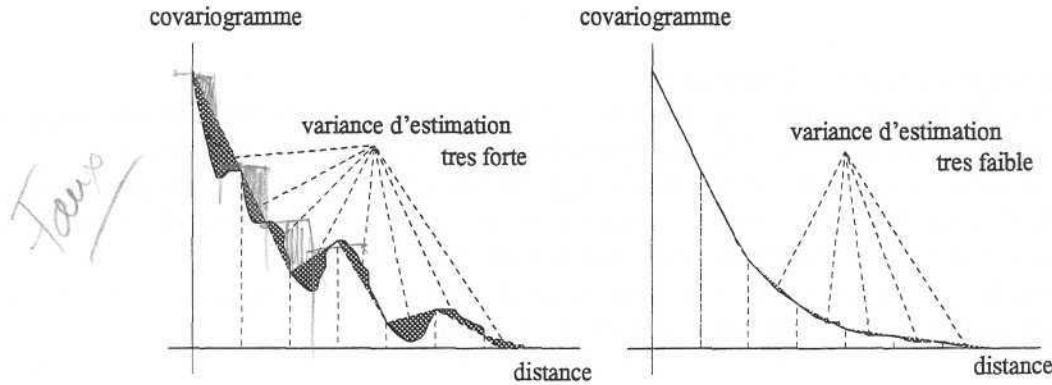


FIG. 3.5 - Variance d'estimation globale en géostatistique transitive. Influence de la régularité du covariogramme.

3.3.2 Krigeage transitif

On aborde ici une tout autre question. Disposant de valeurs de densités en différents points de l'espace, on cherche à estimer la valeur de la variable régionalisée en tous points de l'espace (estimation locale).

Equations du krigeage transitif

En chaque point x de l'espace à cartographier, on se propose d'estimer $z(x)$ à l'aide d'une **moyenne pondérée** des valeurs échantillonnées (zéros compris) à l'intérieur d'un **voisinage** donné de x , soit en des points $x + h_i = x_i$ (Matheron, 1968 et 1978; Rivoirard, 1995):

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

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

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

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

Chercher le minimum de cette fonction quand les λ_i sont inconnus, revient à chercher les valeurs de λ_i qui annulent les dérivées partielles par rapport aux λ_i . Les λ_i solutions vérifient alors le système linéaire suivant :

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

qui comporte autant de lignes que le voisinage du point à estimer comprend d'échantillons. Pour des raisons analogues à celles rencontrées dans le calcul de variance d'estimation globale (produits de densités distantes de h), le covariogramme intervient dans la formulation du **krigeage transitif**. Il permet d'attribuer un **poids** aux points du voisinage en **fonction de la structure spatiale** de la variable régionalisée et en fonction de leurs distances du point à estimer.

Le krigage transitif peut également être conduit **sous contrainte**. On peut en effet choisir les poids λ_i sous la contrainte que l'abondance totale Q soit respectée dans les différentes translations imaginées :

$$\int z^*(x) dx = \int z(x) dx$$

ce qui donne

$$\sum_j \lambda_j = 1$$

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

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

On notera que, avec ou sans contraintes, les poids ne changent pas si l'on multiplie ou si l'on divise le covariogramme par une constante (en considérant par exemple le covariogramme $g(h)/Q$ ou encore $g(h)/Q^2$).

Choix du voisinage

• Voisinage unique.

•• **Principe.** Remarquons tout d'abord qu'il n'y a pas de limitation théorique à la taille de voisinage. A cet égard, on peut décider de faire le krigage en conservant l'ensemble des points du champ. On dira alors qu'il s'agit d'un krigage à **voisinage unique**. En choisissant de procéder ainsi, on juge qu'il est intéressant et justifié qu'un point très éloigné participe à l'estimation. On notera cependant que son poids sera très faible conformément à la valeur du covariogramme à cette distance.

•• **Système dual.** En voisinage unique, il existe une expression **duale** du krigeage où “l’estimateur est une combinaison linéaire des $g(h_i)$ ” (Chauvet, 1994) :

$$z^*(x) = \sum_i b_i g(x_i - x) + c = \sum_i b_i g(h_i) + c$$

qui aboutit au système de krigeage dual suivant :

$$\begin{cases} \sum_i b_i g(h_i - h_j) + c = z(x_j) & \forall j \\ \sum_i b_i = 0 \end{cases}$$

•• **Qualité du covariogramme.** Dans les estimations par krigeage, le covariogramme n’intervient que pour des distances inférieures ou égales à la dimension du voisinage. Ainsi, si le comportement du covariogramme aux grandes distances est jugé peu fiable, on souhaitera probablement réduire la taille du voisinage.

• Voisinage glissant

•• **Principe.** Lorsqu’on ne prend que les points appartenant à un voisinage (non unique) du point à estimer on dit qu’on procède par **voisinage glissant**. Reste alors à choisir la taille, la forme et l’orientation du voisinage.

•• Taille du voisinage.

En utilisant un voisinage assez grand relativement à la taille du champ, on risque de **diluer** les densités de bordure. Cette dilution sera d’ailleurs d’autant plus marquée que les fortes concentrations sont localisées en bordure de champ, comme cela est parfois le cas en halieutique. Une solution consiste à ne réaliser le krigeage que dans un domaine, délimité par nos soins, où l’on juge raisonnable d’obtenir une densité non nulle.

Par contre, avec un voisinage petit, la participation ou non d’une très forte valeur peut modifier notablement l’estimation et il se pose dans ce cas des problèmes de **raccordements**. De ce point de vue, il vaudrait donc mieux chercher à ne pas trop diminuer la taille du voisinage.

La taille du voisinage doit donc, dans la mesure du possible, permettre un **équilibre** entre les problèmes de raccordements et les problèmes de dilution.

•• **Orientation et forme du voisinage.** De façon assez logique on cherchera à utiliser une forme de voisinage qui tienne compte de l’anisotropie de la variable régionalisée telle que révélée par l’analyse covariographique. Dans le cas d’une anisotropie géométrique, on peut par exemple choisir un voisinage en forme d’ellipse dont les caractéristiques (orientation et excentricité) sont identiques à celle de l’ellipse d’anisotropie.

Krigeage et cartographie

Le formalisme transitif ne permet **pas** de proposer une **variance** d’estimation locale. Le krigeage transitif est donc essentiellement un outil de **cartographie**.

Cependant la recherche d’une erreur minimale conduit inévitablement à un **lissage**. En effet, d’une façon générale, l’erreur quadratique s’écrit :

$$\int (z(x) - z^*(x))^2 dx = g(0) - 2 \sum_i \lambda_i g(h_i) + \sum_i \sum_j \lambda_i \lambda_j g(h_i - h_j)$$

Or, si $z^*(x)$ est obtenu par krigeage, on sait que :

$$\sum_j \lambda_j g(h_i - h_j) = g(h_i)$$

Donc,

$$\int (z(x) - z^*(x))^2 dx = \underbrace{g(0)}_{\int z(x)^2 dx} - \underbrace{\sum_i \lambda_i g(h_i)}_{\int z^*(x)^2 dx} > 0$$

d'où :

$$\int z(x)^2 dx > \int z^*(x)^2 dx$$

Cette dernière expression indique bien que la dispersion des valeurs krigées est inférieure à celle de la variable régionalisée.

3.4 Estimations et modélisation

Jusqu'à présent, nous avons rencontré de nombreuses raisons de nous féliciter de l'utilisation d'outils statistiques comme le covariogramme à base de sommes plutôt que de moyennes. La question que j'aborde ici, relative à l'estimation du covariogramme, est par contre rendue plus délicate, précisément à cause du recours aux sommes. On conçoit en effet assez bien que la seule situation confortable pour l'estimation d'une somme est celle d'un plan d'échantillonnage régulier. Or, il existe de nombreux cas où l'échantillonnage est irrégulier. Si l'on souhaite que les statistiques par individu, et le covariogramme en particulier, conservent leur pertinence il faut donc se doter de moyens d'estimations pour tous les types d'échantillonnage.

Pour les statistiques non spatiales (abondance par exemple) nous avons déjà avancé des éléments de solution faisant appel aux surfaces d'influences des échantillons. Cependant, pour les statistiques spatiales comme le covariogramme, une pondération simple ne suffit pas. Il faut en effet tenir compte de l'importance relative des deux points d'appuis.

Dans ce paragraphe, je commencerai par le cas idéal d'un échantillonnage régulier pour ensuite aborder le délicat problème de l'estimation du covariogramme à maille irrégulière. Dans un dernier temps, je rappellerai les contraintes qui pèsent sur le choix d'un modèle de covariogramme.

3.4.1 A maille régulière

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

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

Dans cette estimation la valeur de chaque échantillon compte avec la même surface d'influence. Puisque toutes les surfaces d'influence sont identiques, dans une paire (z_i, z_j) , l'échantillon z_j couvre l'ensemble de sa surface d'influence quand l'échantillon z_i couvre la sienne. Le cas d'une maille régulière constitue donc un cas favorable pour l'estimation.

3.4.2 A maille irrégulière

Estimation du covariogramme par migration sur une grille

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

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

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

Vers une pondération par les surfaces d'influence

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

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

où S_i est la surface d'influence de l'échantillon z_i et S_{i-h} son translaté du vecteur $-h$. Tout se passe comme si la variable s'écrivait sous la forme $z(x) = \sum_i z_i 1_{x \in S_i}$.

Un produit de densités observé *in situ* pour une certaine distance, est donc affecté, dans ce calcul, à toute une gamme de distances; l'éventail des distances étant lié aux formes et positions relatives des surfaces d'influence.

Si maintenant on considère toutes les classes de distance, en faisant la somme du covariogramme, on remarque que la contribution globale d'une paire de densités (z_i, z_j) pour l'ensemble des directions et distances possibles est égale au produit des surfaces d'influence $S_i S_j$:

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

Les développements ultérieurs vont s'inspirer de cette règle.

Estimation du covariogramme par la covariance.

Cette estimation se fait en deux temps:

1) estimation du covariogramme géométrique du champ en informant l'indicatrice du champ sur une grille.

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

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

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

$$g^*(h) = K^*(h) \frac{\sum_{j \sim i+h} z_i z_j S_i S_j}{\sum_{j \sim i+h} S_i S_j}$$

où $j \sim i+h$ donne les paires à peu près distantes de h . Cependant, les classes de distance ne doivent pas être utilisées pour l'estimation à distance strictement nulle, donnée directement par :

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

La pondération par les surfaces d'influence permet de tenir compte éventuellement du caractère préférentiel de l'échantillonnage (plus dense dans les zones *a priori* riches). Mais dans ce cas, la structure expérimentale aux petites distances est biaisée et ne doit pas être calculée.

Une autre précaution concerne la multiplication du covariogramme géométrique et de la covariance. Compte-tenu de son mode d'estimation, le covariogramme géométrique est disponible pour l'ensemble des pas d'une grille fine. Par contre, la covariance est estimée pour des pas beaucoup plus grands qui ne correspondent pas nécessairement à ceux du covariogramme géométrique. Deux solutions sont envisageables :

- On peut prendre la valeur du covariogramme géométrique pour la distance la plus proche possible de celle où la covariance est connue. Cette méthode est valable si la grille de migration est très fine.

- On peut également modéliser le covariogramme géométrique et prendre directement la valeur indiquée par le modèle pour la distance voulue.

Estimation du covariogramme par les régressions.

Ecrire le covariogramme (§II.3.2.3):

$$g(h) = \frac{\sum_i q_i m_i(h)}{\sum_i q_i} = \frac{g(h)}{Q} \quad \leftarrow !$$

suggère d'estimer d'abord les régressions $m_i(h) = m[z(x+h)|z(x) \in C_i]$. Si l'on travaille par classes de distances, on peut, à distance h d'un point z_i , trouver plusieurs valeurs z_j (positives ou nulles, peu importe). A ces valeurs z_j , on affecte des extensions S_j plus ou moins fortes, dont il est naturel de tenir compte. Ainsi, on propose d'estimer $m_i^*(h)$ par l'expression compliquée suivante :

$$m_i^*(h) = \frac{\sum_{z_i \in C_i} S_i \frac{\sum_{j \sim i+h} z_j S_j}{\sum_{j \sim i+h} S_j}}{\sum_{z_i \in C_i} S_i}$$

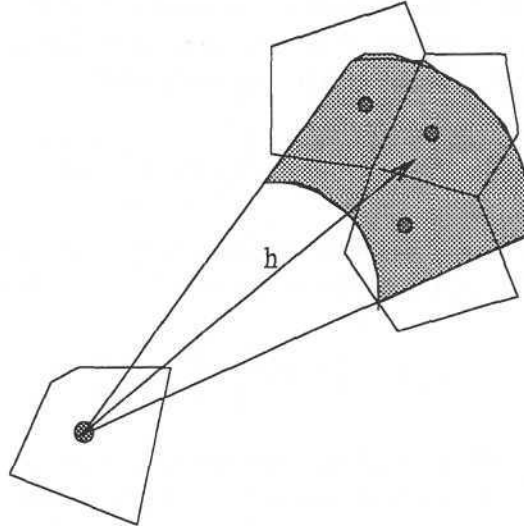


FIG. 3.6 - Tolérance sur la direction de calcul et surfaces d'influence. Estimation des régressions : un échantillon z_i et 3 échantillons z_j dans la classe de distance.

Le traitement des pondérateurs q_i est plus simple. Il s'agit d'estimer l'abondance associée aux densités appartenant à une classe. La question de l'estimation d'abondance a déjà été abordée. On ne fait là que l'appliquer par classe, ce qui donne :

$$q_i^* = \sum_{z_i \in C_i} z_i S_i$$

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

3.4.3 Contraintes pour la modélisation

Le covariogramme expérimental obtenu par l'une ou l'autre de ces méthodes doit ensuite être modélisé. C'est-à-dire ajusté par une courbe qui respecte au mieux les résultats expérimentaux et qui réponde aux contraintes théoriques d'un covariogramme.

Nous avons déjà expliqué les raisons qui poussent en général, le géostatisticien à préférer contrôler manuellement l'ajustement plutôt qu'à le confier à une procédure automatique. Pour autant, l'utilisateur n'est pas totalement libre. Le modèle qu'il choisit doit en effet respecter les propriétés des covariogrammes dont les principales sont rappelées maintenant.

Contraintes théoriques

• **Signe**

Appliqué à une variable positive ou nulle, le covariogramme est, lui aussi, positif ou nul. Il faut donc choisir des **fonctions positives ou nulles** quelle que soit la distance. En particulier, un ajustement polynomial automatique ne conviendra pas, puisqu'il risquerait d'être négatif par endroit, et surtout, comme nous le verrons plus loin, de ne pas être de type positif.

• **Portée**

L'existence d'une **portée finie** conduit souvent à utiliser des **fonctions définies par morceaux** (c'est-à-dire une certaine fonction jusqu'à la portée où elle s'annule d'ailleurs, et 0 ensuite). Les fonctions structurales ayant une pseudo-portée comme le schéma exponentiel ou le schéma gaussien, ne sont donc pas utilisables.

• **Type positif**

Cette propriété constitue une contrainte importante puisqu'elle nous oblige à n'utiliser que des fonctions de **type positif**. Or, démontrer qu'une fonction est de type positif est, en général, très laborieux. On préfère donc utiliser des fonctions dont on a démontré une fois pour toute qu'elles sont de type positif. Cela réduit sensiblement la gamme des fonctions disponibles mais assure la cohérence interne du modèle tout au moins de ce point de vue.

• **Somme unité**

Lorsqu'il est considéré comme la densité de probabilité de la distance entre deux individus aléatoires ($g(h)/Q^2$), le modèle de covariogramme relatif doit être tel que son intégrale sur tout l'espace est égale à 1: $\int \frac{g(h)}{Q^2} dh = 1$

Dans le cas précis d'un modèle sphérique isotrope de palier C et de portée a , cela revient à imposer que :

$$Ca^2\pi = 5Q^2$$

En cas d'anisotropie géométrique de coefficient λ , la portée dépend de la direction

$$a(\theta)^2 = a^2(\cos^2(\theta) + \frac{\sin^2(\theta)}{\lambda})$$

et la condition s'écrit :

$$Ca^2\pi(\lambda^2 + 1) = 10\lambda^2$$

Faux c'est le résultat de l'équation pour une ellipse ou un angle polaire.

$\frac{Ca^2\pi}{5\lambda} = 1$ *Faux*

Contraintes physiques

Le covariogramme doit aussi être cohérent à l'origine, c'est-à-dire respecter les propriétés physiques de continuité de la variable. A cette étape, on doit faire la synthèse des connaissances biologiques et techniques sur l'échantillonnage pour choisir le comportement du covariogramme à l'origine. On se souviendra que selon la technique d'estimation utilisée (migration vers une grille fine), le comportement linéaire du covariogramme à l'origine peut être artificiel.

Contraintes statistiques

Une fois la forme du covariogramme choisie, une fois la fonction structurale intuitée, il reste à spécifier le modèle, c'est-à-dire à choisir la valeur des paramètres qui définissent la fonction. Cette étape correspond à l'inférence du modèle de covariogramme. On évitera donc de choisir des fonctions ayant trop de paramètres. De fait, les fonctions de type positif utilisées en géostatistique, possèdent au plus 3 paramètres (modèle sphérique, exponentiel, gaussien, linéaire, ...).

3.4.4 Covariogrammes: le vrai et le modélisé

Le vrai covariogramme, c'est-à-dire celui qu'on calculerait si on connaissait exhaustivement la réalité, est une fonction très riche, présentant de multiples fluctuations à très grandes échelles (très petites distances). *“Le covariogramme - si nous le connaissions parfaitement - présenterait de petites ondulations, des points singuliers, toute une structure de détails qui fournirait autant d'informations, ou presque, que la donnée de la variable régionalisée elle-même. Mais cette riche structure de détail n'est jamais accessible expérimentalement.”* (Matheron, 1970, pp. 39).

Dans notre démarche, le **vrai covariogramme, très riche**, se trouve donc **modélisé par une fonction beaucoup plus régulière**. En pratique, le lissage se justifie par exemple par le fait que la réalité (i.e. la régionalisation étudiée) aurait pu être légèrement différente et générer une forme un peu différente de covariogramme. En somme, ce passage d'une fonction très complexe à une fonction régulière, *“constitue un équivalent camouflé d'un passage à l'espérance mathématique. Ainsi, les méthodes transitives, qui se voulaient au départ purement géométriques, se révèlent, lorsque l'on analyse les conditions de leur mise en œuvre effective, comme riches d'un contenu probabiliste implicite”* (Matheron, 1970).

Pour autant, ce contenu probabiliste implicite ne justifie pas, à lui seul, le recours aux fonctions aléatoires. Nous avons vu dans la première partie du document, que le cadre probabiliste des fonctions aléatoires est plus exigeant.

3.4.5 Remarques sur les “simulations”

Simulations en géostatistique intrinsèque

Pourquoi chercher à faire des simulations? et de quoi?

Nous avons vu que le krigeage a tendance à lisser la dispersion réelle de la variable régionalisée. Or il importe parfois, non pas d'obtenir une estimation locale la meilleure possible, mais plutôt de reproduire la variabilité du phénomène. Pour cela, on a l'habitude en géostatistique intrinsèque de procéder par simulation.

Partant du principe que plusieurs variables régionalisées peuvent avoir le même variogramme, plusieurs méthodes existent aujourd'hui pour simuler ce que pourrait être une réalisation d'une fonction aléatoire de variogramme connu. En particulier, connaissant la valeur d'un phénomène en un certain nombre de points, on peut estimer et modéliser le variogramme correspondant et contruire des variables régionalisées qui respectent à la fois les valeurs échantillonnées et le modèle de variogramme. Ce sont des simulations conditionnelles.

Une conjecture défavorable

La question est: **sait-on simuler une variable régionalisée de covariogramme connu**, conditionnellement ou non à un réseau de valeurs connues? et cela a-t-il un sens?

Simuler une variable régionalisée de covariogramme donné sous-entend, comme précédemment, que deux variables régionalisées différentes pourraient avoir le même covariogramme. Or selon une conjecture formulée par Matheron, il semblerait que le covariogramme, le vrai, caractérise complètement une variable régionalisée. En somme : une variable régionalisée, un covariogramme.

La question doit donc être reformulée ainsi : sait-on **construire** une variable régionalisée de covariogramme connu ? Mais dans le même temps, la question perd sa pertinence. En effet, dans la pratique, on ne connaît pas le vrai covariogramme faute de ne pas connaître exhaustivement la variable régionalisée. Donc soit on connaît exhaustivement la variable régionalisée et point n'est alors besoin de la re-construire, soit on ne la connaît que partiellement et le vrai covariogramme manque.

Compte-tenu de la conjecture énoncée plus haut, l'utilisation des modèles de covariogramme ne permettrait la construction que d'un nombre très limité de variables régionalisées différentes. Comme cela est souvent le cas, imaginons que le modèle de covariogramme associé à la variable régionalisée à construire soit un schéma sphérique de portée a . Un objet, ou plutôt l'objet, répondant à cette caractéristique est une sphère de diamètre a . Cette sphère est donc la variable régionalisée de covariogramme sphérique et n'a pas grand chose à voir avec une densité de poissons. C'est là une conséquence forte de la différence entre le vrai covariogramme et le covariogramme modélisé.

Chapitre 4

Densité par individu, effet de support et régularisation

Sommaire

*L'effet de support, traité au **paragraphe 1**, est une notion importante de la géostatistique. Je commence par rappeler le principe selon lequel la loi d'une variable dépend du support auquel se font les observations. On montre ensuite l'effet de la régularisation (augmentation du support) sur le covariogramme.*

*Dans le **paragraphe 2** on définit la densité par individu. Cette extension des statistiques par individu à la densité elle-même présente l'avantage de fournir des statistiques sur la densité indépendantes du champ. Ces statistiques ont cependant l'inconvénient de dépendre du support comme cela est indiqué par une comparaison rapide des deux approches possibles, c'est-à-dire "classique" (i.e. par échantillon) ou par individu.*

*Enfin, le **paragraphe 3** utilise le bénéfice conventionnel, dont on rappelle la définition, pour étudier l'effet de support.*

Density per individual, support effect and regularization.

Abstract

*The support effect which is the purpose of the **paragraph 1**, is an important concept proposed by geostatistics. I first explain the principle by which the probability law of a variable depends on the size of support of the observations. The larger the support the more regular the variable. We then analyse how the covariogram varies with the regularization of the variable, that is the increase of the support.*

*In **paragraph 2** we define the density per individual which is an extension of the statistics per individual. This gives the opportunity to calculate statistics about the density itself which are independent of the field. But, these statistics, and in particular the mean*

density per individual, vary with the support. At the end, we compare briefly the two approaches: the "classical" one (density per sample) and the density per individual.

*Finally, in **paragraph 3** it is suggested to use the conventional benefit whose definition is recalled, to study the effect of the support on fish data.*

Après avoir considéré la position, puis la valeur d'un paramètre de l'environnement, d'un individu tiré au hasard, il est logique de s'intéresser à la densité d'un individu aléatoire. Cette notion n'est cependant pas aussi simple que les précédentes. En particulier, la densité par individu, et en particulier sa moyenne, dépend du support.

La notion de support est bien connue en géostatistique minière, mais probablement beaucoup moins en halieutique. Même si le terme de "support" est assez évocateur, il m'a paru tout de même nécessaire de commencer ce chapitre par un rappel sur l'effet de support et la régularisation. Je présenterai par la suite le concept de densité par individu et enfin, dans la troisième partie de ce chapitre, le bénéfice conventionnel qui permet d'étudier l'effet de support.

4.1 Effet de support : principe et définitions

4.1.1 Principe

Le plus souvent, on ne dispose pas expérimentalement des valeurs ponctuelles de densités, mais seulement des valeurs moyennes dans un certain volume. C'est le cas de tous les prélèvements par chalutage par exemple puisqu'on récupère le nombre de poissons se trouvant à l'intérieur de l'aire ou du volume chaluté. "*De tels prélèvements définissent une nouvelle variable, différente de $z(x)$, et dont les caractéristiques dépendent de la taille et de la forme du prélèvement*" (Matheron, 1965, pp. 25). Le volume du prélèvement est appelé **support** et la nouvelle variable sera dite **régularisée de $z(x)$ par le support**. La variable $z(x)$ elle-même est dite à support ponctuel.

En pratique cependant, x désigne généralement le support, considéré comme quasi-ponctuel, de l'échantillon. Cette remarque justifie l'ensemble de ce qui a été présenté jusqu'ici.

On conçoit assez bien qu'une densité présente d'autant moins de valeurs nulles et très fortes que la surface chalutée (i.e. le support) est grande. En somme, on doit s'attendre à ce qu'une densité soit **plus régulière dans sa variation spatiale** et également **moins contrastée dans sa distribution statistique** quand le support augmente; ce qui justifie le terme de variable régularisée. On retiendra donc que **la variable régularisée n'a pas les mêmes caractéristiques que la variable ponctuelle**. C'est l'**effet de support**.

En mélangeant des valeurs obtenues pour des prélèvements de tailles différentes, ce qui est typiquement le cas lorsqu'on utilise des captures commerciales, on mélange en fait, dans le modèle, des variables régionalisées différentes. Reste alors à savoir si l'effet du support est tel qu'il peut être négligé ou non. Pour cela, il faut se doter de moyens permettant d'évaluer l'effet du support.

4.1.2 Régularisation des variables

Notations

Les deux variables régularisées les plus répandues sont la densité et l'abondance de poissons par action de pêche. Elles sont définies respectivement à partir de la variable ponctuelle $z(x)$ par :

$$\begin{cases} z_v(x) = \frac{1}{v} \int_{v(x)} z(x+h)dh & \text{la densité} \\ q_v(x) = \int_{v(x)} z(x+h)dh & \text{l'abondance} \end{cases}$$

où $v(x)$ représente l'aire chalutée centrée en x et v la surface chalutée.

Régularisation du covariogramme

Pour montrer l'effet de la régularisation sur le covariogramme il est pratique d'utiliser les produits de convolution. Je rappelle que le covariogramme est le produit de convolution de $z(x)$ et de sa transposée $\tilde{z}(x) = z(-x)$: $g = z * \tilde{z}$. Ainsi, en remarquant que les deux régularisées sont aussi des produits de convolution, on montre (Matheron, 1965, pp. 26-27) que le covariogramme des régularisées se déduit du covariogramme de la variable ponctuelle par les relations suivantes :

$$\begin{cases} g_{z_v} = \frac{1}{v^2} g * v * \tilde{v} \\ g_{q_v} = g * v * \tilde{v} \end{cases}$$

En particulier, cela implique que $g_{z_v}(0)$, la valeur du covariogramme de la variable régularisée $z_v(x)$ en 0, représente la valeur moyenne du covariogramme $g(h)$ de la variable ponctuelle $z(x)$ lorsque les extrémités du vecteur distance h balayent, chacune, le support de prélèvement. Comme $g(h)$ est borné par $g(0)$, on voit ainsi la propriété importante suivante :

$$g_{z_v}(0) \leq g(0)$$

Comme par ailleurs, l'abondance totale est la même ($\int z(x)dx = \int z_v(x)dx$), on a une manifestation de la régularisation en se souvenant que la valeur du covariogramme relatif à l'origine ($g(0)/Q^2$) mesure le niveau d'agrégation.

Réduction de variance

La variance (au sens statistique) d'une variable régionalisée de moyenne m sur son champ S est donnée par :

$$\begin{aligned} \text{var}(z(x)) &= \frac{1}{S} \int (z(x) - m)^2 dx \\ &= \frac{1}{S} \left(\int z(x)^2 dx - 2m \int z(x) dx + m^2 \right) \\ &= \frac{g(0)}{S} - m^2 \end{aligned}$$

De la même façon, la variance de la variable régularisée s'écrit :

$$\text{var}(z_v(x)) = \frac{g_{z_v}(0)}{S} - m^2$$

Elle est donc **inférieure** à la variance ponctuelle. On montre ainsi que la **variance statistique diminue avec la taille du support**.

Qu'en est-il des statistiques par individu? Et, tout d'abord, voyons quel sens donner à la densité par individu.

4.2 Densité par individu

4.2.1 Définitions

Centre de gravité et inertie des densités

Plutôt que de s'intéresser à la position ou la valeur d'un paramètre de l'environnement à l'endroit d'un individu tiré au hasard, on peut noter la valeur de la densité de poissons et former la **densité par individu** (notée z_I). Cette façon de s'intéresser aux densités donne de fait plus d'importance aux concentrations élevées puisqu'il y a d'autant plus de chances de tirer un individu en x que la densité en x est importante. Mais surtout, elle permet, sans faire référence au champ, de définir une densité **moyenne** par individu, encore appelé **centre de gravité des densités** qui s'écrit simplement comme le rapport entre la valeur du covariogramme à l'origine et l'abondance totale :

$$E(z_I) = \bar{z}_I = \int z(x) \frac{z(x)}{Q} dx = \frac{g(0)}{Q}$$

Le centre de gravité indique à quel niveau de densité se situe en moyenne un individu d'une population. Dans le prolongement, la **variance** de la densité par individu, encore appelée **inertie** de la densité par individu ou, plus simplement, **inertie des densités**, s'exprime de la façon suivante :

$$var(z_I) = \int (z(x) - \bar{z}_I)^2 \frac{z(x)}{Q} dx$$

Surface équivalente

L'intérêt de cette démarche est de disposer, à propos de la densité, de statistiques **indépendantes du champ**. Mais, ayant préféré ne pas délimiter l'aire d'occupation de la population, on manque précisément d'information sur la surface de la population. Il est cependant possible de s'intéresser à la surface qu'occuperait la population si tous ses individus avaient un même comportement moyen, c'est-à-dire si la densité était constante et égale à la densité moyenne $E(z_I)$. L'abondance de la population restant la même, cette **surface équivalente** est donnée très simplement par l'équation suivante :

$$S_e = \frac{Q}{E(z_I)} = \frac{Q^2}{g(0)}$$

Cette expression donne un sens à la portée intégrale vue précédemment (Cf. § II.3.2.1) qui apparaît ici comme la surface équivalente. On notera aussi que l'inverse de la surface équivalente n'est autre que $\frac{g(0)}{Q^2}$, le niveau d'agrégation.

4.2.2 Densité par individu pour une régularisée

Sens physique

Si l'information est disponible à l'échelle d'un coup de chalut, il n'est plus possible de remonter à la densité à l'endroit d'un individu aléatoire; on ne peut rien considérer d'autre que la densité au support des captures.

Imaginons que l'ensemble de la population ait été chaluté. La démarche revient donc à tirer au hasard un individu, à noter à quel coup de chalut il "appartient" et quelle est la densité de ce coup de chalut. C'est cette valeur de densité qui donnera la **densité, au support considéré, d'un individu aléatoire**.

Centre de gravité

En passant d'une variable ponctuelle $z(x)$ à une densité régularisée $z_v(x)$, la teneur moyenne sur le support d'un individu aléatoire $E(z_{v_I})$ est :

$$E(z_{v_I}) = \frac{g_{z_v}(0)}{Q}$$

D'après les résultats énoncés précédemment quant à la régularisation du covariogramme, il apparaît que la densité régularisée moyenne est inférieure à la densité ponctuelle moyenne :

$$E(z_{v_I}) = \frac{g_{z_v}(0)}{Q} \leq \frac{g(0)}{Q} = E(z_I)$$

et ce, d'autant plus que le support est important et/ou que la pente du covariogramme de la densité ponctuelle à l'origine est forte. C'est donc aussi le cas du niveau d'agrégation.

La densité moyenne par individu et l'agrégation diminuent donc avec la taille du support. Cette propriété est due au fait que ces statistiques reposent sur un critère de dispersion puisque faisant intervenir la valeur du covariogramme en 0. En étant sensibles à la dispersion, elles deviennent automatiquement sensibles à la taille du support, ce qui retire un peu d'intérêt.

La densité moyenne usuelle, c'est-à-dire la moyenne des $z_v(x)$ sur un champ qui contiendrait toute la population est, par contre, invariante avec le support. Cela provient de ce qu'on récupère, en moyenne, une quantité de poisson proportionnelle à la taille du support. On peut donc mettre en parallèle deux façons d'appréhender une densité de poissons et en tirer quelques généralités.

4.2.3 Les deux types de loi

Définitions

- La densité par individu z_I est une variable aléatoire définie sur l'ensemble des valeurs de la variable régionalisée. La fonction de répartition $F_{z_I}(z)$ de cette variable aléatoire est définie analytiquement par :

$$1 - F_{z_I}(z) = \frac{1}{Q} \int z(x) \mathbb{I}_{z(x) \geq z} dx = \frac{1}{Q} \int_{S(z)} z(x) dx = \frac{Q(z)}{Q}$$

où $S(z)$ est le champ occupé par les densités **supérieures ou égales** au seuil z , et $Q(z)$ leur abondance. De là, la probabilité pour que la densité par individu soit comprise entre z_1 et z_2 est :

$$\mathcal{P}(z_1 \leq z_I \leq z_2) = \frac{Q(z_1) - Q(z_2)}{Q}$$

c'est-à-dire l'abondance relative de la classe $[z_1, z_2[$. La loi de z_I est donc accessible expérimentalement au travers de l'histogramme des abondances par classe de densité, relativement à l'abondance totale, ou encore l'histogramme des **densités par individu**. Il s'ensuit également que la probabilité infinitésimale $f_{z_I}(z)dz$ construite comme il se doit à l'aide de la densité de probabilité $f_{z_I}(z)$, est égale à l'abondance $q(z)dz$ des densités comprises entre z et $z + dz$:

$$f_{z_I}(z)dz = \frac{q(z)dz}{Q}$$

• Dans un champ donné, l'histogramme usuel du nombre d'échantillons par classe de densité, relativement au nombre total d'échantillons, ou encore l'histogramme des **densités par échantillon**, permet d'estimer la surface relative (de l'aire occupée par) des densités comprises entre z_1 et z_2

$$\mathcal{P}(z_1 \leq z \leq z_2) = \frac{S(z_1) - S(z_2)}{S}$$

ou encore la surface de la classe $[z_1, z_2[$. En conservant les notations précédentes, la fonction de répartition $F_z(z)$ de cette variable est :

$$1 - F_z(z) = \frac{1}{S} \int \mathbb{I}_{z(x) \geq z} dx = \frac{1}{S} \int_{S(z)} dx = \frac{S(z)}{S}$$

et la probabilité infinitésimale est telle que :

$$f_z(z)dz = \frac{dS(z)}{S}$$

où $dS(z)$ est la surface des densités comprises entre z et $z + dz$.

densité	notation	loi	probabilité infinitésimale
par échantillon	$z(x)$	$1 - \frac{S(z)}{S}$	$\frac{dS(z)}{S}$
par individu	z_I	$1 - \frac{Q(z)}{Q}$	$\frac{q(z)dz}{Q}$

TAB. 4.1 - Comparaison des deux types de densités.

Comparaison

Le résumé sous la forme d'un tableau (Tab. 4.1) montre bien la nature de la différence entre les deux approches : les considérations de natures surfaciques (relatives à la surface du champ) sont simplement remplacées par des abondances.

L'approche par échantillon, la plus usuelle, est relative à un champ S . Elle aboutit à des histogrammes très dissymétriques (exemple, Fig. 4.1, du merlu d'âge 0 dans le Golfe de Gascogne) ayant des coefficients de variations élevés et un fort "effet 0" (proportion de densités nulles importante). La moyenne est indépendante de la taille du support. La dispersion est d'autant plus faible que la taille du support est importante.

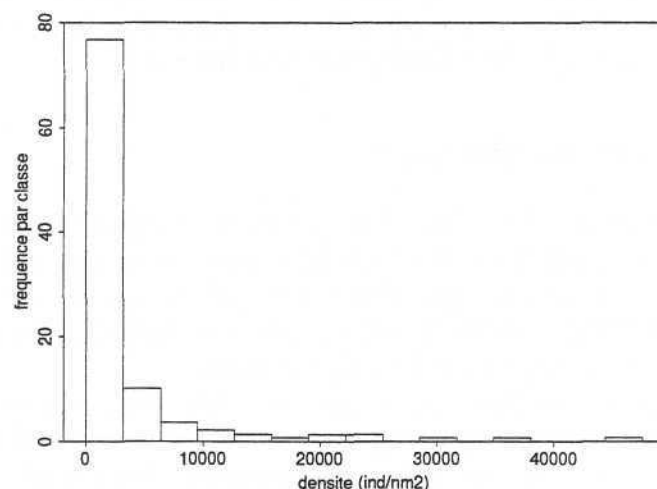


FIG. 4.1 - Exemple d'histogramme des densités par échantillon. Mêmes données que pour la Fig. 4.2

L'approche par individu est indépendante de la définition d'un champ. Elle aboutit à des histogrammes assez homogènes ayant un coefficient de variation faible, avec en particulier une absence d'effet 0 (exemple, Fig. 4.2, du merlu d'âge 0 dans le Golfe de Gascogne). La moyenne et la variance dépendent cependant du support.

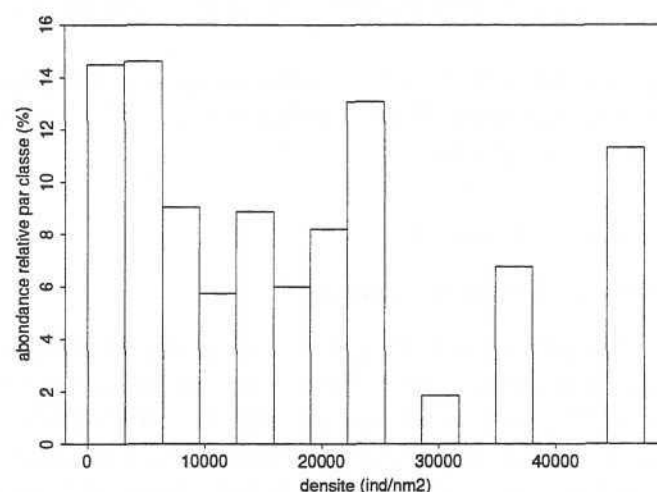


FIG. 4.2 - Exemple d'histogramme des densités par individu. Mêmes données que pour la Fig. 4.3

Le plus intéressant est probablement de mener de front les deux approches. Cela permet en particulier de comparer l'importance d'une classe de densité en termes d'effectif d'une part et en termes d'abondance d'autre part. Par cette comparaison, on dérive vers une autre façon de décrire une distribution qui s'apparente aux **courbes de sélectivité**. Ces courbes ont été développées dans le but d'étudier les effets de supports.

Conformément aux choix méthodologiques que nous avons faits pour analyser des densités de poissons (géostatistique transitive et intégrales d'espace), la courbe de sélectivité la plus appropriée semble être une version transitive du **bénéfice conventionnel**.

4.3 Le bénéfice conventionnel

Les courbes de sélectivité ont été développées dans le contexte des variables aléatoires. Elles sont autant d'alternatives possibles à l'histogramme pour déterminer la loi de probabilité d'une variable aléatoire. Mais surtout, elles aboutissent à la définition de la sélectivité qui constitue un outil approprié pour étudier et modéliser les changements de support. Pour plus de renseignements sur ces fonctions on pourra se reporter à Lantuéjoul (1990).

Les courbes de sélectivité souffrent cependant des mêmes inconvénients que les autres statistiques usuelles, puisqu'elles procèdent par moyenne au sein d'un domaine de l'espace pré-déterminé. On se propose donc, en conservant les notions sous-jacentes à leurs définitions, de construire une version transitive des courbes de sélectivité ; c'est-à-dire de remplacer les moyennes par des sommes.

4.3.1 Définitions

On distingue généralement les trois courbes suivantes :

- le tonnage : $T(z) = S(z)$ qui donne la surface de l'ensemble des densités supérieures ou égales à z . Cependant, on voit bien que l'interprétation de cette fonction fait référence à la surface totale échantillonnée qui est arbitraire.

- l'abondance $Q(z)$ des densités supérieures ou égales à z .

Ces deux fonctions ont déjà été utilisées pour définir les fonctions de répartition des variables $z(x)$ et z_I au paragraphe précédent. Elle peuvent être mises en correspondance pour construire la fonction $Q(S)$.

- le bénéfice conventionnel : $B(z) = Q(z) - zT(z) = Q(z) - zS(z)$ qui est une autre manière d'utiliser les deux premières fonctions. Il correspond par définition à l'abondance des densités comptées à partir du niveau z (Fig. 4.3).

4.3.2 Interprétation et propriétés

Sens "économique" du bénéfice conventionnel

Imaginons que le coût fixe d'un coup de chalut soit connu et qu'il puisse s'exprimer en équivalent densité. Soit z ce coût fixe. Cela veut dire en d'autres termes qu'un coup de chalut n'est rentable qu'à partir du moment où "il" ramène au moins un nombre z de poissons.

Connaissant le coût d'une action de pêche, on cherchera donc à ne récupérer que les densités qui couvrent leur coût, c'est-à-dire à ne chaluter que dans les régions où la densité est au moins égale à cette densité, soit dans la surface $S(z)$. En chalutant l'ensemble de cette partie de la population, le bénéfice qu'on en retire (Fig. 4.3) est conventionnellement égal à la différence entre les gains $Q(z)$ et les coûts $zS(z)$. La différence $Q(z) - zS(z)$ représente donc le **bénéfice conventionnel**

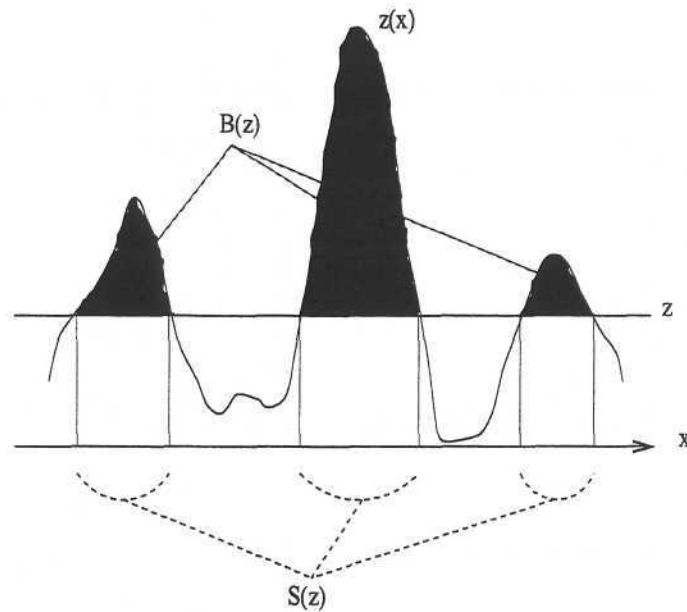


FIG. 4.3 - Signification physique du bénéfice conventionnel au seuil z .

au seuil z . Et, plus généralement, le bénéfice conventionnel est la fonction qui exprime comment évolue ce bénéfice en fonction de la valeur, en densité, du coût.

Propriétés

La fonction $B(z)$ est **décroissante** et convexe. Elle vaut Q en 0 et s'annule passée la densité maximale.

A titre d'exemple, il est intéressant de voir comment évolue le bénéfice conventionnel en fonction du support. Pour cela, j'ai généré 10 000 valeurs log-normales, de moyenne 1 et d'écart-type 2. Supposons que ces valeurs sont des mesures ponctuelles et qu'elles représentent la réalité en continu. J'ai ensuite regroupé ces données par paquets de 2, 4 et 8 valeurs pour simuler trois niveaux de support différents.

L'effet régularisateur de l'augmentation de support est très net sur les histogrammes (Fig. 4.4 a,b,c,d). Au début assez dissymétrique, l'histogramme devient presque gaussien pour le dernier support.

Il apparaît par ailleurs que les courbes de bénéfice sont **hiérarchisées** (Fig. 4.5). Plus le support est grand, plus le bénéfice conventionnel est petit. Ainsi, en modélisant le changement sur les niveaux de support connu, il est possible d'estimer la distribution des densités à un support inconnu. Mais, on retiendra surtout que le bénéfice conventionnel permet de hiérarchiser, et donc de bien différencier, des densités obtenues à différents supports.

Cet exemple démontre bien que le mélange de données de captures obtenues par des coups de chalut de tailles différentes est équivalent, du point de vue statistique, à mélanger des variables de lois différentes. Dans la pratique on prendra donc soin à ne pas mélanger des données obtenues sur des supports trop différents.

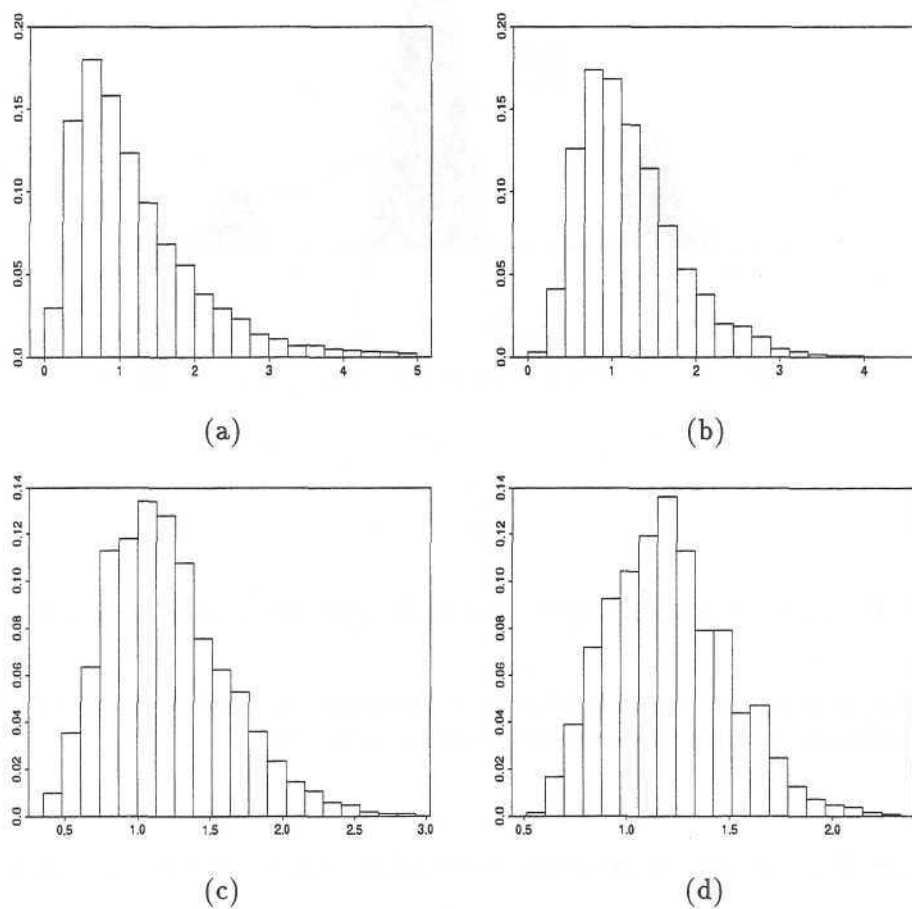


FIG. 4.4 - Régularisation. Histogrammes pour des supports égaux à (a) une fois, (b) deux fois, (c) quatre fois et (d) huit fois le support initial.

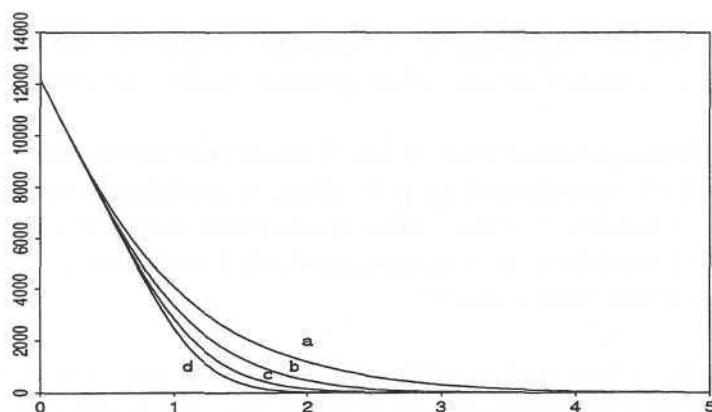


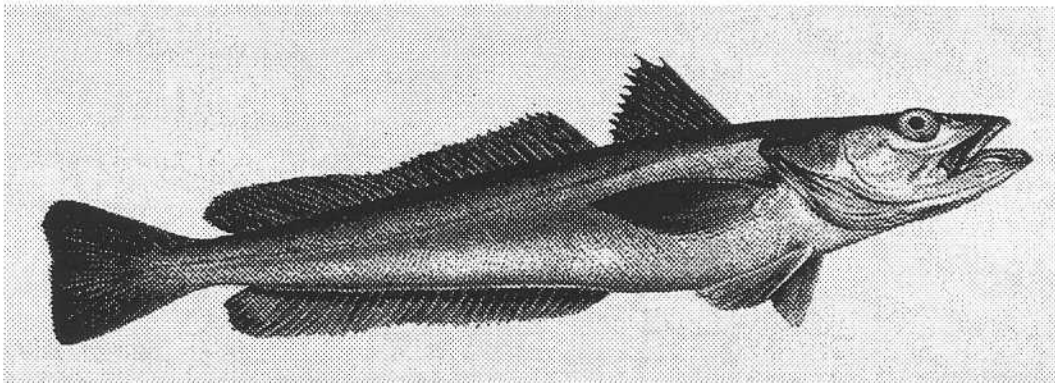
FIG. 4.5 - Régularisation. Bénéfice conventionnel pour différents supports (voir figure ci-dessus).

Troisième partie

**Analyse structurale et cartographie
du merlu du Golfe de Gascogne**

Chapitre 1

Présentation des données



Merlu adulte (Merluccius merluccius)

Sommaire

Dans ce chapitre, on présente tout d'abord les données (mode d'échantillonnage, technique de pêche,...). Les surfaces d'influences sont calculées. Enfin, quelques statistiques élémentaires ainsi que les courbes de sélectivités (bénéfice conventionnel) sont données.

Material

Abstract

In this chapter, we present the data (sampling pattern, fishing technics,...). For basic statistics computations, the surfaces of influence are calculated. By the end, the selectivity curves (conventional benefit) are drawn.

1.1 Situation générale

Le merlu (*Merluccius merluccius*) du Golfe de Gascogne est une espèce commercialement importante et biologiquement bien connue.

En 1996, la France, qui contribue à elle seule pour la moitié des captures communautaires, a bénéficié d'un quota de pêche de 27 000 tonnes de merlu (tous stocks confondus). Outre son importance commerciale, cette espèce a attiré l'attention des chercheurs et des gestionnaires à cause du problème que pose la capture simultanée des langoustines et des juvéniles de cette espèce (merluchons).

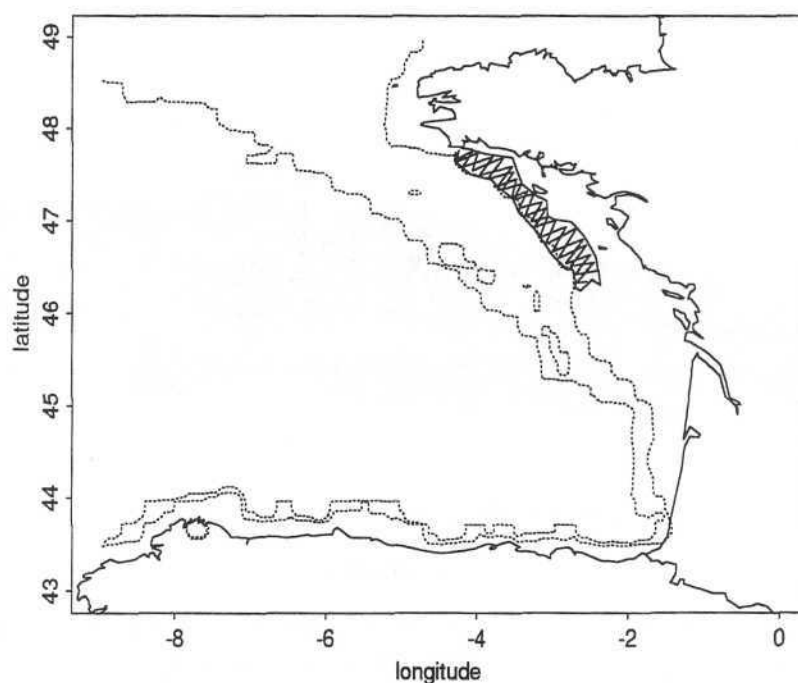


FIG. 1.1 - Golfe de Gascogne. Emplacement de la Grande Vasière. D'après Quéro et al. (1989).

En effet, la Grande Vasière, bande sablo-vaseuse située entre 80 et 140 mètres de profondeur (Fig. 1.1), est l'habitat principal des merlus pendant leurs deux premières années (Dardignac, 1988). Mais, c'est aussi là que se trouvent les langoustines. La capture de ces dernières par des chalutages de fond ne peut donc pas se faire sans captures accessoires importantes de merlus de petites tailles. Des solutions techniques ont été fournies depuis quelques années. Elles reposent sur des doubles chaluts où les jeunes merlus capturés par le chalut supérieur de maillage plus grand, peuvent s'échapper.

Espèce importante et sensible donc, le merlu est l'objet d'une surveillance scientifique importante. Depuis 1987, l'IFREMER (Institut Français de Recherche pour l'Exploitation de la Mer) réalise un programme d'échantillonnage de l'ensemble du Golfe de Gascogne afin d'augmenter les connaissances halieutiques sur l'écologie de la zone. Parmi les "espèces" capturées, le merlu d'âge 0, c'est-à-dire les individus nés à la fin de l'hiver en bordure ouest du plateau continental et arrivés sur la Grande Vasière vers juillet, a retenu notre attention. L'analyse qui suit concerne l'année

1990. Je tiens d'ailleurs à remercier chaleureusement Jean-Charles Poulard (IFREMER-Nantes) de nous avoir transmis ces données.

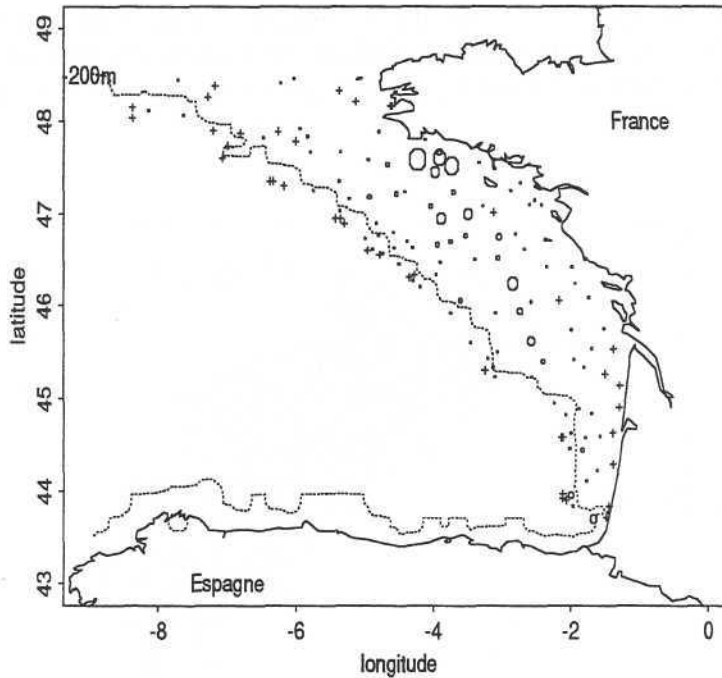


FIG. 1.2 - Représentation proportionnelle des densités de merlu d'âge 0 - Densités nulles représentées par une croix - Golfe de Gascogne 1990 - Minimum = 0, maximum = 47550 ind/nm²

Le plan d'échantillonnage est stratifié selon 8 strates de latitude et 6 niveaux de profondeur. Dans chaque strate, certaines stations sont visitées chaque année et d'autres sont mobiles. On considèrera cependant que les 137 échantillons prélevés en 1990 sont aléatoires uniformes dans la zone d'échantillonnage. L'engin de pêche est un chalut de fond à grande ouverture verticale, qui est l'engin standard pour l'ensemble des relevés scientifiques de densités de poissons de fond en Europe. Les traits de chalut sont réalisés de jour et ont une durée fixe de 30 minutes. Les densités de merlu-0 sont donc exprimées en nombre d'individus par heure de chalutage (ind/h). Pour les calcul d'abondances on utilisera le fait que la taille moyenne du support des densités (i.e la surface moyenne chalutée en une demi-heure) est de 0.02 milles nautiques (Poulard, 1990).

Une représentation proportionnelle des densités (Fig. 1.2) indique que les fortes concentrations de merlu-0 sont réparties essentiellement dans la partie centrale du plateau continental et que l'ensemble du stock de la région semble avoir été échantillonné puisque des densités nulles ont été observées en périphérie. Par ailleurs, on notera la présence de concentrations nulles "à l'intérieur" de la régionalisation des jeunes merlus (au large de l'île de d'Oléron par exemple, ou au Sud de Belle-Ile). Si elle était nécessaire, une éventuelle délimitation du champ de la régionalisation devrait donc statuer sur l'appartenance ou non de telles valeurs à la population.

1.2 Surfaces d'influences

Pour calculer les statistiques par individu sur la base d'un échantillonnage comme celui-ci, il faut d'abord, évaluer les surfaces d'influence de chaque échantillon. Pour respecter les contraintes géographiques et ne pas donner trop d'importance aux échantillons de bordure, chaque polygone d'influence est limité, soit par la côte, soit par un polygone limite égal ici à un carré de 30 milles nautiques de côté (Fig. 1.3). Connaissant les sommets de chaque polygone, il est alors possible, de les représenter, et de calculer la surface d'influence associée à chaque observation.

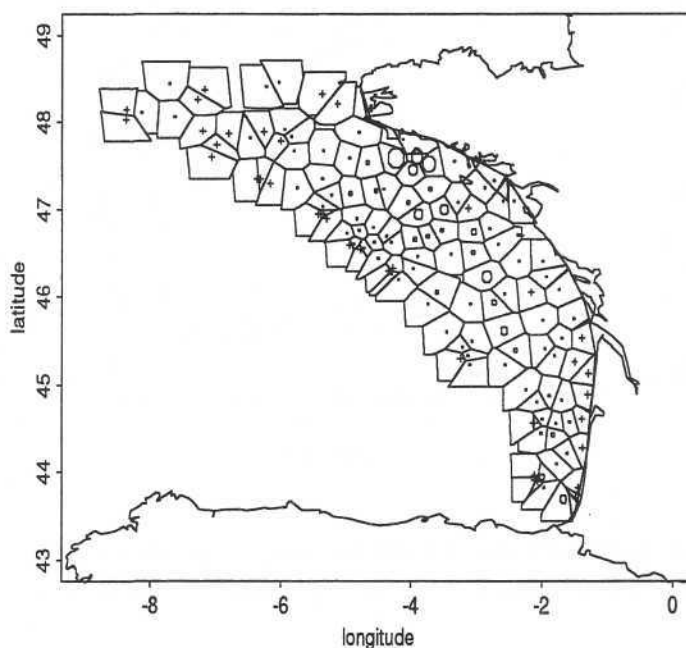


FIG. 1.3 - Représentation des surfaces d'influence.

L'histogramme des surfaces d'influence (Fig. 1.4a), indique une distribution homogène autour d'une valeur modale d'environ 200 nm^2 . La surface d'influence moyenne est 270 nm^2 .

Un contrôle rapide du caractère plus ou moins préférentiel de l'échantillonnage consiste à analyser le nuage de corrélation entre les densités observées et leurs surfaces d'influence (Fig. 1.4b).

On voit ainsi que les densités nulles sont associées à un éventail très large de surfaces d'influence. Il apparaît également que les surfaces d'influence attribuées aux concentrations les plus importantes fluctuent autour de la surface moyenne et ne sont donc pas, en moyenne, plus faibles que pour les autres niveaux de densité. La moyenne des surfaces d'influence pondérées par les densités (250 nm^2), c'est-à-dire leur centre de gravité, n'est d'ailleurs pas très différent de la surface moyenne (270 nm^2). En particulier, la taille moyenne des surfaces d'influence ne diminue pas avec la densité comme cela serait le cas si l'échantillonnage avait été préférentiel (effort d'échantillonnage d'autant plus fort que la densité est forte). Il apparaît donc raisonnable de supposer que l'échantillonnage est non préférentiel.

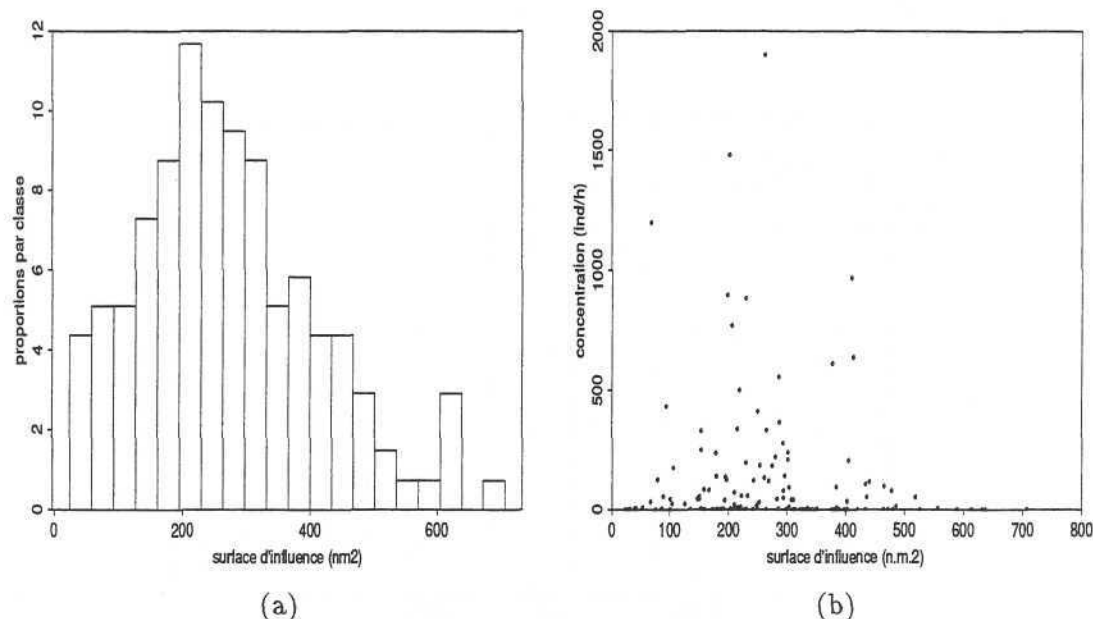


FIG. 1.4- (a) Histogramme des surfaces d'influence. (b) Corrélation entre les densités et les surfaces d'influence.

1.3 Résultats préliminaires

1.3.1 Statistiques élémentaires

Les surfaces d'influence étant exprimées en milles nautiques carrés, les densités de merlu sont transformées en nombre d'individus par mille nautique carré préalablement aux calculs. La quantité globale de merlu d'âge 0 est estimée en 1990 à $Q = 1.1 \cdot 10^8$ individus, soit un peu plus de 100 millions de juvéniles. La densité moyenne par individu au support considéré (c'est-à-dire 0.02 nm^2) est $m = 1.8 \cdot 10^4 \text{ ind/nm}^2$. Cela signifie qu'un individu tiré au hasard, se trouve dans un endroit où la densité, au support 0.02 milles nautiques carrés, est en moyenne de $1.8 \cdot 10^4 \text{ ind/nm}^2$. De plus, l'inertie des densités est estimée à $2.09 \cdot 10^8 \text{ ind}^2/\text{nm}^4$, ce qui conduit à un coefficient de variation de 0.8.

La surface équivalente que l'on peut attribuer à cette régionalisation est 6100 nm^2 , ce qui semble assez peu. Cela signifie en particulier qu'une population dont la densité par individu serait constante et égale à $18\,000 \text{ ind/nm}^2$, n'aurait besoin d'occuper qu'un carré de 80 nm de côté pour avoir la même abondance totale que le merlu-0.

1.3.2 Distributions statistiques

La faiblesse du coefficient de variation (0.8) indique que la distribution de la densité par individu est peu dispersée comme le confirme l'histogramme expérimental des abondances de merlu par classe de densité (Fig. 1.5). En comparant cet histogramme à l'histogramme usuel des densités, c'est-à-dire l'histogramme des densités par échantillon, très dissymétrique (Fig. 1.6), on constate que les fortes concentrations (supérieures à 3000 ind/nm^2), qui ne représentent guère que 0.02% des

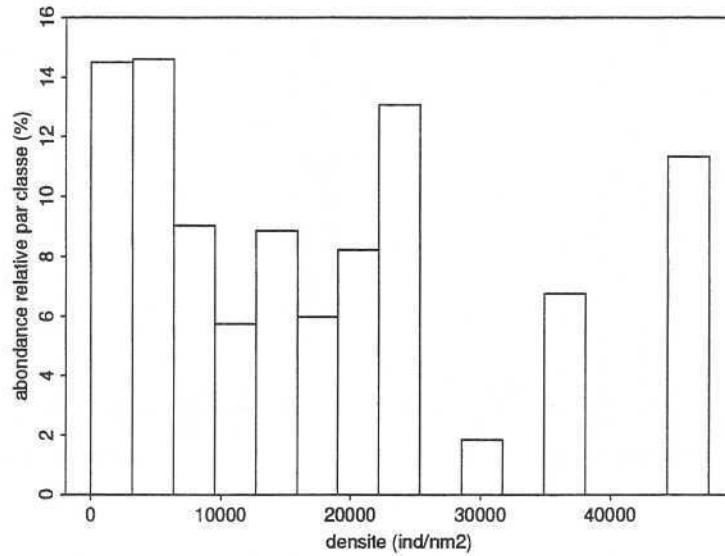


FIG. 1.5 - *Histogramme des abondances relatives par classe de densité. Version expérimentale de la loi des densités par individu.*

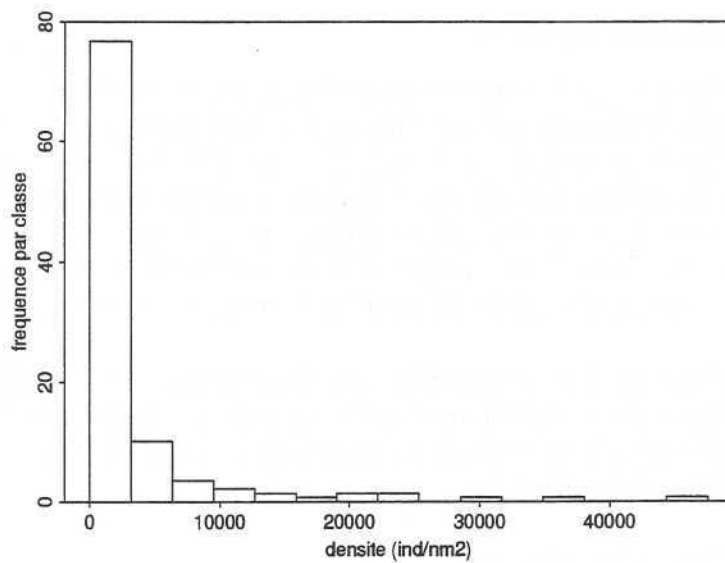


FIG. 1.6 - *Histogramme usuel des fréquences des densités par classe de densité. Version expérimentale de la loi des densités par échantillon.*

mesures, assurent 20% de l'abondance totale. La différence très sensible des histogrammes se traduit également par une différence importante des moyennes de chacune des deux variables considérées. La somme des surfaces d'influence excède légèrement 36 000 milles nautiques carrés de sorte que la moyenne des 137 valeurs d'échantillons vaut environ 3000 ind/nm². Cette moyenne est à rapprocher de la densité moyenne par individu estimée à 18 000 ind/nm² au support de 0.02 nm².

Compte-tenu de la distribution des valeurs, le découpage en classe est assez délicat. Il est donc préférable de construire la fonction de répartition, ou, ce qui est équivalent, les courbes de sélectivité.

1.3.3 Courbes de sélectivité

Le tonnage donne la surface des densités supérieures ou égales à un certain seuil. Il décroît très rapidement avec la densité (Fig. 1.7a). La moitié de la surface échantillonnée (36 000 nm²) est constituée de densités inférieures à 200 ind/nm². Les densités supérieures à 2400 ind/nm² ne représentent que le quart de cette surface, à savoir 10 000 nm².

La fonction $Q(z)$ donne, quant à elle, l'abondance due aux données supérieures ou égales à un seuil fixé. Expérimentalement, cette courbe décroît régulièrement avec ce seuil (Fig. 1.7b). Ainsi, en comparant les courbes $S(z)$ et $Q(z)$ pour une même valeur de z , on obtient la courbe $Q(S)$ (Fig. 1.7c). La forte convexité de cette dernière est reliée à la très forte dissymétrie des données. On notera les deux résultats intéressants suivants. D'une part, la moitié du stock de merlu nés en 1990 se trouve concentrée dans une surface très réduite d'environ 2000 nm². D'autre part, on obtient quasiment la totalité du stock en couvrant une surface de 20 000 nm², soit, la moitié de celle échantillonnée.

Le bénéfice conventionnel expérimental (Fig. 1.7d), chute régulièrement. Des "bénéfices" importants peuvent être retirés jusqu'à des niveaux élevés de "coûts" (exprimés en termes de densités). Cela traduit bien le fait qu'une partie importante du stock est regroupée dans un petit espace.

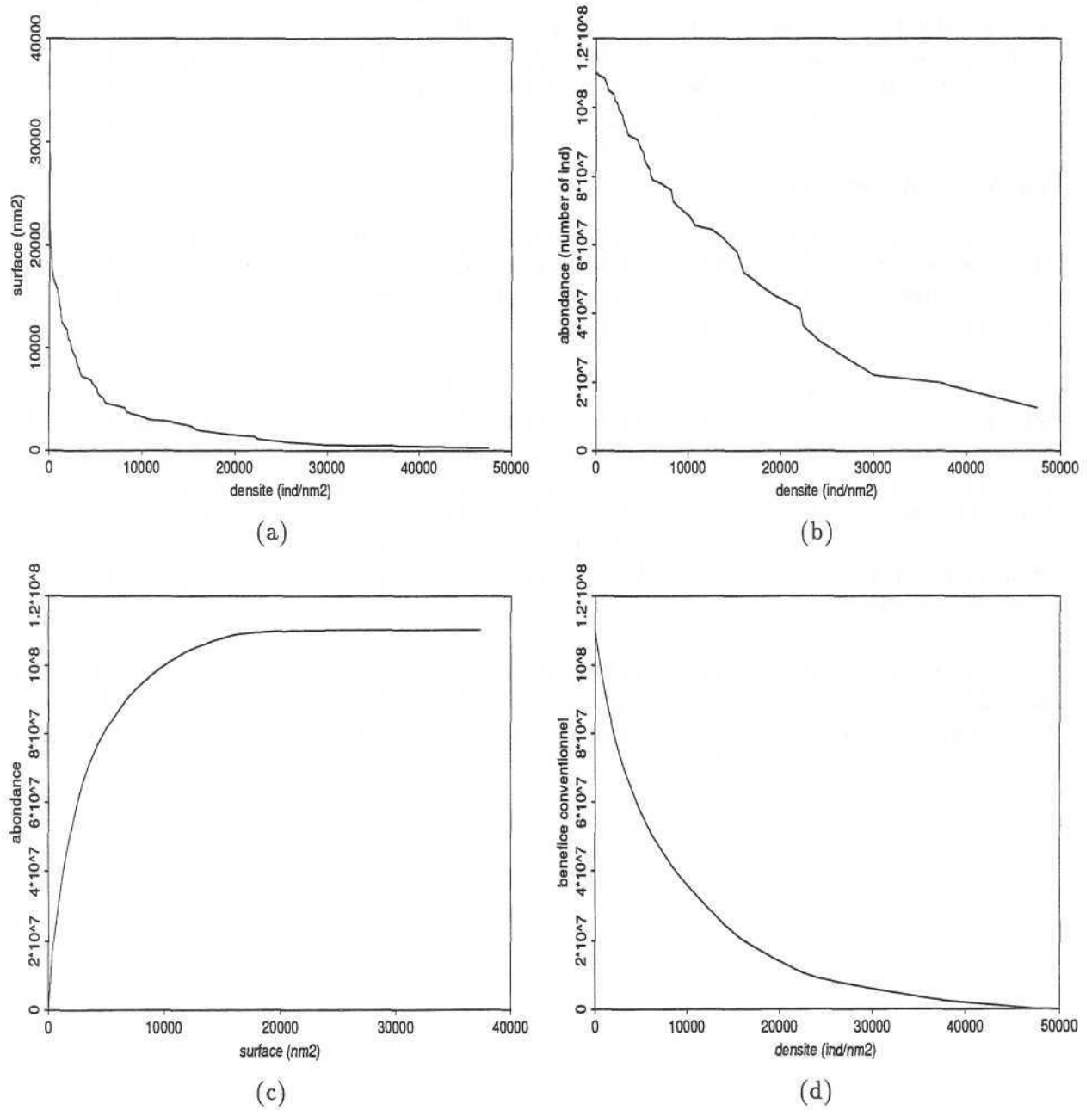


FIG. 1.7 - Courbes de sélectivité.

- (a) Tonnage $S(z)$: surface des densités $z(x) \geq z$ en fonction de z
 (b) Abondance $Q(z)$: abondance des densités $z(x) \geq z$ en fonction de z
 (c) Courbe $Q(S)$: abondance en fonction de la surface.
 (d) Bénéfice conventionnel $B(z)$: $Q(z) - zS(z)$

Chapitre 2

Analyse structurale

Sommaire

*Le **paragraphe 1** est consacré à l'estimation et à l'interprétation du covariogramme des densités de merlu d'âge 0 (Golfe de Gascogne - 1990). La cas particulier du covariogramme géométrique est traité. En outre, ce paragraphe permet de comparer les différents modes d'estimation d'un covariogramme lorsque que le plan d'échantillonnage n'est pas régulier.*

*Dans le **paragraphe 2** la structuration spatiale est décrite à l'aide des régressions expérimentales bivariées. On s'attardera sur l'analyse des régressions pour la première classe de densité.*

*Enfin, le **paragraphe 3** aborde la description structurale sous l'angle des fréquences conditionnelles.*

Structural analysis

Abstract

*The **paragraph 1** deals with the estimation and the interpretation of the covariogram of the hake density (Bay of Biscay - 1990). As a particular case, geometrical covariogram is computed and analysed. Considerations on surface of the population above thresholds are also given. Beside the biological interpretation, this paragraph gives the opportunity to compare the various way of estimating covariograms when the sampling scheme is not regular.*

*In the **paragraph 2** the spatial structuration of hake is described through the experimental bivariable regressions. We focus the analysis on the regression for first level of densities where most of the dissymetries occur.*

*Finally, the **paragraph 3** tackles the structural analysis from the point of view of the conditionnal frequencies.*

Cet ensemble de données sur les densités de merlu nés dans l'année a déjà fait l'objet de nombreuses analyses. Jusque là, les différentes études ont fait appel à une approche de type intrinsèque, c'est-à-dire reposant sur le calcul expérimental de variogrammes. Pierre Petitgas (1991) dans sa thèse a procédé à une comparaison des structures variographiques pour quatre années différentes. Dans une étude "confidentielle" concernant ces mêmes données, Tatiana Koudriavsteva (1993) a cherché à comparer la répartition spatiale de plusieurs espèces présentes dans le Golfe de Gascogne. Enfin, un essai d'analyse variographique a été présenté dans la première partie de cette thèse (Cf. § I.1.3.2). La plupart du temps, les résultats expérimentaux sont difficiles à interpréter.

Dans ce chapitre, je propose une analyse structurale des densités de juvéniles de merlu par une approche transitive. Il s'agira tout d'abord de comparer les différentes façons d'estimer un covariogramme lorsque l'échantillonnage est irrégulier. Ensuite, on cherchera à obtenir une description structurale plus fine que celle proposée par le covariogramme grâce aux outils présentés dans la deuxième partie de la thèse à savoir : les régressions bivariées et les fréquences conditionnelles.

Compte-tenu de l'allure générale de la régionalisation, l'analyse structurale repose sur les quatre directions suivantes, exprimées en degrés trigonométriques à partir de la direction Est-Ouest : 27° , 72° , 116° et 162° .

2.1 Estimations du covariogramme

Je commence par présenter les résultats obtenus en utilisant les différents modes d'estimation d'un covariogramme tels que présentés précédemment (Cf. § II.3.4). L'interprétation structurale sera faite ensuite.

2.1.1 Estimation du covariogramme transitif par migration vers une grille fine

Les données ont été migrées vers une grille de maille 5 mn x 5 mn. La valeur en chacun des nœuds de cette grille correspond à la valeur de l'échantillon le plus proche si ce dernier se trouve à une distance inférieure à 30 milles nautiques (identique à celle retenue pour le calcul des surfaces d'influences). Lorsqu'il n'y a pas d'échantillon suffisamment proche (c'est-à-dire à moins de 30 milles nautiques) ou que le nœud se trouve à terre, on implante une densité nulle. Le covariogramme, relativement au carré de l'abondance totale ($g(h)/Q^2$), est ensuite calculé en appliquant simplement l'algorithme de calcul à maille régulière (Fig. 2.1). Les directions de calcul correspondent, en nombre de maille de grille, aux vecteurs suivants : (2,1), (1,3), (-1,2), et (-3,1). Les pas de calculs sont donc respectivement : 11.2, 15.8, 11.2, et 15.8 milles nautiques.

2.1.2 Estimation du covariogramme géométrique par migration vers une grille fine

De la même façon, on obtient le covariogramme géométrique de l'aire occupée par les merlus en migrant l'indicatrice de la densité vers la grille de 5 mn x 5 mn, et non plus simplement la densité comme précédemment. Cette fonction structurale (Fig. 2.2) permet de décrire la géométrie du stock de merlu d'âge 0 dans le Golfe de Gascogne en 1990.

La valeur en 0 ($30\,000\text{ mn}^2$) est égale à la surface des densités strictement non nulles qui peut être comparée à la somme des surfaces d'influence ($36\,000\text{ mn}^2$). L'extension géographique du stock est différente selon les directions : les portées s'échelonnent entre 130 mn dans la direction 26° et 400 mn dans la direction 116° , la direction d'allongement maximum de la population. La

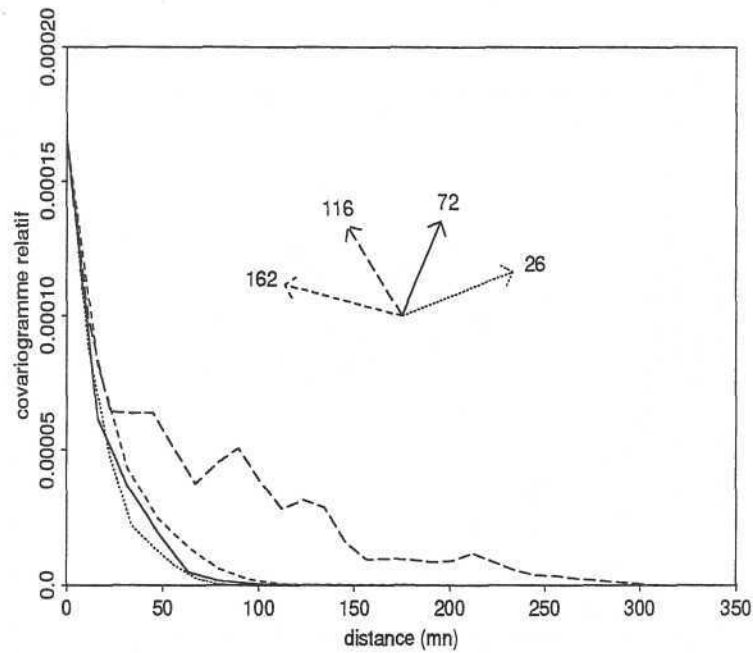


FIG. 2.1 - Covariogramme transitif expérimental exprimé relativement au carré de l'abondance : $g(h)/Q^2$. Migration des densités vers une grille de 5 mn \times 5 mn

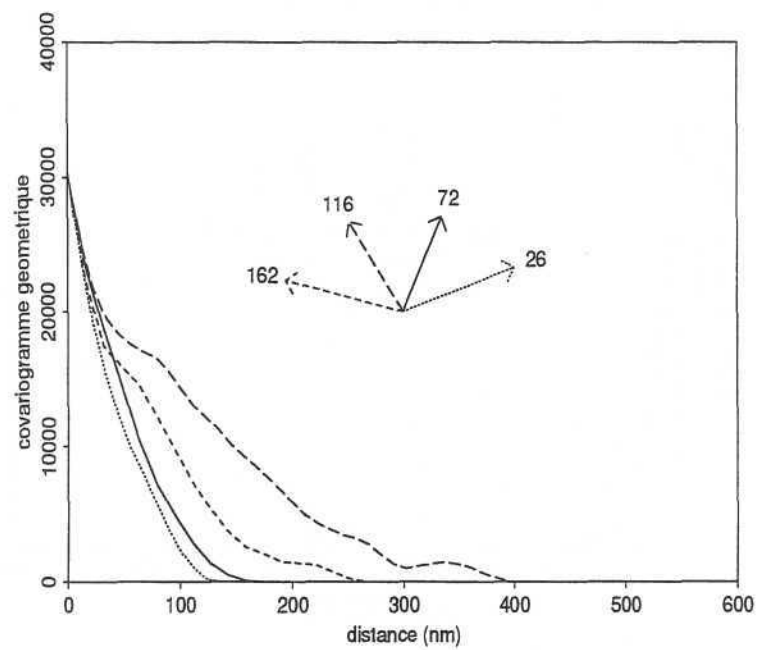
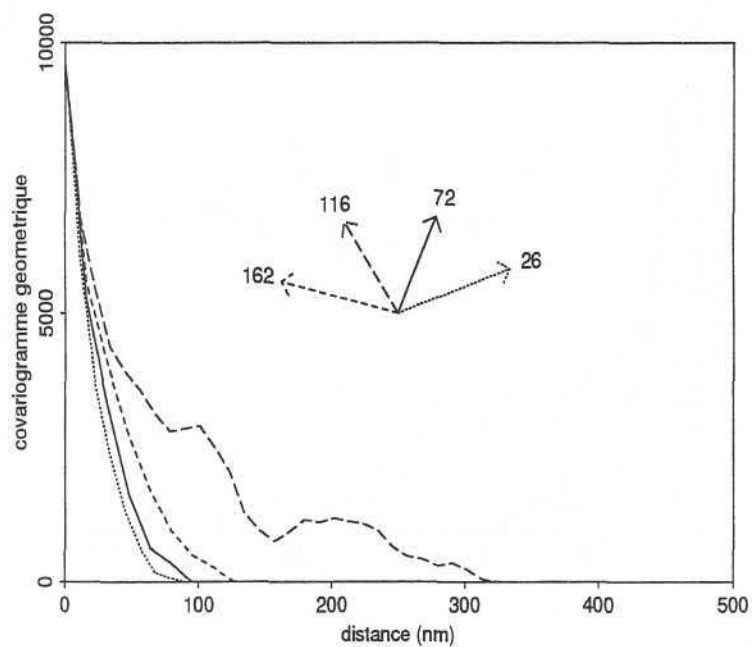
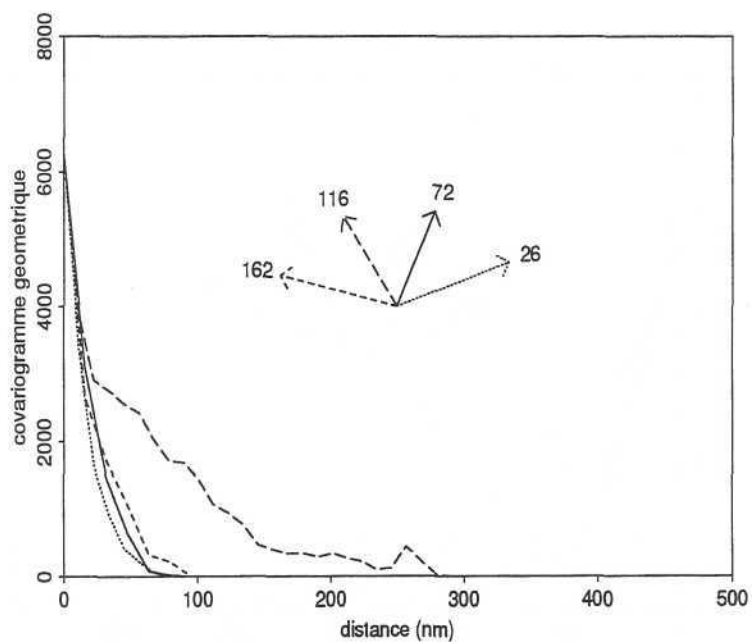


FIG. 2.2 - Covariogramme géométrique expérimental exprimé en mn^2 . Migration de l'indicatrice des densités vers une grille de 5 mn \times 5 mn



(a)



(b)

FIG. 2.3 - Covariogramme géométrique expérimental exprimé en mn^2 pour les ensembles des densités supérieures à : (a) 2500 ind/ mn^2 , (b) 5000 ind/ mn^2 . Migration de l'indicatrice des densités seuillées vers une grille de $5 mn \times 5 mn$

forme globalement elliptique du champ se traduit par une forme de covariogramme grossièrement sphérique. Le comportement du covariogramme géométrique est presque isotrope au voisinage de l'origine pour des distances inférieures à 30 mn environ. L'isotropie s'étend au-delà dans les deux directions 26 et 72°.

Il est, de plus, intéressant de voir comment évolue la géométrie de la partie de la population dépassant un certain seuil. Dans cette optique, j'ai calculé les covariogrammes géométriques pour les seuils 2500 et 5000 ind/mn².

Pour les deux seuils, l'anisotropie du champ est très marquée (Fig. 2.3). Suite au seuillage, seule la partie de la population qui se distribue le long de l'isobathe 120m est conservée. La forme du champ résultant est très allongée (très anisotrope) grossièrement en forme de "poire" ce qui donne cette forme aux covariogrammes géométriques.

2.1.3 Estimation du covariogramme par la covariance

Le covariogramme est estimé ici par le produit du covariogramme géométrique $K^*(h)$ tel qu'il vient d'être estimé, et de la covariance non centrée: $g^*(h) = K^*(h) \times c^*(h)$ (Cf. § II.3.4.2). La covariance non centrée a été calculée pour des classes de distance de 15.8, 22.4, 15.8, 22.4 milles nautiques dans les directions respectives de l'analyse structurale avec une tolérance angulaire de 22.5 degrés (Fig. 2.4).

2.1.4 Estimation du covariogramme par les régressions

Lorsque le covariogramme est décomposé en une somme pondérée de régressions, il devient nécessaire de choisir des classes de distance telles que l'absence de densité ne soit pas fortuite. Autrement dit, il convient de prendre des classes de distance suffisamment grandes pour que, à partir de chaque échantillon z_i , il y ait au moins un échantillon z_j qui appartienne à la classe, à moins qu'on ne tombe en dehors du champ. Cette limitation inhérente à ce mode d'estimation est d'ailleurs d'autant moins contraignante que le plan d'échantillonnage est homogène.

Dans le cas du merlu, on a pu voir qu'en limitant les surfaces d'influence à un carré de 30 milles nautiques de côté, on est parvenu à couvrir l'ensemble du Golfe de Gascogne sans lacune (Fig 1.3). J'ai donc choisi 30 milles nautiques comme pas pour le calcul des régressions dont les résultats sont présentés figure 2.5.

2.1.5 Synthèse et interprétation structurale

Comparaison des résultats

Les différentes estimations du covariogramme transitif montrent globalement les mêmes tendances : décroissance très rapide et isotrope du covariogramme relatif de l'origine au premier pas de calcul et schéma anisotrope au-delà. Cette anisotropie permet de distinguer nettement la direction Nord-Nord-Ouest (116°) des trois autres directions de calcul. La portée dans les 3 directions isotropes avoisine 100 milles nautiques alors que la portée dans la direction d'anisotropie est environ de 300 milles nautiques.

Quelques différences apparaissent cependant entre les différents résultats quant à la structure dans la direction de plus grande continuité (116°). Le covariogramme obtenu par migration et calculé avec un pas assez fin, présente des oscillations atténuées par les autres modes de calculs.

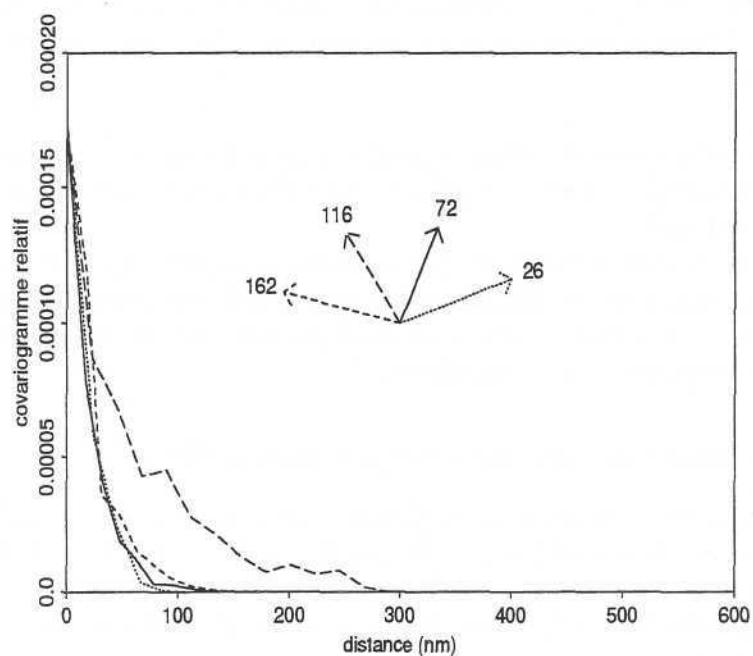


FIG. 2.4 - Covariogramme transitif expérimental exprimé relativement au carré de l'abondance : $g(h)/Q^2$. Estimation par le produit du covariogramme géométrique et de la covariance non centrée.

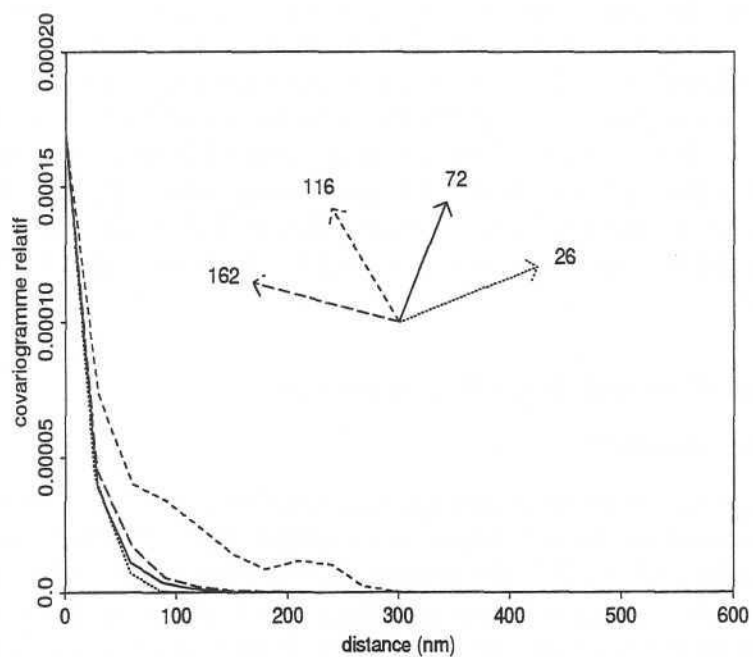


FIG. 2.5 - Covariogramme transitif expérimental exprimé relativement au carré de l'abondance : $g(h)/Q^2$. Estimation par une somme pondérée de régressions.

Interprétation

La valeur en 0 est peu parlante en soi. A titre de comparaison, l'agrégation du merlu-0 telle que mesurée par le covariogramme relatif en 0 ($1.5 \cdot 10^{-4} \text{ mn}^{-2}$) est identique à celle du dernier stade des œufs de maquereaux ($1.2 \cdot 10^{-4} \text{ mn}^{-2}$; Cf. §IV.4.1) On prendra soin dans ce genre de comparaison de s'assurer que les unités sont les mêmes puisque $g(0)/Q^2$ n'est pas un nombre sans dimension.

Par ailleurs, ce qui apparaît graphiquement comme un comportement linéaire et isotrope autour de l'origine, doit être interprété avec précautions. On se souvient en particulier que la migration des densités a pour effet de linéariser les structures expérimentales. Ainsi, la partie du covariogramme comprise entre l'origine et le premier pas de calcul est, soit interpolée (migration), soit inconnue (covariance). En particulier, en changeant le mode de représentation de sorte que la valeur en 0 et la valeur au premier pas de calcul ne sont pas reliées (Fig. 2.6), l'effet de pépité apparaît avec plus de force. De plus, une observation attentive de la représentation proportionnelle (Fig. 1.2), indique qu'il est possible de passer d'une densité nulle à une densité extrêmement forte en 15 milles nautiques environ. L'ensemble de ces considérations nous ont amenés à interpréter le comportement du covariogramme à l'origine comme un effet de pépité (discontinuité) ou une microstructure plutôt que comme une décroissance linéaire.

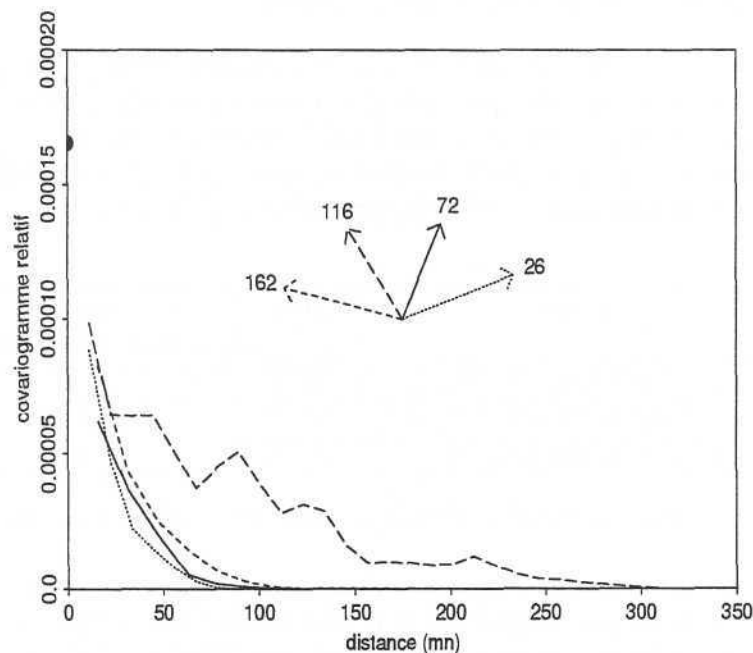


FIG. 2.6 - Covariogramme transitif expérimental exprimé relativement au carré de l'abondance : $g(h)/Q^2$. Mise en évidence de l'effet de pépité.

Ainsi, d'après le comportement du covariogramme relatif autour de l'origine, on montre que, à petite échelle, l'agrégation du merlu est :

- importante
- et d'extension géographique très limitée (pépité); environ 15 milles nautiques.

Les effets de trou que l'on observe dans la direction 116° reposent sur les trois paquets de valeurs

fortes que l'on rencontre respectivement vers 47.5, 47 et 46 degrés de latitude. La réalité de ces trois regroupements n'est pas contestable et correspond par ailleurs plus ou moins aux bancs de vase que l'on trouve dans la région. Faut-il donc en tenir compte lors de la modélisation?

On a vu que les deux derniers modes d'estimations apportaient moins de détails sur ces effets de trou. Par ailleurs, les représentations proportionnelles des densités de merlu d'âge 0 de 1987 à 1990 (Petitgas, 1991), indiquent que cette situation n'est pas systématique. Si les fortes densités ont tendance à toujours être observées dans les mêmes secteurs, leur regroupement est différent d'année en année. De sorte qu'il ne semble pas que ces regroupements soient des éléments structuraux forts et consistants pour les jeunes juvéniles. Enfin, il faut bien admettre que la modélisation d'un covariogramme à effet de trou anisotrope est délicat. Pour toutes ces raisons, on choisira de ne pas tenir compte des effets de trou lors de la modélisation du covariogramme.

Les deux paragraphes suivants permettent de compléter l'analyse structurale. Le covariogramme vient d'être estimé par une moyenne pondérée de régressions. On propose donc d'analyser séparément chacune de ces régressions afin de mieux comprendre le comportement du covariogramme.

2.2 Régressions bivariées par classe

Les régressions ont été calculées avec un pas de 30 milles nautiques et pour trois classes de densités. Pour mémoire, je rappelle que chacune de ces régressions participe au covariogramme avec un poids proportionnel à l'abondance des densités de cette classe soit : $\sum_{z_i \in C_i} z_i S_i$. A l'aide de l'histogramme des abondances par classe de densités (Fig. 1.5), j'ai choisi trois classes de densités de poids à peu près identiques mais d'effectifs très différents (Tab. 2.1).

classe	poids % de Q	effectif nombre d'échantillons
$z(x) \in [0, 10\ 000 \text{ ind/mn}^2[$	40 %	124
$z(x) \in [10\ 000, 20\ 000 \text{ ind/mn}^2[$	23 %	7
$z(x) > 20\ 000 \text{ ind/mn}^2$	37 %	6

TAB. 2.1 - Effectif et abondance pour trois classes de densités

Les résultats sont exprimés par classe de densités pour les 4 directions de calculs (Fig 2.7). Les distances sont comptées positivement dans le sens indiqué par la légende. Pour faciliter l'interprétation, la direction 162° est comptée positivement vers l'Est-Sud-Est, soit -18°. Ainsi, dans les directions -18°, 26° et 72°, une distance positive est synonyme de déplacement vers la côte, alors que les distances négatives indiquent un déplacement vers le large.

Résultats par classe

- Partant d'une valeur de densité appartenant à la première classe (Fig. 2.7a), c'est-à-dire partant de la majorité des échantillons puisque cette classe regroupe 90% des mesures, on voit que la densité augmente en moyenne pour des petites distances (30 milles nautiques environ). Cette augmentation est plus importante pour un déplacement vers la côte (h positif) que pour un déplacement vers le large (h négatif). Elle est par ailleurs isotrope vers la côte et plutôt anisotrope vers le large. Au-delà de 30 milles nautiques la densité moyenne chute brutalement de sorte qu'on

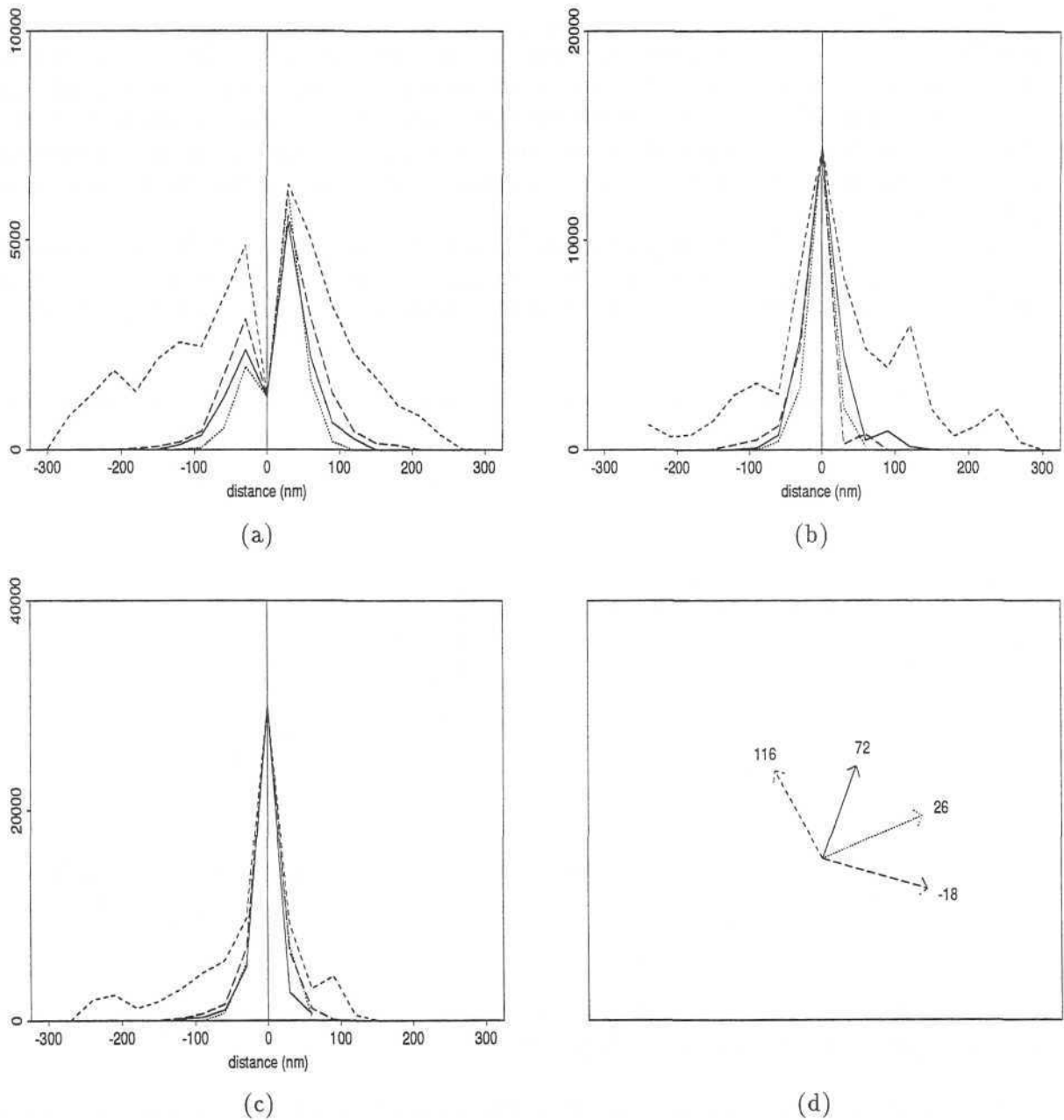


FIG. 2.7 - Forme de la régression $m[z(x+h)|z(x) \in C_i]$ pour trois classes de densités (ind/mn^2).
 (a) : $z(x) \in [0, 10\ 000 \text{ ind}/\text{mn}^2[$
 (b) : $z(x) \in [10\ 000, 20\ 000 \text{ ind}/\text{mn}^2[$
 (c) : $z(x) > 20\ 000 \text{ ind}/\text{mn}^2$
 (d) : directions de calcul en degrés trigonométriques

Rq Faut-il ouvrir ou fermer à gauche la première classe? \Rightarrow Faut-il inclure les 0 dans C_1 $[0, 10\ 000[$ $]0, 10\ 000]$

retrouve, en moyenne, le même niveau de densité en $x + h$ qu'en x , pour des distances de l'ordre de 80 milles nautiques dans les 3 directions isotropes et 160 milles nautiques pour la direction 116°.

Pour faciliter l'interprétation de ces résultats, prenons un exemple simple à une dimension. Imaginons que la distribution des densités soit telle que sur la figure 2.8a. Dans ce cas, la moyenne des densités à une distance h d'une densité inférieure à 5.5 augmente plus rapidement pour un déplacement vers la droite que vers la gauche, c'est-à-dire pour des distances positives. Ainsi, bien que ce soit en allant vers la gauche que l'on rencontre l'augmentation la plus brutale de densités, l'augmentation régulière des densités dans la première moitié du transect génère une croissance plus forte de la régression.

Par conséquent, l'allure des régressions de la figure 2.8a indique que les densités de merluchons augmentent assez régulièrement depuis le bord du plateau continental jusque vers le centre du Golfe de Gascogne où elles sont maximum et diminuent brutalement ensuite en allant vers la côte.

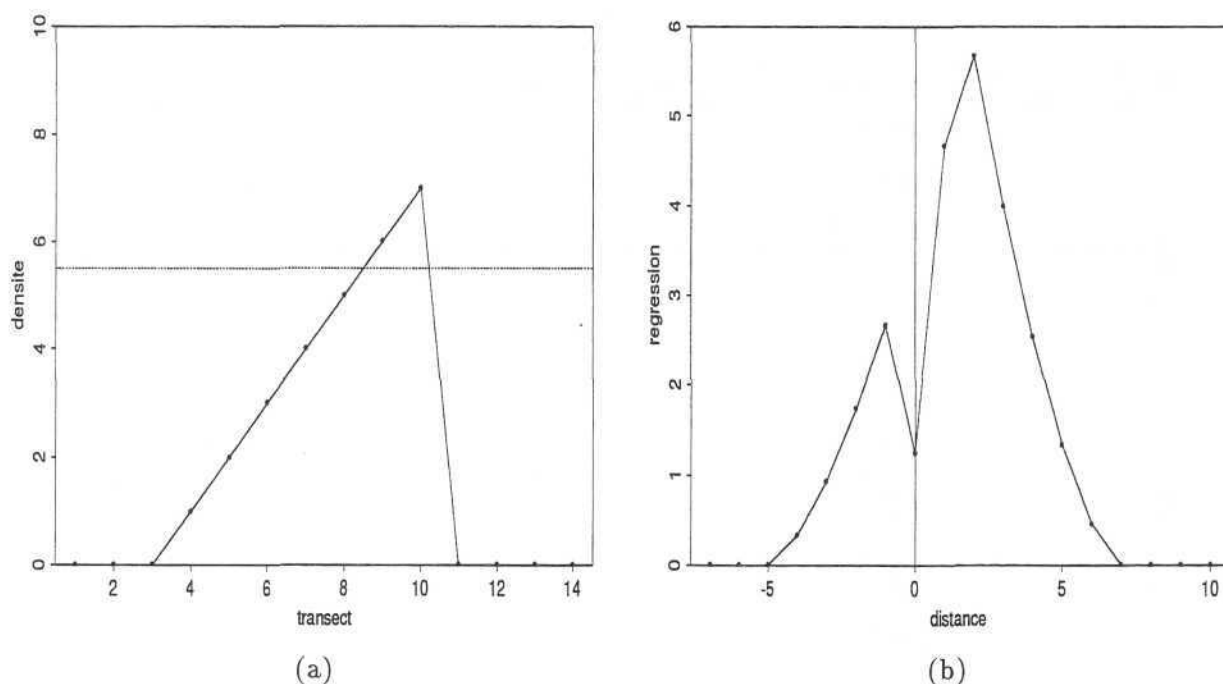


FIG. 2.8 - Exemple de régression dans un cas simple à une dimension. Exemple de la classe des densités inférieures à 5.5 : $m[z(x+h)|z(x) < 5.5]$

- Les régressions pour les deux autres classes donnent l'évolution moyenne des densités en partant d'une densité forte, voire très forte. On ne doit donc pas être surpris qu'elles diminuent très rapidement (Fig 2.8b et c). Ce qui distingue ces deux classes est essentiellement la valeur de la régression à l'origine. En effet, une fois ramenée à la même échelle on peut constater que la moyenne des densités dans la direction de continuité évolue approximativement de la même façon.

Retour au covariogramme

Bien que chaque régression ait un poids identique dans la composition du covariogramme, il apparaît clairement que la forme du covariogramme, et en particulier son comportement à l'origine,

est imposée par les fortes densités.

2.3 Fréquences conditionnelles

Les fréquences conditionnelles de dépassements de seuils sont encore une autre façon de décrire la structuration des densités. J'ai choisi ici quatre couples de seuils basés sur deux niveaux de densités (5000 et 1250 ind/mn²) et conservé les mêmes directions de calcul. Les ensembles de densités respectivement supérieures et inférieures à 1250 ind/mn² sont représentés figure 2.9

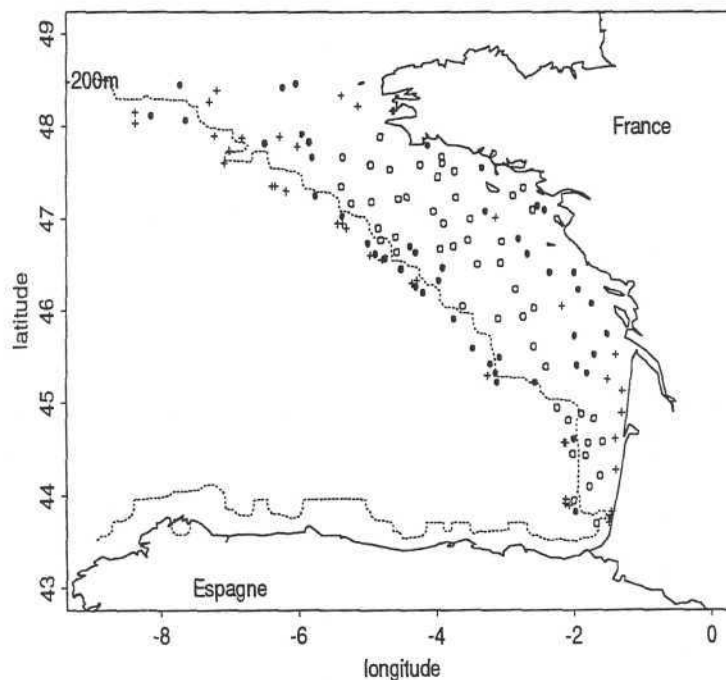


FIG. 2.9 - Ensemble définis par deux coupures.

croix : $z(x) = 0$

cercle plein : $z(x) > 0$ et $z(x) \leq 1250$

cercle vide : $z(x) > 1250$

On constate une diminution très rapide des fréquences aux fortes coupures (Fig. 2.10a). C'est-à-dire que, sachant qu'on se situe en un endroit où la densité est au moins égale à 5000 ind/mn², la probabilité que la densité soit également supérieure à 5000 ind/mn² à 50 milles nautiques de là, chute considérablement. Pour des raisons d'anisotropie, cette chute est d'ailleurs moins rapide dans la direction Nord-Nord-Ouest.

On peut voir également (Fig. 2.10c) que, sachant que la concentration en un point donné dépasse 1250 ind/mn², il y a une chance sur deux pour que la densité soit encore supérieure à ce même seuil à 200 milles nautiques de là dans la direction Nord-Nord-Ouest.

Enfin, le dernier cas de figure s'intéresse à la probabilité de trouver une densité non nulle partant d'une densité au moins égale à 1250 ind/mn². Si le nombre de valeurs intermédiaires (cercles noirs,

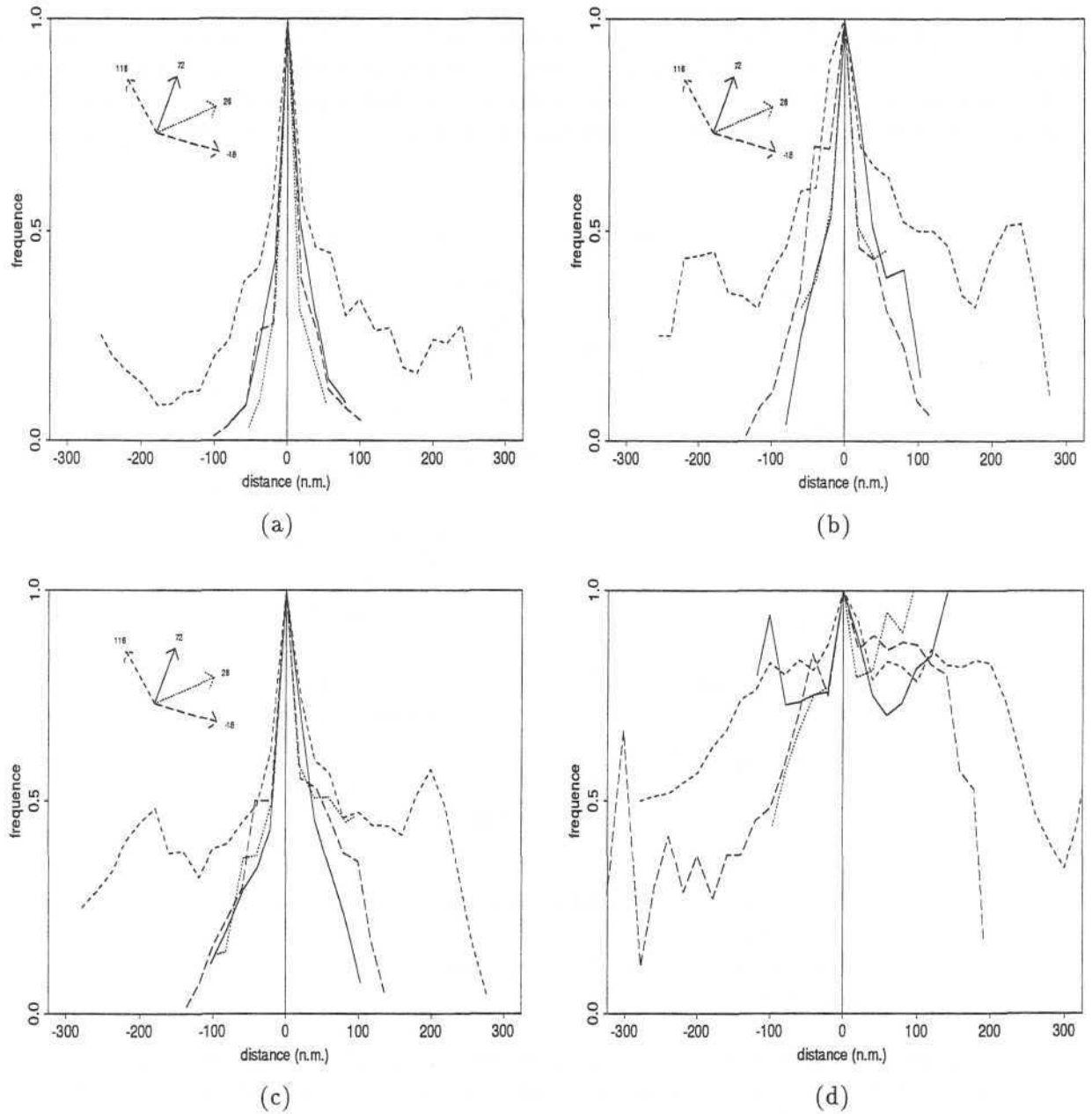


FIG. 2.10 - Fréquences conditionnelles pour différents couples de coupures de densités (ind/mn^2).

(a) : $f(z(x+h) > 5000 | z(x) > 5000)$

(b) : $f(z(x+h) > 1250 | z(x) > 5000)$

(c) : $f(z(x+h) > 1250 | z(x) > 1250)$

(d) : $f(z(x+h) > 0 | z(x) > 1250)$

Fig. 2.9) est à peu près le même de part et d'autre du centre de la régionalisation, on voit cependant que ces valeurs sont regroupées sur une bande assez fine en bordure Ouest alors qu'elles s'étendent jusqu'à la côte de l'autre côté. Cela se traduit sur le graphe des fréquences conditionnelles par une chute plus rapide pour les distances négatives (Fig. 2.10d): partant d'une densité supérieure à 1250 ind/mn^2 , il y a une chance sur deux pour que la densité soit nulle à 100 milles nautiques "à l'Ouest", alors qu'il y a une à deux chance(s) sur dix, pour qu'elle soit nulle à 100 milles nautiques "à l'Est".

Un certain nombre de traits structuraux de la répartition des merluchons en 1990 sont donc ressortis de cette analyse structurale. Suite à cette phase descriptive, on se propose de choisir un modèle de covariogramme et de l'utiliser pour une cartographie des densités.

Chapitre 3

Cartographie

Sommaire

Dans ce chapitre, très court, on commence par choisir un modèle de covariogramme. Ce dernier est ensuite utilisé pour faire une cartographie des densités de merluchons par krigeage transitif. Quelques résultats concernant les poids de krigeage sont présentés.

Mapping

Abstract

In this very short chapter, first of all, a model function is fitted to the experimental covariogram. This model is then used to make a map of the hake density by transitive kriging. Some characteristics of the kriging weights are also given.

L'analyse structurale est une étape importante dans une étude géostatistique. Elle permet en premier lieu de décrire certains traits structuraux importants d'une régionalisation.

Mais l'analyse structurale est aussi le préalable nécessaire au choix d'un modèle structural. L'étape de modélisation peut d'ailleurs se révéler très délicate. En particulier, en transitif, lorsqu'il s'agit de choisir une fonction structurale qui ajuste au mieux un covariogramme expérimental présentant des effets de trou ou bien encore des anisotropies complexes.

Une fois le modèle structural choisi ou inféré, un certain nombre d'opérations statistiques sont possibles. J'ai déjà expliqué comment calculer une variance d'estimation globale (§III.3.3). Partant de là, il est possible de tester ce que donneraient différents plans d'échantillonnage, et en comparant les variances des estimations ainsi obtenues, on peut espérer améliorer l'efficacité de certaines campagnes scientifiques. Ce genre d'étude a déjà été réalisée (Petitgas, 1991), et j'ai donc préféré illustrer une autre manière, moins connue, d'utiliser un modèle de covariogramme. Il s'agit de faire une cartographie des densités de merlu d'âge 0 par krigeage transitif, c'est-à-dire par un mode d'interpolation qui tient compte de la structure spatiale du phénomène.

3.1 Choix d'un modèle de covariogramme

Un modèle a été ajusté au covariogramme expérimental obtenu par migration des densités vers une grille fine. La structure retenue est composée de trois schémas élémentaires (Fig. 3.1) :

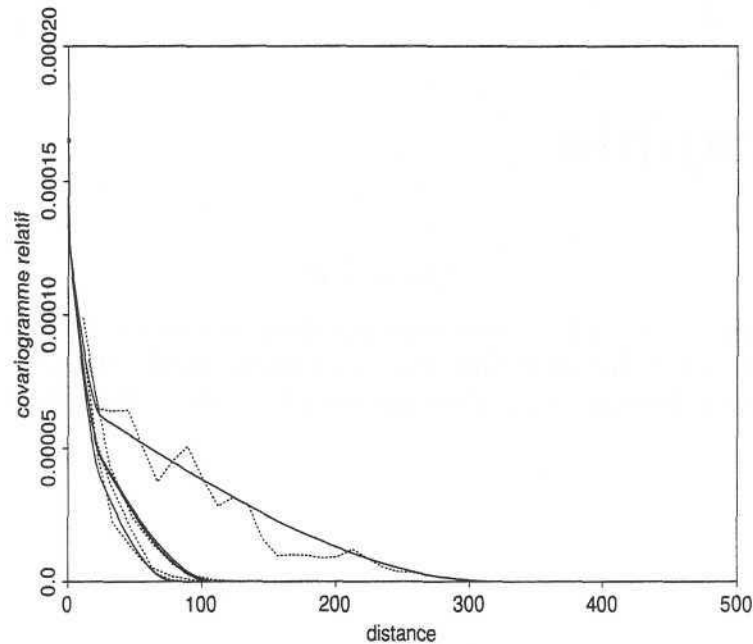


FIG. 3.1 - Modèle de covariogramme (trait plein) et covariogramme expérimental (pointillé) obtenu par migration des densités.

- un effet de pépite de $3.5 \cdot 10^{-5}$
- un schéma sphérique isotrope de palier $6 \cdot 10^{-5}$ et de portée 25 milles nautiques
- et un deuxième schéma sphérique, mais anisotrope cette fois-ci, de palier $7 \cdot 10^{-5}$ et de portée 80 milles nautiques dans la direction 26° et 320 milles nautiques dans la direction orthogonale 116° (coefficient d'anisotropie égal à 4).

3.2 Carte : méthode, résultat et discussion

Le krigeage a été réalisé en voisinage unique sans contrainte sur la somme des poids. La zone krigée est comprise entre la côte à l'Est et le bord du plateau continental à l'Ouest.

Le résultat du krigeage est présenté sous la forme de 3 courbes isodensités pour les niveaux : 1500, 5000 et 10 000 ind/nm² (Fig. 3.2).

En guise de discussion, je me propose de donner quelques indications sur les poids de krigeage. Pour un krigeage dual sans contrainte comme nous avons fait, les poids varient entre $-3.2 \cdot 10^8$ et $3.52 \cdot 10^8$ et leur somme vaut $2.95 \cdot 10^8$. Un grand nombre (la moitié) de ces poids sont cependant voisins de 0 comme l'indique leur histogramme (Fig. 3.3). L'existence de poids négatifs explique pourquoi l'extrapolation, c'est-à-dire le krigeage en dehors de la zone échantillonnée peut conduire à des estimations de densités négatives. C'est une raison supplémentaire pour limiter les calculs à un champ raisonnable.

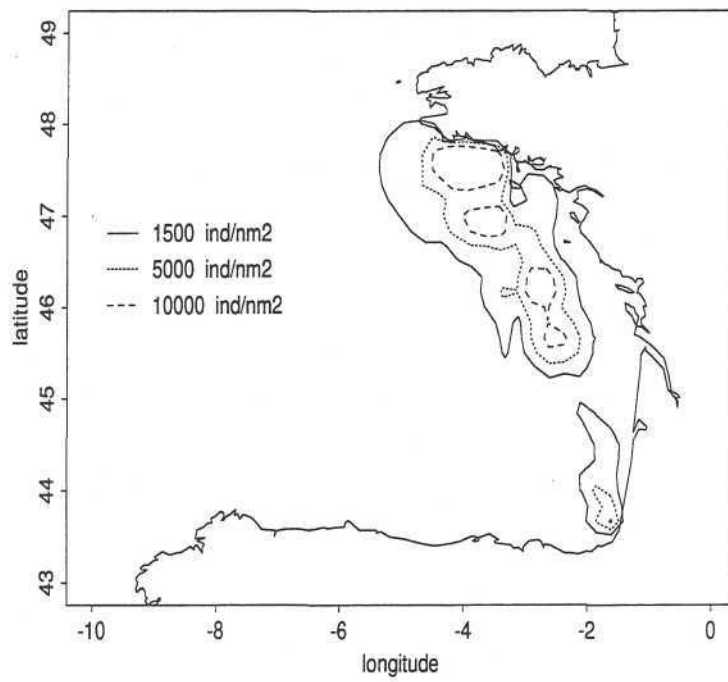


FIG. 3.2 - *Krigeage transitif sans contrainte.*

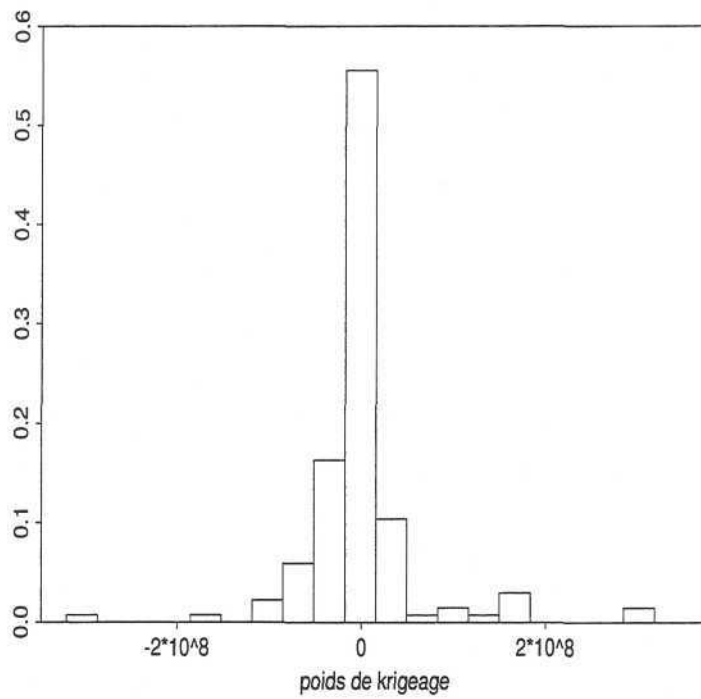


FIG. 3.3 - *Histogramme des poids du krigeage dual*

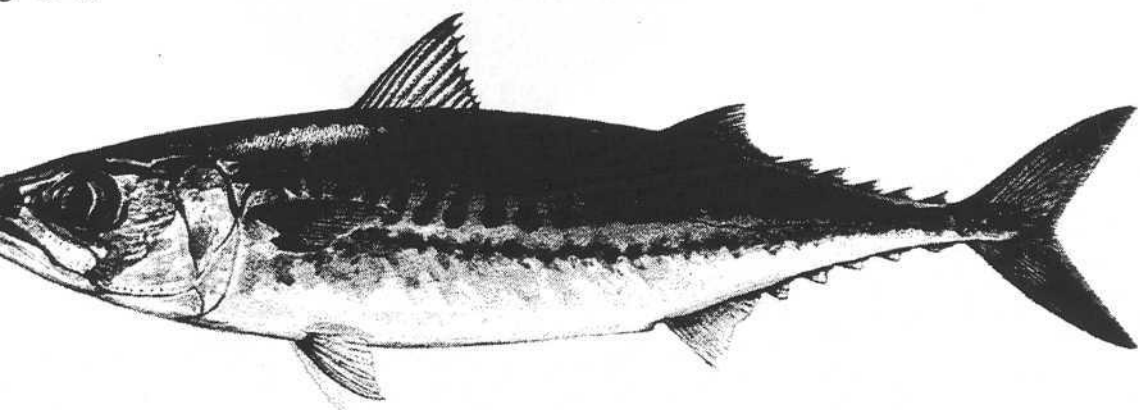
Quatrième partie

Etude de la ponte et des phases larvaires du maquereau

tre 1

ements de biologie halieutique du uereau

e aspects of mackerel biology



Maquereau adulte (Scomber scombrus)

1.1 Importance commerciale du maquereau

Dans les années 60 et 70, le maquereau était très prisé par les chalutiers industriels français autant que la morue, le lieu noir ou le merlan. Les captures françaises étaient supérieures à 400 000 tonnes par an, ce qui faisait du maquereau la deuxième ou troisième espèce française en tonnage débarqués.

Depuis le début des années 80, la moyenne des débarquements français a fortement chuté, plus que le septième rang en tonnage. En 1987 par exemple, où le total de la production avoisinait 15 000 tonnes, le maquereau n'occupait plus que le septième rang en tonnage.

Les raisons de cette baisse de production sont cependant plus à mettre sur le compte de la chute de la vente du maquereau sous criée que sur le compte d'une baisse du niveau des stocks. Le prix du maquereau se situe entre 4 F/kg et 5,50 F/kg depuis le début des années 80, alors que l'état des stocks était très élevé.

que préoccupant dans certaines régions comme la mer du Nord, est satisfaisant. En somme, il y a "le poids sans le prix". En 1993, le maquereau n'appartient pas au groupe des vingt premières espèces en valeurs. En 1995 cependant, avec 39,6 millions de francs, le maquereau était la vingtième espèce française pour un total de 11 785 tonnes débarquées. A titre de comparaison, en 1995, la première espèce en valeur était la sole (351,2 million de francs) et en tonnage, le merlan (23 872 tonnes).

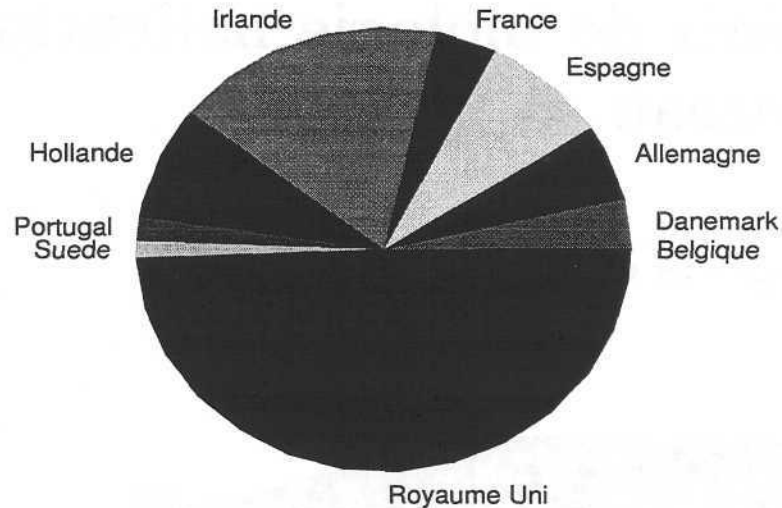


FIG. 1.1 - Répartition du TAC 1996 de maquereau au sein de la CEE.

A l'exclusion de quelques unités, il n'y a pas en France de pêcherie spécialisée dans la capture du maquereau. Il y a par contre en Ecosse et en Norvège un intérêt très fort pour cette espèce. D'après la FAO (Food and Agriculture Organisation), les captures totales de maquereaux dans l'Atlantique Nord-Est étaient, en 1991, de 610 800 tonnes. Plus de la moitié de ces prises étaient le fait des Norvégiens (180 000 tonnes) et des Ecosais (170 000 tonnes).

Cela se retrouve dans la répartition des quotas de pêche au sein de l'Europe. En 1996, le TAC (Total Admissible des Captures) pour l'Atlantique Nord-Est a été fixé à 343 855 tonnes. La part de la France (11 970 tonnes) représente 3,5% du TAC alors que le quota de pêche du Royaume Uni représente 42,5% du TAC.

1.2 Cycle biologique du maquereau

1.2.1 Description sommaire du cycle de vie

Le maquereau (*Scomber scombrus*) dit de l'Atlantique, commun, ou encore européen, est une espèce migratrice et pélagique qui vit en large bancs. On distingue généralement deux stocks : un stock Nord composé de deux populations évoluant entre la mer du Nord et la mer Celtique, et un stock Sud ("the Southern stock") se répartissant le long du plateau continental espagnol et portugais.

Les deux populations du stock Nord se distinguent par leurs cycles de vie. L'une ("the Northern stock"), surexploitée, passe l'hiver le long des côtes norvégiennes et se disperse au printemps vers

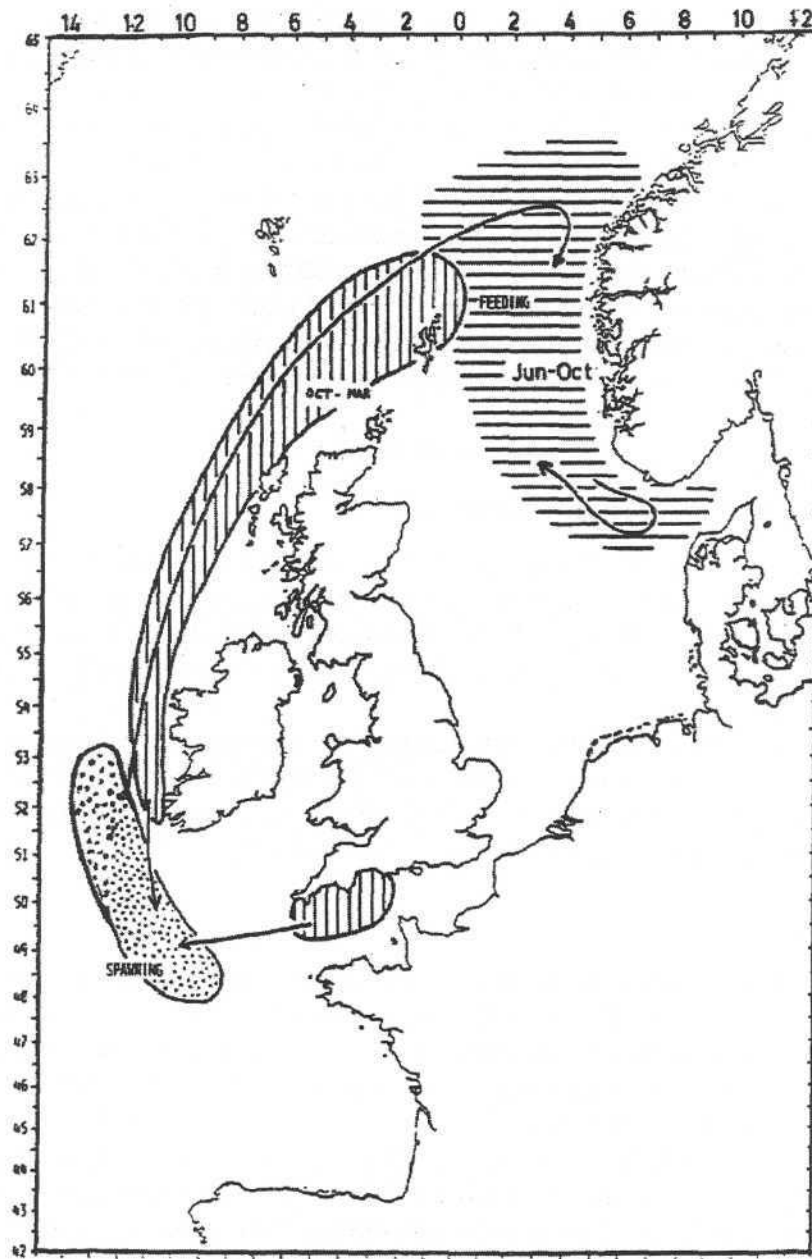


FIG. 1.2 - Schématisation du cycle biologique de la partie Ouest du stock de maquereau dans les années 1980. D'après Anonyme (1990).

la mer du Nord.

L'autre ("the Western stock") passe l'hiver à l'entrée de la mer du Nord et migre le long des accores du plateau continental européen pour rejoindre l'aire de reproduction en mer Celtique (Fig. 1.2). Le déclenchement de la migration n'est pas connu mais il semblerait que la population migre brutalement en bloc. Actuellement, la migration se déclenche courant janvier. Pendant la migration de reproduction les maquereaux ne s'alimentent pas (les contenus stomacaux sont systématiquement vides). Dans le même temps le développement des gonades est amorcé. Une fois sur les lieux de ponte (mer Celtique) vers le mois de Mars, les maquereaux recommencent à s'alimenter. Les femelles pondent leurs œufs en paquets ("batch") avec quelques jours d'intervalle entre chaque. Une femelle produit donc plusieurs paquets d'œufs. Après la reproduction, les maquereaux migrent à nouveau vers la mer du Nord. Le passage sur les lieux de ponte semble par ailleurs suivre un certain ordre. En effet, il a été montré (Lockwood *et al.*, 1981) que la taille moyenne des maquereaux capturés en mer Celtique diminue pendant la saison de ponte. Les individus les plus grands sont les premiers à arriver sur l'aire de ponte, et donc également les premiers à en repartir, alors que les jeunes maquereaux arrivent par la suite. Ce comportement semble très persistant sur les dernières décennies.

1.2.2 Remarques sur le projet SEFOS

Changements de distribution importants.

Il a été mis en évidence (Walsh et Martin, 1986; Anon., 1990) que la partie hivernale du cycle biologique du maquereau avait changé de manière importante depuis les années 80. L'aire de grossissement s'est déplacée vers le Nord et les schéma de migrations ont changé. Toutefois, dans le même temps, le comportement reproductif du maquereau est demeuré remarquablement stable (même zone, même période).

Puisqu'une part importante du cycle biologique du maquereau se passe aux alentours des accores, une partie de la communauté scientifique a souhaité chercher, dans la variabilité de l'hydrologie associée aux accores, la cause de ces modifications. C'est dans cet objectif que le projet SEFOS (Shelf Edge Fishery and Oceanography Study) a été élaboré.

Un seul stock

Une expérience de marquage-recapture a été réalisée dans le cadre du projet SEFOS. En 1994, des maquereaux adultes (10 000) ont été marqués au Sud du Golfe de Gascogne, à la frontière entre la France et l'Espagne, entre le 15 mars et le 20 avril. Au début 1996, au total 70 animaux marqués ont été recapturés. Pour la plupart, ils ont été capturés autour des Shetlands, au nord de la mer du Nord ou sur les côtes norvégiennes.

Les résultats montrent donc très clairement que le stock du sud se mélange (au moins géographiquement) aux stocks nords. Il a d'ailleurs été décidé de ne plus séparer ces stocks à l'avenir, et de gérer l'ensemble de ces trois populations comme un seul stock. On continuera tout de même à distinguer une partie Nord, une partie Ouest et une partie Sud, dont les cycles de vie sont très différents.

Mon travail, ainsi que la suite de cette partie, porte sur la partie Ouest du stock. C'est à cette partie-là du stock que je ferai référence, par défaut, dans la suite du document.

Chapitre 2

Présentation des données et résultats préliminaires

Sommaire

Le but de ce chapitre est de présenter les données issues des campagnes triennales d'estimation de la production d'œufs de maquereau organisées par le CIEM.

*Le **paragraphe 1** permet de présenter la zone et le plan d'échantillonnage ainsi que les modes de prélèvement (chalutages verticaux ou obliques). La liste des variables biologiques et hydrographiques, mesurées in situ, est également décrite.*

*Le **paragraphe 2** passe en revue les différentes transformations qu'ont subies les données de base avant leur analyse: suppression des échantillons aberrants, et passage à une densité d'œufs par unité de surface et par jour. La cohérence des mesures hydrographiques est analysée à l'aide de répliqués. Enfin, les échantillons sont migrés vers une grille régulière.*

*Dans le **paragraphe 3**, je me sers des densités moyennes par mois depuis 1977 pour décrire le déroulement général d'une saison de ponte à l'aide de cartes proportionnelles et de covariogrammes expérimentaux.*

*Enfin, le **paragraphe 4** est l'occasion de tester l'effet de support sur des données réelles, après avoir précisé sa définition dans le contexte des campagnes triennales. L'effet régularisant d'une augmentation du support est illustrée par la diminution de la proportion des valeurs faibles ou nulles sur les histogrammes et par l'évolution du bénéfice conventionnel. Des conclusions sur l'exclusion des données espagnoles pour la suite des analyses sont présentées.*

Material and preliminary results

Abstract

The aim of this chapter is to present the kind of data that come from the triennial mackerel (and horse mackerel) egg production surveys organized by ICES (International Council for the exploration of the Sea).

In paragraph 1, I present the sampling area, the sampling pattern and the fishing technics (vertical or oblique holes). Then the various biological and hydrological variables measured during each survey are described.

The paragraph 2 explains the different transformations that are needed for further analyses like the removal of outliers values and the computation of densities of eggs per unit surface and per day (temperature dependency). The consistency of the hydrological measurements is analysed with regards to the available duplicates. The samples which should be on a regular grid according to the ICES recommendations, are only quasi-regular. That is why they are migrated to the regular scheduled grid.

In paragraph 3 I will use the monthly mean densities since 1977 to describe the general pattern of a spawning season. This will be done simply with proportional representations and with experimental covariograms.

Finally, the paragraph 4 is devoted to analysis of the support effect on real data. I first give the definition of the support in the context of the triennial surveys. I then illustrate the regularization effect of an increase of the support mainly by the decrease of the proportion of low values in the histograms and by the evolution of the conventional benefit. Some conclusions about the removal of the spanish data from the data set are finally stated.

Le maquereau adulte est un poisson assez délicat à capturer. Il n'est par conséquent pas sûr que les captures, commerciales et scientifiques, soient suffisamment représentatives du stock en place pour qu'elles puissent être utilisées pour les estimations annuelles de stock. Pour suivre l'état du stock de maquereau, les œufs étant plus faciles à capturer que les adultes, les scientifiques ont préféré opter pour une démarche basée sur la production d'œufs. Dans cette optique, le CIEM organise depuis 1977, des campagnes internationales pour évaluer l'intensité de la ponte ("Triennial Mackerel Egg Production Surveys").

La démarche procède comme suit : connaissant la quantité d'œufs produit une année donnée, la distribution en taille des femelles, le nombre moyen d'œufs produits par femelle en fonction de leur taille et enfin, le ratio mâle/femelle, il est possible de proposer une estimation du stock de maquereaux adultes qui a généré le niveau de ponte observé. Compte-tenu de la quantité de travail sur le terrain et la quantité d'analyses en laboratoire que nécessite cette démarche, il a été décidé de ne conduire l'estimation qu'une fois **tous les trois ans**. On voit également, que ce procédé nécessite beaucoup de connaissances biologiques et physiologiques sur le maquereau.

La communauté scientifique internationale dispose donc aujourd'hui d'une série de 7 campagnes d'observation de la ponte du maquereau. Je dois à cet égard remercier chaleureusement l'ensemble des instituts de recherche qui ont accepté de mettre à ma disposition 5 années d'observations. Cela comprend les membres du projet SEFOS mais également des instituts extérieurs à ce projet dont le laboratoire de Lowestoft que je tiens à remercier tout particulièrement.

Dans ce chapitre, je commence par présenter le plan d'échantillonnage des campagnes triennales et les modes de prélèvement. Ensuite, j'explique un certain nombre de traitements préliminaires comme la transformation des nombres d'œufs pêchés en densités d'œufs par mètre carré et par jour. Enfin, les deux derniers paragraphes permettront de présenter une vue d'ensemble d'une saison de ponte et de s'intéresser au problème de l'effet de support.

2.1 Echantillonnage

2.1.1 Plan d'échantillonnage

L'échantillonnage est défini pour estimer la production d'œufs au cours d'une saison de ponte. Pour cela, tous les 3 ans depuis 1977, chaque saison est divisée en plusieurs **périodes** d'échantillonnage: 4, 5, ou 6 selon les années (Tab. 2.1). Les deux premières années (1977 et 1980) ne figurent pas dans cette analyse. L'objectif final de ces campagnes est de disposer d'une courbe expérimentale de production d'œufs en fonction du temps et d'obtenir la production totale en intégrant cette courbe.

Les dates de début et fin de chaque période sont exprimées en **jour julien** (numéro du jour de l'année de 1 à 365) de sorte que chaque période est positionnée dans le temps par sa date moyenne (Tab. 2.1). Cela permet de voir que les "périodes" d'échantillonnage ne correspondent pas toujours à la même "période" de l'année. Pour les comparaisons inter-annuelles, il conviendra donc de prendre quelques précautions. En particulier, la période fin Mai / début Juin, correspond selon les années aux périodes d'échantillonnage 3, 2, 3, 2, 3.

Au sein de chaque période, un ou plusieurs navires de un ou plusieurs pays sont chargés de couvrir autant que faire se peut la zone de frai du maquereau, c'est-à-dire de la côte à 100 ou 200 milles nautiques et de 44 à 58 degrés Nord (Fig. 2.1). Compte tenu de l'importance de la surface à échantillonner et de la disponibilité des navires océanographiques, il a été décidé par le groupe de travail du CIEM, d'échantillonner selon une **grille régulière de 0.5 degré de longitude par 0.5 degré de latitude**. Puisque l'intensité de la ponte est maximale autour du mois de Mai, il a également été décidé de privilégier l'échantillonnage à ce moment là. C'est la raison pour laquelle le nombre de bateaux ainsi que le nombre de prélèvements augmentent en milieu de saison (périodes 2 et 3).

Chaque bateau couvre sa zone en faisant des transects longitudinaux parallèles, en allant soit vers le Nord, soit vers le Sud (Fig. 2.1). Les prélèvements se font jour et nuit.

Cependant, la réalité de l'échantillonnage est différente. En outre, plusieurs raisons font que les données ne sont pas régulièrement réparties selon la trame définie au préalable.

En premier lieu, les raisons habituelles inhérentes aux avaries ou aux difficultés météorologiques, ainsi que les problèmes techniques de toutes natures, obligent parfois à ne pas respecter le programme fixé au départ.

Ensuite, il se peut qu'au sein même d'une période d'échantillonnage, deux bateaux échantillonnent, à deux moments différents, une même région. On se rend compte de cela en représentant l'ensemble des données disponibles dans des axes latitude x date (Fig. 2.2). On peut voir par exemple qu'en 1992 (Fig. 2.2d), au cours de la période 2, la zone comprise entre 50 et 51 degrés de latitude Nord, a été échantillonnée par l'Irlande (5) puis par l'Angleterre (4) à 10 jours d'intervalle. La conséquence de ces chevauchements est que plusieurs observations sont parfois disponibles dans un même carré de la grille de prélèvement.

Enfin, il arrive également que des bateaux sur-échantillonnent les zones *a priori* riches en œufs pour augmenter la précision des estimations de production, ou encore, comme en 1989 période 2 (Fig. 2.3), qu'ils couvrent la zone d'échantillonnage en deux passages, l'un Nord-Sud et l'autre Sud-Nord, avec une distance inter-transect double de la "normale".

Pour l'ensemble de ces raisons le **plan d'échantillonnage final n'est que quasi-régulier**. On retiendra de plus, qu'il est la **compilation d'observations réalisées par différents bateaux, à des moments différents**, et qu'il vise, en priorité, l'estimation de la production d'œufs.

	année et période	date début (julien)	date fin (julien)	date moyenne	nombre de données	latitude	pays
1							
2							
3							
4							
5							
6							
7	1983 - 1	77 78	111 112	96 = 6 Avril	110	43.55 - 54.22	4,1
8	1983 - 2	120 121	138 139	128 = 8 Mai	163	43.50 - 53.77	2,7
9	1983 - 3	147 142	168 169	157 = 6 Juin	253	44.25 - 54.25	4,7,2
10	1983 - 4	165 163	175 176	170 = 19 Juin	124	45.23 - 50.77	4,2,6
11	1983 - 5	175	195 196	183 = 2 Juillet	104	46.42 - 53.70	2,6
12	1986 - 1	90 91	117 118	103 = 13 Avril	91	45.25 - 52.12	1
13	1986 - 2	131 132	155 156	146 = 26 Mai	421	43.75 - 59.75	4,7,5,2
14	1986 - 3	165 166	182 183	172 = 21 Juin	273	43.42 - 58.72	2,7
15	1986 - 4	189 190	198 199	192 = 11 Juillet	51	45.25 - 52.25	7
16	1989 - 1	91 92	110 105	99 = 9 Avril	75	45.88 - 54.93	1
17	1989 - 2	113 114	134 135	124 = 4 Mai	140	44.75 - 58.70	2
18	1989 - 3	141 142	157 154	151 = 31 Mai	329	44.73 - 59.75	4,5,6
19	1989 - 4	158 159	175 174	166 = 15 Juin	291	44.75 - 55.75	6,7,2
20	1989 - 5	185 186	200 201	191 = 10 Juillet	88	45.25 - 52.75	7
21	1992 - 1	103 104	125	116 = 26 Avril	187	44.75 - 57.73	1,2
22	1992 - 2	136 137	164 165	152 = 1 Juin	758	44.02 - 57.75	3,4,5,6
23	1992 - 3	167 168	186 187	175 = 24 Juin	188	44.25 - 57.75	7,2
24	1992 - 4	188 189	192 193	190 = 9 Juillet	40	50.25 - 56.25	2
25	1995 - 1	85 83	104 105	96 = 6 Avril	129	44.25 - 51.75	3,1
26	1995 - 2	112 113	137 133	125 = 5 Mai	189	45.25 - 57.75	4,2
27	1995 - 3	137 136	159 160	151 = 31 Mai	232	44.12 - 56.20	4,7,5,3
28	1995 - 4	158 159	183 184	169 = 18 Juin	370	44.12 - 57.77	8,3,5,8,2,7
29	1995 - 5	178 179	197 198	186 = 5 Juillet	146	45.25 - 57.77	7,8,2

TAB. 2.1 - Caractéristiques des périodes d'échantillonnage des années étudiées.

Code pays : 1 = Allemagne, 2 = Ecosse, 3 = Espagne, 4 = Angleterre, 5 = Irlande, 6 = France, 7 = Pays-Bas, 8 = Norvège.

3 à voir dans
add. période. modif

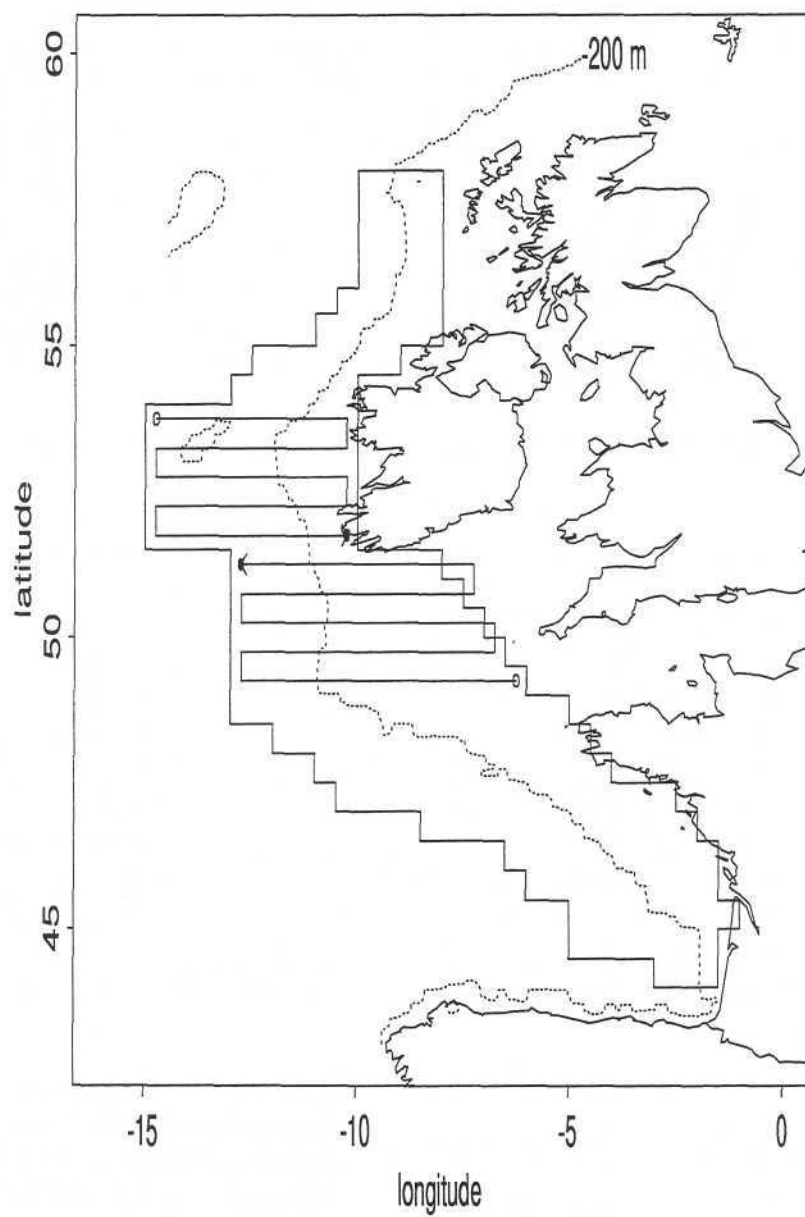


FIG. 2.1 - Zone d'échantillonnage correspondant à la zone de frai du maquereau. Représentation de deux parcours possibles pour l'échantillonnage.

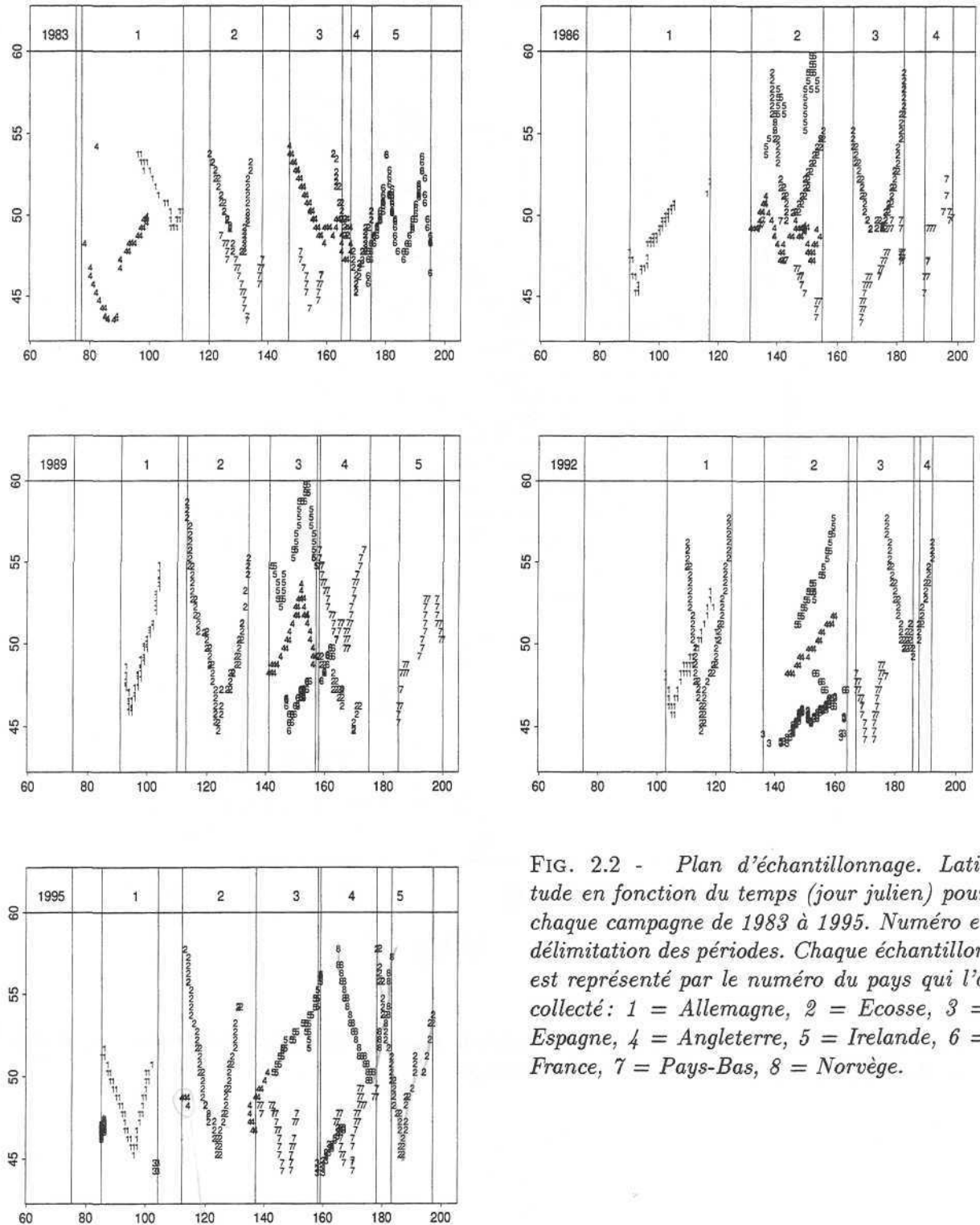


FIG. 2.2 - Plan d'échantillonnage. Latitude en fonction du temps (jour julien) pour chaque campagne de 1983 à 1995. Numéro et délimitation des périodes. Chaque échantillon est représenté par le numéro du pays qui l'a collecté: 1 = Allemagne, 2 = Ecosse, 3 = Espagne, 4 = Angleterre, 5 = Irlande, 6 = France, 7 = Pays-Bas, 8 = Norvège.

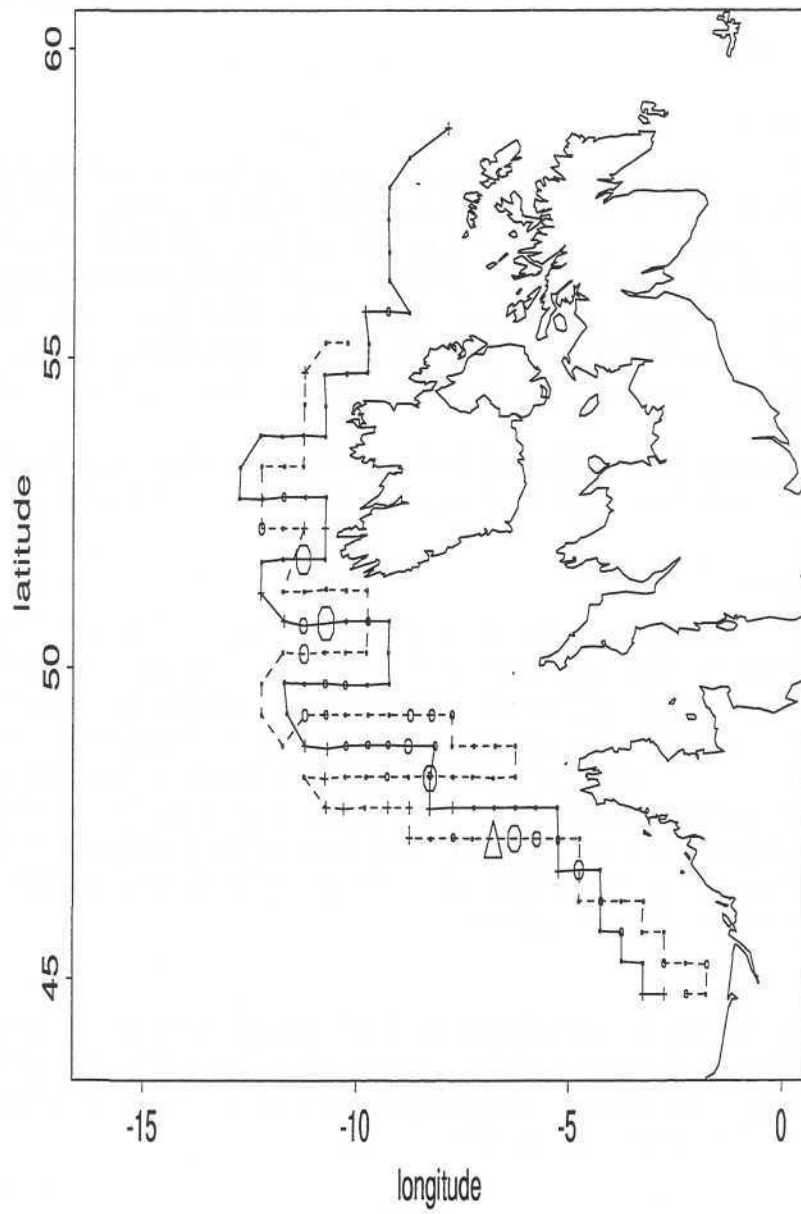


FIG. 2.3- Plan d'échantillonnage correspondant à la période 2 de 1989. Représentation proportionnelle des densités d'œufs de stade 1. Les densités supérieures à 630 œufs/m²/jour sont indiquées par un triangle ; les densités nulles par une croix.

2.1.2 Engins et modes de pêche

Les œufs sont capturés à l'aide de chaluts planctoniques. Il existe de nombreux types de tels chaluts, mais pour l'essentiel, les campagnes triennales du CIEM sont réalisées avec des "Calvet Net", "Bongo Net" ou "GulfIII". Leurs diamètres sont respectivement 24 cm pour le Calvet Net, de 20 à 60 cm pour le Bongo Net et 20 cm (parfois 10 cm) pour le GulfIII. Ils se distinguent par leurs modes d'utilisations.

Chalutages verticaux

Le "Calvet Net" est utilisé en chalutage vertical (Fig. 2.4). Le bateau s'immobilise à la verticale du point à échantillonner; le chalut est descendu jusqu'à un maximum de 200 m de fond puis remonté (phase pendant laquelle se font les captures). La zone couverte par le chalutage est donc très faible. En compensation, la maille de prélèvement est très fortement resserrée. Dans notre cas, seuls les espagnols utilisent cet engin de capture. Les échantillons sont espacés de 3 milles nautiques le long de transects parallèles avec une distance inter-transect de 15 milles nautiques; soit une configuration très différente de celle fixée *a priori* et adoptée par les utilisateurs des autres types d'engins.

Le **volume d'eau filtrée** ou **volume filtré** se déduit simplement à partir de la surface d'ouverture du chalut (s) et de la profondeur atteinte par ce dernier (p): $v = p \times s$.

Chalutages obliques

Le Bongo Net et le GulfIII sont utilisés en chalutage oblique (Fig. 2.5). Le chalutage se fait alors pendant que le bateau fait route à la vitesse standard de 5 nœuds. Le chalut plonge depuis la surface jusqu'à une profondeur maximale de 200 m et remonte ensuite. Les captures se font pendant le filage et le virage. Les temps de traîne sont de l'ordre de 3/4 d'heure de sorte que les distances couvertes par coup de chalut oblique peuvent atteindre une dizaine de kilomètres.

La profondeur maximale atteinte par le chalut, encore appelée profondeur chalutée, est limitée soit par la bathymétrie, soit par la profondeur de la thermocline quand celle-ci est bien nette (au moins 2.5°C en 10 mètres (Anon., 1994)) soit enfin par la profondeur de 200 mètres quand la colonne d'eau n'est pas fortement stratifiée.

- Volume filtré: v

Les chaluts sont équipés de courantomètres. Il est ainsi possible de connaître la "**longueur chalutée**", c'est-à-dire la distance parcourue par le chalut, et par suite, avec la profondeur, la "**longueur de traîne**", c'est-à-dire la longueur parcourue par le navire (l). Connaissant la surface d'ouverture du chalut (s), le volume filtré est:

$$v = l \times s$$

Il existe aujourd'hui un consensus pour considérer que les mesures de volumes filtrés sont entachées d'une erreur importante. Cependant, cette imprécision n'est pas connue même si on retient généralement 10% comme un ordre de grandeur d'erreur possible (J. Nichols, *com. pers.*)

- Volume échantillonné: V

Sous réserve d'homogénéité de la répartition des œufs dans la zone concernée, les chalutages obliques permettent en fait d'estimer la quantité d'individus présents dans la tranche d'eau définie

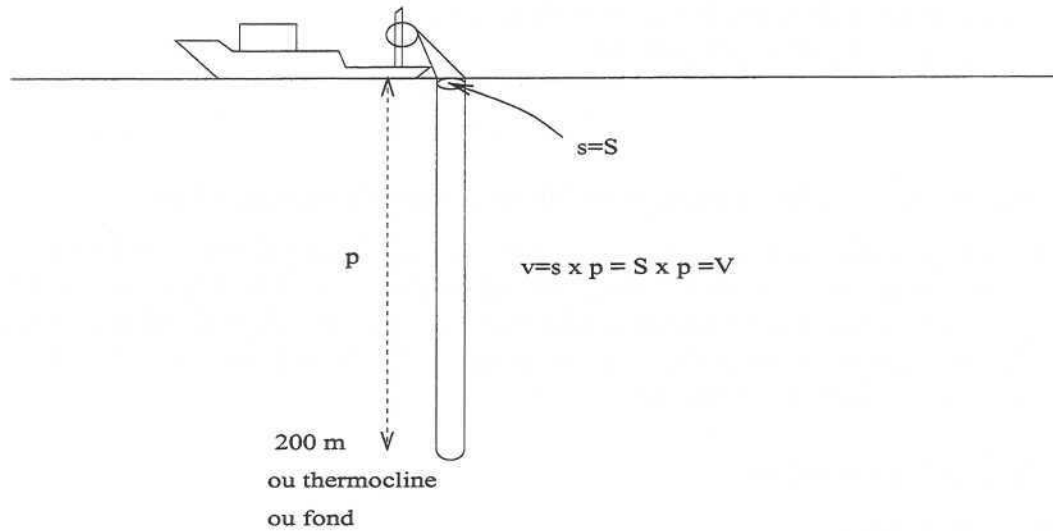


FIG. 2.4 - Chalutage vertical de type "Calvet Net". Notations: profondeur (p), surface du chalut (s), volume filtré (v), volume échantillonné (V), support (S).

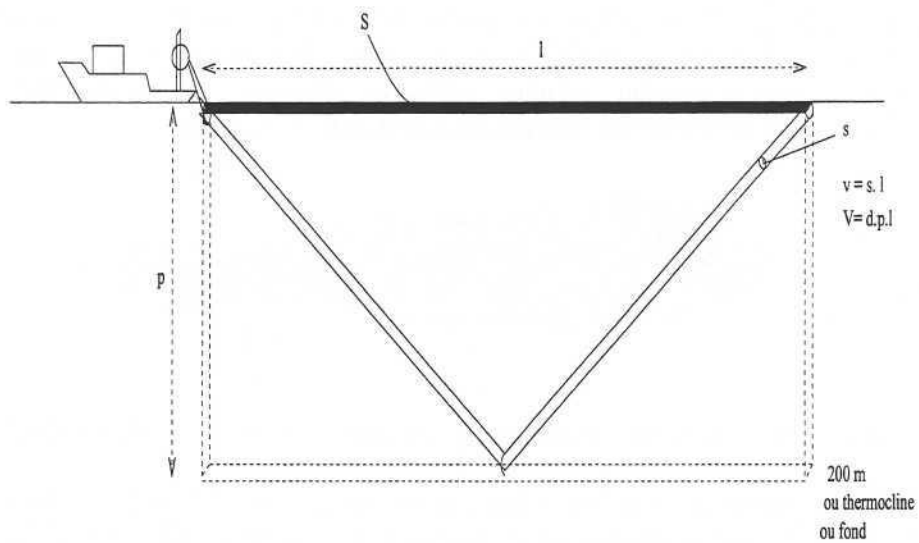


FIG. 2.5 - Chalutage oblique de type "Bongo" ou "Gulf III". Notations: profondeur (p), surface du chalut (s), diamètre du chalut (d), volume filtré (v), volume échantillonné (V), support (S).

par le parallélépipède (Fig. 2.5):

de hauteur, la profondeur échantillonnée (p)
de longueur, la distance parcourue par le navire (l)
de largeur, le diamètre d'ouverture du chalut (d)

de sorte que le volume échantillonné est:

$$V = l \times p \times d$$

Commentaires sur la distribution verticale des œufs de maquereaux

Plusieurs expérimentations (Coombs *et al.*, 1981; Anon, 1993) traitent de la distribution verticale des œufs de maquereau. Il ressort de ces études que les œufs se trouvent dans les 200 mètres sous la surface, ou moins, lorsque la colonne d'eau est fortement stratifiée (thermocline importante). Limiter l'échantillonnage à une profondeur maximale de 200 m ne génère donc pas de biais dans les mesures; tous les œufs sont capturés.

2.1.3 Variables mesurées

Variables biologiques

- Espèces et stades de développement (œufs et larves)

Chaque prélèvement est fixé dans une solution formolée et analysé ensuite à terre. Les œufs sont répartis par espèce: maquereau (*Scomber scombrus*), chinchard (*Trachurus trachurus*), et autres. Il sont ensuite regroupés par stade de développement. Pour le maquereau, le découpage se fait selon la classification en cinq classes proposée par Lockwood *et al.* (1977) et illustrée figure 2.7. Une étude comparative de la détermination des stades de développements des œufs de maquereau a été conduite sous la collaboration du CIEM et montre que les stades I et V sont ceux qui sont reconnus avec le plus de consistance (Anon., 1993). Ce sont donc ces deux stades, notés E1 et E5, que nous choisirons pour décrire l'évolution des stades pré-larvaires.

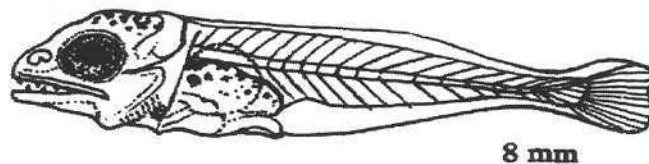


FIG. 2.6 - Larve de maquereau du troisième groupe de taille (8 mm).

Les larves sont également regroupées par espèce: maquereau (*Scomber scombrus*), chinchard (*Trachurus trachurus*), merlu (*Merluccius merluccius*), merlan bleu (*Micromesistius poutassou*), et autres. Pour le maquereau, et en 1989 uniquement, les larves ont également été réparties en trois classes de taille notées L1, L2 et L3 et définies par les bornes suivantes :

- L1: longueur comprise entre 2.5 mm et 3.9 mm
- L2: longueur comprise entre 4 mm et 5.9 mm
- L3: longueur comprise entre 6 mm et 13 mm (Fig. 2.6)

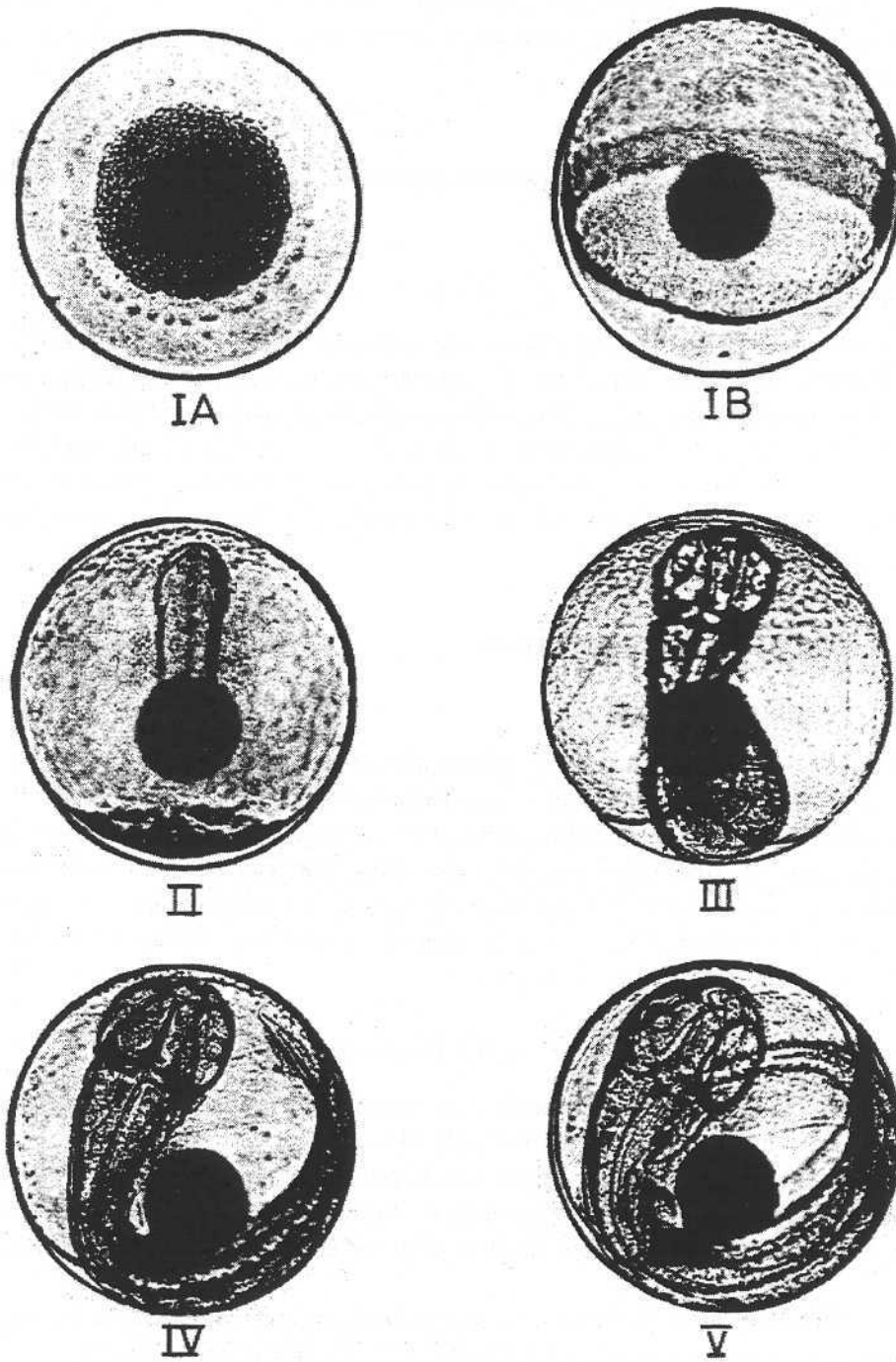


FIG. 2.7 - Illustration des 5 stades de développement des œufs. D'après Lockwood et al., 1977.

- Estimation des densités par unité de volume.

Lorsque le nombre d'œufs capturés est trop important, les densités sont estimées au travers d'un sous-échantillonnage des captures. Les captures de larves sont, par contre, comptées exhaustivement et systématiquement.

- Zones informées.

Seules les densités de larves comprises entre les latitudes 48°00N et 55°00N m'ont été transmises.

Variables hydrographiques

En plus de la profondeur, la température de surface (SST), la température à 20 mètres de profondeur (SST20) susceptible d'indiquer la présence ou non d'une thermocline et la salinité de surface (SSS) sont mesurées aux points de prélèvement. Malheureusement ces variables ne sont pas disponibles en tous les points échantillonnés. Le tableau 2.2 indique, par période, le nombre de données disponibles relatives à ces trois variables physiques. Globalement, depuis 1983, la quantité d'informations a progressé jusqu'à devenir presque parfaite en 1995 où le triplet (SST, SST20) est mesuré systématiquement.

2.2 Traitements préliminaires

2.2.1 Echantillons aberrants

Le travail de détection des échantillons aberrants est en fait réalisé par le groupe de travail du CIEM chargé de l'évaluation du stock de maquereau de l'Ouest européen.

Ainsi, conformément aux recommandations de ce groupe de travail (Anon., 1990) j'ai **supprimé des analyses la première période de 1989**. Les raisons principales sont que le plan d'échantillonnage n'a absolument pas été respecté puisque les prélèvements se répartissent le long des accores uniquement (annexe 2), et que les densités observées sont anormalement élevées sans que les experts aient pu fournir d'explication.

2.2.2 Passage à une densité par unité de surface

Pour supprimer l'effet dû à la profondeur, il convient de cumuler les densités le long de la colonne d'eau et de s'intéresser à la densité d'individus par unité de surface (inds/m²). Dans le cas des chalutages verticaux il s'agit d'un cumul exact, puisque toute la colonne d'eau a été chalutée. Cependant, dans le cas des chalutages obliques, le cumul n'est qu'une estimation de la biomasse cumulée dont l'erreur est d'autant plus faible que la densité se répartit verticalement de manière homogène.

Finalement, connaissant le nombre d'œufs capturés, le volume filtré et la profondeur de la tranche d'eau chalutée, on en déduit une **densité d'œufs par mètre carré**.

2.2.3 Taux journaliers de production

Les cinq stades de développement retenus ici (E1, E5, L1, L2 et L3) n'ont pas, en moyenne, la même durée de vie. Les individus de chacun de ces stades n'ont donc pas la même disponibilité à la capture : plus un stade de développement est long plus il a de "chances" d'être capturé. Il faut par conséquent tenir compte de la durée des 5 stades de développements pour comparer leurs densités.

année et période	nombre de densités d'œufs	nombre de valeurs de SST	nombre de valeurs de SST20	nombre de valeurs de SSS
1983 - 1	110	190	0	108
1983 - 2	163	198	0	110
1983 - 3	253	263	22	202
1983 - 4	124	120	12	111
1983 - 5	104	131	0	7
1986 - 1	91	91	0	0
1986 - 2	421	410	168	271
1986 - 3	273	274	0	140
1986 - 4	51	51	0	0
1989 - 1	75	81	80	80
1989 - 2	140	140	0	140
1989 - 3	329	329	241	241
1989 - 4	291	273	167	177
1989 - 5	88	86	88	0
1992 - 1	187	184	187	180
1992 - 2	758	758	758	399
1992 - 3	188	186	188	186
1992 - 4	40	40	40	40
1995 - 1	129	129	129	129
1995 - 2	189	189	189	189
1995 - 3	232	232	232	123
1995 - 4	370	370	370	354
1995 - 5	146	146	146	146

TAB. 2.2 - Plan d'échantillonnage. Variables physiques: température de surface et à 20 mètres, salinité de surface. Nombre d'observations par période.

Lockwood *et al.* (1981) ont montré que les vitesses de développement des œufs de maquereau dépendent de la température de l'eau selon une relation de type exponentiel. A 12°C par exemple, le premier stade dure 44 heures alors que le cinquième dure 18 heures. A l'aide de leurs abaques et de la température de surface mesurée *in situ*, les densités par unité de surface ont été transformées en **taux de productions journaliers** ou densités journalières exprimés en **nombre d'individus par mètre carré et par jour** (œufs/m²/jour). Par abus de langage, je continuerai dans la suite de ce document à parler de densités pour désigner un taux de production journalière.

La vitesse du développement larvaire est réputée dépendre, elle aussi, de la température. On ne dispose cependant pas de modèle qui décrit cette relation. En concertation avec des experts "ès maquereau" (M. Walsh, J. Fives), nous avons choisi pour les trois classes de taille de larve, les durées suivantes : 5, 10 et 15 jours. Ainsi, à 12°C, 5 semaines s'écoulent entre la fécondation d'un œuf et la fin du dernier stade larvaire (Tab. 2.3).

stade	code	désignation	durée à 12°C	temps écoulé depuis la fécondation
œuf I	E1	de la fécondation à un blastodisque en anneau	44 heures	44 heures
œuf II à IV				
œuf V	E5	queue de l'embryon dépasse la tête	18 heures	6 jours
larve I	L1	2.5 mm < longueur < 4 mm	5 jours	11 jours
larve II	L2	4 mm < longueur < 6 mm	10 jours	3 semaines
larve III	L3	6 mm < longueur < 13 mm	15 jours	5 semaines

TAB. 2.3 - Stade de développements pré-larvaires et larvaires. Désignation, durée à 12°C et temps écoulé depuis la fécondation de l'œuf.

2.2.4 Homogénéité des mesures hydrographiques

Température de surface

Le recouvrement géographique de certaines zones échantillonnées par différents bateaux à quelques jours d'intervalle permet de comparer leurs mesures de SST et d'évaluer la précision des observations hydrographiques. Pour cela, j'ai considéré que deux observations ayant eu lieu le **même jour et distantes de moins de 15 milles nautiques** constituaient un doublon.

En général, on attend des mesures faites par un même navire qu'elles soient cohérentes, c'est-à-dire peu différentes pour des doublons. Sur les 639 doublons relevant d'un même navire, disponibles entre 1983 et 1995 (Fig. 2.8a), l'écart moyen est de 0.22°C, et atteint 2.7°C; les mesures faites par les Pays-Bas (7) sont les plus "bruitées". Sous l'hypothèse que la température de surface n'a pas le "temps" d'évoluer à l'échelle d'une journée et de 15 milles nautiques, on pourra retenir que, **pour un navire, les mesures de température de surface sont fiables, en moyenne, à un quart de degré près.**

Qu'en est-il des 41 répliqués issus de deux navires différents (Fig. 2.8b)? Si l'écart maximum reste à peu près le même (2.5°C), la moyenne des (valeurs absolues) des écarts est maintenant de 0.53°C. Dans une analyse combinée de l'ensemble de ces données on devra donc garder à l'esprit qu'il existe, en moyenne, une **incertitude d'environ 0.5°C** concernant la température de surface.

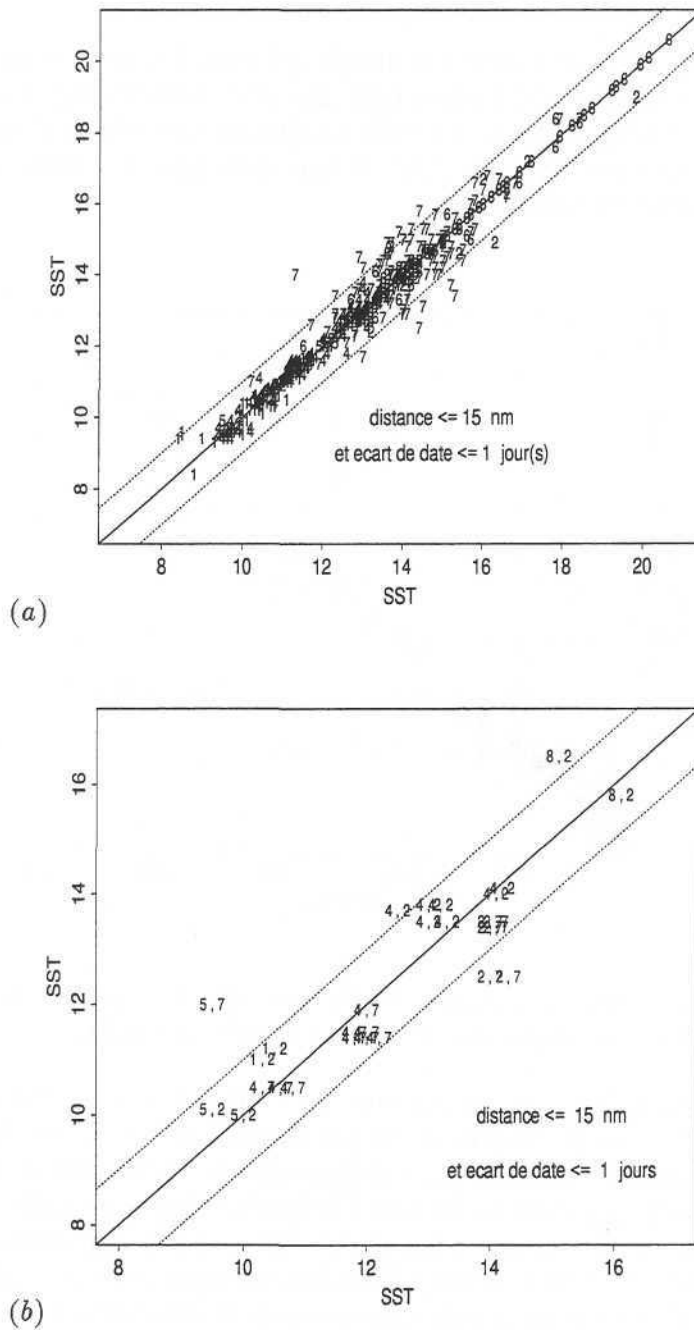


FIG. 2.8 - Comparaisons des SST provenant d'échantillons prélevés le même jour et distants de moins de 15 milles nautiques. Code des pays: 1 = Allemagne, 2 = Ecosse, 3 = Espagne, 4 = Angleterre, 5 = Irlande, 6 = France, 7 = Pays-Bas, 8 = Norvège.

(a) : Doublons pour un même bateau représentés par le code du navire.

(b) Doublons pour des bateaux différents. L'abscisse est fixée par le premier bateau (premier code) et l'ordonnée par le second (deuxième code).

Profondeur

De la même manière, on peut s'intéresser aux fluctuations des mesures de la bathymétrie faites par un même bateau en deux points distants de moins de 5 milles nautiques (il n'y a pas besoin de se fixer ici une limite de temps). De 1983 à 1995, on dispose de 1173 répliqués de ce genre. L'écart (absolu) moyen est important (120 m). Dans de nombreux cas, la différence est très importante (Fig. 2.9). Elle peut atteindre 3700 mètres.

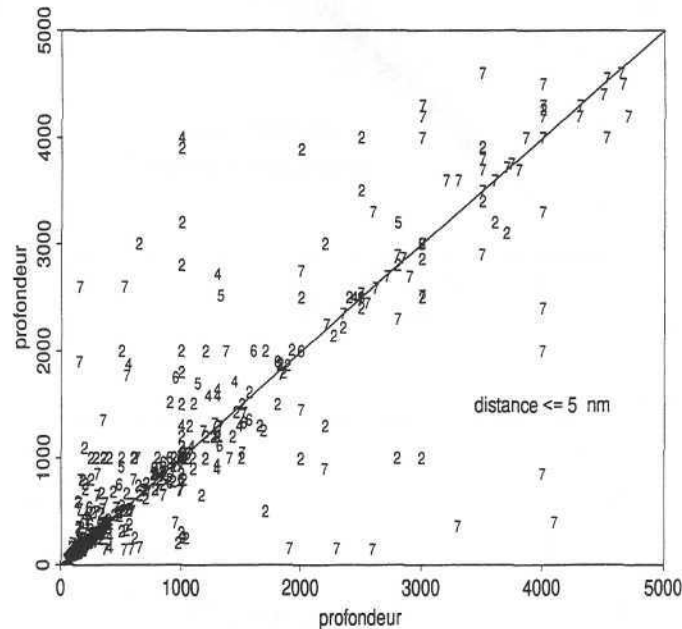


FIG. 2.9 - Répliqués de mesures de profondeur. Mesures effectuées par un même bateau à moins de 5 milles nautiques d'écart. Répliqués représentés par le code du navire.

Ces différences importantes sont, pour partie, dues aux éléments suivants :

- Le mode de mesure. La profondeur est en général mesurée au milieu du coup de chalut par sondage acoustique. Certains navires peuvent avoir indiqué la profondeur en début ou fin de traîne.
- La proximité du talus continental. La zone d'échantillonnage chevauche les accores européens. Or, au delà de 200 m, la profondeur peut chuter extrêmement rapidement. Un petit décalage géographique peut alors être associé à une différence très importante en profondeur.
- Enfin, il reste toujours des erreurs de mesures ou de retranscription possibles.

Pour l'ensemble de ces raisons, j'ai préféré utiliser les profondeurs fournies par des bases de données internationales mises au point et utilisées par les hydrophysiciens. Ces bases de données indiquent la profondeur aux nœuds d'une grille de 5 minutes d'angle.

Température de surface *in situ* et satellitaire

Dans le dernier chapitre de cette partie, nous aurons recours à des mesures de température de surface par satellite pour étudier la possibilité d'un référentiel en température dans la régionalisation de la ponte. Je me propose de tester, sur la base d'un exemple, la plus ou moins bonne correspondance entre les mesures *in situ* et les mesures effectuées par satellite.

Les mesures de température de surface par satellite ont été traitées et fournies par le "Jet Propulsion Laboratory" de la NASA par l'intermédiaire de Steeve Grooms de "Natural Environment Research Council, Computer Service" de l'Université de Plymouth (Angleterre). Elles correspondent aux moyennes hebdomadaires de température de surface par carré de 0.088 degré de côté pour le printemps 1989 (soit 14 semaines en tout, de début Avril à fin Juillet). La représentation en niveau de gris pour les 5 semaines comprises entre le 19 Mai et le 26 Juin 1989, et numérotées de 7 à 11, sont fournies en annexe 2. Malgré le fait que les données sont moyennées sur une semaine, on peut constater que, dans certains cas, la couverture nuageuse reste très importante (semaine 7 et 10 en particulier).

période d'échantillonnage	semaines
2	4 + 5 + moitié 6
3	7 + 8 + moitié 9
4	moitié 9 + 10 + 11
5	moitié 12 + 13

TAB. 2.4 - Correspondance entre les périodes d'échantillonnage et les 14 semaines du printemps de l'année en 1989. Semaine 1: du 2 au 9 Avril. Semaine 14: du 18 au 25 Juillet.

En 1989, les correspondances entre les dates des périodes d'échantillonnage (Tab. 2.1) et les semaines de l'année sont indiquées dans le tableau 2.4.

D'après les figures de l'annexe 2, on peut voir qu'en 1989 la couverture la plus large en température de surface a été réalisée pendant la période 3 et que la semaine n°8 est très peu "polluée" par des nuages. J'ai donc choisi de comparer les mesures faites *in situ* pendant la période 3, aux mesures satellitaires de la huitième semaine. Pour cela j'ai informé une grille régulière de 0.5 x 0.5 degré par :

- la mesure satellitaire la plus proche
- la moyenne des échantillons de température de surface tombant dans un carré de 0.5 x 0.5 degré autour de chacun des nœuds de la grille.

La corrélation entre les deux sources d'information apparaît globalement assez bonne (Fig. 2.10); en particulier, pour les valeurs de température mesurées *in situ* supérieures à 12°C. La variabilité des mesures satellitaires est toutefois plus faible que celle des observations *in situ* et la régression linéaire entre les deux variables montre que le satellite fournit des mesures de SST plus élevées que les observations quand la température est faible et moins élevées quand la température est forte.

Le code du pays responsable de chaque mesure de température de surface a été moyenné et migré vers la grille régulière (une valeur non entière indique alors que plusieurs pays sont à l'origine de la valeur implantée en ce nœud de la grille). Il ressort alors que les principaux décalages entre satellite et mesures effectives concernent les mesures réalisées par le(s) bateau(x) irlandais (n° 5 sur la figure): les mesures *in situ* sont inférieures d'environ 2°C aux mesures satellitaires.

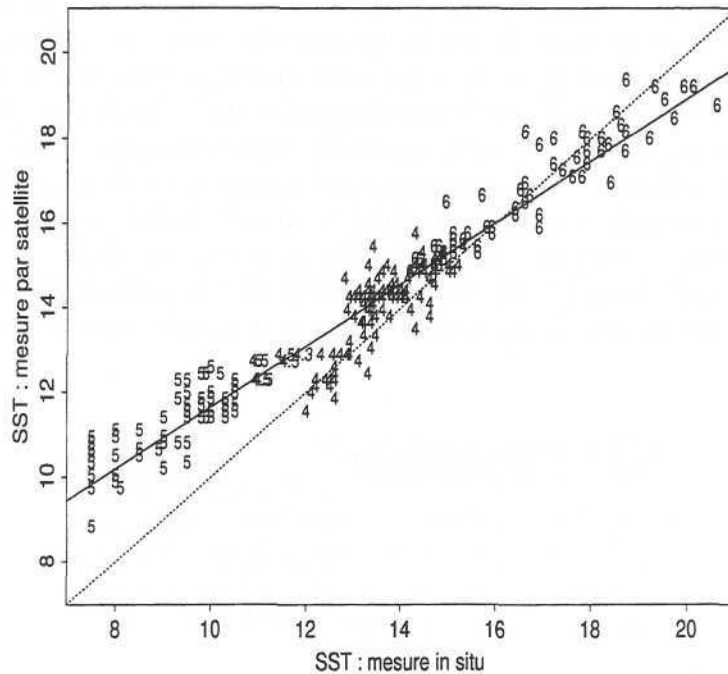


FIG. 2.10 - Comparaison de la température de surface mesurée *in situ* (période 3 1989) et par satellite (8ème semaine). Indication du code du pays responsable de la mesure. Régression linéaire (trait plein) et première bissectrice (trait pointillé).

D'après la représentation proportionnelle des mesures de température de surface réalisées pendant la période 3 (annexe 2), on peut penser que le décalage observé précédemment provient plus d'une erreur de mesures *in situ* que des mesures satellitaires. En effet, en plus du fait que ce décalage apparaît à l'occasion d'un changement de bateau, on peut voir que la région au Nord de l'Irlande présente des températures nettement différentes de celles mesurées légèrement plus au Sud. Cette forte discontinuité spatiale vient renforcer l'idée que les mesures *in situ*, réalisées pendant cette période par les Irlandais, sont probablement sous estimées.

Je retiendrai finalement que, pour des températures intermédiaires (entre 11 et 18°C) le niveau de correspondance entre les deux sources d'information est du même ordre de grandeur que le niveau de précision des mesures de température de surface réalisées en mer. Il reste cependant que les mesures satellitaires sont quasi systématiquement supérieures à celles réalisées en mer. On pourra donc travailler alternativement avec les unes ou les autres tout en sachant que le centre de gravité des températures *in situ* devrait être inférieur à celui des mesures satellitaires d'environ 0.5°C, et leur inertie supérieure.

2.2.5 Migration vers une grille régulière en degrés

Principe

Au sein de chaque période d'échantillonnage, les données sont considérées comme simultanées à une date fixée arbitrairement à la date moyenne des différents prélèvements de la période (Tab. 2.1).

Dans la partie consacrée à l'estimation du covariogramme, nous avons vu que la situation la plus confortable est celle des mailles régulières. Ici, on dispose d'échantillonnages essentiellement réguliers auquel s'ajoutent des perturbations. Par ailleurs, nous venons de voir que le niveau d'imprécision des mesures hydrographiques, *a priori* plus commodes à mesurer que les densités d'œufs, est non négligeable. On doit donc s'attendre à ce que les erreurs de mesures de densités soient plus importantes.

Par conséquent, j'ai choisi de travailler sur la base d'une **grille régulière informée en chacun de ses nœuds par la moyenne des échantillons** se trouvant dans la maille centrée en ce nœud. Ainsi, pour chaque période d'échantillonnage, les données sont migrées et moyennées vers la grille standard de 0.5 degrés de côté.

Les conséquences de cette migration en termes de moyenne et de variance ne sont pas triviales. Conformément aux règles de la régularisation, on doit cependant s'attendre à ce que la variabilité des données moyennées soit inférieure à celle des données brutes, et que leur structure spatiale soit légèrement plus régulière.

Généralités

Les représentations proportionnelles des taux journaliers de production de chacun des stades de développement pour chaque période d'échantillonnage sont présentées dans l'annexe 2.

Les particularités importantes à noter, dans l'ordre des stades de développements, sont les suivants :

- œufs de stade 1 (E1) :

- En 1983 l'échantillonnage ne va pas au-delà de 54°00N. Pendant la période 1, quelques échantillons ont été prélevés sur le plateau continental espagnol et sont retirés des analyses.
- En 1986, la densité la plus forte observée pendant la première période se trouve isolée au Nord de la zone échantillonnée. La dernière période comprend peu d'échantillons.
- En 1989, une densité forte a été observée pendant la période 3, dans la partie Sud de la zone de frai et très au large des accores.
- En 1992, la dernière période d'échantillonnage possède très peu de points. Elle est donc probablement assez peu représentative.
- En 1995, pendant les périodes 2 et 3, des densités non négligeables sont observées très au large des accores.

- œufs de stade 5 (E5) :

- En 1983, la période 5 ne comprend que 5 échantillons.
- En 1989, pendant la période 3, les densités les plus fortes sont observées en bordure Ouest des transects. L'ensemble du stock n'a donc probablement pas été couvert.

- Larves de toutes tailles.

- En 1983, la période 5 n'est pas informée. La période 4 ne comprend que 10 échantillons.

- En 1986, seules les périodes 2 et 3 sont informées.
- Pas d'information en 1992 et 1995.

Calcul des distances

Pour toutes les statistiques faisant référence à la distance entre points (inertie, covariogramme), ou à la surface d'influence des échantillons (abondance), il convient de transformer les axes de la grille en unité de distance comparable.

Une méthode rigoureuse, basée sur la transformation des longitudes en fonction de la latitude, aboutirait à une grille irrégulière. J'ai préféré conserver les avantages de la grille régulière en affectant simplement un même coefficient correcteur à toutes les abscisses de la grille. La latitude moyenne étant d'environ 53 degrés, j'ai multiplié par 0.6 ($\cos(53)=0.6$) les abscisses de la grille.

2.3 Description d'une saison de ponte moyenne

Pour avoir une idée assez grossière du déroulement d'une saison de ponte, je me suis servi des moyennes de productions journalières fournies par Martin Walsh. Ce dernier a calculé, par mois, la moyenne des productions journalières observées par carré de 0.5 degrés de côté entre 1977 et 1995. Certains des carrés contiennent une seule valeur alors que d'autres représentent la moyenne de sept observations. Il est donc préférable de ne pas trop affiner l'analyse de ces données. On se bornera à représenter les productions journalières par des figurés proportionnels et à calculer les covariogrammes expérimentaux de ces distributions.

2.3.1 Représentations proportionnelles

Une fois moyennée sur la période 1977-95, les densités fluctuent entre 0 et 700 œufs/m²/jour et permettent de fixer l'ordre de grandeur d'un certain nombre de caractéristiques. Le maximum de production suit une évolution en forme de cloche au cours de la saison, avec un maximum en Avril (de Mars à Juillet : 150, 700, 460, 250, 56 œufs/m²/jour). La production totale d'œufs a une allure de dôme avec un maximum en Mai où il y a quelques 2 10¹³ œufs produits par jour.

Chaque distribution mensuelle moyenne est représentée avec un figuré proportionnel jusqu'à 300 œufs/m²/jour (Fig. 2.11). Les densités supérieures sont indiquées par un triangle, les densités nulles par une croix, et le centre de gravité de la population d'œufs par le numéro du mois concerné avec une taille proportionnelle à l'abondance observée ce mois-là.

La régularisation (moyenne) qu'ont subie les données se traduit sur les cartes par une succession de distributions spatiales assez régulières. Un examen attentif permet tout de même de voir que des fortes valeurs côtoient des densités faibles ou nulles. On voit par ailleurs que la ponte, au moins jusqu'en Juin, est assez fortement liée aux accores. Enfin, il est intéressant de noter qu'en Mai, les densités observées à l'extrémité Ouest des transects sont parmi les plus fortes. Il n'est donc pas certain que l'ensemble de la production ait été couvert. On verra plus loin dans le document que cela est associé à un déplacement vers l'Ouest de la population au cours de la période 1977 - 19

Le déplacement moyen de la ponte au cours d'une saison est schématisé par l'évolution des centres de gravité. De Mars à Mai, la population se déplace vers le Nord-Ouest en longeant les accores. Ensuite, dans la deuxième moitié de la saison, les œufs sont pondus en moyenne plus au retrait sur le plateau continental et légèrement plus au Sud-Est.

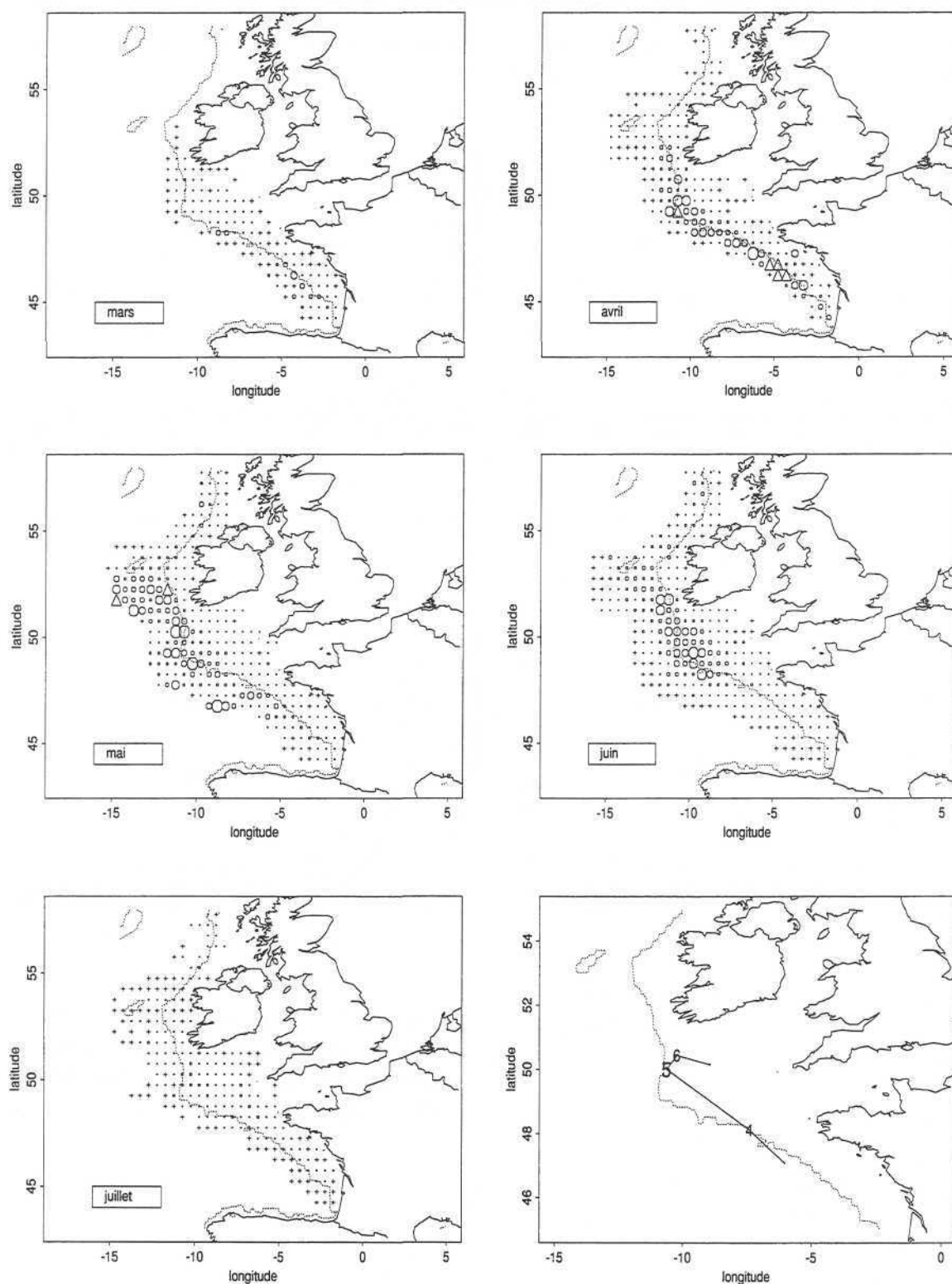


FIG. 2.11 - Représentation proportionnelle des taux de production moyens par mois et par carré de $0.5 \text{ degré} \times 0.5 \text{ degré}$ de 1977 à 1995 : (a) Mars, (b) Avril, (c) Mai, (d) Juin, (e) Juillet. (f) Représentation du centre de gravité de chaque distribution.

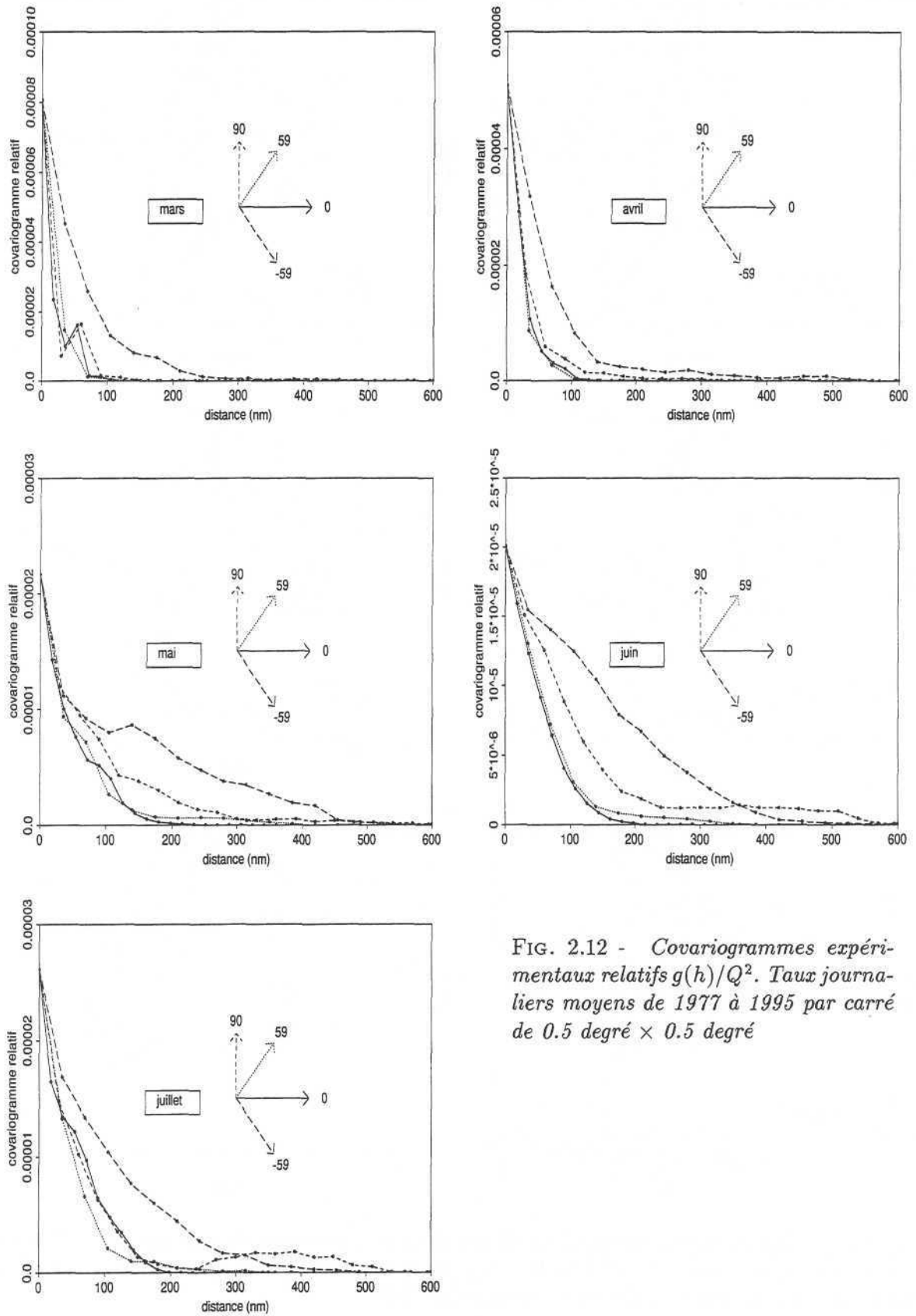


FIG. 2.12 - Covariogrammes expérimentaux relatifs $g(h)/Q^2$. Taux journaliers moyens de 1977 à 1995 par carré de $0.5 \text{ degré} \times 0.5 \text{ degré}$

2.3.2 Covariogrammes moyens

Les covariogrammes expérimentaux sont calculés selon les directions principales et diagonales de la grille, soient : 0, 59, 90 et 121 (-59) degrés. On notera les trois points suivants (Fig. 2.12) :

- La distribution est anisotrope quel que soit le mois considéré. Les trois directions 0, 59 et 90 degrés présentent plus ou moins la même structuration alors que la structure dans la direction Nord-Ouest est différente avec une portée plus longue et une décroissance à l'origine moins rapide. C'est la direction de plus grande continuité spatiale.

- La portée, c'est-à-dire l'extension spatiale de la population, a tendance à augmenter entre Mars et Juillet.

- L'agrégation à petite distance ($g(0)/Q^2$) diminue au cours de la saison.

Il apparaît donc que, en moyenne, les œufs produits en début de saison sont très agrégés avec une faible extension spatiale. Au cours du printemps, l'extension géographique de la population augmente parallèlement à une diminution du niveau d'agrégation.

Ce n'est là qu'une description sommaire de la structure de la population d'œufs. La régularisation affecte l'allure des covariogrammes et nous verrons par la suite que la réalité est moins régulière que cela.

La régularisation provient ici de ce que les valeurs sont le résultat d'une moyenne de plusieurs échantillons. Une autre façon de régulariser consiste, une année donnée, à augmenter le volume d'eau chalutée, soit en agrandissant l'ouverture du chalut soit en allongeant le temps de traîne. Il s'agit alors d'étudier l'effet régularisant d'un changement de support.

2.4 Effet de support

2.4.1 Définition du support

Quel que soit l'engin utilisé (Calvet Net, Gulf III, Bongo) et le mode de pêche (vertical ou oblique), l'information collectée au départ est un nombre d'individus par unité de volume : N/v (Tab. 2.5). Le support dans ce cas là, n'est autre que le volume d'eau chaluté (Tab. 2.5). Il s'exprime en mètre cube et sa définition dépend du mode de pêche.

Mais la variable de travail correspond au cumul des densités le long de la verticale : $\frac{N}{v}p$ (Tab. 2.5). Il convient donc que l'effet de support porte sur cette variable. Le support s'exprime en mètre carré et sa définition dépend du mode de pêche comme cela est expliqué ci-dessous.

Chalutages verticaux

Dans le cas d'un chalutage vertical, le support des densités par unité de surface (densités 2D) correspond à la surface d'ouverture du chalut (Fig 2.4 et Tab. 2.5). C'est le rapport entre le volume filtré (v) et la profondeur (p). En somme, **le support monte avec la densité.**

Chalutages obliques

Dans le cas d'un chalutage oblique, le support ne correspond plus à la surface d'ouverture du chalut (Fig. 2.5). Il est défini par la surface du rectangle de largeur le diamètre d'ouverture du chalut (d) et de longueur la distance parcourue par le navire (l). C'est donc le rapport entre le volume échantillonné (V) et la profondeur échantillonnée (Fig 2.4 et Tab. 2.5). Ainsi, la densité 2D

	3D		2D	
chalutage	support (m^3) S_{3D}	densité $(inds/m^3)$ $z_{3D} = \frac{N}{S_{3D}}$	support (m^2) S_{2D}	densité $(inds/m^2)$ $z_{2D} = \frac{N}{S_{2D}}$
vertical	$v = p.s$	$\frac{N}{v}$	$\frac{v}{p} = s$	$\frac{N}{v} p$
oblique	$v = l.s$	$\frac{N}{v}$	$\frac{V}{p} = l.d$	$\frac{\frac{N}{v} V}{ld} = \frac{N}{v} p$

TAB. 2.5 - Définition des densités à 3 et 2 dimensions et leurs supports correspondants selon le mode de pêche. Le nombre d'individus capturés est noté "N", le volume filtré "v", le volume échantillonné "V", la profondeur atteinte par le chalut "p", la distance parcourue par le navire "l", la surface d'ouverture du chalut "s" et le diamètre d'ouverture du chalut "d".

correspond au nombre estimés d'individus dans V à partir du nombre d'individus capturés dans v, soit $\frac{N}{v}V$, ramené au support 2D, soit :

$$z_{2D} = \frac{\frac{N}{v}V}{ld}$$

2.4.2 Mesure de l'effet de support : campagne 1992 - période 2.

Les situations qui permettent la comparaison de plusieurs supports ne sont pas nombreuses. Tout d'abord les informations nécessaires aux calculs des supports ne sont disponibles que pour les deux dernières années (1992 et 1995). Pour les années précédentes, seul le volume filtré et la profondeur d'échantillonnage sont connues; ce qui est insuffisant pour remonter jusqu'au support pour les chalutages obliques. Par ailleurs, le Calvet Net (chalutage vertical) n'est utilisé que par les Espagnols pendant la deuxième période d'échantillonnage de 1992. C'est donc cette campagne que j'ai utilisée pour analyser l'effet de support sur les densités d'œufs de maquereau de stade I (E1).

Pendant cette campagne, le support de la densité varie entre $0.045 m^2$ (surface d'un Calvet Net dont le diamètre est 24 cm) et $1500 m^2$ (Fig 2.13). J'ai décidé de considérer trois classes de support dont les statistiques élémentaires figurent dans le tableau 2.6. On notera en particulier que les supports de la deuxième classe sont en moyenne 10 000 fois plus grands que ceux de la première classe et deux fois plus petits que ceux de la dernière. Il y a donc une différence énorme en terme de support entre les chalutages verticaux (support de la classe 1) et obliques (supports des classes 2 et 3).

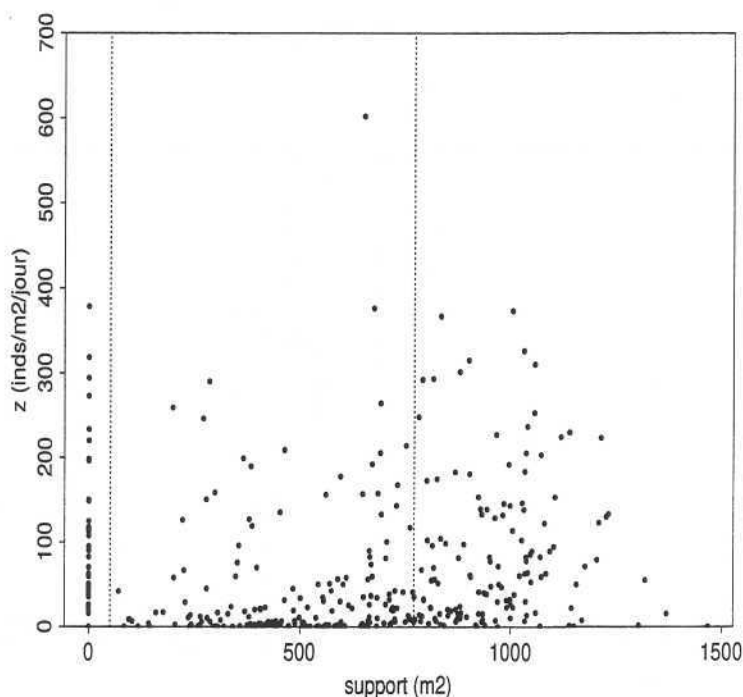


FIG. 2.13 - Corrélation entre le support (m^2) et la densité (z) exprimée en œufs/ m^2 /jour. 1992 période 2. Délimitation par une ligne verticale des trois classes de support.

classe	support (m^2)			n	densités (œufs/ m^2 /jour)			C.V.	abondance (inds/jour)
	min	max	moyenne		min	max	moyenne		
1	0.045	0.045	0.045	446	0	378	14.6	2.9	$1 \cdot 10^{12}$
2	50	770	488	176	0	602	45	1.75	$1.6 \cdot 10^{13}$
3	770	1500	967	136	0	373	84	1.05	$1.6 \cdot 10^{13}$

TAB. 2.6 - Statistiques des différentes classes de support. Support minimum, maximum et moyen; nombre d'échantillons (n); densités minimum, maximum, moyenne et coefficient de variation (C.V.); abondance (en nombre d'œufs de stade I par jour).

La mesure de l'effet de support est rendue difficile par le fait que les données de supports différents n'ont pas été récoltées aux mêmes endroits. En particulier, les échantillons à petits supports ont été prélevés dans une zone à ponte faible (partie Sud du Golfe de Gascogne), alors que les échantillons ayant un grand support proviennent de zones plutôt riches. La comparaison des densités moyennes par classes de support ne présente donc pas d'intérêt ici. On choisira plutôt d'analyser la différence entre supports au travers de l'évolution du coefficient de variation des densités.

L'augmentation du support se traduit par une diminution importante du coefficient de variation (Tab. 2.6). Par ailleurs, l'histogramme des densités pour chacune des classes de support (Fig. 2.14 a,b,c) montre que cette diminution est surtout due à la diminution de la proportion de densités très

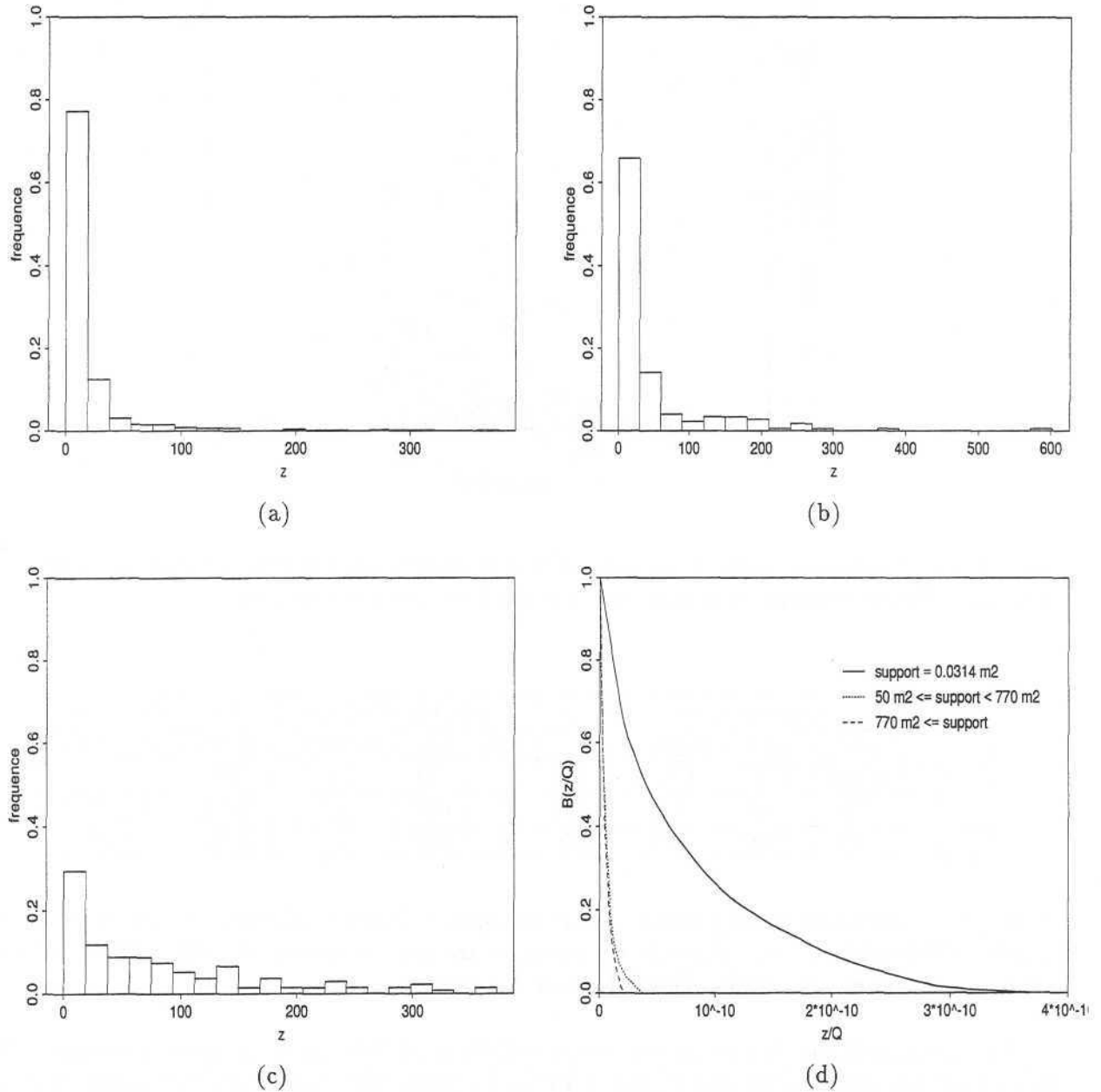


FIG. 2.14 - Effet régularisant d'une augmentation du support.

(a,b,c): Histogramme des densités (z) exprimées en œufs/m²/jour pour chacune des trois classes de support.

(d): Bénéfice conventionnel des densités relatives (z/Q) pour chacune des trois classes de support.

faibles ou nulles. Les densités fluctuent entre 0 et 300 ou 600 selon la classe, mais la proportion de 0 diminue. C'est là une première manifestation de l'effet régularisant d'une augmentation du support.

La comparaison des courbes de bénéfice conventionnel est elle aussi rendue délicate par la provenance géographique des données de différents supports. Une façon de contourner cet obstacle est de considérer la densité relativement à l'abondance (z/Q). Il s'agit donc d'estimer préalablement l'abondance associée aux densités appartenant à chacune des classes de support, en définissant la surface d'influence de chaque échantillon par rapport aux autres échantillons de sa classe. Une fois les densités "standardisées", on voit (Fig. 2.14 d) que le bénéfice conventionnel diminue lentement avec z/Q pour les tous petits supports alors qu'il chute extrêmement rapidement, et de manière identique, pour les deux classes de support associées aux chalutages obliques. Puisque l'abondance associée à ces deux classes est identique, il est possible de vérifier la similitude des courbes de bénéfices conventionnels pour ces deux dernières classes (Fig. 2.15).

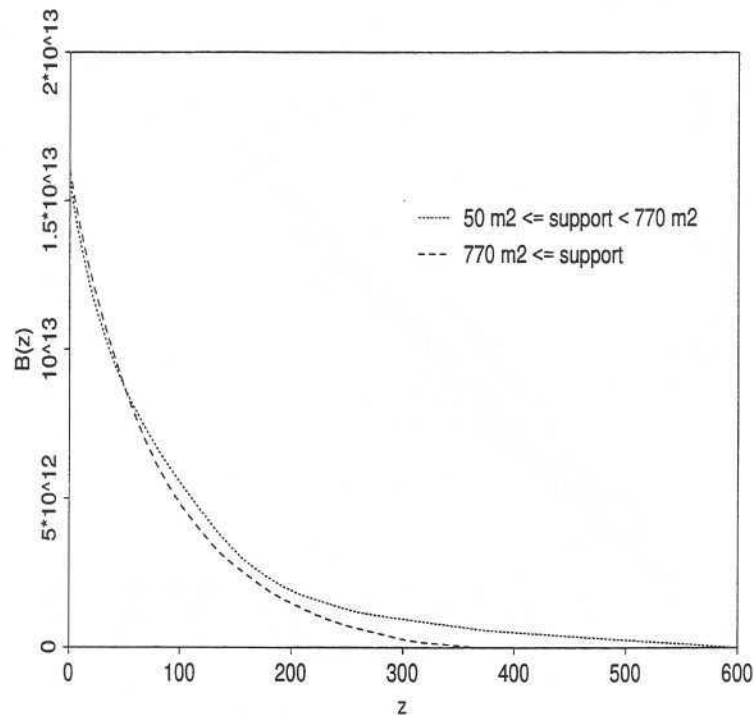


FIG. 2.15 - Effet régularisant d'une augmentation du support. Bénéfice conventionnel des densités (z) pour les deux dernières classes de support (chalutages obliques).

Les conclusions de cette analyse sont les suivantes :

- **Les données espagnoles**, récoltées par chalutage vertical, sont beaucoup plus dissymétriques (sélectives) que celles obtenues par chalutage oblique. **Elles seront donc retirées des analyses.**

- La **différence de sélectivité** observée pour des densités prélevées par **chalutage oblique est négligeable** bien que les supports varient du simple au double. Il semble donc qu'il n'y ait pas d'inconvénient majeur à mélanger l'ensemble des données ayant été récoltées

par chalutage oblique.

- Le cumul des densités le long de la verticale permet d'exprimer toutes les données selon une même unité (nombre total d'œufs se trouvant dans une tranche d'eau) mais **ne permet pas de les standardiser**.

2.4.3 Support moyen par période

Les informations nécessaires au calcul du support des densités pour l'ensemble des périodes échantillonnées, ne sont pas disponibles. Toutefois, lorsque la comparaison est possible (1992 et 1995) il existe une très bonne corrélation entre le volume filtré (en m^3) et le support des densités à 2D (Fig. 2.16).

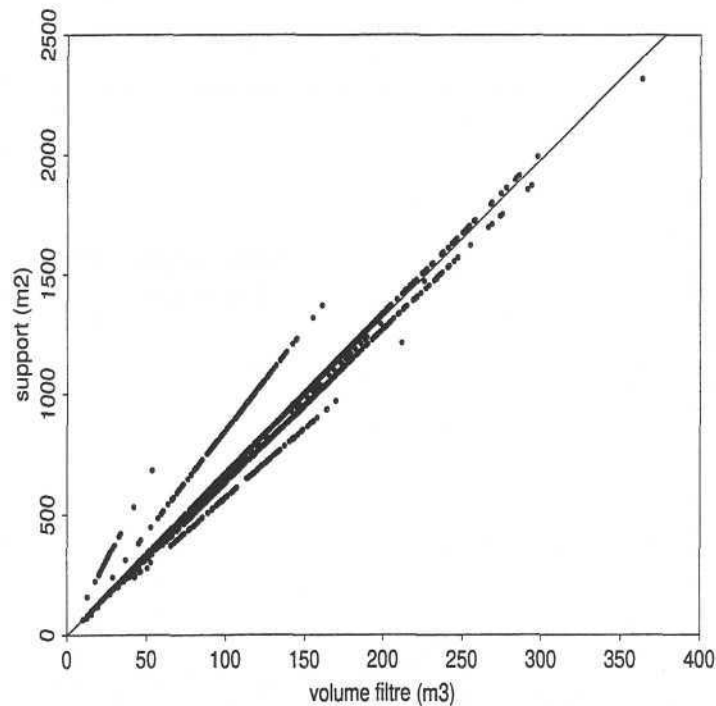


FIG. 2.16 - Corrélation entre le support des densités 2D (en m^2) et le volume filtré (en m^3). Représentation de l'ajustement linéaire par moindres carrés ($support=6.6 \cdot volume$).

Il est donc possible de se servir du volume filtré, connu lui pour l'ensemble des données, pour estimer le support et suivre ainsi l'évolution de la valeur moyenne du support en fonction de la période d'échantillonnage (Fig. 2.17). Globalement le support diminue entre le mois de Mars et la fin de la saison de frai en Juin. Des fluctuations sont observables autour de cette tendance, comme à la fin Mars 1983 où le support semble, en moyenne, plus faible que les autres années (nous reviendrons sur ce point au chapitre suivant).

La diminution du support dans un rapport de 1 à 2 au cours d'une saison de ponte s'explique probablement en partie par la diminution concomitante de la profondeur échantillonnée. En dehors de la limitation naturelle par la bathymétrie, la profondeur échantillonnée est en effet définie par la profondeur de la thermocline quand elle existe, ou bien, par une profondeur maximale de 200

mètres quand la thermocline n'est pas suffisamment nette. La stratification de l'eau, et donc la présence d'une thermocline, se mettant en place pendant le printemps, il est donc logique que la profondeur diminue (Fig. 2.18) et, par conséquent, le support.

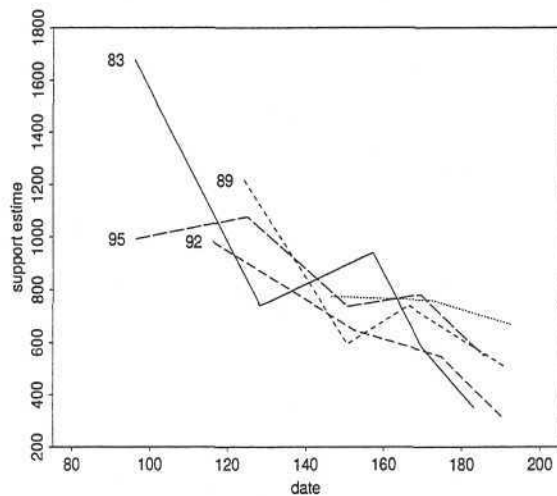


FIG. 2.17 - Estimation du support par le volume filtré. Evolution de la valeur moyenne en fonction des périodes d'échantillonnage de 1983 à 1995.

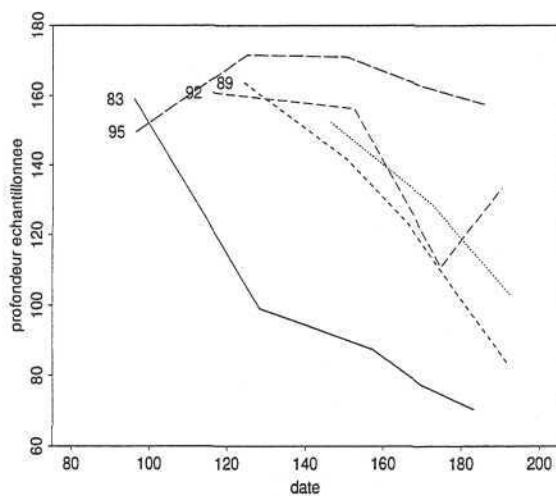


FIG. 2.18 - Profondeur d'échantillonnage. Evolution de la valeur moyenne en fonction des périodes d'échantillonnage de 1983 à 1995.

Chapitre 3

Analyse de la ponte de 1983 à 1995

Sommaire

Dans ce chapitre, les statistiques individuelles permettent de décrire et de comparer au cours du temps, le déroulement de cinq saisons de ponte. Seules les densités d'œufs de premier stade (E1) sont donc utilisées.

*Le **paragraphe 1** ébauche un modèle pour décrire l'évolution de l'abondance en œufs au cours d'une saison, basé sur les valeurs respectives de la densité moyenne et de la surface équivalente.*

*Le **paragraphe 2** rend compte de la position moyenne et de la dispersion des œufs d'après les échantillonnages. Le caractère systématique du déplacement de la population est interprété en termes de flux migratoire.*

*Le **paragraphe 3** poursuit l'analyse par la comparaison des structures spatiales en distinguant les valeurs prises par les covariogrammes expérimentaux à l'origine et pour des distances non nulles.*

*Le **paragraphe 4** s'intéresse à la température de surface moyenne à laquelle se déroule la ponte. Quelques conclusions sur le rôle conditionnant de la température sont avancées sur la base de l'analyse des inertigrammes.*

*Le **paragraphe 5** complète la description des conditions environnementales aux quelles se trouvent les populations d'œufs par l'étude de deux transformées de la profondeur (log-translation et laplacien).*

*Les résultats les plus importants sont finalement synthétisés dans le **paragraphe 6**.*

Analysis of the spawning from 1983 to 1995

Abstract

This chapter concerns the description and the comparison through time of 5 spawning seasons (from 1983 to 1995). As we are dealing with the spawning, only the stage 1 eggs daily productions are being used.

The paragraph 1 gives elements for the definition of a model describing the spawning process. These elements are based on the comparative analysis of the mean density per individual and the equivalent surface.

The paragraph 2 deals with the mean location and the inertia of the egg populations. The systematic movement of the population during each spawning season is interpreted in terms of migration routes.

The paragraph 3 completes the analysis by comparing the egg spatial structures. In particular, we distinguish from the experimental covariograms the value at the origin and the rest of the estimated structure at non zero distances.

The paragraph 4 deals with the mean sea surface temperature at which the spawning occurs. Some preliminary conclusions about the conditioning effect of the temperature on the spawning process are put forward on the basis of the analysis of the inertiograms.

In the paragraph 5 we add some informations about the mean bathymetry at which the eggs are found. This mainly concerns transformations of bathymetry (log-transformation and laplacian) rather than the bathymetry itself.

Finally, the main results are recalled in the paragraph 6.

Dans l'analyse comparative des 5 années d'observations de la ponte dont je dispose, j'ai cherché, en utilisant des statistiques individuelles simples comme la surface équivalente ou le centre de gravité de population, à dégager certaines des caractéristiques persistantes du processus de reproduction. En somme, pour reprendre ce qui a déjà été dit à propos de la recherche de concepts pertinents en statistique appliquée, nous avons cherché à faire parler les données.

3.1 Densité par individu et modèle possible pour la reproduction

3.1.1 Abondance, densité moyenne et surface équivalente

Courbes de production

Les courbes de production qui indiquent l'intensité de la ponte en nombre d'œufs produits par jour en fonction de la date sont, assez classiquement, en **cloche**. Le **pic** apparaît quasi-systématiquement **fin Mai - début Juin**, c'est-à-dire autour du 150ème jour de l'année (Fig. 3.1a). A cette date là, la production fluctue entre 1.5 et 2.5 10^{13} œufs/jour selon les années. Outre le pic de ponte, les début et fin de reproduction se manifestent également toujours approximativement à la même date.

Au-delà de ces éléments reproductibles d'année en année, des différences sont observées. On notera en particulier que la reproduction en 1986 semble avoir été très faible et que le pic de ponte en 1992 a été plus étalé dans le temps, sans pour autant baisser d'intensité (Fig. 3.1a). L'ensemble de ces résultats est décrit et commenté dans les rapports CIEM du groupe de travail "Mackerel and Horse Mackerel Egg Production Workshop" (Anon., 1987; Anon., 1990; Anon., 1993; Anon., 1996).

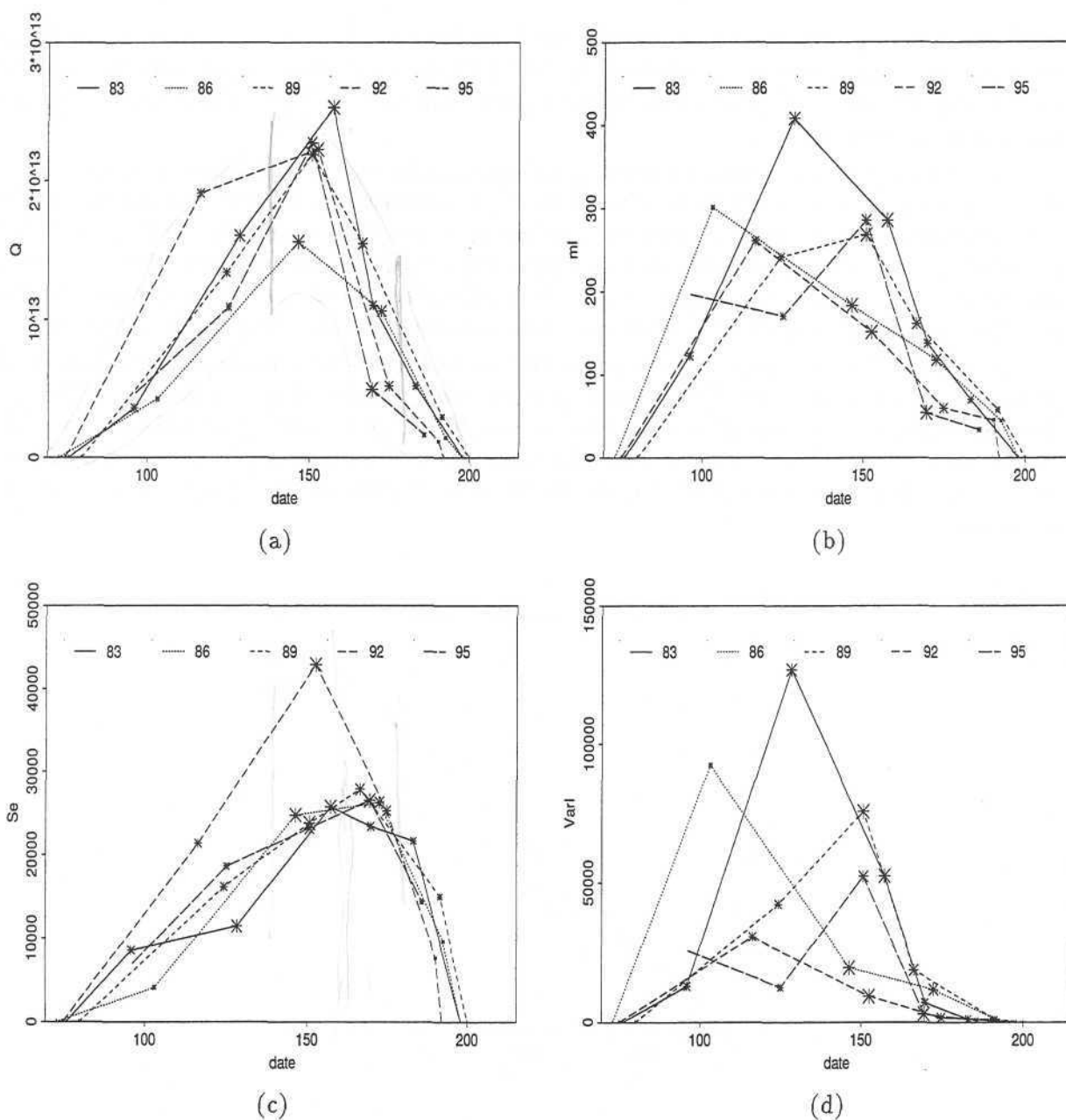


FIG. 3.1 - Statistiques par individu (E1) en fonction de la date moyenne des périodes d'échantillonnage exprimée en jour julien. Les étoiles ont une taille proportionnelle aux nombres d'échantillons utilisés pour les calculs.

(a) : courbes de production donnant le nombre d'œufs de premier stade produit par jour.

(b) : densité moyenne par individu exprimée en nombre œufs/m²/jour.

(c) : surface équivalente (en milles nautiques au carré) ou inverse de l'indice d'agrégation.

(d) : variance de la densité par individu.

Densités moyennes (et variances) par individu.

Chaque année, l'évolution de la densité moyenne par individu suit une courbe en cloche plus ou moins dissymétrique selon les années (Fig. 3.1b). En milieu de saison (jours 120-150), elle varie entre 200 et 400 œufs/m²/jour. Cette dernière valeur, observée en 1983, dépasse de beaucoup les autres densités moyennes.

Au chapitre précédent, nous avons vu qu'autour du 120ème jour de l'année, le support moyen était minimum en 1983. Comme la variabilité, et par construction la densité moyenne par individu, augmente quand le support diminue, une partie du décalage observé en 1983 sur les densités moyennes est probablement imputable à un effet de support. Il convient toutefois d'être prudent, car un effet inverse aurait du être observé fin Mai (jour 150); le support étant, à cette date là, maximum en 1983. Or la densité moyenne par individu reste supérieure aux autres densités moyennes.

Les années suivantes (1986 à 1995), le **maximum de densité moyenne par individu oscille peu** entre 250 et 300 œufs/m²/jour mais **sa date d'apparition dans la saison dépend de l'année**. Ainsi, le maximum de la densité moyenne est observé mi Avril (jour 100) en 1986, début Mai en 1992, et fin Mai en 1989 et 1995. Ce classement des années par la date d'apparition du maximum de densité moyenne par individu sera mis en correspondance plus loin avec la température de surface.

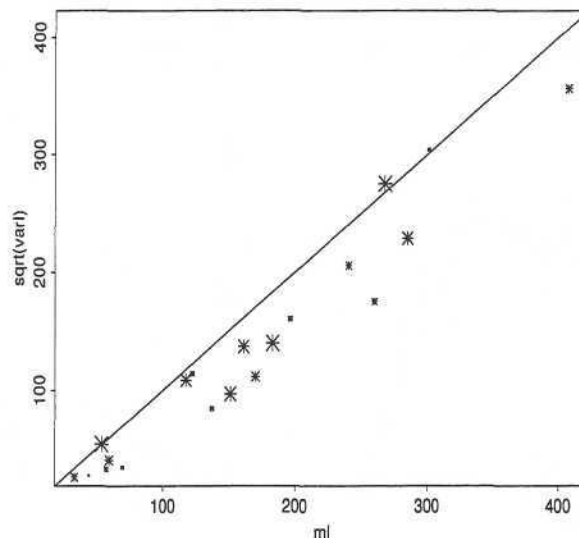


FIG. 3.2 - Nuage de corrélation entre l'écart-type et la moyenne des densités par individu obtenus pour chaque période d'échantillonnage. Maquereau, œufs de premier stade (E1). Position par rapport à la première bissectrice. Les étoiles ont une taille proportionnelle aux nombres d'échantillons utilisés pour les calculs.

La variance par individu (Fig. 3.1d) ne fait qu'accentuer les différences inter-annuelles de densités moyennes: évolution en cloche, date d'apparition des maxima différentes selon les années. Cela provient du fait que l'écart-type des densités par individu augmente proportionnellement à la densité moyenne (Fig. 3.2).

Surfaces équivalentes

L'évolution en cloche, ainsi que la valeur de la surface équivalente au cours d'une saison de ponte sont très persistantes d'année en année, à l'exception de 1992 où la surface équivalente est approximativement double des autres années (Fig. 3.1c). En outre, exceptée 1992, la surface équivalente augmente jusqu'à la troisième semaine de Juin (soit 2 semaines après le pic d'abondance) où elle atteint 25 000 mn² et diminue brutalement ensuite.

A un moment donné de la saison de reproduction, la surface équivalente (ou encore le niveau d'agrégation des œufs) est par conséquent peu sensible aux variations inter-annuelles d'abondance en œufs. Bien que le sens physique de la surface équivalente ne soit pas facile à saisir, il semble donc que cette statistique soit, par sa stabilité inter-annuelle, caractéristique du processus de reproduction du maquereau.

Par ailleurs, dans la logique de ce qui a été dit précédemment, on peut penser qu'une partie du léger décalage observé début Mai 1983 (jour 125) provient d'un support d'échantillonnage en moyenne plus faible ($g(0)/Q^2 = 1/S_e$ augmente et donc S_e diminue quand le support diminue).

Synthèse et modèle possible pour la reproduction

Par définition, l'abondance, à un instant donné de l'année, est le produit de la densité moyenne par la surface équivalente ($Q = S_e \cdot \bar{z}_I$) et la production annuelle d'œufs est fournie par la somme de ces abondances sur l'ensemble de la période de reproduction (intégrale des courbes de productions). Or à part en 1992, les courbes d'évolution des surfaces équivalentes sont apparues peu variables d'année en année. **Les variations inter-annuelles de quantités totales d'œufs sont donc essentiellement liées à des modifications synchrones dans l'évolution des densités moyennes.**

Que savons-nous des fluctuations (inter-annuelles) des courbes de densités moyennes? Nous avons vu que les maxima de ces courbes en cloche oscillent peu autour d'un niveau moyen compris entre 250 et 300 œufs/m²/jour (sauf en 1983) mais que leurs dates d'apparition sont par contre fortement variables d'année en année.

On peut donc postuler que l'abondance en œufs est le résultat de la combinaison de deux processus (en cloche) indépendants. Le premier processus ($S_e(j)$), reproductible d'année en année (sauf exception), concerne la surface équivalente. Il fixe la date d'apparition et la hauteur du maximum de surface équivalente à 25 000 mn² au 175ème jour de l'année. Le deuxième processus ($\bar{z}_I(j, a)$) décrit l'évolution de la densité moyenne par individu. Compte-tenu des résultats précédents, il semble raisonnable de fixer le niveau maximum de densité moyenne (à 280 œufs/m²/jour par exemple) mais de laisser la date de réalisation de ce maximum et la forme de la fonction varier en fonction de l'année. Dans un tel modèle :

$$Q(a) = \sum_j S_e(j) \cdot \bar{z}_I(j, a, \beta)$$

où a représente l'année, j le jour de l'année, et β le paramètre de forme (Fig. 3.3), **l'abondance totale d'œufs s'explique alors par la plus ou moins bonne concomitance entre les pics de surface équivalente et de densité moyenne** et par la plus ou moins forte dissymétrie de la courbe de densité moyenne. En particulier la reproduction du maquereau est d'autant moins bonne que le pic de densité moyenne apparaît tôt dans la saison (i.e. loin du 175ème jour).

Ce modèle simple qui s'apparente en fait au principe connu de "match - mismatch", ne prend pas en compte le fait que, certaines années (comme en 1992 par exemple), la surface équivalente

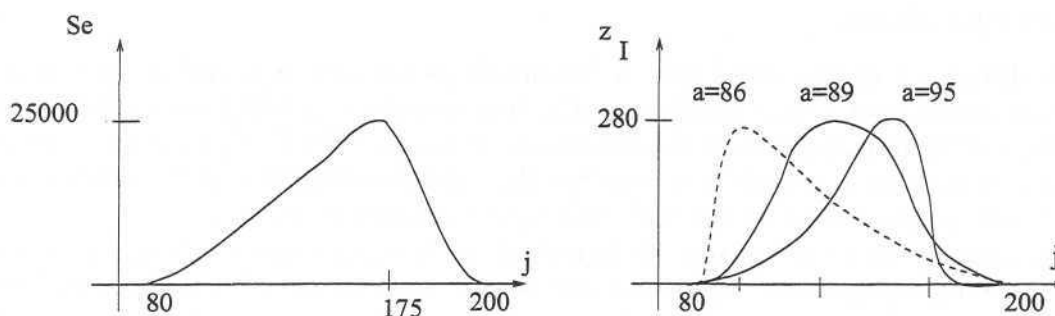


FIG. 3.3 - Modèle simple de reproduction par le produit de deux processus. Un processus d'évolution de la surface équivalente (S_e) persistant d'année en année, et un processus fixant la date d'apparition du maximum de densité moyenne par individu (z_I).

peut être très supérieure à ce qui a été fixé. Mais il fournit cependant des éléments d'interprétations. En particulier, si la surface équivalente de la population d'œufs n'avait pas été exceptionnellement élevée en 1992, la reproduction se serait probablement située à un niveau aussi faible qu'en 1986, puisque le pic de densité moyenne par individu est apparu presque aussi tôt en 1986 et en 1992. En somme, à supposer que l'agrégation des œufs de premier stade nous donne celle des reproducteurs, on peut penser que le changement de comportement agrégatif des géniteurs dans le sens d'une diminution du niveau d'agrégation, a permis d'atteindre un niveau raisonnable de reproduction malgré l'occurrence assez précoce cette année là du pic de densité moyenne.

3.2 Localisation moyenne de la ponte

3.2.1 Persistance des déplacements et flux migratoires

Les déplacements du **centre de gravité des œufs** (E_1) au fil des périodes d'échantillonnage (Fig. 3.4), présentent les mêmes tendances quelle que soit l'année. En début de saison, le centre de gravité de la ponte se situe entre 48 et 49°00N. C'est à ce moment là qu'il occupe sa position la plus méridionale. Puis la ponte se déplace vers le Nord-Ouest, où la production est d'ailleurs maximale (taille des cercles centrés sur les numéros des périodes). Ensuite, la localisation moyenne des œufs se déplace vers le Sud-Est et, par là, vers l'intérieur du plateau continental. En 1992 toutefois, le centre de gravité de la dernière période d'échantillonnage (la période 4) échappe à ce principe.

La latitude du **centre de gravité des échantillons** est en général voisine de 50° (Fig. 3.5), et les modifications d'échantillonnage de période en période se traduisent par des mouvements de faibles amplitudes et essentiellement Nord-Sud. C'est donc que, globalement, les déplacements du centre de gravité de la ponte ne sont pas dus à des modifications synchrones du plan d'échantillonnage.

On notera cependant une exception concernant la quatrième période d'échantillonnage de 1992 où les centres de gravité de la ponte et de l'échantillonnage se trouvent très au Nord. Ainsi, la position "anormale" du centre de gravité de la période 4 de 1992 relativement aux 5 années d'observation peut être expliquée par la position "Nord" des échantillons plus que par celle des œufs.

Le déplacement du centre de gravité des œufs de maquereau semble donc reproductible d'année en année (persistant). Une interprétation possible de ce résultat a trait au **schéma de migration** du maquereau lors de la reproduction. Sur la base des distributions en

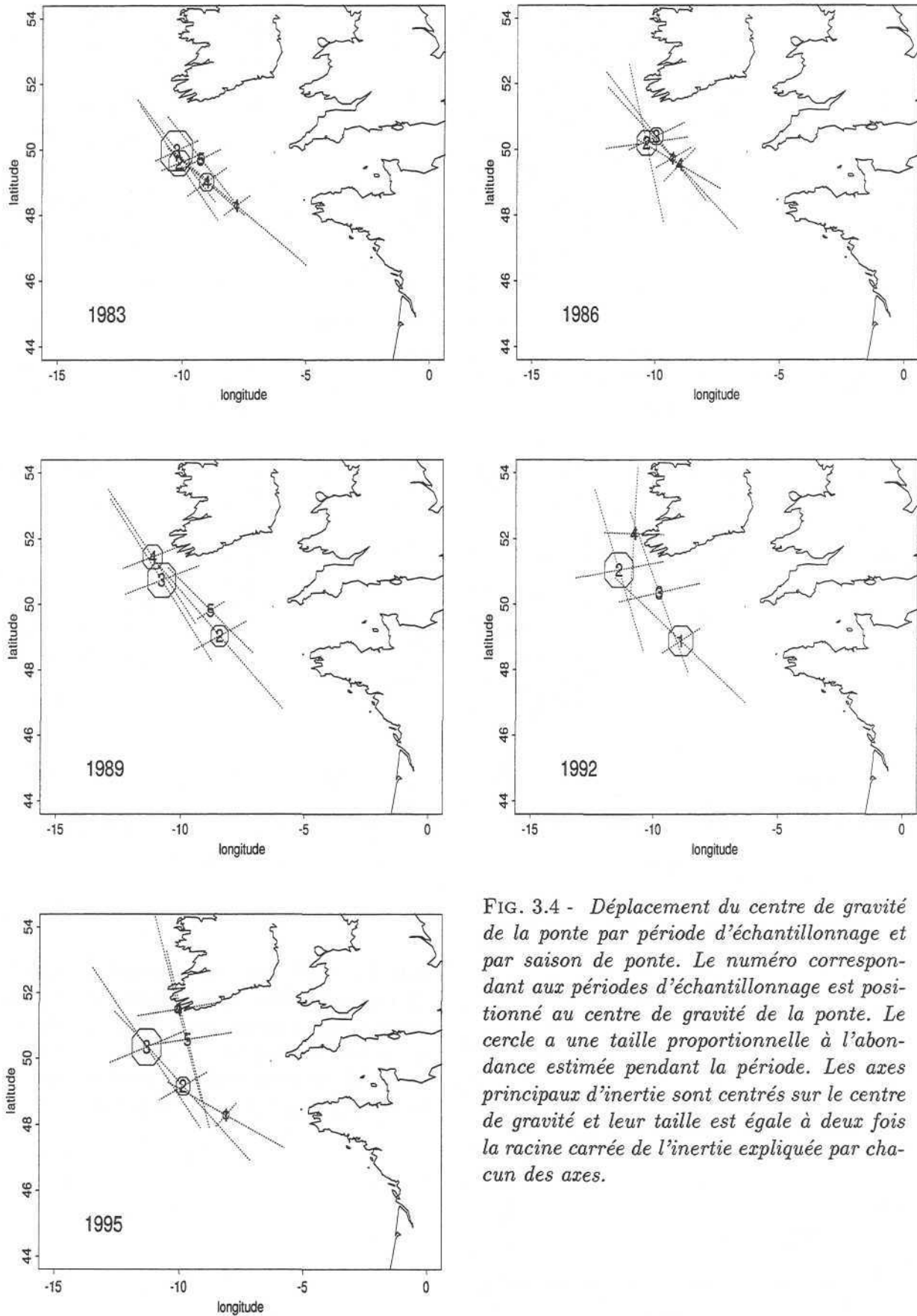


FIG. 3.4 - Déplacement du centre de gravité de la ponte par période d'échantillonnage et par saison de ponte. Le numéro correspondant aux périodes d'échantillonnage est positionné au centre de gravité de la ponte. Le cercle a une taille proportionnelle à l'abondance estimée pendant la période. Les axes principaux d'inertie sont centrés sur le centre de gravité et leur taille est égale à deux fois la racine carrée de l'inertie expliquée par chacun des axes.

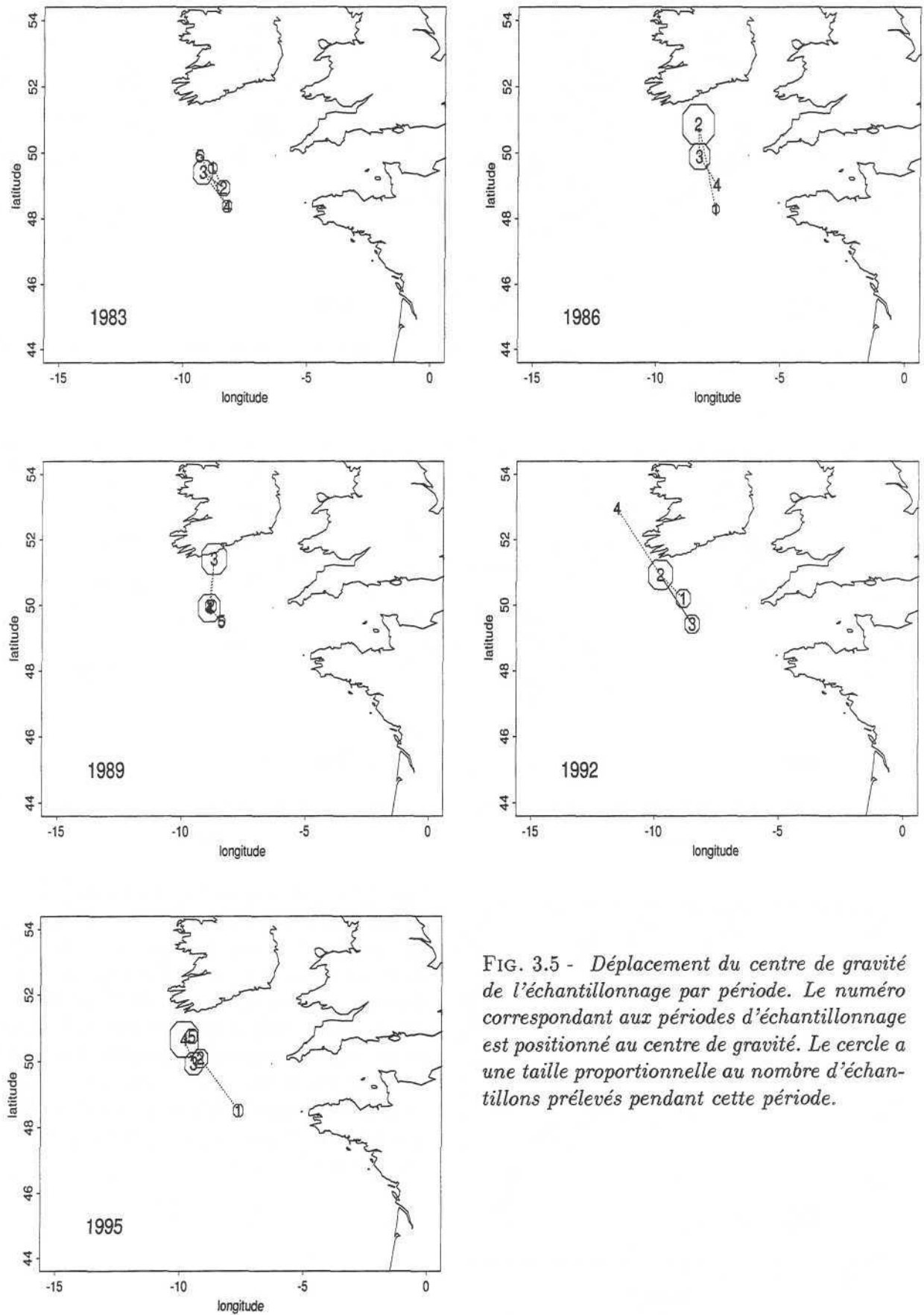


FIG. 3.5 - Déplacement du centre de gravité de l'échantillonnage par période. Le numéro correspondant aux périodes d'échantillonnage est positionné au centre de gravité. Le cercle a une taille proportionnelle au nombre d'échantillons prélevés pendant cette période.

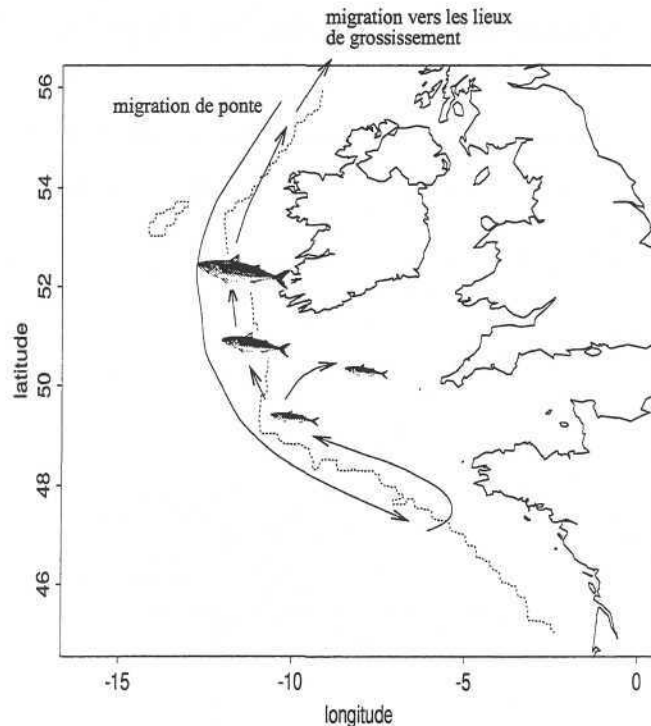


FIG. 3.6 - Schéma de migration de reproduction du maquereau. Passage successif sur la zone de frai de maquereaux de plus en plus jeunes.

taille des maquereaux capturés pendant les campagnes triennales, il apparaît que la taille moyenne des individus présents sur les lieux de ponte diminue au cours de la saison (Dawson, 1986; Eltink, 1987). Le stock de géniteurs, qui arrive du Nord de l'Écosse, séjourne et passe donc sur l'aire de reproduction par ordre d'âge décroissant (Fig. 3.6); les poissons les plus grands étant les premiers sur les lieux de ponte, les premiers à pondre et les premiers à entamer la migration "retour" vers les lieux de grossissement. Selon le principe qui veut que les jeunes individus de certaines espèces calquent leur comportement sur celui de leurs aînés (Corten, 1983), il devient probablement de plus en plus difficile pour les plus jeunes maquereaux de trouver à leur tour la route de migration vers le Nord, puisqu'en fin de saison de reproduction les individus les plus grands sont déjà partis. Ainsi une partie des plus jeunes maquereaux, adoptent probablement un autre comportement et s'orientent vers l'intérieur de la Mer Celtique (Fig. 3.6) ce qui étaye des idées déjà proposées (Lockwood *et al.*, 1981).

3.2.2 Inertie

L'inertie de la distribution des œufs au cours de chaque période d'échantillonnage, décomposée en ses axes principaux (Fig. 3.4), est principalement due aux directions **parallèles aux accores**. En effet, tant que le centre de gravité de la population est voisin de 50° de latitude, l'inertie totale est expliquée pour une grande partie par une direction Nord-Ouest qui correspond à l'orientation des accores dans cette région. Par ailleurs, lorsque la population se trouve en moyenne plus au Nord, dans un secteur où les accores, représentés par l'isobathe des 200 m, sont orientés

Nord-Sud, le premier axe d'inertie prend une direction Nord-Nord-Ouest.

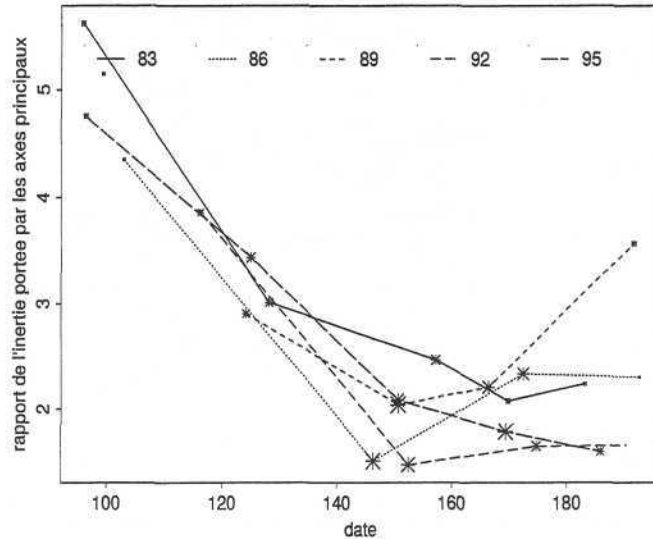


FIG. 3.7 - Isotropisation de l'inertie de la population d'œufs de maquereau ($E1$) au cours du temps. Rapport entre l'inertie expliquée par le premier et le deuxième axe principal.

De plus, au fur et à mesure que la saison avance, l'inertie portée par chacun des deux axes principaux est de plus en plus similaire (Fig. 3.7). Cette tendance est en tous les cas très nette sur la première moitié de la saison de ponte. L'inertie de la distribution des œufs de premier stade semble donc de plus en plus isotrope entre Avril et Juin.

3.2.3 Déplacement vers l'Ouest

La forme en croissant de la distribution des œufs (voir cartes proportionnelles dans l'annexe 2) complique l'interprétation des centres de gravité puisqu'ils peuvent se trouver à l'intérieur du plateau continental alors que les individus se trouvent le long des accores. Par ailleurs, aux périodes de fortes productions d'œufs, la position moyenne de la population a tendance à se déplacer vers l'Ouest entre 1983 et 1995 (Fig. 3.4). Pour confirmer cette tendance, j'ai transformé les coordonnées des points de façon à repérer les échantillons par leur latitude (inchangée) et leur distance par rapport à l'isobathe des 200 mètres (accors). Ainsi, on imagine que les œufs se distribuent le long d'un axe représentant les accors.

Dans ce nouveau repère, où une abscisse positive indique qu'on est sur le plateau continental, l'évolution inter-annuelle des centres de gravité correspondant à la période fin Mai - début Juin (pic de ponte) entre 1983 et 1995 montre très nettement un décalage vers l'Ouest de 2 degrés environ, soit 130 milles nautiques (Fig. 3.8). Dans le même temps, les modifications inter-annuelles d'échantillonnage n'ont engendré qu'un léger déplacement du centre de gravité des échantillons (Fig. 3.9). **Le déplacement de la ponte est donc un déplacement net vers l'Ouest, d'environ 100 milles nautiques.** On notera au passage que l'interprétation de la position moyenne des œufs pendant le début du mois de Juillet est plus délicate. L'évolution des centres de gravité des œufs et des échantillons sont en effet très voisines.

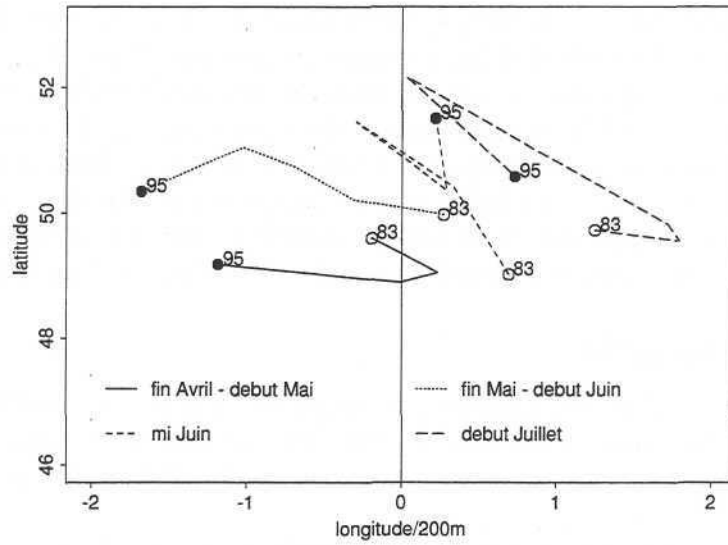


FIG. 3.8 - Position du centre de gravité de la ponte par rapport à l'isobathe des 200 mètres (accres). Evolution inter-annuelle pour différents moments de la saison de ponte (1983 = cercle blanc ; 1995 = cercle noir). Les abscisses positives indiquent un déplacement vers la côte; les négatives, vers la large.

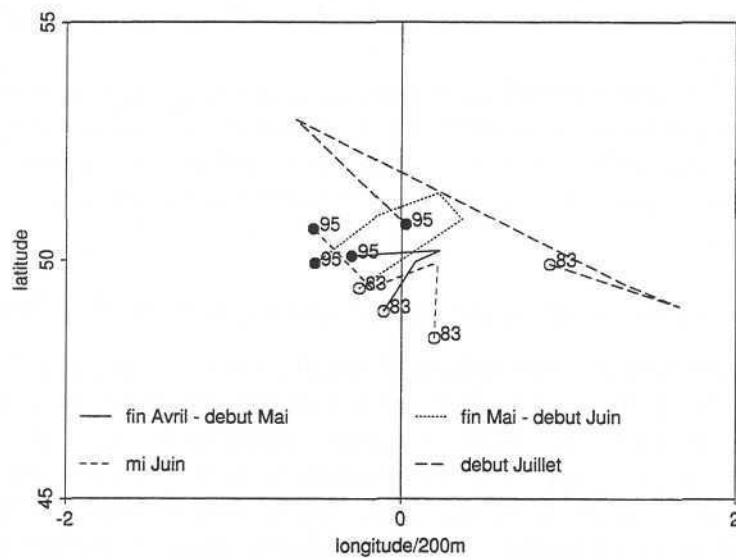


FIG. 3.9 - Position du centre de gravité des échantillons par rapport à l'isobathe des 200 mètres (accres). Evolution inter-annuelle pour différents moments de la saison de ponte (1983 = cercle blanc ; 1995 = cercle noir). Les abscisses positives indiquent un déplacement vers la côte; les négatives, vers le large.

3.3 Structures spatiales

Pour chaque période, le covariogramme expérimental de la distribution des œufs (E1) a été calculé dans quatre directions. Une fois les distances divisées par 20, les structures expérimentales sont représentées le long d'un axe temporel indiquant la date moyenne de la période d'échantillonnage. Cela permet de comparer les structures spatiales à des moments identiques dans la saison de reproduction. Les covariogrammes sont pris ici au sens de la densité de probabilité de la distance entre deux individus aléatoires, c'est-à-dire qu'on représente $g(h)/Q^2$ (Fig. 3.10). Les structures obtenues pour les dernières périodes des années 1983, 1989 et 1995 sont difficilement interprétables compte tenu du très faible nombre d'échantillons prélevés pendant ces campagnes.

3.3.1 Niveau d'agrégation

Selon notre définition, le niveau d'agrégation, donné par la valeur du covariogramme à l'origine, est l'inverse de la surface équivalente. Il n'est donc pas surprenant que l'agrégation des œufs de stade 1 (E1) diminue en début de saison et augmente après le 150^{ème} jour de l'année.

Le reste de la structure spatiale présente deux caractéristiques plus ou moins stables avec le temps : la direction d'anisotropie et l'ordre de grandeur du covariogramme en dehors de l'origine.

3.3.2 Anisotropie

Quelle que soit l'année ou la période de prélèvement, la direction qui présente le **plus de continuité spatiale** est la direction 121° (Nord-Nord-Ouest). En milieu de saison, la portée dans cette direction est en général voisine de 600 milles nautiques. Par ailleurs, le plus souvent la structure spatiale dans les directions 59° et 0° est assez similaire, présentant une portée plus courte de l'ordre de 200 milles nautiques.

La décomposition de l'inertie en deux axes principaux a permis, plus haut, de mettre en évidence une tendance à l'isotropisation de l'inertie de la population d'œufs au cours de chaque saison de ponte. Cependant, rien de tel n'est réellement visible sur les covariogrammes. Les raisons qui permettraient d'expliquer que l'inertie devienne isotrope mais pas les structures spatiales ne sont pas claires à mes yeux.

3.3.3 Stabilité du niveau du covariogramme en dehors de l'origine

La figure 3.11 représente les covariogrammes expérimentaux sans relier les valeurs à l'origine aux premiers pas de calcul. Il apparaît ainsi que le covariogramme des œufs relativement au carré de leur abondance prend à peu près toujours la même valeur quelle que soit l'année ou la période. En effet, passée la chute à l'origine, le covariogramme se situe le plus souvent autour de 0.00002 nm^{-2} pour des distances voisines de 30 nm. Ainsi, relativement au carré de l'abondance, **l'essentiel du changement dans les structures spatiales concerne la valeur du covariogramme à l'origine**. Il est alors tentant d'attribuer au covariogramme relatif obtenu pour des distances supérieures à 30 mn une valeur intrinsèque spécifique, c'est-à-dire propre à l'espèce et donc invariante dans le temps. On pourrait alors enrichir le modèle de reproduction, construit au paragraphe précédent, par la définition d'une structure spatiale composée :

- d'une structure persistante de portée anisotrope (600 mn dans la direction 121 et 200 mn dans la direction orthogonale) et de palier invariant dans le temps. Cette structure s'apparente en fait au covariogramme obtenu pour la période 2 de 1992 puisque celui-ci ne présente pas de discontinuité

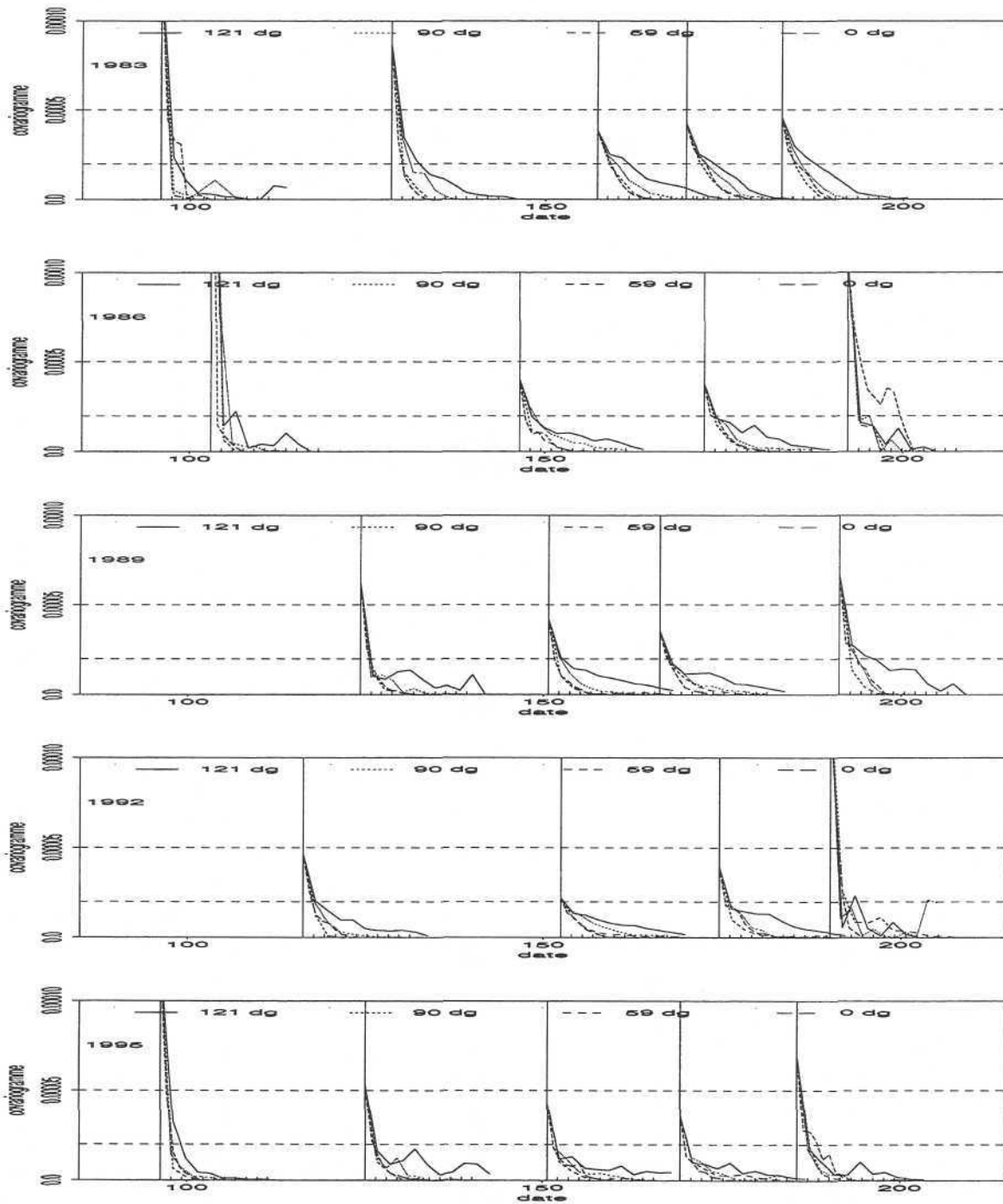


FIG. 3.10 - Covariogrammes expérimentaux relativement au carré de l'abondance par période d'échantillonnage entre 1983 (en haut) et 1995 (en bas) exprimés en mn^{-2} . Œufs de stade 1. Quatre directions de calcul (121, 90, 59 et 0). Une graduation en abscisse représente 30 milles nautiques. Les niveaux 0.00002 et 0.00005 sont indiqués par des lignes horizontales.

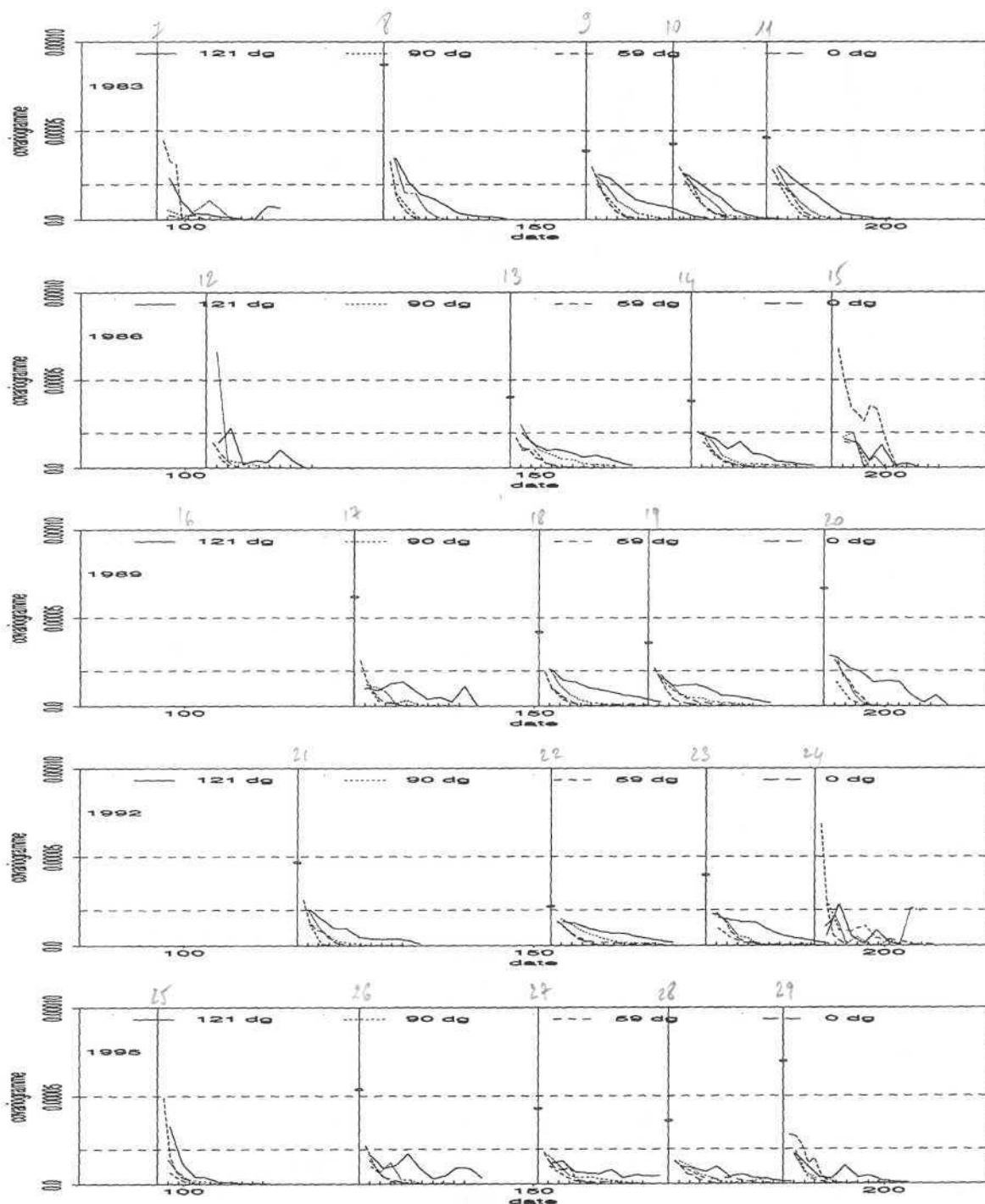


FIG. 3.11 - Covariogrammes expérimentaux relativement au carré de l'abondance par période d'échantillonnage entre 1983 (en haut) et 1995 (en bas) exprimés en mn^{-2} . La valeur à l'origine n'est pas reliée graphiquement à la valeur du covariogramme aux premiers pas de calcul. Œufs de stade 1. Quatre directions de calcul (121, 90, 59 et 0). Une graduation en abscisse représente 30 milles nautiques. Les niveaux 0.00002 et 0.00005 sont indiqués par des lignes horizontales.

à l'origine. La valeur en 0 de ce covariogramme (0.000023) permet de fixer la hauteur du palier de la structure persistante.

- et d'une microstructure modélisée par un effet de pépite $C(j)$ évoluant selon une courbe en cloche au cours de chaque saison de reproduction ($0.000023 + C(j) = 1/S_e(j)$):

$$\frac{g(h)}{Q^2} = C(j) + 0.000023 \cdot g_1\left(\frac{h}{600}\right)$$

Dans ce modèle, la structure spatiale des œufs observée en 1992 durant la deuxième période d'échantillonnage fait figure de situation extrême pour l'espèce. En particulier, la surface équivalente observée à ce moment là, soit 43 000 nm², s'interprète comme la surface équivalente maximum que la population peut atteindre. Toute autre valeur (nettement) supérieure serait en effet associée à une diminution du covariogramme puisque ce dernier doit être, par définition, maximum à l'origine ; ce qui serait incompatible avec l'hypothèse formulée précédemment quant à l'invariance de la partie du covariogramme pour des distances non nulles.

Ce modèle permet également de préciser en quoi la distribution spatiale des œufs a été particulière en 1992: seule l'agrégation a été anormalement petite; le reste de la structure spatiale observée cette année là fut normale, c'est-à-dire conforme à celles observées les autres années.

Afin de lever toute ambiguïté sur la valeur de ce modèle, précisons qu'il n'a vocation qu'à formaliser un certain nombre de tendances qui se dégagent de l'analyse de 5 saisons de reproduction.

3.4 Transformées de la profondeur

3.4.1 Log-translation

La bathymétrie proprement dite présente des variations très fortes et très rapides dans la région de ponte et les résultats obtenus sont très instables. Pour analyser la profondeur moyenne à laquelle se déroule la ponte, j'ai préféré utiliser la transformée logarithmique simple suivante :

$$\text{profondeur transformée} = \log_{10}\left(\frac{\text{profondeur}}{20}\right)$$

de sorte que les profondeurs transformées:

- mesurées sur le plateau continental sont comprises entre 0 et 1
- entre les accores (200 m) et 2000 m sont comprises entre 1 et 2
- supérieures à 2000 m sont transformées en valeurs supérieures à 2.

Le centre de gravité de cette transformée est supérieur à 1.4 (500 m) de début Avril à fin Mai quelle que soit l'année (Fig. 3.12 a). Cela indique que **les œufs de stade 1 se trouvent, en moyenne, plus au large que l'isobathe des 200 m**. Son évolution au cours du temps (Fig. 3.12a) traduit ce que nous avons déjà observé avec le centre de gravité de la distribution des œufs: la profondeur moyenne des œufs augmente pendant la première moitié de la saison et diminue ensuite. Cela correspond au déplacement vers le large de la ponte entre Avril et Mai et à son retour vers le plateau continental après le pic de ponte.

On mesure également la tendance à aller de plus en plus vers le large entre 1983 et 1995 puisque le centre de gravité calculé à la fin Mai passe de 1.4 (500 m) en 1986 à 1.7 (1000 m) en 1995.

L'inertie de cette transformée (Fig. 3.12b) semble, par contre, moins variable d'année en année que le centre de gravité.

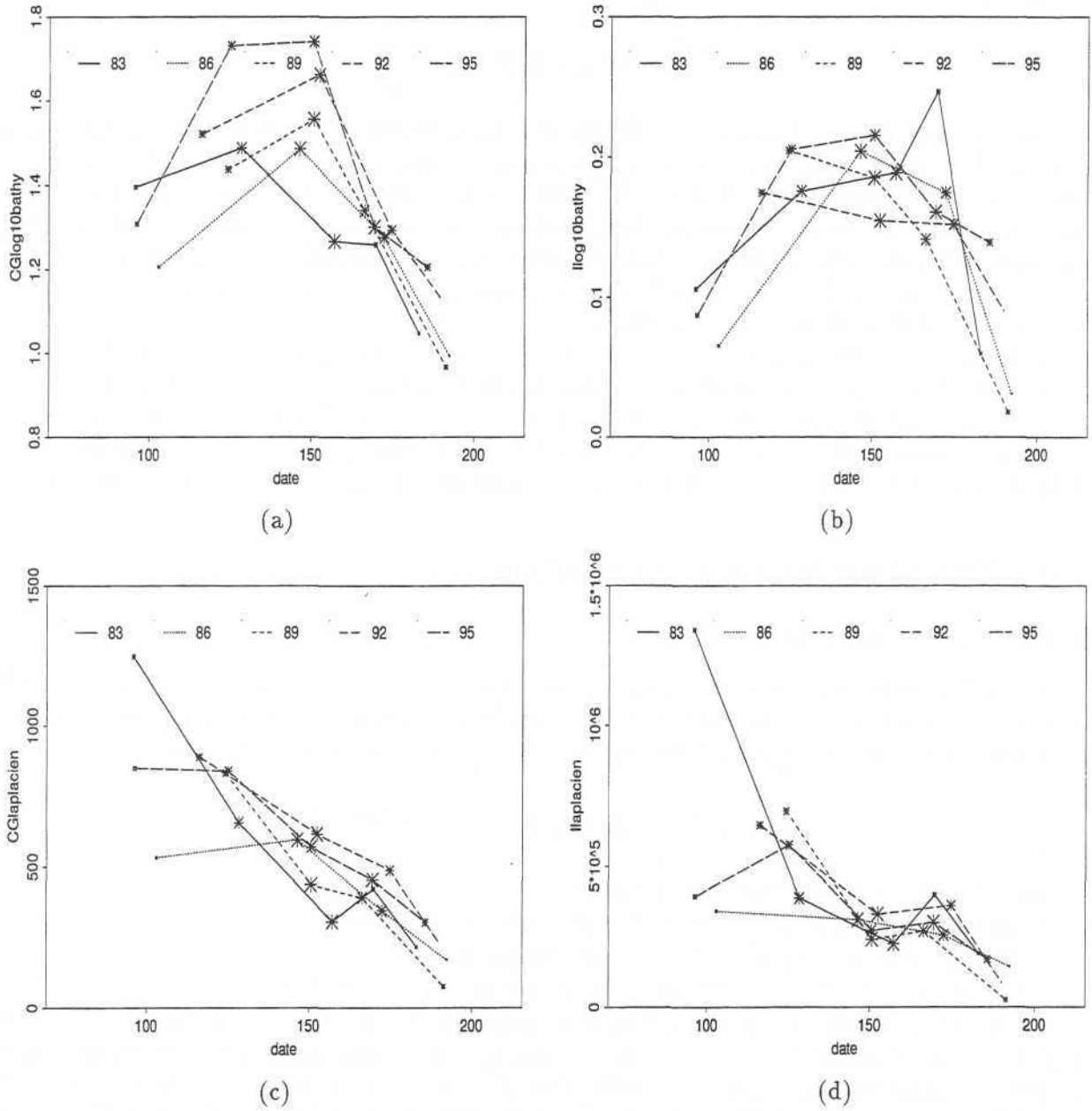


FIG. 3.12 - Centre de gravité (a,c) et inertie (b,d) d'une transformée de la profondeur par individu (œufs de stade 1 - E1) en fonction du temps exprimé en jour julien. Les étoiles ont une taille proportionnelle aux nombres d'échantillons utilisés pour les calculs.

a et b: $\log_{10}\left(\frac{\text{profondeur}}{20}\right)$

c et d: $|\text{laplacien}(\text{profondeur})|$

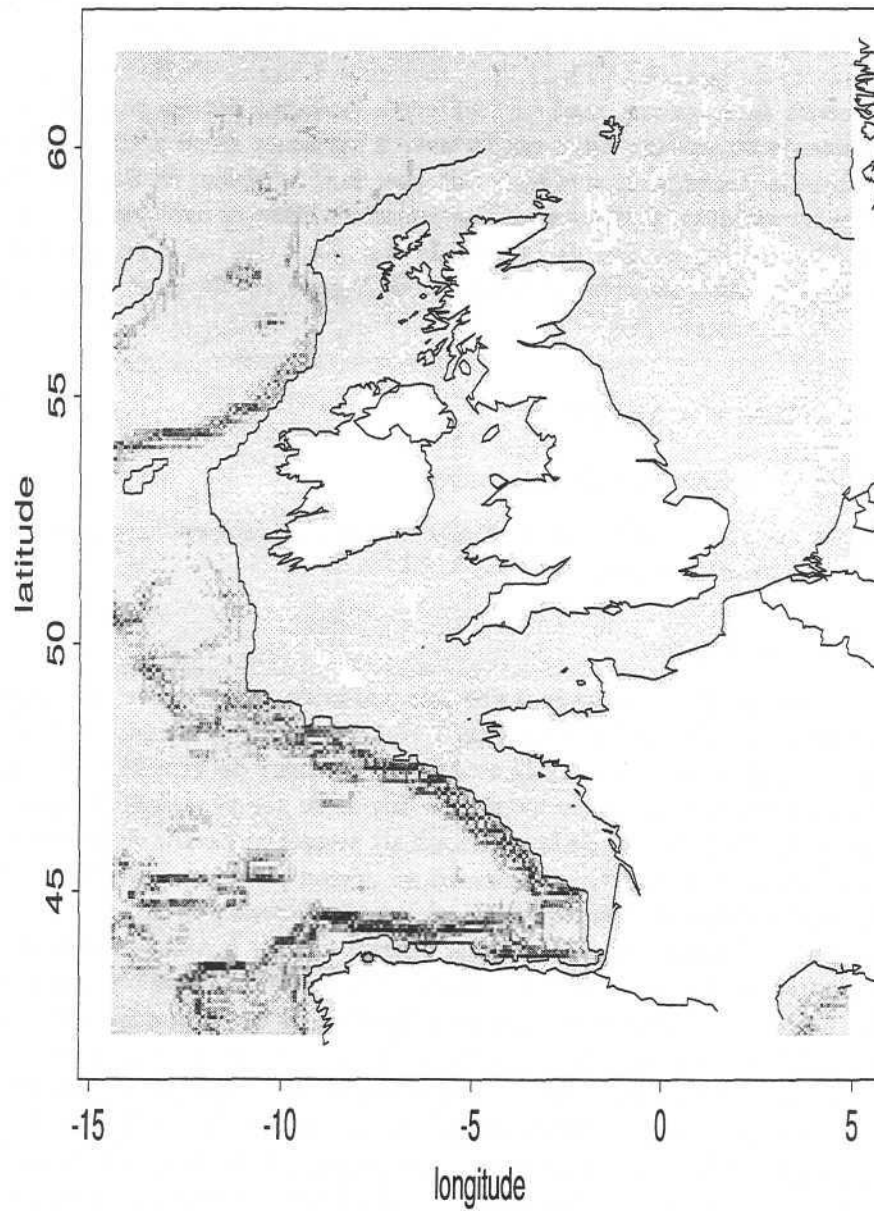


FIG. 3.13 - Représentation du laplacien de la bathymétrie en niveau de gris

3.4.2 Laplacien

Une autre transformée de la bathymétrie *a priori* intéressante est le laplacien de la profondeur. Le laplacien indique les variations de variations. Il est donc d'autant plus fort que la pente varie rapidement (négatif quand la pente diminue et positif quand elle augmente). Ce qui m'intéresse ici sont les changements de pente, j'ai donc considéré la valeur absolue du laplacien. Le laplacien est estimé à partir d'une grille régulière de profondeur en prenant les quatre voisins du point à estimer. Une représentation proportionnelle des laplaciens est présentée figure 3.13.

Le centre de gravité du laplacien (Fig. 3.12c), comme son inertie d'ailleurs (Fig. 3.12d), décroît régulièrement au cours de la saison, sauf en 1986 où le laplacien est un peu plus faible en début de saison. Les géniteurs ne suivent donc pas la zone à laplacien élevé pour s'orienter dans leurs déplacements pendant la ponte. Ceci traduit le fait que, dans la région du Banc de Porcupine (Fig. 3.13), les laplaciens importants se trouvent à l'extérieur du Banc et non pas à l'intérieur, là où se trouvent les œufs. Par contre, au Sud de la zone de frai, pour des latitudes inférieures à 49°00N, les laplaciens élevés coïncident avec des profondeurs de l'ordre de 500 m et donc, avec la présence des œufs.

3.5 Température de surface

3.5.1 Température moyenne de la ponte

Les températures utilisées ici correspondent aux mesures de température de surface réalisées *in situ* par les navires océanographiques.

Evolution au cours du temps

Le centre de gravité de la température par individu augmente au cours de chaque saison de ponte de 10-12°C en début de saison à 14-16°C en fin de saison (Fig. 3.14).

C'est donc que le déplacement vers le Nord-Ouest du centre de gravité de la ponte (Fig. 3.4) n'a pas pour objectif principal de suivre un niveau donné de température. En effet, on aurait pu penser qu'une partie du déplacement de la ponte aurait permis aux œufs de maintenir le centre de gravité de la température de leur environnement à un niveau favorable assez bas, évitant ainsi de subir le réchauffement régional de l'eau en cette période de l'année.

Pour chaque période d'échantillonnage j'ai par ailleurs calculé la température moyenne sur une zone de l'espace centrée sur les accores. J'ai choisi la zone comprise entre plus ou moins 1.5 degrés de part et d'autre des accores et entre les latitudes 48°00 et 52°00N. Le nuage de corrélation (Fig. 3.15) entre la température moyenne sur ce pavé et la température moyenne de la ponte montre qu'il existe une très bonne corrélation entre les deux type de température moyenne. Le décalage maximum observé entre les deux températures moyennes est de 1°C environ. C'est donc que les différences enregistrées sur les centres de gravité de chaque année sont principalement dues aux différences inter-annuelles de température ambiante. En d'autres termes, si le centre de gravité de la température de surface des œufs a été faible en 1986, c'est parce que l'eau était relativement froide, et non parce que les œufs ont été pondus préférentiellement là où l'eau était froide, comparé à l'environnement proche.

La juxtaposition des centres de gravité de la ponte (température et localisation) ainsi que la comparaison de la température ambiante avec la température moyenne des œufs suggère que **la température ne constitue pas un important facteur conditionnant de la distribution des œufs de stade 1**. En particulier, le déplacement systématique de la ponte pendant chaque

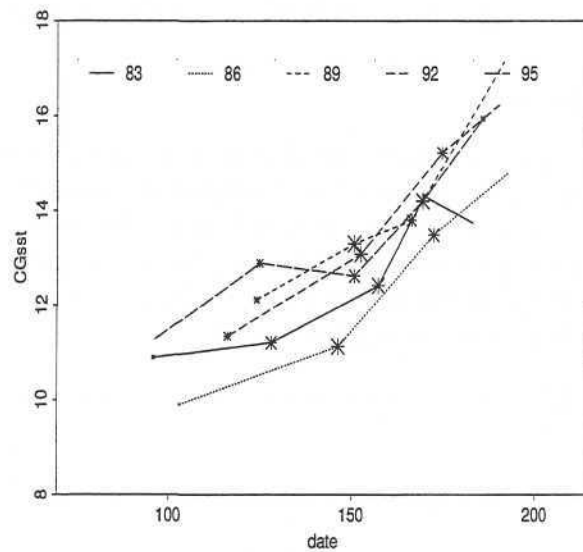


FIG. 3.14 - Evolution du centre de gravité de la température de surface des œufs de stade 1 (E1) au cours de chaque saison. La date est exprimée en jour julien. Les étoiles ont une taille proportionnelle aux nombres d'échantillons utilisés pour les calculs.

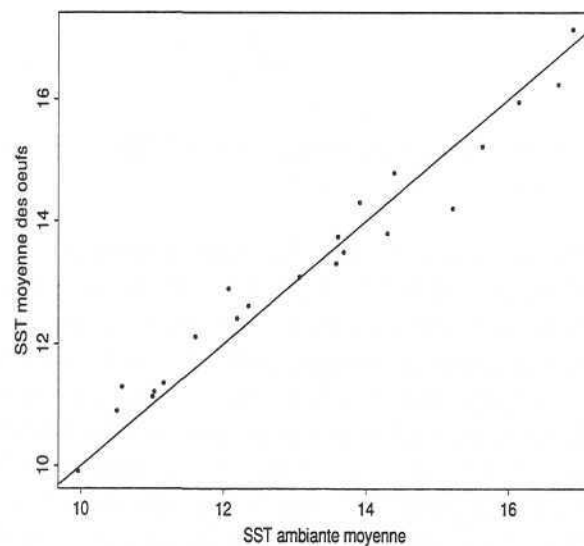


FIG. 3.15 - Nuage de corrélation entre la température de surface ambiante et la température de surface à l'endroit des œufs. En abscisse : température moyenne de l'eau sur la zone comprise entre les latitudes $48^{\circ}00$ et $52^{\circ}00N$ et entre $\pm 1.5^{\circ}$ de part et d'autres de l'isobathe des 200m. En ordonnée : centre de gravité de la température de surface par œufs de stade 1 (E1). La première bissectrice est représentée.

saison de reproduction est probablement, pour l'essentiel, le résultat du flux migratoire des géniteurs (voir plus haut) avant d'être le signe d'une recherche de conditions favorables en température.

Lien avec la densité moyenne

Le nuage de corrélation entre la densité moyenne et la température moyenne par individu (Fig. 3.16) indique une diminution rapide du niveau moyen de densité par individu lorsque la température excède 13.5°C ; au-delà de 14°C la densité moyenne est à son minimum. Les pics de densité moyenne (densité supérieure à 250 œufs/ m^2 /jour) sont donc associés à des températures comprises entre 10 et 13.5°C , sans pour autant qu'une tendance claire se dégage sur cet intervalle de température quant à l'évolution conjointe des deux statistiques.

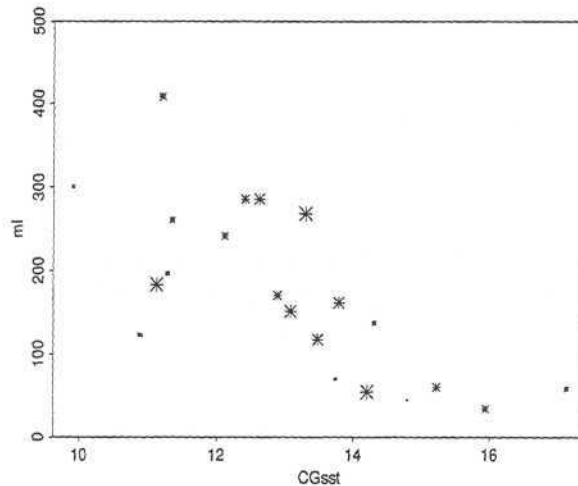


FIG. 3.16 - Corrélation entre la densité moyenne par individu (E1) et le centre de gravité de la température de surface pour E1.

Au 125ème jour de l'année, soit vers le début du mois de Mai, le centre de gravité de la température des œufs est respectivement 10.5 , 11 , 11.7 , 12 et 12.8°C en 1986, 1983, 1992, 1989 et 1995 (Fig. 3.14). Ce classement des années par ordre croissant des centres de gravité des températures correspond grossièrement à celui effectué précédemment sur la base de la date d'apparition du pic de densité moyenne par individu (Cf. §IV.3.1.1). Il ressort donc de cette comparaison que **le pic de densité moyenne semble d'autant plus précoce que la température de l'eau est froide**. Puisque l'abondance en œufs d'une année est d'autant plus faible que le pic de densité moyenne est précoce (voir plus haut), cela signifie aussi que la reproduction serait d'autant plus réussie, en termes de production d'œufs en tous les cas, que la température est voisine de 13°C .

Remarques complémentaires sur la température ambiante

- Profil de température de surface le long des accores

Pour situer le centre de gravité de la température par rapport aux conditions rencontrées par les maquereaux lors de leur migration de reproduction (de l'Ecosse vers le Golfe de Gascogne), j'ai calculé la température moyenne par transect horizontal (latitude fixe) en ne conservant que les prélèvements effectués à moins de 1.5 degrés des accores. Les profils de températures ainsi obtenus

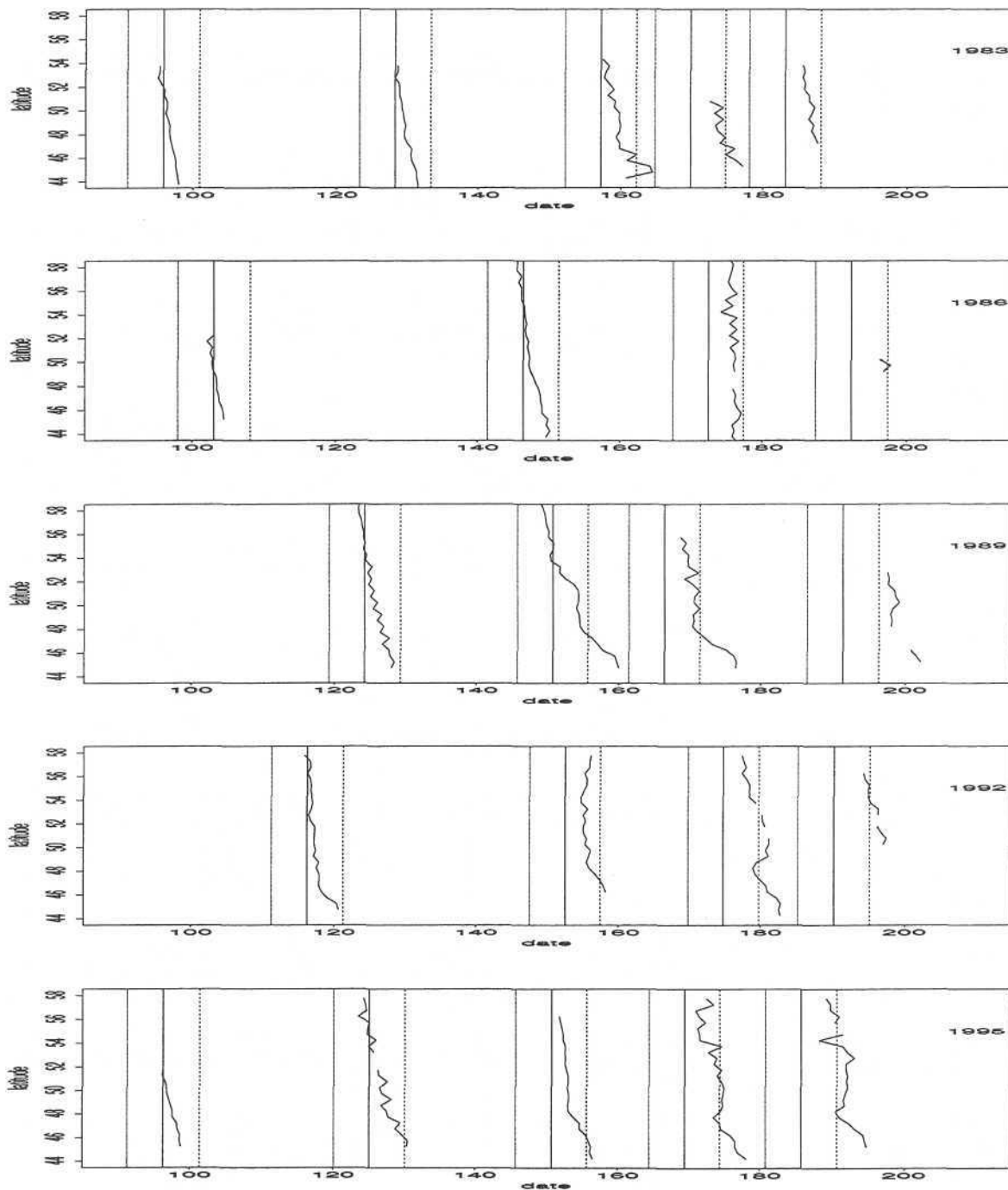


FIG. 3.17 - Evolution de la température de surface le long des accores entre les latitudes $60^{\circ}00N$ et $45^{\circ}00N$. Température moyenne des échantillons situés à moins de 1.5 degrés des accores (isobathe des 200m). Une graduation en abscisse représente $5^{\circ}C$. Le trait plein, situé à la date moyenne de la période, représente $10^{\circ}C$.

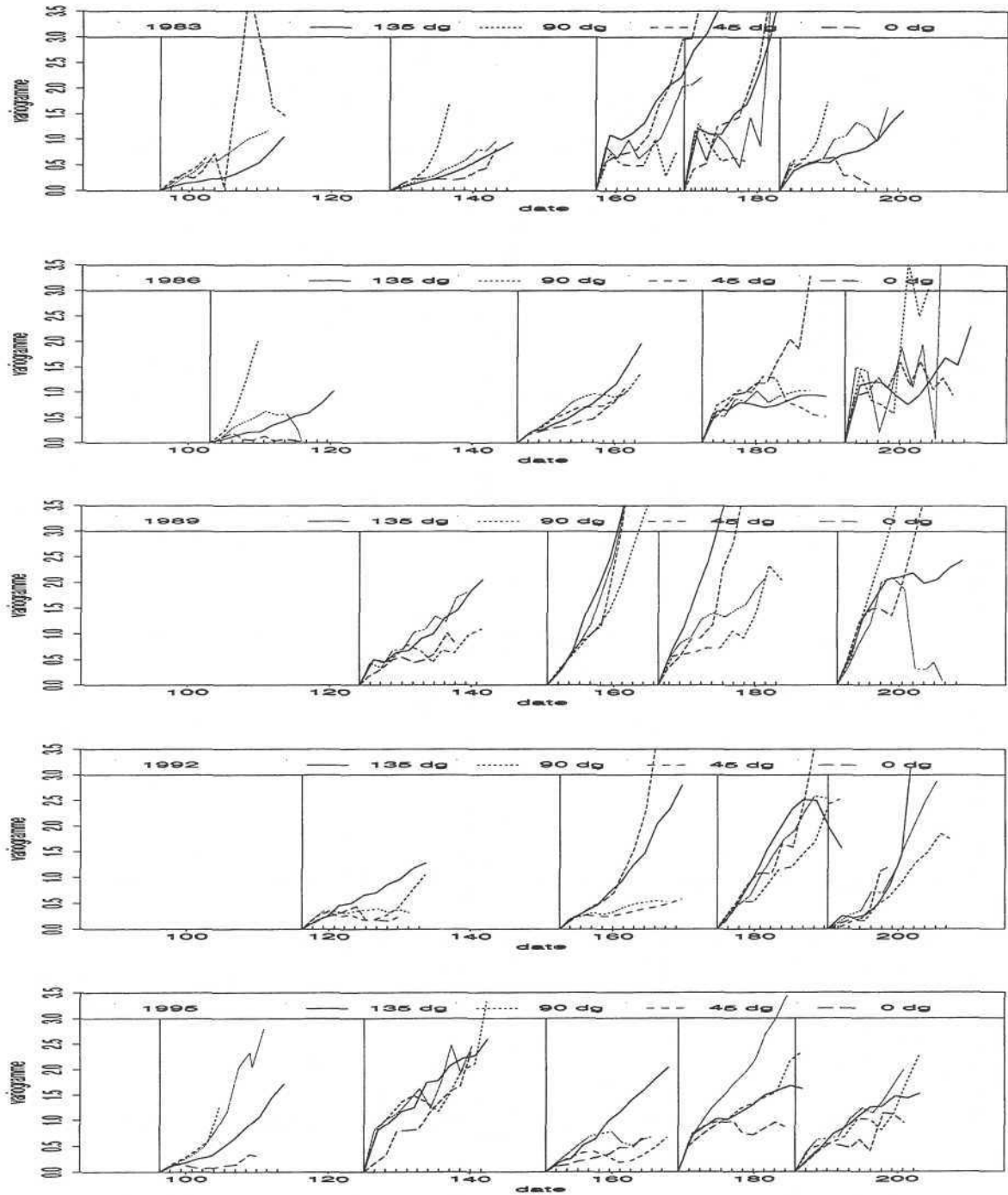


FIG. 3.18 - Variogrammes expérimentaux de la température de surface par période d'échantillonnage. Quatre directions de calcul. Une graduation en abscisse représente 30 milles nautiques.

le long des accores pour chaque période d'échantillonnage sont représentés le long d'un axe temporel (Fig. 3.17).

Cela permet d'évaluer le gradient moyen de la température le long des accores. On trouve ainsi que la température de surface augmente en moyenne en début de saison de 0.4°C par degré de latitude soit $0.007^{\circ}\text{C}/\text{nm}$. De plus, en 1983, 1989 et 1995 un front thermique de l'ordre de 1.5°C par degré de latitude se met en place à la fin du mois de Mai à la latitude de $48-49^{\circ}00\text{N}$, soit à la hauteur du changement d'orientation des accores (Fig. 3.6 par exemple). Ce front n'est donc pas en place quand les premiers maquereaux arrivent du Nord. Ce n'est donc pas lui qui peut expliquer l'arrêt systématique de la migration dans ces latitudes.

Sans que cela ait de valeur explicative claire à mes yeux, on peut noter les deux circonstances suivantes. En 1992, le front thermique était en place dès le début du mois de Mai (jour = 120). C'est l'année où la surface équivalente a été anormalement élevée. Par ailleurs, le front thermique ne semble pas s'être mis en place en 1986, année à faible reproduction et à pic de densité moyenne précoce.

- Structure spatiale

Il est difficile de dégager des tendances claires sur la base des variogrammes expérimentaux de la température de surface (Fig. 3.18). Je soulignerai cependant les points suivants :

- en 1983 et 1986, la pente (discontinuité?) à l'origine augmente avec le temps indiquant que la régionalisation de la température de surface est de plus en plus hétérogène.
- en 1989, la structure est dominée par une non stationnarité Nord-Sud et Est-Ouest.
- en 1992, la période 2 présente une anisotropie assez particulière. La variabilité Nord-Sud et Est-Ouest se stabilise à un niveau très faible dès 60 mn. Dans les deux autres directions, le variogramme indique une dérive.
- en 1995, le variogramme de la période 2 présente une importante discontinuité à l'origine (suivie d'un comportement croissant) comme rarement à cet instant de la saison.

3.5.2 Y-a-t-il un preferendum; et de quelle nature?

Même si, comme il semble que cela soit le cas, la température de surface ne joue pas un rôle fortement conditionnant dans le déroulement de la reproduction, rien n'indique pour l'instant dans quelle mesure et en quoi l'organisation spatiale des œufs est sensible à cette variable. Dans le deuxième chapitre de cette thèse nous avons développé le concept de preferendum en partant du principe que, si les géniteurs ont un preferendum en température, ils doivent se distribuer de façon à minimiser leur inertie en température. Il est alors suggéré de suivre l'évolution de l'inertie des températures de surface par individu lorsque la population est déplacée (translatée) dans l'espace.

La température de surface aux points de données ne suffit pas pour ce genre d'analyses. En 1989, je disposais des moyennes hebdomadaires de température de surface mesurée par satellite. Compte tenu de la date moyenne des périodes d'échantillonnage réalisées en 1989, les semaines intéressantes sont les semaines 4, 5, 8 et 11 (Fig. 3.19 a, b, c et d). J'ai donc étudié les **inertiogrammes de température pour les œufs en 1989**.

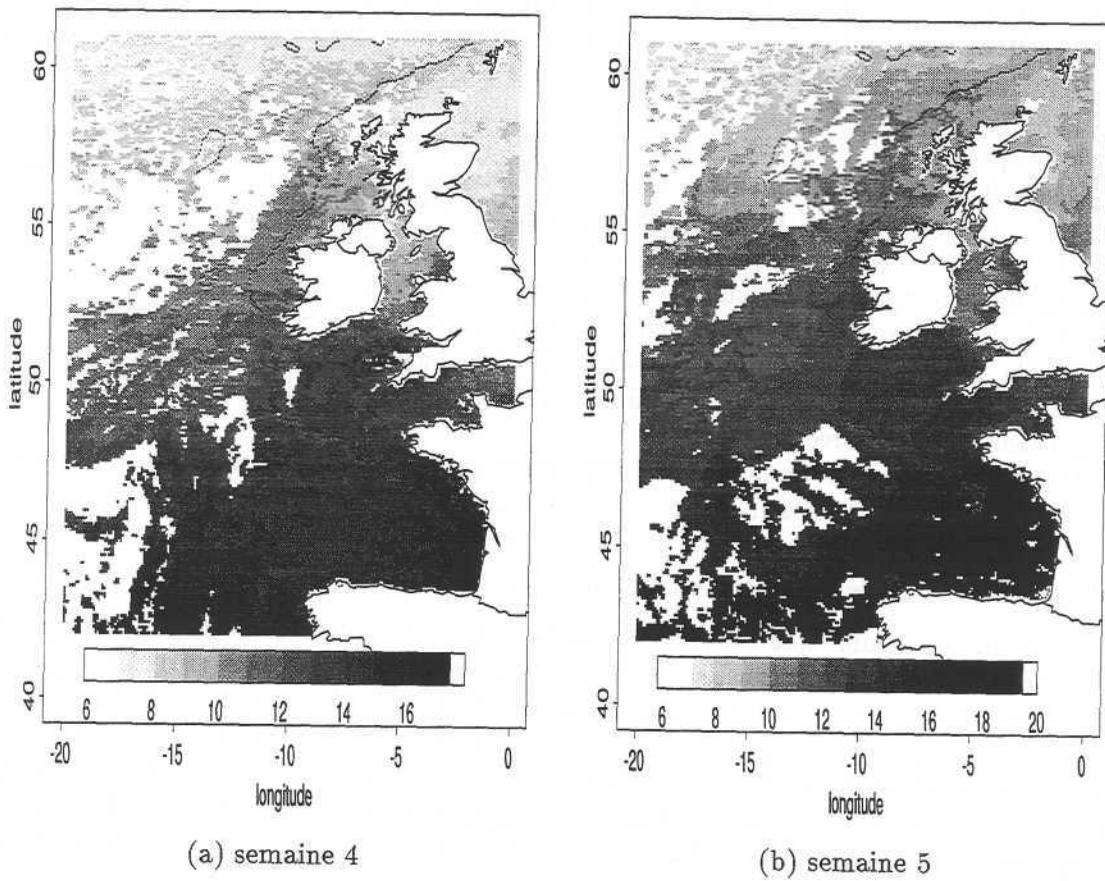


FIG. 3.19 - (a) et (b): Température de surface mesurée par satellite. Moyennes hebdomadaires pour les semaines 4, et 5.

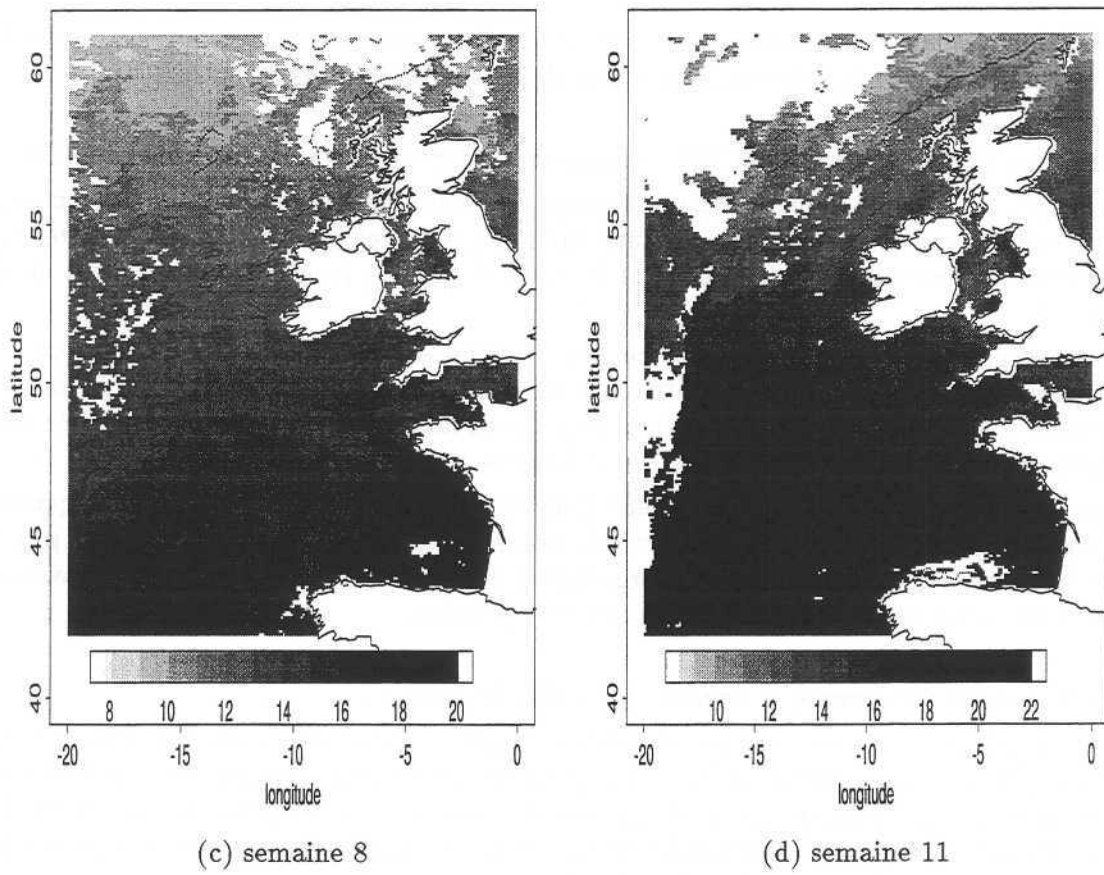


Fig. 3.19 (c) et (d) : Température de surface mesurée par satellite. Moyennes hebdomadaires pour les semaines 8 et 11.

Choix méthodologiques

- Périodes et sous-périodes

Le navire Ecossais responsable des mesures pendant la deuxième période d'échantillonnage de 1989, a effectué un premier balayage Nord-Sud de la zone de frai, en doublant la distance inter-transects, puis un deuxième passage Sud-Nord pour compléter le premier (partie III, Fig. 2.2).

Par ailleurs, les première et deuxième moitiés de cette période d'échantillonnage correspondent respectivement aux semaines 4 et 5 de mesures satellitaires (Tab. 2.4). Or la couverture nuageuse de ces deux semaines est suffisamment faible pour que les données soient utilisables (Fig. 3.19 a et b).

J'ai donc choisi, pour cette analyse assez fine du conditionnement par la température, de décomposer la deuxième période d'échantillonnage en **deux sous-périodes 2.1 et 2.2.**

- Abondance d'œufs par classes de SST : comparaison entre mesures *in situ* et par satellite

Pour chacune des (sous)périodes d'échantillonnage, l'histogramme expérimental de la température par individu est construit avec l'une et l'autre des mesures de température (*in situ* et satellite); pour chaque histogramme on indique également le centre de gravité, l'inertie et le coefficient de variation (Fig. 3.20).

Le centre de gravité estimé à partir des données satellitaires est en moyenne supérieur de 1°C à celui des mesures de terrain (0.4°C et 1.6°C pour les périodes 3 et 4). Les histogrammes ont une forme en cloche mais dans le cas des mesures *in situ*, l'abondance totale se trouve concentrée dans un plus petit nombre de classes de température et l'inertie tend à être plus faible.

Il apparaît donc délicat de mélanger les deux sources d'information. Par soucis d'homogénéité, et compte-tenu des exigences inhérentes aux calculs des inerties translattées, **je n'ai utilisé par la suite que les mesures satellitaires.** L'interprétation des résultats, et des centres de gravité en particulier, devra donc être faite à la lueur de cette comparaison.

- Prise en compte des accores et représentation des résultats

La distribution des œufs est le plus souvent en forme de croissant, fortement associée aux accores. Pour prendre en compte cette caractéristique importante et persistante de la distribution des œufs et pour que les translations aient un sens du point de vue du poisson, l'ensemble des calculs a été réalisé dans un **repère centré sur les accores.** La procédure suivie pour les calculs et leur représentation graphique est donc la suivante :

- les champs de température de surface et les grilles de densités d'œufs sont tout d'abord replacés dans un système d'axes "latitude" x "accores".

- la population d'œufs, ou plus précisément la grille des densités, est translattée jusqu'à 2 degrés dans chaque direction par pas de 0.25 degré. Dans ce repère les mouvements sont donc des combinaisons de déplacements parallèles et perpendiculaires aux accores.

- Finalement, les résultats sont replacés dans le repère géographique normal "latitude" x "longitude" autour du centre de gravité réel de la position des œufs. L'évolution du centre de gravité de la température est représenté par des lignes de niveau tous les 0.5°C (Fig. 3.21). Les inerties (Fig. 3.22) sont représentées proportionnellement à la valeur réelle à l'origine (c'est-à-dire celle obtenue pour la grille d'échantillonnage). Un doublement de l'inertie par rapport à la valeur à l'origine est indiqué par un triangle.

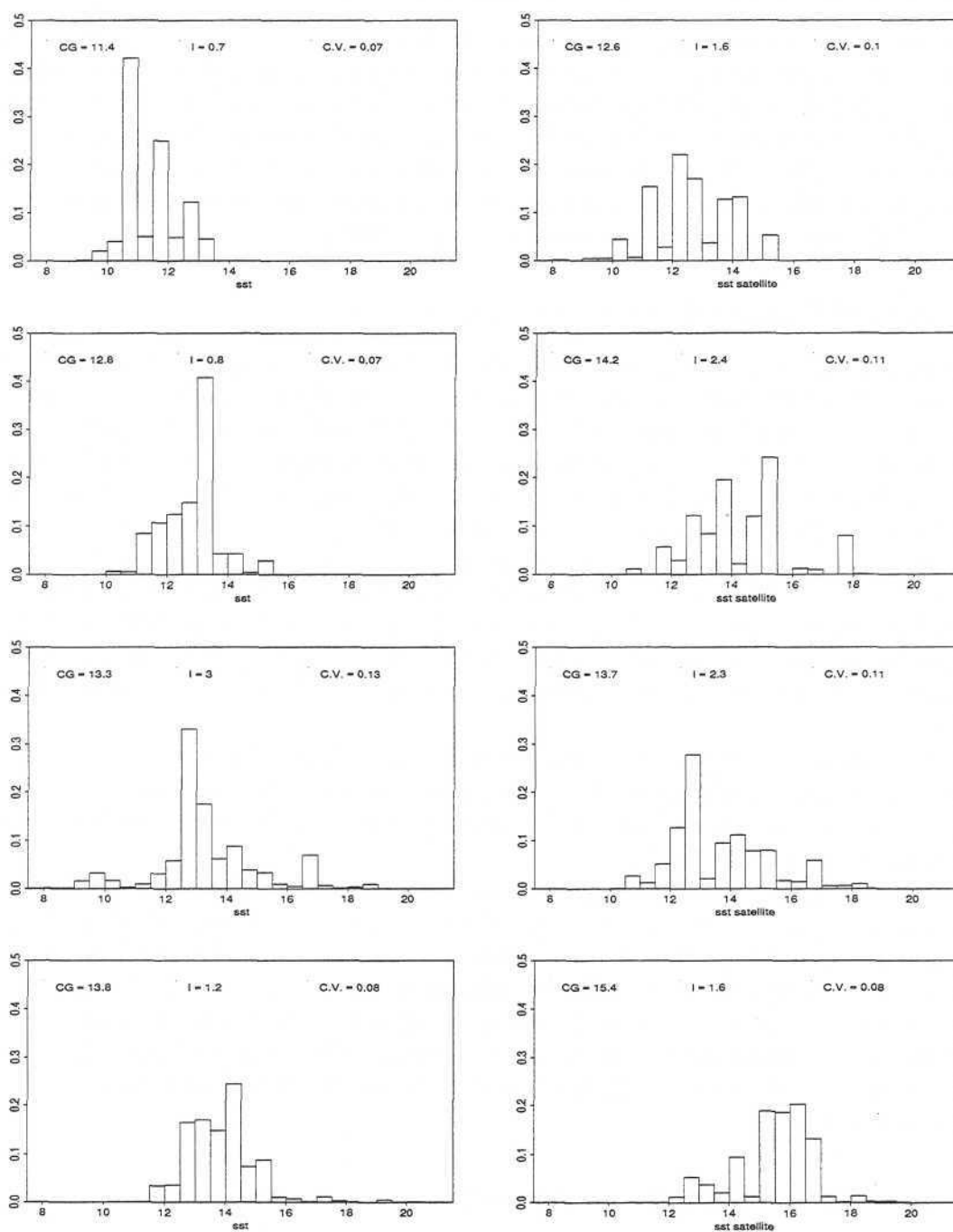


FIG. 3.20 - Histogramme de la température de surface par individu. Œufs de stade 1. **Colonne de gauche** : température de surface mesurée in situ en 1989, périodes 2.1 (en haut), 2.2, 3 et 4 (en bas). **Colonne de droite** : température de surface mesurée par satellite en 1989 pendant les semaines 4 (en haut), 5, 8, et 11 (en bas). Les abondances par classes sont exprimées relativement à l'abondance totale.

Centres de gravité de la SST pour différentes translations

Conformément à l'orientation des courbes de niveaux (Fig. 3.21), on voit que globalement tout déplacement de la population perpendiculairement aux accores ne modifie pas sensiblement son centre de gravité en température de surface, à l'exception du cadran Sud-Est dans le courant de la (sous)période 2.2. Dans ce dernier cas, la forme incurvée des isolignes s'explique par l'intrusion d'une langue d'eau froide le long des accores bien visible sur l'image satellite (Fig. 3.19 b).

Parallèlement aux accores, "se" déplacer d'environ 2 degrés en latitude (120 milles nautiques), modifie de 1°C seulement le centre de gravité de la population.

Inertie de la SST pour différentes translations

Comparée à son niveau de départ, c'est-à-dire à l'inertie des températures à l'endroit réel des œufs, l'inertie de la température ne double que pour des translations importantes le long des accores vers le Sud-Est pendant la période 2.2 ou encore vers le Nord-Ouest lors de la période 4.

On peut également constater que le centre de gravité de la population ne se trouve **jamais dans un puits d'inertie**. Si, selon le principe testé, les œufs avaient été pondus de façon à minimiser leur inertie en température, l'inertie aurait dû augmenter pour toute translation.

Toutefois, à part lors de la période 2.2, les œufs sont distribués de sorte qu'ils sont proches d'une zone où leur inertie aurait été plus forte (s'ils y avaient été). En d'autres termes, un petit déplacement dans une certaine direction, direction qui change d'ailleurs en fonction de la période, est en général associé à une augmentation de l'inertie de la température de surface par œuf. Ainsi en replaçant la population dans son parcours migratoire, on voit que si la population de géniteurs avait été :

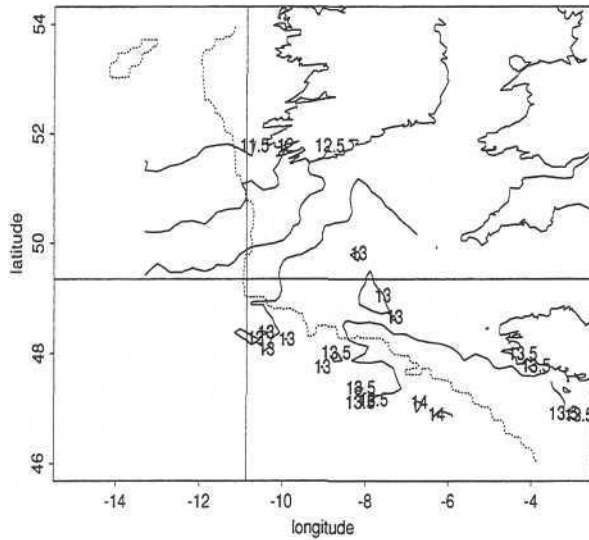
- plus au Sud-Est le long des accores en début de saison (Fig. 3.22 b),
- plus à l'intérieur du plateau continental en milieu de saison (Fig. 3.22 c),
- et plus vers le Nord-Nord-Ouest en fin de saison (Fig. 3.22 d), l'inertie de la température de surface aurait été multipliée par 1.5.

Il serait bien sûr souhaitable de pouvoir vérifier la reproductibilité de ces observations sur d'autres saisons de reproduction. Il reste cependant que ces premiers résultats d'inertie par translation suggèrent un mode possible de conditionnement de la ponte par la température.

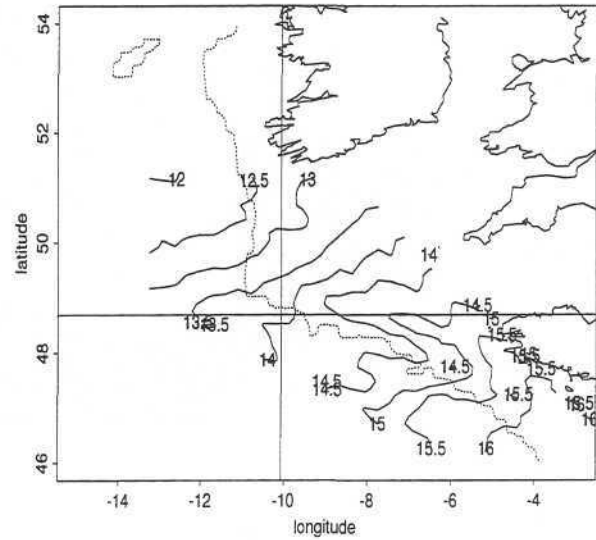
Au sein d'une **plage favorable de température** (le concept de referendum est remplacé par celui d'intervalle favorable), la ponte se fait à proximité d'un "**front virtuel d'inertie**". Il s'agit d'une notion différente de celle de front thermique réel, ce qui explique qu'elle soit mise entre parenthèses. Le front dont je parle est un front virtuel, mis en évidence par des déplacements imaginaires de populations.

Interprétation : dynamique liée au gradient de température

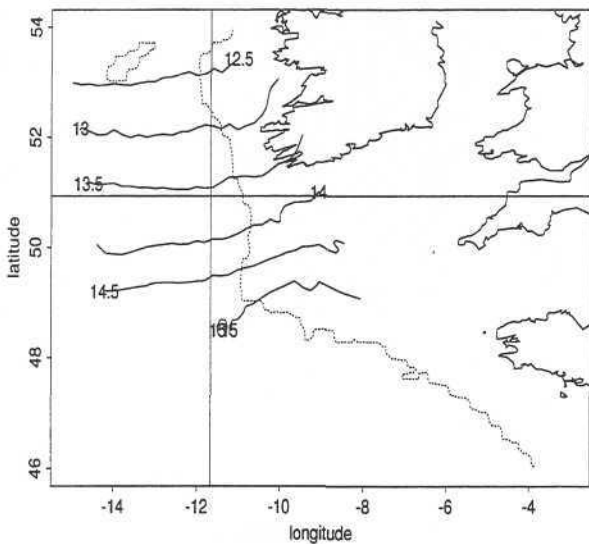
Les œufs sont donc distribués de telle sorte qu'ils minimisent, d'une certaine manière, l'inertie de leur température. Conformément à ce qui a été exposé précédemment (§II.2.3.4), cela est compatible avec un **modèle dynamique** où le déplacement des individus est lié au gradient de la température de surface. Plus précisément, nos résultats indiquent que, pendant la saison de reproduction, les maquereaux se déplacent d'autant plus vite que la combinaison du gradient de température et de l'écart entre la température préférée et la température réelle, est importante. C'est ce mode de **conditionnement du déplacement des individus** qui tend à amener la population à minimiser, à un instant donné, son inertie en température.



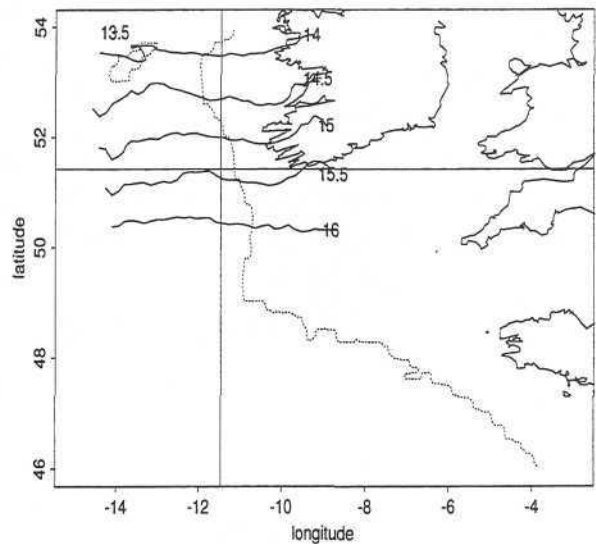
(a) période 2.1 - semaine 4



(b) période 2.2 - semaine 5



(c) période 3 - semaine 8



(d) période 4 - semaine 11

FIG. 3.21 - Centre de gravité de la température de surface pour différentes translations de la population d'oeufs de stade 1. Translations par pas de 0.25 degrés parallèlement et perpendiculairement aux accores jusqu'à un maximum de 2 degrés. Grilles de densités des périodes 2.1, 2.2, 3 et 4 de 1989 associées aux mesures satellitaires des semaines 4, 5, 8 et 11. Lignes de niveaux tous les 0.5°C.

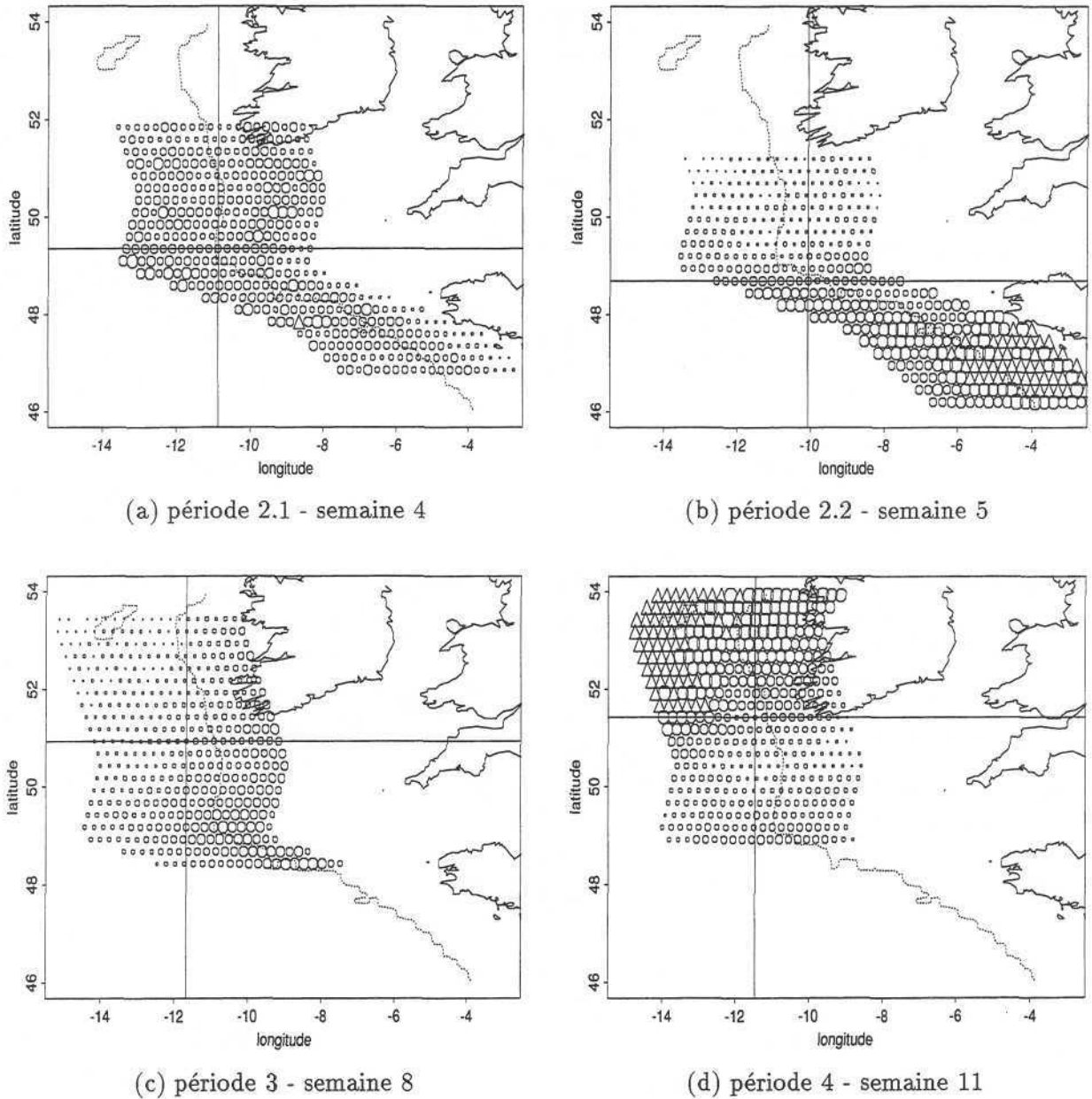


FIG. 3.22 - Inertie de la température de surface pour différentes translations de la population d'oeufs de stade 1. Translations par pas de 0.25 degrés parallèlement et perpendiculairement aux accores jusqu'à un maximum de 2 degrés. Grilles de densités des périodes 2.1, 2.2, 3 et 4 de 1989 associées aux mesures satellitaires des semaines 4, 5, 8 et 11. Représentation proportionnelle par des cercles entre 0 et deux fois la valeur à l'origine. Les inerties supérieures sont représentées par un triangle de taille fixe.

3.6 Synthèse

Il ressort donc les points suivants :

- les changements de structures spatiales concernent essentiellement les distances inférieures à la maille de prélèvement (0.5 degré). Le reste de la structure spatiale est globalement invariante dans le temps relativement à l'abondance observée lors de chaque période d'échantillonnage : elle est anisotrope de portée 600 mn dans la direction 121° et 200 mn dans la direction perpendiculaire.
- l'évolution du comportement agrégatif, ou encore de la surface équivalente, des œufs au cours de chaque saison de ponte est identique quelle que soit l'année (sauf exception en 1992). La surface équivalente augmente jusqu'au 175ème jour de l'année pour atteindre 25 000 nm² et diminue ensuite. Le valeur observée en 1992, soit environ 43 000 mn², est interprétée alors comme un maximum pour l'espèce.
- la densité moyenne par individu atteint son maximum (le plus souvent 250 - 300 œufs/m²/jour) à un moment de la saison variable avec les années : plus il survient tard dans la saison, ou plus précisément plus il survient en même temps que le pic de surface équivalente, et plus le nombre d'œufs pondus sera important. Le pic de densité moyenne semble être d'autant plus précoce que la température est froide (les années froides seraient donc des années de mauvaises reproductions). La densité moyenne chute rapidement quand la température moyenne des œufs dépasse 13.5°C.
- chaque année, le centre de gravité de la ponte se déplace d'abord vers le Nord-Ouest, puis vers le Sud-Est. Ce déplacement n'est pas associé à une recherche de conditions favorables en températures mais traduit plutôt le flux migratoire des géniteurs sur la zone de frai. Entre 1986 et 1995, la position moyenne de la ponte, comme d'ailleurs l'ensemble de la population, s'est déplacée vers l'Ouest de façon significative (100 mn environ) au moment du pic de ponte (fin Mai).
- au sein d'un intervalle favorable, le rôle conditionnant de la température de surface tient à la présence de "fronts virtuels d'inertie". Cela signifie que le critère utilisé pour tester la réalité d'un referendum (inertie de la température) peut augmenter assez rapidement pour un petit déplacement de la population.

Chapitre 4

Etude du développement embryonnaire et larvaire

Sommaire

Ce dernier chapitre est consacré à l'étude des phases embryonnaires et larvaires du maquereau.

*Le **paragraphe 1** cherche à comparer les statistiques individuelles des œufs de stade 1 et 5 entre 1983 et 1995. Des statistiques aussi simples que l'abondance totale ou le centre de gravité permettent de mettre en évidence la densité - dépendance du taux de survie des œufs ou encore l'augmentation de l'agrégation des individus.*

*Le **paragraphe 2** considère le cas particulier de l'année 1989 où nous disposons d'informations précises sur les densités de larves. L'analyse porte sur la comparaison des statistiques individuelles entre stades et traite, par exemple, de l'évolution de la mortalité en fonction de l'âge ou encore du caractère passif ou actif des individus des différents stades.*

Analysis of the early life stages

Abstract

This chapter, the latest, deals with the early stages of mackerel namely the eggs and larval stages.

*The **paragraph 1** looks at the differences between the stage 1 and stage 5 eggs between 1983 and 1995. Simple individual based statistics like the total abundance or the centre of gravity are used. This gives the opportunity to quantify the density - dependency of the mortality or the increase of the aggregation of individuals.*

*The **paragraph 2** concerns the analysis of the year 1989 as we get for that particular year, precise information on larvae. The analysis is nothing but the comparison of the individual based statistics. It reveals some interesting results like the increase of the*

mortality rate with the development of the eggs and larvae. We also get a body of arguments for distinguishing passive individuals (larvae smaller than 4 mm) and active - swimming ones.

En dehors des informations fondamentales concernant la physiologie et l'intensité de la reproduction, les campagnes triennales organisées par le CIEM fournissent des informations sur les densités d'œufs de stade 2 à 5, ou encore les densités de larves. Cela nous donne donc la possibilité de suivre l'évolution des statistiques utilisées pour caractériser la ponte, au cours du développement embryonnaire et larvaire.

La première partie de ce chapitre est consacrée à l'étude comparative du dernier stade de développement embryonnaire, c'est-à-dire le stade 5 des œufs (E5), au premier stade (E1) pour chacune des 5 campagnes triennales (1983 à 1995).

Julie Fives, de l'Université de Galway (Irlande) m'a fourni, pour 1989, les densités de larves réparties en trois groupes de tailles. La deuxième partie de ce chapitre concernera donc l'étude du développement larvaire en 1989.

4.1 Analyse du dernier stade embryonnaire (E5)

4.1.1 Abondance, densité moyenne et surface équivalente

Courbes de production : indices d'une mortalité dépendant de l'agrégation

Le pic d'abondance des œufs de stade 5 (Fig. 4.1 a) est beaucoup plus resserré dans le temps que celui des œufs E1 (Fig. 3.1 a). Le pic d'abondance de E5 survient vers le 150ème jour de l'année, soit fin Mai. La valeur du taux de production à cet instant permet de distinguer trois bonnes années (1983, 1989, 1995) et deux mauvaises années (1986, 1992): $1.5 \cdot 10^{13}$ œufs/jour pour les premières contre $1.2 \cdot 10^{13}$ œufs/jour pour les secondes. Donc 1992, qui est apparue comme une très bonne année pour la production d'œufs, est une mauvaise année pour le développement des œufs. En somme la mortalité embryonnaire a été très importante, et il semble que 1992 soit **une fausse bonne année en matière de reproduction**. Cela conforte d'ailleurs l'analyse et la tentative de modélisation que nous avons faite précédemment où les observations concernant 1992 ont été jugées "a-normales".

On a montré de plus que l'importance de la ponte en 1992 était liée à une diminution du niveau d'agrégation des œufs. Cela pourrait suggérer que la mortalité des œufs a été plus forte parce que leur agrégation était trop faible au départ.

Surface équivalente et agrégation

Les surfaces équivalentes de E5 suivent une évolution en cloche au cours de chaque saison de reproduction (Fig. 4.1 c). Elles présentent plus de disparité inter-annuelle que celles des œufs E1. En particulier, la valeur maximale atteinte chaque année fluctue entre $20\,000 \text{ mn}^2$ et $40\,000 \text{ mn}^2$.

La comparaison des surfaces équivalentes de E1 et E5 par période (Fig. 4.1 d) montre que globalement la surface équivalente diminue avec le développement des œufs. Cela signifie que **l'agrégation des œufs E5 est supérieure, en moyenne, à celle des œufs E1**.

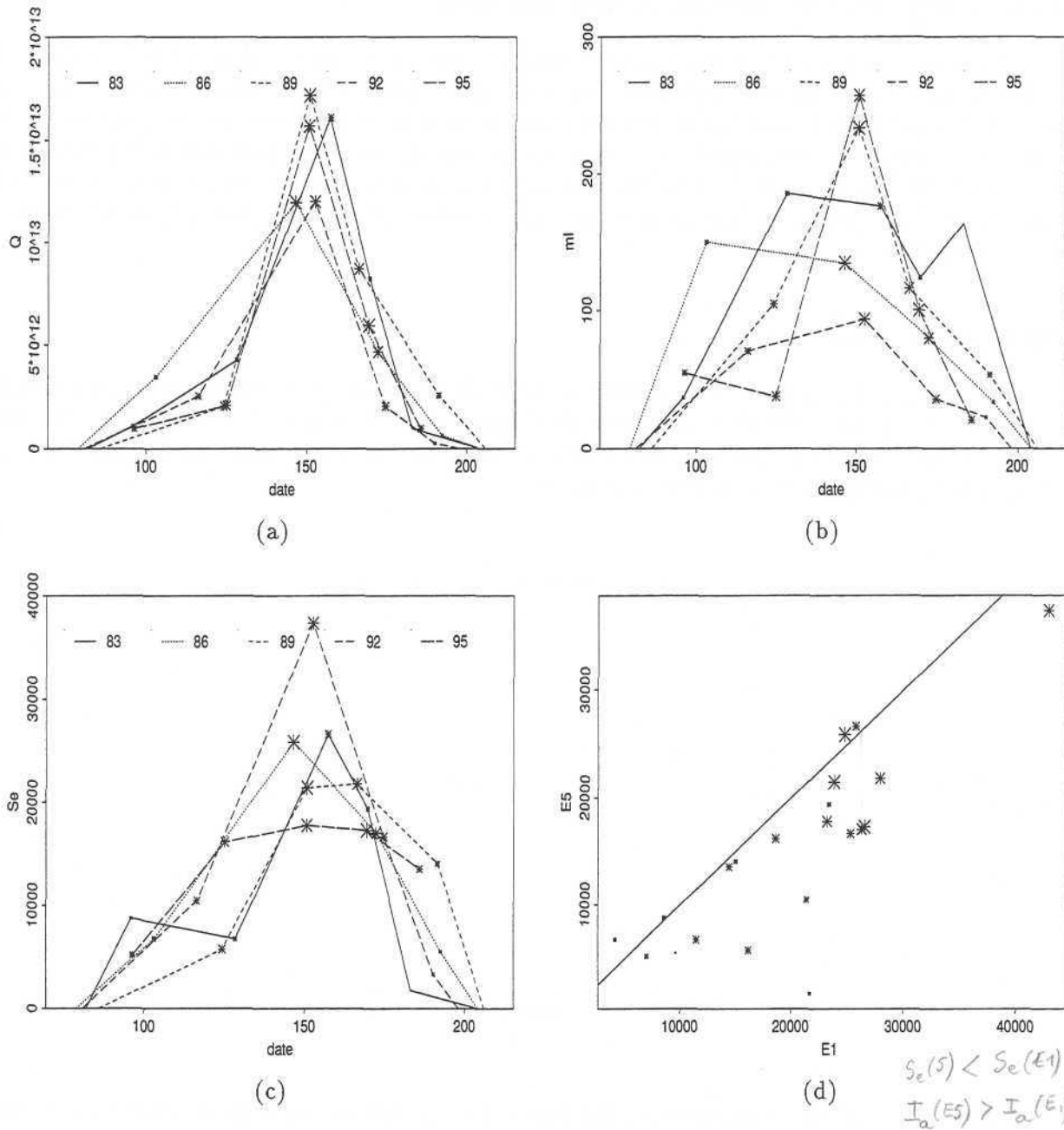


FIG. 4.1 - Statistiques par individu (E5) en fonction de la date moyenne des périodes d'échantillonnage exprimée en jour julien. Les étoiles ont une taille proportionnelle aux nombres d'échantillons utilisés pour les calculs.

- (a) : courbes de production donnant le nombre d'œufs de stade 5 produit par jour.
- (b) : densité moyenne par individu exprimée en nombre œufs/m²/jour.
- (c) : surface équivalente (en milles nautiques au carré) ou inverse de l'indice d'agrégation.
- (d) : Nuage de corrélation entre les surfaces équivalentes des populations d'œufs de stade 1 (abscisse) et 5 (ordonnée).

Densité moyenne : une précision concernant 1992

Lorsqu'elle est à son maximum, la densité moyenne par œuf de stade 5 (Fig. 4.1 b) vaut entre 100 et 250 œufs/m²/jour. Je m'attarderai ici sur le niveau particulièrement bas de la densité moyenne en 1992 qui fait peser la baisse d'abondance observée précédemment sur la densité moyenne plutôt que sur l'agrégation. Cela signifie en outre que la mortalité a été à peu près indépendante du niveau de densité. En effet, si la mortalité avait été plus forte pour les parties peu denses de la population, l'agrégation, relativement à l'abondance totale, aurait augmenté. Ce qui est contraire à nos observations.

Mortalité ou survie totale

L'intégration des courbes de production de E1 et de E5 fournit pour chaque année les quantités totales d'œufs pondus ($Q(E1)$) et d'œufs développés en stade 5 ($Q(E5)$). Le rapport des deux quantités ($Q(E5)/Q(E1)$) permet donc d'estimer la survie totale (naturelle, prédation) des œufs, ou de façon complémentaire, la mortalité totale.

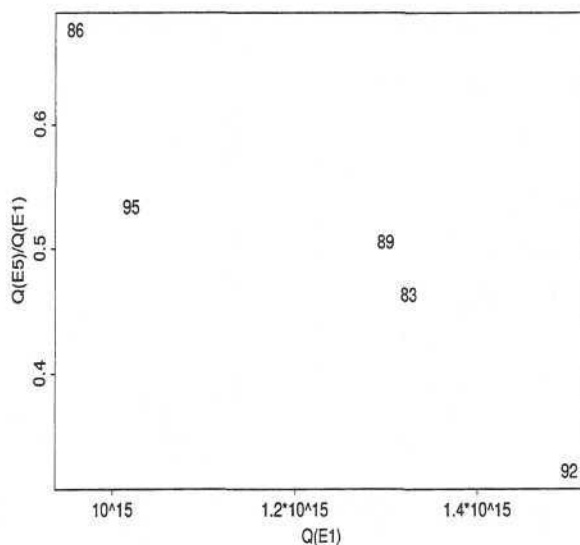


FIG. 4.2 - Taux de survie pré-larvaire. Rapport entre l'abondance des œufs de stade 5 et de stade 1 ($Q(E5)/Q(E1)$) et l'abondance de la ponte ($Q(E1)$).

Selon les années, 32% à 67% des œufs pondus se transforment en œufs de dernier stade (Fig. 4.2). Mais surtout, la mortalité totale pré-larvaire apparaît d'autant plus forte que l'abondance des œufs est importante. Elle est par exemple très faible en 1986 (67% de survie) alors que la ponte est précisément la plus faible des 5 années étudiées. Ce résultat, bien que ne reposant que sur 5 "observations", permet de quantifier la **densité - dépendance du taux de survie** des stades prélarvaires chez le maquereau.

Les mécanismes éco - physiologiques susceptibles d'expliquer cette densité - dépendance pour des particules vivantes, mais passives, ne sont cependant pas bien élucidés à ma connaissance.

4.1.2 Localisation et transport

Position des centres de gravité

La description des distributions d'œufs E5 par les centres de gravité et les axes principaux d'inertie fournit des résultats attendus (Fig. 4.3). En effet, la position moyenne des œufs E5 se déplace globalement conformément à celle de la ponte et l'inertie reste principalement expliquée par une direction parallèle aux accores. En réalité, l'intérêt réside surtout dans la comparaison des centres de gravité de E1 et E5 pour l'estimation des transports.

Transport

Sachant :

- qu'un œuf met environ 6 jours à 12°C pour atteindre le stade 5, soit la moitié ou le tiers de la durée moyenne d'une période d'échantillonnage,
- que les œufs n'ont aucune possibilité de nage,
- et que depuis 1986 (inclus) les œufs de stade 1 et 5 sont connus aux mêmes points (en 1983 les densités d'œufs de stade 5 sont connues en un trop petit nombre de points),

la comparaison des centres de gravité des stades E1 et E5 d'une même période d'échantillonnage devrait permettre d'estimer le transport moyen subi par les œufs pendant leur première semaine de vie.

En toute rigueur pour analyser ce type de transport, il faudrait connaître le centre de gravité des œufs E5 au cours d'une campagne qui aurait lieu une semaine après une première campagne donnant la position moyenne de la ponte; chaque campagne réalisant si possible une image instantanée des distributions. Cela n'est bien évidemment pas possible. Mais, les courbes de production des œufs E1, et dans une moindre mesure des œufs E5, montrent que les niveaux de productions ne changent pas beaucoup en l'espace de 6 jours. De plus, la cartographie des centres de gravité de E5 permet d'estimer la position moyenne des œufs de stade 5 une semaine après l'échantillonnage. En effet, connaissant la distance et le temps qui séparent les centres de gravité de deux périodes successives, il est facile, par une simple règle de trois, d'estimer la position du centre de gravité des œufs de stade 5, 6 jours après chaque campagne.

L'ordre de grandeur du transport à court terme des œufs est alors fourni par la comparaison du centre de gravité de la ponte et du centre de gravité estimé de E5. La vitesse du transport est estimée par le rapport entre la distance (en cm) qui sépare les centres de gravité et le temps (en secondes) qui s'est écoulé entre les deux, c'est-à-dire une semaine.

Enfin, pour que la dimension longitudinale des transports ne soit pas trop influencée par la forme en croissant des distributions spatiales, les calculs ont été effectués dans un repère centré sur les accores. Les résultats sont toutefois représentés dans le repère géographique normal après transformation inverse des coordonnées des centres de gravité. Les distances cependant représentent toujours des déplacements par rapport aux accores.

Les transports présentent des directions et des intensités assez variables de période à période mais aussi d'année en année (Fig. 4.4). Leur vitesse est comprise entre 1 cm/s et 30 cm/s (13 cm/s en moyenne).

En 1986 (Fig. 4.4 a), les transports sont faibles et convergent vers le point -10°00W ; 50°00N, de sorte que la position du centre de gravité de la population d'œufs de stade 5 ne bouge quasiment pas au cours de la saison. Cela suppose que le courant s'est inversé en l'espace de 3 mois.

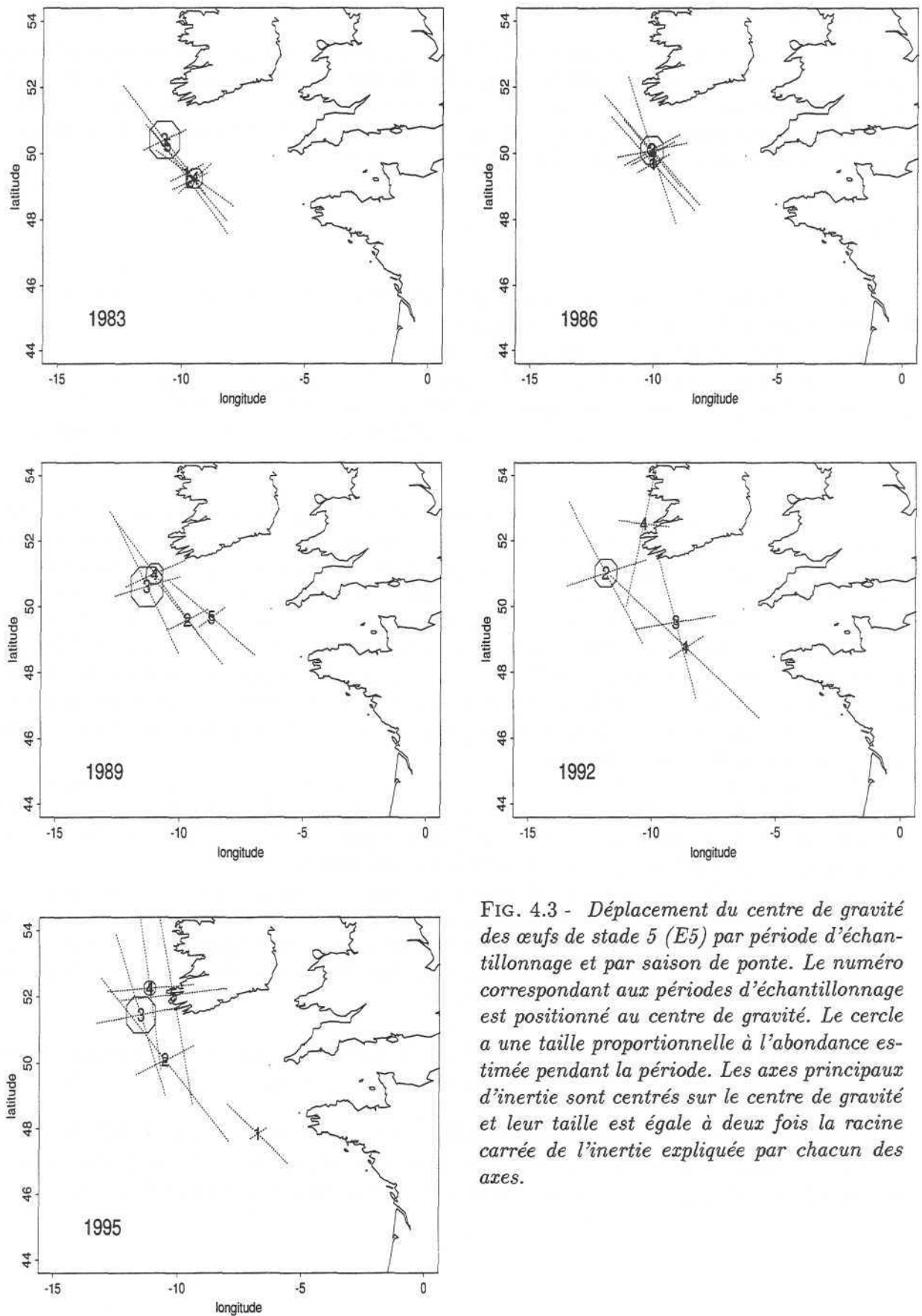


FIG. 4.3 - Déplacement du centre de gravité des œufs de stade 5 (E5) par période d'échantillonnage et par saison de ponte. Le numéro correspondant aux périodes d'échantillonnage est positionné au centre de gravité. Le cercle a une taille proportionnelle à l'abondance estimée pendant la période. Les axes principaux d'inertie sont centrés sur le centre de gravité et leur taille est égale à deux fois la racine carrée de l'inertie expliquée par chacun des axes.

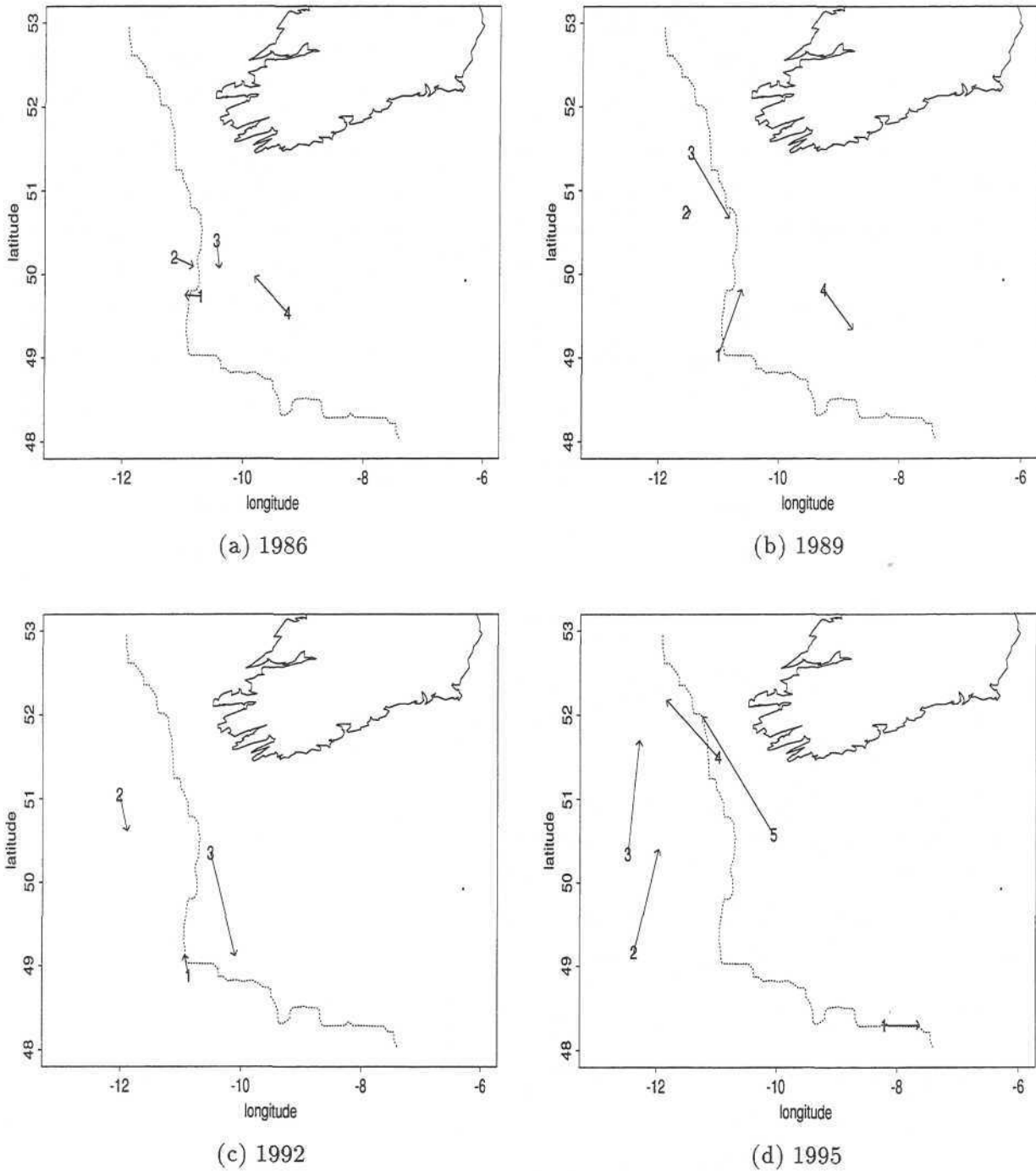


FIG. 4.4 - Estimation du transport des œufs sur une semaine. Déplacement du centre de gravité de la ponte vers le centre de gravité estimé des œufs de stade 5.

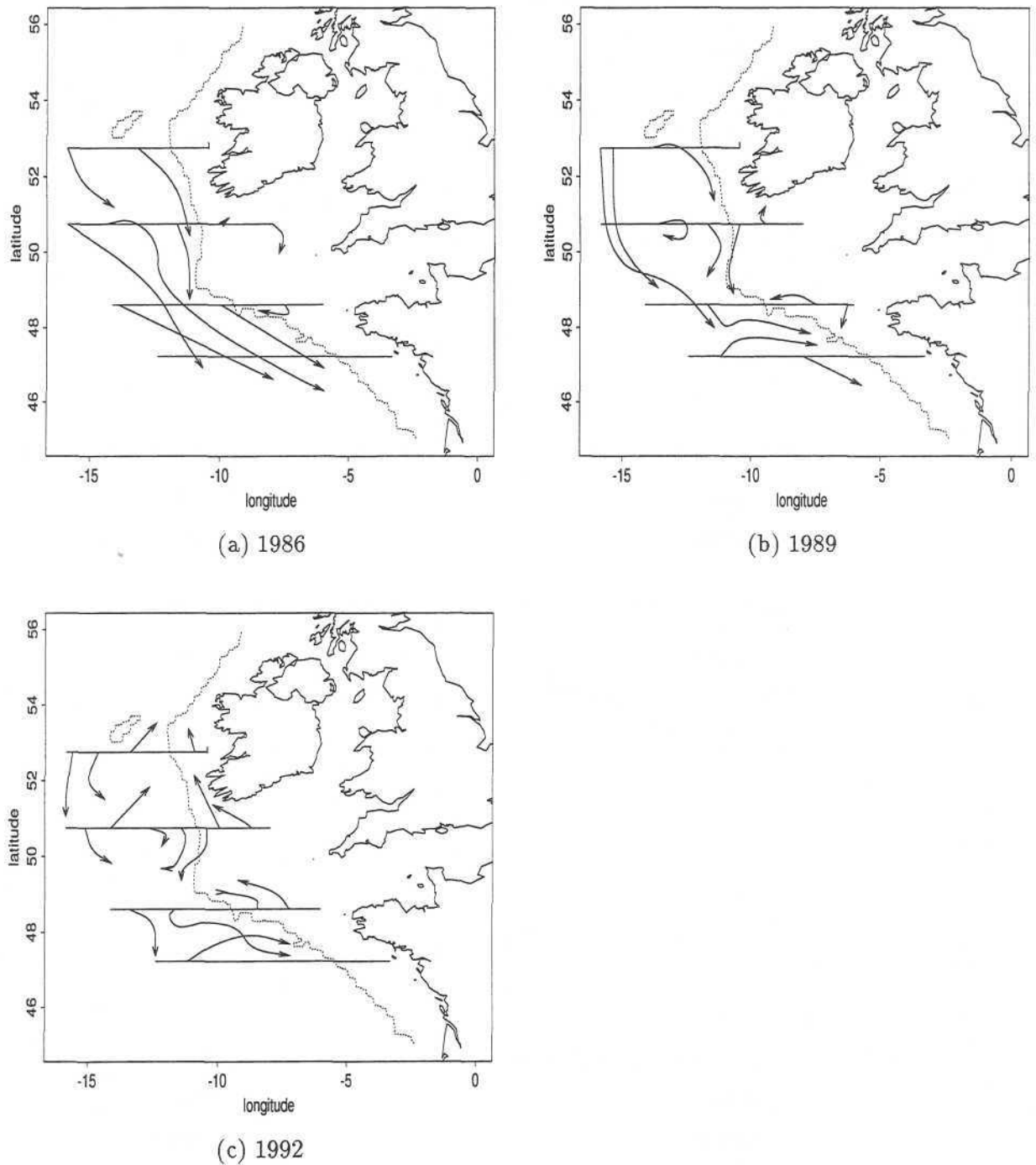


FIG. 4.5 - Estimation du transport des œufs par modélisation des courants. Synthèse des "spaghetti-plots" rapportés dans Walsh et al. (1996). Estimation du chemin suivi par une particule entre le 15 Mai et le 15 Juin à 20 mètres de profondeur en 1986, 1989, et 1992.

En 1989 et 1992 (Fig. 4.4 b et c), le transport, d'intensité moyenne mais d'orientation grossièrement parallèle aux accores, s'inverse également en l'espace de 3 mois. On notera par ailleurs que le transport semble nul pendant la deuxième période de 1989.

Enfin, en 1995 (Fig. 4.4 d), les transports atteignent des intensités importantes quelle que soit la période. Une zone de divergence apparaît vers 49°00N et les transports se font parallèlement aux accores au Sud comme au Nord de cette zone de divergence. En matière de transport, l'année 1995 est donc l'opposée de 1986.

Des résultats récents de modélisation (Bartsch et Coombs, 1996; Kloppmann *et al.*, 1996; Walsh *et al.*, 1996) fournissent également des estimations de transport. Le plus souvent ces transports sont mis en évidence par le suivi de particules au cours de longues périodes de 60 ou 90 jours de modélisation (spaghetti-plots). Le travail de Walsh *et al.* (1996) porte sur des simulations de transports de particules à 20 mètres de profondeur entre le 15 Mai et le 15 Juin des années 1986, 1989 et 1992 (pas 1995) dans la région des accores entre 47°00N et 53°00N. On sait par ailleurs (Coombs *et al.*, 1996) que la distribution verticale des œufs de maquereau a été, ces 20 dernières années, à peu près homogène sur les cinquante premiers mètres de la colonne d'eau.

Les résultats que nous obtenons, qui portent sur 6 jours de transport uniquement et qui font abstraction de toute considération sur la profondeur, sont donc potentiellement comparables avec ceux rapportés dans Walsh *et al.* (1996) et résumés (le plus fidèlement possible) par la figure 4.5.

Dans l'un et l'autre cas, les vitesses évoluent à peu près dans la même gamme de valeur allant d'un transport nul ou presque, à des déplacements de 4° de latitude en 30 jours, soit 30 cm/s. L'orientation des transports est beaucoup plus délicate à comparer et porte à plus de controverses. Je retiendrai les points d'accord et de désaccord suivants :

- Désaccord. Les transports vers le Nord que nous obtenons en début de saison 1989 et 1992 n'ont pas d'équivalent dans le modèle (premières parties des spaghetti-plots). Le transport vers le Nord en période 4 de 1986, bien que se déroulant à l'intérieur du plateau continental, n'existe pas dans les transports simulés (fins des spaghetti-plots ; Fig. 4.5 a). Plus généralement, nos estimations aboutissent à des vitesses de transport assez faibles en 1986, quand le modèle fournit des vitesses très importantes.
- Accord. Les deux modes d'estimations indiquent des transports vers le Sud, globalement parallèles aux accores, en périodes 2 et 3 de chaque année. En 1992, on notera toutefois que les transports estimés pour les deux points de départ voisins des accores à la latitude de 51°00N sont diamétralement opposés, l'un vers le Nord l'autre vers le Sud (Fig. 4.5 c). Enfin, l'absence de transport pendant la période 2 de 1989, correspond au courant circulaire indiqué par le modèle au milieu du Banc de Porcupine (Fig. 4.5 b).

Force est donc de constater qu'il est difficile d'arriver à des conclusions claires en matière de transport.

4.1.3 Remarques sur la température de surface

Malgré des différences dans la localisation moyenne des populations de E1 et E5, le centre de gravité de la température de surface reste identique pour les œufs de stade 1 et 5 (Fig. 4.6 a). L'écart maximum entre leurs centres de gravité est inférieur à 0.5°C. Par contre l'inertie de la température d'un œuf de stade 5 est, en moyenne, plus faible que celle d'un œuf E1 (Fig. 4.6 b); ce qui veut

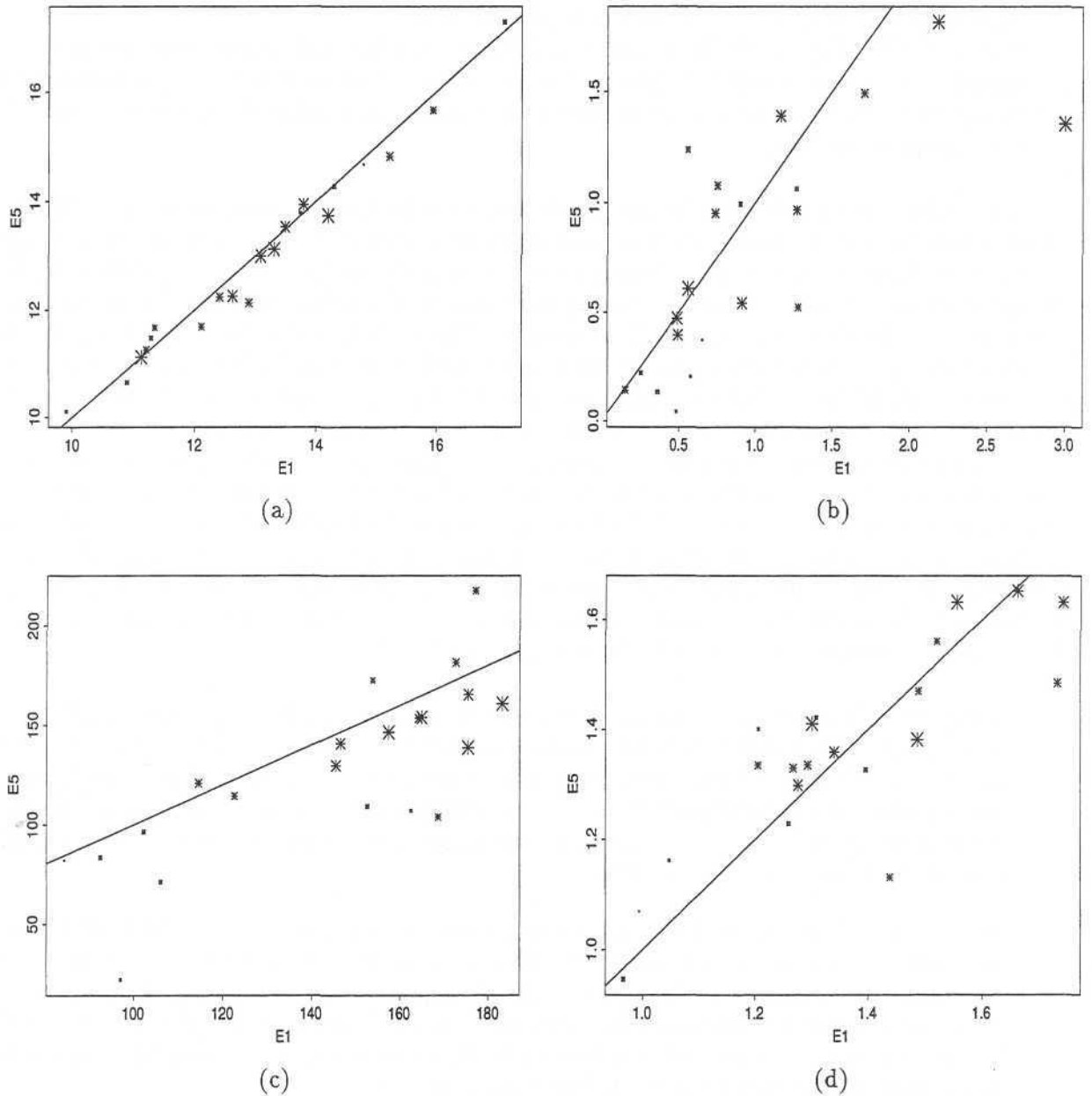


FIG. 4.6 - Corrélation entre statistiques par individu $E5$ et $E1$. Les étoiles ont une taille proportionnelle aux nombres d'échantillons utilisés pour les calculs. La ligne représente la première bissectrice.

(a) : centre de gravité de la température de surface.

(b) : inertie de la température de surface.

(c) : racine carrée de l'inertie de la population.

(d) : centre de gravité de $\log_{10}\left(\frac{\text{profondeur}}{20}\right)$

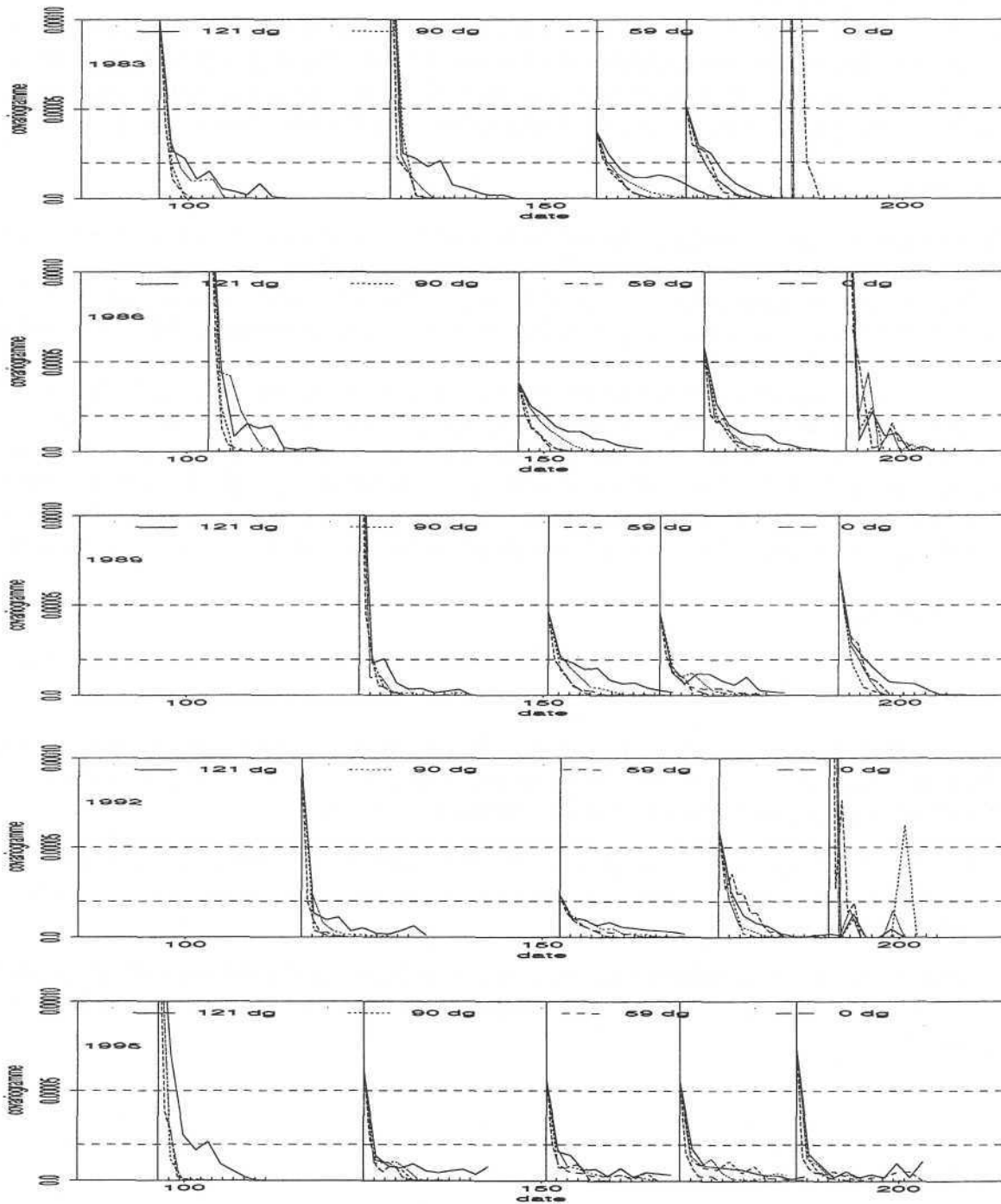


FIG. 4.7 - Covariogrammes expérimentaux ($g(h)/Q^2$) par période d'échantillonnage entre 1983 (en haut) et 1995 (en bas). Œufs de stade 5 (E5). Quatre directions de calcul (121. 90. 59 et 0°). Une graduation en abscisse représente 30 milles nautiques.

dire que les œufs de stade 5 sont plus concentrés autour de leur température moyenne que les œufs qui viennent d'être pondus.

A cela s'ajoute le fait que l'inertie de la distribution spatiale des œufs, c'est-à-dire l'inertie de la population, diminue en général avec le développement des œufs (Fig. 4.6 c). C'est donc qu'au cours de la première semaine de développement, les œufs, en moyenne, augmentent leur agrégation, et diminuent leur inertie; ce qui contribue à faire baisser l'inertie des températures par individu.

4.1.4 Structures spatiales

Des covariogrammes expérimentaux exprimés relativement aux carrés des abondances de E5 par période (Fig. 4.7), je retiendrai les quelques points suivants :

- l'anisotropie est assez nette. La direction de plus forte continuité spatiale est Nord-Nord-Ouest (121°). En milieu de saison, la portée dans cette direction est voisine de 600 milles nautiques (comme pour les œufs de stade 1).
- chaque année la continuité spatiale est meilleure au moment du pic d'abondance puisque c'est à ce moment là que la discontinuité à l'origine est la plus faible (très net en 1986 et 1992).
- en 1995, la valeur à l'origine et la discontinuité sont importantes (le reste du covariogramme se situe à un niveau assez bas). Les œufs E5 semblent particulièrement agrégés: les agrégats représentent une part importante de la structure (valeur à l'origine); il sont, en plus, d'extension spatiale très faible, tout au moins inférieure à la maille de prélèvement, soit 0.5° (discontinuité à l'origine).

4.1.5 Synthèse

L'analyse comparative des stades 1 et 5 du développement embryonnaire a finalement permis de mettre en évidence les points suivants :

- La mortalité embryonnaire augmente avec l'importance de la production d'œufs. En particulier le taux de survie pendant la première semaine peut passer de 30% les années à forte ponte (1992), à plus de 60% les années à faible ponte (1986).
- En moyenne, l'agrégation des œufs d'une semaine (stade 5) est supérieure à celle des œufs qui viennent d'être pondus (stade 1). L'anisotropie et la valeur des portées des covariogrammes sont similaires.
- Après une semaine de développement, le centre de gravité de la température de surface par individu est inchangé. L'inertie de la température de surface a, par contre, tendance à être plus faible.

4.2 Analyse du développement larvaire en 1989

Les densités de larves ne sont malheureusement pas disponibles en deçà de 48°00N et au delà de 55°00N. Pour assurer la cohérence de l'analyse comparative entre les statistiques par individu pour chacun des 5 stades de développement E1, E5, L1, L2, et L3, l'ensemble des statistiques relatives aux stades E1 et E5 ont été recalculées en ne conservant que les données situées entre 48°00N et 55°00N.

4.2.1 Courbes de production et mortalité totale

Conformément aux durées estimées de chacun des stades de développement E5, L1, L2 et L3 (Cf. §IV.2.2.3), les dates de début et de fin de production de chacun d'eux ont été obtenues en ajoutant respectivement 6, 11, 21 et 35 jours aux dates estimées de début et fin de ponte. Il aurait semblé en effet bizarre de faire commencer et finir la production de larves en même temps que celle de la ponte.

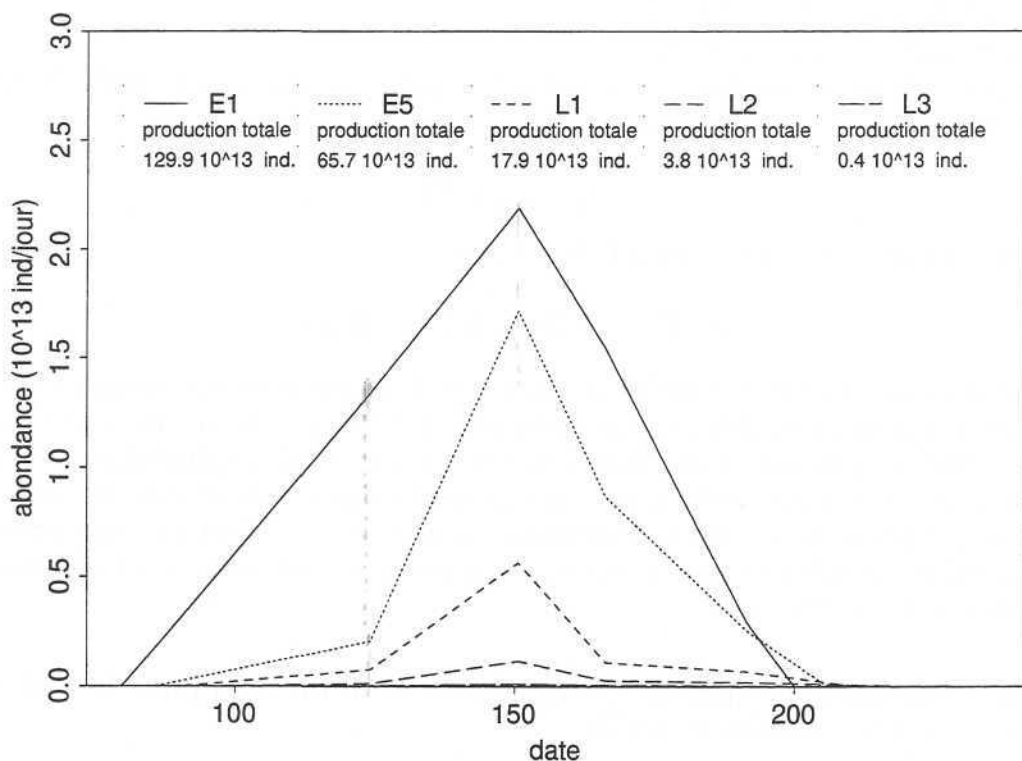


FIG. 4.8 - Courbes de productions des différents stades de développement (abondances en fonction de la date).

La quantité totale d'œufs pondus entre 48°00N et 55°00N est $130 \cdot 10^{13}$ œufs. Cette valeur est très proche de celle obtenue en utilisant l'ensemble des échantillons. La restriction des calculs aux latitudes comprises entre 48°00N et 55°00N, n'a donc pas occasionné ici de biais important.

En intégrant chacune des courbes de production (Fig. 4.8), et en faisant le ratio des abondances totales de deux stades consécutifs, on trouve que :

- 50% des œufs pondus se transforment en œufs de stade 5,
- 27% des œufs de stade 5 se transforment en larve L1,
- 21% des larves L1 se transforment en larves L2
- 10% des larves L2 se transforment en larves L3 tailles,
- au total 0.3% des œufs pondus se transforment en larves de grandes tailles (L3).

$$E1 \xrightarrow{0.50\%} E5 \xrightarrow{0.27\%} L1 \xrightarrow{0.21\%} L2 \xrightarrow{0.10\%} L3$$

$$E1 \xrightarrow{0.003\%} L3$$

Le taux de mortalité est obtenu en rapportant ces écarts d'abondance à la durée de chacun des stades, et plus précisément au temps qui s'écoule entre les dates milieux de chacun des stades. Ce laps de temps est de :

- 4.7 jours (113 heures) entre les stades E1 et E5
- 2.9 jours (69 heures) entre les stades E5 et L1
- 7.5 jours entre les stades L1 et L2
- 12 jours entre les stades L2 et L3.

En supposant le taux de mortalité Z constant par stade, l'abondance d'un stade $i + 1$ est reliée à celle du stade précédent i par la relation :

$$Q_{i+1} = Q_i e^{-Z \cdot t}$$

Les taux de mortalité journaliers sont alors les suivants :

$$E1 \xrightarrow{15\%} E5 \xrightarrow{45\%} L1 \xrightarrow{21\%} L2 \xrightarrow{19\%} L3$$

On voit ainsi qu'en 1989 la mortalité des stades de développement pré-larvaires (15%) est du même ordre de grandeur que celle des stades larvaires (21 et 19%). Elle est par contre **beaucoup plus forte (45%) au moment de l'éclosion**. Cette étape du développement semble donc très délicate et source de nombreux échecs. La légère augmentation de la mortalité entre œufs et larves s'explique en partie par le fait que les réserves alimentaires des embryons (poches vitellines) sont quasiment épuisées au début des phases larvaires. Les larves, contrairement aux embryons, doivent trouver à manger pour survivre.

4.2.2 Comparaisons de quelques statistiques individuelles pour la distinctions entre stades passifs et actifs

Localisation moyenne des individus

Pour chaque stade de développement, la distribution spatiale des individus par période est représentée par leur centre de gravité entouré d'un cercle de taille proportionnelle à l'abondance, et les axes principaux d'inertie (Fig. 4.2.2).

Au moment des pics de production (plus grands cercles sur la figure 4.9) les centres de gravité de E1, E5 et L1 sont positionnés de plus en plus à l'Ouest des accores. En revanche, la position moyenne des deux stades suivants L2 et L3 tend à revenir vers les accores, de sorte que le centre de gravité des larves L3 est très proche de celui des œufs E1.

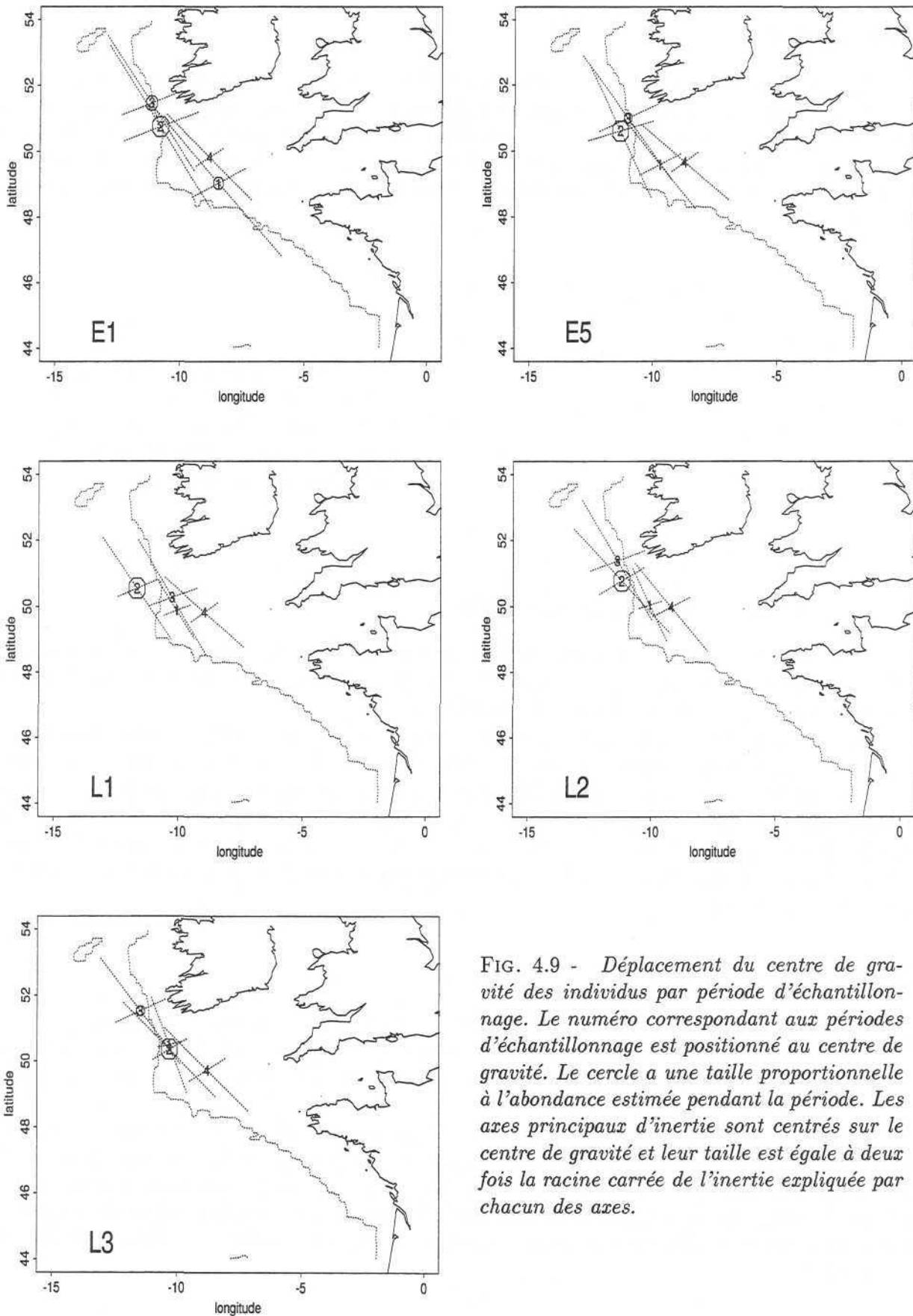


FIG. 4.9 - Déplacement du centre de gravité des individus par période d'échantillonnage. Le numéro correspondant aux périodes d'échantillonnage est positionné au centre de gravité. Le cercle a une taille proportionnelle à l'abondance estimée pendant la période. Les axes principaux d'inertie sont centrés sur le centre de gravité et leur taille est égale à deux fois la racine carrée de l'inertie expliquée par chacun des axes.

Inertie par stade

Quel que soit le stade de développement, l'inertie reste principalement expliquée par une direction parallèle aux accores (Fig. 4.2.2). La taille des axes principaux semble cependant diminuer avec le développement. Cette diminution concerne en fait les trois premiers stades de développement (E1, E5 et L1), comme l'indique les courbes d'évolution de la racine carrée de l'inertie pour chacun des stades (Fig. 4.10). Passé le stade L1, l'inertie de la localisation des individus a tendance à augmenter.

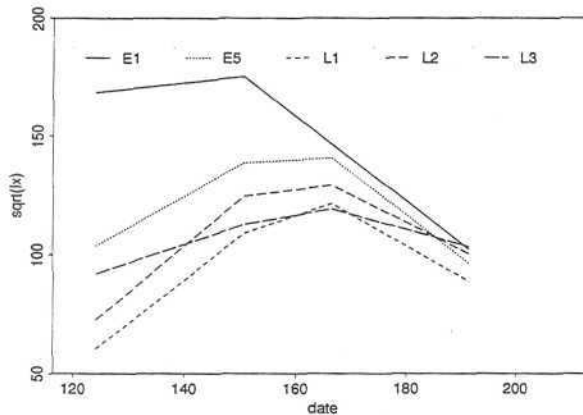


FIG. 4.10 - Statistiques par individu en fonction de la date (jour julien) pour différents stades de développements en 1989. Racine carré de l'inertie de la population.

Densités moyennes et surfaces équivalentes

Du premier (E1) au dernier stade (L3), la densité moyenne par individu chute régulièrement (Fig. 4.11 a). En milieu de saison, elle passe de 270 E1/m²/jour à 1.5 L3/m²/jour, ce qui revient à dire que la densité moyenne a été divisée par 180.

Dans le même temps, la surface équivalente n'a été divisée que par 2 (Fig. 4.11 b). La diminution des quantités totales de stade en stade est donc essentiellement due à la chute des densités moyennes.

Plus précisément, la surface équivalente n'a été divisée par deux qu'entre les stades E1 et L1, passant de 28 000 mn² à 15 000 mn². De la ponte au début du développement larvaire, **l'agrégation des individus a donc doublé**. Au-delà, c'est-à-dire pour les phases ultérieures du développement larvaire (L2 et L3), la surface équivalente augmente à nouveau jusqu'à retrouver le niveau de départ (i.e. celui des œufs).

Température de surface

Dans le prolongement des résultats obtenus en comparant les stades E1 et E5 entre 1983 et 1995, le centre de gravité de la température de surface est identique, à un instant donné, pour tous les stades de développement (Fig. 4.11 c). A un instant donné, l'écart maximum entre les différents centres de gravité est inférieur à 0.5°C.

L'inertie de la température de surface par individu présente par contre des différences importantes entre stades (Fig. 4.11 d). Elle diminue rapidement entre les trois premiers stades E1, E5 et L1. En particulier, en milieu de saison (fin Mai), **l'inertie de la température des larves L1 est 7 fois plus petite que celle des œufs E1**. Là encore, dans la suite du développement larvaire (stades L2 et L3), l'inertie de la température augmente, "au lieu de" continuer à diminuer par exemple.

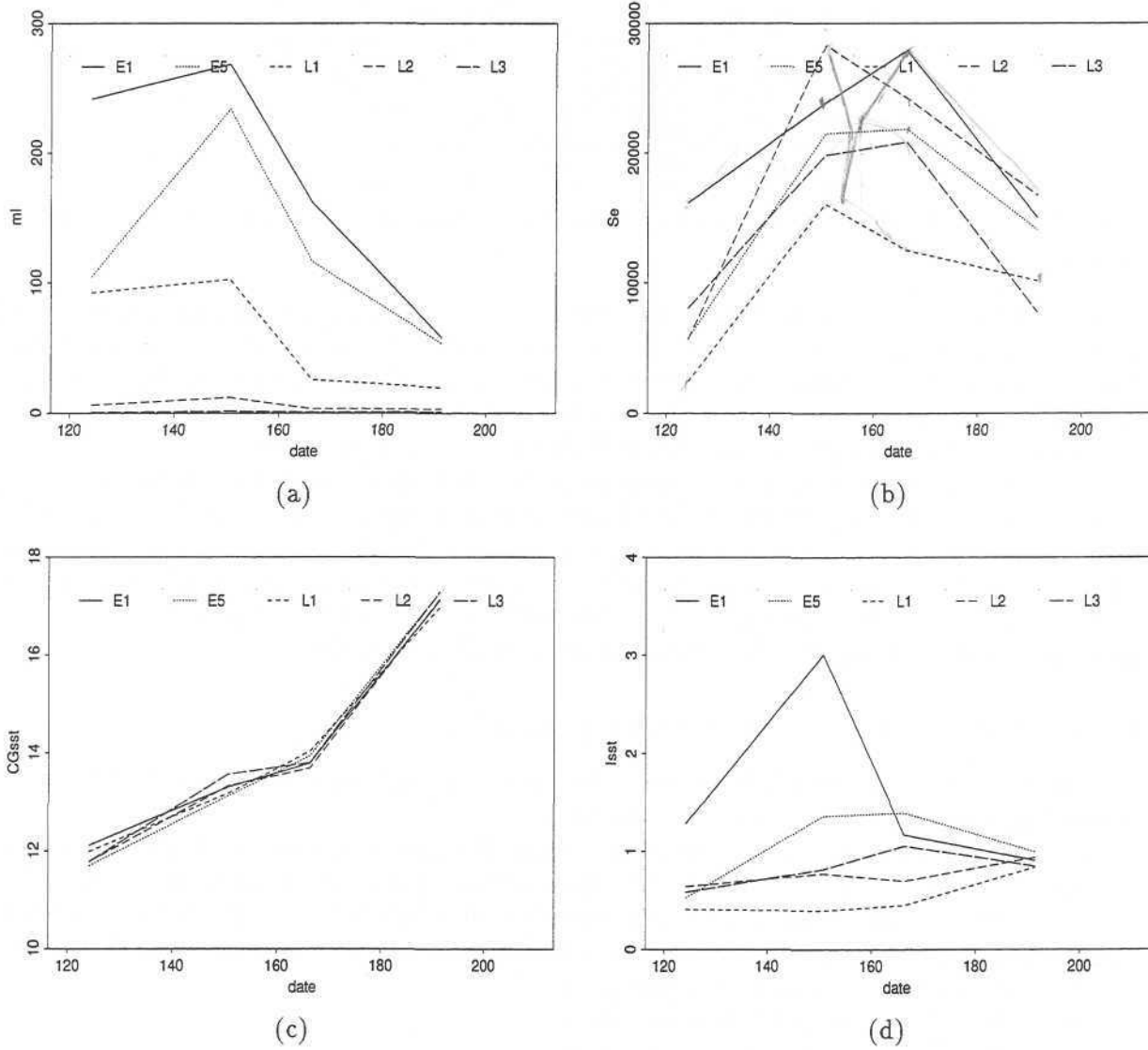


FIG. 4.11 - Statistiques par individu en fonction de la date (jour julien) pour différents stade de développements en 1989. (a) Densité moyenne par individu; (b) Surface équivalente; (c) Centre de gravité de la température de surface par individu; (d) Inertie de la température de surface par individu

Distinction entre stades passifs et actifs

L'ensemble de ces résultats peut être résumé dans le tableau suivant :

statistique	E1 + E5 + L1	L2 + L3
agrégation	↑	↓
localisation moyenne	W ←	→ E
inertie des localisations	↓	↑
SST moyenne	↔	↔
inertie de SST	↓	↑

TAB. 4.1 - Evolution des statistiques individuelles au cours du développement embryonnaire et larvaire.

L'évolution des statistiques individuelles constitue donc un faisceau de raisons pour associer les stades E1 à L1 d'un côté et les stades L2 et L3 d'un autre côté. Deux des trois premiers stades sont bien évidemment passifs, c'est-à-dire incapables de nager. La très petite taille, l'anatomie et la similarité des statistiques des larves L1 avec celles des œufs, semblent indiquer que les larves, au début de leur développement ne sont pas véritablement capables de nager.

Par contre, le retour vers l'Est du centre de gravité de la localisation des individus L2 et L3 et l'augmentation de leur agrégation tendent à prouver que les larves de plus de 4-5 mm sont douées d'une certaine activité. On sait déjà, par des observations en bassin, que les individus de cette taille sont capables de nager. Ce qui n'est pas établi, c'est leur capacité à nager dans l'environnement marin naturel, c'est-à-dire leur capacité à lutter contre le courant. Or on dispose ici d'éléments qui montrent que les larves de 5 mm environ cessent d'être passives.

4.2.3 Transports à moyen terme (2-3 semaines)

L'estimation du transport à l'aide des centres de gravité de la localisation des différents stades de développement nécessite quelques précautions.

En 1989, les dates moyennes des périodes d'échantillonnage 2, 3, 4 et 5 sont espacées de 27, 15 et 25 jours respectivement (voir Tab. 2.1). Cela correspond grossièrement au temps nécessaire à un œuf pour atteindre les stades L2, L1 et à nouveau L2 (voir Tab. 2.3). L'estimation du transport à moyen terme devra donc logiquement porter sur la comparaison entre les centres de gravité de :

- E1 de la période 2 et de L2 de la période 3
- E1 de la période 3 et de L1 de la période 4
- E1 de la période 4 et de L2 de la période 5

Parmi ces comparaisons, seule la deuxième apparaît réellement pertinente dans la mesure où elle ne met en jeu que des stades passifs, ou tout du moins jugés comme tels. On peut toutefois considérer que les centres de gravité des larves L2, même actives, sont utilisables ici.

Pour tenir compte de la forme des distributions spatiales, les centres de gravité sont comptés dans un repère centré sur les accores. Enfin, puisque les comparaisons concernent des périodes d'échantillonnage différentes, il convient de suivre la position moyenne des échantillons en même temps que celles des individus. Connaissant le déplacement des individus ainsi que celui des échantillons on peut alors proposer l'estimation d'un transport net correspondant mathématiquement au

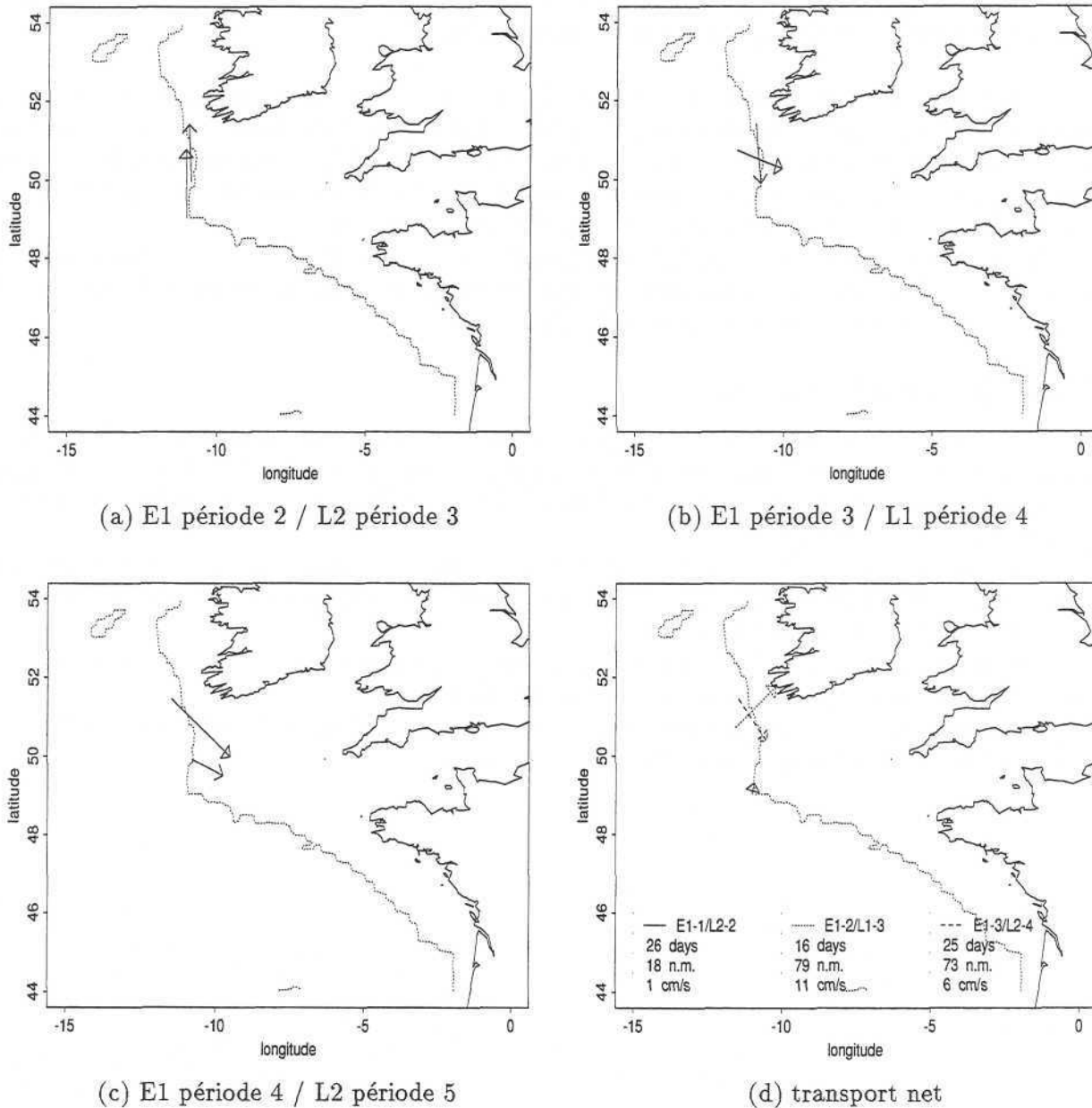


FIG. 4.12 - Estimation du transport pendant le développement des œufs de stade I (E1) en larves. Comparaison des positions du centre de gravité de E1 pour une période "i" et de celui de des larves de stade I ou II pour la période "i+1" selon le laps de temps qui sépare les deux périodes d'échantillonnage. Les flèches dont la pointe est fermée indiquent le déplacement du centre de gravité des individus. Les flèches dont la pointe est ouverte indiquent le déplacement du centre de gravité des échantillons. Estimation du transport net: déplacement du centre de gravité des individus moins celui des échantillons.

déplacement des individus moins celui des échantillons. Les vitesses de transport sont alors simplement obtenues par le rapport entre le module du transport net (en cm) et le temps écoulé entre les périodes d'échantillonnage correspondantes (en secondes).

Les transports bruts (flèches fermées Fig. 4.12 a, b et c) sont cohérents avec les estimations faites préalablement avec les œufs de stade 1 et 5 (Fig. 4.4 b). Cependant, une part importante des déplacements observés sur les individus se retrouve dans le déplacement des échantillons. Cette constatation fragilise les résultats, sans pour autant que l'on puisse statuer sur l'incidence réelle des modifications d'échantillonnage sur l'estimation des transports.

Les transports nets ont de ce fait des intensités et des orientations assez surprenantes. En particulier, le transport net estimé entre les périodes 3 et 4 est dirigé vers le Nord-Est alors que les déplacements bruts vont vers le Sud ou le Sud-Est.

4.2.4 Structures spatiales

L'évolution des surfaces équivalentes ayant déjà été commentée, les structures spatiales expérimentales sont représentées sans relier l'origine au premier pas de calcul (Fig. 4.13). En s'attachant à comparer les structures spatiales des différents stades de développement en milieu de saison de reproduction, on remarque que :

- les portées, et en particulier la portée la plus longue (600 mn dans la direction 121°), ne changent pratiquement pas au cours du développement. L'anisotropie de la population est donc maintenue au cours du développement.

- relativement au carré de l'abondance, la valeur du covariogramme au premier pas de calcul, est voisine de 0.000025 quel que soit le stade de développement.

En somme les structures spatiales ne sont pas nettement différentes d'un stade à l'autre, en dehors de la valeur à l'origine (agrégation).

Attention, les résultats ne sont pas cohérents avec la figure 3.11. p198
 En particulier pour la valeur à l'origine.
 -> Peut-être à cause des limitations en latitude

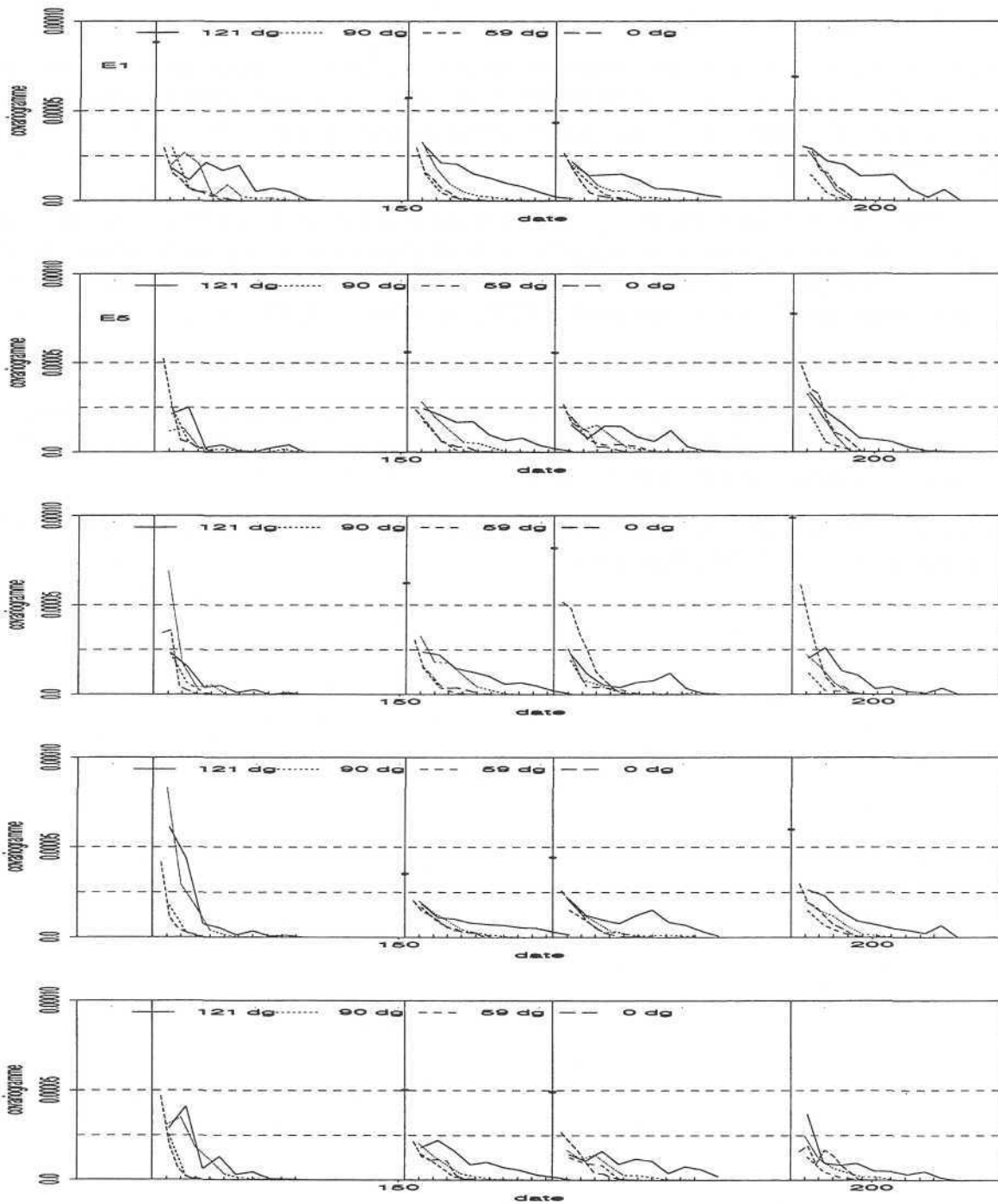


FIG. 4.13 - Covariogrammes expérimentaux par période exprimés relativement au carré de l'abondance de chaque stade. Quatre directions de calcul (121, 90, 59 et 0°). Stades E1 (en haut) à L3 (en bas). Une graduation en abscisse représente 30 milles nautiques. Les deux niveaux 0.00005 et 0.000025 sont représentés par des lignes horizontales.

4.2.5 Synthèse

L'essentiel des changements entre stades que nous avons mis en évidence portent sur les statistiques individuelles non spatiales (centre de gravité et inertie). Les covariogrammes expérimentaux sont apparus en effet assez stables au cours du développement, sauf en 0. Pour le reste, on peut retenir les points suivants :

- en 1989, 10 000 œufs ont donné en moyenne naissance à 3 larves de 6 à 13 mm. La décomposition de cette mortalité totale au cours du développement embryonnaire et larvaire montre que le taux journalier de mortalité est très élevé au moment de l'éclosion (45%) et qu'il augmente légèrement entre les stades pré-larvaires (15%) et larvaires (20%).
- la température moyenne d'un individu est invariante d'un stade de développement à l'autre.
- pendant la phase passive (larves inférieures à 4 mm), l'agrégation des individus augmente, leur inertie diminue et la température de surface à l'endroit des individus est de moins en moins dispersée autour du centre de gravité.
- les larves de plus de 4 mm commencent à avoir une nage active comme en témoigne, par exemple, la baisse de leur agrégation.

Conclusion

En généralisant le principe très simple qui consiste à diviser certains des outils existants de la statistique et de la géostatistique par l'abondance d'une population, nous avons vu que nous disposons d'une méthodologie dotée d'un sens physique clair, donnant en particulier ce que j'ai appelé le point de vue de l'individu, et d'un pouvoir descriptif fort adapté à l'analyse de données spatialisées.

Dans les applications halieutiques que nous avons présentées, nous avons commencé par voir que le covariogramme est un outil structural performant. Evitant toute hypothèse sur l'homogénéité de la variable (ce qui n'est pas le cas du variogramme), il permet de résumer la structure spatiale des individus d'une population, même si l'échantillonnage n'est pas régulier (situation pour laquelle nous avons proposé plusieurs estimations possibles). La modélisation de la structure expérimentale a permis ensuite, en limitant le nombre d'hypothèses, de réaliser une cartographie par krigeage transitif. Les principes de ce krigeage ont été mis en place par Matheron en 1968 mais n'ont été que très peu utilisés jusque là (l'utilisation d'un modèle de covariogramme pour le calcul de variance d'estimation globale lui volant, en quelque sorte "la vedette"). J'espère que ce travail contribuera à développer l'utilisation de cette technique simple de cartographie.

Relativement au carré de l'abondance, le covariogramme a par ailleurs permis de dégager un certain nombre de caractéristiques persistantes du processus de reproduction du maquereau. La notion de surface équivalente, reliée à la valeur du covariogramme à l'origine, bien qu'un peu abstraite au départ, s'est imposée comme un bon critère pour caractériser la distribution des œufs à un instant donné d'une saison de reproduction. L'utilisation d'outils très simples comme les centres de gravité s'est révélée, par ailleurs, assez efficace pour quantifier des principes fondamentaux en écologie comme la densité dépendance du taux de survie au cours des premières semaines du développement d'un œuf.

Enfin, ce travail a initié une nouvelle façon d'appréhender le concept de *preferendum* en écologie halieutique. Au début de cette thèse, la géostatistique multivariable "classique", basée sur l'utilisation de variogrammes croisés, est rapidement apparue inappropriée. Dans le même temps, une réflexion sur l'interprétation des régressions expérimentales dans un contexte spatial nous a conduit à suggérer une approche dynamique pour l'étude de *preferendum*. En outre, pour ce qui concerne la température de surface, il est apparu pertinent, et opératoire, de chercher à savoir, par des translations en bloc, si les individus d'une population étaient distribués de manière à minimiser l'inertie des températures. Par cette technique, encore appelée *inertiogramme* (ou *inertie translatée*), nous avons pu montrer que les maquereaux, au moment de la reproduction, semblaient sensibles à un front virtuel d'inertie. Ces résultats suggèrent de plus que la dynamique d'évolution des maquereaux est alors liée, non pas à la température elle-même, mais au gradient local de la température de surface.

La géostatistique multivariable s'est donc probablement enrichie d'une brique supplémentaire. Mais un travail important reste à faire tant les solutions méthodologiques offertes pour traiter de questions faisant intervenir plusieurs variables sont peu nombreuses.

Enfin, au cours de ce document, quelques prospectives possibles ont été mentionnées. J'ai indiqué en effet que certains des concepts de la dynamique des populations font référence, comme les statistiques individuelles, à un individu tiré au hasard dans la population. Peut-on envisager que derrière cette communauté d'expression, se cache la possibilité d'une dynamique des populations géostatistique?

Bibliographie

Bibliographie

- [1] Anon., 1987. Report of the mackerel egg production workshop. ICES CM 1987/H:2, 58 p.
- [2] Anon., 1990a. Report of the mackerel / horse mackerel egg production workshop. ICES CM 1990/H:2, 89 p.
- [3] Anon., 1990b. Report of the mackerel working group. ICES CM 1990/Assess: 19, 109 p.
- [4] Anon., 1993. Report of the mackerel / horse mackerel egg production workshop. ICES CM 1993/H:4, 142 p.
- [5] Anon., 1994. Report of the mackerel / horse mackerel egg production workshop. ICES CM 1994/H:4, 58 p.
- [6] Anon., 1996. Report of the working group on mackerel and horse mackerel egg surveys. ICES CM 1996/H:2, 146 p.
- [7] Bartsch J, S. Coombs, 1996. A numerical model of the dispersion of blue whiting larvae (*Micromesistius poutassou*) in the western North Atlantic. ICES CM 1996/S:40.
- [8] Bez N., J. Rivoirard, M. Walsh, 1995. On the relation between fish density and sea surface temperature, application to mackerel egg density. ICES CM 1995/D:6 Ref. C,H,L.
- [9] Bez N., J. Rivoirard, M. Walsh, 1995. Approche transitive et densités de poissons. *Cahiers de Géostatistique*, Fasc. 5, pp. 161 - 177.
- [10] Bez N., J. Rivoirard, Ph. Guiblin, M. Walsh, 1996. Covariogram and related tools for fish survey data. Fifth International Geostatistics Congress, Wollongong, Australia (*In press*).
- [11] Chauvet P., 1994. Aide-mémoire de géostatistique linéaire. *Cahiers de géostatistique*, Fascicule 2, ENSMP, France, 146 p.
- [12] Coombs S.H., R.K. Pipe, C.E. Mitchell, 1981. The vertical distribution of eggs and larvae of blue whiting (*Micromesistius poutassou*) and mackerel (*Scomber scombrus*) in the eastern North Atlantic and North Sea. Rapp. P.-v. Reun. Cons. int. Explor. Mer, 178, pp. 188-195.
- [13] Coombs S.H., D. Morgans, N.C. Halliday, 1996. The vertical distribution of eggs and larvae of mackerel (*Scomber scombrus*). ICES CM 1996/S:47.
- [14] Corten A., 1983. Learning process in herring migrations. ICES/CIEM CM 1993/H:18, 8 p.
- [15] Crépon M., 1993. Initiation à la dynamique de l'océan. *Océanis*, Volume 19, Fascicule 2, 110 p.

- [16] Dardignac J., 1988. Les pêcheries du Golfe de Gascogne - Bilan des connaissances. Rapports scientifiques et techniques de l'IFREMER, n° 9.
- [17] Dawson W.A., 1986. Changes in the western mackerel (*Scomber scombrus*) spawning stock composition during the spawning season. *J. Mar. Biol. Ass.*, UK, **66** : pp. 367-247.
- [18] DeAngelis D.L., M.J. Gross (Eds), 1992. Individual based models and approaches in ecology. Chapman & Hall.
- [19] Eltink A., 1987. Changes in the age-size distribution and sex ratio during spawning and migration of western mackerel (*Scomber scombrus*). *J. Cons. int. Explor. Mer.* 44 : pp. 10-22.
- [20] Escoffier B. et J. Pagès, 1990. Analyses factorielles simples et multiples. Objectifs, méthodes et interprétation. Ed. : Dunod, Paris, 267 p.
- [21] Gascuel D., 1993. Efforts et puissances de pêche : redéfinition des concepts et exemple d'application. Premier Forum Halieumétrique, Rennes. In : Les recherches françaises en évaluation quantitative et modélisation des ressources et des systèmes halieutiques, Colloques et séminaires, ORSTOM Ed., 405 p.
- [22] Gohin F., 1987. Analyse géostatistique des champs thermiques de surface de la mer. Thèse Dr. en Géostatistique, ENSMP, France.
- [23] Guiblin Ph., J. Rivoirard et E.J. Simmonds, 1995. Analyse structurale de données à répartitions dissymétriques : exemple du hareng Ecossais. *Cahiers de Géostatistique*, Fasc. **5**, pp. 137 - 159.
- [24] Journel A.G., 1977. Géostatistique minière. Centre de Géostatistique, Fontainebleau, France, 736 p.
- [25] Koudriavsteva T., 1993. Analyse géostatistique des données de chalutage. Rapport de stage S-317 CONFIDENTIEL, Centre de Géostatistique, ENSMP, France, 15 p.
- [26] Kloppmann M., C. Mohn, J. Bartsch, 1996. The effect of currents and hydrography on the distribution of blue whiting eggs and larvae on Porcupine Bank. ICES 1996/S:25.
- [27] Lantuéjoul Ch., 1990. Cours de sélectivité. Cours C-140, Centre de Géostatistique, Fontainebleau, France.
- [28] Lantuéjoul Ch., 1993. Ensembles aléatoires. Cours C-149, Centre de Géostatistique, ENSMP, France, 37 p.
- [29] Laurec A., J.C. Le Guen, 1981. Dynamique des populations marines exploitées. Tome I : Concepts et Modèles. Rapports scientifiques et techniques n° 45-1981, Centre National pour l'EXploitation des Océans, 118 pp.
- [30] Lockwood S.J., J.H. Nichols, S. Coombs, 1977. The development rates of mackerel (*Scomber scombrus*) eggs over a range of temperatures. ICES CM 1977/J:13, 5 p.
- [31] Lockwood S.J., J.H. Nichols, W.A. Dawson, 1981. The estimation of a mackerel (*Scomber scombrus*) spawning stock size by plankton survey. *J. Plankton Res.* **3**(2), pp. 217-233.

- [2] Matheron G., 1963. Les schémas stochastiques intrinsèques et leurs relations avec les schémas transitifs. Note Géostatistique N 47. Juin 1963. N 45, Centre de Géostatistique, ENSMP, Fontainebleau, France, 89 p.
- [33] Matheron G., 1965. Les variables régionalisées et leur estimation. Une application de la théorie des fonctions aléatoires aux sciences de la nature. Masson et Cie Ed., Paris (VIème), 305 p.
- [34] Matheron G., 1968. Le krigeage en représentations transitives. N-67, Février 1968, Centre de Géostatistique, ENSMP, Fontainebleau, France.
- [35] Matheron G., 1970. La théorie des variables régionalisées et ses applications. Les Cahiers du Centre de Morphologie Mathématique, Fascicule 5, ENSMP, France, 212p.
- [36] Matheron, G., 1973. The intrinsic random functions and their applications. *Adv. Appl. Prob.*, Vol. 11, n°1, pp. 184 - 89.
- [37] Matheron G., 1978. Estimer et choisir. Les cahiers de morphologie mathématique, Fascicule 7, ENSMP, France, 175 p.
- [38] Matheron, G., 1989. Estimating and choosing. An essay on probability in practice. Ed.: Springer-Verlag, Berlin, 141 p.
- [39] Ottersen G., K. Michalsen, O. Nakken, 1996. Ambient temperature and distribution of northeast arctic cod. ICES CM 1996/G:13.
- [40] Perry R.I., S.J. Smith, 1994. Identifying habitat associations of marine fishes using survey data: an application to the Northwest Atlantic. *Can. J. of Fish. and Aqua. Sc.* 51: pp. 589-602.
- [41] Petitgas P., 1991. Contributions géostatistique à la biologie des pêches maritimes. Thèse Dr. en Géostatistique, ENSMP, France. 211 p.
- [42] Poulard, 1990. Evaluation des ressources halieutiques de l'Ouest de l'Europe (EVHOE) par campagnes de chalutages programmés. Deuxième phase. Rapport IFREMER 88/1210834/BF.
- [43] Rivoirard J., 1995. Concepts et méthodes de la géostatistique. Cours C-158, Centre de Géostatistique, ENSMP, Fontainebleau, France, 95 p.
- [44] Rivoirard J., Ph. Guiblin, 1996. Global estimation variance in presence of conditioning parameters. Fifth International Geostatistics Congress, Wollongong, Australia (*In press*).
- [45] Saporta G., 1990. Probabilités, analyses des données et statistique. Technip Ed., Paris, 493 p.
- [46] Walsh M., J.H.A. Martin, 1986. Recent changes in the distribution and migrations of the western mackerel stock in relation to hydrographic changes. ICES CM 1986/H:17.
- [47] Walsh M., M. Skogen, D.G. Reid, E. Svendsen, J. Ac. McMillan, 1996. The relationship between the location of western mackerel spawning, larval drift and recruit distributions: a modelling study. ICES CM 1996/S:33.
- [48] Watanabe Y., H. Zenitani, R. Kimura, 1995. Offshore expansion of spawning of the Japanese sardine, *Sardinops melanostictus*, and its implication for egg and larval survival. *Can. J. Fish. Aquat. Sci.*, 53: pp 55 - 61.

Annexes

Annexe 1 : Analyse structurale d'un courant marin

L'objet de cette annexe est de fournir quelques éléments de réflexion sur les aspects théoriques et pratiques de l'analyse structurale d'un vecteur. Cette partie est inspirée d'un travail réalisé au centre de géostatistique sur le courant marin du bord du plateau continental portugais mesuré par des courant-mètres (Rivoirard, 1995; non publié).

1. Nature des difficultés

1.1 Rappels préalables

Plaçons nous directement dans le cadre de la géostatistique intrinsèque. Un champ de vecteurs est alors considéré comme la réalisation d'une fonction aléatoire vectorielle $\vec{V}(x)$ où x est un point de l'espace. Ce qui revient à considérer qu'en chacun des points de l'espace, la valeur de la fonction est un vecteur aléatoire, et non plus une variable aléatoire comme dans le cas d'une fonction aléatoire. Ses composantes $V_1(x)$, $V_2(x)$ et $V_3(x)$, selon un repère orthonormé de notre choix à trois dimensions d'axes u_1, u_2, u_3 sont donc des fonctions aléatoires :

$$\vec{V}(x) = \begin{bmatrix} V_1(x) \\ V_2(x) \\ V_3(x) \end{bmatrix}$$

Une première approche consiste alors à traiter l'analyse de $\vec{V}(x)$ sous l'angle d'un problème multivariable. Dans le cadre stationnaire, l'espérance de $\vec{V}(x)$, qui ne dépend pas de x , est donnée par :

$$E[\vec{V}(x)] = \begin{bmatrix} E[V_1(x)] \\ E[V_2(x)] \\ E[V_3(x)] \end{bmatrix}$$

Ainsi, la stationnarité de la fonction aléatoire vectorielle $\vec{V}(x)$ s'applique à chacune des composantes. C'est ce qui en fait une hypothèse très forte d'ailleurs. Des exemples de champ de vecteurs stationnaire et non stationnaire sont donnés figure 4.14. Par ailleurs, la covariance de $\vec{V}(x)$ correspond en fait à une matrice de variance-covariance de dimension 3×3 , comprenant 3 fonctions de covariances $C_{ii}(h)$ et 6 fonctions de covariances croisées $C_{ij}(h)$ qui ne dépendent, dans le cadre stationnaire, que de la distance h entre les composantes et non de leurs positions dans le champ :

$$C(h) = Cov[\vec{V}(x), \vec{V}(x+h)] = \begin{bmatrix} C_{11}(h) & C_{12}(h) & C_{13}(h) \\ C_{21}(h) & C_{22}(h) & C_{23}(h) \\ C_{31}(h) & C_{32}(h) & C_{33}(h) \end{bmatrix}$$

Une situation stationnaire isotrope, dont je fournis un exemple (Fig. 4.14), correspond à une matrice de variance-covariance identique quelle que soit la direction, c'est-à-dire ne dépendant que du module du vecteur distance h .

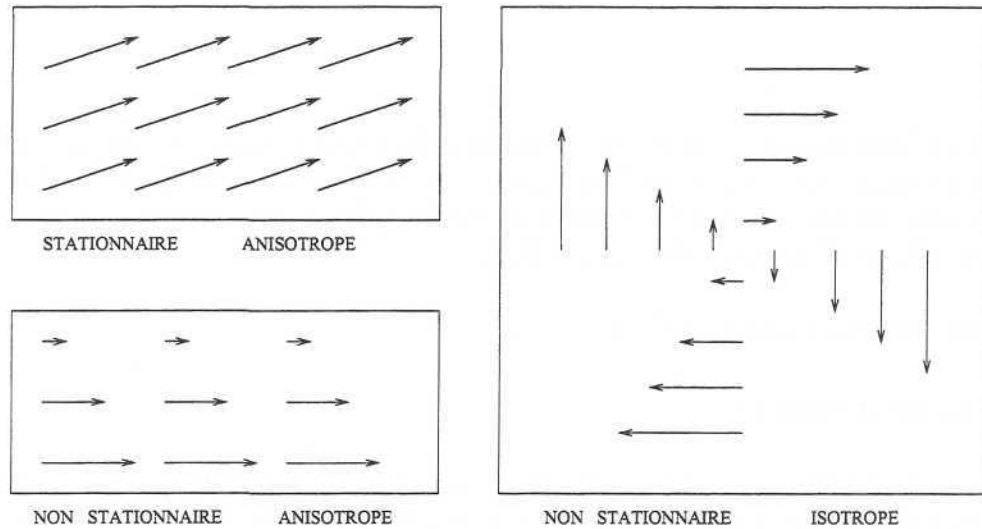


FIG. 4.14 - Exemples de champ de vecteurs.

Si l'hypothèse stationnaire ne convient pas, le cadre moins contraignant de l'hypothèse intrinsèque peut être envisagé. Il conduit à des systèmes équivalents en remplaçant les covariances simples et croisées par les variogrammes correspondants. On pourra noter qu'il est toutefois possible d'employer les variogrammes dans le cadre stationnaire afin d'éviter l'inférence toujours un peu délicate des moyennes nécessaires aux centrages des covariances. Cependant, le variogramme croisé, qui est une fonction symétrique, ne permet pas de décrire les corrélations avec autant de finesse que ne le font les covariances croisées non symétriques.

1.2 Cohérence du modèle structural

Je me contenterai ici de mentionner deux difficultés notoires dans le choix d'un modèle structural.

Il faut tout d'abord respecter la cohérence interne du modèle mathématique. Ce qui, en géostatistique multivariable, conduit à contraindre les paliers des covariances simples et croisées à respecter certaines inégalités (Wackernagel, 1995).

Ensuite, il est souhaitable que le modèle structural soit compatible avec les propriétés physiques du champ de vecteur. En particulier, n'importe quel modèle structural n'est pas compatible avec un gradient; des critères de dérivabilité interviennent ici (Castelier, 1995). On sait par ailleurs en océanographie physique, que, sous l'hypothèse que la densité est constante, la divergence d'un courant est nulle. Il faudrait donc connaître les modèles compatibles avec ce type de propriété (divergence nulle).

1.3 Interprétabilité

Prise dans son ensemble, la matrice de variance-covariance est une information indépendante du repère. On ne comprendrait pas en effet, que la structure spatiale dépende du repère.

Il n'en reste pas moins que la valeur, la signification et l'interprétation de chaque covariance (simple ou croisée) dépendent du choix du système d'axes. Pour s'en convaincre, prenons par exemple la première composante, $V_1(x)$. Cette composante est une entité directionnelle qui représente la partie du phénomène qui est "expliquée" par la direction de l'espace donnée par le premier axe u_1 . On doit donc s'attendre à ce que la direction portée par u_1 soit une direction privilégiée pour cette composante. La covariance $C_{11}(h)$ calculée dans la direction u_1 ne peut donc pas être comparée à la covariance obtenue dans une autre quelconque des directions de l'espace. En somme, ce serait un non-sens de comparer les valeurs expérimentales de la covariance d'une composante vectorielle obtenue dans la direction colinéaire à cette composante et dans une autre direction.

L'ensemble de ces remarques reste vrai pour l'analyse de la corrélation entre une variable vectorielle et une variable scalaire. Cela signifie entre autre, que l'analyse du lien entre une densité de poissons et un gradient de température ne peut se résumer simplement à une analyse multivariable; sauf à supposer, comme cela est d'ailleurs généralement fait, que le gradient n'intervient que par l'intermédiaire de son module et non de sa direction.

Pour une analyse structurale directe, il apparait donc préférable soit d'employer des outils indépendants du choix du repère, soit de prendre la direction de calcul comme axe de référence pour la décomposition du vecteur

2. Outils structuraux pour des vecteurs

2.1 Covariance scalaire et demi produit vectoriel

On s'intéresse ici à deux entités indépendantes du système de référence à savoir le produit scalaire et le produit vectoriel de 2 vecteurs, donnés ici par $\vec{V}(x)$ et $\vec{V}(x+h)$. Pour mémoire, leurs définitions respectives sont :

$$\vec{V}(x) \cdot \vec{V}(x+h) = V_1(x) \cdot V_1(x+h) + V_2(x) \cdot V_2(x+h) + V_3(x) \cdot V_3(x+h)$$

$$\frac{1}{2} \vec{V}(x) \wedge \vec{V}(x+h) = \frac{1}{2} \begin{bmatrix} V_2(x)V_3(x+h) - V_2(x+h)V_3(x) \\ V_3(x)V_1(x+h) - V_3(x+h)V_1(x) \\ V_1(x)V_2(x+h) - V_1(x+h)V_2(x) \end{bmatrix}$$

Dans un cadre stationnaire, la covariance scalaire est définie comme l'espérance du produit scalaire, et vaut de façon immédiate la somme des covariances (non centrées) des différentes composantes :

$$C_{\vec{V}}^s(h) = E[\vec{V}(x) \cdot \vec{V}(x+h)] = C_{11}(h) + C_{22}(h) + C_{33}(h)$$

En terme matriciel, la covariance scalaire est la trace de la matrice de variance-covariance. Dans un cadre un peu plus général du point de vue de la stationnarité, on peut définir de façon équivalente le variogramme "scalaire" comme la somme des variogrammes simples des composantes. Mais, dans ce cas, le terme "scalaire" est un peu usurpé. En effet, le variogramme scalaire est la (demi) espérance du carré de la norme de la différence entre $\vec{V}(x)$ et $\vec{V}(x+h)$:

$$\begin{aligned}
\frac{1}{2}E[\|\vec{V}(x) - \vec{V}(x+h)\|^2] &= \frac{1}{2}E[(V_1(x) - V_1(x+h))^2] \\
&\quad + \frac{1}{2}E[(V_2(x) - V_2(x+h))^2] + \frac{1}{2}E[(V_3(x) - V_3(x+h))^2] \\
&= \gamma_{11}(h) + \gamma_{22}(h) + \gamma_{33}(h)
\end{aligned}$$

D'utilisation classique en mécanique, le produit vectoriel permet de prendre en compte d'éventuelles rotations. Dans le cadre probabiliste des fonctions aléatoires vectorielles, on est naturellement amené à en prendre l'espérance, qui s'exprime, comme nous allons le voir, à l'aide des covariances croisées. En effet, si on s'intéresse par exemple à la troisième composante du demi produit vectoriel, son espérance est donnée par :

$$\frac{1}{2}E[V_1(x).V_2(x+h) - V_2(x).V_1(x+h)] = \frac{C_{12}(h) - C_{21}(h)}{2} = \frac{C_{12}(h) - C_{12}(-h)}{2}$$

qui n'est rien d'autre que la partie impaire (asymétrique) de la covariance croisée. Ainsi, l'espérance du demi produit vectoriel est un vecteur dont les composantes sont les parties impaires des covariances croisées. On ne peut cependant pas développer d'outil équivalent pour le cadre intrinsèque puisque le variogramme croisé est une fonction symétrique qui ignore la partie impaire des covariances croisées :

$$\gamma_{12}(h) = C_{12}(0) - \underbrace{\frac{1}{2}(C_{12}(h) + C_{12}(-h))}_{\text{partie paire}} = \gamma_{12}(-h)$$

A noter au passage que la partie impaire de la covariance croisée entre composantes d'un vecteur est indépendante du repère. A cette partie fixe s'ajoute, selon le repère, une partie paire. Cependant, parmi l'ensemble des repères possibles, il en existe un dans lequel la covariance croisée entre composantes est strictement impaire (ou encore le variogramme croisé identiquement nul). Ce repère se caractérise par le fait que la variance des composantes y est respectivement maximum et minimum, à 2D par exemple.

En se basant sur les produits scalaire et vectoriel, on a donc développé des outils structuraux dont l'interprétation est indépendante du repère utilisé. Dans cette approche, la structure du phénomène est décrite par une rotation moyenne et un produit scalaire moyen en fonction de la distance. Cependant, on a perdu une partie de la richesse des informations contenues dans la matrice 3x3 de variance-covariance puisqu'on ne se sert que de sa trace et de la partie impaire des covariances croisées.

2.2 Covariance colinéaire et orthogonale

Par ailleurs, l'ensemble des outils qui viennent d'être présentés a l'inconvénient de ne pas faire la différence entre deux configurations où seule l'orientation de la distance entre les points d'appui des vecteurs change (Fig. 4.15).

Pour ce faire, une solution consiste à distinguer, pour une direction donnée, la partie de $\vec{V}(x)$ qui lui est colinéaire et, la partie qui lui est orthogonale. Il est à noter que la partie orthogonale est un vecteur à 2 dimensions si l'espace de départ en comprend trois. Son analyse structurale est donc confrontée aux problèmes évoqués au paragraphe précédent et se base donc sur des produits

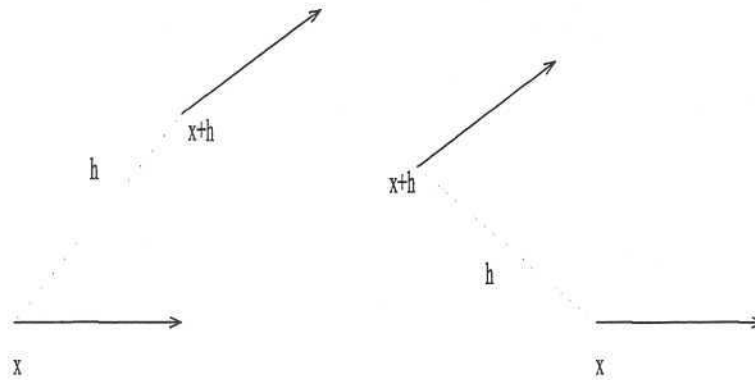


FIG. 4.15 - Deux configurations de vecteurs. Seule l'orientation de la distance entre les points x et $x+h$ change.

scalaire et vectoriel. Ceci permet pour chaque direction de l'espace h , de décomposer l'analyse d'un vecteur en deux : l'analyse du mouvement selon h et l'analyse du mouvement à travers h .

Finalement, l'analyse structurale d'une variable vectorielle, pour être indépendante du choix du repère et pour tenir compte de l'orientation des distances entre vecteurs, devrait comprendre (dans le cadre stationnaire), pour une direction donnée :

- la covariance des parties colinéaires $C_{V \downarrow h}(h)$. La notation $V \downarrow h$ signifie projection de $\vec{V}(x)$ sur h . C'est une fonction de la distance qui schématise la corrélation spatiale du mouvement selon h (Fig. 4.16b)
- la covariance scalaire des parties orthogonales $C_{V \perp h}^s(h)$. Elle décrit la structure spatiale du mouvement au travers de h (Fig. 4.16c). La somme de ces deux structures donne la covariance scalaire du vecteur $\vec{V}(x)$:

$$C_{\vec{V}}^s(h) = C_{V \downarrow h}(h) + C_{V \perp h}^s(h)$$

- l'espérance du produit vectoriel entre la partie de $\vec{V}(x)$ colinéaire à h ($\vec{V}(x) \downarrow h$), et $\vec{V}(x+h)$. Cependant, la partie de $\vec{V}(x+h)$ colinéaire à h ne joue aucun rôle dans le calcul du produit vectoriel puisqu'on s'intéresse ici aux rotations. Seules les parties orthogonales de $\vec{V}(x+h)$ interviennent (Fig. 4.16d). Ainsi, l'espérance du produit vectoriel est donnée par les covariances croisées entre la partie colinéaire en x et les parties orthogonales en $x+h$. Comme on l'a vu précédemment, la partie impaire correspond en espérance au demi produit vectoriel entre les 2 vecteurs.

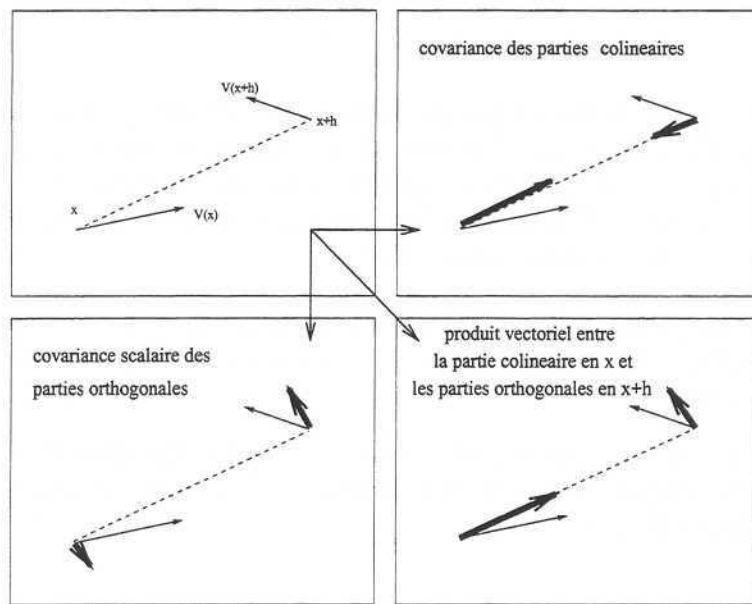


FIG. 4.16 - Décomposition de l'analyse structurale d'un vecteur en plusieurs outils structuraux indépendants du repère et sensibles à l'orientation de la distance entre vecteurs. (b) Covariance des parties colinéaires. (c) Covariance scalaire des parties orthogonales. (d) Produit vectoriel entre la partie colinéaire en x et les parties orthogonales en $x+h$.

Annexe 2 : Représentations proportionnelles

Cette annexe comprend l'ensemble des représentations proportionnelles des densités d'œufs, des densités de larves de maquereaux, et des températures de surface utilisées dans la quatrième partie.

Liste des figures :

- **Œufs de stade 1** : Figures 1 à 5.
 - 1983 (période 1 à 5) : Figure 1
 - 1986 (période 1 à 4) : Figure 2
 - 1989 (période 1 à 5) : Figure 3
 - 1992 (période 1 à 4) : Figure 4
 - 1995 (période 1 à 5) : Figure 5
- **Œufs de stade 5** : Figures 6 à 10.
 - 1983 (période 1 à 5) : Figure 6
 - 1986 (période 1 à 4) : Figure 7
 - 1989 (période 1 à 5) : Figure 8
 - 1992 (période 1 à 4) : Figure 9
 - 1995 (période 1 à 5) : Figure 10
- **Larves toutes taille** : Figures 11 à 13.
 - 1983 (période 1 à 4) : Figure 11
 - 1986 (période 2 à 3) : Figure 12
 - 1989 (période 1 à 5) : Figure 13
- **Larves de la première classe de taille (entre 2.5 et 4 mm) 1989** : Figure 14
- **Larves de la deuxième classe de taille (entre 4 et 6 mm) 1989** : Figure 15
- **Larves de la troisième classe de taille (entre 6 et 13 mm) 1989** : Figure 16
- **Température de surface mesurées *in situ*** : Figure 17 à 21
 - 1983 (période 1 à 5) : Figure 17
 - 1986 (période 1 à 4) : Figure 18
 - 1989 (période 1 à 5) : Figure 19
 - 1992 (période 1 à 4) : Figure 20
 - 1995 (période 1 à 5) : Figure 21

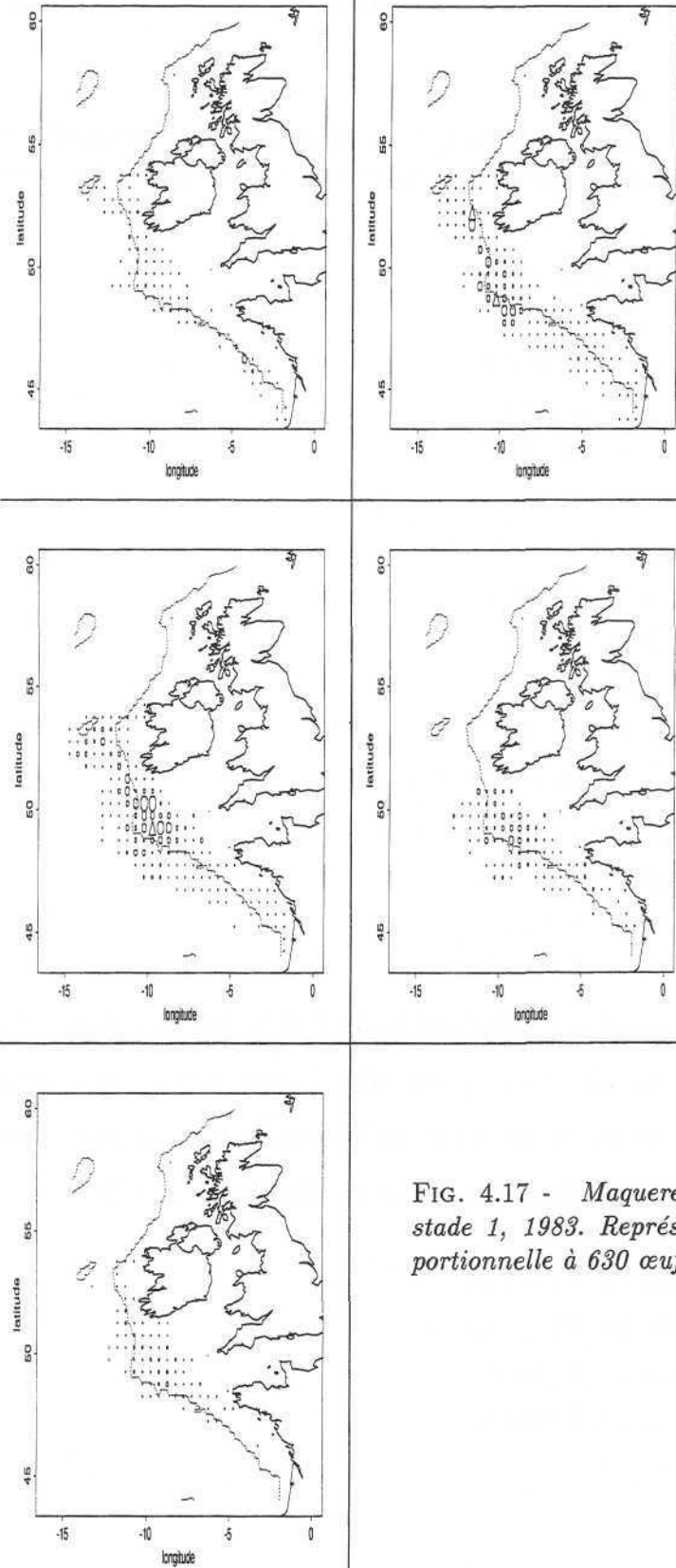


FIG. 4.17 - *Maquereaux*, œufs de stade 1, 1983. Représentation proportionnelle à 630 œufs/m²/jour.

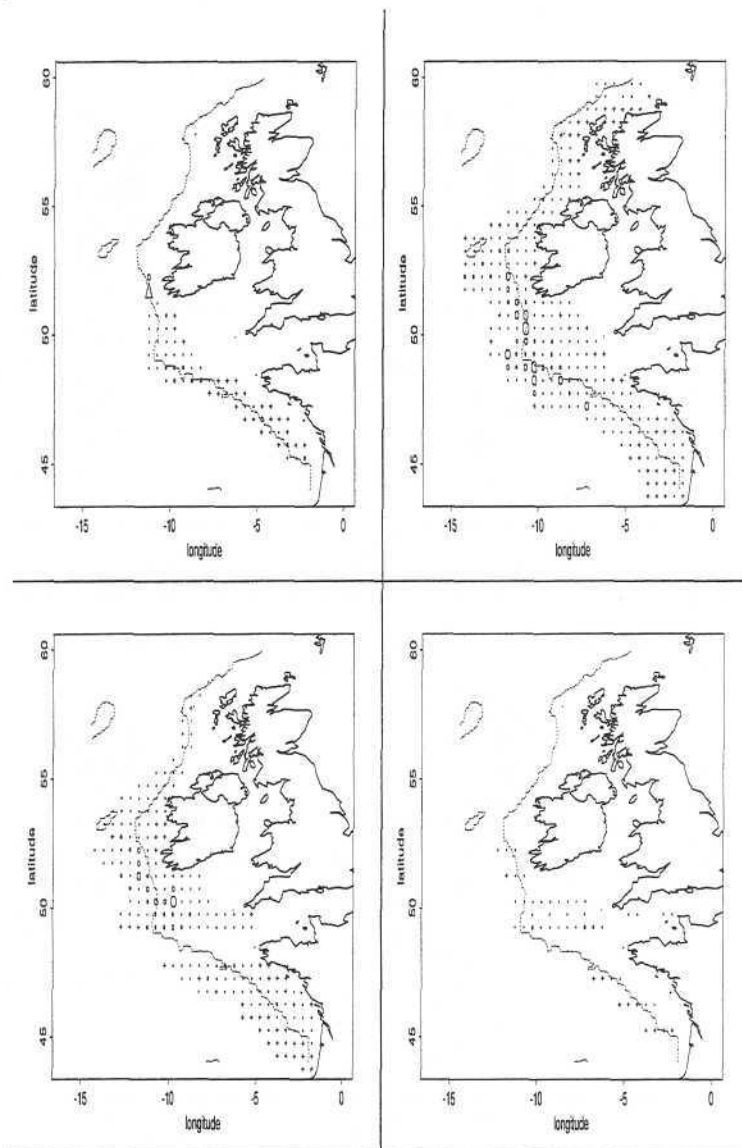


FIG. 4.18 - Maquereaux, œufs de stade 1, 1986. Représentation proportionnelle à 630 œufs/m²/jour.

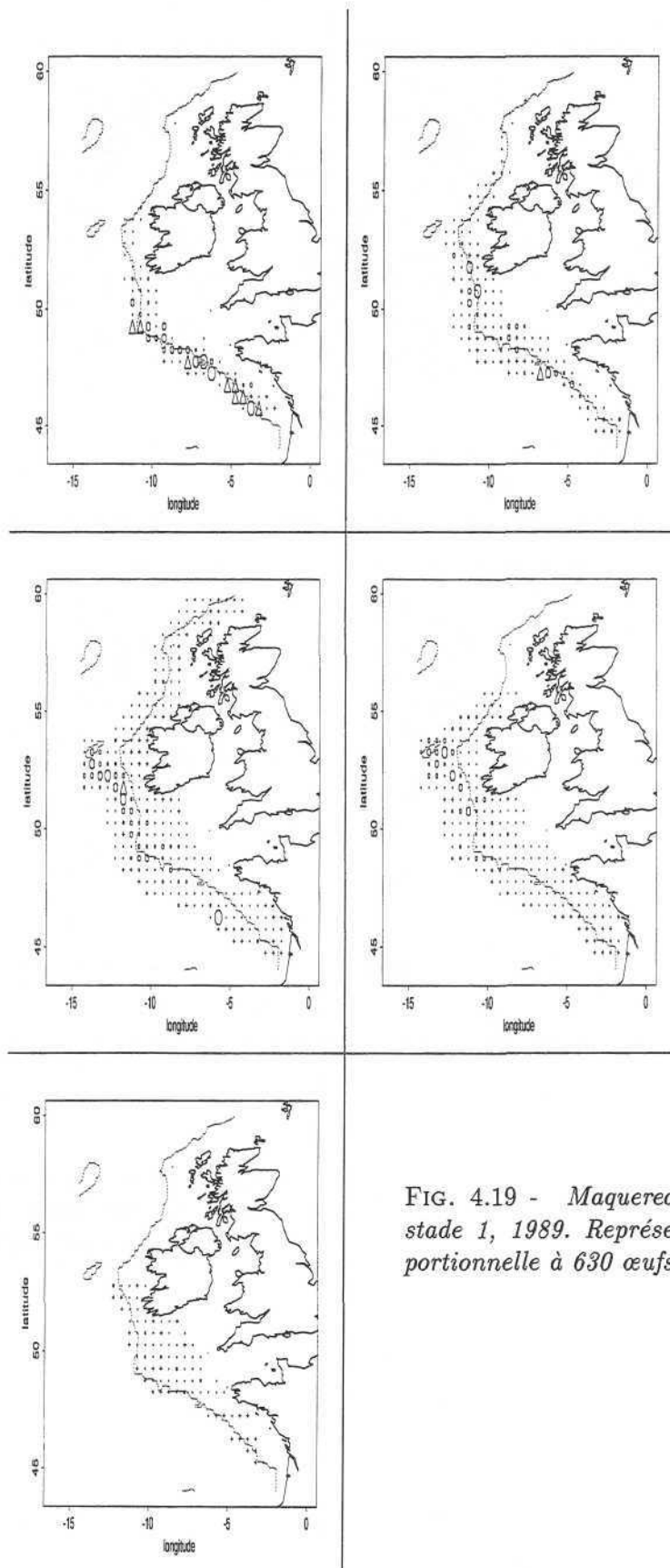


FIG. 4.19 - Maquereaux, œufs de stade 1, 1989. Représentation proportionnelle à 630 œufs/m²/jour.

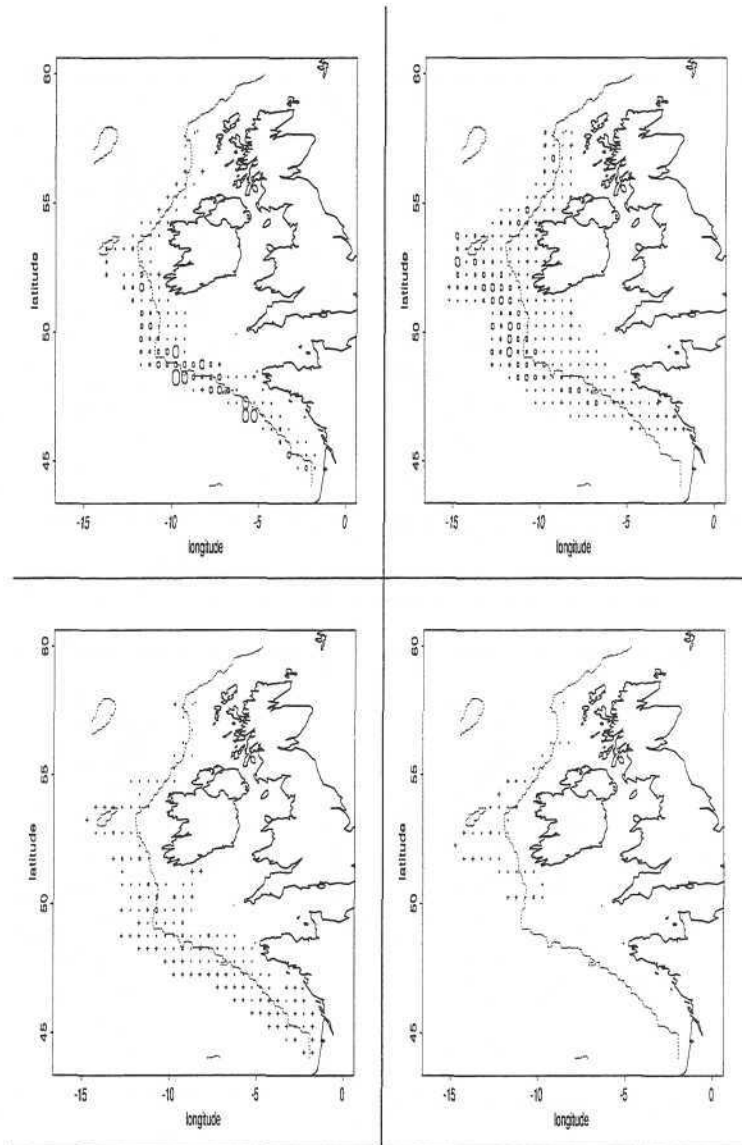


FIG. 4.20 - *Maquereaux*, œufs de stade 1, 1992. Représentation proportionnelle à 630 œufs/m²/jour.

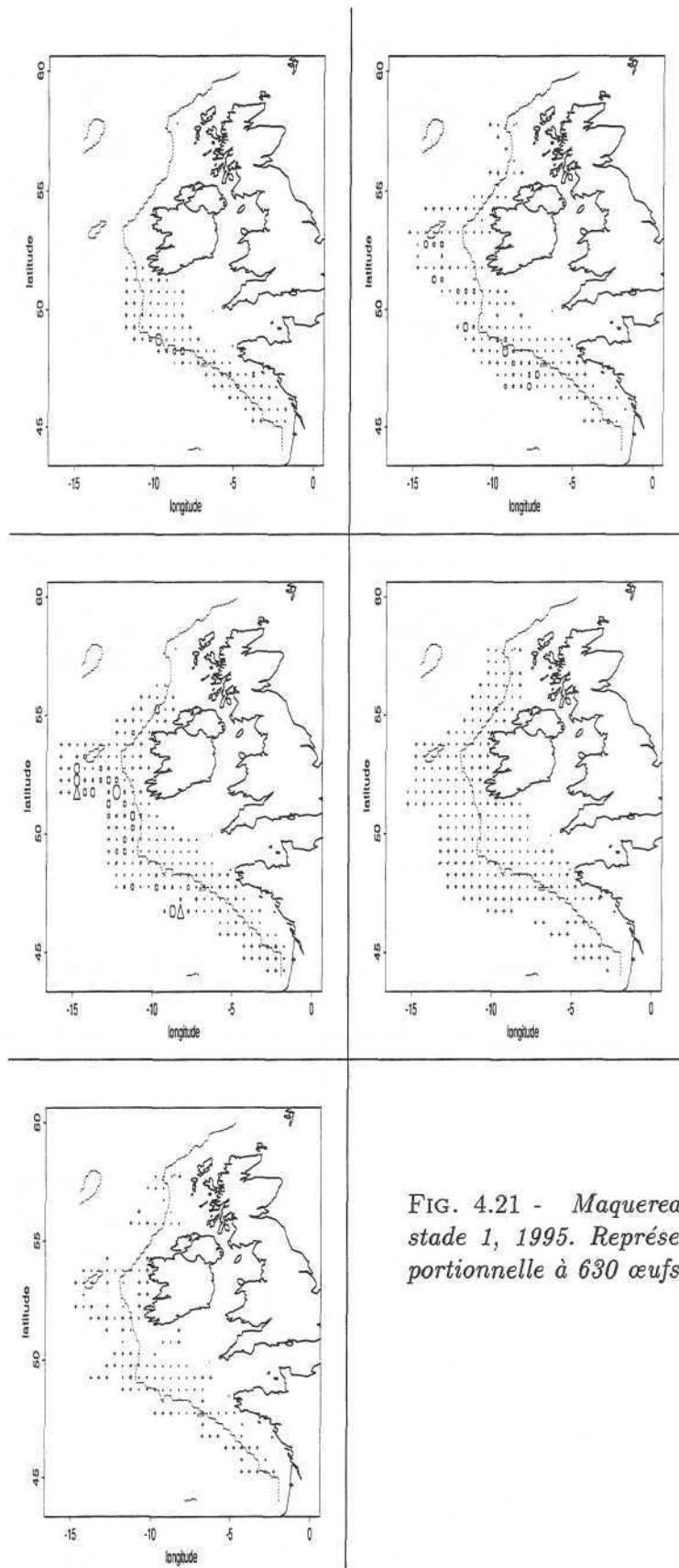
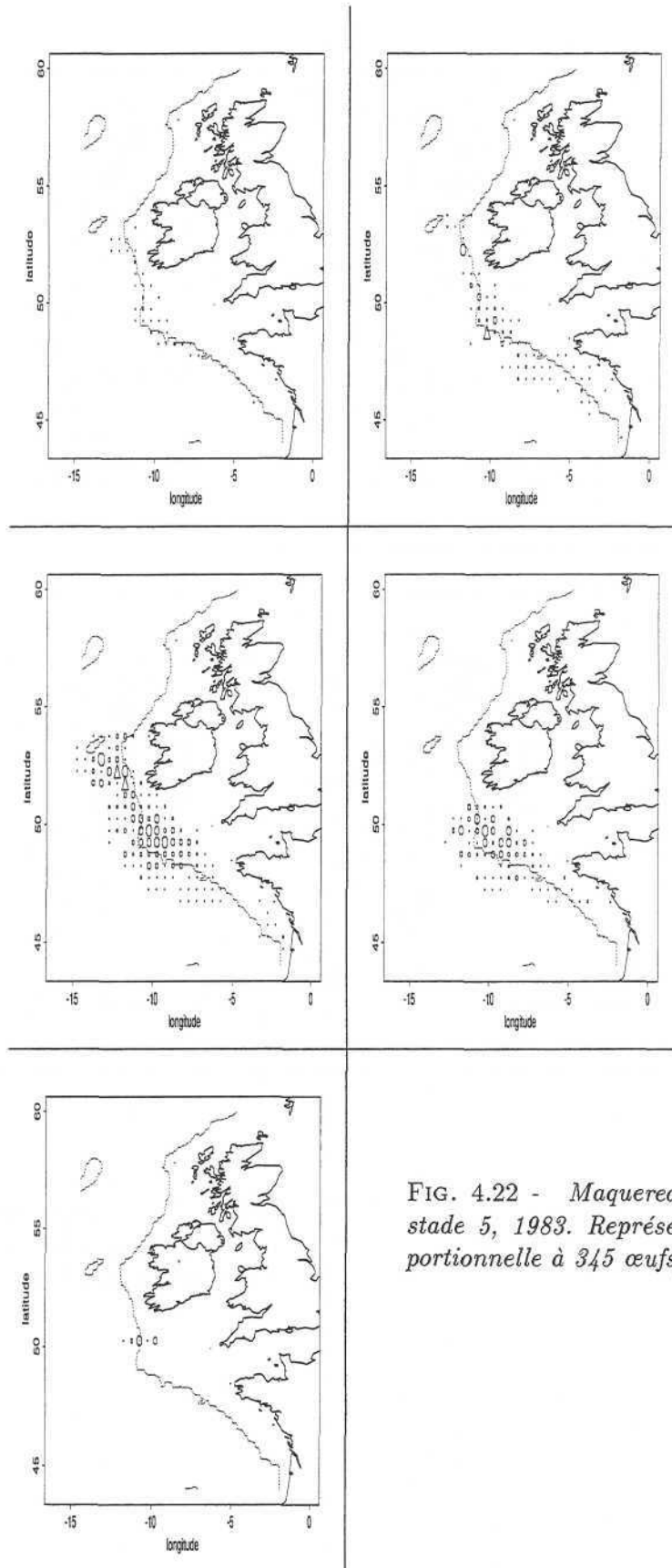


FIG. 4.21 - *Maquereaux*, œufs de stade 1, 1995. Représentation proportionnelle à 630 œufs/m²/jour.



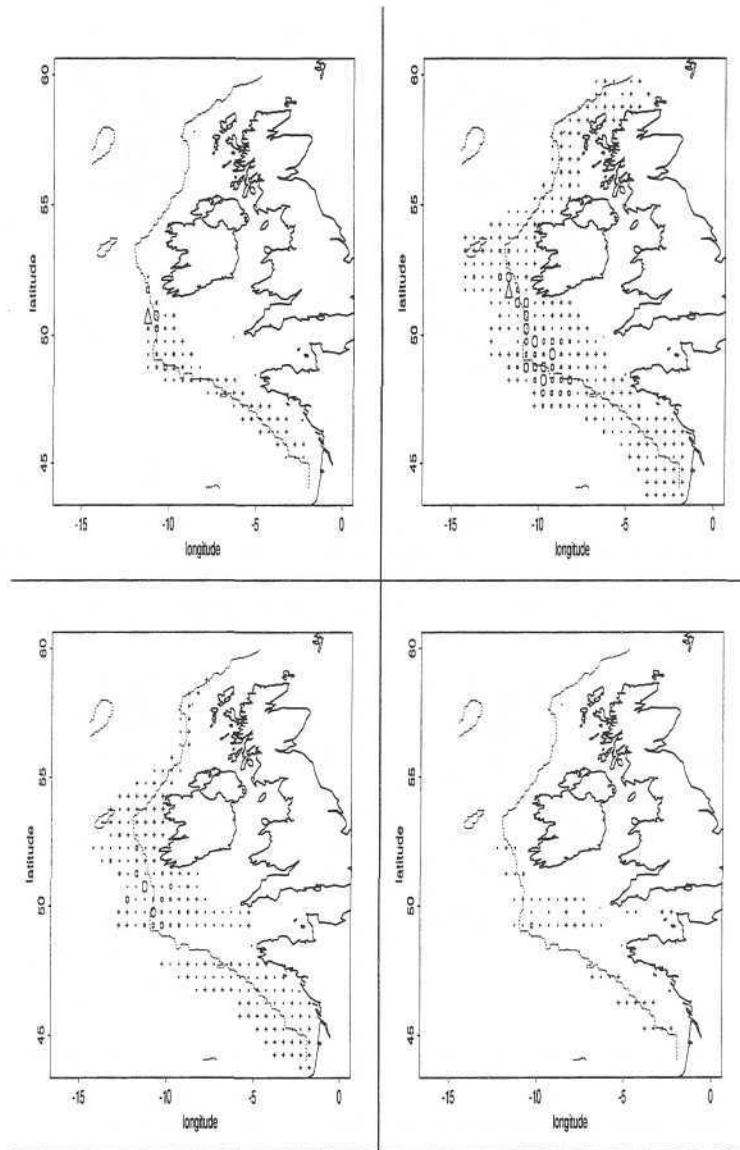


FIG. 4.23 - *Maquereaux*, œufs de stade 5, 1986. Représentation proportionnelle à 345 œufs/m²/jour.

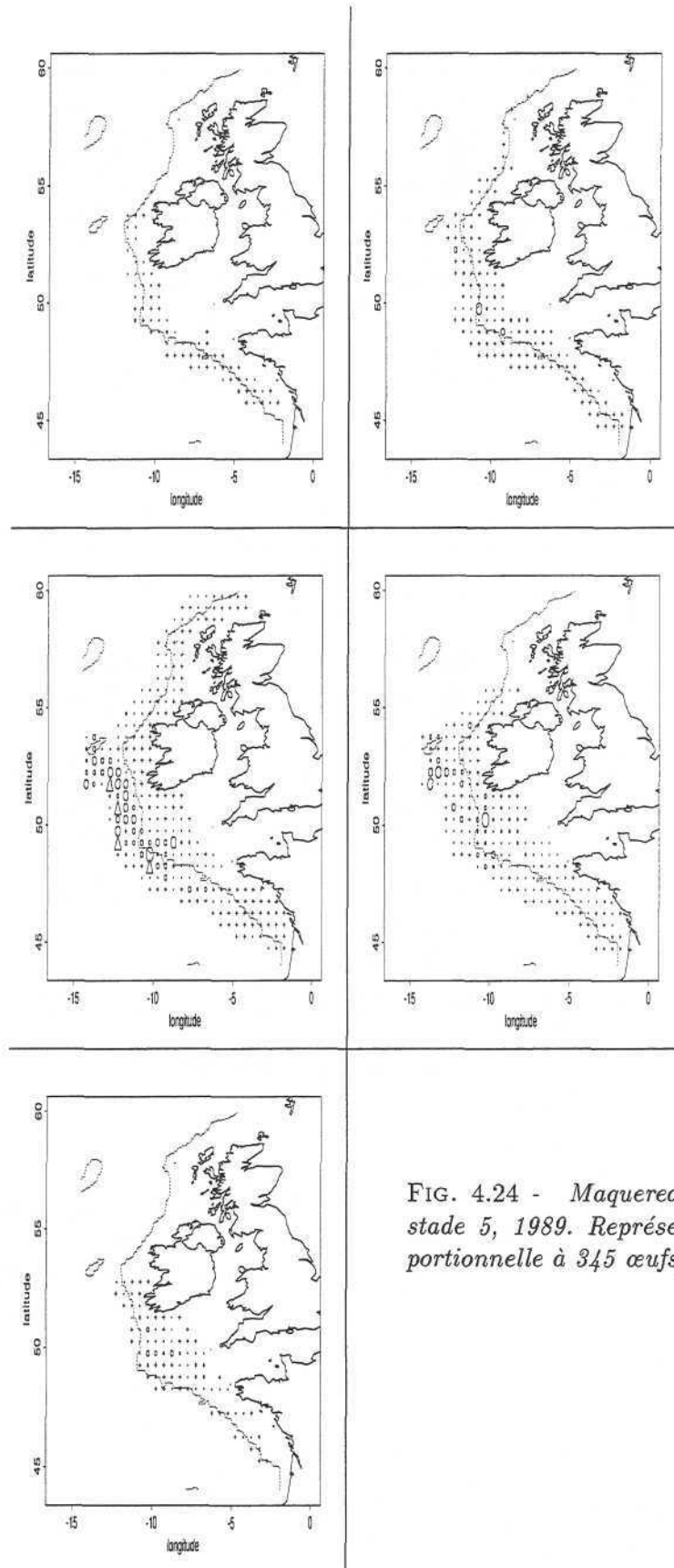


FIG. 4.24 - Maquereaux, œufs de stade 5, 1989. Représentation proportionnelle à 345 œufs/m²/jour.

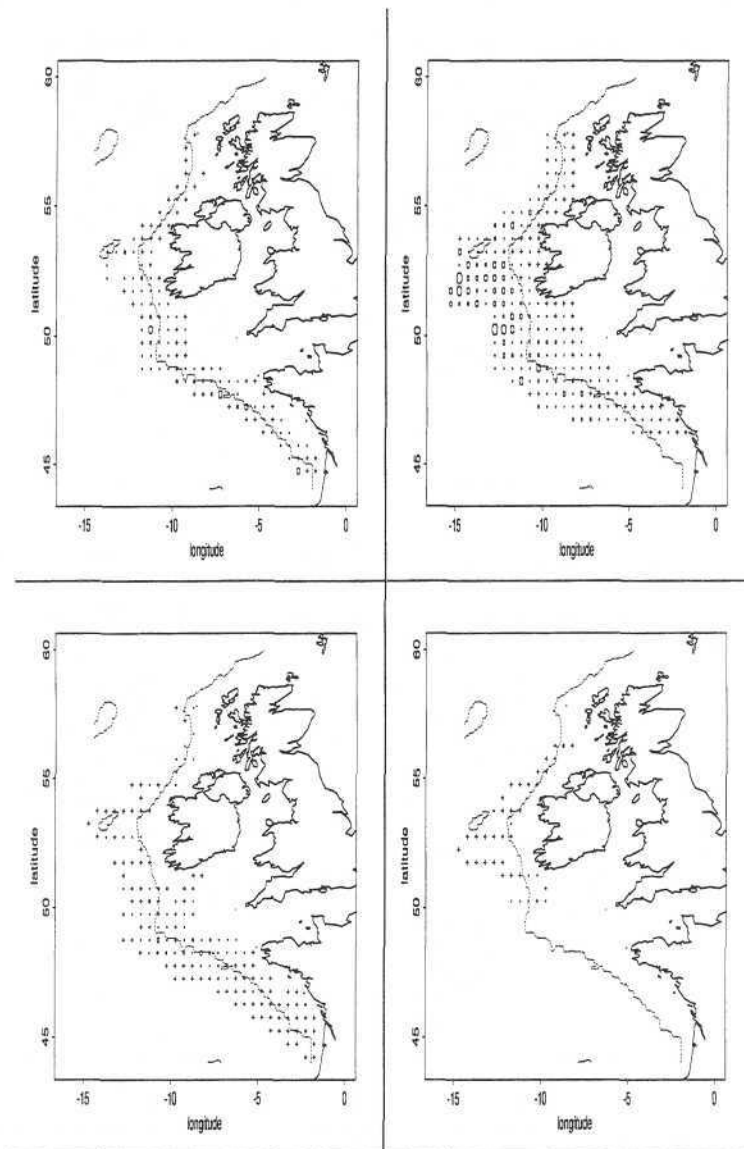


FIG. 4.25 - *Maquereaux*, œufs de stade 5, 1992. Représentation proportionnelle à 345 œufs/m²/jour.

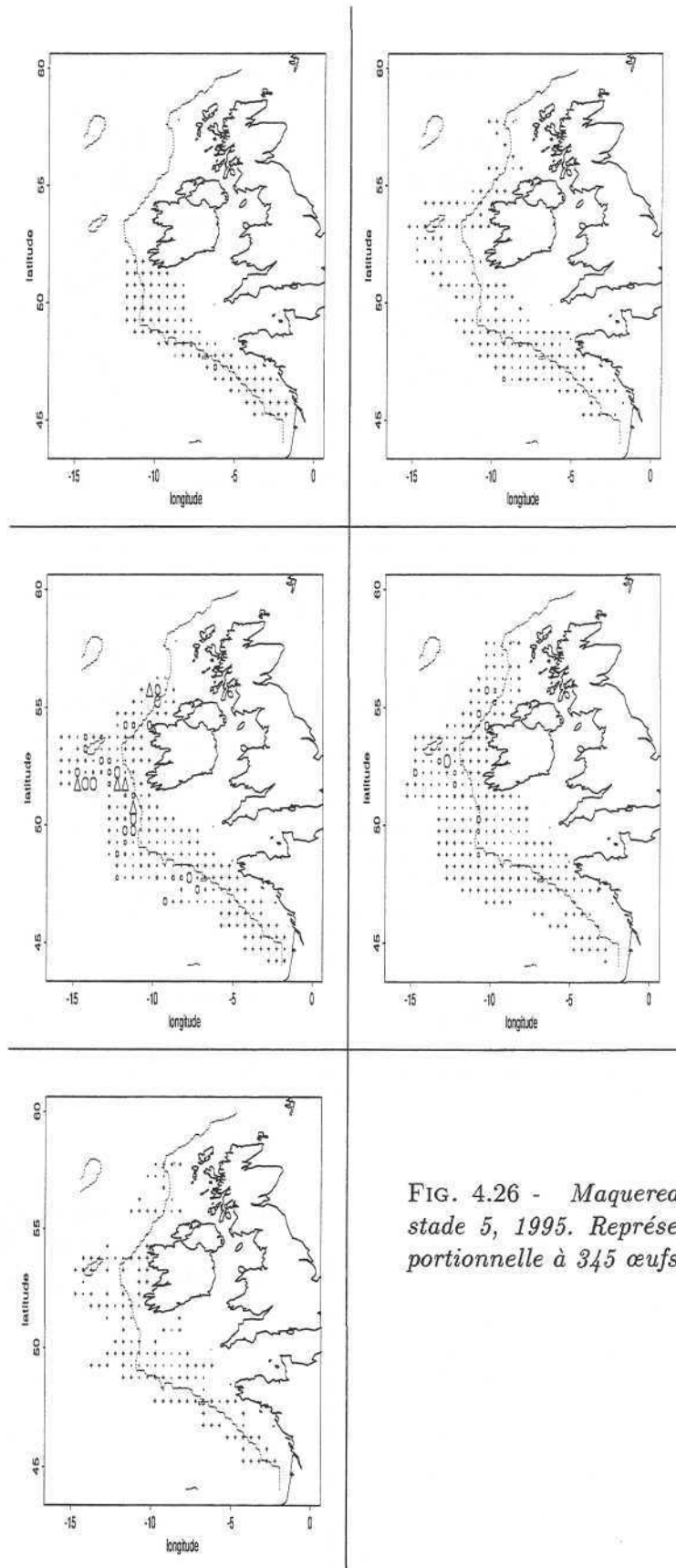


FIG. 4.26 - *Maquereaux*, œufs de stade 5, 1995. Représentation proportionnelle à 345 œufs/m²/jour.

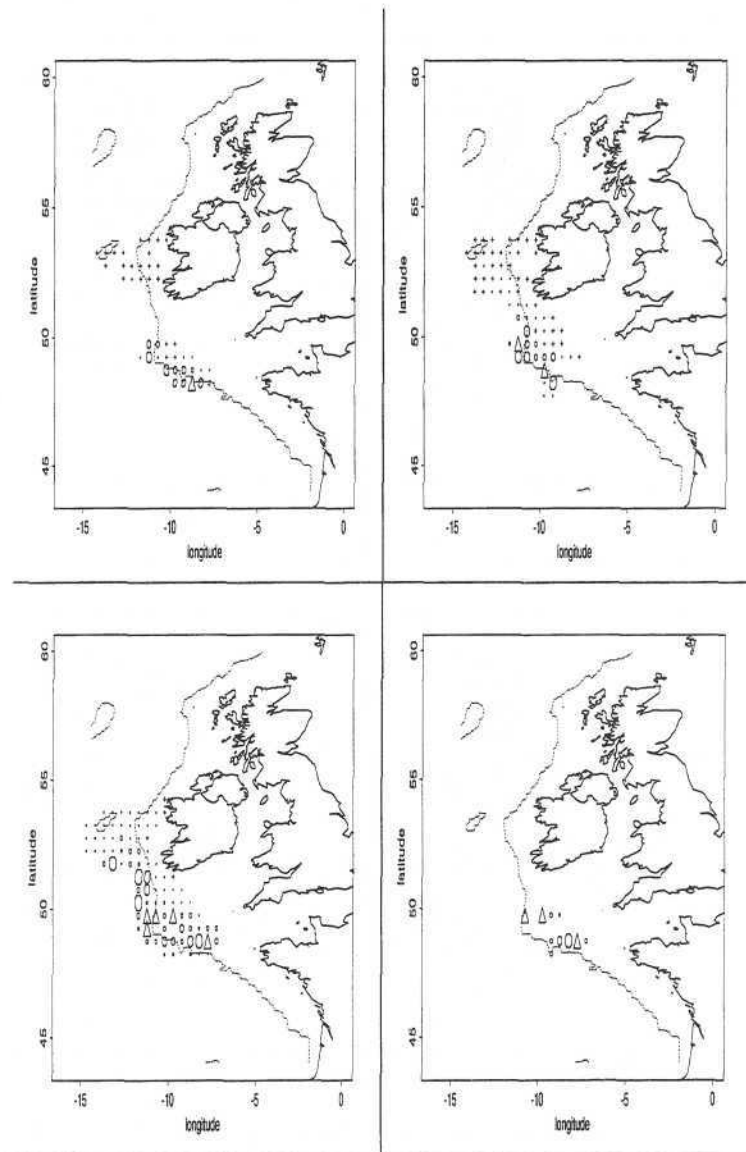


FIG. 4.27 - Maquereaux, larves de toutes tailles, 1983. Représentation proportionnelle à 42 larves/m²/jour.

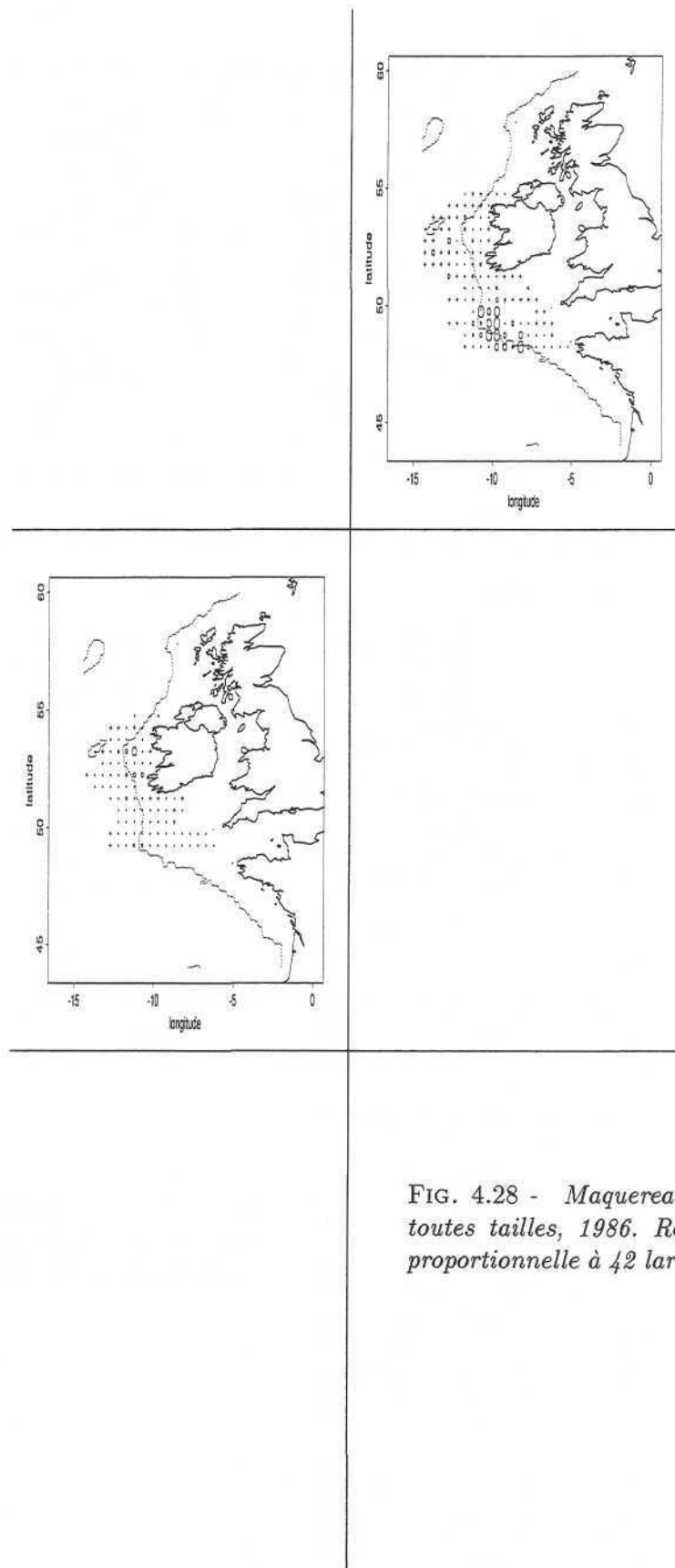


FIG. 4.28 - Maquereaux, larves de toutes tailles, 1986. Représentation proportionnelle à 42 larves/m²/jour.

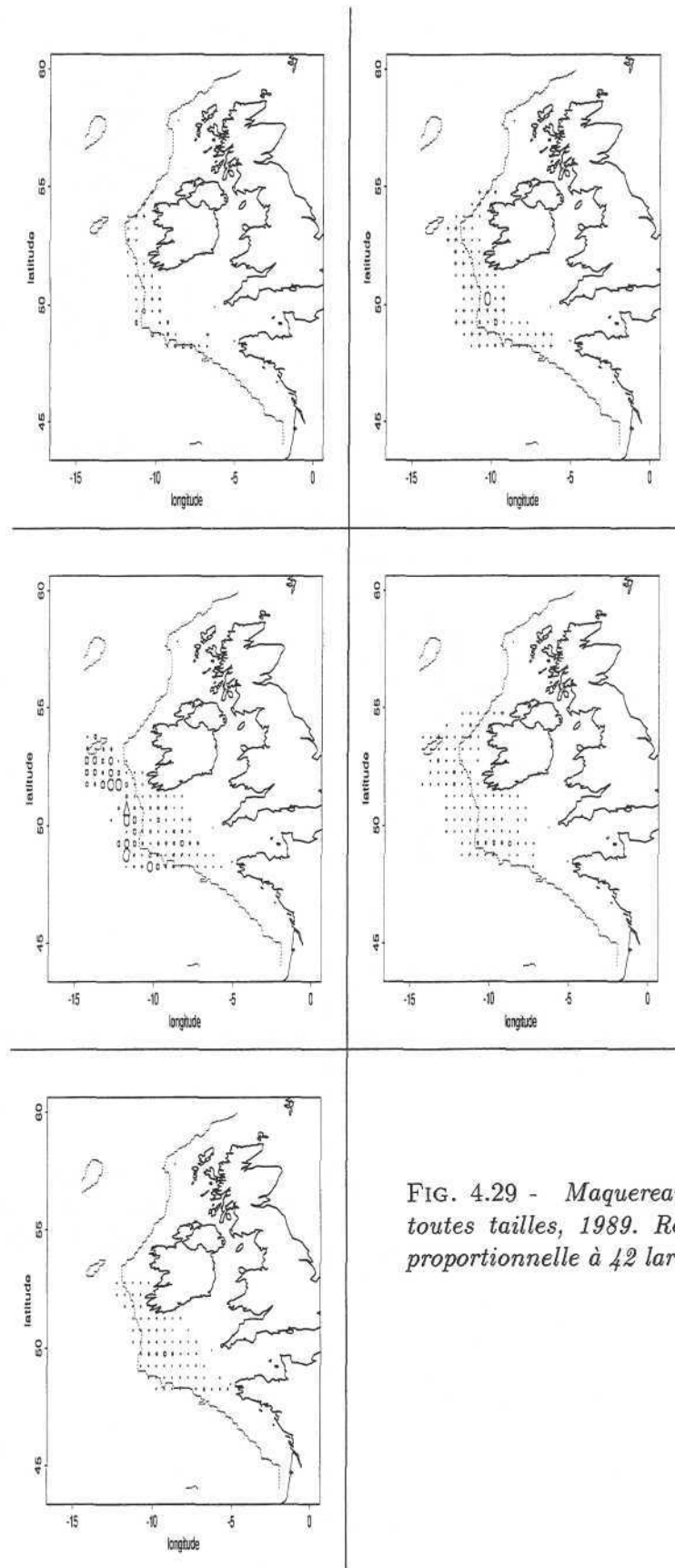


FIG. 4.29 - Maquereaux, larves de toutes tailles, 1989. Représentation proportionnelle à 42 larves/m²/jour.

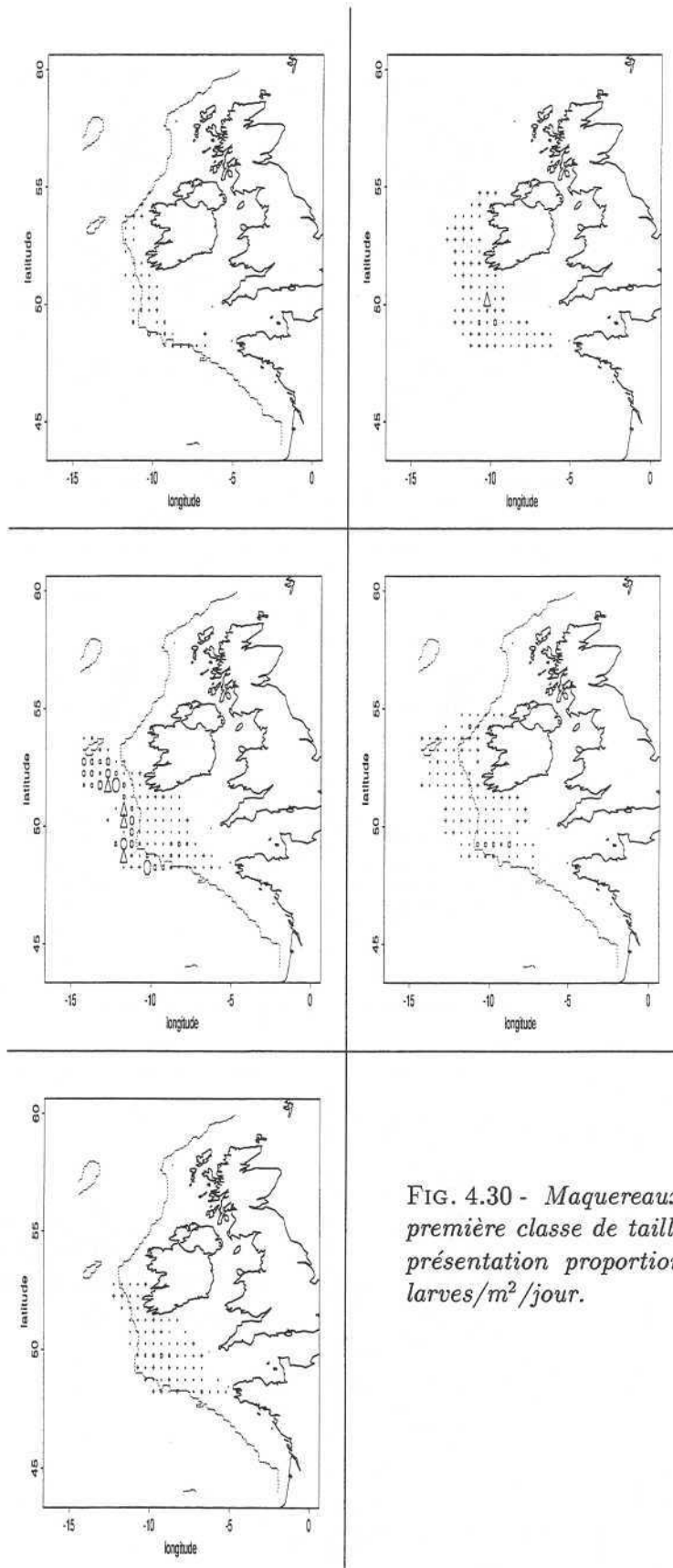


FIG. 4.30 - Maquereaux, larves de la première classe de tailles, 1989. Représentation proportionnelle à 171 larves/m²/jour.

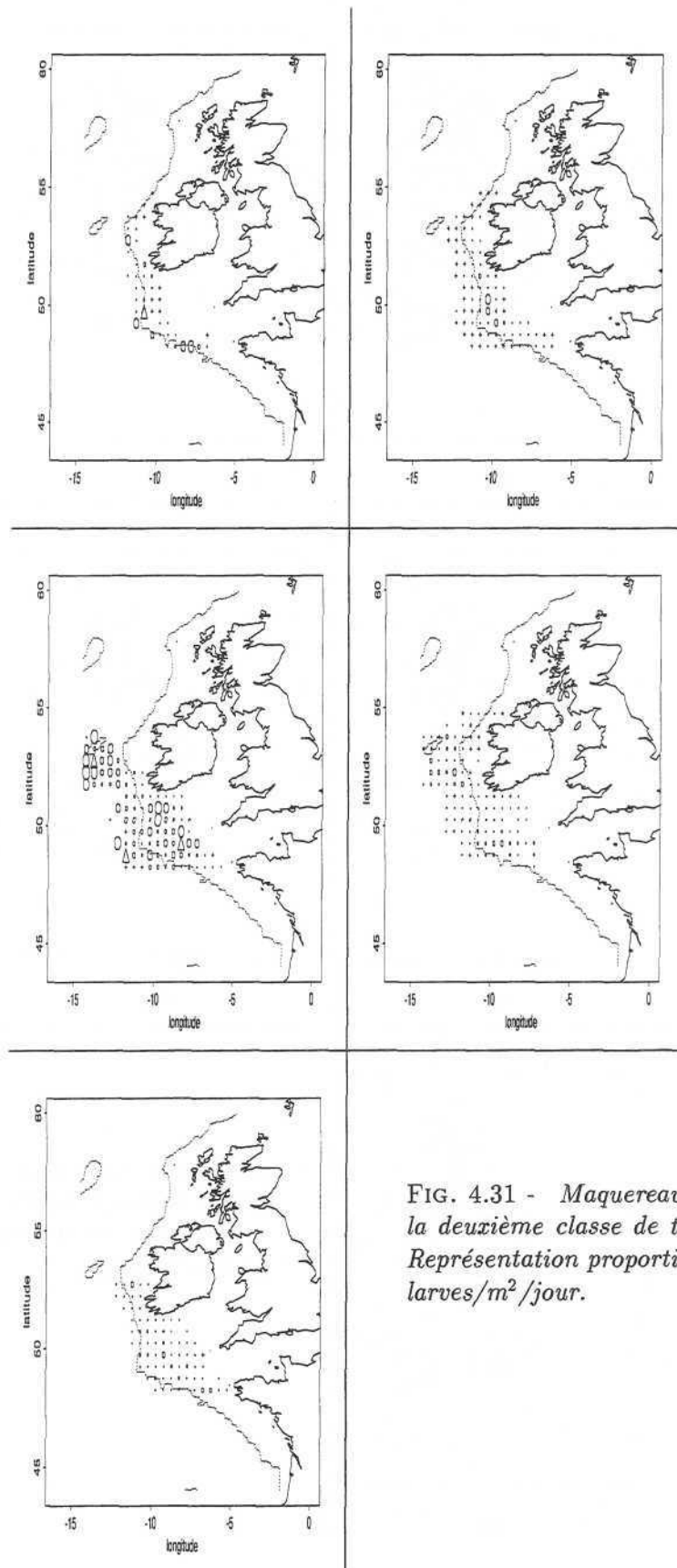


FIG. 4.31 - *Maquereaux*, larves de la deuxième classe de tailles, 1989. Représentation proportionnelle à 21 larves/m²/jour.

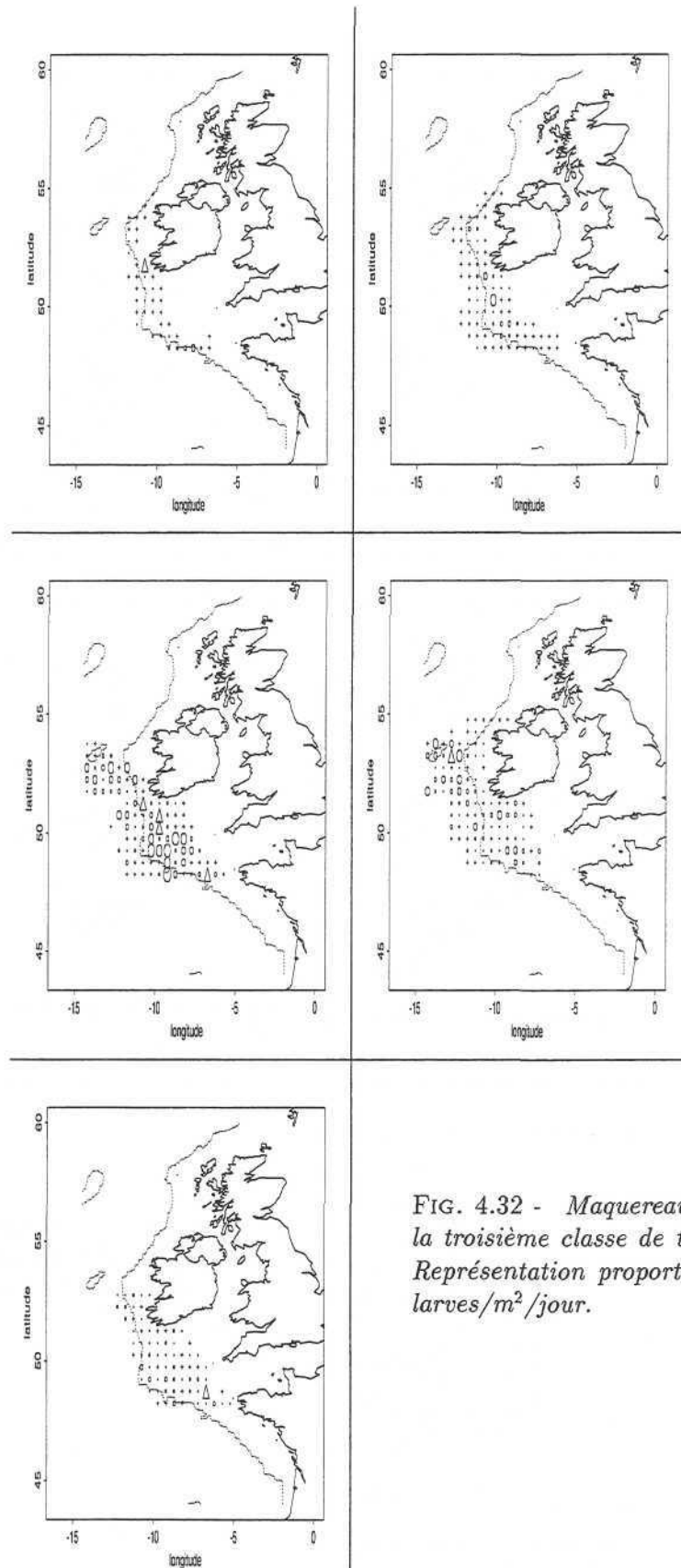


FIG. 4.32 - *Maquereaux*, larves de la troisième classe de tailles, 1989. Représentation proportionnelle à 2 larves/m²/jour.

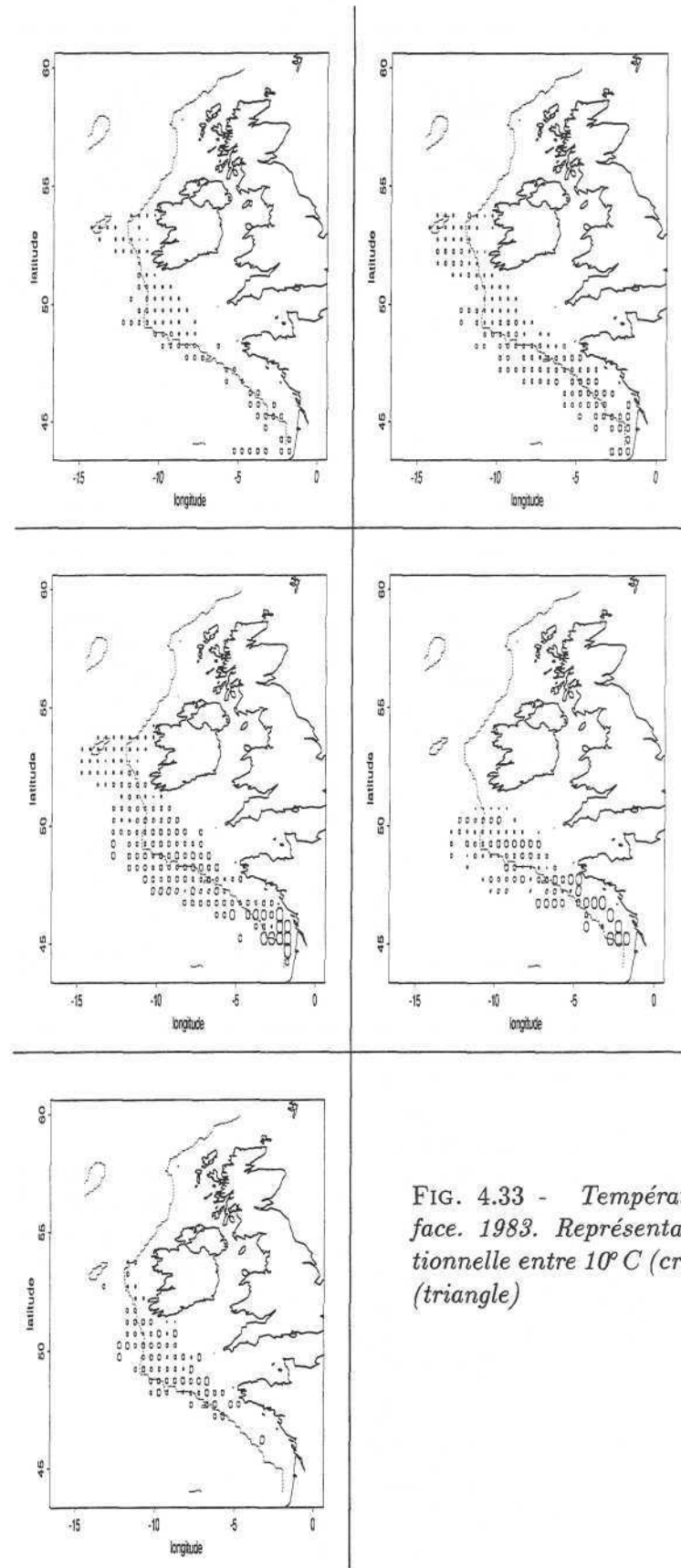


FIG. 4.33 - *Température de surface. 1983. Représentation proportionnelle entre 10°C (croix) et 19°C (triangle)*

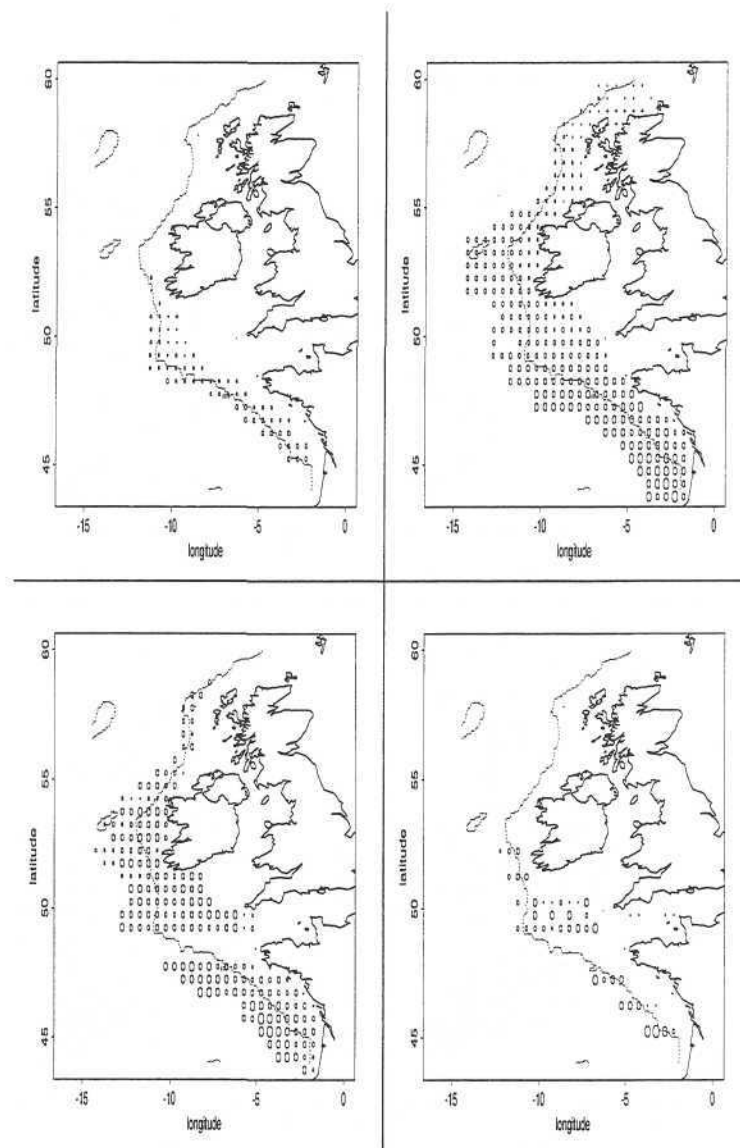


FIG. 4.34 - *Température de surface. 1986. Représentation proportionnelle entre 10° C (croix) et 19° C (triangle)*

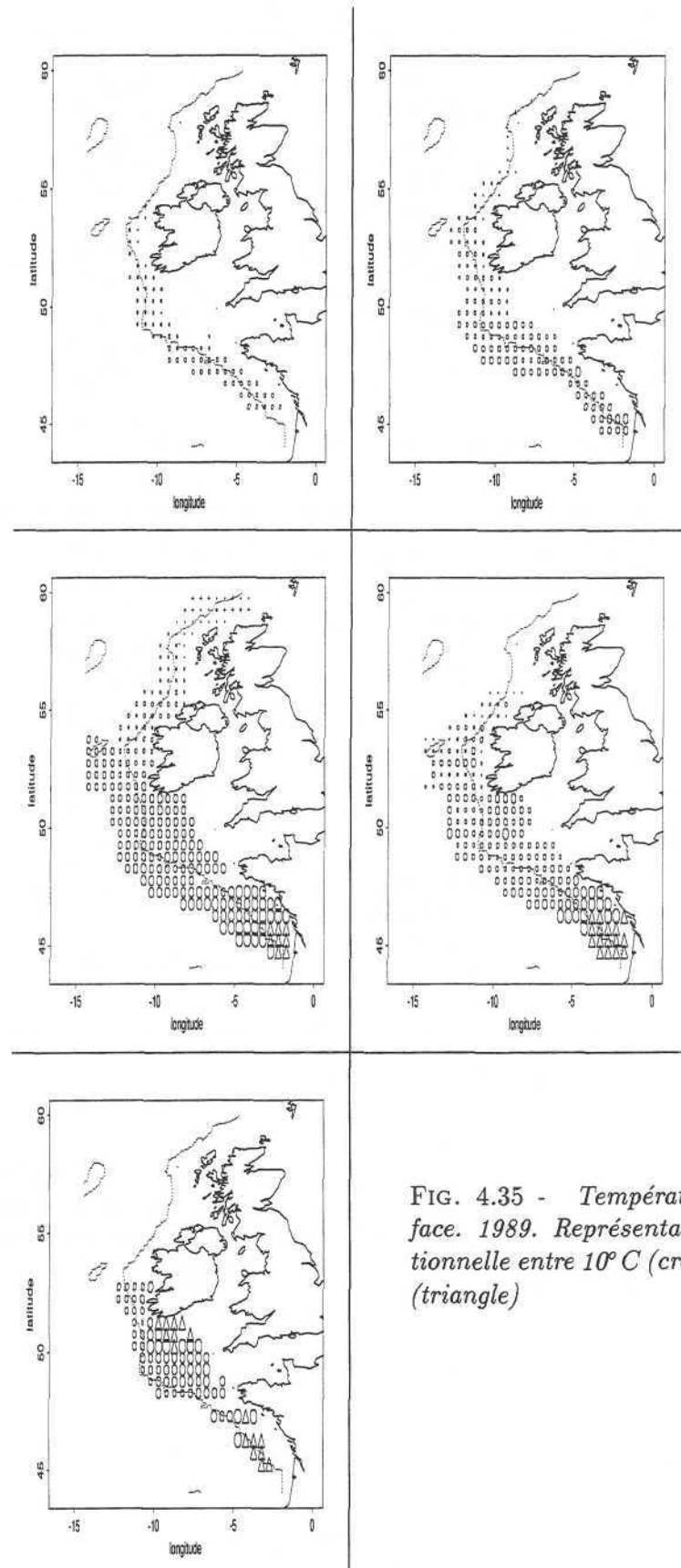


FIG. 4.35 - *Température de surface. 1989. Représentation proportionnelle entre 10°C (croix) et 19°C (triangle)*

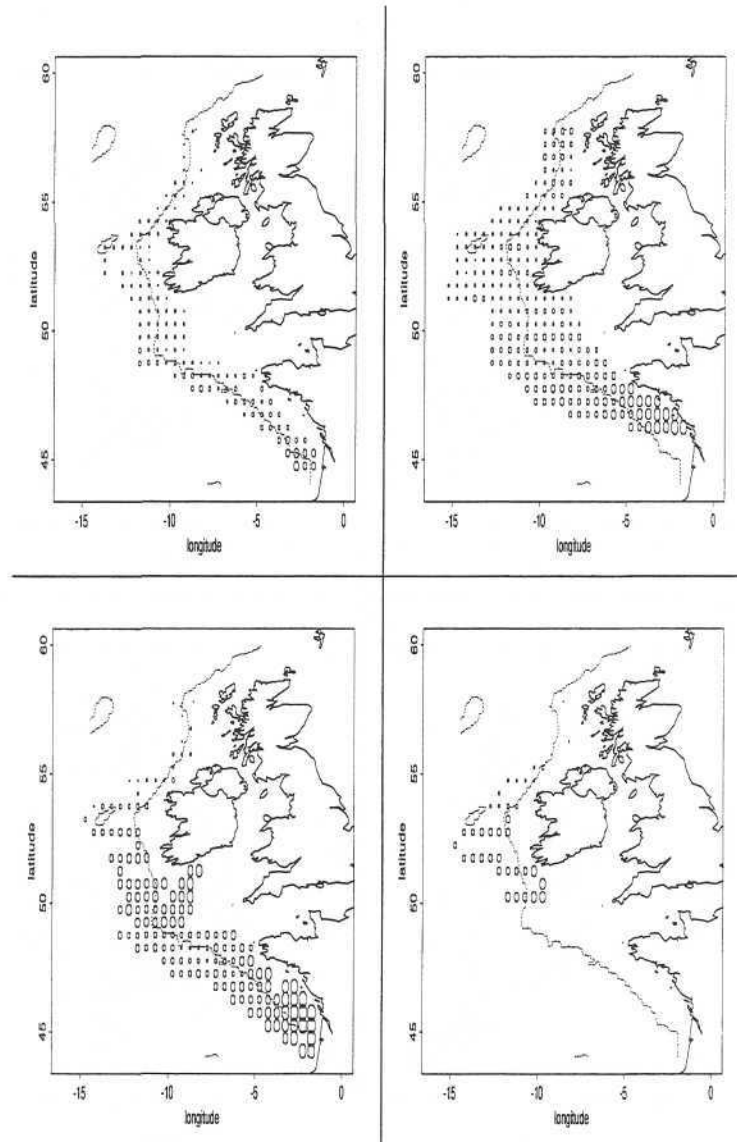


FIG. 4.36 - *Température de surface. 1992. Représentation proportionnelle entre 10° C (croix) et 19° C (triangle)*

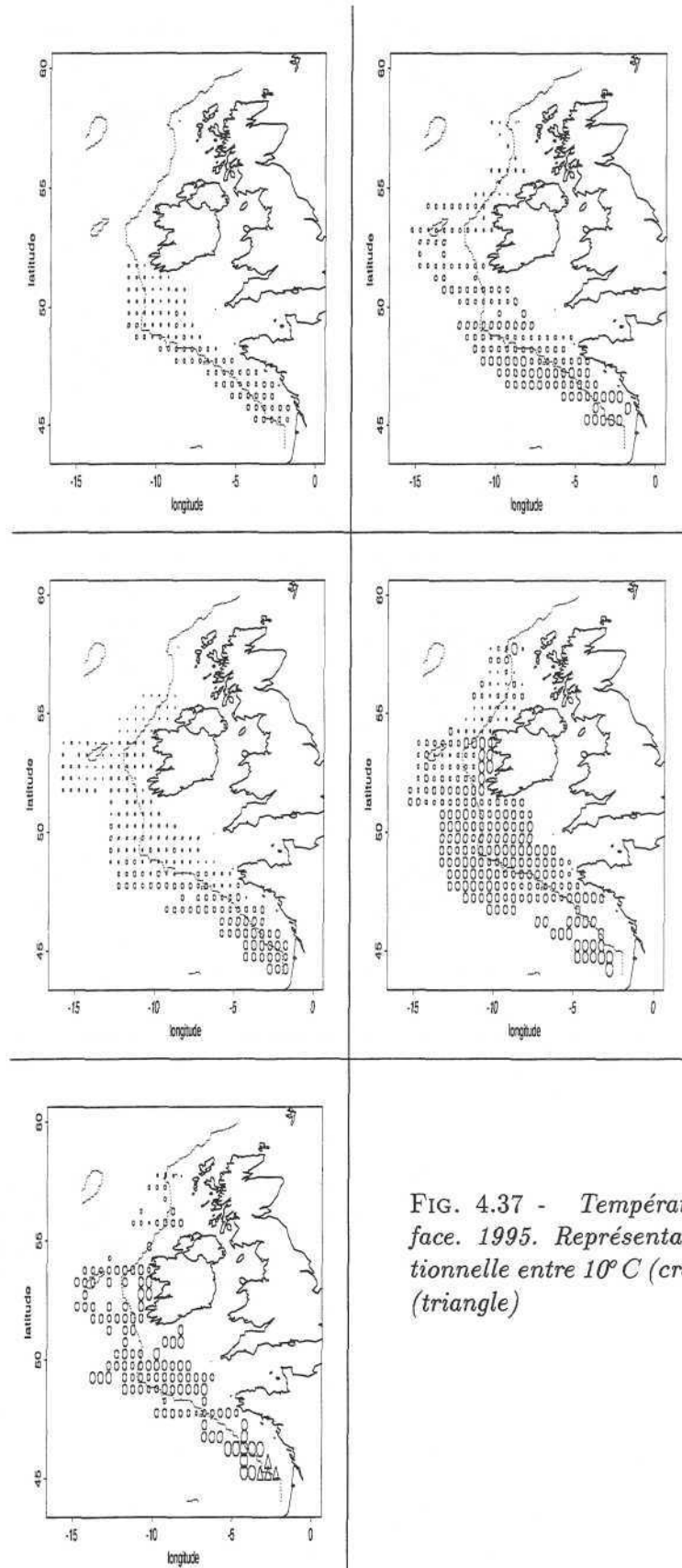
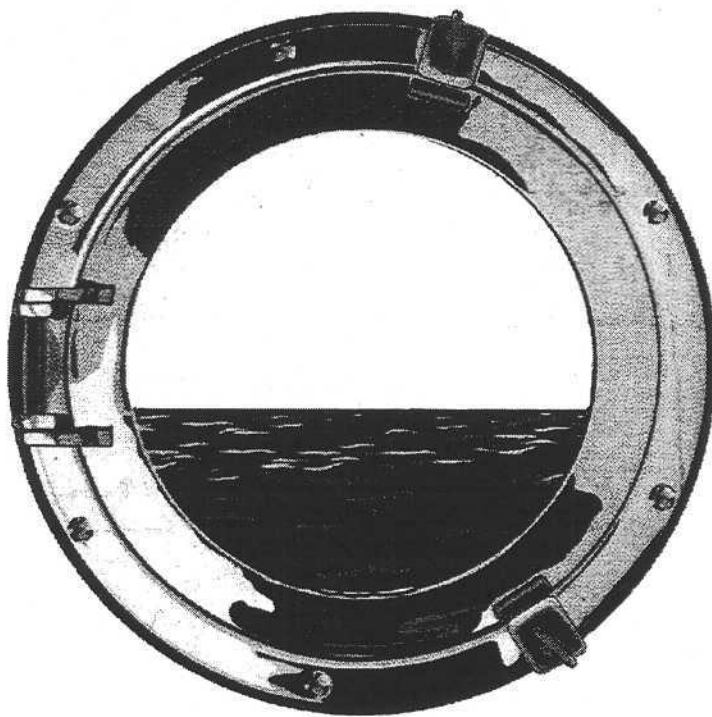


FIG. 4.37 - *Température de surface. 1995. Représentation proportionnelle entre 10°C (croix) et 19°C (triangle)*



Résumé

Comment analyser la distribution spatiale d'espèces marines en relation avec quelques paramètres de l'environnement?

A l'aide de quatre exemples clefs qui permettent de faire le point sur les difficultés de l'analyse structurale en halieutique (densités de poissons, température de surface, courants marins), on discute des choix de modélisation. Il apparaît ainsi que la **géostatistique transitive** est une méthodologie adaptée quand l'abondance totale est le fait de quelques valeurs extrêmement fortes et qu'il est difficile de délimiter un domaine au sein duquel les variations de densités peuvent être supposées homogènes.

De là, on choisit **un individu au hasard dans la population** et on s'intéresse à la probabilité qu'il se trouve en un certain point de l'espace, à une certaine température ou bien dans une partie de la population où la densité vaut une certaine valeur (**statistiques individuelles**).

Une description spatiale, basée sur le **covariogramme**, est obtenue par tirage au sort de deux individus. Des procédures d'estimations du covariogramme adaptées à des échantillonnages irréguliers sont proposées ainsi que les utilisations possibles d'un modèle structural, c'est-à-dire d'un covariogramme modélisé. En particulier, on indique en l'illustrant à partir de densités de merlu du Golfe de Gascogne (France), que le covariogramme contient des éléments structuraux plus fins que ne le laisse envisager sa définition, et on procède à une cartographie par **krigeage transitif**.

L'approche par individu est utilisée pour la description et la modélisation de la répartition des œufs de maquereau pendant la saison de ponte de 1983 à 1995. L'interprétation dans un contexte spatial des régressions expérimentales (en cloche) étant délicate, on préfère tester l'existence d'un preferendum en température de surface en vérifiant que les individus sont distribués de façon à minimiser leur inertie en température (**inertiogrammes**). De ce point de vue, les maquereaux apparaissent contraints dans leur régionalisation par des "fronts virtuels d'inertie".

Abstract

How can the link between the spatial distributions of fish and environmental parameters be analysed?

Four key examples that illustrate the principal difficulties encountered in the structural analysis of fish densities, sea surface temperatures or currents, are used to present some choices for models. **Transitive geostatistics** turned out to be an appropriate methodology when the abundance is due to few extremely high values and when it is difficult to delineate a domain within which variations of density can be assumed homogeneous.

This led to considering **an individual taken at random in the whole population** and to looking at the probability that this individual comes from a given location, a given temperature or a given density slide in the population (**individual-based statistics**).

A spatial interpretation based on the **covariogram** is obtained by considering two individuals drawn at random from the population. Procedures for estimating covariogram in case of irregular sampling are presented for use as structural models. A case-study on hake (Bay of Biscay - France) shows that even though the covariograms are global tools, they incorporate finer structural information than had been imagine earlier, and they made it possible to map the data using **transitif kriging**.

The individual-based approach has been used to describe and model the spatial distribution of mackerel eggs during the spawning seasons from 1983 to 1995. As the bell-shaped experimental regressions proved difficult to interpret in a spatial context, the existence of a sea surface temperature preferendum was tested for by verifying that the distribution of individuals minimizes the inertia of temperature of the population (**inertiograms**). The experimental results show that the spatial distribution of mackerels appears to be limited by a "virtual inertia front".